



文章编号: 1000-4025(2012)02-0409-07

葡萄基因组密码子使用偏好模式研究

续 晨, 蔡小宁, 钱保俐, 贲爱玲

(南京晓庄学院 生物化工与环境工程学院, 南京 211171)

摘 要: 根据完整基因组序列, 运用多元统计分析和对应分析的方法, 探讨了葡萄全基因组序列密码子的使用模式和影响密码子使用的各种可能因素。结果显示: 葡萄密码子偏好性主要受到碱基差异 ($r=0.925$) 和自然选择 ($r=0.193$) 共同作用的影响, 突变压力占了主导因素, 自然选择的作用较小。同时基因长度和蛋白质疏水性也对密码子的偏好性有所影响。确定了葡萄的 20 个最优密码子。

关键词: 分子进化; 密码子用法; 密码子偏好性

中图分类号: Q786 **文献标志码:** A

Codon Usage Bias in *Vitis vinifera*

XU Chen, CAI Xiao-ning, QIAN Bao-li, BEN Ai-ling

(School of Biochemical and Environmental Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China)

Abstract: In this study, trends in synonymous codon usage in *Vitis vinifera* have been firstly examined through the multivariate statistical analysis on 1374 cDNA sequences. The results showed that the genes positions on the primary axis were strongly negatively correlated with GC3s, GC content of individual gene and gene expression level, which indicated that nucleotide composition and gene expression level were the main factors in shaping the codon usage of *V. vinifera*, and the variation in codon usage among genes may be due to mutational bias ($r=0.925$) at the DNA level and natural selection ($r=0.193$) acting at the level of mRNA translation. At the same time, CDS length and the hydrophobicity of each protein were, respectively, significantly correlated with codon bias.

Key words: molecular evolution; codon usage; codon bias

密码子是连接 DNA 和蛋白质的桥梁。每种氨基酸至少对应一个密码子, 最多对应 6 种密码子。编码同一种氨基酸的密码子称为同义密码子。在蛋白质的合成过程中, 同义密码子的使用概率并不相同。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, 此现象被称为密码子偏好性。密码子使用偏好性在不同的生物体中存在着很大的差异^[1], 甚至在同一个生物体基因组的不同基因之间都存在很大的差异。密码子使用偏好性的分析是一个错综复杂的

问题, 因为密码子使用偏好性的产生与基因表达水平^[2]、tRNA 丰度^[3]、基因长度^[4]、GC 含量^[5]等众多因素有关^[6-7], 到目前为止, 密码子的偏好性模式并无定论。密码子偏好性分析由最初的分析方法的建立, 软件的开发, 扩展到对低等生物进行密码子与基因表达关系分析, 当前随着不同高等真核基因组序列的破译, 人们开始关注高等植物、哺乳动物和人的基因组密码子的偏好性, 从而密码子偏好性问题会更加深入到广泛的生物学领域, 用于描述进化规律, 阐释分化机理, 揭示出更多的生物学意义^[8]。通过

收稿日期: 2011-08-08; 修改稿收到日期: 2011-11-21

基金项目: 江苏省高校自然科学基金(08KDJ180011); 南京晓庄学院青年专项基金项目(NXZ09039)

作者简介: 续 晨(1979-), 女, 在读博士, 讲师, 主要从事植物基因工程, 植物分子进化方面的研究。E-mail: xuchendea@hotmail.com

密码子使用偏好性找出密码子使用的一些规律,这将分子水平上的进化研究,转基因中外源基因的改造,新基因的发现具有重要指导意义。

葡萄(*Vitis vinifera*)几乎占全世界水果产量的四分之一,有“水果之神”的称号。Patrick^[9]领导科学家团队完成的葡萄全基因组测序工作为葡萄重要园艺性状功能基因组研究打下基础。葡萄基因组的测序是人类完成全基因组测序的第四种开花植物、第二种木本植物,也是科学家首次成功破译的水果类植物基因组。同时也为分析其密码子使用模式提供了丰富的数据。因此,本研究以葡萄全基因组序列为基础,对可能影响葡萄全基因组密码子偏好性的各种因素进行了分析,试图揭示密码子使用模式与分子进化之间的关系,为提升葡萄育种水平、重要性状研究和分子进化研究等提供基础理论数据。

1 材料和方法

1.1 数据获得

葡萄完整的 cDNA 序列从 NCBI 数据库下载(<http://www.ncbi.nlm.nih.gov/>)。为了做到最低限度的取样误差,编写 C 程序对序列进行筛选处理,提取满足下列条件的 CDS 序列共 1 376 条:选择长度大于或者等于 300 bp(100 个氨基酸),碱基数全长为 3 的倍数,并且拥有正确的起始密码子和终止密码子的序列。

1.2 同义密码子偏好性分析

使用 CodonW(1.4.2)进行密码子偏性分析。密码子偏性的度量指标包括:相对密码子使用度(relative synonymous codon usage, RSCU)、有效密码子数(effective number of codons, Nc)、密码子适应指数(codon adaptation index, CAI),密码子偏好性指数(codon bias index, CBI)和最优密码子使用频率(frequency of optimal codons, Fop)。RSCU 值的计算方法为某一密码子所使用的频率与其在无偏使用时预期频率之间的比值。Nc 值被认为是在评价基因整体密码子偏爱性用法中最具有参考价值的参数之一,目前被广泛用于评价基因密码子偏爱性,其取值范围为 20(每个氨基酸只使用一个密码子的极端情况)到 61(各个密码子均被平均使用)。CAI 值反映编码区同义密码子与密码子最佳使用相符合的程度^[10],取值范围在 0~1 之间,表达量较高的基因具有较高 CAI 值,表达量较低的基因具有较低的 CAI 值。统计全部 CDS 序列的碱基组成参数:鸟嘌呤和胞嘧啶含量(G+C 含量);同义密码子

第 3 位的 G+C 含量(GC3s);同义密码子在第 3 位上腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶的出现频率(A3s、T3s、G3s、C3s),CDS 序列碱基数(Number),该基因所表达蛋白质的疏水性(Gravy),及蛋白质中的芳香族氨基酸的含量(Amoro)。

使用对应分析探究样本各变量之间的关系,对应分析样本中所有基因按密码子的使用频率分布在一个 59 维(64 个密码子去除 3 个终止密码子以及甲硫氨酸和色氨酸的密码子)的向量空间,通过矩阵数据转换,鉴定出影响密码子使用偏性的主要因素。对应分析使用 CodonW(1.4.2)完成,各变量间的相关性分析使用 SPSS 14.0 完成^[11]。

1.3 测定最优密码子

本文首先确定了葡萄基因组中的高频密码子,将检测到的基因如某一密码子相对同义密码子使用频率单值超过 60%或者超过该组同义密码子平均占有频率的 1.5 倍的密码子即为高频密码子^[12]。最后采用 Stenico 等^[13]的高表达优越密码子分析方法,先计算每个基因的密码子有效数(Nc)和相对同义密码子用法(RSCU),然后再根据各 Nc 值确定高表达和低表达样本组,计算出这两个样本组中各个密码子各自的值,最后通过卡方检验确定出高表达基因的优越密码子,确定葡萄全基因组最优密码子。

2 结果与分析

2.1 最优密码子的确定

本研究对葡萄全基因组 1 376 个基因进行了初步分析,结果如表 1 显示,通过高频密码子分析法,通过高频密码子(同义密码子中出现次数最高的密码子)和相对同义密码子使用频率(RSCU)比较直观地反映了密码子的偏好性,确定了 UUG、UCU 等 20 个密码子为高频密码子。使用高频密码子分析法可以明显地看出高频密码子大部分以 A 或 U 结尾。采用高表达最优密码子分析方法^[14],根据高、低表达样本组中相应密码子的出现次数和 RSCU 值,通过计算基因之间同义密码子相对使用频率之差,并经卡平方测验,最终确定了 UUG、UCU 等 20 个密码子为葡萄的高表达优越密码子。它们的使用频率随基因的表达水平升高而增加,且在高、低表达基因样本组间的使用频率差异大,显著性测验结果均呈极显著水平。结果发现所确定的 20 个高表达优越密码子与高频密码子分析法确定的 20 个高频密码子完全相同(表 1 中打 * 的密码子)。这 20 个密码子不仅在葡萄全基因组中出现

频率高,且在高表达基因中的使用频率也高,综合两种评定方法,最终确定了这 20 种密码子为葡萄最优密码子。

2.2 Nc 和 GC3s 的关联分析

有效密码子使用个数(Nc)和 GC3s 的百分比通常用来研究生物体密码子使用的变异,葡萄全基因组中,Nc 值变异范围从 38.240~61.000,平均值为

52.898,标准差为 3.295。GC3s 含量变异范围从 24.50%~82.30%,平均值为 41.46%,标准差为 7.640。图 1 使用 Nc 与 GC3s 绘制散点图(Nc-plot)^[15],图中的连续标准曲线显示了无选择压力而在突变压的情况下 Nc 和 GC3s 的函数关系,如果某基因的密码子偏好是由碱基组成决定,即不受或少受翻译选择影响,则代表该基因的数据点应该落在

表 1 葡萄基因组的密码子使用模式

Table 1 Codon usage of the genome in *V. vinifera*

氨基酸 AA	密码子 Codon	全基因组序列 Genome		高表达序列 High		低表达序列 Low	
		数量 Number	RSCU	数量 Number	RSCU	数量 Number	RSCU
Phe	UUU*	291 490	25.2	1 537	1.39	371	0.65
	UUC	215 288	18.6	670	0.61	768	1.35
Leu	UUA	156 031	13.5	1 022	0.98	111	0.37
	UUG*	289 607	25.1	1 691	1.63	487	1.26
	CUU	255 515	22.1	1 586	1.53	328	0.85
	CUC	156 157	13.5	499	0.48	690	1.79
	CUA	128 918	11.2	645	0.62	165	0.13
	CUG	141 100	12.2	793	0.76	498	1.29
Ile	AUU*	301 597	26.1	1 745	1.64	357	0.98
	AUC	181 341	15.7	612	0.58	523	1.13
	AUA	158 431	13.7	830	0.78	218	0.6
Met	AUG*	301 968	26.1	1 608	1	532	1
Val	GUU*	268 788	23.3	1 827	1.76	381	1
	GUC	126 316	10.9	491	0.47	391	1.04
	GUA	120 281	10.4	758	0.73	146	0.38
	GUG	226 898	19.6	1 078	1.04	596	1.57
Tyr	UAU*	206 846	17.9	1 071	1.46	218	0.66
	UAC	121 432	10.5	392	0.54	113	1.31
TER	UAA	6 828	0.6	28	0.92	31	1.03
	UAG	6 815	0.6	26	0.86	26	0.87
His	CAU*	206 688	17.9	1 294	1.59	206	0.8
	CAC	105 740	9.1	337	0.41	307	1.2
Gln	CAA*	256 139	22.2	1 599	1.09	245	0.72
	CAG	184 009	15.9	1 334	0.91	435	1.28
Asn	AAU*	301 492	26.1	2 445	1.44	386	0.89
	AAC	168 187	14.6	942	0.56	485	1.11
Lys	AAA	337 782	29.2	2 751	1.05	108	0.76
	AAG*	388 631	33.6	2 494	0.95	663	1.24
Ser	UCU*	211 291	21.1	2 109	1.81	386	1.11
	UCC	145 986	2.6	737	0.63	516	1.18
	UCA	215 994	18.7	1 779	1.53	313	0.9
	UCG	64 626	5.6	248	0.21	370	1.06
Pro	CCU*	216 148	18.7	1 264	1.69	321	0.88
	CCC	105 200	9.1	372	0.5	360	0.98
	CCA	211 478	18.3	1 200	1.6	371	1.01
	CCG	60 461	5.2	160	0.21	416	1.13

续表1 Continued Table 1

Thr	ACU*	204 752	17.7	1 284	1.7	286	0.97
	ACC	127 287	11	479	0.63	155	1.51
	ACA	179 134	15.5	1 116	1.18	191	0.66
	ACG	19 960	4.3	110	0.19	243	0.83
Ala	GCU*	267 133	23.1	1 998	1.8	363	0.9
	GCC	155 209	13.1	636	0.57	558	1.38
	GCA	229 251	19.8	1 631	1.47	299	0.71
	GCG	58 533	5.1	176	0.16	393	0.97
Cys	UGU*	125 132	10.8	732	1.23	232	0.87
	UGC	91 398	8.2	463	0.77	300	1.13
TER	UGA	10 455	0.9	37	1.22	33	1.1
Trp	UGG*	180 435	15.6	761	1	351	1
Arg	CGU	69 669	6	191	0.78	118	0.68
	CGC	53 766	4.7	191	0.31	201	0.91
	CGA	76 515	6.6	111	0.65	179	0.82
	CGG	59 827	5.2	291	0.47	230	1.06
Ser	AGU	162 715	14.1	1 378	1.18	187	0.51
	AGC	118 518	10.3	715	0.64	320	0.92
Arg	AGA*	214 862	18.6	1 418	2.25	273	1.25
	AGG	185 039	16	975	1.55	272	1.25

注: * 为最优密码子, RSCU 为同义密码子的使用频率。

Note: Table with a * are optimal codon for each amino acid. Relative frequency of synonymous codon is abbreviated to RSCU.

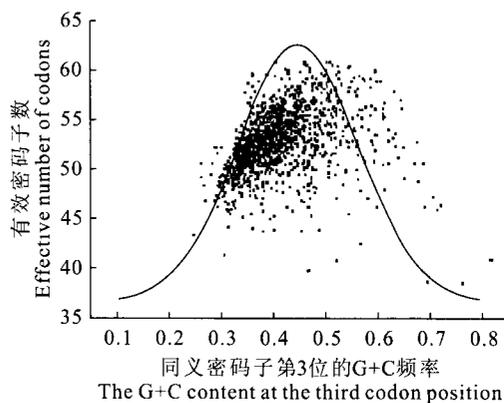


图1 葡萄中 Nc 值和 GC3s 的关系曲线

标准曲线表示了随机密码子用法中 Nc 和 GC3s 的预期曲线

Fig. 1 Nc-plot (Nc values vs GC3s) of the *V. vinifera* genes

The continuous curve represents the expected curve between GC3s and Nc under random codon usage

标准曲线附近;如果某基因中密码子偏好性严重,且与表达水平有较强的显著关系,Nc 值偏低,则数据点远远落在标准曲线下方。从图 1 来看实际观测值大部分远远低于标准曲线值,表明其值明显比期望值低,也就是说,除了 GC3s 含量这一由于中性碱基突变而引起的序列背景差异外,还存在其他因素对密码子的使用也具有重要影响,如弱选择效应等。

2.3 密码子使用的对应分析

为了调查各种可能的不同因素对密码子偏好性的不同影响趋势,作了各种因子的多元统计分析,其中包含了 G+C 含量和碱基在第三位置上的含量(GC3s、G3s、T3s、C3s 等),高表达基因密码子偏性有关的适应指数(CAI、CBI、Fop、Nc)等一些数值之间的相关分析,GC、GC3s、G3s、T3s、C3s 与 CAI、CBI、Fop、Nc 有极显著的相关性,表明碱基组成变异对密码子偏好性有重要影响。

根据 59 个有义密码子的用法,将所有基因所代表的点标定在一个 58 维的超维空间里进行对应分析,两主轴分别能解释基因总变异 12.63% 和 6.41% 的数据,第一轴的贡献率约为第二轴的 2 倍,第一轴总贡献率超过 10%,说明第一轴是密码子偏好性的主要贡献者。因此本研究对第一、二向量主轴(axis1、axis2)与其他因子之间也做了相关分析。从表 2 中可以看出大部分数值为极显著相关,说明密码子偏好性与很多因素有明显相关性。其中第一主轴与所有的数值都呈极显著相关,并且第一轴上基因的位置与 GC3s、T3s、C3s、G3s、GC 含量有很强的线性相关性($r=0.925, P<0.01$; $r=0.815, P<0.01$; $r=0.843, P<0.01$, $r=0.868, P<0.01$),通常 $|r|$ 大于 0.8 时,认为两个变量有很强的线性相关

表 2 各因子的相关系数

Table 2 Correlation coefficients of the factors

	密码子第 3 位上各嘧啶的含量				密码子 适应指数 CAI	密码子 偏好性指数 CBI	最优密码子 使用频率 Fop	有效密码 子数 Nc	密码子第 3 位 为 G+C 的含量 GC3s	G+C 含量 GC	疏水性 Gravy	芳香族氨基酸 Aromo	基因长度 Gene number	向量 1 轴 Axis 1
	T3s	C3s	A3s	G3s										
C3s	-0.787**													
A3s	0.380**	-0.575**												
G3s	-0.175**	0.236**	-0.575**											
CAI	0.036	0.362**	-0.303**	-0.058**										
CBI	-0.463**	0.681**	-0.529**	0.047	0.597**									
Fop	-0.377**	0.613**	-0.184**	0.038	0.703**	0.955**								
Nc	-0.319**	0.319**	-0.156**	0.113**	0.023	0.181**	0.151**							
GC3s	-0.856**	0.812**	0.763**	0.699**	0.195**	0.553**	0.501**	0.290**						
GC	-0.678**	0.612**	-0.703**	0.116**	0.230**	0.573**	0.560**	0.210**						
Gravy	-0.115**	0.159**	-0.212**	-0.119**	-0.104**	0.160**	-0.033	0.082**	0.091**	-0.059*				
Aromo	0.079*	0.237**	-0.091**	-0.050	-0.003	0.021	0.011	0.075**	0.103**	0.158**	0.111**			
Gene number	0.351**	-0.362**	-0.227**	0.153**	0.120**	0.225**	0.215**	0.173	0.319**	-0.311**	-0.065**	-0.161**		
Axis1	-0.813**	0.868**	-0.851**	0.482**	0.193**	0.629**	0.551**	0.312	0.925**	0.815**	0.213**	0.221**	-0.383**	
Axis2	0.059*	-0.067*	0.221**	0.193	0.190**	0.027	0.168**	0.009	0.230**	0.115**	-0.755**	-0.567**	-0.011	0.083**

注:CAI. 密码子适应指数;CBI. 密码子偏好性指数;Fop. 最优密码子使用频率;Nc. 有效密码子数;*和**分别表示相关性达到 0.05 和 0.01 显著水平。

Note:CAI. Codon adaptation index;CBI. Codon bias index;Fop. Frequency of optimal codons;Nc. Effective number of codons;* and ** indicate significant correlation at 0.05 and 0.01 level, respectively.

性。说明影响葡萄基因表达的密码子偏性的主要因素为碱基组成的变异。第一轴虽然与 CAI、CBI、Fop、Nc 也是极显著相关 ($r=0.193, P<0.01$; $r=0.629, P<0.01$; $r=0.551, P<0.01$), 但是很明显 $|r|$ 小于 0.8, 比前一组相关系数偏小, 所以认为基因的表达可能是塑造葡萄基因组密码子使用的次要因素。表明了突变引起的序列差异比自然选择表达对密码子偏好性的影响要大。即不是由单一因素决定密码子的偏好性, 而是碱基突变和自然选择一起决定的。另外, 基因长度及氨基酸的疏水性对密码子第一主轴也有显著影响, 表明这些因素都与密码子使用偏好性相关。其中 CDS 长度与向量主轴呈负相关, 可以解释为在自然选择压力下缩短表达量高的基因对生物体本身有利, 可以减少物质、能量和时间的消耗。相反地, 对于大的基因, 可能存在更多的更严紧的调控因子, 因而在自然选择过程中承受着更大的选择压。

从表 2 可知, 基因长度与 GC3s、Nc 呈极显著负相关 ($r=-0.349, P<0.01$; $r=-0.173, P<0.01$),

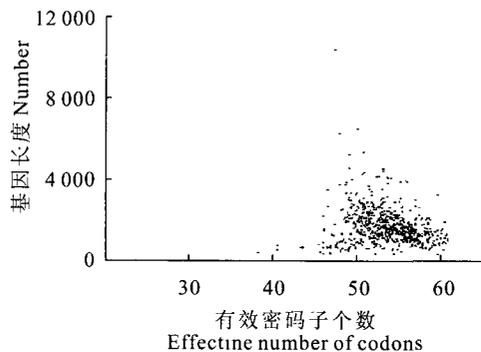


图 2 葡萄基因中基因长度与有效密码子个数使用分布图

Fig. 2 Nc distribution of *V. vinifera* genes

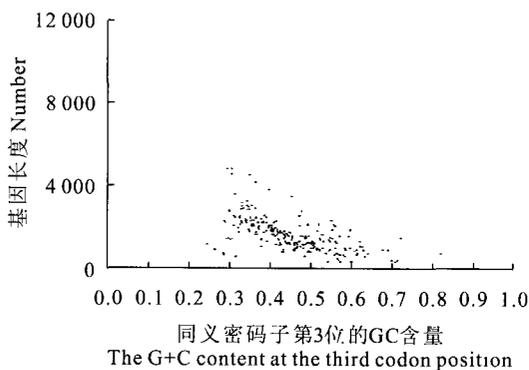


图 3 葡萄基因中的基因长度与 GC3s 分布图

Fig. 3 Compositional distribution of GC3s of *V. vinifera* genes

基因越长能够容纳的密码子越多。如图 2、3 所示, 基因长度越长, GC3s 值越小, 也就是随着基因长度的增加, 以 GC 结尾的密码子的量减少, 表明了随着基因长度的增加, 更倾向于选择使用以 AU 结尾的同义密码子。可能由于越长的基因翻译的时间越长越需要消耗更多的能量, 较长的基因为了节约时间能量需要特异地进行选择使用密码子。因此基因长度也影响着密码子的偏性。

同样从表 2 数据可以看出, 不同基因所形成蛋白质的疏水性也对密码子有一定的影响, 但相关系数很小, 相关性不高。

3 讨论

在长期的进化过程中, 每种生物体都形成了为自己特有的密码子用法。中性理论和“选择-突变-漂变”是密码子偏好性解释的两个经典模型。中性理论认为在密码子第三位上的突变是中性选择的结果, 同义密码子突变并不影响生存适合度, 这种对密码子的选择仅仅和突变有关, 不受自然选择压力影响; “选择-突变-漂变”模型则认为, 突变的发生是有方向的, 同义密码子使用的偏好性反映了对最优密码子的选择和对非同义密码子的“突变-漂变”两方面的作用。在本研究中, 葡萄基因组是偏向使用以 A 或 U 结尾的密码子。从本研究中可以发现葡萄基因序列碱基相关的因素 (T3s、C3s、A3s、G3s、GC3s、GC) 和第一向量主轴间的相关系数均呈极显著相关 ($r=-0.843, P<0.01$; $r=0.868, P<0.01$; $r=-0.735, P<0.01$; $r=0.482, P<0.01$; $r=0.925, P<0.01$; $r=0.815, P<0.01$), 说明密码子使用偏好性受碱基差异较大, 得到葡萄密码子偏好性主要受到碱基差异 ($r=0.925$) 和自然选择 ($r=0.193$) 共同作用的影响, 但是碱基差异占了主导因素, 两者影响程度差异悬殊。虽然自然选择的作用较小, 但是不可忽视。葡萄密码子偏好性是突变压和选择压共同作用的结果, 印证了后者。目前进行了全基因组测序的植物并不多。在水稻^[16]的研究中, 其基因表达水平 ($r=-0.608$) 明显大于碱基组成 ($r=0.344$) 对它的影响。杨树^[17]的研究则是其影响因素很复杂, 都有些相关但相关性较小 ($r<0.3$)。对蝴蝶兰叶绿体的研究^[11]表明, 其受碱基差异 ($r=0.614$) 和自然选择 ($r=0.508$) 共同影响, 并且两者影响程度差别不大。密码子使用模式是一个复杂的问题, 有待更多的全基因组测序结果的分析 and 基因组之间密码子使用的比较。

葡萄大约出现在1 300 000~2 000 000年前,与其他已测序的开花植物(杨树、水稻、拟南芥)相比是更古老的植物,也是较早的栽培植物,其已有7 000多年的栽培史,在进化上除了自然选择和基因突变

外,人工选择的痕迹是否可以在密码子模式中被发现,如何被发现,还是在葡萄进化的历史长河中栽培中的人工选择时间可忽略不计还需进一步探讨。

参考文献:

- [1] ZHANG G Y(张光亚),GE H H(葛慧华),FANG B SH(方柏山). Analysis of synonymous codon usage in two hyperthermophiles[J]. *Journal of Huaqiao University*(华侨大学学报),2006,**22**(2):87-90(in Chinese).
- [2] SHI X F(石秀凡),HUANG J F(黄京飞),LIU SH Q(柳树群),*et al.* The features of synonymous codon bias and GC-content relationship in human genes[J]. *Progress in Biochemistry and Biophysics*(生物化学与生物物理进展),2002,**29**(3):411-414(in Chinese).
- [3] MORIYAMA E N,POWELL J R. Codon usage bias and tRNA abundance in drosophila[J]. *Journal of Molecular Evolution*,1996,**45**(5):514-523.
- [4] KNIGHT R D,FREELAND S J,LANDEWEBER L F. A simple model based on mutation and selection explains trends in codon and amino acid usage and GC composition within and across genomes[J]. *Genome Biol.*,2001,**2**(1):1-12. <http://genomebiology.com/2001/2/1/research/0010.3>.
- [5] CARLINI D B,CHEN Y,STEPHAN W. The relationship between third codon position nucleotide content codon bias,mRNA secondary structure and gene expression in the drosophild alcohol dehydrogenase genes Adh and Adhr[J]. *Genetics*,2001,**159**(2):623-633.
- [6] LIU H M(刘汉梅),HE R(何 瑞),ZHANG H Y(张怀渝),*et al.* Analysis of synonymous codon bias in maize[J]. *Journal of Agricultural Biotechnology*(农业生物技术学报),2010,**18**(3):156-161(in Chinese).
- [7] MARIN A,GONZALEZ F,GUTIERREZ G. Gene length and codon usage bias in *Drosophila melanogaster*,*Saccharomyces cerevisiae* and *Escherichia coli*[J]. *Nucleic Acids Research*,1998,(19):223-226.
- [8] WU X M(吴宪明),WU S F(吴松锋),REN D M(任大明),*et al.* The analysis method and progress in the study of codon bias[J]. *Hereditas*(遗传),2007,**29**(1):120-126(in Chinese).
- [9] OLIVIER J,JEAN M A,BENJAMIN N. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla [J]. *Nature*,2007,449:163-167
- [10] GUPTA S K,BHATTACHARYYA T K,GHOSH T C. Synonymous codon usage in lactococcus lactis:mutational bias versus translational selection[J]. *J. Biomol. Struct. Dyn.*,2004,**21**(1):527-536.
- [11] XU CH(续 晨),BEN A L(贡爱玲),CAI X N(蔡晓宁). Analysis of synonymous codon usage in chloroplast genome of *Phalaenopsis aphrodite*[J]. *Molecular Plant Breeding*(分子植物育种),2010,**8**(5):945-950(in Chinese).
- [12] KAWABE A,MIYASHITA N T. Patterns of codon usage bias in three dicot and four monocot plant species[J]. *Genes & Genetic Systems*,2003,**78**(5):34-35.
- [13] STENICO M,LLOYD A T,SHARP P M. Codon usage in *Caenorhabditis elegans*; delineation of translational selection and mutational biases[J]. *Nucleic Acids Research*,1991,**22**(13):2 137-2 116.
- [14] KAWABE A,MIYASHITA N T. Patterns of codon usage bias in three dicot and four monocot plant species[J]. *Genes & Genetic Systems*,2003,**78**(5):343-352.
- [15] WRIGHT F. The "effective number of codon" used in a gene[J]. *Gene*,1990,87:23-29.
- [16] LIU Q P(刘庆坡),XUE Q ZH(薛庆中). Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. *japonica*)[J]. *Acta Agronomica Sinica*(作物学报),2004,**30**(12):1 220-1 224(in Chinese).
- [17] ZHOU M,LONG W,LI X. Analysis of synonymous codon usage in chloroplast genome of *Populus alba*[J]. *Journal of Forestry Research*,2008,**19**(1):293-297.