

# 缢蛏(*Sinonovacula constricta*)EST-SSR 分布特征 及引物开发利用\*

刘 博 邵艳卿 滕爽爽 柴雪良 肖国强

(浙江省海洋水产养殖研究所 温州 325000; 浙江省近岸水域生物资源开发与保护重点实验室 温州 325000)

**提要** 采用 CAP3 软件对 NCBI 上的 5296 条缢蛏 ESTs 序列进行了微卫星特征分析。结果表明, 经拼接、去冗得到非冗余 EST 序列 3453 条, 含 SSR 位点的 EST 序列 267 条, 共 307 个 SSR 位点, 检出率为 8.89%, 平均每 6.83kb 出现 1 个 SSR 位点。设计了 40 对 EST-SSR 引物并进行 PCR 扩增, 29 对引物能扩增出理想的 PCR 产物, 其中多态性引物 14 对。利用 14 对多态性引物分析了乐清湾缢蛏遗传多样性, 共检测到等位基因数( $N_a$ )61 个, 每个位点的等位基因数为 2—12 个。二核苷酸、三核苷酸和四核苷酸重复是最主要的重复类型, 分别占 15.96%、37.13%和 35.50%。乐清湾缢蛏群体观测杂合度( $H_o$ )、期望杂合度( $H_e$ )和多态信息含量( $PI_C$ )分别为 0.569、0.490 和 0.449, 表明乐清湾缢蛏遗传多样性较丰富。

**关键词** 缢蛏, 表达序列标签(EST), 简单重复序列(SSR)

**中图分类号** S968.3

缢蛏(*Sinonovacula constricta* Lamarck)俗称蛏子, 隶属软体动物门(Mollusca)、瓣鳃纲(Lamellibranchia)、竹蛏科(Solenidae)、缢蛏属(*Sinonovacula*), 为广温广盐性海产双壳类, 具有重要的经济价值, 是我国四大养殖贝类之一。缢蛏养殖历史悠久, 过去主要集中在福建和浙江一带, 随着缢蛏养殖业的不断扩展, 山东沿海等地区开始从福建一带引进苗种。浙江沿海为了能提早出苗, 每年 9 月份也大量从广东湛江和福建云霄等地购买亲本催产育苗。无序的养殖造成缢蛏资源混乱, 盲目引种、移养等使本土自然种质资源混淆不清。因此, 分析人工养殖条件下群体遗传多样性的变化及亲贝个体间遗传差异, 对缢蛏优良品种选育和苗种培育, 以及今后合理保护、开发和利用都有重要指导意义。然而, 缢蛏可有效利用的微卫星标记数量非常有限, 仅 Niu 等(2008)发表了关于缢蛏基因组 SSR 标记开发的初步报告。目前迫切需要开发更多有效的分子标记, 以开展缢蛏遗传学研究。

在众多分子标记中, 简单序列重复(simple sequence repeat, SSR)亦称微卫星, 与其它分子标记相比, 具有多态性高、多等位性、共显性、可重复性高、数量丰富和对基因组有很好的覆盖性等特点(Powell *et al.*, 1996)。微卫星引物的开发可以从公共核酸序列数据库中查询, 但是针对初次研究的物种, 传统的 SSR 引物开发通常需要经过构建基因组文库、探针杂交、克隆、测序来完成, 存在费时费力, 成本较高, 效率较低, 同位素污染等问题(Rassmann *et al.*, 1991)。近年来, 许多贝类大规模的 cDNA 测序并在网上公开释放, 大大增加了基于 EST(expressed sequence tag)的 SSR 标记开发能力。与传统 SSR 标记相比, EST-SSR 标记不仅具有传统 SSR 标记多态性高、共显性、重复性好等特点, 更重要的是它的开发成本较低, 节省了大量人力物力, 而且由于 EST-SSR 来源于表达的基因组区域(Varshney *et al.*, 2005), 可直接反映相关基因的多样性, 同时它在不同物种间有良好的通用

\* 温州市科技兴海项目, S20080019 号; 浙江省近岸水域生物资源开发与保护重点实验室开放基金, J2010009 号; 浙江省海水养殖重点科技创新团队资助, 2010R50025 号; 国家贝类产业技术体系浙江综合试验站资助, CARS-48 号。刘 博, 博士, E-mail: liubomusic@126.com

通讯作者: 柴雪良, 教授级高工, E-mail: cxl-5888@163.com

收稿日期: 2011-10-26, 收修改稿日期: 2011-12-08

性(Yu *et al.*, 2004; Poncet *et al.*, 2006; Jia *et al.*, 2007; 孙国华等, 2010; 上官凌飞等, 2011)。本文从 GenBank 公布的缙蛭 EST 中发掘了含 SSR 的 EST 序列, 并研究了这些 EST-SSR 在缙蛭 EST 中的分布特征, 随机设计合成了部分 EST-SSR 引物, 并以此对乐清湾缙蛭群体进行了遗传结构分析。本研究将为缙蛭种质资源遗传多样性分析、功能基因发掘和分子标记辅助选择育种奠定基础。

## 1 材料与方法

### 1.1 EST-SSR 的搜索与预处理

从 NCBI 公共数据库 dbEST/GenBank (<http://www.ncbi.nlm.nih.gov/dbEST/>)以 FASTA 格式下载了全部 5296 条 EST 序列(截止到 2011 年 9 月 30 日), 利用 CAP3 软件对全部 EST 进行冗余性查找和拼接。然后利用 Pimer3 软件对拼接后的 EST 进行 SSR 位点搜索, 搜索参数设置为重复单位为 2、3、4、5、6 个碱基, 重复次数依次为 6、4、3、3、3 次。

### 1.2 EST-SSR 引物设计

用 Primer3 软件设计 40 对 SSR 引物, 设计的主要参数: 引物长 18—23bp, 最适长度 21bp; PCR 产物 100—300bp, 最适长度 150bp; GC 含量 30%—70%, 最适含量 50%; 退火温度 50—70℃, 最适温度 55℃。设计出来的引物利用 Oligo 软件进行引物评估, 避免引物二聚体、发夹结构和错配等情况的发生。引物由上海生物工程有限公司合成。

### 1.3 DNA 提取、PCR 扩增及产物检测

实验用缙蛭 40 颗, 于 2010 年 10 月采自浙江省乐清湾, 现场取其水管及斧足放入无水乙醇固定保存, 带回实验室备用。缙蛭基因组 DNA 提取参照 Strauss(1989)的方法。PCR 反应体系为 15μl, 包括 10× PCR buffer 1.5μl, 25mmol/L Mg<sup>2+</sup>1.5μl, 2.5mmol/L dNTP 1.2μl, 10μmol/L 上下游引物各 1.5μl, Taq 酶 (5U)0.075μl, DNA 模板 1.5μl, 加 ddH<sub>2</sub>O 至 15μl。PCR 反应程序为: 94℃条件下变性 5min 后, 进入 30 个 PCR 循环: 在 94℃下变性 30s, 复性(温度见表 1)30s, 72℃下延伸 45s, 再最后 72℃下延伸 7min, 并将产物

于 4℃下保存。扩增产物用 1.2%琼脂糖凝胶检测, 鉴定有扩增产物后再用 8%的变性聚丙烯酰胺凝胶电泳检测, 银染。

### 1.4 EST-SSR 多态性检测及数据统计

根据电泳图谱, 并参照引物设计时预期片段的大小统计条带, 应用 POPGENE1.32 软件计算微卫星引物的期望杂合度(Expected heterozygosity,  $H_e$ )和观察杂合度(Observed heterozygosity,  $H_o$ )。多态信息含量(Polymorphic information content, PIC)的计算参照 Botstein 等(1980)的方法, 计算公式如下:

$$PIC = 1 - \sum_{i=1}^n P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2P_i^2 P_j^2$$

其中,  $P_i$ 、 $P_j$  是某个位点第  $i$ 、 $j$  个等位基因的基因频率,  $n$  为某个位点上的等位基因数。

## 2 结果

### 2.1 缙蛭 EST-SSR 分布、发生频率及特点

利用 CAP3 软件将 NCBI 网上公开的 5296 条缙蛭 ESTs 序列进行拼接、去冗余后得到 3453 条 Unigene, 其中单序列(singletons)2837 条, 重叠序列(contigs) 616 条, 拼接总长度为 2096.17kb。在非冗余序列中发现含有 2—6 核苷酸重复的 EST 序列 267 条, 共 307 个 SSRs 位点, 平均每 6.83kb 出现 1 个 SSR 位点, SSR 出现频率为 8.89%。在发现的 307 个 SSRs 位点中, 只含有 1 个 SSR 位点的 EST 为 240 条, 含 2 个 SSRs 位点的 EST 为 21 条, 含 2 个 SSRs 位点的 EST 为 21 条, 含 3 个 SSRs 位点的 EST 为 5 条, 含 4 和 6 个 SSRs 位点的 EST 各 1 条。

在检测出的 307 个 SSR 位点中, 2—6 核苷酸重复都有出现, 其中二核苷酸重复有 5 种类型, 三核苷酸重复有 22 种类型, 四核苷酸重复有 46 种类型, 五核苷酸重复有 22 种类型, 六核苷酸重复有 7 种类型(表 1)。在所有的重复基元类型中含有三核苷酸重复类型的 SSR 数量最多 114 个, 占总 EST-SSR 数量的 37.13%, 其次是含四核苷酸重复类型的 SSR 数量为 109 个, 占总 EST-SSR 的 35.50%, 最少的六核苷酸重复类型的 SSR 占总 EST-SSR 的 2.28% (表 1)。

表 1 EST-SSR 引物的特征  
Tab.1 Distribution characteristics of EST-SSR of *S. constricta*

| 重复基元类型 | 重复基元类型数量 | EST-SSR 数量 | 占总 EST-SSR 比例(%) | EST-SSR 分布频率(%) |
|--------|----------|------------|------------------|-----------------|
| 二核苷酸   | 5        | 49         | 15.96            | 1.42            |
| 三核苷酸   | 22       | 114        | 37.13            | 3.30            |
| 四核苷酸   | 46       | 109        | 35.50            | 3.16            |
| 五核苷酸   | 22       | 28         | 9.12             | 0.81            |
| 六核苷酸   | 7        | 7          | 2.28             | 0.20            |

在含有二核苷酸重复基元的缙蛭 EST 序列中, AT/TA 基序出现频率最高(42.00%), 三核苷酸重复基元中, AAC/GTT 基序出现频率最高(13.16%), 四核苷酸重复基元中, AAAT/ATTT 基序出现频率最高(11.01%)(表 2)。五核苷酸重复基序共有 22 种, 出现频率最高的为 TAAAA/TTTTA, 共出现 3 次; 六核苷酸重复基序共有 7 种, 出现频率均为 1 次。

表 2 缙蛭 EST 中主要重复基元及频率

Tab.2 The major motifs and their frequency in EST of *S. constricta*

| 重复基元类型 | 重复基元      | 数量 | 各重复类型总数 | 频率(%) |
|--------|-----------|----|---------|-------|
| 二核苷酸   | AT/TA     | 21 | 50      | 42.00 |
|        | TG/CA     | 9  |         |       |
|        | AG/CT     | 8  |         |       |
|        | AC/GT     | 6  |         |       |
|        | TC/GA     | 6  |         |       |
| 三核苷酸   | AAC/GTT   | 15 | 114     | 13.16 |
|        | ACA/TGT   | 13 |         |       |
|        | TTG/CAA   | 12 |         |       |
|        | ATG/CAT   | 9  |         |       |
|        | TTA/TAA   | 9  |         |       |
|        | TAT/ATA   | 8  |         |       |
|        | TGA/TCA   | 7  |         |       |
|        | AAG/CTT   | 6  |         |       |
|        | AAT/ATT   | 6  |         |       |
|        | TTC/GAA   | 5  |         |       |
|        | GAT/ATC   | 5  |         |       |
| 四核苷酸   | AAAT/ATTT | 12 | 109     | 11.01 |
|        | AACA/TGTT | 8  |         |       |
|        | TTTA/TAAA | 8  |         |       |
|        | AAAC/GTTT | 8  |         |       |
|        | TATT/AATA | 7  |         |       |
|        | TTAT/ATAA | 7  |         |       |
|        | ACAT/ATGT | 7  |         |       |

由图 1 可知, 在重复基元重复次数相同时, 随着重复基元所含核苷酸数的增加 EST-SSR 的数量减少。另外, 不同 SSR 重复基元的出现频率与其重复次数的相关统计分析结果表明: 随着重复次数(或 SSR 长度)的增加, 基本上二核苷酸、三核苷酸、四核苷酸、五核苷酸及六核苷酸 EST-SSR 序列的频率均呈明显下降趋势, 且当 EST-SSR 长度超过 30bp 时, 二核苷酸、三核苷酸、四核苷酸、五核苷酸及六核苷酸 EST-SSR 序列出现的频率均近零。

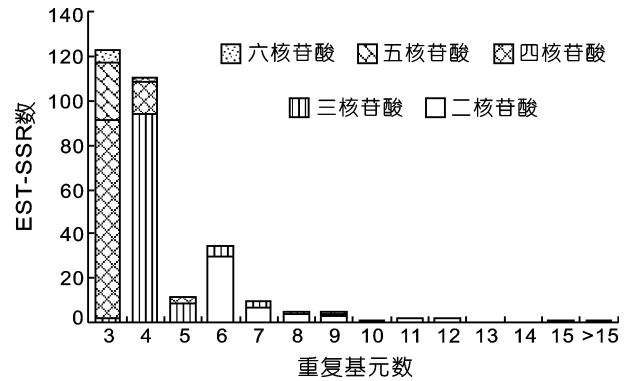


图 1 基于不同重复基元数 EST-SSR 的数量分布

Fig.1 Frequency distribution of the EST-SSR based on the number of motifs

## 2.2 缙蛭 EST-SSR 引物筛选

利用 Primer3 软件, 对 267 条含有 SSR 的 EST 进行引物设计, 共设计了 307 对 EST-SSR 引物。随机选取了 40 对 EST-SSR 引物, 以 8 个乐清湾缙蛭的基因组 DNA 为模板, 依据理论, 退火温度上下波动 5°C 左右为梯度, 对 40 对 EST-SSR 引物进行了初步筛选。研究发现扩增出稳定的可重复条带的引物为 29 对, 占引物总数的 72.5%。通过变性聚丙烯酰胺凝胶电泳结合硝酸银染色检测, 其中 14 对微卫星引物表现出个体多态性(表 3), 多态引物占引物总数的 35%, 多态性引物占有有效引物的比例为 48.28%。

## 2.3 乐清湾缙蛭 EST-SSR 等位基因与遗传变异分析

利用初步筛选出的 14 对多态性 EST-SSR 引物对 40 颗乐清湾缙蛭群体进行遗传变异分析, 结果见表 4。共检测到 61 个等位基因, 每对引物在乐清湾养殖群体中分别检测到 2—12 个数量不等的等位基因, 平均每个位点 4.4 个, 最多的为 12 个(YC-30), 最少的为 2 个(YC-8), 等位基因预期大小在 130—200bp 之间。各位点的平均观测杂合度( $H_o$ )为 0.569, 平均期望杂合度( $H_e$ )为 0.490, 平均多态信息含量(PIC)为 0.449。依据 PIC 判断标准:  $PIC < 0.25$  为低度多态,  $0.25 < PIC < 0.5$  为中度多态,  $PIC > 0.5$  为高度多态(徐美佳等, 2009), 可以看出缙蛭 ESR-SSR 在乐清湾缙蛭群体中表现出较高的多态性, 同时表明乐清湾缙蛭遗传多态性水平较高。

## 3 讨论

EST-SSR 标记来源于基因的编码区, 作为新型的分子标记, 在种质遗传评价、系谱鉴定、连锁图谱构建以及分子标记辅助选择等研究领域展现出了广

表 3 缢蛭部分多态性 EST-SSR 引物序列及参数  
Tab.3 Characters sequences of some polymorphic EST-SSR in *S. constricta*

| 引物    | 重复单元                 | 上游引物序列(5'—3')          | 下游引物序列(5'—3')         | 扩增片段大小 | 退火温度(°C) |
|-------|----------------------|------------------------|-----------------------|--------|----------|
| YC-1  | (ATCT) <sub>5</sub>  | CGTTTTGAACGTTACATTGTT  | TAACTTTCTCTGCAGCTTGAC | 133    | 54       |
| YC-3  | (ACAT) <sub>3</sub>  | TCAAATTGTGATCCGTACTCT  | CACCATTACAAAACATCACCT | 148    | 54       |
| YC-8  | (AAAGT) <sub>3</sub> | CTCATAAGTACCCCTGTGCTA  | AATACCCACGTAACATTCTG  | 150    | 56       |
| YC-10 | (TTTAA) <sub>3</sub> | CTATAAATGGTGTCCGGTTTTG | TAGACCTTCGAACTGTTTGTG | 160    | 55       |
| YC-11 | (TACA) <sub>4</sub>  | ACCAGTTTGTGTCTGCATATT  | GTGATAACATTTGTACACCA  | 150    | 54       |
| YC-12 | (TTAT) <sub>3</sub>  | TCAAGAGATTTTACAGCAAGC  | AACTGTGTCAAATGGTGATTC | 148    | 54       |
| YC-13 | (AAAC) <sub>3</sub>  | TGGTAATGATTGAAAAAGTGG  | TTGGCCAATATGATGTACTCT | 170    | 53       |
| YC-16 | (TAAA) <sub>3</sub>  | GGCCTAATGGTGTATGTTGT   | CATGCATGTTATTTCTTGTCA | 171    | 54       |
| YC-20 | (TGT) <sub>4</sub>   | GTTGTTGTTTGTGCTTTGTT   | AAATATCCAATGACTTTGCAG | 175    | 52       |
| YC-24 | (GATT) <sub>5</sub>  | TCCTTATCAACTTCACAGTGG  | ACTGCATTGCCATATAATCAC | 155    | 55       |
| YC-27 | (AAC) <sub>5</sub>   | CCTATAATGTAGTCCGGTCT   | CCCATAAGATGAGAACAGTGA | 142    | 57       |
| YC-30 | (TGTC) <sub>3</sub>  | TATGTCTGTCTGTCTGCCTGT  | TCTGAAAATAGACGGACTGG  | 167    | 57       |
| YC-33 | (GACA) <sub>3</sub>  | TAAAGAAAACCGAACACAGTT  | TTTTCAGGAATTTACAACAGG | 148    | 52       |
| YC-35 | (AC) <sub>8</sub>    | GAGCCTGATCTATTGCCTTAT  | TGAGAATGTTTCTTCTACAGG | 145    | 55       |

表 4 乐清湾缢蛭群体 EST-SSR 分析统计结果  
Tab.4 Summary of statistics for Yueqing Gulf *S. constricta* population with EST-SSR analysis

| 基因座位  | 等位基因数( $N_a$ ) | 观测杂合度( $H_o$ ) | 期望杂合度( $H_e$ ) | 多态信息含量(PIC) | 登录号        |
|-------|----------------|----------------|----------------|-------------|------------|
| YC-1  | 3              | 0.694          | 0.387          | 0.353       | G0309043.1 |
| YC-3  | 3              | 0.692          | 0.3103         | 0.288       | G0309023.1 |
| YC-8  | 2              | 0.617          | 0.375          | 0.304       | G0308992.1 |
| YC-10 | 4              | 0.600          | 0.337          | 0.308       | G0308934.1 |
| YC-11 | 3              | 0.868          | 0.629          | 0.557       | G0308851.1 |
| YC-12 | 4              | 0.794          | 0.3776         | 0.333       | G0308900.1 |
| YC-13 | 5              | 0.487          | 0.572          | 0.533       | G0308860.1 |
| YC-16 | 4              | 0.678          | 0.651          | 0.591       | G0308806.1 |
| YC-20 | 5              | 0.410          | 0.474          | 0.429       | G0310715.1 |
| YC-24 | 3              | 0.789          | 0.258          | 0.242       | G0309511.1 |
| YC-27 | 3              | 0.666          | 0.486          | 0.438       | G0309427.1 |
| YC-30 | 12             | 0.179          | 0.850          | 0.834       | G0309357.3 |
| YC-33 | 4              | 0.461          | 0.440          | 0.413       | G0309218.1 |
| YC-35 | 6              | 0.025          | 0.710          | 0.662       | G0309189.1 |
| 平均值   | 4.4            | 0.569          | 0.490          | 0.449       |            |

阔的前景。随着 EST 和 cDNA 大规模测序的开展(秦玉明等, 2010), NCBI 的 EST 数据库容量在不断增长, 人们相继在许多海产贝类中开始进行 EST-SSR 引物的开发利用(石耀华等, 2008; Wang *et al.*, 2009; 李云峰等, 2010)。本研究对 5269 条缢蛭 EST 序列进行拼接去冗后, 共得到 Unigene3453 条, 包括单序列(singletons)2837 条, 重叠序列(contigs)616 条, 对这些 EST 序列进行检索, 共在 267 条 EST 序列中发现共 307 个 SSR 位点, 平均每 6.83kb 出现 1 个 SSR 位点, SSR 出现频率为 8.89%, 该比例远高于马氏珠母贝

(石耀华等, 2008)、栉孔扇贝(Wang *et al.*, 2009)和虾夷扇贝(李云峰等, 2010)的 EST-SSR 出现频率 3.84%、0.60%、3.69%。相比之下, 本研究得到的微卫星比例较高。分析差异的原因可能是由于 SSR 查询程序所采用的标准、可利用的 EST 数据库中的数据信息量、组织来源以及物种间的不同所造成的。研究结果表明相对传统基因组 SSR 而言, 利用 NCBI 数据库中公布的缢蛭 EST 序列信息, 开发缢蛭 EST-SSR 标记是切实可行的一条开发捷径。

从目前报道的结果来看, 马氏珠母贝(石耀华等,

2008)和虾夷扇贝(李云峰等, 2010)的 EST-SSR 都是以二核苷酸重复基元为主, 其比例分别为 48.5%、40.54%, 而本研究以三核苷酸重复基元为主, 占 EST-SSR 总数的 37.13%, 该结果与大多数植物的 EST-SSR 以三核苷酸为主要重复基元一致。这是由于 EST 是表达序列, 大部分由外显子组成, 加上密码子以三核苷酸为一个功能单位, 在被翻译成蛋白质时受到移码突变而造成了三核苷酸位移, 但不会对一个表达基因的阅读框造成太大的影响(Metzgar *et al.*, 2000), 该理论已在谷类作物中得到了验证(Temnykh *et al.*, 2001; Gao *et al.*, 2004)。

Eujayl 等(2002)通过对来自基因组和 EST 数据库的 SSR 多态性的比较, 发现来自基因组的 SSR 多态性较高为 53%, 来自 EST 数据库的 SSR 多态性较低为 25%, 但 EST-SSR 的扩增效果较好。本研究也得到了与之基本一致的结果, 本实验开发了 14 对多态性 SSR 分子标记, 多态性出现率(48.28%), 高于缢蛭基因组中 SSR 的多态性出现频率(44.44%), 但来自 EST 序列的 SSR 平均等位基因数(4.4)低于来自基因组 SSR 的平均等位基因数(10.0) (Niu *et al.*, 2008)。本研究随机合成的引物均严格按照标准设计, 但仍有 11 对引物不能扩增出产物, 主要可能有: (1) 引物的序列落在两个外显子上; (2) 基因组上的对应序列含有内含子而不具备 SSR 序列的特征, 扩增不出产物; (3) 设计引物的 EST 序列存在问题, 或 cDNA 克隆比较荒诞(Varshney *et al.*, 2005)。

另外, 本实验出现了部分引物扩增产物大小与预期片段大小不一致的现象, 该情况在栉孔扇贝的 EST-SSR 扩增产物中也有出现(李红蕾等, 2003; Nicot *et al.*, 2004) 研究认为相对于基因组 SSR 扩增, EST-SSR 扩增产物更易偏离预期目的片段, 究其原因, 可能为: (1) 引物非特异性结合; (2) 所扩增的基因序列中出现较小的内含子或插入/缺失突变, Saha 等(2004)通过相关序列分析证实了该结论; (3) EST 序列的复制品分布于相同或不同的染色体; (4) EST 属于多基因家族, 对应的引物设计位于其保守区内。

尽管相对于基因组 SSR 引物来说, EST-SSR 引物多态性偏低, 但其具有开发成本低、扩增效率高、信息含量丰富等优点, 同时 EST 本身是功能基因的一部分序列, 从 EST 中开发出的 SSR 引物理论上可为功能基因提供“绝对”的标记。因此, 这些 EST-SSR 标记将有效用于缢蛭种质资源遗传多样性分析、功能基因发掘和分子标记辅助选择育种奠定坚实基础。

## 参 考 文 献

- 上官凌飞, 李晓颖, 宁 宁等, 2011. 杏 EST-SSR 标记的开发. 园艺学报, 38(1): 43—54
- 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
- 李云峰, 刘卫东, 高翔刚等, 2010. 虾夷扇贝外套膜和肾脏组织 cDNA 文库构建以及 EST 的初步分析. 中国水产科学, 17(3): 578—585
- 李红蕾, 宋林生, 王玲玲等, 2003. 栉孔扇贝 EST 中微卫星标记的筛选. 高技术通讯, 12: 72—75
- 秦玉明, 苏秀榕, 李 晔等, 2010. 缢蛭(*Sinonovacula constricta*) cDNA 文库的构建及肌动蛋白基因的研究. 海洋与湖沼, 41(1): 54—60
- 徐美佳, 张 妍, 鲁翠云等, 2009. 利用 EST-SSR 座位对鲤鱼 4 种生长性状的单标记回归分析. 水产学杂志, 22(1): 15—19
- Botstein D, White R L, Skolnick M *et al.*, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, 32: 314—331
- Eujayl I, Sorrells M E, Baum M *et al.*, 2002. Isolation of EST-derived microsatellite marker for genotyping the A and B genomes of wheat. *Theor Appl Genet*, 104(2—3): 399—407
- Gao L F, Jing R L, Huo N X *et al.*, 2004. One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. *Theor Appl Genet*, 108(7): 1392—1400
- Jia Xiao-ping, Shi Yun-su, Song Yan-cun *et al.*, 2007. Development of EST-SSR in foxtail millet (*Setaria italica*). *Genetic Resources and Crop Evolution*, 54: 233—236
- Metzgar D, Bytof J, Wills C *et al.*, 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res*, 10: 72—80
- Nicot N, Chiquet V, Gandon B *et al.*, 2004. Study of simple sequence repeat (SSR) markers from wheat expressed sequenced tags (ESTs). *Theor Appl Genet*, 109: 800—815
- Niu D H, Li J L, Liu B D, 2008. Polymorphic microsatellite loci for population studies of the razor clam *Sinonovacula constricta*. *Conserv Genet*, 9: 1393—1394
- Poncet V, Rondeau M, Tranchant C *et al.*, 2006. SSR mining in coffee tree EST databases: Potential use of EST-SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics*, 276: 436—449
- Powell W, Machray G C, Provan J, 1996. Polymorphism revealed by simple sequence repeats. *Trends Plant Science*, 1: 215—222
- Rassmann K, Schlötterer C, Tautz D, 1991. Isolation of simple sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis*, 12(2—3): 113—118
- Saha M C, Mina M A, Eujayl I *et al.*, 2004. Tall fescue EST-SSR

- markers with transferability across several grass species. *Theor Appl Genet*, 109: 783—791
- Strauss W M, 1989. Preparation of Genomic DNA from Mammalian Tissues. In: Ausubel F M, Brent R, Kingston R E ed. *Current Protocol in Molecular Biology*. New York, John Wiley and Sons, 221—222
- Temnykh S, DeClerck G, Lukashova A *et al*, 2001. Computational and experimental analysis of microsatellite in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res*, 11: 1441—1452
- Varshney R K, Graner A, Sorrells M E *et al*, 2005. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, 23(1): 48—55
- Wang X, Song B, Qiu X, 2009. Development of EST-SSRs in scallop (*Patinopecten yessoensis*) from sequence database. *Conserv Genet*, 10(4): 1129—1131
- Yu J K, Rota M L, Kantety R V, 2004. EST derived SSR markers for comparative mapping in wheat and rice. *Molecular Genetics and Genomics*, 271: 742—751

## CHARACTERIZATION, DEVELOPMENT AND UTILIZATION OF EST-DERIVED MICROSATELLITES IN *SINONOVACULA CONSTRICTA*

LIU Bo, SHAO Yan-Qing, TENG Shuang-Shuang, CHAI Xue-Liang, XIAO Guo-Qiang  
(Zhejiang Mariculture Research Institute, Wenzhou, 325000; Zhejiang Key Laboratory of Exploitation and Reservation of Coastal Bio-resource, Wenzhou, 325000)

**Abstract** A set of 5296 expressed sequence tags (ESTs) of *Sinonovacula constricta* from NCBI were screened using CAP3.0 software. After assembly, a total of 3453 non-redundant EST sequences were mined and yielded 267 non-redundant sequences which contained 307 putative SSRs accounted for 8.89% of the total number of EST sequences. This is equivalent to one EST-SSR per 6.83kb of sequence. Di-, Tri- and Tetranucleotide repeats EST-SSRs were dominant, counting for 15.96%, 37.13% and 35.50% respectively. Forty primers were designed for partial EST-SSRs and used for PCR amplification with 29 primer pairs showing amplifications while 14 were polymorphic. These were used to conduct SSR analyses of genetic variations of the 40 individual samples of *S. constricta* collected in Yueqing Bay. A total of 61 alleles were detected, with the number of alleles ranging from 2 to 12. The observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) and polymorphism information content ( $PIC$ ) were 0.569, 0.490 and 0.449, respectively. All of these parameters indicate that the *S. constricta* of Yueqing Bay rich genetic variation.

**Key words** *Sinonovacula constricta*, Expressed sequence tags (EST), Simple sequence repeats (SSR)