

引用格式: 柳婷婷, 许丽爱, 胡紫蔚, 等. 菜心叶绿体基因组特征及系统发育分析[J]. 浙江农林大学学报, 2025, 42(X): 1-11.
LIU Tingting, XU Li'ai, HU Ziwei, et al. Characteristics of the chloroplast genome and phylogenetic analysis of flowering Chinese cabbage[J]. Journal of Zhejiang A&F University, 2025, 42(X): 1-11.

菜心叶绿体基因组特征及系统发育分析

柳婷婷^{1,2}, 许丽爱³, 胡紫蔚², 熊兴鹏², 安霞¹, 曹家树²

(1. 浙江省农业科学院 浙江省园林植物与花卉研究所, 浙江 杭州 311251; 2. 浙江大学 蔬菜研究所 细胞与分子生物学实验室, 浙江 杭州 310058; 3. 浙江农林大学 园艺科学学院, 浙江 杭州 311300)

摘要: 【目的】对菜心 *Brassica campestris* var. *parachinensis* 叶绿体基因组进行组装、注释和分析, 揭示其与其他十字花科 Brassicaceae 植物的系统进化关系。【方法】利用 Illumina NovaSeq 6000 平台对菜心叶绿体全基因组测序, 利用生物信息学的手段分析其叶绿体基因组结构、重复序列、核苷酸多样性、密码子偏好性, 以及结合其他 25 种十字花科植物叶绿体基因组序列构建系统发育关系。【结果】菜心叶绿体基因组全长为 153 482 bp, GC 含量为 36.36%, 是典型的四分体结构; 共注释到 132 个编码基因, 其中 86 个蛋白质编码基因、37 个 tRNA 基因、8 个 rRNA 基因, 以及 1 个假基因; 密码子偏好性分析结果显示: 亮氨酸的使用频率最高, 相对同义密码子使用度大于 1 的密码子有 31 个, 多以 A 或者 U 结尾; 共有 37 个散在重复序列, 315 个简单重复序列, 其中单核苷酸重复占比最高, 为 72.70%; 小单拷贝区的核苷酸多样性最高, 反向重复区的核苷酸多样性最低; 系统发育分析结果显示: 菜心与其他白菜类蔬菜亲缘关系最近。【结论】菜心叶绿体基因组为典型的四分体结构, 相对保守, 与普通白菜 *B. rapa* var. *chinensis*、紫菜薹 *B. rapa* var. *purpuraria*、大白菜 *B. rapa* subsp. *pekinensis* 等亲缘关系密切。图 5 表 3 参 39

关键词: 菜心; 叶绿体基因组; 序列特征; 系统进化

中图分类号: S634 文献标志码: A 文章编号: 2095-0756(2025)00-0001-11

Characteristics of the chloroplast genome and phylogenetic analysis of flowering Chinese cabbage

LIU Tingting^{1,2}, XU Li'ai³, HU Ziwei², XIONG Xingpeng², AN Xia¹, CAO Jiashu²

(1. Zhejiang Institute of Landscape Plants and Flowers, Zhejiang Academy of Agricultural Sciences, Hangzhou 311251, Zhejiang, China; 2. Laboratory of Cell & Molecular Biology, Institute of Vegetable Science, Zhejiang University, Hangzhou 310058, Zhejiang, China; 3. College of Horticulture Science, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China)

Abstract: [Objective] This study aims to assemble, annotate, and analyze the chloroplast genome of flowering Chinese cabbage (*Brassica campestris* var. *parachinensis*), and to elucidate its phylogenetic relationships with other Brassicaceae crops. [Method] The complete chloroplast genome of flowering Chinese cabbage was sequenced using the Illumina NovaSeq 6000 platform. Bioinformatics approaches were employed to analyze its genome structure, repetitive sequences, nucleotide diversity, codon usage bias, and phylogenetic relationships by constructing phylogenetic trees with chloroplast genome sequences of other 25 Brassicaceae species. [Result] The chloroplast genome of flowering Chinese cabbage is 153 482 bp in length with a GC content of 36.36%, exhibiting a typical quadripartite structure. In total, 132 functional genes had been identified and

收稿日期: 2025-05-07; 修回日期: 2025-08-30

基金项目: 国家自然科学基金资助项目 (32202506, 32202508)

作者简介: 柳婷婷 (ORCID: 0000-0003-0673-5254), 从事白菜生长发育研究。E-mail: liutt@zaas.ac.cn。通信作者: 曹家树 (ORCID: 0000-0002-8637-987X), 教授, 博士, 从事白菜生长发育研究。E-mail: jshcao@zju.edu.cn

annotated, comprising 86 protein-coding genes, 37 tRNA genes, 8 rRNA genes, and 1 pseudogene. Codon usage bias analysis revealed that leucine (Leu) is the most frequently used amino acid, with 31 codons showing a relative synonymous codon usage (RSCU) value greater than 1, predominantly ending with A or U. Repetitive sequence analysis detected 37 dispersed repeats and 315 simple sequence repeats (SSRs), with mononucleotide repeats predominating (72.70% of total SSRs). The nucleotide diversity of small single copy region (SSC) is the highest, while that of inverted repeat region (IR) is the lowest. Phylogenetic analysis revealed that flowering Chinese cabbage exhibits the closest genetic relationship with leafy Chinese cabbage varieties. [Conclusion] The chloroplast genome of flowering Chinese cabbage exhibits a conserved quadripartite structure and demonstrates close phylogenetic relationships with *B. rapa* var. *chinensis*, *B. rapa* var. *purpuraria*, and *B. rapa* subsp. *pekinensis*. [Ch, 5 fig. 3 tab. 39 ref.]

Key words: flowering Chinese cabbage; chloroplast genome; sequence feature; systematic evolution

叶绿体是植物细胞内的关键细胞器,在光合作用和碳水化合物代谢等过程中扮演重要角色^[1]。近年来,随着高通量测序技术的飞速发展,植物叶绿体基因组的研究也在不断深入。被子植物的叶绿体基因组通常呈现为四分体结构,由 1 个大单拷贝区 (large single copy region, LSC)、1 个小单拷贝区 (small single copy region, SSC) 以及 2 个反向重复区 (inverted repeat regions A 和 B, IRa 和 IRb) 共 4 部分组成^[2]。然而,也有一些植物,如豆科 Leguminosae 的箭筈豌豆 *Vicia sativa*^[3]、三叶草 *Trifolium*^[4] 等由于缺失 1 个反向重复序列,因而其叶绿体基因组结构组成为 LSC、SSC 和 IR;而绿潮藻 *Ulva flexuosa* 的反向重复序列则完全缺失^[5]。叶绿体基因组是母系遗传,具有高度的保守性,不受核基因组的影响,被用来进行植物系统发育、遗传多样性分析等研究^[6]。

十字花科 Brassicaceae 家族十分庞大,约包含 351 个属 3 977 个种^[7]。芸薹属 *Brassica* 作为十字花科植物中的关键成员,包括白菜类 *B. campestris*、甘蓝类 *B. oleracea*、芥菜类 *B. juncea* 和甘蓝型油菜 *B. napus* 等重要的蔬菜作物和油料作物^[8]。白菜作为芸薹属中的一员,经过长期人工选择和环境变化,出现了诸多变种。菜心 *B. campestris* var. *parachinensis* 是白菜种下的一个变种,因口感好、营养丰富而备受消费者青睐,在中国南方地区尤其是华南地区多有栽培^[9-11]。然而,菜心与其他白菜变种以及十字花科成员之间的进化关联尚不明确。叶绿体基因组在一定程度上可呈现植物的进化历程,目前针对十字花科植物叶绿体基因组的研究已取得较为丰硕的成果。早在 20 多年前,模式植物拟南芥 *Arabidopsis thaliana* 的叶绿体基因组即已完成解析工作,为后续十字花科植物的遗传进化研究奠定了重要基础^[12]。近些年来,随着研究的持续深入,紫菜薹 *B. rapa* var. *purpuraria*^[13]、芥菜^[14]、萝卜 *Raphanus sativus*^[15] 等众多十字花科植物的叶绿体基因组也已相继被研究,然而关于菜心叶绿体基因组的研究还未见报道。本研究以菜心品种‘油青四九’ *B. campestris* var. *parpachinensis* ‘Youqing 49’ 为材料,借助高通量测序技术结合生物信息学方法对其叶绿体基因组进行测序、组装并注释,对其基本结构特征等进行分析,明晰了菜心与其他白菜变种以及十字花科其他成员间的进化关系。研究结果有助于弥补菜心叶绿体基因组研究空白,为十字花科植物系统发育分析及遗传资源利用提供科学依据。

1 材料与方法

1.1 材料及测序

本研究所用材料为菜心‘油青四九’,种植于浙江农林大学园艺科学学院气候室 (30°26'N, 119°72'E), 2024 年 8 月取健康植株的叶片,用锡箔纸包好后置于液氮中速冻,取出后保存于 -80 °C 冰箱。使用通用型植物 DNA 提取试剂盒 (D312) 提取总 DNA,使用 Illumina NovaSeq 6000 平台进行测序。

1.2 叶绿体基因组组装及功能注释

使用 fastp v0.23.4^[16] 软件对原始数据过滤获得净数据 (clean data)。采用 GetOrganelle v1.7.7.1^[17] 软件组装叶绿体基因组,分别运用 prodigal v2.6.3^[18]、hmmer v3.1b2^[19] 和 aragorn v1.2.38^[20] 对叶绿体的编码序列 (coding sequence, CDS) 注释,预测 rRNA 和 tRNA。叶绿体基因组图谱使用 OGDRAW^[21] 软件绘制。

1.3 散在重复序列和简单重复序列分析

通过使用 vmatch v2.3.09 (<http://www.vmatch.de/>) 软件鉴定重复序列, 相关参数设定为: 海明距离为 3, 最小长度为 30 bp, 鉴定形式为正向 (forward)、回文 (palindromic)、反向 (reverse)、互补 (complement) 共 4 种。通过 MISA v1.0 工具^[22] 对简单重复序列 (simple sequence repeats, SSRs) 进行分析。参数配置为: 单碱基重复次数不少于 8 次, 双碱基重复不少于 5 次, 而 3 碱基、4 碱基、5 碱基及 6 碱基重复则要求至少出现 3 次。

1.4 叶绿体基因组核苷酸多样性和边界分析

在美国国家生物技术信息中心 (NCBI) 上下载包括拟南芥 (KX551970)、埃塞俄比亚芥 *B. carinata* (MW628493)、芥菜 (KT581449)、油菜 (NC_016734)、黑芥 *B. nigra* (KT878383)、甘蓝 (MT721156)、白菜 (NC_040849) 等 7 种十字花科植物的叶绿体基因组。利用 MAFFT v7.427^[23] 对不同物种间的同源基因序列进行了全局比对分析, 并通过 dnasp5^[24] 工具计算核苷酸多样性 (nucleotide diversity, π)。使用集思慧远的云平台工具 CPJSdraw (<http://cloud.geneioneer.com:9929/#/tool/alltool/detail/296>) 将各物种的叶绿体基因组边界信息进行可视化处理, 使用 Mauve v2.3.1^[25] 进行基因组比对。

1.5 系统进化分析

从 NCBI 数据库下载 25 种十字花科植物的叶绿体基因组序列, 外类群选择番木瓜 *Carica papaya*, 以此来分析系统发育关系。分别使用 MAFFT v7.427^[23] 和 RAxML v8.2.10^[26] 进行多序列比对和进化树构建。

2 结果与分析

2.1 菜心叶绿体基因组基本特征

菜心叶绿体基因组 (GenBank 登录号: PX244600) 呈典型的四分体结构, 长度为 153 482 bp, 包含反向重复序列 a (IRa) 和反向重复序列 b (IRb) 以及大单拷贝区域 (LSC) 和小单拷贝区域 (SSC), 序列长度分别为 26 213、26 213、83 281、17 775 bp (图 1 和表 1)。对碱基组成进行分析, 结果显示: 菜心叶绿体基因组中 A、C、G、T 的含量分别为 31.35%、18.51%、17.85%、32.39%; 整个菜心叶绿体基因组中的 GC 总含量为 36.36%, 其中 IRa 和 IRb 的 GC 含量均为 42.34%, 高于 LSC (34.12%) 和 SSC (29.20%) 中的 GC 含量 (表 1)。

2.2 菜心叶绿体基因的功能注释

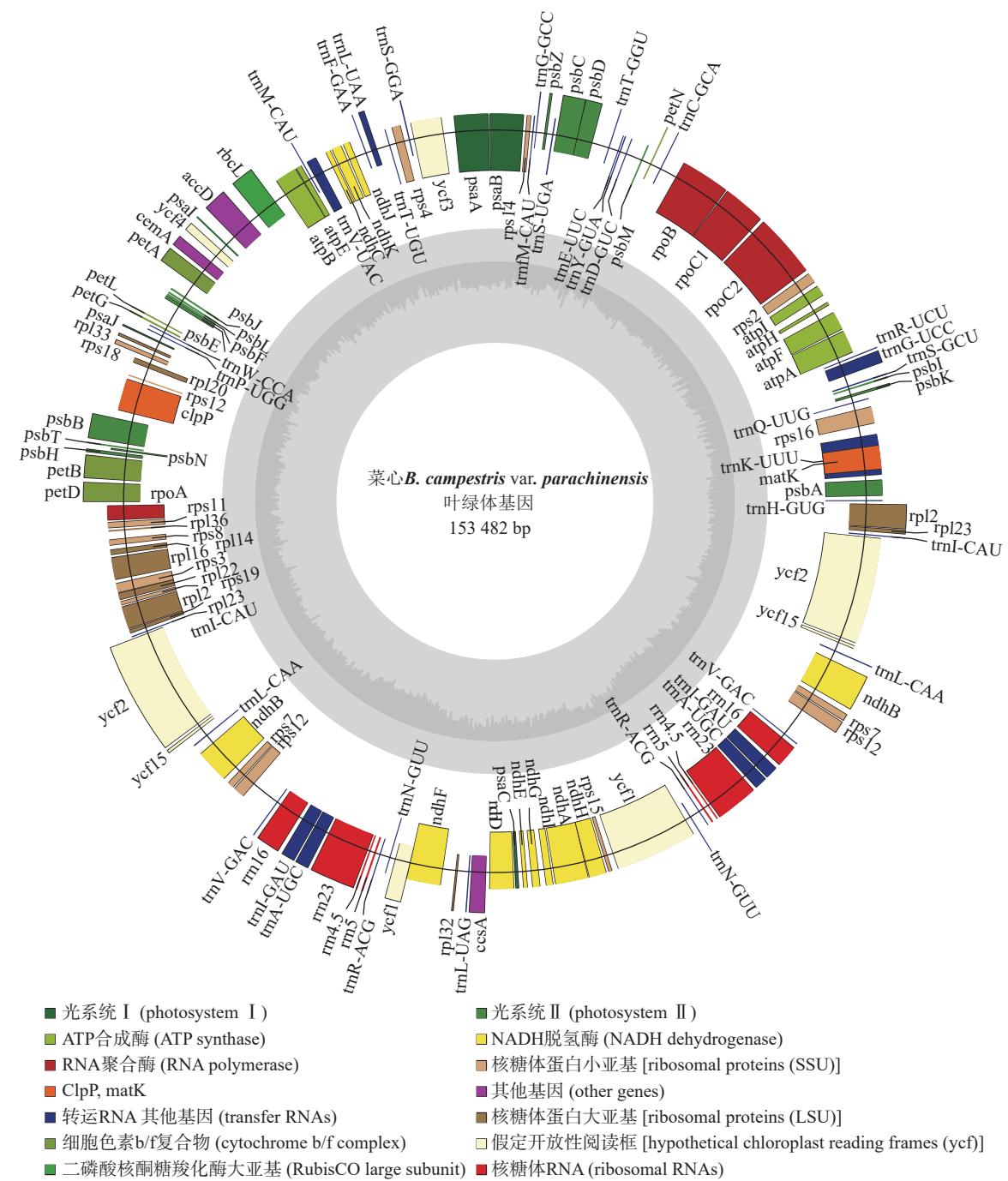
在菜心的叶绿体基因组中, 共发现了 132 个基因, 包括 86 个蛋白质编码基因、37 个 tRNA 基因、8 个 rRNA 基因, 以及 1 个假基因。这些基因的主要功能涉及光合作用和自我复制等过程, 此外还有部分基因的功能尚未明确 (表 2)。在菜心叶绿体基因组中, 72 个蛋白质编码基因、23 个 tRNA 的拷贝数为 1, 7 个蛋白质编码基因、7 个 tRNA、4 个 rRNA 的拷贝数为 2; 11 个蛋白质编码基因、8 个 tRNA 的内含子数量为 1, 4 个蛋白质编码基因的内含子数量为 2 (表 2)。

2.3 密码子偏好性分析

分析菜心叶绿体基因组发现: 总计 22 799 个密码子被用于编码氨基酸, 其中编码亮氨酸 (Leu) 的密码子最为常用, 数量高达 2 421 个; 其次为编码异亮氨酸 (Ile) 和丝氨酸 (Ser) 的密码子, 数量分别为 1 980 和 1 699 个 (表 3)。对其同义密码子使用度 (relative synonymous codon usage, RSCU) 进行分析, 发现 RSCU 大于 1 的密码子有 31 个, 其中以 A 或者 U 结尾的有 29 个; RSCU 小于 1 的密码子有 34 个, 其中以 G 或者 C 结尾的 31 个; 编码色氨酸 (Trp) 的密码子只有 1 个 (UGG), 因此其 RSCU 等于 1; AUG 密码子不仅是起始密码子, 还编码甲硫氨酸 (Met), 因此 RSCU 最高, 为 6.97; 其次为编码 Leu 的 UUA 和编码丙氨酸 (Ala) 的 GCU, RSCU 分别为 2.26 和 1.89; RSCU 最低的为编码 Met 的 GUG 和 UUG, 为 0.01 (表 3)。

2.4 重复序列分析

散在重复序列在基因组里以分散的形式存在。菜心叶绿体基因组中包括正向重复 14 个、回文重复 18 个、反向重复 3 个、互补重复 2 个共 37 个散在重复序列。其中 36 个散在重复序列长度为 30~58 bp, 长度是 30 bp 的散在重复序列数量最多, 共有 11 个, 其次是长度为 32 bp 的散在重复序列, 共有 5 个;



外圈显示正向编码的基因，内圈展示反向编码的基因；不同功能组由彩色条带表示；内部的深灰色环状图代表CG含量的比例；浅灰色环状图则表示AT含量的比例。

图 1 菜心叶绿体基因组图谱

Figure 1 Map of *B. campestris* var. *parachinensis* chloroplast genome

表 1 菜心叶绿体基因组特征

Table 1 Chloroplast genome characteristics of *B. campestris* var. *parachinensis*

区域	A 含量/%	C 含量/%	G 含量/%	T 含量/%	GC 含量/%	碱基长度/bp
LSC	32.05	17.54	16.58	33.83	34.12	83 281
SSC	35.49	15.16	14.04	35.31	29.2	17 775
IRa	28.91	22.03	20.31	28.76	42.34	26 213
IRb	28.76	20.31	22.03	28.91	42.34	26 213
总量	31.35	18.51	17.85	32.29	36.36	153 482

表 2 菜心叶绿体基因组编码基因注释信息

Table 2 Gene annotation of the chloroplast genome of <i>B. campestris</i> var. <i>parachinensis</i>		
基因分类	基因分组	基因列表
光合作用	光系统 I 亚基	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC</i> 、 <i>psaI</i> 、 <i>psaJ</i>
	光系统 II 亚基	<i>psbA</i> 、 <i>psbB</i> 、 <i>psbC</i> 、 <i>psbD</i> 、 <i>psbE</i> 、 <i>psbF</i> 、 <i>psbH</i> 、 <i>psbI</i> 、 <i>psbJ</i> 、 <i>psbK</i> 、 <i>psbL</i> 、 <i>psbM</i> 、 <i>psbN</i> 、 <i>psbT</i> 、 <i>psbZ</i>
	NADH脱氢酶亚基	<i>ndhA</i> *, <i>ndhB</i> *(2)、 <i>ndhC</i> 、 <i>ndhD</i> 、 <i>ndhE</i> 、 <i>ndhF</i> 、 <i>ndhG</i> 、 <i>ndhH</i> 、 <i>ndhI</i> 、 <i>ndhJ</i> 、 <i>ndhK</i>
	色素细胞b/f复合物亚基	<i>petA</i> 、 <i>petB</i> *, <i>petD</i> *, <i>petG</i> 、 <i>petL</i> 、 <i>petN</i>
	ATP合成酶亚基	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF</i> *, <i>atpH</i> 、 <i>atpI</i>
	Rubisco大亚基	<i>rbcL</i>
	核糖体蛋白大亚基	<i>rpl14</i> 、 <i>rpl16</i> *, <i>rpl2</i> *(2)、 <i>rpl20</i> 、 <i>rpl22</i> 、 <i>rpl23</i> (2)、 <i>rpl32</i> 、 <i>rpl33</i> 、 <i>rpl36</i>
	核糖体蛋白小亚基	<i>rps11</i> 、 <i>rps12</i> ***(2)、 <i>rps14</i> 、 <i>rps15</i> 、 <i>rps16</i> *, <i>rps18</i> 、 <i>rps19</i> 、 <i>rps2</i> 、 <i>rps3</i> 、 <i>rps4</i> 、 <i>rps7</i> (2)、 <i>rps8</i>
	RNA聚合酶亚基	<i>rpoA</i> 、 <i>rpoB</i> 、 <i>rpoC1</i> *, <i>rpoC2</i>
	核糖体RNAs	<i>rrn16</i> (2)、 <i>rrn23</i> (2)、 <i>rrn4.5</i> (2)、 <i>rrn5</i> (2)
复制基因	转运RNAs	<i>trnA</i> -UGC*(2)、 <i>trnC</i> -GCA、 <i>trnD</i> -GUC、 <i>trnE</i> -UUC、 <i>trnF</i> -GAA、 <i>trnG</i> -GCC、 <i>trnG</i> -UCC*、 <i>trnH</i> -GUG、 <i>trnI</i> -CAU(2)、 <i>trnI</i> -GAU*(2)、 <i>trnK</i> -UUU*, <i>trnL</i> -CAA(2)、 <i>trnL</i> -UAA*, <i>trnL</i> -UAG、 <i>trnM</i> -CAU、 <i>trnN</i> -GUU(2)、 <i>trnP</i> -UGG、 <i>trnQ</i> -UUG、 <i>trnR</i> -ACG(2)、 <i>trnR</i> -UCU、 <i>trnS</i> -GCU、 <i>trnS</i> -GGA、 <i>trnS</i> -UGA、 <i>trnT</i> -GGU、 <i>trnT</i> -UGU、 <i>trnV</i> -GAC(2)、 <i>trnV</i> -UAC*、 <i>trnW</i> -CCA、 <i>trnY</i> -GUA、 <i>trnY</i> -CAU
其他基因	成熟酶	<i>matK</i>
	蛋白酶	<i>clpP</i> **
	被膜蛋白	<i>cemA</i>
	乙酰辅酶A羧化酶	<i>accD</i>
未知功能基因	色素细胞C合成酶	<i>ccsA</i>
未知功能基因	保守的开放阅读框	# <i>ycf1</i> 、 <i>yef1</i> 、 <i>yef15</i> (2)、 <i>yef2</i> (2)、 <i>yef3</i> **, <i>yef4</i>

说明：(2)代表多拷贝基因；*表示带1个内含子的基因；**表示带2个内含子的基因；#表示假基因。

表 3 菜心叶绿体基因组同义密码子使用度 (RSCU) 分析

Table 3 Relative synonymous codon usage (RSCU) analysis of <i>B. campestris</i> var. <i>parachinensis</i>											
氨基酸	密码子	数量	RSCU	氨基酸	密码子	数量	RSCU	氨基酸	密码子	数量	RSCU
Ter	UAA	48	1.82	Ile	AUC	350	0.53	Arg	AGG	126	0.57
Ter	UAG	19	0.72	Ile	AUU	1005	1.52	Arg	CGA	296	1.35
Ter	UGA	12	0.46	Lys	AAA	997	1.57	Arg	CGC	96	0.44
Ala	GCA	348	1.12	Lys	AAG	276	0.43	Arg	CGG	102	0.47
Ala	GCC	177	0.57	Leu	CUA	323	0.80	Arg	CGU	311	1.42
Ala	GCG	131	0.42	Leu	CUC	149	0.37	Ser	AGC	100	0.35
Ala	GCU	586	1.89	Leu	CUG	139	0.34	Ser	AGU	358	1.26
Cys	UGC	62	0.46	Leu	CUU	498	1.23	Ser	UCA	340	1.20
Cys	UGU	207	1.54	Leu	UUA	872	2.16	Ser	UCC	235	0.83
Asp	GAC	170	0.39	Leu	UUG	440	1.09	Ser	UCG	156	0.55
Asp	GAU	698	1.61	Met	AUG	512	6.97	Ser	UCU	510	1.80
Glu	GAA	928	1.55	Met	GUG	1	0.01	Thr	ACA	367	1.23
Glu	GAG	270	0.45	Met	UUG	1	0.01	Thr	ACC	212	0.71
Phe	UUC	401	0.59	Asn	AAC	244	0.44	Thr	ACG	123	0.41
Phe	UUU	967	1.41	Asn	AAU	855	1.56	Thr	ACU	490	1.64
Gly	GGA	624	1.61	Pro	CCA	266	1.16	Val	GUA	455	1.45
Gly	GGC	153	0.39	Pro	CCC	161	0.70	Val	GUC	153	0.49
Gly	GGG	254	0.66	Pro	CCG	124	0.54	Val	GUG	169	0.54
Gly	GGU	519	1.34	Pro	CCU	366	1.60	Val	GUU	475	1.52
His	CAC	125	0.49	Gln	CAA	642	1.57	Trp	UGG	396	1.00
His	CAU	383	1.51	Gln	CAG	175	0.43	Tyr	UAC	155	0.37
Ile	AUA	625	0.95	Arg	AGA	384	1.75	Tyr	UAU	687	1.63

此外还有 1 个最长的散在重复序列, 长度为 26 213 bp (图 2A)。

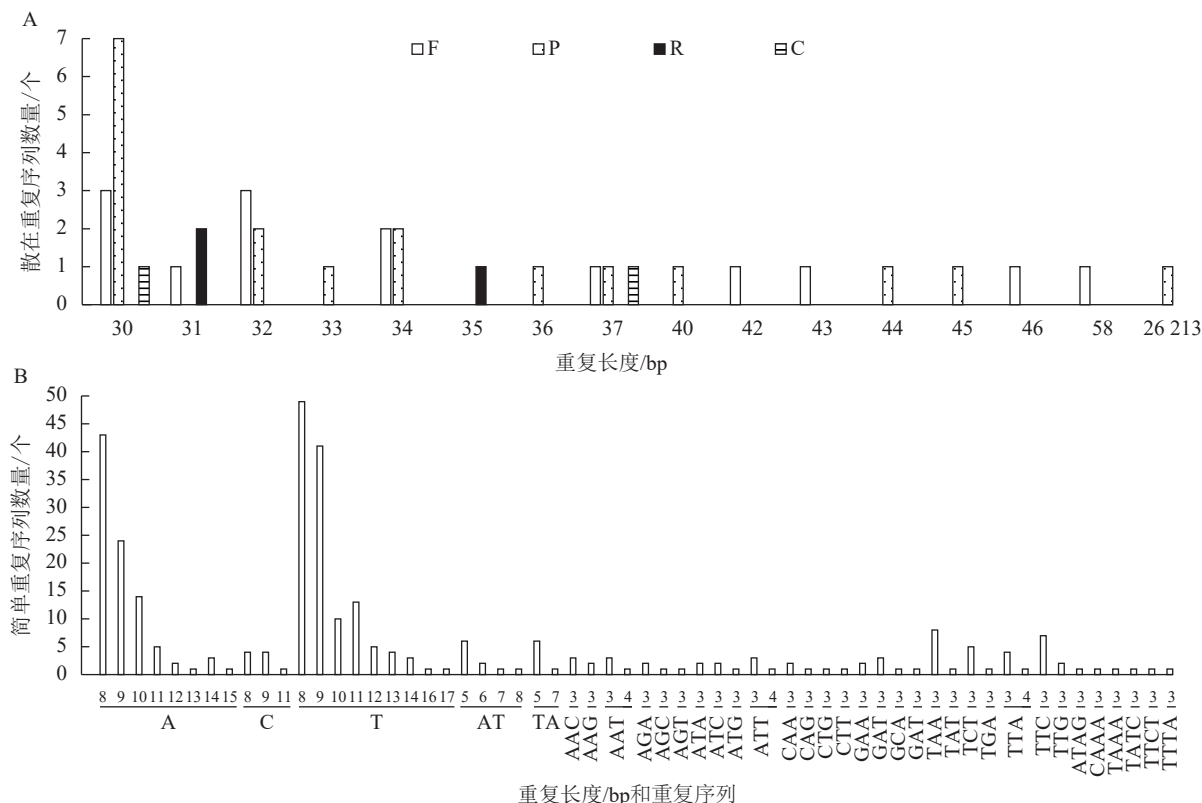


图 2 菜心叶绿体基因组散在重复序列 (A) 和简单重复序列 (B) 分析

Figure 2 Analysis of scattered sequence repeats (A) and simple sequence repeats (B) in the chloroplast genome of *B. campestris* var. *parachinensis*

SSRs 是由 1~6 个核苷酸组成的短序列作为重复单元串联而成的一类序列。在菜心叶绿体基因组中共存在 315 个 SSRs, 其中 LSC 含有 197 个, SSC 含有 72 个, IR 含有 46 个; 在这些 SSRs 中, 单核苷酸重复、双核苷酸重复、3 核苷酸重复和 4 核苷酸重复的数量分别为 229、17、63 和 6 个。进一步分析显示: 在全部 315 个 SSRs 中, 占比排名前 3 位的类型依次为 T(8)、A(8)、T(9), 分别为 15.56%(49)、13.65%(43)、13.0%(41)(图 2B)。

2.5 核苷酸多样性和边界分析

菜心、拟南芥、芥菜等 8 种十字花科植物的叶绿体基因组平均核苷酸多样性为 0.007 2 (图 3)。SSC 区域的平均核苷酸多样性最高 (0.012 3), LSC 次之 (0.007 4), IR 区域的核苷酸多样性最低 (0.002 1)。利用核苷酸多样性筛选出 4 个高变异区域 ($P_i \geq 0.02$), 发现 LSC 区域含有 3 个位点, 分别为 *rps16*、*matK*、*rpl22*, SSC 区域含有 1 个位点, 为 *ccsA*, 其中变异度最高的位点为 *rps16*。

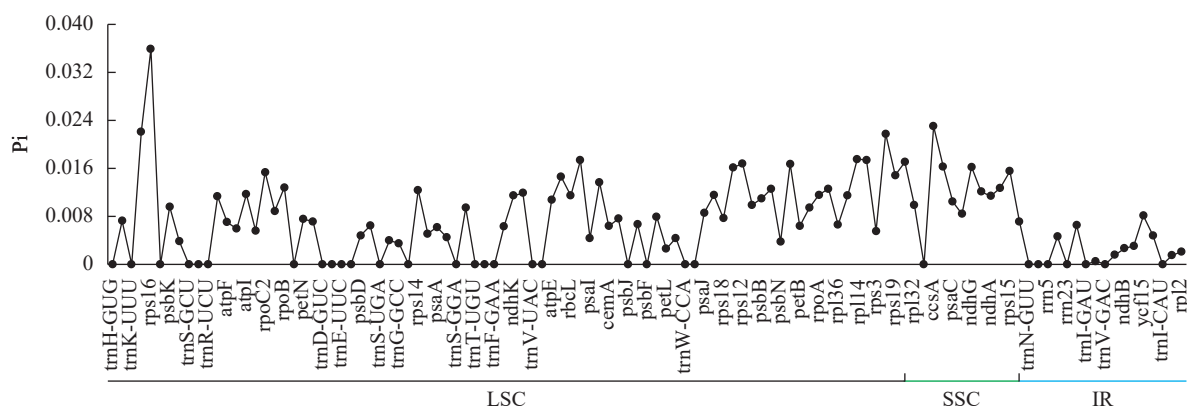


图 3 基因核苷酸多样性 P_i 折线图

Figure 3 Line chart of gene nucleotide diversity (P_i) value

在植物叶绿体基因组的进化过程中，导致其大小差异的主要因素是 IR 边界的扩张与收缩。边界分析结果表明：菜心等 8 种十字花科植物的叶绿体基因组均存在 4 个边界，即 JLB (LSC/IRb)、JSB (IRb/SSC)、JSA (SSC/IRa)、JLA (IRa/LSC)。 *rps19*、*rpl2*、*ycf1*、*ndhF*、*trnN* 和 *trnH* 是位于 IR 边界附近的主要基因。在这 8 种十字花科植物中，JLB 边界均位于 *rps19* 基因编码区内，且在不同植物叶绿体基因组中仅存在 1~2 bp 的位置差异；JSB 边界均位于 *ycf1* 和 *ndhF* 基因编码区内，两者有 36~37 bp 的重叠区域，其中 *ycf1* 基因的大部分区域位于 IRb，只有 2~3 bp 位于 SSC；JSA 边界均位于 *ycf1* 基因编码区内，其中 1 027~1 030 bp 位于 IRa，4 271~4 358 bp 位于 SSC；*trnH-GUG* 均位于 LSC，距离 JLA 边界 2~30 bp (图 4)。以上结果显示菜心等 8 种十字花科植物叶绿体基因组高度保守，IR 边界总体变化幅度较小，仅涉及少数基因。

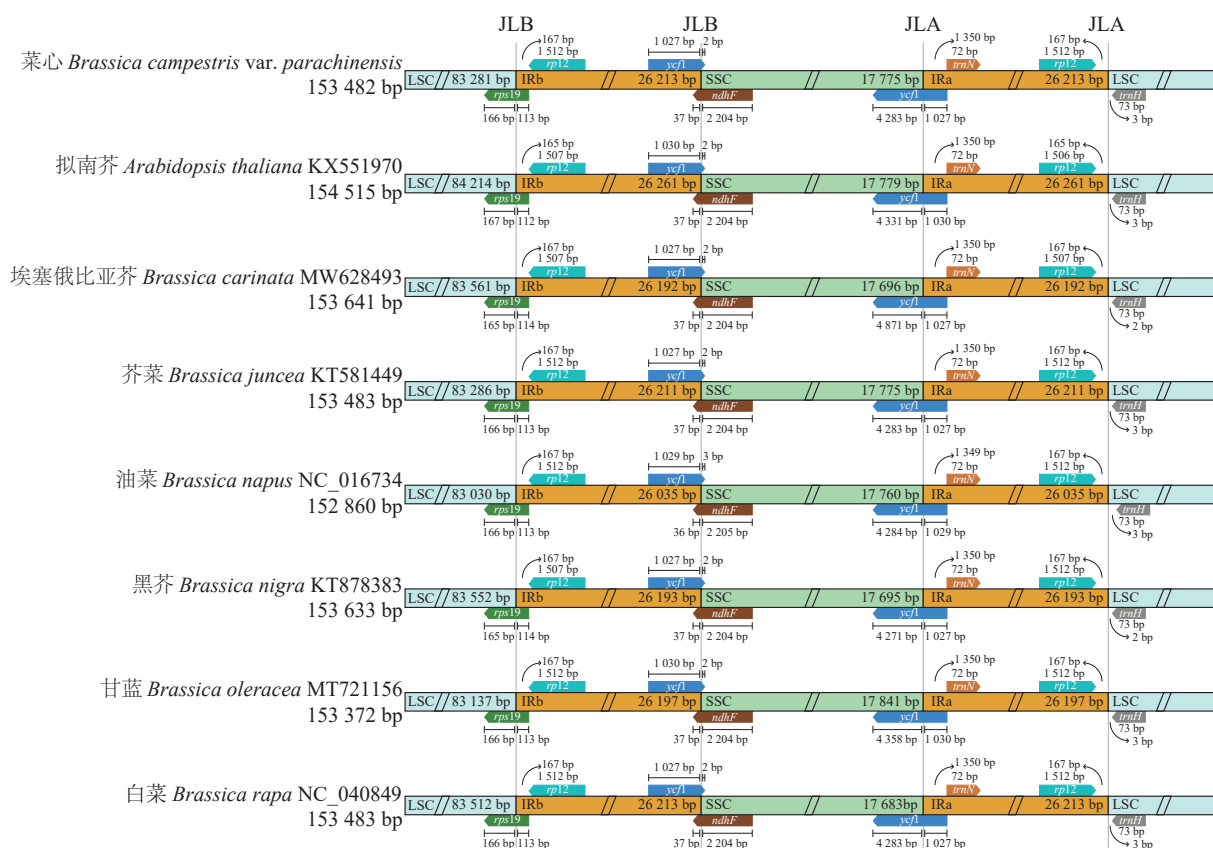


图 4 8 种十字花科植物的 IR/SC 边界分析

Figure 4 IR/SC boundary analysis of 8 Brassicaceae crops

2.6 系统进化分析

为了更全面地了解十字花科植物之间的进化关系，从 NCBI 数据库中下载了 25 种十字花科植物的叶绿体基因组数据。在此基础上，以番木瓜作为外类群构建系统发育树。结果显示：与菜心亲缘关系最近的为普通白菜，其次为紫菜薹、大白菜、芜菁、芥菜、甘蓝、油菜等芸薹属作物；再接下来与菜心亲缘关系近的是萝卜，表明在十字花科植物中，萝卜是与菜心等芸薹属作物进化上接近的外属作物，反之，同为芸薹属的黑芥和埃塞俄比亚芥与菜心的亲缘关系则相对较远。番木瓜作为外类群，与菜心的亲缘关系最远 (图 5)。

3 讨论

叶绿体是植物遗传系统中的半自主性细胞器，其基因组的基因数量、基因组成和基因排列比线粒体基因组和核基因组更为保守^[27]。本研究利用高通量测序分析了菜心的叶绿体基因组，发现菜心叶绿体基因组呈典型的四分体结构，全长为 153 482 bp，GC 含量为 36.36%，与白菜类蔬菜中的普通白菜 *B. rapa* subsp. *chinensis*、大白菜 *B. rapa* subsp. *pekinensis*、紫菜薹的叶绿体基因组长度和 GC 含量一致；比芜菁

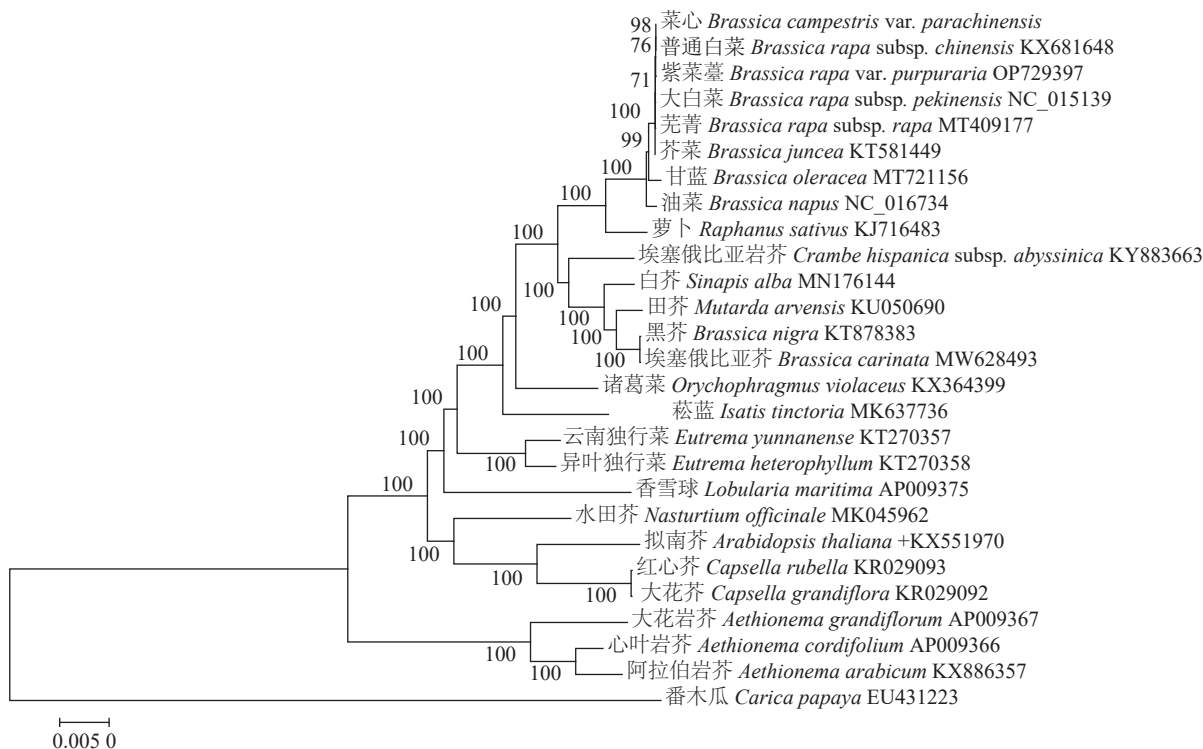


图5 基于叶绿体基因组序列构建的系统发育树

Figure 5 Phylogenetic tree constructed based on chloroplast genome sequences

叶绿体基因组短 139 bp, GC 含量高了 0.02%^[13]; 与芸薹属中的芥菜、黑芥、埃塞俄比亚芥等作物以及模式植物拟南芥的叶绿体基因组长度和 GC 含量差距也不明显^[12, 14, 28-29]。由此可见, 在白菜及芸薹属内, 各个物种的叶绿体基因组相对保守, 这个结果与 ZHAO 等^[30]的研究结果一致。

菜心叶绿体基因组共鉴定出 132 个基因, 其中 45 个基因与光合作用相关, 与已发表的芸薹属物种相比, 其基因种类和数量未表现出显著差异, 表明叶绿体基因组核心功能基因在进化过程中高度保守^[31]。在植物的进化历程中, 不同密码子的使用通常显示出一定的偏好性。在菜心叶绿体基因组中, 使用频率最高的氨基酸为 Leu, 使用频率最低的氨基酸为半胱氨酸 (Cys), 这与凤梨 *Ananas comosus* var. *comosus*^[32]、宝华玉兰 *Magnolia zenii*^[33] 等植物的研究结果一致。相对同义密码子使用频率 (RSCU) 是通过比较某个特定密码子的实际出现频率与其理论预期频率来定义的, 是评估密码子偏好性的有效工具, RSCU 大于 1 代表该密码子的使用具有明显偏好^[34]。在菜心叶绿体基因组中, 绝大多数 (93.55%) RSCU 大于 1 的密码子以 A 或者 U 结尾, 绝大多数 (93.55%) RSCU 小于 1 的密码子以 A 或者 U 结尾, 类似现象在被子植物的叶绿体基因组中普遍存在, 说明叶绿体基因组的密码子使用频率也有高度的保守性^[35-36]。SSRs 被广泛用作构建遗传连锁图谱、种群遗传分析等。ZHAO 等^[30]在芸薹 *Brassica oleracea* var. *gongylodes* 叶绿体基因组中共鉴定出 290 个 SSRs, WU 等^[37]在冬季油菜中发现了 288 个 SSRs, 本研究在菜心叶绿体基因组中共发现 315 个 SSRs, 可为研究芸薹属作物的遗传多样性提供潜候选分子标记。

核苷酸多样性是用来衡量一个群体内遗传变异程度的重要指标, 其值越高说明群体内的遗传多样性越丰富, 能够为种群遗传学提供潜在的分子标记^[38]。本研究利用菜心、拟南芥等 8 种十字花科植物进行比较分析, 发现 8 个物种叶绿体基因组的平均核苷酸多样性为 0.007 2, 不同区域平均核苷酸多样性从大到小依次为 LSC (0.007 4)、SSC (0.012 3)、IR (0.002 1), 说明 IR 相比其他 2 个区域更为保守; 其中核苷酸多样性最高的前 4 个位点依次为位于 LSC 区域的 *rps16*, SSC 区域的 *ccsA*, LSC 区域的 *matK*、*rpl22*, 这些高度变异位点可以用作十字花科物种鉴定的分子标记。在植物基因组进化过程中, IR 区的扩张或收缩是叶绿体基因组结构变异的主要驱动力, 可为物种鉴定和系统发育研究提供分子依据^[39]。对包括菜心在内的 8 个十字花科物种的 IR 边界进行分析, 发现 IR 边界的差异主要跟 *rps19*、*ycf1*、*ndhF*、

trnH 的位置有关, 但整体变化较小, 说明菜心等十字花科植物的叶绿体基因组比较保守。

基于叶绿体基因组, 利用最大似然法对菜心进行的系统发育分析, 为菜心在十字花科中的进化位置提供了依据。本研究结果显示: 菜心与普通白菜、紫菜薹、大白菜、芜菁等白菜类蔬菜亲缘关系最近, 其次为芥菜、甘蓝、油菜等芸薹属作物, 而同为芸薹属的黑芥和埃塞俄比亚芥与菜心的亲缘关系则较远。WANG 等^[31] 基于叶绿体基因组对芥蓝 *B. oleracea* var. *alboglabra* 等十字花科作物进行系统发育分析也发现了同样的现象, 芥蓝与油菜、白菜、芥菜、甘蓝聚为一个类群, 黑芥和埃塞俄比亚芥聚为另一个类群。

4 结论

菜心叶绿体基因组呈典型的四分体结构, 全长为 153 482 bp, 包含 86 个蛋白质编码基因、37 个 tRNA 基因和 8 个 rRNA 基因, 基因功能主要为光合作用、自我复制等; 其基因组密码子多以 A/U 结尾, 共有 37 个散在重复序列, 315 个 SSR; 系统发育分析结果显示: 普通白菜与菜心的亲缘关系最近。本研究有助于深入理解菜心的系统发育关系, 可为菜心及其他十字花科植物的遗传改良、育种策略等提供理论依据。

5 参考文献

- [1] 熊瑶, 童凌斐, 曹岚, 等. 4 种忍冬属植物叶绿体基因组结构特征及系统发育分析[J]. 药学学报, 2024, **59**(11): 3164–3171.
XIONG Yao, TONG Lingfei, CAO Lan, *et al.* Structural characteristics and phylogenetic analysis of chloroplast genomes of four species of *Lonicera* [J]. *Acta Pharmaceutica Sinica*, 2024, **59**(11): 3164–3171.
- [2] YIN Dongmei, WANG Yun, ZHANG Xingguo, *et al.* Development of chloroplast genome resources for peanut (*Arachis hypogaea* L.) and other species of *Arachis* [J/OL]. *Scientific Reports*, 2017, **7**: 11649 [2025-04-30]. DOI: [10.1038/s41598-017-12026-x](https://doi.org/10.1038/s41598-017-12026-x).
- [3] 温静, 郭茂伟, 朱琳, 等. 黄花型扁蓊豆叶绿体基因组及其系统发育分析[J]. 中国草地学报, 2024, **46**(8): 10–18.
WEN Jing, GUO Maowei, ZHU Lin, *et al.* Analysis of chloroplast genome and phylogeny of yellow flowered type *Medicago ruthenica* [J]. *Chinese Journal of Grassland*, 2024, **46**(8): 10–18.
- [4] SVEINSSON S, CRONK Q. Evolutionary origin of highly repetitive plastid genomes within the clover genus (*Trifolium*) [J/OL]. *BMC Evolutionary Biology*, 2014, **14**(1): 228 [2025-04-30]. DOI: [10.1186/s12862-014-0228-6](https://doi.org/10.1186/s12862-014-0228-6).
- [5] CAI Chuner, WANG Lingke, ZHOU Lingjie, *et al.* Complete chloroplast genome of green tide algae *Ulva flexuosa* (Ulvophyceae, Chlorophyta) with comparative analysis [J/OL]. *PLoS One*, 2017, **12**(9): e0184196 [2025-04-30]. DOI: [10.1371/journal.pone.0184196](https://doi.org/10.1371/journal.pone.0184196).
- [6] DANIELL H, LIN C S, YU Ming, *et al.* Chloroplast genomes: diversity, evolution, and applications in genetic engineering [J/OL]. *Genome Biology*, 2016, **17**(1): 134 [2025-04-30]. DOI: [10.1186/s13059-016-1004-2](https://doi.org/10.1186/s13059-016-1004-2).
- [7] LIU Liangmin, DU Xinyu, GUO Cen, *et al.* Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data [J]. *Journal of Systematics and Evolution*, 2021, **59**(3): 442–453.
- [8] YANG Jinghua, LIU Dongyuan, WANG Xiaowu, *et al.* The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection [J]. *Nature Genetics*, 2016, **48**(10): 1225–1232.
- [9] 李晓晴, 杨佳雨, 陈少明, 等. 氮肥与生物炭、有机肥配施对菜心生长及养分积累的影响[J]. 植物营养与肥料学报, 2024, **30**(11): 2150–2160.
LI Xiaoqing, YANG Jiayu, CHEN Shaoming, *et al.* Effects of nitrogen fertilizer combined with biochar or organic fertilizer on the growth and nutrient accumulation of Chinese flowering cabbage [*Brassica campestris* L. ssp. *chinensis* (L.) var. *utilis* Tsen et Lee [J]. *Journal of Plant Nutrition and Fertilizers*, 2024, **30**(11): 2150–2160.
- [10] 原远, 王春政, 周贤玉, 等. 六种菜心氨基酸组成及营养价值评价[J]. 南方农业学报, 2019, **50**(10): 2271–2277.
YUAN Yuan, WANG Chunzheng, ZHOU Xianyu, *et al.* Amino acid composition and nutritional value evaluation of six kinds of flowering Chinese cabbages (*Brassica campestris* L. ssp. *chinensis* var. *utilis* Tsen et Lee) [J]. *Journal of Southern Agriculture*, 2019, **50**(10): 2271–2277.

- [11] 朱云娜, 杨庆庆, 寇尔丰, 等. 水杨酸处理对菜心生长发育及其品质的影响[J]. 北方园艺, 2020(22): 53–59.
ZHU Yunna, YANG Qingqing, KOU Erfeng, *et al.* Effects of salicylic acid on plant growth development and qualities in flowering Chinese cabbage [J]. *Northern Horticulture*, 2020(22): 53–59.
- [12] SATO S, NAKAMURA Y, KANEKO T, *et al.* Complete structure of the chloroplast genome of *Arabidopsis thaliana* [J]. *DNA Research*, 1999, **6**(5): 283–290.
- [13] 王传之, 周贤玉, 李扬眉, 等. 紫菜薹叶绿体全基因组序列及其系统发育分析[J]. 西北农业学报, 2024, **33**(8): 1483–1494.
WANG Chuanzhi, ZHOU Xianyu, LI Yangmei, *et al.* Chloroplast genome sequence and phylogenetic analysis of *Brassica rapa* var. *purpuraria* [J]. *Acta Agriculturae Boreali-occidentalis Sinica*, 2024, **33**(8): 1483–1494.
- [14] PRABHUDAS S K, RAJU B, KANNAN THODI S, *et al.* The complete chloroplast genome sequence of Indian mustard (*Brassica juncea* L.) [J]. *Mitochondrial DNA Part A*, 2016, **27**(6): 4622–4623.
- [15] JEONG Y M, CHUNG W H, MUN J H, *et al.* De novo assembly and characterization of the complete chloroplast genome of radish (*Raphanus sativus* L.) [J]. *Gene*, 2014, **551**(1): 39–48.
- [16] CHEN Shifu, ZHOU Yanqing, CHEN Yaru, *et al.* Fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, **34**(17): i884–i890.
- [17] ANTON B, SERGEY N, DMITRY A, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing [J]. *Journal of Computational Biology*, 2012, **19**(5): 455–477.
- [18] HYATT D, CHEN G L, LoCASCIO P F, *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification [J/OL]. *BMC Bioinformatics*, 2010, **11**(1): 119 [2025-04-30]. DOI: [10.1186/1471-2105-11-119](https://doi.org/10.1186/1471-2105-11-119).
- [19] MISTRY J, FINN R D, EDDY S R, *et al.* Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions [J/OL]. *Nucleic Acids Research*, 2013, **41**(12): e121 [2025-04-30]. DOI: [10.1093/nar/gkt263](https://doi.org/10.1093/nar/gkt263).
- [20] LASLETT D, CANBACK B. ARAGORN a program to detect tRNA genes and tmRNA genes in nucleotide sequences [J]. *Nucleic Acids Research*, 2004, **32**(1): 11–16.
- [21] GREINER S, LEHWARK P, BOCK R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Research*, 2019, **47**(W1): W59–W64.
- [22] BEIER S, THIEL T, MÜNCH T, *et al.* MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 2017, **33**(16): 2583–2585.
- [23] NAKAMURA T, YAMADA K D, TOMII K, *et al.* Parallelization of MAFFT for large-scale multiple sequence alignments [J]. *Bioinformatics*, 2018, **34**(14): 2490–2492.
- [24] LIBRADO P, ROZAS J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data [J]. *Bioinformatics*, 2009, **25**(11): 1451–1452.
- [25] DARLING A C E, MAU B, BLATTNER F R, *et al.* Mauve: multiple alignment of conserved genomic sequence with rearrangements [J]. *Genome Research*, 2004, **14**(7): 1394–1403.
- [26] SILVESTRO D, MICHALAK I. raxmlGUI: a graphical front-end for RAXML [J]. *Organisms Diversity & Evolution*, 2012, **12**(4): 335–337.
- [27] 王杰, 贺文闯, 向坤莉, 等. 基因组时代的植物系统发育研究进展[J]. 浙江农林大学学报, 2023, **40**(1): 227–236.
WANG Jie, HE Wenchuang, XIANG Kunli, *et al.* Advances in plant phylogeny in the genome era [J]. *Journal of Zhejiang A&F University*, 2023, **40**(1): 227–236.
- [28] SEOL Y J, KIM K, KANG S H, *et al.* The complete chloroplast genome of two *Brassica* species, *Brassica nigra* and *B. oleracea* [J]. *Mitochondrial DNA Part A, DNA Mapping, Sequencing, and Analysis*, 2017, **28**(2): 167–168.
- [29] ZHU Bin, GAO Zuomin, QIAN Fang, *et al.* The complete chloroplast genome of a purple Ethiopian rape (*Brassica carinata*: Brassicaceae) from Guizhou Province, China and its phylogenetic analysis [J]. *Mitochondrial DNA Part B*, 2021, **6**(7): 1821–1823.
- [30] ZHAO Mengliang, WU Yanxun, REN Yanjing. Complete chloroplast genome sequence structure and phylogenetic analysis of kohlrabi (*Brassica oleracea* var. *gongylodes* L.) [J/OL]. *Genes*, 2024, **15**(5): 550 [2025-04-30]. DOI: [10.3390/genes15050550](https://doi.org/10.3390/genes15050550).
- [31] WANG Yilin, LIANG Qiannan, ZHANG Chenlu, *et al.* Sequencing and analysis of complete chloroplast genomes provide

- insight into the evolution and phylogeny of Chinese kale (*Brassica oleracea* var. *alboglabra*) [J/OL]. *International Journal of Molecular Sciences*, 2023, **24**(12): 10287 [2025-04-30]. DOI: [10.3390/ijms241210287](https://doi.org/10.3390/ijms241210287).
- [32] REDWAN R M, SAIDIN A, KUMAR S V. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae [J/OL]. *BMC Plant Biology*, 2015, **15**: 196 [2025-04-30]. DOI: [10.1186/s12870-015-0587-1](https://doi.org/10.1186/s12870-015-0587-1).
- [33] LI Yongfu, SYLVESTER S P, LI Meng, *et al.* The complete plastid genome of *Magnolia zenii* and genetic comparison to Magnoliaceae species [J/OL]. *Molecules*, 2019, **24**(2): 261 [2025-04-30]. DOI: [10.3390/molecules24020261](https://doi.org/10.3390/molecules24020261).
- [34] DURET L. tRNA gene number and *Codon* usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes [J]. *Trends in Genetics*, 2000, **16**(7): 287–289.
- [35] 朱梦飞, 胡迎峰, 师雪芹. 濒危植物新绒苔叶绿体基因组特征及系统发育位置分析 [J]. 浙江农林大学学报, 2025, **42**(1): 55–63.
- ZHU Mengfei, HU Yingfeng, SHI Xueqin. Characterization and phylogenetic location analysis of chloroplast of the endangered plant *Neotrichocolea bissetii* [J]. *Journal of Zhejiang A&F University*, 2025, **42**(1): 55–63.
- [36] 蒋思思, 袁军, 周文君, 等. 薄壳山核桃 (*Carya illinoensis*) 叶绿体基因组及其特征分析 [J]. 园艺学报, 2022, **49**(8): 1772–1784.
- JIANG Sisi, YUAN Jun, ZHOU Wenjun, *et al.* Complete chloroplast genome sequence and characteristics analysis of *Carya illinoensis* [J]. *Acta Horticulturae Sinica*, 2022, **49**(8): 1772–1784.
- [37] WU Junyan, MA Xuecai, MA Li, *et al.* Complete chloroplast genome sequence and phylogenetic analysis of winter oil rapeseed (*Brassica rapa* L.) [J]. *Mitochondrial DNA. Part B, Resources*, 2021, **6**(3): 723–731.
- [38] 王云生, 黄宏文, 王瑛. 植物分子群体遗传学研究动态 [J]. 遗传, 2007, **29**(10): 1191–1198.
- WANG Yunsheng, HUANG Hongwen, WANG Ying. Recent progress in plant molecular population genetics [J]. *Hereditas*, 2007, **29**(10): 1191–1198.
- [39] XU Jin, LIU Chen, SONG Yun, *et al.* Comparative analysis of the chloroplast genome for four *Pennisetum* species: molecular structure and phylogenetic relationships [J/OL]. *Frontiers in Genetics*, 2021, **12**: 687844 [2025-04-30]. DOI: [10.3389/fgene.2021.687844](https://doi.org/10.3389/fgene.2021.687844).