

研究报告

Research Report

马泡瓜叶绿体基因组测序

马宗新 *

安徽省阜阳市农业科学院, 阜阳, 236065

* 通信作者, mzxslm965@163.com

摘要 马泡瓜(*Cucumis melo* L. var. *agrestis* Naud.)葫芦科甜瓜属植物, 种子富含油脂, 可榨取优质食用油, 还可为医药和食品工业提供原料。本研究首次开展了马泡瓜叶绿体基因组测序工作。结果表明: 马泡瓜叶绿体基因组与普通甜瓜近似, 物种聚类分析表明与甜瓜、蛇瓜、越瓜最为接近。本研究可以为育种提供理论参考。

关键词 马泡瓜; 叶绿体基因组; 测序

Sequencing of chloroplast genome of *Cucumis melo* L. var. *agrestis* Naud.

Mazon Xin *

Academy of Agricultural Sciences of Fuyang, Fuyang, 236065

* Corresponding author, mzxslm965@163

DOI: 10.5376/mpb.cn.2021.19.0019

Abstract *Cucumis melo* L. var. *agrestis* Naud. belong to the genus *Cucumis* in the Gourd family, it is rich in oil. It can extract high quality edible oil and provide raw materials for medicine and food industry. The chloroplast genome of *Cucumis Melo* L. was sequenced for the first time. The results showed that the chloroplast genome of *Cucumis Melo* L. was similar to that of common melon. This study can provide a theoretical reference for breeding.

Keywords *Cucumis melo* L. var. *agrestis* Naud.; Chloroplast genome; Sequencing

马泡瓜(*Cucumis melo* L. var. *agrestis* Naud.)为葫芦科(Cucurbitaceae)甜瓜属(*Cucumis*)植物, 野生杂草。有学者认为马泡瓜是中国东方薄皮甜瓜的野生祖先(马宗新, 2020a)。种子中含有大量的脂肪, 可以榨油食用, 果实中富含药用成分葫芦素等, 是一种极具开发潜力的新型野生植物(马宗新, 2020b)。目前叶绿体基因组测序小片段准确组装已经可以做到, 但组装较大的完整基因组仍比较困难, 核基因组组装更为困难(Anton et al., 2012)。甜瓜(*Cucumis melo* L.)作为世界最重要的经济作物之一, 叶绿体基因组测序已经获得成功, 甜瓜叶绿体基因组中包含 133 个基因。基因组的总 GC 含量为 36.9% (Zhu et al., 2016)。利用叶绿体基因组分析揭示了越南甜瓜地方品种的遗传多样性(Phan et al., 2010)。利用

基因组序列多态性推断栽培甜瓜的多样性和遗传分化(Katsunori et al., 2013)。并阐述了甜瓜叶绿体基因组全序列及其系统学意义 (Hong et al., 2020)。截止目前, 马泡瓜叶绿体基因组研究还没有相关文献报道。为了了解马泡瓜的叶绿体基因组的组成和基因与植物学性状之间的关系, 为育种遗传改良马泡瓜提供理论技术依据, 采用马泡瓜为实验材料, 利用 Illumina 高通量测序技术, 完成了马泡瓜叶绿体基因组测序工作。

1 结果与分析

1.1 叶绿体基因组测序数据统计及注释

37.45%, Q20 为 96.83% (表 1)。基因总共 133 个,

本文首次发表在《分子植物育种》上, 现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License, 协议对其进行授权, 再次发表与传播

收稿日期: 2021 年 5 月 8 日; 接受日期: 2021 年 5 月 10 日; 发表日期: 2021 年 5 月 17 日

引用格式: 马宗新, 2021, 马泡瓜叶绿体基因组测序, 分子植物育种 (网络版) 19(19): 1-9 (doi: 10.5376/mpb.cn.2021.19.0019) (ZMazon X., 2021, Sequencing of chloroplast genome of *Cucumis melo* L. var. *agrestis* Naud., Fenzi Zhiwu Yuzhong (Molecular Plant Breeding (online)), 19(19): 1-9 (doi: 10.5376/mpb.cn.2021.19.0019))

tRNA37 个, rRNA8 个, mRNA88 个(表 2)。叶绿体的基因功能主要分为光合合成、自我复制等其他一些基因, 此外还有一些功能未知的基因(表 3)。

1.2 叶绿体基因组图谱

马泡瓜叶绿体基因组为共价闭合的双链环状分子(图 1), 全长 156 017 bp, 包括 1 对反向重复(IR)区

(25 797 bp)、1 个大的单拷贝(LSC)区(86 336 bp)和 1 个小的单拷贝(SSR)区(18 087 bp), 全基因组的 GC 含量为 37.45%。6 个 ATP 合酶基因、11 个 NADH 脱氢酶基因。

1.3 RSCU 密码子偏好性

密码子偏好性表明, TGG 单独对应 Trp 编码

表 1 测序数据统计

Table 1 Sequencing data statistics

Sample ID	ReadSum	BaseSum	GC (%)	Q20 (%)	Q30 (%)
Cucumis_melo_L._var._agrestis_Naud.	55968670	16790601000	37.45	96.83	91.75

表 2 叶绿体基因注释信息统计

Table 2 Statistics of chloroplast gene annotation information

#ID	Gene	Trna	Rrna	Mrna	Pseudo
Cucumis_melo_L._var._agrestis_Naud.	133	37	8	88	0

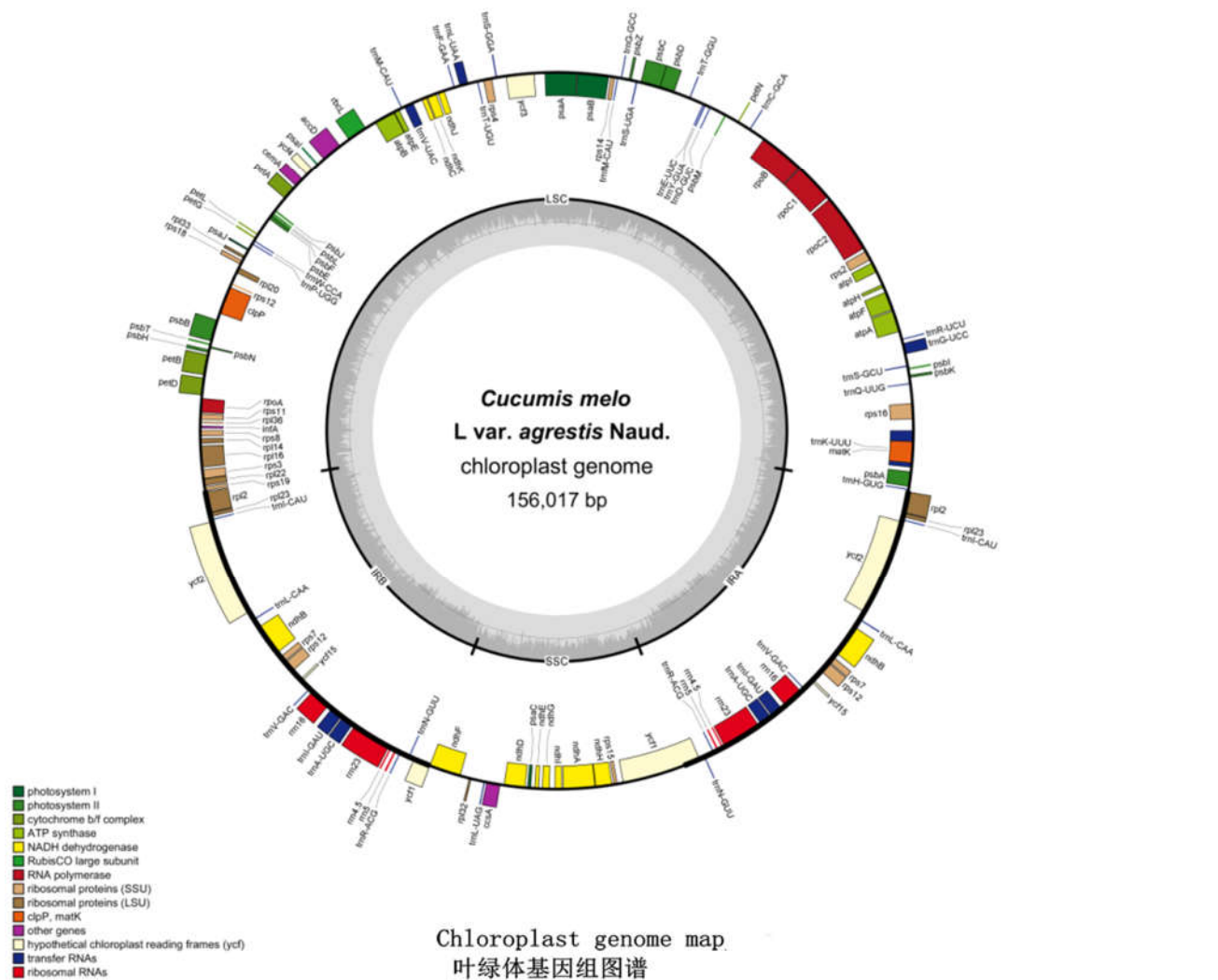


图 1 叶绿体基因组图谱

Figure 1 Chloroplast genome map

表 3 叶绿体基因功能分类统计

Table 3 Classification and Statistics of chloroplast gene function

Category	Gene group	Gene name
分类	基因组	基因
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of NADH dehydrogenase	<i>ndhA *</i> , <i>ndhB * (2)</i> , <i>ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA, petB *</i> , <i>petD *</i> , <i>petG, petL, petN</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF *</i> , <i>atpH, atpI</i>
	Large subunit of rubisco	<i>rbcL</i>
	Subunits photochlorophyllide reductase	–
	Self-replication	Proteins of large ribosomal subunit
Proteins of small ribosomal subunit		<i>rpoA, rpoB, rpoC1*</i> , <i>rpoC2</i>
Subunits of RNA polymerase		<i>rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)</i>
Ribosomal RNAs		<i>trnA -UGC * (2)</i> , <i>trnC -GCA</i> , <i>trnD -GUC</i> , <i>trnE -UUC</i> , <i>trnF -GAA</i> , <i>trnG -GCC</i> , <i>trnG -UCC *</i> , <i>trnH -GUG</i> , <i>trnI -CAU (2)</i> , <i>trnI -GAU *(2)</i> , <i>trnK -UUU *</i> , <i>trnL -CAA (2)</i> , <i>trnL -UAA *</i> , <i>trnL -UAG</i> , <i>trnM -CAU</i> , <i>trnN -GUU (2)</i> , <i>trnP -UGG</i> , <i>trnQ -UUG</i> , <i>trnR -ACG (2)</i> , <i>trnR -UCU</i> , <i>trnS -GCU</i> , <i>trnS -GGA</i> , <i>trnS -UGA</i> , <i>trnT -GGU</i> , <i>trnT -UGU</i> , <i>trnV -GAC (2)</i> , <i>trnV -UAC *</i> , <i>trnW -CCA</i> , <i>trnY -GUA</i> , <i>trnY -CAU</i>
Transfer RNAs		
Other genes		Maturase
	Protease	<i>clpP **</i>
	Envelope membrane protein	<i>cemA</i>
	Acetyl-CoA carboxylase	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Translation initiation factor	<i>infA</i>
	Other	–
Genes of unknown function	Conserved hypothetical chloroplast ORF	<i>ycf1(2), ycf15(2), ycf2(2), ycf3 **</i> , <i>ycf4</i>

(图 1), Met 分别对应 GTG、ATT、ATG 和 ATC 编码。Arg、Leu 和 Scr 都对应 6 个密码子编码(图 2)。密码子偏好性分析结果表明: UAA 和 GCU 数值较高(表 4), 而 GCG、UGC、UAG 和 UGA 数值相对较低(表 4)。

1.4 散在重复序列

散在重复序列结果表明: 正向重复和回文重复数量较多, 而反向重复数量较少, 互补重复没有(图 3)。

1.5 cpSSR 结果

马泡瓜叶绿体基因组中单核苷酸重复 SSR 有 191 个(表 5), 占 66.55%; 二核苷酸 SSR 11 个, 三核苷酸 SSR 有 71 个, 占 24.74%; 四核苷酸、五核苷酸

和六核苷酸 SSR 较少, 分别为 9、3 和 2 个(表 5)。单核苷酸重复中 A、T 核苷酸重复数量最多, C、G 单核苷酸重复数量较少(图 4)。

1.6 KaKs 实验结果

非同义突变率(Ka)和同义突变率(Ks)的比值则说明受到何种选择作用。比值大于 1, 说明受到正选择效应, 小于 1, 说明有纯化选择作用。KaKs 值在 0.4~0.8 之间, 说明有纯化选择作用(表 8)。

1.7 核酸多样性 pi 分析

ssc 和 IR 区域内 pi 值较高, 这也意味着该区域变异程度较大, 在 lsc 区间数值较小, 说明变异程度较小(图 5)。

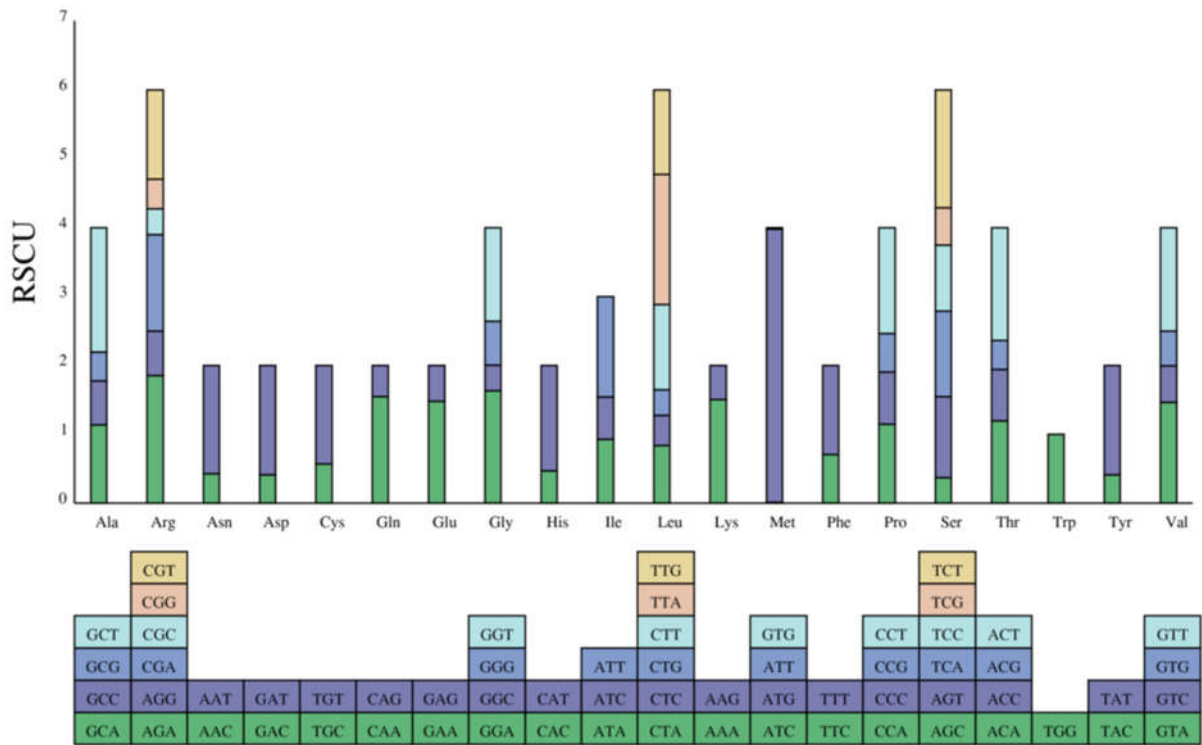


图2 RSCU 密码子偏好性

Figure 2 RSCU codon preference

表4 密码子偏好性结果

Table 4 RSCU codon preference

氨基酸	符号	密码子	数量	偏好性
Amino acid	Symbol	Codon	No.	RSCU
*	Ter	UAA	56	1.909 2
*	Ter	UAG	15	0.511 5
*	Ter	UGA	17	0.579 6
A	Ala	GCA	398	1.136 4
A	Ala	GCC	223	0.636 8
A	Ala	GCG	148	0.422 4
A	Ala	GCU	632	1.804 4
C	Cys	UGC	89	0.566 8
C	Cys	UGU	225	1.433 2

1.8 IRscope 试验结果

叶绿体基因组为环形结构,IR 与 LSC 和 SSC 存在 4 个边界,即 LSC-IRb、IRb-SSC、SSC-IRa 和 IRa-LSC。在基因组进化过程中,IR 边界会发生扩张与收缩,使某些基因进入 IR 区或单拷贝区,使用 Perl 中的 SVG 模块将边界信息可视化(图 6)。几个甜瓜种属之间有微小差异,基因组长度也有略微不同,在四个边界之间长度略有不同,只有细微差别,可能这细微的差异,就决定了不同种属的甜瓜形态各异,性状各异(图 6)。

1.9 叶绿体结构比较

叶绿体结构是一个环形状结构,通过比较近源属的甜瓜品种,结果比较近似。除少数位点不同外,其他近似程度较高(图 7)。

1.10 进化树聚类结果

聚类分析结果表明,甜瓜、越瓜、蛇瓜为一类,而马泡瓜与普通甜瓜、越瓜、蛇瓜亲缘关系较近,而与苦瓜、南瓜、丝瓜等亲缘关系较远(图 8)。

1.11 叶绿体序列同源及共线性分析

使用 Mauve (<http://darlinglab.org/mauve>) 软件默认参数进行基因组比对,序列同源性表明,马泡瓜与现有已知甜瓜、越瓜、黄瓜等作物具有同源相似性,在少数位点上有不同之外,大部分序列近似。特别是与越瓜、普通甜瓜更近似。与苦瓜、丝瓜相距较远(图 9)。

2 讨论与结论

参考甜瓜的基因组测序,本研究获得了马泡瓜这种植物的完整叶绿体测序和基因组图谱。研究表明,马泡瓜与同属物种越瓜(*Cucumis Melo* var. *Conomon*) (156 017 bp)、甜瓜(*Cucumis Melo* var. *Makuwa*) (156 016 bp)、甜瓜亚种(*Cucumis Melo* subsp.

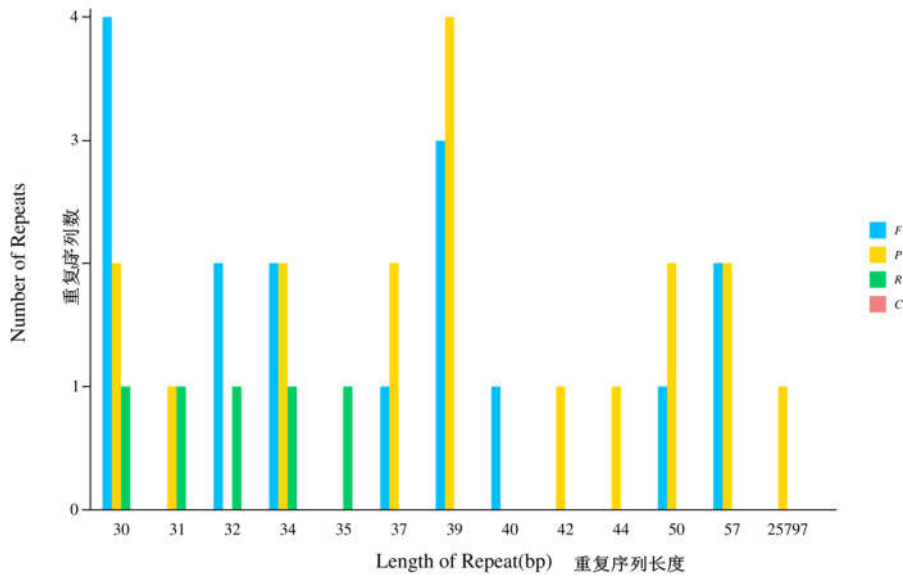


图 3 散在重复序列统计
Figure 3 Statistics of scattered repeated sequences

表 5 SSR 分析结果统计
Table 5 Statistics of SSR analysis results

单位	SSRs 数量	百分率(%)
Unit size	Number of SSRs	Percent (%)
1	191	66.55
2	11	3.83
3	71	24.74
4	9	3.14
5	3	1.05
6	2	0.70

表 6 SSR 结果
Table 6 Results of SSR analysis

ID	SSR nr.	SSR type	SSR	size	start	end
1	1	p2	(TA) 7	14	132	145
1	2	p1	(T) 9	9	1 524	1 532
1	3	c	(A) 8 taagt (A) 8	21	1 675	1 695
1	4	p1	(T) 9	9	2 170	2 178
1	5	p1	(A) 8	8	2 837	2 844
1	6	p3	(TTC) 3	9	3 081	3 089
1	7	p1	(A) 12	12	3 220	3 231
1	8	p3	(AAT) 3	9	3 541	3 549
1	9	p1	(T) 8	8	3 954	3 961

Agrestis) (156 016 bp)相接近。叶绿体序列同源及共线性分析表明马泡瓜与甜瓜和越瓜等比较一致，而与西瓜、南瓜等在 tRNA、少数基因上有很大区别。聚类分析结果表明，马泡瓜与甜瓜最为接近，在甜瓜几个种属中，与普通甜瓜、蛇瓜、越瓜最为接近。马泡瓜内含子数量 21 个，与甜瓜的内含子基本一致(王宏刚, 2017)。这说明葫芦科植物中，马泡瓜与甜瓜有相同的起源，这与前人考证的马泡瓜为中国薄皮甜瓜的野生祖先结论一致(马宗新, 2020a)。叶

绿体基因组具有相对保守性，可以更好地说明物种的进化和亲缘关系。本研究表明马泡瓜在遗传学和分类学上与前人的研究结果一致 (马宗新, 2020a)。甜瓜的为 GC 含量 36.9%。黄瓜的也为 36.9% (Zhu et al., 2016)，马泡瓜的 GC 含量为 37.45%。而 LSC 含量 55.34%；SSC 含量 11.59%，IRs 含量 33.1%与甜瓜基本相同(Zhu et al., 2016)。

表 7 SSR 引物设计结果
Table 7 Result of SSR primer design

#Gene_ID	FPr1(5'-3')	RPr1(5'-3')	PSize	Start	End
1_1	AATCCACTGCCTTGATCCAC	GCATGAACGTAATGCTCACAA	259	88	346
1_2	GGACTCCCAAGCACATGAAT	TCCGACTAGTTCCGGGTTC	280	1 375	1 654
1_3	TCATTAATATGGGTTGCCCG	GAAAAATGCAAGCACGGTTT	114	1 610	1 723
1_4	AGAACCTGTTCTTGCTCCGT	GATCGATTTCGGATTTTG	234	2 025	2 258
1_5	GGAAAATGGATTTCGATTACACA	GTTCAAATCCTTCGCTCCTG	190	2 767	2 956

表 8 叶绿体基因 KaKs 分析统计

Table 8 analysis and Statistics of chloroplast gene KaKs

基因 Gene	方法 Method	Ka/Ks	P-Value
rps12 vs MF536699_rps12-2	MLWL	0.864801	0.565321
petD vs MF536699_petD	MLWL	0.6753	0.159274
rps12-2 vs MF536699_rps12-2	MLWL	0.864801	0.565321
ycf1-2 vs MF536699_ycf1-2	MLWL	0.456595	0.0830734

3 材料与方法

3.1 植物材料

马泡瓜植物新鲜叶片。采集地点为阜阳市农业科学院科技园内。

3.2 生物信息学分析

将 Clean Data 按参考物种的叶绿体基因组序

列进行序列组装,获得叶绿体序列组装结果,对叶绿体序列组装结果进行基因结构注释,制作叶绿体基因组图谱,分析叶绿体基因组 SSR 等基础内容, KaKs 分析、系统发育分析等高级分析内容。使用 fastp (version 0.20.0, <https://github.com/OpenGene/fastp>) 软件对原始数据进行过滤,过滤标准如下:

- (1) 截除 Reads 中的测序接头以及引物序列
- (2) 过滤掉平均质量值小于 Q5 的 reads
- (3) 过滤掉 N 个数大于 5 的 reads

使用 OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 制作叶绿体基因组图谱。由于密码子生物学特点,不同物种其基因组密码子使用率存在着很大的不同。这种同义密码子使用的平等性,叫做密码子的偏好性 (Relative Synonymous Codon Usage, RSCU)。而这种偏好性被认为是自然选择、物种突变和遗传漂变的综合结果。其计算方法

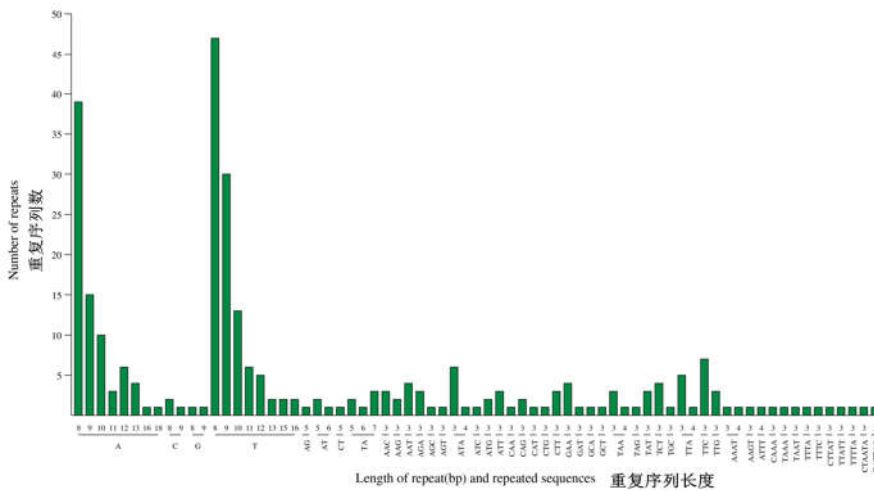


图 4 SSR 各类型数量统计

Figure 4 Quantitative statistics of various SSR types

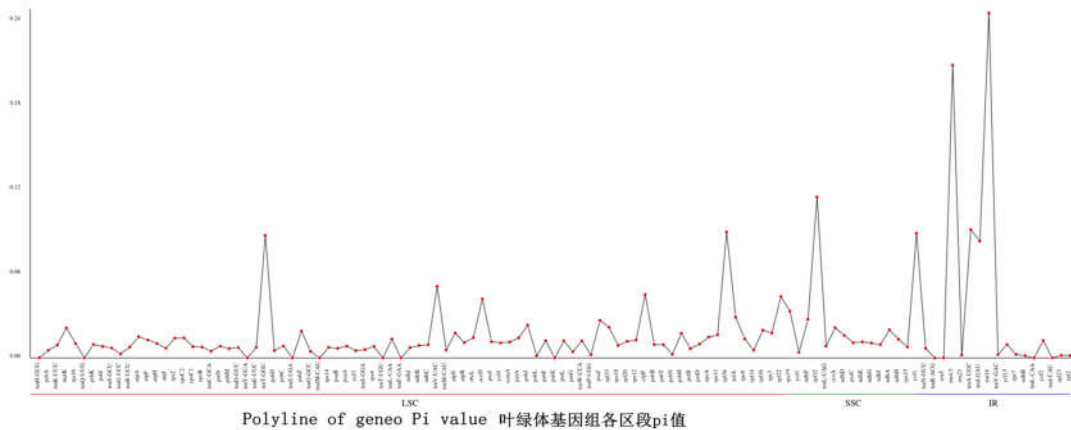


图 5 基因 pi 值折线

Figure 5 Polyline of geneo Pi value

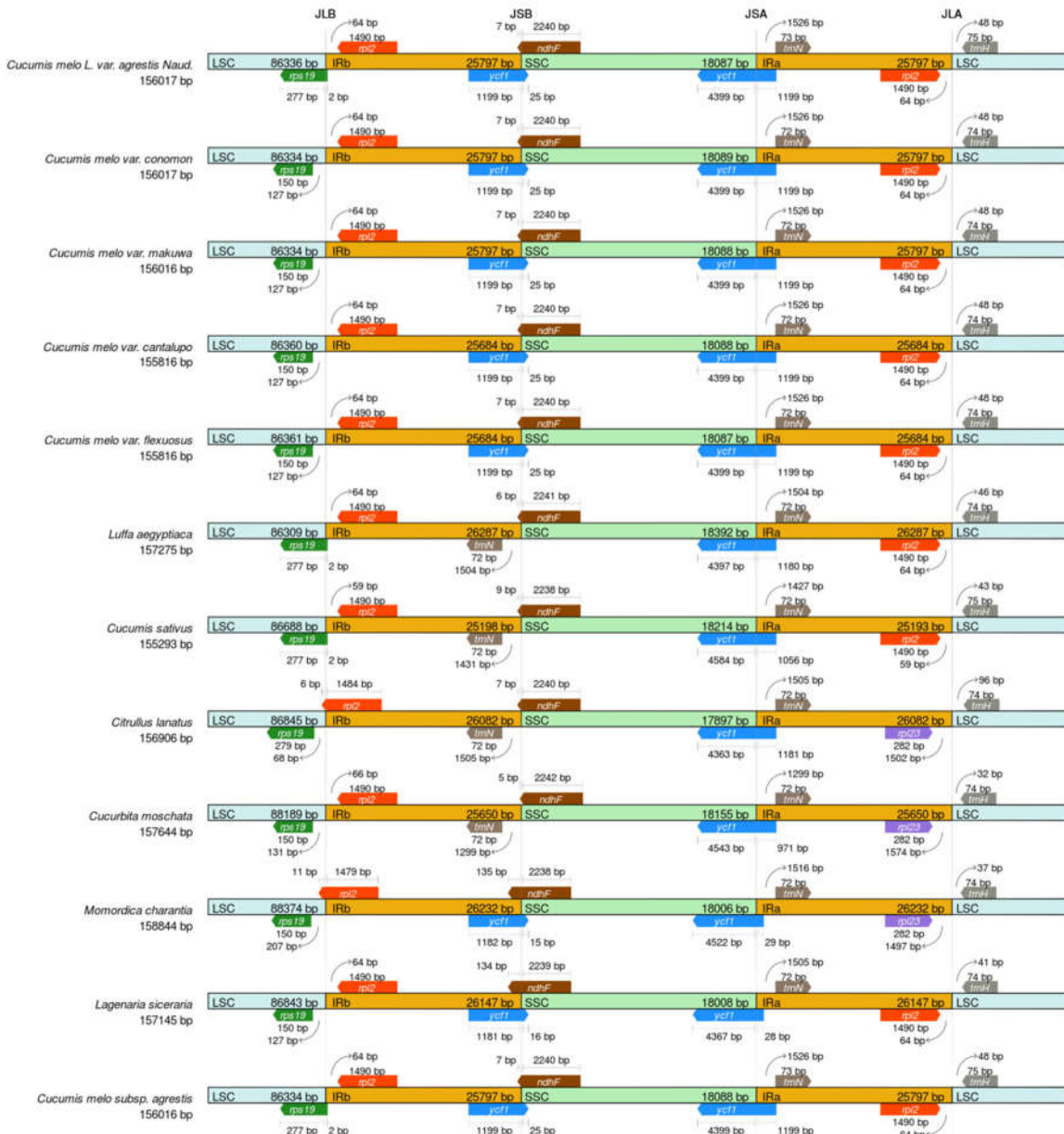


图 6 叶绿体 IR 边界变化
Figure 6 IR boundary change of chloroplast

是:(编码某氨基酸的其中一种密码子数量/编码该氨基酸的所有密码子数量)/(1/编码该氨基酸的密码子种类),及(密码子的实际使用频率/该密码子的理论使用频率)。使用自行编写的 Perl 脚本进行计算。

散在重复序列使用 vmatch v2.3.0 (<http://www.vmatch.de/>)软件结合 Perl 脚本鉴定重复序列。其参数设置为:最小长度(minimum length)=30 bp,海明距离(hamming distance)=3,鉴定形式为四种:正向(Forward)、回文(Palindromic)、反向(Reverse)、互补(Complement)。

叶绿体基因组上的 SSR 标记称之为 cpSSR 标

记。使用 MISA v1.0 软件开展 cpSSR 的分析,参数 1-8 (单碱基重复 8 次及以上),2-5,3-3,4-3,5-3,6-3。

pi (核酸多样性)能揭示不同物种核酸序列的变异大小,变异度较高的区域可以为种群遗传学提供潜在的分子标记。使用 mafft 软件(--auto 模式)对不同物种的同源基因序列进行全局比对,使用 dnasp5 计算每个基因的 pi 值。

使用软件 CGVIEW (http://stothard.afns.ualberta.ca/cgview_server/)默认参数,针对近源物种进行叶绿体基因组结构的比较分析。

默认采用全基因组做进化树分析,将环形序列

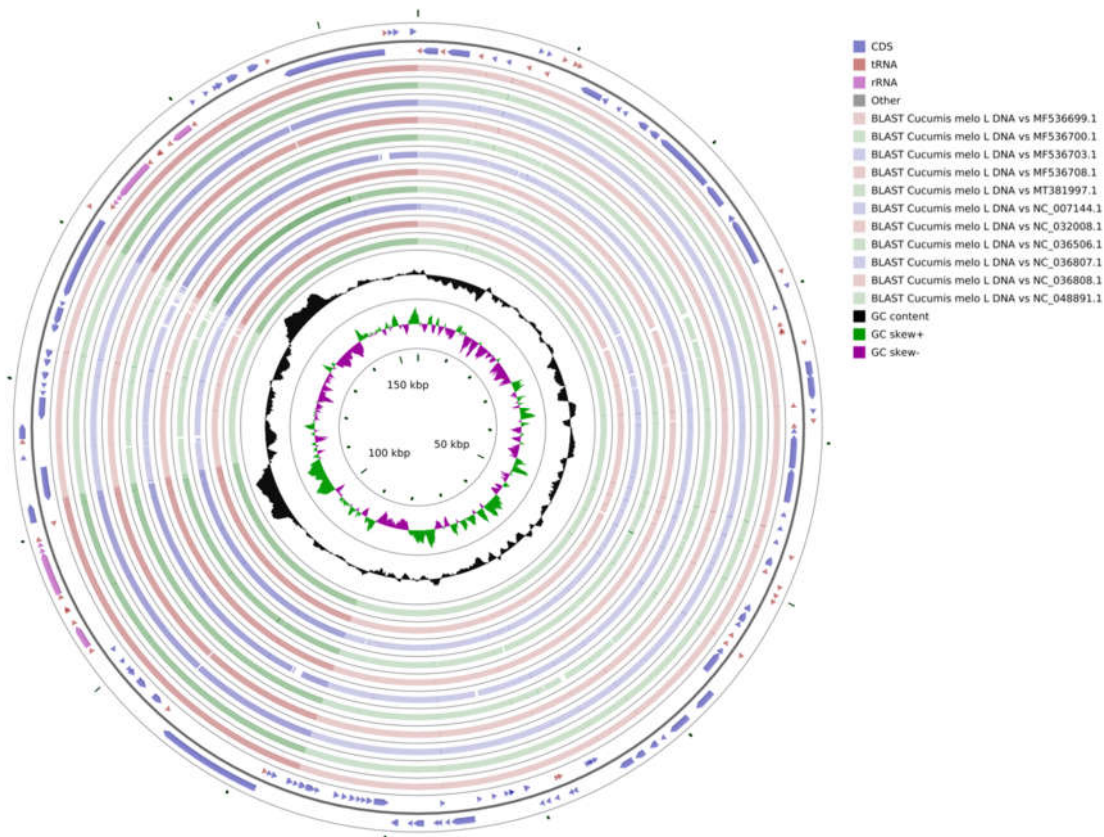


图 7 叶绿体结构比较
Figure 7 Comparative of chloroplast structure

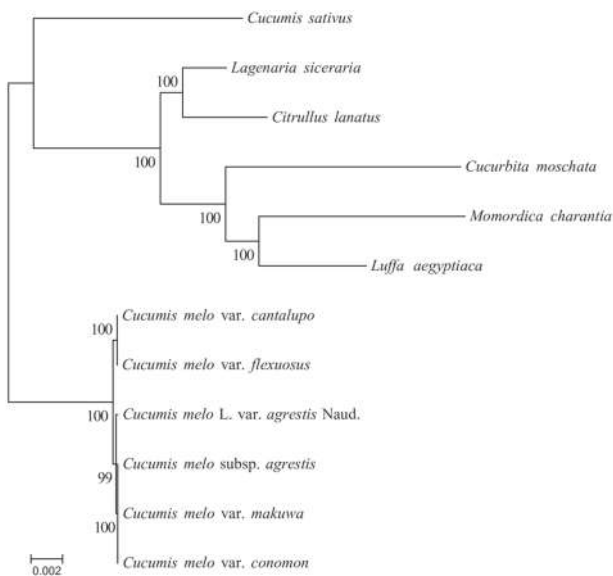


图 8 叶绿体系统进化树
Figure 8 chloroplast system evolution tree

设置相同起点,物种间序列用 MAFFT 软件(v7.427, --auto 模式) 进行多序列比对, 将比对好的数据用 trimAl (v1.4.rev15)修剪, 然后使用 RAxML v8.2.10 (<https://cme.h-its.org/exelixis/software.html>) 软件, 选用 GTRGAMMA 模型, rapid Bootstrap 分析, boot

strap=1 000, 构建最大似然进化树。

3.3 基因组组装

为了降低接下来序列组装的复杂度, 使用 bowtie2 (v2.2.4, <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) very-sensitive-local 模式比对数据库并进行测序。组装核心模块采用 SPAdes [1] v3.10.1 (<http://cab.spbu.ru/software/spades/>) 软件组装叶绿体基因组。对基因组进行 3 个方面的质控来确保组装结果的准确性:

(1) reads 回比基因组, 统计基因组覆盖度、插入片段大小等信息;(2)基因组比对参考序列, 查看基因组的保守与重排等共线性分析;(3)基因组比对参考序列结构信息, 比较两者间的差异。

本项目使用参考序列 Cucumis_melo: MT240-857.1。

作者贡献

马宗新负责方案设计、数据处理和基因测序工作, 并完成论文的写作和修改。本人已经阅读并同意最终的文本。



图9 叶绿体序列同源性

Figure 9 Chloroplast sequence homology

致谢

本研究由安徽省科技厅重点研发项目(17040-704067)资助。

参考文献

Anton B., Sergey N., and Dmitry A., 2012, A new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.*, 19(5): 455-477.

Hong C., Wei P.K., Zhang M.M., and Hou D., 2020, The complete chloroplast genome of *Cucumis melo* L. 'Shengkaihua' (Cucurbitaceae) and its phylogenetic implication, *Mitochondrial DNA Part B*, 5(2): 1253-1254.

Katsunori T., Yukari A., Kenji F., Tatsuya Y., Yasheng A., Hidetaka N., Chun L.L., Hiromichi Y., YoIchiro S., and Kenji K., 2013, Diversification and genetic differentiation of cultivated melon inferred from sequence polymorphism in the chloroplast genome, *Breeding Science*, 63 (2): 183-196.

Ma Z.X., 2020a, Brief history of production of *Cucumis melo* L. var. *agrestis* Naud., *Nongcun Jingji Yu Keji (Rural Economy and Science-Technology)*, 31(7): 76-77. (马宗新, 2020a, 马泡瓜生产简史, *农村经济与科技*, 31(7): 76-77.)

Ma Z.X., 2020b, Investigation on new food resources of a new wild plant, *Nongye Yu Jishu (Agriculture and technology)*, 40(1) : 32-34. (马宗新, 2020b, 一种新型野生植物新食品资源调查研究, *农业与技术*, 40(1): 32-34.)

Phan T.P.N., Yukari A., Tran T.M.H., Katsunori T., Yasheng A., Tatsuya Y., Hidetaka N., Long C. L., and Kenji K., 2010, Genetic diversity in Vietnamese melon landraces revealed by the analyses of morphological traits and nuclear and cytoplasmic molecular markers, *Breeding Science*, 60 (3): 255-266.

Wang H.G., Bai YL., Xu H.J., Liu S.H., Zhang W.J., Liu Q., Zhang X.M., Qiao M.Q., and Wang Y., 2017, Construction and stability of cucumber chloroplast expression vector, *Nankai Daxue Xuebao (Acta Scientiarum Naturalium Universitatis Nankaiensis)*, 50 (5): 61-66. (王宏刚, 白艳玲, 徐海津, 刘盛海, 张文娟, 刘青, 张秀明, 乔明强, 王勇, 2017, 黄瓜叶绿体表达载体构建与稳定性研究, *南开大学学报(自然科学版)*, 50(5): 61-66.)

Zhu Q.L., Gao P., Liu S., Amanullah S., and Luan F.S., 2016, Comparative analysis of single nucleotide polymorphisms in the nuclear, chloroplast, and mitochondrial genomes in identification of phylogenetic association among seven melon (*Cucumis melo* L.) cultivars., *Breeding science*, 66 (5): 711-719.