

DOI: 10.3724/SP.J.1118.2011.00742

## 草鱼三、四核苷酸重复微卫星标记的分离与特征分析

李文升<sup>1,2</sup>, 刘翠<sup>2,3</sup>, 鲁翠云<sup>2</sup>, 李超<sup>2</sup>, 孙效文<sup>2</sup>

1. 上海海洋大学 水产与生命学院, 上海 201306;

2. 中国水产科学研究院 黑龙江水产研究所, 农业部北方鱼类生物工程育种重点开放实验室, 黑龙江 哈尔滨 150070;

3. 哈尔滨师范大学, 黑龙江 哈尔滨 150025

摘要: 采用磁珠富集法、通过质粒检测法结合同位素杂交检测获得草鱼(*Ctenopharyngodon idella*)三、四核苷酸阳性克隆 1 378 个, 测序 705 个, 共得到 846 个微卫星座位, 其中完美型 632 个(占 74.7%), 非完美型 114 个(占 13.48%), 混合型 100 个(占 11.82%)。根据微卫星侧翼序列设计并合成 100 对微卫星引物, 检测结果显示 35 对(35%)引物在邗江野生群体中表现出多态性。选择其中 20 对多态性稳定的引物进行遗传多样性分析, 结果显示, 20 个位点共检测到 212 个等位基因, 平均观察杂合度( $H_o$ )为 0.681 1, 平均期望杂合度( $H_e$ )为 0.797 5, 平均多态信息含量为 0.777 7。结果表明, 所筛选的微卫星标记能够用于草鱼群体遗传学研究。本研究旨在为草鱼的种群遗传结构分析、遗传图谱构建等研究奠定基础。

关键词: 草鱼; 微卫星标记; 微卫星文库; 多态性

中图分类号: S96

文献标志码: A

文章编号: 1005-8737-(2011)04-0742-09

微卫星 DNA(Microsatellites DNA)又称短串联重复序列(Short Tandem Repeats, STRs)或简单重复序列(Simple Sequence Repeats, SSRs), 是广泛分布于真核生物基因组中以 1~6 个核苷酸为基本重复单位的串联重复序列。微卫星在基因组定位方面具有分布密度高、分布均匀等特点, 它一出现就被作为良好的作图标记广泛应用<sup>[1-2]</sup>。微卫星分子标记以其重复性好、共显性、符合孟德尔分离等特点, 迅速应用到群体多样性分析、构建遗传连锁图谱、亲子鉴定等方面<sup>[3]</sup>。近年来, 微卫星标记应用趋向于重复单位较长的微卫星位点, 主要是三碱基和四碱基重复位点, 原因是它们可以减少扩增产物中 *stutter* 带的出现, 并且可以方便于实现扩增产物的自动化分析<sup>[4]</sup>。

草鱼(*Ctenopharyngodon idella*)是中国土著鱼类, 自然分布于长江、珠江及黑龙江等大型水系中,

是一种重要的经济鱼类。目前国内报道的草鱼微卫星多见于重复单位为两碱基的微卫星标记<sup>[3, 5]</sup>。本研究利用磁珠富集法、质粒检测法结合同位素杂交法筛选一批三、四核苷酸重复的微卫星标记, 并用野生草鱼群体对其多态性进行了初步的分析, 为草鱼种群遗传结构分析、遗传图谱构建等打下基础。

### 1 材料与方法

#### 1.1 实验材料

克隆所用的草鱼血液取自黑龙江水产研究所松浦试验站, 野生草鱼群体采集于江苏邗江长江流域, 共 36 尾。实验所用的生物素探针(TGA)<sub>8</sub>、(AGAT)<sub>6</sub> 以及相应的同位素探针均由上海生工生物工程技术服务有限公司合成; 所用的限制性内切酶 *Sau3AI*、*T<sub>4</sub>*DNA 连接酶和 *Taq* DNA 聚合酶

收稿日期: 2010-08-13; 修订日期: 2010-10-30.

基金项目: 国家 973 计划资助项目(2004CB117405).

作者简介: 李文升(1984-), 硕士研究生, 主要从事鱼类遗传育种与繁殖方面的研究. E-mail: liwensheng1984@163.com

通信作者: 孙效文, 研究员, 博士生导师. E-mail: sunxw2002@163.com

购自 MBI 公司; pMD18-T 载体购自宝生物工程(大连)有限公司; Biospin PCR 产物纯化试剂盒购自杭州博日科技有限公司; 同位素( $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ )购自北京福瑞生物工程公司。

## 1.2 草鱼基因组微卫星文库的构建和筛选

**1.2.1 提取、纯化并酶切基因组 DNA** 采取草鱼新鲜血液, 100  $\mu\text{L}$  血液中加入 900  $\mu\text{L}$  裂解液(0.5% 十二烷基肌氨酸钠, 200  $\mu\text{g}/\text{mL}$  蛋白酶 K, 0.5 mol/L EDTA(pH 8.0), 50  $^{\circ}\text{C}$  消化 3 ~ 5 h, 用苯酚-氯仿提取法提取基因组 DNA, 再用透析液[Tris-Cl(pH 8.0) 50 mmol/L, EDTA(pH 8.0)10 mmol/L]透析 24 h 去除 DNA 中盐离子。透析后的 DNA 经过预冷无水乙醇沉淀、70%乙醇洗涤后于室温干燥, 用 0.1 $\times$ TE 溶解后置于 4 $^{\circ}\text{C}$  冰箱备用。用限制性内切酶 *Sau* 3AI(10 U/ $\mu\text{L}$ )的酶切缓冲体系对基因组 DNA 进行不完全酶切, 使 70%以上的酶切片段集中在 250 ~ 1 000 bp 范围内, 采用蔗糖密度梯度(10%、20%、30%、40%)离心(22 000 r/min, 22 h), 收 300 ~ 700 bp 的酶切片段。

**1.2.2 接头的制备和连接** 等比例混合两组寡聚核苷酸链 A(5'-GCGGTACCCGGAAGCTTGG-3')、B(5'-GATCCCAAGCTTCCCGGTACCGC-3'), 95 $^{\circ}\text{C}$  变性 10 min 后经 4 h 缓慢冷却至 10  $^{\circ}\text{C}$ , 形成双链接头:

5' GCGGTACCCGGAAGCTTGG A 链

3' CGCCATGGGCCCTTCGAACCCTAG B 链

建立 20  $\mu\text{L}$  连接体系, 其中 10 $\times$ B $\mu$ ffer 2  $\mu\text{L}$ , 接头 8  $\mu\text{L}$ ,  $T_4$  连接酶 2  $\mu\text{L}$ , 酶切后 DNA 3  $\mu\text{L}$ , ddH $_2$ O 5  $\mu\text{L}$ , 16 $^{\circ}\text{C}$  循环水浴连接 16 h, 连接结束后用 PCR 纯化试剂盒去除多余接头。

**1.2.3 PCR** 建立 25  $\mu\text{L}$  PCR 体系, 包括去接头后 DNA 1  $\mu\text{L}$ , 10 $\times$ Buffer 18  $\mu\text{L}$ , 寡核苷酸链 A(10  $\mu\text{mol}/\text{L}$ ) 0.5  $\mu\text{L}$ , *Taq* DNA 聚合酶 0.5  $\mu\text{L}$ , ddH $_2$ O 5  $\mu\text{L}$ , 在 PE 9700 PCR 仪进行扩增。反应程序为: 95  $^{\circ}\text{C}$  预变性 5 min, 94  $^{\circ}\text{C}$  变性 30 s, 58  $^{\circ}\text{C}$  退火 1 min, 72  $^{\circ}\text{C}$  延伸 2 min, 24 个循环, 最后 72  $^{\circ}\text{C}$  延伸 10 min。反应结束后用纯化试剂盒去除多余 dNTP、引物等, 将 PCR 产物浓缩至 20  $\mu\text{L}$  左右, 琼脂糖电泳检测其浓度和片段大小。

**1.2.4 草鱼微卫星文库的构建** 配制 50  $\mu\text{L}$  杂交体系, 包括 10% SDS 0.5  $\mu\text{L}$ , 20 $\times$ SSC 15  $\mu\text{L}$ , Primer A 5  $\mu\text{L}$ , 生物素探针 1.5  $\mu\text{L}$ (10  $\mu\text{mol}/\text{L}$ )。将混合体系在 50  $^{\circ}\text{C}$  水浴中预热, 取 12  $\mu\text{L}$  PCR I 后的模板 DNA, 95  $^{\circ}\text{C}$  变性 5 min 后加入杂交液中杂交 1 h。

吸取 100  $\mu\text{L}$  磁珠, 200  $\mu\text{L}$  B & W Buffer(10 mmol/L Tris-Cl, 1 mmol/L EDTA, 2 mol/L NaCl)清洗 2 次, 然后加入 200  $\mu\text{L}$  洗液 (6 $\times$ SSC, 0.1% SDS)反复清洗 7 ~ 8 次, 至吹吸时容易散开为止。将磁珠移入杂交液中, 20  $^{\circ}\text{C}$  轻摇 20 min 后吸去溶液。分别用洗液 I、洗液 II(3 $\times$ SSC, 0.1% SDS)、洗液 I(1 $\times$ SSC, 0.1% SDS)清洗磁珠, 去除不含微卫星的 DNA 片段。洗涤方法是: 在室温下用洗液 I 清洗 2 次, 每次静置 10 min; 在 50 $^{\circ}\text{C}$  水浴中用洗液 II 洗 2 次, 每次静置 15 min; 室温下用洗液 I 快速洗 2 次。

洗涤结束后磁珠溶于 30  $\mu\text{L}$  0.1 $\times$ TE 中, 然后转入 200  $\mu\text{L}$  PCR 小管, 95  $^{\circ}\text{C}$  变性 10 min, 释放出含有微卫星序列的单链。再以单链 DNA 为模板进行 PCR II, 反应程序同 PCR I, 反应结束后用纯化试剂盒去除过量 dNTP、引物等, 并浓缩体积至 20  $\mu\text{L}$ , 2%琼脂糖电泳检测结果。

**1.2.5 连接转化** 建立 10  $\mu\text{L}$  连接体系, 包括 pMD 18-T 1  $\mu\text{L}$ , Ligation Solution I 5  $\mu\text{L}$ , Insert DNA 1  $\mu\text{L}$ , ddH $_2$ O 3  $\mu\text{L}$ , 16  $^{\circ}\text{C}$  循环水浴 3 h。将连接产物转入用 CaCl $_2$  制备的感受态大肠杆菌(*Escherichia coli*)DH5 $\alpha$  中。

**1.2.6 质粒二次检测** 配制适量液体培养基(蛋白胨 10 g/1 000 mL, 酵母提取物 5 g/1 000 mL, NaCl 10 g/1 000 mL), 高压灭菌, 在培养基温度约 50 $^{\circ}\text{C}$  时加入氨苄溶液(50 mg/mL), 每 1 000 mL 培养基中加入 2 000  $\mu\text{L}$  氨苄。准备 1.5 mL 灭菌离心管若干, 每个离心管中加入 1 mL 液体培养基, 用灭菌牙签划下菌株后插入离心管中搅动几下后拿出, 37 $^{\circ}\text{C}$  摇床培养过夜。第 2 天将离心管取出, 7 500 r/min 离心去上清, 每管中加入 30  $\mu\text{L}$  溴酚蓝吹打然后加入 20  $\mu\text{L}$  苯酚、20  $\mu\text{L}$  氯仿异戊醇, 剧烈震荡 5 h 后离心(12 000 r/min, 5 min), 吸出液体部分, 1.5%琼脂糖检测。

**1.2.7 同位素二次杂交筛选** 将含重组子的菌株转移到硝酸纤维素膜上, 用相应同位素 ( $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ ) 标记的探针进行二次杂交。

### 1.3 测序与引物设计

将质粒检测后的菌株(135 个, 三碱基菌株 67 个、四碱基菌株 68 个)和同位素杂交二次筛选后的阳性克隆(570 个, 三碱基菌株 315 个、四碱基菌株 255 个)送往上海生物芯片有限公司测序, 返回序列在去除载体序列和接头后用 Primer3.0 在线设计引物, 设计的引物由上海生工生物工程技术服务有限公司合成。

### 1.4 数据处理

微卫星是共显性遗传, 可以直接从聚丙烯电泳上判断个体的基因型, 用 GelPro(Verion 4.5)软件分析各座位等位基因片段大小, 用 PopGene (Verion 3.2)软件统计微卫星座位的等位基因数 (numbers of allele,  $A$ )、有效等位基因数 (effective numbers of allele,  $N_e$ )、观测杂合度 (observed heterozygosity,  $H_o$ )、期望杂合度 (expected heterozygosity,  $H_e$ )。计算公式:

$$\text{有效等位基因数: } N_e = 1 / \sum_{i=1}^n P_i^2;$$

$$\text{平均期望杂合度: } H_e = \sum_{i=1}^r h_i / r;$$

根据 Botstein 等<sup>[7]</sup> (1980)公式计算多态信息含量 (polymorphism information content, PIC):

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

其中  $n$  为某一位点上等位基因数,  $P_i$  为群体中第  $i$  个位点的基因频率;  $h_i$  为第  $i$  个位点的期望杂合度,  $r$  为位点个数,  $H_e$  为群体内平均期望杂合度;  $P_i$ 、 $P_j$  分别为第  $i$  和第  $j$  个等位基因在群体中的频率,  $j=i+1$ 。

## 2 结果与分析

### 2.1 测序结果

质粒检测 200 个菌株(图 1), 得到含插入片段的 135 个菌株, 全部进行测序, 得到测序 135 个, 返回的序列中, 含微卫星序列的有 86 个, 阳性克

隆率为 43%。挑取 2 656 个菌株进行同位素杂交(图 2, 箭头所指的克隆为阳性克隆), 得到 1 243 个阳性克隆。从这些阳性克隆中挑取 570 个进行测序, 成功测序 524 个, 返回的序列中含微卫星的有 486 个, 根据同位素二次杂交结果计算阳性克隆率为 43.39%。两次测序得到的 659 个序列中, 含微卫星的序列有 572 个, 占 86.80%。其中三核苷酸重复克隆 347 个, 获得目的序列 282 个(81.27%); 四核苷酸重复克隆 312 个, 获得目的序列 290 个(92.95%)。

在含微卫星序列的 572 个阳性克隆中, 共有 846 个微卫星座位, 根据 Weber(1990)<sup>[6]</sup>提出的分类标准, 统计得到完美型(perfect)632 个, 占 74.70%; 非完美型(imperfect)114 个, 占 13.48%; 混合型(compound)100 个, 占 11.82%。其中三核苷酸重复的有 370 个微卫星座位(完美型 302 个, 占 81.62%; 非完美型 58 个, 占 15.68%; 混合型 10 个, 占 2.70%); 四核苷酸重复的有 476 个微卫星座位(完美型 330 个, 占 69.33%; 非完美型 56 个, 占 11.76%; 混合型 90 个, 占 18.91%)。所得的微卫星重复数在 5~10 次占 61.28%; 11~20 次占 19.64%; 21~30 次占 6.20%; 30~50 次占 5.45%; 50 以上占 7.42%。

根据郑劲松等<sup>[8]</sup>定义的杂交效率(hybridization efficiency), 即通过探针杂交实际上所获得的与探针序列相符的微卫星序列所占的百分比, 通过统计得到与三核苷酸重复探针相符的有 98 个, 与四核苷酸重复序列相符的有 271 个, 分别占 34.75%、93.45%。三核苷酸重复序列中除去与探针(TGA)<sub>8</sub> 吻合的序列外, 还观察到 TCA/AGT、AGC/TCG、TGC/ACG 重复碱基的序列; 四核苷酸重复序列中除去与探针(AGAT)<sub>6</sub> 相符的序列外, 还有很少的 AGAC/TCTG、CCTA/GGAT、CTGC/GACG 等重复碱基序列; 在三、四核苷酸重复序列中还发现存在一些双核苷酸如 CA/GT、CT/GA 等重复碱基序列。以上统计结果见表 1。

### 2.2 引物设计与筛选及群体遗传结构评估

所测序列在去除载体序列和接头后, 部分微卫星序列因本身结构或侧翼序列太短不能设计引

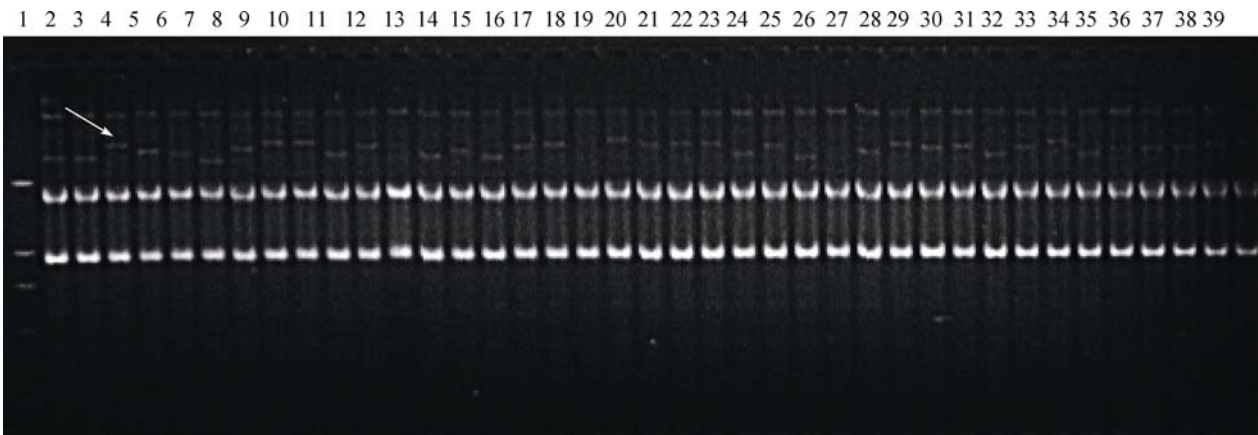


图 1 质粒检测阳性克隆  
箭头所指目标电泳带为检测的含有目的片段的质粒; 1-39: 菌株个体  
Fig.1 Plasmid detected in *Ctenopharyngodon idellus*  
Arrow point to the plasmid of containing targeted fragment; 1-39: strain individuals

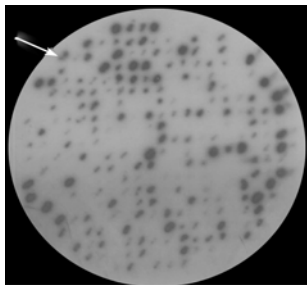


图 2 同位素二次筛选  
箭头示阳性克隆.  
Fig.2 Isotope secondary screening  
Arrow shows positive clones.

物, 其余微卫星利用 Primer3.0 在线设计并选择合成引物 100 对, PCR 扩增检测, 结果 59 对能扩增出特异性条带, 其中 35 对在个体间具有不同程度的多态性。用江苏邗江野生群体对其中 20 对引物 (5 对三核苷酸重复、15 对四核苷酸重复) 的多态性进行了深入评价, 结果共扩增出 212 个等位基因, 各位点扩增片段大小在 104 ~ 310 bp; 等位基因数在 2 ~ 22 之间; 有效等位基因数在 1.433 6 ~ 17.753 6 之间; 观察杂合度在 0.314 3 ~ 0.972 2 之

间; 期望杂合度变动范围为 0.302 4 ~ 0.943 7; 多态信息含量 PIC 在 0.256 7 ~ 0.940 8 之间。根据 Botstein 等<sup>[7]</sup>提出的衡量遗传变异程度高低的多态信息含量指标(PIC>0.5 时, 该位点为高度多态位点; 0.25<PIC<0.5 时, 该位点为中度多态位点; PIC<0.25 时, 该位点为低度多态位点), 其中 20 个位点中有 17 个位点表现出高度的多态性, 另外 3 个位点 *HLJC135*(PIC=0.373 3)、*HLJC172*(PIC=0.256 7)、*HLJC207*(PIC=0.340 3)为中度多态位点。对江苏邗江野生群体的遗传结构进行评估, 结果显示, 等位基因数、有效等位基因数、观察杂合度、期望杂合度、多态信息含量五项多态评估指标分别为 10.600 0、7.510 0、0.681 1、0.797 5、0.777 7。图 3 为 *HLJC137* 标记在邗江草鱼野生群体中的扩增结果, 引物信息见表 2, 统计数据见表 3。

3 讨论

3.1 草鱼三、四核苷酸重复微卫星的筛选

许多研究表明, 在鱼类基因组中, 二碱基微

表 1 草鱼微卫星的不同类型重复序列数量及所占比例(括号中数据)  
Tab.1 Percentages (in brackets) of various repeat sequence types of microsatellite in grass carp

| 探针类型<br>probe type         | 微卫星数目<br>numbers of<br>microsatellite | 完美型<br>perfect | 非完美型<br>imperfec | 混合型<br>compound |
|----------------------------|---------------------------------------|----------------|------------------|-----------------|
| Biotin-(TGA) <sub>8</sub>  | 370                                   | 302 (81.62%)   | 58 (15.68%)      | 10 (2.70%)      |
| Biotin-(AGAT) <sub>6</sub> | 476                                   | 330 (69.33%)   | 56 (11.76%)      | 90 (18.91%)     |

表 2 草鱼部分微卫星引物  
Tab. 2 Some microsatellite primer pairs for *Ctenopharyngodon idellus*

| 位点<br>locus | 引物序列(5'→3')<br>primer sequence(5'→3')                    | 退火温度/℃<br>$T_m$ | 重复单元<br>repeat motif                      | GenBank 注册号<br>GenBank accession No. |
|-------------|----------------------------------------------------------|-----------------|-------------------------------------------|--------------------------------------|
| HLJC128     | F: ATGGGTGGATGTGTGATGG<br>R: TCATGAGGCGTCTGAATCTCT       | 60              | (GATA) <sub>6</sub>                       | HM563708                             |
| HLJC132     | F: GCCATAAAGGTTCCCAGGTT<br>R: GTCCATCAGGCTGGAGAAAAG      | 60              | (CAT) <sub>5</sub>                        | HM563709                             |
| HLJC135     | F: CTGCTCTCCTGTGCTGAGGCTCT<br>R: GACAGCACACGACACCAACT    | 60              | (GAT) <sub>11</sub>                       | HM563710                             |
| HLJC136     | F: CTATTCGCCAAGACCAAAAC<br>R: GCTGTTATTGCTTCTGCTGTTG     | 60              | (CAG) <sub>5</sub>                        | HM563711                             |
| HLJC137     | F: CACCCGCTGACATTCTGATT<br>R: AGCAATTCATATGGCCTTCG       | 60              | (ATCT) <sub>6</sub> (TCCA) <sub>5</sub>   | HM563712                             |
| HLJC140     | F: TCAAGGATTGGGGTAAATTC<br>R: GATGGATGGATGGATACAACG      | 60              | (CAAT) <sub>5</sub> (ATCT) <sub>10</sub>  | HM563713                             |
| HLJC141     | F: CGATGAATGGATGGATGGAT<br>R: TCCCAACTTATTTGGAATCG       | 60              | (GATA) <sub>5</sub> (GATA) <sub>15</sub>  | HM563714                             |
| HLJC142     | F: TCTGGTCACTTTGTTGCTGAA<br>R: CCAGTGAAAGTGTGCGACAG      | 62              | (AC) <sub>25</sub>                        | HM563715                             |
| HLJC143     | F: TTTCAGTGGTGTCCCTTTCC<br>R: TCCACTGATGTGTCCACGTT       | 60              | (AC) <sub>13</sub>                        | HM563716                             |
| HLJC145     | F: CACACCTGAATGAATGACGA<br>R: TGGATTATCGAATGCTGCTG       | 60              | (ATCT) <sub>6</sub>                       | HM563717                             |
| HLJC148     | F: AACAGACGGATGGATGGATG<br>R: CTTTCAAAATGTGGAGTCTTGC     | 60              | (GGAT) <sub>7</sub> (GATA) <sub>10</sub>  | HM563718                             |
| HLJC150     | F: GTGGGGCTTTCATCATGTTT<br>R: GAAGCAGGTGGTTGGAGACT       | 62              | (TG) <sub>15</sub>                        | HM563719                             |
| HLJC151     | F: GATCGACAGACAGACGGACA<br>R: TGCTGGTTAAGATGGTTAGCA      | 60              | (ATAG) <sub>11</sub> (AGAA) <sub>10</sub> | HM563720                             |
| HLJC165     | F: TCAACTCGCTCTCAAATTCTCA<br>R: AGGGTGTGTGGGCTATGTGT     | 60              | (CTAT) <sub>25</sub>                      | HM563721                             |
| HLJC168     | F: TTGGCTTGAAAAAGCCTAGC<br>R: CACACACTTGCATACTCGCATA     | 62              | (GT) <sub>16</sub>                        | HM563722                             |
| HLJC172     | F: CAATAACCCGAAACAGCATT<br>R: CGCTGCACATCACCATAATC       | 56              | (CAT) <sub>5</sub>                        | HM563723                             |
| HLJC174     | F: TTTCAGTCTCCTCCCTCAC<br>R: GGACTTCTCTTGCTGCATCC        | 60              | (TCA) <sub>5</sub>                        | HM563724                             |
| HLJC180     | F: TGTCTGGCTGTCTGTCTGTCTA<br>R: CTCACCTCACTCACTCAGTCACCT | 60              | (GTCTG) <sub>10</sub> (GGCT) <sub>6</sub> | HM563725                             |
| HLJC182     | F: CCGAACGAGCGAACTAACTC<br>R: CCGAACGAAGTCAGTGAGTG       | 60              | (GGCT) <sub>5</sub>                       | HM563726                             |
| HLJC186     | F: GATGGCGACTCCTCTCTGTC<br>R: CAAACTCAAGGTGCGTCTGA       | 60              | (TCA) <sub>13</sub>                       | HM563727                             |
| HLJC193     | F: TCACTCATCCAGAAGCAGCA<br>R: CGGTACCGTCGACGATCTAT       | 60              | (AGC) <sub>9</sub> (CAG) <sub>8</sub>     | HM563728                             |
| HLJC198     | F: GGCAGTAGCAAGTCCCATAC<br>R: TACGCTGCTGCTATTGGTTG       | 60              | (AGC) <sub>5</sub>                        | HM563729                             |
| HLJC199     | F: CGGCCAGTAGTTTGGATTCT<br>R: ACCAAAAAGTCACGTGGAAGG      | 60              | (TCA) <sub>5</sub> (TCA) <sub>5</sub>     | HM563730                             |
| HLJC203     | F: CGCTGCTCCTATCCACAAC<br>R: CCACAGTCTTCTCCAGTCC         | 60              | (TGA) <sub>6</sub>                        | HM563731                             |
| HLJC207     | F: TCCATCCATCCGTAATTGA<br>R: GGATGCACGGATGATAGACA        | 60              | (ATCC) <sub>10</sub>                      | HM563732                             |
| HLJC209     | F: TCAACACATGTGACTAAGTGCAG<br>R: GGAGAGATAGATGAATGGATGGA | 60              | (ATCT) <sub>9</sub>                       | HM563733                             |
| HLJC211     | F: CATCCCAGACCTCTTACATGG<br>R: TTGGCTTTGTTGGTGTGTGT      | 60              | (GT) <sub>5</sub>                         | HM563734                             |
| HLJC213     | F: AGCAGTCATGTTGAAACCACA<br>R: ATCTGACATGGCTCCCAAAT      | 61              | (TG) <sub>19</sub>                        | HM563735                             |

续表 2 Tab. 2 continued

| 位点<br>Locus | 引物序列(5'→3')<br>primer sequence(5'→3')                 | 退火温度/℃<br>$T_m$ | 重复单元<br>repeat motif                                        | GenBank 注册号<br>GenBank accession No. |
|-------------|-------------------------------------------------------|-----------------|-------------------------------------------------------------|--------------------------------------|
| HLJC217     | F: TCTGACCATTTCTATCCATCCA<br>R: CTTGGGATCAACGATAGAACG | 60              | (TCTA) <sub>5</sub>                                         | HM563736                             |
| HLJC220     | F: TCCGAGTCAGAGGAAGAGGA<br>R: AACGGGCTTTATGGTTGTTG    | 60              | (TGA) <sub>8</sub>                                          | HM563737                             |
| HLJC221     | F: GGACAAGGCGTCACTCAGTC<br>R: GAACGCCACTGAACAACCTCA   | 60              | (TATC) <sub>9</sub> (TCTA) <sub>7</sub> (TCTG) <sub>5</sub> | HM563738                             |
| HLJC222     | F: GCATAACACGGTCTGCCTCT<br>R: GGATGGATGGATGCTTGAGT    | 60              | (TATC) <sub>7</sub>                                         | HM563739                             |
| HLJC223     | F: GCCTGTCTACTCAAGCATCCA<br>R: GCATGTTACCCTGAAACCTCA  | 60              | (CTAT) <sub>8</sub>                                         | HM563740                             |
| HLJC225     | F: TCCATCCATGACCACAGCTA<br>R: TGCCATAGCAACTACCCAGA    | 60              | (GATA) <sub>14</sub> (ATGG) <sub>9</sub>                    | HM563741                             |
| HLJC227     | F: TGTTCCACATGCAAACTCTCA<br>R: GACGGACGGATGGATAGATG   | 60              | (TATC) <sub>6</sub>                                         | HM563742                             |

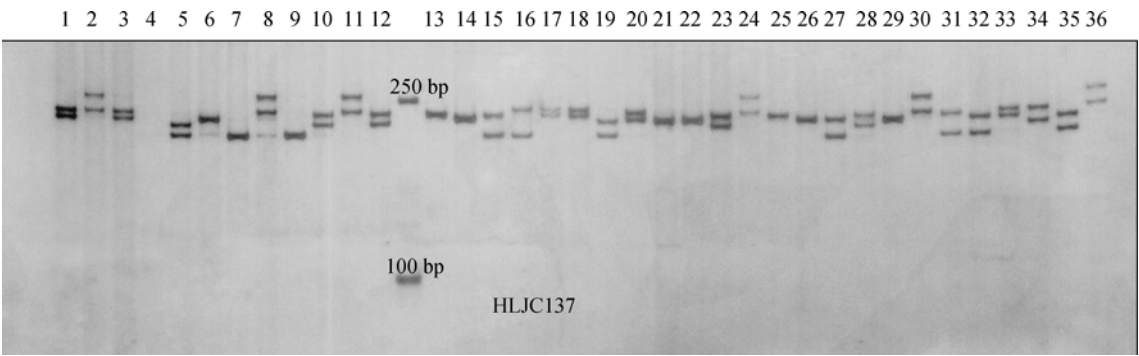


图 3 HLJC137 标记在邗江草鱼野生群体中的扩增  
1–36: 草鱼个体编号  
Fig.3 Amplified result of wild *Ctenopharyngodon idellus* population from Hanjiang River  
at HLJC137 locus  
1–36: individual number

卫星重复占绝对优势, 其中又以 CA/GT 分布广泛, 其次是 CT/GA<sup>[9-12]</sup>。多数鱼类开发的微卫星标记以(CA)等双核苷酸重复为主, 三、四核苷酸重复序列开发较少。研究表明, 在人类基因组中三核苷酸重复序列至少与 16 种遗传疾病的发生有关<sup>[13]</sup>, 并且具有较高的多态性和遗传稳定性<sup>[14]</sup>。但是由于三、四核苷酸重复的微卫星类型在真核生物基因组中所占比例相对较少, 用小片段克隆法分离获得的数量极其有限。本次实验利用三碱基(TGA)<sub>n</sub>和四碱基(AGAT)<sub>n</sub>作为探针, 得到(CAT)<sub>n</sub>、(CAG)<sub>n</sub>、(TCA)<sub>n</sub>、(AGC)<sub>n</sub>、(TGA)<sub>n</sub>三碱基和(GATA)<sub>n</sub>、(ATCT)<sub>n</sub>、(TCCA)<sub>n</sub>、(CAAT)<sub>n</sub>、(GGAT)<sub>n</sub>等四碱基核心序列, 并得到少量的二碱基核心序列。本研究成功开发了 137 个微卫星标记, 包括双核苷酸重复的 24 个标记、三核苷酸重复的 56

个标记和四核苷酸重复的 57 个标记。Toth 等<sup>[15]</sup>的统计结果表明脊椎动物中四核苷酸重复以 AGAT 最为普遍, 而三核苷酸重复的是 CAG。Castano-Senchez 等<sup>[16]</sup>在日本牙鲆中也证实了这一点。本研究中(AGAT)<sub>n</sub>探针的杂交效率(93.45%)明显高于(TGA)<sub>n</sub>探针的杂交效率(34.75%), 原因可能是草鱼基因组中 AGAT 重复的微卫星与 TGA 重复的微卫星相比更为广泛。

目前, 在水产生物中克隆微卫星序列的方法很多, 但主要有 2 种: 经典的小片段基因组 DNA 随机克隆技术和生物素磁珠富集技术。在分离微卫星的研究中, 李琪等<sup>[17-18]</sup>采用小片段克隆法在长牡蛎和皱纹盘鲍中的富集效率(阳性克隆率)分别为 20.5%和 13.1%; 徐鹏等<sup>[19]</sup>用小片段克隆法在中国对虾中从 92 个重组阳性克隆中得到 22

表 3 邗江草鱼野生群体 24 个微卫星位点的多样性指数

Tab. 3 Polymorphic information of 24 microsatellite locus of wild *Ctenopharyngodon idellus* population from Hanjiang River

| 基因座<br>locus | 产物大小/bp<br>product size | 等位基因数<br><i>A</i> | 有效等位基因数<br><i>N<sub>e</sub></i> | 观察杂合度<br><i>H<sub>o</sub></i> | 期望杂合度<br><i>H<sub>e</sub></i> | 多态信息含量<br>PIC |
|--------------|-------------------------|-------------------|---------------------------------|-------------------------------|-------------------------------|---------------|
| HLJC128      | 213–239                 | 10                | 7.2934                          | 0.5294                        | 0.8629                        | 0.8480        |
| HLJC135      | 199–205                 | 2                 | 1.9862                          | 0.4722                        | 0.4965                        | 0.3733        |
| HLJC137      | 221–286                 | 9                 | 6.9209                          | 0.7143                        | 0.8555                        | 0.8383        |
| HLJC140      | 216–230                 | 6                 | 5.5149                          | 0.5278                        | 0.8187                        | 0.7928        |
| HLJC141      | 131–204                 | 17                | 13.2432                         | 0.8571                        | 0.9245                        | 0.9194        |
| HLJC145      | 188–211                 | 7                 | 5.9322                          | 0.8857                        | 0.8314                        | 0.8098        |
| HLJC148      | 176–240                 | 18                | 11.7285                         | 0.9722                        | 0.9147                        | 0.9086        |
| HLJC151      | 104–167                 | 22                | 17.7536                         | 0.8000                        | 0.9437                        | 0.9408        |
| HLJC165      | 175–282                 | 16                | 10.8407                         | 0.7714                        | 0.9078                        | 0.9007        |
| HLJC172      | 236–240                 | 2                 | 1.4336                          | 0.3143                        | 0.3024                        | 0.2567        |
| HLJC186      | 199–232                 | 11                | 8.3904                          | 0.7429                        | 0.8808                        | 0.8691        |
| HLJC199      | 159–212                 | 9                 | 6.7405                          | 0.6176                        | 0.8516                        | 0.8346        |
| HLJC203      | 177–207                 | 10                | 5.6322                          | 0.4286                        | 0.8224                        | 0.7995        |
| HLJC207      | 136–148                 | 3                 | 1.6712                          | 0.4167                        | 0.4016                        | 0.3403        |
| HLJC209      | 136–196                 | 14                | 8.3333                          | 0.8000                        | 0.8800                        | 0.8689        |
| HLJC217      | 186–211                 | 7                 | 4.5037                          | 0.7143                        | 0.7780                        | 0.7472        |
| HLJC221      | 212–310                 | 15                | 10.5603                         | 0.7143                        | 0.9053                        | 0.8974        |
| HLJC222      | 200–244                 | 12                | 8.7189                          | 0.7714                        | 0.8853                        | 0.8745        |
| HLJC223      | 168–228                 | 14                | 7.3574                          | 0.7143                        | 0.8641                        | 0.8508        |
| HLJC225      | 203–228                 | 8                 | 5.6452                          | 0.8571                        | 0.8229                        | 0.8831        |
| 平均值 mean     | —                       | 10.6000           | 7.5100                          | 0.6811                        | 0.7975                        | 0.7777        |

个含微卫星的重组阳性克隆; 刘丽等<sup>[20]</sup>用小片段克隆法从青石板鱼 900 个阳性克隆中得到 33 个可能含微卫星核心序列的 PCR 阳性克隆。由于小片段随机克隆技术获得的阳性克隆率较低, 很难获得足够用于群体遗传结构或建立图谱等研究所需的微卫星标记, 因此本研究采用生物素磁珠富集法结合同位素杂交法分离草鱼三、四核苷酸重复的微卫星克隆, 并用质粒二次检测法检测阳性克隆率, 得到阳性克隆率为 43%, 这和用同位素杂交法得到的阳性克隆率 43.39% 结果相符。理论上用磁珠富集法构建微卫星阳性克隆率就可以达到 100%<sup>[21]</sup>, 但实际上往往达不到, 这可能是操作上的一些原因导致。Gao 等<sup>[22]</sup>研究证实对杂交混合物洗涤的温度和强度是影响富集效率最关键的两个因素, 提高洗涤的温度和强度可以有效的减少 DNA 片段与磁珠之间的非特异性吸附<sup>[23]</sup>, 因此在研究中, 为了得到较高的富集效率, 针对不同研究对象, 对洗涤温度等条件进行探索和优化非

常必要<sup>[24]</sup>。

3.2 四核苷酸重复微卫星的多态性及群体评估效果

本研究中三、四核苷酸微卫星序列的重复数主要集中在 30 次以内, 占 87.12%, 其中重复数在 5~10 次的最多, 占 61.28%。研究表明真核生物微卫星大部分都在 30 次重复以下<sup>[25]</sup>, 本实验的结果也验证了这一观点。在筛选的 24 对有多态的微卫星标记中, 有 4 对双核苷酸重复的标记、5 对三核苷酸重复的标记和 15 对四核苷酸重复的标记。其中双核苷酸重复标记重复数都大于 12, 三核苷酸和四核苷酸重复标记重复数大于 10 的有 10 对。Weber<sup>[6]</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的多态信息含量(PIC), 当  $n \geq 16$  时, 可提供的 PIC 在 0.5 以上, 才可以进行相应的多态性分析。本研究中 4 对双核苷酸重复标记在邗江群体中的 PIC 都大于 0.5, 除了 HLJC150 标记的重复数为 15 外, 其余 3 个标记重复数都在 16 或者 16

以上, 这与 Weber 结论基本一致。在三核苷酸和四核苷酸重复标记中重复数在 5 以上便能在个体间有不同程度的多态性, 微卫星多态性程度与微卫星长度也没有显著性的正比关系。

有效等位基因数( $H_e$ )是反映群体遗传变异大小的一个指标, 其值越接近所检测到的等位基因数的绝对值, 说明等位基因在群体中分布越均匀<sup>[26]</sup>。邗江群体共检测到 276 个等位基因, 有效等位基因数为 186.360 3, 两值之间较大的偏差, 说明所检测的 24 个基因座的等位基因在邗江群体基因组中分布不够均匀, 应适当提高样本数和基因座数目。李鸥等<sup>[27]</sup>发现当取样量大于 40, 标记量大于 25 时, 各遗传参数值趋于稳定。本研究中样本量为 36, 标记量为 24, 都略低于所需的数目, 可能造成了有效等位基因数与等位基因数之间的偏差。群体杂合度的高低反应了群体在多个基因座上的遗传变异及群体遗传多样性丰富度。本研究中邗江草鱼群体  $H_e$  为 0.811 9,  $H_o$  为 0.662 8, 说明邗江草鱼群体的遗传多样性较高。平均期望杂合度和平均观察杂合度存在差异说明个别位点存在无效等位基因, 无效等位基因的存在会导致杂合子的缺失或者纯合子的过剩, 从而使  $H_e$  与  $H_o$  之间出现偏离。在所用的 24 个基因座中, *HLJC140*、*HLJC128*、*HLJC142*、*HLJC168*、*HLJC199*、*HLJC203* 和 *HLJC213* 中  $H_e$  和  $H_o$  存在较大差异, 说明这 7 个基因座存在无效等位基因。

### 3.3 开发意义及应用展望

研究表明, 三、四核苷酸重复的微卫星标记具有更高的多态性<sup>[28]</sup>。本研究所得三碱基微卫星较二碱基微卫星也具有高多态性特征, 例如李鸥<sup>[27]</sup>利用 45 个二碱基微卫星标记在 40 个长江野生草鱼样本中获得的等位基因数是 3.68, 平均期望杂合度是 0.659。本研究利用所得的 24 个三碱基微卫星标记检测 35 个长江野生草鱼样本, 平均有效等位基因数 7.765, 平均期望杂合度 0.811 9。本研究首次从草鱼基因组中开发了三、四核苷酸重复的微卫星标记, 并筛选出种类较多的核心序列, 是较好的微卫星资源, 丰富了草鱼原有标记的类型。经过筛选结果证明具有高度的多态性,

可以用于草鱼群体的遗传变异水平、近交衰退和性状退化等方面的分析, 通过建立分子遗传选育技术方法, 为遗传育种、种质鉴定、种苗放流等提供理论依据, 实现草鱼的种质资源科学管理及持续利用。随着水产经济动物遗传连锁图谱的相继构建, 草鱼遗传连锁图谱的构建也成为可能, 本研究开发的微卫星标记能够增加草鱼遗传连锁图谱的覆盖率, 对于草鱼遗传连锁图谱的构建具有重要意义。

### 参考文献:

- [1] Marklund L, Johansson Moller M, Davis W, et al . A comprehensive linkage map of the pig based on a wild pig-large white intercross[J] . *Animal Genetics*, 1996, 27: 255-269 .
- [2] 樊叶杨, 庄杰云, 吴建利, 等 . 用微卫星标记鉴别水稻粳亚种[J] . *遗传*, 2000, 22(6): 392-394.
- [3] 马海涛, 鲁翠云, 于冬梅, 等 . 草鱼基因组中微卫星分子标记的制备及筛选[J] . *上海水产大学学报*, 2007, 16: 389-393.
- [4] O'Connell M, Wright J M . Microsatellite DNA in fishes[J] . *Rev Fish Biol Fish*, 1997, 7: 331-363.
- [5] 孙效文, 鲁翠云, 梁利群 . 磁珠富集法分离草鱼微卫星分子标记[J] . *农业生物技术学报*, 2005, 13(6): 772-776.
- [6] Weber J L . Informativeness of human(dC-dA)<sub>n</sub>/(dG-dT)<sub>n</sub> polymorphisms[J] . *Genomics*, 1990, 7: 524-530.
- [7] Botstein D, White R L, Skolnick M, et al . Construction of a genetic linkage map in man using restriction fragment length polymorphisms[J] . *Am J Hum Gen*, 1980, 32: 314-331.
- [8] 郑劲松, 廖小林, 童金苟, 等 . 长江江豚微卫星 DNA 分离的初步研究[J] . *水生生物学报*, 2008, 32: 19-25.
- [9] Maguire T L, Edwards K J, Saenger P, et al . Characterization and analysis of microsatellite loci in mangrove species, *Avicennia marina*(Forsk)Vierh . (Avicenniaceae) [J] . *Theor Appl Genet*, 2000, 101: 279-285.
- [10] Brenner S, Elgar G, Sandford R, et al . Characterization of the pufferfish(*Fugu*)genome as a compact model vertebrate genome [J] . *Nature*, 1993, 366: 256-259.
- [11] Serapion J, Kucuktas H, Feng J, et al . Bioinformatic mining of type microsatellites from expressed sequence tags of channel catfish(*Ictalurus punctatus*) [J] . *Mar Biotechnol*, 2004, 6: 364-377.
- [12] Crooijmans R P M A, Beirbooms N A F, Komen J, et al . Microsatellite markers in common carp[J] . *Animal Genetics*, 1997, 28: 129-132.
- [13] 高焕, 孔杰 . 串联重复序列的物种差异及其生物功能[J] . *动物学研究*, 2005, 26(5): 555-564.
- [14] Edwards A, Civitello A, Hammond H A and Caskey C T . DNA Typing and genetic mapping with trimeric and

- tetrametic tandem repeats[J]. Am J Hum Gen, 1991, 49: 746–756.
- [15] Toth G, Gaspari Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis[J]. Gen Res, 2000, 10: 967–981.
- [16] Castano-Sanchez C, Fuji K, Hayashida K, et al. A set of polymorphic trinucleotide and tetranucleotide microsatellite markers for the Japanese flounder (*Paralichthys olivaceus*) [J]. Anim Gen, 2007, 38: 75–76.
- [17] 李琪, 木岛明博. 长牡蛎(*Crassostrea gigas*)微卫星克隆快速分离及特性分析[J]. 海洋与湖沼, 2004, 35: 364–370.
- [18] Li Q, Park C, Kijima A. Isolation and characterization of microsatellite loci in the Pacific abalone (*Haliotis discus hannai*) [J]. J Shellfish Res, 2002, 21: 811–815.
- [19] 徐鹏, 周岭华, 相建海. 中国对虾微卫星 DNA 的筛选[J]. 海洋与湖沼, 2001, 32: 255–259.
- [20] 刘丽, 刘楚吾, 郭昱嵩, 等. 青石斑鱼微卫星 DNA 标记的筛选及群体遗传多样性分析[J]. 中国水产科学, 2008, 15: 22–29.
- [21] Reddy O U K, Pepper A E, Abdurakhmonov I, et al. New dinucleotide and analysis of microsatellite marker resources for cotton research[J]. J Cott Sci, 2001, 5: 103–113.
- [22] Gao G, He G, Li Y. Microsatellite enrichment from AFLP fragments by magnetic beads[J]. Acta Botanica Sinica, 2003, 45(11): 1266–1269.
- [23] Fischer D, Bachmann K. Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.) [J]. Bio Techniques, 1998, 24: 796–802.
- [24] Kandpal R P, Kandpal G, Weissman S M. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers[J]. Proc Nat Acad Sci USA, 1994, 91: 88–92.
- [25] Ellegren H. Microsatellite evolution: a battle between replication slippage and point mutation[J]. Trends Genet, 2002, 18: 70.
- [26] Hines HC, Zikakis JP, Haenlein GF, et al. Linkage relationships among loci of polymorphisms in blood and milk of cattle[J]. Dairy Sci, 1981, 64(14): 71–76.
- [27] 李鸥, 赵莹莹, 郭娜, 等. 草鱼种群 SSR 分析中样本量及标记数量对遗传多度的影响[J]. 动物学研究, 2009, 30(2): 121–130.
- [28] 鲁翠云, 毛瑞鑫, 孙效文, 等. 鲤鱼三、四核苷酸重复微卫星座位的筛选及特征分析[J]. 农业生物技术学报, 2009, 17(6): 919–987.

## Isolation and characterization of tri- and tetranucleotide repeat microsatellites in grass carp (*Ctenopharyngodon idellus*)

LI Wensheng<sup>1,2</sup>, LIU Cui<sup>2,3</sup>, LU Cuiyun<sup>2</sup>, LI Chao<sup>2</sup>, SUN Xiaowen<sup>2</sup>

1. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China

2. Heilongjiang River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Harbin 150070, China

3. College of Life Science and Technology, Harbin Normal University, Heilongjiang Harbin 150025, China

**Abstract:** We isolated microsatellites of grass carp (*Ctenopharyngodon idellus*) by an enrichment by magnetic beads technique. One thousand three hundred and seventy-eight positive clones were obtained at the second screening by hybridization with [ $\gamma$ -<sup>32</sup>P]ATP-labeled probes. Eight hundred and forty-six microsatellites were ultimately isolated from 705 positive clones. There were 632 (74.7%) perfect motifs, 114 (13.48%) imperfect motifs, and 100 (11.82%) compound motifs among the sequences. One hundred and thirty-seven primer pairs were designed according to the flanking sequences, and 100 pairs were synthesized. Thirty-five of the primer pairs (35%) showed polymorphisms in the wild population from the Hanjiang River. Twenty microsatellite markers were assessed for their genetic diversity, and a total of 212 different alleles were found. The average observed heterozygosity ( $H_o$ ) was 0.681 1, the expected heterozygosity ( $H_e$ ) was 0.797 5, and the polymorphism information content (PIC) was 0.777 7. The results suggested that the isolated microsatellites were valid for use in population genetic studies of grass carp.

**Key words:** *Ctenopharyngodon idellus*; microsatellite; polymorphism

**Corresponding author:** SUN Xiaowen. E-mail: sunxw2002@163.com