



分子植物育种
Molecular Plant Breeding
ISSN 1672-416X, CN 46-1068/S

《分子植物育种》网络首发论文

题目： 大麦叶绿体密码子使用偏好性的比较与分析
作者： 郭奇凡, 魏新雨, 林汇源, 张启赢, 李思琦, 刘晓艺, 纪新明, 余奎军, 聂小军
网络首发日期： 2024-08-13
引用格式： 郭奇凡, 魏新雨, 林汇源, 张启赢, 李思琦, 刘晓艺, 纪新明, 余奎军, 聂小军. 大麦叶绿体密码子使用偏好性的比较与分析[JOL]. 分子植物育种. <https://link.cnki.net/urlid/46.1068.S.20240810.1035.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188, CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

研究报告

Research Report

大麦叶绿体密码子使用偏好性的比较与分析

郭奇凡¹ 魏新雨¹ 林汇源¹ 张启赢¹ 李思琦¹ 刘晓艺¹ 纪新明¹ 余奎军^{1,3*} 聂小军^{1,2*}

1 西北农林科技大学农学院, 杨凌, 712100; 2 西北农林科技大学, 旱区作物逆境生物学国家重点实验室, 杨凌, 712100; 3 宁夏农林科学院作物科学研究所, 银川, 750002

* 共同通信作者, nxskj@sohu.com; small@nwsuaf.edu.cn

摘要 密码子偏好性(codon usage bias, CUB)是生物的重要遗传特性之一。叶绿体是植物所特有的细胞器, 具有自身独立的基因组。分析和比较不同植物叶绿体基因组的密码子偏好性, 可为叶绿体的起源和物种进化提供有益信息, 也为细胞器基因工程和生物反应器等研究提供理论基础。为了明确大麦叶绿体基因组密码子偏好性及其影响因素, 本研究以栽培大麦、野生大麦和青稞3种大麦为材料, 对其叶绿体基因组的密码子偏好性进行了比较和分析。结果发现, 栽培大麦与青稞叶绿体基因的密码子使用模式完全一致, 而与野生大麦具有一定的差异, 基本反映了它们三者间的亲缘关系; 3种大麦叶绿体基因的有效密码子数在35~61之间, 平均为49.7; 三者RSUC值大于1的密码子数目都为32个, 且第3位密码子以AU居多; 中性绘图、PR2等分析均发现它们三者的密码子使用模式比较相似, 其偏好性受到自然选择和突变的影响; 在野生大麦、青稞和栽培大麦叶绿体基因组中分别鉴定出9、12和13个最优密码子。本研究比较了3种大麦叶绿体基因组密码子的使用模式, 为进一步叶绿体基因的原核表达和基因工程研究提供了有益信息。

关键词 大麦(*Hordeum vulgare* L.); 叶绿体基因; 最优密码子; 密码子使用偏好性

Comparative Analysis of the Codon Usage Bias of Chloroplast Genomes in Barley

Guo Qifan¹ Wei Xinyu¹ Lin Huiyuan¹ Zhang Qiying¹ Li Siqi¹ Liu Xiaoyi¹ Ji Xinming¹ She Kuijun^{1,3*}
Nie Xiaojun^{1,2*}

1 College of Agronomy, Northwest A&F University, Yangling, 712100; 2 State Key Laboratory of Crop Stress Biology in Arid Areas, Northwest A&F University, Yangling, 712100; 3 Institute of Crop Science, Ningxia Academy of Agricultural and Forestry Sciences, Yinchun, 750002

* Co-corresponding authors, nxskj@sohu.com; small@nwafu.edu.cn

Abstract Codon usage bias (CUB) is an important genetic characteristic formed during the evolution of organisms. Chloroplast is a unique organelle of plants with its own independent genome. Analysis and comparison of codon preference of chloroplast genomes in different plants can provide useful information for the origin of chloroplasts and species evolution, and also provide theoretical basis for organelle genetic engineering and bioreactor research. To understand the codon bias and its influencing factors in barley chloroplast genome, this study systematically compared and analyzed the codon bias of chloroplast genomes of three types of barley, including cultivated barley, wild barley and Tibetan hulless barley. Results showed that, on the whole, the codon usage patterns of the chloroplast

genes of cultivated barley and Tibetan hulless barley were completely consistent, while they displayed some difference from that of wild barley, which was basically accordance with their phylogenetic relationships. The effective codons of the three barley chloroplast genomes ranged from 35 to 61, with the average of 49.7. The numbers of codons with RSUC values greater than 1 were 32, and the third codon was mostly A or U. Furthermore, neutral PR2 and ENC plotting analysis all found that they had the similar codon usage bias patterns, which were mainly resulted by natural selection and mutation. Finally, 9 optimal codons were identified in wild barley chloroplast genome, as well as 12 and 13 optimal codons were identified in the chloroplast genomes of Tibetan hulless barley and cultivated barley respectively. This study conducted the systematic comparison and analysis of codon usage bias of three barley chloroplast genomes, which provided the useful information for further gene expression and genetic engineering research of chloroplast gene in barley.

Keywords Barley (*Hordeum vulgare* L.); Chloroplast genome; Optimal codons; Codon usage bias

大麦(*Hordeum vulgare* L.)是中国乃至全世界最重要的粮食作物之一, 其种植面积和总产量均居全球禾谷类作物第四位。大麦也是最早被人类驯化栽培的作物之一, 对推动人类文明的形成具有重要意义(Sato, 2020)。大麦具有早熟、耐旱、耐瘠薄等特性, 兼具重要的食用、饲用以及工业原料用价值, 在啤酒发酵、工业制淀粉、淀粉酶及优质饲料中有重要的作用(Rani and Bhardwaj, 2021)。近年来, 大麦因其具有高蛋白、高维生素、高纤维、低脂肪、低糖等“三高两低”的营养特性而被作为健康食品而越来越受到重视(Du et al., 2019)。

遗传密码子作为联结生物体遗传信息与蛋白质翻译的桥梁, 在实现遗传信息传递和表达方面具有关键作用(张以忠等, 2022)。在翻译过程中, 能够翻译为同一种氨基酸的不同密码子被称为同义密码子。在不同的物种, 甚至同物种不同的基因中, 用来编码同一氨基酸的密码子会不同, 即同义密码子的使用频率不同, 这种现象被称为密码子偏好性(CUB) (Grantham et al., 1980)。这是生物在长期进化过程中所形成的一种特殊的遗传特性。前人关于密码子偏好性产生的原因及其影响因素已开展了大量研究, 提出了多个假说, 如“选择-突变-漂移”假说和“中性理论”假说(Mazumdar et al., 2017)。“选择-突变-漂移”假说认为不同物种的密码子使用模式会受到最优密码子的选择压力和非同义密码子的突变-漂移的影响, 而二者的共同作用决定了所使用的密码子(Chen et al., 2004; 刘慧等, 2017)。“中性理论”假说认为同义密码子的使用模式不受自然选择压力的影响, 同义密码子的突变属于中性选择(吴正常等, 2013)。此外, 基因组 GC 含量、上下游依赖性核苷酸组成等也会影响密码子的偏好性(Iriarte et al., 2021)。

作为植物所特有的进行光合作用的细胞器, 叶绿体具有独立的基因组, 其大小在 107~218 kb 之间, 能够进行自我复制并编码 70~90 个蛋白质, 以母系遗传方式进行遗传信息传递(Daniell et al., 2016)。随着测序技术的不断发展, 叶绿体基因组被大量解析, 其密码子使用模式和形成原因研究不仅为叶绿体起源和物种进化研究提供了重要信息, 而且为叶绿体基因工程和生物反应器研究奠定了理论依据。有研究通过对 6 种山茶属(*Camellia* L.)叶绿体基因组的密码子分析, 发现它们的密码子使用模式具有一定的相似性, 其中 AAA、ATT、GAA、AAT 和 TTT 使用频率最高, TGA、TAG、TAA、TGC、CGC 使用频率最低, 该研究结果为山茶属的分类学研究提供了理论依据(Li et al., 2019); 还有研究对 18 个野生水稻(*Oryza rufipogon* Griff.)和栽培水稻(*Oryza sativa* L.)叶绿体基因组密码子进行分析, 发现稻属叶绿体基因组的密码子偏好性较低, 密

密码子第 3 位偏好 AT 结尾，其偏好性主要受自然选择和突变压力的影响，其中自然选择可能起主要作用 (Chakraborty et al., 2020)。在芒草(*Miscanthus*)叶绿体基因组中，共检测到 21 个常见的高频密码子和 11 个最佳密码子，其密码子的使用模式主要受到自然选择的影响，并发现拟南芥(*Arabidopsis thaliana*)具有与芒草相似的密码子使用模式，可作为芒草基因异源表达的宿主(Sheng et al., 2021)。但目前有关大麦叶绿体基因组密码子使用模式及其形成原因的研究还未见报道。

鉴于此，本研究以青稞、栽培大麦和野生大麦 3 种大麦的叶绿体基因组为对象，对其密码子的基本特性、使用模式以及偏好性等进行了系统的比较和分析，鉴定它们的最优密码子，明确了其密码子偏好性形成的影响因素，为从叶绿体密码子偏好性了解大麦的进化提供了有益信息，同时也为大麦叶绿体基因的原核表达、生物反应器等研究提供了理论基础。

1 结果与分析

1.1 大麦叶绿体基因组的密码子的基本特性分析

通过对各叶绿体基因组中 CDS 序列进行筛选，删除长度小于 300 bp 的序列，最终筛选到 53 条野生大麦、51 条青稞和 51 条栽培大麦的 CDS 序列用于密码子特性分析(表 1)。野生大麦基因组总 GC 含量为 39.1%，而其编码基因的平均 GC 含量为 38.4%，同时密码子 3 个位置的 GC 含量差异较大，GC₁、GC₂ 和 GC₃ 值分别为 45.7%、38.0%和 30.1%，呈不断降低的趋势，其中 GC₂ 的含量与 GC_{all} 的含量十分接近，与 GC₁、GC₃ 有较大差异，GC_{3S} 值为 27.2%。而青稞和栽培大麦各参数值完全一样，其基因组总 GC 含量均为 39.0%，而编码基因的平均 GC 含量为 37.9%，GC₁、GC₂ 和 GC₃ 的分布情况也为 GC₁>GC₂>GC₃，GC_{3S} 为 27%。ENC 值是用来衡量密码子的偏好程度的指标，数值区间在 20 到 61 之间，其值越接近 20 表示密码子的偏性越强，等于 20 则表示一种密码子仅编码一种氨基酸(完全偏好)，数值越接近 61 表示密码子的偏性越弱，等于 61 则表示密码子使用频率均等(无偏好)(Dilucca et al., 2020)。野生大麦的 ENC 值为 49.86，而青稞和栽培大麦的 ENC 值为 49.66，表明大麦叶绿体基因组密码子偏好性不强。

将 3 种大麦筛选后的基因合并为一条序列，分析其 RSCU 值。结果显示，3 种大麦的不同密码子的 RSCU 值基本一致，表现出相似的密码子使用偏性(表 2)。在 3 种大麦中，均发现了 32 个密码子的 RSCU 值大于 1，且这 32 个密码子中有 29 个密码子以 A/U 碱基结尾，说明大麦叶绿体基因组对这 32 个密码子具有使用偏好性，且偏好使用 A/U 碱基结尾。

表 1 野生大麦，青稞，栽培大麦叶绿体基因组特征比较

Table 1 Comparison of chloroplast genome characteristics of wild barley, Tibetan hulless barley and cultivated barley

物种 Species	大小(bp) Size (bp)	登录号 GenBank No.	CDS	CDS>300 bp	密码子总数 Codon count	GC 含量(%) Content of GC (%)						ENC
						GC _{all}	GC _g	GC ₁	GC ₂	GC ₃	GC _{3s}	
野生大麦 Wild barley	136 537	KC912689	83	53	17 426	39.10	38.40	45.70	38.00	30.10	27.20	49.86
青稞	136 462	KT962228	83	51	17 072	39.00	37.90	46.80	39.70	29.90	27.00	49.66

Tibetan hulless barley												
栽培大麦	136 462	EF115541	83	51	17 072	39.00	37.90	46.80	39.70	29.90	27.00	49.66
Cultivated barley												

表 2 3 种大麦叶绿体基因组密码子的平均相对同义密码子利用率(RSCU 值)

Table 2 Average relative synonymous codon usage (RSCU value) in the 3 barley chloroplast genomes

氨基酸 Amino acid	密码子 Codon	相对密码子使用度 RSCU			氨基酸 Amino acid	密码子 Codon	相对密码子使用度 RSCU		
		野生大麦 Wild barley	青稞 Tibetan hulless barley	栽培大麦 Cultivated barley			野生大麦 Wild barley	青稞 Tibetan hulless barley	栽培大麦 Cultivated barley
Phe	UUU(F)	1.32	1.33	1.33	Tyr	UAU(Y)	1.57	1.59	1.59
	UUC(F)	0.68	0.67	0.67		UAC(Y)	0.43	0.41	0.41
Leu	UUA(L)	2.07	2.10	2.10	His	CAU(H)	1.45	1.46	1.46
	UUG(L)	1.11	1.12	1.12		CAC(H)	0.55	0.54	0.54
	CUU(L)	1.28	1.28	1.28	Gln	CAA(Q)	1.55	1.55	1.55
	CUC(L)	0.39	0.38	0.38		CAG(Q)	0.45	0.45	0.45
Ile	CUA(L)	0.87	0.85	0.85	Asn	AAU(N)	1.50	1.49	1.49
	CUG(L)	0.28	0.28	0.28		AAC(N)	0.50	0.51	0.51
	AUU(I)	1.50	1.51	1.51	Lys	AAA(K)	1.47	1.48	1.48
	AUC(I)	0.55	0.56	0.56		AAG(K)	0.53	0.52	0.52
Met	AUA(I)	0.95	0.93	0.93	Asp	GAU(D)	1.54	1.55	1.55
	AUG(M)	1.00	1.00	1.00		GAC(D)	0.46	0.45	0.45
Val	GUU(V)	1.49	1.51	1.51	Glu	GAA(E)	1.48	1.48	1.48
	GUC(V)	0.49	0.48	0.48		GAG(E)	0.52	0.52	0.52
	GUA(V)	1.51	1.51	1.51	Cys	UGU(C)	1.51	1.51	1.51
	GUG(V)	0.51	0.50	0.50		UGC(C)	0.49	0.49	0.49
Ser	UCU(S)	1.66	1.68	1.68	Trp	UGG(W)	1.00	1.00	1.00
	UCC(S)	1.08	1.07	1.07		Arg	CGU(R)	1.38	1.40
	UCA(S)	1.01	1.03	1.03	CGC(R)		0.54	0.54	0.54
	UCG(S)	0.53	0.52	0.52	CGA(R)		1.26	1.25	1.25
Pro	CCU(P)	1.59	1.60	1.60	CGG(R)		0.41	0.40	0.40
	CCC(P)	0.91	0.90	0.90	Ser	AGU(S)	1.28	1.28	1.28
	CCA(P)	1.08	1.09	1.09		AGC(S)	0.44	0.43	0.43
	CCG(P)	0.42	0.41	0.41	Arg	AGA(R)	1.76	1.75	1.75
Thr	ACU(T)	1.71	1.71	1.71		AGG(R)	0.66	0.65	0.65
	ACC(T)	0.70	0.70	0.70	Gly	GGU(G)	1.25	1.26	1.26
	ACA(T)	1.13	1.13	1.13		GGC(G)	0.47	0.47	0.47
	ACG(T)	0.47	0.47	0.47		GGA(G)	1.57	1.55	1.55
Ala	GCU(A)	1.76	1.77	1.77		GGG(G)	0.71	0.72	0.72
	GCC(A)	0.59	0.58	0.58					
	GCA(A)	1.20	1.20	1.20					

1.2 大麦叶绿体基因组的中性绘图分析

为了揭示密码子 3 个位置之间的关系及其可能的影响因素,进一步对这 3 个大麦叶绿体基因组进行了中性绘图分析(图 1)。结果表明,在青稞和栽培大麦中,GC₁₂ 和 GC₃ 的回归系数为 0.1614,而野生大麦的 GC₁₂ 和 GC₃ 的回归系数为 0.165,3 种大麦的 GC₁₂ 和 GC₃ 均没有显著相关性($P>0.05$),表明突变压力对大麦密码子偏好性的影响较小,自然选择的效应占主导地位。

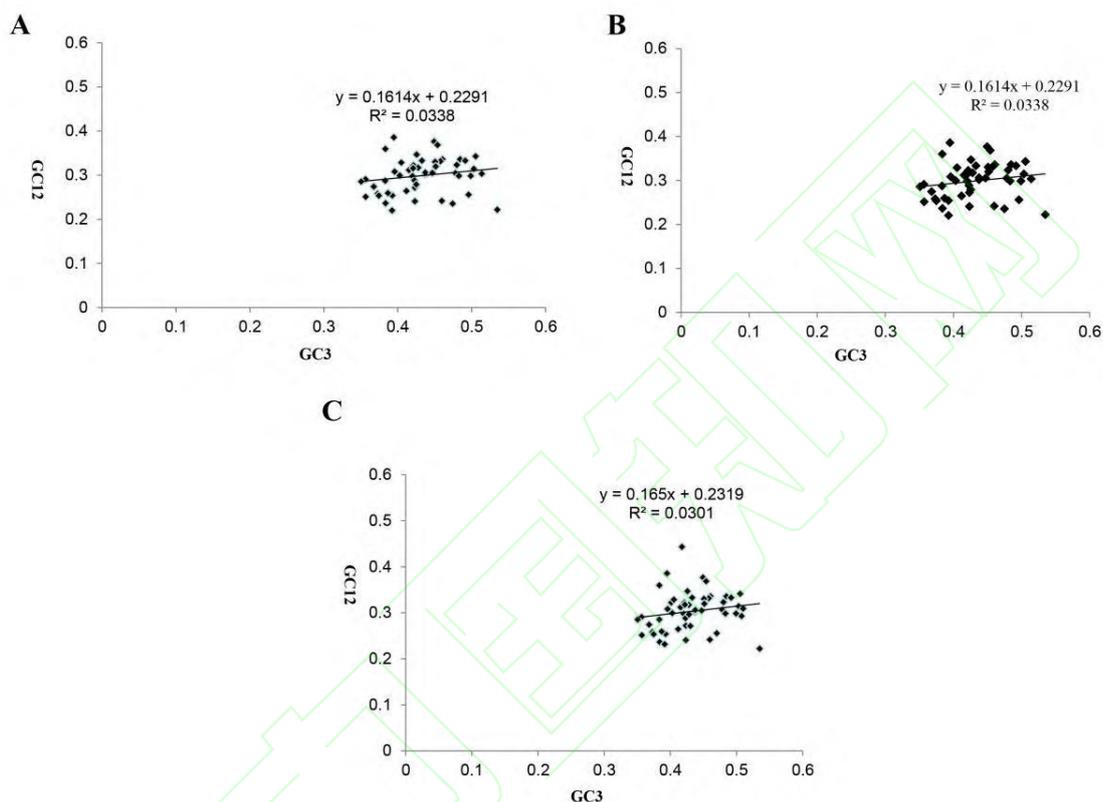


图 1 中性绘图分析

注: A: 青稞; B: 栽培大麦; C: 野生大麦

Figure 1 Neutrality plot analysis

Note: A: Tibetan hulless barley; B: Cultivated barley; C: Wild barley

1.3 大麦叶绿体基因组的 ENC-plot 分析

ENC-plot 分析结果表明,青稞叶绿体基因的 ENC 值多集中在 30~60 之间,偏性较弱,且落在标准曲线的上方及曲线周围的散点多与下方(图 2A),同时其 ENC 和 GC_{3S} 具有显著的正相关性,相关系数为 0.57(图 3A),这表明突变和自然选择都对青稞叶绿体密码子偏性具有一定程度的影响。栽培大麦(图 2B; 图 3B)、野生大麦的情况(图 2C; 图 3C)与青稞类似,ENC 值显示其密码子使用偏性较弱,且散点多落在标准曲线的上方及周围,说明突变和自然选择对栽培和野生大麦叶绿体密码子使用偏性有影响作用,栽培大麦相关系数与青稞一致,具有显著相关性;野生大麦的 ENC 和 GC_{3S} 的相关系数为 0.63,具有显著相关性。这些结果表明,突变和自然选择在这 3 种大麦的叶绿体基因组密码子的使用偏好性中都具有一定程度的作用。

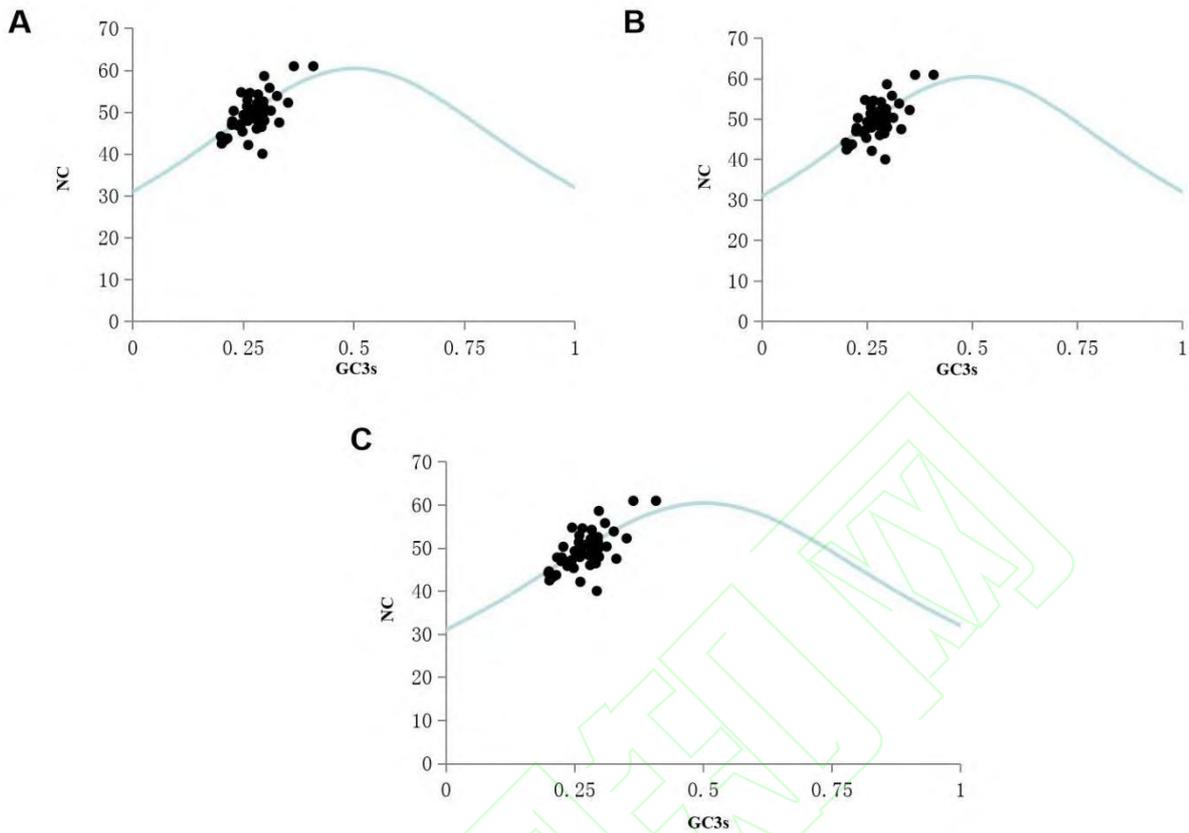


图 2 ENC-plot 绘图分析

注: A: 青稞; B: 栽培大麦; C: 野生大麦

Figure 2 Analysis of ENC and GC3 relationship

Note: A: Tibetan hulless barley; B: Cultivated barley; C: Wild barley

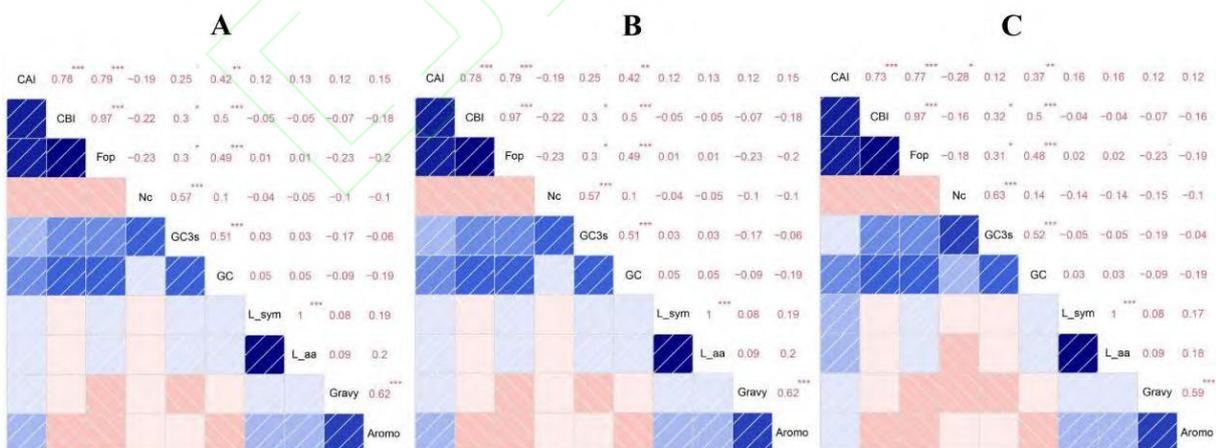


图 3 密码子相关参数的相关性分析

注: A: 青稞; B: 栽培大麦; C: 野生大麦

Figure 3 Correlation analysis of codon-related parameters

Note: A: Tibetan hulless barley; B: Cultivated barley; C: Wild barley

1.4 大麦叶绿体基因组 PR2 偏倚性分析

为了解突变和选择压力对3种大麦叶绿体基因组密码子使用的影响,对第3个密码子位置的核苷酸A、T和C、G的使用频率进行PR2-plot分析。如果密码子的使用模式仅受到突变因素的影响,那么4种碱基在密码子中的分布应为均匀,即在PR2图上所有的散点均匀的分布在4个象限中。在青稞中,基因密集分布在第四象限,在第一象限最稀疏,表明密码子第3位上使用A的频率低于T,使用C的频率低于G,密码子第3位碱基存在使用偏好性,4种碱基中T、G的使用频率更高(图4A)。在栽培大麦(图4B)和野生大麦(图4C)中,与青稞类似,均表现出T、G碱基的使用频率高于A、C的现象。表明在这3种大麦叶绿体基因组中,影响密码子偏好性的因素不只有突变,选择压力、序列长度等其他诸多因素也会对其造成影响。

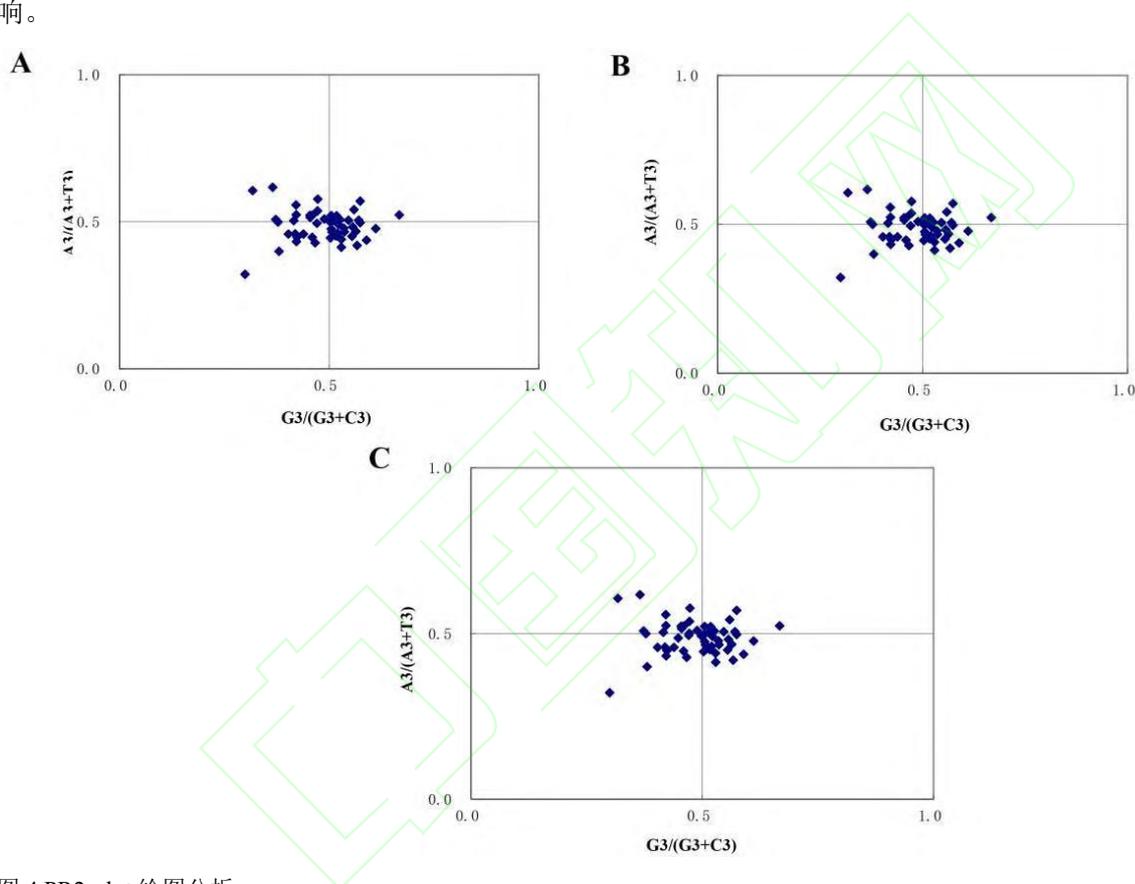


图4 PR2-plot 绘图分析

注: A: 青稞; B: 栽培大麦; C: 野生大麦

Figure 4 PR2-plot analysis of the codons

Note: A: Tibetan hulless barley; B: Cultivated barley; C: Wild barley

1.5 大麦叶绿体基因组最优密码子

对3种大麦叶绿体基因组的最优密码进行分析(表3),结果显示,在野生大麦叶绿体基因组中,共鉴定到了9个最优密码子,分别为UUU、CUA、CUG、AUU、GUA、GCU、GAU、CGC、GGA,其中77.8%(7/9)的最优密码子第3位为A/U;在青稞叶绿体基因组中,共鉴定到了12个最优密码子,分别为UUU、CUA、CUG、CCU、CCG、AUU、GUA、GCU、GAU、CGC、AAA、GGA,其中75%以A/U结尾;而栽培大麦叶绿体基因组中,有13个最优密码子,相比于青稞,多了CGU,其以A/U结尾的最优密码子占比为76.9%。结果表明,3种大麦叶绿体基因组中,大多数最优密码子以A/U结尾,具有明显的A/U偏好性。

表 3 大麦叶绿体基因组最优密码子的鉴定

Table 3 Putative optimal codons in the 3 barley chloroplast genomes

氨基酸	密码子	野生大麦	青稞	栽培大麦	氨基酸	密码子	野生大麦	青稞	栽培大麦
Amino acid	Codon	Wild barley	Tibetan hulless barley	Cultivated barley	Amino acid	Codon	Wild barley	Tibetan hulless barley	Cultivated barley
Phe	UUU	*	*	*	Tyr	UAU			
	UUC					UAC			
Leu	UUA				His	CAU			
	UUG					CAC			
	CUU				Gln	CAA			
CUC				CAG					
Ile	CUA	*	*	*	Asn	AAU			
	CUG	*	*	*		AAC			
	AUU	*	*	*	Lys	AAA		*	*
AUC				AAG					
Met	AUA				Asp	GAU	*	*	*
	AUG					GAC			
Val	GUU				Glu	GAA			
	GUC					GAG			
	GUA	*	*	*	Cys	UGU			
GUG				UGC					
Ser	UCU				Trp	UGG			
	UCC					Arg	CGU		
	UCA				CGC		*	*	*
	UCG				CGA				
Pro	CCU		*	*	CGG				
	CCC				Ser	AGU			
	CCA					AGC			
	CCG		*	*	Arg	AGA			
Thr	ACU					AGG			
	ACC				Gly	GGU			
	ACA					GGC			
	ACG					GGA	*	*	*
Ala	GCU	*	*	*	GGG				
	GCC								
	GCA								
	GCG								

注: *表示该密码子为最优密码子

Note: * represents the optimal codon

2 讨论

密码子偏好性是蛋白编码过程中对同义密码子的不均等使用,是物种特有的属性,可以反映物种进化地位及亲缘关系(张以忠等, 2022)。大量研究发现,基因突变和选择效应是影响密码子偏好性的主要因素,同时蛋白质氨基酸种类、氨基酸序列长度、基因表达水平等也会影响密码子偏好性(王婧等, 2019; Dilucca et al., 2020)。本研究对 3 种大麦叶绿体基因组的密码子偏好性分析发现,青稞和栽培大麦具有完全一致的密码子使用模式,而野生大麦与青稞和栽培大麦的密码子使用模式和偏好性具有一些差异,暗示相比于与野生大麦,青稞和栽培大麦具有更近的亲缘关系,这与前人基于核基因组的进化分析结论一致(Tao et al., 2022)。同时,中性绘图和 PR2 偏倚性分析均发现,3 种大麦叶绿体基因组的密码子偏性主要受自然选择影响,而 ENC-plot 分析表明自然选择和突变都对密码子偏好性有一定的影响。这与小麦(*Triticum aestivum*) (Zhang et al., 2012)、水稻(Chakraborty et al., 2020)、糜子(*Panicum miliaceum*) (刘慧等, 2017)、芒草(Sheng et al., 2021)等物种一致,但与秋茄(*Kandelia candel*) (赵森等, 2020)等受突变影响更大的有所不同,表明叶绿体基因组的密码子偏好性的影响因素在不同物种存在一定的差异,且与进化相关。

本研究分别在野生大麦、青稞和栽培大麦叶绿体基因组中鉴定到 9、12 和 13 个最优密码子,其中野生大麦没有特有的最优密码子,CCU、CCG、AAA 只是青稞和栽培大麦的最优密码子,而 CGU 是栽培大麦所特有的最优密码子,该研究结果表明尽管叶绿体基因组非常保守,但是最优密码子还是有一定的差异,这可能与基因转录表达活性相关(Zhang et al., 2012)。本研究结果显示,3 种大麦叶绿体的最优密码子偏好以 A/U 结尾,占比都在 75%以上。通常,在细菌或古细菌等富含 AU 的物种中,最优密码子偏好于 AU 结尾(Liu et al., 2020)。本研究结果也论证了叶绿体是由细菌通过内共生进化而来的,但有关大麦叶绿体密码子偏好形成的分子机制及其因素还需要进一步的研究。

3 材料与方法

3.1 基因序列的获取

从 NCBI Genebank 数据库中下载栽培大麦、野生大麦、青稞的叶绿体基因组序列, GeneBank 登录号分别为 KC912689、KT962228、EF115541。为提高密码子偏好性分析的准确性,将下载获得的 CDS 序列进行筛选处理,剔除重复的及长度小于 300 bp 的 CDS 序列,用于后续分析(刘慧等, 2017)。

3.2 密码子偏好性相关参数的计算

基于 CodonW 1.4.2 软件对密码子偏好性进行分析,分别计算有效密码子数(effective number of codons, ENC)、基因组密码子总 GC 含量和基因 GC 含量、密码子适应指数(codon adaptation index, CAI)、最优密码子频率(frequency of optimal codons, FOP)、密码子偏性指数(codon bias index, CBI)和相对同义密码子使用度(relative synonymous codon usage, RSCU)。利用在线网站 Emboss explorer 中的 CUSP 功能(<https://www.bioinformatics.nl/cgi-bin/emboss/cusp>)计算密码子第 1、2、3 位的 GC 含量(分别用 GC₁, GC₂, GC₃ 表示),采用 IBM SPSS Statistics 26.0 软件,基于皮尔逊相关系数(Pearson correlation coefficient)计算各参数间的相关性。

3.3 ENC-plot 分析

对于 3 个大麦叶绿体基因组中的各个基因, 根据其 GC_{3S} 值计算其理论 ENC 值, 具体公式为: 理论 $ENC = 2 + GC_{3S} + 29 / (GC_{3S}^2 + (1 - GC_{3S})^2)$ 。以理论 ENC 为纵坐标, GC_{3S} 为横坐标绘制 ENC-plot 标准曲线, 然后根据其实测 ENC 值分析各个基因在标准曲线上的分布。当突变是密码子使用模式的决定因素时, 基因应在标准曲线上方或附近; 相反, 当密码子偏好主要受到自然选择的作用时, 基因将位于标准曲线的下方(Wright, 1990)。

3.4 PR2 (Parity rule 2) 偏倚性分析

通过计算各密码子第 3 碱基位上的 A、T、G、C 的含量(分别以 A_3, T_3, G_3, C_3 表示), 研究其在突变时是否有均等几率出现。以 $G_3 / (G_3 + C_3)$ 为横坐标, $A_3 / (A_3 + T_3)$ 为纵坐标进行 PR2 偏倚分析作图。原点的意义为 $A=T$ 且 $G=C$ 。若编码基因的密码子偏好性完全由突变决定时, $A/T, G/C$ 应该分布均匀, 即出现 $A=T, G=C$, 即图中散点均匀地分布在 4 个象限中。

3.5 中性绘图分析

中性图是 GC_{12} 与 GC_3 的散点图, 通常用于分析突变和选择压力对基因碱基组成的影响(Yang et al., 2014)。若 GC_{12} 与 GC_3 之间没有显著相关性时, 表明密码子的使用偏好性受选择影响更大; 当回归系数接近 1 时, 说明密码子的 3 个碱基之间无明显差异, 其使用偏好性受突变影响更大。

3.6 密码子使用相对频率分析

密码子使用相对频率(RSCU)是密码子实际使用频率与在给定的氨基酸同义密码子家族中同等使用的所有密码子的预期使用频率的比率。当 RSCU 值接近 1.0, 表示氨基酸的所有同义密码子都被均等使用, 缺乏使用偏性; 当 RSCU 值大于 1, 则说明密码子有使用偏好, 小于 1 则说明密码子的使用频率较低(尚明照等, 2011)。

3.7 最优密码子分析

根据 ENC 分析结果对大麦叶绿体基因的 ENC 值进行大小排序, 取最高和最低 10% 的基因分别建立高、低表达基因库, 再计算两个表达库中各个密码子的 RSCU 值和 $\Delta RSCU$, 根据 $\Delta RSCU \geq 0.08$ 为标准筛选得出的密码子即为最优密码子(刘慧等, 2017)。

作者贡献

聂小军和余奎军是本研究的实验设计者; 郭奇凡和魏新雨是实验研究的执行人, 完成数据分析, 论文初稿的写作; 林汇源、张启赢、李思琦、刘晓艺和纪新明参与实验设计, 试验结果分析; 聂小军和余奎军是项目的构思者及负责人, 指导实验设计、数据分析、论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由国家自然科学基金(U22A20453)和国家大学生创新创业计划项目(202110712250)共同资助。

参考文献

- Chakraborty S., Yengkhom S., and Uddin A., 2020, Analysis of codon usage bias of chloroplast genes in *Oryza* species: codon usage of chloroplast genes in *Oryza* species, *Planta*, 252(4): 67.
Chen S.L., Lee W., Hottes A.K., Shapiro L., and McAdams H.H., 2004, Codon usage between genomes is constrained by genome-wide mutational processes, *Proc. Natl. Acad. Sci. USA*, 101(10): 3480-3485.

- Daniell H., Lin C.S., Yu M., and Chang W.J., 2016, Chloroplast genomes: diversity, evolution, and applications in genetic engineering, *Genome Biol.*, 17(1): 134.
- Dilucca M., Pavlopoulou A., Georgakilas A.G., and Pavlopoulou A., 2020, Codon usage bias in radioresistant bacteria, *Gene*, 742: 144554.
- Du B., Meenu M., Liu H.Z., and Xu B.J., 2019, A concise review on the molecular structure and function relationship of β -Glucan, *Int. J. Mol. Sci.*, 20(16): 4032.
- Grantham R., Gautier C., and Gouy M., 1980, Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type, *Nucleic Acids Res.*, 8(9): 1893-1912.
- Iriarte A., Lamolle G., and Musto H., 2021, Codon usage bias: an endless tale, *J. Mol. Evol.*, 89(9-10): 589-593.
- Li W., Zhang C.P., Guo X., Liu Q.H., and Wang K.L., 2019, Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis, *PLoS ONE*, 14(5): e0216645.
- Liu H., Wang M.X., Yue W.J., Xing G.W., Ge L.Q., Nie X.J., and Song W.N., 2017, Analysis of codon usage in the chloroplast genome of broomcorn millet (*Panicum miliaceum* L.), *Zhiwu Kexue Xuebao (Plant Science Journal)*, 35(3): 362-371. (刘慧, 王梦醒, 岳文杰, 邢光伟, 葛玲巧, 聂小军, 宋卫宁, 2017, 糜子叶绿体基因组密码子使用偏性的分析, *植物科学学报*, 35(3): 362-371.)
- Liu H.B., Lu Y.Z., Lan B.L., and Xu J.C., 2020, Codon usage by chloroplast gene is bias in *Hemiptelea davidii*, *J. Genet.*, 99: 8.
- Mazumdar P., Othman R.Y.B., Mebus K., Ramakrishnan N., and Harikrishna J.A., 2017, Codon usage and codon pair patterns in non-grass monocot genomes, *Ann. Bot.*, 120(6): 893-909.
- Rani H., and Bhardwaj R.D., 2021, Quality attributes for barley malt: "The backbone of beer", *J. Food Sci.*, 86(8): 3322-3340.
- Sato K., 2020, History and future perspectives of barley genomics, *DNA Res.*, 27(4): dsaa023.
- Shang M.Z., Liu F., Hua J.P., and Wang K.B., 2011, Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*, *Zhongguo Nongye Kexue (Scientia Agricultura Sinica)*, 44(2): 245-253. (尚明照, 刘方, 华金平, 王坤波, 2011, 陆地棉叶绿体基因组密码子使用偏性的分析, *中国农业科学*, 44(2): 245-253.)
- Sheng J.J., She X., Liu X.Y., Wang J., and Hu Z.L., 2021, Comparative analysis of codon usage patterns in chloroplast genomes of five *Miscanthus* species and related species, *PeerJ*, 9: e12173.
- Tao W.J., Bian J.X., Tang M.Q., Zeng Y., Luo R.H., Ke Q.L., Li T.T., Li Y.H., and Cui L.C., 2022, Genomic insights into positive selection during barley domestication, *BMC Plant Biol.*, 22(1): 267.
- Wang J., Wang T.Y., Wang L.Y., Zhang J.G., and Zeng Y.F., 2019, Assembling and analysis of the whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias, *Xibei Zhiwu Xuebao (Acta Botanica Boreali-Occidentalia Sinica)*, 39(9): 1559-1572. (王婧, 王天翼, 王罗云, 张建国, 曾艳飞, 2019, 沙枣叶绿体全基因组序列及其使用密码子偏性分析, *西北植物学报*, 39(9): 1559-1572.)
- Wright F., 1990, The 'effective number of codons' used in a gene, *Gene*, 87(1): 23-29.
- Wu Z.C., Wang J., Zhao Q.H., Zhu S.P., Zi C., Wu S.L., and Bao W.B., 2013, Analysis of codon bias of lipopolysaccharide-binding protein gene (*LBP*) in pigs (*Sus scrofa*), *Nongye Shengwu Jishu Xuebao (Journal of Agricultural Biotechnology)*, 21(10): 1135-1144. (吴正常, 王靖, 赵乔辉, 朱世平, 警臣, 吴圣龙, 包文斌, 2013, 猪脂多糖结合蛋白基因(*LBP*)的密码子偏好性分析, *农业技术学报*, 21(10): 1135-1144.)
- Yang X., Luo X.N., and Cai X.P., 2014, Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset, *Parasit. Vectors*, 7: 527.
- Zhang Y.R., Nie X.J., Jia X.O., Zhao C.Z., Biradar S.S., Wang L., Du X.H., and Song W.N., 2012, Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family, *Aust. J. Bot.*, 60(5): 461-470.
- Zhang Y.Z., Zeng W.Y., Deng L.Q., Zhang H.C., Liu Q.Y., Zuo T.H., Xie Q.Q., Hu D.K., Yuan C.M., Lian X.P., and Zhu L.Q., 2022, Codon usage bias analysis of S-locus genes SRK, SLG, and SP11/SCR in *Brassica oleracea*, *Zuowu Xuebao (Acta Agronomica Sinica)*, 48(5):1152-1168. (张以忠, 曾文艺, 邓琳琼, 张贺翠, 刘倩莹, 左同鸿, 谢琴琴, 胡煜科, 袁崇墨, 廉小平, 朱利泉, 2022, 甘蓝 S-位点基因 SRK、SLG 和 SP11/SCR 密码子偏好性分析, *作物学报*, 48(5): 1152-1168.)
- Zhao S., Deng L.H., and Chen F., 2020, Codon usage bias of chloroplast genome in *Kandelia obovata*, *Senlin yu Huanjing Xuebao (Journal of Forest and Environment)*, 40(5): 534-541. (赵森, 邓力华, 陈芬, 2020, 秋茄叶绿体基因组密码子使用偏好性分析, *森林与环境学报*, 40(5): 534-541.)