

大竹蛭(*Solen grandis*)cDNA 文库中 微卫星标记的筛选*

乔洪金¹ 刘相全¹ 孙国华¹ 房景辉¹ 韦秀梅¹ 张滔²

(1. 山东省海洋水产研究所山东省海洋生态修复重点实验室 烟台 264006;

2. 鲁东大学生命科学学院 烟台 264025)

提要 利用微卫星查找软件 SSRIT 对大竹蛭 cDNA 文库(2038 条 EST)中 2—6 个碱基重复单元组成的简单序列重复进行了筛选。最少重复次数设定为 5 次,共发现包含微卫星位点的 EST 96 条,占整个 EST 数据库的 4.71%;共发现微卫星位点 103 个,其中含双碱基重复序列 77 个,数量最多,占总数的 74.76%;三、四碱基重复序列分别占微卫星序列总数的 22.33%、2.91%,没有发现五或六碱基的重复。对含有 SSR 位点符合微卫星引物设计的 EST 序列,利用在线引物设计软件 Primer 3 设计合成引物 14 对,以南通野生群体为模板 PCR 扩增和聚丙烯酰胺凝胶电泳发现,其中 5 对有多态性位点。在 5 个微卫星位点上,等位基因的数目从 2—7 个不等,观测杂合度和期望杂合度分别为 0.067—1.000 和 0.066—0.775,香农指数在 0.146—1.545 之间。结果表明,所筛选的微卫星标记可用于遗传分析。

关键词 大竹蛭,微卫星,表达序列标签,等位基因

中图分类号 S968.3

大竹蛭(*Solen grandis*)属于软体动物门、瓣鳃纲、真瓣鳃目、竹蛭科、竹蛭属,个体肥大,足部肌肉特别发达,味极鲜美,入药有滋补、通乳功效,是我国重要的经济贝类。大竹蛭在我国南北沿海均有分布,不少地方已开始试探性地进行人工养殖,在遗传育种方面处于起步阶段。筛选培育具有优良生长性状、抗病性强、抗逆性好的品种,促进大竹蛭养殖业持续健康发展,是目前大竹蛭遗传育种研究的主要方向。

作为遗传育种的一种重要辅助途径,微卫星标记技术在水产养殖业上得到了广泛的应用(McCouch *et al*, 1997; 罗文永等, 2003; 孙效文等, 2008)。微卫星 DNA 亦称短串联重复序列(simple tandem repeats, STRs)或简单序列重复(simple sequence repeats, SSR),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,在真核生物的基因组中,不论是编码区还是非编码区,都广泛而随机地分布着大量微卫星序列(Hamada *et al*, 1982; Tautz *et al*, 1984)。微卫星标记作为第二代分子标记,与其它分子标记技术(如等位酶、RFLP、RAPD 等)相比,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)。微卫星标记技术的主要用途有: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、分子和种群进化。综上所述,微卫星标记技术以其自身的优点应用广泛,是遗传学和基因组学领域的研究热点。

通过随机挑取 cDNA 文库的克隆进行测序所获得的 cDNA 的 5'或 3'端序列称为 EST(表达序列标签, expressed sequence tags),它包含了部分或全部的

* 山东省农业良种工程项目“优质高产抗逆贝类良种选育”资助,2009—2013;水生动物营养与饲料“泰山学者”岗位经费资助,2007—2012;山东省自然科学基金(ZR2012CM037)资助,2012—2015。乔洪金,博士,助理研究员, E-mail: hongjinqiao@gmail.com

通讯作者:刘相全,博士,副研究员, E-mail: emppp@163.com

收稿日期:2012-01-12,收修改稿日期:2012-03-26

mRNA 编码信息, 而且 7%—10%的 EST 序列中包含微卫星序列(SSR), 因而可以从 EST 序列中开发利用新的 SSR 标记, 即 EST-SSR 标记(Adams *et al.*, 1991)。伴随着基因组时代的到来, EST 序列数据飞速增长, 其中包含着大量与特定基因相关联的微卫星序列, 开发这些序列不仅成本低, 而且由于 EST 序列的保守性使其可以在相近的物种中使用, 因而可用于种群遗传和进化分析(Ellis *et al.*, 2007)。目前, 对许多水产动物都已开展了从 EST 数据库中筛选微卫星标记的工作, 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)、刺参(孙国华等, 2010)等, 目前还未见从大竹蛭 EST 中分离 SSR 的报道。本文旨在对大竹蛭 EST 序列中 SSR 的分布、频率和组成进行分析, 开发新的 SSR 分子标记, 并利用这些标记对大竹蛭群体的遗传多样性进行分析, 以期为大竹蛭的种质状况评价和遗传育种研究提供理论依据。

1 材料与方法

1.1 cDNA 文库的构建

大竹蛭(*Solen grandis*)样品取自山东省日照市, 样品于实验室洁净海水内暂养 2d, 10 只个体分别取外套膜、管足、闭壳肌、性腺、肝胰脏等组织混合样品投入液氮中保存。将组织在液氮中充分研磨, 加入适量的 Trizol reagent (GIBCO BRL 公司), 提取总 RNA。将提取的总 RNA 进行 DNA 酶消化后紫外分析测定并进行变性凝胶电泳, 检测总 RNA 的纯度及完整性。采用 QIAGEN 公司的 Oligotex mRNA Kits 进行 mRNA 的分离纯化, 定量测定得率。

文库构建按照 SMART cDNA 文库构建试剂盒(Clontech 公司)的流程进行: 以 poly(A) + mRNA 为模板, 加入反转录引物 SMART IV Oligo nucleotide 和 CDS 3' PCR Primer, 在 PowerScript Reverse Transcriptase 42℃作用下反转录为第一链 cDNA。以 5' PCR Primer 和 CDS 3' PCR Primer 为引物, 在 Advantage Polymerase 作用下扩增双链得到 cDNA, 双链 cDNA 经蛋白酶 K 消化、SfiI 消化酶切及片段分级分离纯化后, 与 pBluescript SK(-)载体于 16℃连接, 连接载体转化感受态宿主菌 DH-5 α 。

随机挑选 2050 个重组克隆测序, 用 phred 软件读取峰图(ABD)文件, 将其转化为序列文件。采用 NCBI

载体查找工具 VecScreen 去除两侧的载体引物, 并用 Bioedit 软件将所有序列转换为 FASTA 格式。

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 SSRIT (<http://www.gramene.org/db/searches/ssrtool>)分析和查找 SSR 位点。查找标准为: 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, 利用引物在线设计软件 Primer 3 (<http://frodo.wi.mit.edu/primer3/>)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 55—65℃, 产物长度控制在 100—400bp 之间, 引物送交上海生工生物技术公司合成。

1.4 大竹蛭 DNA 的提取

随机选取南通大竹蛭群体 30 个, 剖取获得闭壳肌 100mg, 机械剪碎, 加入 700ml CTAB 提取缓冲液(100mmol/L Tris-HCl, pH 8.0, 20mmol/L Na₂EDTA, 1.4mol/L NaCl, 2% CTAB, 0.1% β -巯基乙醇)和终浓度为 100 μ g/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, 70%乙醇洗涤两次, 二倍体积无水乙醇沉淀, TE 溶解, 4℃保存。

1.5 PCR 扩增筛选

以南通大竹蛭群体基因组 DNA 为模板进行 PCR 扩增, 使用 25 μ l PCR 反应体系, 包括: 约 100ng 的基因组 DNA, 正反向引物各 10pmol, 2.5 μ mol MgCl₂, 0.2 μ mol dNTP, 1U Taq DNA 聚合酶。PCR 反应条件为: 94℃ 5min, 然后 94℃ 1min, 退火(退火温度由梯度优化获得)1min, 72℃延伸 1min, 30 个循环, 之后 72℃ 10min, 4℃保温。扩增产物进行 1%琼脂糖凝胶电泳, 对明确有扩增产物的引物通过 12% (W/V)非变性聚丙烯酰胺凝胶电泳和溴化乙锭染色检测其多态性, 筛选多态性微卫星引物。

1.6 数据统计处理

使用凝胶分析软件 Quantity One 进行微卫星 DNA 带型分析和记录, 计算条带分子量范围, 对每个微卫星位点的等位基因的数目进行统计, 采用 POPGENE32 软件(<http://wbiomed.curtin.edu.au/genepop/>)计算每个等位基因的频率和下列参数: 观测杂合度(H_o), 期望杂合度(H_e), Hardy-Weinberg 平衡, 香农多样性指数(I)。

2 结果

2.1 cDNA 文库中微卫星的数量和频率分布

对 2038 条 EST 序列进行了分析, 经过计算机在线软件筛选, 共发现包含微卫星位点的 EST 序列 96 条(5 次以上重复的双碱基至十碱基重复序列), 占总 EST 序列数的 4.71%, 其中 14 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 77 个, 占微卫星总数的 74.76%, 随后依次是三碱基重复序列(22.33%), 四碱基重复序列(2.91%), 未发现五碱基和六碱基的重复。在这些微卫星序列中, (AT/TA)形式在双碱基重复中最为常见, (CAA/TTG)形式在三碱基重复中最为常见, 分别占微卫星序列总数的 23.30%和 5.83%, 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

通过 Primer 3 软件共设计微卫星引物序列 30 对, 其中有 5 对引物得到有效扩增并表现出多态性, 可扩增引物序列及其它参数见表 2。

聚丙烯酰胺凝胶电泳检测显示此 5 个微卫星标记均为多态性位点(图 1), 5 个位点共获得 20 个等位基因, 不同的引物获得的等位基因数为 2—7 个不等, 平均每对引物获得 4 个等位基因。其中位点 sg-2 获得等位基因 7 个, 为等位基因数目最多的位点, sg-3 等位基因数目最少, 仅为 2 个; 多数位点获得 3 个等位基因。其中两对引物严重偏离 Hardy-Weinberg 平衡 ($P < 0.01$)。5 对引物得到的等位基因数目, 等位基因的大小范围, 观测杂合度, 期望杂合度和香农指数见表 2。观测杂合度平均值为 0.393, 期望杂合度平均值为 0.358, 香农指数平均值为 0.706。

3 讨论

近年来, 在水产动物中利用 EST 开发微卫星标记的研究已有较多报道, 对其中几个典型物种的比较结果见表 3。由于采用的筛选标准和库容的不同, 因而产生的数据难免有较大差异, 但其中也有某些共同的特征: (1) 双碱基重复的微卫星序列总是比三碱基重复的多; (2) 双碱基重复序列都以重复单元 (AT/TA)_n 为主要类型。随着水产动物大规模测序时代的到来, 海量的 EST 数据中必然包含着大量的微卫

表 1 大竹蛭 cDNA 文库中微卫星重复序列的类型及分布特点

Tab.1 Types and distribution of microsatellite repeat sequences in cDNA library of *S. grandis*

SSR 类型	重复序列	SSR 位点数量(总 103)	最大重复次数	占总位点百分比
双碱基		77		74.76%
	AT/ta	24	12	
	AC/ca	13	7	
	AG/ga	23	16	
	TC	2	11	
	GT/tg	13	10	
	GC	2	6	
三碱基		23		22.33%
	TTA	2	5	
	GAT	1	13	
	CAA/TTG	6	6	
	ATA	1	12	
	GTG	1	7	
	ACA	2	5	
	TCA/TGA	3	6	
	AGG	1	6	
	GGT/CCA	2	7	
四碱基		3		2.91%
	GAG/CTC	1	5	
	TAC	2	5	
	GGC	1	6	
	CATA	1	5	
	TCTG	1	9	
	TGTA	1	5	

星位点, 因此, 使用统一的筛选标准来筛选这些微卫星标记, 并在不同的物种之间进行横行比较, 才能充分地利用这些数据资源, 挖掘出基因组层面的数据意义。鉴定转录组中的微卫星位点是一种开发多态性 EST 标记的有效方法, EST 标记与微卫星标记多态性的形成方式并不完全相同, EST 标记长度的变化一般是由于 PCR 引物间的序列插入或缺失突变引起的, 而微卫星标记长度的变化是由于 DNA 聚合酶剪切重复单元造成的(Tautz *et al.*, 1986), 但是插入或缺失突变的概率很低, 因此如果能够发现可转录的微卫星序列, 即 EST-SSR, 那么通过扩增该微卫星的区域就很可能可以获得多态性的 EST-SSR 标记(Liu *et al.*, 1999)。这些多态性的 EST-SSR 标记融合了 EST 标记

表 2 大竹蛭 5 对微卫星引物的特性与评价
Tab.2 Characteristics and evaluation of five microsatellite primer pairs of *S. grandis*

位点名称	核心序列	引物序列(5'—3')	退火温度(°C)	等位基因数	等位基因大小范围(bp)	观测杂合度 H_o	期望杂合度 H_e	香农指数
sg-1	(AC) ₂ ct(AC) ₅	Tttggctacagcgatctgaa tctctcccaagccatcaagt	60	3	270—316	0.133	0.128	0.282
sg-2*	(GC) ₆ (GT) ₁₀	caattgtccccatgtaggt cggatacattgaagagcgaaa	60	7	200—264	1.000	0.775	1.545
sg-3	(GTG) ₇	aatgaaaccggtcactttgc caagaggacattcagcagca	59.5	2	131—139	0.067	0.066	1.151
sg-4	(TC) ₆ tt(TC) ₂ tt(TC) ₃	gggactgtccgttacaaaaatg tgacaagcaatcacagcaca	57.7	3	231—257	0.233	0.213	0.408
sg-5*	(TCTG) ₉	gaccgaaaggagaacctctc ctagtagcccctggaccaga	65	5	220—314	0.533	0.606	0.146

注: *表示严重偏离 Hardy-Weinberg 平衡($P < 0.01$)

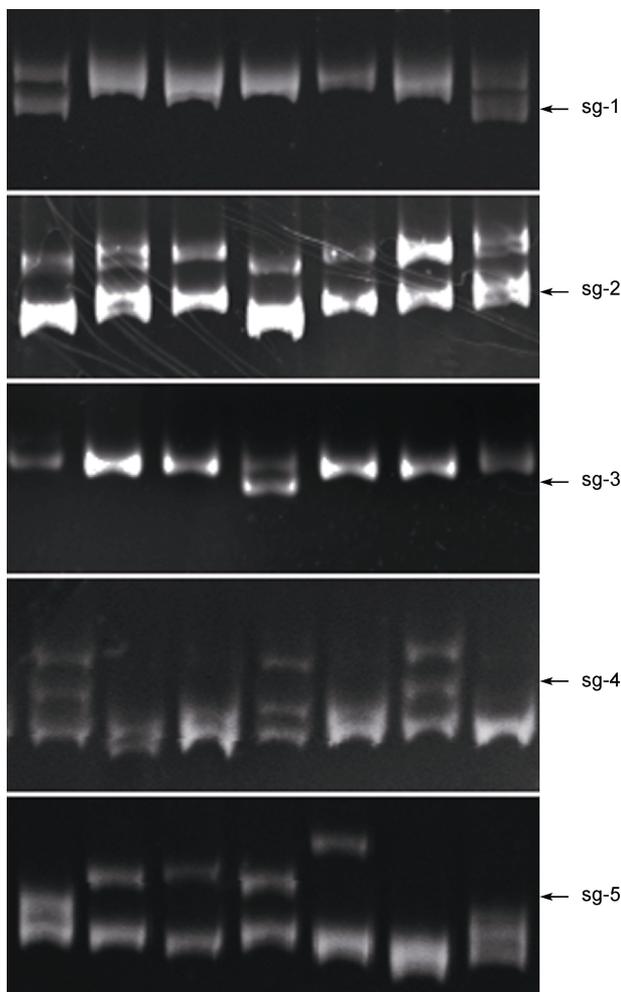


图 1 5 对微卫星引物对大竹蛭部分个体基因组扩增的非变性聚丙烯酰胺凝胶电泳分析

Fig.1 Non-denaturing polyacrylamide gel electrophoresis analysis on several individuals of *S. grandis* with five microsatellite primer pairs

和 SSR 标记的优点,一方面,由于 SSR 标记较高的多态性,在后代中可以获得多样的基因型,因此提高了

整合到遗传图谱中的可能性;另一方面,由于 EST 标记对应着特定的基因,可以在遗传图谱中指示出特定基因的位置,因而在选择育种工作中,可以直观地观察到特定基因的遗传连锁情况,实现分子辅助育种。

关于大竹蛭群体遗传多样性已有报道,张志伟等(2009)利用 6 种同工酶对南通野生型群体的 12 个同工酶位点进行了分析,平均每个位点的等位基因数为 2.5 个,期望杂合度(H_e)为 0.3737,观测杂合度(H_o)为 0.3333,香农多样性指数(I)为 0.6242。Yuan 等(2010)利用磁珠法分离到的 18 对微卫星引物对山东大竹蛭群体的 30 个个体进行了分析,平均每个位点的等位基因数为 14.6 个,期望杂合度(H_e)平均值为 0.886,观测杂合度(H_o)平均值为 0.711。本文以 5 对微卫星引物对南通大竹蛭 30 个个体进行了种群遗传结构的分析,平均每个位点的等位基因数目为 4 个,多样性高于同工酶方法,观测杂合度、期望杂合度、香农指数也略高(表 2),但是显著低于磁珠法获得的微卫星标记。推测原因,可能是同工酶主要以编码区的插入或缺失突变形成多态性位点,但插入和缺失概率低于微卫星序列剪切重复单元产生多态性的方式,因而造成了 EST-SSR 标记的多态性高于同工酶的现象,但是 EST-SSR 毕竟位于保守的编码区,其序列多用于合成功能蛋白,反之,处于非编码区的 SSR,由于编码功能缺失导致其序列多变,因而 EST-SSR 标记的多态性显著低于磁珠法分离的 SSR 标记(以非编码区的 SSR 为主)。

总之,对 EST-SSR 标记的分离宜采用相同的筛选标准以期与其它物种进行比较,EST-SSR 标记具有与功能基因相关联等优点,但其多态性低于非编码区 SSR。本文获得的 5 个多态性微卫星标记适合于大

表 3 4 种水产动物 EST 库中筛选微卫星序列的比较
Tab.3 Comparison of screening microsatellite sequences from four aquaculture animals' ESTs

水产动物	EST 库容	查找标准*	微卫星占 ESTs 的比率(%)	双碱基重复比率(%)	三碱基重复比率(%)	双碱基主要类型	三碱基主要类型
马氏珠母贝	6979	(2/7) (3/5)	3.48	48.50	31.00	(AT/TA) _n	(TAA/ATT) _n (AAG/CTT) _n
中国对虾	10446	(2/5) (3/4)	2.19	63.76	25.33	(AT/TA) _n	(TAA/ATT) _n
刺参	6632	(2/5)	6.48	65.86	26.68	(AT/TA) _n	(TAA/ATT) _n
大竹蛭	2038	(2/5)	4.71	74.76	22.33	(AT/TA) _n (AG/GA) _n	(CAA/TTG) _n

注: *数值表示为重复单元碱基数/最少重复次数

竹蛭种群遗传多样性的分析,并具有在其它相近物种如长竹蛭和缢蛭中作种间遗传比较的潜力。

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
- 孙国华, 杨建敏, 宋志乐等, 2010. 刺参(*Apostichopus japonicus*)EST 序列中微卫星分布分析及其标记的筛选. 海洋与湖沼, 41(1): 133—139
- 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
- 张志伟, 姚国兴, 陈爱华, 2009. 大竹蛭同工酶组织特异性与多态性初步研究. 海洋科学, 33(3): 41—96
- 罗文永, 胡骏, 李晓方, 2003. 微卫星序列及其应用. 遗传, 25(5): 615—619
- 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
- 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
- Ellis J R, Burke J M, 2007. EST-SSRs as a resource for population genetic analyses. Heredity, 99: 125—132
- Hamada H, Petrino M G, Kakunaga T, 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci, 79: 6465—6469
- Liu Z, Attila K, Rex A D, 1999. Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. Mar Biotechnol, 1: 437—447
- McCouch S R, Chen X P, 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, 12(10): 4127—4138
- Tautz, D, Trick M, Dover G A, 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature, 322: 652—656
- Toth G, Gaspari Z, Jurka J, 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research, 10: 967—981
- 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
- Yuan Y, Li Q, Kong L *et al*, 2010. Isolation and characterization of polymorphic microsatellite loci in the grand jackknife clam *Solen grandis* (Bivalvia: Veneroida). Genes & Genomics, 32: 191—197

ISOLATION AND CHARACTERIZATION OF MICROSATELLITE MARKERS FROM cDNA LIBRARY OF *SOLENI GRANDIS*

QIAO Hong-Jin¹, LIU Xiang-Quan¹, SUN Guo-Hua¹, FANG Jing-Hui¹,
WEI Xiu-Mei¹, ZHANG Tao²

(1. *Shandong Marine Fisheries Research Institute, Shandong Provincial Key Laboratory of Restoration for Marine Ecology, Yantai, 264006*; 2. *School of Life Science, Ludong University, Yantai, 264025*)

Abstract The distribution of 2—6bp simple sequence repeats (SSRs) were analyzed from cDNA Library (2038 ESTs) of *Solen grandis* using the software SSRIT. When the minimum repeats were assigned as 5, 96 sequences were found to contain microsatellite sequences (taking 4.71%). Up to 103 microsatellite loci were obtained, among which the dinucleotide repeats motifs were the most abundant type (taking 74.76%), followed by trinucleotide and tetranucleotide repeats of 22.33%, 2.91% respectively, and pentanucleotide and hexanucleotide repeats were not found. 14 primer pairs were designed with Primer 3 software from EST sequences containing microsatellites, and 5 primers had polymorphic PCR products in Nantong wild population by PCR and polyacrylamide gel electrophoresis analysis. In these 5 microsatellite loci, the allele numbers, the observed heterozygosity and expected heterozygosity distributed from 2 to 7, 0.067 to 1.000 and 0.066 to 0.775, respectively. Shannon's diversity index ranged from 0.146 to 1.545. The results showed that 5 microsatellite loci were suitable for genetic analysis.

Key words *Solen grandis*, Microsatellite, EST, Allele