



研究报告

A Letter

几种果树及拟南芥开花基因 *LFY* 密码子偏好性分析

张颖[✉], 樊秀彩[✉], 姜建福[✉], 孙海生[✉], 李民[✉], 张永辉[✉], 刘崇怀[✉]

中国农业科学院郑州果树研究所, 郑州, 450009

✉ 通讯作者: liuchonghuai@caas.net.cn ✉ 作者

分子植物育种, 2011 年, 第 9 卷, 第 94 篇 doi: 10.5376/mpb.cn.2011.09.0094

收稿日期: 2011 年 06 月 20 日

接受日期: 2011 年 07 月 21 日

发表日期: 2011 年 07 月 25 日

这是一篇采用 Creative Commons Attribution License 进行授权的开放取阅论文。只要对本原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

引用格式:

引用格式(中文):

张颖等, 2011, 几种果树及拟南芥开花基因 *LFY* 密码子偏好性分析, 分子植物育种(online) Vol.9 No.94 pp.1673-1679 (doi: 10.5376/mpb.cn.2011.09.0094)

引用格式(英文):

Zhang et al., 2011, Characterization of Codon Usage of *LFY* Genes in Several Fruit Trees, Fenzi Zhiwu Yuzhong (online) (Molecular Plant Breeding) Vol.9 No.94 pp. 1673-1679 (doi: 10.5376/mpb.cn.2011.09.0094)

摘要 由于不同物种在密码子的使用上存在不同的偏性, 而密码子偏性与核酸遗传信息和生物功能载体蛋白质相关联, 具有重要的生物学意义。本研究运用相关生物学软件和数理统计学软件对甜橙、桃、葡萄、芒果、甜杏、苹果、砂梨 *LFY* (terminal flower) 基因的密码子用法进行了分析, 并与模式植物拟南芥的 *LFY* 基因进行了比较。结果表明, 几种果树间的 *LFY* 基因密码子偏性差异不大, 都偏好使用 CG 结尾的密码子, 而拟南芥密码子第三位碱基的分布较均匀。不同物种的 *LFY* 基因在密码子上的差异与其亲缘关系呈正相关, 表现为亲缘关系越近的物种其基因密码子的用法越相似。基于密码子用法的偏好研究结果可以作为物种进化关系和分子机理研究的重要补充。

关键词 *LFY* 基因; 密码子偏性; 果树

Characterization of Codon Usage of *LFY* Genes in Several Fruit Trees

Zhang Ying[✉], Fan Xiucui[✉], Jiang Jianfu[✉], Sun Haisheng[✉], Li Min[✉], Zhang Yonghui[✉], Liu Chonghuai[✉]

Zhengzhou Fruit Research Institute, Chinese Academy of Agriculture Sciences, Zhengzhou, 450009, P.R., China

✉ Corresponding author, liuchonghuai@caas.net.cn; ✉ Authors

Abstract Genetic codons has degeneracy in different species, as it is related to the carrier of genetic information (DNA) and functional molecular (protein) of life, this phenomenon implicates important biological sense. In this study, codon usage of *LFY* genes in *Citrus sinensis*, *Prunus persica*, *Vitis vinifera*, *Mangifera indica*, *Eriobotrya japonica*, *Prunus dulcis*, *Malus*, *Pyrus pyrifolia*, were analyzed and compared to *Arabidopsis thaliana* by software. Results indicated codon bias of *LFY* in fruit trees were little different, and prefer to the codons with C or G at the synonymous position. *LFY* of *Arabidopsis thaliana* prefer to the codons ending with A or U. And it was found that the relationships of species were more closely, and the codon usage patterns of *LFY* were more similar. So the analysis of codon usage was an important and supplementary method to phylogenetic research, and was used to the investigation of the evolutionary relation and molecular mechanism.

Keywords *LFY* gene; Codon bias; fruit tree

研究背景

生物在遗传过程中通过“中心法则”传递遗传信息, DNA 的遗传信息传递给蛋白质, 从而实现生物体的整个生命过程, 这一过程是通过三个碱基的密码子对应一个氨基酸的方式来完成遗传信息的传递。每种氨基酸至少对应一个密码子, 编码同一种氨基酸的密码子称为同义密码子。生物体内普遍存在同义密码子非均衡使用的现象, 即某一物种或某一基因通常倾向于使用一种或几种特定的同义

密码子, 这些密码子被称为最优密码子, 这一现象被称为密码子偏性, 在进行基因的异源表达时, 密码子偏性就显的非常重要。在分子进化、翻译调控方面, 通过对果蝇、禾本科类等物种的研究表明, 不同物种基因中同义密码子使用偏好性的形成与转录相关突变和翻译水平上的选择有关(Wong et al., 2002; Powell et al., 2003)。

早期关于密码子偏性的理论(如: 密码子的选择模式在进化过程中是保守的; 密码子使用的偏爱与

相应的同工 tRNA 浓度相关, 从而有利于翻译的效率和准确性), 其数据支持多来源于单细胞的大肠杆菌和酵母(Ikemura, 1985; 吴宪明等, 2007), 随着新的物种被测序, 大量的核酸序列和蛋白质序列被发现, 研究发现多细胞生物高度的细胞分化和复杂的基因表达调控系统使密码子在选择模式上与单细胞生物存在很大差异, 使密码子偏性研究进入新的发展阶段。Donal A Hickey 对水稻全基因组密码子进行了分析, 结果表明基因的多样性与其含有的核苷酸类别相关, 特别与 GC 含量相关, 同时密码子的偏好性与基因长度成负相关(Wang and Hickey, 2007)。

LFY (terminal flower) 基因是 PEBPs (phosphatylethanolamine-binding proteins, 磷脂酰乙醇胺结合蛋白) 基因家族中的一员, 最初是在拟南芥中发现 *LFY* 基因在花芽形成过程及花序建成方面起到重要调控作用(Shannon and Meeks-Wagner, 1993; Bradley et al., 1997; Ratcliffe et al., 1998), 随后在许多开花植物中发现 *LFY* 基因, 在主要的几种果树中也陆续发现了 *LFY* 基因的同源序列, 如葡萄(Carmona et al., 2002)、苹果、砂梨、枇杷(Esumi et al., 2005) 和芒果(NCBI)等。果树的花芽分化是开花结果的基础, 开花基因是花芽分化的内在因素, 开展花芽分化相关基因的研究, 阐明其表达机理, 对利用分子生物学技术提高果树成花数量和质量具有重要意义。本论文从生物信息学角度分析果树 *LFY* 基因的密码子偏性, 为生物进化和异源基因表达提供理论基础。

1 结果与分析

1.1 *LFY* 基因有效密码子及密码子组成

为了研究不同果树 *LFY* 基因的密码子使用方式, 把来自 8 种果树及拟南芥的 *LFY* 基因片段的长度、ENC 值、GC 含量和 59 个密码子的 RSCU 值分别列表(表 1; 表 2)。由表 2 可知, 不同果树 *LFY* 基因的的 ENC 值为 52.274~56.228, 都在 50 以上, 反映了已知几种果树的 *LFY* 基因均一性较好, 且与拟南芥的 ENC 值 56.404 相近。同时 GC3s 值进一步表明了同义密码子使用方式的均一性, 表 2 中 GC3s 值范围为 0.526~0.645, 标准偏差为 0.038, 同样说明这几种果树 *LFY* 基因均一性较好, 偏好性不强, 稍高于拟南芥的 0.497。

基因的表达水平越高, 要求的翻译效率就越高, 同时密码子的偏好性就越高, 相反基因的表达水平越低, 要求的翻译效率越低, 密码子的偏好性也就随之降低, 所以密码子偏好性的高低可以从一定程度上反映该基因的表达水平。CAI 的值越低, 基因越倾向于低水平表达, 当 CAI 的值低于 0.25 时, mRNA 水平与蛋白质水平的相关性很低(大多数基因如此), CAI 大于 0.5 时, 基因为高表达, mRNA 水平与蛋白质水平的相关性较高(Carbone et al., 2003)。CAI 值从 0.195 到 0.225, 平均值为 0.212, 其中桃和甜杏的 CAI 值为 0.195 和 0.197, 最高的甜橙也只有 0.225, 所以该基因在果树中为低水平表达, 拟南芥的 CAI 值为 0.247, 同样为低水平表达(表 1)。

表 1 开花相关基因 *LFY* 编码的 GC 含量及有效密码子数(ENC)

Table 1 The content of GC and ENC values in *LFY* genes in nine species

物种 Species	登录号 Accession No.	ENC	GC3s	GC	GC1s	GC2s	CAI	长度(bp) Length (bp)
甜橙 <i>Citrus sinensis</i>	AAR01229.1	54.083	0.582	0.554	0.622	0.444	0.225	1197
桃 <i>Prunus persica</i>	ABY78032.1	54.728	0.632	0.565	0.606	0.445	0.195	1248
葡萄 <i>Vitis vinifera</i>	AF378126.1	52.274	0.594	0.568	0.635	0.461	0.212	1209
芒果 <i>Mangifera indica</i>	HQ585993.1	53.341	0.526	0.527	0.602	0.435	0.202	1152
枇杷 <i>Eriobotrya japonica</i>	AY551183.2	56.228	0.645	0.569	0.616	0.433	0.201	1227
甜杏 <i>Prunus dulcis</i>	AY947465.1	54.761	0.632	0.563	0.602	0.441	0.197	1245
苹果 <i>Malus</i>	BAB83097.1	54.797	0.621	0.568	0.620	0.448	0.211	1233
砂梨 <i>Pyrus pyrifolia</i>	BAD10956.1	55.522	0.618	0.568	0.620	0.451	0.214	1224
拟南芥 <i>Arabidopsis aliana</i>	DQ447103.1	56.404	0.479	0.522	0.613	0.456	0.247	1263

1.2 *LFY* 同义密码子的使用频率

利用 Codon W 软件对 *LFY* 基因密码子使用的具体情况进行了分析(表 2), 其中, 甜橙 RSCU 值大于 1 的密码子有 26 个, 桃有 25 个, 葡萄有 25 个, 芒果有 27 个, 枇杷有 29 个, 甜杏有 26 个,

苹果 28 个, 砂梨有 27 个, 拟南芥有 25 个(7 个以 U 结尾, 12 个以 A 结尾)。RSCU 值大于 1 的密码子为基因的偏爱密码子, 果树中 *LFY* 基因 RSCU 值大于 1 的密码子多以 C、G 结尾, 拟南芥中 *LFY* 基因多偏好以 AU 结尾的密码子(表 3)。

表 2 开花相关基因 *LFY* 编码氨基酸的同义密码子中相对使用频率(RSCU)

Table 2 The relative synonymous (RSCU) codon usage of *LFY* genes that coding amino acid

氨基酸 Amino	密码子 Codon	甜橙 <i>Citrus sinensis</i>	桃 <i>Prunus persica</i>	葡萄 <i>Vitis vinifera</i>	芒果 <i>Mangifera indica</i>	枇杷 <i>Eriobotrya japonica</i>	甜杏 <i>Prunus dulcis</i>	苹果 <i>Malus</i>	砂梨 <i>Pyrus pyrifolia</i>	拟南芥 <i>Arabidopsis thaliana</i>
Phe	UUU	0.33	0.62	0.18	0.53	0.62	0.62	0.50	0.50	0.86
Phe	UUC	<u>1.67</u>	0.38	<u>1.82</u>	<u>1.47</u>	<u>1.38</u>	<u>1.38</u>	<u>1.50</u>	<u>1.50</u>	<u>1.14</u>
Leu	UUA	0.36	0.19	0.39	0.22	0.19	0.00	0.00	0.00	1.00
Leu	UUG	<u>1.64</u>	<u>1.31</u>	0.58	<u>1.78</u>	<u>1.50</u>	<u>1.50</u>	<u>1.31</u>	<u>1.31</u>	<u>1.17</u>
Leu	CUU	<u>1.09</u>	<u>1.12</u>	<u>1.35</u>	<u>1.56</u>	<u>1.69</u>	<u>1.12</u>	<u>2.06</u>	<u>2.06</u>	0.82
Leu	CUC	<u>1.27</u>	<u>1.88</u>	<u>1.35</u>	0.67	<u>1.50</u>	<u>1.88</u>	<u>1.31</u>	<u>1.31</u>	<u>1.09</u>
Leu	CUA	0.00	0.94	0.39	0.00	0.38	0.94	0.91	0.19	<u>1.09</u>
Leu	CUG	<u>1.64</u>	0.56	<u>1.94</u>	<u>1.78</u>	0.75	0.56	<u>1.12</u>	<u>1.12</u>	1.00
Ser	UCU	<u>1.50</u>	<u>1.50</u>	0.95	<u>1.76</u>	<u>1.89</u>	<u>1.50</u>	<u>2.33</u>	<u>2.33</u>	<u>1.64</u>
Ser	UCC	0.75	<u>1.20</u>	<u>1.26</u>	0.35	<u>1.26</u>	<u>1.20</u>	1.00	1.00	0.82
Ser	UCA	0.75	0.30	0.32	0.71	0.32	0.30	0.00	0.00	1.09
Ser	UCG	0.75	0.60	0.63	<u>1.06</u>	0.63	0.60	0.67	0.67	0.55
Tyr	UAU	0.71	0.43	0.15	<u>1.33</u>	0.86	0.62	0.77	0.86	0.33
Tyr	UAC	<u>1.29</u>	<u>1.57</u>	<u>1.85</u>	0.67	<u>1.14</u>	<u>1.38</u>	<u>1.23</u>	<u>1.14</u>	<u>1.67</u>
Cys	UGU	0.29	0.86	1.00	0.33	<u>1.14</u>	1.00	<u>1.25</u>	<u>1.25</u>	0.89
Cys	UGC	<u>1.71</u>	<u>1.14</u>	1.00	1.67	0.86	1.00	0.75	0.75	<u>1.11</u>
Pro	CCU	0.38	0.86	0.40	0.95	<u>1.18</u>	0.86	0.53	0.75	<u>1.20</u>
Pro	CCC	0.38	<u>1.14</u>	1.00	0.76	0.71	<u>1.14</u>	<u>1.07</u>	1.00	0.20
Pro	CCA	<u>1.90</u>	0.57	<u>2.00</u>	<u>1.52</u>	<u>1.18</u>	0.57	<u>1.33</u>	1.00	<u>1.20</u>
Pro	CCG	<u>1.33</u>	<u>1.43</u>	0.60	0.76	0.94	<u>1.43</u>	<u>1.07</u>	<u>1.25</u>	<u>1.40</u>
His	CAU	0.91	<u>1.08</u>	<u>1.11</u>	<u>1.25</u>	0.91	0.91	1.00	0.89	<u>1.17</u>
His	CAC	<u>1.09</u>	0.92	0.89	0.75	<u>1.09</u>	<u>1.09</u>	1.00	<u>1.11</u>	0.83
Gln	CAA	0.75	0.94	0.80	0.82	0.59	0.89	0.75	0.88	0.87
Gln	CAG	<u>1.25</u>	<u>1.06</u>	<u>1.20</u>	<u>1.18</u>	<u>1.41</u>	<u>1.11</u>	<u>1.25</u>	<u>1.12</u>	<u>1.13</u>
Arg	CGU	0.86	1.00	0.51	<u>1.33</u>	0.88	1.00	<u>1.12</u>	0.97	<u>2.18</u>
Arg	CGC	0.69	0.67	<u>1.20</u>	0.17	<u>1.24</u>	0.67	<u>1.12</u>	<u>1.35</u>	0.36
Arg	CGA	0.51	0.50	0.51	0.33	0.00	0.50	0.19	0.19	0.55
Arg	CGG	0.51	0.67	0.86	0.50	<u>1.41</u>	0.67	<u>1.31</u>	<u>1.55</u>	0.91
Arg	AGA	<u>1.20</u>	<u>1.17</u>	<u>1.54</u>	<u>1.67</u>	<u>1.06</u>	<u>1.33</u>	0.94	0.77	<u>1.64</u>
Arg	AGG	<u>2.23</u>	<u>2.00</u>	<u>1.37</u>	<u>2.00</u>	<u>1.41</u>	<u>1.83</u>	<u>1.31</u>	<u>1.16</u>	0.36
Ile	AUU	<u>1.91</u>	1.00	0.69	<u>1.07</u>	<u>1.09</u>	1.00	<u>1.36</u>	<u>1.36</u>	0.82
Ile	AUC	0.82	<u>1.25</u>	<u>1.85</u>	0.43	<u>1.36</u>	<u>1.50</u>	<u>1.09</u>	<u>1.09</u>	<u>1.09</u>
Ile	AUA	0.27	0.75	0.46	<u>1.50</u>	0.55	0.50	0.55	0.55	<u>1.09</u>
Thr	ACU	<u>1.09</u>	0.80	0.36	0.92	0.62	0.80	<u>1.25</u>	<u>1.25</u>	0.42
Thr	ACC	<u>1.45</u>	<u>1.87</u>	0.73	0.92	<u>1.54</u>	<u>1.87</u>	1.00	1.00	0.63
Thr	ACA	0.36	0.27	0.73	0.92	0.31	0.27	0.25	0.50	0.84
Thr	ACG	<u>1.09</u>	<u>1.07</u>	<u>2.18</u>	<u>1.23</u>	<u>1.54</u>	<u>1.07</u>	<u>1.50</u>	<u>1.25</u>	<u>2.11</u>
Asn	AAU	<u>1.08</u>	<u>1.14</u>	0.43	0.77	0.71	<u>1.14</u>	0.80	0.67	0.50
Asn	AAC	0.92	0.86	<u>1.57</u>	<u>1.23</u>	<u>1.29</u>	0.86	<u>1.20</u>	<u>1.33</u>	<u>1.50</u>
Lys	AAA	0.55	0.48	0.27	0.32	0.36	0.40	0.38	0.52	0.80

续表 2

Continuing table 2

氨基酸 Amino	密码子 Codon	甜橙 <i>Citrus sinensis</i>	桃 <i>Prunus persica</i>	葡萄 <i>Vitis vinifera</i>	芒果 <i>Mangifera indica</i>	枇杷 <i>Eriobotrya japonica</i>	甜杏 <i>Prunus dulcis</i>	苹果 <i>Malus</i>	砂梨 <i>Pyrus pyrifolia</i>	拟南芥 <i>Arabidopsis thaliana</i>
Lys	AAG	<u>1.45</u>	<u>1.52</u>	<u>1.73</u>	<u>1.68</u>	<u>1.64</u>	<u>1.60</u>	<u>1.62</u>	<u>1.48</u>	<u>1.20</u>
Ser	AGU	0.75	0.30	0.63	0.35	0.32	0.30	0.33	0.00	1.36
Ser	AGC	<u>1.50</u>	<u>2.10</u>	<u>2.21</u>	<u>1.76</u>	<u>1.58</u>	<u>2.10</u>	<u>1.07</u>	<u>2.00</u>	0.55
Val	GUU	0.89	0.67	0.76	1.19	0.86	0.64	0.91	0.91	<u>1.71</u>
Val	GUC	0.44	0.67	0.57	0.15	0.86	0.64	0.91	0.91	0.57
Val	GUA	0.15	0.17	0.76	0.15	0.29	0.32	0.00	0.18	0.38
Val	GUG	<u>2.52</u>	<u>2.50</u>	<u>1.90</u>	<u>2.52</u>	<u>2.00</u>	<u>2.40</u>	<u>2.18</u>	<u>2.00</u>	<u>1.33</u>
Ala	GCU	<u>1.36</u>	0.78	<u>1.04</u>	<u>1.37</u>	0.72	0.73	0.70	0.70	<u>1.89</u>
Ala	GCC	0.77	<u>1.48</u>	<u>1.11</u>	0.59	<u>1.33</u>	1.64	1.30	<u>1.22</u>	0.33
Ala	GCA	0.94	<u>1.13</u>	<u>1.26</u>	<u>1.07</u>	0.82	<u>1.09</u>	0.96	<u>1.04</u>	0.44
Ala	GCG	0.94	0.61	0.59	0.98	<u>1.13</u>	0.55	<u>1.04</u>	<u>1.04</u>	1.33
Asp	GAU	<u>1.24</u>	0.71	1.00	<u>1.36</u>	0.88	0.77	<u>1.15</u>	<u>1.20</u>	0.80
Asp	GAC	0.76	<u>1.29</u>	1.00	0.64	<u>1.12</u>	<u>1.23</u>	0.85	0.80	<u>1.20</u>
Glu	GAA	0.63	0.36	0.46	0.72	0.32	0.40	0.26	0.26	<u>1.03</u>
Glu	GAG	<u>1.37</u>	<u>1.64</u>	<u>1.54</u>	<u>1.28</u>	<u>1.68</u>	<u>1.60</u>	<u>1.74</u>	<u>1.74</u>	0.97
Gly	GGU	<u>1.33</u>	0.88	<u>1.38</u>	<u>1.33</u>	0.98	0.98	0.76	0.76	<u>1.65</u>
Gly	GGC	0.85	<u>1.07</u>	0.00	0.74	0.59	<u>1.07</u>	0.95	0.86	0.96
Gly	GGA	0.61	0.39	<u>1.50</u>	0.74	0.59	0.39	0.86	0.86	0.87
Gly	GGG	<u>1.21</u>	<u>1.66</u>	<u>1.12</u>	<u>1.19</u>	<u>1.85</u>	<u>1.56</u>	<u>1.43</u>	<u>1.52</u>	0.52

注: 数值带下划线表示 RSCU 值大于 1 的密码子为相应果树 *LFY* 的偏好密码

Note: The data underline means the RSCU value is more than 1

表 3 欧氏距离系数检测开花相关基因 *LFY* 间相对密码子相对使用频率

Table 3 Codon usage bias between samples of *LFY* genes evaluating by Squared Euclidean distance

	甜橙 <i>Citrus sinensis</i>	桃 <i>Prunus persica</i>	葡萄 <i>Vitis vinifera</i>	芒果 <i>Mangifera indica</i>	枇杷 <i>Eriobotrya japonica</i>	甜杏 <i>Prunus dulcis</i>	苹果 <i>Malus</i>	砂梨 <i>Pyrus pyrifolia</i>	拟南芥 <i>Arabidopsis thaliana</i>
甜橙 <i>Citrus sinensis</i>	0.000								
桃 <i>Prunus persica</i>	3.213	0.000							
葡萄 <i>Vitis vinifera</i>	3.918	4.024	0.000						
芒果 <i>Mangifera indica</i>	2.652	3.695	4.295	0.000					
枇杷 <i>Eriobotrya japonica</i>	3.187	2.573	3.327	3.560	0.000				
甜杏 <i>Prunus dulcis</i>	3.035	1.240	3.747	3.622	2.237	0.000			
苹果 <i>Malus</i>	3.089	2.947	3.614	3.525	1.776	2.668	0.000		
砂梨 <i>Pyrus pyrifolia</i>	3.410	2.924	3.663	3.681	1.780	2.619	1.313	0.000	
拟南芥 <i>Arabidopsis thaliana</i>	4.500	4.992	4.625	4.257	4.409	4.981	4.452	4.784	0.000

1.3 *LFY* 基因密码子使用频率的差异

根据表 3 中各物种密码子的相对使用频率 (RSCU), 利用 SPSS 分析软件中的欧氏距离算法, 对几种果树及拟南芥的 *LFY* 基因的密码子偏好进行了计算(表 4)和聚类分析(图 1)。各物种间的距离系数能够体现它们在密码子使用上的相似度, 由表 4 可知, 桃和甜杏的距离最小, 为 1.240; 苹果和砂梨的距离也较小, 为 1.313; 而拟南芥与所有果树的距离都较大, 这一结果与图 1 的结果相符。基于 *LFY* 蛋白质序列构建的系统发育树的分支结果与图 1 的结果相一致, 说明密码子的偏性差异与物种的亲缘关系存在一定的相关性, 亲缘关系越近, 密码子偏性差异越小。

表 4 8 种果树及拟南芥的 *LFY* 基因序列登录号

Table 4 The accession number of *LFY* genes of fruit tree and *Arabidopsis thaliana*

编号	序列登录号	物种
No.	Accession number	Species
1	AAR01229.1	甜橙 <i>Citrus sinensis</i>
2	ABY78032.1	桃 <i>Prunus persica</i>
3	AF378126.1	葡萄 <i>Vitis vinifera</i>
4	HQ585993.1	芒果 <i>Mangifera indica</i>
5	AY551183.2	枇杷 <i>Eriobotrya japonica</i>
6	AY947465.1	甜杏 <i>Prunus dulcis</i>
7	BAB83097.1	苹果 <i>Malus</i>
8	BAD10956.1	砂梨 <i>Pyrus pyrifolia</i>
9	DQ447103.1	拟南芥 <i>Arabidopsis thaliana</i>

2 讨论

密码子使用偏性产生的生物学基础目前还不太清楚, 关于偏性与翻译效率也有一些不同的看法, Kanaya 等(2001)通过 5 个真核生物(酵母, 线虫, 果蝇, 非洲爪蟾和人)的密码子使用模式进行对比发现, 密码子偏性有可能和基因组中的 *isochore* 结构和含 GC 二核苷酸的密码子使用有关; 在植物中密码子的偏性分析也得到了类似的结论(Murray et al., 1989); zhang 和 Li (2004)认为基因表达的水平是决定密码子使用偏好的主要因素。

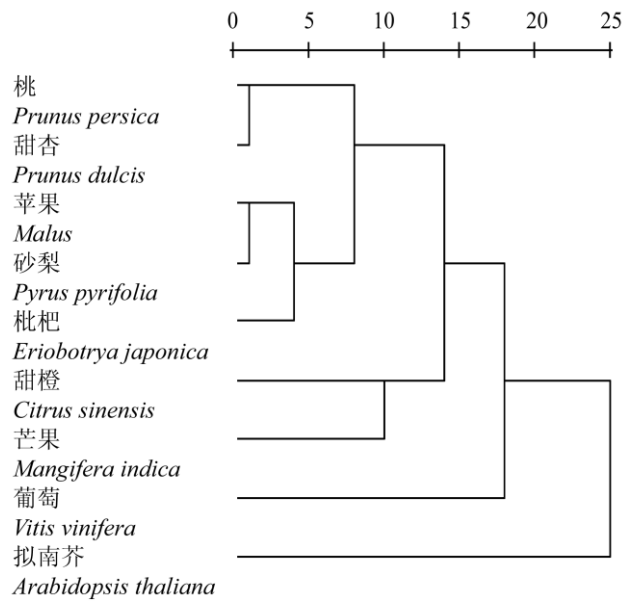


图 1 基于 *LFY* 基因密码子编码氨基酸的相对使用频率聚类
Figure 1 Analysis dendrogram of RSCU for *LFY* genes by cluster

在分析 GC 含量的研究中, 以 60% 为一个分界点, 高于 60% 的为高 GC 含量, 低于 60% 为低 GC 含量(Wang and Hickey, 2007)。本文对几个果树开花基因 *LFY* 的密码子进行偏性分析, 发现几种果树的 *LFY* 基因的 GC 含量均低于 60%, 和拟南芥一起属于低 GC 含量基因(Wang and Hickey, 2007)。*LFY* 基因中密码子 RSCU 值大于 1, 同时偏好一致的氨基酸所对应的密码子有如下几种: 编码氨基酸 Phe 的密码子 UUC、编码氨基酸 Leu 的密码子 UUG/CUU、编码氨基酸 Tyr 的密码子 UAC、编码氨基酸 Gln 的密码子 CAG、编码氨基酸 Arg 密码子 AGG、编码氨基酸 Thr 的密码子 ACG、编码 Lys 的密码子 AAG、编码氨基酸 Val 的密码子 GUG、编码氨基酸 Glu 的密码子 GAG、编码氨基酸 Gly 的密码子 GGG (表 2), 与拟南芥的密码子偏好特征不同(拟南芥中 *LFY* 基因多偏好以 AU 结尾的密码子)。

基因完整编码序列中的密码子用法在生物界广泛存在, 是分子进化中的一种重要现象。密码子是携带 DNA 遗传信息的最小单位, 从密码子偏性的特征角度可以对基因甚至物种的进化关系进行推测, 不同物种或同一物种的不同基因对密码子的使用模式即密码子偏性有所不同, 这种差异与物种亲缘关系存在密切关系, 亲缘关系越远, 密码子偏性差异大, 亲缘关系近则差异小(吴宪明等, 2007; 赵薇等, 2010)。在本文中几种果树密码子偏性之间

的大小(表 3; 图 1)与其进化关系(图 2)远近相符, 基于密码子使用模式差异进行的聚类分析, 显示同一基因不同物种之间的关系与通过蛋白质氨基酸序列构建的系统进化树显示的各物种间的进化距离一致, 说明在基因功能和类型一致的情况下, 进化上亲缘关系较近的物种, 密码子的使用模式相似。目前, 基于分子序列的系统发育分析中, 密码子使用信息的分析结果可从另外的角度为系统发育研究提供更多的参考信息, 作为各种系统发育分析方法的重要补充。

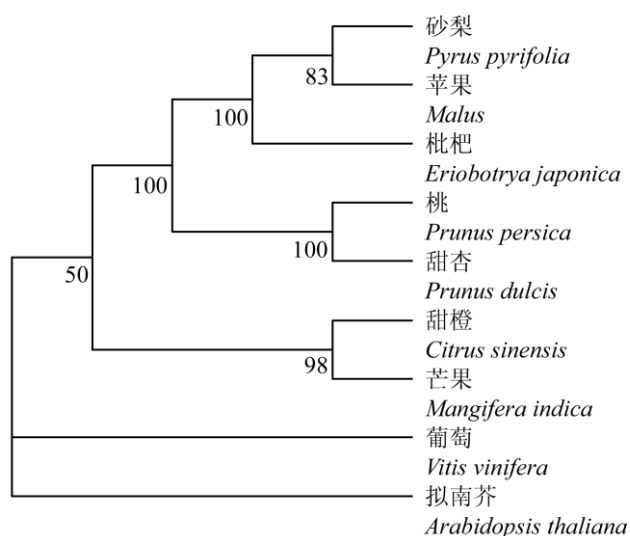


图 2 基于 *LFY* 氨基酸序列的系统分类图

Figure 2 A dendrogram of protein sequences encoding by *LFY* amino acid

3 材料与方法

3.1 序列来源

本研究所采用的 8 种果树及拟南芥的 *LFY* 基因序列均来源于 GenBank (表 4)。

3.2 使用主要软件和程序

密码子分析软件 Codon W (<http://www.molbio1.ox.ac.uk>)和序列比对软件 Invitrogen Vector NTI Advance 11 中 ClustalW progress; 欧洲分子生物学开放软件系统 EMBOSS 中的 CHIPS 和 CUSP 程序; 系统进化树构建软件 MEGA 5.0。

3.2.1 密码子分析中涉及的指标

(1)有效密码子数(effective number of codons, ENC) (Wright, 1990);

(2) GC3s: 基因中 GC 含量和密码子中不同位置的 GC 含量;

(3)密码子适应指数(codon adaptation index, CAI), 密码子偏性达到最大值, 即所有氨基酸均使用最优密码子, CAI 值为 1; 而 CAI 值越小, 密码子偏好程度就越低, 相反偏好程度就越高;

(4)同义密码子相对使用频率(relative synonymous codon usage, RSCU), 指对于特定密码子在编码对应氨基酸相对同义密码子间的概率, 密码子没有偏好时, 该密码子的 RSCU 值等于 1; RSCU 值大于 1, 表明该密码子使用频率较高, 反之使用频率就越低(Sharp and Li, 1986; 刘汉梅等, 2010)。

3.2.2 基因密码子使用概率聚类分析

利用 SPSS 13.0 对 9 条 *LFY* 基因进行了基于密码子使用偏性的聚类。在对该基因密码子使用频率(RSCU)分析时, 把每个基因作为一个对象, 采用 59 个同义密码子(去除起始密码子和 3 个终止密码子)的 RSCU 值。基因间的距离分析利用密码子使用度的欧拉平方距离(Das et al., 2006), 构建距离系数矩阵, 并进行基于该系数的聚类分析。

作者贡献

张颖是本研究的实验设计和实验研究的执行人; 樊秀彩, 姜建福, 孙海生完成数据分析, 论文初稿的写作; 李民, 张永辉参与实验设计, 试验结果分析; 刘崇怀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

本研究由现代农业产业技术体系建设专项资金(No. CARS-30-yz-1)资助。作者感谢两位匿名的同行评审人的评审建议和修改建议。

参考文献

- Bradley D., Ratcliffe O., Vincent C., Carpenter R., and Coen E., 1997, Inflorescence commitment and architecture in *Arabidopsis*, *Science*, 275(5296): 80-83
- Carbone A., Zinovyev A., and Képès F., 2003, Codon adaptation index as a measure of dominating codon bias, *Bioinformatics*, 19(16): 2005-2015
- Carmona M.J., Cubas P., and Martinez-Zapater J.M., 2002, VFL, the grapevine FLORICAULA/LEAFY ortholog, is expressed in meristematic regions independently of their fate, *Plant Physiology*, 130: 68-77
- Das S., Paul S., and Dutta C., 2006, Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydrophathy, *Virus Res.*, 117(2): 227-236

- Esumi T., Tao R., and Yonemori K., 2005, Isolation of LEAFY and TERMINAL FLOWER 1 homologues from six fruit tree species in the subfamily Maloideae of the Rosaceae, *Sex. Plant. Reprod.*, 17: 277-287
- Ikemura T., 1985, Codon usage and tRNA content in unicellular and multicellular organisms, *Mol. Biol. Evol.*, 2(1): 13-34
- Kanaya S.Y.Y., Kinouchi M., Kudo Y., and 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
- Liu H.M., Zhao Y., Gu Y., Zhang H.Y., and Huang Y.B., 2010, Characterization of codon usage of waxy genes in several plants, *Journal of Nuclear Agricultural Sciences*, 24(3): 476-481 (刘汉梅, 赵耀, 顾勇, 张怀渝, 黄玉碧, 2010, 几种植物 waxy 基因的密码子用法特性分析, *核农学报*, 24(3): 476-481)
- Murray E.E., Lotzer J., and Eberle M., 1989, Codon usage in plant genes, *Nucleic Acids Res.*, 17(2): 477-498
- Powell J.R., Sezzi E., Moriyama E.N., Gleason J.M., and Caccone A., 2003, Analysis of a shift in codon usage in *Drosophila*, *J. Mol. Evol.*, 57(Suppl 1): S214-225
- Ratcliffe O.J., Amaya I., Vincent C.A., Rothstein S., Carpenter R., Coen E.S., and Bradley D.J., 1998, A common mechanism controls the life cycle and architecture of plants, *Development*, 125(9): 1609-1615
- Shannon S., and Meeks-Wagner D.R., 1993, Genetic interactions that regulate inflorescence development in *Arabidopsis*, *Plant Cell*, 5: 639-655
- Sharp P.M., and Li W.H., 1986, An evolutionary perspective on synonymous codon usage in unicellular organisms, *J. Mol. Evol.*, 24(1-2): 28-38
- Wang H.C., and Hickey D.A., 2007., Rapid divergence of codon usage patterns within the rice genome, *BMC Evol. Biol.*, 7(Suppl 1): S6
- Wong G.K.S., Wang J., Tao L., Tan J., Zhang J.G., Passey D.A., and Yu J., 2002, Compositional gradients in gramineae genes, *Genome Res.*, 12: 851-856
- Wright F., 1990, The 'effective number of codons' used in a gene, *Gene*, 87(1): 23-29
- Wu X.M., Wu S.F., Ren D.M., Zhu Y.P., and He F.C., 2007, The analysis method and progress in the study of codon bias, *Hereditas*, 29(4): 7 (吴宪明, 吴松峰, 任大明, 朱云平, 贺福初, 2007, 密码子偏性的分析方法及相关研究进展, *遗传*, 29(4): 7)
- Zhang L., and Li W.H., 2004, Mammalian housekeeping genes evolve more slowly than tissue-specific genes, *Mol. Biol. Evol.*, 21(2): 236-249
- Zhao W., Chen B.L., Wang M.Z., and Huang J.Z., 2010, Analysis on codon usage bias of fatty acid delta-6 desaturase genes from different algae, *Journal of Tropical Oceanography*, 29(3): 126-134 (赵薇, 陈必链, 王明兹, 黄建忠, 2010, 藻类 $\Delta 6$ 脂肪酸脱饱和酶基因密码子偏好性分析, *热带海洋学报*, 29(3): 126-134)



5thPublisher是一个致力于科学与文化传播的中文出版平台

在5thPublisher上发表论文, 任何人都可以免费在线取阅您的论文

- ※同行评审, 论文接受严格的高质量的评审
- ※在线发表, 论文一经接受, 即刻在线发表
- ※开放取阅, 任何人都可免费取阅无限使用
- ※快捷搜索, 涵盖谷歌学术搜索与知名数据库
- ※论文版权, 作者拥有版权读者自动授权使用

在线投稿: <http://5th.sophiapublisher.com>