

三疣梭子蟹基因组微卫星特征分析

宋来鹏^{1,2}, 刘萍¹, 李健¹, 刘振辉²

(1 中国水产科学研究院 黄海水产研究所, 山东 青岛 266071; 2 中国海洋大学 海洋生命学院, 山东 青岛 266003)

摘要: 对三疣梭子蟹 (*Portunus trituberculatus*) 部分基因组 DNA 文库测序, 获得了总长度为 622 409 个碱基的基因组 DNA 序列, 从中找到微卫星重复序列 (1~6 bp 重复) 697 个。统计微卫星重复类型, 以两碱基重复数目最多, 为 445 个, 占微卫星序列总数目的 63.84%; 其次是三碱基重复 152 个, 占 21.81%; 再次分别是单碱基重复 45 个, 占 6.46%; 四碱基重复 31 个, 占 4.45%; 五碱基重复 14 个, 占 2.01%; 六碱基重复 10 个, 占 1.43%。在单碱基重复类型中, 重复拷贝类别全部为 A; 两碱基重复类型中, AG 重复数目最多, 其次是 AC 和 AT; 三碱基重复类型中以 ACT 最多, 其次是 AGG 和 AAT; 四碱基重复类型中, AGAC 重复数目最多; 五碱基重复类型中, 以 AACCT 重复拷贝类别最多; 六碱基重复中以 AGGGGA 重复数目最多。GC 重复拷贝类别的重复数目很少, 只发现 1 个 (GenBank 注册号为 EU113241)。[中国水产科学, 2008, 15 (5): 738-744]

关键词: 三疣梭子蟹; 基因组; 微卫星

中图分类号: Q959.223

文献标识码: A

文章编号: 1005-8737-(2008)05-0738-07

三疣梭子蟹 (*Portunus trituberculatus*) 是一种重要的海洋经济动物, 隶属甲壳纲 (Crustacea)、十足目 (Decapoda)、梭子蟹科 (Portunidae)、梭子蟹属 (*Portunus*), 主要分布于中国、朝鲜、日本等的海域^[1-2], 是中国重要的渔业资源。近年来, 出于有效保护蟹类种质资源的目的, 关于其遗传特性、微卫星标记的开发和应用等研究日趋增多。

微卫星 (Microsatellites) 或称简单序列重复 (Simple sequence repeats, SSR), 是指由 1~6 个核苷酸组成的简单串联重复 DNA 序列。迄今研究过的所有生物种类中都发现了它的存在, 并且其分布密度很大^[3]。由于微卫星标记具有多态性丰富、高度杂合且稳定性好、遵循孟德尔分离定律、共显性遗传及易于 PCR 扩增等特点, 因此, 该标记正广泛应用于生物的家系分析、基因连锁分析、遗传图谱构建、种质鉴定及种群遗传多样性分析等许多研究领域。近年来国外学者已相继开发出邓杰内斯蟹 (*Cancer magister*)^[4]、锯缘青蟹 (正罳) (*Scylla serrata*)^[5]、青蟹 (沙罳) (*Scylla paramamosain*)^[6]、远海梭子蟹 (*Portunus pelagicus*)^[7]、蛛雪蟹 (*Chionoecetes opilio*)^[8-9]、中华绒螯蟹 (*Eriocheir sinensis*)^[10] 等重

要经济蟹类的微卫星标记, 并已应用于亲缘关系鉴定和种群遗传多样性研究中, 但关于三疣梭子蟹微卫星特征的研究还鲜有报道。目前国内, 除常玉梅等^[11]报道了中华绒螯蟹微卫星标记的筛选情况外, 很少有蟹类微卫星方面的研究报道。

本研究分析了针对三疣梭子蟹的大量的基因组随机测序结果中微卫星的重复序列, 旨在为阐明全基因组水平上微卫星重复序列的分布特征奠定初步基础, 以期深化对三疣梭子蟹基因组遗传结构特征的认识, 为将来开发微卫星标记以及开展深入的遗传标记辅助育种工作奠定理论基础。

1 材料与方法

1.1 材料来源

本研究所用的三疣梭子蟹于 2005 年 10 月中旬采集自黄海的海州湾。

1.2 三疣梭子蟹部分基因组 DNA 文库构建

提取三疣梭子蟹基因组 DNA, 用 Sau3AI 酶切后, 经低熔点琼脂糖凝胶电泳回收 500~1 500 bp 的片段, 与 PUC19 质粒连接后, 将重组 DNA 转化到大肠杆菌 DH5α 中, 从而建立三疣梭子蟹部分基因组文库。

收稿日期: 2007-09-10; 修订日期: 2007-11-15.

基金项目: 国家高技术研究发展计划 (863 计划) 项目 (2006AA10A406); 国家科技基础条件平台项目 (2006DKA30470); 青岛市科技计划项目 (07-2-3-5-jch).

作者简介: 宋来鹏 (1983-) 男, 硕士, 主要从事海水养殖生物种质资源与遗传多样性研究. E-mail: songlp1983@163.com

通讯作者: 刘萍, 研究员. E-mail: liuping@ysfri.ac.cn

1.3 克隆与测序

通过对此基因组文库 DNA 克隆进行随机测序,共获得了 4 164 个 DNA 随机克隆源序列。利用 Seqman II (DNASTAR 5.0) 软件,除去 PUC19 质粒中的载体序列和克隆 DNA 序列 2 段出现杂峰的污染序列后装配输出。其中在拼接序列时,如果 2 个克隆重合部分的序列为重复序列,则分别按单个克隆处理;只有重合部分的序列不是重复序列时,才把这些克隆序列拼接成 1 个序列来处理。最终得到 709 个 DNA 克隆,每个序列的长度从 500~1 500 bp 不等,代表着 622 409 个碱基的基因组总长度。

1.4 统计方法

1.4.1 分类 根据重复类型、重复数目 (Repeat number)、重复拷贝数 (Repeat copy number) 和重复拷贝类别对序列进行分类。

1.4.2 统计的标准 通过软件 Tandem Repeats Finder (Version 3.21) 对拼装后的克隆序列进行分析,查找微卫星序列。Tandem Repeats Finder 的查找参数如下: alignment parameters (match, mismatch, indel)=(2, 7, 7), minimum alignment score to report repeat=50, maximum period size=1000。利用本实验室编写的 Excel 宏程序对 Tandem Repeats Finder 的初步分析结果进行细化和汇总分析,结果如表 1 所示。

表 1 1~3 碱基重复类型
Tab. 1 Types of 1~3 bp repeats

重复类型 Repeat type	重复拷贝类别 Repeat copy type	对应碱基 Corresponding base
单碱基 Mononucleotide	A	A、T
	C	C、G
	AT	AT、TA
两碱基 Dinucleotide	AG	AG、GA、TC、CT
	AC	AC、CA、TG、GT
	GC	GC、CG
	AAT	AAT、ATA、TAA、TTA、TAT、ATT
	AAC	AAC、ACA、CAA、TTG、TGT、GTT
	AAG	AAG、AGA、GAA、TTC、TCT、CTT
	ATC	ATC、TCA、CAT、TAG、AGT、GTA
三碱基 Trinucleotide	ACG	ACG、CGA、GAC、TGC、GCT、CTG
	ACT	ACT、CTA、TAC、TGA、GAT、ATG
	AGC	AGC、GCA、CAG、TCG、CGT、GTC
	GCC	GCC、CCG、CGC、CGG、GGC、GCG
	AGG	AGG、GGA、GAG、TCC、CCT、CTC
	ACC	ACC、CCA、CAC、TGG、GGT、GTG

四碱基及其以上重复的兼并原则与表 1 相同,因 4~6 碱基重复的类型较为复杂,同时三疣梭子蟹基因组中该类型重复的拷贝数相对较少,这里不再一一列出。

2 结果与分析

2.1 微卫星序列的分类、数目和相应的百分比

通过软件分析,从 622 409 bp 长度的序列中共找到了 827 个重复序列,微卫星重复序列 (1~6 bp 重复) 为 697 个,占重复序列总数目的比例为

84.28%; 小卫星重复序列 (7~100 bp) 为 130 个,占重复序列总数目的比例为 15.72% (另文发表)。平均每 10 万碱基所具有的微卫星重复序列数目约为 112 个。统计微卫星重复类型,以两碱基重复数目最多,为 445 个,占微卫星序列总数目的 63.84%; 其次是三碱基重复 152 个,占 21.81%; 再次分别是单碱基重复 45 个,占 6.46%; 四碱基重复 31 个,占 4.45%; 五碱基重复 14 个,占 2.01%; 六碱基重复 10 个,占 1.43% (图 1)。

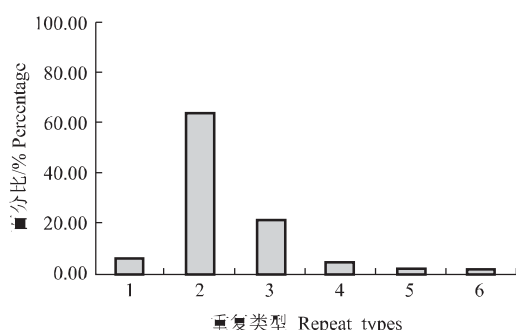


图1 不同重复类型的微卫星序列数目及其百分比

1. 单碱基; 2. 二碱基; 3. 三碱基; 4. 四碱基; 5. 五碱基; 6. 六碱基.

Fig. 1 The number and percentage of repeat types

1. Mononucleotide; 2. dinucleotides; 3. trinucleotides;

4. tetranucleotides; 5. pentanucleotides; 6. hexanucleotides.

在同类型的重复序列中,各重复拷贝类别占的比例也各不相同。在46个单碱基重复类型中,重复拷贝类别全部为A型,没有发现核心序列为C型的重复拷贝类别。两碱基重复类型中,AG型重复拷贝类别最多,为214个,占两碱基总重复序列数目的48.09%;其次是AC和AT,各为187(占两碱基总重复序列数目的42.02%)和43(占两碱基总重复序列数目的9.66%)个。只发现1个核心序列为GC的重复拷贝类别,即(GC)₁₄,占两碱基重复序列总数目的0.23%,该序列已经在GenBank注册,注册号为EU113241。三碱基重复类型中,共发现8种重复拷贝类别,它们分别是ACT(42个)、AGG(35个)、AAT(28个)、ACC(21个)、AAG(9个)、ATC(7个)、AAC(7个)和AGC(3个),其中以ACT最多,其次是AGG和AAT(表2)。

表2 1~3 碱基重复类型中重复拷贝类别及其在所属重复类型中的百分比

Tab.2 Repeat motifs and their percentage in 1-3 bp repeats

	重复类型 Repeat types												
	单碱基 Mononucleotide	两碱基 Dinucleotide				三碱基 Trinucleotide							
	A	AG	AC	AT	GC	ACT	AGG	AAT	ACC	AAG	ATC	AAC	AGC
数目 Number	45	214	187	43	1	42	35	28	21	9	7	7	3
合计 Total	45	445				152							
百分比 /% Percentage	100	48.09	42.02	9.66	0.23	27.63	23.03	18.42	13.82	5.92	4.61	4.61	1.97

在四碱基重复类型中,AGAC型重复拷贝类别最多,共14个(重复拷贝数范围是7.3~102.5,平均拷贝数是28.4);其次是AGAT,共4个(重复拷贝数范围是8.3~43.0,平均拷贝数是22.5)。其他的分别是:AAAT,共3个(重复拷贝数范围是6.3~25.8,平均拷贝数是16.1);ACTG,共3个(重复拷贝数范围是12.3~43.5,平均拷贝数是30.9);CACT共2个(重复拷贝数范围是17.8~21.5,平均拷贝数是19.7);ACTA共1个(重复拷贝数是52.8);ATGA共1个(重复拷贝数是24.5);CCTT共1个(重复拷贝数是13.8);CTCC共1个(重复拷贝数是10.0);CAGG共1个(重复拷贝数是8.8)。

在五碱基重复类型中,AACCT型重复拷贝类别最多,共6个(重复拷贝数范围是8.4~62.6,平均拷贝数是22.3);其次是TAACA共2个(重复拷贝数范围是7.0~9.6,平均拷贝数是8.3)和AGGTG共2个(重复拷贝数范围是7.8~10.8,平

均拷贝数是9.3)。其他的分别是:CCTTG共1个(重复拷贝数是48.2);CACCA共1个(重复拷贝数是14.2);AAATT共1个(重复拷贝数是6.2);TCCAC共1个(重复拷贝数是5.6)。

在六碱基重复类型中,AGGGGA型重复拷贝类别最多,共3个(重复拷贝数范围是5.0~8.5,平均拷贝数是6.4)。其他的分别是:CTCTCC共2个(重复拷贝数范围是6.7~33.7,平均拷贝数是20.2);TCTTCC共2个(重复拷贝数范围是5.2~11.3,平均拷贝数是8.3);TCCTCG共1个(重复拷贝数是10.8);AAAAGA共1个(重复拷贝数是6.5);TACTGC共1个(重复拷贝数是5.3)。

2.2 6种碱基重复类型中各种重复拷贝数的分布

三疣梭子蟹基因组微卫星的不同重复拷贝类别的重复拷贝数变化范围较大,从5~280都有分布,但其主要分布在12~70之间,占全部重复拷贝数的82.64%。与拷贝数的分布趋势相对应,微卫星

序列长度主要分布在 24~72 bp 的长度范围内。单碱基重复拷贝数主要分布在 28~40 和 68~76 两个范围,共 36 个位点,占单碱基重复类型总数 (45) 的 80.00%(图 2); 两碱基重复拷贝数主要分布在 12~36 范围内,共 285 个,占两碱基重复类型总数 (445) 的 64.04%(图 3); 三碱基重复拷贝数主要分布在 8~24 范围内,共 88 个,占三碱基重复类型总数 (152) 的 57.90%(图 4); 四碱基重复拷贝数主要分布在 7~26 范围内,共 23 个,占四碱基重复类型总数 (31) 的 74.19%(图 5); 五碱基重复拷贝数主要分布在 5~12 范围内,共 10 个,占五碱基重复类型总数 (14) 的 71.43%(图 6); 六碱基重复拷贝数主要分布在 4~12 之间,共 9 个,占六碱基重复类型总数 (10) 的 90.00%(图 7)。

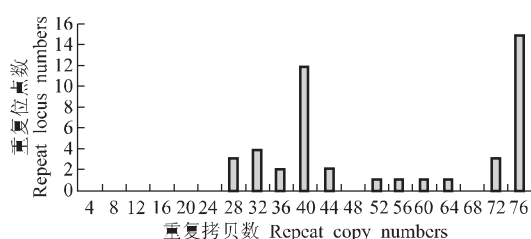


图 2 单碱基重复类型中不同的重复拷贝数所对应的重复位点数

Fig. 2 The repeat locus number corresponding to repeat copy number in mononucleotide-repeat type

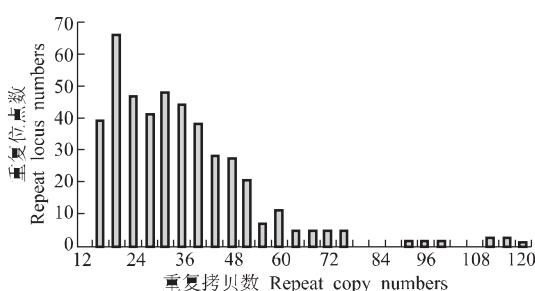


图 3 二碱基重复类型中不同的重复拷贝数所对应的重复位点数

Fig. 3 The repeat locus number corresponding to repeat copy number in dinucleotide-repeat type

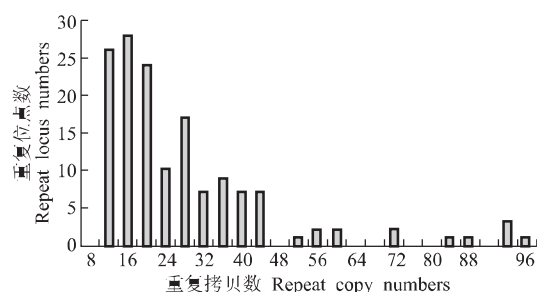


图 4 三碱基重复类型中不同的重复拷贝数所对应的重复位点数

Fig. 4 The repeat locus number corresponding to repeat copy number in trinucleotide-repeat type

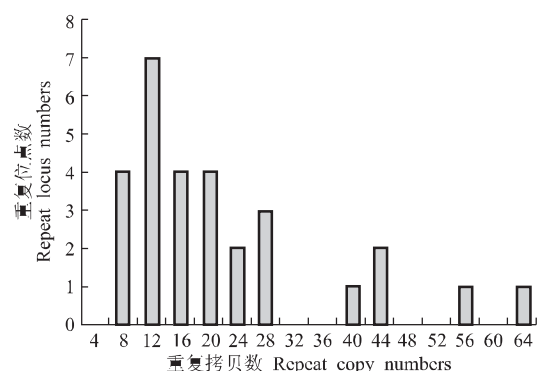


图 5 四碱基重复类型中不同的重复拷贝数所对应的重复位点数

Fig. 5 The repeat locus number corresponding to repeat copy number in tetranucleotide-repeat type

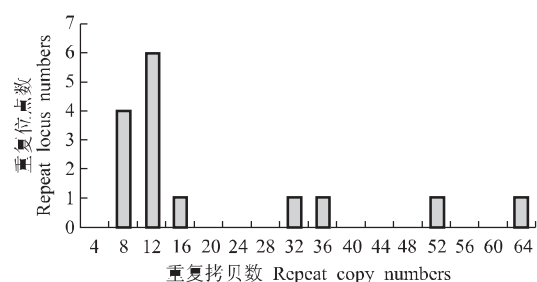


图 6 五碱基重复类型中不同的重复拷贝数所对应的重复位点数

Fig. 6 The repeat locus number corresponding to repeat copy number in pentanucleotide-repeat type

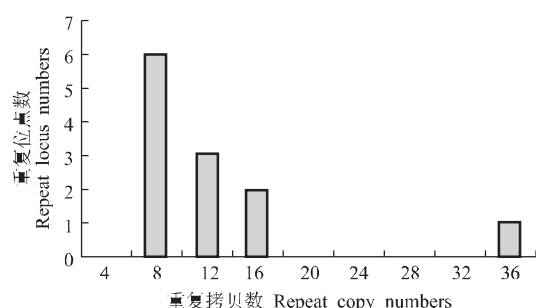


图7 六碱基重复类型中不同的重复拷贝数所对应的重复位点数

Fig. 7 The repeat locus number corresponding to repeat copy number in hexanucleotide-repeat type

3 讨论

3.1 三疣梭子蟹基因组微卫星分布特征

在三疣梭子蟹基因组单碱基重复类型中,46个单碱基重复类型中的重复拷贝类别全部为A型,没有发现核心序列为C型的重复拷贝类别。A型重复拷贝类别数量最多,这与对中国明对虾(*Fenneropenaeus chinensis*)^[12]和其他物种的研究结果相一致^[13-14]。两碱基重复类型中,以AG重复拷贝类别最为丰富,这与中国明对虾的AT^[12]、红鳍东方鲀(*Takifugu rubripes*)的AC^[15]重复拷贝最多不同,与有胚植物和真菌类生物也不同^[14]。

三碱基重复类型中,共发现8种重复拷贝类别,其中以ACT最多,其次是AGG和AAT。中国明对虾以AAT最多,其次是AAG和ATC^[12]。红鳍东方鲀中以AAT最多,其次是AGG和ATC, GCC最少^[15]。在所有的微卫星重复类型中,三碱基重复微卫星是最为特殊的一类,一些三碱基重复与某些遗传性病变密切相关^[16]。由于微卫星的不稳定性,它可以通过改变基因的表达来控制某些遗传疾病的发生^[17]。

四碱基重复类型中,AGAC重复拷贝类别最多,共14个;其次是AGAT,共4个;AAAT共3个,总体数量不是很丰富。AAAY(AAAT、AAAG、AAAC)重复拷贝类别在灵长类和啮齿类中最丰富^[14]。在中国明对虾中,AGAT重复拷贝类别最丰富,AAAY重复拷贝类别的数量总体上也很多^[12]。在红鳍东方鲀基因组中,AGAT重复拷贝类别最丰富,其次是ACAG、AGGT、ACCT,AAAY的数量不是很丰富^[15]。

三疣梭子蟹中五碱基重复类型的种类和数量都较少。在红鳍东方鲀基因组中五碱基重复类型的微卫星数量也较少,AGAGG的分布最为丰富,占到五碱基重复类型总数的46.3%^[15]。在本研究的五碱基重复类型中,AACCT重复拷贝类别最多,共6个;其次是TAACA和AGGTG各2个。关于五碱基重复微卫星的生物学意义报道不多,De Fonzo等^[18]发现五碱基重复可能与一些人类疾病相关。

六碱基重复类型中发现了6种重复拷贝类别,其中AGGGGA重复拷贝类别最多,共3个,但各种重复拷贝类别的拷贝数都很少。在其他生物中对此研究的较少,而且重复拷贝类别也不完全相同^[14]。由此可见,不同生物中各种重复类型中的重复拷贝类别和其重复数目是不同的,这是否与生物进化的程度有关,它们的不同对基因表达调控有何影响,这些都需要进一步地研究。

3.2 两碱基重复类型的分布特征及对GC重复过少的分析

Xu等^[19]研究斑节对虾(*Penaeus monodon*)基因组中两碱基重复类型时得到的初步结果是(CT)_n(即AG)最多,其次才是(AT)_n。高焕等^[12]在中国明对虾的研究中发现,AT重复拷贝类别的频率最高,占两碱基重复总数的42.44%;其次是AC,占两碱基重复数的34.42%。崔建洲等^[15]在对红鳍东方鲀的微卫星序列研究中发现,两碱基重复类型中,以AC重复拷贝类别最为丰富,接下来是AG和AT。在本研究中AG重复拷贝类别的频率最高,占两碱基重复总数的48.09%;其次是AC,占两碱基重复数的42.02%。AG重复拷贝类别含量高的特性与斑节对虾一致^[19]。

GC两碱基重复拷贝类别在所有已经研究过的生物基因组中的含量都很少^[13-14]。到目前为止,在中国明对虾基因组的研究中,除了徐鹏等^[20]发现了1个(GCG)₃的重复(GenBank登录号为AF295791)外,还未发现完全由GC重复拷贝类别组成的重复序列的存在。在本研究中只发现1个核心序列为GC的重复拷贝类别,占两碱基重复序列总数的0.23%(GenBank注册号为EU113241)。Schorderet等^[21]研究了6种脊椎动物基因组后,对此的解释是:由于基因组DNA中的CpG的甲基化,使之成为一个突变的热点,因为甲基化的胞嘧啶C很容易经过脱氨基作用转变成胸腺嘧啶T,而少量的GC又是维持DNA热力学稳定性所必须的。这样的

结果是 GC 重复减少,同时突变后的序列 TG (即 AC 类型) 相应增加,这可以一定程度上解释人类基因组中 AC 重复最多的现象。Lund 等^[22]的研究表明, CpG 的甲基化可以抑制玉米基因组中种子储藏蛋白基因 Zeins 的表达,但这种基因表现出很大的变异性,这种变异就是与胞嘧啶残基 C 脱氨基作用转变成胸腺嘧啶 T 相关。作者认为这可能也是三疣梭子蟹基因组中 GC 含量稀少的原因,因为与之相对应的突变类型 AC 重复的量仅次于 AG 重复。

3.3 微卫星标记技术在三疣梭子蟹分子遗传学研究中的应用前景

微卫星标记技术虽然建立的时间不长,但已经在包括人类到细菌的基因组研究中得到了广泛的应用,并取得了许多重要成果。由于它具有多态性高、共显性、容易用 PCR 检测和结果稳定可靠等特点,因此是一种十分理想的分子标记。同时,由于微卫星 DNA 在生物基因组中具有丰富的长度多态性信息,在群体间和群体内具有变异大、杂合性高、种类多、分布广以及重组率低、容易筛选等优点,因而在甲壳动物蟹类的遗传多样性研究中也得到了日益广泛的应用。

蟹的种类很多,目前仅在小部分蟹类中有微卫星分离的报道^[6-7],微卫星标记在蟹类遗传学研究中的应用还处于初级阶段。本研究表明,在总的重复序列中占绝对优势的两碱基重复拷贝数主要分布在 12~36 的范围内,即三疣梭子蟹基因组中的微卫星长度主要分布在 24~72 bp 的长度范围内,这说明在长期进化中微卫星位点积累的变异相对较大,因而而这些微卫星序列应该具有较为丰富的多态信息。由此可见,随着蟹类分子遗传学研究和分子生物学技术的飞速发展,微卫星标记技术在三疣梭子蟹种群遗传、品系亲缘关系分析、家系分析、亲子鉴定、个体识别、遗传图谱构建、基因连锁分析和遗传多样性保护等研究领域的应用将更加广泛而深入。

致谢:中国水产科学研究院黄海水产研究所育种室栾生博士提供了文中所应用的 Excel 宏程序,并在本文的修改定稿中给予大量指导性意见,在此表示真挚的感谢。

参考文献:

[1] 戴爱云,冯钟琪,宋玉枝,等.三疣梭子蟹渔业生物资源的初步调查[J].动物学杂志,1977,12(2):30-33.

[2] 戴爱云,杨思琼,宋玉枝,等.中国海洋蟹类[M].北京:海洋出版社,1986:213-214.

[3] Zane L, Bargelloni L, Patarnello T. Strategies for microsatellites isolation: a review [J]. Mol Ecol, 2002, 11: 1-16.

[4] Jensen P C, Bentzen P. Isolation and inheritance of microsatellite loci in the Dungeness crab (Brachyura: Cancridae: Cancer magister) [J]. Genome, 2004, 47: 325-331.

[5] Gopurenko D, Hughes J M, Ma J. Identification of polymorphic microsatellite loci in the mud crab *Scylla serrata* (Brachyura: Portunidae) [J]. Mol Ecol Notes, 2002, 2(4): 481-483.

[6] Takano M, Barinova A, Sugaya T, et al. Isolation and characterization of microsatellite DNA markers from mangrove crab, *Scylla paramamosain* [J]. Mol Ecol Notes, 2005, 5(4): 794-795.

[7] Yap E S, Sezmis E, Chaplin J A. Isolation and characterization of microsatellite loci in *Portunus pelagicus* (Crustacea: Portunidae) [J]. Mol Ecol Notes, 2002, 2(1): 30-32.

[8] Puebla O, Parent E, Sevigny J M. New microsatellite markers for the snow crab *Chionoecetes opilio* (Brachyura: Majidae) [J]. Mol Ecol Notes, 2003, 3(4): 644-646.

[9] An H S, Jeong J H, Park J Y. New microsatellite markers for the snow crab *Chionoecetes opilio* (Brachyura: Majidae) [J]. Mol Ecol Notes, 2006, 7(1): 86-88.

[10] Hanfling B, Weetman D. Characterization of microsatellite loci for the Chinese mitten crab, *Eriocheir sinensis* [J]. Mol Ecol Notes, 2003, 3(1): 15-17.

[11] Chang Y M, Liang L Q, Li S W, et al. A set of new microsatellite loci isolated from Chinese mitten crab, *Eriocheir sinensis* [J]. Mol Ecol Notes, 2006, 6(4): 1237-1239.

[12] 高焕,刘萍,孟宪红,等.中国对虾基因组微卫星特征分析[J].海洋与湖沼,2004,35(5):424-431.

[13] Katti M V, Ranjekar P K, Gupta V S. Differential distribution of simple sequence repeats in eukaryotic genome sequences [J]. Mol Biol Evol, 2001, 18: 1161-1167.

[14] Toth, Gspri Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis [J]. Gen Res, 2000, 10(7): 967-981.

[15] 崔建洲,申雪艳,杨官品,等.红鳍东方鲀基因组微卫星特征分析[J].中国海洋大学学报,2006,36(2):249-254.

[16] Pearson C E, Sinden R. Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA [J]. Curr Opin Struct Biol, 1998, 8: 321-330.

[17] Cummings C J, Zoghbi H Y. Fourteen and counting: unraveling trinucleotide repeat diseases [J]. Hum Mol Genet, 2000, 9: 909-916.

- [18] De Fonzo V, Bersani E, Aluffi P F, et al. Are only repeated triplets guilty [J]. J Theor Biol, 1998, 194: 125-142.
- [19] Xu Z, Dhar A K, Wyrzykowski J, et al. Identification of abundant and informative microsatellites from shrimp (*Penaeus monodon*) genome [J]. Anim Genet, 1999, 30 (2): 150-156.
- [20] 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.
- [21] Schorderet D F, Gartler S M. Analysis of CpG suppression in methylated and nonmethylated species [J]. Proc Natl Acad Sci USA, 1992, 89: 957-961.
- [22] Lund G, Lauria M, Guldberg P, et al. Duplication-dependent CG suppression of the seed storage protein genes of maize [J]. Genetics, 2003, 165: 835-848.

Analysis of microsatellite sequences in genome of crab, *Portunus trituberculatus*

SONG Lai-peng^{1,2}, LIU Ping¹, LI Jian¹, LIU Zhen-hui²

(1. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China; 2. Marine Life College, Ocean University of China, Qingdao 266003, China)

Abstract: By sequencing randomly, 4 164 clones of sequences in the genomic library of crab *Portunus trituberculatus* were obtained. Software DNASTAR (Version 5.0) was used to assemble all of the clones in this study. The length of DNA sequences is about 622 409 bp totally. With the help of the bio-software Tandem Repeats Finder (Version 2.02), 697 microsatellite repeat sequences are found in the sequences. In the 697 repeat sequences, the number of dinucleotide repeat is 445, being the richest (63.84%) of all the repeat sequences. The second one is trinucleotide repeat with 152 (21.81%); the third one is mononucleotide repeat with 45 (6.46%); the forth one is tetranucleotide repeat with 31 (4.45%); the fifth one is pentanucleotide repeat with 14 (2.01%); the sixth one is hexanucleotide repeat with 10 (1.43%). Forty-five mononucleotide repeat sequences are all composed of the motif of A, while the motif of C was not found among the mononucleotide repeats. In dinucleotides repeats, the number of AG repeat is 214, accounting for 48.09%; and the numbers of AC and AT repeats are 187 (42.02%) and 43 (9.66%), respectively. Eight classes of repeat sequences that include motifs of ACT, AGG, AAT, ACC, AAG, ATC, AAC and AGC are found in trinucleotides repeat, in which the number of ACT repeats is the largest with 42; the second one is AGG (35); the others are AAT (28), ACC (21), AAG (9), ATC (7), AAC (7) and AGC (3) in turn. AGAC, AACCT and AGGGGA repeats are the richest ones in tetranucleotide-, pentanucleotide- and hexanucleotide-repeat, respectively. Only one GC dinucleotide repeat is found in the study and its GenBank accession number is EU113241. The reason of fewer GC repeat is possibly that methylation of C in CpG islands results in mutation of C-T or that it is difficult to sequence the GC repeat sequences.

Distributions of copy numbers in different types of repeat sequences are as follows: copy numbers of mononucleotide repeats are mainly between 28 and 40 or between 68 and 76, accounting for 80.00% totally; copy numbers of dinucleotides are mainly between 12 and 36, accounting for 64.04%; copy numbers of trinucleotides repeats are mainly between 8 and 24, accounting for 57.90%; copy numbers of tetre-, penta- and hexanucleotides repeats together are mainly between 4 and 12. In general, the length of microsatellite repeat sequences are mainly 24-72 bp. Based on the above, it can be concluded that the nucleotide mutation of microsatellite locations has been accumulated largely in a long term of evolution; and there would be abundant polymorphism in these locations. Therefore, it would be practical to use microsatellite to study the genome of *P. trituberculatus* and the method would be applied to a variety of fields including population differentiation, kinship analysis, linkage analysis, and evolutionary and ecological studies. This study provides base for *P. trituberculatus* microsatellite researches. [Journal of Fishery Sciences of China, 2008, 15 (5): 738-744]

Key words: *Portunus trituberculatus*; microsatellite; genome

Corresponding author: LIU Ping. E-mail: liuping@ysfri.ac.cn