



草鱼全基因组微卫星特征分析与亲子鉴定

黄纬杰^{1,2}, 郭向召³, 张子豪^{1,2}, 董强^{1,2},
熊雪梅^{1,2}, 高泽霞^{1,2*}

- (1. 华中农业大学水产学院, 农业农村部淡水生物繁育重点实验室, 长江经济带大宗水生生物产业绿色发展教育部工程研究中心, 湖北省名优鱼育种与健康养殖工程技术研究中心, 湖北 武汉 430070;
2. 湖北洪山实验室, 湖北 武汉 430070;
3. 广东海大集团股份有限公司畜牧水产研究中心, 广东 广州 511400)

摘要: 为分析草鱼全基因组微卫星特征并开发高多态性微卫星标记亲子鉴定平台, 实验利用已发布的草鱼全基因组序列, 开发高度多态、准确度高、重复单元在 4~6 碱基范围的微卫星标记。结果显示, 在草鱼 900.51 Mb 基因组序列中共筛选到微卫星序列 677 363 个, 总长度 12 835 407 bp, 占全基因组长度的 1.425 4%, 平均跨度为 1 329.43 bp。其中单碱基重复序列出现最多, 占比 52.85%; 二碱基重复序列其次, 占比 31.44%; 再次是四碱基重复序列, 占比 8.05%, 三碱基重复序列占比 6.47%, 五碱基重复序列占比 1.12%; 六碱基重复序列出现最少, 仅占比 0.07%。1~6 碱基重复类型中优势重复单位分别是 A/T、AC/GT、AAT/ATT、AGAT/ATCT、AATAT/ATATT、AACCCCT/AGGGTT。随机挑选重复序列为 4~6 个碱基的 110 个位点设计引物, 在草鱼繁殖亲本群体中扩增, 筛选出 15 对多态性引物, 经 CERVUS 3.0 软件分析 15 个位点平均期望杂合度 0.802~0.959, 模拟分析结果鉴定准确率为 100%, 置信度为 95%。通过对 20 个家系子代 192 个个体进行亲本来源鉴定, 所有子代均成功匹配到正确父母本, 鉴定准确率达到 100%。本研究分析了草鱼全基因组微卫星特征, 并利用多态性微卫星位点进行草鱼亲子鉴定, 为草鱼微卫星的功能和应用研究以及草鱼的育种工作奠定基础。

关键词: 草鱼; 全基因组; 微卫星; 亲子鉴定

中图分类号: Q 785; S 917.4

文献标志码: A

微卫星是由长度为 1~6 bp 的重复单位串联重复组成, 又称为短串联重复序列 (short tandem repeats, STR) 或简单重复序列 (simple sequence repeats, SSR), 20 世纪 80 年代被发现, 广泛且均匀随机分布于真核生物的基因组中^[1]。从基因组水平上对微卫星的种类、数量、丰度等特征的研

究有助于深化对研究物种的基因组认识^[2]。而作为一种应用广泛的遗传分子标记, 微卫星具有丰富的多态性, 主要表现在核苷酸重复单位数目的多态性和等位基因位点的多态性, 并且遵循孟德尔遗传定律, 能稳定地由亲本遗传给子代, 等位基因之间呈现共显性遗传特点^[3]。同时, 微卫星

收稿日期: 2020-11-17 修回日期: 2021-04-01

资助项目: 国家自然科学基金 (31872559); 国家大宗淡水鱼类产业技术体系建设项目 (CARS-46-08); 中国博士后科学基金 (2019M652829)

第一作者: 黄纬杰 (照片), 从事鱼类遗传育种研究, E-mail: huangweijie0306@webmail.hzau.edu.cn

通信作者: 高泽霞, 从事鱼类遗传育种研究, E-mail: gaozx@mail.hzau.edu.cn



还具有信息含量高、操作简便、结果稳定等优点,广泛应用于种群结构研究、进化史重建、亲子鉴定等方面^[4]。随着高通量测序方法的日益普及,测序成本大幅下降,公共全基因组数据库日渐丰富,通过计算机软件挖掘全基因组微卫星数据的微卫星标记开发方法也逐渐普及,相比传统富集文库的微卫星挖掘方法,新的方法更加便利和高效^[5]。在金线鲃属(*Sinocyclocheilus*)^[6]、大黄鱼(*Larimichthys crocea*)^[7]、虾虎鱼科(Gobiidae)^[8]等鱼类上已经基于全基因组序列开展了微卫星序列特征分析和标记开发的工作。

鱼类遗传育种研究对于推动水产产业发展具有重要意义。然而,由于在育种过程中的近交系数增加,遗传多样性降低,种质资源退化严重,育种过程中的亲子鉴定变得尤为重要。亲子鉴定又称亲权鉴定,除了在法医学中有重要作用外,在畜牧业和水产业中也有重要的作用,如育种过程中用于避免近亲繁殖造成的种群衰退、种质资源保护过程中增殖放流效果评估等^[9]。微卫星标记因其信息含量高、操作简便、结果稳定等优点,近年来在许多动物中普遍用于亲子鉴定分析。家猪(*Sus scrofa domestica*)、德州驴(*Equus asinus*)、西门塔尔牛(*Bos taurus*)等畜禽动物育种研究中普遍应用了亲子鉴定分析^[10-12];近年来,亲子鉴定分析也广泛应用于鱼类遗传育种工作中,如哲罗鲑(*Hucho taimen*)^[13]、青鱼(*Mylopharyngodon piceus*)^[14]、圆口铜鱼(*Coreius guichenoti*)等^[15]。

草鱼(*Ctenopharyngodon idella*)隶属鲤科(Cyprinidae)草鱼属(*Ctenopharyngodon*),是草食性鱼类,为我国四大家鱼之一。作为中国产量最大的鱼类,草鱼的育种研究是十分必要的,但目前还没有草鱼的遗传改良品种^[16]。前期,Fu等^[17]已经开发了草鱼微卫星标记,但这些标记主要是通过富集文库的方法筛选获得,并以2碱基重复序列为主,2碱基重复序列的微卫星标记在基因型判读方面极易产生误判,在草鱼亲子鉴定的实际使用过程中准确率也受到一定影响^[18],而4~6碱基重复微卫星多态性更好,在基因分型时判读也更加准确,亲子鉴定结果也更加准确。目前草鱼全基因组序列已经发布^[19],这为在全基因组水平开发多态性高,且基因型容易判读的4~6碱基重复的微卫星标记提供了可能。本实验基于草鱼全基因组序列,对其微卫星位点进行了挖掘与统计分

析,以此为基础筛选多态性微卫星位点并应用于草鱼的亲子鉴定中。

1 材料与方法

1.1 实验材料

全基因组序列来源 草鱼全基因组数据来自于中国科学院国家基因研究中心数据库(<http://www.ncgr.ac.cn/grasscarp>),下载为fasta格式。

草鱼鳍条样品 本实验材料来自于广东海大集团股份有限公司畜牧水产研究中心台山育种基地人工选育草鱼群体,包含草鱼雌性亲本20尾,雄性亲本20尾,家系繁育子代192尾。每尾鱼剪取少量鳍条,用无水乙醇固定,放入-20℃冰箱保存备用。

1.2 微卫星位点挖掘与统计

采用MISA软件对草鱼全基因组序列中的1~6核苷酸重复类型SSR进行检索,检索标准同时包括完美型(perfect)及复合型(compound)SSR重复单元,含单碱基重复微卫星(mononucleotide repeat microsatellites, MNRs)、二碱基重复微卫星(dinucleotide repeat microsatellites, DNRs)、三碱基重复微卫星(trinucleotide repeat microsatellites, TNRs)、四碱基重复微卫星(tetranucleotide repeat microsatellites, TTNRs)、五碱基重复微卫星(pentanucleotide repeat microsatellites, PTNRs)和六碱基重复微卫星类型(hexanucleotide repeat microsatellites, HXNRs)的基序,检索标准为单碱基重复大于10次,二碱基重复大于6次,三碱基以上大于5次,并统计全基因组微卫星分布基本特征。同时检索SSR位点前后200bp的序列便于进行后续的引物设计。

1.3 引物设计

通过MISA软件基于草鱼全基因组序列挖掘微卫星位点并检索前后200bp序列,再配合Primer premier 3.0软件进行引物设计。设计4碱基重复SSR引物40对,5碱基重复SSR引物64对,6碱基重复SSR引物6对,共110对。

1.4 草鱼鳍条样本收集及基因组DNA的提取

草鱼个体基因组DNA提取按照高泽霞^[20]所采用的醋酸铵/异戊醇法,提取的DNA样品用NanoDrop NanoDrop-2000紫外分光光度仪检测浓度和质量,将各样品稀释成100ng/μL的工作液放入-20℃冰箱保存。

1.5 多态性位点筛选

对 110 对引物进行基础筛选, 以随机 5 尾草鱼的混合 DNA 样品作为模板, 设置退火温度为 52~62 °C 的温度梯度, 以确定最佳退火温度和产物片段长度: 排除 PCR 扩增无条带、条带模糊及多条带等效果较差的位点后, 对筛选出的 50 对引物 5'端进行荧光修饰 (FAM、ROX、HEX), 由北京擎科生物科技有限公司合成。将引物分为 3 个 1 组, 分别带有 FAM、ROX 和 HEX 标记, 并以 40 个奠基群体草鱼亲本基因组 DNA 为模板来进行 PCR 扩增。反应体系为 10 μ L 体系, 包括 100 ng/ μ L 草鱼基因组 DNA 0.5 μ L, 2 \times Taq PCR Master Mix (上海翊圣生物科技有限公司) 5 μ L, 上游引物 (浓度 10 pmol/ μ L) 0.25 μ L, 下游引物 (浓度 10 pmol/ μ L) 0.25 μ L, ddH₂O 4 μ L。PCR 反应条件: 94 °C 预变性 5 min; 94 °C 变性 30 s, 适宜退火温度下退火 30 s, 72 °C 下延伸 30 s, 共进行 35 个循环; 最后在 72 °C 下延伸 10 min, 于 4 °C 下保存备用。PCR 产物原液送往武汉擎科生物科技有限公司进行毛细管荧光电泳检测 (ABI3730 毛细管电泳仪), 并用 GeneMarker v1.75 软件读取基因分型结果。

将基因分型数据导入 CERVUS 3.0 软件利用 allele frequency analysis 功能对 50 个位点多态性进行分析并以非亲排除率、观测杂合度 (H_o) 和期望杂合度 (H_e)、Hardy-Weinberg 平衡 (HWE) 检验、无效等位基因频率、多态信息含量 (PIC) 等参数为依据, 挑选质量好的微卫星位点 15 个进行下一步亲子鉴定实验。

1.6 亲子鉴定分析

将筛选出的 15 个位点 (表 1) 在 20 对草鱼亲本和子代 192 尾上扩增, 并进行荧光毛细管电泳检测, 用 GeneMarker v1.75 软件读取每个位点的基因分型结果。将基因分型数据导入 CERVUS 3.0 软件利用 allele frequency analysis 功能分析在 15 个位点这 232 尾个体上的非亲排除率、 H_o 和 H_e 、PIC 等, 并用 P-LOCI 软件模拟分析各位点联合非亲排除率。使用 CERVUS 3.0 软件 simulation of parentage analysis 功能进行 10 000 次模拟亲权鉴定, 并用 parentage analysis 功能对 192 尾子代与 40 尾亲本进行比对鉴定。

2 结果

2.1 草鱼全基因组微卫星特征

微卫星数量、分布、种类特征 在草鱼

约 900.51 Mb 全基因组中共挖掘微卫星重复序列 677 363 个, 总长度 12 835 407 bp, 占全基因组长度的 1.425%, 平均跨度为 1 329.43 bp, 相对丰度为 0.192 (表 2)。按重复类型统计结果显示, 其中单碱基重复序列出现最多为 357 951 个占比 52.85%; 其次是二碱基重复序列 212 973 个, 占比 31.44%; 再次是四碱基重复序列 54 550 个, 占比 8.05%; 三碱基重复序列 43 804 个, 占比 6.47%; 五碱基重复序列 7 600 个, 占比 1.12%; 六碱基重复序列最少, 485 个仅占比 0.07%。

不同重复类型中, 单碱基重复序列中出现最多类型为 A/T, 共 352 024 个, 占单碱基重复序列 98.34%; 二碱基重复序列中出现最多类型为 AC/GT, 共 94 374 个, 占二碱基重复序列 44.31%; 三碱基中出现最多类型为 AAT/ATT, 共 28 004 个, 占比 63.93%; 四碱基中出现最多为 AGAT/ATCT, 共 18 273 个, 占比 33.50%; 五碱基中出现最多为 AATAT/ATATT, 共 2 645 个, 占比 34.80%; 六碱基中为 AACCCCT/AGGGTT, 共 336 个, 占比 69.28% (表 3)。

微卫星重复类型拷贝数分布特征 草鱼全基因组微卫星重复序列的重复拷贝数变化范围很广, 从 5~549 均有分布 (图 1)。单碱基重复序列拷贝数主要集中在 10~29, 占有单碱基重复的 98.11%, 平均拷贝数 14.54; 二碱基重复序列拷贝数主要集中在 6~25, 占有二碱基重复的 96.80%, 平均拷贝数 11.45; 三碱基重复序列拷贝数主要集中在 5~11, 占有三碱基重复的 95.14%, 平均拷贝数 6.76; 四碱基重复序列拷贝数主要集中在 5~10, 占有四碱基重复的 88.48%, 平均拷贝数 7.16; 五碱基重复序列主要集中在 5~10, 占有五碱基重复的 84.93%, 平均拷贝数 7.55; 六碱基重复序列主要集中在 5~7, 占有六碱基重复的 87.42%, 平均拷贝数 6.09。重复单位长度与平均拷贝数为显著负相关 ($r=-0.858 1$, $P=0.028 8$), 即随重复单位长度变长, 平均拷贝数减少。

2.2 多态性微卫星位点筛选

将设计的 110 对引物以随机 5 尾草鱼的混合 DNA 为模板进行梯度 PCR 扩增筛选, 退火温度梯度 52~62 °C, 筛选出条带明亮单一、扩增效果较好的引物共 50 对。对筛选出的引物 5'端进行荧光标记修饰, 分别以草鱼 40 尾奠基选育群体亲本的基因组 DNA 为模板进行 PCR 扩增, 并对产物进行毛细管电泳检测个体在每个位点的基因型。

表 1 15 个草鱼微卫星位点引物基本信息

Tab. 1 Basic information of 15 SSRs primer for *C. idella*

组号 group	序号 no.	重复单元 repeat motif	引物序列 primer sequences	片段大小/bp size range	退火温度/℃ annealing temperature	荧光标记物质 fluorescence label
1	C.I.5-11	CTCTT	F: CTGGGAGGAACGGTTGTCTG R: CATGGCTGCACAACCTCAAACAAAT	265~346	52	HEX
	C.I.4-25	AGAT	F: TGCTGACTTGTTTTGAGCTGTAAG R: AATCAGCTCTACAGAAGACACCAG	228~314	58	ROX
	C.I.4-14	TAGA	F: GGATGAAAAGGGAGAAAACAACCAG R: GATGATTCCCTGATAAACTGCTGGC	90~185	52	FAM
2	C.I.4-19	GATA	F: TATACAATGTACAACGGAGCACTG R: CATTATTCGATGGGGAAACGTT	236~288	62	HEX
	C.I.4-21	GATA	F: ACAGACAGCAAACAGATAGATCGA R: GCTGTGACGAAATGATACTTGCTT	206~284	54	ROX
	C.I.5-5	TATTT	F: CTTTAGTCATGCCCTTCCACTTTC R: AAGCCATACTCTTAATACCAGGGG	185~295	62	FAM
3	C.I.6-6	GGACAT	F: TGAAGTTGAGTTGGTTTTCTGAGC R: TTAAAGCCCTCATTTTGCCATGTG	216~344	58	HEX
	C.I.5-6	ATTTT	F: GTCTCATTTTCATCATCAACGTGCT R: TCATTTTGCAATCTTGCTTCCCC	268~354	56	ROX
	C.I.5-27	TAAAA	F: AGGTGAGTAAATGTCTTATGCGGA R: CGGCTTGACTTATGTAGGTGTTT	131~211	62	FAM
4	C.I.5-34	GAGAA	F: AAGAACAAGCACAGACATGATCAC R: CCTTTCATGTTCCGTGTAGTGTG	252~308	62	HEX
	C.I.4-3	CCAT	F: AAGGGGTCCAGTTTTGTTTGAATC R: GTTGTGATGTAGGCTTTACGTCAC	209~400	62	ROX
	C.I.5-12	GTCCA	F: ATCCAGGAGAGAAGCAGCATAAAA R: CAAACAACCATACAGTTCCACAAAAG	145~287	62	FAM
5	C.I.5-43	TTCTC	F: GGTCAAACCTGCTTAATTACAACCTC R: AGTACACACGAGGAAGAGTTTGAA	205~308	60	HEX
	C.I.4-17	ATCT	F: TTGTTCCCCTTGACAGTTGTACAT R: CACAGTTTTATATACGGCCTGCTG	224~284	62	ROX
	C.I.5-8	AAGAG	F: GCCTGTTTATAAATGCACACACTG R: ACCCTAGCATTCTTGACTCAGTTT	202~280	52	FAM

表 2 草鱼全基因组微卫星特征

Tab. 2 Characteristics of microsatellites in *C. idella*

微卫星特征 characteristics of microsatellites	草鱼 <i>C. idella</i>
检测序列总数/个 total number of sequences examined	164 368
检测序列总长度/bp total size covered by examined sequences	900 506 596
检测到的SSRs总数量/个 total number of SSRs identified	677 363
检测到的SSRs总长度/bp total size of SSRs identified	12 835 407
包含有微卫星重复单元的序列数量/个 number of sequences containing SSRs	31 533
微卫星相对丰度/(bp/Mb) total relative abundance	1 425.35
微卫星平均跨度/bp average distance of SSR	1 329.43

表 3 不同重复类型优势重复单位统计

重复类型 repeat types		优势重复单位 advantage repeat unit	数量/个 number	在该类中占比/% percentages in this type
单碱基	MNRs	A/T	352 024	98.34
二碱基	DNRs	AC/GT	94 374	44.31
三碱基	TNRs	AAT/ATT	28 004	63.93
四碱基	TTNRs	AGAT/ATCT	18 273	33.50
五碱基	PTNRs	AATAT/ATATT	2 645	34.80
六碱基	HXNRs	AACCT/AGGGTT	336	69.28

利用 CERVUS 3.0 软件 allele frequency analysis 功能对这 50 个位点的多态性进行分析, 挑选出 15 对多态性较好的引物 (表 4)。15 对引物的多态性分析结果显示, 在 40 个亲本中, 平均 H_0 为 0.708~0.995, 平均值 0.888; 平均 H_e 为 0.802~0.959, 平均值 0.904; 平均 PIC 为 0.785~0.955, 平均值 0.895; 平均每个位点有 26.4 个等位基因。

2.3 亲子鉴定

将 15 个多态性较好的微卫星位点进一步在 192 尾子代中扩增, 结合 40 尾亲本和 192 尾家系子代的基因型数据, P-LOCI 软件模拟分析结果显示 (图 2), 随微卫星位点数量增加, 亲子鉴定的成功率逐渐上升。单个位点 (C.I.4-3) 鉴定准确率最高可达 6.68%, 2 个位点 (C.I.4-3、C.I.5-12) 最高可达 65.97%, 3 个位点 (C.I.4-14、C.I.4-3、C.I.5-12) 最高可达 96.85%, 当达到 4 个位点 (C.I.4-14、C.I.4-3、C.I.5-12、C.I.5-8) 时鉴定准确率则达到 99.78%, 位点数量继续增加则无限接近 100%。CERVUS 3.0 软件 allele frequency analysis 功能分析 (表 5), H_0 为 0.842~0.991, 平均值 0.915; H_e 为 0.872~0.943, 平均值 0.908; PIC 为 0.857~0.939, 平均值 0.899; 平均每个位点有 20.5 个等位基因。排除概率分析结果显示, 双亲基因型未知时, 单个位点单亲排除失败概率 ($NE-1P$) 为 0.207~0.412, 平均值 0.313; 单亲基因型已知时, 单个位点的排除失败概率 ($NE-2P$) 为 0.116~0.258, 平均值 0.187; 双亲基因型未知时, 单个位点的双亲排除失败概率 ($NE-PP$) 为 0.022~0.099, 平均值 0.059。15 个位点联合排除率为 0.999。

利用 CERVUS 3.0 软件 simulation of parentage analysis 功能进行了 10 000 次亲本已知情况下的模拟测试, 其结果显示, 9 999 次成功匹配, 鉴定准确率为 99.99%, 置信度为 95%。由此结果可

知, 在理想情况下使用这 15 个微卫星位点进行亲权鉴定可以在 95% 置信度时达到 99.99% 鉴定准确率。利用 CERVUS 3.0 软件 parentage analysis 功能, 以 15 个微卫星位点对 192 个草鱼子代样本进行鉴定, 所有子代均成功匹配到正确父母本, 鉴定准确率为 100%。

3 讨论

3.1 草鱼全基因组微卫星位点挖掘及分布特征分析

高通量测序技术广泛应用于各类公共基因组序列测定, 为从基因组水平研究各类生物的微卫星序列提供了便利。本实验从中国科学院国家基因研究中心 (NCGR) 数据库下载草鱼全基因组序列 900.51 Mb, 并利用 MISA 软件从中挖掘出 677 363 个微卫星位点, 平均跨度为 1 329.43 bp, 总长度 12.84 Mb, 占全基因组长度 1.425 4%。与红鳍多纪鲃 (*Takifugu rubripes*) (0.77%)^[21]、猪 (0.85%)^[22] 和北极熊 (*Ursus maritimus*) (0.79%)^[2] 较为相近, 与按蚊 (*Anopheles gambiae*) (2.14%)^[23] 和人类 (*Homo sapiens*) (3.00%)^[24] 相差较大。草鱼全基因组微卫星中, 单碱基重复类型 (52.85%) 最多, 二碱基重复类型 (31.44%) 第二, 四碱基重复类型 (8.05%) 第三, 这与美丽硬骨舌鱼 (*Scleropages formosus*) (51.15%)^[25] 以及大熊猫 (*Ailuropoda melanoleuca*) (45.56%) 和北极熊 (46.78%)^[2] 的重复类型组成相似, 单碱基重复序列占了微卫星总数的大部分。而在红鳍多纪鲃^[21]、金钱鱼 (*Scatophagus argus*)^[26]、西方蜜蜂 (*Apis mellifera*)^[27] 中则是二碱基重复序列为优势类型, 在酵母 (*Saccharomyces cerevisiae*)^[28]、粗糙链孢菌 (*Neurospora crassa*)^[29] 中三碱基重复序列为优势序列。由此可以看出, 优势重复类型在不同物种中差距较大, 且与物种亲缘关系远近无明显关系, 可能是由于不同物种对不同环境的适应和进化上的差异导致了这种不同。

草鱼全基因组微卫星重复序列的重复拷贝数变化范围从 5~549 均有分布, 且重复单位长度与平均拷贝数为显著负相关 ($r=-0.858 1$, $P=0.028 8$)。这与栾生等^[30] 和李红梅^[7] 在研究日本囊对虾 (*Mar-supenaes japonicus*) 和大黄鱼中发现的现象一致, 这种现象可能是由于重复单位长度更长的微卫星受到的选择压力相对更大, 因此在微卫星序列复制过程中产生滑链错配, 短的微卫星更倾向于扩张,

长的微卫星倾向于收缩^[31]。这种突变机制使得微卫星序列的长度更加均衡，不会因为突变频繁而长度过长。在重复单位具体类型上，单碱基、二碱

基、三碱基、四碱基、五碱基及六碱基重复类型中分别以 A/T、AC/GT、AAT/ATT、AGAT/ATCT、AATAT/ATATT 及 AACCT/AGGGT 为出现最

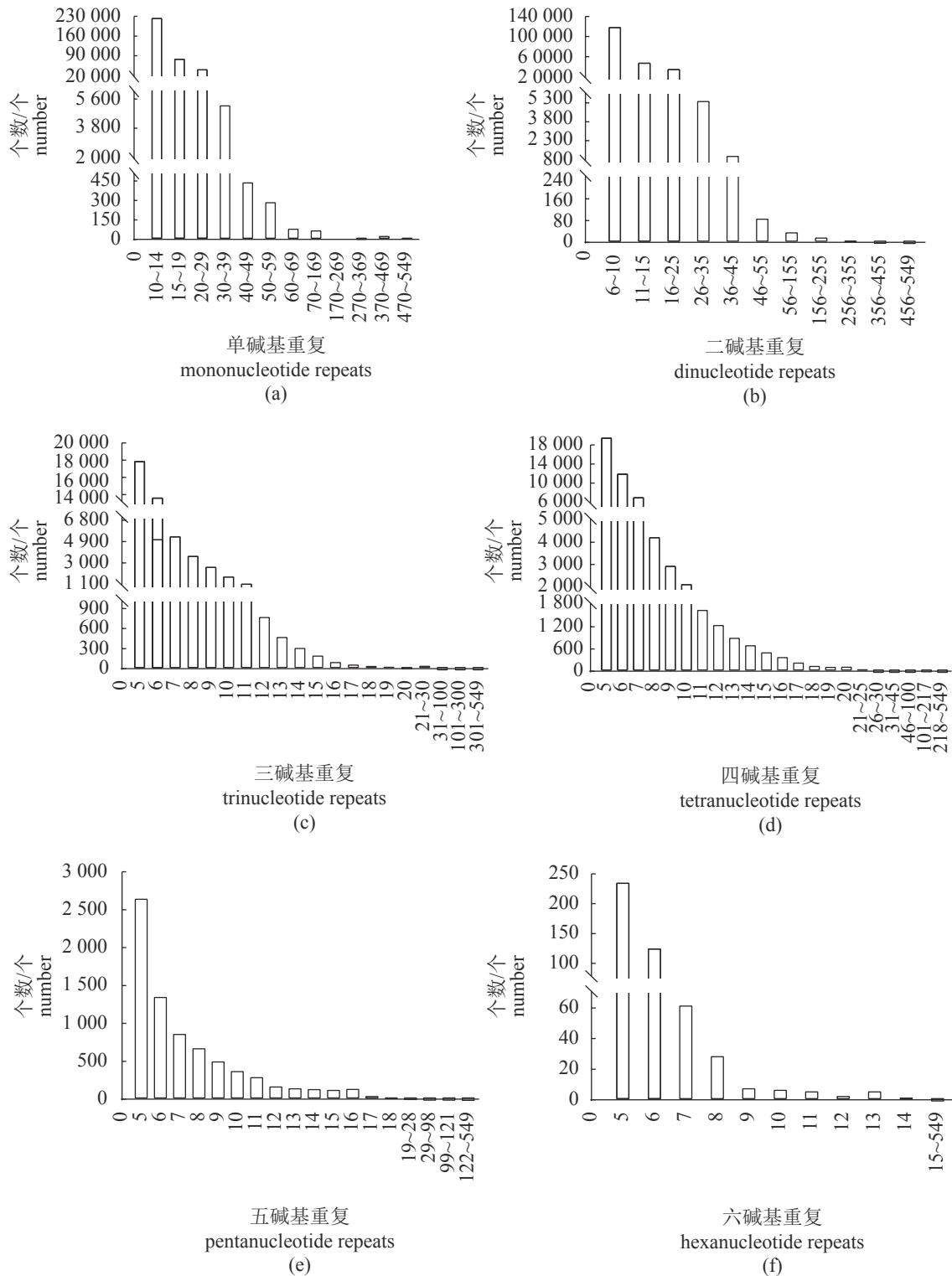


图 1 草鱼全基因组序列中微卫星不同重复类型拷贝数分布

Fig. 1 Distributions of copy numbers in different microsatellites repeat types in the whole genome sequences of *C. idella*

表 4 15 个草鱼微卫星位点多态性信息

Tab. 4 Polymorphism information of 15 SSRs primer in the *C. idella* parental population

位点 locus	等位基因数 K	平均观测杂合度 H_o	平均期望杂合度 H_e	多态信息含量 PIC	样本数 N	Hardy-Weinberg平衡 HWE	无效等位基因频率 $F(null)$
C.I.4-3	47	0.774	0.950	0.945	190	ND	0.101 1
C.I.4-14	25	0.995	0.928	0.920	191	ND	-0.036 9
C.I.4-17	34	0.879	0.945	0.939	190	ND	0.035 9
C.I.4-19	18	0.953	0.911	0.902	191	ND	-0.024 8
C.I.4-21	24	0.900	0.925	0.917	190	ND	0.013 4
C.I.4-25	40	0.921	0.892	0.882	191	ND	-0.021 2
C.I.5-5	24	0.708	0.859	0.848	192	ND	0.108 0
C.I.5-6	43	0.944	0.959	0.955	177	ND	0.006 5
C.I.5-8	18	0.921	0.892	0.880	190	ND	-0.017 9
C.I.5-11	25	0.859	0.915	0.906	192	***	0.029 5
C.I.5-12	21	0.865	0.802	0.785	192	**	-0.056 1
C.I.5-27	20	0.932	0.895	0.883	192	ND	-0.023 4
C.I.5-34	15	0.921	0.893	0.881	189	ND	-0.016 7
C.I.5-43	17	0.903	0.873	0.857	186	ND	-0.018 4
C.I.6-6	25	0.848	0.927	0.919	191	ND	0.044 6

注: ND. 未确定, NS. 无显著偏离 ($P>0.05$); ***. 极其显著偏离 ($P<0.001$), **. 极显著偏离 ($P<0.01$), *. 显著偏离 ($P<0.05$); 下同

Notes: ND. not determined, NS. non-significant deviation ($P>0.05$); ***, extremely significant deviation at $P<0.001$, **, significant deviation at $P<0.01$, *. significant deviation at $P<0.05$; the same below

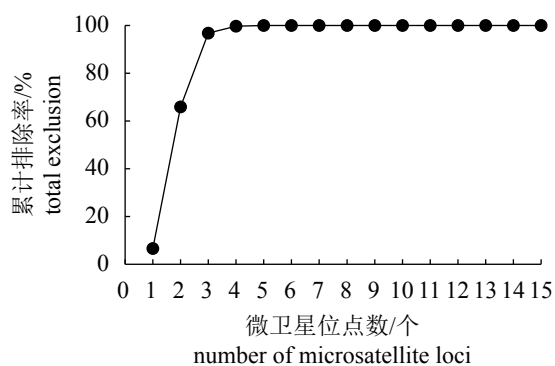


图 2 15 位点累计排除率

Fig. 2 combined probabilities of exclusion calculated over 15 microsatellite loci

多的优势重复单位, 从中可以看出微卫星中对 A 和 T 碱基存在偏向性, 这与猪、蚊子、北极熊、中国明对虾 (*Fenneropenaeus chinensis*)^[32] 中发现的规律相似, 但与红鳍东方鲀中的规律有较大差别。同时在草鱼全基因组微卫星统计中仅发现 CG/GC 重复 361 个, 占比不足 0.05%, 这与红鳍东方鲀中仅发现 1 个以及中国明对虾中不存在该重复的结果类似, 目前对于此现象的研究较少。Schooreer 等^[33] 研究表明, 由于 cpG 的甲基化使得 C 碱基易经过脱氨基过程变为 T 碱基, 因此, 未来对于微卫星序列中的 A/T、C/G 碱基含量和分布

的深入研究将对探索微卫星的功能和应用有着推动作用。

3.2 微卫星位点在草鱼家系鉴定中的应用

本实验中对 20 个繁育家系的父母本在 15 个微卫星位点的基因型进行分析, 基于亲本结果, P-LOCI 分析结果显示, C.I.4-14、C.I.4-3、C.I.5-12 和 C.I.5-8 4 个位点联合即可达到 99.780% 鉴定准确率。并对 20 个家系子代 192 个个体在 15 个位点的基因型进行了分析, CERVUS 3.0 鉴定结果显示, 所有子代均从 20 对候选亲本中匹配到正确亲本, 准确率为 100%。相比于以往采用二/三碱基重复微卫星位点在草鱼家系鉴定中的应用^[34-35], 本实验使用的 4~6 碱基重复微卫星具有更高的鉴定准确率。

本实验在鉴定过程中发现少数个体部分位点出现不符合孟德尔遗传定律的情况, 无法与其亲本基因型匹配。这种现象的出现有许多可能的原因, 不排除子代个体中发生突变的可能性, 但也可能是由于在基因分型的步骤出现分型错误, O'Reilly 等^[36] 发现, 基因分型的步骤中由于每个基因座存在 2%~3% 的分型错误, 导致基因分型的准确率无法达到 100%, 在鉴定过程中多个个体的分型错误就会导致错配。毛细管电泳这一步骤对

表 5 草鱼候选亲本和子代群体遗传多样性参数及排除概率

Tab. 5 Genetic diversity and exclusion probability for parents and offspring populations in *C. idella*

位点 locus	等位 基因数 <i>K</i>	平均观测 杂合度 <i>H_o</i>	平均期望 杂合度 <i>H_e</i>	多态信息 含量 <i>PIC</i>	<i>NE-1P</i>	<i>NE-2P</i>	<i>NE-PP</i>	样本数 <i>N</i>	Hardy-Weinberg 平衡 HWE	无效等位 基因频率 <i>F(null)</i>
C.I.4-3	32	0.862	0.936	0.930	0.232	0.132	0.029	224	ND	0.041 1
C.I.4-14	21	0.856	0.899	0.889	0.341	0.206	0.066	229	NS	0.024 3
C.I.4-17	19	0.957	0.904	0.895	0.328	0.196	0.061	232	NS	-0.030 2
C.I.4-19	13	0.896	0.886	0.874	0.374	0.229	0.079	230	NS	-0.005 2
C.I.4-21	19	0.991	0.929	0.922	0.259	0.149	0.037	229	ND	-0.034 2
C.I.4-25	23	0.930	0.904	0.895	0.322	0.192	0.057	227	NS	-0.016 2
C.I.5-5	22	0.842	0.929	0.923	0.255	0.146	0.036	221	ND	0.048 8
C.I.5-6	23	0.978	0.927	0.921	0.261	0.150	0.059	231	NS	-0.028 7
C.I.5-8	18	0.871	0.909	0.901	0.310	0.184	0.053	232	NS	0.020 0
C.I.5-11	18	0.966	0.918	0.910	0.289	0.169	0.047	232	ND	-0.026 6
C.I.5-12	32	0.965	0.943	0.939	0.207	0.116	0.022	228	NS	-0.013 9
C.I.5-27	20	0.900	0.902	0.893	0.327	0.196	0.059	231	***	-0.001 2
C.I.5-34	19	0.909	0.883	0.870	0.382	0.236	0.084	231	NS	-0.016 4
C.I.5-43	16	0.927	0.876	0.862	0.402	0.250	0.094	232	NS	-0.029 0
C.I.6-6	12	0.878	0.872	0.857	0.412	0.258	0.099	230	NS	-0.005 1

双亲未知时单亲本的联合非亲权排除率 combined non-exclusion probability (first parent): 0.000 000 02

已知单亲时另一亲本的联合非亲权排除率 combined non-exclusion probability (second parent): 8.274E-0012

双亲未知时父母本组合的联合非亲权排除率 combined non-exclusion probability (parent pair): 6.575E-0020

注: *NE-1P*. 双亲未知时单亲本单个位点的非亲权排除率; *NE-2P*. 已知单亲时另一亲本单个位点的非亲权排除率; *NE-PP*. 双亲未知时父母本组合的非亲权排除率

Notes: *NE-1P*. non-exclusion probability for one candidate parent; *NE-2P*. non-exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex; *NE-PP*. non-exclusion probability for a candidate parent pair

基因分型的准确性也有巨大的影响,电泳过程中仪器的长时间封闭运行易造成温度上升,导致不均匀的温度梯度和局部黏度的变化,可能对结果产生影响^[37],因此,在电泳过程中应采用适当的降温手段。同时毛细管荧光电泳在应用于基因分型时,片段之间差距越大分型越精确^[38],二碱基、三碱基重复微卫星标记在基因分型步骤中由于重复单位长度较短,不同重复次数的片段之间差距较小,读取毛细管电泳峰图数据时少次数的重复差距将会难以区分,易出现误读。而四碱基以上重复单位的微卫星标记重复次数相差一次便有4~6 bp的片段长度差距,在基因分型时更容易精确地辨认,分型结果相对更准确,更适合用于开展家系鉴定研究。

微卫星位点的多态性也对鉴定准确性有很大影响,在张春雷等^[13]对哲罗鱼的鉴定过程中发现,位点多态性越丰富则可用越少的微卫星标记达到准确鉴定的目的,同时随着候选亲本的规模扩大鉴定的准确率会降低。PCR过程中出现一些无法

预测的问题也会导致亲本子代基因型不匹配,王鸿霞等^[39]对凡纳滨对虾的亲子鉴定过程中,因PCR过程中出现一些不可避免的问题以及无效等位基因的影响,父母本实际匹配准确率为78%和88%。因此,为保证亲权鉴定的准确性,①尽可能挑选多态性高、无效等位基因频率低的微卫星位点,保证鉴别的有效性;②尽量使用4~6碱基重复微卫星标记,保障基因型的判读准确;③在基因分型时要校准数据,保证基因分型的准确率;④规范PCR过程,保证扩增过程的稳定准确。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

- [1] 卢小宁,马永康,付学言,等. 微卫星标记在鹌鹑遗传育种中的应用[J]. 当代畜禽养殖业, 2019(11): 3-5.
- Lu X N, Ma Y K, FU X Y, *et al.* Application of microsatellite markers in quail genetic breeding[J]. *Modern Livestock and Poultry Breeding Industry*, 2019(11):

- 3-5 (in Chinese).
- [2] 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[J]. 四川动物, 2014, 33(6): 874-878.
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[J]. Sichuan Journal of Zoology, 2014, 33(6): 874-878 (in Chinese).
- [3] 余成晨, 沈玉帮, 徐晓雁, 等. 草鱼生长相关的微卫星标记在选育群体中的验证[J]. 水产学报, 2021, 45(3): 321-332.
Yu C C, Shen Y B, Xu X Y, *et al.* Verification of microsatellite markers associated with growth traits in selected populations of grass carp (*Ctenopharyngodon idella*)[J]. Journal of Fisheries of China, 2021, 45(3): 321-332 (in Chinese).
- [4] Gao L Z, Xu H Y. Comparisons of mutation rate variation at genome-wide microsatellites: evolutionary insights from two cultivated rice and their wild relatives[J]. BMC Evolutionary Biology, 2008, 8(1): 11.
- [5] 杨兵, 林琳, 李纯厚, 等. 基于高通量测序的二长棘鲷微卫星标记开发与评价[J]. 南方水产科学, 2015, 11(4): 116-120.
Yang B, Lin L, Li C H, *et al.* Development and evaluation of microsatellite markers in *Parargyrops edata*[J]. South China Fisheries Science, 2015, 11(4): 116-120 (in Chinese).
- [6] 王军, 罗琦, 段茜, 等. 3种金线鲃属鱼类基因组微卫星的筛选与分析[J]. 贵州师范大学学报(自然科学版), 2019, 37(4): 19-24.
Wang J, Luo Q, Duan X, *et al.* Screening and analysis of microsatellites in the genomes of three *Sinocyclocheilus* fishes[J]. Journal of Guizhou Normal University (Natural Sciences), 2019, 37(4): 19-24 (in Chinese).
- [7] 李红梅. 基于基因组信息的大黄鱼 (*Pseudosciaena crocea*) 微卫星标记开发及应用 [D]. 舟山: 浙江海洋学院, 2014.
Li H M. New microsatellite satellite markers development based on whole genome sequencing information and its application in population genetics in larger yellow croaker[D]. Zhoushan: Zhejiang Ocean University, 2014 (in Chinese).
- [8] 白翠翠, 柳淑芳, 庄志猛. 虾虎鱼科(Gobiidae)基因组微卫星DNA的分布特征[J]. 渔业科学进展, 2016, 37(5): 9-15.
Bai C C, Liu S F, Zhuang Z M. Characteristic analysis of microsatellite DNA in the genome of Gobiidae[J]. Progress in Fishery Sciences, 2016, 37(5): 9-15 (in Chinese).
- [9] 辛苗苗, 张书环, 汪登强, 等. 多倍体中华鲟微卫星亲子鉴定体系的建立[J]. 淡水渔业, 2015, 318(4): 3-9.
Xin M M, Zhang S H, Wang D Q, *et al.* Parentage identification of polyploidy *Acipenser sinensis* based on microsatellite markers[J]. Freshwater Fisheries, 2015, 318(4): 3-9 (in Chinese).
- [10] 马温华, 张颖, 孙启凡, 等. 一种对未知猪检材进行个体识别和亲权鉴定的方法和系统: 中国, 201610876866.8 [P]. 2016-12-21.
Ma W H, Zhang Y, Sun Q F, *et al.* Method and system for performing individual identification and paternity identification on unknown pig subject: CN, 201610876866.8[P]. 2016-12-21 (in Chinese).
- [11] 王敏, 张新浩, 崔冉, 等. 利用微卫星标记鉴定德州驴亲子关系[J]. 中国畜牧兽医, 2019, 46(7): 2003-2011.
Wang M, Zhang X H, Cui R, *et al.* Parentage testing of Dezhou donkey with microsatellite marker[J]. China Animal Husbandry & Veterinary Medicine, 2019, 46(7): 2003-2011 (in Chinese).
- [12] 郭丽平, 徐丽, 朱淼, 等. 西门塔尔牛微卫星亲子鉴定体系的优化[J]. 畜牧兽医学报, 2013, 44(6): 871-879.
Guo L P, Xu L, Zhu M, *et al.* Optimization of microsatellite-based paternity testing system for Simmental[J]. Acta Veterinaria et zootechnica sinica, 2013, 44(6): 871-879 (in Chinese).
- [13] 张春雷, 佟广香, 匡友谊, 等. 哲罗鱼微卫星亲子鉴定的应用[J]. 动物学研究, 2010, 31(4): 395-400.
Zhang C L, Tong G X, Kuang Y Y, *et al.* Applicability of microsatellite DNA markers to the parental identification of *Hucho taimen* (Pallas)[J]. Zoological Research, 2010, 31(4): 395-400 (in Chinese).
- [14] 王丰, 张家华, 沈玉帮, 等. 青鱼野生与养殖群体遗传变异的微卫星分析[J]. 水生生物学报, 2019, 43(5): 939-944.
Wang F, Zhang J H, Shen Y B, *et al.* Microsatellite analysis of genetic variation of wild and cultural populations in black carp *Mylopharyngodon piceus*[J]. Acta Hydrobiologica Sinica, 2019, 43(5): 939-944 (in Chinese).
- [15] 何勇凤, 朱永久, 吴兴兵, 等. 基于微卫星标记的圆口铜鱼亲子鉴定技术[J]. 水生生物学报, 2019, 43(6): 1216-1223.
He Y F, Zhu Y J, Wu X B, *et al.* Parentage analysis of

- Coreius guichenoti* using microsatellites[J]. *Acta Hydrobiologica Sinica*, 2019, 43(6): 1216-1223 (in Chinese).
- [16] 任昆. 草鱼亲缘关系的微卫星鉴定[D]. 上海: 上海海洋大学, 2013.
- Ren K. The parentage assignment of Grass carp (*Ctenopharyngodon idellus*) using microsatellite markers[D]. Shanghai: Shanghai Ocean University, 2013 (in Chinese).
- [17] Fu J J, Shen Y B, Xu X Y, *et al.* Multiplex microsatellite PCR sets for parentage assignment of grass carp (*Ctenopharyngodon idella*)[J]. *Aquaculture International*, 2013, 21(6): 1195-1207.
- [18] Fu J J, Shen Y B, Xu X Y, *et al.* Genetic parameter estimates for growth of grass carp, *Ctenopharyngodon idella*, at 10 and 18 months of age[J]. *Aquaculture*, 2016, 450: 342-348.
- [19] Wang Y P, Lu Y, Zhang Y, *et al.* The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation[J]. *Nature Genetics*, 2015, 47(6): 625-631.
- [20] 高泽霞. 蓝鳃太阳鱼性控和性决定机制以及性别相关分子标记的研究[D]. 武汉: 华中农业大学, 2010.
- Gao Z X. Study on sex determination mechanism, sex control and sex related markers in bluegill sunfish *Lepomis macrochirus*[D]. Wuhan: Huazhong Agriculture University, 2010 (in Chinese).
- [21] 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[J]. 中国海洋大学学报, 2006, 36(2): 249-254, 272.
- Cui J Z, Shen X Y, Yang G P, *et al.* The Analysis of simple sequence repeats in *Takifugu rubripes* genome[J]. *Periodical of Ocean University of China*, 2006, 36(2): 249-254, 272 (in Chinese).
- [22] 戚文华, 蒋雪梅, 肖国生, 等. 猪全基因组中微卫星分布规律[J]. 畜牧与兽医, 2014, 46(8): 9-13.
- Qi W H, Jiang X M, Xiao G S, *et al.* Distribution regularities of microsatellites in the pig genome[J]. *Animal Husbandry & Veterinary Medicine*, 2014, 46(8): 9-13 (in Chinese).
- [23] 余泉友, 李斌, 李关荣, 等. 蚊子全基因组中微卫星的丰度及其分布[J]. *生物化学与生物物理进展*, 2005, 32(5): 435-441.
- Yu Q Y, Li B, Li G R, *et al.* Abundance and distribution of microsatellites in the entire mosquito genome[J]. *Progress in Biochemistry and Biophysics*, 2005, 32(5): 435-441 (in Chinese).
- [24] Subramanian S, Mishra R K, Singh L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions[J]. *Genome Biology*, 2003, 4(2): R13.
- [25] 段永楠, 刘奕, 胡隐昌, 等. 美丽硬仆骨舌鱼全基因组微卫星分布规律特征[J]. *中国农学通报*, 2019, 35(23): 152-158.
- Duan Y N, Liu Y, Hu Y C, *et al.* Distribution regularity of microsatellites in *Scleropages formosus* genome[J]. *Chinese Agricultural Science Bulletin*, 2019, 35(23): 152-158 (in Chinese).
- [26] 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.
- 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 (in Chinese).
- [27] 魏朝明, 孔光耀, 廉振民, 等. 蜜蜂全基因组中微卫星的丰度及其分布[J]. *昆虫知识*, 2007, 44(4): 501-504.
- Wei Z M, Kong G Y, Lian Z M, *et al.* Abundance and distribution of microsatellites in the entire *Apis mellifera* genome[J]. *Chinese Bulletin of Entomology*, 2007, 44(4): 501-504 (in Chinese).
- [28] Katti M V, Ranjekar P K, Gupta V S. Differential distribution of simple sequence repeats in eukaryotic genome sequences[J]. *Molecular Biology and Evolution*, 2001, 18(7): 1161-1167.
- [29] Li C Y, Li J B, Zhou X G, *et al.* Frequency and distribution of microsatellites in the genome of filamentous fungus, *Neurospora crassa*[J]. *Agricultural Sciences in China*, 2005, 4(2): 118-124.
- [30] 栾生, 孔杰, 王清印, 等. 日本囊对虾(*Marsupenaeus japonicus*)基因组微卫星特征分析[J]. *自然科学进展*, 2007, 17(6): 731-740.
- Luan S, Kong J, Wang Q Y, *et al.* The analysis of microsatellite in *Marsupenaeus japonicus* genome[J]. *Progress in Natural Science*, 2007, 17(6): 731-740 (in Chinese).
- [31] Lai Y L, Sun F Z. The relationship between microsatellite slippage mutation rate and the number of repeat units[J]. *Molecular Biology and Evolution*, 2003, 20(12): 2123-2131.
- [32] 高焕, 刘萍, 孟宪红, 等. 中国对虾(*Fenneropenaeus chinensis*)基因组微卫星特征分析[J]. *海洋与湖沼*, 中国水产学会主办 sponsored by China Society of Fisheries

- 2004, 35(5): 424-431.
- Gao H, Liu P, Meng X H, *et al.* Analysis of microsatellite sequences in Chinese shrimp *Fenneropenaeus chinenses* genome[J]. *Oceanologia et Limnologia Sinica*, 2004, 35(5): 424-431 (in Chinese).
- [33] Schorderet D F, Gartler S M. Analysis of CpG suppression in methylated and nonmethylated species[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89(3): 957-961.
- [34] 任昆, 白俊杰, 樊佳佳, 等. 草鱼的微卫星亲权鉴定[J]. *南方农业学报*, 2013, 44(8): 1367-1371.
- Ren K, Bai J J, Fan J J, *et al.* Parentage identification of grass carp (*Ctenopharyngodon idella*) using micro-satellites[J]. *Journal of Southern Agriculture*, 2013, 44(8): 1367-1371 (in Chinese).
- [35] 武栋. 草鱼家系选育中的亲子鉴定研究[J]. *山西水利科技*, 2018(3): 90-93.
- Wu D. Study on parentage test in breeding of grass carp family[J]. *Shanxi Hydrotechnics*, 2018(3): 90-93 (in Chinese).
- [36] O'Reilly P T, Herbinger C, Wright J M. Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites[J]. *Animal Genetics*, 1998, 29(5): 363-370.
- [37] 叶宪曾, 张新祥. 仪器分析教程 [M]. 2 版. 北京: 北京大学出版社, 2007: 386-387.
- Ye X Z, Zhang X X. Tutorial of Instrumental Analysis[M]. 2nd ed. Beijing: Peking University Press, 2007: 356-387 (in Chinese).
- [38] Rehman A, 李玉秋, 赵洪锟, 等. 毛细管电泳在基因分型中的应用[J]. *分子植物育种*, 2018, 16(14): 4750-4755.
- Rehman A, Li Y Q, Zhao H K, *et al.* Application of capillary electrophoresis in genotyping[J]. *Molecular Plant Breeding*, 2018, 16(14): 4750-4755 (in Chinese).
- [39] 王鸿霞, 吴长功, 张留所, 等. 微卫星标记应用于凡纳滨对虾家系鉴别的研究[J]. *遗传*, 2006, 28(2): 179-183.
- Wang H X, Wu C G, Zhang L S, *et al.* The application of microsatellite markers for parentage determination in selective breeding of Pacific white shrimp (*Litopenaeus vannamei*)[J]. *Hereditas (Beijing)*, 2006, 28(2): 179-183 (in Chinese).

Analysis of microsatellite in the entire grass carp (*Ctenopharyngodon idella*) genome and the application in parentage identification

HUANG Weijie^{1,2}, GUO Xiangzhao³, ZHANG Zihao^{1,2}, DONG Qiang^{1,2},
XIONG Xuemei^{1,2}, GAO Zexia^{1,2*}

- (1. Key Laboratory of Freshwater Animal Breeding, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Green Development for Conventional Aquatic Biological Industry in the Yangtze River Economic Belt, Ministry of Education, Engineering Technology Research Center for Fish Breeding and Culture in Hubei Province, College of Fisheries, Huazhong Agricultural University, Wuhan 430070, China;
2. Hubei Hongshan Lab, Wuhan 430070, China;
3. Animal Husbandry and Fisheries Research Center of Guangdong Haid Group Co., Ltd., Guangzhou 511400, China)

Abstract: Grass carp (*Ctenopharyngodon idella*) is the most productive freshwater aquaculture species in China, which is of great importance for Chinese aquaculture industry. Developing molecular marker assisted breeding program for *C. idella* has important theoretical value and practical significance. Microsatellite markers are popular in breeding studies with characterization of wide distribution, large numbers and high polymorphism. In this study, microsatellite loci were mined based on whole genome sequences of *C. idella* and the distribution characteristics of the loci were analyzed. In the meantime, microsatellite markers with 4-6 bases repeats with high polymorphism and high accuracy of paternity test were developed. The results showed that 677 363 microsatellite sequences were found in the 900.51 Mb genome sequence of *C. idella*, with a total length of 12 835 407 bp, accounting for 1.425 4% of the whole genome length. The average distance of SSR was 1 329.43 bp. The mononucleotide repeats motifs (52.85%) were the most common, followed by dinucleotide repeat (31.44%), tetranucleotide repeat (8.05%), trinucleotide repeat (6.47%), pentanucleotide repeat (1.12%), hexanucleotide repeat (0.07%). The preponderance of 1-6 base repeat motifs were A/T, AC/GT, AAT/ATT, AGAT/ATCT, AATAT/ATATT and AACCT/AGGGTT, respectively. The distributions of copy numbers in different microsatellites repeat types showed there was a significant negative correlation between repeats motifs length and mean copy number ($r = -0.858 1$, $P = 0.028 8$). The copy numbers of mononucleotide repeats were focused mainly on 10-29 (98.11%), dinucleotide repeats 6-25 (96.80%), trinucleotide repeats 5-11 (95.14%), tetranucleotide repeats 5-10 (88.48%), pentanucleotide repeats 5-10 (84.93%), hexanucleotide repeats 5-7 (87.42%). The 110 loci with 4-6 base repeats were randomly selected and primers were designed. Among these loci, 50 loci were amplified successfully with mixed DNA of *C. idella* used as a template and 15 loci showed high polymorphism when analyzed in its breeding parent population. Cervus 3.0 analysis results showed that average expected heterozygosity of 15 loci was 0.802-0.959. The accuracy of simulated paternity test was 100% with 95% confidence degree. These 15 loci were used to identify 192 offspring from 20 families, and all offspring matched to the correct parents successfully with the accuracy of identification being 100%. In this study, we analyzed the microsatellite characteristics of *C. idella*, screened out highly polymorphic microsatellite markers with 4-6 base repeats and carried our parental identification of *C. idella* by using these highly polymorphic microsatellite loci. At the same time, it was proved that 4-6-base repeat microsatellite is more suitable for paternity test than traditional 2-3-base repeat. The study summarized the Characteristics of whole genome microsatellites of *C. idella* and laid a foundation for the application of *C. idella* microsatellites in its breeding program.

Key words: *Ctenopharyngodon idella*; genome; microsatellite; parentage identification

Corresponding author: GAO Zexia. E-mail: gaozx@mail.hzau.edu.cn

Funding projects: National Natural Science Foundation of China (31872559); Modern Agriculture Industry Technology System Construction Projects of China (CARS-46-08); China Postdoctoral Science Foundation (2019M652829)