

DOI: 10.11758/yykxjz.20161209001

http://www.yykxjz.cn/

倪守胜, 杨钰, 柳淑芳, 庄志猛. 基于高通量测序的虾夷扇贝基因组微卫星特征分析. 渔业科学进展, 2018, 39(1): 107-113  
Ni SS, Yang Y, Liu SF, Zhuang ZM. Microsatellite analysis of *Patinopecten yessoensis* using next-generation sequencing method. Progress in Fishery Sciences, 2018, 39(1): 107-113

# 基于高通量测序的虾夷扇贝基因组 微卫星特征分析\*

倪守胜<sup>1,2,3</sup> 杨钰<sup>1,2,4</sup> 柳淑芳<sup>1,2①</sup> 庄志猛<sup>1</sup>

- (1. 农业部海洋渔业可持续发展重点实验室 中国水产科学研究院黄海水产研究所 青岛 266071;  
2. 青岛海洋科学与技术国家实验室 海洋渔业科学与食物产出过程功能实验室 青岛 266071;  
3. 上海海洋大学水产与生命学院 上海 201306; 4. 大连海洋大学水产与生命学院 大连 116023)

**摘要** 为全面了解虾夷扇贝(*Patinopecten yessoensis*)微卫星分布频率和数量, 深化对虾夷扇贝基因组的认识, 本研究运用第二代高通量测序技术, 进行虾夷扇贝简化基因组测序(RAD-seq), 从基因组水平阐明虾夷扇贝基因组微卫星特征。结果显示, 简化基因组测序共获得序列总长为 92,551,435 bp, 经过滤筛选, 获得 259,535 个 contig, 其中, 包含微卫星序列 3618 条, 经引物设计共获得 3460 对微卫星引物。统计微卫星序列的重复类型, 其中, 三核苷酸重复单元数量最多(1587 个, 45.87%), 其次是二核苷酸重复(1282 个, 37.05%), 六核苷酸重复单元数量最少(20 个, 0.58%)。在三核苷酸重复中, 以 ATA 重复类型所占比例最高(11.41%), 共计 181 个。此外, 虾夷扇贝同一重复类型的微卫星随着重复数增加其数量相对减少, 而相同重复数的微卫星随着重复单元长度的增加其数量也呈下降趋势, 可见微卫星长度与其数量呈负相关, 表明长度较短的微卫星变异速率较快。本研究结果为认识虾夷扇贝基因组特征和在基因组水平开展种群遗传学研究提供了基础数据。

**关键词** 虾夷扇贝; 二代测序; 基因组; 微卫星

**中图分类号** S917 **文献标识码** A **文章编号** 2095-9869(2018)01-0107-07

微卫星 DNA(Simple sequence repeat, SSR), 是由基本单元为 1~6 个核苷酸所构成的简单重复序列, 是目前应用较为广泛的分子标记之一(O'Connell *et al.*, 1997), 其稳定性和可重复性高, 特异性和共显性强, 现已应用于生物遗传多样性分析、全基因组相关分析、分子标记辅助育种等诸多方面(刘志毅等, 2001; Alam *et al.*, 2005; Cavagnaro *et al.*, 2010; Zhu *et al.*, 2012; 王日芳等, 2017)。海洋生物的微卫星筛选及多态性分析、利用微卫星构建连锁图谱和抗病相关标记

的连锁分析等工作已陆续开展(赵莹莹等, 2006a、b; 李春艳等, 2009; Hou *et al.*, 2011)。

虾夷扇贝(*Patinopecten yessoensis*)自 20 世纪 80 年代引入中国, 经过几十年的培育与发展, 现已成为我国重要的海洋养殖物种。目前, 已有采用分子标记开展虾夷扇贝种群遗传研究的报道, 如同工酶电泳技术(高悦勉等, 2003、2004)、AFLP(鲍相渤等, 2009; 丁君等, 2010)、SSR(Sato *et al.*, 2005; 赵莹莹等, 2006b; 常亚青等, 2007; 李春艳等, 2009; Hou *et al.*, 2011), 而

\* “十二五”农村领域国家科技计划课题(2012BAD18B03)和山东省泰山学者建设工程专项共同资助 [This work was supported by "12th Five-Year" Rural Areas of National Science and Technology Project(2012BAD18B03), and The Scholars of Taishan Special Construction Project in Shandong Province]. 倪守胜, E-mail: shoushengyongji@126.com

① 通讯作者: 柳淑芳, 研究员, E-mail: liusf@ysfri.ac.cn

收稿日期: 2016-12-09, 收修改稿日期: 2016-12-28

对虾夷扇贝基因组水平上的微卫星及其分布特征还知之甚少。本研究采用第二代高通量测序技术对虾夷扇贝进行简化基因组测序(Restriction-site associated DNA sequencing, RAD-seq),通过对基因组中不同类型微卫星的统计与分析,了解其微卫星分布频率和数量,从基因组水平阐明虾夷扇贝基因组 SSR 特征及组成,对进一步利用微卫星开展遗传标记辅助育种和种群遗传学研究具有重要的参考价值。

## 1 材料与方法

### 1.1 材料

虾夷扇贝于 2015 年 7 月采自我国北黄海獐子岛虾夷扇贝底播养殖海域,冰冻保存运送至实验室,解剖处理,取虾夷扇贝闭壳肌组织,存于 $-20^{\circ}\text{C}$ 备用。

### 1.2 基因组 DNA 提取

采用海洋动物组织基因组 DNA 提取试剂盒(离心柱型)(天根,北京)进行基因组 DNA 的提取,琼脂糖凝胶电泳初步检测提取的 DNA 质量,采用核酸蛋白分析仪检测 DNA 浓度与纯度,选取质量较好的 DNA 样品保存于 $-20^{\circ}\text{C}$ 备用。

### 1.3 文库构建及测序

将所获取的虾夷扇贝 DNA 样品送至北京诺禾致源生物信息科技有限公司进行测序。DNA 样品通过随机酶切打断,经末端修复、加 A 尾、加测序接头、纯化、PCR 扩增等步骤完成整个 RAD 文库的制备,构建好的文库通过 Illumina(测序仪)进行双末端测序,测序个数乘以测序序列长度即为原始数据值。对原始数据进行过滤:去除带接头(Adapter)的 reads pair;去除单端测序 read 中含有的 N(未读碱基)含量超过该条 read 长度比例的 10%的 reads pair;去除单端测序 read 中含有的低质量( $Q \leq 5$ )碱基数超过该条 read 长度比例的 50%的 reads pair。对所得数据样本含有的酶识别位点 reads 用 cd-hit-est 软件(Li *et al.*, 2006)进行聚类,并保证每类中 reads 支持数在 10~400 之间。根据聚类结果,用 Velvetopt 软件(Zerbino *et al.*, 2008)进行组装,得到最后的组装序列,并对 125 bp 以下的 contig 进行过滤。

### 1.4 分析方法

使用 SR search 软件(Rozen *et al.*, 2000)对过滤所得的 contig 进行简单重复序列的检测,检测标准如下:(1) SSR 重复的最小长度为 2;(2) SSR 重复的最大长度为 6;(3) SSR 序列的最小长度为 12;(4) SSR

上下游序列长度为 100 bp;(5) 2 个 SSR 的最小距离为 12 bp。使用 Primer 3 软件(Rozen *et al.*, 2000)对鉴定出的含有 SSR 重复基元的序列片段进行引物设计:引物的长度范围为 20~28,最适长度为 24;引物退火温度范围为 60~65 $^{\circ}\text{C}$ ,最适退火温度为 63 $^{\circ}\text{C}$ ,且一对引物退火温度差值不超过 1 $^{\circ}\text{C}$ 。

SSR 分布密度( $D$ )计算公式为:

$$D=N/L$$

式中, $D$ 为不同重复微卫星分布密度(ind./Mb); $L$ 为虾夷扇贝基因组重叠群(contig)总长(Mb); $N$ 为各重复类型微卫星数量(ind.)。

SSR 不同长度的分布频率( $F$ )计算公式为:

$$F=N/T$$

式中, $F$ 为不同长度微卫星的分布频率; $N$ 为各长度微卫星数量(ind.); $T$ 为微卫星总数量(ind.)。

## 2 结果与分析

### 2.1 简化基因组测序及拼接

测序结果共产生原始数据 2.872 G,经过滤后共获得有效数据 2.847 G(表 1)。Illumina HiSeq 产生的序列的碱基质量用 Qphred 表示,Phred 值为 20 和 30,碱基正确识别率分别为 99%和 99.9%。本测序中 Phred 数值大于 20 和 30 的碱基占所有碱基的百分比分别为 96.77%和 92.66%,RAD-Tag 捕获率为 93.07%,数据量符合预先设定,表明建库成功,测序质量合格。

表 1 虾夷扇贝简化基因组测序主要数据  
Tab.1 Mean data of reduced-representation genome sequencing in *P. yessoensis*

类目 Category	数值 Value
原始数据量 Raw base (bp)	2,871,603,900
有效数据量 Clean base (bp)	2,847,335,100
组装总长度 Total contig base (bp)	92,551,435
组装总个数 Total contig number	259,535
组装平均长度 Average contig length (bp)	356

通过聚类组装获得序列总长度为 92,551,435 bp,过滤掉小于 125 bp 的 contig 后获得 259,535 个 contig,其中, $N_{50}$ (将序列从大到小排列,当长度达到组装总长度一半时,contig 的长度)的长度为 419 bp,组装基因组的平均覆盖深度为 14.55 $\times$ ,其中,4 $\times$ 覆盖度在 86.64%以上,GC 含量为 35.63%,拼接结果正常,拼接序列质量较高。

### 2.2 基因组微卫星的数量及分布密度

通过对虾夷扇贝基因组的序列进行微卫星查找, 在虾夷扇贝的高通量测序数据库中进一步搜索含有 2~6 重复单元的 SSR, 从总长为 92,551,435 bp 的 259,535 个 contig 中筛选出 3618 个潜在微卫星位点, 经引物设计共获得 3460 个微卫星引物, 利用微卫星密度计算公式可得, 平均 26,748.97 bp 出现 1 个 SSR (表 2)。由表 2 可以看出, 虾夷扇贝微卫星各重复类型之间 SSR 数量均相差较大, 各类型 SSR 分布密度差异明显。所获得的 3460 条微卫星序列中, 分布最多的重复类型为三核苷酸重复(45.87%); 其次是二核苷酸重复(37.05%)、四核苷酸重复(14.48%)、五核苷酸重复(2.02%); 分布最少的是六核苷酸重复(0.58%)。

表 2 虾夷扇贝不同重复类型微卫星所占比例及分布密度  
Tab.2 The proportion and density of different repeat types in SSR database of *P. yessoensis*

重复类型 Repeat type	数量 Number	百分比 Percentage(%)	密度 Density (Mb)
二核苷酸 Dinucleotide	1282	37.05	13.93
三核苷酸 Trinucleotide	1587	45.87	17.25
四核苷酸 Tetranucleotide	501	14.48	5.45
五核苷酸 Pentanucleotide	70	2.02	0.76
六核苷酸 Hexanucleotide	20	0.58	0.22
总数 Total	3460	100	37.61

### 2.3 基因组微卫星不同重复类型中各重复单元数量与频率

统计微卫星各重复类型中不同重复单元的数量显示, 二核苷酸重复有 12 种重复单元, 三核苷酸重复 59 种, 四核苷酸重复 100 种, 五核苷酸重复 47 种, 六核苷酸重复 20 种。统计微卫星各重复类型中不同重复单元的百分比(表 3), 在二核苷酸重复微卫星中, AT/TA 重复单元所占比例最高(79.17%), 共计 1015 个; 其次是 TG/GT 重复单元 103 个(8.03%)。在三核苷酸重复的微卫星中, ATA 重复单元所占比例最高(11.41%), 共计 181 个; AAT、TAT、ATT、TAA 这 4 种重复单元紧随其后, 分别有 155 个(9.77%)、150 个(9.45%)、137 个(8.36%)和 120 个(7.56%)。四核苷酸重复中每种重复单元所占比例相对均匀, 其中, 数量最多的为 CAAA 重复单元, 共计 48 个, 所占比例为 9.58%。五核苷酸重复中, 数量最多的重复单元为 AAACC, 共计 11 个, 占比 15.71%。六核苷酸重复共计 20 个, 每一种重复单元仅有 1 个。

表 3 虾夷扇贝 SSR 重复单元的数量及百分比

Tab.3 The numbers and percentage of SSR repeat motifs in *P. yessoensis*

重复类型 Repeat type	重复单元 Motif	数量 Number	重复单元百分比 Percentage of each repeat motif (%)
二核苷酸 Dinucleotide	AT/TA	1,015	79.17
	TG/GT	103	8.03
	AC/CA	99	7.72
	CT/TC	31	2.42
	AG/GA	28	2.18
	CG/GC	6	0.47
三核苷酸 Trinucleotide	ATA	181	11.41
	AAT	155	9.77
	TAT	150	9.45
	ATT	137	8.63
	TAA	120	7.56
	TTA	71	4.47
	TTC	62	3.91
	Others	711	44.80
四核苷酸 Tetranucleotide	CAAA	48	9.58
	AAAT	29	5.79
	GACG	26	5.19
	ATAA	24	4.79
	AAAC	18	3.59
	ATCA	18	3.59
	Others	338	67.47
五核苷酸 Pentanucleotide	AAACC	11	15.71
	Others	59	84.29

### 2.4 基因组微卫星变异规律分析

虾夷扇贝基因组微卫星长度范围为 12~36 bp, 平均长度为 14.96 bp。微卫星长度分布频率计算结果显示, 微卫星长度越长其分布频率越低, 其中, 12~16 bp 的微卫星占绝对优势, 比例高达 79.35% (图 1)。

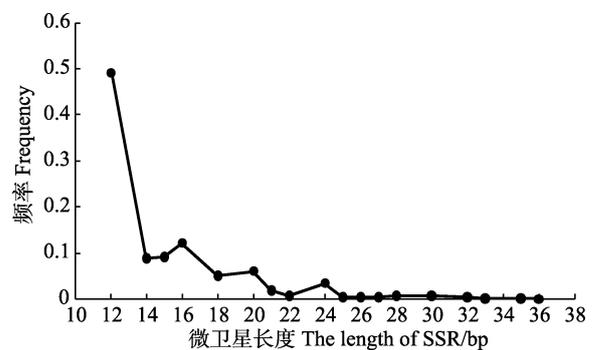


图 1 虾夷扇贝微卫星长度分布及不同长度微卫星频率  
Fig.1 The length distribution and frequency of *P. yessoensis* SSR

针对虾夷扇贝基因组微卫星数量进行分析发现, 当同一重复类型的微卫星随着重复数增加其数量相对减少(表 4 纵向), 而相同重复数的微卫星随着重复

单元长度的增加其数量也呈下降趋势(表 4 横向),且随着不同重复单元长度的增加其重复数范围逐渐减小(表 4)。如,二核苷酸重复单元按照重复数多少可分成 11 种,其中,重复数为 6 的二核苷酸重复单元数量有 695,随着重复数增加其数量逐渐减少,当重复数增加至 16 时该重复单元的数量降至 7;当重复单元长度增加至六核苷酸时,该重复类型数量仅有 20 种。可见,随着重复单元长度的增加,微卫星数量和重复数范围均呈下降趋势,推测虾夷扇贝基因组中长度较短的微卫星变异速率比长度较长的微卫星变异速率快。

表 4 虾夷扇贝 SSR 不同重复数及不同重复类型的数量  
Tab.4 The number of SSR with different repeat types and different copy number in *P. yessoensis*

重复数 Copy number	二核苷酸 Di- nucleotide	三核苷酸 Tri- nucleotide	四核苷酸 Tetra- nucleotide	五核苷酸 Penta- nucleotide	六核苷酸 Hexa- nucleotide
4		1006	318	53	19
5		315	117	14	1
6	695	123	44	1	
7	308	64	12	2	
8	103	43	9		
9	53	13	1		
10	40	17			
11	27	6			
12	14				
13	14				
14	13				
15	8				
16	7				

### 3 讨论

#### 3.1 虾夷扇贝基因组微卫星序列特征分析

本研究基于虾夷扇贝基因组获得有效 SSR 数目为 3460 个,其中出现频率最高的为三核苷酸重复(45.87%),主要以 ATA(11.41%)和 AAT(9.77%)重复单元为主,然后依次是 TAT、ATT、TAA 重复单元;而二核苷酸重复中,则是以 AT 最多,其次为 TA。Hou 等(2011)采用 454 测序技术对虾夷扇贝转录组进行测序,在其转录组中搜索到的 2748 个 SSR 中,三核苷酸重复的频率最高(39.4%),其次是二核苷酸重复(21.1%)、四核苷酸重复(15.5%)、五核苷酸重复(14.6%)和六核苷酸重复(9.4%)。可见,无论是在转录组还是基因组水平上,虾夷扇贝微卫星三核苷酸重复类型的出现频率均最高。

通过二核苷酸、三核苷酸和四核苷酸 3 种重复类型中各重复单元对比研究,在各种长度的重复中,A/T 都较 G/C 含量高很多,这与多数生物基因组微卫星具有的特点相一致,如中国对虾(*Fenneropenaeus chinensis*)微卫星(高焕等,2004),红鳍东方鲀(*Sargassum fusiforme*)微卫星(崔建洲等,2006),三疣梭子蟹(*Portunus trituberculatus*)微卫星(宋来鹏等,2008)。生物基因组微卫星富含 A/T 的可能原因是微卫星序列 A/T 含量高,则  $T_m$  值降低, DNA 链容易解开,通过 DNA 复制滑动机制和重组机制,产生富含 AT 重复类型的机率更高。此外, Schlotterer 等(1992)认为,基因组 DNA 因 CpG 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,使其具有较高的突变可能性,并且少量的 GC 又是维持 DNA 热力学稳定性所必需的,一定程度上可解释生物体基因组微卫星中 G/C 含量如此之少的原因(黄杰等,2015)。还有一种可能是,统计结果中 GC 重复过少也可能与 GC 重复的测序工作比较困难有关(高焕等,2004)。具体原因,现今仍无定论,尚有待进一步的研究探讨。

对虾夷扇贝基因组不同重复类型微卫星的长度变异情况进行分析,发现这些微卫星的数量与重复长度成反比,微卫星的数量随着重复长度的增加而降低。Samadi 等(1998)的模拟分析研究认为,重复长度越长,经受的选择压力越大,则它们的重复数就越少,数量越低。此外,在中国对虾(高焕等,2004)、红鳍东方鲀(崔建洲等,2006)、枸杞(*Lycium barbarum*) (党少飞等,2016)等不同种类的动植物基因组微卫星的研究中也得到了相同的结论。这意味着虾夷扇贝基因组中重复长度较短的微卫星比重复长度较长的微卫星变异速率快,为虾夷扇贝进一步的多态性微卫星标记筛选工作提供选择方向和依据。

#### 3.2 SSR 开发与筛选方法对 SSR 特征分析的影响

目前,比较成熟的微卫星开发与筛选技术有磁珠富集法、构建 cDNA 文库筛选 EST-SSR 法以及基于二代测序的转录组测序和简化基因组测序法。不同方法获得微卫星标记数量不同,所体现出的分布特征也不尽相同。传统的磁珠富集法筛选微卫星标记,获得率较低,每次仅能获得百余个微卫星标记,因其筛选探针多用二核苷酸重复的微卫星序列,在富集过程中,磁珠经过多个洗涤过程将一些因重复数目较少而附着不牢固的微卫星序列洗去,仅能获得较长序列的微卫星标记,因此,不能完全反映其微卫星的分布情况(赵莹莹等,2006a;李春艳等,2009)。李云峰等(2010)通过构建虾夷扇贝 cDNA 文库获得 74 个

EST-SSR, 由于 EST 序列仅能反映转录加工后的 mRNA 信息, 除去了许多内含子和其他序列, 因此, 并不能反映基因组 DNA 的真实信息。随着测序技术的发展, 获取效率高、技术难度低且花费成本与传统方法相近的二代测序被广泛应用到微卫星引物开发中, 基于转录组或基因组进行全部或部分高通量测序的方法可以获得大量的核酸序列信息, 微卫星标记的开发也就水到渠成, 而且二代测序一次性开发标记数据量大, 可以比较完整地反应物种的微卫星分布特征。

### 3.3 与其他贝类微卫星分布特征比较

由虾夷扇贝基因组微卫星特征分析得知, 其三碱基重复类型微卫星数量最多, 在已开展基因组微卫星开发的贝类中(表 5), 栉孔扇贝(*Chlamys farreri*) (Zhang *et al.*, 2008)和三角帆蚌(*Hyriopsis cumingii*) (Bai *et al.*,

2009)微卫星数量最多的同样为三碱基重复类型, 而在马氏珠母贝(*Pinctada martensii*) (王忠良等, 2015)、里氏拟石磺(*Paraoncidium reevesii*) (吴欣等, 2015)、菲律宾蛤仔(*Ruditapes philippinarum*) (闫路路等, 2015)和文蛤(*Meretrix meretrix*) (Wang *et al.*, 2011)中, 却呈现出单核苷酸重复数量最多, 二核苷酸、三核苷酸和四核苷酸重复数量依次递减的趋势。而长牡蛎(*Crassostrea gigas*) (张琳琳等, 2011)的 EST-SSR 则呈现六核苷酸重复数目最多(851), 单核苷酸次之(805), 二核苷酸、三核苷酸和四核苷酸重复依次递减的趋势。由此可见, 在已有贝类微卫星研究中, 各重复类型的微卫星分布并未呈现较明显的规律, 产生这样的结果, 一方面可能是微卫星开发手段不同而导致结果不同, 另一方面, 也可能是不同物种其微卫星组成存在较大差异, 这些都有待进一步研究。

表 5 不同贝类 SSR 分布特征

Tab.5 The characteristics of SSR distribution in different molluscs

物种 Species	SSR 类型 SSR type	平台/方法 Method/Sequencing platform	个数 Numbers	主要特征 Main features	参考文献 References
马氏珠母贝 <i>P.martensii</i>	EST-SSR	Illumin/ Hiseq-2000 平台	9872	单核苷酸重复最多(8042), 二核苷酸重复(1035)、三核苷酸重复(630)、四核苷酸重复(158)依次递减	王忠良等, 2015
文蛤 <i>M.meritrix</i>	EST-SSR	454 平台	2970	未考虑单核苷酸重复, 二核苷酸重复最多(1194), 三核苷酸重复(1123)、四核苷酸重复(357)次之	Wang 等, 2011
三角帆蚌 <i>H.cumingii</i>	EST-SSR	构建 cDNA 文库, 3730 测序	201	未考虑单核苷酸重复, 三核苷酸重复最多(110), 二核苷酸重复次之(83)	Bai 等, 2009
长牡蛎 <i>C.gigas</i>	EST-SSR	NCBI 下载 EST 序列标签	2602	六核苷酸重复最多(851), 单核苷酸重复次之(307), 二核苷酸重复(258)和三核苷酸重复(240)个数相差不多	张琳琳等, 2011
里氏拟石磺 <i>P.reevesii</i>	基因组 SSR	Illumin/ Hiseq-2000 平台	67696	单核苷酸重复最多(53948), 二核苷酸重复(7577)、三核苷酸重复(4579)、四核苷酸重复(1572)依次递减	吴欣等, 2015
菲律宾蛤仔 <i>R.philippinarum</i>	基因组 SSR	454 平台	1601	三核苷酸重复最多(903), 其次为二核苷酸重复(492)和四核苷酸重复(160), 单核苷酸重复较少(22)	闫路路等, 2015

## 参 考 文 献

- Alam MS, Islam MS. Population genetic structure of *Catla catla* (Hamilton) revealed by microsatellite DNA markers. *Aquaculture*, 2005, 246(1-4): 151-160
- Bai Z, Yin Y, Hu S, *et al.* Identification of genes involved in immune response, microsatellite, and SNP markers from expressed sequence tags generated from hemocytes of freshwater pearl mussel (*Hyriopsis cumingii*). *Marine Biotechnology*, 2009, 11(4): 520-530
- Bao XB, Dong Y, He CB, *et al.* Studies of Japanese scallop (*Mizuhopecten yessoensis*) germplasm by using AFLP markers. *Biotechnology Bulletin*, 2009(4): 126-129 [鲍相渤, 董颖, 赫崇波, 等. 基于 AFLP 技术对中国虾夷扇贝群体种质资源的研究. *生物技术通报*, 2009(4): 126-129]
- Cavagnaro PF, Senalik DA, Yang L, *et al.* Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics*, 2010, 11(1): 569
- Chang YQ, Chen XX, Ding J, *et al.* Genetic diversity in five scallop populations of the Japanese scallop (*Patinopecten yessoensis*). *Acta Ecologica Sinica*, 2007, 27(3): 1145-1152 [常亚青, 陈晓霞, 丁君, 等. 虾夷扇贝 (*Patinopecten yessoensis*) 5 个群体的遗传多样性. *生态学报*, 2007, 27(3): 1145-1152]

- Cui JZ, Shen XY, Yang GP, *et al.* Extraction, purification and chemical composition studies on fucodian from *Sargassum fusiforme*. Periodical of Ocean University of China (Natural Science Edition), 2006, 36(2): 249–254 [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. 中国海洋大学学报(自然科学版), 2006, 36(2): 249–254]
- Dang SF, Wang ZL, Zhang DF, *et al.* Characteristic analysis of microsatellites of *Lycium barbarum*. Journal of Northwest Forestry University, 2016, 31(1): 97–102 [党少飞, 王占林, 张得芳, 等. 枸杞基因组微卫星特征分析. 西北林学院学报, 2016, 31(1): 97–102]
- Ding J, Wang T, Chang YQ, *et al.* AFLP analysis on genetic diversity and differentiation of different populations of *Patinopecten yessoensis*. Journal of Yantai University (Natural Science and Engineering Edition), 2010, 23(1): 47–53 [丁君, 王婷, 常亚青, 等. 基于 AFLP 技术的不同群体虾夷扇贝遗传结构及多样性研究. 烟台大学学报(自然科学与工程版), 2010, 23(1): 47–53]
- Gao H, Liu P, Meng XH, *et al.* Analysis of microsatellite sequences in Chinese shrimp *Fenneropenaeus chinensis* genome. Oceanologia et Limnologia Sinica, 2004, 35(5): 424–431 [高焕, 刘萍, 孟宪红, 等. 中国对虾(*Fenneropenaeus chinensis*)基因组微卫星特征分析. 海洋与湖沼, 2004, 35(5): 424–431]
- Gao YM, Li GX, Zhao YL. Genetic structure of Japanese scallop *Patinopecten yessoensis* hatched population in Dalian Sea areas. Journal of Dalian Fisheries University, 2004, 19(2): 142–145 [高悦勉, 李国喜, 赵银丽. 大连沿海虾夷扇贝养殖群体遗传结构的研究. 大连水产学院学报, 2004, 19(2): 142–145]
- Gao YM, Li GX. Biochemical genetic analysis of isozymes from *Patinopecten yessoensis*. Journal of Dalian Fisheries University, 2003, 18(4): 269–272 [高悦勉, 李国喜. 虾夷扇贝同工酶的生化遗传分析. 大连水产学院学报, 2003, 18(4): 269–272]
- Hou R, Bao ZM, Wang S, *et al.* Transcriptome sequencing and de novo analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. PLoS One, 2011, 6(6): e21560
- Huang J, Zhou Y, Liu YZ, *et al.* Characteristics of microsatellites in *Arborophila rufipectus* genome sequences using 454 GS FLX. Sichuan Journal of Zoology, 2015, 34(1): 8–14 [黄杰, 周瑜, 刘与之, 等. 基于 454 GS FLX 高通量测序的四川山鹧鸪基因组微卫星特征分析. 四川动物, 2015, 34(1): 8–14]
- Li CY, Ding J, Chang YQ, *et al.* Isolation of microsatellite markers in Japanese scallop and its application in genetic structure analysis. Journal of Fishery Sciences of China, 2009, 16(1): 39–46 [李春艳, 丁君, 常亚青, 等. 虾夷扇贝微卫星标记的分离及其养殖群体的遗传结构分析. 中国水产科学, 2009, 16(1): 39–46]
- Li WZ, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 2006, 22(13): 1658–1659
- Li YF, Liu WD, Gao XG, *et al.* Construction of cDNA libraries from mantle and kidney of Japanese scallop (*Mizuhopecten yessoensis*) and ESTs analysis. Journal of Fishery Sciences of China, 2010, 17(3): 578–585 [李云峰, 刘卫东, 高祥刚, 等. 虾夷扇贝外套膜和肾脏组织 cDNA 文库构建以及 EST 的初步分析. 中国水产科学, 2010, 17(3): 578–585]
- Liu ZY, Xiang JH. Application of the microsatellite technique for genetic analysis in marine animals. 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. 海洋科学, 2001, 25(6): 11–13]
- O'Connell M, Wright JM. Microsatellite DNA in fishes. Reviews in Fish Biology and Fisheries, 1997, 7(3): 331–363
- Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology, 2000, 132(3): 365–386
- Samadi S, Artiguebielle E, Estoup A, *et al.* Density and variability of dinucleotide microsatellites in the parthenogenetic polyploid snail *Melanoides tuberculata*. Molecular Ecology, 1998, 7(9): 1233–1236
- Sato M, Kawamata K, Zaslavskaya N, *et al.* Development of microsatellite markers for Japanese scallop (*Mizuhopecten yessoensis*) and their application to a population genetic study. Marine Biotechnology, 2005, 7(6): 713–728
- Schlotterer C, Tautz D. Slippage synthesis of simple sequence DNA. Nucleic Acids Research, 1992, 20(2): 211–215
- Song LP, Liu P, Li J, *et al.* Analysis of microsatellite sequences in genome of crab, *Portunus trituberculatus*. Journal of Fishery Sciences of China, 2008, 15(5): 738–744 [宋来鹏, 刘萍, 李健, 等. 三疣梭子蟹基因组微卫星特征分析. 中国水产科学, 2008, 15(5): 738–744]
- Wang HX, Huan P, Lu X, *et al.* Mining of EST-SSR markers in clam *Meretrix meretrix* larvae from 454 shotgun transcriptome. Genes and Genetic Systems, 2011, 86(3): 197–205
- Wang RF, Li J, Li JT, *et al.* Genetic analysis of 33 microsatellite loci in 3 families of *Exopalaemon carinicauda* inbred line. Progress in Fisheries Science, 2017, 38(4): 78–86 [王日芳, 李健, 李吉涛, 等. 脊尾白虾(*Exopalaemon carinicauda*)近交系 3 个家系 33 个微卫星座位的遗传分析. 渔业科学进展, 2017, 38(4): 78–86]
- Wang ZL, Ding J, Xu YH, *et al.* Polymorphism of EST-SSRs from *Pinctada martensii* based on transcriptome datasets. Oceanologia et Limnologia Sinica, 2015, 46(3): 687–693 [王忠良, 丁燊, 许尤厚, 等. 基于转录组数据的马氏珠母贝 EST-SSR 位点的信息分析及其多态性检测. 海洋与湖沼, 2015, 46(3): 687–693]
- Wu X, Shen HD, Wang DF, *et al.* Development of microsatellite markers for *Paraoncidium reevesii* based on transcriptome sequencing Illumina HiseqTM 2000. Marine Sciences, 2015, 39(11): 20–25 [吴欣, 沈和定, 王冬凤, 等. 基于 Illumina HiseqTM 2000 高通量转录组测序的里氏拟石磺 SSR 标记开发. 海洋科学, 2015, 39(11): 20–25]
- Yan LL, Qin YJ, Yan XW, *et al.* Development of microsatellite markers in *Ruditapes philippinarum* using next-generation

- sequencing. *Acta Ecologica Sinica*, 2015, 35(5): 1573–1580 [闫路路, 秦艳杰, 闫喜武, 等. 基于转录组平台的蛤仔微卫星标记筛选. *生态学报*, 2015, 35(5): 1573–1580]
- Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 2008, 18(5): 821–829
- Zhang LL, Bao ZM, Wang S, *et al.* FISH mapping and identification of Zhikong scallop (*Chlamys farreri*) chromosomes. *Marine Biotechnology*, 2008, 10(2): 151–157
- Zhang LL, Li L, Zhang GF. Bioinformatics data mining of EST-tandem repeats of the Pacific oyster (*Crassostrea gigas*). *Marine Sciences*, 2011, 35(4): 9–14 [张琳琳, 李莉, 张国范. 长牡蛎(*Crassostrea gigas*)EST 串联重复序列的组成和分布. *海洋科学*, 2011, 35(4): 9–14]
- Zhao YY, Zhu XC, Sun XW, *et al.* Microsatellite enrichment by magnetic beads in Japanese scallop, *Patinopecten yessoensis*. *Journal of Fishery Sciences of China*, 2006a, 13(5): 749–755 [赵莹莹, 朱晓琛, 孙效文, 等. 磁珠富集法筛选虾夷扇贝微卫星序列. *中国水产科学*, 2006a, 13(5): 749–755]
- Zhao YY, Zhu XC, Sun XW, *et al.* Polymorphic microsatellite loci of the Japanese scallop *Patinopecten yessoensis*. *Acta Zoologica Sinica*, 2006b, 52(1): 229–233 [赵莹莹, 朱晓琛, 孙效文, 等. 虾夷扇贝的多态性微卫星座位. *动物学报*, 2006b, 52(1): 229–233]
- Zhu H, Senalik D, McCown BH, *et al.* Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theoretical and Applied Genetics*, 2012, 124(1): 87–96

(编辑 冯小花)

## Microsatellite Analysis of *Patinopecten yessoensis* Using Next-Generation Sequencing Method

NI Shousheng<sup>1,2,3</sup>, YANG Yu<sup>1,2,4</sup>, LIU Shufang<sup>1,2①</sup>, ZHUANG Zhimeng<sup>1</sup>

(1. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071; 2. Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071;

3. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306;

4. College of Fisheries and Life Science, Dalian Ocean University, Dalian 116023)

**Abstract** Microsatellite has become an important molecular marker for genetic diversity analysis and marker-assisted breeding as its unique advantages, such as high stability and repeatability, strong specificity and codominance. To understand the distribution and frequency of microsatellite of *P. yessoensis* genome, the current study analyzed the microsatellite sequences of *P. yessoensis* genome by establishing and sequencing a RAD library. The sequences were assembled, and SSRs were detected by bioinformatics method. The total length of the DNA sequences of *P. yessoensis* genome is 92,551,435 bp and 3,618 contigs from all filtered 259,535 contigs containing SSR sequences. The 3460 potential SSR loci were identified with the number of repeat motif from 2 to 6 using appropriate amplifying primers. A total of 1587 trinucleotide SSRs (45.87%) were the most common motif, and ATA (11.41% of whole trinucleotide motifs) was the richest motif. The dinucleotide and the tetranucleotide types ranked the second and the third with a proportion of 37.05%(1282) and 14.48%(501), respectively, while the pentanucleotide type accounted for 2.02%(70), and the hexanucleotide type was the least amount with a proportion of 0.58%(20). Interestingly, the abundance of microsatellites of the same repeat type decreased with the increase of copy number, and the abundance of microsatellites of the same copy number decreased with the increase of the repeat unit length. The variation of length distribution frequency, copy numbers and the abundance of the microsatellite suggested that the mutation rate of shorter repeats was larger than that of the longer repeats. These results provide basic information of microsatellite of *P. yessoensis*, which would be useful to study the genome and population genetics of *P. yessoensis*.

**Key words** *Patinopecten yessoensis*; Next-generation sequencing; Genomes; Microsatellite

① Corresponding author: LIU Shufang, E-mail: liusf@ysfri.ac.cn