

草莓 DHAR 基因密码子偏好性分析

刘万达¹, 张丙秀², 魏媛媛², 王禹¹, 董超², 朱延明³

(1. 黑龙江省农业科学院 园艺分院, 黑龙江 哈尔滨 150069; 2. 东北农业大学 园艺学院, 黑龙江 哈尔滨 150030;

3. 东北农业大学 生命科学学院, 黑龙江 哈尔滨 150030)

摘要:以草莓为试材, 运用 CHIPS、CUSP 和 CodonW 程序分析了草莓 DHAR 基因的密码子偏好性, 并与大肠杆菌和酵母的基因组密码子偏好性及 7 种植物的 DHAR 基因密码子偏好性进行了比较, 同时基于 DHAR 基因的密码子使用偏好性进行了系统聚类分析, 以期在作物遗传改良中为草莓 DHAR 基因选择合适的遗传转化受体及进行基因优化提供依据。结果表明: 草莓 DHAR 基因更适合导入苹果等双子叶植物中, 若要提高草莓 DHAR 基因在大肠杆菌或酵母中的表达水平, 还需进行密码子的优化。

关键词:草莓; DHAR; 密码子偏好性; 聚类分析; 基因表达

中图分类号:S 668.4 **文献标识码:**A **文章编号:**1001-0009(2014)14-0092-06

植物组织中的抗坏血酸可为人类提供抗坏血酸源, 同时也是植物体自身代谢必不可少的物质, 抗坏血酸在

第一作者简介:刘万达(1982-), 男, 黑龙江富裕人, 硕士, 助理研究员, 现主要从事寒地果树新品种选育与配套栽培技术的研究及示范推广工作。E-mail: haaslwd@126.com.

责任作者:朱延明(1955-), 男, 博士, 教授, 现主要从事作物耐盐碱与低温分子生物学及基因工程等教学与科研工作。E-mail: ymzhu2001@yahoo.com.cn.

基金项目:农业生物功能基因重点实验室开放课题资助项目(NS-GJ2012-05)。

收稿日期:2014-03-21

植物中作为某些重要酶类的辅因子, 参与抗氧化作用、光合作用、细胞的分裂、生长发育以及信号传递等生理过程^[1]。草莓(*Fragaria ananassa* Duch.) 属蔷薇科(Rosaceae)草莓属(*Fragaria*)多年生草本植物, 其果实香嫩多汁, 除含糖、蛋白质及钙、镁、铁、磷等矿物质较高外, 还含有丰富的抗坏血酸(35~75 mg/100g)。抗坏血酸还原酶(dehydroascorbatereductase, DHAR)在植物抗坏血酸再循环中发挥重要作用。植物细胞内, 抗坏血酸在抗坏血酸氧化酶(AAO)和抗坏血酸过氧化酶(APX)作用下被氧化成单脱氢抗坏血酸(MDHA)和脱氢抗坏血酸(DHA)^[2]。脱氢抗坏血酸还原酶(DHAR)以谷胱甘肽

[8] 姜英林, 冯霄汉. 早熟西洋梨胚培养技术研究[J]. 北方果树, 2012(6): 5-8.

[9] 冯建荣, 吉燕, 樊新民, 等. 早熟蟠桃的胚培养[J]. 石河子大学学报(自然科学版), 2009(1): 23-26.

[10] 潘晓, 陈瑞丹. “淡丰后”梅胚培养影响因素的研究[J]. 北京林业大学学报, 2010(S2): 88-92.

[11] 石文芳, 郝瑞杰, 王史琴, 等. 4 个梅花品种的胚培养及愈伤组织诱导

研究[J]. 中国农学通报, 2012(10): 198-201.

[12] 刘会超, 贾文庆, 王坤. 牡丹胚培养及丛生苗继代培养研究[J]. 北方园艺, 2010(6): 172-174.

[13] 曾端香, 袁涛, 王莲英, 等. 两个牡丹杂交系种子胚培养技术研究[J]. 热带农业科学, 2011(3): 8-12.

[14] 王莹, 何桂梅, 韩丽晓. 紫斑牡丹胚培养及幼苗生长的研究[J]. 湖南农业科学, 2012(9): 103-106.

Research on Cotyledonary Embryo Culture of *Cinnamomum camphora* L.

DU Li

(School of Life Science and Technology, Nanyang Normal University, Nanyang, Henan 473000)

Abstract: The cotyledonary embryo of *C. camphora* as the explants were cultured on MS medium supplemented with the different plant hormones combinations at different concentrations, the effect of plant hormones on the cotyledonary embryo culture was studied. The results showed that the highest germination frequency was 41.7% by using MS+BA 2.0 mg/L+IBA 1.0 mg/L.

Key words: *Cinnamomum camphora* L.; cotyledonary embryo; tissue culture; germination; somatic embryogenesis

(GSH)为底物,催化脱氢抗坏血酸还原为抗坏血酸^[3],使得抗坏血酸得以再生,从而降低了抗坏血酸的从头合成的需求。草莓 DHAR 基因已被克隆出来,这为提高作物抗坏血酸含量的转基因育种以及进一步研究该基因的生理代谢功能提供了基础。由于基因组的碱基组成、蛋白质二级结构和基因表达水平等原因,导致生物基因对同义密码子的使用存在偏好,即使在同一物种中,不同基因的密码子偏好也不同^[4-6]。

现通过在线程序 EMBOSS 及 CodonW,分析了 DHAR 基因的密码子偏好性,并与拟南芥等 7 种植物的 DHAR 基因密码子偏好性进行比较。同时将草莓 DHAR 基因密码子与大肠杆菌及酵母菌的基因组密码子进行了偏好性比较,以期为进一步研究 DHAR 蛋白质的结构和功能奠定基础,并为选择合适的表达系统和基因密码子的优化提供依据和信息。

1 材料与方法

1.1 序列来源和分析软件

密码子分析软件为法国巴斯德研究所生物信息中心(<http://bioweb.pasteur.fr>)的在线软件: CodonW、CHIPS 和 CUSP。草莓 DHAR 基因序列来源于 GenBank(<http://www.ncbi.nlm.nih.gov/genbank>)(登录号:HM045477)。拟南芥、烟草、番茄、葡萄、小麦、水稻等植物的 DHAR 基因的序列来源于 GenBank,见表 1。大肠杆菌和酵母基因组的密码子偏好性数据来源于 Codon Usage data base(<http://www.kazusa.or.jp/codon/>)。

表 1 DHAR 基因的完整编码区序列来源

Table 1 The complete coding sequence of DHAR gene

序列登陆号	物种
GenBank accession No.	Species
AY074785	拟南芥 <i>Arabidopsis thaliana</i>
AY074787.1	烟草 <i>Nicotiana tabacum</i>
DQ322706.2	苹果 <i>Malus sieversii</i> (Led.) Roem
NM-001247893	番茄 <i>Solanum lycopersicum</i>
EU280162.1	葡萄 <i>Vitis vinifera</i>
AY074784.1	小麦 <i>Triticum aestivum</i> Linn
AY074786.1	水稻 <i>Oryza sativa</i>

1.2 试验方法

基因有效密码子数(ENc)的统计,采用软件 CHIPS 进行;基因 CDS 区的 GC 含量、密码子中第 3 位碱基的 GC 含量(GC3s)、相对同义密码子使用度(Relative synonymous codon usage, RSCU)和密码子使用频率,运用 CodonW 和 CUSP 进行计算;DHAR 基因密码子使用偏好性的聚类分析,用 SPSS 软件进行。

2 结果与分析

2.1 密码子偏好性分析

某一特定的密码子在编码对应氨基酸时,相对于同义密码子出现的概率,用同义密码子使用相对概率

表 2 CUSP 和 CodonW 程序分析
草莓 DHAR 基因的密码子偏好性

Table 2 Analysis on codon bias of DHAR gene in

strawberry analyzed using CUSP and CodonW program

密码子 Codon	氨基酸 Amino acid	比例 Fract	频率 Frequency	个数 Number	相对密码子使用度 RSCU
GCA	A	0.167	6.550	3	0.67
GCC	A	0.167	6.550	3	0.67
GCG	A	0.056	2.183	1	0.22
GCT	A	0.611	24.017	11	2.44
TGC	C	0.367	24.017	11	0.73
TGT	C	0.633	41.485	19	1.27
GAC	D	0.400	13.100	6	0.80
GAT	D	0.600	19.651	9	1.20
GAA	E	0.765	28.384	13	1.53
GAG	E	0.235	8.734	4	0.47
TTC	F	0.400	21.834	10	0.80
TTT	F	0.600	32.751	15	1.20
GGA	G	0.250	17.467	8	1.00
GGC	G	0.281	19.651	9	1.12
GGG	G	0.094	6.550	3	0.38
GGT	G	0.375	26.201	12	1.50
CAC	H	0.350	15.284	7	0.70
CAT	H	0.650	28.384	13	1.30
ATA	I	0.100	4.367	2	0.30
ATC	I	0.200	8.734	4	0.60
ATT	I	0.700	30.568	14	2.10
AAA	K	0.647	24.017	11	1.29
AAG	K	0.353	13.100	6	0.71
CTA	L	0.098	8.734	4	0.59
CTC	L	0.122	10.917	5	0.73
CTG	L	0.073	6.550	3	0.44
CTT	L	0.293	26.201	12	1.76
TTA	L	0.146	13.100	6	0.88
TTG	L	0.268	24.017	11	1.61
ATG	M	1.000	15.284	7	1.00
AAC	N	0.438	15.284	7	0.88
AAT	N	0.562	19.651	9	1.12
CCA	P	0.423	24.017	11	1.69
CCC	P	0.115	6.550	3	0.46
CCG	P	0.077	4.367	2	0.31
CCT	P	0.385	21.834	10	1.54
CAA	Q	0.529	19.651	9	1.06
CAG	Q	0.471	17.467	8	0.94
AGA	R	0.500	13.100	6	3.00
AGG	R	0.167	4.367	2	1.00
CGA	R	0.083	2.183	1	0.50
CGC	R	0.167	4.367	2	1.00
CGG	R	0.083	2.183	1	0.50
CGT	R	0.000	0.000	0	0.00
AGC	S	0.071	6.550	3	0.43
AGT	S	0.095	8.734	4	0.57
TCA	S	0.167	15.284	7	1.00
TCC	S	0.190	17.467	8	1.14
TCG	S	0.024	2.183	1	0.14
TCT	S	0.452	41.485	19	2.71
ACA	T	0.238	10.917	5	0.95
ACC	T	0.286	13.100	6	1.14
ACG	T	0.000	0.000	0	0.00
ACT	T	0.476	21.834	10	1.90
GTA	V	0.250	15.284	7	1.00
GTC	V	0.179	10.917	5	0.71
GTG	V	0.214	13.100	6	0.86
GTT	V	0.357	21.834	10	1.43
TGG	W	1.000	17.467	8	1.00
TAC	Y	0.474	19.651	9	0.95
TAT	Y	0.526	21.834	10	1.05
TAA	*	0.111	6.550	3	0.33
TAG	*	0.111	6.550	3	0.33
TGA	*	0.778	45.852	21	2.33

注:大于 1 的 RSCU 值用下划线表示。

(RSCU)表示,因为 RSCU 计算简便、去除了氨基酸组成对密码子使用的影响,所以能更直观地反映出密码子使用的偏好性程度^[7]。当某一密码子的 RSCU 值等于 1 时,说明密码子的使用没有偏好。RSCU 值大于 1,则表明该密码子的使用频率相对较高。各个密码子在编码对应氨基酸的密码子中所占的比例以 Fract 表示,各个比例相加总和为 1。结果表明,草莓 *DHAR* 基因的终止密码子使用 TAA、TAG、TGA、GCT、TGT、GAT、GAA、TTT 等 24 个以 A 或 T 碱基结尾的密码子的 RSCU 值均大于 1,且 Fract 值也较大,为偏好密码子;大多数以 G 或 C 碱基结尾的密码子的 RSCU 值和 Fract 值均较低(表 2)。

2.2 不同植物 *DHAR* 基因密码子偏好性的比较

2.2.1 草莓有效密码子数(ENc)和 GC 含量分析

ENc 值(Effective number of codons)代表着特定基因中同义密码子非均衡使用的偏好程度,是对基因密码子偏好性程度的客观评价标准。ENc 值在 20~61 之间,ENc 值越接近 20,说明密码子偏好性越强。该研究的计算表明 *DHAR* 基因的 ENc 值为 50.279,可见 ENc 值偏大,说明草莓 *DHAR* 基因各密码子在编码氨基酸时出现的频率比较均衡;草莓 *DHAR* 基因编码区 GC 含量较低,为 0.419,GC3s 值则更低,为 0.358,表明在草莓 *DHAR* 整个编码区序列中 A+T 含量大于 G+C,且偏好使用以 A、T 结尾的密码子。

2.2.2 *DHAR* 基因的 ENc、RSCU 及 GC 含量分析

基因的表达水平可根据 ENc 值预测,一般来说,ENc 值越小,表明该基因的表达水平越高,ENc 值小于 30 的可预

测为高表达基因,大于 55 的可预测为低表达基因^[7]。从表 3 可以看出,拟南芥、烟草、苹果、番茄、葡萄在密码子使用上与草莓相似,而单子叶植物水稻、小麦 ENc 值为 56.133 和 55.480。从计算数据来看,拟南芥、草莓、烟草、苹果、番茄、葡萄等双子叶植物 GC 值均低于 0.5,而单子叶植物的 GC 值均高于 0.5,预示着双子叶植物的 *DHAR* 基因对 A、T 具有一定的偏好性,单子叶植物则对 G、C 有一定的偏好性。

表 3 8 个物种 *DHAR* 基因的 ENc 值、GC 值和 GC3s 含量比较

Table 3 Comparison of ENc, GC and GC3s value of *DHAR* gene in 8 species

物种 Species	ENc 值 ENc value	GC 值 GC value	GC3s 含量 GC3s content
草莓 <i>Fragaria ananassa</i>	50.279	0.419	0.358
拟南芥 <i>Arabidopsis thaliana</i>	49.995	0.434	0.487
烟草 <i>Nicotiana tabacum</i>	51.431	0.418	0.378
苹果 <i>Malus sieversii</i> (Led.) Roem	50.596	0.458	0.395
番茄 <i>Solanum lycopersicum</i>	50.151	0.451	0.436
葡萄 <i>Vitis vinifera</i>	50.189	0.446	0.357
小麦 <i>Triticum aestivum</i> Linn	55.480	0.573	0.585
水稻 <i>Oryza sativa</i>	56.133	0.503	0.580

表 4 列出了各物种 *DHAR* 基因的 RSCU,根据 RSCU 可以了解这几个物种 *DHAR* 基因密码子的具体情况。由表 4 可知,这 8 个物种中,草莓的 *DHAR* 基因的密码子无论与单子叶植物还是双子叶植物,都存在类似的使用情况:双子叶植物拟南芥有 33 个密码子的 RSCU 值大于等于 1,烟草有 33 个、苹果有 35 个、番茄有 34 个、葡萄有 34 个;以 A 或 T 碱基结尾 RSCU 值

表 4 物种 *DHAR* 基因的 RSCU

Table 4 RSCU of *DHAR* gene in the species

密码子 Codon	草莓 <i>Fragaria ananassa</i>	拟南芥 <i>Arabidopsis thaliana</i>	烟草 <i>Nicotiana tabacum</i>	苹果 <i>Malus sieversii</i> (Led.) Roem	番茄 <i>Solanum lycopersicum</i>	葡萄 <i>Vitis vinifera</i>	小麦 <i>Triticum aestivum</i> Linn	水稻 <i>Oryza sativa</i>
GCA	0.67	2.00	1.54	0.67	0.31	0.73	1.11	1.54
GCC	0.67	0.50	0.62	1.67	0.62	1.45	1.11	0.92
GCG	0.22	1.00	0.92	0.00	0.00	0.00	1.56	0.15
GCT	2.44	0.50	0.92	1.67	3.08	1.82	0.22	1.38
TGC	0.73	1.00	0.75	1.06	0.18	0.80	1.20	1.16
TGT	1.27	1.00	1.25	0.94	1.82	1.20	0.80	0.84
GAC	0.80	0.29	0.57	1.00	0.92	0.86	1.00	0.40
GAT	1.20	1.71	1.43	1.00	1.08	1.14	1.00	1.60
GAA	1.53	0.00	1.33	1.87	1.09	1.69	0.50	1.62
GAG	0.47	2.00	0.67	0.13	0.91	0.31	1.50	0.38
TTC	0.80	1.33	1.07	0.80	1.00	0.38	0.00	1.00
TTT	1.20	0.67	0.93	1.20	1.00	1.62	0.00	1.00
GGA	1.00	0.57	1.14	1.20	1.20	1.66	1.09	1.53
GGC	1.12	0.00	2.10	1.07	0.40	0.83	0.00	0.94
GGG	0.38	1.71	0.38	0.67	1.60	0.28	1.45	0.47
GGT	1.50	1.71	0.38	1.07	0.80	1.24	1.45	1.06
CAC	0.70	2.00	0.67	1.00	1.00	0.50	1.20	1.08
CAT	1.30	0.00	1.33	1.00	1.00	0.50	0.80	0.92
ATA	0.30	0.92	1.12	0.79	0.63	0.33	1.29	0.75
ATC	0.60	0.92	0.56	0.79	1.26	1.00	1.29	1.88

续表 4

密码子	草莓	拟南芥	烟草	苹果	番茄	葡萄	小麦	水稻
Codon	<i>Fragaria ananassa</i>	<i>Arabidopsis thaliana</i>	<i>Nicotiana tabacum</i>	<i>Malus sieversii</i> (Led.) Roem	<i>Solanum lycopersicum</i>	<i>Vitis vinifera</i>	<i>Triticum aestivum</i> Linn	<i>Oryza sativa</i>
ATT	2.10	1.15	1.31	1.42	1.11	1.67	0.43	0.38
AAA	1.29	1.17	1.52	1.60	0.97	1.25	1.27	2.00
AAG	0.71	0.83	0.48	0.40	1.03	0.75	0.73	0.00
CTA	0.59	0.35	0.00	0.57	0.46	0.67	0.00	1.56
CTC	0.73	0.71	1.36	1.14	1.38	1.33	1.11	0.67
CTG	0.44	1.18	0.82	0.86	0.69	1.33	2.22	1.33
CTT	1.76	1.18	1.36	1.43	1.38	1.56	0.89	1.11
TTA	0.88	0.47	1.36	1.14	0.46	0.89	0.00	0.44
TTG	1.61	2.12	1.09	0.86	1.62	0.22	1.78	0.89
ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AAC	0.88	0.57	0.44	1.33	0.60	0.88	0.00	0.80
AAT	1.12	1.43	1.56	0.67	1.40	1.12	2.00	1.20
CCA	1.69	1.82	1.67	1.33	1.33	1.69	0.65	1.12
CCC	0.46	0.73	1.67	1.33	1.11	0.46	0.97	1.28
CCG	0.31	1.09	0.33	0.00	0.22	0.31	1.51	0.48
CCT	1.54	0.36	0.33	1.33	1.33	1.54	0.86	1.12
CAA	1.06	0.80	1.86	1.83	0.57	1.06	1.20	1.64
CAG	0.94	1.20	0.14	0.17	1.43	0.94	0.80	0.36
AGA	3.00	3.26	1.91	1.04	3.00	3.00	1.38	1.16
AGG	1.00	1.37	0.82	0.00	0.60	1.00	2.12	0.19
CGA	0.50	0.34	1.36	1.83	0.60	0.50	0.88	1.35
CGC	1.00	0.00	0.27	1.30	0.00	1.00	0.62	0.97
CGG	0.50	0.34	1.09	1.30	0.00	0.50	0.38	0.77
CGT	0.00	0.69	0.55	0.52	1.80	1.00	0.62	1.55
AGC	0.43	1.33	0.00	0.51	0.86	0.50	0.81	0.86
AGT	0.57	1.00	1.24	0.69	1.54	0.57	0.23	1.03
TCA	1.00	0.33	0.62	0.51	1.20	1.00	1.62	0.69
TCC	1.14	0.33	1.86	1.54	0.00	1.14	0.81	1.03
TCG	0.14	0.67	0.21	0.17	1.03	0.14	1.38	0.51
TCT	2.71	2.33	2.07	2.57	1.37	2.00	1.15	1.89
ACA	0.95	1.25	0.57	0.80	0.40	1.33	1.14	0.50
ACC	1.14	0.25	1.71	2.00	0.80	1.33	1.71	1.50
ACG	0.00	1.25	0.00	0.00	0.00	0.44	0.57	0.50
ACT	1.90	1.25	1.71	1.20	2.80	0.89	0.57	1.50
GTA	1.00	0.24	1.00	0.80	0.48	0.36	0.00	0.84
GTC	0.71	0.94	0.67	0.40	0.97	2.18	1.07	0.42
GTG	0.86	1.41	0.33	0.80	1.33	0.36	2.40	0.63
GTT	1.43	1.41	2.00	2.00	1.21	1.09	0.53	2.11
TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
TAC	0.95	0.00	1.20	1.60	1.20	0.80	0.00	0.67
TAT	1.05	2.00	0.80	0.40	0.80	1.20	0.00	1.33
TAA	0.33	0.41	0.62	0.67	2.40	1.29	0.20	0.43
TAG	0.33	0.95	0.75	0.67	0.30	0.00	0.40	0.43
TGA	2.33	1.64	1.62	1.67	0.30	1.71	2.40	2.14

大于等于 1 的密码子,拟南芥有 17 个,烟草有 22 个、苹果有 20 个、番茄有 20 个、葡萄有 23 个。单子叶植物小麦和水稻的密码子中,RSCU 值大于等于 1 的分别有 32 和 33 个,其中以 A 或 T 碱基结尾的分别有 13 个和 23 个。

2.2.3 基于密码子使用偏性的系统聚类 依据表 4 中各密码子的相对使用度(RSCU),计算出各物种之间的欧氏平方距离系数。根据欧式平方距离系数进一步分析几个物种 DHAR 基因密码子使用差异性大小,物种间的欧式平方距离系数越大,则表示 2 个物种在密码子

上的使用差异越大。由表 5 可知,小麦和水稻距离较近,这 2 个单子叶植物与其它 6 个双子叶植物距离均较远。各双子叶植物之间距离均大于水稻和小麦之间的距离。可见这种密码子使用偏性的分析结果与传统的植物系统分类是一致的:小麦和水稻均属于单子叶禾本科,而其它双子叶植物却属于不同的科。根据各物种 DHAR 基因密码子使用偏好性的欧式平方距离系数进行聚类分析。在聚类分析过程中,以每个物种的基因作为一个对象,将密码子的相对使用度作为变量。采用 59 个密码子的 RSCU 值对其密码子使用偏好性进行分析。

基因间的距离规定为基因相对密码子使用度的欧式平方距离。例如,对于基因 a 与基因 b,其密码子使用距离 D 的计算公式如下:

$$D_{ab} = \sqrt{\sum_{i=1}^{59} (RSCU_{ai} - RSCU_{bi})^2}.$$

为了使同样样品之间的离差平方和最小,而类与类之间的离差平方和最大,采用了离差平方和法计算类与

表 5 物种 DHAR 基因间密码子使用偏性的欧式平方距离系数

Table 5 Coefficient of absolute squared euclidean distance of codon usage bias among different DHAR genes

物种 Species	草莓 <i>Fragaria ananassa</i>	拟南芥 <i>Arabidopsis thaliana</i>	烟草 <i>Nicotiana tabacum</i>	苹果 <i>Malus sieversii</i> (Led.) Roem	番茄 <i>Solanum lycopersicum</i>	葡萄 <i>Vitis vinifera</i>	小麦 <i>Triticum aestivum</i> Linn	水稻 <i>Oryza sativa</i>
草莓 <i>Fragaria ananassa</i>								
拟南芥 <i>Arabidopsis thaliana</i>	34.100							
烟草 <i>Nicotiana tabacum</i>	20.391	40.659						
苹果 <i>Malus sieversii</i> (Led.) Roem	18.398	53.586	14.741					
番茄 <i>Solanum lycopersicum</i>	28.219	40.537	39.889	41.476				
葡萄 <i>Vitis vinifera</i>	12.730	39.019	25.755	21.162	28.534			
小麦 <i>Triticum aestivum</i> Linn	43.964	30.641	42.653	47.448	52.219	43.101		
水稻 <i>Oryza sativa</i>	23.447	37.696	17.948	13.520	37.077	23.351	1.791	

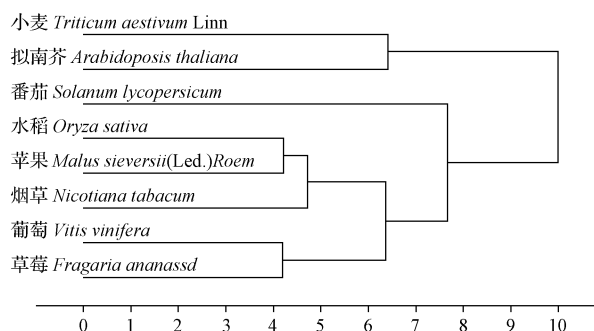


图 1 基于 DHAR 基因 RSCU 的聚类树状图

Fig. 1 Cluster analysis dendrogram of RSCU values of different DHAR genes

2.3 与大肠杆菌、酵母菌的基因组密码子偏好性比较

大肠杆菌作为原核表达系统,酵母作为真核表达系统,二者广泛用于基因表达的研究。为实现草莓 DHAR 基因在这 2 类表达系统中的高效表达,进行了密码子偏好性比较。该试验中衡量物种间密码子使用偏性差异的指标是密码子出现频率的比值。表 5 列出了草莓 DHAR 基因、大肠杆菌和酵母菌基因组中不同密码子出现的频率(以 1/1 000 表示)以及其比值。如果比值在 0.5~2.0 之间,表示二者的密码子偏好性较为接近;比值小于或等于 0.5、大于或等于 2.0 表示偏好性差异较大^[10]。结果表明,草莓 DHAR 基因与大肠杆菌的基因组中不同密码子使用频率值差异较大的有 13 个,与酵母基因组密码子使用频率差值较大的有 23 个。说明酵母表达系统优于大肠杆菌表达系统。但若若要使草莓 DHAR 基因在大肠杆菌或酵母中得到较高水平的表达,还需对这些密码子进行优化。

类间的距离。聚类分析结果见图 1。密码子使用频率的聚类树状图常可用于推测不同物种间、不同基因间在密码子使用方面的相似度或亲缘性^[8]。亲缘关系比较近的物种常表现出相似的密码子使用频率,所以根据密码子使用频率的聚类树状图也可推测不同物种之间的进化关系^[9]。从图 1 可以看出,拟南芥和小麦聚为一类,其它植物聚为一类。

表 6 草莓 DHAR 基因与大肠杆菌基因组和酵母基因组的密码子出现频率

Table 6 Codon frequency of strawberry DHAR gene,

E. coli genome and yeast genome

密码子 Codon	氨基酸 Amino acid	DHAR	大肠杆菌 基因组 <i>E. coli</i>	酵母基 因组 Yeast	DHAR/大 肠杆菌基因组 DHAR/ <i>E. coli</i>	DHAR/酵 母基因组 DHAR/Yeast
GCT	A	24.017	15.4	10.1	1.56	2.38
TGC	C	24.017	6.4	10.1	3.75	2.38
TGT	C	41.485	5.2	6.7	7.98	6.19
GAC	D	13.100	19.2	57.2	0.68	0.23
GAT	D	19.651	32.8	13.5	0.60	1.46
GAA	E	28.384	39.3	13.5	0.72	2.10
GAG	E	8.734	18.7	33.7	0.47	0.26
TTC	F	21.834	15.9	20.2	1.37	1.08
TTT	F	32.751	22.2	13.5	1.48	2.43
GGA	G	17.467	8.9	3.4	2.00	5.14
GGC	G	19.651	28.1	23.6	0.70	0.83
GGG	G	6.550	11.8	3.4	0.56	1.93
GGT	G	26.201	24.2	20.2	1.08	1.30
CAC	H	15.284	9.4	33.7	1.63	0.45
CAT	H	28.384	12.8	3.4	2.21	8.35
ATA	I	4.367	5.5	3.4	0.79	1.28
ATC	I	8.734	23.9	60.6	0.37	0.14
ATT	I	30.568	29.7	13.5	1.03	2.26
AAA	K	24.017	34.0	13.5	0.71	1.78
AAG	K	13.100	11.0	40.4	1.19	0.32
CTA	L	8.734	3.9	6.7	2.24	1.30
CTC	L	10.917	10.5	33.7	1.04	0.32
CTG	L	6.550	51.1	16.8	0.13	0.39
CTT	L	26.201	11.4	6.7	2.30	3.91
TTA	L	13.100	13.8	3.4	0.95	3.85
TTG	L	24.017	13.0	6.7	1.85	3.58
ATG	M	15.284	27.2	16.8	0.56	0.91
AAC	N	15.284	21.7	23.6	0.70	0.65
AAT	N	19.651	19.2	13.5	1.02	1.46

续表 6

密码子 Codon	氨基酸 Amino acid	DHAR	大肠杆菌 基因组 <i>E. coli</i>	酵母基 因组 Yeast	DHAR/大 肠杆菌基因组 DHAR/ <i>E. coli</i>	DHAR/酵 母基因组 DHAR/Yeast
CCA	P	24.017	8.4	3.4	2.86	7.06
CCC	P	6.550	5.6	20.2	1.17	0.32
CCG	P	4.367	22.4	10.1	0.19	0.43
CCT	P	21.834	7.2	3.4	3.03	6.42
CAA	Q	19.651	14.7	23.6	1.34	0.83
CAG	Q	17.467	29.4	20.2	0.59	0.86
AGA	R	13.100	2.9	3.4	4.52	3.85

3 讨论与结论

草莓 DHAR 基因的密码子偏好性分析结果表明,草莓 DHAR 基因的整个编码区序列中,A+T 含量大于 G+C;偏好使用以 A、T 结尾的密码子,较少使用以 G 或 C 结尾的密码子。在草莓 DHAR 基因的密码子中,GCT、TGT、GAT、GAA、TTT 等 24 个以 A 或 T 碱基结尾的密码子,为草莓 DHAR 基因的偏好密码子。

基于 DHAR 基因的密码子使用偏性的系统聚类分析图,预示草莓 DHAR 基因更适合导入双子叶植物中,例如葡萄、苹果等。

要实现目的基因在外源表达系统中的成功表达,必须尽可能的提高目的基因的表达量,密码子优化在提高表达量中起到关键作用。在基因表达的研究中,经常被用来表达外源蛋白的原核表达系统为大肠杆菌,真核表达系统为酵母。该研究中发现草莓 DHAR 基因的密码子偏好性与大肠杆菌和酵母菌的基因组密码子偏好性都有较大差异,预示要实现草莓 DHAR 在以上宿主中的高效表达,尚需对其进行部分密码子的改造。

密码子偏爱性往往是指基因组的偏爱性,该试验仅对 DHAR 基因的密码子偏爱性进行了研究,没有对草莓基因组的密码子偏爱性进行研究,其结果不能代表草莓基因组的密码子使用偏爱性。

参考文献

- [1] 陈坤明,宫海军,王锁民.植物抗坏血酸的生物合成、转运及其生物学功能[J].西北植物学报,2004,24(2):329-336.
- [2] 刘永立,胡海涛,兰大伟.维生素 C 的生物合成及其基因调控研究进展[J].果树学报,2006,23(3):431-436.
- [3] Yabuta Y, Motoki T, Yoshimura K, et al. Thylakoid membrane-bound ascorbate peroxidase is a limiting factor of antioxidant systems under photo-oxidative stress[J]. Planta, 2002, 32: 915-925.
- [4] 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(3):411-414.
- [5] Urrutia A O, Hurst L D. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection[J]. Genetics, 2001, 159(3): 1191-1199.
- [6] Sueoka N, Kawanishi Y. DNA G+C content of the third codon position and codon usage biases of human genes[J]. Gene, 2000, 261(1): 53-62.
- [7] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms[J]. J Mol Evol, 1986, 24(1-2): 28-38.
- [8] Zhou H, Wang H, Huang L F, et al. Heterogeneity in codon usages of sobemovirus genes[J]. Arch Virol, 2005, 150(8): 1591-1605.
- [9] Sharp P M, Cowe E, Higgins D G, et al. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: A review of the considerable within species diversity[J]. Nucl Acids Res, 1988, 16(17): 8207-8211.
- [10] 李平,白云凤,冯瑞云,等.籽粒苋苹果酸酶(NAD-ME)基因密码子偏好性分析[J].应用与环境生物学报,2011,17(1):12-17.

Analysis of Codon Bias of DHAR Gene in Strawberry

LIU Wan-da¹, ZHANG Bing-xiu², WEI Yuan-yuan², WANG Yu¹, DONG Chao², ZHU Yan-ming³

(1. Horticultural Sub-Academy, Heilongjiang Academy of Agricultural Sciences, Harbin, Heilongjiang 150069; 2. College of Horticulture, Northeast Agricultural University, Harbin, Heilongjiang 150030; 3. College of Life Science, Northeast Agricultural University, Harbin, Heilongjiang 150030)

Abstract: Taking strawberry as materials, CHIPS, CUSP and CodonW were used to analyze the codon bias of DHAR gene in strawberry, in order to select appropriate transformation receptor for strawberry DHAR gene in crop genetic improvement and provide reference for genetic optimization. The codon bias of *E. coli* and yeast genome, codon bias of DHAR gene in 7 kinds of plants and strawberry were compared, and based on codon bias of DHAR gene the cluster analysis was made. The results showed that the strawberry DHAR gene was more suitable for introduction to apple or other dicotyledonous plants; to improve the expression level of strawberry DHAR gene in *E. coli* or yeast, codon optimization was required.

Key words: strawberry; DHAR; codon bias; cluster analysis; gene expression