

包国媛,李文辛,杨鑫光,等. 海甜菜线粒体和叶绿体基因组密码子使用偏好性分析[J]. 江苏农业学报, 2023, 39(9): 1804-1817.
doi:10.3969/j.issn.1000-4440.2023.09.002

海甜菜线粒体和叶绿体基因组密码子使用偏好性分析

包国媛¹, 李文辛¹, 杨鑫光¹, 王雅琼^{1,2,3,4}

(1.青海民族大学生态环境与资源学院,青海 西宁 810007; 2.青海省特色经济植物高值化利用重点实验室,青海 西宁 810007; 3.青海省生物技术与分析测试重点实验室,青海 西宁 810007; 4.青藏高原资源化学与生态环境保护国家民委重点实验室,青海 西宁 810007)

摘要: 为明确海甜菜线粒体和叶绿体基因组密码子使用偏好性及影响因素,以海甜菜线粒体基因组 136 条蛋白编码序列(CDS)和叶绿体基因组 52 条蛋白编码序列为数据源,利用 Codon W 和 CUSP 软件,结合密码子使用的中性绘图分析、有效密码子数分析(ENC-plot)、偏倚性分析(PR2-plot),探究海甜菜线粒体和叶绿体基因组密码子使用的偏好性,并明确最优密码子。结果显示:海甜菜线粒体和叶绿体基因组中所有密码子的平均 G+C 碱基含量分别为 43.42% 和 37.92%;密码子第 3 位碱基多为 A 或 U。线粒体和叶绿体基因组有效密码子数(ENC)分别为 36.44~61.00 和 35.00~52.01,平均值为 52.23 和 46.10。自然选择是影响海甜菜线粒体和叶绿体基因组密码子使用偏好性的主要原因;GCU、AGG、GCU、CGA 等 27 个密码子可视为最优密码子。本研究结果为海甜菜线粒体和叶绿体基因组的研究和遗传多样性分析提供了参考依据。

关键词: 海甜菜; 线粒体基因组; 叶绿体基因组; 密码子; 使用偏好性

中图分类号: S556.3 **文献标识码:** A **文章编号:** 1000-4440(2023)09-1804-14

Analysis of codon usage bias in mitochondrial and chloroplast genomes of *Beta vulgaris* subsp.

BAO Guo-yuan¹, LI Wen-xin¹, YANG Xin-guang¹, WANG Ya-qiong^{1,2,3,4}

(1.The College of Ecological Environment and Resources, Qinghai Minzu University, Xining 810007, China; 2.Qinghai Provincial Key Laboratory of High-value Utilization of Characteristic Economic Plants, Xining 810007, China; 3.Qinghai Provincial Biotechnology and Analytical Testing Key Laboratory, Xining 810007, China; 4.Key Laboratory of Resource Chemistry and Eco-environmental Protection in Qinghai-Tibet Plateau, State Ethnic Affairs Commission, Xining 810007, China)

Abstract: In order to clarify the codon usage bias and influencing factors of the mitochondrial and chloroplast genomes of *Beta vulgaris* subsp., 136 protein coding sequences (CDS) of the mitochondrial genome and 52 protein coding sequences of the chloroplast genome of *Beta vulgaris* subsp. were used as data source. Using Codon W and CUSP software, combined with the neutral mapping analysis of codon usage, effective number of

收稿日期:2023-03-29

基金项目: 青海民族大学研究生创新项目(54M2022005);青海民族大学研究生精品示范课程项目(JK-2023-08);2023 年中央科研创新平台建设项目——青海省林业草原生态系统功能维护及可持续开发利用科研创新团队(30160101141);青海民族大学 2023 年大学生创新创业训练计划项目

作者简介: 包国媛(1997-),女,青海互助人,硕士研究生,研究方向为植物资源开发利用。(E-mail)1791564133@qq.com

通讯作者: 王雅琼, (Tel) 18297178818; (E-mail) wangyaqiong727@163.com

codons (ENC-plot) and bias analysis (PR2-plot), the codon usage bias in mitochondrial and chloroplast genomes of *Beta vulgaris* subsp. was explored, and the optimal codon was determined. The results showed that the average G+C content of all codons in the mitochondrial and chloroplast genomes of *Beta vulgaris* subsp. was 43.42% and 37.92%, respectively. The third base of the codon was mostly A or U. The effective number of codons (ENC) of mitochondrial and chloroplast genomes was 36.44~61.00 and 35.00~52.01, respectively, with an average of 52.23 and 46.10. Natural selection was the

main reason affecting the codon usage preference of mitochondrial and chloroplast genomes in *Beta vulgaris* subsp.. Twenty-seven codons, such as GCU, AGG, GCU and CGA, could be regarded as the optimal codons. The results of this study provide a reference for the study of mitochondrial and chloroplast genomes and genetic diversity analysis of *Beta vulgaris* subsp.

Key words: *Beta vulgaris* subsp.; mitochondrial genome; chloroplast genome; codon; usage bias

密码子偏好性是生物基因组进化的一个重要特征^[1],植物密码子偏好性研究对探究植物分子进化和外源蛋白质表达有着重要的意义^[2]。植物在进化过程中密码子的使用会受到环境选择、碱基突变、基因漂变等因素的影响^[3-4]。作为传递生物遗传信息的序列单位,密码子具有简并性,常出现2个或多个密码子编码同一种氨基酸的现象^[5]。编码同一种氨基酸的不同密码子称为同义密码子(synonymous codon)^[6]。64个密码子编码20种氨基酸和3个终止密码子^[6]。同义密码子在不同基因组中使用的频率并不相同,甚至在同一基因组的不同基因中也不相同,这种不平等使用同义密码子的现象被称为同义密码子使用偏好性^[7]。

海甜菜(*Beta vulgaris* subsp.)是所有驯化甜菜的祖先^[8-9],通常被称为野生甜菜,起源于马德拉群岛^[10-11]。海甜菜亚种主要有4个,而全部的栽培种属于*Beta vulgaris* subsp. *vulgaris* 亚种^[12-13]。海甜菜种群在干旱和盐碱化高的地区有较好的适应性^[14]。中国海甜菜栽培种(甜菜)主要分布在东北、西北和华北等地区^[15]。海甜菜根富含多种营养成分、维生素和生物活性物质,具有较高的抗氧化性能,在食品生产中得到了较多的应用^[16-17]。

叶绿体和线粒体是植物能量利用和转化的主要细胞器^[18-19]。线粒体基因组具有重排进化快、DNA插入容易等特点^[20-21],而叶绿体基因组具有较小的进化率和较大的保守性,基因拷贝数较多、单亲遗传等特点^[22]。目前,对植物线粒体基因组的研究主要侧重于基因结构与功能、基因表达水平调控、核质互作及线粒体的起源进化等领域^[23-24],对叶绿体基因组的研究侧重植物叶绿体遗传进化等方面^[25]。

目前海甜菜完整线粒体和叶绿体基因组已经公布^[26],但海甜菜细胞器基因组密码子使用偏好性还缺乏分析。本研究利用Codon W和CUSP软件和海甜菜线粒体和叶绿体细胞器基因组的蛋白质编码序列(CDS),结合密码子使用的中性绘图分析、有效密码子数(ENC-plot)分析、偏倚性(PR2-plot)分析,明确海甜菜细胞器基因组密码子使用的偏好性及最优

密码子。以期在海甜菜基因资源的利用和其他相关研究提供依据。

1 材料和方法

1.1 海甜菜细胞器基因组序列的获取

海甜菜线粒体和叶绿体基因组完整序列均来自NCBI数据库(GenBank接收号为:FP885845.1和ON641300.1)。依据密码子偏好性分析的序列选择标准^[27],在基因编码序列(CDS)中,选择长度>300 bp,以ATG开始,TAG、TAA、TGA结尾,内部不存在终止密码子及重复序列的编码序列进行分析。

1.2 密码子碱基组成分析

利用Codon W软件分析线粒体和叶绿体基因组中符合密码子偏好性分析条件的CDS序列,获得各CDS的有效密码子数(ENC)、同义密码子相对使用度(RSCU)^[28]。利用CUSP在线软件(<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>)分析密码子第1位碱基G+C含量(GC_1)、第2位碱基G+C含量(GC_2)、第3位碱基G+C含量(GC_3)、3位碱基平均G+C含量(GC_{all})、第3位碱基上A、T、C、G碱基的含量 A_3 、 T_3 、 C_3 、 G_3 及同义密码子第3位碱基G+C含量(GC_{3s})等参数^[29-30]。

1.3 中性绘图分析

以线粒体和叶绿体基因密码子 GC_1 和 GC_2 平均值 GC_{12} 为纵坐标, GC_3 为横坐标制作中性图^[31]。若 GC_{12} 与 GC_3 之间显著相关,回归曲线斜率接近于1,表明密码子偏好性的成因主要源于碱基突变,反之则说明密码子偏好性受到自然选择影响较大^[32]。

1.4 ENC-plot 分析

以线粒体和叶绿体基因密码子 GC_3 为横坐标,ENC为纵坐标绘制ENC-plot图。以公式 $ENC_{exp} = 2 + GC_3 + 29/[GC_3 + (1 - GC_3)^2]$ ^[33]绘制标准曲线。以 $(ENC_{exp} - ENC_{obs})/ENC_{exp}$ 计算各基因有效密码子比值(ENC_{obs} 为各基因实际ENC),并进行海甜菜线粒体和叶绿体基因组有效密码子比值频数分布分析。基因密码子ENC取值范围为20~61。当ENC取值为20时,表示每个氨基酸只使用一个密码子,为极

端偏好;*ENC* 取值为 61 时,表示密码子偏向随机使用,不存在使用偏好^[34-35]。

1.5 海甜菜线粒体和叶绿体基因密码子偏倚分析 (PR2-plot)

以海甜菜线粒体和叶绿体基因组 $G_3/(G_3+C_3)$ 为横坐标, $A_3/(A_3+U_3)$ 为纵坐标,制作基因密码子偏倚分析图(PR2-plot)^[28]。PR2-plot 图中心点代表无偏性使用时的密码子状态,即 $A=U$ 且 $C=G$,其余点与中心点的矢量距离则代表各基因密码子碱基偏倚程度和方向^[36]。

1.6 最优密码子的确定

以海甜菜线粒体和叶绿体各基因的 *ENC* 排序,从前后两端各选取 10% 的基因,分别建立高表达基因库和低表达基因库,利用软件 Codon W 计算高表达和低表达基因同义密码子相对使用度(*RSCU*)及其差值($\Delta RSCU$),定义 $\Delta RSCU \geq 0.08$ 且高表达基因 $RSCU \geq 1.00$ 的密码子为最优密码子^[37-38]。

2 结果与分析

2.1 海甜菜密码子特征分析

本研究共筛选得到 136 个海甜菜线粒体基因编码序列(CDS)和 52 个叶绿体基因编码序列。线粒体基因组密码子 3 位碱基平均 G+C 含量(GC_{all})为 43.42%, GC_1 、 GC_2 、 GC_3 及 GC_{3S} 分别为 46.98%、41.21%、42.06%及 42.23%(表 1)。叶绿体基因组密码子 3 位碱基平均 G+C 含量(GC_{all})为 37.92%, GC_1 、 GC_2 、 GC_3 及 GC_{3S} 分别为 46.67%、39.11%、27.97%及 28.13%(表 2)。线粒体基因组 *ENC* 的均值为 52.23,介于 36.44~61.00, $ENC > 45$ 的基因有 121 个;叶绿体基因组有效密码子数(*ENC*)介于 35.00~52.01,均值为 46.10, $ENC > 45$ 的基因有 35 个。本研究中海甜菜线粒体基因组 *ENC* 均值为 52.23,叶绿体均值为 46.10,更加接近 61,说明海甜菜线粒体和叶绿体基因组密码子偏好性较弱。

表 1 海甜菜线粒体基因编码序列(CDS)密码子各位置的 G+C 含量及有效密码子数

Table 1 The G+C content and the effective number of codons at each position of mitochondrial gene coding sequence (CDS) codons in *Beta vulgaris* subsp.

基因	密码子数	GC_1 (%)	GC_2 (%)	GC_3 (%)	GC_{all} (%)	有效密码子数	GC_{3S} (%)
<i>atp1</i>	507	58.19	41.22	35.90	45.10	53.97	36.30
<i>atp4</i>	199	47.24	41.71	36.68	41.88	61.00	37.00
<i>atp8</i>	163	35.58	36.81	41.10	37.83	50.12	42.30
<i>ccmB</i>	207	43.48	41.55	34.78	39.94	54.34	35.00
<i>ccmFC</i>	439	48.75	43.51	43.28	45.18	56.63	44.00
<i>ccmFN</i>	578	49.65	45.85	40.14	45.21	53.62	40.10
<i>cob</i>	394	48.98	40.36	33.25	40.86	50.50	33.40
<i>cox1</i>	525	47.24	42.48	36.95	42.22	51.00	37.20
<i>cox2</i>	261	50.57	37.93	30.65	39.72	45.74	30.40
<i>cox3</i>	266	50.75	41.35	32.33	41.48	51.79	32.10
<i>mat-r</i>	656	51.98	43.29	56.71	50.66	55.94	57.10
<i>nad2</i>	489	41.10	41.92	31.70	38.24	47.30	31.50
<i>nad3</i>	119	42.02	40.34	33.61	38.66	52.17	33.60
<i>nad4</i>	496	43.95	38.31	34.88	39.05	52.16	35.50
<i>nad6</i>	206	41.75	36.41	37.86	38.67	50.99	37.90
<i>nad7</i>	392	56.12	42.09	31.89	43.37	46.01	32.30
<i>nad9</i>	193	49.74	37.82	32.12	39.90	52.58	32.30
<i>orf100a</i>	101	41.58	34.65	24.75	33.66	38.20	23.40
<i>orf100g</i>	101	51.49	37.62	42.57	43.89	42.00	43.20
<i>orf102a</i>	103	49.51	52.43	49.51	50.49	55.93	49.00
<i>orf102e</i>	103	52.43	44.66	42.72	46.60	55.42	41.80
<i>orf103a</i>	104	36.54	38.46	45.19	40.06	48.37	45.80
<i>orf103b</i>	104	50.00	52.88	41.35	48.08	42.81	41.20
<i>orf103c</i>	104	39.42	40.38	48.08	42.63	61.00	48.50

续表 1 Continued 1

基因	密码子数	GC_1 (%)	GC_2 (%)	GC_3 (%)	GC_{all} (%)	有效密码子数	GC_{3S} (%)
<i>orf104a</i>	105	47.62	42.86	46.67	45.71	55.88	47.50
<i>orf105a</i>	106	37.74	56.60	60.38	51.57	55.75	60.20
<i>orf105b</i>	106	49.06	51.89	51.89	50.94	48.26	52.00
<i>orf105c</i>	106	51.89	48.11	40.57	46.86	49.85	41.60
<i>orf105d</i>	106	43.40	42.45	47.17	44.34	54.32	46.60
<i>orf106a</i>	107	45.79	43.93	58.88	49.53	42.22	57.70
<i>orf106b</i>	107	42.06	38.32	54.21	44.86	61.00	55.80
<i>orf106c</i>	107	51.40	36.45	46.73	44.86	61.00	48.50
<i>orf107a</i>	108	47.22	38.89	44.44	43.52	46.06	43.80
<i>orf107b</i>	108	47.22	56.48	56.48	53.40	58.48	56.40
<i>orf107c</i>	108	58.33	40.74	58.33	52.47	53.95	57.30
<i>orf108a</i>	109	38.53	42.20	35.78	38.84	40.54	35.20
<i>orf108b</i>	109	39.45	43.12	49.54	44.04	52.11	50.00
<i>orf108c</i>	109	52.29	45.87	29.36	42.51	51.83	29.00
<i>orf109a</i>	110	35.45	56.36	32.73	41.52	44.33	33.00
<i>orf109c</i>	110	40.00	40.91	45.45	42.12	53.36	46.20
<i>orf110b</i>	111	48.65	44.14	44.14	45.65	61.00	45.40
<i>orf111a</i>	112	44.64	44.64	47.32	45.54	56.64	47.70
<i>orf112b</i>	113	38.94	44.25	41.59	41.59	43.09	40.40
<i>orf113a</i>	114	55.26	36.84	50.00	47.37	59.50	50.00
<i>orf114a</i>	115	54.78	46.09	38.26	46.38	57.27	38.70
<i>orf114b</i>	115	43.48	38.26	50.43	44.06	61.00	51.80
<i>orf114c</i>	115	45.22	35.65	40.00	40.29	51.18	41.70
<i>orf114d</i>	115	46.96	45.22	35.65	42.61	38.32	36.00
<i>orf115a</i>	116	41.38	46.55	37.07	41.67	47.44	37.80
<i>orf115b</i>	116	45.69	39.66	45.69	43.68	48.11	44.20
<i>orf115c</i>	116	36.21	34.48	35.34	35.34	51.19	35.70
<i>orf116a</i>	117	52.14	38.46	40.17	43.59	42.98	41.10
<i>orf116b</i>	117	43.59	50.43	35.90	43.30	50.97	36.20
<i>orf119a</i>	120	35.83	33.33	52.50	40.56	45.35	53.90
<i>orf119b</i>	120	28.33	45.83	26.67	33.61	45.76	25.60
<i>orf120b</i>	121	42.98	34.71	39.67	39.12	39.35	40.20
<i>orf121</i>	122	46.72	48.36	48.36	47.81	58.31	48.70
<i>orf122</i>	123	47.15	43.09	52.85	47.70	55.07	53.80
<i>orf122b</i>	123	59.35	54.47	52.85	55.56	51.99	52.20
<i>orf124</i>	125	52.00	43.20	41.60	45.60	60.54	41.00
<i>orf125a</i>	126	45.24	65.87	57.14	56.08	57.78	57.40
<i>orf125b</i>	126	43.65	45.24	39.68	42.86	61.00	40.50
<i>orf131</i>	132	49.24	47.73	46.21	47.73	57.74	45.80
<i>orf133b</i>	134	47.01	43.28	36.57	42.29	54.06	37.40
<i>orf133c</i>	134	56.72	48.51	41.04	48.76	56.07	39.70
<i>orf134a</i>	135	46.67	40.74	36.30	41.23	52.25	37.40
<i>orf134c</i>	135	54.07	45.93	48.89	49.63	58.36	48.10
<i>orf135a</i>	136	44.12	29.41	35.29	36.27	50.55	34.60
<i>orf136b</i>	137	45.99	37.96	43.80	42.58	55.21	44.40
<i>orf137</i>	138	55.80	34.78	53.62	48.07	49.06	55.40
<i>orf138</i>	139	46.76	48.92	45.32	47.00	51.90	45.40

续表 1 Continued 1

基因	密码子数	GC ₁ (%)	GC ₂ (%)	GC ₃ (%)	GC _{all} (%)	有效密码子数	GC _{3S} (%)
<i>orf140</i>	141	51.06	46.81	41.13	46.34	49.15	41.90
<i>orf143a</i>	144	55.56	49.31	44.44	49.77	53.08	43.60
<i>orf143b</i>	144	45.83	31.25	36.11	37.73	38.49	35.90
<i>orf145d</i>	146	58.22	59.59	59.59	59.13	51.30	60.30
<i>orf146</i>	147	48.98	38.10	44.22	43.76	55.25	45.30
<i>orf147b</i>	148	50.00	35.14	35.14	40.09	52.19	35.70
<i>orf148</i>	149	42.95	34.23	36.24	37.81	47.80	35.90
<i>orf152a</i>	153	45.10	44.44	47.71	45.75	54.35	47.30
<i>orf152b</i>	153	37.91	32.03	48.37	39.43	44.25	47.70
<i>orf155b</i>	156	55.77	44.87	34.62	45.09	48.51	34.90
<i>orf162</i>	163	44.17	28.83	28.22	33.74	43.79	28.10
<i>orf166</i>	167	43.11	44.91	43.11	43.71	55.80	44.70
<i>orf169a</i>	170	49.41	53.53	54.12	52.35	49.90	53.90
<i>orf169c</i>	170	47.65	45.88	45.29	46.27	55.48	46.30
<i>orf174b</i>	175	44.57	41.14	42.29	42.67	61.00	42.90
<i>orf175</i>	176	44.89	36.93	36.36	39.39	45.32	36.30
<i>orf176a</i>	177	52.54	54.80	55.37	54.24	54.22	56.50
<i>orf178b</i>	179	45.81	38.55	39.11	41.15	53.69	39.00
<i>orf187a</i>	188	38.83	35.11	32.45	35.46	48.14	32.40
<i>orf189</i>	190	43.68	42.11	42.63	42.81	46.37	42.80
<i>orf192</i>	193	44.56	36.27	39.90	40.24	52.94	40.40
<i>orf193</i>	194	50.00	34.54	41.24	41.92	56.72	41.60
<i>orf198a</i>	199	60.30	47.24	53.77	53.77	61.00	52.90
<i>orf199</i>	200	51.50	34.50	38.50	41.50	53.65	38.50
<i>orf204</i>	205	36.59	36.59	33.17	35.45	58.26	32.70
<i>orf208</i>	209	44.98	44.02	44.02	44.34	58.21	43.60
<i>orf217a</i>	218	40.83	34.40	45.87	40.37	53.66	46.20
<i>orf217c</i>	218	37.16	36.70	43.58	39.14	56.26	45.00
<i>orf224a</i>	225	47.11	29.33	39.56	38.67	55.02	39.40
<i>orf227</i>	228	51.75	35.96	42.54	43.42	61.00	42.20
<i>orf238</i>	239	43.51	40.59	36.82	40.31	44.90	36.10
<i>orf249</i>	250	50.00	43.20	44.40	45.87	59.59	43.70
<i>orf251</i>	251	43.03	32.27	28.69	34.66	48.13	28.90
<i>orf256</i>	257	49.81	42.02	45.53	45.78	60.46	46.00
<i>orf270</i>	271	51.29	35.42	50.55	45.76	57.43	50.90
<i>orf273</i>	274	48.18	41.61	43.80	44.53	56.34	43.30
<i>orf281</i>	282	49.65	31.56	41.49	40.90	61.00	41.80
<i>orf282</i>	283	55.12	36.75	48.76	46.88	50.28	48.70
<i>orf288</i>	289	44.29	40.14	35.99	40.14	52.57	36.70
<i>orf290</i>	291	46.05	34.02	44.67	41.58	57.10	44.30
<i>orf294</i>	295	53.56	46.10	47.80	49.15	57.50	48.6
<i>orf300</i>	301	48.17	49.83	38.87	45.63	53.46	38.60
<i>orf310a</i>	311	44.69	32.80	40.51	39.34	50.43	41.00
<i>orf315</i>	316	52.85	37.03	53.80	47.89	51.08	53.10
<i>orf317</i>	318	56.29	41.82	49.06	49.06	61.00	48.90
<i>orf329a</i>	330	48.18	33.33	39.09	40.20	52.51	38.80
<i>orf348</i>	349	41.26	29.80	45.56	38.87	53.45	46.30

续表 1 Continued 1

基因	密码子数	GC_1 (%)	GC_2 (%)	GC_3 (%)	GC_{all} (%)	有效密码子数	GC_{3S} (%)
<i>orf352</i>	353	49.29	32.86	41.36	41.17	48.94	40.90
<i>orf393</i>	394	34.26	31.98	38.83	35.03	53.33	39.60
<i>orf399</i>	400	42.75	42.75	35.00	40.17	49.55	34.80
<i>orf496b</i>	497	45.27	32.80	46.68	41.58	58.69	46.60
<i>orf518</i>	519	48.75	44.70	38.34	43.93	49.95	38.40
<i>orf575</i>	576	46.70	42.36	37.67	42.25	51.66	37.80
<i>orf670</i>	671	43.96	32.04	35.77	37.26	48.03	35.50
<i>orf764</i>	765	53.20	39.08	52.94	48.41	52.69	52.90
<i>orf774</i>	775	46.84	32.52	41.29	40.22	52.64	41.40
<i>orf99a</i>	100	60.00	33.00	56.00	49.67	51.36	55.70
<i>orf99b</i>	100	47.00	43.00	43.00	44.33	47.84	44.10
<i>orf99g</i>	100	44.00	42.00	42.00	42.67	49.86	41.70
<i>rpl5</i>	187	50.80	35.29	40.64	42.25	56.99	41.00
<i>rps12</i>	126	53.97	48.41	30.16	44.18	50.58	31.90
<i>rps13</i>	117	47.01	37.61	21.37	35.33	36.44	20.20
<i>rps3</i>	552	42.39	39.31	39.31	40.34	55.74	39.70
<i>rps4</i>	316	41.14	39.56	34.81	38.50	52.59	35.80
<i>rps7</i>	156	54.49	44.87	25.00	41.45	47.27	25.70

GC_1 、 GC_2 、 GC_3 分别为第 1 位、第 2 位、第 3 位碱基 G+C 含量; GC_{all} : 3 位碱基平均 G+C 含量; GC_{3S} : 同义密码子第 3 位碱基 G+C 含量。

表 2 海甜菜叶绿体基因编码序列 (CDS) 密码子各位置的 G+C 含量及有效密码子数

Table 2 The G+C content and the effective number of codons at each position of chloroplast gene coding sequence (CDS) codons in *Beta vulgaris* subsp.

基因	密码子数	GC_1 (%)	GC_2 (%)	GC_3 (%)	GC_{all} (%)	有效密码子数	GC_{3S} (%)
<i>accD</i>	492	39.23	35.57	28.86	34.55	44.37	29.20
<i>atpA</i>	508	56.89	39.76	26.38	41.01	47.21	26.60
<i>atpB</i>	499	55.91	42.08	29.86	42.62	47.15	30.10
<i>atpE</i>	135	48.89	40.74	29.63	39.75	49.82	30.20
<i>atpF</i>	185	46.49	30.27	31.35	36.04	50.76	30.60
<i>atpI</i>	248	50.40	37.50	25.00	37.63	45.41	25.00
<i>ccsA</i>	324	36.42	35.80	22.84	31.69	43.23	22.50
<i>cemA</i>	230	39.13	27.39	30.00	32.17	47.91	30.00
<i>clpP1</i>	197	56.85	37.56	31.98	42.13	48.61	31.90
<i>matK</i>	506	37.55	30.63	26.48	31.55	45.07	26.10
<i>ndhA</i>	364	41.21	39.01	21.43	33.88	41.30	21.90
<i>ndhB</i>	511	41.49	38.75	32.29	37.51	47.98	32.50
<i>ndhC</i>	121	44.63	33.06	23.14	33.61	46.89	23.30
<i>ndhD</i>	384	39.32	37.50	26.04	34.29	43.76	26.00
<i>ndhE</i>	102	41.18	33.33	23.53	32.68	45.75	23.00
<i>ndhF</i>	744	35.35	35.75	21.77	30.96	39.31	21.80
<i>ndhG</i>	177	43.50	33.33	22.03	32.96	46.05	22.20
<i>ndhH</i>	394	50.00	35.79	27.16	37.65	47.78	27.50
<i>ndhI</i>	171	43.86	39.18	22.81	35.28	41.91	22.40
<i>ndhJ</i>	159	51.57	37.74	33.33	40.88	50.88	33.50
<i>ndhK</i>	251	42.23	41.43	29.48	37.72	52.01	30.10
<i>pafI</i>	169	47.34	37.87	31.95	39.05	45.44	32.10
<i>pafII</i>	185	42.70	41.08	30.27	38.02	47.61	30.90
<i>petA</i>	321	52.65	35.83	30.22	39.56	50.85	29.90

续表 2 Continued 2

基因	密码子数	GC_1 (%)	GC_2 (%)	GC_3 (%)	GC_{all} (%)	有效密码子数	GC_{3S} (%)
<i>petB</i>	216	49.54	40.74	30.09	40.12	44.18	29.80
<i>petD</i>	161	50.93	38.51	22.36	37.27	35.00	22.00
<i>psaA</i>	751	52.20	43.68	32.49	42.79	48.57	32.80
<i>psaB</i>	735	48.98	42.99	29.80	40.59	46.21	29.90
<i>psbA</i>	354	49.72	43.50	31.92	41.71	39.94	32.00
<i>psbB</i>	509	54.22	46.37	31.43	44.01	44.96	32.00
<i>psbD</i>	354	52.82	43.50	31.07	42.47	44.38	31.00
<i>rbcL</i>	476	57.77	44.12	28.57	43.49	45.59	28.60
<i>rpl14</i>	122	50.82	39.34	28.69	39.62	46.34	29.10
<i>rpl16</i>	136	51.47	54.41	27.94	44.61	42.47	29.40
<i>rpl2</i>	275	50.55	48.36	32.36	43.76	50.55	32.60
<i>rpl20</i>	129	40.31	35.66	27.91	34.63	43.95	29.40
<i>rpl22</i>	198	44.95	33.33	30.81	36.36	50.17	31.40
<i>rpoA</i>	337	44.51	31.45	27.00	34.32	50.60	26.70
<i>rpoB</i>	1071	50.33	37.82	29.60	39.25	47.23	29.20
<i>rpoC1</i>	678	49.71	37.76	26.55	38.00	47.15	26.70
<i>rpoC2</i>	1366	45.97	37.63	28.70	37.43	47.80	28.50
<i>rps11</i>	139	53.96	56.12	25.18	45.08	43.11	24.60
<i>rps12</i>	124	51.61	47.58	29.84	43.01	47.98	30.40
<i>rps14</i>	101	43.56	46.53	30.69	40.26	44.34	31.20
<i>rps18</i>	102	32.35	42.16	24.51	33.01	38.97	25.30
<i>rps2</i>	237	44.30	43.88	28.69	38.96	49.30	29.30
<i>rps3</i>	219	47.49	34.25	23.29	35.01	51.29	22.90
<i>rps4</i>	202	49.50	39.60	29.21	39.44	49.33	29.20
<i>rps7</i>	156	54.49	44.87	25.00	41.45	47.27	25.70
<i>rps8</i>	135	41.48	40.74	23.70	35.31	37.31	24.40
<i>ycf1</i>	1849	36.99	27.96	23.26	29.40	45.31	23.50
<i>ycf2</i>	2141	41.43	34.28	36.01	37.24	50.93	35.90

GC_1 、 GC_2 、 GC_3 分别为第 1 位、第 2 位、第 3 位碱基 G+C 含量; GC_{all} : 3 位碱基平均 G+C 含量; GC_{3S} : 同义密码子第 3 位碱基 G+C 含量。

线粒体基因组密码子参数间的相关性如表 3 所示。 GC_{all} 和 GC_1 、 GC_2 、 GC_3 都呈极显著相关, GC_1 和 GC_2 呈显著相关, GC_1 和 GC_3 呈极显著相关, GC_2 与 GC_3 呈极显著相关; ENC 与 GC_1 呈显著相关, 与 GC_2 不相关, 与 GC_3 和 GC_{all} 呈极显著相关; ENC 、 GC_1 、 GC_3 和密码子数均不相关, GC_2 与密码子数呈极显著负相关。 ENC 与密码子第 1 位碱基组成密切相关, 与密码子数不相关, 可能的原因是基因序列长度对密码子使用偏好性的影响很小, 这与蔡元保等研究结果一致^[32]。叶绿体基因组密码子参数间的相关性如表 4 所示。 GC_{all} 和 GC_1 、 GC_2 、 GC_3 呈极显著相关, GC_1 和 GC_2 呈极显著相关, 与 GC_3 呈显著相关, GC_2 与 GC_3 无显著相关性; ENC 与 GC_3 呈极显著相关, 与 GC_1 、 GC_2 、 GC_{all} 不相关; 密码子数与 ENC 、 GC_1 、 GC_2 、 GC_3 均不相关。海甜菜线粒体和叶绿体基因组密码子的使用偏好性均受到 G+C 含

量、 ENC 和密码子第 1 位碱基的影响。其中, 线粒体基因组密码子的使用偏好性与 G+C 含量密切相关, 而叶绿体基因组密码子的使用偏好性则与 ENC 和 GC_3 密切相关。

2.2 同义密码子相对使用度 (RSCU) 分析

海甜菜线粒体和叶绿体基因同义密码子相对使用度如表 5 所示。线粒体基因高频密码子 ($RSCU > 1$) 有 28 个, 其中 11 个以 A 结尾, 2 个以 G 结尾, 15 个以 U 结尾; 低频密码子 ($RSCU < 1$) 共有 34 个, 其中 5 个以 A 结尾, 16 个以 C 结尾, 12 个以 G 结尾, 1 个以 U 结尾。叶绿体基因高频密码子有 30 个, 其中 11 个以 A 结尾, 12 个以 U 结尾, 4 个以 G 结尾, 3 个以 C 结尾; $RSCU$ 为 1 的密码子共 2 个, 分别为 AUG 和 UGG; 低频密码子共有 32 个, 这表明海甜菜线粒体和叶绿体基因组中使用度较高的密码子是以 A 和 U 结尾, 此结果与沙枣叶绿体基因组相同^[39]。

表3 海甜菜线粒体基因密码子参数间的相关性

Table 3 Correlation between mitochondrial gene codon parameters in *Beta vulgaris* subsp.

参数	GC_1	GC_2	GC_3	GC_{all}	ENC	密码子数
GC_1	1.000	0.208 *	0.251 **	0.637 **	0.215 *	0.073
GC_2	0.208 *	1.000	0.296 **	0.703 **		-0.227 **
GC_3	0.251 **	0.296 **	1.000	0.777 **	0.403 **	-0.101
GC_{all}	0.637 **	0.703 **	0.777 **	1.000	0.361 **	-0.129
ENC	0.215 *	0.122	0.403 **	0.361 **	1.000	0.092
密码子数	0.073	-0.227 **	-0.101	-0.129	0.092	1.000

GC_1 、 GC_2 、 GC_3 分别为第1位、第2位、第3位碱基的G+C含量; GC_{all} 为3位碱基平均G+C含量, ENC 为有效密码子数。*、**分别表示参数间相关性达显著水平($P<0.05$)和极显著水平($P<0.01$)。

表4 海甜菜叶绿体基因密码子参数间的相关性

Table 4 Correlation between codon parameters of chloroplast gene in *Beta vulgaris* subsp.

参数	GC_1	GC_2	GC_3	GC_{all}	ENC	密码子数
GC_1	1.000	0.489 **	0.350 *	0.859 **	0.252	-0.150
GC_2	0.489 **	1.000	0.190	0.795 **	-0.197	-0.255
GC_3	0.350 *	0.190	1.000	0.571 **	0.527 **	0.168
GC_{all}	0.859 **	0.795 **	0.571 **	1.000	0.191	-0.152
ENC	0.252	-0.197	0.527 **	0.191	1.000	0.135
密码子数	-0.150	-0.255	0.168	-0.152	0.135	1.000

GC_1 、 GC_2 、 GC_3 分别为第1位、第2位、第3位碱基的G+C含量; GC_{all} 为3位碱基平均G+C含量, ENC 为有效密码子数。*、**分别表示参数间相关性达显著水平($P<0.05$)和极显著水平($P<0.01$)。

表5 海甜菜线粒体基因和叶绿体基因同义密码子相对使用度(RSCU)

Table 5 Relative synonymous codon usage (RSCU) of mitochondrial and chloroplast genes in *Beta vulgaris* subsp.

氨基酸	密码子	线粒体 RSCU	叶绿体 RSCU	氨基酸	密码子	线粒体 RSCU	叶绿体 RSCU
终止子	UAA	1.411 8	1.615 5	甲硫氨酸	AUG	1.000 0	1.000 0
	UGA	0.904 5	0.692 4	天冬酰胺	AAC	0.744 4	0.453 6
	UAG	0.683 7	0.692 4		AAU	1.255 6	1.546 4
丙氨酸	GCG	0.588 4	0.488 8	脯氨酸	CCC	0.886 4	1.647 2
	GCU	1.538 4	1.714 8		CCA	1.017 6	1.155 2
	GCA	0.935 2	1.112 4		CCU	1.455 6	0.653 2
	GCC	0.938 0	0.683 6		CCG	0.640 4	0.544 4
半胱氨酸	UGC	0.881 6	0.509 0	谷氨酰胺	CAA	1.282 4	1.566 2
	UGU	1.118 4	1.491 0		CAG	0.717 6	0.433 8
天冬氨酸	GAC	0.703 2	0.418 0	精氨酸	CGA	1.072 2	0.424 2
	GAU	1.296 8	1.582 0		CGC	0.639 0	0.389 4
谷氨酸	GAG	0.791 0	0.464 6		AGG	1.009 2	1.701 0
	GAA	1.209 0	1.535 4		CGU	0.936 0	1.434 6
苯丙氨酸	UUU	1.116 0	1.388 6		AGA	1.569 0	1.464 0
	UUC	0.884 0	0.611 4		CGG	0.775 2	0.586 8
甘氨酸	GGA	1.337 2	1.585 2	丝氨酸	AGU	1.056 0	0.971 4
	GGC	0.662 8	0.434 8		UCC	0.941 4	0.592 8
	GGU	1.156 0	1.321 2		UCA	1.056 0	1.260 0
	GGG	0.844 0	0.659 2		AGC	0.861 6	0.351 0

续表 5 Continued 5

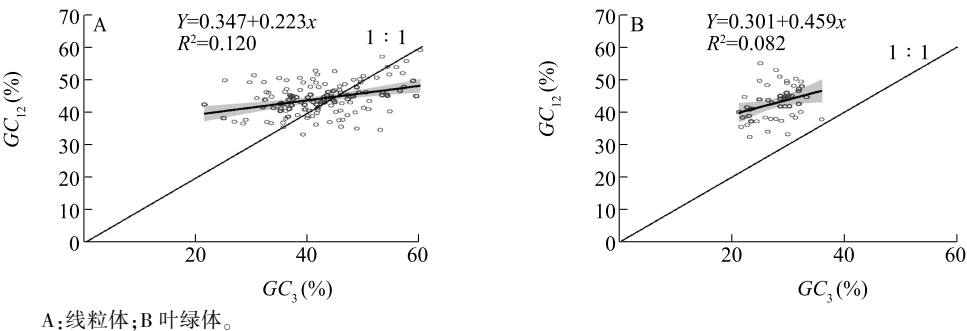
氨基酸	密码子	线粒体 <i>RSCU</i>	叶绿体 <i>RSCU</i>	氨基酸	密码子	线粒体 <i>RSCU</i>	叶绿体 <i>RSCU</i>
组氨酸	CAU	1.427 0	1.489 6	苏氨酸	UCU	1.443 0	1.139 4
	CAC	0.573 0	0.510 4		UCG	0.641 4	1.685 4
异亮氨酸	AUA	0.847 8	0.955 2		ACC	0.945 2	0.449 2
	AUC	0.909 0	0.532 5		ACA	1.028 4	0.704 8
亮氨酸	AUU	1.243 2	1.512 6		ACG	0.574 0	1.591 6
	AAA	1.057 4	1.572 6	ACU	1.452 4	1.254 4	
	AAG	0.942 6	0.427 4	缬氨酸	GUA	0.995 2	1.514 0
亮氨酸	CUC	0.781 2	0.373 2	GUC	0.847 6	1.553 6	
	CUU	1.350 0	1.263 6	GUG	0.876 4	0.524 0	
	CUG	0.640 8	0.290 4	GUU	1.280 8	0.408 4	
	UUA	1.264 8	2.065 2	色氨酸	UGG	1.000 0	1.000 0
	UUG	1.115 4	1.191 6	酪氨酸	UAC	0.684 0	1.626 2
	CUA	0.847 2	0.815 4	UAU	1.316 0	0.373 8	

海甜菜线粒体和叶绿体基因组中除甲硫氨酸 (Met) 和色氨酸 (Trp) 分别由 1 个密码子 AUG 和 UGG 编码之外, 剩余氨基酸由 2~6 个密码子编码, 且都表现出明显的使用偏好性。在编码亮氨酸 (Leu) 的 6 个密码子中, UUA 的 *RSCU* 最大, 线粒体基因组中为 1.264 8, 叶绿体基因组中为 2.065 2, 表明 UUA 具有很高的偏好性, 是海甜菜线粒体和叶绿体基因组中常用的密码子。

2.3 海甜菜线粒体和叶绿体基因密码子中性绘图分析

海甜菜线粒体基因组 GC_{12} 的取值范围在

33.12%~58.91%, GC_3 的取值范围在 21.37%~60.38%, GC_{12} 与 GC_3 的相关系数为 0.346, 无显著相关性 (图 1A)。这说明海甜菜线粒体基因组密码子第 3 位与第 1、2 位碱基突变模式不相同, 比起碱基突变密码子偏好性更易受到自然选择的影响, 这与睡莲属植物的结果相似^[40]。叶绿体基因组 GC_{12} 的取值范围在 32.46%~55.04%, GC_3 的取值范围在 21.43%~36.01%, GC_{12} 与 GC_3 的相关系数为 0.286, 也无显著相关性 (图 1B)。此外, 图 1B 所有点均在 1:1 线上方, 说明所有基因 $GC_{12}>GC_3$, 即密码子第 3 位的 G+C 含量偏低, 密码子偏好性受自然选择的影响大。



A: 线粒体; B: 叶绿体。

图 1 海甜菜线粒体和叶绿体基因密码子中性绘图

Fig.1 The neutral mapping of mitochondrial and chloroplast gene codons in *Beta vulgaris* subsp.

2.4 ENC-plot 分析

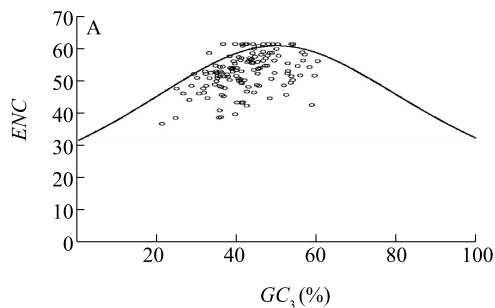
海甜菜线粒体和叶绿体基因组有效密码子比值频率分布及 *ENC* 与 GC_3 的关系图 (ENC-plot) 分别如表 6 和图 2 所示。海甜菜线粒体基因 *ENC* 大多偏离标准曲线 (图 2A), 仅有 39 个基因有效密码子比值频率分布在 -0.05~0.05 区, 这 39 个基因与预期 *ENC* 较

接近, 其余的 97 个基因分布在 -0.05~0.05 之外。叶绿体基因 *ENC* 也大多数偏离标准曲线 (图 2B), 分布在 -0.05~0.05 的基因 14 个, 其余 38 个基因分布在 -0.05~0.05 之外。与叶绿体基因组相比, 线粒体基因组有更多基因远离标准曲线, 表明线粒体基因组密码子偏好性更易受到碱基突变影响。

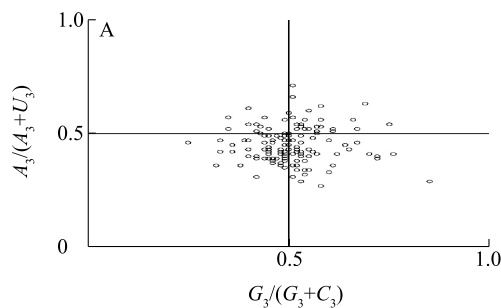
表 6 有效密码子比值频数分布

Table 6 Distribution of ENC ratio frequency

组段区间	频数		频率(%)	
	线粒体	叶绿体	线粒体	叶绿体
$[-0.15, -0.05)$	3	1	2	2
$[-0.05, 0.05)$	39	14	29	27
$[0.05, 0.15)$	63	27	46	52
$[0.15, 0.25)$	21	9	15	17
$[0.25, 0.35)$	10	1	7	2
总计	136	52	100	100



A: 线粒体; B: 叶绿体。

图 2 海甜菜线粒体和叶绿体基因组有效密码子数(ENC)与GC₃的关系图(ENC-plot)Fig.2 The relationship between GC₃ and the effective number of codons (ENC) in mitochondrial and chloroplast genomes of *Beta vulgaris* subsp.

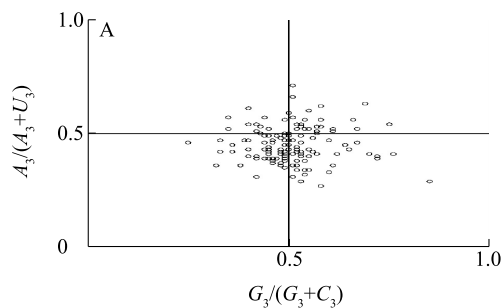
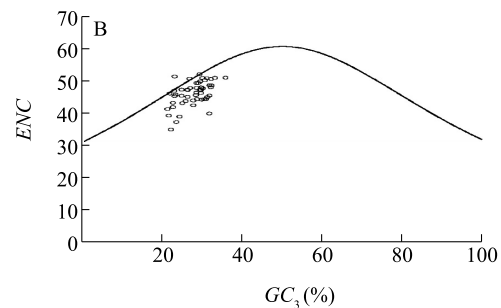
A: 线粒体; B: 叶绿体。

图 3 海甜菜线粒体和叶绿体基因密码子偏倚性分析(PR2-plot)

Fig.3 Analysis of codon bias of mitochondrial and chloroplast genomes in *Beta vulgaris* subsp.

2.5 偏倚分析 (PR2-PLot)

海甜菜线粒体和叶绿体各基因密码子第 3 位上碱基偏倚性如图 3 所示。各基因不均匀地分布在 4 个不同区域,大部分位于平面图的下半部和右半部。这说明海甜菜线粒体和叶绿体基因密码子第 3 位上 U(T) 碱基出现频率大于 A 碱基, G 碱基出现频率大于 C 碱基。海甜菜细胞器基因组密码子偏好性不但受到了突变的影响,还受到了自然选择的影响。



2.6 最优密码子确定

海甜菜线粒体和叶绿体高、低表达基因的数量及同义密码子相对使用度如表 7 所示。从表中可以看出,线粒体和叶绿体 $\Delta RSCU$ 不小于 0.08 的高表达优越密码子分别有 18 个和 17 个。线粒体基因组中以 A、U、C、G 结尾的优越密码子分别有 6 个、10 个、1 个和 1 个;线粒体 $\Delta RSCU$ 不小于 0.30 且小于 0.50 的优越密码子有 2 个, $\Delta RSCU$ 不小于 0.50 的优越密码子有 6 个。叶绿体基因组中以 A、U、C、G 结尾的优越密码子分别有 5 个、12 个、0 个和 0 个;叶绿体 $\Delta RSCU$

不小于 0.30 且小于 0.50 的优越密码子有 4 个; $\Delta RSCU$ 不小于 0.50 的优越密码子有 8 个。海甜菜线粒体和叶绿体基因组密码子中共有 27 个既满足高频率也满足高表达的条件,可视为最优密码子。

3 讨论与结论

海甜菜线粒体和叶绿体基因密码子的平均 G+C 含量分别为 43.42% 和 37.92%, 第 3 位碱基 G+C 含量 (GC_3) 大多在 50% 以下,说明海甜菜线粒体和叶绿体基因密码子第 3 位碱基以 A、U 为主,这与水

表 7 海甜菜线粒体和叶绿体各基因的 *RSCU* 和最优密码子Table 7 *RSCU* and optimal codons of mitochondrial and chloroplast genes in *Beta vulgaris* subsp.

氨基酸	密码子		高表达基因 密码子数		高表达基因同义 密码子相对使用度 (<i>RSCU</i>)		低表达基因 密码子数		低表达基因同义 密码子相对使用度 (<i>RSCU</i>)		$\Delta RSCU$	
	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体
丙氨酸	GCA	GCA	11	24	0.71	0.96	31	28	0.98	1.20	-0.27	-0.24
	GCC	GCC	12	9	0.77	0.36	28	23	0.89	0.99	-0.12	-0.63
	GCG	GCG	11	11	0.71	0.44	29	9	0.92	0.39	-0.21	0.05
	GCU ***	GCU ***	28	56	1.81	2.24	38	33	1.21	1.42	0.60	0.82
精氨酸	AGA	AGA	24	21	1.55	1.73	41	58	1.59	1.76	-0.04	-0.24
	AGG **	AGG	23	5	1.48	0.41	26	33	1.01	1.00	0.47	-0.03
	CGA **	CGA	21	24	1.35	1.97	26	44	1.01	1.33	0.34	-0.59
	CGC	CGC *	3	5	0.19	0.41	23	11	0.89	0.33	-0.70	0.64
	CGG	CGG	7	3	0.45	0.25	20	18	0.77	0.55	-0.32	0.08
	CGU	CGU	15	15	0.97	1.23	19	34	0.74	1.03	0.23	-0.30
天冬酰胺	AAC	AAC	17	20	0.76	0.49	39	32	0.92	0.38	-0.16	0.11
	AAU *	AAU	28	61	1.24	1.51	46	138	1.08	1.62	0.16	-0.11
天冬氨酸	GAC	GAC	12	10	0.37	0.45	38	29	1.00	0.34	-0.63	0.11
	GAU ***	GAU	53	34	1.63	1.55	38	142	1.00	1.66	0.63	-0.11
半胱氨酸	UGC	UGC	11	0	0.67	0	17	12	0.85	0.63	-0.18	-0.63
	UGU *	UGU ***	22	11	1.33	2.00	23	26	1.15	1.37	0.18	0.63
谷氨酰胺	CAA	CAA *	35	28	1.27	1.75	50	91	1.23	1.48	0.04	0.27
	CAG	CAG	20	4	0.73	0.25	31	32	0.77	0.52	-0.04	-0.27
谷氨酸	GAA	GAA *	47	40	1.24	1.43	90	117	1.19	1.31	0.05	0.12
	GAG	GAG	29	16	0.76	0.57	61	61	0.81	0.69	-0.05	-0.12
甘氨酸	GGA *	GGA	26	33	1.39	1.23	33	58	1.16	1.63	0.23	-0.40
	GGC	GGC	14	10	0.75	0.37	21	16	0.74	0.45	0.01	-0.08
	GGG	GGG	14	10	0.75	0.37	29	29	1.02	0.82	-0.27	-0.45
	GGU	GGU ***	21	54	1.12	2.02	31	39	1.09	1.10	0.03	0.92
组氨酸	CAC	CAC	8	5	0.50	0.40	22	20	0.77	0.61	-0.27	-0.21
	CAU *	CAU *	24	20	1.50	1.60	35	46	1.23	1.39	0.27	0.21
异亮氨酸	AUA *	AUA	39	50	1.03	1.00	35	85	0.77	1.00	0.26	0
	AUC	AUC	28	17	0.74	0.34	47	56	1.04	0.66	-0.30	-0.32
	AUU	AUU **	47	83	1.24	1.66	54	115	1.19	1.35	0.05	0.31
亮氨酸	CUA	CUA	23	17	0.68	0.61	32	51	0.72	0.94	-0.04	-0.33
	CUC	CUC	27	2	0.80	0.07	41	32	0.93	0.59	-0.13	-0.52
	CUG	CUG	20	4	0.59	0.14	33	23	0.75	0.43	-0.16	-0.29
	CUU *	CUU	56	35	1.66	1.26	67	82	1.52	1.52	0.14	-0.26
	UUA *	UUA ***	41	81	1.21	2.91	46	62	1.04	1.15	0.17	1.76
	UUG	UUG	36	28	1.06	1.01	46	74	1.04	1.37	0.02	-0.36
赖氨酸	AAA *	AAA **	65	45	1.14	1.64	43	138	0.91	1.33	0.23	0.31
	AAG	AAG	49	10	0.86	0.36	51	69	1.09	0.67	-0.23	-0.31

续表 7 Continued 7

氨基酸	密码子		高表达基因 密码子数		高表达基因同义 密码子相对使用度 (RSCU)		低表达基因 密码子数		低表达基因同义 密码子相对使用度 (RSCU)		$\Delta RSCU$	
	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体	线粒体	叶绿体
甲硫氨酸	AUG	AUG	41	49	1.00	1.00	44	57	1.00	1.00	0	0
苯丙氨酸	UUC	UUC	36	33	0.97	0.58	61	86	0.87	0.93	0.10	-0.35
	UUU	UUU **	38	80	1.03	1.42	79	99	1.13	1.07	-0.10	0.35
脯氨酸	CCA *	CCA	19	18	1.09	1.11	29	37	0.95	1.08	0.14	0.03
	CCC	CCC	4	6	0.23	0.37	34	25	1.11	0.73	-0.88	-0.36
	CCG	CCG	11	9	0.63	0.55	21	27	0.69	0.79	-0.06	-0.24
	CCU ***	CCU ***	36	32	2.06	1.97	38	48	1.25	1.40	0.81	0.57
丝氨酸	AGC	AGC	23	4	0.86	0.21	30	14	0.99	0.29	-0.13	-0.08
	AGU	AGU **	27	27	1.01	1.42	35	48	1.15	0.99	-0.14	0.43
	UCA	UCA	29	13	1.09	0.68	37	61	1.22	1.26	-0.13	-0.58
	UCC *	UCC	28	13	1.05	0.68	25	55	0.82	1.14	0.23	-0.46
	UCG	UCG	10	14	0.38	0.74	23	37	0.76	0.77	-0.38	-0.03
	UCU ***	UCU ***	43	43	1.61	2.26	32	75	1.05	1.55	0.56	0.71
苏氨酸	ACA	ACA	21	18	1.08	0.90	29	45	1.12	1.34	-0.04	-0.44
	ACC	ACC	11	15	0.56	0.75	25	21	0.96	0.63	-0.40	0.12
	ACG	ACG	10	5	0.51	0.25	19	21	0.73	0.63	-0.22	-0.38
	ACU ***	ACU ***	36	42	1.85	2.10	31	47	1.19	1.40	0.66	0.70
色氨酸	UGG	UGG	22	30	1.00	1.00	32	55	1.00	1.00	0	0
酪氨酸	UAC	UAC	15	10	0.62	0.32	24	24	0.66	0.40	-0.04	-0.08
	UAU	UAU *	33	53	1.38	1.68	49	95	1.34	1.60	0.04	0.08
缬氨酸	GUA	GUA ***	13	32	0.57	1.68	21	42	0.79	1.17	-0.22	0.51
	GUC	GUC	15	5	0.65	0.26	26	17	0.98	0.48	-0.33	-0.22
	GUG	GUG	21	5	0.91	0.26	23	27	0.87	0.76	0.04	-0.50
	GUU ***	GUU *	43	34	1.87	1.79	36	57	1.36	1.59	0.51	0.20

* 表示 $0.08 \leq \Delta RSCU < 0.30$ 水平上的高表达密码子; ** 表示 $0.30 \leq \Delta RSCU < 0.50$ 水平上的高表达密码子; *** 表示 $\Delta RSCU \geq 0.50$ 水平上的高表达密码子。

稻线粒体基因组^[41]、玉米线粒体基因组^[42],大粒香水稻叶绿体基因组^[43]、苟当3号水稻叶绿体基因组^[44]、白羊草叶绿体基因组^[45]的密码子偏好性一致。由此可以看出植物细胞器基因组进化趋势大致相同。

RSCU 是评价生物细胞器基因组密码子使用模式的重要指标^[46]。本研究中海甜菜线粒体和叶绿体基因组 $RSCU > 1.00$ 的高频密码子有 28 个和 30 个,其中线粒体基因组有 26 个,叶绿体有 23 个以 A/U 结尾;这与紫花苜蓿 (*Medicago sativa*) 叶绿体基因组密码子使用模式^[47]一致。中性绘图分析、ENC-plot 分析、

PR2-plot 分析结果均显示海甜菜线粒体和叶绿体基因组密码子受自然选择的影响较大^[24],与人参属植物 (*Panax Linn*)^[48]、乌头属药用植物 (*Aconitum*)^[49] 叶绿体基因组密码子一致,不同于杨树 (*Populus przewalskii* Maxim.) 叶绿体基因组密码子主要受碱基突变的影响^[50],说明物种密码子偏好性影响因素的多元化^[51-58]。

本研究初步揭示了海甜菜线粒体和叶绿体基因组密码子偏好性形成的主要原因是自然选择,筛选出 27 个最优密码子。本研究结果为探索海甜菜的分子特性和遗传多样性,弄清基因的进化压力以及分子育种具有重要意义。

参考文献:

- [1] 黄 兴. 多头带缘虫功能基因的原核表达及蛋白特性分析[D]. 成都:四川农业大学,2016.
- [2] 冯 展,江 媛,郑 燕,等. 肉苁蓉属植物叶绿体基因组密码子偏好性分析[J]. 中草药,2023,54(5):1540-1550.
- [3] ROMERO H, ZAVALA A, MUSTO H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces[J]. *Nucleic Acids Research*, 2000,28(10):2084-2090.
- [4] XU C, CAI X, CHEN Q, et al. Factors affecting synonymous codon usage bias in chloroplast genome of *oncidium gower ramsey* [J]. *Evolutionary Bioinformatics*,2011,7(7):271-278.
- [5] 中国科学院中国植物志委员会. 中国植物志第 22 卷:被子植物门双子叶植物纲壳斗科榆科马尾树科[M]. 北京:科学出版社,1998:263-298.
- [6] 郝新艳,刘嘉伟,米福贵. 杂花苜蓿叶绿体基因组密码子使用偏好性分析[J/OL]. 分子植物育种. <https://kns.cnki.net/kcms2/detail/46.1068.S.20230524.1400.012.html>.
- [7] LIU H, LU Y, LAN B, et al. Codon usage by chloroplast gene is bias in *Hemiptelea davidii*[J]. *Journal of Genetics*,2020,99:1-11.
- [8] WEI L, HE J, JIA X, et al. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution[J]. *BMC Evolutionary Biology*,2014,14:1-12.
- [9] PANELLA L, LEWELLEN R T. Broadening the genetic base of sugar beet: introgression from wild relatives[J]. *Euphytica*,2007,154:383-400.
- [10] CASTRO S, ROMEIRAS M M, CASTRO M, et al. Hidden diversity in wild Beta taxa from Portugal: insights from genome size and ploidy level estimations using flow cytometry[J]. *Plant Science*, 2013,207:72-78.
- [11] ASCARIN F, NÓBREGA, LEITE I S, et al. Assessing the diversity of sea beet (*Beta vulgaris* L. ssp. *maritima*) populations[J]. *Journal of Agricultural Science and Technology*, 2021, 23(3):685-698.
- [12] VINCENT H, WIERSEMA J, KELL S, et al. A prioritized crop wild relative inventory to help underpin global food security[J]. *Biological Conservation*,2013,167:265-275.
- [13] 邹 奕,吴则东,兴 旺,等. 甜菜种质资源遗传多样性研究进展[J]. 中国糖料,2018,40(5):73-76,80.
- [14] HANELT P, BUTTNER R, MANSFELD R. Mansfeld's encyclopedia of agricultural and horticultural crops (except Ornamentals) [M]. New York:Springer,2001.
- [15] OLIVEIRA S P A, NASCIMENTO H, SAMPAIO K B, et al. A review on bioactive compounds of beet (*Beta vulgaris* L. subsp. *vulgaris*) with special emphasis on their beneficial effects on gut microbiota and gastrointestinal health[J]. *Critical Reviews in Food Science and Nutrition*,2021,61(12):2022-2033.
- [16] RIBEIRO I C, PINHERIRO C, RIBEIRO C M, et al. Genetic diversity and physiological performance of Portuguese wild beet (*Beta vulgaris* spp. *maritima*) from three contrasting habitats [J]. *Frontiers in Plant Science*,2016,7:1293.
- [17] BANGAR S P, SING H A, CHAUDHARY V, et al. Bectroot as a novel ingredient for its versatile food applications[J]. *Critical Review in Food Sci and Nutri*,2023,63(26):8403-8427.
- [18] 王军伟,魏佑营,邱 红,等. 盐胁迫对不同品种菠菜种子萌发特性的影响[J]. 山东农业科学,2007(6):48-50,53.
- [19] 陶晓丽,王彦荣,刘志鹏. 牧草叶绿体基因组研究进展[J]. 草业科学,2015,32(6):978-987.
- [20] 赵 稳. 人干扰素 α -2b (HuIFN α -2b) 基因的烟草细胞核转化及其叶绿体定点表达载体的构建[D]. 武汉:华中科技大学,2010.
- [21] 赵晶晶,汪琪璇,蔺 欣,等. 线粒体功能障碍相关遗传性非综合征型听力损失的研究进展[J]. 中华耳科学杂志,2018,16(2):136-140.
- [22] 唐向民,杨守臻,陈怀珠,等. 栽培大豆和野生大豆线粒体基因组密码子使用偏性的比较分析[J]. 广西植物,2020,40(7):926-934.
- [23] 吴 茜,姜 梅,陈海梅,等. 旋覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组比较分析和系统发育研究[J]. 药学学报,2020,55(5):1042-1049.
- [24] 夏玉玲. 中国野桑蚕线粒体基因组研究[D]. 重庆:西南农业大学,2003.
- [25] KONG F, SUN P, CAO M, et al. Complete mitochondrial genome of *Pyropia yezoensis*: reasserting the revision of genus *Porphyra*. [J]. *Mitochondrial DNA*,2014,25(5):335-336.
- [26] 潘韵佳. 五倍子基原物种 DNA 条形码鉴定及叶绿体基因组解析[D]. 北京:北京协和医学院,2021.
- [27] DARRAC A, VARRE J S, MARECHAL-DROUARD L, et al. Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis[J]. *Genome Biology and Evolution*, 2011,3:723-736.
- [28] 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):35927.
- [29] 原晓龙,李云琴,张劲峰,等. 降香黄檀叶绿体基因组密码子偏好性分析[J]. 广西植物,2019,41(4):622-630.
- [30] 辛雅萱,黎若竹,李 鑫,等. 杠果叶绿体基因组密码子使用偏好性分析[J]. 中南林业科技大学学报,2021,41(9):148-156,165.
- [31] 王 飞,赵文植,董章宏,等. 扁核木属植物叶绿体基因组特征分析[J]. 热带作物学报,2022,43(9):1759-1770.
- [32] 蔡元保,杨祥燕. 澳洲坚果壳种叶绿体基因组的密码子使用偏好性及其影响因素分析[J]. 植物科学学报,2022,40(2):229-239.
- [33] 赖瑞联,陈 瑾,冯 新,等. 橄榄叶绿体基因组密码子偏好性特征[J]. 福建农林大学学报(自然科学版),2022,51(4):502-509.
- [34] 杨祥燕,蔡元保,谭秦亮,等. 菠萝叶绿体基因组密码子偏好性

- 分析[J]. 热带作物学报, 2022, 43(3): 439-446.
- [35] 徐盼. 鸭疫里默氏杆菌 OmpA 蛋白的免疫原性研究[D]. 成都: 四川农业大学, 2014.
- [36] WRIGHT F. The effective number of codons' used in a gene.[J]. Gene, 1990, 87(1): 23-29.
- [37] 赵洋, 杨阳, 刘振, 等. 茶树密码子用法分析[J]. 茶叶科学, 2011, 31(4): 319-325.
- [38] 田春育, 武自念, 李贤松, 等. 扁蓊豆叶绿体基因组密码子偏好性分析[J]. 草地学报, 2021, 29(12): 2678-2684.
- [39] 王婧, 王天翼, 王罗云, 等. 沙枣叶绿体全基因组序列及其使用密码子偏性分析[J]. 西北植物学报, 2019, 39(9): 1559-1572.
- [40] 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.
- [41] 金刚, 王丽萍, 龙凌云, 等. 普通野生稻线粒体蛋白质编码基因密码子使用偏好性的分析[J]. 植物科学学报, 2019, 37(2): 188-197.
- [42] 侯梦薇, 汪文睿, 曹言勇, 等. 玉米线粒体基因组密码子偏好性分析[J/OL]. 分子植物育种. <https://kns.cnki.net/kcms/detail/46.1068.S.20211209.1844.010.html>.
- [43] 吴朝昕, 刘雪薇, 李祖军, 等. 大粒香水稻叶绿体基因组特征分析[J]. 广西植物, 2022, 42(11): 1830-1839.
- [44] 吴朝昕, 徐海峰, 刘雪薇, 等. 苟当 3 号水稻叶绿体基因组特征分析[J/OL]. 分子植物育种. <https://kns.cnki.net/kcms/detail/46.1068.S.20220118.1548.006.html>.
- [45] 高守舆, 李钰莹, 杨志青, 等. 白羊草叶绿体基因组密码子使用偏好性分析[J/OL]. 草业学报, 2023, 32(7): 85-95.
- [46] 牛俊梅, 王昕玥, 岳加蕊, 等. 前胡 (*Peucedanum praeruptorum* Dunn) 叶绿体基因组密码子偏好性分析[J/OL]. 分子植物育种. <https://kns.cnki.net/kcms/detail/46.1068.S.20221011.1111.002.html>.
- [47] 喻凤, 韩明. 紫花苜蓿叶绿体基因组密码子偏好性分析[J]. 广西植物, 2021, 41(12): 2069-2076.
- [48] 卢志宏, 田文勇, 杨传东. 人参属植物叶绿体基因组特征和密码子偏好性分析[J/OL]. 分子植物育种. <https://kns.cnki.net/kcms/detail/46.1068.S.20220720.1016.002.html>.
- [49] 樊东昌, 穆赢通, 贾俊英, 等. 乌头属药用植物叶绿体基因组密码子特征和系统发育分析[J/OL]. 分子植物育种. <https://kns.cnki.net/kcms/detail/46.1068.S.20220711.1339.002.html>.
- [50] ZHOU M, LONG W, LI X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants[J]. Forestry Studies in China, 2008, 10: 235-242.
- [51] 刘潮, 韩利红, 盛巧, 等. 8 种植物类甜蛋白家族基因进化及密码子特征分析[J]. 江苏农业科学, 2022, 50(10): 44-51.
- [52] 王飞, 辛雅萱, 董章宏, 等. 无刺龙舌兰叶绿体基因组特征及密码子偏好性分析[J]. 南方农业学报, 2022, 53(4): 1030-1039.
- [53] 刘潮, 韩利红, 吴丽芳, 等. 辣椒基因组密码子使用偏性分析[J]. 江苏农业科学, 2022, 50(5): 16-22.
- [54] 王占军, 吴子琦, 王朝霞, 等. 3 个茶树品种 *WOX* 基因家族的进化及密码子偏好性比较[J]. 南京林业大学学报(自然科学版), 2022, 46(2): 71-80.
- [55] 何积翠, 喻达辉, 白丽蓉. 瓜螺线粒体全基因组密码子偏好性分析[J]. 南方农业学报, 2022, 53(1): 191-199.
- [56] 张海霞, 王玉道, 许雪妮. 苜蓿质膜内在蛋白编码基因 *MsPIP*s 家族的密码子偏好性分析[J]. 江苏农业学报, 2021, 37(6): 1393-1399.
- [57] 李凤, 辛静, 辛雅萱, 等. 楸叶绿体基因组密码子偏性分析[J]. 南方农业学报, 2021, 52(10): 2735-2743.
- [58] 肖明昆, 严炜, 熊贤坤, 等. 云南樟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(6): 127-134.

(责任编辑: 石春林)