

不同种类柑橘的密码子用法分析

胡桂兵^{1,2}, 张上隆¹, 徐昌杰¹, 林顺权²

(¹浙江大学果树科学研究所, 浙江 杭州 310029; ²华南农业大学园艺生物技术研究所, 广东 广州 510642)

摘要: 采用高频密码子分析法, 对甜橙 *Citrus sinensis*、温州蜜柑 *C. unshiu*、葡萄柚 *C. paradisi* 和柠檬 *C. limon* 等 4 种柑橘的蛋白质编码基因序列 (codon DNA sequence, CDS) 行了分析, 计算出了柑橘同义密码子相对使用频率 (relative frequency of synonymous codon, RFSC), 确定出了 4 种柑橘的高频率密码子, 发现不同种类柑橘偏爱密码子稍有差别。

关键词: 柑橘; 密码子用法; 高频密码子; 密码子偏爱性

中图分类号: S666; Q755 文献标识码: A 文章编号: 1001-411X(2006)01-0013-04

Analysis of Codon Usage Between Different Citrus Species

HU Gui-bing^{1,2}, ZHANG Shang-long¹, XU Chang-jie¹, LIN Shun-quan²

(¹ Fruit Science Institute of Zhejiang University, Hangzhou 310029, China;

² Institute of Biotechnology in Horticultural Plants, South China Agric. Univ., Guangzhou 510642, China)

Abstract: Codon DNA sequence (CDS) of four different citrus species, including sweet orange (*Citrus sinensis*), satsuma mandarin (*C. unshiu*), grapefruit (*C. paradisi*) and lemon (*C. limon*) were used to analyze the relative frequency of synonymous codon (RFSC), and high-frequency codons were selected by high-frequency codon analysis. The results indicated that the codon preference was little different in four citrus species.

Key words: citrus; codon usage; high-frequency codon; codon preference

密码子是生物体内信息传递的基本环节。由于密码子具有简并性, 因此一个氨基酸通常由 2 个以上的密码子编码。密码子的这种简并性特征造成了不同物种密码子使用偏爱性不同, 即编码同一个氨基酸的密码子在不同物种中具有不同的使用频率。这种现象会导致外源基因在宿主细胞中的表达量降低。分析表达外源基因物种的密码子用法, 可为提高外源基因在该物种中的表达提供依据。

密码子用法分析方法主要有高表达优越密码子分析法^[1]和高频密码子分析法^[2]等。高表达优越密码子分析法不仅步骤繁琐, 而且对于目的生物遗传背景要求高, 分析起来非常不便; 而高频密码子分析法克服了高表达优越密码子分析法的缺点, 结果可靠。

柑橘种类繁多, 是我国乃至世界上最重要的水

果之一, 近几年包括转基因在内的柑橘生物技术方面的研究取得了很大进展, 已有多篇关于柑橘衰退病毒外壳蛋白基因、几丁质酶基因、绿色荧光蛋白基因、柞蚕抗菌肽 D 基因、人表皮生长因子基因、酵母 HAL2 基因、Bt 基因以及肝炎疫苗基因等外源基因转入不同种类柑橘的报道^[3-4]。了解不同种类柑橘之间密码子用法是否存在差异, 将更加有助于提高所转化的基因在柑橘中的表达效率, 有鉴于此, 笔者采用较为简便、可靠的高频密码子分析法对主要种类柑橘的密码子用法进行了研究。

1 材料与方法

1.1 不同种类柑橘基因的获取

在 GenBank 中检索甜橙 *Citrus sinensis* (简称

收稿日期: 2005-04-04

作者简介: 胡桂兵 (1969-), 男, 副教授, 在取博士研究生; 通讯作者: 张上隆 (1938-), 男, 教授, E-mail: shlzhang@zju.edu.cn

基金项目: 国家自然科学基金 (30100125); 浙江省自然科学基金 (301291); 广东省科技计划项目 (2002A208020202; 2003A2010202)

CS)、温州蜜柑 *C. unshiu* (简称 CU)、葡萄柚 *C. paradisi* (简称 CP) 和柠檬 *C. limon* (简称 CL) 等主要柑橘种类的蛋白质和核苷酸数据, 采用 Paul 和 E-

lizabeth 的筛选原则^[5] 从中筛选用于密码子分析的蛋白质编码基因序列 (codon DNA sequence, 简称 CDS), 作为不同种类柑橘密码子分析的样本 (表 1)。

表 1 不同种类柑橘密码子分析的基因样本

Tab. 1 The gene samples for analysis of codon usage in different citrus species

柑橘种类 species	基因数 genes	GenBank 中的登录号 accession number in GenBank
甜橙 <i>C. sinensis</i>	66	AB009350、AB009351、AB011794、AB011795、AF000135、AF000136、AF090336、AF092088、AF169241、AF196966、AF240787、AF255013、AF255014、AF269158、AF321533、AF369931、AF433643、AF441124、AJ000081、AJ011095、AJ012550、AJ012551、AJ012696、AJ243739、AJ319761、AJ319762、AJ582678、AJ617703、AJ628975、AJ865015、AJ865589、AJ866184、AJ867222、AY029198、AY040710、AY040711、AY094582、AY098891、AY098894、AY242385、AY243478、AY297793、AY338974、AY338975、AY338976、AY344244、AY344245、AY519363、AY533827、AY581048、AY644699、AY679167、AY679168、U38914、U82973、U82975、U82976、U82977、X66377、Y08618、Z46824、Z46944、Z47793、Z48450、Z70032、Z82983
温州蜜柑 <i>C. unshiu</i>	51	AB004247、AB005023、AB007818、AB007819、AB008100、AB008101、AB011796、AB011797、AB011798、AB016204、AB016809、AB022091、AB022092、AB024274、AB024275、AB024276、AB024277、AB025778、AB027456、AB029401、AB033758、AB036922、AB036923、AB036924、AB037106、AB037975、AB046992、AB056126、AB072343、AB074885、AB075547、AB076608、AB110636、AB110637、AB110639、AB110640、AB110641、AB115767、AF184597、AF184598、AF220218、AF296158、AF315289、AF320905、AF320906、AY026338、AY166796、AY166797、AY166798、AY561840、AY579887
葡萄柚 <i>C. paradisi</i>	32	AF095520、AF095521、AF152246、AF152892、AF283532、AF283533、AF283534、AF283535、AF283536、AF283537、AF364515、AF369930 (含 2 条全长基因)、AF372617、AF378333、AF411120、AF486650、AJ001304、AY028416、AY032975、AY158153、AY160772、AY271308、AY383619、CQ795868、CQ795870、CQ813505、CQ813506、CQ813508、Y12689、Z32632、Z97064
柠檬 <i>C. limon</i>	15	AF073507、AF165939、AF176669、AF184068、AF318938、AF514286、AF514287、AF514288、AF514289、AX025477、AX025512、AX025514、AX025516、AY226999、U43338

1.2 不同种类柑橘蛋白质编码基因序列的同义密码子相对使用频率 (relative frequency of synonymous codon, RFSC) 的计算

计算公式为某一密码子在样本中的实际观察数与该密码子对应的氨基酸在样本中的实际观察数的比值。RFSC 反映的是样本中各个同义密码子的使用频率, 它避免了由于样本选择中各基因的长短与氨基酸丰度的影响。

1.3 高频密码子 (high-frequency condon, HF) 的确定

根据 RFSC 的计算结果, 利用高频密码子分析法^[2] 来筛选不同种类柑橘基因的偏爱密码子。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。

2 结果与分析

根据 1.2 所述的计算方法, 对甜橙的 66 个基因共 25 468 个密码子、温州蜜柑的 51 个基因共 21 467

个密码子、葡萄柚的 32 个基因共 12 414 个密码子以及柠檬的 15 个基因共 6 935 个密码子的 RFSC 进行了统计, 结果如表 2 所示。根据 1.3 所述的筛选标准, 共筛选出甜橙的高频密码子 7 个、温州蜜柑的 14 个、葡萄柚的 12 个、柠檬的 12 个 (表 2), 从表 2 及偏爱密码子的分析结果 (表 3) 中不难看出, 不同种类柑橘偏爱的密码子虽有所差别, 但从密码子使用频率整体来看, 种类之间的差别很小, 共性的密码子占绝大多数, 如终止子均偏爱用 TAA 这个密码子、氨基酸 Asp 均偏爱用 GAT 密码子、Arg 均偏爱用 AGG 和 AGA 密码子、Val 均偏爱用 GTT 密码子; 一些氨基酸使用的密码子, 如 Ala 中的 GCT 密码子、Phe 中的 TTT、Gly 中的 GGT、His 中的 CAT、Ile 中的 ATT、Lys 中的 AAG、Asn 中的 AAT、Pro 中的 CCT、Gln 中的 CAA、Thr 中的 ACT 以及 Tyr 中的 TAT, 均是不同种类柑橘偏爱或使用频率较高的密码子。密码子使用频率差异较大的, 仅有 Cys、Glu 和 Leu 等极少数几个氨基酸。

表 2 不同种类柑橘的密码子使用频率¹⁾
Tab.2 The frequency codon usage in different citrus species

氨基酸 amino acid	密码子 codon	甜橙 <i>C. sinensis</i>			温州蜜柑 <i>C. unshiu</i>			葡萄柚 <i>C. paradisi</i>			柠檬 <i>C. limon</i>		
		<i>N</i> ²⁾	RFSC	1/1000	<i>N</i> ²⁾	RFSC	1/1000	<i>N</i> ²⁾	RFSC	1/1000	<i>N</i> ²⁾	RFSC	1/1000
STOP	TAG	9	13.64	0.35	8	15.69	0.37	5	15.62	0.40	0	0.00	0
	TAA	33	50.00	1.30	29	56.86	1.35	16	50.00	1.29	8	53.33	1.15
A(Ala)	TGA	24	36.36	0.94	14	27.45	0.65	11	34.38	0.89	7	46.67	1.01
	GCG	218	11.03	8.56	153	9.54	7.13	65	7.82	5.24	55	10.64	7.93
	GCA	543	27.48	21.32	461	28.74	21.47	278	33.45	22.39	175	33.85	25.23
C(Cys)	GCT	766	38.77	30.08	707	44.08	32.93	344	41.40	27.71	193	37.33	27.83
	GCC	449	22.72	17.63	283	17.64	13.18	144	17.33	11.60	94	18.18	13.55
	TGT	203	45.82	7.97	165	52.55	7.69	116	57.71	9.34	46	46.00	6.63
D(Asp)	TGC	240	54.18	9.42	149	47.45	6.94	85	42.29	6.85	54	54.00	7.79
	GAT	807	62.17	31.69	845	70.83	39.36	490	71.53	39.47	279	69.75	40.23
E(Glu)	GAC	491	37.83	19.28	348	29.17	16.21	195	28.47	15.71	121	30.25	17.45
	GAA	663	45.60	26.03	640	46.82	29.81	438	53.28	35.28	236	52.21	34.03
F(Phe)	GAG	791	54.40	31.06	727	53.18	33.87	384	46.72	30.93	216	47.79	31.15
	TTT	589	50.60	23.13	525	55.56	24.46	330	59.67	26.58	172	57.52	24.80
G(Gly)	TTC	575	49.40	22.58	420	44.44	19.56	223	40.32	17.96	127	42.47	18.31
	GGT	507	28.14	19.91	504	31.42	23.48	260	32.74	20.94	114	25.79	16.44
	GGG	363	20.14	14.25	269	16.77	12.53	133	16.75	10.71	81	18.33	11.68
H(His)	GGA	447	24.81	17.55	494	30.80	23.01	238	29.97	19.17	147	33.26	21.20
	GGC	485	26.91	19.04	337	21.01	15.70	163	20.53	13.13	100	22.62	14.42
	CAT	306	53.78	12.02	303	60.48	14.11	161	54.76	12.97	116	61.38	16.73
I(Ile)	CAC	263	46.22	10.33	198	39.52	9.22	133	45.24	10.71	73	38.62	10.53
	ATA	299	21.68	11.74	279	22.81	13.00	175	24.65	14.10	91	21.98	13.12
K(Lys)	ATT	631	45.76	24.78	613	50.12	28.56	361	50.84	29.08	203	49.03	29.27
	ATC	449	32.56	17.63	331	27.06	15.42	174	24.51	14.02	120	28.99	17.30
	AAA	637	42.90	25.01	523	44.82	24.36	397	47.20	31.98	206	47.80	29.70
L(Leu)	AAG	848	57.10	33.30	644	55.18	30.00	444	52.79	35.77	225	52.20	32.44
	CTA	205	8.30	8.05	180	8.42	8.38	128	10.00	10.31	57	8.41	8.22
	TTA	331	13.40	13.00	225	10.52	10.48	189	14.76	15.22	87	12.83	12.55
M(Met)	TTG	565	22.87	22.18	555	25.95	25.85	288	22.50	23.20	189	27.88	27.25
	CTT	631	25.55	24.78	609	28.47	28.37	329	25.70	26.50	161	23.75	23.22
	CTC	404	16.36	15.86	285	13.32	13.28	170	13.28	13.69	86	12.68	12.40
N(Asn)	CTG	334	13.52	13.11	285	13.32	13.28	176	13.75	14.18	98	14.45	14.13
	ATG	647	100.00	25.40	567	100.00	26.41	310	100.00	24.97	172	100.00	24.80
	AAT	603	51.06	23.68	526	61.02	24.50	321	60.91	25.86	195	62.90	28.12
P(Pro)	AAC	578	48.94	22.70	336	38.98	15.65	206	39.09	16.59	115	37.10	16.58
	CCG	190	14.91	7.46	117	11.39	5.45	55	9.34	4.43	38	12.58	5.48
	CCA	390	30.61	15.31	333	32.42	15.51	206	34.97	16.59	88	29.14	12.69
Q(Gln)	CCT	416	32.65	16.33	401	39.04	18.68	227	38.54	18.29	124	41.06	17.88
	CCC	278	21.82	10.92	176	17.14	8.20	101	17.15	8.14	52	17.22	7.50
	CAA	489	54.09	19.20	339	50.45	15.79	228	55.61	18.37	153	64.02	22.06
R(Arg)	CAG	415	45.91	16.29	333	49.55	15.51	182	44.39	14.66	86	35.98	12.40
	CGT	141	11.26	5.54	202	17.99	9.41	74	12.44	5.96	55	17.57	7.93
	CGA	119	9.50	4.67	131	11.66	6.10	71	11.93	5.72	30	9.58	4.33
S(Ser)	CGC	129	10.30	5.07	86	7.66	4.01	55	9.24	4.43	20	6.39	2.88
	AGG	374	29.87	14.69	290	25.82	13.51	152	25.55	12.24	87	27.80	12.55
	AGA	383	30.59	15.04	304	27.07	14.16	202	33.95	16.27	95	30.35	13.70
T(Thr)	CGG	106	8.47	4.16	110	9.80	5.12	41	6.89	3.30	26	8.31	3.75
	TCG	229	12.15	8.99	153	9.43	7.13	59	6.84	4.75	44	9.18	6.34
	TCA	423	22.45	16.61	394	24.28	18.35	228	26.45	18.37	129	26.93	18.60
V(Val)	ACT	248	13.16	9.74	252	15.53	11.74	143	16.59	11.52	84	17.54	12.11
	TCT	409	21.71	16.06	420	25.88	19.56	228	26.45	18.37	97	20.25	13.99
	TCC	257	13.64	10.09	175	10.78	8.15	92	10.67	7.41	57	11.90	8.22
W(Trp)	AGC	318	16.88	12.49	229	14.11	10.67	112	12.99	9.02	68	14.20	9.81
	ACG	197	14.34	7.74	96	9.50	4.47	69	11.08	5.56	40	11.02	5.77
	ACT	485	35.30	19.04	384	37.98	17.89	231	37.08	18.61	126	34.71	18.17
Y(Tyr)	ACA	395	28.75	15.51	302	29.87	14.07	206	33.06	16.59	115	31.68	16.58
	ACC	297	21.61	11.66	229	22.65	10.67	117	18.78	9.42	82	22.59	11.82
	CTA	201	11.57	7.89	208	14.30	9.69	127	14.14	10.23	65	13.40	9.37
Z(Cys)	GTT	710	40.85	27.88	608	41.82	28.32	398	44.32	32.06	213	43.92	30.71
	CTC	308	17.72	12.09	237	16.30	11.04	146	16.26	11.76	79	16.29	11.39
	GTG	519	29.86	20.38	401	27.58	18.68	227	25.28	18.29	128	26.39	18.46
Y(Tyr)	TGG	359	100.00	14.10	270	100.00	12.58	149	100.00	12.00	98	100.00	14.13
	TAT	380	50.73	14.92	447	59.60	20.82	244	59.80	19.66	145	61.18	20.91
TAC	TAC	369	49.27	14.49	303	40.40	14.11	164	40.20	13.21	92	38.82	13.27

1)带“□”符号的为高频密码子;2)*N*为每种密码子在所统计的所有基因中出现的次数

表3 不同种类柑橘偏爱的密码子

Tab.3 The preference codon in different citrus species

氨基酸 amino acid	甜橙 <i>C. sinensis</i>	温州蜜柑 <i>C. unshiu</i>	葡萄柚 <i>C. paradisi</i>	柠檬 <i>C. limon</i>
A (Ala)	GCT [†]	GCT [†]	GCT [†]	GCT, GCA
C (Cys)	TGC	TGT	TGT	TGC
D (Asp)	GAT [†]	GAT [†]	GAT [†]	GAT [†]
E (Glu)	GAG	GAG	GAA	GAA
F (Phe)	TTT	TTT	TTT	TTT
G (Gly)	GCT, GGC	GGT, GGA	GCT, GCA	GGT, GGA
H (His)	CAT	CAT [†]	CAT	CAT [†]
I (Ile)	ATT	ATT [†]	ATT [†]	ATT
K (Lys)	AAG	AAG	AAG	AAG
L (Leu)	CTT [†]	CTT [†] , TTG [†]	CTT [†]	TTG [†]
M (Met)	ATG	ATG	ATG	ATG
N (Asn)	AAT	AAT [†]	AAT [†]	AAT [†]
P (Pro)	CCT, CCA	CCT [†]	CCT [†]	CCT [†]
Q (Gln)	CAA	CAA	CAA	CAA [†]
R (Arg)	AGG [†] , AGA [†]	AGG [†] , AGA [†]	AGG [†] , AGA [†]	AGG [†] , AGA [†]
S (Ser)	TCA, TCT	TCT [†]	TCA [†] , TCT [†]	TCA [†]
T (Thr)	ACT, ACA	ACT [†]	ACT, ACA	ACT, ACA
V (Val)	GTT [†]	GTT [†]	GTT [†]	GTT [†]
W (Trp)	TGG	TGG	TGG	TGG
Y (Tyr)	TAT	TAT	TAT	TAT [†]
STOP	TAA [†]	TAA [†]	TAA [†]	TAA [†]

1) 带“[†]”符号的为偏爱密码子,其余为使用频率较高的密码子

3 讨论

传统的密码子用法的计算方法为高表达优越密码子分析法^[1],此法需要先计算每个基因的密码子有效数(effective number of codons,简称为Nc或ENC)和相对同义密码子用法(relative synonymous codon usage,简称RSCU),然后再根据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RSCU值,最后通过t-检验确定出高表达基因的优越密码子。高表达优越密码子分析法不仅计算方法十分繁琐,而且对于目的生物遗传背景要求高,分析起来十分不便。而高频密码子分析法^[2],方法简便、快速,本文的研究结果也再次证明了这一点。

不同种类柑橘偏爱密码子差别很小,这符合它们之间固有的亲缘关系。研究结果提示我们,用不同种类柑橘中任何一种柑橘的偏爱密码子所设计的外源基因在其他柑橘中也可以使用,这种“通用性”可以减少试验的重复性、节约有限的资源,是十分有意义的。

通过密码子改造提高动物、微生物或植物基因在大肠杆菌或酵母中的表达已有不少成功的例子^[6-10]在某些植物中也有一些报道^[11-12],但在柑橘中鲜见报道,因此分析结果对通过密码子优化手段

提高外源基因在不同种类柑橘中的表达效率具有一定的指导意义。当然,由于基因数据库中登录的柑橘基因资源有限,本研究结果可靠性会受到一定的影响;另外,外源基因的表达受多种因素的影响,如DNA修饰水平、翻译后蛋白的定位以及转基因方式等^[13],是否满足密码子的偏爱性只是影响目的基因表达的因素之一,我们不能过高估计密码子优化的效果,提高外源基因表达的效率还有很长的路要走。

参考文献:

- [1] FRANK W. The effective number of codons usage used in a gene[J]. Gene, 1990, 87:23-29.
- [2] 林涛,倪志华,沈明山,等. 高频密码子分析法及其在烟草密码子分析中的应用[J]. 厦门大学学报:自然科学版,2002,41(5):551-554.
- [3] 蒋迪,徐昌杰,陈大明,等. 柑橘转基因研究的现状及展望[J]. 果树学报,2002,19(1):48-52.
- [4] 胡蓉,魏泓,陈善春,等. 甲型肝炎病毒衣壳蛋白融合基因植物表达载体的构建与遗传转化柑橘的研究[J]. 遗传,2004,26(4):425-431.
- [5] PAUL M S, ELIZABETH C. Synonymous codon usage in *Saccharomyces cerevisiae*[J]. Yeast,1991, 7:657-678.
- [6] 李秀兰,戴祝英,张双全,等. 家蚕抗菌肽 CM IV 基因结构改造及表达产物的研究[J]. 中国生物化学与分子生物学报,1999,15(3):387-391.
- [7] VALENTINO S J T, CZIFERSZKY A E, BERGQUIST P L, et al. Codon optimization of xylanase gene *xynB* from the thermophilic bacterium *Dictyoglomus thermophilum* for expression in the filamentous fungus *Trichoderma reesei* [J]. FEMS Microbiology Letters, 2000, 190:13-19.
- [8] 张雅丽,杨国庆,郭英. 密码子优化的牛精蛋白基因在大肠杆菌中的表达[J]. 高技术通讯,2002(4):42-46.
- [9] 韩雪清,刘湘涛,张永国,等. 猪瘟病流行毒株 E2 基因密码子优化及在酵母中的高效表达[J]. 微生物学报,2003,43(5):560-568.
- [10] 张素芳,贾赞,蔡梅红,等. Magainin 和 cecA-mil 抗菌肽基因的密码子优化及在毕赤酵母中的高效表达[J]. 中国生物工程杂志,2004,24(7):93-97.
- [11] 黄其满,毛立群,黄卫红,等. 转人工合成 GFM CryIA 基因烟草表现明显杀虫活性[J]. 植物学报,1998,40(3):228-233.
- [12] PANAHL M, ALLIL Z, CHENG Xiong-ying, et al. Recombinant protein expression plasmids optimized for industrial *E. coli* fermentation and plant systems produced biologically active human insulin-like growth factor-1 in transgenic rice and tobacco plants[J]. Transgenic Research, 2004,13(3):245-259.
- [13] 皮灿辉,易自力,王志成. 提高转基因植物外源基因表达效率的途径[J]. 中国生物工程杂志,2003,23(1):1-4.

【责任编辑 柴 焰】