

DOI: 10.12131/20230086

文章编号: 2095-0780-(2023)05-0123-11

# 基于基因组 survey 数据的疣吻沙蚕微卫星特征分析及多态标记开发

杨尉, 司圆圆, 许瑞雯, 陈兴汉  
阳江职业技术学院, 广东 阳江 529566

**摘要:** 为了解疣吻沙蚕 (*Tylorrhynchus heterochaetus*) 基因组信息并高效地开发微卫星标记, 指导其种质资源保护与新品种的遗传改良研究, 采用低深度高通量测序开展全基因组 survey, k-mer 分析估计疣吻沙蚕基因组大小为 759.53 Mb, 杂合率 1.41%, 重复序列比例 45.92%; 初步组装获得 2 181 621 条 scaffold, 全长为 840 375 821 bp。在基因组序列中检测到 130 216 个微卫星位点, 丰度为 154.9 个·Mb<sup>-1</sup>。微卫星重复次数集中在 4~18 拷贝; 单碱基重复比例最高 (35.00%), 二碱基 (32.48%)、三碱基 (14.42%) 次之; 二碱基、三碱基优势基序分别是 AT/AT、AAT/ATT, 表现出 A/T 碱基优势。从随机挑选的 50 对引物中筛选到 15 对多态标记, 在 30 尾样本中共检测到 87 个等位基因, 等位基因数 ( $N_a$ ) 为 2.000~12.000 (平均 5.800), 有效等位基因数 ( $N_e$ ) 为 1.164~6.713 (平均 3.328), 期望杂合度 ( $H_e$ ) 为 0.141~0.789 (平均 0.561), 多态信息含量 (PIC) 为 0.136~0.776 (平均 0.511); 其中 13 个为高度或中度多态性位点, 在遗传分析中有较高的实用价值。结果表明, 疣吻沙蚕基因组为复杂基因组, 其微卫星位点类型丰富且具备良好的多态性潜能, 可为种质资源评价、群体遗传学及分子育种研究提供有效的标记资源。

**关键词:** 疣吻沙蚕; 全基因组 survey; 简单重复序列; 多态性位点

中图分类号: S 917.4

文献标志码: A

开放科学 (资源服务) 标识码 (OSID):



## Characterization of microsatellites and polymorphic marker development in ragworm (*Tylorrhynchus heterochaetus*) based on genome survey data

YANG Wei, SI Yuanyuan, XU Ruiwen, CHEN Xinghan

Yangjiang Polytechnic, Yangjiang 529566, China

**Abstract:** In order to understand the genomic information of *Tylorrhynchus heterochaetus* and efficiently develop microsatellite markers, so as to guide the conservation of its germplasm resources and genetic improvement of new varieties, we conducted a whole-genome survey by using low depth high-throughput sequencing. A total of 57.48 Gb of clean data were generated after the quality control of raw data. K-mer analysis estimates that the genome size of *T. heterochaetus* was 759.53 Mb; the heterozygosity rate was 1.41%; the proportion of repetitive sequences was 45.92%. Preliminary assembly obtained 2 181 621 scaffolds with a total length of 840 375 821 bp. A total of 130 216 microsatellite loci were detected with a density of 154.9 loci per Mb. The repeated number of microsatellite units largely ranged from 4 to 18. The ratio of mononucleotide loci was the highest (35.00%),

收稿日期: 2023-04-25; 修回日期: 2023-06-29

基金项目: 广东省自然科学基金项目 (2022A1515011231); 广东省普通高校创新团队项目 (2021KCXTD054); 广东省农业科技社会化服务与成果集成示范项目 (2023B0202010011); 广东省农业农村厅农业科研类及技术推广示范类项目 (0202020014); 阳江职业技术学院自然科学基金重点项目 (2022kjzd01)

作者简介: 杨尉 (1984—), 男, 讲师, 博士, 研究方向为水产动物种质资源创新。E-mail: yangwei516@163.com

通信作者: 陈兴汉 (1978—), 男, 教授, 博士, 研究方向为海洋河口渔业资源保护与利用。E-mail: chenxh1978@163.com

followed by those of dinucleotide (32.48%) and trinucleotide (14.42%) loci. AT/AT and AAT/ATT motifs were dominant in dinucleotide and trinucleotide loci, respectively, indicating an A/T dominance. Fifteen polymorphic loci were identified from 50 randomly selected primers, and 87 alleles were amplified in a *T. heterochaetus* population containing 30 individuals. The number of alleles per locus ranged from 2.000 to 12.000, with an mean of 5.800. The effective allele number ( $N_e$ ) and expected heterozygosity ( $H_e$ ) ranged from 1.164 to 6.713 and from 0.141 to 0.789, with means of 3.328 and 0.561, respectively. The polymorphic information content (PIC) ranged from 0.136 to 0.776, with a mean of 0.511. Thirteen loci were found to be highly or moderately polymorphic, having high practical value in genetic analysis. In conclusion, *T. heterochaetus* genome is a complex genome, and its microsatellites have a rich variety and high polymorphic potential. The results can provide effective marker resources for germplasm resource evaluation, population genetics and molecular breeding research.

**Keywords:** *Tylorrhynchus heterochaetus*; Whole-genome survey; Simple sequence repeat; Polymorphic loci

疣吻沙蚕 (*Tylorrhynchus heterochaetus*) 又名疵吻沙蚕, 属环节动物门、多毛纲、叶须虫目、沙蚕科, 俗称禾虫、流蜞, 在中国东南沿海河口地区的泥沙质浅滩或稻田中广泛分布<sup>[1-2]</sup>。作为一种经济价值极高的多毛类, 其味道鲜美且营养丰富, 是中国广东、广西、福建、香港、澳门及东南亚各地独具特色的水产品, 素有“水中冬虫夏草”之美誉<sup>[2-3]</sup>。更重要的是, 疣吻沙蚕与水稻具有天然的共生关系, 发展疣吻沙蚕稻田综合种养模式可显著增加水稻种植的经济效益<sup>[4]</sup>。“水稻+疣吻沙蚕”生态综合种养技术正日趋成熟, 自2021年起连续3年入选广东省农业主推技术, 在助力乡村振兴、保障国家粮食安全等方面应用潜力巨大。然而, 疣吻沙蚕具有生境狭窄、种群地理隔离、群体恢复力低等物种特性, 受栖息地破坏、过度捕捞、环境污染等因素的影响其自然种群呈逐年衰退的趋势, 在有些地区甚至已经绝迹<sup>[4-5]</sup>。在突破疣吻沙蚕全人工繁殖技术的基础上, 近年已实现其苗种的规模化培育, 增养殖业的发展以及技术的革新培育了大量的人工繁育群体, 但因遗传育种研究的滞后, 疣吻沙蚕种质开始出现生长性能及抗逆性下降等退化现象<sup>[4]</sup>。为建立疣吻沙蚕资源管理策略并促进其可持续开发利用, 特别是避免常年的产业化增养殖引起遗传结构单一化, 防止人工繁育群体污染自然种群, 有必要采用分子手段科学地分析及评估其种质现状, 揭示种群遗传多样性水平与遗传结构, 并以此指导新品种的培育与遗传改良研究。

目前, 国内外针对疣吻沙蚕的研究主要集中在形态学、生活史、繁殖生物学、增养殖技术、营养与活性成分分析等方面<sup>[3,5]</sup>。Chen等<sup>[5-6]</sup>完成了疣吻沙蚕线粒体基因组测序, 并基于线粒体COI序列分析了7个地理群体的遗传结构, 除此之外, 遗传学相关的研究鲜有报道; 分子遗传信息的匮乏制

约了疣吻沙蚕资源保护利用研究的深入开展。微卫星标记具有多态性丰富、共显性遗传、基因组覆盖广泛等优点<sup>[7]</sup>, 该技术操作简便、检测周期短、稳定性高、重复性好, 已广泛应用于谱系鉴定、亲缘关系检测、遗传多样性分析、种质资源评价、遗传图谱构建及基因定位等领域<sup>[8]</sup>。基因组 survey 是利用高通量测序技术实施小片段、低深度测序, 然后基于 Lander-Waterman 模型进行 k-mer 分析, 并根据 k-mer 频率和深度的统计结果评估物种基因组的大小与复杂程度等关键特征<sup>[9]</sup>。基因组 survey 不仅能为全基因组测序及高质量组装提供科学依据, 而且对无参考基因组序列的物种而言, 利用基因组 survey 数据大规模开发微卫星标记是当前最高效的策略之一, 较传统方法具有效率高、周期短、成本低的优势<sup>[9]</sup>。基因组微卫星鉴定研究在水产经济动物中已广泛开展, 筛选的标记获得了良好的遗传分析效果。例如, 张永德等<sup>[9]</sup>在卵形鲳鲹 (*Trachinotus ovatus*) 基因组 survey 数据中检测到 190 121 个微卫星位点并成功开发了多态分子标记, 为全基因组测序与组装、渔业资源保护利用及良种选育奠定了基础; 上官清等<sup>[10]</sup>分析了斑鳢 (*Channa maculata*) 基因组微卫星特征, 并筛选到 20 个多态性位点用于群体遗传多样性和遗传结构分析, 为该物种的遗传监测、亲缘关系鉴定、种质资源养护及管理提供了技术支持。

本研究对疣吻沙蚕基因组进行低深度高通量测序, 通过 k-mer 分析预测基因组大小、杂合度和重复比例等信息, 对测序数据初步组装后搜索组装序列中的微卫星位点, 分析其特征与分布规律, 初步验证标记的有效性和多态性, 以期指导基因组精细图谱的绘制, 并为群体遗传学研究提供可靠的标记资源。

## 1 材料与amp;方法

### 1.1 实验材料

疣吻沙蚕采集自阳江市广东阳海农业技术发展有限公司疣吻沙蚕增殖试验基地 (111°55'23"E、21°49'16"N) 保种的越南海防群体。用于基因组测序的疣吻沙蚕幼体用适量无菌水洗涤3次, 解剖后剪取体壁肌肉组织装入2 mL冻存管, 液氮速冻后置于-80 °C保存。微卫星标记多态性验证群体的30尾个体经无菌水洗涤后, 解剖剪取体壁肌肉组织, 置于体积分数为95%的乙醇中在-20 °C下保存。

### 1.2 基因组测序及生物信息学分析

运用苯酚-氯仿法提取疣吻沙蚕基因组DNA, 经1.0% (w) 琼脂糖凝胶电泳检测完整度后, 用NanoDrop 2000超微量分光光度计 (ThermoFisher, 美国) 检测浓度和纯度。利用Covaris超声波破碎仪将基因组DNA片段化, 筛选合适长度的DNA片段, 经末端修复、加A尾、添加测序接头、纯化及PCR扩增等步骤建立350 bp小片段文库。测序文库经Qubit 2.0 (ThermoFisher, 美国) 和Agilent 2100 Bioanalyzer (Agilent, 美国) 检测浓度和插入片段大小后, 用Illumina HiSeq™ X Ten平台进行双末端测序。基因组测序工作委托北京诺不致源科技股份有限公司完成。原始数据经质控和过滤后, 用GCE 1.0.0<sup>[11]</sup> 软件对有效数据进行k-mer分析。运用SOAPdenovo 2.01<sup>[12]</sup> 软件, 选择k-mer=41将有效数据组装至contig和scaffold级别, 并统计GC含量和覆盖深度等信息。

### 1.3 微卫星位点搜索与引物设计

使用MISA (Microsatellite identification tool) 软件 (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) 在长度大于500 bp的序列中搜索微卫星位点。运行参数为: 重复基序长度1~6 bp, 单碱基重复次数≥12次, 二碱基重复次数≥6次, 三碱基、四碱基重复次数≥5次, 五碱基、六碱基重复次数≥4次; 将间隔区域长度小于100 bp的相邻微卫星归为1个复合型位点。基于微卫星侧翼序列, 用Primer 5<sup>[13]</sup> 批量设计引物, 主要参数为: 引物长度18~27 bp, 扩增产物100~300 bp, 退火温度 ( $T_m$ ) 55~65 °C, GC含量40%~60%, 正、反向引物退火温差≤5 °C; 尽量避免出现发卡结构、二聚体、错配和引物二聚体; 每个微卫星位点生成3~5对候选引物。随机选取50对微卫星引物委托

生工生物工程 (上海) 股份有限公司合成。

### 1.4 多态微卫星标记的验证与筛选

使用天根生化科技 (北京) 有限公司的海洋动物组织基因组DNA提取试剂盒从疣吻沙蚕体壁肌肉组织中提取基因组DNA。PCR反应体系为20 μL, 正、反向引物 (10 μmol·L<sup>-1</sup>) 各0.5 μL, DNA模板 (15 ng·μL<sup>-1</sup>) 1.0 μL, 2×PCR Mix 10 μL, 用超纯水补至20 μL。扩增反应由Bio-Rad MyCycler Thermal Cycler (Bio-Rad, 美国) 完成, 运行程序为: 95 °C预变性5 min; 95 °C变性30 s, 59~60 °C退火30 s, 72 °C延伸30 s, 25个循环; 72 °C延伸5 min。PCR产物先进行1.5% (w) 琼脂糖凝胶电泳分析, 参照预期片段大小筛选出有效扩增引物。有效扩增引物加荧光接头后 (正向引物5'端添加FAM荧光素), 在30尾疣吻沙蚕个体的基因组DNA中扩增以验证其多态性。PCR产物送至上海翼禾应用生物技术有限公司用3730XL测序分析仪 (Applied Biosystems, 美国) 进行毛细管电泳分析, 使用GeneMapper 3.2软件 (Applied Biosystems, 美国) 进行基因分型。

### 1.5 数据分析

使用Excel 2016软件完成微卫星位点分布特征信息的统计分析和图表绘制。微卫星发生频率=含微卫星的序列总数/序列总数×100%; 微卫星出现频率=微卫星总数/序列总数×100%; 微卫星丰度 (个·Mb<sup>-1</sup>)=微卫星总数/序列总长度<sup>[14]</sup>。用GenAlEx 6.5软件<sup>[15]</sup> 计算等位基因数 ( $N_a$ )、有效等位基因数 ( $N_e$ )、观测杂合度 ( $H_o$ )、期望杂合度 ( $H_e$ )、多态信息含量 (PIC)。用Genepop在线软件 (<https://genepop.curtin.edu.au/>) 检验位点间的连锁不平衡及群体的哈迪-温伯格平衡 (Hardy-Weinberg Equilibrium, HWE), 并用Bonferroni法对显著性阈值进行校正。

## 2 结果

### 2.1 疣吻沙蚕的基因组 survey 测序

低深度高通量测序产生的原始数据经质控后共获得57.48 Gb有效数据, 碱基错误率为0.05%, Q20和Q30分别为95.56%和89.70%, 表明基因组测序质量较高 (表1)。对有效数据进行k-mer分析 (k=17), 结果显示在深度为57时出现主峰值 (图1), 总k-mer为44 257 233 158, 排除错误k-mer的误差影响后得到修正的基因组大小为759.53 Mb,

杂合率为 1.41%，重复序列比例为 45.92%。选择 k-mer=41 将有效读段初步组装至 contig 和 scaffold 水平，最终获得的 contig 总长度为 821 637 022

bp，最大长度为 84 713 bp，N50 为 548 bp；scaffold 总长度为 840 375 821 bp，最大长度为 89 326 bp，N50 为 662 bp。

表1 疣吻沙蚕基因组 survey 测序数据统计  
Table 1 Statistics of genomic survey sequencing data of *T. heterochaetus*

测序文库 Sequencing library	原始数据量 Raw base/Gb	有效数据比 Effective rate/%	有效数据量 Clean base/Gb	碱基错误率 Error rate/%	Q20/%	Q30/%	GC 含量 GC content/%
L1	33.12	99.68	33.01	0.04	96.23	90.77	39.13
L2	24.53	99.75	24.47	0.05	94.88	88.63	39.00
总计 Total	57.65	—	57.48	—	—	—	—
均值 Mean	—	99.72	—	0.05	95.56	89.70	39.07

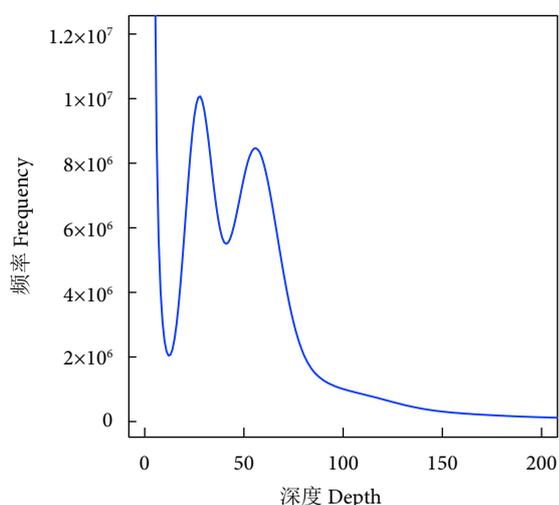


图1 疣吻沙蚕基因组 k-mer 种类频率分布

Fig. 1 Frequency distribution of k-mer species in genome of *T. heterochaetus*

## 2.2 疣吻沙蚕基因组微卫星位点特征

用 MISA 软件在 109 881 条组装序列中检测到 130 216 个微卫星位点，长度共计 2 341 179 bp；微卫星发生频率 5.04%，出现频率 5.97%，分布丰度为 154.9 个·Mb<sup>-1</sup>。疣吻沙蚕微卫星位点以单碱基和二碱基重复最为丰富，分别有 45 582 和 42 298 条，各占 35.00% 和 32.48%；其次是三碱基重复 (18 782 条)，占 14.42%；六碱基数量最少，仅占 2.44% (图 2)。微卫星序列的重复数范围为 4~56 拷贝，主要集中在 4~18 拷贝，重复 19 次及以上的有 3 703 条，仅占 3.50%；单碱基重复以 12~16 次最常见 (40 706 条，89.30%)，二碱基重复集中在 6~18 次 (42 035 条，99.38%)，三碱基重复以 5~12 次为主 (18 526 条，98.64%)，四碱基重复以 5~10 次为主 (12 173 条，98.91%)，而五碱基和六碱基重复主要为 4~8 次 (共 11 176 条，99.37%) (图 2—图 3)。

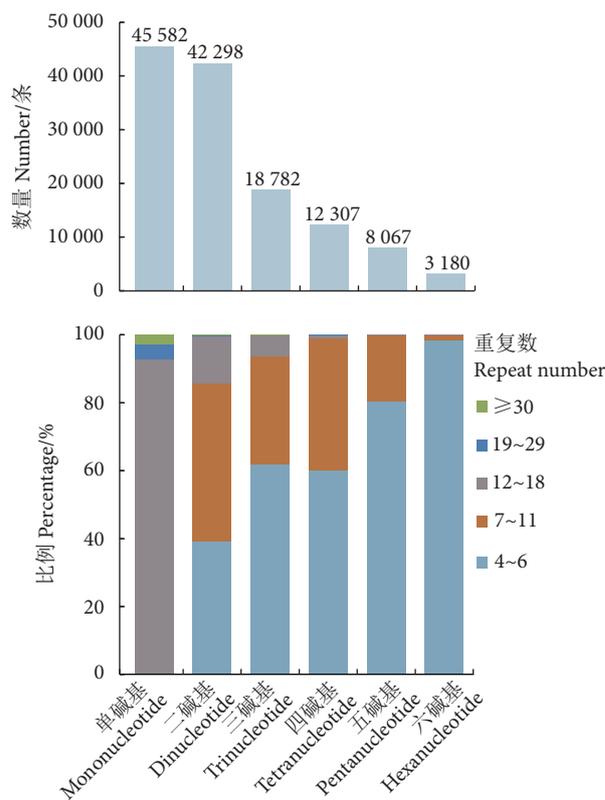


图2 疣吻沙蚕基因组 6 种类型微卫星的数量与比例

Fig. 2 Number and proportion of six motif types of microsatellite loci in genome of *T. heterochaetus*

疣吻沙蚕基因组微卫星共包含 320 种重复基序类型：单碱基 2 种、二碱基 4 种、三碱基 10 种、四碱基 31 种、五碱基 91 种、六碱基 182 种。单碱基重复基序拷贝数集中在 12~15 次，以 C/G 为主，占比 58.02%；二碱基重复拷贝数多为 6~10 次，以 AT/AT 最为丰富，占比 62.38%，其次是 AC/GT (27.37%)、AG/CT (10.22%)，而 CG/CG 数量稀少 (0.02%)；在三碱基重复中，拷贝数多为 5~12 次，占比最高的是 AAT/ATT，有 6 253 条 (33.29%)，其

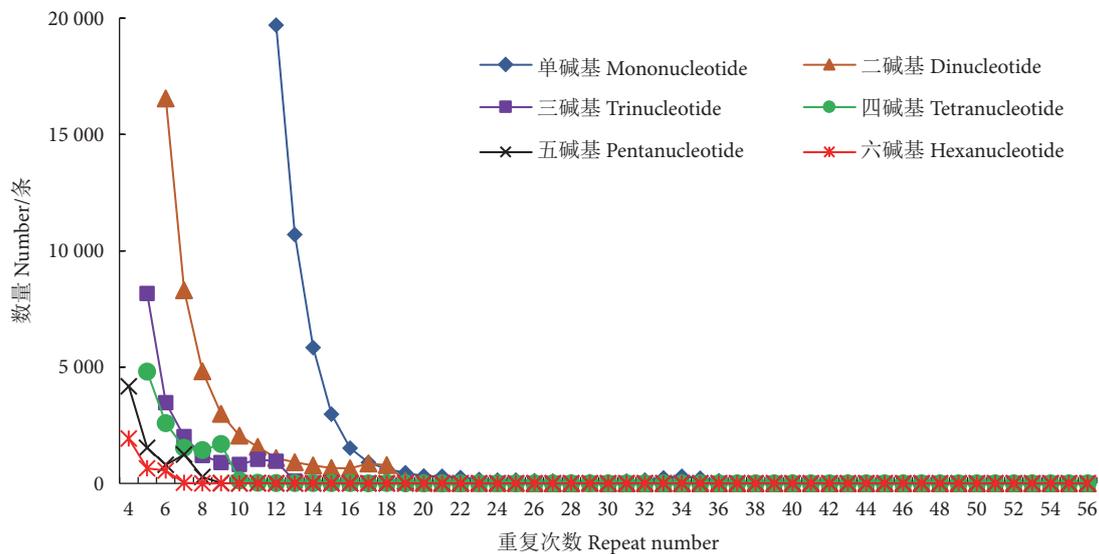


图3 疣吻沙蚕基因组微卫星重复数分布特征

Fig. 3 Distribution pattern of microsatellite repeat number in genome of *T. heterochaetus*

次是 ATC/GAT (32.15%), 而 CCG/CGG 数量最少, 仅占 0.06%; 四碱基拷贝数多为 5~8 次, 优势重复基序是 AAAT/ATTT, 有 3 567 条 (28.98%), 其次是 AATC/GATT (14.08%)、ACTC/GAGT (9.60%)、ATCC/GGAT (7.38%); 五碱基、六碱基拷贝数集中在 4~7 次, 优势重复基序分别是 AAAAT/ATTTT (17.07%) 和 AACCT/AGGGT (11.73%) (图 4)。

### 2.3 疣吻沙蚕基因组微卫星位点长度特征

根据重复序列长度可将微卫星位点分为两类: 一类是长度达 20 bp 及以上的高度多态 I 型, 另一类则是长度介于 12~19 bp 的中度多态 II 型<sup>[16]</sup>。疣吻沙蚕基因组微卫星位点的长度分布区间为 12~336 bp, I 型微卫星位点有 43 240 条, 占 33.21%, 剩余的 II 型位点占 66.79%; 绝大部分 I 型微卫星位点的长度为 20~39 bp, 超过 50 bp 的位点仅占 0.51% (图 5-a)。进一步分析发现, 在 I 型微卫星位点中, 二、三、四、五碱基是最主要的重复类型, 占 85.92%, 在后续多态标记的筛选中有较高的开发价值 (图 5-b)。

### 2.4 疣吻沙蚕基因组微卫星多态标记验证筛选

使用 Primer 5 成功对 37 370 个微卫星位点设计了引物。随机选取 50 个位点合成引物并进行 PCR 验证, 共获得 41 对 (82%) 有效扩增引物, 表明 MISA 软件鉴定的微卫星位点具有较高有效性。多态性筛选结果显示, 15 个 (30%) 微卫星位点的引物表现出稳定且可重复的多态性 (表 2, 图 6)。在 30 尾疣吻沙蚕中, 15 对引物共检测到 87 个等

位基因, 平均等位基因数 5.800, 等位基因频率为 0.060~0.400, 其中 ThGM021 位点的等位基因数最少 ( $N_a=2.000$ ), ThGM004 位点的等位基因数最多 ( $N_a=12.000$ )。  $N_e$  为 1.164~6.713, 平均值为 3.328;  $H_o$  为 0.050~0.879, 平均值为 0.487;  $H_e$  为 0.141~0.789, 平均值为 0.561; PIC 为 0.136~0.776, 平均值为 0.511 (表 3)。在 15 个位点中, 8 个属高度多态性位点 ( $PIC>0.5$ ), 5 个属中度多态性位点 ( $0.25<PIC\leq 0.5$ ), 2 个属低度多态性位点 ( $PIC\leq 0.25$ ), 表明该群体的遗传多样性较丰富。经 Bonferroni 校正后, 有 3 个位点 (ThGM006、ThGM011、ThGM040) 偏离 HWE, 其中 ThGM040 属于高度多态性位点。各位点间无连锁不平衡现象。

## 3 讨论

### 3.1 疣吻沙蚕基因组基本特征

不同物种间基因组的大小和复杂程度有明显差异, 会直接影响到测序策略的选择及基因组的组装效果。因此, 进行全基因组测序前须先评估物种基因组的基本特征。疣吻沙蚕基因组 survey 测序及 k-mer 分析估计其基因组为 759.53 Mb, 远大于水蛭 (*Helobdella robusta*)<sup>[17]</sup> 和宽体金线蛭 (*Whitmania Pigra*)<sup>[18]</sup> 等所有已报道的蛭纲物种, 也较多毛纲的海蠕虫 (*Capitella teleta*)<sup>[17]</sup>、欧文虫 (*Owenia fusiformis*)<sup>[19]</sup>、巨型管虫 (*Riftia pachyptila*)<sup>[20]</sup>、*Lamellibrachia luymesii*<sup>[21]</sup> 的大; 与寡毛纲通俗腔蚓 (*Metaphire vulgaris*)<sup>[22]</sup> 及多毛纲搓稚虫 (*Streblospio benedicti*)<sup>[23]</sup> 的大小 (约 0.7 Gb) 相当; 但明显小于

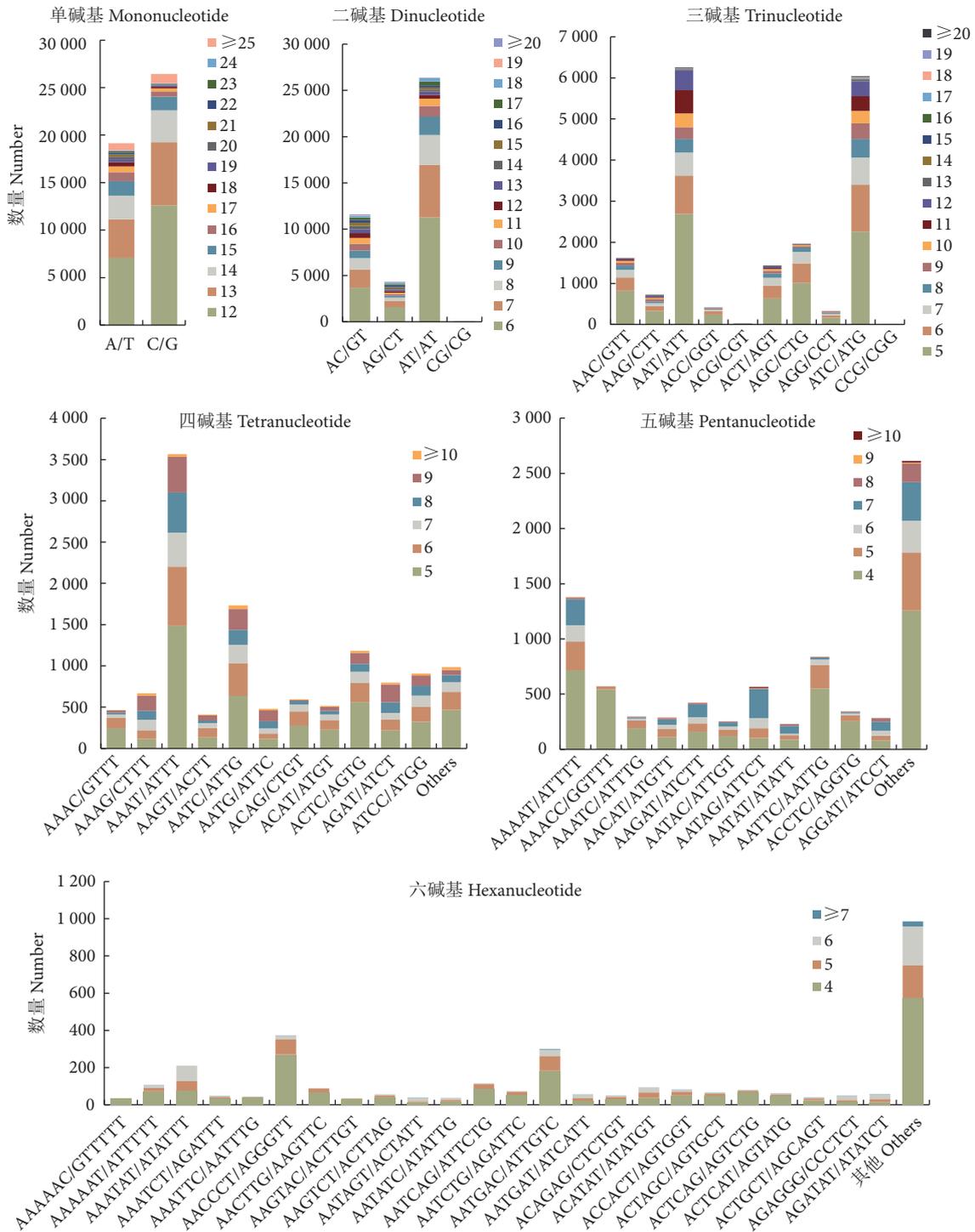


图4 疣吻沙蚕基因组微卫星重复基序类型分布特征

Fig. 4 Distribution pattern of microsatellite motif types in genome of *T. heterochaetus*

安德爱胜蚓 (*Eisenia andrei*)<sup>[24]</sup>、赤子爱胜蚓 (*E. fetida*)<sup>[25]</sup>，以及多毛纲的旋鳃虫 (*Spirobranchus lamarki*)<sup>[26]</sup> 和深海管虫 (*Paraescarpia echinospica*)<sup>[27]</sup> (1.0~1.3 Gb)。物种间基因组大小的差异性在一定程度上与重复序列比例的高低有关。疣吻沙蚕基因组的重复序列比例 (45.92%) 低于旋鳃虫<sup>[26]</sup> 和深海管虫<sup>[27]</sup>，但显著高于海蠕虫<sup>[17]</sup>、巨型管虫<sup>[20]</sup>，以及

所有已公布的蛭纲动物 (重复比例均低于 34%)<sup>[15-18]</sup>；其基因组杂合率 (1.41%) 与安德爱胜蚓<sup>[24]</sup> 和赤子爱胜蚓<sup>[25]</sup> 相当，远远高于 *L. luymesii* (0.60%)<sup>[21]</sup>、搓稚虫 (0.29%)<sup>[23]</sup> 与深海管虫 (0.63%)<sup>[27]</sup>，是目前已报道的杂合率最高的多毛类。此外，组装的 contig 和 scaffolds 总长分别为 821 637 022 和 840 375 821 bp，N50 分别为 548 和 662 bp，提示过高的杂合率

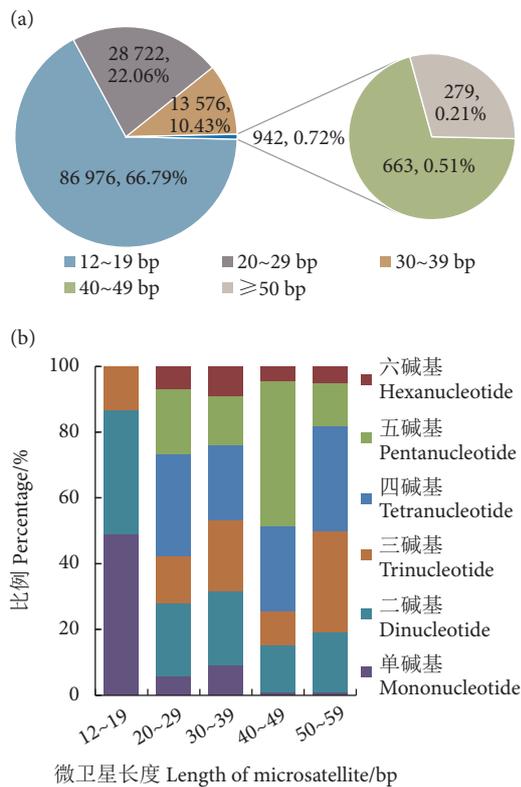


图5 疣吻沙蚕基因组微卫星长度分布特征

注: a. 不同长度区间微卫星数量及比例; b. 不同类型微卫星长度分布特征。

Fig. 5 Distribution pattern of length of microsatellite loci genome of *T. heterochaetus*

Note: a. Number and percentage of microsatellite loci at different length intervals; b. Length distribution of the six motif types of microsatellite loci.

会造成组装序列长度大于预估基因组大小, 并导致显著偏低的 N50 指标<sup>[28]</sup>。综上所述, 疣吻沙蚕基因组属复杂基因组类型, 绘制染色体级别高质量全基因组图谱应优先考虑“PacBio+Illumina+Hi-C”策略。该研究数据为后续全基因组测序与组装提供了基础资料。

### 3.2 疣吻沙蚕基因组微卫星分布特征

水生动物基因组中单碱基重复微卫星占优势的现象较少, 已报道的有中华绒螯蟹 (*Eriocheir sinensis*)<sup>[29]</sup>、鲤 (*Cyprinus carpio*)<sup>[30]</sup>、脊尾白虾 (*Exopalaemon carinicauda*)<sup>[31]</sup> 和胡鲇 (*Clarias batrachus*)<sup>[32]</sup>。在疣吻沙蚕基因组中检测到 130 216 个微卫星位点, 位点数量随重复基序碱基数的增加而迅速减少, 单碱基重复微卫星最丰富 (35.00%), 其次为二碱基重复 (32.48%)、三碱基重复 (14.42%), 这与以二碱基或三碱基为优势类型的鱼<sup>[9,33-35]</sup>、虾<sup>[36-37]</sup>、贝类<sup>[38]</sup> 明显不同, 特别是与亲缘关系较近的蛭类、寡毛类相比也表现出较大差异。宽体金线蛭<sup>[39]</sup>、

天锡杜拉蚓 (*Drawida gisti*)<sup>[40]</sup> 基因组中三碱基重复占绝对优势, 主要基序类型分别是 AAT、ATA 和 ATT、AAT。王斌等<sup>[41]</sup> 对 4 种蛭类的转录组数据进行分析, 同样发现三碱基重复微卫星占优势, 在宽体金线蛭中甚至高达 68.00%。作为最特殊的微卫星重复类型, 三碱基重复微卫星可形成复杂的环-折叠构型来稳定 DNA 结构, 从而更有利于转录过程中的解旋和蛋白质识别<sup>[33]</sup>。此外, 疣吻沙蚕转录组中二碱基 AT/TA 重复基序频率 (32.19%) 最高, 单碱基 A/T 重复类型 (21.17%) 其次, 之后是三碱基 AGC/GCT (29.06%) 基序 (未发表数据), 表明微卫星类型特征在转录组和基因组水平上存在差异, 类似的差异性在圆鳍鱼 (*Cyclopterus lumpus*)<sup>[42]</sup>、凡纳滨对虾<sup>[36,43]</sup>、宽体金线蛭<sup>[39,41]</sup> 中也被证实。不同物种表现出不同的微卫星优势类型分布规律, 可能与其进化水平有关。

疣吻沙蚕基因组微卫星核心区集中在 4~18 拷贝, 随着重复次数的增加, 微卫星数量呈显著降低趋势, 这可能与重复单元长度的不断增加使得微卫星稳定性降低或基序高频次重复导致更高的突变率有关<sup>[44]</sup>。进一步分析发现, 二碱基至六碱基微卫星重复单元中 A/T 含量显著高于 G/C 含量, 即微卫星序列表现出明显的 A/T 碱基优势, 这与许多水生动物如脊尾白虾<sup>[31]</sup>、凡纳滨对虾<sup>[36]</sup>、虾夷扇贝 (*Mizuhopecten yessoensis*)<sup>[38]</sup>、宽体金线蛭<sup>[39]</sup> 及旧金山湾卤虫 (*Artemia franciscana*)<sup>[45]</sup> 中的研究结果一致。早期观点认为, 微卫星富含 A/T 的原因可能是由于 CpG 甲基化后的 C 易脱氨基转变为 T, 导致 G/C 比例不断缩小, 而突变的 A/T 碱基类型相应增多<sup>[33]</sup>; 后来却发现这也可能与微卫星位点的产生方式, 即与 DNA 的复制滑动存在一定关系<sup>[29,46]</sup>。微卫星富含 A/T 的原因可能是 A/T 含量高则  $T_m$  值降低, 其序列容易发生 DNA 解链, 并通过复制滑动机制和重组机制产生高 A/T 含量的重复类型的概率更高<sup>[38]</sup>。

### 3.3 疣吻沙蚕微卫星标记开发及其遗传分析效果

微卫星是重要的分子遗传学研究工具, 开发多态标记是其广泛应用的关键。与传统方法相比, 基于高通量测序技术开发微卫星标记更具优势, 产生的基因组或转录组数据是标记开发的重要资源。然而与转录组微卫星相比, 基因组微卫星的多态性往往更高且分布更广泛, 可获得更优的基因组覆盖

表2 疣吻沙蚕 15 对多态微卫星引物信息  
Table 2 Information of 15 polymorphic microsatellite loci in genome of *T. heterochaetus*

位点 Locus	引物序列 (5'—3') Primer sequence (5'—3')	重复单元 Repeat unit	产物大小 Size/bp	退火温度 Annealing temperature/°C
ThGM004	F: TGCTGCTACTGCTACAGCTACTATG R: CTGACAAAGTTTGGTGGCTG	(TAC) <sub>18</sub>	289	60.0
ThGM006	F: TGAAAATTAGTGTGATTTGTCC R: AGCCAACCAGAACATGAACA	(CA) <sub>11</sub>	260	59.0
ThGM011	F: AACTGGACTAAGGCTATCAAAAA R: CTTGGGGTTCATGCATCATT	(AG) <sub>17</sub>	220	59.0
ThGM015	F: TTGGTTGTTATCCATGCACC R: AGACAGCAGTGAAATAGCACCA	(TAT) <sub>12</sub>	279	59.5
ThGM017	F: ATTCGATAAGCATTCCACCG R: CTTGGTAGCTGGCCTGTCTC	(ATGG) <sub>8</sub>	215	60.0
ThGM021	F: TCGGAAATGAGAAGTGAGCA R: TGCCTGTGTGGAATACCAAG	(TA) <sub>10</sub>	277	60.0
ThGM024	F: ACCTGTCCACCCGTCATTTA R: CCTTTAGGGGATGGCTACAA	(TAT) <sub>14</sub>	294	59.5
ThGM029	F: GAGCAAAATATTCAAGTTGGCA R: TTGTTTGTTCATATCTTCTAAAGAGCA	(ATT) <sub>12</sub>	243	59.0
ThGM033	F: GGAGTGGGGAGGATTTAGC R: CCATGTACAGCATTACAGCCA	(TG) <sub>18</sub>	277	60.0
ThGM035	F: GTAAGGGCAAGGGTTGTGAA R: ACCGTTACCCTAACCCCAAC	(AG) <sub>13</sub>	226	60.0
ThGM038	F: TTACCCTGCCATCCTACCAG R: CTATTCTGCCAGTGGTCGCT	(TG) <sub>20</sub>	157	60.0
ThGM040	F: GGATCCAGAAGGGGTAAAGC R: GTTGGTCATGTTCTGTTGC	(TTA) <sub>11</sub>	239	59.5
ThGM041	F: ACCAGCTGCTAGAGGCAGAC R: TTAGGTCCTCACCCAGGGAT	(ATG) <sub>7</sub>	260	60.0
ThGM043	F: AAAAGCAAGTGGTAACACAAAATG R: CATTGGGCTCTGGGAATAAA	(TCAT) <sub>11</sub>	272	59.5
ThGM047	F: CGACCTGCGGATTTAATTTG R: ATATCTTGGCGCGGATAG	(TGG) <sub>12</sub>	148	60.0

注: F. 正向引物; R. 反向引物。

Note: F. Forward primer; R. Reverse primer.

率<sup>[42]</sup>。Temnykh 等<sup>[16]</sup>认为长度 $\geq 20$  bp 的微卫星位点多态性较高, 长度 10~20 bp 的为中等多态性, 小于 10 bp 的多态性低。在疣吻沙蚕基因组中共筛选到 43 240 条序列长度 $\geq 20$  bp 的 I 型微卫星位点, 占比超过 33%, 远高于其转录组序列中

20% 的 I 型位点比例 (未发表数据)。挑选 50 个微卫星位点进行有效性及多态性验证, 有 41 个 (82%) 位点的引物可扩增出特异性条带, 其中 15 个 (30%) 表现出稳定且可重复的多态性。疣吻沙蚕基因组多态微卫星标记的筛选成功率与卵形鲳鲹<sup>[9]</sup>、凡纳滨

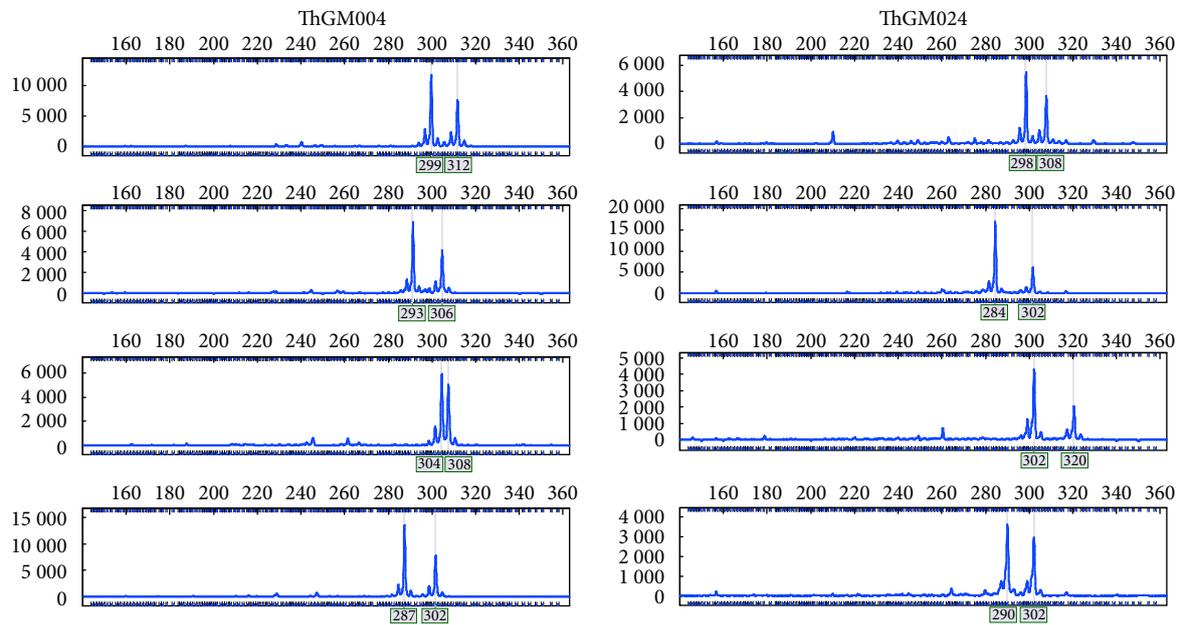


图6 部分多态微卫星标记的毛细管电泳分型结果

Fig. 6 A set of polymorphic microsatellite loci visualized by high-resolution capillary electrophoresis

表3 15 个多态微卫星位点在疣吻沙蚕群体中的遗传特征

Table 3 Genetic characteristics of 15 polymorphic microsatellite loci in a *T. heterochaetus* population

位点 Locus	等位基因数 $N_a$	有效等位基因数 $N_e$	观测杂合度 $H_o$	期望杂合度 $H_e$	多态信息含量 PIC	哈迪-温伯格平衡的 $P$ 值 $P_{HWE}$
ThGM004	12	6.672	0.697	0.775	0.726	0.275
ThGM006	4	1.642	0.367	0.388	0.372	0.001*
ThGM011	5	1.608	0.257	0.349	0.377	0.026*
ThGM015	8	4.933	0.438	0.789	0.776	0.225
ThGM017	3	1.521	0.066	0.271	0.245	0.148
ThGM021	2	1.593	0.367	0.508	0.375	1.000
ThGM024	11	6.713	0.879	0.742	0.682	0.541
ThGM029	10	5.647	0.697	0.658	0.599	0.140
ThGM033	5	3.102	0.167	0.772	0.720	0.069
ThGM035	3	2.164	0.050	0.141	0.136	0.086
ThGM038	8	4.878	0.576	0.545	0.489	0.221
ThGM040	4	2.441	0.733	0.718	0.652	0.008*
ThGM041	3	2.155	0.417	0.431	0.336	0.503
ThGM043	5	2.727	0.724	0.682	0.617	0.148
ThGM047	4	2.224	0.867	0.642	0.569	0.267
均值 Mean	5.800	3.328	0.487	0.561	0.511	—

注: \*. Bonferroni 法校正后显著偏离哈迪-温伯格平衡 ( $P < 0.05$ );  $n = 30$ 。

Note: \*. Significant departure from Hardy-Weinberg equilibrium after Bonferroni's correction ( $P < 0.05$ );  $n = 30$ .

对虾<sup>[36]</sup>、宽体金线蛭<sup>[39]</sup>中的结果相当。由此可见, 疣吻沙蚕微卫星标记筛选成功率较高, 获得的微卫星位点具有良好的开发潜力, 是丰富且可靠的标记资源。

用微卫星标记开展群体遗传分析时, 一般认为有效等位基因数越接近观测等位基因数, 群体等位基因分布就越均匀; 然而在实际分析中, 通常会将全部条带视为有效等位基因, 无效等位基因过剩

便会造成等位基因分布不均<sup>[47]</sup>。从疣吻沙蚕基因组筛选的 15 个多态微卫星位点中, 仅 3 个位点(ThGM021、ThGM035、ThGM041)的等位基因数接近有效等位基因数, 提示大部分位的等位基因分布不均, 这可能是因为验证群体的样本容量偏小引起了主效等位基因的缺失<sup>[9]</sup>。有研究表明, 群体的最小样本容量一般与遗传分析选用的参数有关。利用  $H_e$ 、PIC 以及香农指数等评估遗传多样性时, 群体样本容量达到 27 便能使分析结果接近总体水平的 95%; 但选择  $N_a$  时, 样本容量则需达到 52 以上<sup>[48]</sup>。疣吻沙蚕的群体样本量为 30, 故而选择  $H_e$ 、PIC 可更好地评价微卫星位点的多态性。15 个位点的平均  $H_o$  (0.487)、平均  $H_e$  (0.561) 和平均 PIC (0.511) 共同表明, 疣吻沙蚕验证群体具有较丰富的遗传多样性, 这与 Chen 等<sup>[5]</sup>利用 COI 标记的分析结果并不完全一致, 可能是因为微卫星属于核标记, 多态性更高且受选择性作用更小, 其揭示遗传变异的灵敏度高于线粒体标记。通常认为, PIC 越接近 1, 则群体杂合个体的比例越大、多态性越高: PIC < 0.25 为低度多态性, 0.25 ≤ PIC < 0.5 为中度多态性, PIC ≥ 0.5 为高度多态性<sup>[47]</sup>。在 15 个多态微卫星位点中, 仅 2 个为低度多态性, 其余 13 个 (86.7%) 均为高度或中度多态性。因此, 基于疣吻沙蚕基因组序列筛选的微卫星标记多态性高且遗传信息丰富, 可为群体遗传分析、种质资源评价、分子育种研究提供优质工具。

#### 4 结论

疣吻沙蚕基因组属于高杂合、高重复的复杂基因组, 其微卫星位点的类型丰富且具备较好的多态性潜能, 可作为有效资源用于微卫星标记的大规模开发, 对种质资源评价与保护利用、种群遗传学以及分子育种研究具有实际价值。

#### 参考文献:

[1] YANG Z Q, SUNIL C, JAYACHANDRAN M, et al. Anti-fatigue effect of aqueous extract of Hechong (*Tylorrhynchus heterochaetus*) via AMPK linked pathway[J]. Food Chem Toxicol, 2020, 135: 111043.

[2] 苏跃朋, 黄敏, 崔阔鹏. 珠江河口区禾虫产业技术现状及增养殖效益分析 [J]. 海洋与渔业, 2016(10): 64-67.

[3] ZHANG W X, WANG Z X, GANESAN K, et al. Antioxidant activities of aqueous extracts and protein hydrolysates from marine worm Hechong (*Tylorrhynchus heterochaeta*)[J]. Foods, 2022,

11(13): 1837.

[4] 杨尉, 陈兴汉. 疣吻沙蚕-水稻生态复合种养技术要点及效益分析 [J]. 南方农业, 2022, 16(20): 17-20, 24.

[5] CHEN X H, YANG S, YANG W, et al. First genetic assessment of brackish water polychaete *Tylorrhynchus heterochaetus*: mitochondrial COI sequences reveal strong genetic differentiation and population expansion in samples collected from southeast China and north Vietnam[J]. Zool Res, 2020, 41(1): 61-69.

[6] CHEN X H, LI M M, LIU H P, et al. Mitochondrial genome of the polychaete *Tylorrhynchus heterochaetus* (Phyllodocida, Nereididae)[J]. Mitochondrial DNA A, 2016, 27(5): 3372-3373.

[7] CHEN H, LI X, WANG Y, et al. *De novo* transcriptomic characterization enables novel microsatellite identification and marker development in *Betta splendens*[J]. Life, 2021, 11(8): 803.

[8] 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.

[9] 张永德, 文露婷, 罗洪林, 等. 卵形鲳鲹基因组调研及其 SSR 分子标记的开发应用 [J]. 南方农业学报, 2020, 51(5): 983-994.

[10] 上官清, 陈昆慈, 刘海洋, 等. 斑鳢基因组中微卫星分布特征及野生种群遗传结构分析 [J]. 南方水产科学, 2020, 16(3): 47-60.

[11] LIU B H, SHI Y J, YUAN J Y, et al. Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects[J]. Quant Biol, 2013, 35(s1-3): 62-67.

[12] LUO R B, LIU B H, XIE Y L, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler[J]. GigaScience, 2012, 1(1): 18.

[13] LALITHA S. Primer premier 5[J]. Biotech Softw Internet Rep, 2000, 1(6): 270-272.

[14] 刘玉萍, 王棋, 黄新蕊, 等. 基于高通量测序的带鱼肌肉组织转录组微卫星信息分析 [J]. 南方农业学报, 2022, 53(3): 725-734.

[15] PEAKALL R, SMOUSE P E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research: an update[J]. Bioinformatics, 2012, 28(19): 2537-2539.

[16] TEMNYKH S, DECLERCK G, LUKASHOVA A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential[J]. Genome Res, 2001, 11(8): 1441-1452.

[17] SIMAKOV O, MARLETAZ F, CHO S J, et al. Insights into bilaterian evolution from three spiralian genomes[J]. Nature, 2013, 493(7433): 526-531.

[18] TONG L, DAI S X, KONG D J, et al. The genome of medicinal leech (*Whitmania pigra*) and comparative genomic study for exploration of bioactive ingredients[J]. BMC Genom, 2022, 23(1): 76.

[19] MARTÍN-ZAMORA F M, LIANG Y, GUYNES K, et al. Annelid functional genomics reveal the origins of bilaterian life cycles[J]. Nature, 2023, 615(7950): 105-110.

[20] de OLIVEIRA A L, MITCHELL J, GIRGUIS P, et al. Novel insights on obligate symbiont lifestyle and adaptation to chemosyn-

- thetic environment as revealed by the giant tubeworm genome[J]. *Mol Biol Evol*, 2022, 39(1): msab347.
- [21] LI Y, TASSIA M G, WAITS D S, et al. Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm *Lamellibrachia luymesii*[J]. *BMC Biol*, 2019, 17(1): 91.
- [22] JIN F, ZHOU Z L, GUO Q, et al. High-quality genome assembly of *Metaphire vulgaris*[J]. *PeerJ*, 2020, 8: e10313.
- [23] ZAKAS C, HARRY N D, SCHOLL E H, et al. The genome of the poecilognous Annelid *Streblospio benedicti*[J]. *Genome Biol Evol*, 2022, 14(2): evac008.
- [24] SHAO Y, WANG X B, ZHANG J J, et al. Genome and single-cell RNA-sequencing of the earthworm *Eisenia andrei* identifies cellular mechanisms underlying regeneration[J]. *Nat Commun*, 2020, 11(1): 2656.
- [25] ZWARYCZ A S, NOSSA C W, PUTNAM N H, et al. Timing and scope of genomic expansion within Annelida: evidence from homeoboxes in the genome of the earthworm *Eisenia fetida*[J]. *Genome Biol Evol*, 2016, 8(1): 271-281.
- [26] KENNY N J, NAMIGAI E K O, MARLÉTAZ F, et al. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patellogastropoda) and *Spirobranchus (Pomatoceros) lamarcki* (Annelida, Serpulida)[J]. *Mar Genom*, 2015, 24(2): 139-146.
- [27] SUN Y N, SUN J, YANG Y, et al. Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*[J]. *Mol Biol Evol*, 2021, 38(10): 4116-4134.
- [28] 高胜寒, 禹海英, 吴双阳, 等. 复杂基因组测序技术研究进展 [J]. *遗传*, 2018, 40(11): 944-963.
- [29] 徐杰杰, 毕宜慧, 程景颢, 等. 中华绒螯蟹 (*Eriocheir sinensis*) 全基因组微卫星分布特征研究 [J]. *基因组学与应用生物学*, 2021, 40(Z2): 2422-2429.
- [30] 梁霞, 王慧琪, 马宇璇, 等. 鲤鱼 (*Cyprinus carpio*) 全基因组微卫星分布特征研究 [J]. *南京师大学报 (自然科学版)*, 2021, 44(3): 103-111.
- [31] ZHANG Q, ZHANG C S, YU Y, et al. Characteristic analysis of simple sequence repeats in the ridgetail white prawn *Exopalaemon carinicauda* genome and its application in parentage assignment[J]. *J World Aquacult Soc*, 2020, 51(3): 690-701.
- [32] SRIVASTAVA S, KUSHWAHA B, PRAKASH J, et al. Development and characterization of genic SSR markers from low depth genome sequence of *Clarias batrachus* (Magur)[J]. *J Genet*, 2016, 95(3): 603-609.
- [33] 彭冶, 李杰, 王涛, 等. 瓦氏黄颡鱼全基因组微卫星的分布特征及其定位的初步研究 [J]. *南方水产科学*, 2022, 18(1): 90-98.
- [34] XU S Y, SONG N, XIAO S J, et al. Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus* [J]. *Biosci Rep*, 2020, 40(2): BSR20192252.
- [35] 王九龙, 李洪莉, 尹硕, 等. 绿鳍马面鲀全基因组微卫星分布特征 [J]. *烟台大学学报 (自然科学与工程版)*, 2022, 35(3): 285-293.
- [36] 王佳佳, 王琼, 秦桢, 等. 凡纳滨对虾全基因组 SSR 标记开发及不同养殖群体的遗传多样性分析 [J]. *水产学报*, 2023, 47(6): 64-74.
- [37] SUN J X, PENG G H, XIONG L J, et al. Genome-wide SSR marker development and application in genetic diversity analysis of the red swamp crayfish, *Procambarus clarkii* (Girard, 1852) in China [J]. *Crustaceana*, 2021, 94(2): 189-205.
- [38] 倪守胜, 杨钰, 柳淑芳, 等. 基于高通量测序的虾夷扇贝基因组微卫星特征分析 [J]. *渔业科学进展*, 2018, 39(1): 107-113.
- [39] 熊良伟, 王帅兵, 岳丽佳, 等. 宽体金线蛭基因组 SSR 序列特征分析及其分子标记开发 [J]. *南方农业学报*, 2018, 49(11): 2298-2303.
- [40] LIU H Y, ZHANG Y F, WANG G B, et al. Development and characterization of microsatellite markers in the earthworm *Drawida gisti* Michaelsen, 1931 and cross-amplification in two other congeners[J]. *Mol Biol Rep*, 2020, 47(10): 8265-8269.
- [41] 王斌, 孙静, 刘凌云, 等. 蛭类转录组中 EST-SSR 分析及抗凝血相关分子标记的挖掘 [J]. *中草药*, 2017, 48(1): 172-178.
- [42] MADUNA S N, VIVIAN-SMITH A, JÓNSDÓTTIR Ó D B, et al. Genome- and transcriptome-derived microsatellite loci in lumpfish *Cyclopterus lumpus*: molecular tools for aquaculture, conservation and fisheries management[J]. *Sci Rep*, 2020, 10(1): 559.
- [43] 李强勇, 李旻, 曾地刚, 等. 凡纳滨对虾微卫星分子标记的开发及不同养殖家系遗传多态性分析 [J]. *南方农业学报*, 2020, 51(2): 429-436.
- [44] WIERDL M, DOMINSKA M, PETES T D. Microsatellite instability in yeast: dependence on the length of the microsatellite[J]. *Genetics*, 1997, 146(3): 769-779.
- [45] JO E, LEE S J, CHOI E, et al. Whole genome survey and microsatellite motif identification of *Artemia franciscana*[J]. *Biosci Rep*, 2021, 41(3): BSR20203868.
- [46] SCHLÖTTERER C, TAUTZ D. Slippage synthesis of simple sequence DNA[J]. *Nucleic Acids Res*, 1992, 20(2): 211-215.
- [47] 马军, 刘嘉鑫, 江智景, 等. 基于 RNA-seq 数据的密斑刺鲃 SSR 分子标记开发及鉴定 [J]. *南方水产科学*, 2020, 16(1): 127-136.
- [48] 朱维岳, 周桃英, 钟明, 等. 基于遗传多样性和空间遗传结构的野生大豆居群采样策略 [J]. *复旦学报 (自然科学版)*, 2006, 45(3): 321-327.