

DOI: 10.13718/j.cnki.xdzk.2023.10.005

李国良, 张鸿, 林赵森, 等. 叶菜型甘薯叶绿体基因组及其特征分析 [J]. 西南大学学报(自然科学版), 2023, 45(10): 43-53.

叶菜型甘薯叶绿体基因组及其特征分析

李国良, 张鸿, 林赵森, 许泳清, 许国春, 李华伟,
纪荣昌, 罗文彬, 邱永祥, 邱思鑫, 汤浩

福建省农业科学院 作物研究所/农业农村部南方薯类观测实验站, 福州 350013

摘要: 为探明叶菜型甘薯叶绿体基因组特征及其与番薯属植物的亲缘关系, 以叶菜型甘薯‘福菜薯 18 号’为试验材料, 利用 BGISEQ-500 平台和 Oxford Nanopore Technologies 单分子测序技术对全基因组进行建库测序, 并组装其叶绿体基因组。结果显示: 叶菜型甘薯叶绿体基因组全长 161 387 bp, 具有典型的环状四分体结构, 其大单拷贝区 (large single copy, LSC)、小单拷贝区 (small single copy, SSC) 和 2 个反向重复序列 (inverted repeat, IR) 的长度分别为 87 597, 12 052 和 30 869 bp。注释共得到 132 个基因, 包含 87 个蛋白编码基因, 8 个 rRNA 基因, 37 个 tRNA 基因。在叶菜型甘薯叶绿体基因组中共搜索到 54 个 SSR 位点, 其中单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复基序个数分别为 32, 4, 3, 11, 2 和 2 个。系统进化分析表明, 叶菜型甘薯与甘薯四倍体野生种 *Ipomoea tabascana* 和二倍体野生种 *Ipomoea trifida* 具有较近的亲缘关系。

关键词: 叶菜型甘薯; 叶绿体基因组; SSR; 系统发育分析

中图分类号: S531

文献标志码: A

文章编号: 1673-9868(2023)10-0043-11

开放科学(资源服务)标识码(OSID):



Analysis of Chloroplast Genomic Characteristics of Leafy Sweetpotato

LI Guoliang, ZHANG Hong, LIN Zhaomiao, XU Yongqing,
XU Guochun, LI Huawei, JI Rongchang, LUO Wenbin,
QIU Yongxiang, QIU Sixin, TANG Hao

Institute of Crop Sciences, Fujian Academy of Agricultural Sciences/Scientific Observing and Experimental Station of Tuber and Root Crops in South China, Ministry of Agriculture and Rural Affairs, Fuzhou 350013, China

收稿日期: 2023-03-06

基金项目: 财政部和农业农村部 国家现代农业产业技术体系项目(CARS-10-B14); 福建省人民政府-中国农业科学院农业高质量发展超越“5511”协同创新工程项目(KXXYJBG0039); 福建省属公益类科研院所基本科研专项(2021R1031011)。

作者简介: 李国良, 硕士, 助理研究员, 主要从事甘薯抗逆基因功能研究。

通信作者: 邱思鑫, 博士, 研究员。

Abstract: In order to find out the characteristics of chloroplast genome of leafy sweetpotato and the genetic relationship of the genus *Ipomoea*, the whole chloroplast genome of leafy sweetpotato ‘Fucaishu18’ was sequenced and assembled using the BGISEQ 500 platform and single molecule sequencing technology of Oxford Nanopore Technologies. The results showed that the leafy sweetpotato chloroplast genome was 161 387 bp in length, with a typical quadripartite structure of circular molecule, including a pair of inverted repeats (IRA and IRB) of 30 869 bp that were separated by large and small single copy regions (LSC and SSC) of 87 596 bp and 12 052 bp, respectively. A total of 132 genes were annotated, including 87 protein-coding genes, 8 rRNA genes, and 37 tRNA genes. We found 54 SSR loci in the leafy sweetpotato chloroplast genome, of which the number of mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeat motifs were 32, 4, 3, 11, 2 and 2, respectively. The phylogenetic analysis showed that the leafy sweetpotato had a relatively close relationship with the wild tetraploid species *Ipomoea tabascanana* and the wild diploid species *Ipomoea trifida*.

Key words: leafy sweetpotato; chloroplast genome; SSR; phylogenetic analysis

叶绿体是绿色植物进行光合作用最重要的细胞器，是许多基本物质和次生代谢物合成的重要场地^[1]。叶绿体蛋白虽然绝大多数是由核基因编码，但也有 100 多种蛋白是由叶绿体基因组(cpDNA)编码的。cpDNA 序列长度为 130~150 kb，具有典型的双链环状结构，由 1 个大单拷贝区(LSC)，1 个小单拷贝区(SSC)和两个反向重复区(IR)组成。大多数的 cpDNA 是母系遗传，但也有一部分植物是父系遗传或双亲遗传。相对于核基因组的复杂多样性，cpDNA 结构简单，序列高度保守，不同物种或同一物种不同个体间仅存在着局部区域的序列变异，因此 cpDNA 更易解析，更有利于研究植物的分类与进化^[2-4]。

甘薯是我国重要的粮食作物、饲料作物和食品加工业、化工业的原料作物^[5]。叶菜型甘薯是一类以鲜幼嫩茎叶作蔬菜用的甘薯品种，与普通甘薯相比，其茎叶产量较高^[6]。叶绿体基因组的完整解析是准确研究植物系统进化关系、发掘基因功能的更有效而可靠的手段^[7]。叶菜型甘薯的父本或母本往往来源于普通甘薯，其叶绿体基因组序列与普通甘薯有什么差别目前尚未清楚，甘薯种间叶绿体基因组之间存在多少碱基差异目前也不清楚。本研究以叶菜型甘薯‘福菜薯 18 号’为材料，通过序列拼接获得完整的叶绿体基因组，并利用生物信息学方法进行分析，为叶菜型甘薯的亲缘关系和甘薯种间关系奠定基础。

1 材料与方法

1.1 试验材料

以叶菜型甘薯‘福菜薯 18 号’为试验材料，其嫩梢幼叶于 2021 年 7 月采自‘福菜薯 18 号’甘薯组培苗。

1.2 DNA 的提取及测序

利用 DNA 提取试剂盒(南京诺维赞生物科技有限公司)提取甘薯组织总 DNA^[8]，用 1.5% 琼脂糖凝胶电泳和 Nanodrop 2000 检测甘薯总 DNA 的质量和完整性，DNA 质量合格后进行上机测试。

二代测序实验流程按照 BGISEQ-500 的标准程序执行，样品基因组 DNA 检测合格后，用超声波法将 DNA 片段化，然后对片段化的 DNA 进行纯化，末端修复，3'端加 A，连接测序接头，构建测序文库，文库质检合格后用 BGISEQ-500 平台进行测序。

三代测序采用 Oxford Nanopore Technologies 公司的建库试剂盒进行测序文库构建，文库检验合格后上机测序。

1.3 基因组组装与注释

使用 Fastp 软件对原始数据进行过滤,去除其中的接头序列及低质量序列,获取高质量的序列数据,按参考物种的叶绿体基因组序列进行组装,得到叶绿体基因组序列组装结果^[9].使用 minimap2 将三代测序 reads 比对 NCBI 旋花科所有叶绿体基因组数据,提取比对长度大于 5 000 bp 的 reads 用于后续组装^[10].使用 bowtie2 将二代测序 reads 比对广州佰数生物科技有限公司自建的叶绿体基因组数据库,将比对上的 reads 用于后续组装^[11].组装软件使用 Unicycler version: v 0.4.8,将上述提取到的叶绿体候选三代和二代 reads 用于叶绿体基因组组装^[12].

使用 GeSeq 软件对叶绿体基因序列进行注释^[13];利用 tRNAscan-SE 在线网站对 tRNA 进行注释^[14],利用 RNAmmer 1.2 Server(<http://www.cbs.dtu.dk/services/RNAmmer/>)对 rRNA 进行注释,经人工修正后获得最终的注释结果;最后使用 OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>)软件生成基因组物理图谱^[15].

1.4 叶绿体基因组数据分析

根据 Sharp 等^[16]计算方法对组装好的叶菜型甘薯叶绿体基因组序列进行密码子偏好性(Relative Synonymous Codon Usage, RSCU)值统计和分析.采用 MISA 软件(<http://pgrc.ipk-gatersleben.de/misa/>)对组装好的叶菜型甘薯叶绿体基因组序列进行微卫星扫描^[17],利用 Tandem repeats finder v 4.04 软件检测串联重复序列,默认参数参考文献^[18].

由于叶绿体基因组组装过程中发现有两种类型的叶绿体,即叶绿体基因组在 SSR 区域具有正反两种方向的结构,因此利用 dottup 软件对两种类型的叶绿体基因组进行共线性分析^[19].

1.5 番薯属叶绿体基因组比较

从 NCBI 数据库下载番薯属 *Ipomoea trifida* (NC_034670), *Ipomoea cordatoteiloba* (NC_041204), *Ipomoea lacunosa* (NC_037912), *Ipomoea cynanchifolia* (NC_041203) 和 *Ipomoea ramosissima* (NC_041205) 等 45 个种植物叶绿体基因组序列与叶菜型甘薯进行聚类分析,利用其与两个外群物种全叶绿体序列构建进化树.使用软件 mafft(默认参数)进行序列比对^[20],而后利用 fasttree 软件构建 ML 进化树^[21].

1.6 甘薯种间叶绿体基因组 SNP 和 Indel 分析

从 NCBI 数据库下载 8 个甘薯叶绿体基因组数据,将 9 个基因组序列用 mafft 对比后,以‘福菜薯 18 号’叶绿体基因组为参考序列进行 SNP 和 Indel 分析.

2 结果与分析

2.1 叶菜型甘薯叶绿体基因组的基本特征

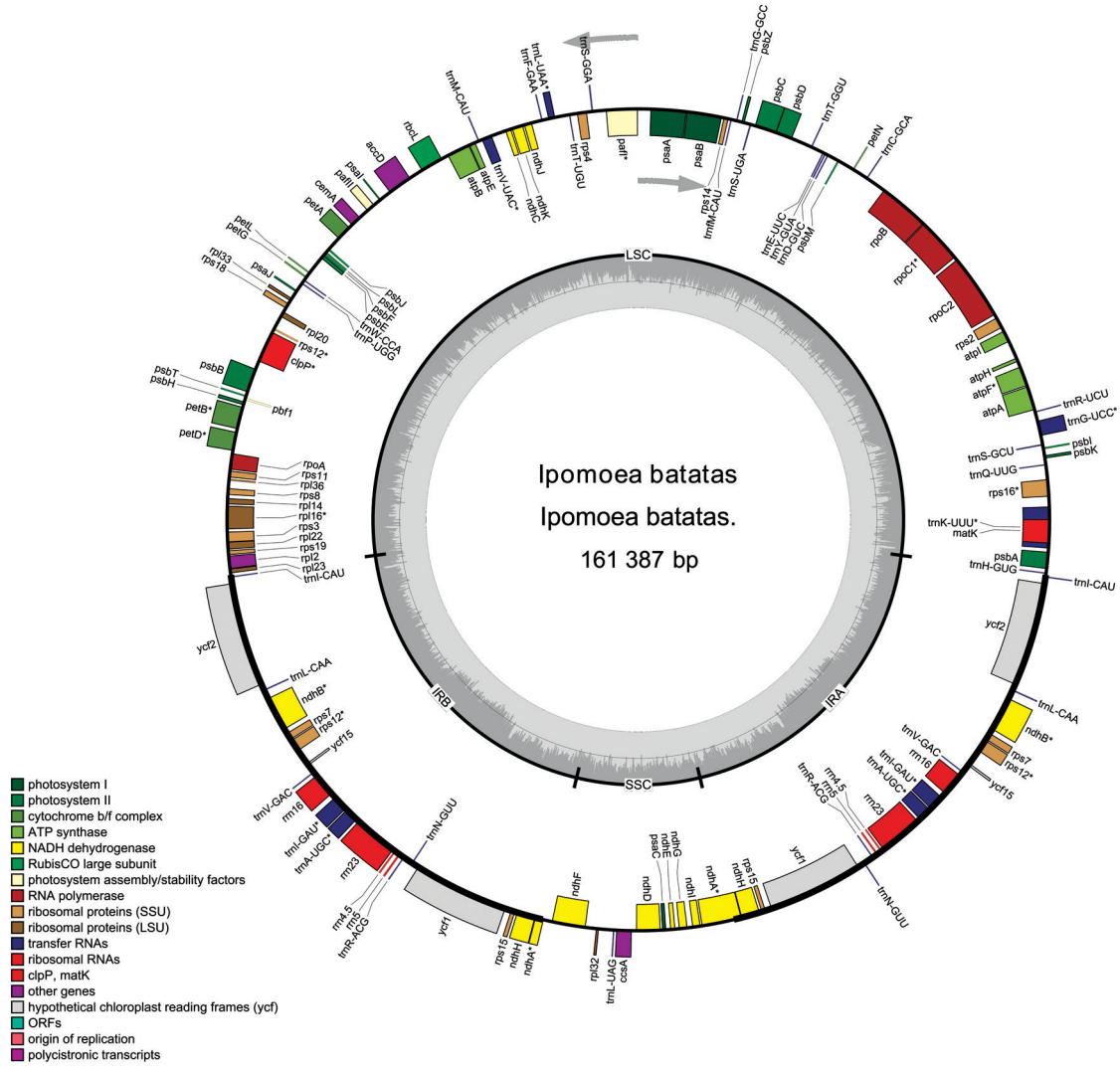
叶菜型甘薯的 cpDNA 序列全长为 161 387 bp,由大单拷贝区域(LSC, 87 597 bp),小单拷贝区域(SSC, 12 052 bp)及两个反向重复区域(IRA 和 IRB, 30 869 bp)4 个部分构成(图 1).组装注释好的叶绿体基因序列提交至 GenBank,获得序列登录号 OM808940.基因注释结果表明:叶菜型甘薯 cpDNA 具有 132 个功能基因,包括 87 个蛋白编码基因、8 个 rRNA 基因和 37 个 tRNA 基因.其中,18 个基因在 IR 区域复制,包括 7 个蛋白编码基因(*ycf1*, *ycf2*, *ycf15*, *ndhB*, *ndhH*, *rps7* 和 *rps15*),7 个 tRNA 基因(*trnI*-CAU, *trnL*-CAA, *trnV*-GAC, *trnI*-GAU, *trnA*-UGC, *trnR*-ACG 和 *trnN*-GUU)和 4 个 rRNA 基因(*rrn4.5*, *rrn5*, *rrn16* 和 *rrn23*).叶菜型甘薯 cpDNA 总的 GC 质量分

数为 37.54%，AT 质量分数为 62.46%。

经过统计分析，叶菜型甘薯 cpDNA 中 20 个基因包含内含子，其中，11 个蛋白编码基因和 7 个 tRNA 基因含有 1 个内含子，2 个蛋白编码基因(*psaI* 和 *clpP*)含有 2 个内含子(表 1)。 *rps12* 有 2 个拷贝，每个拷贝具有 3 个外显子，且两个拷贝共享第 1 个外显子，第 1 个外显子位于 LSC 区域，另外 2 个外显子位于 IR 区域。

2.2 叶菜型甘薯密码子的偏好性

对于不同的生物体蛋白质结构组成，即使编码氨基酸的密码子相同，但是对于氨基酸的同义密码子使用频率却是不相等的，这种同义密码子使用频率的不相等就是密码子偏好性(RSCU)。 叶菜型甘薯 cpDNA 中 RSCU 值大于 1.00 的密码子为 32 个，其中大多数以 A 或 T 结尾，仅 3 个以 G 结尾(ATG，TTG，TGG)。 编码亮氨酸(Leu)的密码子数量最多，为 2 950 个，占比 10.35%；半胱氨酸(Cys)出现的次数最少，为 332，占比 1.16%(表 2)。 这与大多数被子植物叶绿体基因组密码子使用偏好一致。



内侧基因顺时针转录，外侧基因逆时针转录。

图 1 叶菜型甘薯的 cpDNA 图谱

表 1 甘薯叶绿体基因组注释基因列表

基因分类	基因分组	基因名称
光合系统基因	光系统 I 基因	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	光系统 II 基因	<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>
	细胞色素复合物	<i>petN</i> , <i>petA</i> , <i>psbF</i> , <i>psbE</i> , <i>petG</i> , <i>petB^a</i> , <i>petD^a</i> , <i>petL</i>
	ATP 酶基因	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF^a</i> , <i>atpH</i> , <i>atpI</i>
	NADH 氧化还原酶基因	<i>ndhA^a</i> , <i>ndhB^{a,c}</i> , <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>
	二磷酸核酮糖羧化酶大亚基基因	<i>rbcL</i>
自我复制基因	RNA 聚合酶亚基基因 RNA	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1^a</i> , <i>rpoC2</i>
	核糖体蛋白小亚基基因	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7^c</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12^{b,c}</i> , <i>rps14</i> ,
		<i>rps15</i> , <i>rps16^a</i> , <i>rps18</i> , <i>rps19^c</i>
	核糖体蛋白大亚基基因	<i>rpl2</i> , <i>rpl14</i> , <i>rpl16^a</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23^c</i> , <i>rpl32</i> , <i>rpl33</i> ,
		<i>rpl36</i>
	核糖体核糖核酸基因	<i>rrn4</i> . 5 ^c , <i>rrn5^c</i> , <i>rrn16^c</i> , <i>rrn23^c</i>
		<i>trnA</i> -UGC ^{a,c} , <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnfM</i> -
其他基因	转移 RNA 基因	CAU, <i>trnF</i> -GAA, <i>trnG</i> -GCC, <i>trnG</i> -UCC ^a , <i>trnH</i> -GUG, <i>trnI</i> -
		CAU, <i>trnI</i> -GAU ^{a,c} , <i>trnK</i> -UUU ^a , <i>trnL</i> -CAA, <i>trnL</i> -UAA,
		<i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU, <i>trnP</i> -UGG, <i>trnQ</i> -UUG,
		<i>trnR</i> -ACG, <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, <i>trnV</i> -UAC ^a , <i>trnW</i> -CCA,
		<i>trnY</i> -GUA
其他基因	翻译起始密码子基因	<i>infA</i>
	成熟酶 K 基因	<i>matK</i>
	乙酰辅酶 A 羧化酶基因	<i>accD</i>
	膜蛋白基因	<i>cemA</i>
	细胞色素合成基因	<i>ccsA</i>
	蛋白酶基因	<i>clpP^b</i>
	假定叶绿体阅读框	<i>ycf1^c</i> , <i>ycf2^c</i> , <i>ycf15</i> , <i>pafl^b</i> , <i>paflII</i>

注：a 表示基因含有 1 个内含子，b 表示基因含有 2 个内含子，c 表示含有 2 个拷贝基因。

表 2 叶菜型甘薯各氨基酸同义密码子偏好性

氨基酸	密码子	RSCU 值	数量/个	氨基酸	密码子	RSCU 值	数量/个
丙氨酸 Ala(A)	GCA	1. 076	406	赖氨酸 Lys(K)	AAA	1. 499	1 266
	GCC	0. 615	232		AAG	0. 501	423
	GCG	0. 482	182	甲硫氨酸 Met(M)	ACG	0. 006	1
	GCT	1. 826	689		ATG	3. 950	637
精氨酸 Arg(R)	AGA	1. 780	508		ATT	0. 012	2
	AGG	0. 645	184		GTG	0. 031	5

续表 2

氨基酸	密码子	RSCU 值	数量/个	氨基酸	密码子	RSCU 值	数量/个
	CGA	1.391	397	苯丙氨酸 Phe(F)	TTC	0.684	574
	CGC	0.456	130		TTT	1.316	1104
	CGG	0.484	138	脯氨酸 Pro(P)	CCA	1.118	322
	CGT	1.244	355		CCC	0.882	254
天冬酰胺 Asn(N)	AAC	0.461	321		CCG	0.549	158
	AAT	1.539	1071		CCT	1.451	418
天冬氨酸 Asp(D)	GAC	0.400	232	丝氨酸 Ser(S)	AGC	0.478	179
	GAT	1.600	927		AGT	1.230	461
半胱氨酸 Cys(C)	TGC	0.548	91		TCA	1.102	413
	TGT	1.452	241		TCC	0.985	369
谷氨酰胺 Gln(Q)	CAA	1.554	832		TCG	0.619	232
	CAG	0.446	239		TCT	1.585	594
谷氨酸 Glu(E)	GAA	1.471	1194	终止子 TER	TAA	1.793	52
	GAG	0.529	429		TAG	0.586	17
甘氨酸 Gly(G)	GGA	1.531	695		TGA	0.621	18
	GGC	0.419	190	苏氨酸 Thr(T)	ACA	1.139	429
	GGG	0.811	368		ACC	0.778	293
	GGT	1.240	563		ACG	0.393	148
组氨酸 His(H)	CAC	0.505	166		ACT	1.691	637
	CAT	1.495	491	色氨酸 Trp(W)	TGG	1.000	469
异亮氨酸 Ile(I)	ATA	0.916	723	酪氨酸 Tyr(Y)	TAC	0.387	187
	ATC	0.602	475		TAT	1.613	780
	ATT	1.483	1171	缬氨酸 Val(V)	GTA	1.454	532
亮氨酸 Leu(L)	CTA	0.828	407		GTC	0.473	173
	CTC	0.417	205		GTG	0.601	220
	CTG	0.384	189		GTT	1.473	539
	CTT	1.340	659				
	TTA	1.810	890				
	TTG	1.220	600				

2.3 叶菜型甘薯 cpDNA 简单重复序列组成

SSR(Simple Sequence Repeats) 是一类由 1~6 个核苷酸为重复单位组成的长达几十个核苷酸的串联重复序列, 每个 SSR 两侧的序列一般是相对保守的单拷贝序列. 从叶菜型甘薯叶绿体基因组中共鉴定到 54 个 SSR 位点, 其中, 单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复分别有 32, 4, 3, 11, 2 和 2 个(表 3), 且 32 个单核苷酸重复均由 A 或 T 组成.

表 3 54 个 cpDNA 在叶菜型甘薯叶绿体基因组上的分布

编号	SSR 类型	SSR	片段大 小/bp	起始/ bp	终止/ bp	编号	SSR 类型	SSR	片段大 小/bp	起始/ bp	终止/ bp
1	p1	(T)13	13	174	186	28	p3	(GGA)4	12	59 833	59 844
2	p1	(A)11	11	1 499	1 509	29	p1	(T)10	10	61 307	61 316
3	p1	(A)12	12	1 627	1 638	30	p2	(TA)6	12	63 682	63 693
4	p4	(TTGT)3	12	6 482	6 493	31	p1	(A)12	12	65 204	65 215
5	p1	(A)15	15	7 849	7 863	32	p1	(T)10	10	68 002	68 011
6	p1	(A)11	11	8 103	8 113	33	p4	(AATA)3	12	68 611	68 622
7	p4	(AATA)3	12	9 787	9 798	34	p4	(TTTC)3	12	70 897	70 908
8	p1	(T)10	10	9 974	9 983	35	p1	(A)13	13	71 118	71 130
9	p1	(T)10	10	10 322	10 331	36	p1	(T)10	10	71 636	71 645
10	p1	(T)10	10	12 680	12 689	37	p1	(T)11	11	72 788	72 798
11	p1	(T)11	11	13 335	13 345	38	p3	(AGA)4	12	73 321	73 332
12	p4	(CAAT)3	12	13 896	13 907	39	p1	(T)10	10	73 678	73 687
13	p1	(A)10	10	15 855	15 864	40	p1	(A)11	11	74 034	74 044
14	p1	(T)13	13	18 901	18 913	41	p1	(T)11	11	77 369	77 379
15	p1	(T)10	10	26 702	26 711	42	p4	(AAAT)3	12	78 552	78 563
16	p1	(A)10	10	28 254	28 263	43	p6	(AATCAA)3	18	88 064	88 081
17	p1	(T)13	13	30 193	30 205	44	p4	(TATC)3	12	93 626	93 637
18	p4	(TCAA)3	12	30 875	30 886	45	p5	(TTCTA)4	20	100 375	100 394
19	p1	(T)10	10	31 059	31 068	46	p1	(A)12	12	110 849	110 860
20	p3	(TTC)4	12	36 291	36 302	47	p1	(A)10	10	111 260	111 269
21	p1	(T)10	10	36 505	36 514	48	p4	(ATAG)3	12	121 006	121 017
22	p1	(A)13	13	37 715	37 727	49	c	(T)10...(A)10	64	130 241	130 304
23	p1	(A)11	11	47 936	47 946	50	p1	(T)10	10	137 716	137 725
24	c	(T)10...(T)13	71	48 444	48 514	51	p1	(T)12	12	138 125	138 136
25	p4	(GAAA)3	12	48 969	48 980	52	p5	(AATAG)4	20	148 589	148 608
26	p1	(T)12	12	50 231	50 242	53	p4	(AGAT)3	12	155 347	155 358
27	c	(T)10...(T)13	97	56 116	56 212	54	p6	(ATTTTG)3	18	160 901	160 918

2.4 系统进化分析

从 NCBI 上下载旋花科甘薯近缘种植物的叶绿体基因组序列，对全基因组序列的共有蛋白编码基因进行聚类分析，并以 *Distimake quinquefolius* 和 *Operculina macrocarpa* 为外类群。结果显示，叶菜型甘薯与普通甘薯‘clm’和甘薯四倍体野生种 *Ipomoea tabascan*a 聚为一类，普通甘薯‘徐薯 18 号’和甘薯二倍体野生种 *Ipomoea trifida* 聚为一类，同时聚为一个大类。旋花科的黄毛银背藤也与番薯属植物聚为一类，两个外群植物 *Distimake quinquefolius* 和 *Operculina macrocarpa* 聚为一类，与番薯属植物分为两个进化支(图 2)。

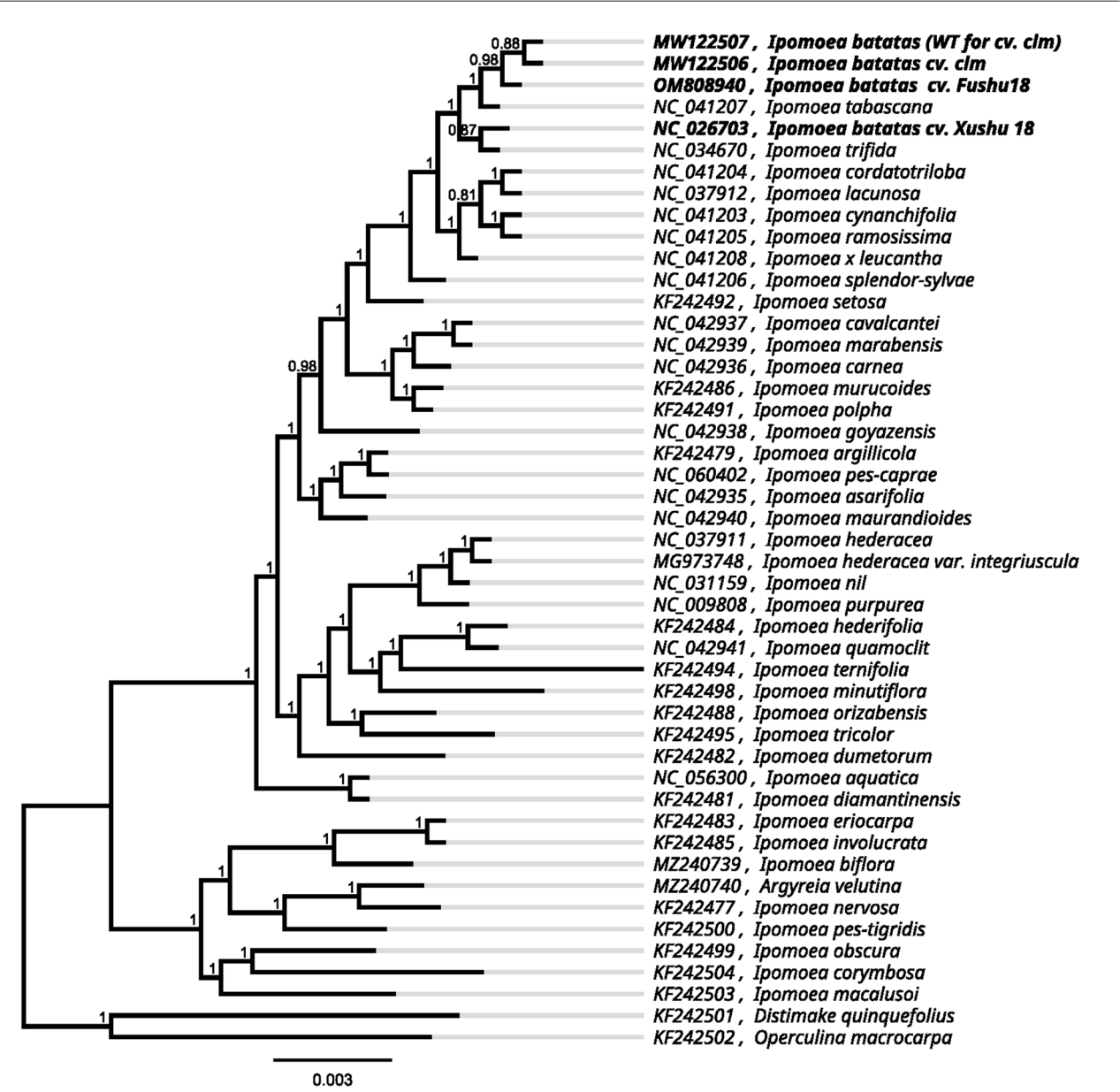


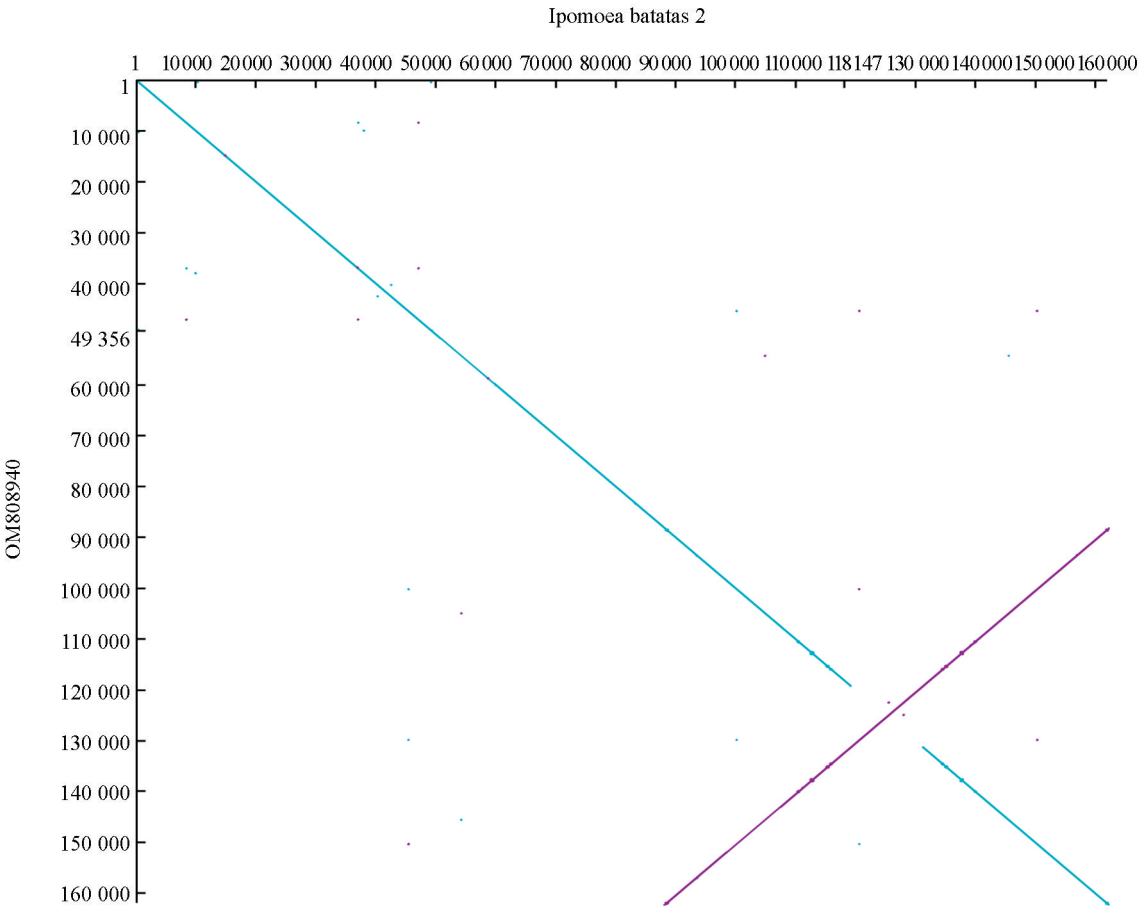
图 2 基于蛋白编码基因构建 45 个物种的系统进化树

2.5 叶菜型甘薯两种类型的叶绿体基因组

由于叶绿体基因组组装过程中发现有两种类型的叶绿体，即叶绿体基因组在 SSR 区域具有正反两种方向的结构，利用 dottup 软件对两种类型的叶绿体基因组进行共线性分析，其中蓝色是正向共线性，紫红色是反向共线性(图 3)。

2.6 甘薯品种间叶绿体基因组的 SNP 和 Indel 分析

以‘福菜薯 18 号’为对照，对甘薯品种间叶绿体基因组进行种内 SNP 分析，发现有 199 个 SNP 位点，其中有 118 个位点位于编码区中，编码区包含 *matK*, *rpoC2*, *psaB*, *accD*, *psbL*, *rps8*, *ycf1*, *ycf2*, *ndhB*, *ndhC*, *ndhE*, *ndhF*, *ndhH* 等基因，*ycf1* 和 *ycf2* 具有较多的 SNP 位点，其余位于非编码区中。‘福菜薯 18 号’叶绿体基因组与其他甘薯相比，有 121 个片段缺失，其中有 37 个属于 SSR 位点缺失；有 146 个片段插入，其中有 66 个属于 SSR 位点插入；另外还有 7 个长片段替换(数据略)。



右下角的蓝色为共同的 IRA 序列，左上角的长蓝色为共同的 LSC+IRB 序列，gap 部分为反向共线性的 SSR，由于 IRA 和 IRB 本身是反向共线性的，所有两种结构的叶绿体呈现 IRA-SSC-IRB 反向共线性(红色).

图 3 两种类型叶绿体 dotplot 图

3 讨论与结论

20 世纪 70 年代末，双脱氧终止法标志着第一代测序技术的诞生，实现了对 DNA 序列的测序与分析，由于这种方法测序通量低、自动化水平差等缺点，限制了其在转录组学和基因组学的发展。第二代测序是边合成边测序，通过捕捉末端新合成的碱基来获得待测 DNA 片段的序列，实现了高通量和自动化测定，极大地提高了测序速度，但由于二代测序包含 PCR 扩增等过程，可能会引入模板迁移等假阳性，而且二代测序读长普遍较短，也限制了其应用。三代测序技术以 PacBio 公司的单分子实时测序技术 (Single Molecule Real Time Sequencing, SMRT-seq) 和 Oxford Nanopore Technologies 的纳米孔单分子测序技术为代表，与前两代测序技术相比，其最大的特点就是单分子实时测序，测序过程无需进行 PCR 扩增，可以实现长片段序列测定，但缺点是通量相对较小，测序成本较高^[22-23]。本文为提高叶菜型甘薯叶绿体基因组测定序列的准确性，采用了二代和三代测序数据相结合，克服了测序技术本身的不足，保证序列拼接组装的准确性。

甘薯属于旋花科番薯属植物，本研究从 NCBI 上下载了 45 个番薯属植物的叶绿体基因组数据，包括普通甘薯‘徐薯 18 号’等叶绿体基因组数据并进行序列对比，并以 *Distimake quinquefolius* 和 *Operculina macrocarpa* 为外类群构建系统进化树，结果显示，叶菜型甘薯与甘薯四倍体野生种 *Ipomoea tabascana* 和

甘薯二倍体野生种 *Ipomoea trifida* 聚为一类。Srisuwan 等^[24]通过细胞遗传学方法认为甘薯栽培种与野生种 *Ipomoea trifida* 具有更为密切关系, *Ipomoea trifida* 是甘薯栽培种和四倍体 *Ipomoea tabascana* 的祖先, 与叶绿体基因组进化相一致。有研究表明, 番薯属植物 *Ipomoea purpurea* 与其近缘种 *Ipomoea alba* 和 *Ipomoea nil* 在花青素合成基因中有约 1% 的核苷酸多态性, 符合分子进化的标准中性模型^[25]。甘薯种内 SNP 分析结果表明, 甘薯叶绿体基因组总长度有所差异, ‘徐薯 18 号’的叶绿体基因组长度为 161 303 bp^[26], 甘薯‘clm’及其对照分别为 161 393 bp 和 161 429 bp^[27], 甘薯叶绿体基因中存在单核苷酸突变, 也存在长片段替换、缺失和插入, 这些碱基差异有些是位于编码区基因 *matK*, *rpoC2*, *psaB*, *accD*, *rbcL* 中, 其中 *matK* 和 *rbcL* 是植物 DNA 条形码的核心序列^[28], 另外一些编码基因是否可以成为番薯属的 DNA 条形码有待更多数据的支持。

参考文献:

- [1] 李保珠, 赵孝亮, 彭雷. 植物叶绿体发育及调控研究进展 [J]. 植物学报, 2014, 49(3): 337-345.
- [2] 程慧, 葛春峰, 张红, 等. 果树叶绿体基因组测序及系统发育研究进展 [J]. 核农学报, 2018, 32(1): 58-69.
- [3] 杨亚蒙, 焦健, 樊秀彩, 等. 桑叶葡萄叶绿体基因组及其特征分析 [J]. 园艺学报, 2019, 46(4): 635-648.
- [4] 宋芸, 贾孟君, 曹亚萍, 等. 连翘叶绿体基因组特征分析 [J]. 园艺学报, 2022, 49(1): 187-199.
- [5] 王璐璐, 黄雨, 傅玉凡, 等. 甘薯淀粉相关性状变异性的分析 [J]. 西南大学学报(自然科学版), 2022, 44(2): 39-47.
- [6] 李静, 傅玉凡, 黄雨, 等. 10 个叶菜型甘薯品种茎尖性状的分析与评价 [J]. 西南大学学报(自然科学版), 2022, 44(4): 45-53.
- [7] 胡颖, 王茜, 张新新, 等. 叶绿体 DNA 标记在谱系地理学中的应用研究进展 [J]. 生物多样性, 2019, 27(2): 219-234.
- [8] PAHLICH E, GERLITZ C. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue [J]. Phytochemistry, 1980, 19: 11-13.
- [9] CHEN S F, ZHOU Y Q, CHEN Y R, et al. Fastp: an Ultra-fast All-in-one FASTQ Preprocessor [J]. Bioinformatics, 2018, 34(17): i884-i890.
- [10] LI H. Minimap2: Pairwise Alignment for Nucleotide Sequences [J]. Bioinformatics, 2018, 34(18): 3094-3100.
- [11] LANGMEAD B, SALZBERG S L. Fast Gapped-read Alignment with Bowtie 2 [J]. Nature Methods, 2012, 9(4): 357-359.
- [12] WICK R R, JUDD L M, GORRIE C L, et al. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads [J]. PLoS Computational Biology, 2017, 13(6): e1005595.
- [13] TILLICH M, LEHWARK P, PELLIZZER T, et al. GeSeq-versatile and Accurate Annotation of Organelle Genomes [J]. Nucleic Acids Research, 2017, 45(W1): W6-W11.
- [14] LOWE T M, EDDY S R. TRNAscan-SE: a Program for Improved Detection of Transfer RNA Genes in Genomic Sequence [J]. Nucleic Acids Research, 1997, 25(5): 955-964.
- [15] LOHSE M, DRECHSEL O, BOCK R. Organellar Genome DRAW (OGDRAW): a Tool for the Easy Generation of High-quality Custom Graphical Maps of Plastid and Mitochondrial Genomes [J]. Current Genetics, 2007, 52(5): 267-274.
- [16] SHARP P M, LI W H. Codon Usage in Regulatory Genes in *Escherichia coli* does not Reflect Selection for ‘Rare’ Codons [J]. Nucleic Acids Research, 1986, 14(19): 7737-7749.

[17] BEIER S, THIEL T, MÜNCH T, et al. MISA-web: a Web Server for Microsatellite Prediction [J]. Bioinformatics, 2017, 33(16): 2583-2585.

[18] BENSON G. Tandem Repeats Finder: a Program to Analyze DNA Sequences [J]. Nucleic Acids Research, 1999, 27(2): 573-580.

[19] RICE P, LONGDEN I, BLEASBY A. EMBOSS: The European Molecular Biology Open Software Suite [J]. Trends in Genetics: TIG, 2000, 16(6): 276-277.

[20] KATOH K, STANDLEY D M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability [J]. Molecular Biology and Evolution, 2013, 30(4): 772-780.

[21] PRICE M N, DEHAL P S, ARKIN A P. Fast Tree 2-approximately Maximum-likelihood Trees for Large Alignments [J]. PLoS One, 2010, 5(3): e9490.

[22] 李玉梅, 李书娴, 李向上, 等. 第三代测序技术在转录组学研究中的应用 [J]. 生命科学仪器, 2018, 16(S1): 114-121, 113.

[23] 战斌慧, 周雪平. 高通量测序技术在植物及昆虫病毒检测中的应用 [J]. 植物保护, 2018, 44(5): 120-126, 167.

[24] SRISUWAN S, SIHACHAKR D, SILJAK-YAKOVLEV S. The Origin and Evolution of Sweet Potato (*Ipomoea batatas* Lam.) and Its Wild Relatives through the Cytogenetic Approaches [J]. Plant Science, 2006, 171(3): 424-433.

[25] GONZALES A M, FANG Z, DURBIN M L, et al. Nucleotide Sequence Diversity of Floral Pigment Genes in Mexican Populations of *Ipomoea purpurea* (Morning Glory) Accord with a Neutral Model of Evolution [J]. Journal of Heredity, 2012, 103(6): 863-872.

[26] YAN L, LAI X J, LI X D, et al. Analyses of the Complete Genome and Gene Expression of Chloroplast of Sweet Potato (*Ipomoea batata*) [J]. PLoS One, 2015, 10(4): e0124083.

[27] ZOU H D, ZHANG X J, CHEN J Y, et al. Complete Chloroplast Genome of a Novel Chlorophyll-deficient Mutant (*clm*) in Sweetpotato (*Ipomoea batatas* L.) [J]. Mitochondrial DNA Part B, 2021, 6(3): 968-969.

[28] LI F W, KUO L Y, ROTHFELS C J, et al. rbcL and matK Earn Two Thumbs up as the Core dna Barcode for Ferns [J]. PLoS One, 2011, 6(10): e26597.

责任编辑 周仁惠