

殷剑美, 王立, 蒋璐, 等. 芋 [*Colocasia esculenta* (L.) Schott] 的倍性水平及基因组特征解析[J]. 江苏农业学报, 2019, 35(6): 1284-1291.

doi: 10.3969/j.issn.1000-4440.2019.06.003

芋 [*Colocasia esculenta* (L.) Schott] 的倍性水平及基因组特征解析

殷剑美, 王立, 蒋璐, 韩晓勇, 郭文琦, 李春宏, 张培通

(江苏省农业科学院经济作物研究所江苏省农业生物学重点实验室, 江苏 南京 210014)

摘要: 采用流式细胞术, 对 2 个芋 [*Colocasia esculenta* (L.) Schott] 品种苏芋 1 号和香蕉芋的倍性水平和 DNA 含量进行了测定, 2 倍基因组 DNA 含量分别为 7.71 pg 和 5.25 pg, 倍性分别为三倍体和二倍体。同时通过 HiSeq 2500 测序, 分别获得 240.53 Gb 和 87.69 Gb 数据, 芋基因组大小为 2.08~2.38 Gb, GC 含量为 42.57%~42.75%, 杂合度为 0.52%~2.29%。根据苏芋 1 号的基因组数据, 组装得到基因组序列为 2.27 Gb, Contig N50 为 359 bp; 共注释获得 24 302 个基因, 99.43% 的基因可注释到 6 个公共数据库; 筛选出 928 987 个 SSR 位点。

关键词: 芋; 流式细胞术; 基因组测序; 基因功能注释; 倍性水平; 基因组特征

中图分类号: S632.3 **文献标识码:** A **文章编号:** 1000-4440(2019)06-1284-08

Analysis of ploidy level and genome characterization on taro [*Colocasia esculenta* (L.) Schott]

YIN Jian-mei, WANG Li, JIANG Lu, HAN Xiao-yong, GUO Wen-qi, LI Chun-hong, ZHANG Pei-tong

(Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Provincial Key Laboratory of Agrobiotechnology, Nanjing 210014, China)

Abstract: In this study, the ploidy levels and DNA contents of two *Colocasia esculenta* varieties, Suyu No.1 and Xiangjiaoyu from China, were determined by flow cytometry (FCM). The 2C DNA contents of them were 7.71 pg and 5.25 pg, respectively, which indicating that their ploidy levels were triploid and diploid. In total, 240.53 Gb and 87.69 Gb sequences were obtained through HiSeq 2500 sequencing methods, respectively. The genome size of *C. esculenta* was in the range of 2.08–2.38 Gb. The GC contents ranged from 42.57% to 42.75%, and the heterozygosity was in the range of 0.52%–2.29%. According to the genomic data of Suyu No.1, the assembled genome sequence was about 2.27 Gb, and contig N50 was 359 bp. A total of 24 302 genes were generated, and 24 163 (about 99.43%) gene sequences were successfully annotated from six public databases. A total of 928 987 SSR loci were screened out.

Key words: *Colocasia esculenta*; flow cytometry; genome sequencing; gene function annotation; ploidy level; genomic characterization

收稿日期: 2019-04-01

基金项目: 国家自然科学基金项目 (31501776); 江苏省农业生物学重点实验室项目 (4911707Z201707)

作者简介: 殷剑美 (1977-), 女, 江苏丹阳人, 博士, 研究员, 主要从事药食同源作物育种与分子生物学研究。(Tel) 025-84390860; (E-mail) yinjm2006@jaas.ac.cn

通讯作者: 张培通, (E-mail) ptzhang1965@163.com

芋 [*Colocasia esculenta* (L.) Schott] 是天南星科芋属的多年生单子叶草本湿生植物, 具地下球茎, 常作一年生栽培。芋的主要食用部分为地下球茎, 因其富含淀粉, 是热带和亚热带地区最古老的主要粮食作物之一^[1], 在全球蔬菜消费量中居第 14 位^[2]。芋中的非淀粉性多糖 (TPS) 具有增强细胞免疫和体

液免疫的功能^[3],芋还具有较好的食用、药用、观赏和文化价值^[4-5],越来越受到消费者青睐。根据联合国粮农组织(FAO)统计,中国芋的栽培面积居世界第4位,主要分布于江苏、山东、浙江、广西、广东和云南等省^[6]。中国芋资源非常丰富,栽培品种较多,仅江苏省就有7个芋地方特色品种被认定为中国地理标志农产品。

截至2019年3月,还未见芋属物种完整基因组信息的报道,在NCBI网站上芋属也只有1 279个DNA或RNA序列(<https://www.ncbi.nlm.nih.gov/nucore/>),以及有关系统发育和种群研究的41个序列集(<https://www.ncbi.nlm.nih.gov/popset/>)和9个表达序列标记(EST)(<https://www.ncbi.nlm.nih.gov/nucest/>)。基因组数据的缺乏严重制约了芋的基础研究和产业发展。而基因组大小及其特征对于许多领域的研究至关重要,包括分类学和进化分析^[7],尤其是对基因克隆和基因组测序计划^[8]。

目前常用的基因组大小测定方法为流式细胞术(Flow cytometry, FCM)^[9],该方法简单、快速、灵敏、价格低廉,可检测单个细胞倍性以及细胞核DNA含量^[10-11],已被应用于多种植物的基因组大小评估^[12-14]。但是由于流式细胞术在检测植物样品时会受到植物细胞结构及次生代谢产物的影响,其检测结果的准确性有所降低^[15]。高通量测序技术(Next-generation sequencing, NGS)则克服了内源性物质对植物基因组大小评估的影响^[16],在评估基因组大小外还可同时获得更多的物种信息,如杂合率、重复序列、GC含量等^[17],该技术通量高、速度快、信息量大、准确性好^[18]。随着测序成本大幅降低,K-mer分析成为基因组评估的重要手段^[19-22]。

芋一般以二倍体和三倍体形式存在^[21-25],倍性复杂,从外观特征难以区分,基因组信息也相对缺乏。因此本研究采用流式细胞术和K-mer分析相结合,对芋DNA含量和倍性进行检测,以期获得更好的效果,同时对芋的基因组特征进行评估,包括基因组大小、杂合度和GC含量,进一步整合已有转录组数据^[26]获得更准确的基因预测和注释以及SSR标记,为今后芋基因组从头测序及基因组信息的更好利用奠定基础。

1 材料与方法

1.1 试验材料

选取2个江苏省著名地方特色芋品种:多子芋代表性品种苏芋1号,由靖江香沙芋(中国地理标志农产品)经系统选育而成;魁芋类代表性品种香蕉芋。流式细胞试验对照为玉米CE-777(*Zea mays* L.),种子由Dolezel博士(Institute of Experimental Botany, Olomouc, Czech Republic)提供。所有材料均种植于江苏省农业科学院试验基地。

1.2 试验方法

1.2.1 流式细胞分析 取0.5 cm²新鲜叶片样品置于含有1 ml WPB缓冲液(0.2 mol/L Tris·HCl, 4 mmol/L MgCl₂·6H₂O, 2 mmol/L EDTA Na₂·2H₂O, 86 mmol/L NaCl, 10 mmol/L Na₂S₂O₅, 1% PVP-10, 1% Triton X-100, pH 7.5)^[27]的培养皿中,用锋利刀片迅速将叶片切碎,40 μm细胞滤网(Biologix group limited)过滤除去悬浮物,加入50 μl PI(Propidium iodide, 碘化丙啶, 50 μg/ml, 含500 μg/ml RNA酶)染色10 min,上机检测PI的发射荧光强度。实验操作均在冰上进行。PI是一种荧光染料,能够嵌入双链DNA的碱基对中,在着色过程中嵌入量与DNA含量成正比,因而可以用荧光强度表示DNA的相对含量。流式细胞仪(FACS Jazz™, 美国BD公司)采用488 nm的蓝光激发,每个待测样本至少收集10 000个细胞,重复3次。测定所得的图像和数据由流式细胞仪自带软件FACS™进行处理分析。玉米CE-777(2C=5.43 pg)^[28]作为内标,样品2C值(pg)=(样品平均峰值/内标平均峰值)×5.43,基因组大小计算公式为1 pg=978 Mb。

1.2.2 基因组测序和基因组特征分析 基因组测序采用Illumina HiSeq平台完成。采用SOAPdenovo软件(<http://soap.genomics.org.cn/soapdenovo.html>)进行de novo组装,利用K-mer分析估计基因组大小(基因组大小=K-mer数量/深度分布峰值)、重复序列和杂合率。用SOAP软件将过滤之后的读长(reads)比对到组装序列上,获得碱基深度,以10 kb为窗口,在序列上无重复前进,计算每个窗口的平均深度与GC含量,作出GC_depth点图。GC含量计算公式:GC=(C+G)/(A+C+G+T)×100%。

1.2.3 SSR标记开发 利用MISA(<http://pgrc.ipk-gatersleben.de/misa/misa.html>)进行基因组SSR分

析。筛选微卫星重复单元的参数设定为:重复单元为 2 到 6 个核苷酸基序,二核苷酸重复 6 次及以上,其他重复单元重复 5 次及以上。

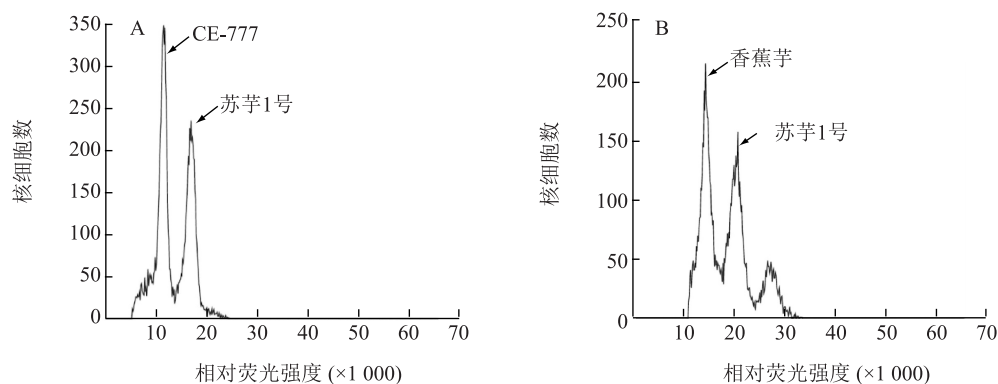
1.2.4 基因预测和注释 基因预测主要采用转录组预测、从头预测和基于同源物种预测 3 种不同策略,然后利用 EVM v1.1.1^[29] 软件对预测结果进行整合。其中整合 Unigene 进行转录组预测;使用 Genscan^[30]、GlimmerHMM v3.0.4^[31]、GeneID v1.4^[32]、SNAP(version 2006-07-28)^[33] 进行从头预测;使用 GeMoMa v1.3.1^[34] 进行基于同源物种的预测,主要使用了拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)和高粱(*Solanum tuberosum*) 3 个物种;最后利用 EVM v1.1.1 整合上述 3 种方法得到预测结果。对预测得到的基因序列与 NR、KOG、GO、KEGG、TrEM-

BL 等功能数据库进行 BLAST v2.2.31^[35] 比对,并进行基因的 KEGG 通路注释分析、KOG 功能注释分析、GO 功能注释分析等基因功能注释分析。

2 结果与分析

2.1 芋的倍性和基因组 DNA 含量

苏芋 1 号和香蕉芋的 PI 发射荧光强度测定结果表明,苏芋 1 号 DNA 含量峰的峰值是对照 CE-777 的 1.42 倍(图 1-A),是香蕉芋峰值的 1.48 倍(图 1-B)。对照 CE-777 的 2C 值约为 5.43 pg,则苏芋 1 号约为 7.71 pg,香蕉芋约为 5.25 pg,可知苏芋 1 号为三倍体,香蕉芋为二倍体。综合两者结果,芋基因组 DNA 含量为 2.57~2.62 pg,而基因组大小为 2.51~2.56 Gb(1 pg = 978 Mb)(表 1)。



A: 苏芋 1 号与玉米 CE-777 的混合样品; B: 苏芋 1 号与香蕉芋的混合样品。

图 1 芋的流式细胞分析结果

Fig.1 Analysis results of taro by flow cytometry

表 1 芋倍性和 DNA 含量统计

Table 1 Statistics on ploidy and DNA content of taro

品种	倍性	2C 值 (pg)	基因组 DNA 含量 (pg)	基因组大小 (Gb)
苏芋 1 号	$2n=3x$	7.71	2.57	2.51
香蕉芋	$2n=2x$	5.25	2.62	2.56

2.2 芋的基因组大小、杂合率、重复序列含量及 GC 含量

苏芋 1 号测序共获得数据 240.53 Gb,总测序深度 110.16, Q20 比例达到 97.0%, Q30 比例达到 92.7%;香蕉芋测序获得数据 87.69 Gb,总测序深度 36.92, Q20 比例达到 96.8%, Q30 比例达到 92.6%(表 2)。

表 2 苏芋 1 号和香蕉芋序列数据统计

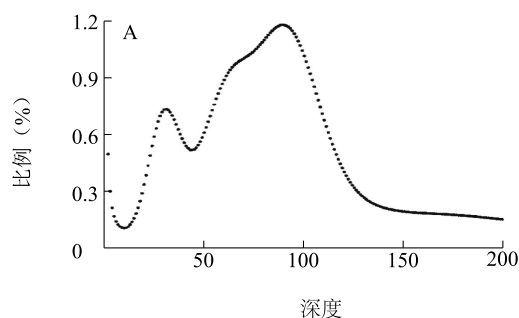
Table 2 Statistics of on sequence data of Suyu No. 1 and Xiangjiaoyu

品种	文库大小 (bp)	数据 (Gb)	深度	Q20 (%)	Q30 (%)
苏芋 1 号	270	240.53	110.16	97.0	92.7
香蕉芋	270	87.69	36.92	96.8	92.6

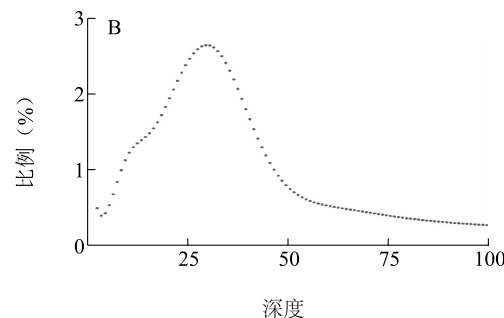
Q20 和 Q30 分别表示碱基识别正确率为 99% 和 99.9%。

由苏芋 1 号 Kmer 分布图(图 2A)可知, Kmer 存在 3 个峰,主峰(Kmer 深度为 90)、2/3 峰(Kmer 深度为 58)和 1/3 峰(Kmer 深度为 30),从 Kmer 深度的倍数关系推断苏芋 1 号是三倍体($2n=3x$)。根据三倍体物种基因组大小计算方式,以 1/3 峰位置的 Kmer 深度 30 计算得到三套染色体($3x$)大小约为

6.24 Gb, 则单套基因组大小为 2.08 Gb; 以主峰 Kmer 深度 90 估计基因组大小约为 2.18 Gb, 即单套基因组大小约为 2.18 Gb。由香蕉芋 Kmer 分布图



(图 2B) 可知, 香蕉芋为二倍体, Kmer 主峰深度为 31, 单套基因组大小为 2.38 Gb。综合两者结果, 芋的基因组大小在 2.08Gb 至 2.38 Gb 之间。



A: 苏芋 1 号; B: 香蕉芋。纵坐标表示该深度处的频率/所有深度的总频率的比例。基因组大小 = K-mer 数量/峰深度。

图 2 芋的 Kmer 分布图

Fig.2 Distribution of Kmer analysis results in taro

苏芋 1 号和香蕉芋的杂合率分别为 2.29% 和 0.52%, 重复序列含量分别为 77.03% 和 73.97%。

将基因组测序结果用 SOAP 软件过滤之后的 reads 比对到组装序列上, 获得碱基深度, 作出苏芋 1 号的 GC-depth 点图。由图 3 可知, GC-depth 明显分为 3 层, 推测苏芋 1 号为三倍体。对 reads 的 A、T、C、G 碱基含量进行分析, 结果表明其中 G 含量为 467 718 565 个, C 含量为 476 195 565 个, 综合 (ATGC) 含量为 2 217 137 760 个, 则 GC 含量为 42.57%。

2.3 芋的基因组组装分析

对测序数据过滤掉低质量数据后, 利用 SOAPdenovo 软件进行初步组装, 总共获得长度在 200 bp 以上的重叠群 (Contig) 7 582 094 条, 总长度为 2.22 Gb, Contig N50 长度为 359 bp, Contig N90 长度为 135 bp, 其中最长的 Contig 长度为 33 381 bp, GC 含量为 42.57%。将 Contig 进一步拼接, 共得到 4 237 691 条骨架 (Scaffold), 基因组序列总长度为 2.27 Gb, 其中 Scaffold N50 长度为 654 bp, Scaffold N90 长度为 246 bp, 最长的 Scaffold 长度为 34 394 bp, Gap 总长度为 49 887 191 bp。

2.4 芋的简单重复序列

利用 MISA 软件对初步组装基因组序列进行 SSR 分析 (表 3), 从 344 172 345 9 bp 基因组序列中鉴定获得 928 987 个 SSR 位点, 其中单碱基重复 SSR 数目占基因碱基序列总数的 58.66%。双碱基重复 SSR 占 31.46%, 三碱基重复 SSR 占 9.25%, 四碱基重复 SSR 占 0.45%, 五碱基重复 SSR 占 0.14%, 六碱基重复 SSR 占 0.03%。

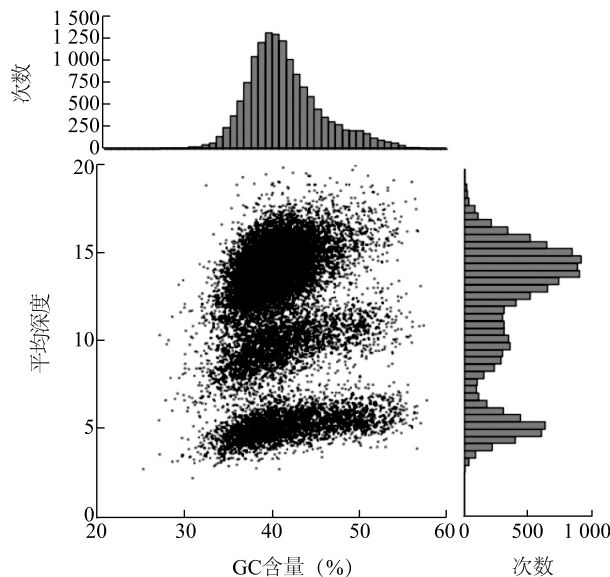


图 3 苏芋 1 号的 GC-depth 点图

Fig.3 GC content and average sequencing depth of *C. esculenta* Suya No.1

2.5 芋的基因预测和功能注释

对苏芋 1 号基因组中大于 1 kb 以上的序列进行基因预测, 最终得到 24 302 个基因 (表 4)。

预测得到的基因中, 总共有 99.43% 的基因可以注释到 NR 等数据库 (表 5), 其中 13 532 个基因注释到 KOG 数据库, 被分为 25 个功能大类 (图 4), 基因数目最多的 3 类分别为“基本功能预测”类有 5 250 个基因, 占 35.53%; “翻译后修饰”类有 1 254 个基因, 占 8.49%; “信号传导机制”类有 1 133 个基因, 占 7.67%。同样有 11 349 个基因被注释到 GO 数据库 (图 5), 其中细胞组成大类中有关“细胞”的

基因数目最多,有5 181个,占 46.27%;分子功能大类中,“催化活性”类基因数目最多,为5 335个,占 47.01%;生物功能大类中,“代谢过程”类有5 752个基因,占 50.68%。另外有6 823个基因被定位到 125 条 KEGG 通路中。

表 3 苏芋 1 号基因组 SSR 分析

Table 3 Simple sequence repeat (SSR) analysis of the genome in Suyu No.1

类型	数目	百分比(%)
评估的序列数目	13 105 284	—
评估的序列总碱基数目(bp)	3 441 723 459	—
识别的 SSR 总数	92 8987	100.00
包含 SSR 的序列数目	774 840	83.41
包含 1 个以上 SSR 的序列数目	106 223	11.43
以复合物形式存在的 SSR 数目	58 796	6.33
单碱基重复	544 987	58.66
双碱基重复	292 281	31.46
三碱基重复	85 961	9.25
四碱基重复	4 170	0.45
五碱基重复	1 345	0.14
六碱基重复	243	0.03

表 4 苏芋 1 号基因预测

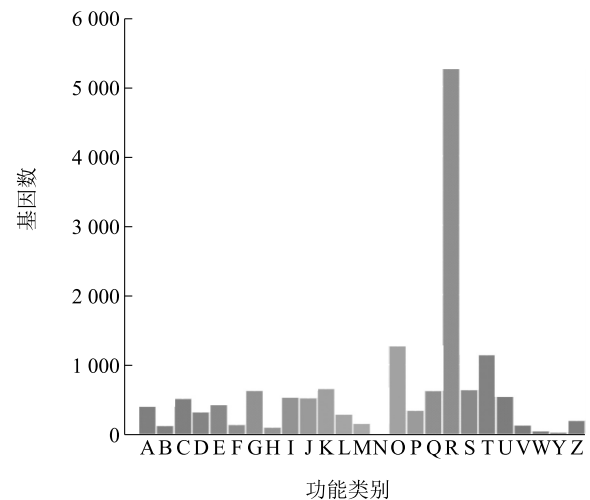
Table 4 Gene prediction of Suyu No.1

方法	软件	物种	基因数目
从头预测	Genscan	—	37 996
	GlimmerHMM	—	52 805
	GeneID	—	62 457
	SNAP	—	61 438
同源物种	GeMoMa	<i>Arabidopsisthaliana</i>	16 224
		<i>Oryzasativa</i>	22 457
		<i>Solanum tuberosum</i>	25 339
整合	EVM	—	24 302

表 5 苏芋 1 号基因功能注释

Table 5 Annotation of gene function in Suyu No.1

数据库	注释基因数目	比例(%)
KOG	13 532	55.68
GO	11 349	46.70
KEGG	6 823	28.08
TrEMBL	24 021	98.84
NR	23 954	98.57
All	24 163	99.43



A: RNA 加工和修饰; B: 染色体结构和动力学; C: 能量生成和转化; D: 细胞循环控制、细胞分裂、染色体分离; E: 氨基酸转运和代谢; F: 核苷酸转运和代谢; G: 碳水化合物转运和代谢; H: 合酶转运和代谢; I: 脂类转运和代谢; J: 翻译、核糖体结构和增殖; K: 转录; L: 复制、重组和修复; M: 细胞壁/细胞膜/包装合成; N: 细胞运动; O: 翻译后修饰、蛋白质折叠、配体; P: 无机离子转运和代谢; Q: 次生代谢产物的合成、转运和代谢; R: 基本功能预测; S: 未知功能; T: 信号传导机制; U: 细胞内交换、分泌和脉管转运; V: 防御机制; W: 细胞外结构; Y: 细胞核结构; Z: 细胞骨架。

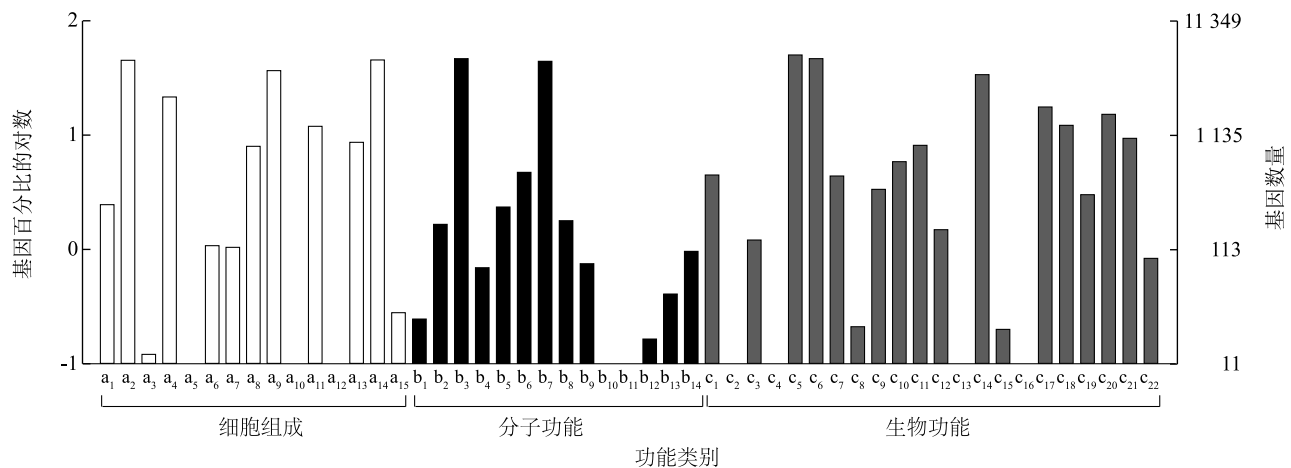
图 4 芋 KOG 功能分类

Fig.4 Functional classification of the *Colocasia esculenta* genes according to KOG criteria

3 讨论

3.1 芋资源倍性的鉴定

中国芋资源丰富,按照芋的分蘖习性可分为魁芋(母芋发达味美,子芋少),多子芋(子芋多且品质好)和多头芋(母芋群生,子芋甚少)。在南方,以魁芋类型居多,如广西荔浦芋、福建漳州黄肉芋等。而在长江流域,主要以多子芋品种居多,如国家地理标志农产品乌柄红芽的靖江香沙芋、紫柄白芽的乌骨芋。黄新芳等^[36]研究结果显示,中国芋中魁芋的染色体倍性为 $2n = 2x = 28$,多头芋和魁子兼用芋为 $2n = 3x = 42$,而多子芋一般为 $2n = 3x = 42$ 。但由于芋对环境的适应性导致其形态发生了多种变化,包括叶形、叶色和球茎肉色等性状的多样化。对于外观相似、产地来源不同的芋品种,仅仅根据形态学特征难以进行有效区分。Isshikis 等^[37]根据天冬氨酸转氨酶(AAT)、莽草酸脱氢酶(SKDH)和 6-磷酸葡萄糖异构酶(GPI)等 3 种同工酶的带型,对孟加拉国



a₁:细胞外区域;a₂:细胞;a₃:拟核;a₄:细胞膜;a₅:病毒颗粒;a₆:细胞连接;a₇:膜封闭腔;a₈:大分子复合物;a₉:细胞器;a₁₀:细胞外区部分;a₁₁:细胞器部分;a₁₂:病毒颗粒部分;a₁₃:细胞膜部分;a₁₄:细胞部分;a₁₅:超分子复合物。b₁:转录因子活性,蛋白质结合;b₂:核酸结合转录因子活性;b₃:催化活性;b₄:信号转导活性;b₅:结构分子活性;b₆:转运活性;b₇:结合;b₈:电子载体活性;b₉:抗氧化活性;b₁₀:金属伴侣活性;b₁₁:蛋白质标签;b₁₂:营养库活性;b₁₃:分子转导活性;b₁₄:分子功能调节。c₁:复制;c₂:细胞杀伤;c₃:免疫系统过程;c₄:习性;c₅:代谢过程;c₆:细胞过程;c₇:生殖过程;c₈:生物粘附;c₉:信号;c₁₀:多细胞器官过程;c₁₁:发育过程;c₁₂:生长;c₁₃:移动;c₁₄:单一生物过程;c₁₅:生物相;c₁₆:节律过程;c₁₇:应激反应;c₁₈:定位;c₁₉:多生物过程;c₂₀:生物调节;c₂₁:细胞成分组织或生物发生;c₂₂:脱毒。

图5 芋 GO 功能分类

Fig.5 Functional classification of *C. esculenta* genes based on the GO annotations

的野生芋资源进行了鉴定,发现二倍体和三倍体品种间存在较大的基因组相似性,这大大增加了从表观性状对芋资源倍性进行区分的难度。传统的细胞学观察可以对芋的染色体数进行有效鉴定和区分,但需要检测者具有丰富的细胞学经验,且耗时长、效率低。目前中国大部分芋资源的基因组倍性研究结果较为模糊和混乱。江苏省处于南北交界的长江下游,芋品种丰富多样,在倍性水平和 DNA 含量的检测和鉴定上也存在一定的困难。根据本研究结果,利用流式细胞术检测多子芋品种苏芋 1 号为三倍体,魁芋品种香蕉芋为二倍体,这与黄新芳等^[35]研究结果较为一致,因此更为简单、快捷的流式细胞检测方法可用于芋的倍性初步鉴定。但由于芋倍性的多样性和基因组大小的复杂性,还需其他方法来提高其检测的准确性。

3.2 芋基因组大小评估

评估基因组大小的方法主要有 3 种:物理作图法^[38]、流式细胞术和高通量测序技术。其中,物理作图法繁琐,耗时长,不适用于无性繁殖作物。流式细胞术比较简便、快速,但准确性会受多种因素影响。高通量测序技术几乎不受外在因素影响,准确性较高,随着测序成本的大幅降低,已广泛应用于植

物基因组大小的预测。在本研究中,用流式细胞仪评估的芋基因组大小为 2.51~2.56 Gb,与 K-mer 的分析结果 2.08~2.38 Gb 较为一致,但仍有偏差。进一步说明流式细胞检测可用于基因组大小的初步快速评估,但要提高其准确性还需结合更准确的测序结果。

3.3 芋基因组特征

苏芋 1 号基因组组装获得的总长度为 2.27 Gb,而实际其三倍体基因组为 3.27~3.92 Gb,组装结果偏小,可能是由于二代测序的短序列导致,也可能跟芋基因组重复序列较多、倍性复杂等因素有关。芋 GC 含量为 42.66%,高于马铃薯(34.8%~36.0%)^[29]和白几内亚山药(35.83%),低于二穗短柄草(46.40%)和水稻(43.57%)^[39]。芋基因组的 GC 含量处于中等水平,而过高的 GC 含量(>65%)和过低的 GC 含量(<25%)均可能会导致 Illumina 测序平台上的序列偏差,从而严重影响基因组的组装^[40]。根据芋三倍体和二倍体的测序结果,基因组序列杂合度为 0.52%~2.29%,重复序列含量为 73.97%~77.03%,芋基因组表现为具有较高重复度、较高杂合度和较大基因组的一个复杂基因组,这也将为芋全基因组的从头测序组装带来难度。

3.4 芋基因功能注释

基因预测共获得24 302个基因,其中有 99.43% 的基因可注释到 NR 等 6 个公共数据库中。KEGG 分析结果显示 125 条通路共检测到6 823条基因序列。较低比例的基因注释可能是由于芋公共序列资源的严重缺乏或存在非编码转录本所致。芋的许多特定功能基因,与非模式生物同源基因的相似性较低,增加了在公共数据库中寻找匹配基因的难度^[41]。在13 105 284个序列中,有 7.09% (928 987 个 SSR) 的序列存在重复位点,其中双核苷酸重复序列的丰度最高,大大丰富了芋的 SSR 标记数目。

在某作物基因组未知情况下,可采用二代测序技术进行初步测序,通过对测序数据分析得到基因组大小、杂合度、重复序列比例等信息,获得作物基因组整体概况的调研图。芋基因组调研图表达了芋的基因组特征,不仅为芋全基因组 *denovo* 测序和 SSR 标记的筛选提供了依据,还对芋及芋属甚至整个天南星科物种的基因组研究提供了丰富的分子数据,为天南星科作物分子水平研究的快速发展奠定了基础。

参考文献:

- [1] LEBOT V. Tropical root and tuber crops: cassava, sweet potato, yams and aroids[M]. Wallingford, UK: CABI, 2009.
- [2] RAO V R, MATTEWS P J, EYZAGUIRRE P B, et al. The global diversity of taro: ethnobotany and conservation[M]. Rome, Italy: Bioversity International, 2010.
- [3] 赵国华,陈宗道,王 赞. 芋头多糖的理化性质及体内免疫调节活性研究[J]. 中国食品学报,2002,2(3):21-25.
- [4] DAS A, DAS A B. Karyotype analysis of ten draught resistant cultivars of Indian taro - *Colocasia esculenta* cv. antiquorum Schott [J]. Nucleus, 2014, 57:113-120.
- [5] GRIMALDI I M, MUTHUKUMARAN S, TOZZI G, et al. Literary evidence for taro in the ancient Mediterranean: A chronology of names and uses in a multilingual world[J]. PLoS ONE, 2018, 13(6): e0198333.
- [6] 柯卫东,黄新芳,李建洪,等. 我国水生蔬菜科研与生产发展概况[J]. 长江蔬菜,2015(14):33-37.
- [7] KRON P, SUDA J, HUSBAND B C. Applications of flow cytometry to evolutionary and population biology[J]. Annual Review of Ecology Evolution & Systematics, 2007, 38:847-876.
- [8] RABINOWICZ P D, BENNETZEN J L. The maize genome as a model for efficient sequence analysis of large plant genomes[J]. Current Opinion in Plant Biology, 2006, 9:149-156.
- [9] DOLEZEL J, GREILHUBER J, SUDA J. Estimation of nuclear DNA content in plants using flow cytometry[J]. Nature Protocols, 2007, 2(9): 2233-2244.
- [10] MIYASHITA T, ARAKI H, HOSHINO Y. Ploidy distribution and DNA content variations of *Lonicera caerulea* (caprifoliaceae) in Japan[J]. Journal of Plant Research, 2011, 124(1): 1-9.
- [11] OBIDIEGWU J, RODRIGUEZ E, ENEOBONG E. Ploidy levels of *Dioscorea alata* L. germplasm determined by flow cytometry [J]. Genetic Resources and Crop Evolution, 2010, 57(3): 351-356.
- [12] 李秋实,徐 江,朱英杰,等. 基于流式细胞技术的灵芝基因组大小估测[J]. 菌物学报,2013,32(5):899-906.
- [13] 邓果特,刘清波,蒋建雄,等. 五节芒基因组大小测定[J]. 植物遗传资源学报,2013,14(2):339-341.
- [14] 李 祯,伊贤贵,顾 宇,等. 山樱花基因组大小的测定[J]. 南京林业大学学报(自然科学版),2014,38(S1):17-19.
- [15] 汪 艳,肖 媛,刘 伟,等. 流式细胞仪检测高等植物细胞核 DNA 含量的方法[J]. 植物科学学报,2015,33(1):126-131.
- [16] ZHANG T Z, HU Y, JIANG W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement[J]. Nature Biotechnology, 2015, 33(5): 531-537.
- [17] 唐 其,马小军,莫长明,等. 罗汉果全基因组 Survey 分析[J]. 广西植物,2015,35(6):786-791.
- [18] 唐玉娟,黄国弟,罗世杏,等. 芒果 2 个不同花芽分化时期转录组分析[J]. 南方农业学报, 2018,49(7):1257-1264.
- [19] CHEN S L, XU J, LIU C, et al. Genome sequence of the model medicinal mushroom *Ganoderma lucidum* [J]. Nature Communications, 2012, 3(2): 913-921.
- [20] XU H, SONG J, LUO H, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza* [J]. Molecular Plant, 2016, 9(6): 949-952.
- [21] WEI Z, HU Y Y, SUI Z H, et al. Genome survey sequencing and genetic background characterization of *gracilariopsis lemaneiformis* (Rhodophyta) based on next-generation sequencing [J]. PLoS ONE, 2013, 8:e69909.
- [22] 张小燕,刘志香,廖保生,等. 基于本草基因组学应用流式细胞术和高通量测序技术检测人参基因组大小[J]. 世界科学技术—中医药现代化,2017,19(10):1724-1728.
- [23] COATES D J, YEN D E, GAFFEY P M. Chromosome variation in taro, *Colocasia esculenta*: implications for origin in the Pacific[J]. Cytologia, 1988, 53:551-560.
- [24] 曹利民,龙春林. 中国芋属植物染色体数目及 5 个种的核型报道[J]. 云南植物研究,2004,26(3):310-316.
- [25] LI H, ZHU G H, BOYCE P C. Flora of China[M]. Louris: Missouri Botanical Garden Press, 2010:71-75.
- [26] WANG L, YIN J, ZHANG P, et al. *De novo* assembly and characterization of transcriptome and microsatellite marker development for taro [*Colocasia esculenta* (L.) Schott.][J]. International Journal of Genetics & Molecular Biology, 2017, 9:26-36.
- [27] TIAN X, ZHOU X, GONG N. Applications of flow cytometry in plant research-Analysis of nuclear DNA content and ploidy level in plant cells[J]. Chinese Agricultural Science Bulletin, 2011, 27:

- 21-27.
- [28] DOLEZEL J, BARTOS J, VOGLMAYR H, et al. Nuclear DNA content and genome size of trout and human[J]. Cytometry A, 2003, 51:127-128.
- [29] HAAS B J, SALZBERG S L, ZHU W, et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments[J]. Genome Biology, 2008, 9: 7.
- [30] BURGE C, KARLIN S. Prediction of complete gene structures in human genomic DNA[J]. Journal of Molecular Biology, 1997, 268: 78-94.
- [31] MAJOROS W H, PERTEA M, SALZBERG S L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders[J]. Bioinformatics, 2004, 20:2878-2879.
- [32] BLANCO E, PARRA G, GUIGO R. Using geneid to identify genes[M]. Current Protocols in Bioinformatics, Unit 4.3. New York: John Wiley & Sons Inc, 2007.
- [33] KORF I. Gene finding in novel genomes[J]. BMC Bioinformatics, 2004, 5:59.
- [34] JENS K, MICHAEL W, ERICKSON J L, et al. Using intron position conservation for homology-based gene prediction[J]. Nucleic Acids Research, 2016, 44:89-89.
- [35] LETUNIC I, COPLEY R R, SCHMIDT S, et al. SMART 4.0: towards genomic data integration[J]. Nucleic Acids Research, 2004, 32:142-144.
- [36] 黄新芳,柯卫东,刘义满,等. 芋种质资源染色体倍性鉴定[J]. 中国蔬菜,2012(16):42-46.
- [37] ISSHIKIS, NAKADAM, NAKAMURAN, et al. Genetic systems and variations of isozymes in wild taro (*Colocasia esculenta* Schott) from Bangladesh[J]. Journal of the Japanese Society for Horticultural Science, 1995, 64(1): 113-119.
- [38] HUANG Z H, XU J, XIAO S M, et al. Comparative optical genome analysis of two pangolin species: *Manis pentadactyla* and *Manis javanica*[J]. Giga science, 2016, 5(1): 1-5.
- [39] TAMIRU M, NATSUME S, TAKAGI H, et al. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination[J]. BMC Biology, 2017, 15:86.
- [40] AIRD D, ROSS M G, CHEN W S, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries[J]. Genome Biology, 2011, 12:18.
- [41] ELLEGREN H. Genome sequencing and population genomics in non-model organisms[J]. Trends in Ecology & Evolution, 2014, 29:51-63.

(责任编辑:张震林)