

基于基因组 survey 的横带髯鲷(*Hapalogenys analis*) 微卫星位点筛选与特征分析*

刘 焘¹ 陈冰洁² 史会来³ 平洪领³ 张 涛³ 瞿印权¹ 高天翔^{1①}

(1. 浙江海洋大学水产学院 浙江舟山 316022; 2. 汕头大学长江新闻与传播学院 广东汕头 515063; 3. 浙江省海洋水产研究所 浙江省海水增殖重点实验室 浙江舟山 316021)

摘要 探究横带髯鲷(*Hapalogenys analis*)的种质资源情况以及群体的遗传多样性和遗传结构, 对于其增殖养殖实践尤为重要。微卫星分子标记在研究物种种质情况及遗传信息方面有很大优势。采用基因组 survey 测序方法开发横带髯鲷微卫星位点, 结果表明, 横带髯鲷基因组大小为 543 Mb; 微卫星位点丰富, 共检测到 280 378 个完美型微卫星位点, 相对丰度为 415.17 个/Mb。二核苷酸重复是最丰富的微卫星类型(56.05%), 其次为单核苷酸(29.20%)、三核苷酸(10.28%)、四核苷酸(2.99%)、五核苷酸(1.03%)、六核苷酸(0.45%)重复, 短序列重复类型占 95.53%。重复单元中 A/T, TG/CA 为优势重复单元, 分别占总微卫星位点数的 22.37%和 21.98%; 10 次重复和 6 次重复的微卫星位点数量最多, 占横带髯鲷基因组微卫星总数的 14.60%和 12.46%。利用筛选出的 20 对多态性微卫星位点扩增一个供试群体进行群体遗传学检测。结果表明, 各位点等位基因数从 5 到 13 不等, 均值为 8.75 个; 平均观测杂合度(H_o)和期望杂合度(H_e)分别为 0.808 和 0.863; 多态信息含量(PIC)均值为 0.826, 且均大于 0.5。经 Bonferroni 校正后有 5 个位点显著偏离哈迪-温伯格平衡($P < 0.001$); 有 4 个位点存在无效等位基因。研究结果中的微卫星位点可用于今后横带髯鲷遗传育种和群体遗传学的研究。

关键词 横带髯鲷; 基因组 survey; 微卫星

中图分类号 S965; Q953; Q789 doi: 10.11693/hyhz20221000281

横带髯鲷(*Hapalogenys analis*)属于鲈形目(Perciformes)、石鲈科(Pomadasyidae)、髯鲷属(*Hapalogenys*), 俗称“十六枚”, 系浅海中下层岩礁性鱼类, 在我国沿海均有分布, 是极具发展潜力的增殖养殖品种, 也是名贵的水产观赏鱼类(赵盛龙等, 2016)。作为重要的海水经济鱼类, 其野生资源匮乏, 渔获量不大, 在市场上供不应求, 近年来已经开始人工养殖。关于横带髯鲷的基础研究已有一些报道, 包括生理生化(史海东等, 2004)、基础生物学(许文军等, 2002; 姚海富等, 2006; 张涛等, 2020; 平洪领等, 2021)、遗传学(喻子牛等, 1994, 1995; 王世锋等, 2008; 徐田军

等, 2010)等方面的研究。迄今尚未见组学方面的研究报道, 这极大地制约了后续横带髯鲷的繁养和保护工作。因此, 为了更好地开展横带髯鲷的增殖养殖研究, 探究其种质资源情况以及群体的遗传多样性和遗传结构尤为重要。

微卫星分子标记在研究物种种质情况及遗传信息方面有很大优势, 因微卫星 DNA 具有高度多态性与共显性、数量丰富且易于检测的特点(Tautz, 1989; Madesis *et al*, 2013), 已被广泛应用到遗传多样性检测(Cheng *et al*, 2015)和遗传种质鉴定(Lü *et al*, 2013)等多个领域。横带髯鲷微卫星标记开发未见有相关研

* 浙江省重点研发计划项目, 2021C2047 号; 浙江省“三农九方”科技协作计划项目, 2022SNJF073 号; 浙江省科技厅院所专项, HYS-CZ-202208 号; 舟山市公益类科技项目, 2021C31012 号, 2022C31056 号。刘 焘, 硕士研究生, E-mail: liu77tao15@163.com

① 通信作者: 高天翔, 博士生导师, 教授, E-mail: gaotianxiang0611@163.com

收稿日期: 2022-10-30, 收修改稿日期: 2022-12-06

究, 截至目前, GenBank 登录的横带髯鲷微卫星序列为 0 条, 其数量严重不足, 亟需开发横带髯鲷微卫星多态性标记。在过去的研究中, 微卫星标记的筛选开发主要依靠构建微卫星富集文库, 其操作复杂、成本高且效率低, 不够便利(Zane *et al*, 2002)。近年来, 随着高通量测序技术的日渐成熟, 动植物微卫星标记的批量开发拥有了基因组水平的全新开发策略(Csencsics *et al*, 2010; Capobianchi *et al*, 2013)。基因组 survey 测序为非模式生物基因组范围内微卫星位点的检测提供了技术支持, 同时也提供了生物体基因组基础信息, 包括其基因组大小、杂合度水平和重复率等信息的评估。目前, 由于基因组中含有大量的微卫星序列且检出率高, 基于该测序方法开发微卫星位点已经在多种水生生物(Lu *et al*, 2017; Chen *et al*, 2020; Xu *et al*, 2020)中得到应用。

本研究旨在使用基因组 survey 技术评估横带髯鲷基因组大小, 解析基因组范围内微卫星序列的组成和分布特征, 从中筛选开发具有多态性的微卫星位点, 并利用一个供试群体检测所开发的微卫星位点的适用性。新开发的微卫星位点可应用于今后横带髯鲷遗传育种和群体遗传学的研究。

1 材料与方法

1.1 材料

用于基因组 survey 测序的 1 尾横带髯鲷样品采自浙江舟山海域。用于多态性微卫星位点分离与鉴定及群体遗传多态度检测的横带髯鲷供试群体为采自浙江舟山海域的野生群体, 共 32 尾。将肌肉等组织液氮速冻, 放置于 -80°C 超低温冰箱中冷冻保存。

1.2 基因组 DNA 提取和 survey 测序

使用常规的酚-氯仿方法抽提基因组 DNA, 总 DNA 经过 RNase A 处理获得高纯度、无 RNA 污染的 DNA。利用微量分光光度计测定基因组 DNA, 由 D_{260}/D_{280} 和浓度($\text{ng}/\mu\text{L}$)判定 DNA 的纯度和浓度。构建 350 bp 基因组 survey 测序文库, 采用 Illumina HiSeq 2000 高通量测序仪进行双末端测序, 由武汉古奥基因科技有限公司完成。

1.3 微卫星序列的获取

将测序所得 clean data 进行基因组 de novo 组装并获得拼接序列, 使用微卫星识别工具 MISA 软件(Beier *et al*, 2017)搜索微卫星位点, 筛选参数设置为: 筛选单核苷酸(mononucleotide)、二核苷酸(dinucleotide)、三核苷酸(trinucleotide)、四核苷酸(tetranucleotide)、

五核苷酸(pentanucleotide)和六核苷酸(hexanucleotide)的微卫星位点, 对应的最低重复次数分别为 10 次、6 次、5 次、5 次、5 次和 5 次, 侧翼长度 ≥ 10 bp, 两个微卫星序列间的间隔大于 12 bp。参考 Weber(1990)评价微卫星的标准, 将所有微卫星位点区分为三种类型: 完美型、非完美型和复合型。

1.4 多态性微卫星位点分析

基于微卫星位点所在序列, 采用 Primer3 v2.37 软件(Koressaar *et al*, 2007; Untergasser *et al*, 2012; Kõressaar *et al*, 2018)设计微卫星引物, 参数设置为默认值。从设计出的引物中随机挑选 150 对合成, 以横带髯鲷基因组 DNA 为模板进行 PCR 扩增。引物合成由上海生工生物公司完成。PCR 扩增反应体系为 25 μL , 包括: 1 μL 模板 DNA(50 ng), 2.5 μL 10*buffer, 2 μL dNTPs, 正反引物各 1 μL (10 $\mu\text{mol/L}$), 17.5 μL ddH₂O, 0.15 μL Easy Taq DNA 聚合酶(TaKaRa, 北京)。PCR 扩增程序设置为: 94°C 预变性 5 min、 94°C 变性 30 s、最佳退火温度退火 30 s、 72°C 延伸 30 s、共 32 个循环, 最后 72°C 延伸 10 min。所有引物均进行了最佳退火温度的筛选, 退火温度梯度设为 $52\sim 60^{\circ}\text{C}$ 。PCR 扩增产物通过 8% 非变性聚丙烯酰胺凝胶电泳和银染法检测。

1.5 数据分析与统计

测序所得原始数据经质控, 过滤低质量数据后, 保留 clean data 使用 Jellyfish v2.01 (Marçais *et al*, 2011) 和 GenomeScope v1.0 软件(Vurture *et al*, 2017)进行 K-mer 分析, 以评估基因组大小、杂合率以及重复序列信息。横带髯鲷基因组草图通过使用 SOAPdenovo v2.2.10 软件(Luo *et al*, 2015)基于 de Bruijn 算法组装获得。基因组大小估算公式为: 基因组大小 $G = K\text{-mer 数量}/\text{峰值深度}$ (Marçais *et al*, 2011)。

微卫星序列特征分析与统计: 使用 Excel Microsatellite Toolkit (Shaibi *et al*, 2008)计算等位基因数(N_a)、观测杂合度(H_o)、期望杂合度(H_e)以及多态性信息含量(PIC)等遗传参数; 使用 Micro-Checker (Van Oosterhout *et al*, 2004)检验无效等位基因; Genepop v4.0 (Rousset, 2008)对位点进行哈迪-温伯格平衡检测。

2 结果

2.1 横带髯鲷基因组组装与基因组大小预测

基因组 survey 测序共获得 clean data 83.02 G。Q20、Q30 和 GC 含量分别为 97.54%、93.1% 和 43.65%, 测序深度约 158 $\times$ 。对过滤后的高质量数据随机抽取 10 000 对 Reads 数据, 进行 NT 库比对, 前 80% 的物

种成功比对到横带髯鲷的近缘物种 DNA 上, 证明该文库数据中不含有明显的外源污染。碱基频率分布检查无 AT、GC 分离现象。拼接共获得 Contigs 总长度 644 630 724 bp, 组装后获得 Scaffolds 总长度 675 336 059 bp, Scaffolds N50 长度为 1 589 bp (表 1)。K-mer

值设为 17 进行分析, 17-mer 深度为 114, 对应 K-mer 数量为 64 483 008 615, 预估横带髯鲷基因组大小约为 543 Mb。分析显示横带髯鲷基因组杂合率为 0.45%, 重复率为 26.78%。综上表明, 横带髯鲷基因组文库测序质量正常, 可用于后续信息分析。

表 1 横带髯鲷基因组拼接结果
Tab.1 The result of genome assembly of *H. analis*

类别	总长度/bp	序列总数	≥ 2 kb	最大长度/bp	N50/bp	GC 含量/%
Contigs	644 630 724	1 645 244	37 450	19 715	682	42.8
Scaffolds	675 336 059	1 115 882	59 651	137 823	1 589	40.9

2.2 微卫星序列特征分析

在横带髯鲷基因组组装获得的 Scaffolds 数据中共检测到 280 378 个完美型微卫星位点。微卫星出现频率为 25.13%, 微卫星序列相对丰度为 415.17 个/Mb, 即平均每 2.4 Kb 出现一个微卫星位点。微卫星总长度为 5 824 206 bp, 平均长度为 20.77 bp。对微卫星的组成类型和数量进行分析, 结果表明: 二核苷酸重复是横带髯鲷基因组最丰富的微卫星类型, 占比 56.05%; 其次为单核苷酸重复与三核苷酸重复, 分别占比 29.20%和 10.28%, 上述三种重复类型占微卫星总量的 95.53%。而其余重复类型占比很低, 四核苷酸

(2.99%)、五核苷酸 (1.03%)、六核苷酸(0.45%)这三种重复仅占微卫星总量的 4.47% (表 2)。

横带髯鲷基因组 survey 测序共检测到重复单元 1 253 种, 五、六核苷酸重复单元种类最丰富, 分别为 484 和 515 种。在单核苷酸重复中 A/T 为优势重复单元, 达到了 62 714 个, 占全部微卫星位点数的 22.37%; 二核苷酸以 TG/CA 为主, 占全部微卫星位点数的 21.98%; 三核苷酸重复中占比较高的重复单元有 AAT/ATT; 四核苷酸重复中出现频率较高的是 AATG/CATT; 五核苷酸重复以 ATCTT/AAGAT 为主, 六核苷酸重复中 CCTAAC/GTTAGG 重复单元数量较多(表 3)。

表 2 横带髯鲷基因组中微卫星位点的分布特征
Tab.2 Distribution pattern of microsatellite loci in *H. analis* genome

重复类型	数量	比例/%	相对丰度/(个/Mb)	平均长度/bp	总长度/bp	平均距离/(kb/个)
单核苷酸	81 867	29.20	121.22	12.37	1 012 772	8.25
二核苷酸	157 144	56.05	232.69	24.24	3 808 858	4.30
三核苷酸	28 828	10.28	42.69	20.96	604 320	23.42
四核苷酸	8 378	2.99	12.41	29.55	247 624	80.61
五核苷酸	2 886	1.03	4.27	32.79	94 640	234.00
六核苷酸	1 275	0.45	1.89	43.91	55 992	529.67
总计	280 378	100.00	415.17	20.77	5 824 206	2.40

表 3 横带髯鲷基因组微卫星中优势和劣势重复单元类型
Tab.3 Dominant and disadvantaged repeat unit types in *H. analis* genome microsatellites

重复类型	种类	最多重复单元			最少重复单元		
		单元类型	位点数量	比例/%	单元类型	位点数量	比例/%
单核苷酸	2	A/T	62 714	22.37	C/G	19 153	6.83
二核苷酸	8	TG/CA	61 641	21.98	GC/CG	164	0.06
三核苷酸	30	AAT/ATT	2 725	0.97	ACG/CGT	22	
四核苷酸	214	AATG/CATT	577	0.20			
五核苷酸	484	ATCTT/AAGAT	328	0.12			
六核苷酸	515	CCTAAC/GTTAGG	131	0.04			

各核苷酸类型的微卫星在重复次数变异上也存在巨大差别。如图 1 所示, 10 次重复和 6 次重复的微卫星位点数量最多, 分别占横带髯鲷基因组微卫星总数的 14.60%和 12.46%。从重复次数与对应微卫星位点数量变化趋势看, 低重复次数的微卫星序列数量较多, 高重复次数的微卫星序列数量较少。随着重复次数的增多, 核苷酸重复类型的位点数在重复次数为 6 时达到第一次峰值, 重复次数为 10 时达到第二次峰值, 后呈递减趋势。

2.3 横带髯鲷多态微卫星位点分离及多态性分析

本研究共筛选 150 对微卫星引物, 其中有 84 对可稳定扩增出清晰明亮的条带, 具有多态性的微卫

星标记共 20 对, 包括 14 个二核苷酸重复、3 个三核苷酸重复和 3 个四核苷酸重复位点。图 2 展示供试群体在 HDZD-46 位点的扩增结果。20 个微卫星位点共检测到等位基因 175 个, 各位点等位基因数从 5~13 不等, 平均等位基因数为 8.75 个, 观测杂合度(H_o)范围为 0.500~0.917, 平均值为 0.808; 期望杂合度(H_e)范围为 0.719~0.918, 平均值为 0.863; 多态信息含量 (PIC)在 0.660~0.890 之间, 平均值为 0.826。经 Bonferroni 校正后有 5 个位点(HDZD-16、HDZD-66、HDZD-68、HDZD-117 和 HDZD-119)显著偏离哈迪-温伯格平衡。有 4 个位点(HDZD-16、HDZD-68、HDZD-117 和 HDZD-119)存在无效等位基因(表 4)。

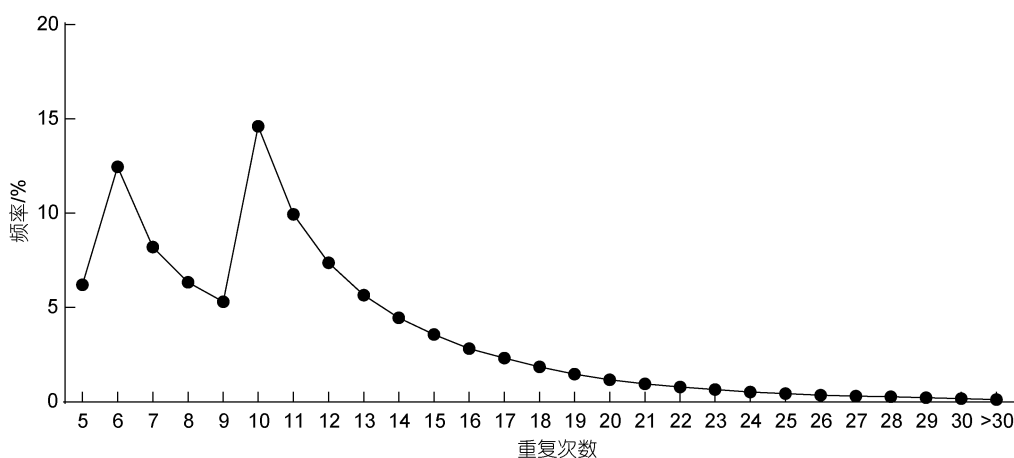


图 1 横带髯鲷基因组微卫星重复次数频率分布
Fig.1 Frequency distribution of genome microsatellite repeats in *H. analis*

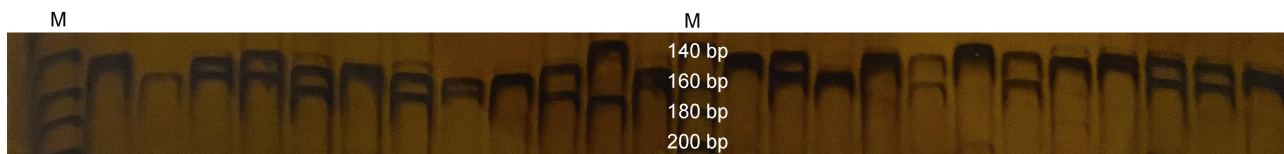


图 2 横带髯鲷在 HDZD-46 位点的聚丙烯酰胺电泳图谱
Fig.2 Polyacrylamide gel electrophoresis at HDZD-46 in *H. analis*

表 4 横带髯鲷 20 对微卫星多态位点特征
Tab.4 Characterization of 20 pairs of polymorphic loci in microsatellite of *H. analis*

位点	核心序列	引物序列(5'-3')	最适退火温度 /°C	片段区间 /bp	等位 基因数	观测 杂合度	期望 杂合度	多态信 息含量	哈迪- 温伯格平衡
HDZD-2	(AC)6	F: CCACCGCATCATTTATTCCT R: CTTCCGAGATGGATGAGAGC	59.4	270~278	7	0.542	0.859	0.822	0.005
HDZD-16	(CT)9	F: TTGTCTACGAGCGTCCTGTG R: GGGTGGAGACAGATGAAAGC	52.5	231~233	9	0.917	0.882	0.848	0.000*
HDZD-42	(CT)12	F: GCGACTTCAGCAGCAATATG R: TGTAATGCTGGTGCGAGAG	56.9	188~201	8	0.625	0.849	0.809	0.002
HDZD-43	(TA)6	F: CACAGCTGACGCTAAACCAC R: CCATCTCTCCACTCCTCTCG	56.9	176~279	5	0.500	0.719	0.660	0.033
HDZD-46	(TGTA)5	F: AAATGAACGGGAGACACGAG R: GTGTTTCTTGAGCCTCTGGC	55.1	145~229	7	0.542	0.812	0.766	0.003

续表

位点	核心序列	引物序列(5'-3')	最适退火温度 /°C	片段区间 /bp	等位 基因数	观测 杂合度	期望 杂合度	多态信 息含量	哈迪- 温伯格平衡
HDZD-47	(GT)20	F: GGCTCTACAGGCCTCTTCCT R: TCTGACAGCATTGGCAGTTC	55.1	194~223	10	0.958	0.893	0.862	0.081
HDZD-53	(GT)16	F: AATTCATGCAGAAGGGATGC R: CCATCATGAGACCAATGTGC	53.5	225~261	10	0.958	0.901	0.871	0.433
HDZD-66	(TG)10	F: CCGCCATTTAACACCTGTCT R: TACTCCTTGTCCTGCTGTG	53.5	204~213	8	0.875	0.866	0.829	0.000*
HDZD-68	(AC)14	F: ACTACTGCTGGCCATTGCT R: CTCTGCACCAAGACGTGTGT	53.5	226~245	11	0.875	0.918	0.890	0.000*
HDZD-79	(AC)12	F: AACGTGCAGACACAGAAACG R: CGAGGGTTTGGATTATGCAG	55.1	142~180	10	0.750	0.871	0.837	0.006
HDZD-87	(AC)21	F: ATTATGTCCAGTTTGCAGGC R: GCAACGATACTGCTCCCTTC	55.1	247~260	9	0.958	0.858	0.823	0.016
HDZD-102	(GT)17	F: ATCCTGAGAATCCCAAGCCT R: GGCCATCTCCTAATCCAGT	55.1	200~260	7	0.833	0.850	0.809	0.176
HDZD-103	(TCTG)7	F: GGAGGGAAAGAGTGGAGGTC R: CAGCGCTTCCTATGGAGAAC	58.4	240~269	9	0.833	0.885	0.853	0.354
HDZD-104	(TTG)7	F: TTGCAAAAGTGACGGAGACTG R: GGGAGAGAGAGTGATGCTGG	55.1	168~257	6	0.917	0.803	0.754	0.003
HDZD-106	(TG)27	F: TCTCGAGATGTTTGCCTCT R: CACGCCTACACAAGCAGATG	59.4	268~275	11	0.917	0.895	0.865	0.015
HDZD-109	(GT)12	F: TGCTGGCAGTGTCTCAGAGT R: ACTGACAGTCATGCAGTGGG	55.1	243~273	8	0.833	0.873	0.838	0.011
HDZD-116	(AAC)5	F: CTATAGCCTGCTCCACTGCC R: CTCCTCGGAGTTTGAGTTGC	58.4	244~266	9	0.917	0.859	0.823	0.018
HDZD-117	(AC)19	F: ACTACCTGGTCAGGTGGCAG R: TTAGTGGGCTTGGTGTCTC	58.4	212~264	9	0.792	0.878	0.844	0.000*
HDZD-119	(TATC)5	F: AACTGCAGGATGTTCTTGGG R: CTCGTCCTTGCATCAGGAAT	56.9	267~271	9	0.875	0.889	0.861	0.000*
HDZD-135	(AAT)8	F: CCATTGATGGGAGGTGAATC R: TGTATTTCTGCTGGTGGCTG	56.9	156~182	13	0.750	0.897	0.868	0.003

注: *经 Bonferroni 校正显著偏离哈迪-温伯格平衡($P < 0.001$)

3 讨论

高通量测序的发展为横带髯鲷这样的非模式生物的遗传学研究提供了很大的便利, 基于基因组筛选微卫星位点逐渐成为开发微卫星位点的重要手段之一。本研究首次对横带髯鲷基因组进行 survey 测序, K-mer 分析结果预测横带髯鲷的基因组大小为 543 Mb。截至目前统计, 硬骨鱼类中红鳍东方鲀(*Takifugu rubripes*)的基因组最小, 为 322.5 Mb (Aparicio *et al.*, 2002); 基因组最大的为非洲肺鱼 (*Protopterus annectens*) 达到了 40 G (Wang *et al.*, 2021)。而大部分鱼类的基因组大小均小于 1 G, 如半滑舌鲷 (*Cynoglossus semilaevis*) 477 Mb (Chen *et al.*, 2014)、大黄鱼 (*Larimichthys crocea*) 728 Mb (Wu *et al.*, 2014)、中国鲳 (*Sillago sinica*) 534 Mb (Xu *et al.*, 2018) 和罗非鱼 (*Oreochromis niloticus*) 815 Mb (Brawand *et al.*, 2014) 等。而基因组大小的增加与简单重复序列的数量增加有关 (Hancock, 1996), 横带髯鲷基因组小于 1

G, 与大部分鱼类无太大差异, 推测其简单重复序列的数量应该与基因组大小相近的物种无明显差异。除基因组大小外, 本研究组装得到的其他横带髯鲷基因组草图信息, 如 GC 含量、Scaffolds 长度等也为以后的基因组方面的深入研究奠定了基础。

利用 de novo 基因组高通量测序技术, 组装得到 Scaffolds 数据中共检测到 280 378 个完美型微卫星位点, 低于在斑尾刺虾虎鱼 (*Acanthogobius ommaturus*) 和多鳞鲳 (*Sillago sihama*) 中检测到微卫星位点的数量 (Chen *et al.*, 2020; Qiu *et al.*, 2020), 分别为 349 138 个和 630 028 个; 其出现频率为 25.13%, 低于龙头鱼 (*Harpadon nehereus*) (40.47%), 相对丰度为 415.17 个 /Mb, 低于龙头鱼 (753.12 /Mb) (黄新芯等, 2021), 高于褐菖鲢 (*Sebastiscus marmoratus*) (314.6 /Mb) (Xu *et al.*, 2018)。其中, 龙头鱼 (1 315 Mb) (Yang *et al.*, 2021)、斑尾刺虾虎鱼 (928.01 Mb) 和褐菖鲢 (800 Mb) 的基因组大于横带髯鲷, 多鳞鲳 (522.6 Mb) 的基因组小于横带髯鲷, 有学者认为, 基因组越大其微卫星的大小或

者相对丰度也随之增大(Crollius *et al.*, 2000; Schlötterer *et al.*, 2000), 但经过对比, 我们认为这样的推论欠妥。横带髯鲷微卫星组成类型结果分析表明, 随着核苷酸数量增多, 微卫星数量呈递减现象。其中二核苷酸重复最为丰富, 这与真核生物中二核苷酸重复类型的微卫星最为常见的普遍规律一致(罗文永等, 2003); 单核苷酸重复数量仅次于二核苷酸重复, 但是单核苷酸重复作为分子标记的用途不大, 一般在设计引物时不纳入考虑范围; 其余核苷酸重复数量相对较少。此外, 核苷酸重复类型的数量占比与物种进化水平有紧密的联系, 低级重复类型数量多表明物种进化水平高, 高级重复类型占比大则表明物种具较短的进化时间或者变异频率低(Harr *et al.*, 2000; Li *et al.*, 2017; Lei *et al.*, 2021)。在横带髯鲷的微卫星组成中, 一、二、三核苷酸重复类型占比达微卫星总量的 95.53%, 我们认为这可能是由于横带髯鲷的进化历程长造成的。这提示我们在大批量开发微卫星时, 二、三核苷酸重复类型应当是主要的筛选对象, 这些类型的微卫星位点具较高的分子标记开发潜力。

在横带髯鲷基因组中, 所有微卫星位点中共有 1 253 种重复单元, 随着核苷酸数量增多, 各类型的重复单元种类呈递增趋势, 但其数量呈递减趋势。单核苷酸重复中 A/T 含量丰富, C/G 较少; 二核苷酸重复中 TG/CA 含量较多, GC/CG 含量较少; 三核苷酸重复中 ACG/CGT 含量极少。三种重复类型中 GC 含量均为少数, 这一特点是由于脊椎动物基因组中 GC 碱基重复较为罕见, 可能与基因组中 CpG 岛甲基化、胞苷酸 C 易通过脱氨基作用转变为胸腺嘧啶 T 有关(Schorderet *et al.*, 1992)。此外, 在比较各重复单元的重复次数时, 我们发现重复次数为 6 时, 微卫星数量达到第一次峰值; 重复次数为 10 时的微卫星数量最多, 达到第二次峰值, 在第二次峰值过后, 微卫星位点数量随着碱基重复单元的重复次数的增加而减少。这一结果不符合大多数真核生物微卫星数目随碱基重复单元的重复次数增加呈指数下降趋势的一般规律。结合重复单元与重复次数, 在横带髯鲷微卫星位点开发中, 优先选择优势重复单元与重复次数较高的微卫星是提高微卫星多态性有效的方法, 这在筛选多态性微卫星位点的结果中也得到了印证。

本研究共筛选出 20 对具多态性的微卫星位点, 并应用于供试群体遗传多样性检测。对于微卫星多态性的评价通常根据 Botstein 等(1980)提出的标准进行, 即当 $PIC > 0.5$ 时, 位点为高度多态; 当 $0.25 < PIC < 0.5$

时, 位点为中度多态; 当 $PIC < 0.25$ 时, 位点为低度多态。据此, 本研究中的 20 个微卫星位点均呈现出较高的多态性。杂合度同样是反映物种遗传多样性水平的重要指标, 当杂合度介于 0.5~0.8 之间时, 可认为供试群体具有较高的遗传多样性。本研究中的 20 个微卫星位点平均期望杂合度与观测杂合度均达到了 0.8, 说明供试群体杂合性高, 遗传多样性高。检测 20 个微卫星位点中有 5 个位点显著偏离哈迪-温伯格平衡, 其中有 4 个位点检测到无效等位基因, 推测无效等位基因存在可能是导致偏离哈迪-温伯格平衡的重要原因。但微卫星位点扩增出现无效等位基因是不能通过一个群体的遗传分析得以解决的, 还需通过家系分析来确认, 以获取更精确的微卫星位点信息(宋忠魁等, 2013)。

4 结论

本研究首次使用基因组 survey 测序方法构建了横带髯鲷的基因组草图, 评估了基因组大小, 并在基因组范围内识别开发了 20 对具有较高多态性的微卫星位点。研究结果可用于今后横带髯鲷群体遗传学和遗传育种的研究。

参 考 文 献

- 王世锋, 杜佳莹, 苏永全, 等, 2008. 厦门近海横带髯鲷野生群体遗传结构分析[J]. 厦门大学学报(自然科学版), 47(1): 88-93.
- 平洪领, 张涛, 史会来, 等, 2021. 横带髯鲷早期生长发育特征[J]. 中国水产科学, 28(3): 276-287.
- 史海东, 毛国民, 王海岳, 2004. 温度和盐度对横带髯鲷胚胎发育的影响[J]. 上海水产大学学报, 13(3): 230-234.
- 许文军, 辛剑, 张学舒, 等, 2002. 横带髯鲷 *Hapalogenys mucronatus* (Eyedoux et Souleyet) 白点虫病的防治研究[J]. 现代渔业信息, 17(11): 24-25, 13.
- 宋忠魁, 聂振平, 谢达, 等, 2013. 拟穴青蟹 (*Scylla paramamosain*) 微卫星位点开发——基于 (AG)₁₂ 探针杂交[J]. 海洋与湖沼, 44(1): 171-176.
- 张涛, 平洪领, 史会来, 等, 2020. 舟山海域不同规格横带髯鲷营养成分分析及品质评价[J]. 浙江海洋大学学报(自然科学版), 39(6): 502-508.
- 罗文永, 胡骏, 李晓方, 2003. 微卫星序列及其应用[J]. 遗传, 25(5): 615-619.
- 赵盛龙, 徐汉祥, 钟俊生, 等, 2016. 浙江海洋鱼类志[M]. 杭州: 浙江科学技术出版社: 660-661.
- 姚海富, 毛国民, 史海东, 2006. 横带髯鲷 *Hapalogenys mucronatus* (Eyedoux et Souleyet) 亲鱼培育、产卵和孵化的初步试验[J]. 现代渔业信息, 21(4): 15-17.
- 徐田军, 王健鑫, 孙悦娜, 等, 2010. 基于髯鲷属鱼类 Cyt b 基因全序列探讨髯鲷属在鲈总科的分类地位[J]. 动物分类学报, 35(3): 530-536.

- 黄新芯, 蒋艳琳, 蒋小姿, 等, 2021. 基于高通量转录组测序技术的龙头鱼微卫星信息分析[J]. 浙江海洋大学学报(自然科学版), 40(3): 189-197.
- 喻子牛, 孔晓瑜, 徐文武, 等, 1994. 斜带髯鲷 *Hapalogenys nitens* (Richardson)和横带髯鲷 *H. mucronatus* (Eydoux et Souleyet)的核型[J]. 青岛海洋大学学报, 24(2): 175-180.
- 喻子牛, 孔晓瑜, 谢宗墉, 1995. 山东近海 21 种经济鱼类的核型研究[J]. 中国水产科学, 2(2): 1-6.
- APARICIO S, CHAPMAN J, STUPKA E, *et al*, 2002. Whole-Genome shotgun assembly and analysis of the genome of *Fugu rubripes* [J]. Science, 297(5585): 1301-1310.
- BEIER S, THIEL T, MÜNCH T, *et al*, 2017. MISA-web: a web server for microsatellite prediction [J]. Bioinformatics, 33(16): 2583-2585.
- BOTSTEIN D, WHITE R L, SKOLNICK M, *et al*, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms [J]. American Journal of Human Genetics, 32(3): 314-331.
- BRAWAND D, WAGNER C E, LI Y I, *et al*, 2014. The genomic substrate for adaptive radiation in African cichlid fish [J]. Nature, 513(7518): 375-381.
- CAPOBIANCHI M R, GIOMBINI E, ROZERA G, 2013. Next-generation sequencing technology in clinical virology [J]. Clinical Microbiology and Infection, 19(1): 15-22.
- CHEN B J, SUN Z C, LOU F R, *et al*, 2020. Genomic characteristics and profile of microsatellite primers for *Acanthogobius ommaturus* by genome survey sequencing [J]. Bioscience Reports, 40(11): BSR20201295.
- CHEN S L, ZHANG G J, SHAO C W, *et al*, 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle [J]. Nature Genetics, 46(3): 253-260.
- CHENG J, YANAGIMOTO T, SONG N, *et al*, 2015. Population genetic structure of chub mackerel *Scomber japonicus* in the Northwestern Pacific inferred from microsatellite analysis [J]. Molecular Biology Reports, 42(2): 373-382.
- CROLLIUS H R, JAILLON O, DASILVA C, *et al*, 2000. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis* [J]. Genome Research, 10(7): 939-949.
- CSENSICS D, BRODBECK S, HOLDEREGGER R, 2010. Cost-effective, species-specific microsatellite development for the endangered Dwarf Bulrush (*Typha minima*) using next-generation sequencing technology [J]. Journal of Heredity, 101(6): 789-793.
- HANCOCK J M, 1996. Simple sequences and the expanding genome [J]. BioEssays, 18(5): 421-425.
- HARR B, SCHLÖTTERER C, 2000. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation [J]. Genetics, 155(3): 1213-1220.
- KÖRESSAAR T, LEPAMETS M, KAPLINSKI L, *et al*, 2018. Primer3_masker: integrating masking of template sequence with primer design software [J]. Bioinformatics, 34(11): 1937-1938.
- KORESSAAR T, REMM M, 2007. Enhancements and modifications of primer design program Primer3 [J]. Bioinformatics, 23(10): 1289-1291.
- LEI Y, ZHOU Y, PRICE M, *et al*, 2021. Genome-wide characterization of microsatellite DNA in fishes: survey and analysis of their abundance and frequency in genome-specific regions [J]. BMC Genomics, 22(1): 421.
- LI Z G, CHEN F, HUANG C H, *et al*, 2017. Genome-wide mapping and characterization of microsatellites in the swamp eel genome [J]. Scientific Reports, 7(1): 3157.
- LU X, LUAN S, KONG J, *et al*, 2017. Genome-wide mining, characterization, and development of microsatellite markers in *Marsupenaeus japonicus* by genome survey sequencing [J]. Chinese Journal of Oceanology and Limnology, 35(1): 203-214.
- LÜ Z M, LI H M, LIU L Q, *et al*, 2013. Rapid development of microsatellite markers from the large yellow croaker (*Pseudosciaena crocea*) using next generation DNA sequencing technology [J]. Biochemical Systematics and Ecology, 51: 314-319.
- LUO R B, LIU B H, XIE Y L, *et al*, 2015. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler [J]. GigaScience, 4(1): 30.
- MADESIS P, GANOPOULOS I, TSAFTARIS A, 2013. Microsatellites: evolution and contribution [M] // KANTARTZI S K. Microsatellites: Methods and Protocols. Totowa: Humana Press: 1-13.
- MARÇAIS G, KINGSFORD C, 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers [J]. Bioinformatics, 27(6): 764-770.
- QIU B X, FANG S B, IKHWANUDDIN M, *et al*, 2020. Genome survey and development of polymorphic microsatellite loci for *Sillago sihama* based on Illumina sequencing technology [J]. Molecular Biology Reports, 47(4): 3011-3017.
- ROUSSET F, 2008. GENEPOP' 007: a complete re-implementation of the GENEPOP software for Windows and Linux [J]. Molecular Ecology Notes, 8(1): 103-106.
- SCHLÖTTERER C, HARR B, 2000. *Drosophila virilis* has long and highly polymorphic microsatellites [J]. Molecular Biology and Evolution, 17(11): 1641-1646.
- SCHORDERET D F, GARTLER S M, 1992. Analysis of CpG suppression in methylated and nonmethylated species [J]. Proceedings of the National Academy of Sciences of the United States of America, 89(3): 957-961.
- SHAIBI T, LATTORFF H, MORITZ R, 2008. A microsatellite DNA toolkit for studying population structure in *Apis mellifera* [J]. Molecular Ecology Resources, 8(5): 1034-1036.
- TAUTZ D, 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers [J]. Nucleic Acids Research, 17(16): 6463-6471.
- UNTERGASSER A, CUTCUTACHE I, KORESSAAR T, *et al*, 2012. Primer3—new capabilities and interfaces [J]. Nucleic Acids Research, 40(15): e115.
- VAN OOSTERHOUT C, HUTCHINSON W F, WILLS D P M, *et*

- al, 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data [J]. Molecular Ecology Notes, 4(3): 535-538.
- VURTURE G W, SEDLAZECK F J, NATTESTAD M, *et al*, 2017. GenomeScope: fast reference-free genome profiling from short reads [J]. Bioinformatics, 33(14): 2202-2204.
- WANG K, WANG J, ZHU C L, *et al*, 2021. African lungfish genome sheds light on the vertebrate water-to-land transition [J]. Cell, 184(5): 1362-1376.
- WEBER J L, 1990. Informativeness of human (dC-dA)_n(dG-dT)_n polymorphisms [J]. Genomics, 7(4): 524-530.
- XU S Y, SONG N, XIAO S J, *et al*, 2020. Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus* [J]. Bioscience Reports, 40(2): BSR20192252.
- XU S Y, XIAO S J, ZHU S L, *et al*, 2018. A draft genome assembly of the Chinese sillago (*Sillago sinica*), the first reference genome for Sillaginidae fishes [J]. GigaScience, 7(9): giy108.
- WU C W, ZHANG D, KAN M Y, *et al*, 2014. The draft genome of the large yellow croaker reveals well-developed innate immunity [J]. Nature Communications, 5: 5227.
- YANG T Y, HUANG X X, NING Z J, *et al*, 2021. Genome-wide survey reveals the microsatellite characteristics and phylogenetic relationships of *Harpadon nehereus* [J]. Current Issues in Molecular Biology, 43(3): 1282-1292.
- ZANE L, BARGELLONI L, PATARNELLO T, 2002. Strategies for microsatellite isolation: a review [J]. Molecular Ecology Notes, 11(1): 1-16.

ISOLATION AND ANALYSIS OF *HAPALOGENYS ANALIS* MICROSATELLITES BY GENOME SURVEY

LIU Tao¹, CHEN Bing-Jie², SHI Hui-Lai³, PING Hong-Ling³, ZHANG Tao³, QU Yin-Quan¹, GAO Tian-Xiang¹

(1. Fishery College, Zhejiang Ocean University, Zhoushan 316022, China; 2. Cheung Kong School of Journalism and Communication, Shantou University, Shantou 515063, China; 3. Zhejiang Province Key Laboratory of Mariculture and Enhancement, Zhejiang Marine Fisheries Research Institute, Zhoushan 316021, China)

Abstract It is important to explore the germplasm resources, genetic diversity, and genetic structure of *Hapalogenys analis* for its breeding research. Microsatellite molecular markers have great advantages in studying germplasm and genetic information of species. Genome survey sequencing was used to develop microsatellite loci. The results show that the genome size of *H. analis* was 543 Mb. A total of 280 378 perfect microsatellite loci with a relative abundance of 415.17 per Mb were detected. The dinucleotide motif was the most abundant microsatellite type (56.05%), followed by mono- (29.20%), tri- (10.28%), tetra- (2.99%), penta- (1.03%), and hexanucleotide repeats (0.45%). The short sequence repeats accounted for 95.53%. A/T and TG/CA were the dominant repeat units, accounting for 22.37% and 21.98% of the total number of microsatellite loci, respectively. Among those microsatellite motifs, the copy number of repeat dominated by 10 and 6, accounting for 14.60% and 12.46%, respectively. In addition, 20 polymorphic microsatellites were isolated to amplify a sampling population. The results show that the number of alleles per locus ranged from 5 to 13 and the mean allelic richness was 8.75, the mean observed heterozygosity (H_o) was 0.808, and the expected heterozygosity (H_e) was 0.863. The mean polymorphic information content (PIC) was 0.826, and all of them were above 0.5. After the Bonferroni correction, five loci were deviated significantly from the Hardy-Weinberg equilibrium ($P < 0.001$). And four loci were involved in PCR amplification of null alleles. The microsatellite loci of genome will be applied in the future study of genetic breeding and population genetics in this species.

Key words *Hapalogenys analis*; genome survey; microsatellites