

文章编号: 1000-0615(2003)03-0213-06

# 从中国对虾 ESTs 中筛选微卫星标记的研究

徐 鹏, 周令华, 田丽萍, 相建海

(中国科学院海洋研究所, 山东 青岛 266071)

**摘要:** 利用生物信息学方法在含有 10446 个中国对虾 ESTs 的数据库中进行微卫星序列的筛选, 共发现微卫星序列 229 个, 占整个 ESTs 数据库的 2.19%, 其中含双碱基重复序列 146 个和 3 碱基重复序列 58 个, 分别占在 ESTs 数据库中发现微卫星序列总数的 63.76% 和 25.33%, 大部分发现的微卫星序列均为 Perfect 形式的重复序列。根据筛选得到的微卫星序列设计并合成引物 19 对进行多态性检测, 在有扩增产物的 16 对引物中, 首次筛选得到 8 个中国对虾微卫星标记, 并对这些微卫星标记进行了等位基因频率、观测杂合度、期望杂合度、PIC 值等统计学指标的评价。结果表明, 在 8 个微卫星位点上, 等位基因的数目从 5 到 15 不等, 等位基因长度从 165~305bp, 期望杂合度和多态性信息含量分别为 0.59 到 0.89 和 0.56 到 0.88, 表明这 8 个中国对虾微卫星标记完全适合于遗传分析。

**关键词:** 中国对虾; 表达序列标签; 微卫星

中图分类号: S917.4 文献标识码: A

## Isolation of microsatellite markers from *Fenneropenaeus chinensis* ESTs

XU Peng, ZHOU Ling-hua, TIAN Li-ping, XIANG Jian-hai

(Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China)

**Abstract:** By using the on-line bioinformatics software Repeatmasker, the ESTs database of Chinese shrimp was screened. 229 microsatellite sequences were obtained from 10446 ESTs sequences, about 2.19% in the whole database. A total of 146 dinucleotide repeats and 58 trinucleotide repeats were found in these 229 microsatellite sequences and most of them were perfect repeats, with (AT)<sub>n</sub> in dinucleotide repeats and (TAA)<sub>n</sub> in trinucleotide repeats being the most common, about 28.82% and 17.03% in the total microsatellite sequences respectively. Nineteen primer pairs were designed from microsatellites using computer program Primer Premier (Version 5.00). Among the 16 primers that can be detected PCR products by electrophoresis, eight primers were polymorphic in about 20 Chinese shrimp samples by electrophoresis analysis. In these eight microsatellite loci, the alleles numbers distributed from 5 to 15, and the range of these alleles were between 165 to 305bp, similar to their predicted size. The expected heterozygosity and the PIC values of them distributed from 0.59 to 0.89 and 0.56 to 0.88, respectively. These values showed that these 8 microsatellite loci were suitable for genetic analysis.

**Key words:** *Fenneropenaeus chinensis*; expressed sequence tag (EST); microsatellite

收稿日期: 2002-09-24

资助项目: 国家重点基础研究发展项目 (G1999012007)、国家自然科学基金项目 (39910113)、国家高技术研究发展计划 (2001AA621050)

作者简介: 徐 鹏 (1977-), 男, 主要从事海洋动物分子遗传学研究。Tel: 0532-2898571

通讯作者: 相建海 (1946-), 男, 研究员, 主要从事海洋动物生殖与遗传学研究。Tel: 0532-2898571

微卫星标记是近年来迅速发展起来的一种遗传标记技术,已经成为遗传学领域和基因组学领域的研究热点。微卫星(microsatellite)是指以少数几个核苷酸(一般为2~6个)为基本重复单位的多次串联重复的DNA序列,亦称简单序列重复(simple sequence repeats, SSR)、短串联重复(short tandem repeats, STR)、简单序列长度多态性(simple sequence length polymorphism, SLP)等。微卫星核心序列突变率相对较高( $10^{-5} \sim 10^{-3}$ ),造成微卫星DNA核心序列重复次数的增加或减少,即微卫星DNA长度的变化,这是微卫星的多态性的根源<sup>[1]</sup>。微卫星DNA广泛而随机分布于真核生物的基因组中,与其他分子标记技术相比,微卫星标记具有更高的多态性,并且符合孟德尔遗传模式,呈共显性遗传,可以区分纯合显性个体和杂合显性个体,因此,在种群遗传结构分析、连锁图谱构建以及系谱认证等工作中,微卫星标记技术得到了广泛的应用。

中国对虾(*Penaeus chinensis*)是中国近海地方特有种,也是中国重要的经济虾类之一。中国对虾养殖业在我国的海水养殖业中占有举足轻重的地位。与陆生家养动物相比,中国对虾微卫星标记研究刚刚起步,目前仅有从小片段基因组文库中筛选获取微卫星序列的报道,还没有关于微卫星标记筛选和应用的报道<sup>[2,3]</sup>。在对虾科中,目前仅在一些重要的经济种类中如斑节对虾(*Penaeus monodon*)和凡纳滨对虾(*Litopenaeus vannamei*)等中筛选得到了一定数量的微卫星标记,并在种群遗传结构分析等方面进行了应用<sup>[4-9]</sup>。

ESTs即表达序列标签(expressed sequence tag),是指通过对cDNA文库测序得到的DNA序列,它能反映mRNA的信息,在功能基因的筛选和遗传图谱构建等工作中具有重要意义。从ESTs数据库中筛选微卫星标记的工作,在许多物种中已见报道,如甘蔗、水稻和玉米等重要农作物<sup>[10-14]</sup>。与构建小片段基因组文库进行微卫星序列的筛选相比,从ESTs数据库中获得微卫星序列要经济得多,可以节省大量的科研经费,是对ESTs数据资源的充分利用。本研究从中国对虾ESTs数据库进行了微卫星序列的查找和分析,设计引物,筛选微卫星标记,现报道如下。

## 1 材料和方法

### 1.1 中国对虾 ESTs 中微卫星序列的查找

构建的中国对虾ESTs数据库共由10466条cDNA序列组成<sup>[15]</sup>,这些序列来自中国对虾头胸甲组织构建的cDNA文库。使用Repeatmasker软件从这些ESTs中进行了微卫星序列的查找。查找中使用的标准是:7次以上重复的双碱基重复序列或5次以上重复的3碱基重复序列。

### 1.2 微卫星引物设计

从获得的微卫星序列中,选取符合引物设计的序列(一般选取重复次数在12次以上的双碱基重复序列或重复次数在5次以上的3碱基重复序列),使用引物设计软件Oligo(Version 6.31)和Primer Premier(Version 5.00)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100~300bp之间,并且最好避免二级结构。引物在上海基康生物技术公司合成。

### 1.3 模板 DNA 的提取

实验中所用中国对虾样品为购买自青岛市场的海捕野生对虾,运回实验室后于-80℃保存。模板DNA提取自中国对虾尾节肌肉,采用常规方法酚/氯仿抽提,乙醇沉淀获得。使用前经紫外分光光度计检测纯度,-20℃备用。

### 1.4 PCR 扩增与电泳检测和筛选

PCR反应体系(25μL)包含约100ng的基因组DNA,15pmol的引物,1×PCR Buffer,2.5μmolMgCl<sub>2</sub>,

0.2 μmol dNTP, Taq DNA 聚合酶 1U。PCR 反应条件为: 94 °C 4min, 然后 94 °C 40s, 退火温度(表 2) 40s, 72 °C 延长 40s, 一共 35 个循环, 之后 72 °C 10min, 4 °C 保温。

各对引物的最适退火温度通过梯度 PCR 结果确定。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。在聚丙烯酰胺凝胶电泳中使用粘合硅烷涂布玻璃板, 使凝胶紧密贴附在玻璃板上。待凝胶干燥后使用扫描仪记录电泳图谱。

淘汰单态性微卫星引物, 对多态性微卫星的电泳图谱使用凝胶分析软件 ImageMaster 1D Elite (Version 3.01) 进行微卫星带型分析并记录。

### 1.5 数据分析与评价

根据凝胶分析软件 ImageMaster 1D Elite (Version 3.01) 比照分子量标准, 对每个微卫星位点等位基因的长度 (base pairs) 进行分析, 并对数量进行统计, 计算每个等位基因的频率, 并计算下列统计量:

杂合度 (H), 用于衡量微卫星标记的信息含量程度。被研究的微卫星基因标记处于哈代-温伯格平衡时, 其杂合子个体的期望频率称为个体的杂合度, 当在一个标记位点的 m 个等位基因的频率分别为  $P_1, P_2, \dots, P_m$  时, 有:

$$\sum_{i=1}^m P_i = 1, H = 1 - \sum_{i=1}^m P_i^2$$

多态性信息含量 (polymorphic information content, PIC) 由下列公式计算:

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

## 2 结果

### 2.1 从 ESTs 数据库中筛选微卫星序列

在 10466 个 ESTs 序列中, 经过计算机筛选结合人工筛选共发现微卫星序列 229 个, 占整个 ESTs 数据库的 2.19%, 其中含有双碱基重复序列 146 个和 3 碱基重复序列 58 个, 分别占在 ESTs 数据库中发现微卫星序列总数的 63.76% 和 25.33%, 大部分发现的微卫星序列均为 Perfect 形式的重复序列。在这些微卫星序列中, (AT)<sub>n</sub> 形式在双碱基重复中最为常见, (TAA)<sub>n</sub> 形式在 3 碱基重复中最为常见, 分别占微卫星序列总数的 28.82% 和 17.03%。表 1 是从 ESTs 数据库中发现的微卫星序列的具体结果。

### 2.2 引物设计和多态性检测结果

根据以上微卫星序列, 共选取 19 条微卫星序列进行了引物的设计、合成与多态性检测, 其中 14 对引物来源于双碱基重复序列, 2 对引物来源于 3 碱基重复序列, 2 对引物来源于 4 碱基重复序列, 还有 1 对引物来源于 6 碱基重复序列。

经过 PCR 扩增、聚丙烯酰胺凝胶电泳和硝酸银染色, 共有 16 对引物有清晰产物, 经过多态性分析, 最终筛选出多态性微卫星引物 8 对。图 1 为部分微卫星位点电泳照片。

表 1 从中国对虾 ESTs 数据库中筛选所得的微卫星序列  
Tab.1 Microsatellite sequences isolated from ESTs database of *Penaeus chinensis*

| 重复类型 motif          | 数量 number | 百分比 (%) percentage |
|---------------------|-----------|--------------------|
| 2 碱基                |           |                    |
| di-                 |           |                    |
| (AT) <sub>n</sub>   | 66        | 28.82              |
| (TG) <sub>n</sub>   | 56        | 24.45              |
| (TC) <sub>n</sub>   | 24        | 10.48              |
| 3 碱基                |           |                    |
| tri-                |           |                    |
| (TAA) <sub>n</sub>  | 39        | 17.03              |
| (TTC) <sub>n</sub>  | 4         | 1.75               |
| (CAG) <sub>n</sub>  | 4         | 1.75               |
| (ATG) <sub>n</sub>  | 3         | 1.31               |
| (ACT) <sub>n</sub>  | 3         | 1.31               |
| 4 碱基                |           |                    |
| tetra-              |           |                    |
| (TAAA) <sub>n</sub> | 3         | 1.31               |
| (TAGA) <sub>n</sub> | 3         | 1.31               |
| (TATG) <sub>n</sub> | 5         | 2.18               |

注: 百分比小于 1% 的重复序列类型未在本表中列出  
Notes: The motifs less than 1% are not listed in this table

### 2.3 微卫星引物评价

使用凝胶分析软件 ImageMaster 1D Elite (Version 3.01) (Pharmacia Biotech Ltd.), 对每个微卫星位点的等位基因数目进行统计, 计算各个等位基因频率, 基因型频率, 并根据材料与方法中的公式计算杂合度(H)和多态性信息含量(PIC)。结果如表2所示。

### 3 讨论

实验共在 ESTs 数据库中发现各种类型的微卫星序列 229 个, 占 ESTs 总数的 2.19%, 这个比例和从许多其他物种的 ESTs 数据库中筛选出来的微卫星序列的比例大体相符, 如 Cordeiro 等<sup>[11]</sup> 在甘蔗 (*Saccharum* spp.) 的含有 8678 个序列的 ESTs 数据库中发现了近 250 个微卫星序列, 比例大约是 2.88% 左右; Sukhwinder 等<sup>[13]</sup> 在水稻、玉米和高粱等物种的 20000 多条 ESTs 序列中共筛选出了 250 余条微卫星序列, 比例约为 1.25% 左右; Attila 等<sup>[14]</sup> 在斑点叉尾 (*Italus punctatus*) 1909 个 ESTs 中发现了 89 条微卫星序列, 比例为 4.66%。可见, 在各个物种的 ESTs 中, 简单重复序列的含量都在 1%~5% 左右。

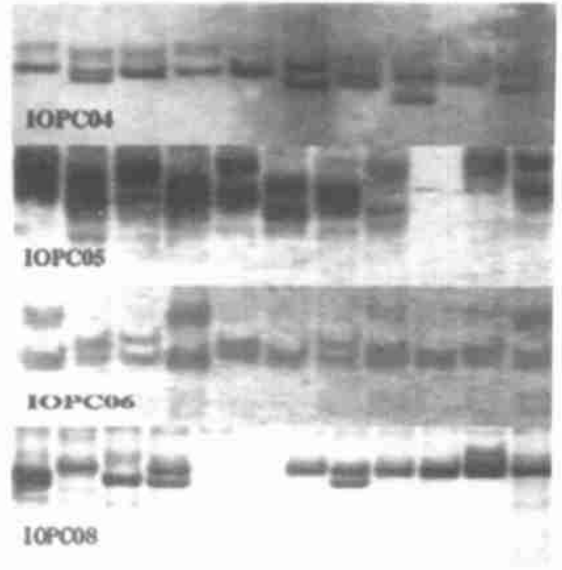


图1 部分微卫星引物 PCR 扩增、聚丙烯酰胺凝胶电泳银染照片 (示多态性, 各泳道为不同个体的中国对虾)

Fig. 1 The silver staining PAGE image of 4 microsatellite markers (Polymorphism, different lanes show different individuals)

表2 中国对虾 8 对微卫星引物及其评价

Tab. 2 Eight pairs of microsatellite primers and their evaluation

| 引物号<br>no. | 引物序列(5'-3')<br>primer sequences(5'-3')         | 核心序列<br>core sequences                    | 温度(°C)<br>T | 等位基因个数<br>(分布范围 bp)<br>alleles number(range) | H <sub>o</sub> | H <sub>e</sub> | PIC  |
|------------|--|---|-------------|--|----------------|----------------|------|
| IOPC04     | TCTGGAAAG AAATG AAAAGT<br>AATACAACAATCCTTTAGTC | (TTA) <sub>14</sub>                       | 55          | 10(220~292)                                  | 0.79           | 0.77           | 0.75 |
| IOPC05     | TTGCTTTAATGGTTGCTG<br>TCTACCAAGAATGGAGTG       | (TTA) <sub>14</sub>                       | 55          | 14(212~294)                                  | 0.96           | 0.82           | 0.80 |
| IOPC06     | TGACAACATAGCCAAGGAC<br>GCTATTCTGTGCTATTTC      | (TAAA) <sub>5</sub>                       | 55          | 15(199~342)                                  | 0.79           | 0.87           | 0.86 |
| IOPC07     | CAGGGCCGGCGAGTAGAA<br>CCACGCACGGTTCATTTCG      | (AG) <sub>29</sub>                        | 52          | 7(224~294)                                   | 0.59           | 0.59           | 0.56 |
| IOPC08     | TTCTTCGCCAGGAAA CAG<br>CGCGTACACAAAGCATA       | (TC) <sub>19</sub> ... (TC) <sub>33</sub> | 50          | 5(273~305)                                   | 0.19           | 0.62           | 0.58 |
| IOPC11     | GATCACATT AACGAGACAA<br>TA AATGTTATCGTTAGGGA   | (CA) <sub>7</sub> A (TA) <sub>22</sub>    | 52          | 9(165~215)                                   | 0.80           | 0.81           | 0.79 |
| IOPC14     | CGACCATTTTCGGTGTTC<br>GCTGCGATAATTGAGACG       | (AG) <sub>37</sub>                        | 55          | 13(191~259)                                  | 0.83           | 0.89           | 0.88 |
| IOPC17     | AATAATTGGG GAGGCCGAG<br>ATGGCAAGCGGAAGAATG     | (TG) <sub>36</sub>                        | 50          | 10(227~277)                                  | 1.00           | 0.88           | 0.87 |

注: T, 最适退火温度; H<sub>o</sub>, 观测杂合度; H<sub>e</sub>, 期望杂合度; PIC, 多态性信息含量

Notes: T means optimum annealing temperature; H<sub>o</sub> means observed heterozygosity; H<sub>e</sub> means expected heterozygosity; PIC means polymorphic information content

实验中,最普遍的微卫星类型是双碱基重复序列,其中最多的为 $(AT)_n$ ,这与双碱基重复序列在基因组中分布几率较高的事实是一致的,也和大多数研究者在不同物种之中的实验结果相符。尽管 Cordeiro 等<sup>[13]</sup>在甘蔗的 ESTs 数据库中发现 90% 以上的微卫星序列都是 3 碱基重复序列,但是,可以认为只是一种特例,在 ESTs 中分布最多的微卫星形式应该和在基因组中的分布是一样的,以双碱基重复序列最多。

微卫星的核心序列有各种类型,关于何种类型的微卫星序列筛选出多态性微卫星标记的可能性较大,一直以来众说纷纭。Pongsomboon 等<sup>[7]</sup>在斑节对虾中对 3 碱基和 4 碱基为基本重复序列的微卫星进行了标记筛选,在 7 对引物中筛选出 6 对微卫星标记,表明在 3 碱基和 4 碱基微卫星中筛选效率较高;Whan 等<sup>[12]</sup>在斑节对虾 ESTs 数据库中筛选出的两个微卫星位点的核心重复单位为 3 碱基和 5 碱基的;Xu 等<sup>[8,9]</sup>在斑节对虾中进行种群遗传学分析中使用的具有较强多态性的微卫星标记也多为 3 碱基或 4 碱基为基本重复单位的微卫星序列。在实验中,由于所得微卫星序列中主要形式为 2 碱基重复序列,且多为 Perfect 类型,所以主要选取了 2 碱基和 3 碱基为基本重复单位的微卫星序列进行了引物筛选,结果表明多态性比例较高。另外,尽量避免了选择 $(AT)_n$ 为核心重复单位的微卫星序列用于引物设计,这是为了避免在扩增中微卫星序列出现较多的二级结构而造成 PCR 效率下降,以及在电泳分离中造成较大误差。

在微卫星核心序列的重复次数方面,Weber<sup>[16]</sup>认为,只有在双碱基重复序列重复次数大于 12 次时,微卫星标记才有可能表现出较高 PIC 值,才可以进行相应的多态性分析。Xu 等<sup>[8]</sup>曾对斑节对虾进行微卫星引物设计,最后结果表明,凡是核心序列重复次数较少的微卫星标记,其结果或是单态性的,或等位基因数目非常少,PIC 值也偏低。本实验中所选用的微卫星序列其核心序列重复次数基本上都是在一个非常高的水平上,避免由于微卫星核心序列过短,造成微卫星标记筛选中多态性引物比例过低而造成浪费。

尽管在 EST 数据库中,微卫星资源并不丰富,而且大量的序列不完整,不适合引物设计,但是,从 EST 数据库中进行微卫星标记的筛选工作,不仅可以不经过构建小片段基因组文库等繁琐步骤,就得到适合于遗传学分析的微卫星标记,节省了大量的时间和经费,而且避免了筛选过程中的同位素操作,同时也是对 ESTs 数据库资源的充分利用。

## 参考文献:

- [1] Liu Z Y, Xiang J H. Application of microsatellite technique for genetic analysis in marine animals [J]. Marine Sciences, 2001, 25(6): 11-13. [刘志毅, 相建海. 微卫星 DNA 分子标记在海洋动物遗传分析中的应用 [J]. 海洋科学, 2001, 25(6): 11-13.]
- [2] Xu P, Zhou L H, Xiang J H. Rapid screening recombinant positive clones containing microsatellites of *Penaeus chinensis* with PCR [J]. J Fish China, 2001, 25(2): 127-130. [徐鹏, 周令华, 相建海. 用 PCR 法快速筛选中国对虾含微卫星序列的重组阳性克隆 [J]. 水产学报, 2001, 25(2): 127-130.]
- [3] Xu P, Zhou L H, Xiang J H. Isolating microsatellite DNA of Chinese shrimp *Penaeus chinensis* [J]. Oceanol et Limnol Sin, 2001, 32(3): 255-259. [徐鹏, 周令华, 相建海. 中国对虾微卫星 DNA 的筛选 [J]. 海洋与湖沼, 2001, 32(3): 255-259.]
- [4] Ball A O, Leonard S, Chapman R W. Characterization of  $(GT)_n$  microsatellites from native white shrimp (*Penaeus setiferus*) [J]. Mol Ecol, 1998, 7(9): 1251-1253.
- [5] Garcia D K, Dhar A K, Alcivar- Warren A. Molecular analysis of a RAPD marker (B20) reveals two microsatellites and differential mRNA expression in *Penaeus vannamei* [J]. Mol Mar Biol Biotechnol, 1996, 5(1): 71-83.
- [6] Wolfus Greg M, Garcia Denise K, Alcivar- Warren Acacia. Application of the microsatellite technique for analyzing genetic diversity in shrimp breeding programs [J]. Aquac, 1997, 152: 35-47.
- [7] Pongsomboon, Whan V, Moore S S, et al. Characterization of Tri- and tetranucleotide microsatellite in the black tiger prawn, *Penaeus monodon* [J]. Science Asia, 2000, 26(1): 1-6.
- [8] Xu Z, Dhar A K, Wyrzykowski J, et al. Identification of abundant and information microsatellites from shrimp (*Penaeus monodon*) genome [J]. Animal Genetics, 1999, 30: 1-7.
- [9] Xu Z K, Primavera J H, Pena L D, et al. Genetic diversity of wild and cultured black tiger shrimp (*Penaeus monodon*) in the Philippines using microsatellites [J]. Aquac, 2001, 199: 13-40.

- [10] Scott K D, Egger P S. Analysis of SSRs derived from grape ESTs [J]. *Theor Appl Genet*, 2000, 100: 723– 726.
- [11] 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 Science*, 2001, 160: 1115– 1123.
- [12] Whan V A, Wilson K J, Moore S S, et al. Two polymorphic microsatellite markers from novel *Penaeus monodon* ESTs [J]. *Animal Genetics*, 2000, 31: 140– 157.
- [13] Singh S, Li W L, Brown-Guedra G, et al. Simple sequence repeats from wheat ESTs are cross-transferable to rice, maize, and *Sorghum* [C]. *Plant, Animal & Microbe Genomes X Conference*, 2002, 125.
- [14] Karsi Attila, Cao D F, Feng J N, et al. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin [C]. *Plant, Animal & Microbe Genomes X Conference*, 2002, 236.
- [15] Xiang J H, Wang B, Liu B, et al. Over 10000 expressed sequence tags from *Penaeus chinensis* [C]. *Plant, Animal & Microbe Genomes X Conference*, 2002, 16.
- [16] Weber J L. Informativeness of human (dC-dA)n, (dG-dT)n polymorphisms [J]. *Genomics*, 1990, 7: 524– 530.