

基于SLAF-seq技术的木薯SNP标记开发

俞奔驰, 韦丽君*, 雷开文, 宋恩亮, 马崇熙

广西壮族自治区亚热带作物研究所, 广西南宁530001

摘要: 开发大量木薯(*Manihot esculenta*)特异分子标记可促进木薯种质资源中优良基因的挖掘和利用。本研究以39份木薯种质资源为材料, 利用特异性位点扩增片段测序(specific-locus amplified fragment sequencing, SLAF-seq)技术进行测序, 结果显示: 本次实验双端比对效率为90.83%, 酶切效率为91.18%; 通过测序获得304.30 Mb的读长数据, 各样本读长范围在2 123 094~15 102 046之间, 对照数据量最小, 仅395 790个读长; 样品测序质量值(Q_{30})在95.62%~96.60%之间, 均在95%以上; GC含量在41.27%~44.96%之间, 平均GC含量为42.89%, 对照的GC含量为42.93%。本研究共开发出725 220个SLAF标签, 样本的平均测序深度为20.98倍(\times), 多态性SLAF标签446 789个, 共得到2 504 553个群体SNP标记。可见, SLAF-seq技术适用于木薯SNP位点的开发, 其效率较高。

关键词: 木薯; 分子标记; SLAF-seq; SNP

木薯(*Manihot esculenta*)是全球重要的薯类作物, 既是热带地区的粮食安全保障, 也是重要的淀粉工业原材料, 保障其稳产、高产、优质一直是育种家不变的研究主题(张鹏等2013)。木薯遗传背景复杂, 基因高度杂合, 后代性状分离严重, 并且具有自交不亲和、种子数量少等特性, 给育种工作带来了极大的麻烦。木薯大多数农艺性状, 如产量、品质和一些与抗性相关的性状都表现为数量性状遗传, 而且它们之间还常常存在一些负相关性, 因此, 分子标记辅助选择对加快木薯育种进程意义重大。

木薯可以在世界上多个国家和地区种植, 具有丰富的性状遗传多样性与地理适应性, 充分利用挖掘这些遗传信息可以有效地提高木薯育种质量。Fregene等(1997)用TMS30572与CM2177-2杂交产生的F1代株系为材料, 用132个限制性片段长度多态性(restriction fragment length polymorphism, RFLP)标记、30个随机扩增多态性DNA (random amplification of polymorphic DNA, RAPD)标记、3个简单重复序列(simple sequence repeat, SSR)标记和3个同工酶标记构建了木薯遗传连锁图谱。近20年来, 通过木薯科研工作者的长期努力, 一系列木薯基因组特有的SSR标记、表达序列标签-简单重复序列(expressed sequence tag-simple sequence repeat, EST-SSR)标记也相继被报道。Mba等(2001)基于SSR技术开发获得了172个木薯SSR标记, 其中的36个标记被定位到已有的RFLP连锁图谱上。

李杰等(2006)用SSR标记和集群分离分析法(bulked segregant analysis, BSA)构建基因池分析木薯的遗传多样性, 筛选到19对SSR引物在各个性状的相对基因池之间具有多态性扩增产物。Sraphet等(2011)用SSR和EST-SSR引物对木薯分离群体进行遗传分析并构建了含SSR/EST-SSR位点的遗传连锁图。彭静茹等(2012a, b)同时用内部简单重复序列(inter-simple sequence repeat, ISSR)和SSR技术对39份木薯种质资源进行遗传多态性研究。

随着高通量测序技术的快速发展, 其简便快捷、低成本的特点在遗传研究中具有了重要的地位, 大规模基因分型结合高通量测序技术为基因挖掘提供了有力的技术保障(Ogura和Busch 2015)。Rabbi等(2014)通过测序基因分型(genotyping by sequencing, GBS)技术获得了第一张木薯高密度单核苷酸多态性(single-nucleotide polymorphism, SNP)遗传图谱, 其中连锁的SNP标记数量达到6 756个, 并且将木薯普通花叶病毒(*Cassava common mosaic virus*)抗性基因CMD2精细定位到木薯基因组的一个支架(scaffold)上。Pootakham等(2014)对木薯进行转录组测序和基因型分析, 构建了一张包含2 110个EST-SNP的遗传连锁图, 并在连锁群LG7和LG10定位了两个木薯淀粉糊化特性的数量性状

收稿 2018-03-20 修定 2018-05-15

资助 广西科技计划项目(桂科AB16380080和桂科AA16380002)和广西自然科学基金(2017GXNSFBA198172)。

* 通讯作者(weilijun2002@126.com)。

基因座(quantitative trait locus, QTL)。Xia等(2014)用扩增片段单核苷酸多态性与甲基化(amplified-fragment single nucleotide polymorphism and methylation, AFSM)技术构建了一张含有3 032个AFSM标记的木薯遗传连锁图。可见, 通过高通量测序技术的EST-SSR和SNP等重要分子标记的开发和利用, 木薯传统育种速度和筛选效率提高。

特异性位点扩增片段测序(specific-locus amplified fragment sequencing, SLAF-seq)即是由第二代高通量测序技术发展而来的。它以目标物种的生物信息学分析为基础设计标记开发方案, 构建SLAF-seq文库, 筛选特异性长度片段进行高通量测序, 获得海量序列, 再通过软件分析比对, 获得多态性SLAF标签并在多态性SLAF标签上开发大量特异性SNP位点, 利用群体内具有代表性的高质量SNP进行群体多态性分析。SLAF-seq技术具有通量高、准确性高、可重复性好、成本低和周期短的特点(Sun等2013), 可应用于全基因组关联分析、遗传图谱构建和QTL定位、种质资源鉴定、基因表达调控、系统进化分析等分子生物学研究中(Zhang等2013)。目前, 基于SLAF-seq技术的SNP位点开发已在甘薯(*Ipomoea batatas*) (苏文瑾等2016)、甘蓝型油菜(*Brassica napus*) (王伟等2016)和火龙果(*Hylocereus undulatus*) (俞超等2017)等植物上有了报道, 这些研究结果证实了SLAF-seq技术的效率远高于SSR、扩增片段长度多态性(amplified fragment length polymorphism, AFLP)和RAPD等分子标记技术。目前, 通过SLAF-seq技术在木薯上开发大量SNP标记尚未见报道。本研究将SLAF-seq技术应用于木薯上, 通过简化基因组测序的方式获取木薯大量多态性SLAF标签, 进而开发一批特异性强、稳定性高的SNP标记, 为木薯特异SNP标记的开发、资源鉴定分析、重要农艺性状的关联分析, 以及进一步加快木薯良种选育步伐提供参考。

1 材料与方法

1.1 试验材料及其基因组DNA提取

利用广西木薯研究所木薯育种课题组木薯种质资源圃中的木薯(*Manihot esculenta* Crantz) ‘新选048’ (XX048)、‘南植199’ (NJ199)、‘桂热3号’

(GR3)、‘桂热891’ (GR891)、‘桂热911’ (GR911)、‘华南5号’ (SC5)、‘华南8号’ (SC8)、‘华南124’ (SC124)、‘华南205’ (SC205)及30份木薯新品系为材料, 采集上述材料的健康、成熟叶片并迅速置于液氮中冷冻保存。采用十六烷基三甲基溴化铵(cetyltrimethylammonium ammonium bromide, CTAB)法提取材料的DNA, 用1%琼脂糖电泳检测DNA的完整性。用NanoDrop分光光度计检测DNA的浓度。按照高通量测序的要求, 每个样本的DNA总量 $\geq 1.5 \mu\text{g}$, 浓度 $\geq 20 \text{ ng} \cdot \mu\text{L}^{-1}$; 样品纯度($\text{OD}_{260/280}$)为1.8~2.2。

1.2 SLAF文库构建

选择已测序完成的木薯基因组作参考基因组(下载地址: <https://www.ncbi.nlm.nih.gov/assembly/GCF-001659605.1>), 组装出的基因组大小为582 Mb, GC含量为34.50%。根据木薯基因组大小以及GC含量等信息作为参考进行电子酶切预测, 得到两个限制性内切酶HaeIII和Hpy166II, 利用上述两种酶对检测合格的各样品基因组DNA分别进行酶切。对酶切片段(SLAF标签)进行3'端加上单核苷酸A处理, 连接双接头(Kozich等2013), 对所获得的片段进行PCR扩增、纯化、混样、切胶和选取目的片段。

1.3 测序及产出数据的质量分析

为保证项目分析质量, 本项目采用读100 bp \times 2作为后续的数据评估和分析数据。文库质检合格后用Illumina HiSeq X Ten进行测序。测序得到的序列去接头、去低质量阅读框和去污处理, 得到干净序列。为评估建库实验的准确性, 选用‘日本晴’水稻(*Oryza sativa* ssp. *japonica*)作为对照, 进行相同的处理参与建库和测序。‘日本晴’水稻基因组大小为374.30 Mb, 下载地址: <http://rapdb.dna.affrc.go.jp/>。

1.4 SLAF标签获得和SNP标记开发

利用BWA软件(Li和Durbin 2009)将测序读长比对到参考基因组上, 统计不同染色体上的SLAF标签和多态性SLAF标签。根据序列相似度将各个木薯样品的序列进行聚类, 聚类在一起的序列来源于同一个SLAF标签, 同一SLAF标签在不同样品间的序列相似度高于不同SLAF标签的相似度。一个SLAF标签在不同样品间的序列有差异就可定义

为多态性SLAF标签。使用SAM tools (Li 等2009) 和GATK (McKenna 等2010)两种方法开发SNP, 以两种方法得到的SNP标记交集作为最终可靠的SNP标记数据集。

2 实验结果

2.1 建库评估

对木薯参考基因组进行电子酶切预测, 根据位于重复序列的酶切片段比例尽可能低、酶切片段在基因组上尽量均匀分布、酶切片段长度与具体实验体系的吻合程度和最终获得酶切片段(SLAF标签)数满足预期标签数的酶切方案选择原则(Davey等2013), 确定限制性内切酶为*Hae*III+*Hpy*166II, 酶切片段长度为264~364 bp的序列定义为SLAF标签, 预测可得到235 555个SLAF标签。

本研究通过对对照测序数据的评估监控实验过程是否正常, 确定酶切方案实施的有效性。利用SOAP软件(Li等2009)将对照的测序读长与参考基因组进行比对, 一条序列两端在参考基因组上的比对跨度介于50~1 000 bp的读长(paired-end mapped reads)占总读长的比例为90.83%, 一条序列两端在参考基因组上的比对跨度<50 bp或>1 000 bp的读长(single-end mapped reads)占总读长的比例为1.01%, 未比对到基因组上的读长(unmapped reads)占总读长的比例为8.16%, 比对效率基本正常。

酶切效率是评价SLAF-seq实验是否成功的一个关键指标。基因组上的复杂结构区域(如环状结构域、连续酶切位点等)、基因组DNA样品纯度较低、酶切时间不足等因素都可能影响限制性内切酶的活性, 导致部分酶切位点未被切开。通过统计测序读长插入片段中残留酶切位点的比例来确定酶切效率的高低, 对照数据的测序读长插入片段中残留酶切位点的比例为8.82%, 酶切效率为91.18%, 表明酶切反应正常。本研究中对照数据的双端比对效率为90.83%, 酶切效率为91.18%, 说明SLAF建库正常。

根据对照双端比对序列在基因组中的位置, 计算SLAF标签的实际长度, 绘制对照读长插入片段的长度分布图(图1), 并估测实际的片段选择范围。可见序列的插入片段长度在预期的范围之内, 说明测序质量正常, 采用的测序方法可信度高。

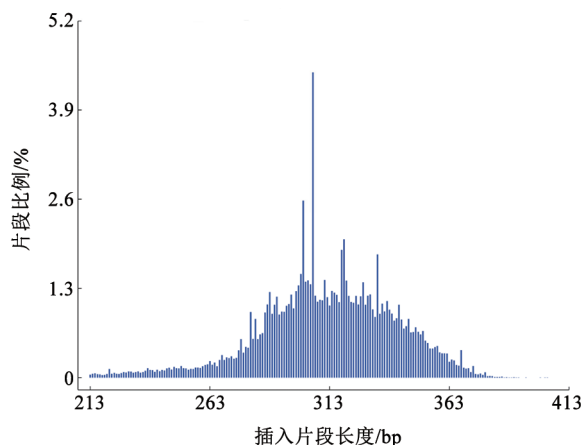


图1 对照序列插入片段分布

Fig.1 Distribution of insert fragments of control sequences

2.2 测序数据统计与评估

为保证项目分析质量, 本项目采用读长100 bp×2作为后续的数据评估和分析数据, 以‘日本晴’水稻的数据为对照评估试建库的准确性。经Illumina HiSeq X Ten平台进行测序, 共获得304.30 Mb的读长数据, 测序后各样品所获得的读长数目在2 123 094~15 020 893之间。测序质量值(Q)是评估高通量测序单碱基错误率的重要指标, 测序质量值越高对应的碱基测序错误率越低。本研究中, 测序质量值≥30的碱基所占比例(Q₃₀)在95.62%~96.60%间, 平均Q₃₀为96.19%, 说明测序碱基错误率低。测序获得GC含量在41.27%~44.96%间, 平均GC含量为42.89%, 对照的GC含量为42.93%。GC含量普遍低(表1), 说明达到测序要求。

2.3 SLAF标签与SNP标记的开发

从39个样本材料中共计获得10 863 426个SLAF标签, 样本测序结果覆盖了基因组的比例在0.640%~1.198%间, 标签的平均测序深度为20.98×, 样本测序数据详细信息见表2。

通过BWA软件将SLAF标签定位到参考基因组上, 统计不同染色体上的SLAF标签和多态性SLAF标签。本研究预测获得SLAF标签数235 555个, 实际所有样品SLAF标签数725 220个, 其中多态性SLAF标签有446 789个, 详细数据见表3。

使用GATK和SAM tools两种方法对得到的446 789个多态性SLAF标签进一步进行SNP标记的开发, 以两种方法得到的SNP标记交集作为最终

表1 不同样本测序数据统计
Table 1 Statistics of sequencing data in different samples

样品号	读长数	$Q_{30}/\%$	GC比例/%	样品号	读长数	$Q_{30}/\%$	GC比例/%
XX048	13 418 431	96.57	42.62	12	3 522 818	96.36	41.34
NJ199	9 074 264	96.33	42.16	13	6 042 775	96.02	44.96
GR3	12 984 951	96.20	42.51	14	5 434 370	96.40	44.89
GR891	15 020 893	96.44	43.35	15	2 372 704	95.90	42.77
GR911	11 932 445	96.51	42.47	16	2 778 779	96.06	42.61
SC5	11 996 413	96.60	43.53	17	8 224 974	96.10	42.04
SC8	9 288 732	96.24	42.86	18	12 517 480	95.97	41.83
SC124	15 102 046	96.47	42.95	19	6 669 124	96.36	41.63
SC205	12 385 282	95.95	43.28	20	3 539 518	96.04	42.41
1	7 086 254	96.28	41.27	21	8 210 632	96.27	41.47
2	2 123 094	96.16	42.09	22	7 343 741	96.27	41.39
3	8 336 079	96.31	42.88	23	6 278 800	95.99	42.12
4	6 733 556	95.62	41.50	24	3 843 362	96.18	42.05
5	8 619 275	96.24	43.09	25	8 321 294	96.25	43.09
6	6 650 604	96.08	44.46	26	7 410 426	95.95	44.88
7	8 491 392	96.14	42.78	27	5 411 142	96.19	43.71
8	6 285 538	95.77	44.85	28	6 126 594	96.31	41.46
9	7 674 461	96.22	41.89	29	6 988 510	95.91	44.55
10	6 261 894	96.28	44.00	30	7 205 162	96.27	44.18
11	5 193 908	96.19	44.82	对照	395 790	96.18	42.93

可靠的SNP标记数据集。从39个样本材料中共计获得45 566 089个群体SNP, 样本完整度在30.84%~58.79%间, 样本平均完整度为46.65%, 样本杂合率在10.02%~17.45%间, 样本平均杂合率为14.08%(表2)。根据完整度>0.8、MAF (次要基因型频率)>0.05过滤, 共筛选到2 504 553个高度一致的群体SNP。

根据SLAF标签和SNP标记在染色体上的分布, 绘制SLAF标签和SNP标签的染色体分布图, 具体分布情况分别见图2和3。按照1 M的大小对基因组进行了划分, 每个条带内的SLAF标签数、SNP标记数越多, 颜色越深; SLAF标签数、SNP标记数越少, 颜色越浅。由图2和3可知, SLAF标签和SNP标记在染色体上分布较均匀。

3 讨论

木薯是遗传背景较复杂的作物, 通过传统杂交育种周期长、费工费力、效率低, 采用分子育种技术提升木薯新品种培育已成为当前木薯育种的重要方法, 各国的木薯遗传改良已经逐步引入以分子标记为基础的基因型选择技术, 开发大量木薯特异分子标记可促进木薯种质资源中优良基

因的挖掘和利用, 不断丰富木薯优良基因资源, 为进一步加快木薯良种选育步伐提供强有力的保障。如木薯EST-SSR (Sraphet等2011)、EST-SNP、SSR (Rabbi等2012)、微卫星AFLP (Whankaew等2012)等技术的开发和利用, 提高了木薯传统育种速度和筛选效率。然而与水稻等大作物相比, 有关木薯分子标记的研究积累还十分有限。如何利用我国丰富的木薯种质资源, 大力开展木薯遗传育种特别是分子育种的基础研究, 加强自主知识产权的品种和种质创新对我国木薯产业的可持续健康发展具有十分重要的意义。

3.1 SLAF-seq技术开发木薯分子标记效率高

SLAF-seq技术是以生物信息学方法和高通量测序技术为基础, 通过设置特定的预酶切方案获得特异长度的片段, 再对特定的片段进行测序。运用这种技术开发的标记不但密度大、特异性和准确率高、稳定性和可重复性好、有效标记多, 而且开发成本低于常规分子标记。木薯为二倍体, 其全基因组大小为769 Mb, 与已完成基因组测序的拟南芥(*Arabidopsis thaliana*)、水稻、高粱(*Sorghum bicolor*)相比, 木薯基因组结构最大的特点在于杂合度高, 利用常规方法开发木薯基因组特异

表2 SLAF和SNP标签的数据统计

Table 2 Data statistics of SLAF and SNP tags

样品编号	SLAF标签信息				SNP标签信息		
	SLAF标签数	覆盖度/%	总测序深度	平均测序深度/ \times	SNP标签数	完整度/%	杂合率/%
XX048	333 381	1.145	10 574 459	31.72	1 366 875	54.58	15.98
NJ199	317 222	1.090	6 965 590	21.96	1 315 560	52.53	15.97
GR3	348 710	1.198	9 922 235	28.45	1 472 530	58.79	15.46
GR891	344 628	1.184	12 141 111	35.23	1 440 374	57.51	16.35
GR911	317 626	1.091	9 476 478	29.84	1 332 865	53.22	14.77
SC5	319 392	1.097	9 752 866	30.54	1 345 286	53.71	13.85
SC8	315 804	1.085	7 367 274	23.33	1 313 251	52.43	14.55
SC124	337 634	1.160	12 139 891	35.96	1 415 587	56.52	16.56
SC205	327 637	1.126	9 492 141	28.97	1 380 760	55.13	16.18
1	279 907	0.962	5 703 081	20.37	1 134 721	45.31	13.31
2	186 158	0.640	1 649 420	8.86	772 319	30.84	10.63
3	279 397	0.960	6 541 541	23.41	1 173 482	46.85	14.68
4	271 991	0.934	5 247 156	19.29	1 096 879	43.80	13.60
5	286 589	0.985	6 564 391	22.91	1 217 429	48.61	17.45
6	255 981	0.879	4 998 441	19.53	1 108 603	44.26	13.27
7	293 105	1.007	6 574 772	22.43	1 206 634	48.18	15.12
8	254 525	0.874	4 786 059	18.81	1 097 637	43.83	13.66
9	253 669	0.872	4 767 173	18.79	1 115 110	44.52	14.54
10	256 362	0.881	3 968 057	15.48	1 152 884	46.03	14.35
11	281 303	0.966	6 077 683	21.61	1 148 135	45.84	13.68
12	229 178	0.787	2 759 853	12.04	949 259	37.90	12.60
13	266 716	0.916	4 648 902	17.43	1 168 341	46.65	13.50
14	257 244	0.884	4 130 287	16.06	1 153 634	46.06	14.26
15	197 338	0.678	1 852 377	9.39	812 040	32.42	11.26
16	201 245	0.691	2 147 396	10.67	819 609	32.72	10.02
17	290 640	0.999	6 525 911	22.45	1 180 139	47.12	13.18
18	333 919	1.147	9 798 621	29.34	1 374 845	54.89	14.84
19	277 546	0.954	5 385 703	19.40	1 134 397	45.29	13.46
20	229 310	0.788	2 775 082	12.10	953 139	38.06	12.40
21	292 883	1.006	6 402 360	21.86	1 177 459	47.01	14.68
22	287 344	0.987	5 939 690	20.67	1 159 661	46.30	13.74
23	273 152	0.938	4 979 378	18.23	1 128 888	45.07	13.85
24	228 211	0.784	3 033 062	13.29	953 408	38.07	12.37
25	283 591	0.974	6 494 749	22.90	1 223 700	48.86	14.58
26	275 475	0.946	5 710 575	20.73	1 196 699	47.78	14.20
27	263 506	0.905	4 262 367	16.18	1 104 359	44.09	13.73
28	274 020	0.941	4 842 815	17.67	1 091 253	43.57	13.43
29	265 416	0.912	5 410 771	20.39	1 144 157	45.68	13.59
30	275 671	0.947	5 519 795	20.02	1 234 181	49.28	15.60

分子标记耗时长, 难度大。本研究利用SLAF-seq技术共获得725 220个木薯SLAF标签, 标签平均测序深度为20.98 \times , 其中多态性SLAF标签有446 789个, 合计开发获得2 504 553个SNP标记, 所获得数据量足够用来进行特异性SNP标记的验证与开发, 而通过常规方法获得木薯分子标记的效率很低。

如韦祖生等(2008)采用EST-SSR技术对76份木薯材料进行遗传多样性分析, 共获得135个多态性位点; Sraphet等(2011)利用EST-SSR技术对‘Huay Bong 60’与‘Hanatee’杂交产生的F₁株系进行多态性研究, 共开发出640个新的SSR标记物, 进一步的多态性研究筛选出199个SSRs和168个EST-SSRs的新多态

表3 SLAF标签在各染色体上的数量统计
Table 3 Quantitative statistics of SLAF tags on each chromosome

染色体编号	染色体长度/bp	预测SLAF标签数	预测SLAF平均间距/bp	实际所有样品SLAF标签数	实际所有样品SLAF平均间距/bp	实际所有样品多态性SLAF标签数	实际所有样品多态性SLAF平均间距/bp
CM004387.1	34 959 721	14 258	2 452	42 384	824	25 844	1 353
CM004388.1	32 417 782	13 582	2 387	40 203	806	25 514	1 271
CM004389.1	29 412 403	11 494	2 559	34 938	841	21 474	1 370
CM004390.1	28 735 499	11 284	2 547	34 711	827	21 494	1 337
CM004391.1	28 438 989	11 847	2 401	35 548	800	21 468	1 325
CM004392.1	27 939 960	11 545	2 420	34 545	808	21 859	1 278
CM004393.1	27 069 033	10 531	2 570	32 604	830	20 548	1 317
CM004394.1	34 000 172	13 806	2 463	43 197	787	26 839	1 267
CM004395.1	29 413 690	11 435	2 572	35 208	835	21 858	1 346
CM004396.1	26 335 333	10 278	2 562	31 154	845	19 575	1 345
CM004397.1	27 369 018	10 382	2 636	32 672	837	20 466	1 337
CM004398.1	31 596 830	12 879	2 453	39 353	802	23 238	1 360
CM004399.1	28 119 335	11 072	2 540	34 026	826	21 004	1 339
CM004400.1	24 855 542	10 230	2 430	30 730	808	18 946	1 312
CM004401.1	26 192 760	10 976	2 386	32 265	811	19 555	1 339
CM004402.1	28 979 753	11 893	2 437	35 567	814	21 198	1 367
CM004403.1	27 373 735	11 354	2 411	37 452	730	23 102	1 185
CM004404.1	25 231 592	10 117	2 494	32 140	785	20 379	1 238
其他	63 669 956	26 592	2 394	86 523	735	52 428	1 214
总计	582 117 524	235 555	2 471	725 220	802	446 789	1 303

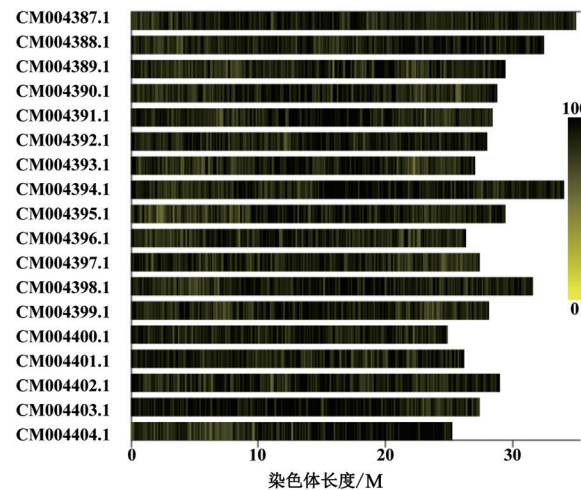


图2 SLAF标签在各染色体上的分布

Fig.2 SLAF tags distribution on each chromosome
图中每一个黄色条带代表一条染色体, 颜色越深的区域即标签越集中分布的区域, 图3同。

性标记; 孙湘来等(2011)利用KU50与‘华南124’杂交建立了含240个单株的木薯F1分离群体, 利用EST-SSR技术对该F1群体进行多态性分析, 共获得269个多态性标记。夏秀忠等(2005)利用AFLP分

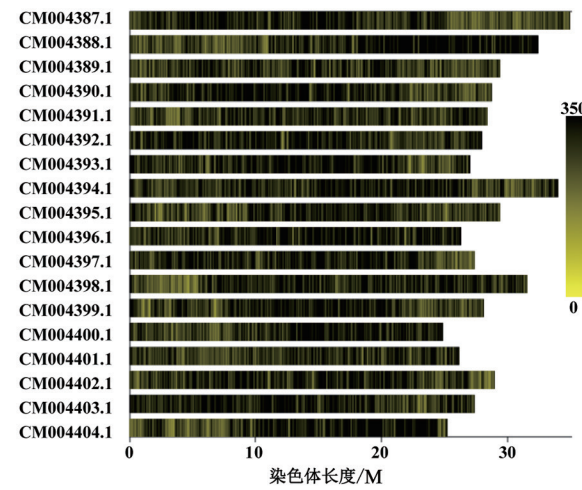


图3 SNP标签在各染色体上的分布

Fig.3 SNP tags distribution on each chromosome
子标记技术对21个木薯品种进行基因组DNA多态性分析, 扩增的产物中的有效带数为117条; 漆雪琳(2011)对‘华南124’的4个干旱胁迫时间点进行了cDNA-AFLP分析, 从3 500个有效扩增的基因片段中, 得到了117个干旱胁迫差异表达的转录获得片段; 王宇阳(2013)采用改良型cDNA- AFLP技术对

木薯KU50和‘华南124’不同程度干旱胁迫的样品进行分析, 总共扩增出条带483个, 其中差异表达条带106个; Whankaew等(2012)基于微卫星AFLP技术对48份木薯进行分子标记开发研究, 获得多态性微卫星位点9个, 每个多态性位点的等位基因在3~15个之间, 总等位基因数32个。赵凯和蒋盛军(2011)以30个块根淀粉差异木薯品种为材料, 利用18条RAPD引物扩增出131个条带, 其中多态性条带126个; 祝天成等(2015)对木薯细菌性枯萎病原菌的RAPD分析, 共扩增到100个RAPD谱带, 其中多态性带有87条。彭靖茹等(2012a)对39份木薯种质资源的ISSR标记多态性进行检测, 共获得70个扩增带形。可见SLAF-seq技术可以很好地应用于木薯分子标记的开发, 效率远高于上述的SSR、AFLP、RAPD和ISSR等分子标记技术。

3.2 基于SLAF-seq技术的木薯SNP位点开发应用

SNP标记作为第三代标记, 具有遗传稳定性高、分布广、数量多、易自动化分析等优点, 但全基因组深度重测序进行SNP开发所需的高额费用, 限制了其在相关领域的广泛应用。SLAF-seq技术结合了位点特异性扩增和高通量测序技术, 只需对特定的DNA酶切片段进行测序, 而非全基因组测序, 大大降低了测序的工作量和费用, 非常适合于快速、大量SNP标记的开发。自SLAF-seq技术产生以来已经运用在多种作物的SNP位点挖掘中, 并取得了显著的效果。如苏文瑾等(2016)用SLAF-seq技术对300份甘薯种质资源进行甘薯SNP位点开发, 共计获得795 794个高质量SNP位点; 王伟等(2016)采用SLAF-seq技术开发甘蓝型油菜霜霉病抗性SNP位点研究, 共开发出132 519个SLAF标签, 获得264 256个SNP位点, 经过筛选最终得到29 385个特异性SNP用于后续关联分析研究; 俞超等(2017)以不同品种红心火龙果为材料, 利用SLAF-seq技术开发获得了120 294个群体SNP, 其中高度一致性的群体SNP 51 859个。

宋红艳(2014)从‘岩生木薯’、‘W14’、‘SC205’和‘SC8’的转录组数据中分别挖掘到SNP 79 354、187 629、83 462和92 909个; 肖急祥(2016)对木薯包含谷氨酰胺合成酶beta-Grasp域的序列进行SNP分析, 发现序列存在35个多态性位点和21个插入缺失标记(InDel)变异。本研究开发共得到2 504 553

个群体SNP标记, 是迄今为止开发效率较高的分子标记, 也是迄今为止从木薯中获得最多的SNP标记。这些标记的获得可以为后续研究木薯品种间共有或特异性SNP标记的开发提供基础, 也可以利用所开发的SNP标记, 联系不同个体的表现型, 关联定位与某些性状紧密连锁的分子标记, 进而可以实现优良基因的精细定位, 开拓木薯种质资源保护、新品种开发等相关领域的研究。

参考文献(References)

- Davey JW, Cezard T, Fuentes-Utrilla P, et al (2013). Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol*, 22 (11): 3151–3164
- Fregene M, Angel F, Gomez R, et al (1997). A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theor Appl Genet*, 95 (3): 431–441
- Kozich JJ, Westcott SL, Baxter NT, et al (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microb*, 79 (17): 5112–5120
- Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25 (14): 1754–1760
- Li J (2006). QTL mapping for high starch content and other important economic traits in cassava (*Manihot esculenta* Crantz) (dissertation). Danzhou, Hainan: South China University of Tropical Agriculture (in Chinese with English abstract) [李杰(2006). 木薯高淀粉等重要经济性状相关基因的QTL标记(学位论文). 海南儋州: 华南热带农业大学]
- Li R, Yu C, Li Y, et al (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25 (15): 1966–1967
- Mba REC, Stephenson P, Edwards K, et al (2001). Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theor Appl Genet*, 102 (1): 21–31
- McKenna A, Hanna M, Banks E, et al (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 20 (9): 1297–1303
- Ogura T, Busch W (2015). From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. *Curr Opin Plant Biol*, 23: 98–108
- Peng J, Ma Z, Li P, et al (2012). Genetic polymorphism of cassava germplasm SSR molecular markers. *Chin Agr Sci Bull*, 28 (6): 58–62 (in Chinese with English abstract) [彭

- 靖茹, 马增风, 黎萍等(2012). 木薯种质资源遗传多态性SSR分子标记的研究. 中国农学通报, 28 (6): 58–62]
- Peng JR, Gan ZY, Li P, et al (2012). Research on genetic polymorphism of cassava germplasms and ISSR molecular markers. J Anhui Agr Sci, 40 (7): 3877–3880 (in Chinese with English abstract) [彭靖茹, 甘志勇, 黎萍等(2012). 木薯种质资源遗传多态性ISSR分子标记的研究. 安徽农业科学, 40 (7): 3877–3880]
- Pootakham W, Shearman JR, Ruang-arereate P, et al (2014). Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). PLoS ONE, 9 (12): e116028
- Qi X (2011). Isolation of drought stress related genes through cDNA-AFLP technique in *Manihot esculenta* Crantz (dissertation). Haikou: Hainan University (in Chinese with English abstract) [漆雪琳(2011). 利用cDNA-AFLP技术分离木薯干旱胁迫响应相关基因(学位论文). 海口: 海南大学]
- Rabbi IY, Hamblin MT, Kumar PL, et al (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Res, 186: 87–96
- Rabbi IY, Kulembeka HP, Masumba E, et al (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). Theor Appl Genet, 125 (2): 329–342
- Song H (2014). The study on transcriptome of tuberous root in wild species and cultivar of cassava (dissertation). Haikou: Hainan University (in Chinese with English abstract) [宋红艳(2014). 木薯野生种和栽培种块根转录组研究(学位论文). 海口: 海南大学]
- Sraphet S, Boonchanawiwat A, Thanyasiriwat T, et al (2011). SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). Theor Appl Genet, 122 (6): 1161–1170
- Su WJ, Zhao N, Lei J, et al (2016). SNP sites developed by specific length amplification fragment sequencing (SLAF-seq) in sweetpotato. Sci Agr Sin, 49 (1): 27–34 (in Chinese with English abstract) [苏文瑾, 赵宁, 雷剑等(2016). 基于SLAF-seq技术的甘薯SNP位点开发. 中国农业科学, 49 (1): 27–34]
- Sun X, Chen X, Wen M, et al (2011). Based on EST-SSR construct the molecular genetic linkage map of cassava. Chin Agr Sci Bull, 27 (9): 231–236 (in Chinese with English abstract) [孙湘来, 陈新, 文明富等(2011). 基于EST-SSR的木薯分子标记遗传连锁图谱的构建. 中国农学通报, 27 (9): 231–236]
- Sun X, Liu D, Zhang X, et al (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. PLoS ONE, 8 (3): e58700
- Wang W, Liu F, Ren L, et al (2016). Resistance SNP development to downy mildew in *Brassica napus* using SLAF-seq technique. Chin J Oil Crop Sci, 38 (5): 555–562 (in Chinese with English abstract) [王伟, 刘凡, 任莉等(2016). 采用SLAF-seq技术开发甘蓝型油菜霜霉病抗性SNP位点. 中国油料作物学报, 38 (5): 555–562]
- Wang Y (2013). Establishment and application of a modified cDNA-AFLP technique to screen drought stress-related genes in cassava (*Mahihot esculenta* Crantz) (dissertation). Haikou: Hainan University (in Chinese with English abstract) [王宇阳(2013). 建立并利用改良型cDNA-AFLP技术筛选木薯干旱胁迫相关基因(学位论文). 海口: 海南大学]
- Wei Z, Xia Z, Li K, et al (2008). Analysis of genetic diversity of cassava gene pool by EST-SSR markers. Chin J Trop Crops, 29 (3): 304–309 (in Chinese with English abstract) [韦祖生, 夏志强, 李开绵等(2008). 木薯种质库遗传多样性的EST-SSR标记. 热带作物学报, 29 (3): 304–309]
- Whankaew S, Sraphet S, Thaikert R, et al (2012). Characterization of microsatellite markers in cassava based on microsatellite-AFLP technique. Genet Mol Res, 11 (2): 1319–1326
- Xia XZ, Feng D, Qin Y (2005). AFLP analysis of genetic diversity within and among Guangxi main cassava cultivars. Southwest China J Agr Sci, 18 (6): 818–821 (in Chinese) [夏秀忠, 冯斗, 覃艳(2005). AFLP对广西主栽木薯品种的遗传多态性分析. 西南农业学报, 18 (6): 818–821]
- Xia Z, Zou M, Zhang S, et al (2014). AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping. Sci Rep, 4: 7300
- Xiao J (2016). Screening of nitrogen efficient germplasm and minning of elite allele in cassava (dissertation). Haikou: Hainan University (in Chinese with English abstract) [肖急祥(2016). 木薯氮高效种质筛选及优等位变异挖掘(学位论文). 海口: 海南大学]
- Yu C, Cheng Y, Wang C, et al (2017). SNP sites development by specific length amplification fragment sequencing (SLAF-seq) and genetic analysis in red pitaya. Chin J Trop Crops, 38 (4): 591–596 (in Chinese with English abstract) [俞超, 陈煜, 汪财生等(2017). 基于SLAF-seq技术的红心火龙果SNP位点开发及遗传分析. 热带作物学报, 38 (4): 591–596]
- Zhang P, An D, Ma Q, et al (2013). Key scientific questions and recent advances in cassava molecular breeding. Sci Sin Vitae, 43 (12): 1082–1089 (in Chinese) [张鹏, 安冬, 马秋香等(2013). 木薯分子育种中若干基本科学问题的思考与研究进展. 中国科学: 生命科学, 43 (12): 1082–1089]
- Zhang Y, Wang L, Xin H, et al (2013). Construction of a

- high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol*, 13: 141
- Zhao K, Jiang S (2011). Genetic background research on RAPD of cassava with different root starch phosphorus contents. *Chin Agr Sci Bull*, 27 (9): 237–241 (in Chinese with English abstract) [赵凯, 蒋盛军(2016). 块根淀粉磷
- 差异木薯种质资源的RAPD遗传背景研究. *中国农学通报*, 27 (9): 237–241]
- Zhu TC, Shi T, Li CP, et al (2015). RAPD analysis of *Xanthomonas axonopodis* pv. *manihotis*. *Guangdong Agr Sci*, 42 (7): 117–121 (in Chinese with English abstract) [祝天成, 时涛, 李超萍等(2015). 木薯细菌性枯萎病原菌的RAPD分析. *广东农业科学*, 42 (7): 117–121]

Development of SNP markers in cassava (*Manihot esculenta* Crantz) based on SLAF-seq technology

YU Ben-Chi, WEI Li-Jun*, LEI Kai-Wen, SONG En-Liang, MA Chong-Xi

Guangxi Subtropical Crops Research Institute, Nanning 530001, China

Abstract: The development of a large number of specific molecular markers can promote the exploitation and utilization of the excellent genes in cassava (*Manihot esculenta* Crantz) germplasm resources. Specific-locus amplified fragment sequencing (SLAF-seq) technology was employed to sequence the genomes of 39 cassava varieties/lines in this study. Results show that paired-end alignment efficiency and cleavage efficiency were 90.83% and 91.18%, respectively. 304.30 Mb reads were obtained by sequencing, the reads of the samples varied from 2 123 094 to 15 102 046, and that of the control was the smallest, only 395 790. The sequencing quality value (Q_{30}) changed from 95.62% to 96.60%, all above 95%. The GC content of the samples changed from 41.27% to 44.96%, the average value was 42.89%, and that of the control was 42.93%. Finally, 725 220 SLAF labels were developed, among them 446 789 were polymorphic. The average sequencing depth of the samples was 20.98 ×, and 2 504 553 SNP markers were found. Obviously, SLAF-seq technology is suitable for developing more efficient SNP sites in cassava than ever.

Key words: cassava; molecular marker; SLAF-seq; SNP

Received 2018-03-20 Accepted 2018-05-15

This work was supported by Science and Technology Program in Guangxi Province (Guike AB16380080 and Guike AA16380002), and the Project of Natural Science Founding in Guangxi Province (2017GXNSFBA198172).

*Corresponding author (weilijun2002@126.com).