

基于 Illumina HiseqTM 2000 高通量转录组测序的里氏拟石磺 SSR 标记开发

吴 欣, 沈和定, 王冬凤, 刘 宸, 杨铁柱, 刁 亚

(上海海洋大学 省部共建水产种质资源发掘与利用教育部重点实验室, 上海 201306)

摘要: 为给石磺科贝类系统进化研究和资源保育工作提供更有用的分子标记, 使用 Illumina HiseqTM 2000 高通量测序技术对里氏拟石磺进行转录组测序, 以所得的拼接序列为基础, 利用 MISA 和 SSR Hunter1.3 软件进行微卫星序列的查找, 对选出的 51 个微卫星位点进行引物设计, 经 PCR 扩增得到单一清晰扩增条带的位点 24 个, 研究湛江里氏拟石磺群体扩增。结果表明, 其中 14 个位点表现多态性。14 个多态性位点得到的等位基因数为 2~4, 其中观测杂合度(H_o)和期望杂合度(H_e)分别为 0.031 3~0.714 3 和 0.031 3~0.708 4, 多态性信息含量(PIC)为 0.089 5~0.640 5, 14 个位点中的 2 个位点表现为高度多态性($PIC > 0.5$), 10 个位点表现为中度多态性($0.5 > PIC > 0.25$), 另外 2 个位点表现为低度多态性($PIC < 0.25$)。经 Sequential Bonferroni 校正的 Hardy-Weinberg 平衡检验, 9 个位点不偏离平衡($P > 0.05$)。开发的 SSR 标记可用于里氏拟石磺群体遗传学和石磺科贝类间的系统发生研究, 并为里氏拟石磺的群体种质保护提供数据支持。

关键词: 里氏拟石磺; 转录组; 微卫星标记

中图分类号: Q953+.1 文献标识码: A

文章编号: 1000-3096(2015)11-0020-06

doi: 10.11759/hyxx20150803002

石磺是软体动物门(Mollusca)、腹足纲(Gastropoda)、肺螺亚纲(Pulmonata)、石磺总科(Onchidioidea)、石磺科(Onchidiidae)的一群贝类, 主要分布于热带或亚热带地区的海岸、滩涂、岩礁、红树林、潮间带地区, 是海洋与陆地的过渡带生物^[1-2]。我国的石磺主要分布在东南沿海的潮间带高潮区, 已报道易采集的主要有 4 种, 包括瘤背石磺(*Onchidium struma*)、平疣桑葚石磺(*Platevindex mortoni*)、紫色疣石磺(*Peronia verruculata*)、里氏拟石磺(*Paraoncidium reevesii*)。学者对于石磺的研究主要集中在瘤背石磺, 对后 3 种研究报道较少。其中, 里氏拟石磺的研究主要集中在形态学^[1-3]、化学成分测定^[4-6]、线粒体基因组及群体多样性的分析^[7-9]几个方面, 未见利用分子标记研究其遗传多样性的报道。

微卫星, 又称简单重复序列(SSR)^[10], 具有高稳定性、高多态性、引物通用性、位点特异性、检测方便和呈共显性遗传等特点, 在水产动物中主要用于遗传连锁图谱的制备、性状分析、QTL 定位、群体遗传学、进化遗传、种质鉴定与亲权分析、品种培育等方面^[11]。里氏拟石磺作为海洋向陆地进化的过渡物种之一, 在研究海洋无脊椎动物向陆地辐射进

化中具有重要作用^[12-13], 本研究利用 Illumina HiseqTM 2000 高通量转录组测序对里氏拟石磺进行微卫星(SSR)分子标记开发, 为研究其群体遗传学和石磺科不同种贝类间的系统发生提供基础数据。

1 材料和方法

1.1 实验材料

实验材料于 2015 年 5 月采自广东湛江海滩, 共 37 个, 样本经鉴定后均保存于 95%无水乙醇并置于 -20℃ 冰箱中备用。

1.2 实验方法

1.2.1 里氏拟石磺转录组测序结果

随机挑选 6 只于 2012 年 10 月采自湛江的里氏拟石磺个体进行活体解剖, 将解剖得到的不同组织保存于 RNA later(Qiagen: 76104)中用于构建转录组

收稿日期: 2015-08-03; 修回日期: 2015-10-11

基金项目: 国家自然科学基金(41276157), 上海高校水产学一流学科建设

作者简介: 吴欣(1990-)男, 河南南阳人, 硕士研究生, 主要从事海洋贝类分子标记开发, E-mail:wuxin2009001@163.com; 沈和定, 通信作者, 教授, 博士, 主要从事海洋生物学研究, E-mail:hdshen@shou.edu.cn

测序文库并送晶能公司(Genergy Biotechnology)进行 Illumina HiseqTM 2000 双末端测序及生物信息学分析,共获得原始数据 9.161 GB,有效数据 8.617 GB。通过 de novo 拼接获得了长度大于 100 bp 的转录物 233 636 条;Unigene N50 长度 485 bp,经 GO 注释分析有 124 598(53.33%)条归入生物学过程,51 891 (22.21%)条归入分子功能,57 136(24.46%)条归入细胞组分,对大于 200 bp 的拼接序列与 Nr 数据库进行比对,22 558 条序列得到蛋白注释,匹配蛋白数目为 2 910 个,对注释后的数据进行分析,共涉及物质及能量代谢、信号传导、转录调控及免疫应答等诸多生理生化过程。

1.2.2 微卫星序列的来源

根据晶能公司(Genergy Biotechnology)基于 Illumina HiseqTM 2000 测序平台所得的拼接序列,经 MISA(<http://pgarc.ipk-gatersleben.de/misa/>)对微卫星位点进行分析,对于非混合型位点筛选时,设置条件为单碱基重复至少 10 次,2 碱基重复至少 6 次,3~6 碱基重复至少 5 次,在此基础上,如同一序列中两个 SSR 位点间距离小于 50 bp,则认为这两个 SSR 位点组成一个混合型 SSR 位点,在设计引物前利用 SSR Hunter1.3^[14]对经 MISA 筛选过的含有微卫星位点的序列进行再次筛选,保证序列的前后侧翼有足够的长度用于设计引物,筛选条件和第一次一样,最终筛选出 51 条含有微卫星位点并适合设计引物的序列。

1.2.3 SSR 引物设计与合成

利用 Primer Premier 5.0(<http://www.premierbio.com/>)软件对上一步所得的 51 条拼接序列进行引物设计,引物长度为 19~26 bp,GC 含量在 40%~60%之间,也可放宽至 30%~70%,引物 Tm 值小于 72℃,且上下游引物 Tm 值差别不能大于 2℃,扩增引物长度在 100~300 bp,所设计引物在侧翼保守序列内。本实验共设计 51 对引物,所有引物均由生工生物工程(上海)股份有限公司合成。

1.2.4 样品 DNA 的提取

取出之前保存于-20℃冰箱样品,根据天根生化科技(北京)有限公司的海洋动物基因组提取试剂盒(DP324)的要求,在每只样品腹足处各取 30mg 组织块用于 DNA 提取,提取后的 DNA 置于-20℃冰箱暂存备用。

1.2.5 PCR 扩增及产物检测

PCR 反应体系为 10μL,包括上下游引物各 0.3 μL,

2×Taq PCR Master Mix 5μL, ddH₂O 4μL, DNA 模板 0.4 μL。PCR 反应条件为:94℃预变性 5 min;94℃ 30 s,退火 30 s,72℃ 30 s,共 40 个循环;72℃延伸 10 min。产物保存于 4℃。引物初筛 PCR 产物使用 1%的琼脂糖凝胶电泳检测,God view 染色,电泳条件为:115 V,35 min,电泳缓冲液为 1×TAE,凝胶成像系统拍照。引物进行群体检测时,PCR 扩增产物 8%的聚丙烯酰胺凝胶电泳进行分离,电泳条件为:电压 110 V,电泳时间 6 h,电泳缓冲液为 1×TBE;电泳结束后用硝酸银对凝胶染色并使用相机进行拍照。

1.2.6 里氏拟石磺 SSR 引物的筛选及群体检测

随机选取 5 个里氏拟石磺个体,提取其 DNA 并将其混合作为引物特异性和最适退火温度筛选时的模板 DNA,根据每对引物的 Tm 值设置间隔为 1.5℃的温度梯度,共设置 8 个温度梯度,用 1%的琼脂糖凝胶电泳进行检测,将扩增清晰,条带单一的引物留用,确定每对引物的最适退火温度,以便进行后续检测。

选取 32 个湛江群体的里氏拟石磺个体,提取 DNA 作为模板,使用初筛后留用的引物进行 PCR 扩增,PCR 反应条件及产物检测方法遵照上述方法,并对产物进行拍照留存。

1.2.7 数据统计和分析

对湛江群体里氏拟石磺的 PCR 扩增结果进行统计时,将每个位点的扩增条带按从小到大的顺序标记为 A, B, C...,并统计出每个位点在进行群体检测时的基因型,使用 Cervus3.0^[15]软件计算每个微卫星位点的等位基因数(number of allele, Na),观测杂合度(observed heterozygosity, Ho),期望杂合度(expected heterozygosity, He),多态性信息含量(polymorphic information content, PIC),使用 Genepop4.2^[16]进行每个位点的 Hardy-Weinberg 平衡的检测。

2 结果分析

2.1 转录组中 SSR 位点的软件分析结果

利用 MISA 软件在里氏拟石磺转录组中查找 SSR 位点的结果,如表 1 所示。

2.2 里氏拟石磺 SSR 筛选结果

51 对引物在经过琼脂糖凝胶电泳的初筛之后有 24 对扩增出清晰单一的条带,扩增率为 47%。将这 24 对引物在湛江群体 32 个个体中做进一步的聚丙烯

表 1 SSR 分析结果

Tab.1 Output statistics of SSR

SSR 标记重复类型	拼接序列个数
查找到的 SSR 总个数	67696
含有 SSR 位点序列个数	46164
含有多于 1 个 SSR 位点的序列数	14423
单碱基型 SSR 个数	53948
双碱基型 SSR 个数	7577
3 碱基型 SSR 个数	4579
4 碱基型 SSR 个数	1572
5 碱基型 SSR 个数	14
6 碱基型 SSR 个数	6

酰胺凝胶电泳筛选, 有 14 个位点表现为多态性, 其余位点表现为单态性。在表现出多态性的 14 个位点中, 7 个位点的等位基因为 4, 4 个位点的等位基因为 3, 3 个位点表现为 2 等位基因。14 个位点中, 1 个位

点为五碱基重复, 2 个位点为四碱基重复, 11 个位点为三碱基重复, 且四, 五碱基重复的次数均为 5, 三碱基的重复次数均为 6。14 对引物中最低退火温度为 50.5℃, 最高退火温度为 61℃, 扩增片段的长度在 90~370 bp。

2.3 里氏拟石磺湛江群体遗传多样性分析

对 14 对 SSR 引物在取自湛江群体的 32 个体扩增片段进行统计分析, 14 个位点的等位基因数在 2~4 之间, 其中 H_o 和 H_e 分别在 0.031 3~0.714 3 和 0.031 3~0.708 4 之间, 多态性信息含量(PIC)在 0.089 5~0.640 5 之间, 其中 2 个微卫星位点表现为高度多态性 ($PIC > 0.5$), 10 个位点表现为中度多态性 ($0.5 > PIC > 0.25$), 2 个位点表现为低度多态性 ($PIC < 0.25$)。经 Sequential Bonferroni 校正的 Hardy-Weinberg 平衡检验, 9 个位点不偏离平衡 ($P > 0.05$), 其余位点均显著偏离平衡 ($0.05 > P$, 表 2)。

表 2 14 对 SSR 引物在 32 个里氏拟石磺扩增特性的描述

Tab.2 Primer sequences and characterization of 14 polymorphic microsatellite loci in 32 *Paraoncidium reevesii*

Locus	Primers: (5'→3')	Repeat motif	Ta(°C)	Size range(bp)	Na	H_o	H_e	PIC	P value HWE
PR4	F: CTTGTTACACCCCGCT CCC R: CTGTGACATAATCAA TGCTGCTCT	(GCCCT)5	60	90~100	4	0.7143	0.7084	0.6405	0.0349
PR5	F: TGATGCGTCAATACA AGGGTG R: AGTTCGGAGTGCGGA ATGGT	(ATGC)5	50.5	172~185	3	0.3750	0.3229	0.2888	1.0000
PR6	F: AGTTACAGGTCCAGA CAATG R: CTAAGACGGCAATCC AATAC	(AATA)5	56.3	250~275	2	0.6944	0.5035	0.3733	0.0408
PR9	F: TAAAGAAGACAGAA ACAGGCAACA R: CAACGGTCATAAAGA CTAAAGGAAT	(ATC)6	61	128~161	4	0.4063	0.3566	0.3333	1.0000
PR10	F: TCTGCTGAACTTGCC ATTGT R: TCCTCTACCAGGGCT CAAAG	(ATT)6	57	131~150	4	0.3929	0.3468	0.3227	1.0000
PR12	F: TCCACCTTCTTTATTA CCACCC R: GCCACGCAGTATTGA TTGTT	(TAT)6	56	122~136	2	0.0313	0.0313	0.0303	1.0000
PR14	F: TTCTTCTTCAGGGAG GTAAA R: TCAAGTTCTTCATTC GGTTC	(CAT)6	55	161~193	3	0.4444	0.5524	0.4408	0.1739

续表

Locus	Primers: (5'→3')	Repeat motif	Ta(°C)	Size range(bp)	Na	Ho	He	PIC	P value HWE
PR18	F: TGTGTGGTTGCCTGCCT TTCA R: ATCGGCTGGCTTGCC TGTTT	(GTA)6	56	150~162	2	0.3214	0.4929	0.3669	0.1125
PR20	F: CTGTCTGTTTGGCTG TGATG R: AATGAAGTCTTGCGT AGATGAA	(TTG)6	53	185~280	4	0.4375	0.4172	0.3801	0.0555
PR22	F: GACAAACGACGAGC GACAAA R: TGGAAGAAGGACAA GGAGGAGTA	(TGC)6	52	120~195	3	0.4583	0.5496	0.4802	0.3018
PR28	F: GGTCCCTCTGTGGAT TTGGT R: AAGGATTCTGGCTTC GGTGA	(GAT)6	56	195~270	4	0.6129	0.6171	0.5315	0.0459
PR37	F: CATTCGCCCTCTGTTC TATCC R: TGTCAAACGGTGCTT CTCCTA	(ATC)6	61	260~370	3	0.4333	0.3655	0.3262	1.0000
PR46	F: ATAATGATGCTGGAA CTGAT R: AGGATAACAAACTG GGACTT	(ATT)6	52	250~300	4	0.6250	0.4673	0.3799	0.0304
PR50	F: TCGGGTTTCCTTGTCT TGTA R: ATCTCCATCACCTTG GCAGT	(ATT)6	58	102~175	4	0.0625	0.0923	0.0895	0.0469

Locus: 位点; Ta: 退火温度, Annealing temperature; Na: 等位基因数 numbers of alleles; P value: P 值 the test for deviation from HWE

3 讨论和结论

SSR 是一种应用广泛的分子标记。利用高通量测序技术对里氏拟石磺进行微卫星分子标记的开发不仅比传统方法更为快捷, 花费更少, 而且在获得微卫星位点的同时可以查找对应 EST 序列的功能性, 与基因关联性较强。相比于基因组 SSR, 基于转录组 EST 序列开发的 SSR 多态性较低^[17], 本实验共获得 14 个具有多态性的微卫星位点, 其中 2 个微卫星位点表现为高度多态性(PIC>0.5), 10 个位点表现为中度多态性(0.5>PIC>0.25), 2 个位点表现为低度多态性(PIC<0.25), 这表明湛江里氏拟石磺野生群体的遗传多态性处于中度多态性水平。14 个位点中有 5 个位点偏离平衡($P<0.05$), 可能的原因是种群中有个体产生基因漂变, 或者出现了无效等位基因^[18]。

作为海洋无脊椎动物向陆地延伸的过渡物种,

石磺科贝类认为是研究海洋贝类向陆地辐射生活的最好代表。而利用微卫星分子标记来研究水产动物的系统进化已经得到广泛引用: 汪桂玲^[19]等曾利用微卫星分析五大湖野生三角帆蚌群体的遗传多样性和种间发生关系; 董秋芬^[20]等利用 13 个微卫星位点对 9 种石斑鱼的种系发生关系进行研究。这些都为利用 SSR 标记研究石磺科贝类间种系发生提供了借鉴。本研究中利用高通量测序技术开发的 14 个微卫星位点, 不仅为研究里氏拟石磺的遗传多样性提供了基础数据, 也可用于石磺科贝类间种系发生关系的研究中。

参考文献:

- [1] 吴旭峰, 沈和定, 吴文健, 等. 我国华东沿海 4 种石磺形态学比较[J]. 动物学杂志, 2010, 6: 92-100.
- [2] Bouchet P, Rocroi J. Classification and nomenclator of

- gastropod families [J]. *Malacologia*, 2005, 47(1): 2.
- [3] 王冬凤, 沈和定, 吴欣. 石磺科 3 种贝类皮肤显微结构比较[J]. *动物学杂志*, 2015, 50(3): 437-444.
- [4] 黄金田, 王爱民. 瘤背石磺营养成分分析及品质评价[J]. *海洋科学*, 2008, 32(11): 29-35.
- [5] 孙变娜, 沈和定, 吴洪喜, 等. 四种石磺科贝类的胆甾醇含量测定[J]. *海洋科学*, 2015, 39(1): 24-28.
- [6] Sun B, Shen H, Wu H, et al. Determination of Chemical Constituents of the Marine Pulmonate Slug, *Paraoncidium reevesii* [J]. *Tropical Journal of Pharmaceutical Research*, 2015, 13(12): 2071-2074.
- [7] Zhu H C, Wei L L, Shen H D, et al. Complete mitochondrial genome of *Paraoncidium reevesii* (Gastropoda, Pulmonata, Systellommatophora, Onchidiidae)[J]. *Mitochondrial DNA*, 2012, 23(5): 379-381.
- [8] Sun B, Chen C, Shen H, et al. Species diversity of Onchidiidae (Eupulmonata: Heterobranchia) on the mainland of China based on molecular data[J]. *Molluscan Research*, 2014, 34(1): 62-70.
- [9] 周娜, 沈和定, 陈诚. 基于线粒体 CO I 基因的中国沿海和泰国普吉岛里氏拟石磺群体遗传结构分析[J]. *水产学报*, 2014, 01: 33-40.
- [10] Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers[J]. *Nucleic Acids Research*, 1989, 17 (16): 6463-6471.
- [11] 孙效文, 张晓锋, 赵莹莹, 等. 水产生物微卫星标记技术研究进展及其应用[J]. *中国水产科学*, 2008, 15(4): 689-703.
- [12] Tillier S. A new mountain Platevindex from Philippine islands (Pulmonata: Onchidiidae) [J]. *Journal of Molluscan Studies*, 1983, 49(supp12A): 198-202.
- [13] Klussmann-Kolb A, Dinapoli A, Kuhn K, et al. From sea to land and beyond—new insights into the evolution of euthyneuran Gastropoda (Mollusca)[J]. *BMC Evolutionary Biology*, 2008, 8(1): 57.
- [14] 李强, 万建民. SSRHunter, 一个本地化的 SSR 位点搜索软件的开发[J]. *遗传*, 2005, 05: 808-810.
- [15] Kalinowski S T, Taper M L, Marshall T C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment [J]. *Molecular Ecology*, 2007, 16(5): 1099-1106.
- [16] Rousset F. genepop'007: a complete re implementation of the genepop software for Windows and Linux [J]. *Molecular Ecology Resources*, 2008, 8(1): 103-106.
- [17] 程小毛, 黄晓霞. SSR 标记开发及其在植物中的应用[J]. *中国农学通报*, 2011, 27(5): 304-307.
- [18] Liu N, Li S, Zhang J. Isolation and characterization of 16 polymorphic microsatellite loci in the leopard coral grouper *Plectropomus leopardus*[J]. *Conservation Genetics Resources*, 2013, 5(4): 1067-1069.
- [19] 汪桂玲, 袁一鸣, 李家乐. 中国五大湖三角帆蚌群体遗传多样性及亲缘关系的 SSR 分析[J]. *水产学报*, 2007, 31(2): 152-158.
- [20] 董秋芬, 刘楚吾, 郭昱嵩, 等. 9 种石斑鱼遗传多样性和系统发生关系的微卫星分析[J]. *遗传*, 2007, 29(7): 837-843.

Development of microsatellite markers for *Paraoncidium reevesii* based on transcriptome sequencing Illumina His-seq™ 2000

WU Xin, SHEN He-ding, WANG Dong-feng, LIU Chen, YANG Tie-zhu, DIAO Ya

(Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China)

Received: Aug., 3, 2015

Key words: *Paraoncidium reevesii*; transcriptome; microsatellite marker

Abstract: Repeats in the transcriptome sequences of *Paraoncidium reevesii* were searched using MISA and SSR Hunter1.3, 51 sequences with SSR were selected for primer designing. 24 primers were produced clearly defined bands after preparatory screening. The further test in Zhanjiang population indicated that 14 loci were showed polymorphism. The allele numbers per locus ranged from 2 to 4 and the observed and expected heterozygosity ranged from 0.0313 to 0.7143 and from 0.0313 to 0.7084, respectively. The polymorphic information contents (PIC) of each locus were between 0.0895 and 0.6405. Among 14 loci, 2 had a $PIC > 0.5$, 10 had a $0.5 > PIC > 0.25$, 2 had a $PIC < 0.25$. 9 loci were detected not deviated from HWE ($P > 0.05$) in 14 after Bonferroni correction. No pairs of loci showed significant linkage disequilibrium. These loci can be used for the research of population genetic and phylogenetic analysis of Onchidiidae. Furthermore, it will provide molecular data to conservation of *P. reevesii* resources.

(本文编辑: 李晓燕)