

文冠果基因组密码子偏好性分析

雷佳欣,张丽娟,高丹丹,田晓静,刘珊珊*

(西北民族大学 生命科学与工程学院,甘肃 兰州 730030)

摘要:为解析文冠果基因组密码子的偏好性使用模式,提高基因异源表达效率,利用 CodonW 1.4.4 和 Python 程序对文冠果基因组 21,294 条编码序列的密码子进行中性绘图、ENC-plot、PR2-plot 和对应性分析。结果表明,文冠果基因组密码子不同位置上的 GC 含量表现为 $GC_1(50.13\%) > GC_3(40.22\%) > GC_2(42.86\%)$, $RSCU > 1$ 的 29 个密码子中以 A 或 U 结尾的有 25 个,通过 RSCU 确定了 22 个最优密码子,全部以 A/U 结尾,说明文冠果基因组密码子末位碱基较偏好使用 A/U;文冠果基因组密码子使用偏好性主要受到自然选择和突变压力因素的影响。通过比较发现文冠果密码子偏好性与拟南芥、小麦、烟草、杨树、小立碗藓、大肠杆菌及酿酒酵母等物种存在不同程度的差异,其中与拟南芥和烟草差异最小。研究结果不仅可为文冠果基因组学的研究提供理论基础,还为实现文冠果的异源基因表达提供了指导。

关键词:文冠果;基因组;密码子偏好性;最优密码子

中图分类号:S759.33

文献标志码:A

文章编号:1001-7461(2023)04-0104-07

Codon Usage Bias of *Xanthoceras sorbifolium* Genome

LEI Jia-xin, ZHANG Li-juan, GAO Dan-dan, TIAN Xiao-jing, LIU Shan-shan*

(College of Life Sciences and Engineering, Northwest Minzu University, Lanzhou 730030, Gansu, China)

Abstract: To explore the pattern of codon usage bias in *Xanthoceras sorbifolium* genome and to improve heterologous expression efficiency of the genes, we adopted CodonW1.4.4 and Python to conduct neutral plot, ENC-plot, PR2-plot, and correspondence analysis for codons of 21 294 coding sequences in *X. sorbifolium* genome. The results showed that the GC content at different positions of the codons in the *X. sorbifolium* genome was $GC_1(50.13\%) > GC_3(40.22\%) > GC_2(42.86\%)$. Among the 30 codons with $RSCU > 1$, 25 ended with A or U, and 22 optimal codons were determined by RSCU, all of which ended with A/U, indicating that the codon terminal bases of *X. sorbifolium* genome preferred to A/U. The codon usage bias of *X. sorbifolium* genome was mainly affected by natural selection and mutation pressure. The codon usage bias of *X. sorbifolium* genome had differences in various degree with *Arabidopsis thaliana*, *Triticum aestivum*, *Nicotiana tabacum*, *Populus trichocarpa*, *Escherichia coli*, and *Saccharomyces cerevisiae*, the differences with *A. thaliana* and *N. tabacum* were the smallest. Our results not only lay a theoretical basis for the follow-up research on *X. sorbifolium* genome, but also provide guidance for the realization of heterologous gene expression in *X. sorbifolium*.

Key words: *Xanthoceras sorbifolium*; genome; codon usage bias; optimal codons

生物体的遗传信息通过 DNA 复制、转录、翻译等中心法则描述的过程从 DNA 传递到 RNA,最终

形成具有生物功能的蛋白质,其间以 3 个核苷酸为一组的三联体密码转译为蛋白质中氨基酸序列的环

收稿日期:2022-10-09 修回日期:2022-11-23

基金项目:西北民族大学引进人才项目(xbmuyjrc202222, xbmuyjrc202223);中央高校基本科研业务费项目(31920210055)。

第一作者:雷佳欣。研究方向:油料作物遗传育种。E-mail:1979053584@qq.com

*通信作者:刘珊珊,博士,讲师。研究方向:遗传育种与分子生物学。E-mail:liushsh@xbmu.edu.cn

节极其重要。密码子的简并性会导致同义密码子的产生,在基因翻译过程中同义密码子之间的使用频率存在显著差异的现象称为密码子使用偏好性(codon usage bias, CUB)。由于同义突变不改变蛋白质的氨基酸序列,因此对细胞功能、有机体适应性或进化也没有影响。但在进化过程中被选择时,这种没有任何功能后果的同义或“沉默”突变会导致基因组中的同义密码子使用发生变化^[1]。

研究表明生物体进化过程中的突变偏差和自然选择是产生密码子偏好性的重要因素,植物在不同的环境条件下表现出基因表达、生理和应激反应的变化,从而形成独特的密码子偏好性,而同一物种或亲缘关系较近的物种密码子使用模式相似^[2]。基因组的编码序列是基因产物的蓝图,可提供有关基因功能和生物体进化的宝贵信息。因此,开展基因的碱基组成、密码子使用模式以及影响密码子偏好性因素的全基因组研究在确定生物体的基因起源、分子进化、蛋白表达与生产等方面有着重大科学意义^[3]。

文冠果(*Xanthoceras sorbifolium*)属于无患子科(Sapindaceae)文冠果属(*Xanthoceras*),原产于中国北方黄土高原地区,是我国特有的第三纪孑遗油料树种^[4]。文冠果种子含油率达45%~50%,是生物柴油的优质原料;其种仁中丰富的不饱和脂肪酸(约90%)构成了接近食用油的脂肪酸合理配比,因此获得“北方油茶”的美称^[5]。目前,文冠果的全基因组序列信息已经公布^[6],本研究通过生物信息学方法在全基因组水平上分析文冠果基因组的密码子使用偏好性及其影响因素,筛选出最优密码子,以期通过基因工程实现文冠果基因在其他物种中的定点整合和高效表达,加深对文冠果基因功能的了解,为文冠果的基因组学研究及遗传改良提供重要基础。

1 材料与方法

1.1 文冠果基因组数据的获取

文冠果基因组序列及注释信息下载自NCBI基因组数据库(<https://www.ncbi.nlm.nih.gov/data-hub/genome/>)。为减小样本误差,提高密码子偏好性分析的准确度,利用Python工具筛选满足下列条件的编码序列(coding sequence, CDS)序列:(1)序列必须由A、T、C、G四种碱基组成;(2)序列碱基数是3的倍数;(3)具有正确的起始和终止密码子;(4)序列长度 ≥ 300 bp。最终筛选得到了21,294个符合分析条件的CDS序列文件。

1.2 密码子碱基组分析

利用Python编译的程序计算文冠果CDS序列的GC总含量(GC_{all})和密码子第1、第2、第3位碱

基组成中的GC含量(分别用GC₁、GC₂和GC₃表示)及同义密码子第3位的GC含量(用GC_{3s}表示)。利用软件CodonW1.4.4计算相对同义密码子使用度(relative synonymous codon usage, RSCU)、有效密码子数(effective number of codon, ENC)、密码子适应指数(codon adaptation index, CAI)、氨基酸总数(L_{aa})等参数^[7]。

1.3 Neutrality-plot 分析

Neutrality-plot分析分别以GC₃和GC₁₂[(GC₁+GC₂)/2]为横、纵坐标绘制散点图,分析2个参数的相关性。若两者显著相关,说明密码子第1、2、3位碱基组成相似,突变是影响其偏好性的主要因素;若GC₃与GC₁₂相关性不显著,表明自然选择是影响密码子偏好性的主要因素^[7]。

1.4 ENC-plot 绘图分析

运用R语言,以GC_{3s}为横坐标,ENC实际值为纵坐标绘制二维散点图;并根据公式 $ENC = 2 + GC_{3s} + 29 / [GC_{3s}^2 + (1 - GC_{3s})^2]$ 计算ENC期望值,在散点图中绘制ENC值的标准曲线^[8]。使用ENC比值频率(ENC比值=(ENC期望-ENC实际)/ENC期望)可以更加直观地判断影响密码子偏好性的主要因素。

1.5 PR2-plot 分析

密码子第3位碱基(分别用A₃、T₃、G₃、C₃表示)组成偏差对密码子偏好性产生的影响使用PR2-plot(碱基奇偶偏好分析)进行分析。以G₃/(G₃+C₃)为横坐标,以A₃/(A₃+T₃)为纵坐标进行绘图分析。中心点位置表示A=T且C=G,即表明基因2条互补链间不存在任何突变或自然选择上的偏倚,中心点到某个基因点的矢量距表示该基因的偏倚程度和方向^[9]。

1.6 最优密码子分析

根据ENC值对文冠果CDS序列进行排序,分别挑选ENC值最高和最低的10%基因构建高表达和低表达库,并计算出二者的RSCU差值($\Delta RSCU$)。将RSCU值 >1 的密码子定义为高频密码子, $\Delta RSCU$ 值 ≥ 0.08 的密码子定义为高表达密码子,同时满足RSCU值 >1 且 $\Delta RSCU \geq 0.08$ 密码子作为最优密码子^[9]。

1.7 对应分析

对应性分析(correspondence analysis, COA)是一种常用的多变量统计技术,广泛用于分析不同编码序列密码子的变异趋势。使用CodonW对密码子的RSCU值进行对应分析,将样本中所有基因绘制在59维向量空间中,通过分析基因向量以及基因间值的变异情况,可判断密码子偏好性的主要影响

因素^[10]。

1.8 文冠果与其他几种生物密码子偏好性比较

下载常见的异源表达宿主及主要植物代表类群拟南芥(*Arabidopsis thaliana*)、杨树(*Populus trichocarpa*)、烟草(*Nicotiana tabacum*)、小麦(*Triticum aestivum*)、大肠杆菌(*Escherichia coli*)及酿酒酵母(*Saccharomyces cerevisiae*)等物种的基因组密码子使用数据(Codon Usage Database: <http://www.kazusa.or.jp/codon/>),与文冠果基因组密码子使用情况进行比较。当文冠果与其他生物的密码子使用频率的比值 ≥ 2 或 ≤ 0.5 时,表明2种生物的密码子使用偏倚差异显著。

2 结果与分析

2.1 文冠果密码子使用情况分析

文冠果基因组整个编码区的 GC_{all} 含量平均值为 44.48%,其中,密码子第1~3位的 GC 含量平均值分别为 50.13%、40.22%、42.86%,大多数基因3位密码子的 GC 含量呈非均匀分布,分布趋势为 $GC_1 > GC_3 > GC_2$,表明文冠果基因组密码子末位碱基偏好使用 A 或 U(图1)。

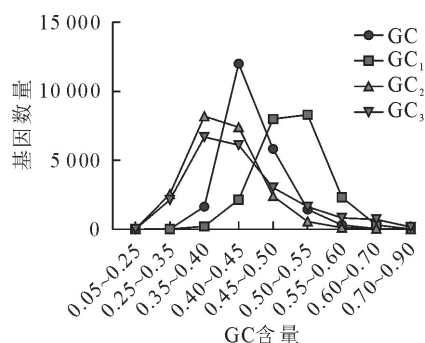


图1 文冠果基因组 GC 含量

Fig. 1 GC content of *X. sorbifolia* genome

RSCU 分析显示,文冠果编码基因中共有 29 个密码子的 $RSCU > 1$,分别为 UGG、AAG、GAA、CAA、CGA、GUG、AGA、UGU、UUU、AGU、UAU、AAU、CGU、GGU、CAU、GGA、UCA、GCA、ACA、UUG、GAU、AUU、ACU、CCA、CCU、UCU、CUU、GUU、GCU,其中以 A 结尾有 9 个,以 U 结尾有 16 个,表明文冠果基因组密码子偏好以 A 或 U 结尾(图2)。

2.2 Neutrality-plot 分析

Neutrality-plot 分析结果如图3所示,图中黑色直线为拟合的线性回归线,该直线的斜率越大,表示密码子偏好性主要受到突变的影响,反之说明自然选择的影响越强。结果显示 GC_{12} 和 GC_3 的取值范围分别在 0.282~0.752 和 0.143~0.87, GC_{12} 与 GC_3 的相关系数为 0.176,相关性较弱,表明大多数

基因的第3位密码子与1、2位密码子的碱基组成具有差异。回归系数为 0.076,斜率较小。因此,自然选择可能是对文冠果基因组密码子偏好性影响较大的因素。

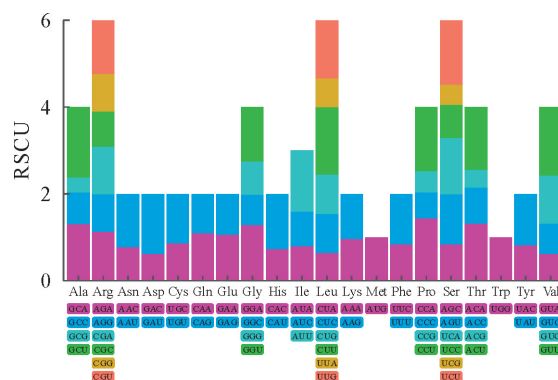


图2 文冠果基因组 RSCU

Fig. 2 RSCU in *X. sorbifolia* genome

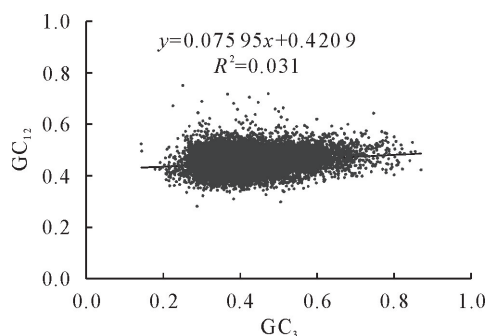


图3 中性绘图分析

Fig. 3 Neutrality plot

2.3 ENC-plot 分析

文冠果的 ENC 值介于 29.37~61,平均值为 52.71。ENC-plot 分析结果显示多数基因位点都偏离标准曲线,说明选择压力是影响文冠果密码子偏好性的主要因素(图4)。ENC 频数分布结果表明,−0.05~0.05 有 6,621 个基因,这些基因的实际 ENC 值与预期值差异小;其余基因都分布在 −0.05~0.05 之外(表1)。这些数据影响文冠果基因组编码序列的密码子偏好性的主要因素是选择压力,部分基因还受突变的影响。

2.4 PR2-plot 分析

PR2-plot 分析结果表明所有基因 $G_3/(G_3 + C_3)$ 、 $A_3/(A_3 + T_3)$ 平均取值的分布位置为(0.533, 0.444),即文冠果基因第3位密码子在总体上 A 与 T、C 与 G 的比例接近。但从图5可看出很多基因位点并非聚集在中心点,大部分基因分布于图中的右下方区域,这表明大部分基因第3位密码子的4种碱基使用频率存在不一致性,即选择压力影响文冠果基因密码子偏好性的同时,其他因素也会产生一定的影响。

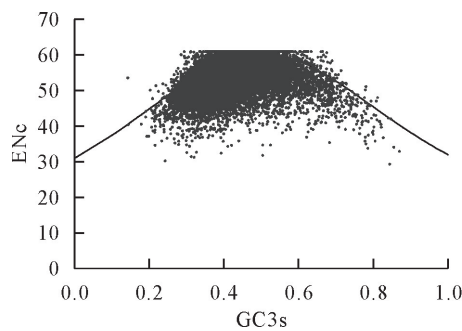


图 4 ENC-plot 分析
Fig. 4 ENC-plot analysis

表 1 ENC 比值频数分布
Table 1 Distribution of ENC ratios

组限	组中值	频数	频率
-0.35~-0.25	-0.3	2	0.000 09
-0.25~-0.15	-0.2	3	0.000 14
-0.15~-0.05	-0.1	219	0.010 28
-0.05~0.05	0.0	6 621	0.310 93
0.05~0.15	0.1	12 254	0.575 47
0.15~0.25	0.2	1 919	0.090 12
0.25~0.35	0.3	255	0.011 98
0.35~0.45	0.4	19	0.000 89
0.45~0.55	0.5	2	0.000 09
合计		21 294	1.000 00

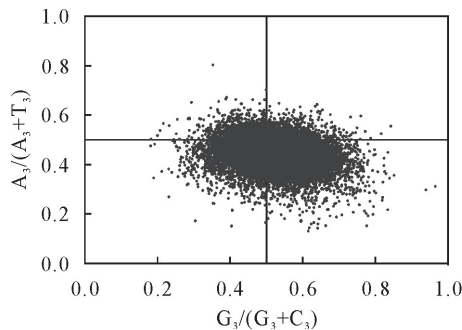


图 5 PR2-plot 分析
Fig. 5 PR2-plot

2.5 最优密码子

根据 RSCU 值和 Δ RSCU 值,最终筛选出了 22 个最优密码子,分别为 GCA、GCU、UGU、GAU、UUU、GGU、CAU、AUU、CUU、AAU、CCA、CCU、CAA、AGA、CGU、AGU、UCA、UCU、ACA、ACU、GUU、UAU,其中以 A 结尾的有 6 个,以 U 结尾的有 16 个(表 2)。

2.6 对应分析

基于 RSCU 值的对应分析结果见图 6,为探究 GC 含量对密码子使用偏好性的影响,按照 GC 含量给不同的基因标记颜色。结果显示 GC 含量 $>60\%$ 的基因主要分布在轴左侧,GC 含量 $<45\%$ 的基因位点较为集中的分布在坐标轴右侧,而 GC 含量在 $45\% \sim 60\%$ 的大部分基因位点则较为分散地分布在

坐标轴中央偏左。对应分析表明第 1 轴占总变异的 14.2% ,第 2、3、4 轴分别存在 5.18% 、 4.97% 和 3.96% 的变异,前 4 个向量轴累计贡献变异 28.31% 。轴 1 是影响密码子使用偏好性的主要因素。为了确定导致文冠果基因沿轴 1 和轴 2 分散的因素,对轴 1 与 ENC、GC_{all}、GC3s 及 CAI 等参数进行相关性分析,发现基因在第 1 轴上的坐标值与 ENC($r = -0.31, P < 0.01$)、GC3s($r = -0.95, P < 0.01$)、GC_{all}($r = -0.82, P < 0.01$)和 CAI($r = -0.32, P < 0.01$)均具有显著的负相关关系。表明选择压力、基因突变均可影响文冠果基因组密码子的使用偏好性,CAI 评估的基因表达水平对其使用模式也具有一定贡献。

2.7 文冠果与其他物种密码子使用情况比较

文冠果基因组密码子使用频率与其他物种进行比较的结果显示,文冠果与拟南芥和烟草等模式物种密码子使用频率比值全部在 $0.5 \sim 2$,表明其密码子使用偏好性相似;而与杨树、酿酒酵母、大肠杆菌及小麦分别存在 1、4、23、20 个使用频率比值 ≥ 2 或 ≤ 0.5 的密码子(图 7),表明文冠果与这些物种的同义密码子偏好性存在一定差异。

3 结论与讨论

密码子末位碱基的同义突变一般不引起氨基酸的变化,相较前两位密码子受到的选择压力小,所以 GC₃ 常被用作分析密码子偏好性的重要参数^[11-12]。文冠果基因组密码子的 GC 和 GC₃ 平均含量分别为 44.48% 和 42.86% ,表明文冠果基因组编码序列偏好使用 A、U 及 A/U 结尾的密码子。这与鸟王茶(*Camellia sinensis*)^[13]、榕树(*Ficus*)^[14]、苦马豆(*Sphaerophysa salsula*)^[15]和睡莲(*Nymphaea*)^[16]等双子叶植物的密码子使用特征研究结果一致,但与小麦(*Triticum aestivum*)^[17]、水稻(*Oryza sativa*)^[18]和玉米(*Zea mays*)^[19]等单子叶植物的密码子使用模式具有差异。

中性绘图分析结果表明,自然选择对塑造文冠果基因组密码子使用偏好性具有重要影响。GC₁₂ 与 GC₃ 之间没有显著相关性,符合分子进化的中性理论^[20]。另外,ENC-plot、PR2-plot 和对应分析等结果都表明文冠果基因组密码子的使用模式形成过程较复杂,除了自然选择和突变压力的作用,基因表达等其他因素对其也有一定影响。虽然目前普遍认为产生密码子使用偏好性的主要原因是自然选择和突变压力,但在不同的物种中影响密码子偏好性形成的主导作用力不同。如橄榄(*Canarium album*)^[21]、唐古特白刺(*Nitraria tangutorum*)^[22]、

表 2 文冠果基因组最优密码子分析
Table 2 Optimal codon analysis of *X. sorbifolium* genome

氨基酸	密码子	高表达库		低表达库		ΔRSCU
		数目	RSCU	数目	RSCU	
Ala	<u>GCA</u> *	13 275	1.272	12 435	1.130	0.142
	GCC	7 164	0.686	10 261	0.933	−0.247
	GCG	3 241	0.311	6 269	0.570	−0.259
	<u>GCU</u> **	18 068	1.731	15 037	1.367	0.364
Cys	UGC	4 945	0.778	6 142	0.986	−0.208
	<u>UGU</u> *	7 772	1.222	6 317	1.014	0.208
Asp	GAC	9 707	0.559	13 633	0.809	−0.250
	<u>GAU</u> *	25 040	1.441	20 073	1.191	0.250
Glu	<u>GAA</u>	22 789	1.047	18 668	0.983	0.064
	GAG	20 730	0.953	19 320	1.017	−0.064
Phe	UUC	10 443	0.778	14 446	0.996	−0.218
	<u>UUU</u> *	16 416	1.222	14 557	1.004	0.218
Gly	<u>GGA</u>	13 353	1.248	12 263	1.177	0.071
	GGC	6 636	0.620	9 144	0.878	−0.258
	GGG	8 270	0.773	8 460	0.812	−0.039
	<u>GGU</u> *	14 552	1.360	11 811	1.134	0.226
His	CAC	4 898	0.646	7 317	0.947	−0.301
	<u>CAU</u> **	10 277	1.354	8 143	1.053	0.301
Ile	AUA	8 167	0.722	9 037	0.791	−0.069
	AUC	8 661	0.765	11 342	0.992	−0.227
	<u>AUU</u> *	17 126	1.513	13 917	1.217	0.296
Lys	AAA	19 928	0.938	17 880	0.937	0.001
	<u>AAG</u>	22 555	1.062	20 283	1.063	−0.001
Leu	CUA	5 141	0.61	6 383	0.643	−0.033
	CUC	7 615	0.903	11 103	1.118	−0.215
	CUG	7 286	0.864	9 640	0.97	−0.106
Leu	<u>CUU</u> **	13 689	1.623	12 610	1.269	0.354
	UUA	7 964	0.622	7 078	0.653	−0.031
	<u>UUG</u>	17 640	1.378	14 600	1.347	0.031
Met	<u>AUG</u>	16 794	1.000	16 071	1.000	0.000
Asn	AAC	10 645	0.710	13 484	0.954	−0.244
	<u>AAU</u> *	19 358	1.290	14 791	1.046	0.244
Pro	<u>CCA</u> *	11 698	1.551	10 060	1.253	0.298
	CCC	3 850	0.510	5 805	0.723	−0.213
	CCG	3 084	0.409	6 213	0.774	−0.365
	<u>CCU</u> *	11 543	1.530	10 039	1.250	0.280
Gln	<u>CAA</u> *	13 308	1.145	12 043	1.059	0.086
	CAG	9 938	0.855	10 696	0.941	−0.086
Arg	<u>AGA</u> *	13 007	1.181	8 633	1.075	0.106
	AGG	9 020	0.819	7 431	0.925	−0.106
Arg	<u>CGA</u>	2 602	1.020	4 290	0.995	0.025
	CGC	2 065	0.809	4 170	0.967	−0.158
	CGG	2 059	0.807	4 109	0.953	−0.146
	<u>CGU</u> *	3 482	1.364	4 681	1.085	0.279
Ser	AGC	6 688	0.782	8 269	0.983	−0.201
	<u>AGU</u> *	10 414	1.218	8 549	1.017	0.201
Ser	<u>UCA</u> *	13 469	1.303	10 739	1.101	0.202
	UCC	7 274	0.703	9 259	0.949	−0.246
	UCG	4 082	0.395	7 032	0.721	−0.326
	<u>UCU</u> **	16 536	1.599	11 997	1.23	0.369
Thr	<u>ACA</u> *	10 000	1.291	9 473	1.14	0.151
	ACC	6 398	0.826	8 550	1.029	−0.203
	ACG	2 732	0.353	5 084	0.612	−0.259
	<u>ACU</u> **	11 859	1.531	10 124	1.219	0.312
Val	GUA	5 985	0.552	6 483	0.595	−0.043
	GUC	6 660	0.614	9 591	0.880	−0.266
	<u>GUG</u>	12 606	1.162	13 056	1.198	−0.036
	<u>GUU</u> **	18 127	1.672	14 464	1.327	0.345
Trp	<u>UGG</u>	7 742	1.000	8 684	1.000	0.000
Tyr	UAC	6 654	0.765	9 409	0.988	−0.223
	<u>UAU</u> *	10 742	1.235	9 635	1.012	0.223

注：下划线表示基因 RSCU>1 的密码子，“*”代表 ΔRSCU≥0.08，“**”代表 ΔRSCU>0.3，加粗的密码子为最优密码子。

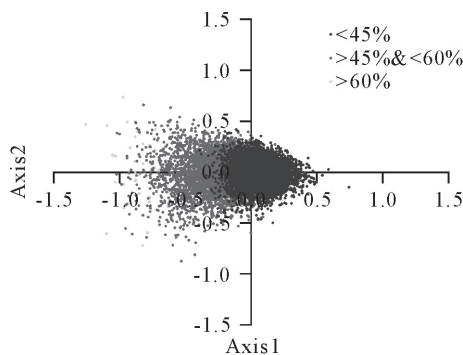
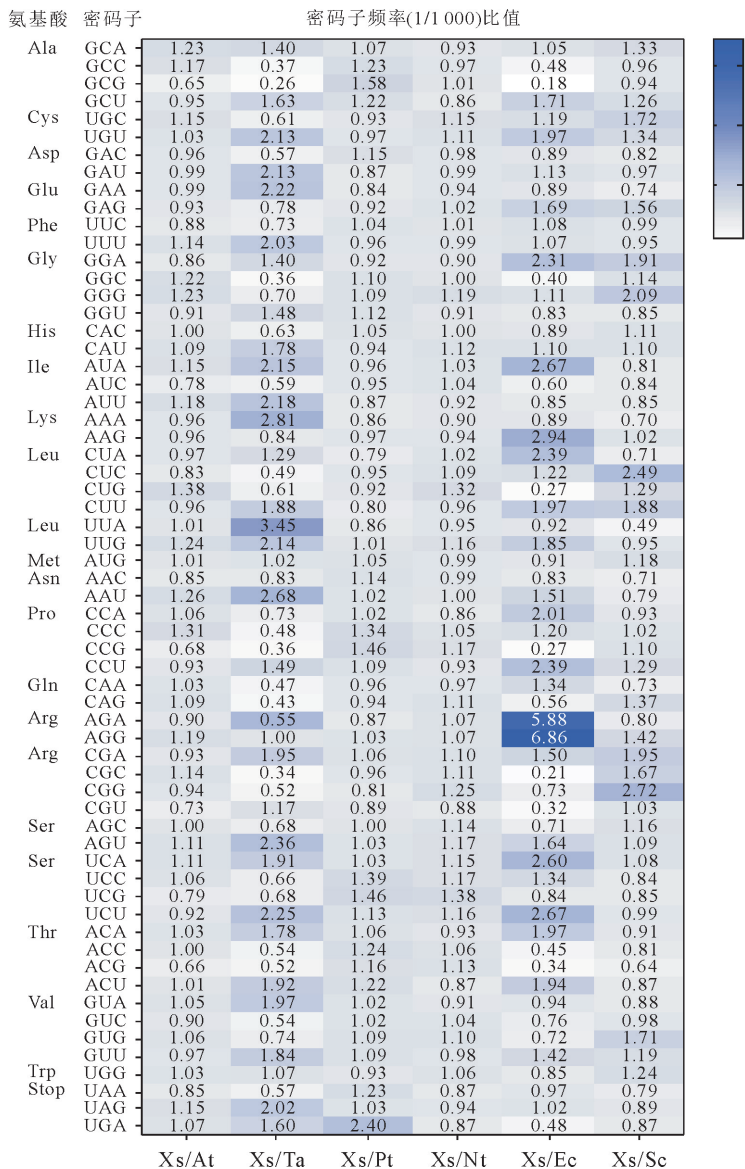


图 6 RSCU 的对应分析

Fig. 6 Corresponding analysis based on RSCU

西南桦 (*Betula alnoides*)^[23]、薏苡 (*Coix lachryma*)^[24] 等物种的密码子使用模式主要受到碱基突变压力的影响;而在榕树 (*Ficus*)^[14]、银杏 (*Ginkgo biloba*)^[25] 和川桑 (*Morus notabilis*)^[26] 中,自然选择对其的密码子使用模式起主导作用,受突变压力的影响较小。

基于 RSCU 分析共筛选出 22 个符合条件的最优密码子,其中以 A 结尾的有 6 个,以 U 结尾的有 16 个,密码子使用模式为 NNA、NNU。此外,密码子使用频率比较结果显示,文冠果与拟南芥、杨树和烟草等双子叶植物的密码子使用偏好性相似,但和



Xs:文冠果 *X. sorbifolium*; At:拟南芥 *A. thaliana*; Ta:小麦 *T. aestivum*; Pt:杨树 *P. trichocarpa*; Nt:烟草 *N. tabacum*; Ec:大肠杆菌 *E. coli*; Sc:酿酒酵母 *S. cerevisiae*。

图 7 文冠果与其他物种密码子偏好性比较

Fig. 7 Comparison of codon usage baise between *X. sorbifolium* and other species

单子叶植物小麦的密码子使用模式差异较大。因此,拟南芥和烟草更适合作为验证文冠果功能基因时的遗传转化的载体。与常用的异源表达宿主大肠杆菌和酿酒酵母相比,文冠果密码子偏好与酿酒酵母相似而而大肠杆菌差异较大。文冠果基因在外源系统中表达时,可以选择在酵母表达系统中进行,若

在大肠杆菌等系统中表达时,可适当根据密码子偏好性进行碱基的优化,从而提高表达量。

本研究以文冠果基因组为数据来源,揭示了文冠果基因组密码子使用特征及最优密码子,这些结果对选择文冠果基因异源表达的受体及表达系统提供了指导,对进一步阐明文冠果基因组功能、解析文冠果的进化过程也具有重要意义。

参考文献:

- [1] SUPEK F. The code of silence; widespread associations between synonymous codon biases and gene function[J]. Journal of Molecular Evolution, 2016, 82(1): 65-73.
- [2] PARVATHY S T, UDAYASURIYAN V, BHADANA V. Codon usage bias[J]. Molecular Biology Reports, 2022, 49: 539-565.
- [3] CHANEY J L, CLARK P L. Roles for synonymous codon usage in protein biogenesis[J]. Annual Review of Biophysics, 2015, 44: 143-166.
- [4] 贺晓慧, 宁小莉, 郭彦龙, 等. 我国文冠果分布的气候地理特征及未来气候情景下适宜种植区预测[J]. 农业现代化研究, 2019, 40(2): 316-324.
- [5] 高述民, 马凯, 杜希华, 等. 文冠果 (*Xanthoceras sorbifolia*) 研究进展[J]. 植物学通报, 2002, 19(3): 296-301.
GAO S M, MA K, DU X H, *et al.* Advances in research on *Xanthoceras sorbifolia* [J]. Chinese Bulletin of Botany, 2002, 19 (3): 296-301. (in Chinese)
- [6] LIU H, YAN X M, WANG X R, *et al.* Centromere-specific retrotransposons and very-long-chain fatty acid biosynthesis in the genome of yellowhorn (*Xanthoceras sorbifolium*, Sapindaceae), an oil-producing tree with significant drought resistance[J]. Frontiers in Plant Science, 2021, 12: 766389.
- [7] 周显臻, 曹支敏, 于丹. 落叶松-杨栅锈菌基因组密码子使用偏好分析[J]. 菌物学报, 2020, 39(2): 289-297.
- [8] WRIGHT F. The 'effective number of codons' used in a gene [J]. Gene, 1990, 87(1): 23-29.
- [9] 李江飞, 王瑜, 颜廷雨, 等. 云南油杉叶绿体基因组密码子偏好性分析[J]. 中南林业科技大学学报, 2022, 42(4): 30-39.
LI J F, WANG Y, YAN T Y, *et al.* Analysis on codon usage bias of *Keteleeria evelyniana* chloroplast genome[J]. Journal of Central South University of Forestry & Technology, 2022, 42 (4): 30-39. (in Chinese)
- [10] PERRIÈRE G, THIOULOUSE J. Use and misuse of correspondence analysis in codon usage studies[J]. Nucleic Acids Research, 2002, 30(20): 4548-4555.
- [11] LIU Q, FENG Y, XUE Q. Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa* [J]. Mitochondrion, 2004, 4(4): 313-320.
- [12] LIU Q, XUE Q. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species[J]. Journal of Genetics, 2005, 84(1): 55-62.
- [13] 滕腾, 赵懿琛, 赵德刚. 鸟王茶基因密码子偏好性分析[J]. 基因组学与应用生物学, 2021, 40(2): 795-801.
TENG T, ZHAO Y C, ZHAO D G. Analysis of genetic codon preference of *Camellia sinensis* var. *niaowangensis* [J]. Genomics and Applied Biology, 2021, 40(2): 795-801. (in Chinese)
- [14] 夏晞, 彭劲谕, 王大玮, 等. 3种榕属叶绿体基因组密码子偏好性分析[J]. 西北林学院学报, 2022, 37(5): 88-94.
XIA X, PENG J Y, WANG D W, *et al.* Codon usage bias in chloroplast genomes of three *Ficus* species [J]. Journal of Northwest Forestry University, 2022, 37(5): 88-94. (in Chinese)
- [15] 梁湘兰, 郭松. 苦马豆叶绿体基因组密码子偏好性分析[J]. 西北林学院学报, 2022, 37(2): 121-126.
LIANG X L, GUO S. Codon usage bias in the chloroplast genome of *Sphaerophysa salsula* [J]. Journal of Northwest Forestry University, 2022, 37(2): 121-126. (in Chinese)
- [16] 毛立彦, 黄秋伟, 龙凌云, 等. 7种睡莲属植物叶绿体基因组密码子偏好性分析[J]. 西北林学院学报, 2022, 37(2): 98-107.
MAO L Y, HUANG Q W, LONG L Y, *et al.* Comparative analysis of codon usage bias in chloroplast genomes of seven *Nymphaea* species [J]. Journal of Northwest Forestry University, 2022, 37(2): 98-107. (in Chinese)
- [17] ZHANG W J, ZHOU J, LI Z F, *et al.* Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L [J]. Journal of Integrative Plant Biology, 2007, 49(2): 246-254.
- [18] LIU Q, FENG Y, DONG H, *et al.* Synonymous codon usage bias in *Oryza sativa* [J]. Plant Science, 2004, 167 (1): 101-105.
- [19] FENNOY S L, BAILEY-SERRES J. Synonymous codon usage in *Zea mays* L. nuclear genes is varied by levels of C and G-ending codons [J]. Nucleic Acids Research, 1993, 21(23): 5294-5300.
- [20] SHARP P M, STENICO M, PEDEN J F, *et al.* Codon usage: mutational bias, translational selection, or both? [J]. Biochemical Society Transactions, 1993, 21(4): 835-841.
- [21] 赖瑞联, 冯新, 陈瑾, 等. 橄榄转录组密码子使用偏好性及其影响因素[J]. 核农学报, 2019, 33(1): 31-38.
LAI R L, FENG X, CHEN J, *et al.* Codon usage preference of chloroplast genome of *Canarium album* [J]. Journal of Nuclear Agriculture, 2019, 33 (1): 31-38. (in Chinese)
- [22] 张得芳, 夏涛. 基于转录组的唐古特白刺密码子偏好性分析[J]. 干旱区资源与环境, 2021, 35(12): 163-170.
- [23] 原晓龙, 李云琴, 王毅, 等. 西南桦叶绿体基因组密码子偏好性分析[J]. 基因组学与应用生物学, 2020, 39(12): 5758-5764.
YUAN X L, LI Y Q, WANG Y, *et al.* Analysis of codon usage in the chloroplast genome of *Betula alnoides* [J]. Genomics and Applied Biology, 2020, 39(12): 5758-5764. (in Chinese)
- [24] 张重菊, 赵懿琛, 赵德刚. 薏苡转录组中密码子使用偏好性分析[J]. 种子, 2022, 41(4): 13-19.
ZHANG Z J, ZHAO Y C, ZHAO D G. Analysis of codons use preference of transcriptome in *Coix lachryma-jobi* L. [J]. Seed, 2022, 41(4): 13-19. (in Chinese)
- [25] HE B, DONG H, JIANG C, *et al.* Analysis of codon usage patterns in *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C-ending [J]. Scientific Reports, 2016, 6 (1): 1-11.
- [26] WEN Y, ZOU Z, LI H, *et al.* Analysis of codon usage patterns in *Morus notabilis* based on genome and transcriptome data [J]. Genome, 2017, 60(6): 473-484.