

从绵羊 UniGene 数据库中筛选 微卫星标记的初步研究

闫秋良^{1,2}, 张英汉¹, 李宏滨², 魏彩虹², 杜立新²

(1 西北农林科技大学 动物科技学院, 陕西 杨凌 712100; 2 中国农业科学院 北京畜牧兽医研究所, 北京 100094)

[摘 要] 为了利用生物信息学方法筛选绵羊微卫星标记, 利用 SSRIT 和 SSRFinder 软件对绵羊 UniGene 数据库的 4 081 个簇进行搜索, 筛选绵羊 EST-SSRs 标记。结果表明, 从绵羊 UniGene 数据库中搜索到微卫星 136 个, 含有微卫星的序列 121 个, 占整个 EST 序列数据库的 3.0%, 其中双碱基重复 47 个, 三碱基重复 54 个, 四碱基重复 3 个, 五碱基重复 4 个, 六碱基重复 28 个。在这些微卫星序列中, AC/TG 重复在双碱基类型中最丰富, CTG 重复在三碱基类型中最常见, 分别占双碱基和三碱基微卫星序列总数的 64% 和 29%。根据筛选到的微卫星序列设计并合成引物 30 对, 在设计的 30 对引物中, 20 对引物有扩增产物, 且条带清晰, 其中有 2 对引物在小尾寒羊品种内呈现多态。

[关键词] 绵羊; 表达序列分析; 微卫星; SSR

[中图分类号] S826.2

[文献标识码] A

[文章编号] 1671-9387(2007)11-0015-04

Analysis of microsatellite markers from sheep UniGene

YAN Qiu-liang^{1,2}, ZHANG Ying-han¹, LI Hong-bing², WEI Cai-hong², DU Li-xin²

(1 College of Animal Sciences and Technology, Northwest A & F University, Yangling, Shaanxi 712100, China;

2 Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100094, China)

Abstract: In this study, SSRs markers were developed by bioinformatics methods and a total number of 4 081 sheep unigene sequences were downloaded. These UniGene sequences were screened for the presence of perfect microsatellites by using SSRIT and SSRFinder soft tools. A total number of 136 SSRs were identified from 121 EST sequences. The frequency of EST containing SSR is 3.0%. The trinucleotide repeat motif is the most abundant SSR, accounting for 39.71%, followed by 34.56% for dinucleotide repeats. Among the dinucleotide repeats, AC/TG was the most abundant repeat motif, accounting for 64% of all dinucleotide repeats. While CTG was the most abundant type in the trinucleotide repeat types, accounting for 29%. 30 primer pairs were designed from microsatellite and 20 pairs could show clear PCR products by electrophoresis for further polymorphism analysis. The result in this paper provided a basis for the development and further application of EST-SSR markers in sheep.

Key words: sheep; expressed sequence tags; microsatellite; simple sequence repeat

微卫星 (Microsatellites) 又称简单序列重复 (Simple sequence repeat, SSR), 是指以 1~6 个核苷酸为基本单元的串联重复序列。与其他分子标记

技术相比, SSR 标记具有多态信息含量高、共显性遗传、技术简单、重复性好、特异性强等优点^[1], 已广泛应用于遗传图谱构建、遗传多样性分析、分子进化、

[收稿日期] 2007-07-16

[基金项目] 国家“863”计划项目(2006AA10Z199); 农林动植物育种工程项目(2006BAD01A11)

[作者简介] 闫秋良(1974-), 女, 山西石楼人, 在读博士, 主要从事生物技术与家畜育种研究。

[通讯作者] 杜立新(1956-), 男, 陕西华阴人, 教授, 博士生导师, 主要从事分子生物技术与家畜遗传育种研究。

E-mail: lxdu@263.net

比较基因组学和系统分类等方面。传统基因组来源的 SSR 标记开发需要构建基因组文库、探针杂交和克隆测序等复杂的操作程序,不仅费时费力,而且开发成本很高。

表达序列标签 (Expressed sequence tag, EST) 是指,通过对 cDNA 文库中随机挑取的单克隆进行大规模测序所获得 cDNA 的 5' 或 3' 端序列,长度一般为 150 ~ 500 bp,它是基因的“窗口”,一个 EST 代表生物某一时期的某种组织或细胞的一个表达基因^[2]。近年来,随着 EST 计划在不同物种间的扩展,大量快速增长的数据为 SSR 标记的开发提供了丰富的来源。根据建立 SSR 标记序列性质的不同,SSR 标记可分为基因组 SSR (Genomic SSR, gSSR) 和表达序列标签 SSR (Expressed sequence tag SSR, EST-SSR) 等。与 gSSR 标记相比,从 EST 数据库中获得 EST-SSR 标记更经济,而且 EST-SSR 标记反映了基因的编码部分,可以直接获得基因表达的信息,为功能基因提供“绝对”的标记,这有可能对决定重要表型性状的等位基因进行直接鉴定,省去了 SSR 引物开发过程中的克隆和测序步骤,充分利用了现有测序数据,降低了开发成本。从 EST 中筛选微卫星标记的工作,在小麦^[3-4]、大麦^[5]、黑麦^[6]、甘蔗^[7-8]和葡萄^[9]等植物物种中已有报道;在动物物种中,仅在鲶鱼^[10]、大虾^[11]和斑马鱼^[12]中有报道。

本研究利用现有的绵羊 EST 资源,对绵羊 UniGene 数据库进行 SSR 搜寻和信息分析,以了解绵羊 EST 资源的特性,以期利用绵羊 EST 建立 EST-SSR 标记及在绵羊遗传育种中的应用奠定基础。

1 材料与方法

1.1 绵羊 UniGene 序列来源

绵羊 UniGene 序列来源于 NCBI 的 UniGene 数据库,截止日期 2006-03-31,共获得 4 081 个 UniGene 簇,含有 EST 序列 58 353 条,所有这些序列以 FASTA 格式独立保存,以便进一步分析。

1.2 绵羊 EST-SSR 的筛选

利用 SSRIT 和 SSRFinder 软件,从上述 UniGene 簇中查找微卫星序列。查找标准是重复 7 次以上的双碱基序列,重复 6 次以上的三碱基序列,重复 5 次以上的四碱基序列,重复 4 次以上的五碱基序列和重复 3 次以上的六碱基序列。搜索的重复序列为完全重复类型。单核苷酸重复序列未进行选择,因为一般认为它不是有用的多态标记。

1.3 微卫星引物的设计

从获得的微卫星序列中,使用引物设计软件 Oligo (Version 6.31) 和 Primer Premier (Version 5.00) 设计 30 对 EST-SSR 引物,引物由上海生工生物技术公司合成(表 1)。

表 1 试验设计的 30 对微卫星引物

Table 1 30 pairs of primers of microsatellites

引物序号 Primer number	上游引物 (5'→3') Upstream primer (5'→3')	下游引物 (3'→5') Downstream primer (3'→5')
Sri1	CTTTGATGGTGCGGTACTCTT	A GGA GCCCTAACTTGCTTGA GT
Sri2	TCTTGACCGTA GGGGA TAGA G	A TCTCTCTA TCA GTCGCTCGC
Sri3	ACCGGCA GCA GTA GAAGAA	GAGCGTGACACTGA GAGAA G
Sri4	GCTCAGTTTACACCTCAAGTCC	ATGGA TGGA GGA GCCTGGTA
Sri5	TCTCCCGTGGACACCTCA	AA GGCA TGA GCGGTTGA G
Sri6	CCACCTACCGCAACTCCAT	TTCTTCCTCACCTGCA GC
Sri7	CCAA GTTCTACGACCTGTTCC	CTGTCCCA TTGCCTCGTAA
Sri8	CCTTTCCCTTAA TGTGCCAAG	TTTTCCCA GTCACGACGT
Sri9	AGTAACGATGGCAGCGAG	AGATGGA GAGCCTGGA GA
Sri10	AGGACCTCGGGA TGA ACT	CCA TTGCAAGGTA GGTGTC
Sri11	CTTAAGACACGCA GCTCG	TGCA GCCTTCTTGACCTT
Sri12	TTCTGACCCGCCTGCTTT	GTCCACACTCA TAACGGCA
Sri13	CCCACA TCTGACAGCCTATAC	GAA GGACTTGGA TGGA TGGA
Sri14	TCTACGGTGGAA GCGACTA	ACCCCTA TAGCCACTCCCAAG
Sri15	AACGAGGAGGCA TGTTTTG	GGAA GAGTTGGAACACAGGA
Sri16	AGGGCA TCGCTCCA GAAGA	GCCAAA GGTGGGCAACAACA
Sri17	TCGGGCA GCACATGTA GTA	TGA GAGCCCTA GTGGAA T
Sri18	GATTGACACTACTCCCGGAT	GTA TCTATGCTGCTGCTCGTC
Sri19	CGCTCCGTTCTGACA ACTC	GGCTGCCTCTCCTGCTCCT
Sri20	GCTCAA GACTCGCAACTACTAA	CAGGCAATGAGAGCTA TAA GG

续表 1 Continued of table 1

引物序号 Primer number	上游引物 (5'3') Upstream primer (5'3')	下游引物 (3'5') Downstream primer (3'5')
Sri21	CACCACCA TGCA GCC GA GAC	CA GA GTGCTCCTCGCCA TTGT
Sri22	A GCCCTTGCGCACGCCTACT	CA GA GGCCACCA CGGA GCA T
Sri23	A GTGTGCTGGTGCCCTA TCA	AACA TTCCGGACGCTCA TC
Sri24	GA ACTGTCA TCGCTGCTGTG	CAA TCCTCCCAA TCA TTACCTCAA
Sri25	AA GGGCCTGA AGCGAAAG	CCGCA GTGTGTTCACTACCA GC
Sri26	GA ATGGAGTGGCTGGGAGG	CTGCTTCGGGTCGGTGTTG
Sri27	TGGTGAA GTCAAACAACCCAC	A TTCTCTCGGACACACGTGA
Sri28	GCTACAACCCTGGCACAA	GGAACCCA GCTTTGA GAG
Sri29	TGAATCA GCA TCTCCTCTG	GACGGCCA GTGAA TTGTA
Sri30	CATGCCACA ACTAAGACTCGACA	GACACCA GCTGAAACGAACACA

引物设计的原则为: EST 序列长度大于 100 bp; SSR 序列的开始和结束位置分别距 5 和 3 端不少于 20 bp; 引物长度控制在 19 ~ 23 bp, GC 含量为 40 % ~ 60 %, 理论退火温度 (T_m) 为 59 ~ 63 , 而且上游和下游引物的 T_m 值相差 ≤ 5 ; PCR 扩增产物长度为 100 ~ 300 bp。尽量避免引物二聚体结构、发卡结构、错配以及连续 6 个碱基配对的出现。

1.4 动物材料及基因组 DNA 的提取

93 只小尾寒羊取自山东嘉祥种羊场。取小尾寒羊耳组织和血液, 参考分子克隆实验指南(第 3 版)的方法, 并优化模板提取 DNA。用分光光度计检测 DNA 浓度, - 20 保存备用。

1.5 PCR 扩增与电泳检测

PCR 反应体系为 (10 μ L): 1 U *Taq* DNA Polymerase 0.4 μ L, 10 $\times$ Buffer 1 μ L, 20 ng 基因组 DNA 1.0 μ L, 2.5 mmol/L dNTP 混合物 1 μ L, 20 μ mol 引物 0.5 μ L。PCR 反应程序为: 95 变性 5 min; 94 变性 30 s, 退火 30 s (温度依引物而定), 72 延伸 30 s, 共 30 个循环; 最后 72 延伸 10

min, 4 保存。各对引物的最适温度通过梯度 PCR 结果而定。PCR 产物用 1.5 % 琼脂糖凝胶电泳分离, 检测扩增产物的可用性, 对有清晰条带的引物进行聚丙烯酰胺凝胶电泳分离和银染检测其有无多态性。

2 结果与分析

2.1 绵羊 EST-SSR 的筛选

在 4 081 个绵羊 Uni Gene 簇中, 经过计算机筛选共发现微卫星 136 个, 含有微卫星的序列 121 个, 占整个 EST 序列数据库的 3.0 %, 其中双碱基序列 47 个, 三碱基序列 54 个, 四碱基序列 3 个, 五碱基序列 4 个, 六碱基序列 28 个。在这些微卫星序列中, AC/ TG 重复在双碱基类型中最丰富, CTG 重复在三碱基类型中最常见, 分别占双碱基和三碱基微卫星序列总数的 64 % 和 29 %。

2.2 EST-SSR 扩增结果的检测

表 2 结果表明, 10 对引物无扩增产物, 20 对引物有产物且条带清晰, 其中有 2 对引物在小尾寒羊品种内呈现多态, 有 18 对引物无多态。

表 2 20 对 EST-SSR 引物在小尾寒羊品种内的扩增情况及其多态性

Table 2 Amplification and polymorphisms showed in small-tailed han sheep by 20 pairs of EST-SSR primer					
引物序号 Primer number	重复序列类型 Type of repeat	多态或单态 Poly-or mono-morphism	引物序号 Primer number	重复序列类型 Type of repeat	多态或单态 Poly-or mono-morphism
Sri1	(ac) 20	单态 Monomorphism	Sri14	(gcg) 7	单态 Monomorphism
Sri2	(ac) 23	单态 Monomorphism	Sri15	(cac) 6	单态 Monomorphism
Sri3	(ct) 23	多态 Polymorphism	Sri16	(aga) 6	单态 Monomorphism
Sri5	(ac) 12	多态 Polymorphism	Sri18	(gag) 6	单态 Monomorphism
Sri6	(cag) 7	单态 Monomorphism	Sri19	(cac) 6	单态 Monomorphism
Sri7	(cac) 7	单态 Monomorphism	Sri22	(gt) 7	单态 Monomorphism
Sri9	(gt) 14	单态 Monomorphism	Sri23	(tgccc) 3	单态 Monomorphism
Sri10	(gt) 24	单态 Monomorphism	Sri25	(gga) 7	单态 Monomorphism
Sri11	(gca) 7	单态 Monomorphism	Sri26	(ctg) 6	单态 Monomorphism
Sri12	(tg) 21	单态 Monomorphism	Sri27	(tg) 9	单态 Monomorphism

3 讨 论

本试验对 4 081 个绵羊 Uni Gene 簇进行了搜

索, 发现各种类型的微卫星序列 136 个, 占整个 EST 序列数据库的 3.0 %, 这与从许多其他物种的 ESTs 数据库中筛选的微卫星序列的比例相一

致^[7,13-16]。从目前报道的结果来看,大多数植物的 EST-SSR 是以三核苷酸重复为主^[9,17-19],动物则都是以二核苷酸重复为主^[11-12,20-22]。本研究结果表明,在绵羊 EST-SSR 中三核苷酸重复类型最丰富,这与上述动物物种的研究结果不同,但与植物的相同。这可能是由于,一方面,随着物种的不同,不同重复类型的丰富程度也会随之改变所致^[23];另一方面,与本研究绵羊 EST 数目相对较少有一定关系,具体原因有待于进一步深入研究。本研究结果还表明,筛选的微卫星序列中,双碱基序列 47 个,三碱基序列 54 个,重复序列较小的类型占主导地位,这可能是由于重复序列越长,突变率就越高,因此稳定性就越差所致^[24]。本研究中,AC/ TG 重复在双碱基类型中最丰富,CTG 重复在三碱基类型中最常见,分别占双碱基和三碱基微卫星序列总数的 64% 和 29%,这与鲮鱼^[10]和斑马鱼^[12]中报道的结果相一致,但与在植物^[4,25]中报道的结果有一定差异。这可能与植物中某些氨基酸的频率较动物高有关^[23]。在三碱基重复类型中,GCT 基序是三碱基重复类型最丰富的类型,这与 Li 等^[26]的报道一致。

本研究根据绵羊 EST 序列设计引物,并对小尾寒羊基因组 DNA 进行 PCR 扩增,在设计的 30 对引物中有 67% 的引物能够扩增出产物,仅有 2 对引物在小尾寒羊品种内呈现多态,这与 Thiel 等^[5]报道的结果相一致。

在绵羊 EST 数据库中,虽然微卫星资源并不丰富,且大量的序列不完整,大部分不适合引物设计,但是从 EST 中发掘 SSR 标记,可以节省大量时间和经费,同时也是对 EST 数据资源的充分利用,为功能基因组和比较基因组的研究奠定了基础。

[参考文献]

- [1] Gupta P K, Balyan H S, Sharma P C, et al. Microsatellites in plants: a new class of molecular markers[J]. Current Sciences, 1996, 70: 45-54.
- [2] Adams M D, Kelley J M, Gocayne J D, et al. Complementary DNA sequencing: expressed sequence tags and human genome project[J]. Science, 1991, 252(5013): 1651-1656.
- [3] Peng J H, Lapitan N L. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers[J]. Funct Integr Genomics, 2005, 5: 80-96.
- [4] Eujayl I, Sorrells M E, Baum M, et al. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat[J]. Theoretical and Applied Genetics, 2002, 104: 399-407.
- [5] Thiel T, Michalek W, Varshney K, 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: 411-422.
- [6] Hackauf B, Wehling P. Identification of microsatellite polymorphisms in an expressed portion of the rye genome [J]. Plant Breed, 2002, 121: 17-25.
- [7] Cordeiro G M, Casu R, McIntyre C L, et al. Microsatellite markers from sugarcane (*Saccharum spp.*) ESTs cross transferable to erianthus and sorghum [J]. Plant Sci, 2001, 160(6): 1115-1123.
- [8] Pinto L R, Oliveira K M, Ulian E C, et al. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats [J]. Genome, 2004, 47: 795-804.
- [9] Scott K D, Eggler P, Seaton G, et al. Analysis of SSRs derived from grape ESTs [J]. Theoretical and Applied Genetics, 2000, 100: 723-726.
- [10] Serapion J, Kucuktas H, Feng J, et al. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*) [J]. Marine Biotechnology, 2004, 6: 364-377.
- [11] Perez F, Ortiz J, Zhinaula M, et al. Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus* [J]. Marine Biotechnology (New York, NY), 2005, 7: 554-569.
- [12] Ju Z L, Wells M C, Martinez A, et al. An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, Fundulus, and Xiphophorus [J]. In Silico Biology, 2005, 5: 439-463.
- [13] 徐 鹏. 中国对虾微卫星标记的筛选 [D]. 青岛: 中国科学院海洋研究所, 2002.
- [14] Kantety R V, La R M, Matthews D E, et al. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. Plant Molecular Biology, 2002, 48: 501-510.
- [15] Temnykh S, DeClerck 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: 1441-1452.
- [16] Nicot N, Chiquet V, Gandon B, et al. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs) [J]. Theoretical and Applied Genetics, 2004, 109: 800-805.
- [17] Gupta P K, Rustgi S, Sharma S, et al. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat [J]. Molecular Genetics and Genomics, 2003, 270: 315-323.
- [18] Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. Genetics, 2000, 156: 847-854.

(下转第 23 页)

- 2000:1-77.
- [7] Wieslaw S,Bradford S P,Jeremy G. *Yersinia pestis* Yop secretion protein F:Purification and protective efficacy against bubonic plague[J]. Protein Expression and Purification,2005,42:166-172.
- [8] Haim G,Sara C,Tamar B,et al. Effective protective immunity to *Yersinia pestis* infection conferred by DNA vaccine coding for derivatives of the F1 capsular antigen[J]. Infect Immun,2003,71:374-383.
- [9] Jyls M,Kelly A D,David S B,et al. Immunization of mice with YscF provides protection from *Yersinia pestis* infections[J]. BMC Microbiol,2005,38(5):1-10.
- [10] Taff J,Jeffrey J A,Sony L C,et al. Intranasal protollinTM/F1-V vaccine elicits respiratory and serum antibody responses and protects mice against lethal aerosolized plague infection[J]. Vaccine,2005,38(5):1-10.
- [11] Heath D G,Anderson G W,Mauro J M,et al. Protection against experimental bubonic and pneumonic plague by a recombinant capsular F1-V antigen fusion protein vaccine[J]. Vaccine,1998,16(11-12):1131-1137.
- [12] Straley S C,Skrzypek E,Plano G V. Yops of *Yersinia* spp. pathogenic for humans [J]. Infect Immun,1993,61:3105-3110.
- [13] Blocker A,Jouihri N,Larquet E,et al. Structure and composition of the *Shigella flexneri* "needle complex",a part of its type III secreton[J]. Mol Microbiol,2001,39(3):652-663.
- [14] Philipovski A V,Cowan C,Wulff-Strobel C R,et al. Antibody against V antigen prevents Yop-dependent growth of *Yersinia pestis*[J]. Infect Immun,2005,73:1532-1542.
- [15] Titball R W,Williamson E D. *Yersinia pestis* (plague) vaccines[J]. Expert Opin Biol Ther,2004,4(6):965-973.

(上接第 18 页)

- [19] Varshney R K,Thiel T,Stein N,et al. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species[J]. Cellular & Molecular Biology Letters,2002,7:537-546.
- [20] Yue G H,Orban L. Microsatellites from genes show polymorphism in two related *Oreochromis species*[J]. Molecular Ecology notes,2002,2:99-100.
- [21] Rohrer G A,Fahrenkrug S C,Nonneman D,et al. Mapping microsatellite markers Identified in porcine EST sequences[J]. Animal Genetics,2002,33:372-376.
- [22] Yue G H,Ho M Y,Orban L,et al. Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp[J]. Aquaculture,2004,234:85-98.
- [23] T Óh G,G Áp ái Z,Jurka J. Microsatellites in different eukaryotic genomes:survey and analysis[J]. Genome Research,2000,10:1967-1981.
- [24] Wierdl M,Dominska M,Petes T D. Microsatellite instability in yeast: dependence on the length of the microsatellite[J]. Genetics,1997,146:769-779.
- [25] Morgante M,Hanafey M,Powell W. Microsatellites are referentially associated with nonrepetitive DNA in plant genomes[J]. Nature Genetics,2002,30:194-200.
- [26] Li Y C,Korol A B,Fahima T,et al. Microsatellites within genes:structure,function,and evolution[J]. Mol Biol Evol,2004,21(6):991-1007.