

DOI: 10.12264/JFSC2021-0426

绒杜父鱼全基因组 survey 分析及微卫星分布特征

赵蕊蕊, 徐胜勇

浙江海洋大学水产学院, 浙江 舟山 316022

摘要: 绒杜父鱼(*Hemitripterus villosus*)为西北太平洋近海重要的经济鱼类, 近年来其渔业资源呈现过度开发及衰退的趋势。为评估绒杜父鱼基因组基本信息, 开发基因组范围内遗传标记, 以期为资源管理和保护工作提供参考, 本研究对绒杜父鱼开展全基因组 survey 分析, k-mer 分析表明绒杜父鱼基因组大小约为 713.18 Mb, 杂合率为 0.26%, 重复序列的比例为 38.61%。基因组初步组装结果显示, Contig N50 和 Scaffold N50 分别为 7433 bp、19388 bp。在整个基因组范围内, 总共检测到 583498 个微卫星位点, 相对丰度为 1010 个/Mb。其中二碱基重复比例最高(55.61%), 单碱基重复次之(33.39%)。在二、三碱基重复中主要是 AC/GT 和 AGG/CCT 类型, 微卫星重复拷贝数主要集中在 5~17 次。研究结果表明, 绒杜父鱼基因组为简单基因组, 可以用“ Illumina+PacBio+Hi-C”的测序策略组装出高质量全基因组序列, 而经过筛选的微卫星位点也能为后续的遗传学研究提供有效的分子标记。本研究结果可以为绒杜父鱼进化生物学和遗传学研究提供基础资料。

关键词: 绒杜父鱼; 高通量测序技术; 全基因组 survey; 微卫星标记

中图分类号: S931

文献标志码: A

文章编号: 1005-8737-(2022)07-0994-08

高通量测序技术(high-throughput sequencing, HTS)的发展为大规模基因组学研究提供了技术先导^[1], 使得鱼类基因组学研究从斑马鱼等模式物种不断深入^[2], 目前已报道了多达 50 余种鱼类^[3]的全基因组测序结果。基于 Illumina 测序技术的全基因组 survey 分析, 即基因组大小和复杂程度的研究, 是通过对没有参考基因组序列的物种进行小片段低深度测序, 基于 Lander-Waterman 模型进行 k-mer 分析, 根据 k-mer 频率和深度的统计结果, 在较低的测序成本下预估基因组的大小、杂合度及重复序列等信息, 得出的结果可进一步指导后续的建库策略和测序数据量, 为高质量的全基因组组装提供基础资料^[3-8]。在海洋鱼类中, 张永德等^[5]对卵形鲳鲹(*Trachinotus ovatus*)基因组学进行了初步研究, 预估基因组大小为 642.68 Mb, 杂合率和重复序列比例分别为 0.31%、30.19%, 为

其高质量全基因组组装工作提供基础资料; Xu 等^[6]对褐斑鮗(*Platycephalus* sp.1)雌雄个体开展全基因组 survey 分析, 测得雌雄个体基因组大小, 并推断褐斑鮗可能为雄异配型(XX/XY), 为后续比较基因组学和进化生物学研究提供科学资料。同时, 全基因组 survey 分析也可在全基因组范围内识别并开发微卫星标记(microsatellite markers), 微卫星标记已广泛应用于种群遗传学分析、遗传图谱构建、亲缘关系鉴定等^[9]研究领域。相对于传统微卫星标记的开发流程, 高通量测序数据下的微卫星标记开发工作具有低成本、高效率、高产出等优点, 相关研究已在动植物中广泛开展, 例如王耀蝶等^[10]利用全基因组 survey 数据识别出 390967 个微卫星标记, 为金钱鱼(*Scatophagus argus*)种质资源管理、种群遗传学及保护生物学提供大量的数据支撑; Zhou 等^[11]从盾叶薯蓣(*Dios-*

收稿日期: 2021-09-08; 修订日期: 2022-01-06。

基金项目: 国家重点研发计划项目(2017YFA0604902); 舟山市科技局专项(2019C21027)。

作者简介: 赵蕊蕊(1997-), 女, 硕士研究生, 研究方向为渔业资源养护与利用. E-mail: 3245828525@qq.com

通信作者: 徐胜勇, 副研究员, 研究方向为渔业资源生物学. E-mail: kevin890223@163.com

corea zingiberensis)基因组中检测到 419372 个微卫星位点, 为遗传图谱构建工作提供数据和思路。

绒杜父鱼(*Hemitripterus villosus*)隶属于鲈形目(Perciformes)、杜父鱼亚目(Cottoidei)、八角鱼科(Agonidae)、绒杜父鱼亚科(Hemitripteridae), 属于底栖肉食性鱼类, 主要分布在西北太平洋近海海域^[12]。因其肉质紧实鲜美、鱼子营养价值高^[13], 成为中国、韩国、俄罗斯等近海重要经济鱼类^[12,14]。随着鲟、鳇类自然资源的日渐衰退^[15], 绒杜父鱼可作为重要的替代鱼种用于鱼子酱产业^[16]。然而, 受近海海域过度捕捞、生态环境恶化等的影响^[17-18], 绒杜父鱼资源量也呈现衰退趋势^[16], 因此, 亟需开展绒杜父鱼种质管理和资源保护工作。目前, 俄罗斯、韩国等国已陆续开展了有关绒杜父鱼的研究, 而中国的相关研究较少, 且研究内容以生理生态^[19]及生物学^[20-23]为主, 分子遗传学数据和信息的缺乏^[14,24]严重限制了绒杜父鱼进化生物学和种质资源保护工作的开展, 同时对其渔业资源管理工作也具有一定的影响。本研究利用高通量测序技术对绒杜父鱼开展全基因组 survey 分析, 评估其基因组大小、杂合度及重复序列比例等基本信息, 为绒杜父鱼进化和保护生物学研究提供基础资料; 同时基于高通量测序数据在全基因组范围内识别微卫星标记, 为绒杜父鱼遗传学和种质资源研究提供合适的分子标记。

1 材料与方法

1.1 样本采集

本研究所用绒杜父鱼样品于 2020 年 3 月采自浙江舟山近海海域。采用形态学和 DNA 条形码两种方法进行物种鉴定。形态学指标(如鳍条数、侧线鳞数等)及形态学描述参考《中国海洋鱼类》^[25]; 选择线粒体 12S ribosomal RNA (12S rRNA)片段为 DNA 条形码^[26], 通过与 National Center for Biotechnology Information (NCBI)数据库中相关序列信息进行 BLAST 分析, 将覆盖度(query cover)和一致性(percent identity)均大于 99.0% 作为物种鉴定的主要依据。经形态学和 DNA 条形码鉴定后, 取肌肉组织用 95%乙醇溶液固定后, 存于 -80 °C超低温冰箱中待用。

1.2 高通量测序

使用苯酚-氯仿法提取绒杜父鱼基因组 DNA。使用 Covaris 超声波破碎仪对基因组 DNA 进行片段化处理, 筛选合适长度的 DNA 片段, 经末端修复、添加测序接头后构建 350 bp 的 DNA 文库, 之后使用 Illumina HiSeq4000 测序平台进行双末端(paired-end)的测序。高通量测序文库构建和测序工作均委托北京诺禾致源科技股份有限公司进行。对原始下机数据(raw data)进行数据质控和过滤后, 最终得到有效数据(clean data)用于后续生物信息学分析。

1.3 数据分析

使用 GCE V1.0.0^[27]软件进行 k-mer 分析, 基于 k-mer 分析的结果, 估算基因组的大小并评估杂合率和重复序列比例。使用 SOAPdenovo v2.01^[28]软件将 clean data 拼接组装至 Contig 和 Scaffold 级别($k=41$), 并统计组装结果信息。

使用 MISA 脚本 (MISA, <http://pgrc.ipk-gatersleben.de/misa/>)对组装好的基因组序列进行微卫星位点筛选, 统计 1~6 碱基重复的数量、频率等信息, 在设置参数时分别设定 10、6、5、5、5、5 为 1~6 碱基重复最小值。使用 Primer3 软件^[29]进行引物设计, 并随机选取 10 对引物, 经引物合成、PCR 扩增后, 利用琼脂糖凝胶电泳验证微卫星位点和对应引物的有效性。

2 结果与分析

2.1 测序数据量统计

本研究采用 NGS 技术进行测序, 对该物种插入片段为 350 bp 的文库进行基因组测序, 共获得原始数据(raw data) 52.36 Gb, 过滤之后共获得有效数据(clean data) 52.04 Gb, GC 含量为 43.13%。测序数据错误率为 0.04%, 测序质量值 Q20 和 Q30 分别为 95.99% 和 90.85%, 表明测序质量较高, 满足后续基因组 survey 分析要求。

2.2 17-mer 分析

对 clean data 进行 k-mer 分析以评估绒杜父鱼基因组大小及其杂合率和重复比例。选取 k-mer 的大小为 17, 通过统计分析, 预估绒杜父鱼基因组大小为 728.52 Mb, 排除错误 k-mer 的误差影响

后, 得到修正的基因组大小为 713.18 Mb, 基因组杂合率为 0.26%, 重复序列的比例为 38.61%。

2.3 基因组初步组装

使用 SOAPdenovo 软件对绒杜父鱼全基因组进行初步组装。选择 k-mer=41 将基因组组装至 Contig 和 Scaffold 级别。绒杜父鱼基因组初步组装结果见表 1。最终组装的绒杜父鱼基因组全长为 577 386 707 bp, 约为预估基因组大小的 80.96%。其中 Contig N50 和 Scaffold N50 的长度分别是 7433 bp 和 19 388 bp, Scaffold 长度大于 100 bp 以及大于 2 kb 的序列数量分别为 205 339 和 42 158。初步组装的绒杜父鱼基因组 GC 含量为 43.49%。

表 1 绒杜父鱼基因组初步组装结果

Tab. 1 Preliminary assembly results of the genome of *Hemitripterus villosus*

k-mer	Contigs		Scaffolds	
	长度/bp size	数目 number	长度/bp size	数目 number
N90	953	101087	1701	46063
N50	7433	19686	19388	7394
序列总长 total size	568984670		577386707	
序列条数 total number		311514		205339
序列条数(≥ 100 bp) total number (≥ 100 bp)		293307		205339
序列条数(≥ 2 kp) total number (≥ 2 kp)		65160		42158

2.4 基因组微卫星位点分析

利用 MISA 软件对绒杜父鱼初步组装的基因组序列开展微卫星标记筛选和识别。结果共检测到 583498 个微卫星位点, 检测到的微卫星位点分布在 97637 条序列中, 其微卫星相对丰度估计为 1010/Mb 个微卫星位点。对各类型微卫星位点进行统计, 结果显示 33.39% (194850/583498) 的微卫星标记为单碱基, 55.61% (324502/583498) 为二碱基, 8.27% (48264/583498) 为三碱基, 2.19% (12800/583498) 为四碱基, 0.30% (1769/583498) 为五碱基, 0.23% (1313/583498) 为六碱基。其中二碱基重复的数量是最高的, 六碱基重复的数量是最

低的, 整体呈现出重复位点的频率与重复基序长度成反比的结果。这与 Xu 等^[8]、Srivastava 等^[30]的研究结果相一致, 均显示微卫星标记以二碱基重复为主。绒杜父鱼基因组中的微卫星重复次数的变化范围为 5~70 次, 其中二碱基重复 AC/GT 重复次数范围最广(6~70 次), 重复次数在 5~17 处(微卫星数量>20000)较为集中(图 1), 占所检测到的微卫星位点的比例约为 88.41%。6 次重复的微卫星数量最多, 为 75671, 其次为 10 次重复, 为 75334 (图 1b)。绒杜父鱼微卫星分布情况总体呈现随重复次数的增加微卫星的数目下降的趋势, 这与 Chen 等^[31]的综述结论相一致。在绒杜父鱼基因组中, 同种碱基基序中不同的微卫星位点也有很大差别(图 2)。考虑到五碱基和六碱基重复微卫星数量相对较少, 作者对单碱基重复数量大于 5000 (重复次数为 10~17 次)、二碱基重复数量大于 5000 (重复次数为 6~17 次)、三碱基重复数量大于 1000 (重复次数为 5~12 次)、四碱基重复数量大于 1000 (重复次数为 5~9 次), 结果显示在单碱基重复中, A/T 重复单元频率较 C/G 重复单元高得多(图 2a); 二碱基重复中, AC/GT 所占比例最高, 而 CG/CG 比例最低(图 2b); 三碱基重复中, AGG/CCT 数量最多, 其次为 AAT/ATT, 而 ACG/CGT 数量最少(图 2c); 四碱基重复中, ACAG/CTGT、ACAT/ATGT、ACGC/CGTG、AGTA/ATCT、AGGG/CCCT 和 ATCC/ATGG 的数量均大于 1000, 而 AACG/CGTT、AAGC/CTTG、ATCG/ATCG、CCGG/CCGG 数量较少, 均小于 10 (图 2d)。为验证识别的微卫星位点的有效性, 我们随机选取了 10 个微卫星位点进行 PCR 扩增和电泳检测, 琼脂糖凝胶电泳结果显示 PCR 扩增成功率 90%, 表明 MISA 软件识别的微卫星位点具有一定的有效性, 然而微卫星位点的多态性仍需通过进一步群体分析确定。

3 讨论

3.1 绒杜父鱼基因组的基本特征

本研究首次采用全基因组 survey, 对绒杜父鱼全基因组信息进行评估, 同时在基因组范围内识别出大量微卫星位点, 研究结果可为绒杜父鱼

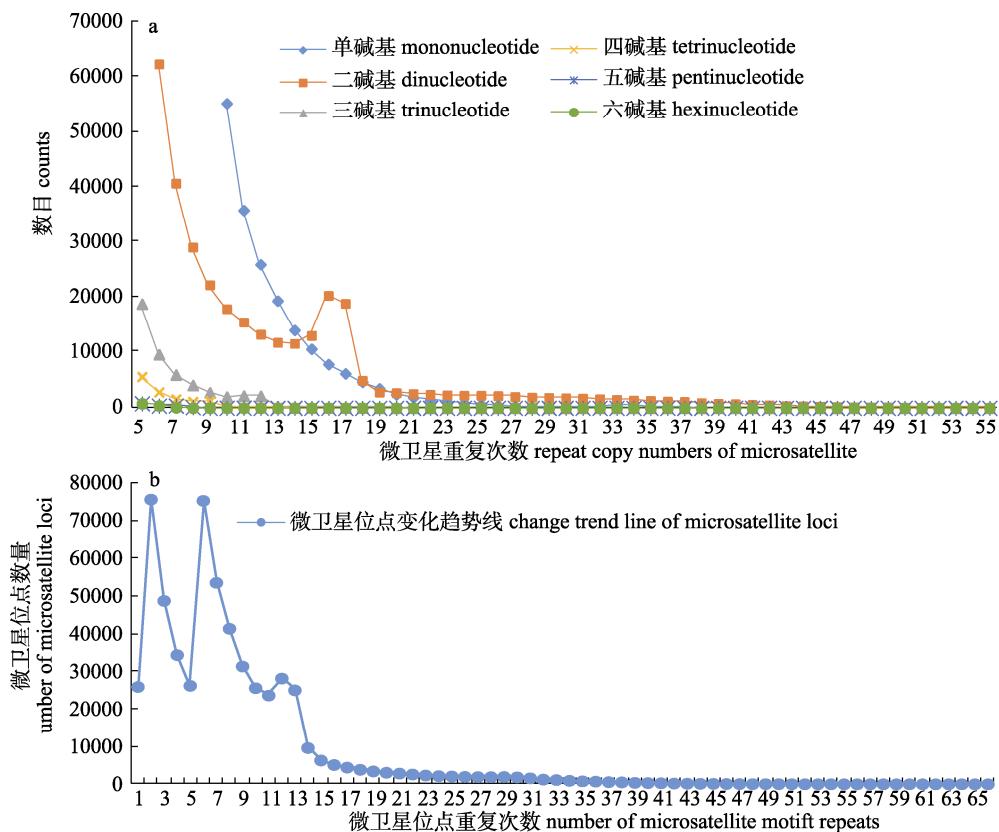


图 1 绒杜父鱼基因组各类型碱基微卫星重复次数变化曲线

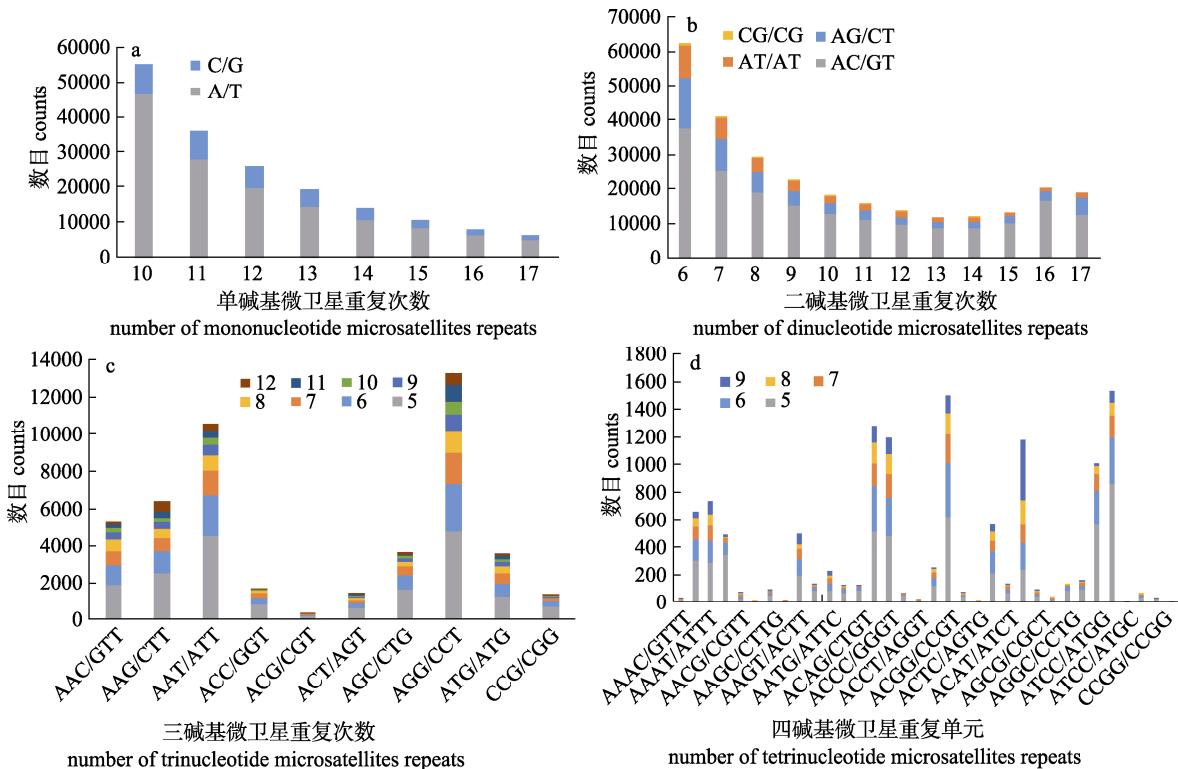
Fig. 1 Repeated frequency statistics of different types of microsatellites in the genome of *Hemitripterus villosus*

图 2 绒杜父鱼基因组 1-4 碱基微卫星重复次数统计

Fig. 2 Repeated frequency of 1-4 types of microsatellites in the genome of *Hemitripterus villosus*

进化生物学、保护遗传学和渔业资源管理等工作提供参考资料。k-mer 分析结果表明, 绒杜父鱼基因组大小为 713.18 Mb, 物种的差异造成不同物种相关基因组信息的不同, 相较于已报道的近缘种海洋鱼类基因组大小, 绒杜父鱼基因组较褐菖鲉(*Sebastiscus marmoratus*)^[8](812.86 Mb)、许氏平鲉(*Sebastes schlegelii*)(846.36 Mb)、朝鲜平鲉(*Sebastes koreanus*)(832.53 Mb)、金斑平鲉(*Sebastes nudus*)^[32](813.12 Mb)的基因组较小, 但与褐斑鲬(*Platycephalus sp.1*)^[6](雌: 684.15 Mb; 雄: 674.96 Mb)、松江鲈(*Trachidermus fasciatus*)^[33](652 Mb)相比, 基因组较大, 这可能与以上海洋鱼类基因组重复序列比例有关。相比较而言, 绒杜父鱼基因组的重复序列比例为 38.61%, 较褐斑鲬^[6](雌: 29.96%; 雄: 29.23%)高, 但均低于菖鲉属及平鲉属鱼类(重复序列比例均大于 39.65%)^[6,8,32]。此外, 绒杜父鱼杂合率为 0.26%, 其杂合度相较于褐斑鲬^[6](雌: 0.76%, 雄: 0.77%)和金斑平鲉(0.31%)较低, 但与许氏平鲉(0.22%)、朝鲜平鲉^[32](0.20%)相当。基因组 survey 评估结果表明, 绒杜父鱼基因组应属于简单基因组类型。本研究基因组 survey 分析结果可以为后续的绒杜父鱼高质量基因组组装工作提供基础资料。

绒杜父鱼的初步组装结果表明, Contig N50 和 Scaffold N50 分别为 7433 bp、19388 bp。组装效果与朝鲜平鲉(Contig N50 7261 bp, Scaffold N50 16255 bp)相当^[32], 与其他近缘种基因组初步组装结果相比较好^[6,8,32], 尤其较褐菖鲉和褐斑鲬的组装效果好(Contig N50<1790 bp, Scaffold N50<4362 bp)^[6,8]。但与其他已报道的硬骨鱼类基因组组装效果^[3-4,7]相比较, 绒杜父鱼基因组初步组装的 N50 长度较短, 整体序列组装的连续性较差, 尚不能满足其进化生物学研究工作的要求, 建议后续采用“ Illumina+PacBio+Hi-C”组合技术构建染色体水平的全基因组序列。然而, 本研究获得的组装序列仍可用于开发基因组范围内 SNP 位点及超保守序列元件(ultraconserved element, UCE), 以开展杜父鱼科鱼类系统基因组学研究^[34]。

3.2 绒杜父鱼的 SSR 分布特征

本研究获得绒杜父鱼基因组中微卫星的相对

丰度为 1010 个/Mb, 与已报道的金钱鱼(*Scatophagus argus*)^[10](653 个/Mb)、中华绒螯蟹(*Eriocheir sinensis*)^[35](572 个/Mb)、褐菖鲉^[8](314.6 个/Mb)、卵形鲳鲹^[5](295.8 个/Mb)等研究相比较高, 考虑到以上研究中未统计单碱基重复微卫星的数量, 笔者推断绒杜父鱼较高的微卫星相对丰度可能与单碱基重复数量有关, 进一步去除单碱基重复数量(194 850 个)后, 统计得到微卫星相对丰度约为 673 个/Mb, 这与金钱鱼基因组微卫星丰度^[10]相当, 但仍较褐菖鲉等鱼类高得多。较高的微卫星相对丰度可能与基因组重复序列比例及组装质量有关。通常认为重复序列比例较高的基因组所含微卫星数量较多, 但由于组装质量较差, 序列片段化程度较高, 可能会导致识别的微卫星位点较少。例如, 绒杜父鱼基因组重复序列比例较褐菖鲉低, 但由于褐菖鲉组装质量较差(Contig N50=674 bp), 其识别的微卫星数量相对较少^[8]。绒杜父鱼基因组微卫星分布以二碱基重复占据绝对优势(55.61%), 同 Labb  等^[36]提出在多数基因组中短碱基重复基元(1~3)的序列更为丰富相一致, 且与已发表的金钱鱼^[10]、中华绒螯蟹^[35]等的研究结果相同。绒杜父鱼基因组范围的微卫星位点主要集中在 5~17 次重复, 表明微卫星数量随着重复次数增加而逐渐减少的趋势, 这可能与微卫星位点长度的不断增加使得稳定性降低或由于碱基元不断重复出现更高的突变现象有关^[37]。另外, 本研究共识别到微卫星位点 583498 个, 其中单碱基的重复基元中 A/T 占多数, 这与松江鲈^[33]、中华绒螯蟹^[35]等的研究结果相同, 倪守胜等^[38]提出 A/T 重复比例更高的原因是 DNA 的复制滑动和重组机制。绒杜父鱼基因组微卫星位点中频率最高的是二碱基重复, 主要以 AC/GT、AG/CT 重复单元为主, 而 CG/GC 仅占二碱基重复的 0.31%。在脊椎动物中, GT 和 AC 是二碱基重复基元中最常见的^[39], 这可能是绒杜父鱼基因组微卫星位点中 AC/GT 数量最多的原因。本研究利用生物信息学方法在全基因组范围内识别绒杜父鱼微卫星位点, 研究结果不仅可以为后续微卫星开发提供基础信息, 也能为其基因组特征的研究提供材料。

参考文献:

- [1] Schuster S C. Next-generation sequencing transforms today's biology[J]. *Nature Methods*, 2008, 5(1): 16-18.
- [2] Kettleborough R N W, Busch-Nentwich E M, Harvey S A, et al. A systematic genome-wide analysis of zebrafish protein-coding gene function[J]. *Nature*, 2013, 496(7446): 494-497.
- [3] Han Z F. Whole genome gene MAP construction and sex-determining genes mapping in *Nibea albiflora*[D]. Xiamen: Jimei University, 2019. [韩兆方. 黄姑鱼基因组精细图谱绘制及性别决定候选基因定位研究[D]. 厦门: 集美大学, 2019.]
- [4] Wu C W, Zhang D, Kan M Y, et al. The draft genome of the large yellow croaker reveals well-developed innate immunity[J]. *Nature Communications*, 2014, 5: 5227.
- [5] Zhang Y D, Wen L T, Luo H L, et al. Genome survey and development of SSR molecular markers for *Trachinotus ovatus*[J]. *Journal of Southern Agriculture*, 2020, 51(5): 983-994. [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.]
- [6] Xu S Y, Zhang H, Gao T X. Comprehensive whole genome survey analyses of male and female brown-spotted Flathead fish *Platycephalus* sp. 1[J]. *Genomics*, 2020, 112(6): 4742-4748.
- [7] Liu K, Xu D P, Li J, et al. Whole genome sequencing of Chinese clearhead icefish, *Protosalanx hyalocranius*[J]. *GigaScience*, 2017, 6(4): giw012.
- [8] Xu S Y, Song N, Xiao S J, et al. Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus*[J]. *Bioscience Reports*, 2020, 40(2): BSR20192252.
- [9] Song H, Guo X M, Sun L N, et al. The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia[J]. *BMC Biology*, 2021, 19(1): 15.
- [10] Wang Y R, Yang W, Ren X L, et al. Distribution patterns of microsatellites and development of polymorphic markers from *Scatophagus argus* genome[J]. *Journal of Guangdong Ocean University*, 2020, 40(4): 7-14. [王耀嵘, 杨尉, 任席林, 等. 金钱鱼基因组微卫星分布特征分析及多态性标记开发[J]. 广东海洋大学学报, 2020, 40(4): 7-14.]
- [11] Zhou W, Li B, Li L, et al. Genome survey sequencing of *Dioscorea zingiberensis*[J]. *Genome*, 2018, 61(8): 567-574.
- [12] Munehara H. Utilization of polychaete tubes as spawning substrate by the sea raven *Hemitripterus villosus* (Scorpaeniformes)[J]. *Environmental Biology of Fishes*, 1992, 33(4): 395-398.
- [13] Yang J J, Jiang Z Q, Zuo R T, et al. Nutritional analysis and evaluation on eggs of *Hemitripterus villosus*[J]. *Chinese Journal of Animal Nutrition*, 2014, 26(4): 1103-1110. [杨晶晶, 姜志强, 左然涛, 等. 绒杜父鱼卵营养成分分析及评价[J]. 动物营养学报, 2014, 26(4): 1103-1110.]
- [14] Wang M L, Liu G, Wang Z G, et al. Characterization and phylogenetic analysis of the mitochondrial genome of *Hemitripterus villosus*[J]. *Mitochondrial DNA Part B*, 2020, 5(2): 1457-1458.
- [15] Gao X, Brosse S, Chen Y B, et al. Effects of damming on population sustainability of Chinese sturgeon, *Acipenser sinensis*: Evaluation of optimal conservation measures[J]. *Environmental Biology of Fishes*, 2009, 86(2): 325-336.
- [16] Du J Y. Ecological features and stock enhancement prospects of *Hemitripterus villosus*[J]. *Beijing Fisheries*, 2006, 5: 12-15. [杜佳垠. 太平洋绒杜父鱼生态学特点与增养殖前景[J]. 北京水产, 2006, 5: 12-15.]
- [17] Srinivasan U T, Watson R, Sumaila U R. Global fisheries losses at the exclusive economic zone level, 1950 to present[J]. *Marine Policy*, 2012, 36(2): 544-549.
- [18] Halpern B S, Longo C, Hardy D, et al. An index to assess the health and benefits of the global ocean[J]. *Nature*, 2012, 488(7413): 615-620.
- [19] Lee B Y, Surh J H. Seasonal variations in the nutritional compositions and heavy metals in two demersal fish, *Liparis tessellatus* (cubed snailfish) and *Hemitripterus villosus* (shaggy sea raven)[J]. *Korean Journal of Food Science and Technology*, 2011, 43(3): 282-290.
- [20] Antonenko D V, Pushchina O I, Solomatov S F, et al. Distribution and some features of biology of sea raven *Hemitripterus villosus* (Hemipteridae) in the northwestern part of the Sea of Japan[J]. *Journal of Ichthyology*, 2010, 50(8): 640-649.
- [21] Tokranov A M, Orlov A M. Biological features of the sea raven *Hemitripterus villosus* (Hemipteridae) in Pacific waters off the northern Kuril Islands and southeastern Kamchatka[J]. *Journal of Ichthyology*, 2006, 46(9): 743-751.
- [22] Sato M. Histological observations on the cutaneous processes on the head of *Azuma emmion* and *Hemitripterus villosus*[J]. *Japanese Journal of Ichthyology*, 1977, 24(1): 12-16.
- [23] Markevich A I. Dynamics of the reproduction period of the sea raven *Hemitripterus villosus* (Hemipteridae) in the southwestern part of Peter the Great Bay: Results of 20-year monitoring (1997-2016)[J]. *Journal of Ichthyology*, 2019, 59(1): 31-37.
- [24] Jang S H, Lee J W, Kim J K. Molecular and morphometric variations in the sea raven, *Hemitripterus villosus* from Korea, with its implication on fisheries management[J].

- Ocean Science Journal, 2019, 54(3): 419-433.
- [25] Chen D G, Zhang M Z. Marine Fishes of China[M]. Qingdao: China Ocean University Press, 2016: 851. [陈大刚, 张美昭. 中国海洋鱼类[M]. 青岛: 中国海洋大学出版社, 2016: 851.]
- [26] Miya M, Sato Y, Fukunaga T, et al. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species[J]. Royal Society Open Science, 2015, 2(7): 150088.
- [27] 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]. Quantitative Biology, 2013, 35(s1–3): 62-67.
- [28] 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.
- [29] Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers[J]. Methods in Molecular Biology (Clifton, NJ), 2000, 132: 365-386.
- [30] 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]. Journal of Genetics, 2016, 95(3): 603-609.
- [31] Chen M, Tan Z Y, Zeng G M, et al. Comprehensive analysis of simple sequence repeats in pre-miRNAs[J]. Molecular Biology and Evolution, 2010, 27(10): 2227-2232.
- [32] Xu S Y, Zhao L L, Xiao S J, et al. Whole genome resequencing data for three rockfish species of *Sebastodes*[J]. Scientific Data, 2019, 6: 97.
- [33] Zhu K H. Analysis of the whole genome sequence of *Trachidermus fasciatus* and its preliminary study on salinity adaptation mechanism[D]. Zhoushan: Zhejiang Ocean University, 2020. [朱科桦. 松江鲈(*Trachidermus fasciatus*)全基因组序列分析及盐度适应机制初探[D]. 舟山: 浙江海洋大学, 2020.]
- [34] Faircloth B C, McCormack J E, Crawford N G, et al. Ultra-conserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales[J]. Systematic Biology, 2012, 61(5): 717-726.
- [35] Xu J J, Bi Y H, Cheng J H, et al. Study on distribution characteristics of whole genome microsatellite of *Eriocheir sinensis*[J]. Genomics and Applied Biology, 2021, 40(Z2): 2422-2429. [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.]
- [36] Labbé J, Murat C, Morin E, et al. Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers[J]. Current Genetics, 2011, 57(2): 75-88.
- [37] 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.
- [38] Ni S S, Yang Y, Liu S F, et al. Microsatellite analysis of *Patinopecten yessoensis* using next-generation sequencing method[J]. Progress in Fishery Sciences, 2018, 39(1): 107-113. [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] Zardoya R, Vollmer D M, Craddock C, et al. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes)[J]. Proceedings Biological Sciences, 1996, 263(1376): 1589-1598.

Whole-genome analysis and microsatellite distribution characteristics of *Hemitripterus villosus*

ZHAO Ruirui, XU Shengyong

Fishery College, Zhejiang Ocean University, Zhoushan 316022, China

Abstract: The sea raven, *Hemitripterus villosus*, is a common species in the northwestern Pacific Ocean. Owing to the significant role of the sea raven in the flesh and caviar industries, it has become one of the economically important species in China, Korea, and Russia. Owing to the decrease in resources for sturgeons in the caviar industry, sea raven is one of the best alternative candidates. However, due to anthropogenic activities, such as overfishing and pollution, the fishery resources for sea raven have been overexploited and have declined in recent years. Further, due to limited biological and genetic information on this species in Chinese coastal waters, the development of resource management and conservation, as well as fisheries and related processing industries, have been markedly restricted. To obtain useful genetic information to implement sound management and conservation plan, whole-genome analysis based on high-throughput sequencing data was performed to estimate basic genomic information and develop genome-wide genetic markers of this species at low cost. Genomic information was estimated based on the k-mer distribution pattern ($k=17$) using GCE v1.0.0 software. The genome sequence was then assembled to contig and scaffold levels using SOAPdenovo v2.01 software, with $k=41$. Subsequently, genome-wide microsatellite motifs were identified using MISA software. Primers were designed using Primer3 software. A total of 52.04 Gb of clean sequencing data was generated. K-mer analysis revealed that the genome size of *H. villosus* was approximately 713.18 Mb, with a heterozygosity rate of 0.26% and repeat sequence rate of 38.61%. Based on preliminary genome assembly results, the Contig N50 and Scaffold N50 were 7433 bp and 19 388 bp, respectively, with a final sequence size of approximately 577 Mb and GC content of 43.49%. The levels of continuity and completeness of this assembly were low; hence, further studies are needed for improved high-quality genome assembly. Based on the assembled genome sequence, 583 498 microsatellite motifs were detected throughout the genome, with a relative abundance of 1010 microsatellite motifs per Mb. Among them, the ratio of dinucleotides was the highest (55.61%), followed by that of mononucleotides (33.39%). AC/GT and AGG/CCT were dominant in the dinucleotides and trinucleotide repeats, respectively, and the repeated copy number of microsatellites mainly ranged from 5 to 17 times. To preliminarily validate the effectiveness of identifying microsatellites, ten motifs were randomly selected for primer design. After PCR amplification and electrophoresis detection, nine microsatellite motifs were successfully amplified. However, further confirmation is still needed to validate the effectiveness of identifying microsatellites and their polymorphisms at the population level. The genome survey results revealed a simple genome of *H. villosus*, which could be assembled using the sequencing strategy of “Illumina+PacBio+Hi-C” to generate a high-quality chromosomal level genome assembly. The screened microsatellite loci could also be used as effective molecular markers for subsequent genetic studies. Owing to the limitations of our study, further analyses are needed to improve the quality of genome assembly, confirm the effectiveness of the identified microsatellite loci in population genetic studies, and conduct comparative analyses between samples from different geographic localities. The findings of this study provide not only basic information for evolutionary biology and genetic assessments of the sea raven, *H. villosus*, but also scientific references for fishery and germplasm resource management and conservation from a genetic perspective.

Key words: *Hemitripterus villosus*; high-throughput sequencing; whole-genome analysis; microsatellite marker

Corresponding author: Xu Shengyong. E-mail: kevin890223@163.com