

玉米全基因组批量化 SSR 多态性标记 开发及 PCR 扩增验证

史典义^{1,2}, 邓佳¹, 孙明伟¹, 文轩涛¹, 汪洋¹

(1. 内江师范学院生命科学学院, 四川 内江 641100; 2. 沱江流域特色农业资源四川省科技资源共享服务平台, 四川 内江 641100)

摘要: 简单序列重复在真核生物的基因组中含量非常丰富, 且常常随机、均匀分布于整个基因组中。SSR 标记广泛应用于动、植物基因定位、群体遗传多样性分析、遗传连锁图谱构建、指纹图谱构建、种质鉴定、分子标记辅助选择等方向。随着大量全基因组序列的公布, NCBI 等数据库中可利用的序列信息量激增。但在常规分子生物学研究中, 仍然延续传统的小规模 SSR 标记开发模式, 开发标记数量非常有限、效率较低。结合 MISA 脚本指令、Primer3 引物设计软件、e-PCR 特异性扩增分析软件, 对玉米全基因组水平 SSR 位点进行检索、引物设计、特异性扩增分析以及扩增长度差异分析, 开发高密度的 SSR 标记。通过引物合成、PCR 扩增和琼脂糖凝胶电泳, 从常规分子生物学水平验证所开发标记的可靠性。

关键词: 玉米; 基因组; 批量化; SSR; 多态性; 分子标记

中图分类号: S513.035.3

文献标识码: A

Development of SSR Polymorphic Markers at Genome-wide Level in Maize and Validation by PCR Amplification

SHI Dian-yi^{1,2}, DENG Jia¹, SUN Ming-wei¹, WEN Xuan-tao¹, WANG Yang¹

(1. College of Life Sciences, Neijiang Normal University, Neijiang 641100;

2. Sichuan Science and Technology Resources Sharing Service Platform of Special Agricultural
Resources in Tuojiang River Basin, Neijiang 641100, China)

Abstract: Simple sequence repeats (SSR) are abundant in eukaryotic genomes and are often randomly and uniformly distributed throughout the whole genome. SSR markers are widely used in animal and plant gene mapping, population genetic diversity analysis, genetic linkage map construction, fingerprint map construction, germplasm identification, molecular marker-assisted selection and other directions. With the publication of large numbers of whole genome sequences, the sequence information available in databases such as NCBI has grown dramatically. However, in conventional molecular biology research, the traditional small-scale marker development model is still continuing, and the markers developed are very limited due to low efficiency. Combining MISA scripts, Primer3 and e-PCR software to search SSR locus, design primers, and to analyze the specific amplification, and calculate their length differences, high-density SSR markers in maize genome were successfully developed. Finally, through primer synthesis, PCR amplification and agarose gel electrophoresis, the reliability of the developed marker was verified by conventional methods of molecular biology.

Key words: Maize; Genome; Batch; SSR; Polymorphism; Molecular marker

录用日期: 2022-04-29

基金项目: 内江师范学院 2020 年大学生创新创业训练项目“全基因组 fasta 序列批量化 SSR 引物设计流程开发”(X2020047)、内江师范学院线上线下混合式一流课程建设专项“生物信息学”(JK202021)

作者简介: 史典义(1978-), 山东烟台人, 博士, 主要从事植物、微生物分子机理及育种研究。E-mail: onaboard@163.com

微卫星重复序列, 又称为简单序列重复 (SSR, Simple Sequence Repeats)、短串联重复序列 (Short Tandem Repeat, STR), 在原核生物、真核生物以及真核生物的细胞器 DNA 中广泛存在^[1]。尤其在真核生物的基因组中, SSR 的含量非常丰富, 而且常常随机、均匀分布于整个基因组 DNA 中。SSR 的突变率在不同物种以及同一物种的不同位点或同一位点的

不同等位基因间存在很大差异^[2]。根据SSR两端的保守性序列设计特异引物,通过PCR技术将其间的SSR DNA序列扩增出来,再利用电泳分析技术就可以得到其长度的多态性^[3]。

SSR标记具有数量大、分布广泛均匀、多态性高、检测方便等特点。随着越来越多的全基因组参考序列的公布,SSR标记已经被广泛应用于动、植物基因定位、群体遗传多样性分析、遗传连锁图谱构建、指纹图谱构建、种质鉴定、分子标记辅助选择等方向^[4-6]。在过去的20年中,SSR是动植物基因分型最为广泛使用的标记^[7]。目前,有多种SSR位点分析工具,如在Windows系统下操作的SSRhunter^[8]以及GMATA^[9]。Thiel等^[10]使用Perl脚本开发的MISA工具(MicroSAteellite identification tool)以及Wang等^[11]使用Perl和Java开发的GMATo等,可在Linux系统下批量化分析SSR位点。其中,应用最为广泛的是MISA工具,MISA工具是一套开源的Perl语言脚本,其运行指令简单,运行速度快,可本地化操作,且可与多种工具联用,已成为大部分SSR研究者的首选工具^[12]。

随着Illumina、PacBio为代表的高通量测序、第3代测序技术的发展,大量全基因组序列被公布,NCBI等数据库中可利用的序列信息量激增。但在常规分子生物学研究中,仍然延续传统的小规模SSR标记开发模式,开发标记数量非常有限、效率较低;另一方面,完全依托高通量测序获得的SNP(Single Nucleotide Polymorphism)标记信息,进行理论化的多态性分析,SNP通常无法像SSR标记一样通过常规的PCR扩增进行验证。本文结合MISA脚本指令、Primer3^[13]引物设计软件、e-PCR^[14]特异性扩增分析软件,对玉米全基因组水平SSR位点进行搜索、引物设计、特异性扩增分析以及扩增长度差异分析,成功开发高密度的SSR标记。通过引物合成、PCR扩增和琼脂糖凝胶电泳,从常规分子生物学实验水平验证所开发标记的可靠性。

1 材料与方法

1.1 软件下载与安装

SSR分析软件MISA(version 2.1)的3个Perl脚本,即misa.pl、primer3_in.pl、prim_output.pl均下载于<http://pgrc.ipk-gatersleben.de/misa/>,运行MISA脚本的软件为Perl 5(version 30),服务器操作系统为Ubuntu(version 20.04)。引物设计软件Primer3下载自<https://pkgs.org/download/primer3>。e-PCR(version 2.3.12)软件通过Ubuntu系统中的apt install指令进

行安装。

1.2 玉米全基因组fasta序列下载

在NCBI Genome(<https://www.ncbi.nlm.nih.gov/genome/>)数据库网站中下载玉米纯合自交系B73和Mo17的全基因组fasta序列。B73全基因组fasta序列下载自ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/005/GCF_000005005.2_B73_RefGen_v4/GCF_000005005.2_B73_RefGen_v4_genomic.fna.gz。Mo17全基因组fasta序列下载自ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/003/185/045/GCA_03185045.1_Zm-Mo17-REFERENCE-CAU-1.0/GCA_03185045.1_Zm-Mo17-REFERENCE-CAU-1.0_genomic.fna.gz。

1.3 SSR引物设计及位点特异性扩增与多态性分子标记的筛选

使用Primer3(version 2.5.0)软件进行SSR引物设计,参数设置为PRIMER_PRODUCT_SIZE_RANGE=150-30, PRIMER_MIN_GC=40.0, PRIMER_MAX_GC=60.0, PRIMER_MIN_SIZE=20, PRIMER_MAX_SIZE=28, PRIMER_OPT_SIZE=22,其余参数采用默认值。对设计成功的SSR引物,使用e-PCR软件分析其在基因组上的扩增特异性,参数设置为:-u +, -w 9, -f 1, -m 50, D=100-500, N=2, G=2, T=3。分别分析引物在B73、Mo17参考基因组上的扩增特异性之后,筛选出在能够同时在两个参考基因组上特异性扩增的引物,并根据各自的理论扩增长度,进一步筛选出扩增长度差异在3、20、50 bp及以上的多态性引物。

1.4 SSR标记的冗余性分析及染色体物理图谱画图分析

根据MaizeGDB(<https://www.maizegdb.org>)数据库公布的2034个SSR位点的上下游引物碱基信息,使用e-PCR软件与新开发的45162个SSR标记位点两侧各500 bp的序列片段进行特异性扩增分析,从而筛选出与已报道的2034个SSR标记存在冗余的位点。参数设置为:-u +, -w 9, -f 1, -m 50, D=100-1000, N=2, G=2, T=3。使用R包LinkageMapView(version 2.1.2)对新开发的长度差异绝对值 ≥ 3 bp的SSR标记进行物理图谱画图分析。对长度差异绝对值 ≥ 50 bp的SSR特异性扩增位点,使用MapChart(version 2.32)软件进行物理图谱画图分析。

1.5 玉米自交系材料准备与DNA提取

用于PCR扩增验证的实验材料,包括自交系B73、Mo17以及B73×Mo17、B73×B104、B104×Mo17这3个杂交F₁代群体,由西南大学农学与生物科技学院刘朝显副教授保种并配制。收获后的干种子经

55℃温水浸泡30 min后,播种于营养钵中,并在光照培养箱中,24℃,14 h/10 h光暗条件下培养2周左右。

取新鲜玉米叶片约0.1 g,用液氮研磨至粉末状,加入700 μ L 2.0%的CTAB提取液(使用前加入终浓度为0.5%的 β -巯基乙醇),65℃浸提60 min。使用体积比为24:1的氯仿:异戊醇萃取去除蛋白。上清液用2/3体积的-20℃预冷异丙醇沉淀析出DNA。干燥后的DNA,加入50 μ L 1.0 $\times$ 无菌TE缓冲液充分溶解,并加入1 μ L 10 mg/mL的RNase A溶液,25℃孵育30 min以消化处理残留的RNA。

1.6 PCR扩增与琼脂糖凝胶电泳验证

用1X TBE缓冲液配制1.2%琼脂糖凝胶,待凝胶自然冷却至60℃左右,按照每100 mL凝胶加5 μ L的比例加入Gel-Green(D0143,上海碧云天生物技术有限公司)核酸染料,混匀后倒入胶槽并自然冷却20 min以上。取5 μ L PCR扩增产物与适量的10X loading buffer(P022-01,南京诺唯赞生物科技股份有限公司)混匀后加入琼脂糖凝胶电泳孔。取5 μ L DL2000 Plus DNA Marker(MD101-01,南京诺唯赞生

物科技股份有限公司)加入点样孔,作为分子量参考标准。在100 V电压、100 mA电流的条件下电泳约45 min,然后转至凝胶成像系统(Gel Doc XR System,美国伯乐)下进行拍照记录。

2 结果与分析

2.1 分析流程

通过NCBI Genome、MaizeGDB数据库下载全基因组fasta序列,获得了玉米自交系B73、Mo17的全基因组序列。使用MISA软件对B73、Mo17基因组序列进行SSR位点搜索。根据B73序列SSR位点信息,使用Primer3进行引物设计,并使用e-PCR软件分析其在B73、Mo17序列中的扩增特异性。筛选出能够同时在B73和Mo17基因组序列中能够特异性扩增的引物。根据特异性扩增引物在两个自交系中的理论扩增长度,比较长度差异,从而筛选出多态性标记,最后通过PCR扩增和琼脂糖凝胶电泳验证新开发标记的可靠性(图1)。

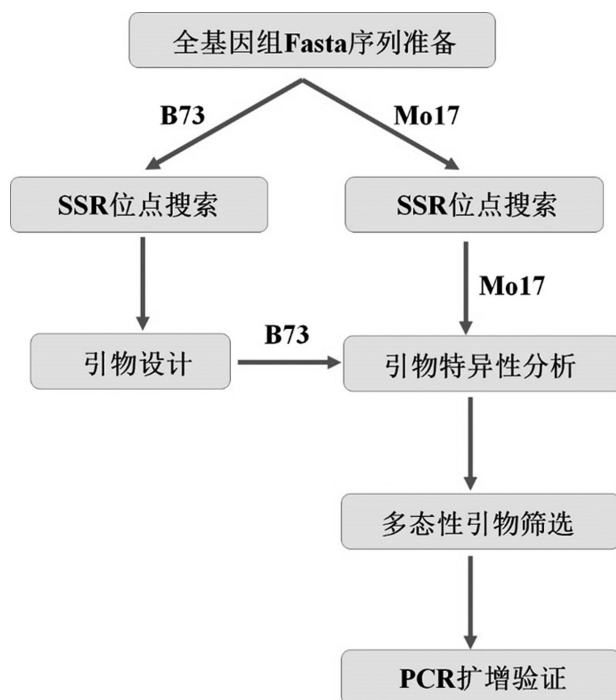


图1 全基因组多态性SSR标记开发分析流程

Fig.1 Analysis workflow for the development of genome-wide polymorphic SSR markers

2.2 玉米全基因组SSR位点信息统计

玉米自交系B73、Mo17染色体平均长度分别为(210.63 $\pm$ 49.12)Mb和(210.45 $\pm$ 49.66)Mb(图2)。每条染色体平均SSR位点数量,包括重复序列在1~6个碱基(p1-p6型)以及复合型(c型)SSR,两个自交系各为(19 558.20 $\pm$ 4 785.31)个和(19 306.40 $\pm$ 4 717.93)个,

平均分布密度为(92.70 $\pm$ 1.81)个/Mb和(91.61 $\pm$ 1.80)个/Mb(表1)。在B73和Mo17基因组染色体上,随着碱基重复数量增多,对应的SSR位点数量依次减少,其中单碱基重复序列最多,然后是双碱基重复、三碱基重复;四至六碱基重复数量则显著下降到百位数、十位数(表2、表3、图2)。

表1 玉米自交系 B73 和 Mo17 各染色体长度及 SSR 位点数量

Table 1 Chromosome length and corresponding number of SSR loci in maize inbred lines B73 and Mo17

项 目 Item	Inbred	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	平均值 Mean value
染色体长度(Mb)	B73	307.04	244.44	235.67	246.99	223.90	174.03	182.38	181.12	159.77	150.98	210.63±49.12
	Mo17	306.18	243.31	240.96	249.75	220.38	168.72	178.57	182.61	164.94	149.04	210.45±49.66
SSR 位点数量(个)	B73	29 462	22 919	21 670	22 399	20 895	16 354	16 721	16 513	14 431	14 218	19 558.20±4 785.31
	Mo17	28 611	22 569	21 873	22 638	20 694	15 789	16 243	16 117	14 872	13 658	19 306.40±4 717.93
SSR 分布密度(个/Mb)	B73	95.95	93.76	91.95	90.69	93.32	93.97	91.68	91.17	90.32	94.17	92.70±1.81
	Mo17	93.45	92.76	90.78	90.64	93.90	93.58	90.96	88.26	90.17	91.64	91.61±1.80

表2 玉米自交系 B73 各染色体 SSR 类型及数量

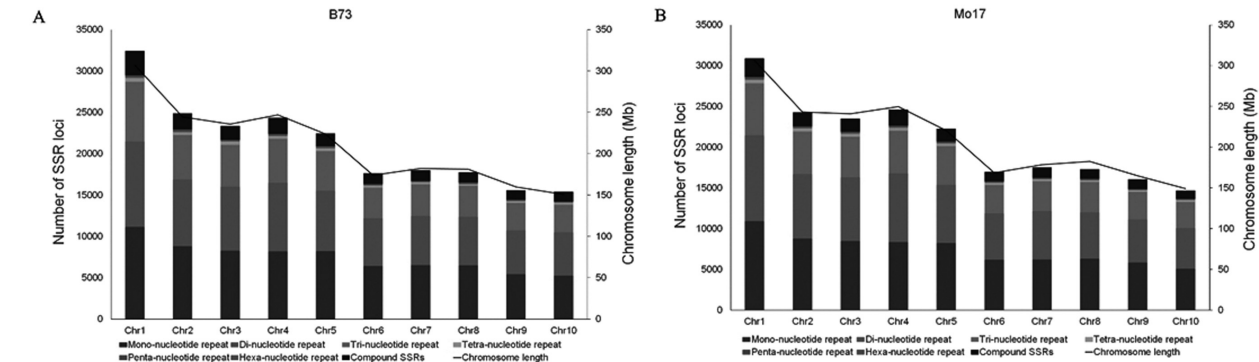
Table 2 Types and quantities of SSR loci on each chromosome of maize inbred line B73

项目 Item	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
Mono-nucleotide repeat	11 153	8 777	8 245	8 169	8 208	6 390	6 509	6 470	5 442	5 249
Di-nucleotide repeat	10 284	8 075	7 731	8 302	7 264	5 773	5 939	5 865	5 252	5 197
Tri-nucleotide repeat	7 234	5 393	5 083	5 283	4 824	3 725	3 814	3 759	3 341	3 365
Tetra-nucleotide repeat	505	404	422	405	373	295	303	248	247	282
Penta-nucleotide repeat	176	147	123	127	139	98	92	112	83	74
Hexa-nucleotide repeat	110	123	66	113	87	73	64	59	66	51
Compound SSRs	2 956	1 943	1 685	1 932	1 566	1 301	1 288	1 252	1 145	1 200

表3 玉米自交系 Mo17 各染色体 SSR 类型及数量

Table 3 Types and quantities of SSR loci on each chromosome of maize inbred line Mo17

项目 Item	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
Mono-nucleotide repeat	10 914	8 786	8 477	8 329	8 237	6 174	6 199	6 318	5 853	5 068
Di-nucleotide repeat	10 506	7 881	7 793	8 419	7 100	5 679	5 915	5 648	5 253	4 975
Tri-nucleotide repeat	6 381	5 231	4 988	5 268	4 748	3 516	3 691	3 724	3 386	3 226
Tetra-nucleotide repeat	495	437	428	390	384	276	281	256	252	262
Penta-nucleotide repeat	187	135	127	127	128	92	92	111	78	73
Hexa-nucleotide repeat	128	99	60	105	97	52	65	60	50	54
Compound SSRs	2 273	1 715	1 624	1 973	1 550	1 191	1 283	1 191	1 171	1 037



注:A;玉米自交系 B73 各染色体 SSR 类型及数量;B:玉米自交系 Mo17 各染色体 SSR 类型及数量。
Note: A, Types and quantities of SSR loci on each chromosome of maize inbred line B73; B, Types and quantities of SSR loci on each chromosome of maize inbred line Mo17.

图2 玉米自交系 B73 和 Mo17 各染色体长度及 SSR 位点数量比较

Fig.2 Comparison of chromosome length and number of SSR loci between maize inbred lines B73 and Mo17

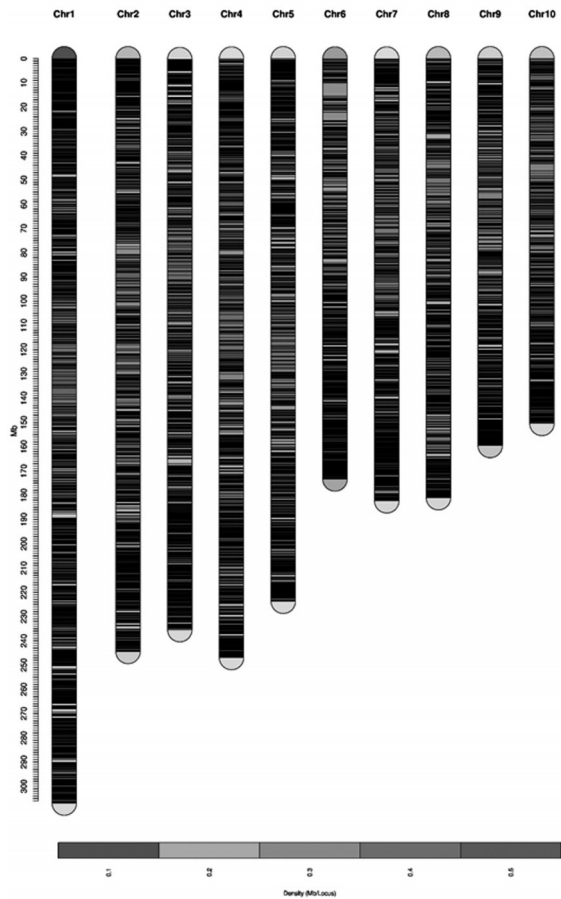
2.3 位点特异性扩增分析及多态性SSR标记筛选

对设计成功的SSR引物进行特异性扩增分析,分别在B73和Mo17中获得了平均每条染色体(6 921.10±1 808.04)和(6 813.90±1 765.09)个特异性扩增标记,两者平均每条染色体上共同存在(4 516.20±1 203.22)个特异性扩增SSR位点(表4)。SSR标记的冗余性分析结果发现,有446个与已报道的SSR位点存在冗余,剩余44 222个为新设计的SSR标记位点。比较SSR标记在两个自交系中的理

论扩增长度,长度差异绝对值≥3 bp、≥20 bp、≥50 bp的SSR特异性扩增位点,平均每条染色体有(1 737.50 ±509.45)、(397.30±116.62)和(143.50±41.72)个(表4、图3、图4)。长度差异绝对值≥3 bp的位点在10条染色体上的分布密度较高,平均每个位点占据的染色体长度在0.5 Mb以下,且这些位点在各染色体的上分布相对比较均匀,只在部分染色体上存在较大片段的空缺,如6号染色体短臂部分(图3)。

表4 位点特异性扩增与多态性分子标记的筛选
Table 4 Site-specific amplification and screening of polymorphic markers

项 目 Item	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	平均值 Mean value
B73特异性扩增位点	10 751	8 094	7 872	7 571	7 634	5 628	5 869	5 810	5 189	4 793	6 921.10±1 808.04
Mo17特异性性扩增位点	10 564	7 792	7 817	7 570	7 468	5 547	5 881	5 600	5 192	4 708	6 813.90±1 765.09
共有特异性位点	6 927	5 155	5 136	5 165	5 162	3 575	3 813	3 941	3 379	2 909	4 516.20±1 203.22
非冗余共有特异性位点	6 804	5 051	5 011	5 057	5 062	3 490	3 733	3 866	3 304	2 844	4 422.20±1 184.24
多态性位点(≥3 bp)	2 922	1 940	2 001	1 880	1 828	1 457	1 515	1 354	1 343	1 135	1 737.50 ±509.45
多态性位点(≥20 bp)	662	459	447	442	402	324	382	295	295	265	397.30±116.62
多态性位点(≥50 bp)	237	177	150	163	144	111	131	119	106	97	143.50±41.72



注:标尺为染色体物理距离(Mb);差值在3 bp及以上的多态性SSR位点用黑色横线标注在染色体上。

Note: The scale indicates the physical distance of chromosomes(Mb); the polymorphic SSR loci with a difference of ≥3 bp were marked with black horizontal lines on the chromosomes.

图3 差值在3 bp及以上多态性SSR位点的染色体分布
Fig.3 Chromosomal distribution of polymorphic SSR loci with a difference of ≥3 bp

2.4 PCR扩增及琼脂糖凝胶电泳验证

对 1 435 个长度差异绝对值 ≥ 50 bp 的多态性 SSR 标记,使用 MapChart 软件画图。结果显示,这些

SSR 特异性扩增位点在染色体上分布相对均匀,但部分区段尤其是着丝粒附近标记密度相对稀疏(图4)。在每条染色体短臂、长臂上各选择一对引物

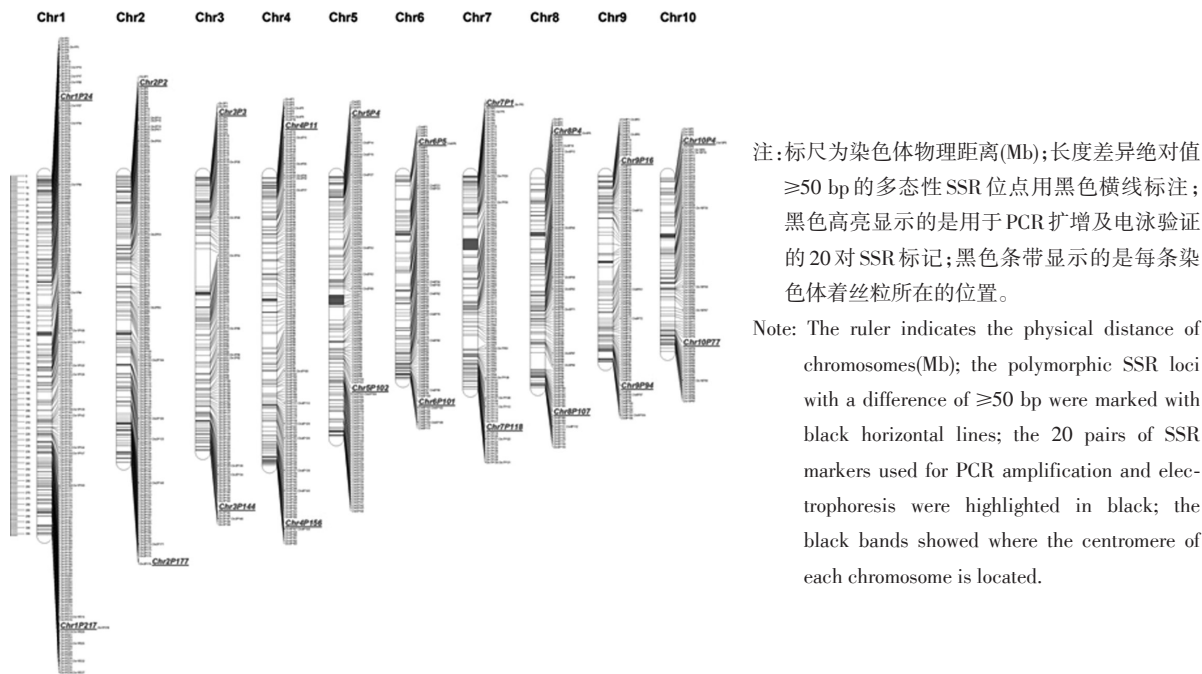
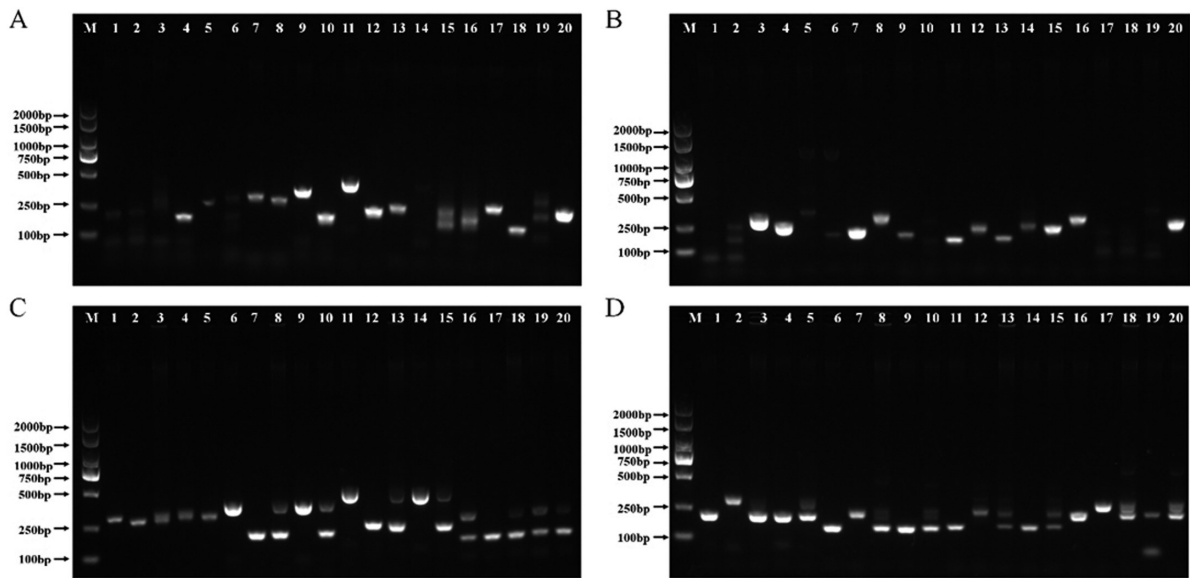


图4 差值在 50 bp 及以上多态性 SSR 位点的染色体分布
Fig.4 Chromosomal distribution of polymorphic SSR loci with a difference of ≥ 50 bp



注:A:Chr 1~Chr 5 染色体短臂、长臂各 1 对 SSR 引物扩增结果;B:Chr 6~Chr 10 染色体短臂、长臂各 1 对 SSR 引物扩增结果;C:A 图电泳结果中部分引物在群体样本中的扩增结果;D:B 图电泳结果中部分引物在群体样本中的扩增结果。验证的群体依次为 B73、Mo17、 F_1 (B73 $\times$ Mo17)、 F_1 (B73 $\times$ B104)、 F_1 (B104 $\times$ Mo17)。

Note: A, Amplification results of SSR primers on the short and long arms of Chr 1 to Chr 5 chromosomes; B, Amplification results of SSR primers on the short and long arms of Chr 6 to Chr 10 chromosomes; C, Electrophoresis results of some selected primers from A in the population samples; D, Electrophoresis results of some selected primers from B in the population samples. The populations used for verification are B73, Mo17, F_1 (B73 $\times$ Mo17), F_1 (B73 $\times$ B104), and F_1 (B104 $\times$ Mo17).

图5 多态性 SSR 引物的 PCR 扩增验证
Fig.5 PCR amplification and validation of polymorphic SSR markers

进行PCR扩增验证(表5)。除部分引物扩增效果不佳外,大部分引物在B73和Mo17两样本中的扩增条带存在明显差异(图5)。选择扩增条带明亮、长度差异明显的几对引物,在杂交F₁代群体中进行验证,显示出较好的基因型区分效果。

表5 用于PCR扩增验证的20对SSR标记在B73基因组中的信息

Table 5 The genomic information of 20 pairs of SSR markers used for PCR amplification and validation in B73 inbred

编 号	上、下游引物序列	染色体位置(bp)	SSR重复单位	理论扩增长度(bp)
No.	Upstream and downstream primer sequences	Chromosome location	SSR repeat units	Theoretical amplification length
1	5'-CACATGTGATATGTTAGCGGGT-3' 5'-AACGAGTACCCGAAAATGGATA-3'	Chr1: 12553090-12553280	(AT)10	190
2	5'-TCAAACCTAAGGGCAAAGTCTGAG-3' 5'-AATCAGTTTGTCATGATGCAG-3'	Chr1: 286999283-286999547	(TTA)5	264
3	5'-TGGGATACGTGGCTATCCTATG-3' 5'-TCAGGGTGTTATAGTCGCTGAA-3'	Chr2: 254032-254259	(TA)40	227
4	5'-TCTGCATGACAATGCTAGACCT-3' 5'-AGTTCATCACATCCAAGTGTGC-3'	Chr2: 244297474-244297826	(CTT)32	352
5	5'-ACCACCATGTTAACACCATGTC-3' 5'-GATGGTGAGATCTCTTTGGAGG-3'	Chr3: 2132722-2133064	(TC)7	342
6	5'-AAGAAAAGAGGGTGACTGATGC-3' 5'-ACTAAAGTGC GCGTTAGCTAGG-3'	Chr3: 230887862-230888307	(ATGC)5	445
7	5'-TTGAGAGAGCCAACATCACAGT-3' 5'-GTCAATCTTCCGAAGGTGTTTC-3'	Chr4: 8121773-8122009	(CCCCTC)5	236
8	5'-GGTGGTCTTTGAGGATTTCTTG-3' 5'-CCTATCTTAGGGATGGGTGATG-3'	Chr4: 11759529-11759929	(CG)6	400
9	5'-GAGGGGTTAACAGTTAGCGTTG-3' 5'-GCCACTTCTCGTTCTCGATACT-3'	Chr5: 1963575-1963853	(TTTGA)5	278
10	5'-GTTAGGGTCCAAGCCTAAAACC-3' 5'-GTCGTACCCTCTTTCACCAAAC-3'	Chr5: 166647943-166648284	(AG)7	341
11	5'-CATTGATCAGCTCTGTACCAGG-3' 5'-GAATCTTTTGAGGATGGTGTC-3'	Chr6: 4343716-4343921	(AT)45	205
12	5'-GCCTCACTAGTCTCGTGATTGA-3' 5'-ACGTCGTGTCTCCCATAGAACT-3'	Chr6: 169232207-169232512	(TCC)5	305
13	5'-GTGTTTAGGAGTGTGGGGGTA-3' 5'-TTTTATAGACATCCGGACCTGG-3'	Chr7: 309533-309849	(ACAT)8	316
14	5'-AAAGAAGAGAGAGGGAAATGCC-3' 5'-CATATATGCTCTGCGTGTTTCG-3'	Chr7: 169184507-169184716	(AG)41	209
15	5'-TTCCACTTCATGTCTTCCCT-3' 5'-TATACTTGTGGATAGCGGAGA-3'	Chr8: 11855755-11855935	(TA)50	180
16	5'-ATCAGCTTGTTCCTCACATGAA-3' 5'-AACTTTGTAAAGCAGGGGATGA-3'	Chr8: 171842786-171842934	(GA)46	148
17	5'-GCATGTTGCGATGCTAAATAGA-3' 5'-CGTGGATATTCTCTCAAAATGGT-3'	Chr9: 10838884-10839052	(TAA)28	168
18	5'-CGTTGTACACTCGTCTCGGAT-3' 5'-TGCTAATAAGCTTGCTTTGCTG-3'	Chr9: 148901961-148902183	(TC)35	222
19	5'-TGCTCTGACGATTCTCGAATAA-3' 5'-ACTATTGAAAAACAGGGCTCC-3'	Chr10: 4915077-4915274	(TA)80	197
20	5'-TACATCTAATGGGCTGATGCTG-3' 5'-GAAGAAGAATCCAAGCTCATGC-3'	Chr10: 131918813-131919133	(AG)22	320

3 结论与讨论

SSR 标记具有丰富的多态性和高度保守的物种特异性,在群体遗传多样性分析、基因定位、遗传连锁图谱构建、指纹图谱构建、种质鉴定、分子标记辅助选择等方面应用广泛。基于全基因组序列开发 SSR 标记,可显著提高开发效率,完善种质鉴定标准^[15-17],并拓宽应用范围。对于尚未建立有效种质鉴定标准的物种,可以利用已公布的基因组序列信息,或者通过高通量测序、高效筛选多态性 SSR 位点,开发高质量的分子标记^[18-20]。

3.1 全基因组多态性 SSR 引物筛选拓展了可选标记范围

随着高通量测序、单分子荧光测序技术的发展和应用,NCBI Genome、Ensembl(<https://asia.ensembl.org/index.html>)等数据库公布了大量物种全基因组序列信息。MaizeGDB 数据库中也已公布了玉米 B73、Mo17、B107 等 37 个自交系材料的全基因组序列信息。近 10 年来,SNP 标记已成为群体遗传学研究的首选标记,如通过 GWAS(Genome-Wide Association Study)在全基因组水平,开展大样本量、高密度遗传标记分型,从而寻找复杂性状相关的遗传基因^[21,22]。虽然 SNP 比 SSR 具有更为密集的优点,但其在常规分子生物学验证实验中仍然无法取代 SSR 标记。随着可利用的全基因组序列信息越来越丰富,在全基因组水平开发多态性 SSR 标记变得可行,但相关研究仍然比较匮乏。以玉米为例,MaizeGDB 中公布的 SSR 标记共有 2 034 个,且大部分是根据 EST、CDS 等转录本序列获得的^[23]。本研究在全基因组水平搜索玉米自交系 B73、Mo17 中的 SSR 位点,并通过 e-PCR 进行特异性扩增分析,有效筛选获得了 44 222 个新设计的 SSR 标记位点。比较 SSR 引物在两个自交系中的理论扩增长度,分别获得长度差异绝对值 ≥ 3 bp、 ≥ 20 bp、 ≥ 50 bp 的 SSR 特异性扩增位点 17 375、3 973 和 1 435 个。这些多态性 SSR 引物的筛选有效拓展了玉米可选标记的范围。

3.2 引物扩增特异性分析及扩增长度差异分析提高了标记筛选效率

与传统 SSR 标记开发方法相比,通过软件大批量搜索 SSR 位点、特异性扩增分析以及扩增长度差异分析,可有效降低后续验证的时间成本和实验成本。本研究在每条染色体短臂、长臂上各选择一对引物进行 PCR 扩增验证,除部分引物扩增效果不佳外,大部分引物在 B73 和 Mo17 两样本中的扩增条带存在明显差异。通过比较分析,筛选扩增长度差异

较大的 SSR 标记,使用普通的琼脂糖凝胶电泳即可快速实现基因分型,比使用聚丙烯酰胺电泳检测更为快速、高效。需要注意的是,虽然理论分析可以有效提高后续实验验证的成功几率,但由于数据库中公布的基因组序列,与各实验室保有的实验材料的基因序列之间存在一定差异,所以电泳验证结果与理论分析也并非完全一致。这种差异可能是由于各实验室在自交系自交保种过程中存在一定程度的基因型分离,并在后续保种过程中得以纯合造成的。所以,通过实验验证是开发 SSR 标记不可缺少的重要环节。

SSR 标记验证方便、稳定性高,未来仍将广泛适用于动、植物遗传和基因组研究。对全基因组 fasta 序列进行分析,可提高引物开发的效率,高效筛选多态性 SSR 标记,完善种质鉴定标准,并拓宽应用范围。此外,玉米全基因组 SSR 多态性引物筛选对其他物种 SSR 标记的开发具有较好的借鉴意义。

参考文献:

- [1] Vieira M L, Santini L, Diniz A L, et al. Microsatellite markers: what they mean and why they are so useful[J]. Genet Mol Biol, 2016, 39(3): 312-328.
- [2] Ellegren H. Microsatellites: simple sequences with complex evolution [J]. Nat Rev. Genet, 2004, 5(6): 435-445.
- [3] Ramsay L, Macaulay M, Cardle L, et al. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley[J]. Plant J., 1999, 17(4): 415-425.
- [4] Garcia F A A, Kido A E, Meza N A, et al. Development of an integrated genetic map of a sugarcane(*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases[J]. Theor. Appl Genet, 2006, 112(2): 298-314.
- [5] Jonah P M, Bello L L, Lucky O, et al. Review: the importance of molecular markers in plant breeding programmes[J]. Glob J. Sci. Front Res, 2011, 11(5): eV-vers1.
- [6] Souza L M, Gazaffi R, Mantello C C, et al. QTL Mapping of growth-related traits in a full-sib family of rubber tree(*Hevea brasiliensis*) evaluated in a sub-tropical climate[J]. PLoS One, 2013, 8(4): e61238.
- [7] Mason A S. SSR Genotyping. In: Batley J(ed) plant genotyping[M]. Springer, New York, 2015.
- [8] 李强,万建民. SSRHunter, 一个本地化的 SSR 位点搜索软件的开发[J]. 遗传, 2005, 27(5): 808-810.
Li Q, Wan J M. SSRHunter: Development of a local searching software for SSR sites[J]. Hereditas(Beijing), 2005, 27(5): 808-810. (in Chinese)
- [9] Wang X W, Wang L. GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing[J]. Front Plant Sci., 2016, 7: 1350.
- [10] Thiel T, Michalek W, Varshney R K, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-

- markers in barley(*Hordeum vulgare* L.) [J]. Theor. Appl. Genet., 2003, 106(3): 411–422.
- [11] Wang X W, Lu P, Luo Z P. GMATo: A novel tool for the identification and analysis of microsatellites in large genomes[J]. Bioinformatics, 2013, 9(10): 541–544.
- [12] 王 希,陈 丽,赵春雷. 利用MISA工具对不同序列进行SSR标记位点挖掘的探讨[J]. 中国农学通报, 2016, 32(10): 150–156.
- Wang X, Chen L, Zhao C L. Mining SSR molecular marker sites with MISA tool for different types of sequences[J]. Chinese Agricultural Science Bulletin, 2016, 32(10): 150–156. (in Chinese)
- [13] Untergasser A, Cutcutache I, Koressaar T, et al. Primer3—new capabilities and interfaces[J]. Nucleic Acids Res., 2012, 40(15): e115.
- [14] Rotmistrovsky K, Jang W, Schuler G D. A web server for performing electronic PCR[J]. Nucleic Acids Res., 2004, 32: W108–112.
- [15] 王风格,赵久然,郭景伦,等. 中国玉米新品种DNA指纹库建立系列研究 I. 玉米品种纯度及真伪鉴定中SSR技术标准实验体系的建立[J]. 玉米科学, 2003, 11(1): 3–6.
- Wang F G, Zhao J R, Guo J L, et al. Series of research on establishing DNA fingerprinting pool of Chinese new maize cultivars I. The establish of a standard SSR system fitting for maize cultivars' identification[J]. Journal of Maize Sciences, 2003, 11(1): 3–6. (in Chinese)
- [16] 赵久然,王风格,郭景伦,等. 中国玉米新品种DNA指纹库建立系列研究 II. 适于玉米自交系和杂交种指纹图谱绘制的SSR核心引物的确定[J]. 玉米科学, 2003, 11(2): 3–5, 8.
- Zhao J R, Wang F G, Guo J L, et al. Series of research on establishing DNA fingerprinting pool of Chinese new maize cultivars II. Confirmation of a set of SSR core primer pairs[J]. Journal of Maize Sciences, 2003, 11(2): 3–5, 8. (in Chinese)
- [17] 王风格,赵久然,余花娣,等. 中国玉米新品种DNA指纹库建立系列研究 III. 多重PCR技术在玉米SSR引物扩增中的应用[J]. 玉米科学, 2003, 11(4): 3–6.
- Wang F G, Zhao J R, She H D, et al. Series of research on establishing DNA fingerprinting pool of Chinese new maize cultivars III. Multiplex PCR applied in maize SSR amplification[J]. Journal of Maize Sciences, 2003, 11(4): 3–6. (in Chinese)
- [18] Yang T, Bao S Y, Ford R, et al. High-throughput novel microsatellite marker of faba bean via next generation sequencing[J]. BMC Genomics, 2012, 13: 602.
- [19] Al-Faifi S A, Migdadi H M, Algamdi S S, et al. Development of genomic Simple Sequence Repeats(SSR) by enrichment libraries in date palm[J]. Methods Mol Biol, 2017, 1638: 315–337.
- [20] Jian Y Q, Yan W Y, Xu J F, et al. Genome-wide simple sequence repeat markers in potato: abundance, distribution, composition, and polymorphism[J]. DNA Res., 2021, 28(6): dsab020.
- [21] Li H, Peng Z Y, Yang X H, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels [J]. Nat Genet, 2013, 45(1): 43–50.
- [22] Wang B B, Lin Z C, Li X, et al. Genome-wide selection and genetic improvement during modern maize breeding[J]. Nat Genet, 2020, 52(6): 565–571.
- [23] Sharopova N, McMullen M D, Schultz L, et al. Development and mapping of SSR markers for maize[J]. Plant Mol Biol, 2002, 48(5–6): 463–481.

(责任编辑:朴红梅)