

致病疫霉基因组序列微卫星的发掘与开发分析

张宏磊, 朱杰华, 胡珍珠, 杨志辉

(河北农业大学 植物保护学院, 河北 保定 071000)

摘要:利用 SciRoKo 软件, 确定了致病疫霉整个基因组及其不同区域完美型和复合型微卫星的数量, 并分析其分布规律。结果表明: 在致病疫霉基因组中共发现了 5 010 个完美型微卫星位点, 其出现频率平均为 26.35 个/Mb, 最偏好基元类型为 ACGG、AT 和 AAG; 外显子区域仅含 3~6 个碱基的微卫星基元, 其中 3 和 6 个碱基的微卫星基元占 93.68%; 复合型位点 275 个, 82.18% 为中断完美类型。通过 SciRoKo 软件对致病疫霉基因组不同区域的序列进行分析, 共获得了 430 个适合开发高多态性微卫星标记位点, 其中包括 83 个外显子编码区微卫星位点和 264 个非编码区位点; 分别对 32 个编码区和 105 非编码区的微卫星位点进行验证, 结果表明其开发成功率和多态性明显提高。

关键词:致病疫霉; 基因组; 微卫星

中图分类号:Q 813 **文献标识码:**A **文章编号:**1001-0009(2013)17-0110-05

植物病害的发生和防控与群体结构及遗传进化密切相关^[1], 而微卫星标记是研究生物遗传学、群体动态演替和多样性很好的分子标记。目前微卫星标记开发方法主要有 5 种, 即构建和筛选基因组文库法^[2]、富集微卫星文库法^[3-4]、ISSR-PCR 分离方法^[5]、结合 AFLP 技术的微卫星序列分离法^[6]以及 DNA 序列分析法。无论基于何种方法, 其快速获得微卫星位点及其两端序列是微卫星开发的一个根本问题。近年来, 随着 DNA 测序技术的进步和基因组计划的大力推进, 从 GenBank 中可以方便、快速地获得大量基因组序列。同时, 生物信息技术也产生了大量基于基因组序列的微卫星筛选, 如 SciRoKo 3.3^[7]、MISA^[8]、TRF^[9]、SSRIT^[10]和 Sputnik^[11]等。这为高通量获得微卫星及两侧翼序列提供了前所未有的机遇。

该课题组前期曾对致病疫霉基因组微卫星标记进行了开发, 从随机抽取的 35 个微卫星位点中仅发现有 4 个有效微卫星标记, 且其中 3 个无多态性。而在 30 个无效引物中, 有 28 对引物在 PCR 扩增中出现 3 个条带, 对于二倍体的致病疫霉来讲是由于基因组中存在大量重

复序列而导致该微卫星具有多个拷贝数, 还有 2 个引物对在 PCR 扩增中部分菌株无条带。因此, 针对致病疫霉含有 74% 的重复序列的基因组, 如何通过序列分析以提高多态性微卫星标记的开发效率是该研究拟解决的一个核心问题。

微卫星序列分析软件 SciRoKo 3.3 不仅能搜索完美和复合型微卫星, 且能统计、筛选微卫星两侧翼序列及引物设计功能, 这为分析微卫星及两侧翼序列是否能设计引物及是否为多拷贝提供了技术支持。此外, 微卫星变异率与其序列长度呈正相关^[12], 在对人类^[13]、水稻及其它生物^[14]微卫星开发研究中也表明微卫星序列总长度与多态性密切相关, 这也为致病疫霉高多态性标记开发提供了参考。该研究通过 SciRoKo 3.3 软件对致病疫霉基因组微卫星序列进行分析, 以期提高微卫星的开发效率、节约开发时间和经费投入, 同时也可作为高重复性基因组中微卫星开发提供可靠的分析方法。

1 材料与方法

1.1 致病疫霉基因组序列

自 *Phytophthora infestans* Database (<http://www.broadinstitute.org>) 下载致病疫霉基因组、基因编码区序列、外显子序列、flank5' 及 flank3' 序列信息。

1.2 基因间及内含子数据组的确定

已公布的致病疫霉序列信息中只注释了基因组、结构基因序列、外显子及结构基因两侧翼序列, 因此在统计基因间及内含子微卫星位点信息时, 其基因间数据

第一作者简介:张宏磊(1986-), 男, 河北唐山人, 硕士研究生, 研究方向为植物病理学。E-mail: bdzhl2013@163.com.

责任作者:杨志辉(1975-), 男, 河北藁城人, 博士, 副教授, 硕士生导师, 研究方向为植物病理学。E-mail: bdyzh2002@yahoo.com.cn.

基金项目:国家自然科学基金资助项目(31000833)。

收稿日期:2013-04-15

组由整个基因组与结构基因及其两侧翼数据组差值获得,内含子数据组由结构基因与外显子数据组差值获得。

1.3 微卫星位点搜索的参数设置

微卫星序列分析软件为 SciRoKo 3.3^[7]。对完美及复合型的微卫星位点搜索的参数设置:完美选项,最小重复数3,最短重复长度15 bp,2个微卫星位点最大距离设置为10 bp^[15-16]。此外,由于微卫星基元类型中 AC=CA=GT=TG,该研究除复合型微卫星要观察其左右微卫星重复基元的特征不标准化外,其余均标准化,即将4个类型合并为AC类型。

1.4 基因组内适合微卫星分子标记开发的位点界定及参数设置

适合微卫星标记开发的位点需满足3个条件。(1)多态性高:微卫星序列总长度 ≥ 20 bp^[10]。(2)PCR产物内只包含1个微卫星位点:相邻完美微卫星位点距离 ≥ 20 bp。(3)单拷贝:将基因组内相同基元类型的微卫星及两侧翼各200 bp序列用 Clustal X 软件进行序列比

对,选取不具同源性的序列,而这些非同源序列经过 GenBank 中引物特异性检测程序,进一步确定其是否为单拷贝。

2 结果与分析

2.1 基因组及不同区域完美微卫星分析

2.1.1 数量分析 由表1可知,在已公布的190.30 Mb的致病疫霉基因组内共发现完美微卫星5 010个,其中微卫星重复基元长度为1~6个碱基的个数依次为295、846、1 490、1 071、616和692,其中3个碱基基元重复的微卫星最多,占总数29.74%。较特殊的是外显子中含基元长度为3~6 bp的微卫星,且3、6 bp类型占该区域总数的93.68%,而4、5 bp类型的分别仅有7和16个。基因组不同区域微卫星密度不同,内含子区域最大为39.52个/Mb,其次是基因上、下游两侧翼及基因间,其频率分别为33.40个/Mb、31.87个/Mb及25.93个/Mb,而外显子区域频率最低为16.03个/Mb。内含子区域是外显子微卫星出现频率的2.47倍。

表1 基因组及不同区域完美微卫星的分布

Table 1 Loci and distribution of microsatellites in the genome and its different regions

区域划分及大小 Regions and size/Mb		单碱基 Mono	二碱基 Din	三碱基 Tri	四碱基 Tetra	五碱基 Penta	六碱基 Hexa	合计 Total
基因组 Genome (190.13 Mb)	数量 No./个	295	846	1 490	1 071	616	692	5 010
	频率 Frequency/个·Mb ⁻¹	1.55	4.45	7.84	5.63	3.24	3.64	26.35
	比例 Percent/%	5.89	16.88	29.74	21.38	12.30	13.81	
基因间区 Intergenic (126.44 Mb)	数量 No./个	223	653	760	926	319	398	3279
	频率 Frequency/个·Mb ⁻¹	1.76	5.16	6.02	7.32	2.52	3.15	25.93
	比例 Percent/%	6.80	19.91	23.18	28.24	9.73	12.14	
基因上游侧翼 Flank5' (17.99 Mb)	数量 No./个	29	75	211	68	140	78	601
	频率 Frequency/个·Mb ⁻¹	1.61	4.17	11.72	3.78	7.78	4.34	33.40
	比例 Percent/%	4.83	12.48	35.11	11.31	23.29	12.98	
基因下游侧翼 Flank3' (17.98 Mb)	数量 No./个	36	56	253	48	104	76	573
	频率 Frequency/个·Mb ⁻¹	2.00	3.11	14.07	2.67	5.78	4.23	31.87
	比例 Percent/%	6.28	9.77	44.15	8.38	18.15	13.26	
外显子 Exon (22.71 Mb)	数量 No./个	0	0	225	7	16	116	364
	频率 Frequency/个·Mb ⁻¹	0	0	9.91	0.31	0.7	5.11	16.03
	比例 Percent/%	0	0	61.81	1.92	4.40	31.87	
内含子 Intron (5.01 Mb)	数量 No./个	7	58	50	21	36	26	198
	频率 Frequency/个·Mb ⁻¹	1.40	11.57	9.97	4.19	7.18	5.19	39.52
	比例 Percent/%	3.54	29.29	25.25	10.61	18.18	13.13	

注:单碱基-单碱基重复,二碱基-2个碱基基元重复,三碱基-3个碱基基元重复,四碱基-4个碱基基元重复,五碱基-5个碱基基元重复,六碱基-6个碱基基元重复。

Note: Mono-monomucleotide, din-dinnucleotide, tri-trinucleotide, tetra-tetra nucleotide, penta-pentanucleotide, hexa-hexanucleotide.

2.1.2 微卫星基元分析 从表2可以看出,致病疫霉基因组中共发现微卫星基元类型272种,出现次数最多的4种基元重复类型为ACGG、AT、AAG及ACG,其数量依次是504、368、268及257个,总计1 397个,占基因组搜索发现5 010个完美微卫星的27.88%。将微卫星标准化以后,单碱基微卫星只有A和C2种类型,在基因组中单碱基类型微卫星C基元多于A基元,如基因组中C基元137个,而A基元仅86个。基因组不同区域较偏好微卫星基元类型不同(单碱基微

卫星除外),如在基因组中发现的504个ACGG和368个AT较偏好基元类型中,在基因间区的数量分别为481个和336个,分别占总数的95%和91%;而在基因组中发现的3个碱基重复的较偏好基因类型AAG,其总数为268个,基因区存在182个,占总数的70%,其中外显子出现数量最多共73个,占基因区该类型微卫星总数的40.11%。

2.2 基因组复合型微卫星分析

2个及2个以上完美微卫星之间的距离在10 bp及

表 2 基因组内不同区域较偏好的微卫星基元类型

Table 2 The perfect microsatellite motifs in the genome and its different regions

区域 Regions	单碱基 Mono	二碱基 Din	三碱基 Tri	四碱基 Tetra	五碱基 Penta	六碱基 Hexa	总计 Total
基因组 Genome	类型 Motifs/(个) ^a	C(174)	AT(368)	AAG(268)	ACGG(504)	ACAGT(70)	AACTC(92)
			AG(247)	ACG(257)	ACTC(194)	AAGCG(66)	ACCAGC(67)
		A(121)	AC(231)	AGC(242)	ACAG(115)	AAAAT(48)	AAGGAC(57)
				ACT(229)	AAAT(43)	ATGCC(30)	ACACGT(33)
基因间 Intergenic	类型 Motifs/(个) ^a	C(137)	AT(336)	ACG(174)	ACGG(481)	AAGCG(39)	AACTC(92)
			AG(198)	AGC(123)	ACTC(173)	ATGCC(25)	ACCAGC(59)
		A(86)	AC(165)	AAC(96)	ACAG(109)	AAAAT(20)	AAGGAC(55)
				AAG(86)	AAAT(25)	AAAAC(17)	ACACGT(28)
基因上游侧翼 Flank5'	类型 Motifs/(个) ^a	C(16)	AG(45)	ACT(78)	ACTC(13)	ACAGT(39)	ATGTAC(6)
			AC(19)	AAG(42)	ACGG(11)	AAAAT(11)	AAGAGC(4)
		A(13)	AT(11)	AGC(27)	ACCC(10)	AAGCG(10)	AAAACG(3)
				ACG(17)	AAGC(6)	AGAGC(3)	ACCAGC(3)
基因下游侧翼 Flank3'	类型 Motifs/(个) ^a	C(19)	AG(25)	ACT(101)	AAAT(9)	AAGCG(14)	ACAGCT(5)
			AT(16)	AAG(49)	ACGG(8)	ACAGT(13)	ATGTAC(4)
		A(17)	AC(15)	AGC(29)	AAAC(8)	AAAAT(9)	ACCAGC(4)
				ACG(26)	ACTC(6)	AAAAC(7)	ATCGTC(4)
外显子 Exons	类型 Motifs/(个) ^a			AAG(73)	ACGG(2)	ACCGC(2)	ACGAGG(9)
				AGC(55)	AAGC(2)	ACTCG(2)	ACAGCT(6)
				ACG(34)	AGCG(1)	AGCGC(2)	AACAGC(6)
				AGG(26)	ATCG(1)	AGCCG(1)	ATCGTC(5)
内含子 Introns	类型 Motifs/(个) ^a		AC(30)	AAG(18)	AAAC(4)	AAAAT(7)	ACAGCT(2)
			AG(22)	ACG(10)	ACGG(3)	ACAGT(4)	ATGTAC(1)
			AT(6)	ACG(6)	AAAT(3)	AAAAC(3)	ATCTGC(1)
				ACT(4)	ACTC(2)	AGCCG(2)	ATCTCG(1)

注:(个)^a:微卫星基元的数量。Note:(Counts)^a:The amount of motif.

以内被当做 1 个复合型微卫星,包括中断完美、组合及中断组合 3 种类型。在致病疫霉基因组内复合型微卫星在基因区域出现频率较低,仅占 18.18%,因此在对其复合型微卫星分析中,以致病疫霉基因组为整体。致病疫霉基因组中共发现复合型微卫星 275 个,共 47 种类型。由表 3 可知,其中,中断完美类型微卫星 226 个,共有 18 种类型,AT-AT 类型出现 124 个,占中断完美类型总数的 54.87%,其频率最大。其次,是 AGTG-AGTG、ACTC-ACTC、AG-AG 和 TC-TC,数量分别为 30、20、17、16 个,所占比例依次是 13.27%、8.85%、7.52% 和 7.08%。

表 3 基因组中断完美类型的微卫星类型及特征

Table 3 Types and characteristics of interrupt perfect microsatellite in the genome

中断完美类型 Type of interrupt perfect	数量 No./个	平均距离 Average distance/bp	频率 Frequency/个·Mb ⁻¹
AT-AT	124	3.13	0.65
AGTG-AGTG	30	3.20	0.16
ACTC-ACTC	20	3.90	0.11
AG-AG	17	2.35	0.09
TC-TC	16	3.13	0.08
ACT-ACT	4	4.00	0.02
AAC-AAC	2	2.00	0.01
ATCT-ATCT	2.00	4.50	0.01
AGT-AGT	2	3.00	0.01
...
合计 Total	226	3.17	1.185

由表 4 可知,中断组合型有 49 个,共有 29 种类型,其中出现频率最大的为 AAGT-AGTG、TCG-AGT 和 TG-AG,其数量分别为 7、6 和 5 个,在该类型中所占比例分别为 14.29%、12.24% 和 10.20%。在该类型中同一组合类型 2 个微卫星基元很相似,研究发现该类型中右侧基元都可以通过左侧基元碱基的替代、插入或缺失获得,如 AAGT-AGTG 类型中 AGTG 可通过 AAGT 的碱基 A 与碱基 G 的替换获得,TTC-TC 类型中 TC 可通过 TTC 类型中碱基 T 的缺失获得,C-AC 类型中 AC 可通过 C 类型中碱基 A 的插入获得,这与 Kofler 等^[15]在哺乳动物、鱼类、鸟类及昆虫 8 种生物中发现复合型微卫星中不同的基元重复类型相似的现象相同,因此,试

表 4 基因组中组合及中断组合微卫星类型及特征

Table 4 Types and characteristics of compound and interrupted compound microsatellites in the genome

组合及中断组合类型 Type of compound and interrupted one	数量 No./个	平均距离 Average distance/bp	频率 Frequency/个·Mb ⁻¹
AAGT-AGTG	7	3.0	0.04
TCG-AGT	6	1.0	0.03
TG-AG	5	2.2	0.03
ACAG-AC	3	6.0	0.02
ATAG-AG	2	8.0	0.01
TG-TCTG	2	3.5	0.01
TTG-TGC	2	1.0	0.01
...
合计 Total	49	2.76	0.26

验推测该菌组合型微卫星也大都起源于完美微卫星。

2.3 基因组内适合开发微卫星分子标记的位点分析

从表 5 可以看出,在致病疫霉基因组中发现的 5 010 个完美微卫星中,适合微卫星开发的 CLASSI 仅占 1 236 个。通过 PCR 对微卫星位点进行检测时,所需引物设计长度一般不少于 20 bp,为确保 PCR 产物仅含有单个微卫星位点,因此需去除微卫星位点间距 < 20 bp 的微卫星。在致病疫霉 CLASSI 类型微卫星中共发现 542 个间距 < 20 bp 的微卫星,至此符合微卫星开发的位点还有 694 个。再对 694 个含有微卫星位点的序列进行比对分析,其中发现含多拷贝微卫星位点 264 个。通过上述 3

表 5 基因组及不同区域内适宜开发微卫星标记的位点

Table 5 The suitable loci as microsatellite markers in the genome

区域 Regions	类型 I CLASSI/个	完美-20 Perfect-20/个	多拷贝 Multi-copy/个	适合开发微卫星标记 Suitable for microsatellite markers/个	频率 Frequency/个·Mb ⁻¹	所占比例 Percent/%
基因间 Intergenic	850	406	251	155	1.23	18.23
基因上游侧翼 Flank5'	142	96	7	89	4.95	62.68
外显子 Exons	85	84	1	83	3.65	97.65
内含子 Introns	29	26	2	24	4.79	82.76
基因下游侧翼 Flank3'	130	82	3	79	4.39	60.77
合计 Total	1 236	694	264	430	2.26	34.79

2.4 微卫星多态性验证

自基因组中获得 430 个适合开发的微卫星位点中,从编码的外显子中随机选取了 32 个,而非编码区域选取了 105 个微卫星位点,进行引物开发和 PCR 验证。在 32 个编码区中,仅有 1 个位点未扩增出任何 PCR 产物,而其余 31 个均可扩增出 1~2 条清晰的条带,其微卫星开发成功率为 96.88%,而多态性的比率为 54.84%。而非编码区随机选取的 105 个微卫星位点中,55 个位点可扩增出 1~2 清晰条带,其开发成功率为 52.38%,较未分析前的 11.43% 提高了 4.58 倍,其多态性的微卫星占 60.0%。而非编码区的微卫星在验证中仍存在多个条带主要是因为目前公布的基因组覆盖度不够,仍有大量重复序列未公布,因此在序列分析中无法将其去除。有关上述微卫星的具体验证数据及引物序列见后续文章。

3 讨论

该研究对致病疫霉基因组及不同区域完美和复合型微卫星进行了全面的发掘与分析,确定了该病原菌基因组中不同类型微卫星的数量及其分布规律,并筛选出了适合该病原菌微卫星分子标记开发的位点。然而,在对致病疫霉基因组微卫星开发过程中发现了 2 个值得关注的科学问题,即 2 编码区域为何出现基元长度非 3 或 6 bp 的微卫星及如何增加致病疫霉微卫星标记开发的成功率。

3.1 编码区微卫星特征分析

外显子序列属于编码区,理论上基元长度为 1、2、4 和 5 bp 的微卫星的出现频率会极低,因为此类微卫星重

步筛选自 5 010 个微卫星位点中最终获得了 430 个适合高多态性微卫星位点,其中外显子中适合微卫星标记开发的位点占该区域微卫星位点总数的比例最大为 97.65%,其次是内含子、基因上游及下游侧翼,其所占比例分别为 82.76%、62.68% 和 60.77%,基因间区域最小仅占 18.23%。这表明在编码区及其邻近区域开发微卫星标记有效率高,而非编码区相对较低。此外,还发现较偏好的基元类型微卫星,出现多拷贝的频率较大,以基元 ACGG、AT 和 AAG 类型微卫星为例,发现有 85% 以上为多拷贝,因此选用较偏好微卫星开发标记时,应谨慎考虑。

复单元数目的变化会导致基因发生移码突变,造成其所编码的蛋白功能改变或缺失,对生物存活和演化造成不利影响^[17]。然而,在致病疫霉外显子区域发现了基元长度为 4 和 5 bp 的微卫星,同时双色腊蘑外显子序列(微卫星搜索条件为基元长度 1~6 bp,重复数阈值依次为 [16,5,5,5,5,5])^[18] 及在灵长类、啮齿类、节肢动物、有胚植物及一些真菌的外显子序列中(微卫星搜索条件为其序列总长度 ≥ 12 bp)^[14] 均发现了 1、2、4 和 5 bp 的微卫星。推测其原因可能是(1)不同物种编码区微卫星阈值不同^[12]; (2)所搜索出的基元长度为 1、2、4 和 5 bp 的微卫星为“假微卫星”; (3)基元长度为 1、2、4 和 5 bp 的微卫星发生 DNA 复制滑移的个体,由于其编码蛋白功能的缺失,导致死亡,而未被检测,其原因还需进一步确定。

3.2 基因组内适合开发微卫星标记的位点

为了提高致病疫霉基因组中有效且多态性较高的微卫星分子标记开发率,将微卫星筛选条件设置为:完美微卫星,其序列长度 ≥ 20 bp; 相邻微卫星最小间距为 20 bp; 单拷贝。同等长度条件,完美微卫星变异率大于非完美微卫星^[19],微卫星序列长度大于等于 20 bp 的微卫星为高变量标记微卫星^[10],且其长度越大越易形成较稳定和复杂的非 B-DNA 构象增加滑移突变率^[12],同时考虑同一基因座内尽量避免含 2 个及以上微卫星位点,例如,(AG)₁₄ TC(ATCG)₈。在做基因型分析时,不同菌株间 AG 重复基元减少 2 个重复与 ATCG 重复基元减少 1 个重复,其扩增片段长度相等,会造成试验结果的失真,结合引物设计的是适宜长度范围,确定了致病疫霉基因组内适合微卫星标记开发的位点的前 2 个条件。然而,

致病疫霉基因组内重复序列比其它疫霉较多^[20],这就可能会增加出现多拷贝微卫星出现的几率,利用这些位点开发微卫星标记可能会出现 1 个菌株内同一基因座多等位基因的假象,影响试验结果的分析。因而在满足其高变异率及特异性的基础上,还要保证其单拷贝。但由于目前致病疫霉基因组只公布 1 条且不完全的局限性,该研究无法确定一些微卫星位点两侧翼序列是否保守以及在未公布序列中是否存在其它拷贝。因此,筛选出的是一个较大的范围,至于适合做微卫星标记的位点还需进一步开发验证。

参考文献

- [1] Goodwin S B. The population genetics of *Phytophthora*[J]. Phytopathology, 1997, 87(4): 462-473.
- [2] Rassmann K, Schlötterer C, Tautz D. Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting[J]. Electrophoresis, 1991, 12(2-3): 113-118.
- [3] Ostrander E A, Jong P M, Rine J, et al. Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences[J]. Proceedings of the National Academy of Sciences, 1992, 89(8): 3419-3423.
- [4] Kandpal R P, Kandpal G, Weissman S M. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers[J]. Proceedings of the National Academy of Sciences, 1994, 91(1): 88-92.
- [5] Zietkiewicz E, Rafalski A, Labuda D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification[J]. Genomics, 1994, 20(2): 176-183.
- [6] Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation; a review[J]. Molecular Ecology, 2002, 11(1): 1-16.
- [7] Kofler R, Schlötterer C, Lelley T. SciRoKo: a new tool for whole genome microsatellite search and investigation[J]. Bioinformatics, 2007, 23(13): 1683-1685.
- [8] Thiel T, Michalek W, Varshney R, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)[J]. Theoretical and Applied Genetics, 2003, 106(3): 411-422.
- [9] Benson G. Tandem repeats finder: a program to analyze DNA sequences[J]. Nucleic Acids Research, 1999, 27(2): 573-580.
- [10] Temnykh S, Decker G, Lukashova A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential[J]. Genome Research, 2001, 11(8): 1441-1452.
- [11] Blanchette M, Kent W J, Riemer C, et al. Aligning multiple genomic sequences with the threaded blockset aligner[J]. Genome Research, 2004, 14(4): 708-715.
- [12] Bhargava A, Fuentes F F. Mutational dynamics of microsatellites[J]. Molecular Biotechnology, 2010, 44(3): 250-266.
- [13] Weber J L. Informativeness of human (dC-dA)n • (dG-dT)n polymorphisms[J]. Genomics, 1990, 7(4): 524-530.
- [14] Cho Y G, Ishii T, Temnykh S, et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.)[J]. Theoretical and Applied Genetics, 2000, 100(5): 713-722.
- [15] Kofler R, Jan Bartos J, Gong L, et al. Development of microsatellite markers specific for the short arm of rye (*Secale cereale* L.) chromosome 1[J]. Theoretical and Applied Genetics, 2008, 117(6): 915-926.
- [16] 李小白, 张明龙, 崔海瑞. 油菜 EST 资源的 SSR 信息分析[J]. 中国油料作物科学, 2007, 29(1): 20-25.
- [17] Kelkar Y D, Eckert K A, Chiaromonte F, et al. A matter of life or death: How microsatellites emerge in and vanish from the human genome[J]. Genome Research, 2011, 21(12): 2038-2048.
- [18] Labbe J, Murat C, Morin E, et al. Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers[J]. Current Genetics, 2011, 57(2): 75-88.
- [19] Jin L, Macaubas C, Hallmayer J, et al. Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence[J]. Proceedings of the National Academy of Sciences, 1996, 93(26): 15285-15288.
- [20] Haas B J, Kamoun S, Zody M C, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*[J]. Nature, 2009, 461(7262): 393-398.

Detection and Analysis of Microsatellites in the *Phytophthora infestans* Genome

ZHANG Hong-lei, ZHU Jie-hua, HU Zhen-zhu, YANG Zhi-hui

(Department of Plant Pathology, Agricultural University of Hebei, Baoding, Hebei 071000)

Abstract: The amount and distribution of pure and compound microsatellite in the whole genome and its different regions of *Phytophthora infestans* were identified by software SciRoKo. The results showed that, a total of 5 010 loci for perfect microsatellites were detected in the genome, whose density was 26.35 SSRs/Mb, and the perfect motifs were ACGG, AT and AAG. Only motif for 3 to 6 bases occurred in the exons, in which the percent of microsatellites with 3 and 6 bases motif accounted for 93.68%. The number of compound microsatellites were 275, accounting for 82.18% was interrupted perfect microsatellites. The loci suitable for microsatellite markers were identified according to the software of SciRoKo. At last, 430 loci suitable for the development of microsatellites were acquired, in which the coding regions were 83 and non-coding regions were 264. In the coding region and non-coding regions, which were respectively selected 32 and 105 microsatellite loci for validation. The results showed that the success rate of its development and polymorphism significantly increased. This provided a reliable method for the development of microsatellite in genome of high repetition.

Key words: *Phytophthora infestans*; genome; microsatellite