

数据分析

Data Analysis

鲨烯合酶基因的密码子偏性分析

何闪 张淑杰 修乐山 邢朝斌*

河北联合大学生命科学学院,唐山,063000

* 通讯作者, xzbheuu@126.com

摘要 为了分析鲨烯合酶(squalene synthase, SS)基因密码子的使用方式及其影响因素,利用 codon W 和 SPSS 16.0 软件对 47 条来自不同物种的 SS 基因进行多元统计分析、对应性分析。SS 基因密码子 1~3 位碱基的 GC 含量(GC1, GC2 和 GC3)依次为 51.33%、34.65%和 54.37%, 3 个位点的 GC 含量均呈极显著相关关系($p < 0.01$)。对应性分析的结果表明,第 1 轴显示 30.71% 的差异,有效密码子数和 GC3、GC1 和 GC2 的均值与 GC3 之间的相关性均达极显著水平($p < 0.01$)。筛选出的 26 个最优密码子的第 3 位碱基均为 G 或 C。以 MEGA 5.0 构建的基于 SS 蛋白质序列的进化树比基于 RSCU 的聚类更符合传统的系统发育观点。SS 基因密码子偏好以 G/C 结尾,使用模式受选择和突变影响,突变对密码子偏好影响较大。

关键词 鲨烯合酶, 密码子偏性, 多元分析, 聚类分析

Codon Usage Analysis in Squalene Synthase Gene

He Shan Zhang Shujie Xiu Leshan Xing Zhaobin*

College of Life Science, Hebei United University, Tangshan, 063000

* Corresponding author, xzbheuu@126.com

DOI: 10.3969/gab.032.000232

Abstract To analyze the codon usage of squalene synthase (SS) gene and the influence factors, multivariate analysis and corresponding analysis of SS genes from different species were analyzed by software codon W and SPSS 16.0. GC content at three positions of codons (GC1, GC2 and GC3) in turns was 51.33%, 34.65% and 54.37%, and corresponded significantly. GC1 and GC2 have a significant correlation coefficient ($p < 0.01$). Corresponding result showed that, Axis 1 showed 30.71% variation. Effective number of codons (ENC) and GC3 were significantly correlated ($p < 0.01$). The correlation between GC12 and GC3 were significant ($p < 0.01$). 26 synonymous codons were the optimum codons which were all ended by C or G. The 3rd base composition preferred G/C in SS genes. Compared with cluster analysis dendrogram of RSCU for SS genes, the evolutionary tree build based on protein sequences encoded by SS genes through MEGA 5.0 software fit traditional viewpoint higher. The third position for all codons in SS preferred to end with A and T. Synonymous codon usage pattern formed under effect of mutation pressure and natural selection, mutation pressure influenced higher in SS gene.

Keywords Squalene synthase, Codon usage bias, Multivariate analysis, Cluster analysis

鲨烯合酶(squalene synthase, SS)是生物体内胆固醇合成、代谢途径中的一种重要的酶(Gu et al., 1998)。在药用植物的次生代谢中,SS 也是三萜皂苷和甾醇生物合成过程中的关键酶,SS 能够催化 2 分子的法尼基焦磷酸(farnesylpyrophosphate, FPP)合成 1 分子的鲨烯(squalene) (Busquets et al., 2008; Kim et al., 2011; Han et al., 2012)。目前已经克隆了包括人(Sum-

mersa et al., 1993)、拟南芥(Busquets et al., 2008)、裂殖酵母(Robinson et al., 1993)等在内的众多物种的 SS 基因。通过转基因技术将 SS 转入柴胡(*Bupleurum falcatum*)后,能够提高其三萜皂苷和植物甾醇的含量(Kim et al., 2011),通过抑制人参(*Panax ginseng*) SS 的表达则可降低三萜类的积累量(Han et al., 2012)。

不同物种或是同一物种的不同基因其同义密码

子使用是不均衡的,即密码子的使用具有偏性(杨春亮等,2012)。而密码子偏性的存在,则可对基因的外源表达产生影响,造成外源基因的表达量低下甚至不表达(谢磊等,2004;赵武等,2008),如拟南芥中的2个不同构型的SS基因在大肠杆菌中进行体外表达,效率较低(Busquets et al., 2008)。根据宿主密码子的偏爱性对外源基因进行了优化,可以在不改变氨基酸序列的基础上提高目的蛋白的表达量,为研究蛋白的纯化和功能做出贡献(Zhou et al., 2012;唐龙盘等,2011;张正平等,2010)。通过优化第2、5和6位密码子的4个碱基能够明显的提高甘油脱氢酶基因(glycerol dehydrogenase)在大肠杆菌中的表达水平(唐龙盘等,2011)。通过定点突变 α -糖基转移酶基因(α -cyclodextrin glycosyltransferase) α -糖基转移酶产量达到了最高值的48.9 U/mL(Zhou et al., 2012)。因此,密码子优化是改善目的基因异源表达水平的一种有效方法。本文通过对SS的密码子使用偏性进行分析,构建系统进化树,为下一步开展体外高效表达SS蛋白,研究SS的生物功能以及调节植物等的次生代谢等提供理论参考。

1 结果与分析

1.1 各物种SS基因的GC含量和ENC值

SS基因密码子各位置的GC含量、编码氨基酸有效密码子数(effective number of codons, ENC)和密码子数目的相关关系如表1所示。SS基因的总GC含量为46.78%,GC1、GC2和GC3分别为51.33%、34.65%和54.37%。整个编码区的GC含量与GC1、GC2、GC3之间的相关性均达到了极显著水平($p < 0.01$)。GC1、GC2与GC3之间相关性达到极显著水平($p < 0.01$),且表现为 $GC3 > GC1 > GC2$,说明SS基因密码子的第1、2、3位碱基的组成较相似(r 分别为0.626, 0.853和0.555),不存在显著差异(表2)。由表2中GC与GC1、GC2和GC3之间的相关系数(r 分别为0.912, 0.656和0.987)可知GC与GC1、GC2之间的相关性更强。ENC的最小值为32.30,最大值为61,均值为51.03,ENC值均比较大,密码子存在偏性但不明显,且ENC值和密码子第1、2和3位的碱基的相关性未达到显著水平(表2)。

1.2 各物种SS基因密码子的偏性

1.2.1 ENC和GC3的相关性分析

以47个物种SS的GC3值和ENC值分别为横、

纵坐标绘图(图1)。由图可知,SS沿期望值曲线分布或落在期望值曲线的下方,但大多数距离期望值曲线较近,只有少数远离期望值曲线。大多数的SS基因偏性较弱,少数基因偏性较强。说明SS的密码子偏性形成受突变和选择的共同作用,其中碱基突变的影响更强烈。

1.2.2 各物种GC12和GC3的关系

各物种SS基因的中性绘图分析结果如图2所示,多数SS基因主要分布于对角线下方。GC3的分布范围在0.1314~0.8174之间,GC12的分布范围在0.3584~0.5065之间,GC12分布较为紧凑。GC3分布范围较大,主要是由于白色念珠菌(*Candida albicans*)的GC3值(GC3为0.1314,GC12为0.3664)偏小引起的。GC12与GC3的相关性双尾检验达到极显著水平($p < 0.01$),回归曲线斜率为0.1497,说明GC12与GC3的相关性很强,密码子3个位置的碱基组成无差异,密码子的偏好受突变的影响,选择在密码子偏性形成过程中作用较弱,突变在此过程中的作用较强。

1.3 对应性分析

以同义密码子相对使用频率(relative synonymous codon usage, RSCU)为基础,进行对应性分析,

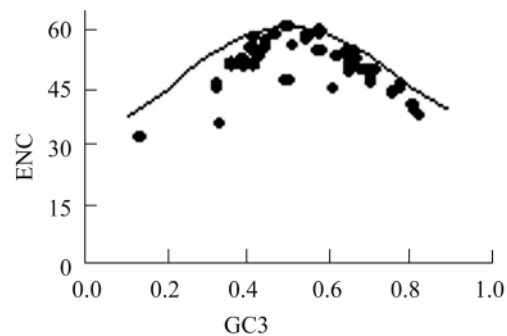


图1 ENC和GC3的相关性分析
Figure 1 Correlation analysis of ENC and GC3

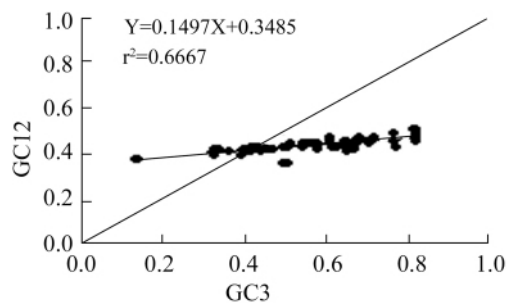


图2中性绘图分析
Figure 2 Neutrality plot analysis

表 1 各物种 SS 基因编码区各部位的 GC 含量及 ENC 值

Table 1 GC content of different positions of codon and ENC in SS genes

编号 Serial No.	GC	GC1	GC2	GC3	ENC	编号 Serial No.	GC	GC1	GC2	GC3	ENC
1	0.46	0.54	0.34	0.51	56.19	25	0.38	0.48	0.34	0.33	35.65
2	0.48	0.53	0.34	0.57	59.18	26	0.50	0.53	0.34	0.64	54.06
3	0.53	0.55	0.36	0.68	49.53	27	0.43	0.50	0.34	0.45	56.25
4	0.55	0.55	0.35	0.75	44.27	28	0.39	0.49	0.32	0.36	51.14
5	0.51	0.56	0.34	0.65	51.18	29	0.42	0.49	0.35	0.43	54.48
6	0.29	0.42	0.32	0.13	32.30	30	0.39	0.46	0.33	0.39	52.46
7	0.55	0.54	0.40	0.71	49.77	31	0.42	0.49	0.34	0.42	55.47
8	0.61	0.60	0.41	0.81	40.88	32	0.51	0.49	0.39	0.65	54.76
9	0.55	0.54	0.33	0.77	45.42	33	0.41	0.46	0.34	0.41	57.69
10	0.52	0.55	0.36	0.66	52.72	34	0.42	0.48	0.33	0.44	55.40
11	0.53	0.53	0.35	0.70	46.41	35	0.40	0.47	0.34	0.39	50.66
12	0.45	0.54	0.33	0.49	61.00	36	0.41	0.48	0.35	0.41	55.41
13	0.47	0.51	0.33	0.57	54.81	37	0.42	0.48	0.35	0.45	57.17
14	0.49	0.51	0.30	0.65	49.01	38	0.42	0.49	0.34	0.45	55.15
15	0.40	0.43	0.28	0.50	46.77	39	0.41	0.49	0.33	0.40	55.33
16	0.37	0.48	0.31	0.32	45.10	40	0.43	0.51	0.36	0.41	51.36
17	0.53	0.55	0.35	0.70	48.28	41	0.43	0.49	0.34	0.47	58.77
18	0.58	0.57	0.36	0.82	37.66	42	0.47	0.51	0.36	0.54	57.01
19	0.60	0.60	0.39	0.82	39.74	43	0.42	0.49	0.35	0.42	55.05
20	0.51	0.55	0.36	0.61	44.72	44	0.42	0.49	0.35	0.42	54.50
21	0.58	0.58	0.38	0.76	43.20	45	0.41	0.50	0.33	0.41	54.49
22	0.50	0.52	0.35	0.61	53.20	46	0.42	0.48	0.35	0.43	53.15
23	0.50	0.52	0.35	0.62	53.11	47	0.48	0.53	0.35	0.55	58.60
24	0.51	0.52	0.34	0.66	50.06						

注: 表中编号与表 3 中编号相对应

Note: The serial number in the table was same with that in table 3

表 2 SS 基因各相关参数的相关性分析

Table 2 Correlation analysis of each gene's related parameters

	GC	GC1	GC2	GC3
GC1	0.912**	-	-	-
GC2	0.656**	0.626**	-	-
GC3	0.987**	0.853**	0.555**	-
ENC	-0.218	-0.203	-0.115	-0.219

注: **: $p < 0.01$ 极显著水平

Note: **: $p < 0.01$ super significant

第 1 轴显示了 30.71% 的差异, 第 2、3、4 轴分别显示了 13.52%、10.80% 和 9.40% 的差异。前 4 轴差异合计 64.43%, 以第 1 轴为主, 其余各轴差异不明显。第 1 轴与 GC3 双尾检验达到极显著水平 ($p < 0.01$), 呈极显著正相关关系 ($r = 0.989$)。第 1 轴与密码子适应指数 (codon adaptation index, CAI) 和密码子偏好指数

(codon bias index, CBI) 的相关性达到双尾侧检验极显著水平 ($p < 0.01$, r 分别为 0.758 和 0.766)。说明 GC3 对 SS 基因密码子的偏性有较大的影响, 同时物种差别等其它因素对 SS 基因密码子的偏性也有一定影响, 其使用模式形成过程较为复杂。

根据各物种的进化水平, 将 SS 基因分为: 微生物类的 SS 基因 (编号 1~20, 共 20 条)、动物类的 SS 基因 (编号 21~26, 共 6 条) 和植物类的 SS 基因 (编号 27~47, 共 21 条)。以第 1 轴为横坐标, 第 2 轴为纵坐标将各物种 SS 基因分布于平面中。除了动物中的 *Naegleria gruberi* 的 SS 基因, 微生物中的白色念珠菌和酿酒酵母 (*Saccharomyces cerevisiae*) 的 SS 基因相对分散度较大外, 其它动物、植物和微生物的 SS 基因则集中分布 (图 3)。这说明动物、植物、微生物类的各物种 SS 基因的碱基构成更为相似, SS 基因各自分布更为集中。

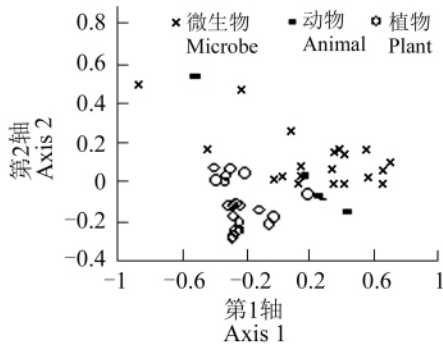


图3 基于 RSCU 的对应性分析
Figure 3 Corresponding analysis based on RSCU

1.4 最优密码子分析

通过比较 47 条不同物种的 SS 基因，根据其 ENC 值的大小，筛选高、低表达基因，高表达基因的 ENC 均值为 49.59，低表达基因的 ENC 均值为 46.82，分别计算高表达和低表达基因中各密码子

的 RSCU 值，确定了 SS 基因中存在除了 AUG、GUG、UAG、UGG、AGG、GGG 以外的 26 个以 G/C 为第 3 位的密码子为最优密码子。SS 基因最优密码子使用频率(frequency of optimal codons index, FOP)的范围为 0.31~0.54 之间，均值为 0.41，最优密码子的使用频率较高。SS 基因偏好使用 G/C 结尾的密码子。

1.5 基于 SS 基因密码子偏性的聚类分析

在对基因密码子使用频率进行聚类分析的过程中，将每一条基因作为一个对象，将密码子的 RSCU 值作为变量，利用 SPSS 16.0 进行聚类分析，聚类结果见图 4A。由 MEGA 5.0 的 neighbor-joining 连接法构建的 47 条 SS 蛋白的系统发育树如图 4B 所示，20 种微生物首先聚为一支，动植物和原虫的亲缘关系较近聚为一支，随后和微生物聚为一个大的分支。通

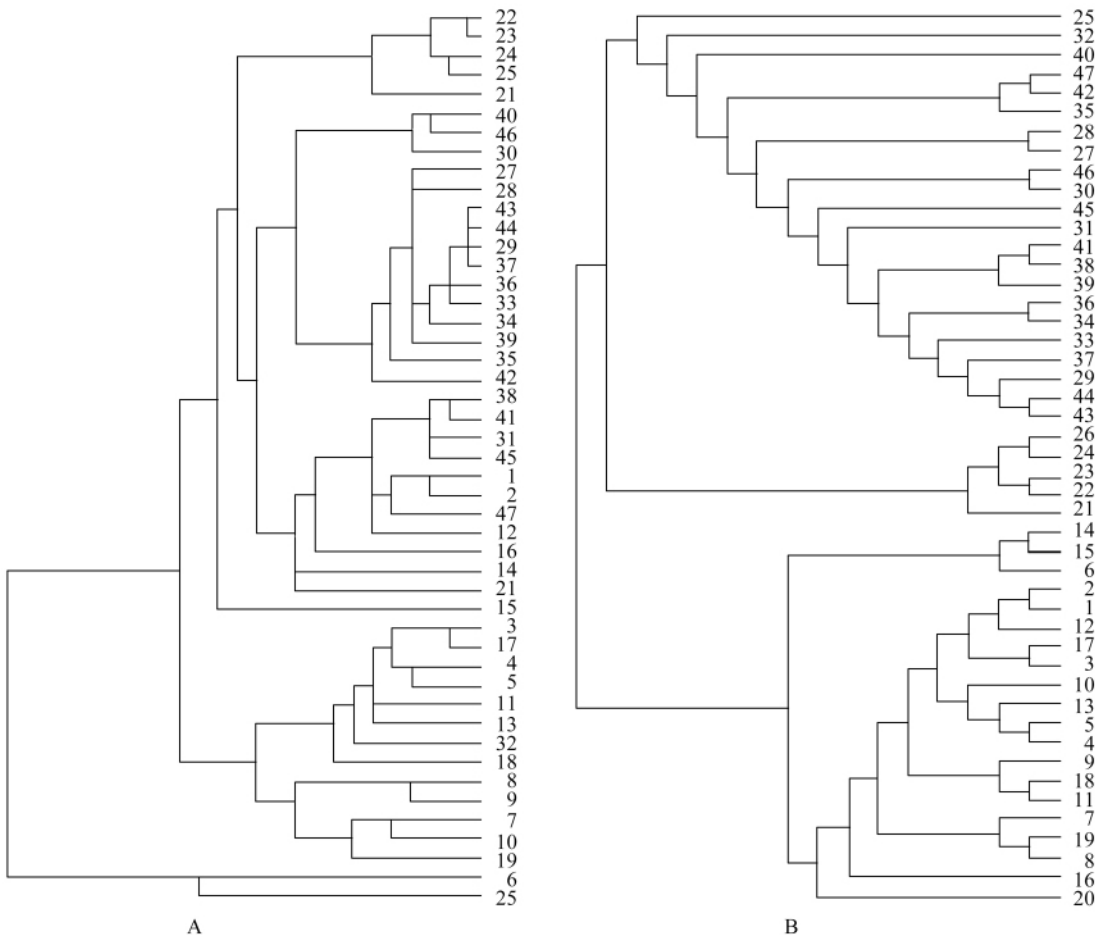


图 4 SS 基因构建的系统发育树
注: A: 基于 SS 相对密码子使用度的聚类图; B: 基于 SS 基因编码蛋白质序列分析所得的进化分支图; 各编号对应物种的 SS 基因见表 1

Figure 4 Evolutional tree of SS genes

Note: A: Cluster analysis dendrogram of RSCU for SS genes; B: A dendrogram of protein sequences encoded by SS genes; The serial number respond to SS genes showed in table 1

过比较系统发育树的结果可知,在2种不同方法构建的系统发育树中,动物物种的5条SS基因都聚为一支,且聚类结果相同,植物类和微生物类物种的SS基因大体上聚类结果一致。但是基于RSCU的聚类结果显示,微生物类物种中的*Ajellomyces capsulatus*和*A. dermatitidis*的SS基因和植物类玉米聚为一类,而基于氨基酸序列的聚类结果则显示,上述2个物种与其它微生物类物种聚为一个大类,随后和动植物聚为一类。另外,在依据RSCU进行的聚类分析中,*N. gruberi*和白色念珠菌聚为一类,和其它的物种距离均较远,而在基于氨基酸序列的进化树中则和植物聚为一类。

2 讨论

大肠杆菌中的稀有密码子包括AGA、AGC、AUA、CCG、CCU、CUC、CGA和GUC,共8个(Grosjean and Fiers, 1982),若外源基因中包含这些稀有密码子时,会显著降低其表达效率(谢磊等, 2004)。SS基因的所有密码子中,RSCU>1.0的CUC、GUC、CCU、CGA、AGC、AGA(RSCU分别为1.35, 1.17, 1.08, 1.01, 1.04和1.29),这些密码子属于大肠杆菌表达系统的稀有密码子,因此利用大肠杆菌作为受体进行原核外源表达SS基因时,应将其优化为大肠杆菌的常用密码子,以提高外源基因的表达量。在以酵母为受体菌,进行真核外源表达SS基因时,同样应该根据酵母的密码子使用频率分布信息(<http://www.kazusa.or.jp/codon>),对外源基因的对应密码子进行优化才能获得良好的体外表达。这种改造外源基因的密码子以消除其存在的宿主中的稀有密码子,而提高表达效率的研究,已在其它基因中被证实。如通过定点突变扩展青霉(*Penicillium expansum*)脂肪酶基因中毕赤酵母的稀有密码子后,目的蛋白的表达量比野生基因重组子提高了2.3~2.5倍(张正平等, 2010)。因此,根据对SS基因密码子偏性分析的结果,结合宿主密码子偏性的特点进行优化后,对提高SS基因的外源表达效率也同样适用。

生物体内密码子偏性的存在是选择、突变和遗传漂变相互平衡的结果(Bulmer, 1987; 晁岳恩等, 2011)。密码子的使用并不是随机的,不同基因的密码子使用偏性存在不同的特点(Guo et al., 2012)。偏爱使用以A或T碱基结尾的密码子的基因有双子叶植物*waxy*基因(刘汉梅等, 2010),重要的抗寒调控转录因子ICE1(时慧等, 2012)等。偏爱使用以G或C碱

基结尾的密码子的有单子叶植物的*waxy*基因(刘汉梅等, 2010)和大豆光周期抑制开花的重要抑制因子RAV (related to ABI3/VP1) (杨春亮等, 2012)。SS基因密码子分析结果显示了SS基因偏好于使用以G或C碱基结尾的密码子,与上述的*waxy*和RAV基因的密码子偏性较为相似。SS基因在突变、选择等各种因素的作用下,产生了密码子偏性,其中突变的作用比选择的作用更强,这符合单基因密码子偏性形成的原因,多数是突变作用大于选择作用的特点(刘汉梅等, 2010)。

以不同物种SS基因的RSCU为基础构建的系统进化树和基于SS蛋白差异构建的系统发育树分析结果的总体趋势一致,但不完全相同。这种基于RSCU与蛋白序列不完全相同的现象在其它研究中也出现过(时慧等, 2012; 晁岳恩等, 2011)。造成聚类分析结果有偏差的原因可能是由于单个基因在物种中的突变,使密码子偏性出现大的不同,通过增加分析的基因类型和聚类时使用更多的指标,来弥补分析单一因素聚类的不足。基于RSCU的聚类分析可以作为传统的进化树建树的参考。

鲨烯合酶在初级代谢和次级代谢中均扮演着重要的角色。其催化合成的三萜皂苷类往往是有生理、药理活性的物质。本研究通过对SS基因密码子偏性进行分析,明确了不同物种中SS密码子使用的共同特点,为SS基因进行密码子优化,规避使用宿主稀有密码子和提高体外表达的效率提供借鉴。

3 材料和方法

3.1 目的基因的获取和软件

本研究所采用的各物种SS基因的全长编码区序列(coding DNA sequence, CDS)信息均来自GenBank。参照文献(时慧等, 2012)进行密码子分析时采用的CDS筛选原则,筛选符合条件的SS基因。根据该标准,共获得了47条来自不同物种的SS基因作为本研究的分析样本(表3)。

使用在线分析软件codon W(<http://mobylye.pasteur.fr/cgi-bin/portal.py?#forms:codonw>)和CUSP(<http://mobylye.pasteur.fr/cgi-bin/portal.py?#forms:cusp>),统计GC1、GC2、GC3和GC值等指标,同时统计CAI、RSCU、ENC、CBI和FOP等值。

3.2 同义密码子偏性分析

通过codon W 1.4.4在线软件计算样本的总体GC含量,记为GC。同时利用该软件计算RSCU、

表 3 SS 基因完整编码区核苷酸序列的来源
Table 3 Sources of coding sequences of SS genes

编号 Serial number	物种 Species	GenBank 登录号 Accession number	编号 Serial number	物种 Species	GenBank 登录号 Accession number
1	<i>Ajellomyces capsulatus</i>	XM_001536763	25	<i>Naegleria gruberi</i>	HM582092
2	<i>Ajellomyces dermatitidis</i>	XM_002620790	26	<i>Rattus norvegicus</i>	M95591
3	<i>Arthroderma gypseum</i>	XM_003170236	27	<i>Arabidopsis thaliana</i>	NM_119630
4	<i>Aspergillus niger</i>	XM_001395238	28	<i>Arabidopsis thaliana</i>	NM_119631
5	<i>Aspergillus oryzae</i>	XM_001827183	29	<i>Aralia elata</i>	GU354313
6	<i>Candida albicans</i>	D89610	30	<i>Artemisia annua</i>	AY445506
7	<i>Coprinopsis cinerea</i>	XM_001831035	31	<i>Azadirachta indica</i>	JQ327160
8	<i>Ganoderma lucidum</i>	DQ494674	32	<i>Botryococcus braunii</i>	AF205791
9	<i>Magnaporthe oryzae</i>	XM_003709815	33	<i>Bupleurum falcatum</i>	AY964186
10	<i>Mycosphaerella graminicola</i>	XM_003856154	34	<i>Centella asiatica</i>	AY787628
11	<i>Neurospora crassa</i>	XM_954724	35	<i>Chlorophytum borivillianum</i>	JQ975033
12	<i>Paracoccidioides brasiliensis</i>	XM_002788989	36	<i>Eleutherococcus senticosus</i>	HQ456917
13	<i>Penicillium minioluteum</i>	JN853774	37	<i>Eleutherococcus senticosus</i>	JN714465
14	<i>Pichia jadinii</i>	B012604	38	<i>Glycyrrhiza glabra</i>	D86409
15	<i>Saccharomyces cerevisiae</i>	X59959	39	<i>Gynostemma pentaphyllum</i>	FJ906799
16	<i>Schizosaccharomyces pombe</i>	L06071	40	<i>Huperzia serrata</i>	JQ004938
17	<i>Trichophyton rubrum</i>	XM_003233967	41	<i>Lotus japonicus</i>	AB102688
18	<i>Verticillium albo-atrum</i>	XM_003002687	42	<i>Oryza sativa</i>	AB007501
19	<i>Wolfiporia cocos</i>	JN688664	43	<i>Panax notoginseng</i>	DQ186630
20	<i>Yarrowia lipolytica</i>	AF092497	44	<i>Panax quinquefolius</i>	GU997681
21	<i>Danio rerio</i>	NM_001202525	45	<i>Polygala tenuifolia</i>	DQ672339
22	<i>Homo sapiens</i>	L06105	46	<i>Salvia miltiorrhiza</i>	FJ768961
23	<i>Homo sapiens</i>	S76822	47	<i>Zea mays</i>	NM_001111369
24	<i>Mus musculus</i>	D29016			

ENC、CAI、CBI 和 FOP，并进行对应性分析。利用 CUSP 软件分别计算密码子第 1、2 和 3 位碱基的 GC 含量，分别记作 GC1、GC2、GC3。使用 SPSS 16.0 软件分析各变量之间的相关性。

ENC 值是目前公认在评价基因整体密码子偏性用法中最具有参考价值的参数，该值用于衡量某条基因或某个物种的密码子偏好程度，其取值为 20~61，密码子偏性越大，该值越小，反之亦然(时慧等, 2012)。GC 测量的是基因中 G 和 C 的含量。GC3 则计算密码子第三个碱基中出现 G 或 C 的频率。

3.3 ENC 与 GC3 的相关性分析

利用 ENC 和 GC3 值构建散点图。同时构建标准曲线，即密码子偏性仅由碱基组成决定时的基因预期值，计算公式为： $ENC = 2 + GC3 + 29/[GC3^2 + (1 - GC3)^2]$ 。ENC 和 GC3 的相关关系可以用来分析密码子的使用受到突变和选择影响的幅度。当某基因的密码子偏性仅受突变影响时则分布于曲线附近，

若其受选择影响较大则分布于曲线下方较远的位置(尚明照等, 2011)。

3.4 中性绘图分析

将各物种 SS 基因的 GC12 (GC1 和 GC2 的平均值)作为纵坐标，GC3 作为横坐标进行绘图，分析 SS 基因密码子的 3 个位点碱基组成之间的相关性，确定影响密码子偏性的因素。如果 GC3 与 GC12 相关性显著，说明 3 个位点的碱基组成相同，碱基突变在密码子偏性的形成中影响较大。否则，基因组 GC 含量高度保守，密码子不同位点的碱基组成不同，密码子的偏好特点受选择的影响较大。

3.5 最优密码子分析

将 SS 基因中具最高和最低 ENC 值的各 10%样本分别作为高、低表达组。计算高、低表达组的 RSCU 差值，两者间的 RSCU 差值 > 0.08 的密码子确定为高表达最优密码子(尚明照等, 2011)。

3.6 各物种 SS 基因的聚类分析

利用 SPSS 16.0 软件分析各条 SS 基因的 RSCU 值,进行基于密码子使用偏性的聚类分析。以密码子为分组,采用 59 个密码子(去除编码蛋氨酸的密码子 AUG 和编码色氨酸的密码子 UGG 以及 3 个终止密码子 UAA, UAG, UGA)的 RSCU 值。将各条 SS 基因的 RSCU 值构建数据库,对密码子使用概率分析时,将每一条基因作为一个对象,RSCU 作为变量,利用各密码子 RSCU 值的欧式距离平方测定组间距离,把最接近的 2 类合并为一个新类,最后分为 n-1 类,输出树状结构图。

使用 MEGA 5.0 软件,采用 neighbor-joining 连接法,构建 47 条 SS 基因编码蛋白质的氨基酸序列的系统进化树。

作者贡献

本文在邢朝斌副教授的指导下,由何闪完成研究的设计、实施、数据整理以及论文撰写,修乐山进行数据汇总和整理,经由张淑杰教授共同修改后完成。

致谢

本研究由国家自然科学基金(30701086)、河北省自然科学基金(C2009001252)和河北省自然科学基金-石药集团医药联合研究基金(H2012401006)共同资助。在实验过程中得到了龙月红老师和吴鹏老师的大力帮助和支持,也得到了河北联合大学生命科学学院周秘、柴丽花、朱金丽、李宝财等同学的有益建议和帮助,在此,一并表示衷心的感谢。

参考文献

- Bulmer M., 1987, Coevolution of codon usage and transfer RNA abundance, *Nature*, 325(6106): 728-730
- Busquets A., Keim V., Closa M., Arco A.D., Boronat A., Arro M., and Ferrer A., 2008, *Arabidopsis thaliana* contains a single gene encoding squalene synthase, *Plant Mol. Biol.*, 67(1-2): 25-36
- Chao Y.E., Chang Y., Wang M.F., He S.L., Zhao X.L., and Lei Z. S., 2012, Codon usage bias and cluster analysis on chloroplastic genes from seven crop species, *Huabei Nongxuebao (Acta Agriculture Boreali-Sinica)*, 27(4): 60-64 (晁岳恩, 常阳, 王美芳, 何盛莲, 赵献林, 雷振生, 2012, 7 种作物叶绿体基因的密码子偏好性及聚类分析, *华北农学报*, 27(4): 60-64)
- Chao Y.E., Wu Z.Q., Yang H.M., He N., and Yang P., 2011,

- from 11 plant species, *Henong Xuebao (Journal of Nuclear Agricultural Sciences)*, 25(5): 927-932 (晁岳恩, 吴政卿, 杨会民, 何宁, 杨攀, 2011, 11 种植物 *psbA* 基因的密码子偏好性及聚类分析, *核农学报*, 25(5): 927-932)
- Grosjean H., and Fiers W., 1982, Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes, *Gene*, 18(3): 199-209
- Gu P., Ishii Y., Spencer T.A., and Shechter I., 1998, Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase, *The Journal of Biological Chemistry*, 273(20): 12515-12525
- Guo F.B., Ye Y.N., Zhao H.L., Lin D., and Wei W., 2012, Universal pattern and diverse strengths of successive synonymous codon bias in three domains of life, particularly among prokaryotic genomes, *DNA Research*, 19(6): 477-485
- Han J.Y., In J.G., Kwon Y.S., and Choi Y.E., 2012, Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*, *Phytochemistry*, 71(1): 36-46
- Kim Y.S., Cho J.H., Park S., Cho J.H., Han J.Y., Back K., and Choi Y.E., 2011, Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum*, *Planta*, 233(2): 343-355
- 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, *Henong Xuebao (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)
- Robinson G.W., Tsay Y.H., Kienzle B.K., Smith-Monroy C.A., and Bishop R.W., 1993, Conservation between human and fungal squalene synthetases: Similarities in structure, function, and regulation, *Mol. Cell Biol.*, 13(5): 2706-2717
- Shang M.Z., Liu F., Hua J.P., and Wang K.B., 2011, Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*, *Zhongguo Nongye Kexue (Scientia Agricultura Sinica)*, 44(2): 245-253 (尚明照, 刘方, 华金平, 王坤波, 2011, 陆地棉叶绿体基因组密码子使用偏性的分析, *中国农业科学*, 44(2): 245-253)
- Shi H., Wang Y., Yang L.C., and Ding Z.T., 2012, Analysis of codon bias of the cold regulated transcription factor ICE1 in tea plant, *Yuanyi Xuebao (Acta Horticulturae Sinica)*, 39(7): 1341-1352 (时慧, 王玉, 杨路成, 丁兆堂, 2012, 茶树抗寒调控转录因子 ICE1 密码子偏性分析, *园艺学报*, 39(7): 1341-1352)
- Summers C., Karst F., and Charles A.D., 1993, Cloning, expression and characterisation of the cDNA encoding human

- hepatic squalene synthase, and its relationship to phytoene synthase, *Gene*, 136(1-2): 185-192
- Tang L.P., Yu J.C., Dai D.F., and Fang B.S., 2011, Expression of glycerol dehydrogenase gene in *Escherichia coli* by codon optimization, *Weishengwu Xuebao (Acta Microbiologica Sinica)* 51(4): 504-509 (唐龙盘, 余劲聪, 戴丹凤, 方柏山, 2011, 甘油脱氢酶基因在大肠杆菌中的密码子优化表达, *微生物学报*, 51(4): 504-509)
- Xie L., Sun J.B., Zhang S.Q., and Huang J.S., 2004, Research progress in the *E. coli* expression system, *Huanan Redai Nongye Daxue Xuebao (Journal of South China University of Tropical Agriculture)*, 10(2): 16-20 (谢磊, 孙建波, 张世清, 黄俊生, 2004, 大肠杆菌表达系统及其研究进展, *华南热带农业大学学报*, 10(2): 16-20)
- Yang C.L., Wang L., Wu B., Zhao L., Zhang X.L., and Zhong S. Q., 2012, Codon bias analysis of soybean *GmRAV* gene, *Dongbei Nongye Daxue Xuebao (Journal of Northeast Agricultural University)*, 43(7): 34-41 (杨春亮, 王良, 武斌, 赵琳, 张晓丽, 钟淑琦, 2012, 大豆 *GmRAV* 基因的密码子偏好性分析, *东北农业大学学报*, 43(7): 34-41)
- Zhang Z.P., Yang J.K., Xu L., Liu Y., and Yan Y.J., 2010, Cloning, codon optimization and expression of mature lipase gene *Penicillium expansum*, *Weishengwu Xuebao (Acta Microbiologica Sinica)*, 50(2): 228-235 (张正平, 杨江科, 徐莉, 刘云, 闫云君, 2010, 扩展青霉脂肪酶基因克隆、密码子优化及表达, *微生物学报*, 50(2): 228-235)
- Zhao W., Li B., Liang J.X., Liang B.Z., Yao R.Y., Huang A.G., He Y., and Jiang Y.W., 2008, Codon preference analysis of principal non-structural and structural proteins of porcine parvovirus, *Guangxi Nongye Shengwu Kexue (Journal of Guangxi Agriculture and Biology Science)*, 27(4): 299-303, 314 (赵武, 李斌, 梁家幸, 梁保忠, 姚瑞英, 黄安国, 何颖, 蒋玉雯, 2008, 猪细小病毒主要非结构蛋白与结构蛋白的密码子偏爱性分析, *广西农业生物科学*, 27(4): 299-303, 314)
- Zhou J., Liu H., Du G., Li J., and Chen J., 2012, Production of alpha-cyclodextrin glycosyltransferase in *Bacillus megaterium* MS941 by systematic codon usage optimization, *J. Agric. Food Chem.*, 41(60): 10285-10292