DOI: 10.3724/SP.J.1006.2021.04067

基于 SNP 和 InDel 标记的巴西木薯遗传多样性与群体遗传结构分析

孙倩^{1,2} 邹枚伶² 张辰笈^{2,4} 江思容^{2,5} Eder Jorge de Oliveira⁶
张圣奎⁷ 夏志强^{2,3,4,*} 王文泉^{2,3,4,5,*} 李有志^{1,*}

¹ 广西大学生命科学与技术学院 / 亚热带农业生物资源保护与利用国家重点实验室, 广西南宁 530004; ² 中国热带农业科学院热带生物技术研究所, 海南海口 571101; ³ 中国热带农业科学院热带生物组学大数据中心, 海南海口 571101; ⁴ 海南大学, 海南海口 570203; ⁵ 南京农业大学, 江苏南京 210095; ⁶ Embrapa Mandioca e Fruticultura, Cruz das Almas, Bahia 44380-000, Brazil; ⁷ 齐鲁工业大学, 山东济南 250306

摘要: 为了对巴西木薯种质资源进行遗传多样性、亲缘关系和群体遗传结构分析, 本研究利用了 7946 个 SNPs 和 1997 个 InDels 分子标记, 通过 ADMIXTURE 软件进行群体结构分析, GCTA 软件进行主成分分析。结果显示, 巴西木薯被划分为 9 个亚群。这与利用 PHYLIP 进行的聚类分析结果大概一致, 其中亚群 1、亚群 2、亚群 4、亚群 6 和亚群 8 能较好地分别聚在一起, 而其他亚群中的样品大致能聚在一起, 且样品间有一定的交叉。巴西木薯种质资源遗传多样性指数(0.274)高于中国、尼日利亚等, 其中巴西木薯亚群 5 具有相对较高的遗传多样性水平(0.29)。巴西木薯各亚群的群体遗传分化程度较低(群体分化指数在 0.03~0.15 之间), 但高于中国木薯种质资源的群体分化指数。各木薯材料间的遗传距离变幅为 0.084~0.297, 平均遗传距离为 0.228。本研究结果可为后续关联分析发掘优良等位基因及引种提供依据。

关键词: 木薯; SNP; InDel; 遗传多样性; 群体结构

Genetic diversity and population structure analysis by SNP and InDel markers of cassava in Brazil

SUN Qian^{1,2}, ZOU Mei-Ling², ZHANG Chen-Ji^{2,4}, JIANG Si-Rong^{2,5}, Eder Jorge de Oliveira⁶, ZHANG Sheng-Kui⁷, XIA Zhi-Qiang^{2,3,4,*}, WANG Wen-Quan^{2,3,4,5,*}, and LI You-Zhi^{1,*}

¹ College of Life Science and Technology / State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi University, Nanning 530004, Guangxi, China; ² Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, China; ³ Tropical Bio-omics and Big-Data Center, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, China; ⁴ Hainan University, Haikou 570203, Hainan, China; ⁵ Nanjing Agricultural University, Nanjing 210095, Jiangsu, China; ⁶ Embrapa Mandioca e Fruticultura, Cruz das Almas, Bahia 44380-000, Brazil; ⁷ Qilu University of Technology, Jinan 250306, Shandong, China

Abstract: As a typical tropical crop, cassava (*Manihot esculenta* Crantz) has the characteristics of drought resistance, barren resistance, high biomass and so on. In addition to being used for food and forage, it can also be used for production, processing and starch extraction. Due to highly heterozygous cassava genome, breeding is more difficult. Enriching the genetic diversity of cassava germplasm, comprehensively evaluating its genetic background and traits, and discovering superior alleles that control ex-

本研究由国家重点研发计划项目(2019YFD1001100), 国家自然科学基金-CG 联合基金项目(31861143005), 滇桂黔石漠化地区特色作物产业发展关键技术集成示范项目(SMH2019-2021)和中国热带农业科学院基本科研业务费专项(1630052019022)资助。

This study was supported by the National Key Research and Development Project (2019YFD1001100), the National Natural Science Foundation of China-CG Joint Fund (31861143005), the Key Technology Integration Demonstration Project of Characteristic Crop Industry Development in the Rocky Desertification Area of Yunnan, Guangxi and Guizhou (SMH2019-2021), and the Special Fund for Basic Scientific Research Operating Expenses of the Chinese Academy of Tropical Agricultural Sciences (1630052019022).

* 通信作者(Corresponding authors): 李有志, E-mail: dyzl@gxu.edu.cn; 王文泉, E-mail: wangwenquan@itbb.org.cn; 夏志强, E-mail: xiazhiqiang@itbb.org.cn

第一作者联系方式: E-mail: sunqian9595@163.com

Received (收稿日期): 2020-03-13; Accepted (接受日期): 2020-08-19; Published online (网络出版日期): 2020-09-15.

URL: <https://kns.cnki.net/kcms/detail/11.1809.S.20200915.0855.002.html>

cellent traits are of great significance for cassava breeding in the future. In order to analyze the genetic diversity, genetic relationship and population structure of cassava germplasm in Brazil, 7946 SNPs and 1997 InDels molecular markers were used. Population structure analysis was performed by ADMIXTURE software, and principal component analysis was performed by GCTA software. Brazilian cassava was divided into nine subgroups, and was roughly consistent with the results of cluster analysis using PHYLIP. Among them, subgroup 1, subgroup 2, subgroup 4, subgroup 6, and subgroup 8 could be clustered together respectively, while the samples of other subgroups could be roughly clustered, and there was a certain cross between the samples. The genetic diversity of cassava germplasm in Brazil (0.274) was higher than the genetic diversity level of cassava germplasm in China and Nigeria. Subgroup 5 of Brazil cassava had a relatively high genetic diversity (0.29). The genetic differentiation of subgroups was low (the genetic differentiation vary from 0.03 to 0.15), but higher than domestic cassava germplasm. The genetic distance between cassava accessions varied from 0.084 to 0.297, with the average of 0.228. The results of this study can provide a basis for subsequent association analysis to identify great alleles and introduction.

Keywords: cassava; SNP; InDel; genetic diversity; population structure

木薯 (*Manihot esculenta* Crantz) 是大戟科 (Euphorbiaceae) 木薯属 (*Manihot* P. Mill.) 的多年生灌木植物, 具有高生物量、抗贫瘠、抗病虫能力强等特点, 被广泛种植于亚、非、美三洲等多个国家或地区^[1]。木薯起源于亚马逊河流域, 于 19 世纪 20 年代传入中国, 最初种植于广东省, 之后逐渐在海南、广西、贵州、云南等地大量种植。它是世界三大薯类作物之一, 同时也是世界上第六大粮食作物, 仅次于小麦、水稻、玉米、马铃薯和大麦^[2]。其用途广泛, 除可食用、饲用外, 还可用于生产加工, 如造纸、纺织、生物燃料等^[3]。木薯块根还可用于提取淀粉, 加工成薯条、面包, 以及生产燃料乙醇等; 茎秆可用来进行木薯繁殖、粉碎还田、做燃料等; 叶片可作蔬菜食用或喂鱼、养蚕等^[4]。

木薯基因组具有高度杂合的特性, 原因是其异花授粉, 且长期进行无性繁殖。由于基因组的高度杂合, 从而增加了木薯遗传变异的多样性, 这些多样性可为木薯育种人员提供更多可选择的良好亲本, 但同时由于木薯的基因组高度杂合、亲缘关系不清晰、遗传改良周期长等特点也增加了育种的工作难度^[5]。目前已有一些利用相关序列扩增多态性 (sequence-related amplified polymorphism, SRAP)、简单重复序列 (simple sequence repeat, SSR)、扩增片段长度多态性 (amplified fragment length polymorphism, AFLP)、单核苷酸多态性 (single nucleotide polymorphisms, SNP) 等分子标记进行木薯遗传多样性的研究。Fregene 等^[6]利用 SSR 标记对来源于哥伦比亚、巴西和秘鲁等地的木薯地方品种的种质资源多样性评价发现, 不同国家来源的木薯种质的遗传多样性水平都很高, 其中来自巴西和哥伦比亚材料的基因多样性水平最高。Alex 等^[7]利用 13 对 SSR

标记对巴西多地的传统甜木薯品种的群体结构和遗传多样性评估结果显示, 该群体的遗传多样性平均值为 0.5407, 范围为 0.3138 (GA21)~0.6502 (GA140), 表明该群体的遗传变异性宽泛。Carvalho 等^[8]采用 SSR 标记和 RAPD 标记的研究中发现, 巴西的木薯种质资源的地理来源和遗传聚类具有显著正相关关系。

丰富木薯种质资源的遗传多样性, 并对其遗传背景和性状进行综合评价, 发掘控制优良性状的优异等位基因, 对今后木薯育种具有重大意义。全基因组关联分析 (genome-wide association study, GWAS) 能够鉴定目的表型性状与遗传标记或基因间的关系, 并检测出控制相关性状的优良等位基因位点^[9]。而进行关联分析需要先评估实验群体的遗传多样性、遗传结构及亲缘关系^[10]。但目前利用 SNP 和 InDel 标记对木薯进行遗传多样性、亲缘关系及群体结构分析等的相关研究还鲜为报道。

本研究拟利用 SNP 和 InDel 分子标记, 对由巴西 Embrapa 机构提供的来源于巴西多地的 192 份木薯种质资源进行遗传多样性和群体结构分析。本研究将为以后木薯育种亲本选配提供材料和理论指导, 也可为进一步通过关联分析发掘控制木薯种质中优良性状的优异等位基因提供理论依据, 从而促进利用分子标记辅助选择技术培育木薯新品种。

1 材料与方法

1.1 试验材料

供试木薯材料共 192 份, 均为巴西栽培种木薯 (表 1)。

表 1 192 份木薯栽培种
Table 1 List of 192 cassava cultivars

编号 No.	名称 Name	编号 No.	名称 Name	编号 No.	名称 Name	编号 No.	名称 Name	编号 No.	名称 Name
001	9624-09	033	BGM0400	065	BGM0783	097	BGM1291	129	BGM1608
002	98150-06	034	BGM0405	066	BGM0788	098	BGM1313	130	BGM1615
003	BGM0020	035	BGM0408	067	BGM0807	099	BGM1324	131	BGM1622
004	BGM0042	036	BGM0425	068	BGM0886	100	BGM1327	132	BGM1626
005	BGM0045	037	BGM0428	069	BGM0889	101	BGM1328	133	BGM1640
006	BGM0073	038	BGM0436	070	BGM0896	102	BGM1364	134	BGM1662
007	BGM0087	039	BGM0443	071	BGM0905	103	BGM1370	135	BGM1667
008	BGM0103	040	BGM0451	072	BGM0917	104	BGM1376	136	BGM1668
009	BGM0135	041	BGM0452	073	BGM0930	105	BGM1387	137	BGM1672
010	BGM0145	042	BGM0467	074	BGM0972	106	BGM1396	138	BGM1679
011	BGM0152	043	BGM0495	075	BGM0976	107	BGM1409	139	BGM1682
012	BGM0163	044	BGM0501	076	BGM0982	108	BGM1412	140	BGM1684
013	BGM0179	045	BGM0507	077	BGM0989	109	BGM1429	141	BGM1690
014	BGM0188	046	BGM0523	078	BGM0993	110	BGM1437	142	BGM1697
015	BGM0206	047	BGM0540	079	BGM1042	111	BGM1440	143	BGM1698
016	BGM0209	048	BGM0542	080	BGM1050	112	BGM1458	144	BGM1704
017	BGM0211	049	BGM0543	081	BGM1081	113	BGM1485	145	BGM1706
018	BGM0213	050	BGM0544	082	BGM1106	114	BGM1487	146	BGM1715
019	BGM0226	051	BGM0550	083	BGM1110	115	BGM1490	147	BGM1722
020	BGM0250	052	BGM0551	084	BGM1127	116	BGM1495	148	BGM1728
021	BGM0261	053	BGM0562	085	BGM1153	117	BGM1510	149	BGM1732
022	BGM0264	054	BGM0564	086	BGM1164	118	BGM1524	150	BGM1750
023	BGM0269	055	BGM0574	087	BGM1165	119	BGM1535	151	BGM1757
024	BGM0283	056	BGM0600	088	BGM1171	120	BGM1537	152	BGM1760
025	BGM0288	057	BGM0620	089	BGM1174	121	BGM1539	153	BGM1761
026	BGM0331	058	BGM0624	090	BGM1178	122	BGM1549	154	BGM1763
027	BGM0336	059	BGM0640	091	BGM1180	123	BGM1552	155	BGM1794
028	BGM0338	060	BGM0654	092	BGM1193	124	BGM1576	156	BGM1814
029	BGM0352	061	BGM0670	093	BGM1202	125	BGM1581	157	BGM1835
030	BGM0361	062	BGM0678	094	BGM1226	126	BGM1590	158	BGM1850
031	BGM0390	063	BGM0726	095	BGM1252	127	BGM1593	159	BGM1883
032	BGM0394	064	BGM0752	096	BGM1281	128	BGM1596	160	BGM1884
								161	BGM1942
								162	BGM1957
								163	BGM2022
								164	BGM2028
								165	BGM2041
								166	BGM2047
								167	BGM2052
								168	BGM2061
								169	BGM2062
								170	BGM2063
								171	BGM2071
								172	BGM2078
								173	BGM2081
								174	BGM2082
								175	BGM2083
								176	BRS Amansa Burro
								177	BRS Caipira
								178	BRS Dourada
								179	BRS Formosa
								180	BRS Gema Ovo
								181	BRS Jari
								182	BRS Kiriris
								183	BRS Mulatinha
								184	BRS Tapioqueira
								185	BRS Verdinha
								186	Cascuda
								187	Corrente
								188	Eucalipto
								189	Fécula Branca
								190	IAC90
								191	Olho Junto
								192	Valencia

1.2 木薯基因组 DNA 提取及建库测序

采用改良的 CTAB 法^[11]提取木薯叶片的基因组 DNA, 经 1%的琼脂糖凝胶电泳检测及浓度测定后, 将工作液浓度稀释到 100 ng μL^{-1} , -20°C 保存。然后利用 AFSM^[12]技术对 192 份木薯 DNA 样品使用 96 孔 PCR 板分别构建 *EcoR* I-*Msp* I 和 *EcoR* I-*Hpa* II 文库, 单克隆检测达到要求后, 将相应的 *EcoR* I-*Msp* I 和 *EcoR* I-*Hpa* II 文库按 1 : 1 比例混合成 1 个文库, 总共 192 个样本, 构建了 2 个 AFSM 文库, 并利用 Hiseq 2500 对构建好的测序文库进行双端 150

bp 测序。

1.3 SNP 和 InDels 分子标记检测

利用 Perl 脚本(<http://afsmseq.sourceforge.net/>)对原始测序数据进行过滤, 同时统计测序得到的总 reads 数, 再将 reads 根据 AFSM 技术设计的 Barcodes 分配到每个个体中, 并统计每个个体的 reads 数^[12]。使用 Bowtie 2 软件^[13]将优化后的测序 reads 比对到木薯 AM560 参考基因组^[14], 再使用 SAMtools^[15]和 VCFtools (<http://vcftools.sourceforge.net/>)检测 SNP 和 InDels 位点。基于木薯 AM560 参考基因组 V6.1,

利用 snpEff 软件^[16]识别其变异位置(间隔区、非翻译区/UTR、基因上游区或基因下游区)、变异类型(同义突变和错义突变、移码突变和非移码突变), 同时对其进行注释。

1.4 巴西木薯群体结构、遗传多样性及群体分化分析

先利用 PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>) 计算样品的遗传距离矩阵, 然后用 Notepad++ 软件将遗传距离矩阵的文件调整成合适的格式, 采用邻接法构建系统进化树结构, 生成 tree 文件后, 再使用 iTOL (<https://itol.embl.de/>) 绘制进化树图。通过 GCTA 软件^[17]利用检测出的 SNP 对参试木薯群体材料进行主成分分析(PCA)。再使用 R 软件计算各个主成分的向量, 绘制 PCA 散点图。

此外, 使用 ADMIXTURE 软件^[18]进行群体结构分析, 估算出最佳群体亚群数。先用 PLINK 软件^[19]调整 ADMIXTURE 软件的输入文件格式, 并输入文件, 然后将亚群数 K 值范围设置为 1~12, 根据得到的 cross-validation error 值选择合适的亚群数 K 值, 利用各个材料占各亚群的遗传成分系数(Q)构成群体遗传结构矩阵。

利用 VCFtools 软件(<https://vcftools.github.io/index.html>) 计算群体遗传多样性指数(π)和群体分化指数(F_{st})^[20]。根据 Wright 的研究, 当群体分化指数

(F_{st}) 等于 0 或 1 时, 分别表明亚群间没有分化, 或亚群间完全分化。而当 $0 < F_{st} < 0.05$ 、 $0.05 < F_{st} < 0.15$ 、 $0.15 < F_{st} < 0.25$ 、或 $0.25 < F_{st} < 1$ 时, 则分别表明亚群间具有较弱、中等、比较强或非常强的遗传分化^[21]。

2 结果与分析

2.1 巴西木薯群体基因型分析

通过对 192 份巴西栽培种木薯基因组 DNA 进行 AFSM 建库及测序, 总共得到了 155 G 数据, 过滤后得到 134 G 数据, 893,020,018 条 reads。再利用木薯参考基因组 AM560 V6.1, 通过 SAMtools 和 VCFtools 软件对 192 份木薯样品基因组进行扫描, 得到 796,006 个 SNPs 和 116,821 个 InDels。

通过哈迪温伯格检测($HWE > 0.001$)、次等位基因频率(MAF) > 0.05 过滤, 并舍去低质量的变异位点后, 仅保留了 9443 个高质量的变异位点(7946 个 SNPs 和 1997 个 InDels)用于后续分析。其中, 3287 个 SNPs 和 InDels 位于基因间隔区, 4005 个 SNPs 和 InDels 位于基因上游区, 471 个 SNPs 和 InDels 位于基因下游区, 2 个 SNPs 和 InDels 位于 5'端 UTR。845 个 SNPs 和 InDels 属于错义突变, 745 个 SNPs 和 InDels 属于同义突变, 417 个 SNPs 和 InDels 属于移码突变, 另有 171 个 SNPs 和 InDels 属于其他类型突变(表 2)。

表 2 SNPs 和 InDels 的统计

Table 2 Summary of SNPs and InDels

总数 Total	间隔区 Intergenic region	基因上游区 Upstream gene region	基因下游区 Downstream gene region	5'端 UTR 5' prime UTR variant	错义突变 Missense variant	同义突变 Synonymous variant	移码突变 Frame shift variant	其他类型 Other types
9443	3287	4005	471	2	845	745	417	171

2.2 巴西木薯群体结构分析

通过 ADMIXTURE 软件利用 9943 个高质量的 SNPs 和 InDels 分子标记对 192 份巴西栽培种木薯进行群体遗传结构分析。将亚群数 K 值范围设置为 1~12, 计算不同 K 值下的交叉验证错误率(cross-validation error, CV error)。当 K 从 1 到 2 时, CV error 值迅速减小; K 从 2 到 4 时, CV error 值又逐渐增加; 当 K 从 4 到 9 时, CV error 值逐渐减小并趋于平缓; 当 K 大于 9 时, CV error 值又出现一定的增幅(图 1-a)。说明在 K 等于 9 时, CV error 值最小, 因此巴西栽培种木薯群体可分为 9 个亚群(Subgroup

1~Subgroup 9)。

192 份巴西木薯可以被分为 9 个亚群, 再根据每个个体在这 9 个亚群的 Q 值, 将每个个体归类到 Q 值最大所在的亚群(图 1-b)。9 个亚群中分别含有 3 份、22 份、27 份、6 份、20 份、25 份、24 份、12 份和 53 份材料。

主成分分析以所有的高质量 SNPs 和 InDels 为基础, 通过 R 软件分析绘图, 得到如下结果: 该木薯群体的 9 个亚群在 PC1 轴上可以看出一定的分布差距, 大部分亚群可以聚类在一起, 该结果说明聚类结果与群体结构的划分具有一致性(图 1-c)。

图 1 192 份木薯品种的群体遗传结构

Fig. 1 Population genetic structure of 192 cassava cultivars

a: 利用 ADMIXTURE 软件对 192 份木薯材料进行群体结构分析, 计算 K 为 1~12 时的 CV error。b: $K=9$ 时的群体结构, 在该群体结构中, 每个个体用一根带有 9 种不同颜色的线表示, 根据颜色的占比推断该品种属于哪个亚群, 红、橙、黄、绿、蓝绿、湖蓝、深蓝、紫、紫红色分别代表亚群 1~9。c: 利用高质量变异位点对 192 份巴西木薯进行主成分分析, 每一个点代表一份样品, 橙色、桔色、黄色、草绿、深绿、天蓝、深蓝、紫和紫红色分别代表亚群 1~9 (根据 ADMIXTURE 软件推断的结果)。

a: ADMIXTURE was used to analyze the population structure of 192 cassava materials, and the cross-validation error was calculated when K was 1~12. b: The population structure at $K=9$, in this population structure, each individual is represented by a line with multiple different colors, and the subgroup to which the individual belongs is inferred based on the proportion of the color, the nine color regions arranged from left to right in order represent subgroups 1 to 9, respectively. c: Principal component analysis (PCA) of the group, one point represents an individual, nine colors and their corresponding shapes represent different subgroups.

由图 2 可知, 聚类结果与群体结构的划分相一致, 亚群 1、亚群 2、亚群 4、亚群 6 和亚群 8 能较好地分别聚在一起, 而其他亚群样品大致能聚在一起, 且样品间有一定的交叉。巴西木薯各栽培种之间并未聚类到一起, 可能是由于木薯栽培历史比较短, 来源于巴西多地的木薯栽培种还未产生明显的分化。

2.3 巴西栽培种木薯遗传多样性分析

利用 9943 个高质量的 SNPs 和 InDels, 通过计算遗传多样性指数(π), 评估巴西栽培种木薯群体和各个亚群的遗传多样性。通过 vcftools 计算发现, 巴西栽培种木薯群体的遗传多样性指数为 0.274, 亚群 1~9 的遗传多样性指数在 0.192~0.289 之间, 其中亚群 1 具有最低的遗传多样性指数(0.192), 亚群 7、亚群 2、亚群 6 和亚群 3 具有相对较高的遗传多样性指数, 分别达到 0.284、0.281、0.264 和 0.261, 而亚

群 5 具有 9 个亚群中最高遗传多样性指数(0.289)

图 2 192 份巴西栽培种木薯系统进化树图

Fig. 2 Phylogenetic tree of 192 cassava cultivars from Brazil

淡蓝色、深蓝色、草绿色、深绿色、肉色、橙色、紫红色、红色和紫色分别代表亚群 1~9。

Light blue, dark blue, grass green, dark green, flesh, orange, fuchsia, red, and purple represent subgroup 1 to 9, respectively.

(表 3)。说明巴西栽培种木薯群体具有相对较高的遗传多样性水平。

利用群体分化指数(F_{st})评估巴西栽培木薯亚群间的差异程度(表 4)发现, 除亚群 1 和亚群 4 之间有较强遗传分化外, 其他亚群之间均为中等或较弱遗传分化, 亚群间的遗传分化指数在 0.031~0.152 之间。其中, 亚群 8 与其他各亚群之间均为中等分化; 而除亚群 4 与亚群 1 外, 亚群 4 和亚群 1 分别与其他各亚群之间也均为中等分化。表明, 除亚群 1 与亚群 4 间遗传分化较强、亲缘关系较远外, 其余各亚群间的为中等或较弱遗传分化程度, 即亚群间的亲缘关系相对均较近。

另外, 本研究对试验所用的巴西栽培木薯的遗传距离分析发现, 这些木薯种质间的遗传距离为 0.084~0.297, 平均遗传距离为 0.228。其中, BGM 1883 与 Valencia 遗传距离最近(0.084); BGM0640 与

BRSJari 遗传距离最远(0.297)。

表 3 遗传多样性指数(π)的统计
Table 3 Statistics of genetic diversity index (π)

亚群 Subgroup	π
1	0.192
2	0.281
3	0.261
4	0.221
5	0.289
6	0.264
7	0.284
8	0.209
9	0.234
整体 Entirety	0.274
平均 Average	0.248

表 4 群体分化指数(F_{st})的统计
Table 4 Statistics of population differentiation index (F_{st})

亚群 Subgroup	1	2	3	4	5	6	7	8	9
1									
2	0.065								
3	0.081	0.061							
4	0.152	0.073	0.112						
5	0.053	0.032	0.069	0.062					
6	0.058	0.045	0.033	0.086	0.041				
7	0.059	0.038	0.054	0.078	0.031	0.045			
8	0.139	0.085	0.076	0.119	0.084	0.056	0.080		
9	0.058	0.043	0.046	0.083	0.048	0.034	0.038	0.061	

3 讨论

3.1 木薯群体结构分析

研究表明, 基因型与性状之间会产生假关联, 其原因可能是群体结构分层、等位基因分布不均等^[22]。为了消除造成关联分析不准确的因素, 我们需要先对试验群体进行群体结构分析。本研究利用 ADMIXTURE 软件对巴西木薯自然群体的群体结构分析表明, 当 $K=9$ 时, CV error 值最小, 由此将 192 份木薯种质划分为 9 个亚群, 该结果与聚类分析、主成分分析的结果大概相符, 它们之间相互补充及印证, 说明该木薯群体的遗传结构较为可靠。在这 9 个亚群中, 群体分化指数在 0.03~0.15 之间, 且大部分亚群间的群体分化指数均小于 0.09, 表明该木薯群体存在一个中等偏弱的遗传分化。前人的研究结

果显示, 中国热带农业科学院热带品种资源种质圃收集的 158 份木薯种质的群体分化指数在 0.03~0.07 之间^[23]; 在其他地区的栽培木薯中, 群体分化指数在 0.01~0.05 之间^[14], 表明国内收集的木薯遗传分化程度较低。比较看来, 本研究中的巴西木薯种质的群体分化指数高于国内收集的木薯种质, 可挑选优质巴西木薯品种并引进中国, 从而丰富已有的木薯种质资源。

3.2 木薯遗传多样性分析

本研究对遗传多样性指数(π)进行了计算, 从而评估该木薯群体的遗传多样性。Ramu 等^[14]的研究表明, 来源于不同地区(尼日利亚、哥伦比亚、巴西等地)的国外栽培木薯的遗传多样性指数为 0.0036, 低于其祖先 (*M. esculenta* ssp. *flabellifolia*, $\pi = 0.0051$); Fregene 等^[6]对来源于哥伦比亚、巴西和秘

鲁等地的木薯地方品种的种质资源多样性评价发现,巴西和哥伦比亚的木薯种质具有最高的遗传多样性水平;在张圣奎对中国热带农业科学院热带品种资源种质圃收集的 158 份木薯种质的遗传多样性研究中发现,该群体的遗传多样性指数为 1.21×10^{-4} ,表明该群体的遗传多样性较低^[23],同时也表明目前国内的木薯种质资源丰富度较为缺乏。本研究发现,该巴西木薯群体中各亚群的遗传多样性指数在 0.19~0.29 之间,平均遗传多样性指数为 0.248,说明巴西木薯群体的遗传多样性较为丰富,可引进部分优良巴西种质以丰富国内的木薯种质资源。除此之外,样本之间的亲缘关系也会对关联分析的结果造成一定的影响。本研究对 192 份木薯种质间的遗传距离进行分析,从而评估不同材料之间的亲缘关系,结果发现这些木薯种质的平均遗传距离为 0.228。

4 结论

本研究利用 9943 个高质量的 SNPs 和 InDels 对 192 份巴西 Embrapa 机构提供的木薯种质进行了和群体遗传结构分析。遗传多样性分析结果显示,巴西木薯群体的遗传多样性水平较为丰富,高于中国和哥伦比亚等地区;群体遗传结构分析结果显示,该群体被划分为 9 个亚群,此结果与主成分分析及聚类分析结果基本一致。另外,该木薯群体的分化程度较低,但高于国内的木薯种质资源。遗传距离分析显示, BGM1883 与 Valencia 遗传距离最近, BGM0640 与 BRSJari 遗传距离最远。该研究将为之后关联分析挖掘优良基因及引进优良巴西木薯种质提供依据。

References

- [1] El-Sharkawy M A. Cassava biology and physiology. *Plant Mol Biol*, 2004, 56: 481–501.
- [2] Zidenga T, Leyva-Guerrero E, Moon H, Siritunga D, Sayre R. Extending cassava root shelf life via reduction of reactive oxygen species production. *Plant Physiol*, 2012, 159: 1396–1407.
- [3] Uchchukwu-Agua A D, Caleb O J, Opara U L. Postharvest handling and storage of fresh cassava root and products: a review. *Food Bioproc Technol*, 2015, 8: 729–748.
- [4] 严华兵, 叶建秋, 李开绵. 中国木薯育种研究进展. *中国农学通报*, 2015, 31(15): 63–70.
Yan H B, Ye J Q, Li K M. Progress of cassava breeding in China. *Chin Agric Sci Bull*, 2015, 31(15): 63–70 (in Chinese with English abstract).
- [5] 李月仙, 刘倩, 严伟, 宋记明, 段春芳, 沈绍斌, 张林辉, 姜予强, 刘光华. 木薯种质资源在云南的形态多样性及其形态标记聚类分析. *华中农业大学学报*, 2018, 37(3): 10–18.
Li Y X, Liu Q, Yan W, Song J M, Duan C F, Shen S B, Zhang L H, Lou Y Q, Liu G H. Morphological diversity and clustering analysis of cassava germplasm resources in Yunnan province. *J Huazhong Agric Univ*, 2018, 37(3): 10–18 (in Chinese with English abstract).
- [6] Fregene M A, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon A G O, Dean R, Kresovich S. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Theor Appl Genet*, 2003, 107: 1083–1093.
- [7] Ortiz A H T, Rocha V P C, Moiana L D, Gonçalves-Vidigal M C, Galvão M Z, Filho P S V. Population structure and genetic diversity in sweet cassava cultivars from Paraná, Brazil. *Plant Mol Biol Rep*, 2016, 34: 1153–1166.
- [8] Carvalho L J C B, Schaal B A. Assessing genetic diversity in the cassava (*Manihot esculenta* Crantz) germplasm collection in Brazil using PCR-based markers. *Euphytica*, 2001, 120: 133–142.
- [9] Pradhan S K, Barik S R, Sahoo A, Mohapatra S, Nayak D K, Mahender A, Meher J, Anandan A, Pandit E. Population structure, genetic diversity and molecular marker trait association analysis for high temperature stress tolerance in rice. *PLoS One*, 2016, 11: e0160027.
- [10] Dhanapala A P, Rayb J D, Singhe S K, Hoyos-Villegas V, Smith J R, Purcelle L C, King C A, Fritsch F B. Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. *Plant Genome*, 2015, 8: 1–15.
- [11] Murray M G, Thompson W F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*, 1980, 8: 4321–4326.
- [12] Xia Z Q, Zou M L, Zhang S K, Feng B X, Wang W Q. AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping. *Sci Rep*, 2014, 4: 7300.
- [13] Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 2012, 9: 357–359.
- [14] Ramu P, Esuma W, Kawuki R, Rabbi I Y, Egesi C, Bredeson J V, Bart R S, Verma J, Buckler E S, Lu F. Cassava hapmap: managing genetic load in a clonal crop species. *bioRxiv*, 2016. doi: <https://doi.org/10.1101/077123>.
- [15] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25: 2078–2079.
- [16] Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L, Land S J, Lu X Y, Ruden D M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 2012, 6: 80–92.
- [17] Yang J, Lee S H, Goddard M E, Visscher P M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*, 2011, 88: 76–82.
- [18] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 2009, 19: 1655–1664.
- [19] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M A R, Bender D, Maller J, Sklar P, Bakker P I W D, Daly M J, Sham P C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 2007, 81:

- 559–575.
- [20] Danecek P, Auton A, Abecasis G, Albers C A, Banks E, DePristo M A, Handsaker R E, Lunter G, Marth G T, Sherry S T, McVean G, Durbin R. The variant call format and VCFtools. *Bioinformatics*, 2011, 27: 2156–2158.
- [21] Wright S. *Evolution and the Genetics of Population 4: Variability Within and Among Natural Populations*. Chicago & London: the University of Chicago Press, 1978. p 157.
- [22] Flint-Garcia S A, Thornsberry J M, Buckler E S. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol*, 2003, 54: 357–374.
- [23] 张圣奎. 木薯种质资源综合评价及主要农艺性状的全基因组关联分析. 华中农业大学博士学位论文, 湖北武汉, 2018.
- Zhang S K. *Phenotyping and Genome-wide Association Studies of Important Agronomic Traits in Cassava (*Manihot esculenta* Cranz)*. PhD Dissertation of Huazhong Agricultural University, Wuhan, Hubei, China, 2018 (in Chinese with English abstract).