

文章编号: 1673 1689(2010)05-0755-10

基于 RSCU 与 QRSCU 的 F/10 和 G/11 木聚糖酶家族密码子偏好性的比较

赵静静¹, 齐斌^{1,2}, 莉娟^{1,3}, 唐旭清^{*1}

(1. 江南大学理学院, 江苏 无锡 214122; 2. 无锡职业技术学院, 江苏 无锡 214122; 3. 无锡市青山高级中学, 江苏 无锡 214100)

摘 要: 分别计算了 F/10 和 G/11 木聚糖酶家族的同义密码子的相对使用度和拟相对使用度, 即 RSCU 和 QRSCU. 通过分析和比较得到, 基于拟氨基酸基因编码方法能更明显的展示出密码子家族中对同义密码子的一致偏好性. 也就是说, 基于拟氨基酸基因编码方法下的 F/10 与 G/11 木聚糖酶家族更偏好使用密码子-反密码子结合作用强的密码子, 恰好是以 g/c 结尾的密码子, 与前人的偏好性研究结果一致, 而且进一步验证了拟氨基酸基因编码方法与密码子偏好性的研究结果密切相关.

关键词: 木聚糖酶; 拟氨基酸基因编码; 密码子偏好性; RSCU; QRSCU

中图分类号: O 29; TP 391; Q 612

文献标识码: A

Based on RSCU and QRSCU Research Codon Bias of F/10 and G/11 Xylanase

ZHAO Jing-jing¹, QI Bin^{1,2}, DING Li-juan^{1,3}, TANG Xu-qing^{*1}

(1. School of Science, Jiangnan University, Wuxi, 214122, China; 2. Wuxi Institute of Technology, Wuxi 214122, China; 3. Wuxi Qingshan Senior Middle School, Wuxi 214100, China)

Abstract: According to the work of Shi Xiufan et al. (2000 Chinese Science Bulletin 45 2520) and Zhu Ping et al. (Chinese physics B, 1(2009)), this study computes the relative usage degree of the synonymous codon of F/10 and G/11 Xylanase: (RSCU and QRSCU). Through the analysis and comparison, it was found that the quasar amino acid shown more consistent preference to the synonymous codon. This result indicated that F/10 and G/11 Xylanase prefer to use the codons with strong combination of the codon/anticodon, just the codons ending with g/c based on the classification of quasar amino acids. The results present here demonstrated a good accordance with the results that achieved on the 78 human genes and further verified codon preference closely related to quasar amino acids coding method.

Key words: xylanase, quasar amino acids, codon bias, RSCU, QRSCU

收稿日期: 2009-06-22

基金项目: 财政部专项资金项目(200709017); 江南大学创新团队发展计划项目。

* 通信作者: 唐旭清(1963-), 男, 安徽望江人, 工学博士, 副教授, 主要从事智能计算、生物信息学及数学模型方面的研究。Email: txq5139@jiangnan.edu.cn

木聚糖酶是可将木聚糖降解成低聚木糖和木糖的一类酶的总称, 它可以将木聚糖降解为寡聚木糖、木糖和少量单糖, 其主要分为 F/10 和 G/11 两个家族。F/10 和 G/11 木聚糖酶家族在空间结构上属于两种不同的折叠类型。F/10 家族木聚糖酶三维结构为 $(\alpha/\beta)_8$ 折叠桶类型。G/11 家族木聚糖酶三维结构为 β 折叠片为主所构成的单个结构域。目前木聚糖酶广泛应用于工业领域, 比如制浆造纸、饲料加工、酿酒、烘焙食品、制备功能性低聚糖、临床、环境保护等领域。

由于木聚糖酶在工业应用方面的重要作用, 长期以来, 人们对其做了大量研究^[1-5]。在对 F/10 和 G/11 木聚糖酶的研究中, 其重要的内容之一便是对其同义密码子的研究。比如, 刘亮伟等人^[1,3,5] 用生物信息学的方法, 研究了 F/10 和 G/11 木聚糖酶氨基酸二联体同最适温度、最适 pH 值的关系、木聚糖酶的分子进化以及用主成分分析的方法对 F/10 和 G/11 木聚糖酶分类等一系列的问题。作者则是综合了石秀凡等人^[6] 基于氨基酸编码方法下对同义密码子偏好性及朱平等^[7-8] 基于拟氨基酸基因编码方法下对同义密码子的偏好性的研究成果, 运用两种编码方法, 进行编程, 分别计算 RSCU 和 QRSCU, 得出两种编码方法下同义密码子都偏好使用以 g/c 结尾的密码子, 而且基于拟氨基酸基因编码方法下, 其结论更加明显。

1 材料与方法

1.1 数据库的构建与编程

根据文献[1]中 Swiss-Prot 数据库中木聚糖酶的登录号, 寻找 GenBank 和 EMBL-EBI 数据库中对应的编码序列, 建立木聚糖酶基因的编码序列库。文中取了 F/10 木聚糖酶家族的 28 条序列, 总共包含 13 790 个密码子, GC 占编码区碱基总数的 48.2%。G/11 木聚糖酶家族的 33 条序列, 总共包含 12 077 个密码子, GC 占编码区碱基总数的 50.6%。

F/10 木聚糖酶的编码序列库(28 条序列):
O59859, P07528, P26223, P29417, P33559, P40942, P40943, P40944, P45703, P51584, P56588, Q60041, Q60042, Q00177, Q60037, Q12603, P14768, P07986, P10478, P23556, P23360, P07529, P48789, P49942, P23030, O60206, P23557。

G/11 木聚糖酶的编码序列库(33 条序列):

O43097, P00694, P09850, P17137, P18429, P23031, P26220, P26515, P29127, P33557, P33558, P35811, P36217, P36218, P45705, P45796, P48791, P48824, P54865, P55328, P55329, P55330, P55331, P55332, P55333, P55334, P55335, P77853, P83513, Q06562, Q12667, Q53317。

文中所用的数据都是用 C 语言编程完成的。

1.2 密码子的相对使用度

1.2.1 氨基酸编码方法及 RSCU 同义密码子是指编码同一种氨基酸的两个或两个以上的密码子。自遗传密码破译以来大量的研究发现, 同义密码子的使用与基因表达存在着密切的关系。作者用密码子的相对使用度来研究密码子的偏好性。

由文献[6]知, 基于氨基酸编码方法下的同义密码子的相对使用度(RSCU) 计算公式为: $RSCU = \text{同义密码子的观测数} / \text{同义密码子的期望数}$, 同义密码子的期望数= 氨基酸的同义密码子出现次数/ 编码此种氨基酸的同义密码子数。

1.2.2 拟氨基酸基因编码方法及 QRSCU 基因组遗传密码字典自确立以来, 在 40 年的科学研究中被不断的扩充和完善。作为生命组成不可缺少的一个部分, 该遗传密码显示出了它在进化上的复杂性和适应性, 但也充分地显示了人类在遗传密码认识上的局限性。正因为这些原因, 人们完全揭示遗传密码内在特性的道路还很漫长, 还需要进行不断的研究和尝试。目前, 有学者从另一角度对此进行了大量探讨, 其中拟氨基酸基因编码方法就是一个重要形式。

定义 1: 令 $WC = \{tgg, tgt, tga, tgc\}$, $RS = \{agg, agt, aga, agc\}$, $LF = \{ttg, ttt, tta, ttc\}$, $MI = \{atg, att, ata, atc\}$, $ED = \{gag, gat, gaa, gac\}$, $-Y = \{tag, tat, taa, tac\}$, $KN = \{aag, aat, aaa, aac\}$, $QH = \{cag, cat, caa, cac\}$, $S' = \{tcg, tct, tca, tcc\}$, $R' = \{cgg, cgt, cga, cgc\}$, $L' = \{ctg, ctt, cta, ctc\}$ 。Let $ZU = \{G, V, ED, A, WC, LF, -Y, S', RS, MI, KN, T, R', L', QH, P\}$, 则 ZU 称为拟氨基酸集, 这里, -Y 为新终止子, 但是 $(tga \in) WC$ 不是终止子^[7]。

由文献[8]知, 基于拟氨基酸基因编码方法下同义密码子的相对使用度(QRSCU) 也相应被提出, 其计算公式为: $QRSCU = \text{同义密码子的观测数} / \text{同义密码子的期望数}$, 同义密码子期望数= 拟氨基酸的同义密码子出现次数/ 编码此种拟氨基酸的同义密码子数。

1.3 密码字的偏好性强度划分

与[1]的强度划分相似, 计算出两个家族中编码序列在氨基酸编码方法和拟氨基酸基因编码方法下密码子偏好性数值后, 为了区分两个木聚糖酶家族在两种编码方法下密码子偏好性差异及分析密码子偏好使用情况, 分别对其进行了强度划分. 因此做如下规定:

- 1) 基于氨基酸编码方法下以 G/11 家族中偏好性最高值 2.0 和最低值 0.181 8 为标准, 将这两个数据之间的差距划分为 4 个等级, 划分如下: A 无偏好性: $RSCU \leq 1.0$; B 低度偏好性: $1.0 < RSCU < 1.15$; C 中度偏好性: $1.15 \leq RSCU \leq 1.3$; D 高度偏好性: $RSCU \geq 1.3$ 。
- 2) 基于拟氨基酸基因编码方法下以 F/10 家族

中偏好性最高值 2.808 899 和 G/11 家族中偏好性最低值 0.010 8 为标准, 将这两个数据之间的差距划分为 4 个等级, 划分如下: A 无偏好性: $QRSCU \leq 1.05$; B 低度偏好性: $1.05 \leq QRSCU \leq 1.3$; C 中度偏好性: $1.3 \leq QRSCU \leq 1.4$; D 高度偏好性: $QRSCU \geq 1.4$ 。见表 3、4。

2 结果与分析

根据 1.1 节中所建立的数据库, 包括 F/10 木聚糖酶的编码序列库 (28 条序列) 和 G/11 木聚糖酶的编码序列库 (33 条序列), 按 1.2 节所给出的公式进行计算, 得出相对使用度 RSCU 和拟相对使用度 QRSCU 表, 见表 1~ 2。

表 1 基于氨基酸编码方法下 F/10 和 G/11 的密码子偏好情况
Tab.1 Codon bias of F/10 and G/11 based on amino acids coding method

氨基酸	密码子	F/10 木聚糖酶 密码子数 RSCU		G/11 木聚糖酶 密码子数 RSCU	
Val	总数(total)	974		767	
	gtt	258	1.059 55	229	1.194 3
	gtg	270	1.108 83	168	0.876 1
	gtc	262	1.075 98	237	1.235 98
	gta	184	0.755 65	133	0.693 6
Ile	总数(total)	813		479	
	att	311	1.147 6	174	1.089 8
	atc	348	1.284	250	1.565 76
	ata	154	0.568 3	55	0.344 5
Met	atg	263	1	231	1
Leu	总数(total)	1 008		542	
	ctt	239	1.422 62	143	1.583
	ctg	240	1.428 57	109	1.206 6
	ctc	212	1.261 9	137	1.516 6
	cta	41	0.244 0	21	0.232 5
	ttg	164	0.650 8	53	0.586 7
	tta	112	0.666 7	79	0.874 5
Phe	总数(total)	576		422	
	ttt	283	0.982 64	159	0.753 6
	ttc	293	1.017 4	263	1.246 4
Ala	总数(total)	1 143		827	
	gct	294	1.028 87	306	1.480
	gcg	231	0.808 4	111	0.534 9
	gcc	316	1.105 86	229	1.107 6
	gca	302	1.056 87	181	0.875 5

续表 1

氨基酸	密码子	F/10 木聚糖酶		G/11 木聚糖酶	
		密码子数 RSCU		密码子数 RSCU	
Thr	总数(total)	786		1113	
	act	147	0.748 1	275	0.988 3
	acg	180	0.916	205	0.736 75
	acc	259	1.318 1	401	1.441
	aca	200	1.017 8	232	0.833 8
Pro	总数(total)	609		406	
	cct	140	0.919 5	121	1.192 1
	cgg	196	1.287 4	97	0.955 7
	ccc	136	0.893 3	92	0.906 4
	cca	137	0.899 8	96	0.945 8
Ser	总数(total)	960		1177	
	tct	131	0.818 75	223	1.136 79
	tcg	123	0.768 75	119	0.606 63
	tcc	194	1.212 5	225	1.146 98
	tca	142	0.887 5	100	0.509 77
	agt	137	0.856 25	196	0.999 15
	agc	233	1.456 25	314	1.600 68
Gly	总数(total)	962		1366	
	ggt	237	0.985 45	479	1.402 6
	ggg	93	0.386 7	99	0.289 9
	ggc	319	1.326 4	537	1.572 5
	gga	313	1.301 45	251	0.735
Arg	总数(total)	521		372	
	cgt	68	0.783 1	99	1.596 8
	cgg	50	0.735 29	34	0.548 39
	cgc	114	1.312 9	95	1.532 258
	cga	33	0.380 04	30	0.483 87
	agg	92	1.059 5	22	0.354 84
	aga	164	1.888 7	92	1.483 9
Cys	总数(total)	125		126	
	tgt	53	0.848	68	1.079 4
	tgc	72	1.152	58	0.920 6
Trp	tgg	316	1	380	1
Asp	总数(total)	844		566	
	gat	438	1.037 9	293	1.035 3
	gac	406	0.962	273	0.964 7

续表 1

氨基酸	密码子	F/10 木聚糖酶		G/11 木聚糖酶	
		密码子数 RSCU		密码子数 RSCU	
Asn	总数(total)	866		989	
	aat	360	0.831 4	341	0.689 6
	aac	506	1.168 59	648	1.310 4
His	总数(total)	258		163	
	cat	113	0.875 97	72	0.883 4
	cac	145	1.124 03	91	1.116 6
Tyr	总数(total)	607		716	
	tat	275	0.906 09	265	0.740 2
	tac	332	1.093 9	451	1.259 8
Glu	总数(total)	755		378	
	gag	309	0.818 54	183	0.968 3
	gaa	446	1.181 46	195	1.031 7
Gln	总数(total)	515		571	
	cag	321	1.246 6	431	1.509 63
	caa	194	0.753 4	140	0.490 36
Lys	总数(total)	861		453	
	aag	447	1.038 3	325	1.434 9
	aaa	414	0.961 7	128	0.565 1
Stop	总数(total)	28		33	
	tga	9	0.964 3	9	0.818 2
	tag	8	0.857 1	2	0.181 8
	taa	11	1.1786	22	2

表 2 基于拟氨基酸基因编码方法下 F/ 10 和 G/ 11 的密码子偏好情况
Tab.2 Codon bias of F/ 10 and G/ 11 based on quasi- amino acids coding method

拟氨基酸	密码子	F/10 木聚糖酶		G/11 木聚糖酶	
		密码子数 QRSCU		密码子数 QRSCU	
G	总数	962		1 366	
	ggg	93	0.386 7	99	0.289 9
	ggt	237	0.985 45	479	1.402 6
	gga	313	1.301 45	251	0.735
	ggc	319	1.326 4	537	1.572 5
WC	总数	450		515	
	tgg	316	2.808 89	380	1.112 7
	tgt	53	0.471 1	68	0.239 4
	tga	9	0.08	9	0.031 7
	tgc	72	0.64	58	1.508 1
RS	总数	626		624	
	agg	92	0.587 86	22	0.141 0
	agt	137	0.875 4	196	1.256 4
	aga	164	1.047 9	92	0.589 7
	agc	233	1.488 8	314	2.012 8

续表 2

拟氨基酸	密码子	F/10 木聚糖酶		G/11 木聚糖酶	
		密码子数 QRSCU		密码子数 QRSCU	
R´	总数	265		258	
	cgg	50	0.754 72	34	0.527 1
	cgt	68	1.026 4	99	1.534 9
	cga	33	0.498	30	0.465 1
	cgc	114	1.720 8	95	1.472 9
V	总数	974		767	
	gtg	270	1.108 83	168	0.876 1
	gtt	258	1.059 55	229	1.194 3
	gta	184	0.755 65	133	0.693 6
	gtc	262	1.075 98	237	1.235 98
LF	总数	852		554	
	ttg	164	0.769 95	53	0.382 7
	ttt	283	1.328 6	159	1.148 0
	tta	112	0.525 8	79	0.570 4
	ttc	293	1.375 59	263	1.898 9
MI	总数	1 076		710	
	atg	263	0.977 7	231	1.301 4
	att	311	1.156 1	174	0.980 3
	ata	154	0.572 5	55	0.309 9
	atc	348	1.293 7	250	1.408 45
L´	总数	732		410	
	ctg	240	1.311 5	109	1.063 4
	ctt	239	1.306 0	143	1.395 1
	cta	41	0.224 0	21	0.204 9
	ctc	212	1.158 5	137	1.336 6
ED	总数	1599		944	
	gag	309	0.773	183	0.775 4
	gat	438	1.095 7	293	1.241 5
	gaa	446	1.115 7	195	0.826 3
	gac	406	1.015 6	273	1.156 8
– Y	总数	626		740	
	tag	8	0.051 1	2	0.010 8
	tat	275	1.757 2	265	1.432 4
	taa	11	0.070 3	22	0.118 9
	tac	332	2.121 4	451	2.437 8
KN	总数	1 727		1 442	
	aag	447	1.035 3	325	0.901 5
	aat	360	0.833 8	341	0.945 9
	aaa	414	0.958 89	128	0.355 1
	aac	506	1.171 97	648	1.797 5

续表 2

拟氨基酸	密码子	F/ 10 木聚糖酶		G/ 11 木聚糖酶	
		密码子数 QRSCU		密码子数 QRSCU	
QH	总数	773		734	
	cag	321	1. 661	431	2. 348 8
	cat	113	0. 584 7	72	0. 392 4
	caa	194	1. 003 9	140	0. 762 9
	cac	145	0. 750 3	91	0. 495 9
A	总数	1143		827	
	gcg	231	0. 808 4	111	0. 536 9
	gct	294	1. 028 87	306	1. 480
	gca	302	1. 056 87	181	0. 875 5
	gcc	316	1. 105 86	229	1. 107 6
S´	总数	590		667	
	tcg	123	0. 833 9	119	0. 713 6
	tct	131	0. 888 1	223	1. 337 3
	tca	142	0. 962 7	100	0. 599 7
	tcc	194	1. 315 25	225	1. 349 3
T	总数	786		1 113	
	acg	180	0. 916	205	0. 7367 5
	act	147	0. 748 1	275	0. 988 3
	aca	200	1. 017 8	232	0. 833 8
	acc	259	1. 318 1	401	1. 441
P	总数	609		406	
	ccg	196	1. 287 4	97	0. 955 7
	cct	140	0. 919 5	121	1. 192 1
	cca	137	0. 899 8	96	0. 945 8
	ccc	136	0. 893 3	92	0. 906 4

高度和中度偏好密码子类型见表 3~ 4。

表 3 基于氨基酸编码方法下 F/ 10 与 G/ 11 木聚糖酶密码子偏好性
Tab.3 Codon bias of F/ 10 and G/ 11 xylanase based on amino acids coding

木聚糖酶	项目	高	中	低
F/ 10 家族	密码子	cgc, agc, ctg, ctt, ggc, acc, gga, aga	c cg, atc, etc, cag, gaa, taa, aac, tcc, tgc	att, gtg, gcc, tac, gtc, gtt, gca, aag, gat, gct, aca, ttc, agg, cac
	a/ u 结尾	3	2	6
	g/ c 结尾	5	7	8
G/ 11 家族	密码子	taa, agc, ggc, ctt, at c, cgt, etc, cag- gct, cgc, acc, aag, ggt, aac, aga	tac, ttc, gtc, ctg, gtt, cct	tcc, tct, cac, gcc, att, tgt, gat, gaa
	a/ u 结尾	6	2	5
	g/ c 结尾	9	4	3

表 4 基于拟氨基酸基因编码方法下 F/10 木聚糖酶密码子偏好性
Tab.4 Codon bias of F/10 and G/11 xylanase based on quasi amino acids coding

木聚糖酶	项目	高	中	低
F/10 家族	密码子	tgg, tac, tat, cgc, cag, agc	ttc, ttt, ggc, acc, tcc, ctg, ctt, gga	atc, ccg, aac, ctc, att, gaa, glg, gcc, gat, gtc, gtt, gca
	a/u 结尾	1	3	5
	g/c 结尾	5	5	7
G/11 家族	密码子	tac, cag, agc, ttc, aac, ggc, cgttgc, gct, cgc, acc, tat, atc, ggt	tcc, tct, ctt, ctc, atg	agt, gat, gtc, gtt, cct, gac, ttt, tgg, gcc, ctg
	a/u 结尾	4	2	5
	g/c 结尾	10	3	3

2.1 基于氨基酸编码方法下的密码子偏好情况分析

由表 1 所列出的基于氨基酸编码方法下 F/10 木聚糖酶家族和 G/11 木聚糖酶家族的同义密码子的相对使用情况可以看出:

1) 在 F/10 木聚糖酶家族中除了{R, D, E, Stop} 家族偏好使用 a/t 结尾的密码子外,所有的氨基酸均偏好使用以 g/c 结尾的密码子。而在 G/11 木聚糖酶家族中除了以 g/c 结尾的密码子偏好使用外,有 6 种氨基酸偏好使用以 t 结尾的密码子。

2) F/10 与 G/11 木聚糖酶家族中,在由两个密码子组成的氨基酸中,其中{a, t, c}{a}{c}所含密码子数量远大于{a, t, c}{a}{t}; {g}{a}{t}所含密码子数量远大于{g}{a}{c}; {a, c}{a}{g}所含密码子数量远大于{a, c}{a}{a}; {g}{a}{a}所含密码子数量远大于{g}{a}{g}。

由表 3 所列出的氨基酸编码方法下 F/10 与 G/11 木聚糖酶家族的密码子的偏好性还可以得出如下结论:

1) 两个家族都偏好使用以 g/c 结尾的密码子,与以前的研究结果一致。F/10 家族中除了 ctg, gga 外高度偏好性的密码子都包含在 G/11 中。这说明 F/10 相对集中在几种主要的密码子,而 G/11 密码子偏好性种类更加多样,这与分子进化的结果相互印证。

2) F/10 家族 17 个主要偏好性密码子(高度、中度)中有 12 个以 g/c 结尾,只有 5 个以 a/t 结尾,显示了氨基酸偏好使用以 g/c 结尾的密码子。G/11 家族 21 个主要偏好性密码子(高度、中度)中有 13 个以 g/c 结尾,只有 8 个以 a/t 结尾,也显示了密码子偏好使用以 g/c 结尾的密码子。当然,这种效

果还不是太好。

3) 虽然高度偏好的密码子 g/c 结尾的大于 a/t 结尾的,但随着密码子偏好程度降低,以 a/t 结尾的密码子数目与以 g/c 结尾的密码子数目趋于接近,而且以上结果与[1]中结果一致。

2.2 基于拟氨基酸基因编码方法下的密码子偏好情况分析

由表 2 所列出的拟氨基酸基因编码方法下 F/10 与 G/11 木聚糖酶家族的密码子偏好性可以看出:

1) F/10 和 G/11 木聚糖酶家族在{G, RS, S', -Y, T, KN, QH, LF, MI, V, L', WC} 这些拟氨基酸家族中都偏好使用以 g/c 结尾的密码子。拟氨基酸{A, R', P} 在 F/10 木聚糖酶家族中偏好使用以 g/c 结尾的密码子,而在 G/11 木聚糖酶家族偏好使用以 t 结尾的密码子,以 g/c 结尾的密码子次之。

2) F/10 木聚糖酶家族中除了{ED} 外,所有拟氨基酸都偏好使用以 g/c 结尾的密码子, G/11 木聚糖酶家族中除了{R', ED, A, P} 外,所有拟氨基酸都偏好使用以 g/c 结尾的密码子。

3) F/10 与 G/11 木聚糖酶家族中,当密码子的第一位碱基为 a 时,则拟氨基酸偏好使用以 c 结尾的密码子。

4) F/10 木聚糖酶家族中,除了第二位碱基为 g 外,当第一位碱基为 t 时,拟氨基酸偏好使用以 c 结尾的密码子。而 G/11 木聚糖酶家族中,当第一位碱基为 t 时,所有的拟氨基酸都偏好使用以 c 结尾的密码子。

5) F/10 与 G/11 木聚糖酶家族中,当第一位碱基为 g 时,除了第二位碱基为 t 的密码子外,其它的都避免使用以 g 结尾的密码子。

由表 4 所列出的拟氨基酸基因编码方法下 F/10 与 G/11 木聚糖酶家族的密码子偏好性还可以得出如下结论:

- 1) 两个家族都偏好使用以 g/c 结尾的密码子, 这与以往研究结果一致。F/10 家族中除了 tgg 外高度偏好性的密码子都包含在 G/11 中。说明 F/10 相对集中在几种主要的密码子, 而 G/11 密码子偏好性种类更加多样, 这与分子进化的结果相互印证。
- 2) F/10 家族 14 个主要偏好性密码子中(高度、中度)有 10 个以 g/c 结尾, 只有 4 个以 a/t 结尾, 显示了拟氨基酸基因偏好使用以 g/c 结尾的密码子。G/11 家族 19 个主要偏好性密码子中(高度、

中度)有 13 个以 g/c 结尾, 只有 6 个以 a/t 结尾, 也显示了拟氨基酸偏好使用以 g/c 结尾的密码子, 且该结果比在氨基酸编码方法下效果还要好;

- 3) 虽然高度偏好的密码子 g/c 结尾的大于 a/t 结尾的, 但随着密码子偏好程度降低, 以 a/t 结尾的密码子数目与以 g/c 结尾的密码子数目趋于接近, 而且以上结果也与[1]中结果一致。

3 结 语

首先, 根据上面的研究, 给出基于氨基酸编码方法与基于拟氨基酸基因编码方法下的密码子偏好情况对比见表 5。

表 5 两种编码方法下密码子偏好性对比
Tab.5 Codon bias comparison of two kinds of coding method

碱基偏好	氨基酸编码		拟氨基酸基因编码	
	F/10 木聚糖酶	G/11 木聚糖酶	F/10 木聚糖酶	G/11 木聚糖酶
c	L, F, A, T, H, S, G, C, N, Y	V, I, F, H, N, T, S, G, Y	G, RS, A, S; - Y, T, KN, R; LF, MI	G, RS, LF, - Y, T, S; KN, WC, V, L; MI
g	K, V, L, P, M, W, Q	K, Q, M, W	P, WC, V, L; QH	QH
t	R, D, Stop	L, A, P, C, R, D, Stop		ED, A, P, R'
a	E	E	ED	

- 1) 基于氨基酸编码方法下 F/10 家族有{D, E, Stop} 3 种氨基酸偏好使用以 a/t 结尾的密码子, 其余的都偏好使用 g/c 结尾的密码子。基于拟氨基酸基因编码方法下 F/10 家族只有{ED} 偏好使用以 a/t 结尾的密码子, 其余的都偏好使用 g/c 结尾的密码子;

- 2) 基于氨基酸编码方法下 G/11 家族有{L, A, P, R, C, D, E, Stop} 八种氨基酸偏好使用以 a/t 结尾的密码子外, 其余的都偏好使用 g/c 结尾的密码子。基于拟氨基酸基因编码方法下 G/11 家族中除了{R', ED, A, P} 四种拟氨基酸偏好使用以 a/t 结尾的密码子外, 其余的都偏好使用 g/c 结尾的密码子;

- 3) 两种编码方法都体现了氨基酸更偏好使用以 g/c 结尾的密码子, 而且以 c 结尾的密码子比以 g 结尾的密码子更偏好使用, 且两种编码方法下 F/10 木聚糖酶的密码子偏好性都优于 G/11 木聚糖酶的密码子偏好性;

- 4) 由表 5 我们还可以得出基于拟氨基酸基因编码方法比基于氨基酸编码方法能更好的体现密码子偏好性, 与文献[8]的结论一致。

同义密码子的使用并非随机的均匀的, 其具有一定的偏好性, 作者通过计算 RSCU 与 QRSCU 分析了 F/10 及 G/11 木聚糖酶家族密码子偏好性, 结果发现:

- 1) 密码子偏好性与生物进化密切相关。由表 3 与表 4 可知, 两种编码方法下几乎全部的 F/10 高度偏好性密码子都包含在 G/11 中, G/11 比 F/10 家族中偏好性密码子种类多。同样的由表 5 可以看出无论是基于哪种编码方法 F/10 家族都比 G/11 家族更偏好使用以 g/c 结尾的密码子。这些性质均可用文献[2]中分子进化的观点说明, 因为 F/10 在进化时间上相对于 G/11 较长, 所以其偏好性趋于均衡;

- 2) 拟氨基酸基因编码方法更能体现密码子的偏好性。文中的表格显示了基于拟氨基酸基因编码方法能更好的体现密码子偏好性, 与文献[8]的结论一致。该结论充分验证了拟氨基酸分类对于 F/10 和 G/11 木聚糖酶家族在偏好性上是充分显现的, 且与分子进化的观点是一致的, 从而对研究蛋白质结构并预测蛋白质的相互作用的分子机制有着重要帮助。

参考文献(References):

- [1] 刘亮伟, 秦天苍, 翟继, 等. F/10 及 G/11 木聚糖酶家族密码子偏好性分析[J]. 河南农业大学学报, 2008, 42: 223–227.
LIU Liang wei, QIN Tian cang, ZHAI Ji, et al. Analysis on codon bias in F/10 and G/11 xylanase[J]. **Journal of Henan Agricultural University**, 2008, 42: 223–227. (in Chinese)
- [2] 刘亮伟, 秦天苍, 王宝, 等. 木聚糖酶的分子进化[J]. 食品与生物技术学报, 2007, 26: 110–116.
LIU Liang wei, QIN Tian cang, WANG Bao, et al. Molecular evolution of xylanase[J]. **Journal of Food Science and Biotechnology**, 2007, 26: 110–116. (in Chinese)
- [3] 苏玉春, 陈光, 白晶, 等. 木聚糖酶的产生条件优化[J]. 吉林农业大学学报, 2008, 30: 793–796.
SU Yu chun, CHEN Guang, BAI Jing, et al. Production and growth condition of xylanase from *Aspergillus niger*[J]. **Journal of Jilin Agricultural University**, 2008, 30: 793–796. (in Chinese)
- [4] 邵蔚蓝, 毛忠贵, 薛业敏. 极端耐热木聚糖酶基因在大肠杆菌中的高效表达[J]. 食品与发酵工业, 2003, 29: 20–25.
SHAO Wei lan, MAO Zhong gui, XUE Ye min. Expression of xylanase B gene of *thermotoga maritima* in *Escherichia coli*[J]. 2003, 29: 20–25. (in Chinese)
- [5] 刘亮伟, 秦天苍, 刘新育, 等. 热稳定性木聚糖酶结构模拟及分析[J]. 河南农业大学学报, 2007, 41: 304–308.
LIU Liang wei, QIN Tian cang, LIU Xin yu, et al. Modelling and analysis of the structure of thermo stable xylanase[J]. **Journal of Henan Agricultural University**, 2007, 41: 304–308. (in Chinese)
- [6] 管维红, 徐振源, 朱平. 用非线性预测方法研究蛋白质序列的特性(II)[J]. 食品与生物技术学报, 2008, 27(2): 71–75.
GUAN Wei hong, XU Zhen yuan, ZHU Ping. Nonlinear Prediction Analysis of Properties in protein sequences(II)[J]. **Journal of Food Science and Biotechnology**, 2008, 27(1): 103–105. (in Chinese)
- [7] ZHU Ping, TANG Xu qing, XU Zhen yuan. The structure analysis of protein sequence based on the quasi amino acids code[J]. **Chinese physics B**, 2009, 1: 363–367.
- [8] 朱平, 高雷, 徐振源. 基于拟氨基酸编码方法下的同义密码子的偏好性仍与结合强度密切相关[J]. 物理学报, 2009, 6: 714–719.
ZHU Ping, GAO Lei, XU Zhen yuan. The usage degree of synonymous codon is close correlated with the strength of combination based on the quasi amino acid coding[J]. **Acta Physica Sinica**, 2009, 6: 714–719. (in Chinese)
- [9] 石秀凡, 黄京飞, 柳树群, 等. 人类基因同义密码子偏好的特征以及与基因GC含量的关系[J]. 生物化学与生物物理进展, 2002, 29: 411–414.
SHI Xiu fan, HUANG Jing fei, LIU Shu qun, et al. The features of synonymous codon bias and GC content relationship in human gene[J]. **Prog Biochem Biophys**, 2002, 29: 411–414. (in Chinese)
- [10] 杨潇, 胡军, 陈祥贵. 人类蛋白编码基因外显子谱和同义密码子偏好的研究[J]. 西华大学学报: 自然科学版, 2008, 27.
YANG Xiao, HU Jun, CHEN Xiang gui. Study on the relationship between exonic pattern and codon bias of human protein coding gene[J]. **Journal of Xihua University: Natural Science**, 2008, 27(2): 79–82. (in Chinese)
- [11] 顾万君, 马建民, 周童, 等. 不同结构的蛋白编码基因的密码子偏好性研究[J]. 生物物理学报, 2002, 18: 81–86.
GU Wan jun, MA Jian ming, ZHOU Tong, et al. Codon usage in genes coding for protein with different tertiary structures[J]. **Acta Biophysica Sinica**, 2002, 18: 81–86. (in Chinese)
- [12] LIU LW, ZHANG J, CHEN B, et al. Principal component analysis in F/10 and G/11 xylanase[J]. **Biochem Biophys Res Commun**, 2004, 322: 277–280.
- [13] Xie T, Ding D. The relationship between synonymous codon usage and protein structure[J]. **FEBS letters**, 1998, 434: 93–96.
- [14] 管维红, 徐振源, 朱平. 用非线性预测方法研究蛋白质序列的特性(I)[J]. 食品与生物技术学报, 2008, 27(1): 71–75.
GUAN Wei hong, XU Zhen yuan, ZHU Ping. Nonlinear prediction analysis of properties in protein sequences(I)[J]. **Journal of Food Science and Biotechnology**, 2008, 27(1): 71–75. (in Chinese)

(责任编辑:李春丽)