

文章编号:1000-0615(2007)02-0137-08

## 日本囊对虾基因组小卫星的特征分析

栾生, 孔杰, 王清印, 高焕, 王伟继, 张庆文

(中国水产科学研究院黄海水产研究所农业部海洋渔业资源可持续利用重点开放实验室, 山东 青岛 266071)

**摘要:**为了分析日本囊对虾基因组中小卫星序列的分布特征并开发小卫星标记,构建了日本囊对虾随机基因组文库,对其中片段长度为400~1 000 bp的3 395个克隆测序,共获得总长度为1 606 711 bp的DNA序列。在序列拼接后的2 815个克隆中,筛选到487个小卫星序列,其序列总长度占测序序列总长度的3.97%。小卫星序列中,以12 bp重复单位的序列数量最多(8.62%),总体趋势表现为重复单位越长,相应的重复序列数目越少( $R = -0.826, P < 0.01$ )。小卫星重复单位拷贝数分布范围以8 bp重复单位最广(3.1~47.6),其次是10 bp(2.5~44),再次是12 bp(2.2~28)。平均拷贝数最高的三种重复类型分别为8 bp重复单位(13.45),32 bp(9.32)和7 bp(9.28)。小卫星序列重复单位的拷贝数分布范围2~48,集中分布在2~30,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。小卫星序列中,AT含量大于50%的序列占66%小卫星AT含量与相应的序列数目不相关( $R = 0.063, P < 0.05$ )。小卫星侧翼序列分析表明,随着小卫星GC含量的升高,其两端侧翼序列GC含量表现为不断增大的趋势( $R = 0.731, P < 0.01$ ),小卫星序列的产生与其两端侧翼序列的碱基组成可能存在一定的关系。

**关键词:**日本囊对虾;串联重复序列;小卫星;重复单位;拷贝数

中图分类号:Q 785;S 917.4

文献标识码:A

### Analysis of minisatellites in the genome of *Marsupenaeus japonicus*

LUAN Sheng, KONG Jie, WANG Qin-yin, GAO Huan, WANG Wei-ji, ZHANG Qing-wen

(Key Laboratory for Sustainable Utilization of Marine Fisheries Resources Certificated by the Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China)

**Abstract:** In this paper, random genome library of Kuruma prawn *Marsupenaeus japonicus* was constructed and the lengths of 3 395 sequenced clones were from 400 bp to 1 000 bp. After the 2 815 sequences being spliced, 487 minisatellites were screened in the 1 606 711 bp Kuruma prawn genome sequences. Their cumulative length occupied 3.79% of the total length of DNA sequences. In the minisatellite sequences, twelve-nucleotide repeats were the most frequent type, accounting for 8.62% of the total number of minisatellites. It showed that the number of sequences decreased with the increase of the length of its repeat unit ( $R = -0.826, P < 0.01$ ). Eight-nucleotide repeat had the largest range of copy number of repeat unit (3.1-47.6), the following were ten-nucleotide repeat type (2.5-44) and twelve-nucleotide (2.2-28), respectively. Descending three repeat types in mean copy number of repeat unit were eight-nucleotide repeat

收稿日期:2006-08-28

资助项目:国家“八六三”高新技术发展计划(2005AA603210,2003AA603032)

作者简介:栾生(1978-),男,山东安丘人,博士,主要从事遗传育种研究。E-mail:luan\_sheng@163.com

通讯作者:孔杰,E-mail:kongjie@sina.com

(13.45), thirty-two-nucleotide (9.32) and seven-nucleotide repeat (9.28), respectively. The range of copy number of repeat unit varied from 2 to 48, and the copy number mostly ranged from 2 to 30. Moreover, it was showed that the number of corresponding minisatellites decreased as copy number of repeat unit increased. Minisatellite was AT-rich sequence because the number of minisatellites whose AT content was more than 50% occupied 66% of the number of all minisatellites. The correlation between AT content and the corresponding sequence number was little and not significant ( $R = 0.063$ ,  $P > 0.05$ ) in minisatellites. At the same time, GC content of minisatellite increased with the increase of GC content in the corresponding flanking sequences, which might imply that base composition of the flanking sequence has an influence on the origin and evolution of minisatellites.

**Key words:** *Marsupenaeus japonicus*; minisatellite; repeat unit; copy number

小卫星(minisatellite)是指由7~100 bp左右的重复单位组成的重复序列,由于其串联重复单位的数目在不同个体基因组的不同位点上数目都不同,因此又称为可变数目串联重复序列(variable number tandem repeats, VNTRs)<sup>[1]</sup>。小卫星在染色体上分布于常染色质区,在原核生物和真核生物的基因组中都有大量分布<sup>[2-4]</sup>。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。

由于小卫星序列中重复单位数目变异能够产生高度多态性,因此常被用作探针应用于DNA指纹图谱分析。通常一个小卫星探针能检测到10~20个高度多态位点,多态信息含量高(0.7~0.9)。Jeffreys等<sup>[5]</sup>应用人源小卫星33.6、33.15作探针对人进行DNA指纹分析,发现其DNA指纹图谱具有高度变异性,多位点性及稳定的遗传性,因此在动植物遗传学研究上得到了广泛的应用。小卫星标记主要应用于个体识别与亲缘关系鉴定,种群内遗传变异与群体间遗传分化的研究以及连锁分析等几个方面。在水生动物的小卫星DNA指纹图谱研究,鱼类是其应用较为广泛的种类。鳙<sup>[6]</sup>、鲢<sup>[7]</sup>和奥利亚罗非鱼<sup>[8]</sup>的小卫星指纹图谱分析显示了种群特异性和个体特异性。水生动物的DNA指纹图谱分析中,由于自身小卫星探针的缺乏,其分析所用小卫星探针一般来自于人的33.6和33.15两个探针<sup>[6-8]</sup>。由于物种基因组以及小卫星组成的差异,借助于人源小卫星会降低小卫星探针的灵敏度。王进科等<sup>[8]</sup>通过人源33.6和33.15小卫星探针对奥利亚罗非鱼的DNA指纹图谱分析表明,平均检测出的条带数目为7.18,多态性较低。

目前,水生生物基因组研究工作的相对落后,仅中国对虾、牙鲆和扇贝等物种中开展了微卫星和小卫星序列分析的研究工作<sup>[9-13]</sup>。日本囊对虾(*Marsupenaeus japonicus*)在我国水产养殖中具有重要地位,是我国北方重要的养殖品种之一,目前关于其小卫星序列在基因组中的组成与分布特征的研究还未见报道。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,在目前仍没有开展日本囊对虾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进行分析,对初步了解日本囊对虾基因组小卫星序列的分布特征,以及进一步进行小卫星标记的开发等具有的意义。

## 1 材料和方法

### 1.1 材料来源

日本囊对虾取自广东汕头的野生群体,取样时间为2004年9月。2005年3月-2005年7月期间,对其部分基因组序列进行了测序。测序方法为:提取基因组DNA,采用超声波随机打断,经过低熔点琼脂糖凝胶电泳回收400~1500 bp左右的片段,与PUC19质粒连接,将重组DNA转化到大肠杆菌DH5 $\alpha$ 中,利用蓝白斑筛选阳性克隆,进而建立日本囊对虾部分基因组文库。对以上建立的部分基因组文库进行双向测序,共获得3395个DNA克隆序列,经过Seqman II软件进行序列拼装,最终获得2815个独立的克隆序列。每个克隆序列的长度从400~1000 bp不等,所有克隆序列的总长度为1606711 bp。

### 1.2 小卫星重复序列的判定标准及其分类方法

通过软件Tandem Repeats Finder (Version 3.21)对拼装后的序列进行分析,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标准为:重复序列中重复单位的长度在 7~80 bp,拷贝数目 2 个及 2 个以上。

为了衡量小卫星重复类型变异水平的高低,这里引入变异系数概念,变异系数的计算公式为:

$$C \cdot V(\%) = \frac{S}{\bar{x}} \times 100$$

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。

## 2 结果

### 2.1 小卫星重复序列在基因组中的分布特征

在长度为 1 606 711 bp 的日本囊对虾基因组序列中,共筛选到了 487 个小卫星序列,序列总长度为 60 581 bp,约占基因组串联重复序列总长度的 35.8%,约占测序序列总长度的 3.79%,即平

均每 1 000 bp 核苷酸序列中包含 37.9 bp 的小卫星序列。487 个小卫星重复序列中,总共筛选到 442 种重复类型,平均 1.1 个重复序列中就能够发现一种重复类型。不同长度重复单位类型的重复序列数目分布情况为(图 1):以 12 bp 重复单位的序列数目最多(42 个),占小卫星重复序列总数目的 8.62%,9 bp 重复单位次之(27 个),比例为 5.54%,总体趋势表现为随着重复单位长度的增加,其重复序列数目降低,且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(3 429 bp),20 bp 重复单位次之(2 688 bp),二者间也存在极显著负相关,但是相关系数较小( $R = -0.560, P < 0.01$ )。

### 2.2 小卫星重复单位拷贝数分布及变异能力

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。487 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 3.1~47.6;其次是 10 bp 重复,范围在 2.5~44;再次是 12 bp 重复,范围在 2.2~28。平均拷贝数最高的三种重复类型分别为 8 bp 重复(13.45),32 bp 重复(9.32)和 7 bp 重复(9.28)。

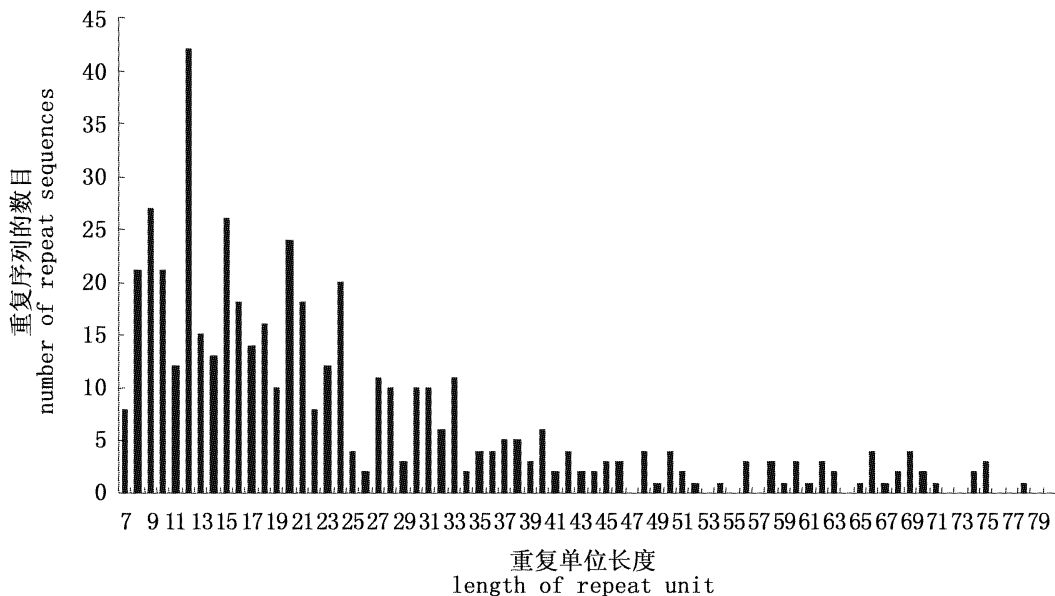


图 1 不同长度的重复单位组成的重复序列的数量分布

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(%) of total no. of mini-satellite	占串联重复序列总数百分比(%) of tandem repeats	序列长度 sequence length	占小卫星总长度百分比(%) of total length of mini satellite	占串联重复序列总长度百分比(%) of total length of tandem repeat	拷贝数范围 range of copy numbers	累积拷贝数 accumulative copy number	平均拷贝数 average of copy numbers	拷贝数变异系数(%) CV of copy number
7	8	8	1.64	0.4624	517	0.85	0.3055	4~22.1	74.20	9.28	66.08
8	17	21	4.31	1.2139	2254	3.72	1.3321	3.1~47.6	282.50	13.45	79.28
9	15	27	5.54	1.5607	1977	3.26	1.1684	2.9~22	221.30	8.20	55.38
10	19	21	4.31	1.2139	1825	3.01	1.0785	2.5~44	185.60	8.84	105.25
11	10	12	2.46	0.6936	614	1.01	0.3629	2.7~9.5	55.90	4.66	50.03
12	35	42	8.62	2.4277	3429	5.66	2.0265	2.2~28	289.60	6.90	88.31
13	14	15	3.08	0.8671	1046	1.73	0.6182	1.9~19.6	83.10	5.54	106.80
14	13	13	2.67	0.7514	1164	1.92	0.6879	2.1~16.9	84.70	6.52	81.34
15	23	26	5.34	1.5029	1301	2.15	0.7689	1.9~8.4	86.80	3.34	43.58
16	17	18	3.70	1.0405	1625	2.68	0.9604	1.9~23	102.40	5.69	112.62
17	14	14	2.87	0.8092	1201	1.98	0.7098	1.9~24.1	71.30	5.09	141.90
18	16	16	3.29	0.9249	1400	2.31	0.8274	2~11.2	77.60	4.85	51.68
19	10	10	2.05	0.5780	586	0.97	0.3463	1.9~7.7	30.90	3.09	55.50
20	24	24	4.93	1.3873	2688	4.44	1.5886	2~15.9	134.50	5.60	72.09
21	17	18	3.70	1.0405	1531	2.53	0.9048	1.9~14.9	74.70	4.15	91.77
22	8	8	1.64	0.4624	1192	1.97	0.7045	1.9~13.5	54.30	6.79	66.55
23	11	12	2.46	0.6936	1391	2.30	0.8221	1.9~16.4	61.20	5.10	84.25
24	19	20	4.11	1.1561	2553	4.21	1.5088	2.1~20.4	105.70	5.29	99.19
25	4	4	0.82	0.2312	586	0.97	0.3463	1.9~16.4	23.80	5.95	117.30
26	2	2	0.41	0.1156	155	0.26	0.0916	2.5~3.5	6.00	3.00	23.57
27	11	11	2.26	0.6358	2037	3.36	1.2038	2.1~21	75.00	6.82	96.95
28	10	10	2.05	0.5780	1747	2.88	1.0325	2~16.2	62.50	6.25	66.93
29	3	3	0.62	0.1734	181	0.30	0.1070	2.1~2.1	6.30	2.10	0.00
30	8	10	2.05	0.5780	1907	3.15	1.1270	2~19.4	65.30	6.53	104.10
31	10	10	2.05	0.5780	2248	3.71	1.3285	2~17	72.90	7.29	65.86
32	5	6	1.23	0.3468	1778	2.93	1.0508	3.4~17.6	55.90	9.32	68.64
33	10	11	2.26	0.6358	2457	4.06	1.4521	1.9~15.3	74.60	6.78	74.20
34	2	2	0.41	0.1156	213	0.35	0.1259	2.8~3.4	6.20	3.10	13.69
35	4	4	0.82	0.2312	860	1.42	0.5082	3.7~9.5	24.90	6.23	41.05
36	3	4	0.82	0.2312	408	0.67	0.2411	2.1~3.2	11.40	2.85	18.23
37	4	5	1.03	0.2890	645	1.06	0.3812	2~7.7	17.40	3.48	68.19
38	4	5	1.03	0.2890	1003	1.66	0.5928	2.7~11.6	26.70	5.34	68.65
39	3	3	0.62	0.1734	679	1.12	0.4013	2~13.2	17.40	5.80	110.51
40	5	6	1.23	0.3468	668	1.10	0.3948	1.9~4.5	16.70	2.78	45.07
41	2	2	0.41	0.1156	278	0.46	0.1643	2.2~4.5	6.70	3.35	48.55
42	4	4	0.82	0.2312	359	0.59	0.2122	2~2.5	8.70	2.18	10.86
43	2	2	0.41	0.1156	391	0.65	0.2311	2.9~6.1	9.00	4.50	50.28
44	2	2	0.41	0.1156	397	0.66	0.2346	4.1~4.9	9.00	4.50	12.57
45	3	3	0.62	0.1734	726	1.20	0.4291	2~11.3	16.30	5.43	93.96
46	3	3	0.62	0.1734	618	1.02	0.3652	2.2~9	13.60	4.53	85.36
48	4	4	0.82	0.2312	890	1.47	0.5260	2.1~6.6	18.60	4.65	46.67
49	1	1	0.21	0.0578	167	0.28	0.0987	3.4~3.4	3.40	3.40	
50	4	4	0.82	0.2312	552	0.91	0.3262	2~4.2	11.00	2.75	36.30
51	2	2	0.41	0.1156	222	0.37	0.1312	2.1~2.3	4.40	2.20	6.43
52	1	1	0.21	0.0578	132	0.22	0.0780	2.6~2.6	2.60	2.60	
54	1	1	0.21	0.0578	489	0.81	0.2890	9~9	9.00	9.00	
56	3	3	0.62	0.1734	987	1.63	0.5833	2~11.8	17.70	5.90	88.09

· 续表 1 ·

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(%) of total no. of mini- satellite	占串联 重复 序列总数 百分比(%) of no. of tandem repeats	序列长 度 sequence length	占小卫星 总长度 百分比(%) of total length of mini satellite	占串联 重复序列 总长度 百分比(%) of total length of tandem repeat	拷贝数 范围 range of copy numbers	累积 拷贝数 accumulative copy number	平均拷 贝数 average of copy numbers	拷贝数 变异 系数(%) CV of copy number
58	3	3	0.62	0.1734	781	1.29	0.4616	2~9.4	13.60	4.53	93.00
59	1	1	0.21	0.0578	118	0.19	0.0697	2~2	2.00	2.00	
60	3	3	0.62	0.1734	404	0.67	0.2388	1.9~2.8	6.70	2.23	22.09
61	1	1	0.21	0.0578	140	0.23	0.0827	2.3~2.3	2.30	2.30	
62	3	3	0.62	0.1734	1075	1.77	0.6353	3.8~7.4	17.30	5.77	31.61
63	2	2	0.41	0.1156	467	0.77	0.2760	2~5.4	7.40	3.70	64.98
65	1	1	0.21	0.0578	212	0.35	0.1253	3.2~3.2	3.20	3.20	
66	4	4	0.82	0.2312	917	1.51	0.5419	2~4.5	14.00	3.50	30.86
67	1	1	0.21	0.0578	234	0.39	0.1383	3.5~3.5	3.50	3.50	
68	2	2	0.41	0.1156	422	0.70	0.2494	2.4~3.9	6.30	3.15	33.67
69	4	4	0.82	0.2312	1757	2.90	1.0384	3.2~9.4	25.50	6.38	45.61
70	2	2	0.41	0.1156	340	0.56	0.2009	2.4~2.4	4.80	2.40	0.00
71	1	1	0.21	0.0578	166	0.27	0.0981	2.4~2.4	2.40	2.40	
74	1	2	0.41	0.1156	586	0.97	0.3463	4~4	8.00	4.00	0.00
75	2	3	0.62	0.1734	725	1.20	0.4285	1.9~5.8	9.70	3.23	68.76
78	1	1	0.21	0.0578	163	0.27	0.0963	2.1~2.1	2.10	2.10	
total	442	487			60581					2898.10	

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。小卫星不同重复单位长度类型重复单位拷贝数的变异能力最强的前 6 种类型分别是:17 bp (141.9), 25 bp (117.3), 16 bp (112.62), 39 bp (110.51), 13 bp (106.8), 10 bp (105.25) (表 1)。统计检验表明,随着小卫星序列重复单位长度的增加,其变异系数存在着降低的趋势

( $R = -0.446, P < 0.01$ )。小卫星序列中各重复单位的拷贝数分布范围 2~48,集中分布在 2~30。不同拷贝数目所对应的重复序列数量的分布情况为(图 2):拷贝数目为 3 的重复单位所组成的重复序列数目最多(135 个),其次是拷贝数目为 4 的重复序列(83 个),表现为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( $R = -0.806, P < 0.01$ )。

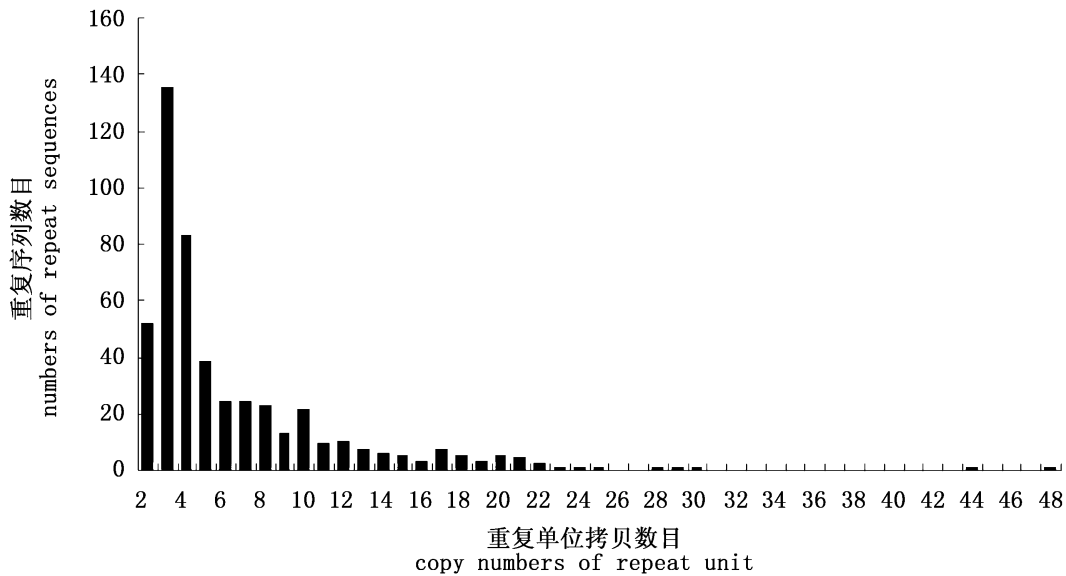


图 2 不同拷贝数目所对应的重复序列数目  
Fig. 2 The numbers of repeat sequences corresponding to the copy numbers of repeat units

### 2.3 小卫星重复单位的碱基组成特征

为了进一步分析小卫星序列 AT 含量的分布特征,对不同 AT 含量范围内的小卫星序列数量进行了统计分析(图 3)。以 AT 含量为 0.4~0.5 间的小卫星序列数目为最多,约占全部小卫星序列数目的 20%;AT 含量为 0.6~0.7 间的小卫星序列数目次之,所占比例为 19%。487 个小卫星序列中,AT 含量大于 0.5 的小卫星序列所占的比例为 66%,因此在一定程度上小卫星序列为富含 AT 序列。

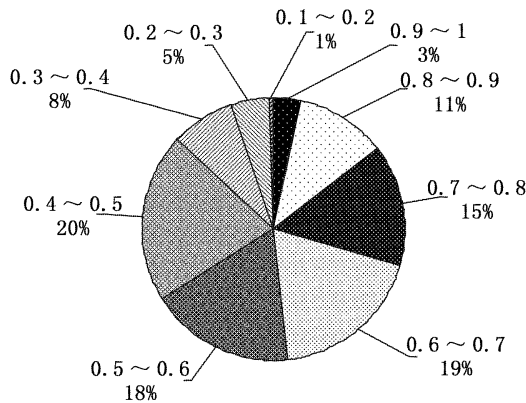


图 3 小卫星序列 AT 含量的分布特征  
Fig. 3 The distributing character of AT content of repeat sequence

### 2.4 小卫星侧翼序列分析

对小卫星序列两端的侧翼序列各取 100 个碱基,计算其 GC 含量,同时,计算小卫星序列 GC 含量。根据小卫星序列 GC 含量的不同,对小卫星进行分类,分析小卫星 GC 含量与其两端侧翼序列 GC 含量的关系(图 4)。结果表明,随着小卫星 GC 含量的升高,其两端侧翼序列 GC 含量表现为不断增大的趋势。GC 含量为 0.07 的小卫星序列其两端的侧翼序列 GC 含量平均值最低(0.155),GC 含量为 0.83 的小卫星序列其两端的侧翼序列 GC 含量平均值最大(0.7),且二者间的这种分布趋势表现为极显著的正相关关系( $R=0.731, P<0.01$ )。这在一定程度暗示着小卫星序列的产生与其两端侧翼序列的碱基组成间可能有着一定的关系。

## 3 讨论

### 3.1 小卫星序列的分布特征

在长度为 1 606 711 bp 的日本囊对虾基因组

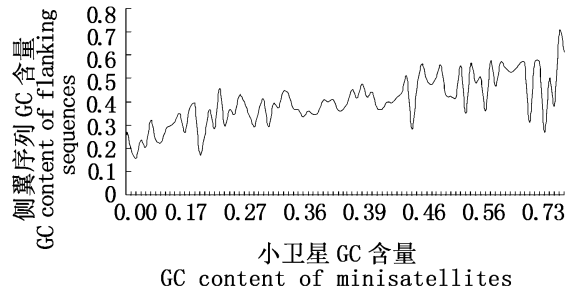


图 4 小卫星序列与侧翼序列 GC 含量的分布特征  
Fig. 4 The distribution character of GC content of minisatellite motifs and its the flanking regions

序列中,共筛选到了 487 个小卫星序列,序列总长度为 60 581 bp,约占基因组测序串联重复序列总长度的 35.8%,约占测序序列总长度的 3.79%,即平均每 1 000 bp 核苷酸序列中包含 37.9 bp 的小卫星序列。分析结果表明,日本囊对虾基因组中的小卫星序列的分布有两个显著的特点,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显著的负相关。高焕等<sup>[13]</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,小卫星序列中也以 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。从中国对虾和日本囊对虾在进化树上同属于对虾科这一角度考虑,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种属内的一致性。

小卫星的起源和产生机制,是当前研究的一个热点和难点,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。Prasad 等<sup>[14]</sup>和 Bachtrog 等<sup>[15]</sup>分别对家蚕和果蝇基因组进行了微卫星分布特征分析,结果表明,随着重复单位序列 GC 含量的增加,其侧翼序列 GC 含量也表现出升高的趋势,这说明序列本身的碱基组成对微卫星的产生和进化可能有一定的影响。如果序列 GC 含量高,那么产生富含 GC 重复的微卫星的机率要大于其他重复类型。借助于这一思路,分析了小卫星碱基组成与其侧翼序列碱基组

成的关系,结果表明,随着小卫星 GC 含量的升高,其两端侧翼序列 GC 含量表现为不断增大的趋势,并且二者之间表现为极显著的正相关关系。小卫星序列的碱基组成同样也影响了特定类型小卫星的产生。

### 3.2 小卫星重复序列的碱基组成特征及 DNA 指纹图谱的开发

日本囊对虾基因组小卫星 AT 含量分析表明,487 个小卫星序列中,AT 含量大于 0.5 的小卫星序列所占的比例为 66%,因此在一定程度上小卫星序列为富含 AT 序列。进一步分析小卫星序列富含 AT 的原因,很大程度上可能是因为日本囊对虾基因组序列 AT 含量的丰富造成的。基因组随机测序结果表明:日本囊对虾基因组富含 AT,其 AT 含量约为 60.63%,基因组较高的 AT 含量是小卫星序列富含 AT 的一个基础。中国对虾的小卫星序列 AT 含量分析结果表明其小卫星序列为富含 AT 序列<sup>[13]</sup>。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>[16]</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>[5]</sup>、大肠杆菌重组  $\chi$  信号顺序(GCTGGTGG)<sup>[17]</sup>、染色体端粒重复序列 TTTAGGG<sup>[18]</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>[19]</sup>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>[20]</sup>。这可能是小卫星在进化上存在着种属内的一致性的另外一个方面的体现。

当前水产动物小卫星探针的开发,如鳙、鲢和奥利亚罗非鱼等<sup>[6-8]</sup>,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,在天鹅、狗、猫和绵羊上得到了成功应用<sup>[21]</sup>。但是考虑到日本囊对虾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日本囊对虾小卫星探针的开发可能并不合适。更多的工作,应该基于现有小卫星序列,进行小卫星探针的开发和指纹图谱的构建。

### 参考文献:

- [1] Nakamura Y, Leppert M, O'Connell P, *et al.* Variable number of tandem repeat (VNTR) markers for human gene mapping [J]. *Science*, 1987, 235: 1616 - 1622.
- [2] Klevytska A M, Price L B, Schupp J M, *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.
- [3] van Belkum A, Scherer S, van Leeuwen W, *et al.* Variable number of tandem repeats in clinical strains of *Haemophilus influenzae* [J]. *Infection and Immunity*, 1997, 65(12): 5017 - 5027.
- [4] van Belkum A, Scherer S, van Alphen L, *et al.* Short sequence DNA repeats in prokaryotic genomes [J]. *Microbiol Mol Biol Rev*, 1998, 62(2): 275 - 293.
- [5] Jeffreys A J, Wilson V, Thein S L. Hypervariable minisatellite regions in human DNA [J]. *Nature*, 1985, 314: 67 - 73.
- [6] 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.
- [7] 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.
- [8] 王进科,夏德全,吴婷婷. 中国莫基群奥利亚罗非鱼遗传多态性的 DNA 指纹图谱[J]. *南京农业大学学报*, 2000, 23(3): 61 - 63.
- [9] 徐 鹏,周令华,田丽萍,等. 从中国对虾 ESTs 中筛选微卫星标记的研究[J]. *水产学报*, 2003, 27 (3): 213 - 218.
- [10] 陈 徽. 牙鲆微卫星分子标记的筛选及多态性检测[D]. 青岛:中国海洋大学硕士学位论文, 2005.
- [11] 孔 杰,高 焕. 中国明对虾基因组串联重复序列分析[J]. *科学通报*, 2005, 50(13): 1340 - 1347.
- [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] 高 焕. 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.
- [14] Prasad M D, Muthulakshmi M, Madhu M, *et al.* Survey and analysis of microsatellites in the silkworm, *Bombyx mori*: frequency, distribution,

- mutations, marker potential and their conservation in heterologous species [J]. *Genetics*, 2005, 169: 197 - 214.
- [15] Bachtrog D, Weiss S, Zangerl B, *et al.* Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome [J]. *Mol Biol Evol*, 1999, 16: 602 - 610.
- [16] 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.
- [17] Smith G R, Kunes S M, Schultz D W, *et al.* Structure of chi hotspots of generalized recombination [J]. *Cell*, 1981, 24 (2): 429 - 436.
- [18] 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.
- [19] Meyer W, Maszewska K, Sorrell T C. PCR fingerprinting: a convenient molecular tool to distinguish between *Candida dubliniensis* and *Candida albicans* [J]. *Med Mycol*, 2001, 39 (2): 185 - 193.
- [20] Bois P R. Hypermutable minisatellites, a human affair [J]. *Genomics*, 2003, 81 (4): 349 - 355.
- [21] 吴易雄, 晏光荣, 曹迎春. 小卫星 DNA 标记的研究及其在动物遗传育种中的应用 [J]. *广东畜牧兽医科技*, 2001, 26 (4): 18 - 20.