

文章编号:1000-0615(2008)06-0838-09

## 三疣梭子蟹基因组小卫星特征分析

宋来鹏<sup>1,2</sup>, 刘萍<sup>1</sup>, 李健<sup>1</sup>, 刘振辉<sup>2</sup>

(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:** 通过构建三疣梭子蟹部分基因组文库,对获得的4 164个克隆测序,共获得总长度为622 409 bp的DNA序列。在序列拼接后长度500~1 500 bp的709个克隆中,筛选到130个小卫星序列,其序列总长度占测序序列总长度的2.55%。小卫星序列中,以12 bp重复单位的序列数量最多(10.77%),总体趋势表现为重复单位越长,相应的重复序列数目越少( $R = -0.663, P < 0.01$ )。小卫星重复单位拷贝数分布范围以8 bp重复单位最广为3.9~66.5;其次是13 bp重复范围在2.0~40.6;再次是26 bp重复,范围在2.3~21.0。平均拷贝数最高的三种重复类型分别为8 bp重复(19.96),25 bp重复(16.00)和22 bp重复(15.85)。小卫星序列中各重复单位的拷贝数分布范围2~66.5,集中分布在2~25,且表现为拷贝数目越大,其相应的重复序列数目越低的趋势。130个重复序列分别由123种重复单位所组成,因而小卫星重复序列的类型很多。初步分成三类:2种碱基组成类别、3种碱基组成类别和4种碱基组成类别,并进一步根据各个重复序列中所含有的碱基种类的数量从大到小排列这些碱基而分成若干小类。从这些分类中可以看出,三疣梭子蟹基因组中的小卫星整体上是富含A/T的重复序列,并揭示了其与微卫星重复序列之间的关系,即一部分小卫星重复序列可能起源于微卫星序列。

**关键词:** 三疣梭子蟹;基因组;重复序列;小卫星

**中图分类号:** S 917      **文献标识码:** A

基因组中的卫星序列一般可以分为两类:微卫星序列(microsatellite sequence)和小卫星序列(minisatellite sequence)<sup>[1]</sup>。一般微卫星序列是指由1~6 bp重复单位组成的重复序列,小卫星序列是指由7~100 bp左右的重复单位组成的重复序列<sup>[2-3]</sup>。其中小卫星序列,由于其串联重复单位的数目在不同个体基因组的相同位点上数目都不同,被称为可变数目串联重复序列(Variable number tandem repeats, VNTRs)<sup>[4]</sup>。这些小卫星序列不仅大量分布于真核生物的基因组中,也大量存在于许多原核生物的基因组中<sup>[5-7]</sup>,因而这些序列在基因组中的生物功能越来越受重视。

伴随着各种生物基因组测序的进展,基于基因组水平的重复序列的调查工作也随之展开。Vergnaud<sup>[8]</sup>等人已经建立了包括人类、秀丽隐杆线虫(*Caenorhabditis elegans*)、拟南芥(*Arabidopsis thaliana*)和一些原核生物的同向重复序列数据(<http://minisatellites.U-psud.Fr>)。由于水生生物基因组研究工作的相对落后,到目前为止,这方面的研究工作还相对较少,仅见在中国对虾(*Fenneropenaeus chinensis*)<sup>[9]</sup>、日本囊对虾(*Marsupenaeus japonicus*)<sup>[10]</sup>、牙鲆(*Paralichthys olivaceus*)<sup>[11]</sup>和栉孔扇贝(*Chlamys farreri*)<sup>[12]</sup>等物种中开展了微卫星和小卫星序

收稿日期:2007-09-12

资助项目:国家“八六三”高技术研究发展计划(2006AA10A406);国家自然科学基金(30871933);青岛市科技计划项目(07-2-3-5-jch)

作者简介:宋来鹏(1983-)男,山东聊城人,硕士,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,E-mail:liuping@ysfri.ac.cn

列分析的研究工作。

三疣梭子蟹 (*Portunus trituberculatus*) 是一种重要的海洋经济动物,分布于中国、朝鲜、日本等海域<sup>[13-14]</sup>,是中国重要的渔业资源。虽然关于三疣梭子蟹的微卫星、小卫星特征情况的研究国内外还鲜有报道,但是在其他蟹类方面,近年来国外学者已相继开发出邓杰内斯蟹 (*Cancer magister*)<sup>[15]</sup>,锯缘青蟹 (正 罇) (*Scylla serrata*)<sup>[16]</sup>,锯缘青蟹 (沙 罇) (*Scylla paramamosain*)<sup>[17]</sup>,远海梭子蟹 (*Portunus pelagicus*)<sup>[18]</sup>,蛛雪蟹 (*Chionoecetes opilio*)<sup>[19-20]</sup>,中华绒螯蟹 (*Eriocheir sinensis* Milne-Edwards)<sup>[21]</sup>等重要经济蟹类的微卫星标记,并已应用于亲缘关系鉴定、种群多样性研究中。鉴于小卫星序列在了解整个基因组序列特征及其遗传进化方面的重要作用,因而在目前仍没有开展三疣梭子蟹基因组大规模系统测序的前提下,通过对随机测序获得的 DNA 序列中的小卫星重复序列的分析,对于初步了解三疣梭子蟹基因组序列的特征,以及开发抗病、抗逆方面的小卫星标记等具有重要的指导意义。

## 1 材料与方 法

### 1.1 材料来源

三疣梭子蟹基因组 DNA 文库构建见文献 [1],对测序的序列进行分析。

### 1.2 小卫星重复序列的判定标准和分类方法

通过软件 Tandem Repeats Finder (Version 3.21)对拼装后的克隆序列进行分析,查找小卫星序列。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 = 1 000。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。判定是否为小卫星重复序列的人工细化分析的标准如下:重复序列中重复单位的长度在 7~80 bp,拷贝数目 ≥ 2 个。设定一个小卫星序列中的最高众数重复单位序列作为标准重复单位,且与标准重复单位相比较,各重复单位平均碱基匹配率(即一致性)在 70%以上。

使用变异系数(coefficient of variation)来衡量小卫星重复类型的变异水平高低,变异系数的

计算公式为:

$$CV = \frac{S}{\bar{x}}$$

其中,  $S$  为小卫星拷贝数的标准差,  $\bar{x}$  为小卫星序列拷贝数平均值。变异系数  $CV$  可以消除单位和(或)平均数不同对两个或多个资料变异程度比较的影响,能够真实的反应重复类型变异能力的大小,便于进行类型间的比较。

根据小卫星重复序列中碱基组成种类和各碱基的数目划分小卫星的类别。由于不同小卫星重复序列中碱基种类和数量的不同,而且几乎每一个重复序列都是由各自的重复单位组成的,因而这给小卫星的归类带来了困难。在这方面,通常根据小卫星序列中碱基组成的不同,简单的归为 A-T 含量丰富的和 G-C 含量丰富的序列。栾生等<sup>[10]</sup>在日本囊对虾、高焕等<sup>[22]</sup>在中国对虾的研究中采用了将小卫星序列中碱基组成成分成三类的研究方法,本研究也采用此种方法。

首先根据小卫星序列中碱基组成种类分成三类<sup>[22]</sup>,两碱基组成类别(重复序列是由 2 种碱基组合形成的),三碱基组成类别(重复序列是由 3 种碱基组合形成的)和四碱基组成类别(重复序列是由 4 种碱基组合形成的);其次对每种类别再进一步分类,方法如下类推:如 CCTGCCTTCCCTGCCTTC CCTGCCTTC CCTGCCTTC 这个重复序列是由 CCTGCCTTC 重复单位重复 4 次组成的,共由 C、T、G3 种碱基组成,它们在这个重复序列中的数目分别为 20、12 和 4,则把此小卫星重复序列归为 CTG 类,即根据整个重复序列碱基数目多少依次排列它们(相同碱基数目的按照 ATCG 顺序排列),考虑到 DNA 双链互补的情况,CTG 类也代表了 GAC 类。如有简并小卫星重复单位,即小卫星重复单位与标准重复单位相比有个别碱基发生突变,如上述小卫星重复单位 CCTGCCTTC 在 T 碱基处发生突变成 CCAGCCTTC,也按上述方法归类,则上面的重复序列相应的应由 CTG 类归为 CTGA 类(GACT 类)。

## 2 结果

### 2.1 小卫星重复序列在基因组中的分布特征

通过软件分析,从 622 409 个碱基长度的序列

中共找到了 827 个重复序列,小卫星重复序列为 130 个,占重复序列总数目的比例为 15.72%。平均每 10 万碱基所具有的小卫星重复序列数目约为 21 个。

130 个小卫星重复序列中,总共筛选到了 123 种重复类型,平均 1.06 个重复序列中就能够发现一种重复类型。不同长度重复单位的重复序列数目分布情况为:以 12 bp 重复单位的序列数目最多(14 个),占小卫星重复序列总数目的 10.77% (图 1),20 bp 重复单位次之(11 个),比例为 8.46%,总体趋势表现为随着重复单位长度的增加,相应的重复序列数目降低,统计相关分析结果表明二者间存在着显著负相关( $r = -0.663$ ,  $P < 0.01$ )。小卫星序列总长度为 15 892 bp,不同长度重复单位类型的重复序列累计长度以 26 bp 重复单位的序列长度为最高(2 082 bp),其次为 8 bp 重复单位(1 120 bp),12 bp 重复单位(1 067 bp),31 bp 重复单位(1 005 bp),20 bp 重复单位(995 bp),统计相关检验表明二者存在着负相关关系,但是相关系数较小( $r = -0.458$ ,  $P < 0.01$ )。

**2.2 小卫星重复单位拷贝数分布及变异能力分析** 根据重复单位长度划分的小卫星各种序列类型的重复单位拷贝数、平均拷贝数和拷贝数范

围统计结果见表 1。130 个小卫星重复序列中,重复单位拷贝数范围以 8 bp 重复拷贝数范围最广,为 3.9~66.5;其次是 13 bp 重复,范围在 2.0~40.6;再次是 26 bp 重复,范围在 2.3~21.0。平均拷贝数最高的 3 种重复类型分别为 8 bp 重复(19.96),25 bp 重复(16.00)和 22 bp 重复(15.85)。

变异系数是衡量观测值变异程度的一个统计量,变异系数越大,这种重复类型的变异能力越大。我们计算了小卫星各种类型重复单位拷贝数的变异系数,分析结果见表 1。变异能力最强的前 5 种类型分别是:13 bp (165.43),8 bp (108.85),20 bp (97.94),23 bp (95.05),21 bp (82.37)。小卫星序列重复单位长度与变异系数的相关分析表明,尽管二者的相关性表现为显著,但是相关系数较小( $r = -0.309$ ,  $P < 0.01$ )。小卫星序列中各重复单位的拷贝数分布范围 2~66.5,集中分布在 2~25,不同拷贝数目所对应的重复序列数量的分布情况为:拷贝数目为 3 的重复单位所组成的重复序列数目最多(43 个),其次是拷贝数目为 4 的重复序列(15 个),随着拷贝数目的增加,由其所组成的重复序列的数目呈递减的趋势,二者呈显著的负相关( $r = -0.592$ ,  $P < 0.01$ )(图 2)。

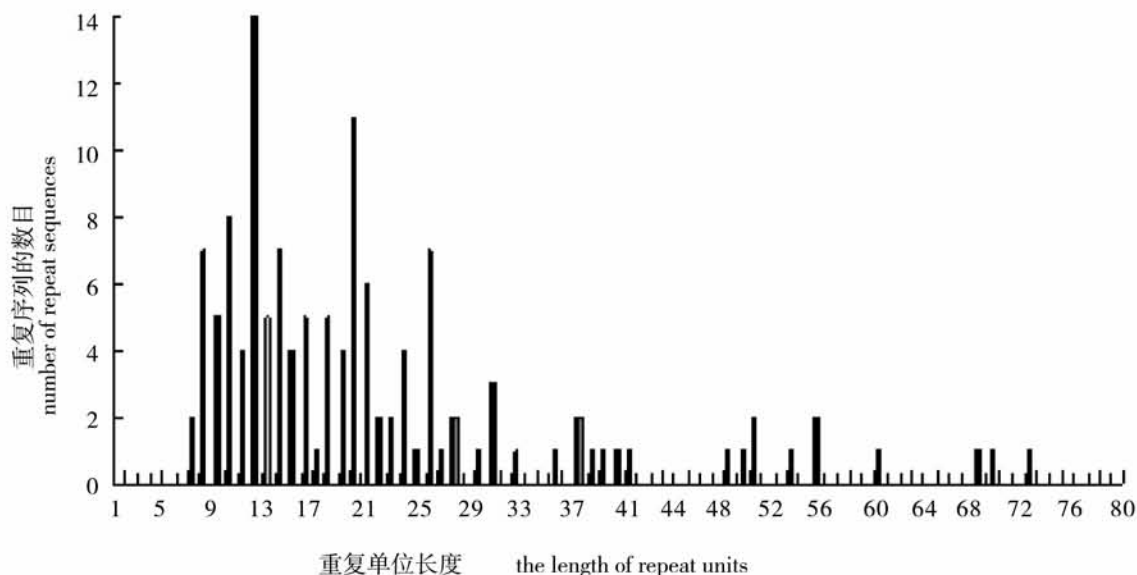


图 1 不同长度的重复单位组成的重复序列的数量分布

Fig. 1 The distributions of repeat sequences composed of different repeat units

表 1 小卫星重复序列的频率和分布特征

Tab. 1 Frequency and distribution of minisatellite repeat types

重复单位 长度类型 length of repeat unit	重复类型 数量 No. of repeat types	小卫星 数量 No. of mini- satellite	占小卫星总 数百分比% per of total No. of mini- satellite	序列长度 sequence length	占小卫星总 长度百分比% per of total length of mini-satellite	拷贝数范围 range of copy number	累积拷贝数 accumul- ative copy numbers	平均 拷贝数 average of copy numbers	拷贝数 变异系 CV of copy numbers
7	2	2	1.54	92	0.58	4.7~8.9	13.60	6.80	43.67
8	7	7	5.38	1120	7.05	3.9~66.5	139.70	19.96	108.85
9	5	5	3.85	255	1.60	3.3~7.4	27.90	5.58	39.12
10	7	8	6.15	367	2.31	3.1~9.1	36.40	4.55	45.92
11	4	4	3.08	526	3.31	3.7~20.8	49.80	12.45	73.54
12	14	14	10.77	1067	6.71	2.1~16.5	89.50	6.39	79.50
13	5	5	3.85	681	4.29	2.0~40.6	51.30	10.26	165.43
14	3	7	5.38	224	1.41	2.3~2.3	16.1	2.30	0.00
15	4	4	3.08	264	1.66	2.3~8.8	18.00	4.50	67.57
16	5	5	3.85	217	1.37	1.9~4.3	13.70	2.74	35.81
17	1	1	0.77	35	0.22	2.1	2.1	2.1	
18	5	5	3.85	244	1.54	1.9~3.6	13.20	2.64	29.78
19	4	4	3.08	235	1.48	2.5~4.7	14.30	3.58	29.91
20	11	11	8.46	995	6.26	1.9~17.2	49.20	4.47	97.94
21	5	6	4.62	614	3.86	2.0~10.6	29.20	4.87	82.37
22	1	2	1.54	695	4.37	8.4~23.3	31.70	15.85	66.47
23	2	2	1.54	286	1.80	2.0~10.2	12.20	6.10	95.05
24	4	4	3.08	272	1.71	2.3~3.1	11.30	2.83	12.70
25	1	1	0.77	390	2.45	16.0	16.00	16.00	
26	7	7	5.38	2082	13.10	2.3~21.0	81.60	11.66	61.93
27	1	1	0.77	62	0.39	2.3	2.30	2.30	
28	2	2	1.54	143	0.90	2.2~2.9	5.10	2.55	19.41
30	1	1	0.77	182	1.15	6.1	6.10	6.10	
31	3	3	2.31	1005	6.32	2.3~17.4	32.50	10.83	71.47
33	1	1	0.77	64	0.40	1.9	1.90	1.90	
36	1	1	0.77	189	1.19	5.3	5.30	5.30	
38	2	2	1.54	265	1.67	2.6~4.4	7.00	3.50	36.37
39	1	1	0.77	118	0.74	3.0	3.00	3.00	
40	1	1	0.77	414	2.61	10.4	10.40	10.40	
41	1	1	0.77	111	0.70	2.7	2.70	2.70	
42	1	1	0.77	93	0.59	2.2	2.20	2.20	
49	1	1	0.77	369	2.32	7.6	7.60	7.60	
50	1	1	0.77	100	0.63	2.0	2.00	2.00	
51	2	2	1.54	433	2.72	2.5~5.9	8.40	4.20	57.24
54	1	1	0.77	129	0.81	2.4	2.40	2.40	
56	2	2	0.77	224	1.41	2.0~2.0	2.00	2.00	0.00
61	1	1	0.77	272	1.71	4.5	4.50	4.50	
69	1	1	0.77	352	2.21	5.0	5.00	5.00	
70	1	1	0.77	159	1.00	2.3	2.30	2.30	
73	1	1	0.77	547	3.44	7.6	7.60	7.60	

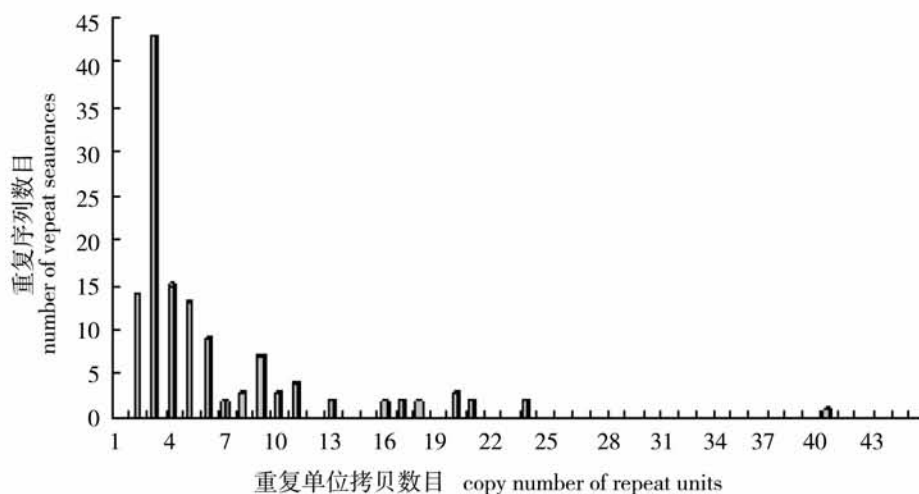


图2 不同拷贝数目所对应的重复序列数目

Fig. 2 The numbers of repeat sequences corresponding to the copy numbers of repeat units

### 2.3 小卫星重复单位的碱基组成特征

在查找到的130个小卫星重复序列中,有4个同一个重复单位所组成的不同拷贝数的重复序列(如重复单位TGTATTTACCTAGT与

TACAAGTAGGTTAAA被归为一类),即在这些重复序列中,130个重复序列代表了123种重复单位所组成的重复序列。对这些序列按照材料和方法中的统计原则进行分类(表2)。

表2 小卫星重复序列的碱基组成及其分类

Tab. 2 The bp composition and classification of minisatellite repeat sequences

两碱基组成类别 dinucleotide minisatellite sequences		三碱基组成类别 trinucleotide minisatellite sequences		四碱基组成类别 tetranucleotide minisatellite sequences	
亚类型 sub-type	数目 numbers	亚类型 sub-type	数目 numbers	亚类型 sub-type	数目 numbers
AT/TA	1	ATC/TAG	3	ATCG/TAGC	13
AC/TG	4	ATG/TAC	6	ATGC/TACG	15
AG/TC	3	ACT/TGA	5	ACTG/TGAC	10
CT/GA	3	AGT/TCA	1	AGTC/TCAG	9
CA/GT	2	ACG/TGC	1	ACGT/TGCA	1
		AGC/TCG	1	AGCT/TCGA	2
		CAT/GTA	2	CTAG/GATC	10
		CTA/GAT	2	CTGA/GACT	4
		CTG/GAC	3	CGTA/GCAT	1
		CAT/GTA	1	AGAC/TCTG	1
		TACG/ATGC	15	CATG/GTAC	10
		TCGC/AGCG	1		
总计 total	13		25		92
			130		
百分比 percentage	10.00%		19.23%		70.77%

从表 2 中可以看出,三疣梭子蟹基因组小卫星序列主要以四碱基组成类别组成,达到 70.77%。两碱基组成类别最少,只有 10.00%。在两碱基组成类别的重复序列中,富含 A/T 的序列为  $AT + AC + AG = 1 + 4 + 3 = 8$ ,富含 G/C 的序列为  $CT + CA = 5^{[22]}$ ,富含 A/T 的序列大于富含 G/C 的序列;在三碱基组成类别的序列中,富含 A/T 序列数目可以保守的估计为  $ATC + ATG + ACT + AGT = 3 + 6 + 5 + 1 = 15$ ,远远大于富含 G/C 的序列数目(约  $CTG = 3$ );在四碱基组成类别的重复序列中,富含 A/T 的序列约为  $ATCG + ATGC + ACTG + AGTC = 13 + 15 + 10 + 9 = 47$ ,也明显大于富含 G/C 的序列,因而总体上,三疣梭子蟹基因组中的小卫星是富含 A/T 的重复序列,主要以 ATN 和 ATNN(N 代表 G 或 C)类别为主<sup>[22]</sup>。

### 3 讨论

#### 3.1 三疣梭子蟹小卫星序列的分布特征

本研究通过建立三疣梭子蟹部分基因组文库,并对此基因组文库 DNA 克隆片段的随机测序(每个克隆 DNA 序列均由双向测序得到),获得了代表着 622 409 个碱基的基因组总长度。在这些基因组序列中,共筛选到了 130 个小卫星序列,序列总长度为 15 892 bp,约占基因组测序串联重复序列总长度的 15.72%,约占测序序列总长度的 2.55%,即平均每 1 000 bp 核苷酸序列中包含 25.5 bp 的小卫星序列。

从结果中可以发现,三疣梭子蟹部分基因组文库中的小卫星序列的分布有两个较为显著的特点,一是随着重复单位长度的增加,其所对应的重复序列的数目在减少;另一个特点是小卫星序列重复单位拷贝数主要集中在低拷贝区(2~22),而且重复单位拷贝数目与小卫星序列数目间存在着较为显著的负相关。高焕等<sup>[22]</sup>对中国对虾随机基因组小卫星序列特征分析结果表明,重复单位长度与其重复序列数目间存在着负相关,小卫星序列中以 12 bp 重复单位的序列数目为最多,在重复单位低拷贝数(2~21)范围内分布着大部分的小卫星序列。栾生等<sup>[10]</sup>对日本囊对虾小卫星序列的特征分析表明,随着重复单位长度的增加,相应的重复序列数目降低,统计相关分析结果表明二者间存在着极显著负相关( $r = -0.826$ ,

$P < 0.01$ )。不同长度重复单位类型的重复序列累计长度的分布情况与序列数目的分布趋势类似,也以 12 bp 重复单位的序列长度为最高。考虑到中国对虾、日本囊对虾与三疣梭子蟹一样都同属于甲壳纲动物,这个分析结果在一定程度上显示出小卫星在进化上存在着的一致性。

简纪常等<sup>[23]</sup>对鳙的一个小卫星测序的结果表明其长度达到了约 1 900 bp,我们研究中的小卫星序列的长度可能受到了测序的影响,因为所有的单个序列克隆的长度约 300~800 个碱基,即使拼接以后最大长度也不过 1 000 bp 左右,而且个别较长的小卫星序列的 5'端测序不完整,这也说明了可能还有更长小卫星序列的存在。本研究在三疣梭子蟹基因组中发现一个由 TTGTGTTTTGCAGCATTTAATAGTAGGA-AGC 31 个碱基为核心序列长度为 538 bp 和一个由 CTACCTGGTGACACCCAACCTGGTGCA-CCCACAACCTCCCTCACACACCACAATATGCCACTAACTTAACAGCC 73 个碱基为核心序列长度为 547 bp 的小卫星序列,这些由单一较长重复单位(30 bp 以上)重复多次组成的长达 500 bp 以上的重复序列,在其它水生生物的研究中报道的并不多。

Hancock<sup>[24]</sup>认为重复序列的数量与基因组大小是有关联的,重复序列越多,基因组越大。在基因组很小的噬菌体 M13 中,已经有了小卫星重复序列的存在<sup>[25]</sup>,挪威云杉[ *Picea abies* (L.) Karst ]的线粒体中也发现了小卫星序列的存在<sup>[26]</sup>。Klevytska 等<sup>[27]</sup>对鼠疫杆菌( *Yersinia pestis* )基因组中的重复序列研究表明,每万碱基中重复序列的数量为 2 118 个,9 bp 以下的重复序列数量约为  $2.18 \times 84\% = 1.83$  个,即每万碱基小卫星的数量平均约为  $2.18 - 1.83 = 0.35$  个。在水生动物甲壳类中,高焕等<sup>[22]</sup>对中国明对虾的基因序列分析表明其每万碱基约含有 6.2 个小卫星序列。栾生等<sup>[10]</sup>对日本囊对虾的基因序列分析表明其每 1 000 bp 核苷酸序列中包含 37.9 bp 的小卫星序列。在本研究中,三疣梭子蟹平均每 1 000 bp 核苷酸序列中包含 25.5 bp 的小卫星序列。

#### 3.2 三疣梭子蟹小卫星序列的碱基组成特征及其分类

三疣梭子蟹基因组 130 个小卫星重复序列中,四碱基组成类别的序列占 70.07%,可见

三疣梭子蟹基因组小卫星重复序列主要是四碱基组成类别组成的。人源小卫星 33.15 的核心序列 AGAGGTGTGGCAGGTGG<sup>[28]</sup>、大肠杆菌重组  $\chi$  信号顺序(GCTGGTGG)<sup>[29]</sup>、染色体端粒重复序列 TTTAGGG<sup>[30]</sup>、噬菌体 M13 中的一个常被用来做 DNA 指纹图谱鉴定的核心序列“GAGGGTGGCGGTTCT”<sup>[25]</sup>都是富含 GC 的序列,人类基因组中的小卫星也富含 GC 序列<sup>[31]</sup>。本研究中,从统计结果中我们可以看出,在 123 种重复单元组成的重复序列中,与几个重要的小卫星重复序列相比较,没有发现与它们相同或者相似的序列,并且三疣梭子蟹基因组小卫星序列总体上是富含 AT 的。这个结果与高焕等<sup>[22]</sup>在中国对虾和栾生等<sup>[10]</sup>在日本囊对虾中的研究结果是一致的。

简纪常等<sup>[23]</sup>对单个小卫星中的序列进一步归类说明时,根据其碱基组成的特点,将其分成富含 A、T 和其他等三类。而对于基于基因组调查性质的小卫星序列的描述上,这些归类明显不能全面描述基因组小卫星的特征。我们根据高焕等<sup>[22]</sup>对重复序列中碱基组成的种类及其相应数量的多少从大到小的排序方法进行归类。相对来讲,此归类有以下几点好处:1)可以初步看出重复序列中碱基的组成种类;2)可以对整体小卫星序列中的各碱基含量进行估计,如从表 2 中我们可以估计出三疣梭子蟹基因组中的小卫星重复序列整体上是富含 A/T 的序列;3)由此而分出的类别较为简单明了。

**3.3 小卫星的起源和产生机制及 DNA 指纹图谱的开发** 小卫星的产生机理涉及到减数分裂过程中等位基因、非等位基因间的不等交换,有丝分裂中姐妹染色体间的不等交换,以及引起单一重复单位增加或缺失的复制时期 DNA 滑动等多种因素,是目前研究的一个热点。由于小卫星序列中重复单位数目变异能够产生高度的多态性,因此常被用作探针应用在 DNA 指纹图谱分析中。一个探针能检测到 10~20 个高度多态位点,多态信息含量高(0.7~0.9)。高焕等<sup>[22]</sup>通过对中国明对虾的研究认为虽然还没有文献进一步明确的研究报道,但可以初步认为一些小卫星序列明显是在微卫星重复单元的基础上进一步发展而来的。这是否揭示了基因组的进化模式是微卫星-小卫星-简并小卫星序列-更复杂的序列-

有功能的序列,直至有功能的新基因,还需要进一步深入的研究。

从来自牛体内的一种寄生虫 *Theileria parva* 基因组筛选的两个小卫星探针,其序列为富含 AT<sup>[32]</sup>。然而,在哺乳动物、鸟类和植物基因组上,经典的重要小卫星重复序列,如人源小卫星 33.15 的核心序列 AGAGGTGTGGCAGGTGG<sup>[28]</sup>、大肠杆菌重组  $\chi$  信号顺序(GCTGGTGG)<sup>[29]</sup>、染色体端粒重复序列 TTTAGGG<sup>[30]</sup>、噬菌体 M13<sup>[25]</sup>中的一个常被用来做 DNA 指纹图谱鉴定的核心序列“GAGGGTGGCGGTTCT”等,它们均为富含 GC 的序列,很少富含 AT 序列的探针。人类基因组中的小卫星也富含 GC 序列<sup>[31]</sup>。因此,对三疣梭子蟹小卫星标记的研究迫在眉睫。

目前,水产动物小卫星探针的开发,如简纪常等<sup>[32]</sup>在鳙、王进科等<sup>[33-34]</sup>在中华绒螯蟹和白东清等<sup>[35]</sup>在花鲢方面的研究,主要是借助于 M13、人源 33.6 和 33.15 等几个探针来进行小卫星探针的筛选,这也是当前各个物种进行小卫星指纹图谱的主要技术路线。这些方法在天鹅、狗、猫和绵羊上得到了成功应用。但是考虑到三疣梭子蟹基因组小卫星为富含 AT 序列,应用现有的几个经典的小卫星探针进行三疣梭子蟹小卫星探针的开发可能并不合适。更多的工作,应该基于筛选的现有小卫星序列,进行小卫星探针的开发和指纹图谱的构建工作。这些也是我们下一步工作的重点。

中国水产科学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] Ramel C. Mini-and microsatellites [J]. Environ Health Perspect, 1997, 4(Suppl): 781-789.
- [2] Ingavale S, Kaur R, Aggarwal P, et al. A minisatellite sequence within the propeptide region of the vacuolar carboxypeptidase Y gene of *Schizosaccharomyces pombe* [J]. Bacteriol, 1998, 180(14): 3727-3729.
- [3] Jauert P, Edmiston S, Conway K, et al. RAD1 controls the meiotic expansion of the human HRAS1 minisatellite in *Saccharomyces cerevisiae* [J]. Molecular and Cellular Biology, 2002, 22(3):

- 953 - 964.
- [4] Nakamura Y, Leppert M, Connell P, *et al.* Variable number of tandem repeat (VNTR) markers for human genemapping[J]. *Science*,1987, 235(4796):1616 - 1622.
- [5] Klevytska A, Price L, Schupp J, *et al.* Identification and characterization of variable number tandem repeats in the *Yersinia pestis* genome[J]. *Journal of Clinical Microbiology*, 2001, 39(9):3179 - 3185.
- [6] van Belkum A, Scherer S, van Leeuwen W, *et al.* Variable number of tandem repeats in clinical strains of *Haemophilus influenzae* [J]. *Infect Immun*, 1997,65(12):5017 - 5027.
- [7] van Belkum A, Scherer S, Alphen L, *et al.* Short sequence DNA repeats in prokaryotic genomes[J]. *Microbiol Mol Biol Rev*,1998,62(2):275 - 293.
- [8] Vergnaud G, Denoeud F. Minisatellites; mutability and genome architecture [J]. *Genome Research*, 2000,10(7):899 - 907.
- [9] 高 焕,刘 萍,孟宪红,等.中国对虾(*Farmeropenaeus chinensis*)基因组微卫星特征分析[J]. *海洋与湖沼* [J]. 2004,35(5):424 - 431.
- [10] 栾 生,孔 杰,王清印,等.日本囊对虾基因组小卫星的特征分析 [J]. *水产学报*, 2007, 31 (2): 137 - 144.
- [11] 陈 徽. 牙鲆微卫星分子标记的筛选及多态性检测 [D]. 中国海洋大学: 硕士学位论文,2005.
- [12] 李红蕾, 宋林生, 王玲玲, 等. 栉孔扇贝 EST 中微卫星标记的筛选 [J]. *高技术通讯*, 2003, 12: 72 - 75.
- [13] 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*, 2003, (2): 30 - 33.
- [14] 戴爱云,杨思谅,宋玉枝,等.中国海洋蟹类[M].北京:海洋出版社,1986:213 - 214.
- [15] Pamela C, Paul Bentzen. Isolation and inheritance of microsatellite loci in the Dungeness crab (Brachyura: Cancridae: *Cancer magister*) [J]. *Genome*,2004,47:325 - 331.
- [16] Masatsugu Takano, Anna Barinova, Takuma Sugaya, *et al.* Isolation and characterization of microsatellite DNA markers from mangrove crab, *Scylla paramamosain* [J]. *Molecular Ecology Notes*, 2005,5:794 - 795.
- [17] David G, Jane M, Jing M. Identification of polymorphic microsatellite loci in the mud crab *Scylla serrata* (Brachyura: Portunidae) [J]. *Molecular Ecology Notes*,2002,2: 481 - 483.
- [18] Yap, Sezmis E, Chaplin J. Isolation and characterization of microsatellite loci in *Portunus pelagicus* (Crustacea: Portunidae) [J]. *Molecular Ecology Notes*,2002,2(1):30 32.
- [19] Puebla O, Parent E, Svigny J. New microsatellite markers for the snow crab *Chionoecetes opilio* (Brachyura: Majidae) [J]. *Molecular Ecology Notes*,2003,3(4):644 646.
- [20] An H,Jeong J, Park J. New microsatellite markers for the snow crab *Chionoecetes opilio* (Brachyura: Majidae) [J]. *Molecular Ecology Notes*, 2006, 7 (1):86 88.
- [21] H nfling B, Weetman D. Characterization of microsatellite loci for the Chinese mitten crab, *Eriocheir sinensis* [J]. *Molecular Ecology Notes*, 2003,3(1),15 - 17.
- [22] 高 焕,孔 杰.中国明对虾基因组小卫星重复序列分析[J]. *动物学报*,2005,51(1):101 - 107.
- [23] 简纪常,夏德全. 鱸小卫星 pBC174 的序列结构特性分析[J]. *中国水产科学*,2002,9(2):186 - 189.
- [24] Hancock J. Genome size and the accumulation of simple sequence repeats; implications of new data from genome sequencing projects [J]. *Genetica*, 2002,115(1):93 - 103.
- [25] Meyer W, Maszewska K, Sorrell T. PCR fingerprinting: a convenient molecular tool to distinguish between *Candida dubliniensis* and *Candida albicans* [J]. *Med Mycol*, 2001, 39 (2): 185 - 193.
- [26] Bastien D, Favre J, Collignon A, *et al.* Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [ *Picea abies* (L.) Karst][J]. *Theor. Appl. Genet*, 2003, 107(3):574 - 580.
- [27] Klevytska A, Price L, Schupp J, *et al.* Identification and characterization of variable number tandem repeats in the *Yersinia pestis* genome[J]. *Journal of Clinical Microbiology*, 2001, 39 (9):3179 - 3185.
- [28] Jeffreys A, Wilson V, Thein S. Hypervariable minisatellite regions in human DNA[J]. *Nature*, 1985,314(7):67 - 73.
- [29] Smith G, Kunes S, Schultz D, *et al.* Structure of chi hotspots of generalized recombination[J]. *Cell*, 1981,24(2):429 - 436.



- [30] Vitturi R, Colomba, Gianguzza P, *et al.* Chromosomal location of ribosomal DNA (rDNA), (GATA)<sub>n</sub> and (TTAGGG)<sub>n</sub> telomeric repeats in the neogastropod *Fasciolaria lignaria* (Mollusca: Prosobranchia) [J]. *Genetica*, 2000, 108(3): 253-257.
- [31] Bois P. Hypermutable minisatellites, a human affair [J]. *Genomics*, 2003, 81(4): 349-355.
- [32] Bishop R, Morzaria S, Gobright E. Linkage of two distinct AT-rich minisatellites at multiple loci in the genome of *Theileria parva* [J]. *Gene*, 1998, 216(2): 245-254.
- [32] 简纪常, 夏德全. 鳙小卫星 DNA 的克隆[J]. *中国水产科学*, 1999, 6(4): 18-20.
- [33] 王进科, 周刚, 曹文明, 等. 用小卫星探针 33.6 对中华绒螯蟹遗传多态性的 DNA 指纹图谱研究 [J]. *大连水产学院学报*, 2001, 16(2): 92-98.
- [34] 王进科, 周刚, 曹文明, 等. 中华绒螯蟹 DNA 指纹图谱的初步研究 [J]. *水产养殖*, 2000, 5: 24-27.
- [35] 白东清, 乔秀亭, 龙良启, 等. 花鲢的 DNA 指纹分析 [J]. *天津农学院学报*, 2001, 8(4): 11-14.

## Analysis of minisatellites in crab *Portunus trituberculatus* genome

SONG Lai-peng<sup>1,2</sup>, LIU Ping<sup>1</sup>, LI Jian<sup>1</sup>, LIU Zhen-hui<sup>2</sup>

(1. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071;

2. Marine Life College, Ocean University of China, Qingdao 266003)

**Abstract:** In this paper, genome library of crab *Portunus trituberculatus* was constructed. By sequencing randomly, 4 164 clones of sequences in the genomic library of crab *Portunus trituberculatus* were obtained. Then, using software DNASTAR (Version 5.0) to assemble all of the clones, 709 clones independent of each other, were made in which the length of DNA sequences is about 622 409 bp in total. With the help of the bio-soft Tandem Repeats Finder (Version 2.02), 130 minisatellites were screened in the crab's genome DNA sequences. Their cumulative length occupied 2.55% of the total length of DNA sequences. In the minisatellite sequences, twelve-nucleotide repeats were the most frequent type, accounting for 10.77% of the total number of minisatellites. It showed that the number of sequences decreased with the length of its repeat unit ( $R = -0.663, P < 0.01$ ). Eight-nucleotide repeat had the largest range of copy number of repeat unit (3.9-66.5), the following were thirteen-nucleotide repeat type (2.0-40.6) and twenty-six-nucleotide (2.3-21.0), respectively. Descending three repeat types in mean copy number of repeat unit were eight-nucleotide repeat (19.96), twenty-five-nucleotide (16.00) and twenty-two-nucleotide (15.85), respectively. The range of copy number of repeat unit varied from 2 to 66.5, and the copy number mostly ranged from 2 to 25. Moreover, it was showed that the number of corresponding minisatellites decreased as copy number of repeat unit increased. In this research, the 130 minisatellite sequences are composed of 123 kinds of repeat units, so it is very difficult to classify the minisatellite sequences. The minisatellite sequences can be classified into three types: dinucleotide minisatellite sequences composed of two kinds of nucleotide, trinucleotide minisatellite sequences composed of three kinds of nucleotide and tetranucleotide minisatellite sequences composed of four kinds of nucleotide. Further, all of above sequences can be divided into many sub-types according to the composition of nucleotide and their number from large to small. Totally, the minisatellite sequences in *Portunus trituberculatus* are A/T rich. This paper also discussed the genesis and evolution of minisatellite repeat sequences, and it is believed that the minisatellite repeats may come from the microsatellite repeats. It could be very practical to use minisatellite to study the genome of *Portunus trituberculatus* and would be applied to a variety of fields.

**Key words:** *Portunus trituberculatus*; genome; repeat sequences; minisatellites