



渔业科学进展

Progress in Fishery Sciences

ISSN 2095-9869, CN 37-1466/S

《渔业科学进展》网络首发论文

题目: 5 种石斑鱼全基因组微卫星筛选与特征分析
作者: 王旭蕾, 高进, 齐鑫, 王永波, 陈傅晓, 刘金叶, 符书源
DOI: 10.19663/j.issn2095-9869.20230109002
收稿日期: 2023-01-09
网络首发日期: 2023-07-21
引用格式: 王旭蕾, 高进, 齐鑫, 王永波, 陈傅晓, 刘金叶, 符书源. 5 种石斑鱼全基因组微卫星筛选与特征分析[J/OL]. 渔业科学进展.
<https://doi.org/10.19663/j.issn2095-9869.20230109002>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: 10.19663/j.issn2095-9869.20230109002

<http://www.yykxjz.cn/>

王旭蕾, 高进, 齐鑫, 王永波, 陈傅晓, 刘金叶, 符书源. 5 种石斑鱼全基因组微卫星筛选与特征分析. 渔业科学进展, 2023, 44(*): 00–00

WANG X L, GAO J, QI X, WANG Y B, CHEN F X, LIU J Y, FU S Y. Screening and characteristics analysis of microsatellites in the whole genome of five groupers. Progress in Fishery Sciences, 2023, 44(*): 00–00

5 种石斑鱼全基因组微卫星筛选与特征分析*

王旭蕾^{1,3} 高 进^{1,2,4} 齐 鑫³ 王永波^{1,2}
陈傅晓^{1,2} 刘金叶¹ 符书源^{1,2,4①}

(1. 海南省海洋与渔业科学院 海南 海口 571126; 2. 海南热带海洋学院崖州湾创新研究院
海南 三亚 572025; 3. 中国海洋大学海水养殖教育部重点实验室 山东 青岛 266003;
4. 海南省热带海水养殖工程技术研究中心 海南 海口 571126)

摘要 为了解赤点石斑鱼(*Epinephelus akaara*)、斜带石斑鱼(*E. coioides*)、云纹石斑鱼(*E. moara*)、棕点石斑鱼(*E. fuscoguttatus*)、鞍带石斑鱼(*E. lanceolatus*) 5 种石斑鱼全基因组中微卫星的分布特征, 本研究使用 Micro-Satellite (MISA) 软件对公共数据库中获取的 5 种石斑鱼全基因组序列进行微卫星筛选, 分析了微卫星重复类型、重复拷贝类别及核心拷贝数的分布特征。结果显示, 在 5 种石斑鱼全基因组中, 均筛选出超过 28 万个微卫星位点, 相对丰度介于 271~296 个/Mb 之间, 平均长度在 22 bp 左右, 微卫星数量在全基因组中的占比为 0.59%~0.67%, 其重复类型数量、占比和相对丰度的分布趋势一致, 二碱基重复最多, 其次为单碱基, 且随着重复单元碱基数目的增加而减少。重复拷贝类别 A、AC、AAT、AAG、AGC、AATC、AAAT、AGAT、AATG、AGAGG、AAAAT、AAGAT、ACAGAG、AAANNN、AANNNN (N 为除 A 外其他 3 种碱基) 为优势类别。不同重复类型微卫星的拷贝数变化范围较大, 但每种重复类型的拷贝数变化趋势一致, 即随着拷贝数增加, 微卫星数目随之递减, 拷贝数为 6 和 12 时微卫星数目出现峰值。其中, 各个重复类型中均有数量尤为突出的位点: 鞍带石斑鱼中 T、TA 和 AGACAG 分别重复了 502、803 和 48 次, 云纹石斑鱼中 GAG、CACT 和 CCACA 分别重复了 205、652 和 111 次。5 种石斑鱼全基因组微卫星分布特征基本一致, 鞍带石斑鱼和云纹石斑鱼中存在与其他 3 种石斑鱼显著差异的重复位点。本研究可为 5 种石斑鱼高质量微卫星分子标记开发提供数据参考, 并为其基因组进化、遗传变异、亲缘关系和新品种选育等方面的工作奠定基础。

关键词 石斑鱼; 全基因组; 微卫星; 分布特征

中图分类号 S917.4 **文献标识码** A **文章编号** 2095-9869(2023)00-0000-10

石斑鱼广泛分布于全球热带、亚热带的暖水性中下层水域, 是一种珊瑚礁鱼类。石斑鱼营养丰富、市场价值高、消费需求大, 为大宗海洋经济鱼类, 深受消费者欢迎, 养殖潜力巨大(Ge et al, 2019)。《中国渔

* 海南省重大科技计划(ZDKJ2021017)、海南省科技计划三亚崖州湾科技城联合项目(320LH030)和海南省省级部门预算项目(KYL-2022-08)共同资助。王旭蕾, E-mail: wangxulei98@foxmail.com

① 通信作者: 符书源, 副研究员, E-mail: fulank23@sohu.com

收稿日期: 2023-01-09, 收修改稿日期: 2023-03-12

业统计年鉴》数据显示,我国石斑鱼养殖产量逐年上升,2021年产量超过20万t,同比上一年增长6.29%(中国渔业统计年鉴,2022),位居我国海水养殖鱼类产量前十,我国现已成为世界范围内的石斑鱼主产国(Rimmer *et al*, 2019)。然而,随着养殖规模的扩大,石斑鱼产业出现了种质退化、优质苗种匮乏等诸多问题(Shapawi *et al*, 2019),亟需开展其重要经济性性状遗传解析和良种选育工作。目前,石斑鱼的良种选育主要采用选择育种和杂交育种等传统育种方式(Rimmer *et al*, 2019)。随着分子生物技术的快速发展和高通量测序成本的急剧下降,SSR (Microsatellite)、SNP (single nucleotide polymorphism)等分子标记相继出现(黎裕等, 1999),分子标记开发逐渐成为石斑鱼性状解析和遗传育种的重要手段。其中,微卫星因其遗传共显性、多态性高、信息含量丰富、种间可转移性好、可视化和基因组覆盖率高等优点(Deng *et al*, 2014),广泛应用于物种区分、群体遗传结构分析、遗传多样性评估、亲子鉴定和标记辅助选择育种等研究。

石斑鱼作为重要的海水经济鱼类,其分子标记被广泛开发并应用。樊欣等(2022)利用SSR标记,鉴定了斜带石斑鱼(*Epinephelus coioides*)人工雌核生殖子代的遗传物质来源。Xiao等(2018)结合荧光AFLP标记技术揭示了七带石斑鱼(*Hyporhamphus septemfasciatus*)养殖群体的遗传多样性和种群遗传结构。蒙子宁等(2007)利用RAPD技术分析了斜带石斑鱼繁育亲本与子代的遗传多样性。唐江等(2018)利用微卫星标记分析了新品种云龙石斑鱼[鞍带石斑鱼(*E. lanceolatus*) (♂)×云纹石斑鱼(*E. moara*) (♀)]与其父母本群体间的遗传变异情况。目前,石斑鱼分子标记开发数量十分有限,在育种上的应用也鲜有报道。因此,开发大量具有实用价值的分子标记对提高石斑鱼育种效率具有重大意义。

近年来,随着高通量测序技术的快速发展和测序成本的下降,全基因组和转录组分析促进了石斑鱼分子标记的开发(Táth *et al*, 2000; 高峰涛等, 2017)。目前,赤点石斑鱼(*E. akaara*)、斜带石斑鱼、棕点石斑鱼(*E. fuscoguttatus*)、鞍带石斑鱼、云纹石斑鱼已完成全基因组测序。为开发与经济性性状相关联的分子标记,构建分子辅助育种技术体系,本研究基于全基因组数据对5种石斑鱼SSR进行筛选,并对其种类、数量进一步比较分析,以期对石斑鱼生长、抗逆和抗病等性状的新品种选育奠定基础。

1 材料与方法

1.1 微卫星位点筛选

从CNCB (<https://www.cngb.org/>)、NCBI (<https://www.ncbi.nlm.nih.gov/>)、DDBJ (<https://www.ddbj.nig.ac.jp/>)与Dryad (<https://datadryad.org/stash>)数据库中获取5种石斑鱼的基因组序列(Ge *et al*, 2019; Zhou *et al*, 2019, 2021; Yang *et al*, 2021),详细信息见表1。利用Micro-Satellite (MISA)软件(<https://webblast.ipkgatersleben.de/misa/>)(Beier *et al*, 2017)进行5种石斑鱼基因组微卫星序列的筛选:重复单元:1~6 bp,单碱基重复拷贝数 ≥ 12 ,二碱基重复拷贝数 ≥ 6 ,三、四、五和六碱基重复拷贝数均 ≥ 5 。

1.2 数据分析

导出MISA检索数据,利用自编脚本整合数据,挑选完美型微卫星,同时利用Excel软件统计5种石斑鱼基因组完美型微卫星序列总长度、相对丰度(个/Mb;微卫星数量/基因组总长度)、SSR平均距离(bp/个;基因组总长度/微卫星数量)、重复单元类型[微卫星重复拷贝序列中的碱基,如(AT)₁₆中的AT]、重复类型(单、二、三、四、五和六碱基重复)、核心拷贝数[重复单元出现的次数,如:(AT)₁₆中的16]、重复拷贝类别(表示重复类别具体由哪些碱基组成)。根据碱基互补配对原则和起始碱基顺序的排列差异,将同类重复兼并为同一种重复拷贝类别,如AGC类别包括AGC、GCA、CAG、GCT、TGC、CTG,AGCG类别包括GCGA、CGAG、GAGC、AGCG、TCGC、CTCG、GCTC、CGCT(Jurka *et al*, 1995)。

2 结果与分析

2.1 5种石斑鱼基因组各重复类型微卫星总体分布特征

从公共数据库中获取5种石斑鱼全基因组数据(表1),使用MISA软件与自编脚本筛选整合其完美型微卫星序列,统计结果见表2。5种石斑鱼基因组大小均超过1G,微卫星相对丰度介于271~296个/Mb之间,总长度为6.30~7.06 Mb,平均长度为22 bp左右,各统计参数总体差别不大,分布规律一致。5种石斑鱼各重复类型的数量、占比、相对丰度排序一致,皆为二碱基重复最多,其次为单碱基,并随着重复单元碱基数目的增加而减少。

2.2 5 种石斑鱼基因组微卫星重复拷贝类别特征

占比稍有差异, 但整体趋势一致(表 4)。单碱基重复类别中, A、C 类别间数目差距巨大, A 类别重复的特征发现, 不同重复类别和相同重复类别的数量、

表 1 5 种石斑鱼全基因组信息

Tab.1 The information of whole genome in the five species of grouper

物种 Species	登录号 Reference or Accession	基因组大小 Genome size	Contig N50	组装质量 BUSCO analysis	组装水平 Assembly level	发表时间 Publish Date
赤点石斑鱼 <i>E. akaara</i>	DRYAD: 4398b9f	1.136 Gb	5.25 Mb	96.80%	Chromosome	2019/6/25
点带石斑鱼 <i>E. coioides</i>	NCBI:GCA_900536245.1	1.024 Gb	2.49 Mb	—	Scaffold	2019/12/5
棕点石斑鱼 <i>E. fuscoguttatus</i>	NCBI:GCA_011397635.1	1.047 Gb	13.8 Mb	97.49%	Chromosome	2020/2/22
鞍带石斑鱼 <i>E. lanceolatus</i>	NCBI:GCF_005281545.1	1.087 Gb	119.9 kb	96.40%	Chromosome	2019/5/13
云纹石斑鱼 <i>E. moara</i>	NCBI:GCA_006386435.1	1.084 Gb	2.22 Mb	95.60%	Chromosome	2020/5/26

表 2 5 种石斑鱼微卫星数量及分布特征

Tab.2 Amount and distribution of microsatellite in the five species of grouper genome

项目 Items	赤点石斑鱼 <i>E. akaara</i>	斜带石斑鱼 <i>E. coioides</i>	棕点石斑鱼 <i>E. fuscoguttatus</i>	鞍带石斑鱼 <i>E. lanceolatus</i>	云纹石斑鱼 <i>E. moara</i>
基因组大小 Genome size /Mb	1 135.73	1 023.56	1 047	1 087.42	1 083.59
微卫星总数 Microsatellites count	307 537	284 438	310 028	297 402	305 214
微卫星总长度 Total length of microsatellites /bp	6 877 655	6 304 559	7 056 845	6 421 185	6 854 645
平均长度 Average of length /bp	22.36	22.16	22.76	21.59	22.46
占基因组序列比例 Proportion of genome sequence	0.61%	0.62%	0.67%	0.59%	0.63%
相对丰度 Relative abundance /(pieces/Mb)	270.78	277.89	296.11	273.49	281.67
单碱基数目及占比 Number and proportion of mononucleotide	74 443 (24.21%)	79 860 (28.08%)	94 003 (30.32%)	90 363 (30.38%)	87 946 (28.81%)
二碱基数目及占比 Number and proportion of dinucleotide	166 081(54.00%)	146 363(51.46%)	151 984(49.02%)	146 940(49.41%)	152 421(49.94%)
三碱基数目及占比 Number and proportion of trinucleotide	43 688(14.21%)	35 492(12.48%)	37769(12.18%)	36 367(12.23%)	37 942(12.43%)
四碱基数目及占比 Number and proportion of tetranucleotide	18 531(6.03%)	17 563(6.17%)	20 970(6.76%)	18 583(6.25%)	20 430(6.69%)
五碱基数目及占比 Number and proportion of pentanucleotide	4 069(1.32%)	4 664(1.64%)	5 015(1.62%)	4 839(1.63%)	5 679(1.86%)
六碱基数目及占比 Number and proportion of hexanucleotide	725(0.24%)	496(0.17%)	287(0.09%)	310(0.10%)	796(0.26%)

数量在 5 种石斑鱼基因组中占据绝对优势, 在该类别中占比达 90.00%左右(表 3)。在二碱基重复类别中, AC 类别数量占据绝对优势, 在该重复类别中占比接近 80.00%, 同时, 是所有重复拷贝类别中数量最多

的类别(表 4); CG 类别含量最少, 在 5 种石斑鱼二碱基重复类别中仅占 0.04%~0.10%。三碱基重复类别中, 数量占前三的类别为 AAT、AGG 和 AGC, 比例分别为 29.62%~31.41%、17.38%~18.69%、12.55%~

16.26%。四碱基重复类别中, AATC、AAAT、AGAT、AATG 为优势类别, 4 种类别数量排序稍有不同, 但差距不大; 其中, 云纹石斑鱼中 AATC 类别的含量位居所有重复类别数量前十名。五碱基优势重复类别为 AGAGG、AAAAT、AAGAT; 其中, 赤点石斑鱼的重复类别数量排序稍与其他 4 种鱼类不一致。六碱基优势重复类别各不相同, 但多为 AAANN(N 代表除 A 以外的其他 3 种碱基)、AANNNN 和 ACAGAG。

2.3 5 种石斑鱼基因组微卫星序列重复拷贝数分布特征

通过分析 5 种石斑鱼基因组微卫星重复拷贝数, 发现不同重复类型微卫星的拷贝数变化具有显著差

异, 但每种重复类型的重复拷贝数变化趋势一致, 微卫星数量均随拷贝数的增加而减少(图 1)。5 种石斑鱼(赤点石斑鱼、斜带石斑鱼、棕点石斑鱼、鞍带石斑鱼和云纹石斑鱼)单碱基重复拷贝数均集中于 12~25 次, 分别占单碱基重复数量的 99.35%、96.54%、97.54%、95.39%和 97.34%。二碱基重复中, 拷贝数主要集中在 6~32 次之间, 占二碱基重复数量的 97.53%、98.91%、98.58%、99.59%和 97.79%; 其中, 分布数量在 11~14 次之间小幅度上涨。在三碱基重复中, 拷贝数主要集中于 5~16 次, 分别占 99.33%、99.18%、99.65%、98.99%和 99.34%。四碱基重复拷贝数主要集中在 5~17 次之间, 分别占 99.51%、

表 3 各重复类型中前 3 种优势重复拷贝类别

Tab.3 The top three dominant duplicated copy categories for each duplication type in the five species of grouper genome

碱基 Bases	赤点石斑鱼 <i>E. akaara</i>	斜带石斑鱼 <i>E. coioides</i>	棕点石斑鱼 <i>E. fuscoguttatus</i>	鞍带石斑鱼 <i>E. lanceolatus</i>	云纹石斑鱼 <i>E. moara</i>
单碱基 Mononucleotide	A(67960)91.29%	A(69052)86.47%	A(85760)91.23%	A(80222)88.78%	A(78792)89.59%
二碱基 Dinucleotide	C(6483)8.71%	C(10808)13.53%	C(8243)8.77%	C(10141)11.22%	C(9154)10.41%
	AC(131654)79.27%	AC(111722)76.33%	AC(114330)75.23%	AC(113806)77.45%	AC(114533)75.14%
	AG(20244)12.19%	AG(20214)13.81%	AG(23389)15.39%	AG(20993)14.29%	AG(23630)15.50%
三碱基 Trinucleotide	AT(14121)8.50%	AT(14362)9.81%	AT(14210)9.35%	AT(12046)8.20%	AT(14109)9.26%
	AAT(13119)30.03%	AAT(10846)30.56%	AAT(11187)29.62%	AAT(11423)31.41%	AAT(11697)30.83%
	AGG(7595)17.38%	AGG(6616)18.64%	AGG(6749)17.87%	AGG(6674)18.35%	AGG(7090)18.69%
四碱基 Tetranucleotide	AGC(7102)16.26%	AGC(4762)13.42%	AGC(5503)14.57%	AGC(4564)12.55%	AGC(4994)13.16%
	AATC(3684)19.88%	AATC(2850)16.23%	AATC(3417)16.29%	AATC(3136)16.88%	AATC(4593)22.48%
	AAAT(2356)12.71%	AAAT(2203)12.54%	AAAT(3332)15.89%	AAAT(2627)14.14%	AAAT(2338)11.44%
五碱基 Pentanucleotide	AGAT(2291)12.36%	AGAT(1954)11.13%	AATG(2606)12.43%	AGAT(2045)11.00%	AGAT(2085)10.21%
	AGAGG(492)12.09%	AGAGG(598)12.82%	AGAGG(527)10.51%	AGAGG(586)12.11%	AGAGG(660)11.62%
	AAAAT(320)7.86%	AAGAT(357)7.65%	AAGAT(441)8.79%	AAGAT(409)8.45%	AAAAT(533)9.39%
六碱基 Hexanucleotide	AAGAT(297)7.30%	AAAAT(319)6.84%	AAAAT(435)8.67%	AAAAT(402)8.31%	AAGAT(482)8.49%
	AAATAT(178)24.55%	AATCTG(52)10.48%	AAATCT(37)12.89%	AATCAG(20)6.45%	AATCTG(68)8.54%
	AAAATT(110)15.17%	AAATCT(46)9.27%	AAATAT(25)8.71%	ACAGAG(19)6.13%	AAATCT(67)8.42%

表 4 5 种石斑鱼中出现次数最多的 10 种重复拷贝类别

Tab.4 The top 10 duplicated copy categories observed in the most frequently occurring five species of grouper

项目 Items	赤点石斑鱼 <i>E. akaara</i>	斜带石斑鱼 <i>E. coioides</i>	棕点石斑鱼 <i>E. fuscoguttatus</i>	鞍带石斑鱼 <i>E. lanceolatus</i>	云纹石斑鱼 <i>E. moara</i>
计数和 百分比 Counts and percentage	AC(131654)79.27%	AC(111722)76.33%	AC(114330)75.23%	AC(113806)77.45%	AC(114533)75.14%
	A(67960)91.29%	A(69052)86.47%	A(85760)91.23%	A(80222)88.78%	A(78792)89.59%
	AG(20244)12.19%	AG(20214)13.81%	AG(23389)15.39%	AG(20993)14.29%	AG(23630)15.50%
	AT(14121)8.50%	AT(14362)9.81%	AT(14210)9.35%	AT(12046)8.20%	AT(14109)9.26%
	AAT(13119)30.03%	AAT(10846)30.56%	AAT(11187)29.62%	AAT(11423)31.41%	AAT(11697)30.83%
	AGG(7595)17.38%	C(10808)13.53%	C(8243)8.77%	C(10141)11.22%	C(9154)10.41%
	AGC(7102)16.26%	AGG(6616)18.64%	AGG(6749)17.87%	AGG(6674)18.35%	AGG(7090)18.69%

C(6483)8.71%	AGC(4762)13.42%	AGC(5503)14.57%	AGC(4564)12.55%	AGC(4994)13.16%
AAG(4708)10.78%	AAG(3741)10.54%	AAG(4191)11.10%	AAG(4021)11.06%	AATC(4593)22.48%
AAC(4129)9.45%	AAC(3493)9.84%	AAC(4026)10.66%	AAC(3714)10.21%	AAG(4250)11.20%

注: 括号内数字为该类别微卫星数量; 百分数为该类别所占该类型微卫星百分比。

Note: The number in parentheses refers to the number of microsatellites in this category, and the percentage refers to microsatellites in this category.

99.41%、99.36%、99.81%和 99.65%; 其中, AGAT、和 AAAG 重复拷贝数较大时分布数量较多。五碱基重复中, 拷贝数主要集中在 5~14 次之间, 分别占 98.21%、99.06%、99.52%、99.90%和 99.56%; 其中, AAGAG、AATAT 和 AGAGG 重复拷贝数较大时, 分布数量较多。在六碱基重复中, 拷贝数主要集中于 5~12 次之间, 分别占 100%、99.60%、100%、99.68%和 99.75% (图 1)。

不同重复类型微卫星的重复拷贝数分布(图 2)显示, 拷贝数在 6 和 12 次时微卫星分布数量出现峰值, 并随着拷贝数的增加而递减; 其中, 微卫星数量在重复拷贝数为 6 次时最多, 其数量在 4~5 万之间, 分别占总微卫星数量的 16.75%、15.77%、14.78%、15.80%和 16.00%。此外, 重复拷贝数最大的 SSR 位点均分布于鞍带石斑鱼和云纹石斑鱼中。鞍带石斑鱼中重复拷贝数最大的位点为单碱基重复 T、二碱基重复 TA 和六碱基重复 AGACAG, 分别重复 502、803 和 48 次。云纹石斑鱼的三碱基重复 GAG、四碱基重复 CACT、五碱基重复 CCACA 为重复拷贝数最大的位点, 分别重复 205、652 和 111 次。

3 讨论

随着下一代测序(next generation sequencing,

NGS)技术的进步, 全基因组、转录组序列等数据迅速积累, 越来越多的微卫星被快速识别并应用。军曹鱼(*Rachycentron canadum*)(马骞等, 2023)、团头鲂(*Megalobrama amblycephala*)(张芹等, 2022)、凡纳滨对虾(*Litopenaeus vannamei*)(王佳佳等, 2022)、红鳍东方鲀(*Takifugu rubripes*)(崔建洲等, 2006)、卵形鲳鲹(*Trachinotus ovatus*)(张永德等, 2020)、黄颡鱼(*Pelteobagrus fulvidraco*)(徐杰杰等, 2021)、金钱鱼(*Scatophagus argus*)(王耀嵘等, 2020)、斑鳢(*Channa maculata*)(上官清等, 2020)等都进行了基因组微卫星分子标记的筛选。本研究利用已发布的石斑鱼全基因组数据, 筛选出大量微卫星位点。相比于 RFLP (restriction fragment length polymorphism) (Ramirez et al, 2006) 和 AFLP (amplified fragment length polymorphism) (Xiao et al, 2018)等一次性能开发 9 和 600 个微卫星标记的传统方式, 全基因组数据筛选微卫星数量多达 28 万个以上, 数量更多, 经济快捷, 且覆盖范围广。

3.1 5 种石斑鱼微卫星全基因组分布分析

5 种石斑鱼全基因组数据筛选出的微卫星序列总长分别占各自基因组序列的 0.59%~0.67% (表 2), 低于人类(*Homo sapiens*)(3%)(Subramanian et al, 2003)、牛(*Bos taurus*)(4.7%)、绵羊(*Ovis aries*)(4.8%) (戚文华等,

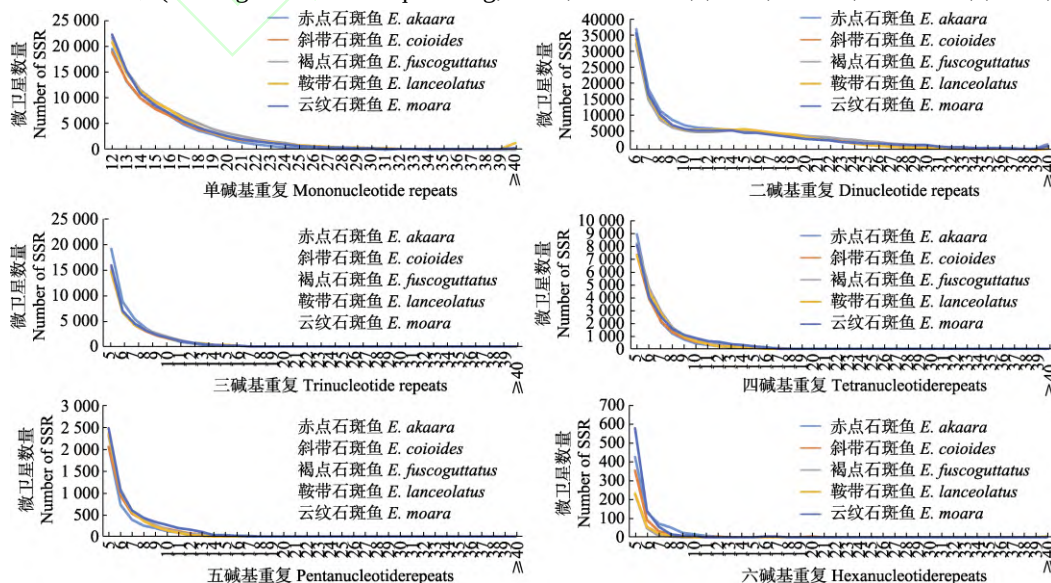


图 1 5 种石斑鱼中微卫星重复类型拷贝数分布

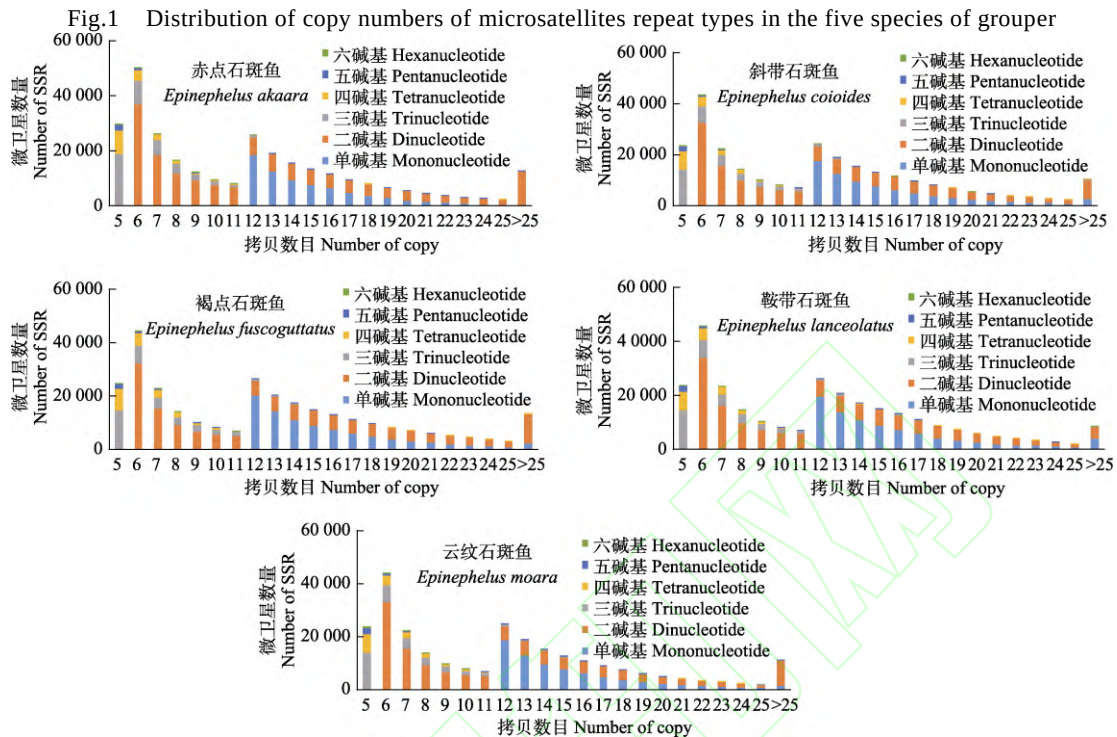


图 2 5 种石斑鱼不同微卫星重复类型拷贝数分布

Fig.2 Distribution of copy numbers of different microsatellite repeat types in five species groupers

2013)等哺乳动物所占比例;但棕点石斑与大熊猫(*Ailuropoda melanoleuca*)(0.64%)(李午佼等, 2014)相比,所占比例较高。这种现象可能归因于基因组大小和基因组内碱基组成及排列方式的差异(Hancock, 1996)。在鱼类中,黄颡鱼(1.8%)(徐杰杰等, 2020)、斑点叉尾鲷(*Ictalurus punctatus*)(1.45%)(唐荣叶等, 2022)、草鱼(*Ctenopharyngodon idella*)(1.43%)(黄纬杰等, 2022)、红鳍东方鲀(0.77%)(崔建洲等, 2006)等鱼类微卫星在基因组序列中的比例均比石斑鱼高,表明 5 种石斑鱼基因组重复序列中可能存在稀疏性(Xu et al, 2020)。

在 5 种石斑鱼基因组中,除单碱基外,随着微卫星重复单元碱基数目增加,微卫星分布频率均逐渐下降。与斑点叉尾鲷(唐荣叶等, 2022)、斑鳢(上官清等, 2020)、草鱼(黄纬杰等, 2022)等鱼类及与牛、绵羊(戚文华等, 2013)、大熊猫、北极熊(李午佼等, 2014)等哺乳动物中单碱基重复的微卫星优势类型不同,5 种石斑鱼基因组中二碱基微卫星数目最多,占微卫星总数的 49.02%~54.00%。导致此现象的原因之一,是所用微卫星筛选软件及筛选规则不同。本研究采用碱基数目均 ≥ 12 bp 原则,而其他动物的筛选规则采用单碱基从 10 bp 开始筛选(若本研究采用 ≥ 10 bp 筛选原则,

则优势重复类型为单碱基)。但在豚鹿(耿广耀等, 2022)等哺乳动物和金钱鱼(王耀嵘等, 2020)、凡纳滨对虾(王佳佳等, 2022)等水生动物中,二碱基微卫星是数量最具优势的碱基类型,与本研究结果一致。由此可见,不同物种的优势微卫星重复类型不同,而物种在微卫星分布上的差异一定程度上可以表现为物种基因组的进化现象。虽然优势类型不同,但各重复类型分布比例未有较大变化,表现出微卫星的进化保守性。

3.2 各重复拷贝类别微卫星特征分析

5 种石斑鱼中, A 类别的数量在单碱基重复中占据绝对优势,其比例为 90%左右,与人类、大熊猫、北极熊、黑青斑河豚(*Tetraodon nigroviridis*)、斑点叉尾鲷、草鱼等情况相似。这可能与上述动物中大量穿插的 Alu 和 LINE-1 等反转录序列和假基因有关(Edwards et al, 1998)。在二碱基重复中, AC、AG、AT 类别在 5 种石斑鱼该重复拷贝类别的微卫星分布数量排行前三,而 CG 类别的含量最低,仅占据二碱基重复数量的 0.04%~0.10%, CG 类别重复在大部分物种中数量都极少(Táth et al, 2000)。这是因为不同物种基因组内 4 种碱基组成含量本身具有差异性,另一方面也可能是不同碱基存在结构上的问题(Edwards et al, 1998)。在三碱基重复中,排名前三的类别均为

AAT、AGG、AGC, 其中, AAT 的高频率分布与人类、大熊猫、凡纳滨对虾、草鱼、红鳍东方鲀等大多数动物一致, AGG 与 AGC 在大多数物种中分布数量较少。有研究表明, AGG 是参与其他物种早期生长发育的许多转录因子已知结合位点, 而 AGG 类别中的 GGA 和 GAA 重复序列能够减弱大鼠多聚免疫球蛋白受体基因的表达调控(徐杰杰等, 2021)。除此以外, AGC 类别的碱基重复多态性变化与遗传疾病直接相关, 具有进化和医学研究意义。研究发现, 在 AGC 类别的 CTG 重复序列中, CTG 作为肌营养不良症肌强直蛋白激酶(DMPK)基因 3' UTR 区的组成成分, 其异常扩增会引起人 I 型强直性肌营养不良症(DM1)(Brook *et al.*, 1992), 同时出现男性不育、白内障、智力障碍、心律不齐等症状(Ergoli *et al.*, 2020)。因此, AGG、AGC 类别的高频率分布很可能在石斑鱼免疫、疾病等基因的表达调控中发挥重要作用。在四、五、六碱基重复类别中, AAAN、AAAAN 和 AAAAAN 类别均为各碱基重复类别中优势碱基类别的前 3 名, 其中, AAAN 类别在哺乳动物中含量十分丰富(Tóth *et al.*, 2000); 而 AAAAN 在灵长类动物中分布较多, 在啮齿类动物中较少。

3.3 不同微卫星重复类型的拷贝数特征分析

5 种石斑鱼全基因组中, 各重复类型的微卫星总数均在重复拷贝数为 6 或 12 次时出现峰值, 并均随着重复拷贝数的增加, 微卫星分布数逐渐减少(图 2)。这与绒杜父鱼(*Hemitripterus villosus*)、金钱鱼、卵形鲳鲹等鱼类全基因组中微卫星的变化规律一致。这种各重复类型微卫星含量随序列长度变化的规律性现象, 主要是由于微卫星的序列长度与其稳定性呈负相关, 而微卫星的突变率与其重复拷贝数呈正相关, 即微卫星序列越长, 其稳定性降低, 突变频率升高, 导致数量越少(Wierdl *et al.*, 1997); 另一方面, 单碱基和二碱基分别从 ≥ 12 和 ≥ 6 次重复开始筛选, 因此, 在这 2 个点形成了峰值。由此可见, 不同的筛选原则导致不同物种得到的统计结果不同。然而, 目前没有统一的微卫星筛选原则, 在各个物种比较时出现结果与实际不符或无法比较的情况。

在比较 5 种石斑鱼各重复类型中重复拷贝数的分布时发现, 鞍带石斑鱼(T、TA 和 AGACAG)和云纹石斑鱼(GAG、CACT、和 CCACA)存在重复拷贝数最大的微卫星位点。有研究表明, AT 是高等真核生物中最具扩展性的重复类型, 反映了鞍带石斑鱼和云

纹石斑鱼对应位点上的微卫星在进化过程中可能更容易扩张(Tóth *et al.*, 2000)。该重复位点也可能在物种进化和生长等自身发展起到重要作用, 对物种的适应性和生存能力产生影响。其中, GAG 重复序列是胃癌中具有高度微卫星不稳定性的突变靶点(Ge *et al.*, 2019), 其多态性的变化可能会导致疾病的发生。重复拷贝数的增加可能代表着相应功能的变化, 因此, 对核心拷贝数最高的重复类型进行进一步的研究, 有助于探索石斑鱼的进化机制和功能表达。

目前, 少见关于石斑鱼全基因组筛选微卫星的报道, 而利用其他方式筛选出的微卫星标记多应用于石斑鱼遗传多样性的研究(Xiao *et al.*, 2018; 蒙子宁等, 2007), 在育种方面的应用研究较少。本研究利用生物信息学软件, 在 5 种石斑鱼全基因组中筛选符合条件的微卫星, 为进一步研究其基因组特征、进化、表达奠定基础。同时, 利用在石斑鱼全基因组中筛选得到的微卫星位置信息, 可为批量开发大量微卫星引物, 为石斑鱼功能标记筛选、亲子鉴定及新品种培育等育种工作提供基础资料。

参 考 文 献

- BEIER S, THIEL T, MÜNCH T, *et al.* MISA-web: A web server for microsatellite prediction. *Bioinformatics*, 2017, 33(16): 2583–2585
- BROOK J D, MCCURRACH M E, HARLEY H G, *et al.* Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, 1992, 68(4), 799–808
- Bureau of Fisheries, Ministry of Agriculture and Rural Affairs, National Fisheries Technology Extension Center, China Society of Fisheries. China fishery statistical yearbook. Beijing: China Agriculture Press, 2022, 22 [农业农村部渔业渔政管理局, 全国水产技术推广总站, 中国水产学会. 中国渔业统计年鉴. 北京: 中国农业出版社, 2022, 22]
- CUI J Z, SHEN X Y, YANG G P, *et al.* The analysis of simple sequence repeats in *Takifugu rubripes* genome. *Periodical of Ocean University of China (Natural Sciences)*, 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]
- DENG Y W, LEI Q N, TIAN Q L, *et al.* De novo assembly, gene annotation, and simple sequence repeat marker development using Illumina paired-end transcriptome sequences in the pearl oyster *Pinctada maxima*. *Bioscience, Biotechnology, and Biochemistry*, 2014, 78(10): 1685–1692
- EDWARDS Y J, ELGAR G, CLARK M S, *et al.* The identification and characterization of microsatellites in the

- compact genome of the Japanese puffer fish, *Fugu rubripes*: Perspectives in functional and comparative genomic analyses. *Journal of Molecular Biology*, 1998, 278(4): 843–854
- ERGOLI M, VENDITTI M, PICILLO E, *et al.* Study of expression of genes potentially responsible for reduced fitness in patients with myotonic dystrophy type 1 and identification of new biomarkers of testicular function. *Molecular Reproduction and Development*, 2020, 87(1): 45–52
- FAN X, ZHANG W W, WU Y C, *et al.* Introduction of diploid gynogenesis in orange-spot grouper *Epinephelus coioides* by giant grouper *Epinephelus lanceolatus* sper. *Acta Hydrobiologica Sinica*, 2022, 46(11): 1675–1683 [樊欣, 张维炜, 吴廷昌, 等. 鞍带石斑鱼精子诱导斜带石斑鱼雌核生殖. *水生生物学报*, 2022, 46(11): 1675–1683]
- GAO F T, SHAO C W, CUI Z K, *et al.* Development and population genetic diversity analysis of microsatellite markers in *Epinephelus awoara*. *Periodical of Ocean University of China (Natural Sciences)*, 2017, 47(4): 52–57 [高峰涛, 邵长伟, 崔忠凯, 等. 基于高通量测序的青石斑鱼基因组微卫星开发及评价. *中国海洋大学学报(自然科学版)*, 2017, 47(4): 52–57]
- GE H, LIN K B, SHEN M, *et al.* De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Molecular Ecology Resources*, 2019, 19(6): 1461–1469
- GENG G Y, YOU Y Y, LIU Q X. SSR Molecular marker development and genetic diversity analysis of *Axis porcinus*. *Chinese Journal of Wildlife*, 2022, 43(3): 816–820 [耿广耀, 由玉岩, 刘群秀. 豚鹿 SSR 标记开发与遗传多样性分析. *野生动物学报*, 2022, 43(3): 816–820]
- HANCOCK J M. Simple sequences in a "minimal" genome. *Nature Reviews Genetics*, 1996, 14(1): 14–15
- HUANG W J, GUO X Z, ZHANG Z H, *et al.* Analysis of microsatellite in the entire grass carp (*Ctenopharyngodon idella*) genome and the application in parentage identification. *Journal of Fisheries of China*, 2022, 46(2): 161–172 [黄纬杰, 郭向召, 张子豪, 等. 草鱼全基因组微卫星特征分析与亲子鉴定. *水产学报*, 2022, 46(2): 161–172]
- JURKA J, PETHIYAGODA C. Simple repetitive DNA sequences from primates: Compilation and analysis. *Journal of Molecular Evolution*, 1995, 40(2): 120–126
- LI W J, LI Y Z, DU L M, *et al.* Comparative analysis of microsatellite sequences distribution in the genome of giant panda and polar bear. *Sichuan Journal of Zoology*, 2014, 33(6): 874–878 [李午佼, 李玉芝, 杜联明, 等. 大熊猫和北极熊基因组微卫星分布特征比较分析. *四川动物*, 2014, 33(6): 874–878]
- LI Y, JIA J Z, WANG T Y. Types of molecular markers and their development. *Biotechnology Information*, 1999, 15(4): 21–24 [黎裕, 贾继增, 王天宇. 分子标记的种类及其发展. *生物技术通报*, 1999, 15(4): 21–24]
- MA Q, WU Y W, WANG L Y, *et al.* Screening and characterization of polymorphic SSR markers based on whole genome sequencing of cobia (*Rachycentron canadum*). *Progress in Fishery Sciences*, 2023, 44(4): 1–11 [马骞, 吴雨薇, 王刘永, 等. 军曹鱼全基因组微卫星特征分析与多态性标记的筛选及应用. *渔业科学进展*, 2023, 44(4): 1–11]
- MENG Z N, YANG L P, WU F, *et al.* Genetic diversity in cultured stocks of *Epinephelus coioides* by RAPD analysis. *Journal of Tropical Oceanography*, 2007, 26(2): 44–48 [蒙子宁, 杨丽萍, 吴丰, 等. 斜带石斑鱼养殖群体遗传多样性的 RAPD 分析. *热带海洋学报*, 2007, 26(2): 44–48]
- QI W H, JIANG X M, XIAO G S, *et al.* Seeking and bioinformatics analysis of microsatellite sequence in the genomes of cow and sheep. *Acta Veterinaria et Zootechnica Sinica*, 2013, 44(11): 1724–1733 [戚文华, 蒋雪梅, 肖国生, 等. 牛和绵羊全基因组微卫星序列的搜索及其生物信息学分析. *畜牧兽医学报*, 2013, 44(11): 1724–1733]
- RAMIREZ M A, PATRICIA-ACEVEDO J, PLANAS S. New microsatellite resources for groupers (Serranidae). *Molecular Ecology Notes*, 2006, 6(3): 813–817
- RIMMER M A, GLAMUZINA B. A review of grouper (Family Serranidae: Subfamily Epinephelinae) aquaculture from a sustainability science perspective. *Reviews in Aquaculture*, 2019, 11: 58–87
- SHANGGUAN Q, CHEN K C, LIU H Y, *et al.* Characteristics of micorsatellites and genetic structure of wild *Channa maculate*. *South China Fisheries Science*, 2020, 16(3): 47–60 [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. *南方水产科学*, 2020, 16(3): 47–60]
- SHAPAWI R, ABDULLAH F C, SENOO S, *et al.* Nutrition, growth and resilience of tiger grouper (*Epinephelus fuscoguttatus*) × giant grouper (*Epinephelus lanceolatus*) hybrid- A review. *Reviews in Aquaculture*, 2019, 11(4): 1285–1296
- SUBRAMANIAN S, MISHRA R K, SINGH L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions. *Genome Biology*, 2003, 4(2): R13
- TANG J, TIAN Y S, LI Z T, *et al.* Analysis of genetic characters in *Epinephelus moara*, *E. lanceolatus* and their hybrids. *Journal of Agricultural Biotechnology*, 2018, 26(5): 819–829 [唐江, 田永胜, 李振通, 等. 云纹石斑鱼和鞍带石斑鱼及其杂交后代遗传性状分析. *农业生物技术学报*, 2018, 26(5): 819–829]
- TANG R Y, SU M Y, YANG W S, *et al.* Analysis of microsatellite distribution characteristics in the channel catfish (*Ictalurus punctatus*) genome. *Progress in Fishery Sciences*, 2022, 43(2): 89–97 [唐荣叶, 苏孟园, 杨汶珊, 等. 斑点叉尾鲷全基因组微卫星分布特征分析. *渔业科学进展*, 2022, 43(2): 89–97]
- TÓTH G, GÁSPÁRI Z, JURKA J. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome*

- Research, 2000, 10(7): 967–981
- WANG J J, WANG Q, QIN Z, *et al.* Development of SSR markers from genomic data for *Litopenaeus vannamei* and analysis of genetic diversity in different cultured populations. *Journal of Fisheries of China*, 2022 [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. *水产学报*, 2022]
- WANG Y R, YANG W, REN X L, *et al.* Distribution patterns of microsatellites and development of polymorphic markers from *Scatophagus argus* genome. *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. *广东海洋大学学报*, 2020, 40(4): 7–14]
- WIERDL M, DOMINSKA M, PETES T D. Microsatellite instability in yeast: Dependence on the length of the microsatellite. *Genetics*, 1997, 146(3): 769–779
- XIAO Y S, GUAN S G, LIU Q H, *et al.* Genetic characteristics of broodstock and offspring of the seven-band grouper (*Hyporhamphus septemfasciatus*) using fluorescent-AFLP markers. *Journal of the Marine Biological Association of the United Kingdom*, 2018, 98(2): 261–267
- XU J J, ZHENG X, LI J, *et al.* Distribution characteristics of whole genome microsatellite of *Pelteobagrus fulvidraco*. *Genomics and Applied Biology*, 2020, 39(12): 5488–5498 [徐杰杰, 郑翔, 李杰, 等. 黄颡鱼 (*Pelteobagrus fulvidraco*) 全基因组微卫星分布特征分析. *基因组学与应用生物学*, 2020, 39(12): 5488–5498]
- XU J J, ZHENG X, ZHANG X Y, *et al.* Analysis of distribution characteristics of microsatellites in four genomes of puffer fish. *Genomics and Applied Biology*, 2021, 40(4): 1441–1451 [徐杰杰, 郑翔, 张鑫宇, 等. 4 种河鲀全基因组微卫星分布特征分析. *基因组学与应用生物学*, 2021, 40(4): 1441–1451]
- YANG Y, WANG T, CHEN J F, *et al.* Whole-genome sequencing of brown-marbled grouper (*Epinephelus fuscoguttatus*) provides insights into adaptive evolution and growth differences. *Molecular Ecology Resources*, 2021, 22(2): 711–723
- ZHANG Q, WANG Y H, WANG B K, *et al.* Characterization of genomic microsatellites and development of SSR markers of *Megalobrama amblycephala*. *Jiangsu Agricultural Sciences*, 2022, 50(3): 79–85 [张芹, 王延晖, 王冰柯, 等. 团头鲂基因组微卫星特征及 SSR 位点开发. *江苏农业科学*, 2022, 50(3): 79–85]
- ZHANG Y D, WEN L T, LUO H L, *et al.* Genome survey and development of SSR molecular markers for *Trachinotus ovatus*. *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. *南方农业学报*, 2020, 51(5): 983–994]
- ZHOU Q, GAO H Y, XU H, *et al.* A chromosomal-scale reference genome of the kelp grouper *Epinephelus moara*. *Marine Biotechnology*, 2021, 23(1): 12–16
- ZHOU Q, GAO H Y, ZHANG Y, *et al.* A chromosome-level genome assembly of the giant grouper (*Epinephelus lanceolatus*) provides insights into its innate immunity and rapid growth. *Molecular Ecology Resources*, 2019, 19(5): 1322–1332

(编辑 冯小花)

Screening and Characteristics Analysis of Microsatellites in the Whole Genome of Five Groupers

WANG Xulei^{1,3}, GAO Jin^{1,2,4}, QI Xin³, WANG Yongbo^{1,2}, CHEN Fuxiao^{1,2}, LIU Jinye¹, FU Shuyuan^{1,2,4}①

(1. Hainan Academy of Ocean and Fisheries Sciences, Haikou 571126, China;

2. Yazhou Bay Innovation Institute, Hainan Tropical Ocean University, Sanya 572025, China;

3. Ocean University of China, Key Laboratory of Mariculture, Ministry of Education, Qingdao 266003, China;

4. Hainan Provincial Engineering Research Center for Tropical Sea-farming, Haikou 571126, China)

Abstract Grouper, a species of coral reef fish, exhibits a wide geographical distribution within the warm waters of the tropical and subtropical regions across the globe, primarily inhabiting the middle and lower layers of water. Characterized as a substantial marine economic fish, grouper possesses considerable nutritional value, boasts a high market worth, and garners significant consumer demand. Its popularity among consumers is attributed to its inherent attributes, and it holds immense potential for further cultivation and breeding endeavors. This study utilized micro-satellite (MISA) software to

① Corresponding author: FU Shuyuan, E-mail: fulank23@sohu.com

investigate the distribution characteristics of microsatellites in the genomes of five grouper species (*Epinephelus akaara*, *E. coioides*, *E. fuscoguttatus*, *E. lanceolatus*, and *E. moara*). A custom script was developed to analyze the screening results, and statistical analyses were conducted on the microsatellite repeat types, duplicate copy types, and core copy numbers in the genomes of the five grouper species. Over 280 000 microsatellite sites were identified from the entire genomes of the five grouper species. The relative abundance of microsatellites ranged from 271–296, with a total length ranging from 6.30–7.06 Mb. The average length of the microsatellites was approximately 22 bp, and their proportion in the genomes ranged from 0.59%–0.67%. These results provide insights into the distribution characteristics of microsatellites in the genomes of these five grouper species and can inform future studies on their genomic architecture and evolution. The repetitive types of microsatellites were analyzed in terms of number, proportion, and relative abundance. The number, proportion, and relative abundance of repetitive types followed a consistent pattern, with the highest number of double base repeats, followed by single base repeats. This pattern decreased as the number of repeat units increased. A, AC, AAT, AAG, AGC, AATC, AAAT, AGAT, AATG, AGAG, AAAAT, AAGAT, ACAGAG, AAANN, and AANNNN (N represents any of the three bases except A) were the most dominant types of each duplicate copy type. Type A accounted for 90.00% of single base repeats, while type AC was the most dominant in double base repeats, accounting for nearly 80.00%. Interestingly, the content of the CG duplication category was the least, accounting for only 0.04%–0.10% in the five grouper species. This may be owing to the fact that the composition content of the four bases in the different species' genomes is different, and there may be structural problems with different bases. The results of this study provide insight into the distribution characteristics of microsatellites in the genomes of these five grouper species. The high frequency distribution of AGG and AGC in the dominant types of triple base repeats may play a crucial role in regulating genes involved in immunity, disease, and other genes in groupers. Previous studies show that AGG is a well-known binding site for numerous transcription factors involved in early growth and development of various species. Additionally, the change of base repeat polymorphism of the AGC category is directly linked to genetic diseases and holds significant evolutionary and medical research value. AAAN, AAAAN, and AAAAAAN are dominant repeat types that are widely distributed in mammals among the four, five, and six base repeat types, respectively. Different types of microsatellites show significant variability in the number of core copy numbers. Nevertheless, the number of duplicate copies of each type of microsatellite exhibit a consistent trend in the five groupers, and the number of microsatellites decrease with an increase in the number of duplicate copies. The analysis of microsatellite distribution revealed several key findings. First, over 95% of single base repeat copies were concentrated in a range of 12 to 25 times. The main number of copies for two base repeats ranged from 6 to 32 times, with a small peak between 11 and 14 copies, and decreasing numbers with increasing copies. The number of copies for four and five base repeats was mainly concentrated in the ranges of 5–16 and 5–17, and 5–14, respectively. Notably, AGAT, AAAG, AAGAG, AATAT, and AGAGG repeats exhibited a large number of copies, even when the number of copies was high. The increase in copy number may represent changes in polymorphism at these loci that may lead to disease or changes in corresponding functions. Overall, these findings provide important insights into the distribution and potential functional significance of microsatellites in the genome of the studied species. The distribution characteristics of microsatellites in the genomes of the five groupers provide a valuable basis to understand the evolutionary mechanisms and functional expression of these species. The distribution of duplicate copy numbers of each type of duplication displays two peaks at 6 and 12 repetitions, with the number of microsatellites decreasing with increasing numbers of core copies. Some duplication types show particularly prominent numbers in specific species, such as T, TA, and AGACAG in the saddle belt grouper, and GAG, CACT, and CCACA in the moara grouper. These variations highlight the importance of exploring the role of microsatellite loci to develop a better understanding of the genetic distance and kinship among the five groupers. This analysis lays the groundwork to develop high-quality microsatellite molecular markers, and facilitates the selection of favorable varieties and the development of new varieties. In general, these research results provide important data to understand the genomic characteristics of the five groupers and helps to conduct advanced genetic research on these species.

Key words Grouper (*Epinephelus*); Whole genome; Microsatellite; Distribution characteristics

