

大豆基因组中的微卫星标记^{*}

刘 峰 陈受宜

(中国科学院遗传研究所植物生物技术实验室, 北京 100101)

摘 要

微卫星 DNA 是一种简单重复序列 (Simple Sequence Repeat, SSR), 其核心单位由 2-5 个核苷酸组成, 两侧一般是保守序列。由于它具有共显性, 多态性高, 可进行 PCR 扩增分析, 既简单又经济, 因此是一种很有价值的分子标记。实验证明大豆的微卫星 DNA 随机分布于基因组中, 其核心单位主要是 (AT)_n, (ATT)_n 在人类基因组中占很大比例的 (CA)_n 则很少在大豆中出现。平均每一个微卫星座位有 7-10 个等位基因, 最高可达 26 个。大豆的微卫星标记可扩充现有的 RFLP 图谱, 广泛应用于基因型鉴定, 基因和 QTL 分析, 分子标记辅助育种和家系分析等。

关键词 微卫星标记; 简单序列重复; 等位基因多样性; 分子图谱

前 言

微卫星 DNA 是一种简单重复序列 (SSR), 其核心单位由 2-5 个核苷酸组成, 两侧一般是保守序列。由于它具有共显性, 多态性高, 可进行 PCR 扩增分析, 既简单又经济, 因此是一种很有价值的分子标记。微卫星标记最早是在人类中发现^[17-38], 现已证明其广泛分布在真核生物中, 如哺乳动物^[22], 鸟类^[9], 鱼类^[16], 昆虫^[33], 酵母^[33] 和许多单子叶和双子叶植物中^[24-37]。微卫星座位的最大优点之一在于等位基因的多样性, 这点对于遗传标记来说非常有用。根据微卫星重复单位两侧的保守序列设计特异性引物, 通过 PCR 扩增微卫星的等位基因^[38]。由于微卫星标记是基因组中的特异性座位, 是一种最富信息和多态性的序列标记位点 (STS), 因此又被称为序列标记微卫星位点 (sequence-tagged microsatellite sites, STMS)^[5]。等位基因的差异 (又称简单序列长度多态性, SSLP), 通常是由于微卫星重复单位的数目不同产生的。这种多态性可通过 PCR 分析, 在高分辨率琼脂糖凝胶或聚丙烯酰胺胶上很容易检测, 是一种共显性标记。微卫星标记的一个缺点是得到

* 收稿日期 1998-04-24

This paper was received on April 24, 1998.

有用的简单序列重复引物花费较高,筛选到阳性克隆后要进行测序,然后再合成特异性引物。然而,一旦座位确定,引物序列公开后,这些标记就可以广泛交流使同行受益。另外,使用测序胶检测等位基因大小差异一般认为是比较费事的。但近年来运用琼脂糖凝胶电泳分析的报道越来越多,并且发展了许多检测等位基因大小的技术。Vignal等(1993)报道在一个泳道里可连续加样,同时检测 10–20 个微卫星座位^[36]。Ziegler等(1992)发明了荧光标记和自动检测 PCR 片段大小的技术^[42]。Diwan等利用该方法对大豆的遗传变异进行了研究^[14]。使用单泳道和多泳道系统毛细管电泳,能快速分离 PCR 片段并检测其大小^[10]。ISSRs(inter-simple sequence repeat polymorphisms)也是一种基于微卫星 DNA 的分子标记,其使用重复序列锚定引物可以扩增出微卫星重复序列之间片段长度的多态性。ISSRs 标记也能检测品种之间的遗传多样性^[43]。由于不需要事先知道靶 DNA 的序列,所以 ISSRs 标记在某些方面比 SSR 标记更为容易些^[35],但 ISSRs 标记检测多态性的效率比微卫星标记低得多。Nagaoka等(1997)用 100 对引物在两个小麦材料中分析发现只有 33 个能检测到多态性。产生的 49 个 ISSR 片段只有 9 个(19%)能定位在现有的 RFLP 连锁图上^[26]。

人们已对许多高等植物的微卫星 DNA 进行了初步研究,如大豆^[2,3],水稻^[1,28,39,41],大麦^[4],小麦^[30],玉米^[4],番茄^[7],葡萄^[34],林树^[11,15],向日葵^[8],十字花科品种^[6,29]等。水稻中已定位了 120 个微卫星标记^[28],而大豆中发现的微卫星标记还较少^[3]。一般认为微卫星重复序列两端的特异性序列在属内是保守的,从而为属内不同作物的比较基因组研究打下了基础。现在已发现在亲缘关系比较远的物种之间也能找到共用的微卫星标记(本实验室未发表的数据)。在人类和小鼠中已构建了由微卫星标记构建的高度饱和的遗传图谱^[12,13]。随着微卫星标记覆盖率的增加,其作为遗传标记的价值也日益明显。当务之急是要快速发展重要农作物的微卫星标记的遗传图谱。

1 大豆的微卫星标记

Akkaya等(1992)首先对大豆中微卫星 DNA 进行了研究,证明其不仅大量存在并且具有高水平的多态性^[2]。发现在大豆基因组中,(AT)_n和(ATT)_n比例较高,而在人类基因组中丰富的(CA)_n微卫星 DNA 含量较少。用 6 对 SSR 引物对 43 个大豆品种进行了 PCR 扩增,其中三个位点在每个品种里只有一条带,每个座位平均有 7 个等位基因。F₂ 代分析证明其是共显性分离,符合孟德尔遗传,并估算每 40kb 就有一个(AT)_n微卫星。

Morgante和 Olivieri(1993)已从报道的大豆 DNA 序列中发现了丰富的二核苷酸和三核苷酸重复序列^[24]。Maughan等(1995)用 5 个微卫星引物在 94 个样品中检测到了 79 个等位基因^[23]。F₂ 分离分析表明其中 4 个定位在 4 个独立的位点,每个位点的等位基因数目在 5–21 之间(平均 15.8 个)。野生品种中的等位基因多样性高于栽培品种。他们认为简单序列重复的多态性是由于减数分裂时的错配和不平等交换造成了微卫星座位中串联重复单位数目的不同。Jiang等(1995)用三个(AT)_n,4 个(ATT)_n在 96 个大豆材料中进行了分析,结果表明每个座位有 11–26 个等位基因,只有两个品种有相同的微卫星等位基因带型,调查发现这两个品种有相似的家系。在 96 个材料中,7 个标记的基因多样性平均为 0.87,远远高于大豆的 RFLP 标记,与人类微卫星标记的平均值相似^[19]。

Akkaya等(1995)^[3]用 40 个微卫星标记在近等基因系 F₂ 代的 60 个单株中的研究表

明,其中 34个微卫星标记定位在 29个连锁群中的 18个,与 9个传统性状,两个同工酶标记连锁。一般来说微卫星随机分布在大豆基因组中,但也发现有两个集中区域,分别包含 5个座位(覆盖 23.4cM)和 4个座位(覆盖 33.6cM) Bell和 Ecker(1994)^[6]在拟南芥中定位了 30个微卫星座位,发现 5条染色体中有三条的微卫星座位比预期要少。此外还发现两个集中区域,各包含有 4个微卫星座位,尽管表面上有些偏离随机分布,但他们认为在得出随机或非随机分布结论之前应用更多的标记进行定位。人类基因组中,微卫星标记,主要是 (CA)_n,其分布也不尽相同,从 21号染色体的 0.28个标记/Mb到 17号染色体的 0.75个标记/Mb。最近, Broun和 Tanksley(1996)在番茄中发现 (GA)_n, (ATT)_n, (GATA)_n成簇集中在着丝粒附近^[7]。

在大豆中用 RFLP整合遗传图谱有两个困难。一是大豆中的 RFLP座位的等位基因很少多于两个。由于这些等位基因具有不对称的频率,即 $P > 0.9$, $Q < 0.1$,任何两个基因型在特定 RFLP座位具有多态性的可能性很小^[20]。特别对于由改良的大豆品种创建的作图群体而言。Shoemaker和 Specht(1995)的图谱中的 365个 RFLP标记,其中只有 118个在 Akkaya等(1995)的作图群体中具有多态性^[32]。二是大豆的 RFLP中通常为多个 DNA条带模式,其原因可能是由于大豆来源于古四倍体^[18]。在一个群体中某一条带呈现多态,而在另一群体中则是不同的条带或新加的条带。这样在确定 RFLP座位时,不仅要根据使用的探针和限制性内切酶,还要考虑分离条带的分子量。所以在用 RFLP标记比较来源于大豆不同作图群体的遗传图谱时非常困难。而微卫星标记的出现则很容易解决这一问题。

2 大豆微卫星标记的发展

微卫星 DNA可从 GenBank等数据库中查询那些富含两个或三个核苷酸简单重复序列(SSR)的序列,选择那些大于或等于 20bp(即 10个二核苷酸的重复单位)的 DNA片段。微卫星 DNA也可用菌落杂交从基因组文库中筛选得到^[31]。基因组文库的 DNA片段大小在 400–600bp之间,可用各种内切酶或机械剪切获得。³²P标记的寡聚核苷酸探针(oligo-CT, oligo-AT和 oligo-ATT)与膜杂交,洗膜时根据杂交片段的种类和大小相应调整温度和时间,然后压 X光底片曝光。第一轮筛选得到的阳性克隆经纯化后再用含有相应 SSR序列的寡聚核苷酸探针进行第二轮筛选确证^[3]。含有微卫星 DNA的阳性克隆经测序后根据两端保守序列设计特异性引物。由于靶序列中 (G-C)含量少导致 T_m 值较低,一般为 $53.5 \pm 1.0^\circ\text{C}$ 。合成引物时一般要选择相同或相似的 T_m 值,这样对大量位点可使用一套标准的 PCR扩增程序。新合成的引物进行 PCR验证,扩增产物为 100–300bp,多于一条带的引物不用。

3 大豆微卫星标记的应用

3.1 遗传图谱的构建 由于微卫星标记是基因组中的特异性位点,是一种共显性遗传的 STS标记,所以很容易在不同组合的遗传图谱间进行标记的转移。在遗传图谱上定位的 STS标记是沟通基因组遗传图谱和物理图谱的中介。Olson等(1989)认为使用 STS标记可使人类基因组作图标准化^[27]。

3.2 基因和 QTL分析 微卫星标记和其它 PCR标记如 AFLP, RAPD等相结合可快速构建遗传图谱,并能与已有的 RFLP图谱整合。日益表明,微卫星标记和许多有用基

因及具有重要农艺性状的 QTL 连锁。Yu 等 (1994) 发现微卫星标记 SM176 与大豆花叶病毒抗性基因 *Rsv* 紧密连锁, 遗传距离只有 0.5 cM^[40]。Mudge 等 (1997) 报道有两个微卫星标记与大豆抗孢囊线虫主效座位连锁, 较近的为 2 个 cM^[25]。Lin 等 (1997) 在大豆中发现微卫星标记 SAT335 与控制缺铁黄化症的主效 QTL 紧密连锁^[21]。通过与目的基因或 QTL 主效基因紧密连锁的微卫星标记, 可直接在 BAC 库中筛选 BAC 克隆构建重叠群, 进而为最终克隆基因打下基础。

3.3 品种鉴定和种质保存 尽管 RFLP 也可用于鉴定品种差异, 但由于其多态性不高, 每个 RFLP 座位的等位基因数目较少, 单个大豆 RFLP 探针提供的信息量很少。由于微卫星有高水平的多态性, 能提供更多信息, 所以对发展大豆材料的独特 DNA 指纹图谱特别有用。从而能更有效地对具有优良性状的种质资源进行长期保存。

3.4 家系分析和标记辅助育种 微卫星标记能揭示丰富的等位基因多样性, 区分亚种之间的差异, 确定品种间的相关程度, 并且能可靠地示踪大豆家系中单基因或 QTL 的走向。它提供的信息可以提高亲本的选择效率和品种发展。

参 考 文 献

- [1] Akagi H, Yokozeki Y, Inagaki A. and Fugimura T. 1996, Microsatellite DNA markers for rice chromosomes. *Theor Appl Genet* 93 1071- 1077
- [2] Akkaya M S, Shoemaker RC, Specht JE, Bhagwat AA, and Cregan PB. 1992. Length polymorphism of simple sequence repeat DNA in soybean. *Genetics* 132 1131- 1139
- [3] Akkaya M S, Bhagwat AA, and Cregan PB. 1995, Integration of simple sequence repeat DNA markers into a soybean linkage map. *Crop Sci* 35 1439- 1445
- [4] Becker J, Heun M. 1995, Barley microsatellite Allele variation and mapping *Plant Mol Biol* 27 835- 845
- [5] Beckmann JS. and Soller M. 1990. Toward a unified approach to the genetic mapping of eukaryotes based on sequence- tagged microsatellite sites. *Bio/Technology* 8 930- 932
- [6] Bell CJ, Ecker JR. 1994, Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. *Genomics* 19 137- 144
- [7] Broun P, Tanksley SD. 1996. Characterization and genetic mapping of simple repeat sequence in the tomato genome. *Mol Gen Genet* 250, 1 39- 49
- [8] Brunel D. 1994. A microsatellite marker in *Helianthus annuus* L. *Plant Mol Biol* 24 397- 400
- [9] Cheng HH, Crittenden LB. 1994, Microsatellite markers for genetic mapping in the chicken. *Poult Sci* 73 539- 546
- [10] Clark SM, Mathies RA. 1993. High- speed parallel separation of DNA restriction fragments using capillary array electrophoresis. *Anal Biochem* 215 163- 170
- [11] Condit R, Hubbell SP. 1991. Abundance and DNA sequence of two- base repeat regions in tropical tree genome. *Genome* 34 66- 71
- [12] Dib C, Faure S, Fizames C, Samson D, Drouot N, Vigal A, Millasseau P, Mare S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, and Weissenbach J. 1996, A comprehensive genetic map of the human genome based on 5264 microsatellites *Nature* 380 152- 154
- [13] Dietrich W F, Miller JC, Steen RG, Merchant M, Damronboles D, Husain Z, Dredge R, Daly M J, Ingalls KA, O'Connor T J, Evans CA, De Angelis NM, Levinson DM, Kryglyak L, Goodman N, Copeland N G, Jenkins N A, Hawkins T L, Stein L, Page DC, and Lander ES. 1996, A comprehensive genetic map of the

- mouse genome Nature 380 149– 152
- [14] Diwan N and Cregan PB. 1997. Automated sizing of fluorescent- labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor Appl Genet 95 723– 733
- [15] Dow BD Ashley MV. Howe HF. 1995. Characterization of highly variable (GA- CT)_n microsatellite in the bur oak *Quercus macrocarpa*. Theor Appl Genet 91 137– 141
- [16] Estoup A. Presa P. Krieg F. Vaiman D. and Guyomard R. 1993. (CT)_n and (GT)_n microsatellite A new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity 71 488– 496
- [17] Hamada H. Kakunaga T. 1982. Potential Z- DNA forming sequences are highly dispersed in the human genome. Nature 298 396– 398
- [18] Hymowitz T. and Singh RJ. 1987. Taxonomy and Speciation, p 23– 48 In Wilcox RJ(ed). Soybeans Improvement, production, and uses, 2ed. Agron Monogr. 16. ASA- CSSA- SSSA. Madison, WI
- [19] Jiang RW, Akkaya MS, Bhagwat AA, Lavi U. and Cregan PB, 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90 43– 48
- [20] Keim P. Beavis W. Schupp J. and Freestone R. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor Appl Genet 85 205– 212
- [21] Lin S, Gianzio S. and Shoemaker R. 1997. Mapping genetic loci for iron deficiency chlorosis in soybean. Molecular Breeding 3 219– 229
- [22] Love JM. Knight AM. Mcleer MA. and Todd JA. 1990. Towards construction of a high resolution map of the mouse genome using PCR- analyzed microsatellites. Nucl Acids Res 18, 14 4123– 4130
- [23] Maughan PJ. Saghai Maroof MA, and Buss GR 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean Genome 38 715– 723
- [24] Morgante M. Olivieri AM. 1993. PCR- amplified microsatellites as markers in plant genetics. Plant J 1 175– 182
- [25] Mudge J Cregan PB. Kenworthy JP. Kenworthy WJ. Olf JH. and Young ND. 1997. Two Microsatellite Markers That Flank the Major Soybean Cyst Nematode Resistance Locus. Crop Sci 37 1611– 1615
- [26] Nagaoka T. and Ogihara Y. 1997. Applicability of inter- simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theor Appl Genet 94 597– 602
- [27] Olson M. Hood L. Cantor D. and Bostein D. 1989. A common language for physical mapping of the human genome. Science 245 1434– 1435
- [28] Panaud O, Chen X. and McCoudh SR 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.) Mol Gen Genet 252 597– 607
- [29] Poulsen GB. Kahl G. Weising K. 1993. Abundance and polymorphism of simple repetitive DNA sequences in *Brassica napus* L. Theor Appl Genet 85 994– 1000
- [30] Roder M. Plaschke J. Koenig SU. Boerner A. Sorrells ME. Tanksley SD. Galal MW. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Gene 246 327– 333
- [31] Sambrook J. Fritsch EF and Maniatis T. 1989. Molecular Cloning A Laboratory Manual. Second Ed. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY.
- [32] Shoemaker RC and Specht JE. 1995. Integration of the Soybean Molecular and Classical Genetic Linkage Groups. Crop Sci, 35 436– 446
- [33] Tatuz D. Ranz M. 1984. Simple sequences are ubiquitous repetitive components of eukarotic genomes. Nucl Acids Res 12 4127– 4138
- [34] Thomas MR, Scott NS. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites(STSs). Theor Appl Genet 86 986– 990
- [35] Tsumura, Ohbak, and Strauss SH. 1996. Diversity and inheritance of inter- simple sequence repeat polymorphisms in Douglas- fir (*Pseudotsuga menziesii*) and Sugi (*Cryptomeria japonica*). Theor Appl Genet

92 40– 45

- [36] Viganl AG. Gyapay G. Hazan J. Nguyen S. Dupraz C. Cheron N. Becuwe N. Tranchant M. Weissenbach J. 1993. Nonradioactive multiplex procedure for genotyping of microsatellite markers. p 211– 221. In Adolph KW: *Methods in molecular genetics- Gene and Chromosome analysis- Part A*. Academic Press, New York.
- [37] Wang Z. Weber JL. Zhong G. Tanksley SD. 1994, Survey of plant short tandem DNA repeats. *Theor Appl Genet* 88 1– 6
- [38] Weber JL. May PE. 1989, Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44 388– 396
- [39] Wu K-S, Tanksley SD. 1993. Abundance, Polymorphism and genetic mapping of microsatellites in rice. *Mol Gen Genet* 241 225– 235
- [40] Yu YG. Saghai Maroof MA. Buss GR, Maughan DJ and Tolin SA. 1994, RFLP and Microsatellite Mapping of a Gene for Soybean Mosaic Virus Resistance. *Phytopathology* 84, 1 60– 64
- [41] Zhao X, Kochert G. 1992. Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.) *Mol Gen Genet* 231 353– 359
- [42] Zegle JS. Su Y, Corcoran KP, Nie L, Mayrand PE, Hoff LB, McBride LJ. Kronick MN and Diehl SR 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* 14 1026– 1031
- [43] Zekiewicz E, Rafalski A. Labuda. 1994, Genome fingerprinting by simple sequence repeat (SSR) – anchored polymerase chain reaction amplification. *Genomics* 20 176– 183

MICROSATELLITE MARKERS IN THE SOYBEAN GENOME

Liu Feng Chen Shouyi

(*Plant Biotechnology Laboratory, Institute of Genetics,
Chinese Academy of Sciences, Beijing 100101*)

Abstract

Microsatellites are simple, tandemly repeated di- to penta-nucleotide sequence motifs flanked by inique sequences. They are valuable as genetic markers because they are co-dominant, detect high levels of allelic diversity and are easily and economically assayed by PCR. Experiments reveal that microsatellites are distributed throughout the soybean genome. The most abundant microsatellite motifs reported in soybean are (AT)_n and (ATT)_n, while (GA)_n is most abundant in the human genome. There are 7– 10 alleles at a single SSR locus, up to 26 alleles /locus. SSR markers can be expected to complement existing RFLP maps, are useful for genotype identification, gene and QTL analysis, marker-assisted selection in breeding and pedigree analysis.

Key words Microsatellite marker; Simple sequence repeat; Allelic diversity; Molecular mapping