

农业生物技术

草菇密码子偏好性分析

蒋玮^{1,2}, 吕贝贝², 何建华^{1,2}, 王金斌^{1,2}, 吴潇^{1,2}, 武国干^{1,2}, 鲍大鹏³, 陈明杰³,
张劲松³, 谭琦³, 唐雪明^{1,2}

1 上海市农业科学院生物技术研究所, 上海 201106

2 上海市农业遗传育种重点实验室, 上海 201106

3 国家食用菌工程技术研究中心, 上海 201106

蒋玮, 吕贝贝, 何建华, 等. 草菇密码子偏好性分析. 生物工程学报, 2014, 30(9): 1424–1435.

Jiang W, Lü BB, He JH, et al. Codon usage bias in the straw mushroom *Volvariella volvacea*. Chin J Biotech, 2014, 30(9): 1424–1435.

摘要: 以草菇全基因组编码序列为研究对象, 利用软件 CodonW1.4.2 分析草菇基因组密码子使用模式, 确定了草菇的 24 个最优密码子。利用 Create a condon usage table (CUSP) 程序分析计算草菇密码子使用频率, 并将它与人、酵母、拟南芥、小鼠、斑马鱼、果蝇 6 个代表性物种及灰盖鬼伞、双孢蘑菇、香菇、平菇 4 个食用菌进行比较。结果显示草菇密码子偏好性与人、酵母、拟南芥、小鼠、斑马鱼、果蝇和平菇都有较大的差异, 与灰盖鬼伞、双孢蘑菇、香菇的密码子偏好性差异较小。利用软件 SPSS16.0 聚类分析表明密码子偏好性差异大小在一定程度上反映物种间的进化关系, 可作为研究物种进化关系的参考。首次以食用菌全基因组为分析对象, 解析草菇的密码子偏好性, 并将其与其他生物进行比较, 这些将为不同来源的外源基因在草菇中的异源表达提供重要参考。

关键词: 草菇全基因组, 最优密码子, 密码子偏好性

Received: December 9, 2013; **Accepted:** February 28, 2014

Supported by: The Shanghai Academy of Agricultural Sciences (No. 2013(23)), the China Postdoctoral Science Foundation Grant (No. 2013M531197), Shanghai Postdoctoral Sustentation Fund Grant (No. 13R21421000), the Key Project of Developing Agriculture through Science and Technology of Shanghai Municipal Agricultural Commission (No. (2011) 1-2).

Corresponding author: Xueming Tang. Tel/Fax: +86-21-62202746; E-mail: xueming70@foxmail.com

上海农业科学院青年基金 (No. 2013(23)), 国家博士后基金 (No. 2013M531197), 上海市博士后基金 (No. 13R21421000), 上海市科技兴农重点项目 (No. 沪农科攻字(2011)第 1-2 号) 资助。

网络出版时间: 2014-03-25

网络出版地址: <http://www.cnki.net/kcms/doi/10.13345/j.cjb.130623.html>

Codon usage bias in the straw mushroom *Volvariella volvacea*

Wei Jiang^{1,2}, Beibei Lü², Jianhua He^{1,2}, Jinbin Wang^{1,2}, Xiao Wu^{1,2}, Guogan Wu^{1,2},
Dapeng Bao³, Mingjie Chen³, Jinsong Zhang³, Qi Tan³, and Xueming Tang^{1,2}

¹ Biotechnology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai 201106, China

² Shanghai Key Laboratory of Agricultural Genetics and Breeding, Shanghai 201106, China

³ National Engineering Research Center of Edible Fungi, Shanghai 201106, China

Abstract: We analyzed the whole genome coding sequence of *Volvariella volvacea* to study the pattern utilization of codons by Codon W 1.4.2. As results, 24 optimal codons were identified. Moreover, the frequency of codons usage was calculated by CUSP program. We compared the frequency of codons usage of *V. volvacea* with other organisms including 6 modal value species (*Homo sapiens*, *Saccharomys cerevisiae*, *Arabidopsis thaliana*, *Mus musculus*, *Danio rerio* and *Drosophila melanogaster*) and 4 edible fungi (*Coprinopsis cinerea*, *Agaricus bisporus*, *Lentinula edodes* and *Pleurotus ostreatus*). We found that there were less differences in 3 edible fungi (excluding *Pleurotus ostreatus*) than 6 modal value species, comparing with the frequency of codons usage of *V. volvacea*. With software SPSS16.0, cluster analysis which showed differences in the size of codon bias, reflects the evolutionary relationships between species, which can be used as a reference of evolutionary relationships of species. This was the first time for analysis the codon preference among the whole coding sequences of edible fungi, serving as theoretical basis to apply genetic engineering of *V. volvacea*.

Keywords: *Volvariella volvacea*, optimal codons, codon usage bias

在绝大多数生物中，密码子并不是均等地被使用，长期的进化过程中，某一物种或基因都会形成对特定的同义密码子的使用偏好，这些被优先使用的密码子被称为该物种或基因的最优密码子(Optimal codon)，这种现象称为密码子的使用偏好性(Codon bias)。分析密码子的使用偏好性对提高外源基因在宿主中的表达水平具有重要的意义。近几年，通过优化密码子的途径提高外源基因在微生物、植物、动物等宿主生物中的表达效率的研究屡见不鲜^[1-7]，其前提便是对宿主生物的密码子偏好性的分析，而酵母密码子偏好性的解析更是为酵母的基因工程技术化应用带来了突破，得到大量应用^[8-14]。

食用菌作为营养、健康的食物代表，日益受到人们的关注，但是食用菌中密码子偏好性分析研究仍然少之又少。其中包括双孢蘑菇^[15]和银耳^[16]，但是他们的分析只是基于部分表达序列，以某一食用菌全基因组序列为分析对象的密码子偏好性研究还尚未报道。草菇 *Volvariella volvacea* 作为一种重要的常见食用菌，在中国有着悠久的栽培历史，因其味道鲜美，营养丰富，在中国及东南亚地区很受欢迎。本研究通过分析测序得到的草菇全基因组编码序列数据，了解草菇基因密码子用法特征，并与不同的代表性物种进行比较，为草菇基因选择合适的表达系统，优化密码子提高基因表达量等提供重要理论基础。

1 材料与方法

1.1 材料

草菇全基因组序列由测序得到,利用真核基因预测软件 Genemark-ES 预测,共获得 11 097 个蛋白编码基因^[17]。剔除 380 条长度小于 300 bp 的序列,剩余 10 717 条序列用于本研究的各项分析^[18]。在 GenBank (<http://www.ncbi.nlm.nih.gov/>) 中检索草菇 mRNA 序列,从中筛选用于比较分析的蛋白质编码序列 40 条(Accession No.: AF329731.1, AF329732.1, AY316307.1, AY855921.1, DQ888226.1, EF143259.1, FJ536255.1, FJ913469.1, GU376797.1, HM237364.1, HM366445.1, HM366446.1, JN830959.1, JN861640.1, JQ478424.1, JQ478425.1, JQ922270.1, JX402409.1, JX402410.1, JX402411.1, KC520537.1, KC748445.1, KC748452.1, KF365489.1, KF365492.1, KF365493.1, KF365495.1, KF365499.1, KF528321.1, KF528322.1, KF528324.1, KF528325.1, KF528326.1, KF528327.1, KF528328.1, KF528329.1, KF528330.1, KF528331.1, KF528332.1, KF528333.1)。

1.2 方法

1.2.1 草菇基因同义密码子偏好性分析

运用 CodonW1.4.2 软件进行分析相对密码子使用度(Relative synonymous codon usage, RSCU)和有效密码子数(Effective number of codons, Nc)。

1.2.2 草菇高表达最优密码子确定

采用 Sharp 等^[19]的方法,先计算每个基因的 Nc 值并进行排列,从这一排列的两端抽取 10% 的样本分别组成高、低表达样本组。然后各自计算这两个样本组中各个密码子的 RSCU 值,最后通过卡方检验确定出高表达基因的优越密码子,确定草菇基因组最优密码子。

1.2.3 草菇与其他物种密码子偏好性比较

运用 CUSP 程序计算草菇基因各密码子的

使用频率,并与从 Codon Usage Database (<http://www.kazusa.or.jp/codon/>) 中获得的人 *Homo sapiens*、酵母 *Saccharomys cerevisiae*、拟南芥 *Arabidopsis thaliana*、小鼠 *Mus musculus*、斑马鱼 *Danio rerio* 果蝇 *Drosophila melanogaster*、灰盖鬼伞 *Coprinopsis cinerea*、双孢蘑菇 *Agaricus bisporus*、香菇 *Lentinula edodes*、平菇 *Pleurotus ostreatus* 的密码子使用频率进行比较。

1.2.4 基于密码子使用偏好性的聚类

利用 SPSS16.0 对草菇及其他 10 种物种进行基于密码子使用偏好性的欧氏平方距离系数进行聚类分析。在聚类过程中,将每个物种作为一个对象,密码子使用频率作为变量。采用 59 个同义密码子(不包括起始密码子 ATG 和色氨酸 TGG 以及 3 个终止密码子 TAG、TAA、TGA) 的密码子使用频率值对其进行分析。物种间的距离规定为同义密码子使用频率的欧氏平方距离,根据两两间的距离构建距离系数矩阵,进行聚类分析。

2 结果与分析

2.1 草菇基因同义密码子相对使用度分析

运用 CodonW 软件分析了 10 717 条草菇蛋白质基因编码序列的 RSCU 值,结果如表 1 所示。有偏好性(RSCU>1)的密码子共计 35 个,其中以 G/C 结尾的 19 个,以 A/T 结尾的 16 个。各氨基酸 RSCU 值最高的密码子分别是: TTC (Phe)、CTC (Leu)、ATC (Ile)、GTC (Val)、TCC (Ser)、CCA (Pro)、ACC (Thr)、GCT (Ala)、TAC (Tyr)、CAC (His)、CAA (Gln)、AAC (Asn)、AAG (Lys)、GAT (Asp)、GAG (Glu)、TGC (Cys)、CGC (Arg)、GGT (Gly)、TGA (终止密码子)。这

19 个密码子以 G/C 结尾的 13 个, 以 A/T 结尾的 6 个。以上结果表明, 草菇偏好于以 G/C 结尾的密码子。

2.2 最优密码子的确定

一般地, 密码子的有效数(Nc)与基因表达量有关, 高表达基因密码子偏好性强, Nc 值小^[20]。本研究以草菇基因组中各基因的 Nc 值排列获取高、低表达基因样本, 通过计算基因之间同义

密码子相对使用频率之差, 并经卡方测验, 最终确定了 TTC、CTC、CTG 等 24 个密码子为草菇的高表达优越密码子(表 2)。它们在高、低表达基因样本组间的使用频率差异较大, 显著性测验结果均呈极显著水平。通过分析 40 条已知草菇蛋白质编码序列所得的最优密码子与分析草菇全基因组序列所得的最优密码子, 经比较可见后者所得结果可靠且更全面(表 2)。

表 1 草菇蛋白编码基因的密码子用法

Table 1 The codon usage of *Volvariella volvacea* coding sequences

Amino acid	Codon	Number	RSCU	Amino acid	Codon	Number	RSCU
Phe	TTT	84 403	0.80	Tyr	TAT	76 195	0.96
	TTC	127 001	1.20		TAC	82 272	1.04
Leu	TTA	44 737	0.50	TER	TAA	3 572	1.00
	TTG	112 221	1.26		TAG	3 083	0.86
	CTT	97 431	1.09	His	CAT	71 864	0.95
	CTC	132 311	1.49		CAC	80 068	1.05
	CTA	58 939	0.66		Gln	CAA	122 693
CTG	88 765	1.00	CAG	104 087		0.92	
Ile	ATT	105 952	1.08	Asn	AAT	100 323	0.95
	ATC	130 319	1.32		AAC	111 126	1.05
	ATA	59 050	0.60	Lys	AAA	101 877	0.79
Met	ATG	109 879	1.00		AAG	156 658	1.21
	Val	GTT	100 062		1.13	Asp	GAT
GTC		116 837	1.31	GAC	147 000		0.95
GTA		53 191	0.60	Glu	GAA	159 420	0.94
GTG		85 317	0.96		GAG	178 154	1.06
Ser		TCT	95 926		1.13	Cys	TGT
	TCC	101 521	1.19	TGC	39 300		1.15
	TCA	86 852	1.02	TER	TGA	4 062	1.14
	TCG	85 633	1.01		Trp	TGG	85 336
Pro	CCT	110 175	1.13	Arg	CGT	57 001	0.97
	CCC	100 227	1.02		CGC	66 425	1.13
	CCA	112 465	1.15		CGA	61 330	1.04
	CCG	68 387	0.70		CGG	44 313	0.75
Thr	ACT	83 294	0.94	Ser	AGT	62 347	0.73
	ACC	100 547	1.13		AGC	78 226	0.92
	ACA	90 149	1.01	Arg	AGA	58 634	1.00
	ACG	81 729	0.92		AGG	64 896	1.10
Ala	GCT	129 272	1.12	Gly	GGT	103 654	1.09
	GCC	117 078	1.01		GGC	99 764	1.05
	GCA	125 320	1.08		GGA	101 809	1.07
	GCG	91 698	0.79		GGG	76 592	0.80

表 2 草菇中高/低表达样本组的密码子用法与最优密码子的确定和验证

Table 2 Codon usage of high/low expressed sample group in *Volvariella volvacea* and confirmation of optimal codons in *Volvariella volvacea*

Amino acid	Codon	High		Low		Amino acid	Codon	High		Low	
		Number	RSCU	Number	RSCU			Number	RSCU	Number	RSCU
Phe	TTT	4 632	0.55	7 732	0.96	Tyr	TAT	3 986	0.61	7 206	1.14
	TTC*	12 221	1.45	8 328	1.04		TAC*+	9 027	1.39	5 451	0.86
Leu	TTA	959	0.15	5 289	0.76	TER	TAA	412	1.15	342	0.96
	TTG	6 973	1.10	9 083	1.30		TAG	304	0.85	277	0.78
	CTT	5 156	0.81	7 669	1.10	His	CAT	3 229	0.58	7 129	1.12
	CTC*+	16 530	2.60	6 852	0.98		CAC*	7 847	1.42	5 617	0.88
	CTA	1 956	0.31	6 328	0.91	Gln	CAA	7 422	0.95	12 331	1.16
	CTG*	6 628	1.04	6 621	0.95		CAG*	8 204	1.05	8 975	0.84
Ile	ATT	5 822	0.80	8 930	1.13	Asn	AAT	4 613	0.54	10 162	1.11
	ATC*+	14 195	1.95	7 885	1.00		AAC*	12 512	1.46	8 091	0.89
	ATA	1 798	0.25	6 914	0.87	Lys	AAA	4 957	0.51	11 074	0.92
	ATG	7 669	1.00	9 972	1.00		AAG*+	14 350	1.49	13 074	1.08
Met	GTT	6 345	0.88	8 269	1.19	Asp	GAT	9 093	0.79	15 572	1.17
	GTC*+	14 522	2.01	6 750	0.97		GAC*+	13 830	1.21	11 084	0.83
Val	GTA	2 014	0.28	5 601	0.81	Glu	GAA	8 097	0.69	16 324	1.06
	GTG	5 951	0.83	7 123	1.03		GAG*	15 525	1.31	14 616	0.94
Ser	TCT	5 914	0.99	8 812	1.17	Cys	TGT	1 310	0.52	2 858	1.10
	TCC*+	10 848	1.81	6 634	0.88		TGC*+	3 711	1.48	2 347	0.90
	TCA	3 387	0.57	9 891	1.32	TER	TGA	355	0.99	452	1.27
	TCG*	6 286	1.05	7 008	0.93		Trp	TGG	6 558	1.00	6 765
Pro	CCT	6 693	0.93	11 042	1.20	Arg	CGT*	4 213	1.04	4 056	0.79
	CCC*+	11 261	1.56	7 395	0.80		CGC*+	7 570	1.87	3 698	0.72
	CCA	5 467	0.76	12 609	1.37		CGA	2 814	0.69	5 969	1.17
	CCG*	5 491	0.76	5 886	0.64		CGG	2 707	0.67	3 765	0.74
Thr	ACT	5 443	0.78	7 530	0.98	Ser	AGT	2 564	0.43	6 797	0.90
	ACC*+	11 776	1.69	6 610	0.86		AGC*	6 963	1.16	5 958	0.79
	ACA	4 483	0.64	9 747	1.27	Arg	AGA	2 598	0.64	7 126	1.40
	ACG	6 134	0.88	6 840	0.89		AGG	4 451	1.10	6 033	1.18
Ala	GCT	10 157	1.03	11 213	1.15	Gly	GGT	9 214	1.11	9 030	1.08
	GCC*+	13 865	1.41	7 820	0.80		GGC*+	12 046	1.46	6 600	0.79
	GCA	7 221	0.74	12 839	1.32		GGA	6 376	0.77	1 069	1.28
	GCG*	8 044	0.82	7 115	0.73		GGG	5 441	0.66	7 106	0.85

Note: Codon usage was compared using Chi squared contingency test to identify optimal codons that occur significantly more often ($P < 0.01$) are indicated with a *. Optimal codons of known 40 coding sequences of *Volvariella volvacea* are indicated with a +.

2.3 草菇与代表性物种密码子偏好性比较

草菇与人、酵母、拟南芥、小鼠、斑马鱼、果蝇等6种物种密码子使用频率进行比较(表3),其中,V/H、V/Y、V/A、V/M、V/D、V/F分别表示草菇与人、酵母、拟南芥、小鼠、斑马鱼、果蝇的每种密码子使用频率比值,密码子使用频率

若在0.5–2.0之间,表明这两个物种对该密码子的偏好性较接近,若 ≥ 2.0 或 ≤ 0.5 ,则表明偏好性差异较大^[21]。草菇与人、酵母、拟南芥、小鼠、斑马鱼、果蝇的比值中, ≥ 2.0 或 ≤ 0.5 的密码子分别有7、12、3、7、3、5种,表明草菇与这些物种之间存在不同程度的密码子偏好性差异。

表3 草菇与代表性物种密码子偏好性比较

Table 3 Comparison of codon preference between the *Volvariella volvacea* and other representative organisms

Amino acid	Codon	Frequency of codon usage (%)							Comparison of codon preference					
		<i>V. volvacea</i>	<i>H. sapiens</i>	<i>S. cerevisiae</i>	<i>A. thaliana</i>	<i>M. musculus</i>	<i>D. rerio</i>	<i>D. melanogaster</i>	V/H	V/Y	V/A	V/M	V/D	V/F
Phe	TTT	14.6	17.6	26.1	21.8	17.2	18.2	13.2	0.83	0.56	0.67	0.85	0.80	1.11
	TTC	22.0	20.3	18.4	20.7	21.8	20.8	21.8	1.08	1.19	1.06	1.01	1.06	1.01
Leu	TTA	7.7	7.7	26.2	12.7	6.7	7.0	4.5	1.01	0.30	0.61	1.16	1.11	1.72
	TTG	19.4	12.9	27.2	20.9	13.4	12.3	16.1	1.50	0.71	0.93	1.45	1.58	1.21
	CTT	16.9	13.2	12.3	24.1	13.4	12.7	9.0	1.28	1.37	0.70	1.26	1.33	1.87
	CTC	22.9	19.6	5.4	16.1	20.2	17.0	13.8	1.17	4.24	1.42	1.13	1.35	1.66
	CTA	10.2	7.2	13.4	9.9	8.1	6.2	8.2	1.42	0.76	1.03	1.26	1.64	1.24
	CTG	15.4	39.6	10.5	9.8	39.5	37.6	38.2	0.39	1.46	1.57	0.39	0.41	0.40
Ile	ATT	18.3	16.0	30.1	21.5	15.4	16.5	16.6	1.15	0.61	0.85	1.19	1.11	1.10
	ATC	22.5	20.8	17.2	18.5	22.5	23.7	22.9	1.08	1.31	1.22	1.00	0.95	0.98
	ATA	10.2	7.5	17.8	12.6	7.4	7.7	9.5	1.36	0.57	0.81	1.38	1.33	1.08
Met	ATG	19.0	22.0	20.9	24.5	22.8	25.5	23.6	0.86	0.91	0.78	0.83	0.75	0.81
Val	GTT	17.3	11.0	22.1	27.2	10.7	14.1	11.0	1.57	0.78	0.64	1.62	1.23	1.57
	GTC	20.2	14.5	11.8	12.8	15.4	14.8	13.9	1.39	1.71	1.58	1.31	1.37	1.45
	GTA	9.2	7.1	11.8	9.9	7.4	6.7	6.4	1.30	0.78	0.93	1.24	1.37	1.44
	GTG	14.8	28.1	10.8	17.4	28.4	28.3	27.8	0.53	1.37	0.85	0.52	0.52	0.53
Ser	TCT	16.6	15.2	23.5	25.2	16.2	16.9	7.0	1.09	0.71	0.66	1.02	0.98	2.37
	TCC	17.6	17.7	14.2	11.2	18.1	15.2	19.6	0.99	1.24	1.57	0.97	1.16	0.90
	TCA	15.0	12.2	18.7	18.3	11.8	13.2	7.8	1.23	0.80	0.82	1.27	1.14	1.93
	TCG	14.8	4.4	8.6	9.3	4.2	5.6	16.6	3.37	1.72	1.59	3.53	2.65	0.89
Pro	CCT	19.1	17.5	13.5	18.7	18.4	16.6	6.9	1.09	1.41	1.02	1.04	1.15	2.76
	CCC	17.3	19.8	6.8	5.3	18.2	12.7	18.1	0.88	2.55	3.27	0.95	1.37	0.96
	CCA	19.5	16.9	18.3	16.1	17.3	15.7	13.5	1.15	1.06	1.21	1.12	1.24	1.44
	CCG	11.8	6.9	5.3	8.6	6.2	8.2	15.8	1.71	2.23	1.38	1.91	1.44	0.75
Thr	ACT	14.4	13.1	20.3	17.5	13.7	14.5	9.5	1.10	0.71	0.82	1.05	0.99	1.52
	ACC	17.4	18.9	12.7	10.3	19.0	16.2	21.3	0.92	1.37	1.69	0.92	1.07	0.82
	ACA	15.6	15.1	17.8	15.7	16.0	17.0	11.0	1.03	0.88	0.99	0.97	0.92	1.42

续表 3

Ala	ACG	14.1	6.1	8.0	7.7	5.6	7.4	14.4	2.32	1.77	1.84	2.52	1.91	0.98
	GCT	22.4	18.4	21.2	28.3	20.0	20.9	14.4	1.22	1.05	0.79	1.12	1.07	1.55
	GCC	20.3	27.7	12.6	10.3	26.0	19.5	33.6	0.73	1.61	1.97	0.78	1.04	0.60
Try	GCA	21.7	15.8	16.2	17.5	15.8	16.6	12.8	1.37	1.34	1.24	1.37	1.31	1.69
	GCG	15.9	7.4	6.2	9.0	6.4	8.6	14.0	2.14	2.56	1.76	2.48	1.84	1.13
	TAT	13.2	12.2	18.8	14.6	12.2	12.6	10.8	1.08	0.70	0.90	1.08	1.05	1.22
TER	TAC	14.2	15.3	14.8	13.7	16.1	17.0	18.4	0.93	0.96	1.04	0.88	0.84	0.77
	TAA	0.6	1.0	1.1	0.9	1.0	1.1	0.8	0.62	0.56	0.69	0.62	0.56	0.77
His	TAG	0.5	0.8	0.5	0.5	0.8	0.6	0.7	0.67	1.07	1.07	0.67	0.89	0.76
	CAT	12.4	10.9	13.6	13.8	10.6	10.9	10.8	1.14	0.91	0.90	1.17	1.14	1.15
Gln	CAC	13.9	15.1	7.8	8.7	15.3	14.8	16.2	0.92	1.78	1.59	0.91	0.94	0.85
	CAA	21.2	12.3	27.3	19.4	12.0	11.8	15.6	1.73	0.78	1.09	1.77	1.80	1.36
Asn	CAG	18.0	34.2	12.1	15.2	34.1	33.5	36.1	0.53	1.49	1.18	0.53	0.54	0.50
	AAT	17.4	17.0	35.7	22.3	15.6	16.3	21.0	1.02	0.49	0.78	1.11	1.06	0.83
Lys	AAC	19.2	19.1	24.8	20.9	20.3	24.1	26.2	1.01	0.78	0.92	0.95	0.80	0.73
	AAA	17.6	24.4	41.9	30.8	21.9	29.3	17.0	0.72	0.42	0.57	0.80	0.60	1.04
Asp	AAG	27.1	31.9	30.8	32.7	33.6	30.7	39.5	0.85	0.88	0.83	0.81	0.88	0.69
	GAT	28.2	21.8	37.6	36.6	21.0	24.8	27.6	1.29	0.75	0.77	1.34	1.14	1.02
Glu	GAC	25.4	25.1	20.2	17.2	26.0	27.8	24.6	1.01	1.26	1.48	0.98	0.91	1.03
	GAA	27.6	29.0	45.6	34.3	27.0	24.4	21.1	0.95	0.60	0.80	1.02	1.13	1.31
Cys	GAG	30.8	39.6	19.2	32.2	39.4	42.8	42.5	0.78	1.61	0.96	0.78	0.72	0.73
	TGT	5.1	10.6	8.1	10.5	11.4	11.3	5.4	0.48	0.63	0.48	0.44	0.45	0.94
TER	TGC	6.8	12.6	4.8	7.2	12.3	11.2	13.2	0.54	1.42	0.94	0.55	0.61	0.52
	TGA	0.7	1.6	0.7	1.2	1.6	1.4	0.5	0.44	1.00	0.59	0.44	0.50	1.41
Trp	TGG	14.8	13.2	10.4	12.5	12.5	11.6	9.9	1.12	1.42	1.18	1.18	1.27	1.49
Arg	CGT	9.9	4.5	6.4	9.0	4.7	6.9	8.8	2.19	1.54	1.10	2.10	1.43	1.12
	CGC	11.5	10.4	2.6	3.8	9.4	9.6	18.0	1.10	4.42	3.02	1.22	1.20	0.64
	CGA	10.6	6.2	3.0	6.3	6.6	6.7	8.4	1.71	3.54	1.68	1.61	1.58	1.26
Ser	CGG	7.7	11.4	1.7	4.9	10.2	6.6	8.2	0.67	4.51	1.56	0.75	1.16	0.93
	AGT	10.8	12.1	14.2	14.0	12.7	13.2	11.5	0.89	0.76	0.77	0.85	0.82	0.94
	AGC	13.5	19.5	9.8	11.3	19.7	18.4	20.4	0.69	1.38	1.20	0.69	0.74	0.66
Arg	AGA	10.1	12.2	21.3	19.0	12.1	14.3	5.1	0.83	0.48	0.53	0.84	0.71	1.99
	AGG	11.2	12.0	9.2	11.0	12.2	10.2	6.3	0.94	1.22	1.02	0.92	1.10	1.78
Gly	GGT	17.9	10.8	23.9	22.2	11.4	13.7	13.3	1.66	0.75	0.81	1.57	1.31	1.35
	GGC	17.3	22.2	9.8	9.2	21.2	17.2	26.7	0.78	1.76	1.88	0.81	1.00	0.65
	GGA	17.6	16.5	10.9	24.2	16.8	21.5	18.0	1.07	1.62	0.73	1.05	0.82	0.98
	GGG	13.2	16.5	6.0	10.2	15.2	10.0	4.7	0.80	2.21	1.30	0.87	1.32	2.82

2.4 草菇与其他食用菌密码子偏好性比较

草菇与灰盖鬼伞、双孢蘑菇、香菇、平菇 4 种食用菌密码子使用频率进行比较 (表 4), 其中, V/C、V/A、V/L、V/P 分别表示草菇与灰盖鬼伞、

双孢蘑菇、香菇、平菇的每种密码子使用频率比值, ≥ 2.0 或 ≤ 0.5 的密码子分别有 1、2、1、4 种, 表明草菇与灰盖鬼伞、双孢蘑菇、香菇的密码子偏好性差异较小, 与平菇的密码子偏好性差异较大。

表 4 草菇与其他食用菌密码子偏好性比较

Table 4 Comparison of codon preference between the *Volvariella volvacea* and other edible fungi

Amino acid	Codon	Frequency of codon usage (%)					Comparison of codon preference			
		<i>V. volvacea</i>	<i>C. cinerea</i>	<i>A. bisporus</i>	<i>L. edodes</i>	<i>P. ostreatus</i>	V/C	V/A	V/L	V/P
Phe	TTT	14.6	14.0	10.8	16.0	9.0	1.04	1.35	0.91	1.62
	TTC	22.0	27.4	30.5	25.1	37.0	0.80	0.72	0.88	0.59
Leu	TTA	7.7	4.4	3.4	7.3	3.3	1.75	2.26	1.05	2.33
	TTG	19.4	15.2	17.8	18.0	16.7	1.28	1.09	1.08	1.16
	CTT	16.9	19.8	17.7	21.5	18.3	0.85	0.95	0.79	0.92
	CTC	22.9	26.6	29.5	21.2	23.9	0.86	0.78	1.08	0.96
	CTA	10.2	5.8	5.7	6.4	4.8	1.76	1.79	1.59	2.13
	CTG	15.4	9.5	14.4	10.6	13.9	1.62	1.07	1.45	1.11
Ile	ATT	18.3	22.9	18.0	21.0	17.9	0.80	1.02	0.87	1.02
	ATC	22.5	28.4	29.9	21.9	31.4	0.79	0.75	1.03	0.72
	ATA	10.2	5.6	4.6	8.4	4.9	1.82	2.22	1.21	2.08
Met	ATG	19.0	18.4	19.6	17.1	19.1	1.03	0.97	1.11	0.99
Val	GTT	17.3	23.6	17.6	22.4	17.7	0.73	0.98	0.77	0.98
	GTC	20.2	31.0	27.6	23.4	28.5	0.65	0.73	0.86	0.71
	GTA	9.2	7.3	5.0	10.3	5.8	1.26	1.84	0.89	1.59
	GTG	14.8	9.7	12.9	12.3	12.7	1.53	1.15	1.20	1.17
Ser	TCT	16.6	17.8	16.1	22.3	14.6	0.93	1.03	0.74	1.14
	TCC	17.6	14.3	20.9	15.9	15.8	1.23	0.84	1.11	1.11
	TCA	15.0	9.8	10.3	13.7	8.9	1.53	1.46	1.09	1.69
	TCG	14.8	10.5	18.8	12.5	17.7	1.41	0.79	1.18	0.84
Pro	CCT	19.1	19.3	23.3	24.1	20.0	0.99	0.82	0.79	0.96
	CCC	17.3	16.5	20.7	13.4	18.9	1.05	0.84	1.29	0.92
	CCA	19.5	9.0	15.2	14.5	15.8	2.17	1.28	1.34	1.23
	CCG	11.8	8.0	12.1	10.5	9.0	1.48	0.98	1.12	1.31
Thr	ACT	14.4	21.0	14.6	22.4	17.8	0.69	0.99	0.64	0.81
	ACC	17.4	21.7	23.4	20.7	23.1	0.80	0.74	0.84	0.75
	ACA	15.6	11.5	10.2	14.4	11.8	1.36	1.53	1.08	1.32
	ACG	14.1	9.2	12.1	12.1	21.2	1.53	1.17	1.17	0.67

续表 4

Ala	GCT	22.4	32.2	25.2	31.8	26.3	0.70	0.89	0.70	0.85
	GCC	20.3	25.7	28.0	19.8	26.3	0.79	0.73	1.03	0.77
	GCA	21.7	16.8	15.3	19.5	14.7	1.29	1.42	1.11	1.48
	GCG	15.9	11.6	16.0	11.7	16.9	1.37	0.99	1.36	0.94
Try	TAT	13.2	13.0	8.5	16.0	9.9	1.02	1.55	0.83	1.33
	TAC	14.2	16.1	16.4	15.8	16.8	0.88	0.87	0.90	0.85
TER	TAA	0.6	1.1	0.9	0.7	1.3	0.55	0.67	0.86	0.46
	TAG	0.5	0.8	0.5	0.7	0.4	0.63	1.00	0.71	1.25
His	CAT	12.4	12.1	10.8	13.7	9.3	1.02	1.15	0.91	1.33
	CAC	13.9	11.1	11.9	11.2	12.9	1.25	1.17	1.24	1.08
Gln	CAA	21.2	20.5	17.8	21.0	19.2	1.03	1.19	1.01	1.10
	CAG	18.0	14.5	19.1	16.9	17.0	1.24	0.94	1.07	1.06
Asn	AAT	17.4	21.3	12.0	22.6	15.2	0.82	1.45	0.77	1.14
	AAC	19.2	28.1	26.5	23.9	32.4	0.68	0.72	0.80	0.59
Lys	AAA	17.6	21.0	15.6	17.1	14.5	0.84	1.13	1.03	1.21
	AAG	27.1	26.9	33.7	19.9	28.7	1.01	0.80	1.36	0.94
Asp	GAT	28.2	30.5	25.7	31.4	25.1	0.92	1.10	0.90	1.12
	GAC	25.4	25.0	34.7	27.9	35.5	1.02	0.73	0.91	0.72
Glu	GAA	27.6	26.5	27.1	26.4	19.8	1.04	1.02	1.05	1.39
	GAG	30.8	22.7	31.0	20.6	24.0	1.36	0.99	1.50	1.28
Cys	TGT	5.1	6.8	4.4	7.7	5.1	0.75	1.16	0.66	1.00
	TGC	6.8	9.3	8.3	7.9	11.2	0.73	0.82	0.86	0.61
TER	TGA	0.7	1.0	0.9	0.7	0.6	0.70	0.78	1.00	1.17
Trp	TGG	14.8	14.8	13.6	14.7	13.7	1.00	1.09	1.01	1.08
Arg	CGT	9.9	11.2	9.4	11.9	7.7	0.88	1.05	0.83	1.29
	CGC	11.5	9.2	13.9	7.6	11.6	1.25	0.83	1.51	0.99
	CGA	10.6	6.3	11.6	9.1	7.3	1.68	0.91	1.16	1.45
	CGG	7.7	4.0	7.2	5.1	3.9	1.93	1.07	1.51	1.97
Ser	AGT	10.8	10.3	7.5	11.1	10.5	1.05	1.44	0.97	1.03
	AGC	13.5	11.2	12.8	11.6	13.2	1.21	1.05	1.16	1.02
Arg	AGA	10.1	5.1	5.2	5.9	6.1	1.98	1.94	1.71	1.66
	AGG	11.2	6.9	8.0	4.4	6.3	1.62	1.40	2.55	1.78
Gly	GGT	17.9	32.4	16.9	23.5	21.0	0.55	1.06	0.76	0.85
	GGC	17.3	26.4	19.5	17.6	26.9	0.66	0.89	0.98	0.64
	GGA	17.6	21.5	13.1	23.1	16.9	0.82	1.34	0.76	1.04
	GGG	13.2	8.1	8.4	9.8	12.4	1.63	1.57	1.35	1.06

2.5 物种间密码子偏好性的聚类分析

根据表 3 及表 4 中各物种密码子的使用频率, 利用 SPSS16.0 进行聚类分析, 聚类结果见

图 1。从图 1 可以看出, 利用密码子使用频率得出的聚类图接近这 11 个物种的系统分类, 在一定程度上反映了物种间的进化关系。

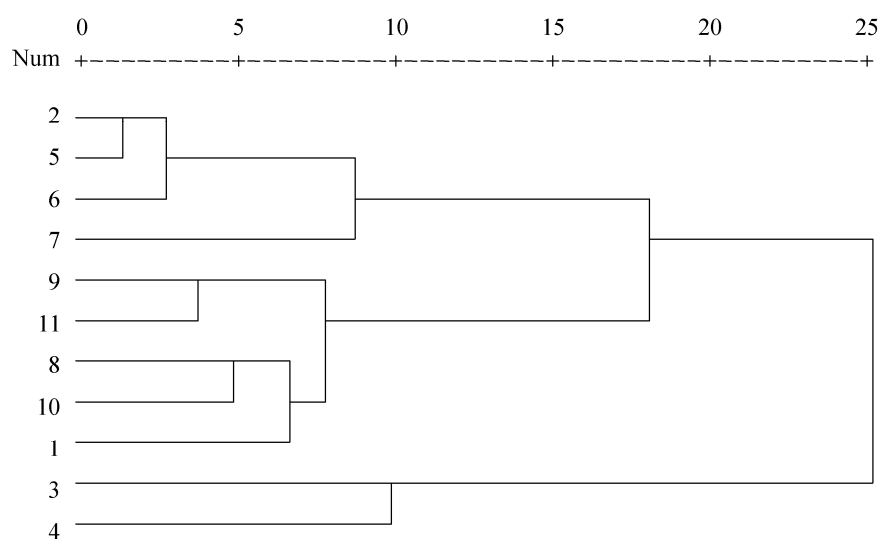


图 1 基于物种密码子使用频率的聚类图

Fig. 1 Cluster analysis dendrogram of frequency of codon usage for 11 species. 1: *Volvariella volvacea*; 2: *Homo sapiens*; 3: *Saccharomyces cerevisiae*; 4: *Arabidopsis thaliana*; 5: *Mus musculus*; 6: *Danio rerio*; 7: *Drosophila melanogaster*; 8: *Coprinopsis cinerea*; 9: *Agaricus bisporus*; 10: *Lentinula edodes*; 11: *Pleurotus ostreatus*.

3 讨论

在长期的进化过程中, 由于进化环境和选择压力的不同, 不同物种经历了不同的进化历程, 所以任何一个物种为适应其基因组环境, 都要形成特定的符合其基因组的密码子使用法则。不同生物物种密码子偏性有一定差别, 将外源基因转入受体生物时, 由于转入基因的密码子和宿主基因组密码子使用偏好不同, 易引起甲基化, 从而引发转基因沉默或转基因表达量降低^[22]。因此, 在外源基因转入前, 应按照宿主物种的密码子使用偏好性对其进行优化和改造。本研究对草菇全基因组蛋白编码基因的密码子偏好性进行分析, 结果表明草菇基因密

码子偏爱使用 G/C 结尾的密码子, 这种偏好可能是由于以 G/C 结尾的密码子具有较高的结合能, 有利于保证翻译的准确性^[23]。

草菇密码子偏好性与人、酵母、拟南芥、小鼠、斑马鱼、果蝇、平菇都具有不同程度的差异, 如果这些生物的基因要在草菇中表达, 则应该首先进行密码子优化。草菇与灰盖鬼伞、双孢蘑菇、香菇的密码子偏好性差异较小。如果这些生物的基因要在草菇中表达, 只需优化个别密码子。聚类结果表明密码子偏好性差异大小在一定程度上反映物种间的进化关系, 可作为研究物种进化关系的参考。但基于物种密码子使用频率的聚类结果与传统分类存在一定的差异, 这可能是参数选择单一的问题, 若与

其他参数联用,可能会更贴切地反映物种间的进化关系^[24]。

目前,通过密码子优化方法提高外源基因在食用菌中的表达量的研究还鲜有报道。本研究中确定了草菇的24个最优密码子,对于今后草菇转基因过程中对构建合适的转基因表达系统具有重要的指导意义,在目的表达基因中针对草菇所偏好的密码子进行优化改造,从而提高目的蛋白的表达量,进而使草菇获得更多有利性状,对草菇品种进行定向改良具有十分重要的理论和现实意义。

REFERENCES

- [1] Zhou ZL, Lin ZM, Geng LL, et al. Comparison of codon optimizations of *cryIAh1* gene in rice. *Chin J Biotech*, 2012, 28(10): 1184–1194 (in Chinese).
周宗梁, 林智敏, 耿丽丽, 等. 水稻中 *cryIAh1* 基因密码子优化方案的比较. *生物工程学报*, 2012, 28(10): 1184–1194.
- [2] Zelasko S, Palaria A, Das A. Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems. *Protein Express Purif*, 2013, 92(1): 77–87.
- [3] Stephen JS, Jeremy DV, Robert M, et al. Expression of chicken parvovirus VP2 in chicken embryo fibroblasts requires codon optimization for production of naked DNA and vectored meleagrid herpesvirus type 1 vaccines. *Virus Genes*, 2013, 47(2): 259–267.
- [4] Rasheda J, Muhammad SK, Yusuf Z, et al. Codon optimization of *cryIAb* gene for hyper expression in plant organelles. *Mol Biol Rep*, 2010, 37(2): 1011–1017.
- [5] Mizuki T, Masafumi T, Takahiro S, et al. Transcripts of a heterologous gene encoding mite allergen Der f 7 are stabilized by codon optimization in *Aspergillus oryzae*. *Appl Microbial Biot*, 2012, 96(5): 1275–1282.
- [6] Kyoko HT, Mpanja N, Tadayoshi H, et al. High-level accumulation of recombinant miraculin protein in transgenic tomatoes expressing a synthetic *miraculin* gene with optimized codon usage terminated by the native miraculin terminator. *Plant Cell Rep*, 2011, 30(1): 113–124.
- [7] Soohyun P, Seung PP, Jinwon L. Expression of codon-optimized phosphoenolpyruvate carboxylase gene from *Glaciecola* sp. HTCC2999 in *Escherichia coli* and its application for C4 chemical production. *Appl Biochem Biotech*, 2012, 167(7): 1845–1853.
- [8] Mirzaei SA, Tabatabaei M, Sephezadeh Z. Secretory expression and purification of a soluble NADH cytochrome b5 reductase enzyme from *Mucor racemosus* in *Pichia pastoris* based on codon usage adaptation. *Biotechnol Lett*, 2010, 32(11): 1705–1711.
- [9] Andrea M, Roland W, Anton G, et al. Expression of lignocellulolytic enzymes in *Pichia pastoris*. *Micro Cell Fact*, 2012(11): 61.
- [10] Fu XY, Zhao W, Xiong AS, et al. High expression of recombinant *Streptomyces* sp. S38 xylanase in *Pichia pastoris* by codon optimization and analysis of its biochemical properties. *Mol Biol Rep*, 2011, 38(8): 4991–4997.
- [11] Christoph S, Alexander G, Iris K, et al. Simple and efficient expression of *Agaricus meleagris* pyranose dehydrogenase in *Pichia pastoris*. *Appl Microbiol Biot*, 2012, 94(3): 695–704.
- [12] Sandra A, Jozef N, Margit W. High-level expression of *Rhodotorula gracilis* D-amino acid oxidase in *Pichia pastoris*. *Biotechnol Lett*, 2011, 33(3): 557–563.
- [13] Ruth C, Zuellig T, Mellitzer A, et al. Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*. *Syst Synth Biol*, 2010, 4(3): 181–191.
- [14] Li CL, Han L, Zheng ZY, et al. Optimization of plant des-pGlu1-Brazzein gene according to yeasty biased codons and its expression in *Pichia pastoris*. *Chin J Biotech*, 2011, 27(8): 1158–1163 (in

- Chinese).
- 李春丽, 韩露, 郑振宇, 等. 植物甜蛋白 *des-pGlu1-Brazzein* 的密码子优化及其在毕氏酵母中的表达. 生物工程学报, 2011, 27(8): 1158–1163.
- [15] Rene VV, Peter JS, Yvonne M. A codon usage table for *Agaricus bisporus*. Cultivated Mushroom Research (CMR) Newsletter, 1993(1): 50–52.
- [16] Huang XX. Analysis of codon usage bias and promoter cloning of *Tremella fuciformis* [D]. Fuzhou: Fujian Agriculture and Forestry University, 2011 (in Chinese).
- 黄晓星. 银耳密码子偏好性分析及启动子克隆 [D]. 福州: 福建农林大学, 2011.
- [17] Bao DP, Zhao GP, Tan Q, et al. Draft sequence of the *Volvariella volvacea* genome. Acta Edulis Fungi, 2010, 17(1): 1–2 (in Chinese).
- 鲍大鹏, 赵国屏, 谭琦, 等. 草菇全基因组框架图. 食用菌学报, 2010, 17(1): 1–2.
- [18] Paul MS, Elizabeth C. Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast, 1991, 7(7): 657–678.
- [19] Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res, 1987, 15(3): 1281–1295.
- [20] Wu XM, Wu SF, Ren DM, et al. The analysis method and progress in the study of codon bias. Hereditas, 2007, 29(4): 420–426 (in Chinese).
- 吴宪明, 吴松锋, 任大明, 等. 密码子偏性的分析方法及相关研究进展. 遗传, 2007, 29(4): 420–426.
- [21] Fan SH, Guo AG, Shan LW, et al. Analysis of genetic code preference in *Arabidopsis thaliana*. Prog Biochem Biophys, 2003, 30(2): 221–225 (in Chinese).
- 范三红, 郭蔼光, 单丽伟, 等. 拟南芥基因密码子偏爱性分析. 生物化学与生物物理进展, 2003, 30(2): 221–225.
- [22] Perlak FJ, Deaton RW, Armstrong TA, et al. Insect resistant cotton plants. Biol Technol, 1990(8): 939–943.
- [23] Shi XF, Huang JF, Liu SQ, et al. The features of synonymous codon bias and GC-content relationship in human genes. Prog Biochem Biophys, 2002, 29(3): 411–414 (in Chinese).
- 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. 生物化学与生物物理进展, 2002, 29(3): 411–414.
- [24] Chao YE, Wu ZQ, Yang HM, et al. Cluster analysis and codon usage bias studies on *psbA* genes from 11 plant species. Acta Agric Nucl Sin, 2011, 25(5): 927–932 (in Chinese).
- 晁岳恩, 吴政卿, 杨会民, 等. 11 种植物 *psbA* 基因的密码子偏好性及聚类分析. 核农学报, 2011, 25(5): 927–932.

(本文责编 陈宏宇)