

刺参(*Apostichopus japonicus*)EST 序列中微卫星 分布分析及其标记的筛选*

孙国华¹ 杨建敏¹ 宋志乐² 孙孝德^{1,3} 王卫军¹ 张宇¹

(1. 山东省海洋水产研究所 烟台 264006; 2. 烟台市芝罘区渔业技术推广站 烟台 264001;
3. 上海海洋大学水产与生命学院 上海 201306)

提要 利用微卫星查找软件对现已公布的 6632 条刺参 ESTs 的数据库中 2—6 个碱基重复单元组成的简单序列重复进行筛选, 进而对其微卫星的丰度和分布进行比较研究。共发现微卫星序列 416 个, 占整个 ESTs 数据库的 6.48%; 其中含双碱基重复序列 146 个, 数量最多, 占 ESTs 数据库中发现微卫星序列总数的 65.86%; 三、四、五、六碱基重复序列分别占微卫星序列总数的 26.68%、1.68%、0.24%、5.53%。对含有 SSR 位点符合微卫星引物设计的 EST 序列利用 Primer 软件结合人工方法设计合成引物 21 对, 其中 13 对有扩增产物且均为多态位点。扩增产物变性聚丙烯酰胺胶电泳, 统计每个微卫星标记等位基因数目, 计算等位基因频率、观测杂合度、期望杂合度等统计学指标的评价。结果表明, 在 13 个微卫星位点上, 等位基因的数目从 3 到 8 不等, 等位基因长度从 175—382bp, 观测杂合度和期望杂合度分别为 0.158—0.650 和 0.198—0.841, 所筛选微卫星标记可于遗传分析。

关键词 刺参, 微卫星, 表达序列标签, 等位基因

中图分类号 Q346

刺参(*Apostichopus japonicus*)隶属棘皮动物门、海参纲、楯手目、刺参科, 含有丰富的蛋白质, 粘性多糖及多种微量元素, 具有很高的营养保健和药用价值(樊绘曾, 2001; 沈鸣, 2001), 是我国重要的海水养殖经济种类。然而一方面刺参野生自然资源过度捕捞开发受到破坏, 另一方面刺参养殖长期近亲繁殖、累代养殖, 导致近亲衰退, 种质逐年下降(常亚青等, 2006; 黄华伟等, 2007), 对现有刺参种质资源遗传多样性进行监测评价, 筛选培育具有优势生长良性状、抗病能力强、抗逆性好的优良品种, 是解决这一系列问题, 保持刺参养殖业持续健康发展的迫切需要。现代分子标记技术的飞速发展, 为解决上述问题提供了有效的辅助和途径。

微卫星标记是近年来迅速发展起来的一种分子

标记技术, 已经成为遗传学领域和基因组学领域的研究热点。微卫星 DNA(Microsatellites DNA)是指以少数几个核苷酸(2—6 个)为单位多次重复的简单序列, 亦称简单序列重复(simple sequence repeats, SSR)、短串联重复(short tandem repeats, STR)等。微卫星 DNA 广泛而随机分布于真核生物的基因组中(Hamada *et al.*, 1982; Tautz *et al.*, 1984)。作为第二代分子标记手段, 与其它分子标记技术相比, 微卫星标记具有等位基因突变快、多态性高、杂合度大、信息含量丰富、呈共显性等显著优点(Toth *et al.*, 2000; Xu *et al.*, 2001), 因此, 在个体亲缘关系鉴定、群体遗传结构分析, 连锁图谱构建及谱系分析、遗传病鉴定、分子进化等方面, 微卫星标记技术得到了广泛的应用(McCouch *et al.*, 1997; 罗文永等, 2003; 孙效文等,

* 山东省良种工程项目“速生抗病耐高温刺参良种选育”, 2008—2011; 山东省优秀中青年科学家奖励基金(博士基金)项目, 2008BS06004 号; 山东省科技攻关项目“海参新品系选育及构建”, 2008GG10005023 号。孙国华, 博士, 助理研究员, E-mail: sgh_smile@163.com, sghqingdao@hotmail.com

通讯作者: 杨建敏, 博士, 研究员, E-mail: ladderup@126.com

收稿日期: 2009-06-23, 收修改稿日期: 2009-08-18

2008)。

EST(表达序列标签, expressed sequence tags)是指通过对 cDNA 文库随机挑取的克隆进行大规模测序所获得的 cDNA 的 5'或 3'端序列,反映 mRNA 的信息(Adams *et al*, 1991)。研究表明,7%—10%的 EST 序列中存在简单重复序列,这为从 EST 序列中开发利用新的 SSR 标记提供了依据,从 ESTs 数据库中筛选微卫星标记的工作,在许多水产动物中已见报道,如中国对虾(徐鹏等, 2003)、栉孔扇贝(李红蕾等, 2003)、马氏珠母贝(石耀华等, 2008)等。美国国立信息中心 GeneBank 数据库中已登录了 6632 条刺参的 EST 序列,目前还没有对这些 EST 中 SSR 的分布和频率等进行分析的报道,本文旨在对该 EST 序列中 SSR 的分布、频率和组成进行分析,为进一步开发新的 SSR 分子标记和刺参基因编码区 SSR 的功能研究提供有价值的信息。

1 材料与方法

1.1 序列来源

从 NCBI (<http://www.ncbi.nlm.nih.gov/>)GeneBank EST 数据库中,以 *Apostichopus japonicus* 为关键词,用 FASTA 格式下载 6632 条刺参的 EST 序列,以 2009 年 6 月 NCBI 公布的数据为准。

1.2 微卫星位点查找

获得 FASTA 格式的 EST 序列后,利用在线查找软件 (<http://www.gramene.org/db/searches/ssrtool>)分析和查找 SSR 位点。查找中使用的标准是:5 次以上重复的双碱基至十碱基重复序列。

1.3 微卫星引物设计

从搜索获得的微卫星序列中,选取符合引物设计的序列(重复次数在 12 次以上的双碱基重复序列或重复次数在 8 次以上的三碱基以上重复序列),利用引物设计软件 PrimerPremier (Version 5.00)结合人工方法进行微卫星引物的设计。引物长度一般控制在 18—22bp 之间,GC 含量一般在 40%—70%, T_m 值控制在 45—65,产物长度控制在 100—400bp 之间,引物在上海生工生物技术公司合成。

1.4 刺参 DNA 的提取

随机选取中国烟台刺参群体和俄罗斯刺参群体共 60 个,剖取获得纵肌 100mg,机械剪碎,加入 700ml CTAB 提取缓冲液(100mmol/L Tris-HCl pH 8.0, 20mmol/L EDTA- Na_2 , 1.4mol/L NaCl, 2% CTAB, 0.1% β -巯基乙醇)和终浓度为 100 $\mu\text{g}/\text{ml}$ 蛋白酶 K, 55 消

化 3h 或 37 过夜,等体积酚氯仿(酚:氯仿 = 1:1)、氯仿抽提,二倍体积乙醇沉淀,TE 溶解,4 保存。

1.5 PCR 扩增筛选

以中国烟台刺参群体和俄罗斯刺参群体基因组 DNA 为模板,采用设计引物序列进行 PCR 扩增,25 μl PCR 反应体系,约 100ng 的基因组 DNA,10pmol 的引物,2.5 μmol MgCl_2 , 0.2 μmol dNTP, *Taq* DNA 聚合酶 1U。PCR 反应条件为:94 4min,然后 94 40s,退火温度(根据梯度 PCR 获得)40s,72 延长 40s,30 个循环,之后 72 10min,4 保温。

1%琼脂糖凝胶电泳,对明确有扩增产物的引物通过 8%(W/V)聚丙烯酰胺凝胶电泳和硝酸银染色检测其多态性,筛选多态性微卫星引物。

1.6 数据统计处理

将 PAGE 电泳结果进行图像扫描,根据 PAGE 电泳图谱,使用凝胶分析软件 Quantity One 进行微卫星 DNA 带型分析及记录。对每个微卫星位点的等位基因的数目进行统计,计算每个等位基因的频率和下列参数:观测杂合度(H_o): H_o 为杂合子观察数与样本含量之比;期望杂合度(H_e): $H_e = 1 - \sum P_i^2$, P_i 为该位点上第 i 个等位基因的频率;软件采用 POPGENE 在线版本(<http://wbiomed.curtin.edu.au/genepop/>)。

2 结果

2.1 EST 数据库中含有微卫星的数量和频率

从公共数据库中获得刺参 EST 序列 6632 条,4.11Mb,平均长度 619bp。经过计算机在线软件筛选共发现微卫星位点 416 个(5 次以上重复的双碱基至十碱基重复序列),占 6632 条刺参 EST 序列的 6.48%,其中 39 条 EST 序列含有两个或两个以上微卫星位点,大部分发现的微卫星序列均为 Perfect 形式的重复序列。

EST 中 5 种基本重复单元的微卫星重复中,以双碱基重复为主导型,其序列丰度最高,共 274 个,占微卫星总数量的 65.86%,随后依次是三碱基重复序列(26.68%)、六碱基重复序列(5.53%)、四碱基重复序列(1.68%)、五碱基重复序列(0.24%)。在这些微卫星序列中,(AT/TA)形式在双碱基重复中最为常见,(AAT/ATT/TAA/TTA)形式在三碱基重复中最为常见,分别占微卫星序列总数的 31.97%和 9.13%,各重复序列类型及占总数量比例见表 1。

2.2 微卫星位点引物设计、筛选及评价

SSR 重复序列中重复次数在 12 次以上的双碱基重复序列或重复次数在 8 次以上的三碱基以上重复

表 1 刺参 EST 中重复序列类型及分布特点

Tab.1 The distribution of microsatellite sequences in EST data of *A. japonicus*

SSR 类型	重复序列	SSR 位点数量 (总 416)	最大重复次数	各类型位点数	占总位点数 百分比			
六碱基	gaccac	23	8	23	5.53%			
五碱基	atttt	1	6	1	0.24%			
四碱基	atac	2	5	7	1.68%			
	ttta	1	5					
	gtat	3	13					
	gata	1	5					
三碱基	aat/att/taa/tta	38	8	111	26.68%			
	gat/atc/tag/cta	12	14					
	aag/ctt/gaa	7	6					
	aac/gtt/ /ttg	6	5					
	tca/tga/act	10	5					
	cca/tgg	3	5					
	aga/tct	3	7					
	ata/tat	11	8					
	gtg/cac	4	12					
	atg/cat	2	6					
	aca	11	6					
	agc	2	5					
	tgc	1	6					
	agg	1	6					
	二碱基	AT/ta	133			25	274	65.86%
		Ga/ag	31			9		
Tc/ct		49	50					
Ac/ca		33	8					
Gt/tg		27	10					
Gc		1	5					

序列(重复序列长度 20bp 以上)共 45 条, 除去其中两端序列较短不能设计引物及相近序列, 通过 Primer 软件及人工共设计微卫星引物序列 21 对。21 对引物中有 13 对得到有效扩增, 占合成总引物数量的 52.4%, 可扩增引物序列及其它参数见表 2。

13 对引物聚丙烯酰胺凝胶电泳检测显示此 13 个微卫星标记均为多态性位点(图 1), 13 个位点共获得 66 个等位基因, 不同的引物获得的等位基因数为 3—8 个不等, 平均每对引物获得 5.08 个等位基因。其中位点 SSR-E3 获得等位 8 个, 为等位基因数目最多的位点。等位基因数最多; SSR-E10 等位基因数目最少, 仅为 3 个; 多数位点获得 4 个等位基因。13 对引物的得到的等位基因数目、等位基因的大小范围、观测杂合度和期望杂合度见表 2。

3 讨论

在本研究中, 对已公布的刺参 EST 序列中其 2—6 个核苷酸重复 SSR 的分布、碱基组成和频率进行了分析, 共从刺参 EST 数据库中发现各种类型的微卫星序列 416 个, 占 ESTs 总数的 6.48%。EST 数据库中微卫星分布的频率在各个物种中不尽相同, 赵美琼等(2008)分析玉米 EST 数据库中微卫星数量占总序列的 7.25%, Karsi 等(2002)发现鲑鱼 ESTs 中微卫星序列比例为 4.66%, 石耀华等(2008)分析含微卫星的 EST 数占马氏珠母贝 EST 总数的 3.48%, 徐鹏等(2003)发现中国对虾微卫星序列占整个 ESTs 数据库的 2.19%, 而李红蕾等(2003)分析了栉孔扇贝 EST 中微卫星比例为 0.6%。分析比例差异的原因, 作者认为主要有三点:

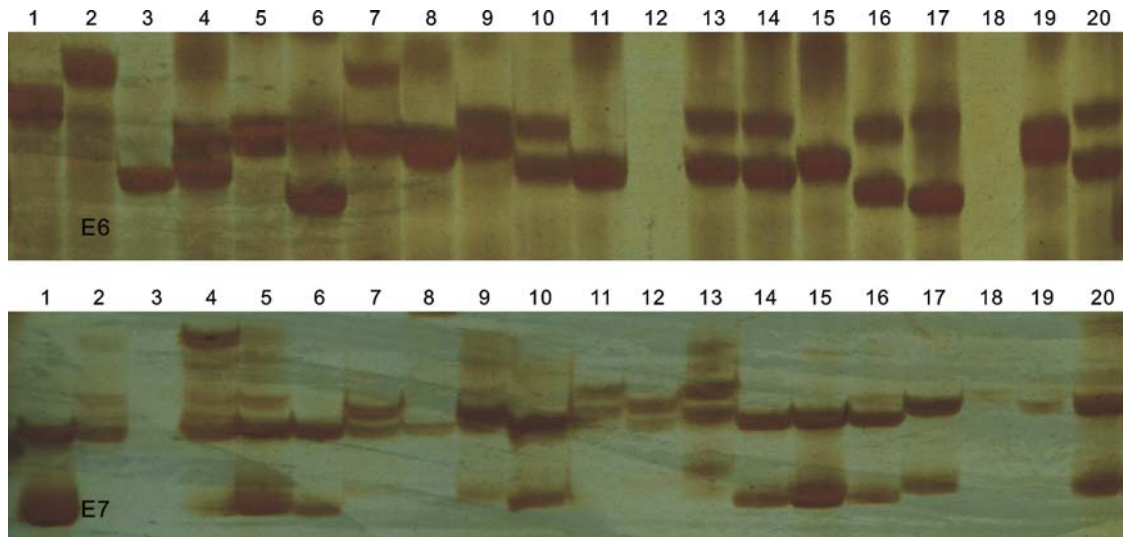


图 1 引物 SSR-E6、SSR-E7 部分个体 PCR 扩增聚丙烯酰胺凝胶银染照片

Fig.1 The silver staining PAGE image of PCR result of microsatellite marker SSR-E6 and SSR-E7

一是不同研究者对微卫星的定义不同,这也是目前研究微卫星很有争议的地方,对于查找的标准,不同研究者有不同的选择。Becker 等(1995)所用的依据两核苷酸重复 $n = 5$, 三核苷酸重复 $n = 4$, 李红蕾等(2003)分析栉孔扇贝的筛选标准是 7 次以上重复的双碱基序列或 5 次以上的三碱基重复序列,而本研究查找微卫星中使用的标准是 5 次以上重复的双碱基至十碱基重复序列。这种状况也造成了不同研究结果之间的可比性降低。

二是各个物种 EST 数据库中序列特性如序列来源及序列读长的不同。石耀华等(2008)分析马氏珠母贝珍珠囊、心脏、外套膜、鳃、血液、足 cDNA 文库 EST 序列中含微卫星的序列所占比例分别为 6.16%、5.98%、5.67%、2.90%、1.84%、1.67%。李红蕾等(2003)发现构建栉孔扇贝文库 EST 读长为 392.52bp,其中含有微卫星的序列总 ESTs 的 0.6%,而海湾扇贝 ESTs 的测序读长为 462.01bp,其中含有微卫星的序列总 ESTs 的 1.43%。

三是物种的不同。不同物种基因组的结构组成不同,大小容量不同,基因序列数量不同,碱基组成含量也不同,因而对微卫星丰度及不同类型微卫星的数量有影响。

本研究中刺参微卫星 2—6 碱基各微卫星重复类型中双碱基重复数量最多,占微卫星总数的 65.86%,随后依次是三碱基重复、六碱基重复序列,比较其它同类研究,马氏珠母贝(石耀华等,2008)、中国对虾(徐鹏等,2003)、长牡蛎(李琪等,2004)EST 数

据库中双碱基重复占主导地位,栉孔扇贝(李红蕾等,2003)、曼氏血吸虫(唐远菊等,2007)、甜瓜(胡建斌等,2009)EST 数据库中三碱基重复是数量最多类型,在玉米(赵美琼等,2008)、蜜蜂(李斌等,2004)、赤拟谷盗(张琳琳等,2008)、苹果(张俊娥,2008)中则是六碱基重复是主导类型。Varshney 等(2002)通过研究多种谷类植物 EST 中发现三碱基重复序列占主导地位,并认为这是 DNA 复制时偏爱三碱基重复的微卫星,因为三碱基是密码子的组成形式,减少或增加一个碱基会增加错配引起移码突变的机会,但是综合多方面研究来看,这无法解释其它重复类型占主导的情况。至于不同物种 EST 数据库中不同类型的微卫星重复占主导的具体原因,还有待于进一步研究。

在刺参双碱基和三碱基各重复类型中, AT/ta 及 aat/att/taa/tta 数量最多,说明富含以富含 A(或 T)的重复单元占主导地位,微卫星偏爱富含 A 或 T 的序列而富含 G 或 C 的序列微卫星出现频率较低。这与其他研究者在其它不同物种之中的实验结果相符,如中国对虾(徐鹏等,2003)、马氏珠母贝(石耀华等,2008)、蜜蜂(李斌等,2004)、曼氏血吸虫(唐远菊等,2007)等,但是水稻(Kantety *et al*, 2002)、玉米(赵美琼等,2008)、苹果(张俊娥,2008)、条斑紫菜(刘必谦等,2005)等物种中的研究结果有差异,推测这可能与不同编码基因的碱基组成偏好及体内甲基化酶活性有关。

设计合成的 21 对引物中,其中 13 对得到有效扩增,并且部分引物在所扩增 20 个个体中有个别个体无扩增现象。对于无扩增产物的引物对,除去引物自

身发夹结构、形成二聚体等因素之外,作者分析由于引物是根据 EST 序列设计,引物序列有跨越内含子的很大可能,因而无法产生有效扩增。对于个别个体中无扩增产物的情况,可能是有些个体在引物结合位置出现了变异,导致引物结合位点的改变或丧失,结果无法扩增出产物,属于该标记位点在个体间多态性的表现。在 13 个微卫星位点上,等位基因的数目从 3 到 8 不等,等位基因长度从 175—382bp,观测杂合度和期望杂合度分别为 0.158—0.650 和 0.198—0.841,所筛选微卫星标记可于遗传分析。

近年来,从 EST 数据库中获得含有微卫星的序列并且设计引物进行微卫星标记筛选在国际上已有成功报道(李永强等,2004)。海洋生物的分子生物学研究起步较晚,刺参的分子生物学及基因组研究更是刚刚开始,GenBank 中刺参的 EST 序列较其它物种相比而言数量比较少,微卫星研究尚未见报道。EST 中微卫星标记的开发不仅省去了传统基因组微卫星开发,杂交、克隆、测序等费时费力繁琐步骤,并且由于 EST 是功能基因的表达片段,在其中发现的微卫星标记直接和功能基因相关,也有可能和一些生产性状相关联,这些特点对以后的遗传图谱构建及基因功能分析有很高的应用价值。虽然大多数物种构建序列文库,获得 EST 序列的目的是为了解组织或者个体的基因表达情况,但从 EST 中发现和挖掘微卫星序列可以更为充分地利用这些数据资源,获得微卫星引物应用于图谱构建和遗传分析并可直接用于基因功能和比较基因组学的研究。

参 考 文 献

- 石耀华,洪葵,郭希明等,2008. 马氏珠母贝 EST 微卫星的筛选. 水产学报, 32(2): 174—181
- 刘必谦,曾庆国,骆其君,2005. 条斑紫菜(*Porphyra yezoensis*)dbEST 中筛选微卫星位点及引物种间转移扩增. 海洋与湖沼, 36(3): 248—254
- 孙效文,张晓锋,2008. 水产生物微卫星标记技术研究进展及其应用. 中国水产科学, 15(4): 689—703
- 李琪,木岛明博,2004. 长牡蛎(*Crassostrea gigas*)微卫星克隆快速分离及特性分析. 海洋与湖沼, 35(4): 364—370
- 李斌,夏庆友,鲁成等,2004. 蜜蜂 EST 中的微卫星分析. 遗传学报, 31(10): 1089—1094
- 李永强,李宏伟,高丽锋,2004. 基于表达序列标签的微卫星标记(F_ST-SSRs)研究进展. 植物遗传资源学报, 5(1): 91—95
- 李红蕾,宋林生,王玲玲,2003. 栉孔扇贝 EST 中微卫星标记的筛选. 高技术通讯, 13(12): 72—75
- 沈鸣,2001. 海参的化学成分和药理研究进展. 中成药, 23(10): 758—761
- 张俊娥,2008. 苹果 EST 中微卫星分析. 广西农业生物科学, 127(14): 378—380
- 张琳琳,魏朝明,廉振民等,2008. 赤拟谷盗全基因组和 EST 中微卫星的丰度. 昆虫知识, 45(1): 38—45
- 罗文永,胡骏,李晓方,2003. 微卫星序列及其应用. 遗传, 25(5): 615—619
- 赵美琼,宋燕,范静华等,2008. 玉米 EST 序列中微卫星的频率和分布. 云南农业大学学报, 23(5): 585—589
- 胡建斌,刘颖,王兰菊等,2009. 甜瓜 EST 序列中微卫星的分布特征. 植物生理学通讯, 45(3): 258—262
- 徐鹏,周令华,田丽萍等,2003. 从中国对虾 ESTs 中筛选微卫星标记的研究. 水产学报, 27(3): 213—218
- 唐远菊,罗洪林,聂奎,2007. 曼氏血吸虫丁中的微卫星分析. 中国预防兽医学报, 29(8): 629—633
- 黄华伟,王印庚,2007. 海参养殖的现状、存在问题与前景展望. 中国水产, 10: 50—53
- 常亚青,隋锡林,李俊,2006. 刺参增养殖业现状、存在问题与展望. 水产科学, 25(4): 198—201
- 樊绘曾,2001. 海参:海中人参. 中国海洋药物, 4(1): 37—44
- Adams M D, Kelley J M, Gocayne J D, 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science, 252(5013): 1651—1656
- Becker J, Heun M, 1995. Barley microsatellites: allele variation and mapping. Plant Mol Biol, 27(4): 835—845
- Hamada H, Petrino M G, 1982. Kakunaga A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci USA, 79: 6465—6469
- Kantety R V, La Rota M, Matthews D E *et al*, 2002. Data mining for simple sequence repeats in expressed tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology, 48: 501—510
- Karsi A, Cao D, Li P *et al*, 2002. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. Gene, 285(1—2): 157—168
- McCouch S R, Chen X, Panaud O *et al*, 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol, 35: 89—99
- Tautz D, Renz M, 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res, 25(10): 4127—4138
- Toth G, Gaspari Z, Jurka J, 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research, 10: 967—981
- Varshney R K, Thiel T, Stein N *et al*, 2002. In silico analysis on frequency and distribution of microsatellites in ests of some cereal species. Cell Mol Biol Lett, 7(2A): 537—546
- Xu Z, Primavera J P, Pena L D *et al*, 2001. Genetic diversity of wild and cultured Black Tiger Shrimp (*Penaeus monodon*) in the Philippines using microsatellites. Aquaculture, 199: 13—40

ISOLATION OF MICROSATELLITE MARKERS FROM *APOSTICHOPUS JAPONICUS* ESTs

SUN Guo-Hua¹, YANG Jian-Min¹, SONG Zhi-Le², SUN Xiao-De^{1,3},
WANG Wei-Jun¹, ZHANG Yu¹

(1. Shandong Marine Fisheries Research Institute, Yantai, 264006; 2. Zhifu Fishery Technical Extension Station, Yantai, 264001;
3. School of Fishery and Life Science, Shanghai Ocean University, Shanghai, 201306)

Abstract Frequency and density of 2—6bp simple sequence repeats (SSRs) were analyzed from *Apostichopus japonicus* EST (expressed sequence tag) database comprising 6632 sequences. The result shows that 416 microsatellite sequences (taking 6.48%) were obtained. The dinucleotide repeats motifs appeared to be the most abundant type in *A. japonicus* ESTs (taking 65.86%), followed by trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats at 26.68%, 1.68%, 0.24%, and 5.53%, respectively. 21 primer pairs were designed with Primer software and manual-searching method from EST sequences containing microsatellites. 13 primers had polymorphic PCR products in Chinese and Russian sea cucumber samples by electrophoresis analysis. In the 13 microsatellite loci, the alleles numbers distributed from 3 to 8, and the range of these alleles were between 175 to 382bp. The observed heterozygosity and expected heterozygosity distributed from 0.158—0.650 and 0.198—0.841, respectively, which showed that these 13 microsatellite loci were suitable for genetic analysis.

Key words *Apostichopus japonicus*, Microsatellite, Expressed sequence tag (EST), Allele