

黄瓜叶绿体基因组全序列微卫星 分布特征与标记开发

胡建斌 李建吾* 梁芳芳 刘 颖

(河南农业大学园艺学院, 郑州 450002)

摘要 对黄瓜叶绿体基因组 (chloroplast genomic DNA, cpDNA) 全序列中微卫星 (chloroplast simple sequence repeat, cpSSR) 的分布特征进行了全面分析。结果发现, cpSSR 主要分布在 25~40 kb 和 100~130 kb 两个区域, 主要以完全重复和间隔重复为主。单碱基重复最多, 占总数的 79.3%, 二碱基重复占 10.1%, 三碱基和三碱基以上重复仅占 10.6%。单碱基重复主要以 (A)_n 和 (T)_n 为主, 二者共占总数的 78.8%, 二碱基重复主要是 (AT)_n, 占总数的 9.3%。在重复次数分布上, 完全重复的 (A)_n 和 (T)_n 均以最低重复次数为主, 而间隔重复的 (A)_n 和 (T)_n 以 14 次或 15 次为多。设计了 6 对引物, 对 7 份黄瓜自交系 cpDNA 进行扩增。结果发现, 6 对引物均有扩增产物, 每对引物平均能检测到 2.2 个等位基因。几乎所有引物都能在甜瓜、西瓜、葫芦和栝楼中通用, 5 对引物在 1~4 种远缘种中有扩增产物。上述结果表明, 从黄瓜 cpDNA 中发掘 cpSSR 标记是一条可行的途径。

关键词 黄瓜; 叶绿体基因组; 微卫星; 标记

微卫星序列 (microsatellite or simple sequence repeat, SSR) 不仅存在真核生物的核基因组中, 在叶绿体基因组中也有广泛分布。Powell 等^[1]首先在扭叶松、樟子松和黑松中筛选到一批叶绿体微卫星 (chloroplast simple sequence repeat, cpSSR) 标记, 并成功用于松科植物的种质资源研究。与核基因 DNA 相比, 叶绿体基因组 (chloroplast genome DNA, cpDNA) 呈单亲遗传模式, 而且具有分子质量小、多拷贝、结构简单等特点, 这非常有利于以 cpSSR 标记对叶绿体基因组进行遗传分析^[2]。此外, cpSSR 标记还兼有核基因组 SSR 标记的共显性、超变异性、多等位性等特点, 可在更高的分类水平上鉴定植物种质的亲缘关系, 描述物种区域和个体水平的遗传差异。

cpSSR 技术与核基因组 SSR 技术原理相同, 均可通过构建基因组文库、筛选和测序获得, 但 cpSSR 标记的开发更为困难, 因为从植物组织提取高纯度的 cpDNA 十分困难。充分利用 cpSSR 侧翼序列的保守性、跨物种开发 cpSSR 引物是一条经济、便利的途径^[3]。例如, 从栎树中开发的 cpSSR 引物可用于大部分壳斗科植物和部分槭树科、桦木科和蔷薇科植物^[4]。目前, 烟草、水稻、玉米、黑松等高等植物 cpDNA 全序列已被测出, 并在网上公开释放, 有望开发出大量的 cpSSR 标记。

黄瓜 (*Cucumis sativus*) 是葫芦科重要的蔬菜作物

之一, 在世界各地均有栽培。黄瓜遗传背景狭窄^[5], 因而常规的分子标记难以在育种中提供大量有效的遗传信息, 不利于有用资源的挖掘。随机扩增多态 DNA (random amplified polymorphic DNA, RAPD)、扩增片段长度多态性 (amplified fragment length polymorphism, AFLP)、SSR 等标记已用于黄瓜资源研究, 然而 cpSSR 标记的开发与应用还未见报道。黄瓜 cpDNA 全序列是国际葫芦科基因组计划的研究成果之一^[6], 本研究通过分析黄瓜 cpSSR 在 cpDNA 中的分布特征, 开发 cpSSR 标记, 并进行通用性研究。

1 材料与方法

1.1 cpSSR 的搜索

2008 年 5 月 20 日从美国国家生物技术信息中心网站 (<http://www.ncbi.nlm.nih.gov/entrez>) 以 FASTA 格式下载黄瓜叶绿体基因组全序列 (DQ865975), 利用 “Tandem repeats finder” (<http://tandem.bu.edu/trf/trf.html>) 查找单碱基微卫星, 多碱基微卫星的搜索使用 “简单重复序列识别工具” SSRIT (<http://www.>

收稿日期: 2008-09-08 接受日期: 2008-12-04

国家高技术研究发展计划 (863 计划) (No.2007AA10Z100)、河南省农业科技成果转化项目 (No.072201110017) 和河南农业大学博士启动基金项目 (No.30400247) 资助

* 通讯作者。Tel: 0371-63554959, E-mail: lijw555@sohu.com

gramene.org/db/searches/ssrtool)。搜索条件为: 含有完全重复的一、二和三碱基的完全重复微卫星最小重复数分别为 10、5 和 4, 四碱基和四碱基以上的微卫星最小重复次数均为 3, 间隔重复和复合重复的微卫星单碱基最小重复数为 5, 二碱基和二碱基以上最小重复数均为 3。间隔重复和复合重复的具体定义见 Jarne 等^[7]的报道。

1.2 cpSSR 引物开发

利用 Primer 3.0 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)对包含有 cpSSR 的序列进行引物设计, 引物配对序列避免微卫星的出现。引物设计主要参数为: 引物长 18~22 bp, 最适 20 bp; 引物退火温度 50~60 ℃, 左右引物退火温度之差在 5 ℃以内; GC% 含量 30%~70%, 最适 50%; PCR 预期产物长 100~300 bp。尽量避免引物二级结构的出现。

1.3 cpSSR 引物筛选与通用性分析

7 份黄瓜(*Cucumis sativus*)自交系[分别包含 3 份华北型、3 份华南型黄瓜和 1 份近缘野生种(西双版纳黄瓜)(*C. sativus* var. *xishuangbannanensis*)]用于 cpSSR 引物的筛选。此外, 4 种葫芦科植物[甜瓜(*C. melo*)、西瓜(*Citrullus lanatus*)、葫芦(*Lagenaria siceraria*)和栝楼(*Trichosanthes kirilowii*)]和 4 种远缘植物[水稻(*Oryza sativa*)、玉米(*Zea mays*)、小白菜(*Brassica campestris* ssp. *chinensis*)和番茄(*Lycopersicon esculentum*)]用于 cpSSR 引物通用性检测。所有材料 DNA 提取采用 CTAB 法^[8]。PCR 扩增反应体系及参数参考 Cheng 等^[9]的报道。扩增产物用含有 0.5 μg/

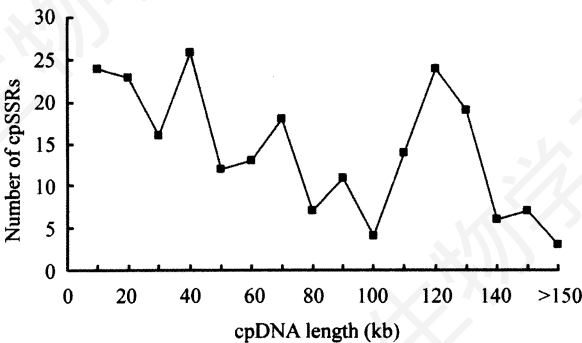


Fig. 1 Distribution of cpSSRs in cpDNA of cucumber

ml 溴化乙锭的 3% 琼脂糖胶凝胶电泳、分离。

2 结果

2.1 叶绿体 cpSSR 的分布

黄瓜 cpDNA 全序列长 155 525 bp, 共搜索出 227 个微卫星, 其序列总长度为 1 993 bp (不包括间隔重复中的间隔序列), 占基因组全长的 1.73%。cpSSR 的分布范围较广, 从 25 bp 至 155 500 bp 均有分布, 但分布密度并不均匀, 有集中区与稀疏区之分。黄瓜 cpSSR 主要集中分布在两个区域, 即 25~40 kb 和 100~130 kb, 这两个区域总长度占基因组全长的 45.0%, 分别分布有 90 个和 59 个 cpSSR, 占微卫星总数的 65.6%。80~100 kb 和 130 kb~ 序列末端是微卫星分布稀疏区, 两个区域共分布着 34 个 cpSSR, 仅占总数的 15.0% (图 1)。

2.2 重复基元的分布

Table 1 Repeat motifs in cpDNA of cucumber

	Repeat types	Major repeat motifs	Number of cpSSRs	Percentage (%)
Complete repeats	Mononucleotides	A (26), T (27)	54	23.8
	Dinucleotides	AT (8)	10	4.4
	Trinucleotides	AAT (3)	5	2.2
	Tetranucleotides	AAAT (4), AATT (2)	9	4.0
	Pentanucleotides	AAAAT (2)	4	1.8
	Total		82	36.1
Interrupted repeats	Mononucleotides	A (53), T (69)	122	53.7
	Dinucleotides	AT (12)	12	5.3
	Trinucleotides	AAT (4)	4	1.8
	Tetranucleotides	AAAT (1), AATT (1)	2	0.9
	Total		140	61.7
Compound repeats	Mononucleotides	A (1), T (3)	4	1.8
	Dinucleotides	AT (1)	1	0.4
	Total		5	2.2
Grand total			227	100.0

Note: The number of cpSSRs related to the motifs was bracketed.

在搜索到的 227 个 cpSSR 中, 完全重复的微卫星有 82 个, 间隔重复的微卫星 140 个, 复合重复的微卫星仅有 5 个, 三者分别占 cpSSR 总数的 36.1%、61.7% 和 2.2%, 这说明黄瓜叶绿体基因组中的 cpSSR 主要以完全重复和间隔重复为主。共发现单碱基微卫星 180 个(包括完全、间断和复合重复), 占有微卫星总数的 79.3%, 且以不完全重复为主(122 个, 占 53.7%)。二碱基微卫星 23 个, 占总数的 10.1%。三碱基和三碱基以上的微卫星共有 24 个, 占 10.6%。A 和 T 是单碱基微卫星的主要重复单元, 分别占 35.2% 和 43.6%, C 只出现了 2 次, G 则没有出现。二至五碱基微卫星共包括 13 种单元。其中, 二碱基以 AT (21 个, 占 9.3%) 为主, 三碱基以 AAT (7 个, 占 3.1%) 为主, 四碱基以 AAAT (5 个, 占 2.2%) 为主, 五碱基微卫星均为完全重复, 其中 AAAAT 出现了 2 次, 占总数的 0.9% (表 1)。

2.3 重复次数分布

基序重复次数的变异引起位点长度的变化是产生 cpSSR 多态性的主要原因。图 2 和图 3 揭示了黄瓜 cpDNA 中主要重复单元 A 和 T 重复次数的分布情况。在单碱基 A 微卫星中, 完全重复次数从 10 到 17 不等, 且主要以 10 次重复为主(9 个), 随着重复次数的增加 ($(A)_n$ 数量逐渐减少, $(A)_{17}$ 仅有 1 个; 间隔重复次数的分布呈倒“V”字型, 14 次重复的微卫星数量最多(9 个), 随着重复次数的增多或减少, $(A)_n$ 数量均为下降(图 2)。微卫星 $(T)_n$ 重复次数的分布情况与 $(A)_n$ 十分相似(图 3), 完全重复数最多为 16 次, 间隔重复数最多为 20 次。完全重复 $(T)_{10}$ 数量为 11, 间隔重复的 $(T)_n$ 在 14 次和 15 次上分布最多, 均为 10 个。此外, 二至五碱基微卫星绝大部分都以最低重复数为主。

2.4 cpSSR 引物的开发

利用 primer 3.0 软件对随机挑选的单碱基 $(A)_n$ 和 $(T)_n$ 以及二碱基 $(AT)_n$ 共 15 个微卫星进行引物设计, 由于微卫星侧翼序列过短、GC 含量偏高或偏低等原因, 只获得 6 对引物。通过对 7 份黄瓜材料进行 PCR 扩增, 结果发现, 6 对引物均有扩增产物, 产物大小为 100~300 bp, 基本符合预期大小, 说明多态性主要是由 cpSSR 长度变化所引起。每对引物能检测到 1~3 个等位基因, 平均 2.2 个。除 CC04 外, 其他引物均能检测到 2 个或 2 个以上的等位基因(表 2)。图 4 为引物 CC03 在 7 份黄瓜材料中的扩增结果。

2.5 cpSSR 的通用性

以 6 对 cpSSR 引物对 4 种近缘种和 4 种远缘种

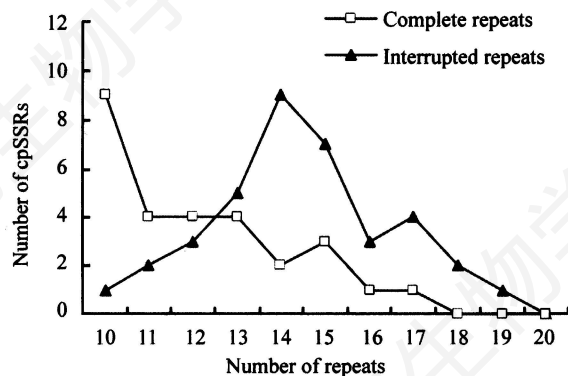


Fig. 2 Distribution of the repeat number of mononucleotide A

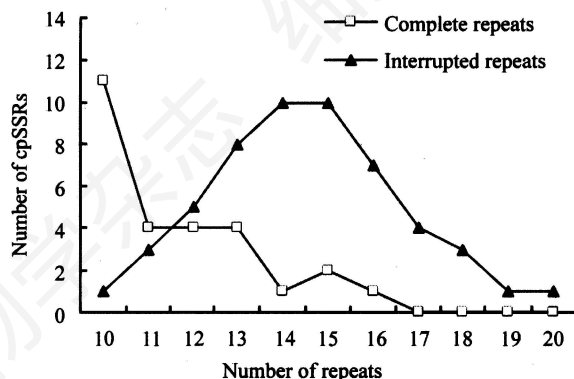


Fig. 3 Distribution of the repeat number of mononucleotide T

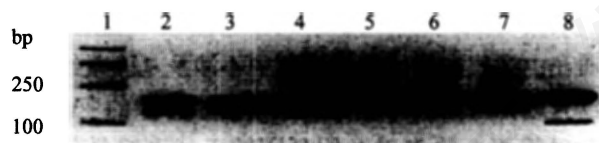


Fig. 4 Band pattern generated by primer CC03

1: DNA marker DL2000; 2-4: north China ecotype; 5-7: south China ecotype; 8: wild relative species (*C. sativus* var. *xishuangbannanensis*).

分别进行 PCR 扩增, 以研究引物的通用性。各引物的扩增结果见表 3。CC01~CC04 和 CC06 在 4 种葫芦科植物中均能扩增出带谱, CC05 在除葫芦以外的 3 种葫芦科植物中有扩增产物, 显示了良好的种间通用性。CC03 在所测试的 4 种远缘种中均能扩增出带谱, 其通用性最强。引物 CC02、CC04、CC05 和 CC06 在 1~3 种远缘种中具有通用性。CC01 只对葫芦科物种通用, 而在远缘种中没有扩增产物。

3 讨论

cpDNA 具有保守性、单亲遗传等特点, 遗传物

Table 2 The cpSSR primers with amplified fragments in cucumber

Primer code	Repeat motifs	Primer sequences (5' - 3')	Expected product size (bp)	Annealing temperature (°C)	Number of alleles
CC01	(A) ₁₁	AAAATGGGCAATCCTGAG ATTCCTTTACTAACGCAACG	117	53	2
CC02	(T) ₁₁	CGTTACGGAAAGGAGGAAGA GAGACAATCGGAAGATGGGA	194	55	2
CC03	(A) ₁₆	AAGACCAAAGAACATTACAT AAGGGAAATAGACTACTCAAA	148	50	3
CC04	(TA) ₆	CGCCTCCTTTATTACAGTTCTAT TCCGTATTGATTTCGTTTCCAT	273	55	1
CC05	(A) ₁₃ ...(A) ₇	CAGGATCAAACCTATGGGACT GGAATCTGGGCTCTTCTACG	112	55	2
CC06	(T) ₉ ...(T) ₇	CAAAATCAACCGAGGTAATGTA ATAAACGAGGGTGTGGATAACT	226	55	3

Note: CC05 and CC06 represent the primer pairs flanking compound repeats. “...” means interrupted bases.

Table 3 Transferability of the cpSSR primers to other plant species

Primer code	Melon	Watermelon	Gourd	Snakegourd	Rice	Maize	Pakchoi	Tomato
CC01	P (1)	P (1)	P (2)	P (2)	N	N	N	N
CC02	P (2)	P (1)	P (2)	P (1)	P (1)	P (1)	N	N
CC03	P (2)	P (2)	P (1)	P (2)	P (1)	P (2)	P (1)	P (1)
CC04	P (1)	P (1)	P (1)	P (1)	N	N	P (1)	N
CC05	P (1)	P (2)	N	P (1)	P (1)	P (1)	N	N
CC06	P (2)	P (1)	P (2)	P (1)	P (1)	P (1)	P (1)	N

N: No amplification; P: Amplification with allele number in bracket. Materials: melon “yilishabai”; watermelon “yuyixinmoyu”; gourd “guoguo’hulu”; Snakegourd represented by primitive species; rice “fangxinyihao”; maize “yudan998”; pakchoi “rekangqing”; tomato “fendounvhuang”.

质在由亲代向子代传递过程中不涉及基因重组,因此,没有选择压力。即使因突变等因素产生的变异,也不会受重组、缺失以及假基因的干扰而容易保存^[10]。因此, cpDNA 有独立的进化路线,具有 cpDNA 特征的 cpSSR 技术在植物群体结构分析、种群分类、物种演化等研究领域有明显的优势。由于采用构建文库的方法开发 cpSSR 标记费时费力,我们采取了与 EST-SSR 标记开发相似的方法^[11],从公共数据库中开发 cpSSR 标记。目前, GenBank 中存放有大量的高等植物 cpDNA 部分序列和部分植物的 cpDNA 完全序列,这是开发 cpSSR 标记的宝贵资源。

明确 cpSSR 在 cpDNA 中的分布规律,是开发 cpSSR 标记的首要工作。cpSSR 贯穿黄瓜整个 cpDNA 序列,但分布不均匀。对黄瓜叶绿体基因组全序列分析发现, cpSSR 主要分布于黄瓜 cpDNA 的大单拷贝区(large single-copy region, LSC)和小单拷贝区(small single copy sequence, SSC),即两个集中区,而这两个区域也是叶绿体为数不多的基因主要分布区。在两个反向重复序列(inverted repeat, IR)中

cpSSR 分布甚少^[6]。基因组进化研究认为,新基因的产生源于重复序列, LSC 和 SSC 分布较多的 SSR 位点,可能是它们比 IR 区变异度大的原因之一^[12]。

黄瓜 cpSSR 较为简单,主要以低重复的单碱基为主,其中 A 和 T 是绝对优势重复单元,与其他高等植物十分相似^[13,14]。在所发现的 21 个二碱基重复中, (AT)_n 出现了 19 次,而高等植物中核基因组 SSR 单元类型较多,例如, (AG)_n、(AT)_n、(AAG)_n 等都是常见的重复类型^[15]。在所发现的高级单元重复中, (AAT)_n、(AAAT)_n 等出现频率相对较多,这可能也与 (A)_n 和 (T)_n 位点变异的结果有关。cpSSR 这种低丰度、以单碱基重复为主的分布特点,可能与 cpDNA 本身的特点和进化方向有关。

cpDNA 序列相当保守,其进化速率仅为核基因组的 1/5^[12],因此, cpSSR 标记在近缘种甚至远缘种中具有很好的通用性^[4,16]。我们设计的 6 对 cpSSR 引物几乎均可在 4 种近缘种中通用, 5 对引物可在 1~4 种远缘种中通用,显示了良好的通用性。这无疑扩大引物使用范围、节约引物开发成本,对于一些研

究基础薄弱的植物的 cpSSR 标记开发具有很好的启示。尽管 cpSSR 标记通用性好, 但多态性比核基因组 SSR 标记低。例如, Watcharawongpaiboon 等^[17]从黄瓜 cDNA 文库中开发的 SSR 引物平均每对能检测到 3.6 个等位基因, 最大数达 7 个。而本研究中平均每对引物只能检测到 2.2 个等位基因, 多态性明显低于核基因组 SSR 标记, 除了 cpDNA 保守性外因素之外, 本研究中的测试材料较少, 可能降低了某些等位基因的出现频率。对于 cpSSR 位点多态性产生机制, 目前还存在争论, 有人认为与非间隔重复数目无直接关系, 而 Weising 等^[16]认为可能是 A 或 T 的单核苷酸重复引起的长度多态性, 也可能是微卫星处在变异度相对较大的非编码区。因此, 他们建议在 cpSSR 引物设过程中, 引物对中应至少有一条引物落在非编码区, 以提高标记的多态性。

通过对黄瓜 cpSSR 分布规律与应用的研究, 不仅加深了对叶绿体进化的认识, 还有助于对葫芦科甚至其他科属植物 cpSSR 标记的开发, 以进一步促进植物种质资源的研究。目前, 包括黄瓜在内的葫芦科 cpSSR 的开发工作在我们实验室已经全面展开。

参考文献(References)

- [1] Powell W, Morgante M, McDevitt R, *et al.* Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines, *Proc Natl Acad Sci USA*, 1995, 92(17): 7759-7763
- [2] Powell W, Morgante M, Andre C, *et al.* Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome, *Curr Biol*, 1995, 5(9): 1023-1029
- [3] Olmstead RG, Palmer JD. Chloroplast DNA systematics: a review of methods and data analysis, *Am J Bot*, 1994, 81: 1205-1224
- [4] Deguilloux MF, Dumolin-Lapègue S, *et al.* A set of primers for the amplification of chloroplast microsatellites in *Quercus*, *Mol Ecol Note*, 2003, 3(1): 24-27
- [5] Dijkhuizen A, Meglic V, Staub JE, *et al.* Linkages among RFLP, RAPD, isozyme, disease-resistance, and morphological markers in narrow and wide crosses of cucumber, *Theor Appl Genet*, 1994, 89(1): 42-48
- [6] Chung SM, Gordon VS, Staub JE. Sequencing cucumber (*Cucumis sativus* L.) chloroplast genomes identifies differences between chilling-tolerant and -susceptible cucumber lines, *Genome*, 2007, 50(2): 215-225
- [7] Jarne P, Lagoda PJJ. Microsatellites, from molecules to populations and back, *Trends Ecol Evol*, 1996, 11(10): 424-429
- [8] Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue, *Focus*, 1990, 12(1): 13-15
- [9] Cheng YJ, Guo WW, Deng XX. cpSSR: a new tool to analyze chloroplast genome to *Citrus* somatic hybrids, *Acta Bot Sin*, 2003, 45(8): 906-909
- [10] Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution, *Trends Ecol Evol*, 2001, 16(3): 142-147
- [11] Kong Q, Xiang C, Yu Z. Development of EST-SSRs in *Cucumis sativus* from sequence database, *Mol Ecol Note*, 2006, 6(4): 1234-1236
- [12] Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs, *Proc Natl Acad Sci USA*, 1987, 84(24): 9054-9058
- [13] Bryan GJ, McNicoll J, Ramsay G, *et al.* Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants, *Theor Appl Genet*, 1999, 99(5): 859-867
- [14] Ishii T, McCouch SR. Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species, *Theor Appl Genet*, 2000, 100(8): 1257-1266
- [15] Wang Z, Weber JL, Zhong G, *et al.* Survey of plant short tandem DNA repeats, *Theor Appl Genet*, 1994, 88(1): 1-6
- [16] Weising K, Gardner RC. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms, *Genome*, 1999, 42(1): 9-19
- [17] Watcharawongpaiboon N, Chunwongse J. Development and characterization of microsatellite markers from an enriched genomic library of cucumber (*Cucumis sativus*), *Plant Breed*, 2008, 127(1): 74-81

Characterization of Microsatellites in Complete Chloroplast Genome of Cucumber (*Cucumis sativus*) and Marker Exploitation

Jian-Bin Hu, Jian-Wu Li*, Fang-Fang Liang, Ying Liu

(College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China)

Abstract Characterization of microsatellites in complete sequence of chloroplast genome (cpDNA) of cucumber (*Cucumis sativus*) was analyzed in the round. The results showed that chloroplast simple sequence repeats (cpSSR) were mainly distributed in the two regions of cpDNA, 25–40 kb and 100–130 kb, most of the cpSSRs belonging to complete repeats and interrupted repeats. Of these cpSSRs, mononucleotide repeats were maximum types with the frequency of 79.3%, and dinucleotide repeats accounted for 10.1%, and all of the trinucleotide repeats and those repeats over trinucleotide in length only accounted for 10.6%. (A)_n and (T)_n were major repeats in mononucleotides, totally accounting for 78.8%. (AT)_n was dominant dinucleotide repeats with frequency of 9.3%. As for repeat number, most of the complete repeats of (A)_n and (T)_n had the lowest repeat number, while 14 or 15 repeats were major for interrupted repeats of (A)_n and (T)_n. Six primer pairs were designed and used for PCR amplification of the cpDNA of 7 cucumber inbred lines. The results showed that all primer pairs had amplified fragments in cucumber materials and each of the primers detected mean 2.2 alleles. All of the primer pairs were almost transferable to melon, watermelon, gourd and *Trichosanthes kirilowii*. Five primer pairs could successfully amplify in 1–4 distantly related species. These results indicated that exploitation of cpSSR markers from cucumber cpDNA is practicable.

Key words cucumber (*Cucumis sativus*); chloroplast genome; microsatellites; makers

Received: September 8, 2008 Accepted: December 4, 2008

This work was supported by the National High Technology Research and Development Program of China (863 Program) (No. 2007AA10Z100), the Henan Agricultural Science and Technology Achievements Transformation Fund Program (No.072201110017) and the Scientific Research Foundation for Doctoral Program of Henan Agricultural University (No.30400247)

*Corresponding author. Tel: 86-371-63554959, E-mail: lijw555@sohu.com