安徽农业大学学报, 2026, 53(4): 644-651
Journal of Anhui Agricultural University [DOI] 10.13610/j.cnki.1672-352x.20260915.001

崔冠伍，杨 坤，李 陈，等. 长吻鮠全基因组微卫星分布特征研究[J]. 安徽农业大学学报, 2026, 53(4): 644-651.
Cui Guanwu, Yang Kun, Li Chen, et al. Study on the microsatellite distribution characteristics in the whole genome of Chinese longsnout catfish (*Leiocassis longirostris*)[J]. Journal of Anhui Agricultural University, 2026, 53(4): 644-651.

长吻鮠全基因组微卫星分布特征研究

崔冠伍[1]，杨 坤[2]，李 陈[1]，孙 雨[1]，汪 鹏[1]，赵秀侠[2]，
高 娜[2]，崔 凯[2]，梁阳阳[2]，卢文轩[2]，方 婷[2*]
(1. 安徽省港航建设投资集团有限公司，安徽 合肥 230051；
2. 安徽省农业科学院水产研究所/水产增养殖安徽省重点实验室，安徽 合肥 230001)

摘 要:【目的】本研究基于 NCBI 数据库已公开的长吻鮠（*Leiocassis longirostris*）全基因组数据，利用 MISA 软件开展其全基因组水平微卫星的筛选与分布特征分析。【方法】联合基因组 General Feature Format（GFF）注释文件与 MISA 输出文件，对基因编码区外显子中嵌有微卫星序列的基因进行 Gene Ontology（GO）功能分类注释和 Kyoto Encyclopedia of Genes and Genomes（KEGG）通路富集分析。【结果】在长吻鮠基因组中共筛选出 1 180 545 个微卫星，总长度为 28 807 417 bp，占基因组序列总长度的 4.14%。其中，二碱基类型数量最多（占比 66.59%），其次为单碱基类型（占比 18.08%），六碱基类型数量最少（占比 0.07%）。【结论】初步定位结果显示，有 1 904 个微卫星分布在外显子区，涉及 1 705 个基因。GO 富集最显著的条目为 RNA 结合活性和信号转导过程，提示微卫星关联基因在核酸互作及细胞信号调控中具有重要作用。KEGG 通路富集分析显示前 5 条最显著富集的通路为 MAPK 信号通路、cAMP 信号通路、PI3K-Akt 信号通路、钙信号通路和黏着斑信号通路，这些通路均与细胞信号传导、生长调控及应激响应密切相关。研究结果可为长吻鮠微卫星标记的开发提供基础资料。

关键词: 长吻鮠；微卫星；全基因组分布
中图分类号: S917.4 **文献标志码:** A **文章编号:** 1672−352X (2026)04−0644−08

Study on the microsatellite distribution characteristics in the whole genome of Chinese longsnout catfish (*Leiocassis longirostris*)

Cui Guanwu[1], Yang Kun[2], Li Chen[1], Sun Yu[1], Wang Peng[1], Zhao Xiuxia[2],
Gao Na[2], Cui Kai[2], Liang Yangyang[2], Lu Wenxuan[2], Fang Ting[2]
(1. Anhui Provincial Port & Shipping Construction Investment Group Co., Ltd, Hefei 230051, China; 2. Anhui Key Laboratory of Freshwater Aquaculture and Enhancement, Institute of Fisheries Research, Anhui Academy of Agricultural Sciences, Hefei 230001, China)

Abstract: [**Objective**] Based on the publicly available whole-genome data of *Leiocassis longirostris* in the National Center for Biotechnology Information (NCBI) database, this study aimed to screen microsatellites and analyze their distribution characteristics at the whole-genome level using MISA software. [**Method**] By integrating the genome's General Feature Format (GFF) annotation files and MISA output files, Gene Ontology (GO) functional classification annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed on genes harboring microsatellite sequences within the exons of coding regions. [**Result**] A total of 1 180 545 microsatellites were identified in the genome of *L. longirostris*, with a total length of 28 807 417 bp, accounting for 4.14% of the total genome sequence length. Among these, dinucleotide microsatellites were the most abundant (66.59%), followed by mononucleotide ones (18.08%), while hexanucleotide microsatellites were the least frequent (0.07%). [**Conclusion**] The preliminary localization results revealed that 1 904 microsatellites were distributed in

收稿日期: 2026-01-21；**录用日期:** 2026-04-22；**网络出版日期:** 2026-09-16
基金项目: 淮河干流航道（三河尖至蚌埠闸段）整治工程对淮河淮南段长吻鮠国家级水产种质资源保护区影响生态修复研究项目（2024STBC01）；安徽省重点水域水生生物资源监测项目（2024BFAFZ02936）；安徽省现代农业产业技术体系建设专项（皖农科函[2021]711 号）
作者简介: 崔冠伍，工程师，E-mail：984493306@qq.com
* **通信作者:** 方 婷，博士，副研究员， E-mail：fangting_candy@163.com

exonic regions, involving 1 705 genes. The most significantly enriched GO terms were RNA binding activity and signal transduction processes, suggesting the core role of microsatellite-associated genes in nucleic acid interactions and cellular signal regulation. KEGG pathway enrichment analysis showed that the top five most significantly enriched pathways were the MAPK signaling pathway, cAMP signaling pathway, PI3K-Akt signaling pathway, calcium signaling pathway, and focal adhesion signaling pathway, all of which are closely related to cellular signal transduction, growth regulation, and stress response. These findings provide fundamental data for the development of microsatellite markers in *L. longirostris*.

Keywords: *Leiocassis longirostris*; microsatellite; genome-wide distribution

微卫星（Microsatellite）是由 1～6 个核苷酸为基本单位串联重复的简单重复序列（Simple Sequence Repeat，SSR）[1]，广泛分布于真核生物[2]、原核生物[3]以及病毒[4]基因组中。微卫星具有多态性高、突变率适中、遗传稳定性强等特点，在物种遗传多样性评估与种质资源保护[5-6]、分子育种与标记辅助选择[7-8]、遗传连锁图谱构建与基因定位[9-10]等研究中已有广泛应用。

长吻鮠（*Leiocassis longirostris*）隶属于鲇形目、鲿科、鮠属，又名江团、淮王鱼等，主要分布于长江干支流及其流域的大型湖泊[11]。受过度捕捞、水利工程建设、水域污染等影响，长吻鮠野生资源量锐减[12]。人工养殖已成为满足市场需求的主要途径，但苗种供应不足、品质参差不齐等问题亟待解决[11]。微卫星标记可精准解析长吻鮠种群的遗传结构、遗传多样性水平及种质资源亲缘关系[13-15]，为优质亲本筛选、良种选育、苗种纯度鉴定提供核心理论依据与技术支撑，最终实现苗种供应能力提升与品质标准化。现已报道的长吻鮠微卫星研究主要侧重于微卫星标记的开发[13]和种群遗传多样性分析[12-16]，其他研究有限[17-18]。目前，长吻鮠全基因组序列已经发布[19]，这为在全基因组水平开发高多态性微卫星标记提供了可能。

截至目前，尚未见关于长吻鮠全基因组微卫星分布特征的报道。本研究利用 MISA（MIcroSAtellite identification tool）微卫星筛选软件，从长吻鮠基因组数据中挖掘微卫星信息，并针对外显子区含有微卫星的基因开展 GO 功能注释和 KEGG 通路富集分析，以深化对相关微卫星在基因序列中功能定位的认识。研究结果可为开发长吻鮠微卫星标记提供基础资料，进而为其种质资源评估和遗传育种研究等提供数据支持。

1　材料与方法

1.1　基因组序列

从 NCBI Genome 数据库下载长吻鮠基因组，登录号为 GCA_041146805.1，基因组的大小为 695.9 Mb[19]。

1.2　微卫星筛选与统计

利用 MISA 2.1 软件对下载的长吻鮠全基因组序列进行扫描，筛选符合检索条件的完整型微卫星。参照 MISA 软件默认参数标准，将单碱基、二碱基微卫星的最小重复次数分别设定为 10 次和 6 次，三至六碱基微卫星的最小重复次数则统一设定为 5 次；统计分析中，将可循环序列与碱基互补序列合并归为同一类型。使用 MISA 软件筛选后，利用 Excel 对检索结果进行汇总和分析。

1.3　GO 注释和 KEGG 富集分析

利用长吻鮠基因组 GFF 注释文件提取 CDS 区域的起始和终止信息，同时从 MISA 软件生成的 misa.fasta.misa 文件中获取 SSR 位点的起始和终止信息；接着利用 bedtools intersect 工具（命令：bedtools intersect -a cds.locus -b ssr.locus -wa -wb > cds _and_ ssr. overlap）获得两者的交叠信息；再通过 less 与 awk 筛选出 SSR 起始和终止位置完全落入 CDS 区域的基因（命令：less -SN cds_and_ssr.overlap | awk '$2<=$6 && $3>=$7' > ssr_in_cds）。利用 DIAMOND 软件，将注释得到的基因集对应的蛋白序列与 KEGG 蛋白数据库进行比对搜索，获取基因集关联的 KEGG 基础注释信息；随后使用 KOBAS 软件（http://kobas.cbi.pku.edu.cn/），将上述 KEGG 注释信息与 KEGG ORTHOLOGY（KO）数据库及 PATHWAY 通路数据库进行关联匹配。InterPro 数据库注释采用 InterProScan 工具基于默认参数进行；GO term 注释通过 InterPro 数据库注释结果映射获得。最后使用 clusterProfiler 对筛选得到的基因集进行 GO 和 KEGG 富集分析，以明确目标基因的功能分类及参与的核心代谢通路。

2　结果与分析

2.1　长吻鮠各重复类型微卫星分布特征

长吻鮠全基因组长度为 695.9 Mb（695 900 828 bp），从中筛选获得 1～6 碱基重复完整型微卫星数量为 1 180 545 个，总长度为 28 807 417 bp，占长吻鮠全基因组序列长度的 4.14%。其中，二碱基类型

微卫星数量最多，共计 786 139 个，占微卫星总数的 66.59%;其次是单碱基类型微卫星有 213 441 个，占比 18.08%；六碱基类型微卫星数量最少，仅有 832 个，占比 0.07%（表 1）。

2.2 长吻鮠微卫星核心序列拷贝数分布

长吻鮠基因组中所有类型微卫星均呈现低拷贝数占绝对主导的左偏分布，高拷贝数微卫星在各类型中占比均较低（图 1）。同时，微卫星的整体数量随重复单元长度增加而显著下降，这与长链重复单元的复制稳定性更低、错配修复系统更易纠正其变异有关[20]。各类型微卫星的优势拷贝数存在差异：单碱基类型、二碱基类型以及三至六碱基类型微卫星数量最多的拷贝数分别为 10 次、6 次和 5 次。随后，微卫星数目随重复拷贝数增加急剧下降，其中四碱基、五碱基和六碱基重复 20 次以上的微卫星数量极少，仅零星分布。

表 1 长吻鮠基因组中不同类型微卫星统计

Table 1 Summary of different types of microsatellites in the genome of *L. longirostris*

重复类型	数量/个	占比/%	总长度/bp	占比/%	平均长度/bp	频率/（个·Mb^{-1}）	密度/（bp·Mb^{-1}）
单碱基	213 441	18.08	2 823 290	9.80	13.23	306.71	4 057.03
二碱基	786 139	66.59	20 289 978	70.43	25.81	1 129.67	29 156.42
三碱基	76 148	6.45	2 205 567	7.66	28.96	109.42	3 169.37
四碱基	94 985	8.05	3 113 556	10.81	32.78	136.49	4 474.14
五碱基	9 000	0.76	345 050	1.20	38.34	12.93	495.83
六碱基	832	0.07	29 976	0.10	36.03	1.20	43.08

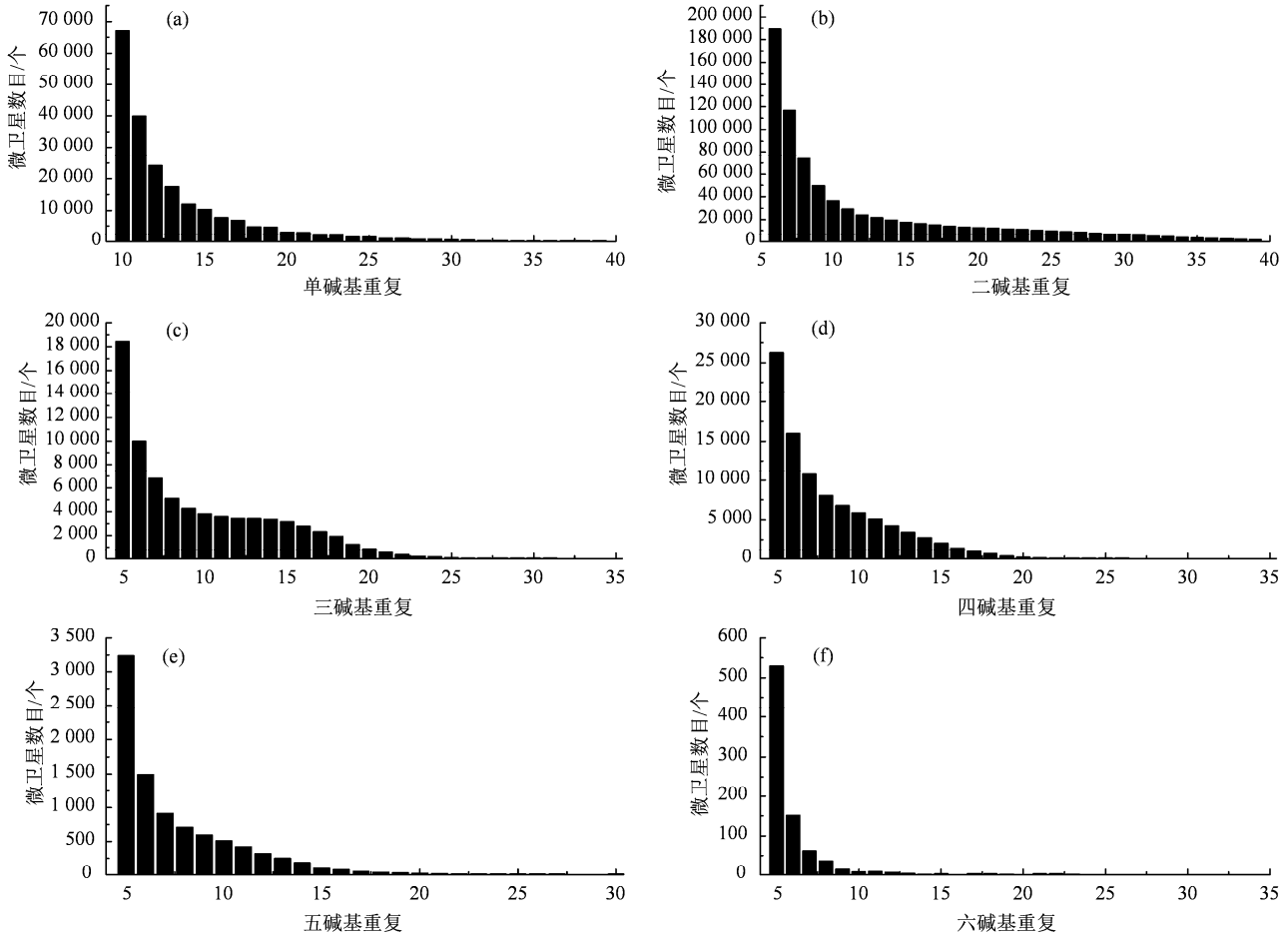


图 1 长吻鮠各类型微卫星重复拷贝数分布

Figure 1 Distribution of copy numbers of various types of microsatellites in *L. longirostris*

2.3 长吻鮠微卫星各重复类型特征

长吻鮠全基因组中出现频率最高的前 10 种微卫星重复类型的统计结果见表 2。

单碱基类型中，A 类别（A/T）微卫星有 209 403 个，占该类型总数目的 98.11%；二碱基类型中，AC 类别微卫星有 521 292 个，占该类型总数目的 66.31%，此外 AG 类别（210 650 个）和 AT 类别（53 440 个），分别占该类型总数目的 26.80%和 6.80%；

三碱基类型中，44 029 个 AAT 类别微卫星占该类型总数目的 57.82%，其次是 ATC 类别（9 110 个）和 AAG 类别（7 977 个），分别占三碱基总数目的 11.96%和 10.48%；四碱基类型中，筛选出 AGAT 类别 16 548 个，占该类型总数目的 17.42%，其次为 AAAT 类别（16 107 个）和 ACAG 类别（12 308 个），分别占该类型的 16.96%和 12.96%。五、六碱基类型微卫星未在出现次数排名前 10 的碱基类别列表中，表明其在长吻鮠基因组微卫星中所占比例较低。五碱基类型中以 AAGAG 类别有 1 248 个，占五碱基总数目的 13.87%，而六碱基类型中 AACCCT 类别仅有 169 个，占六碱基总数目的 20.31%（表 3）。

表 2　出现次数排名前 10 的重复碱基类型

Table 2　Top ten types of SSRs with the highest repeat copy numbers

SSR 类别	SSR 出现次数	占该碱基类型比例/%
AC	521 292	66.31
AG	210 650	26.80
A	209 403	98.11
AT	53 440	6.80
AAT	44 029	57.82
AGAT	16 548	17.42
AAAT	16 107	16.96
ACAG	12 308	12.96
ATC	9 110	11.96
AAG	7 977	10.48

表 3　长吻鮠基因组中各碱基重复类型中前 3 优势碱基类别

Table 3　Top three dominant base classes in each base repeat type in the *L. longirostris* genome

单碱基	二碱基	三碱基	四碱基	五碱基	六碱基
A	AC	AAT	AGAT	AAGAG	AACCCT
209 403[a]	521 292[a]	44 029[a]	16 548[a]	1 248[a]	169[a]
98.11%[b]	66.31%[b]	57.82%[b]	17.42%[b]	13.87%[b]	20.31%[b]
C	AG	ATC	AAAT	AATAT	AAATGT
4 038[a]	210 650[a]	9 110[a]	16 107[a]	1 016[a]	75[a]
1.89%[b]	26.80%[b]	11.96%[b]	16.96%[b]	11.29%[b]	9.01%[b]
	AT	AAG	ACAG	ACACT	ACACTC
	53 440[a]	7 977[a]	12 308[a]	615[a]	65[a]
	6.80%[b]	10.48%[b]	12.96%[b]	6.83%[b]	7.81%[b]

注：a. 微卫星的个数；b. 该类微卫星占其碱基类型的比例。

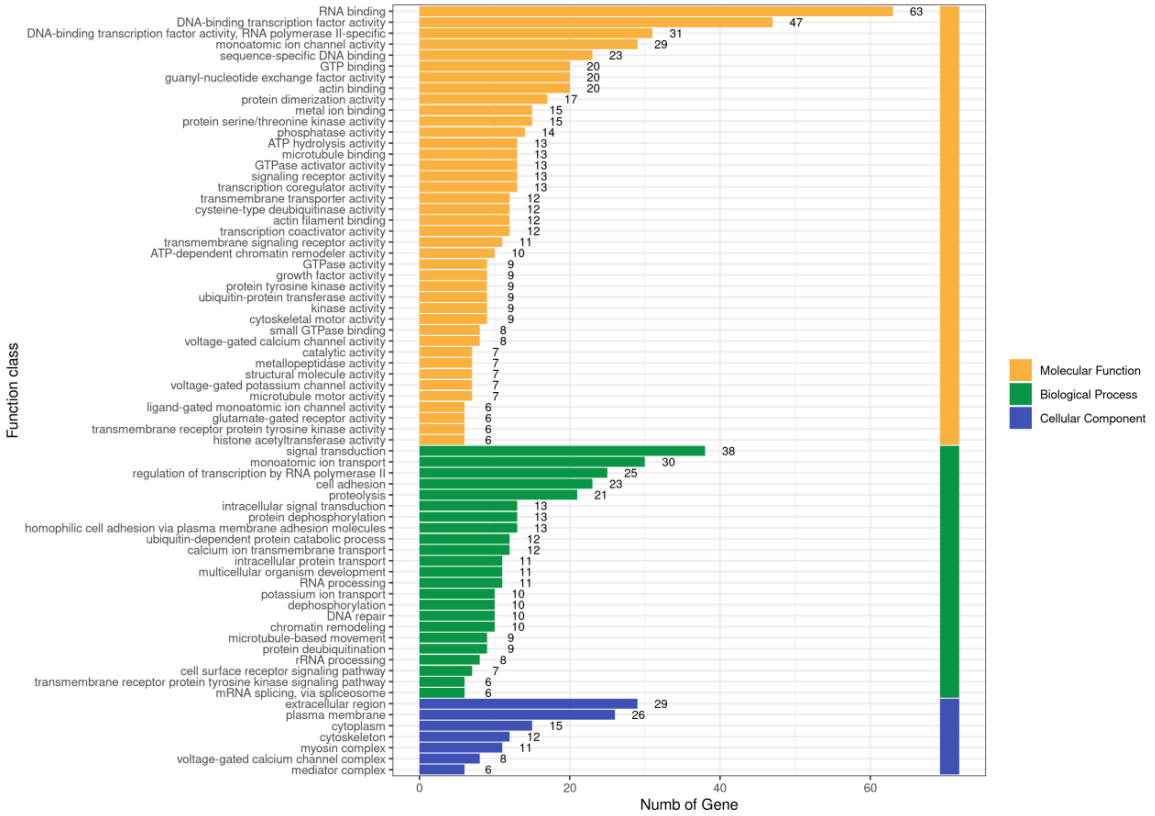


图 2　长吻鮠微卫星分布于外显子的基因 GO 功能注释

Figure 2　GO function annotation of genes with microsatellites in exons in *L. longirostris*

2.4 GO 功能注释和 KEGG 通路富集分析

微卫星定位结果表明，有 1 904 个微卫星序列定位于外显子区域，涉及 1 705 个基因。对相关编码基因开展 GO 注释分析，累计得到 342 个 GO 条目，以分子功能（Molecular function）类条目最多，共 173 个，在总条目中占比 50.58%，主要参与 RNA 结合（GO：0003723）和 DNA 结合转录因子活性（GO：0003700）等；其次是生物过程（Biological process），共 135 个条目，占比 39.47%，主要参与信号转导（GO：0007165）和单原子离子转运（GO：0006811）；此外，细胞成分（Cellular component）类条目最少，34 个条目中主要涉及细胞外区域（GO：0005576）和细胞膜（GO：0005886）（图 2）。GO 功能富集分析中，显著性较高的前 10 个条目主要参与基因表达调控及 RNA 代谢等相关过程（表 4）。

表 4 长吻鮠微卫星分布于外显子的基因 GO 富集

Table 4 GO enrichment of genes with microsatellites located in exons in *L. longirostris*

条目	GO 条目	*P* 值
转录共激活因子活性	0003713	4.68×10^{-7}
转录共调节因子活性	0003712	1.63×10^{-6}
RNA 结合	0003723	2.56×10^{-5}
染色质重塑	0006338	2.95×10^{-5}
ATP 依赖性染色质重塑因子活性	0140658	1.91×10^{-4}
组蛋白乙酰转移酶复合物	0000123	5.26×10^{-4}
组蛋白乙酰转移酶活性	0004402	6.13×10^{-4}
RNA 聚合酶 II 介导的转录调控	0006357	1.06×10^{-3}
钙离子跨膜转运	0070588	1.11×10^{-3}
电压门控钙离子通道活性	0005245	1.33×10^{-3}

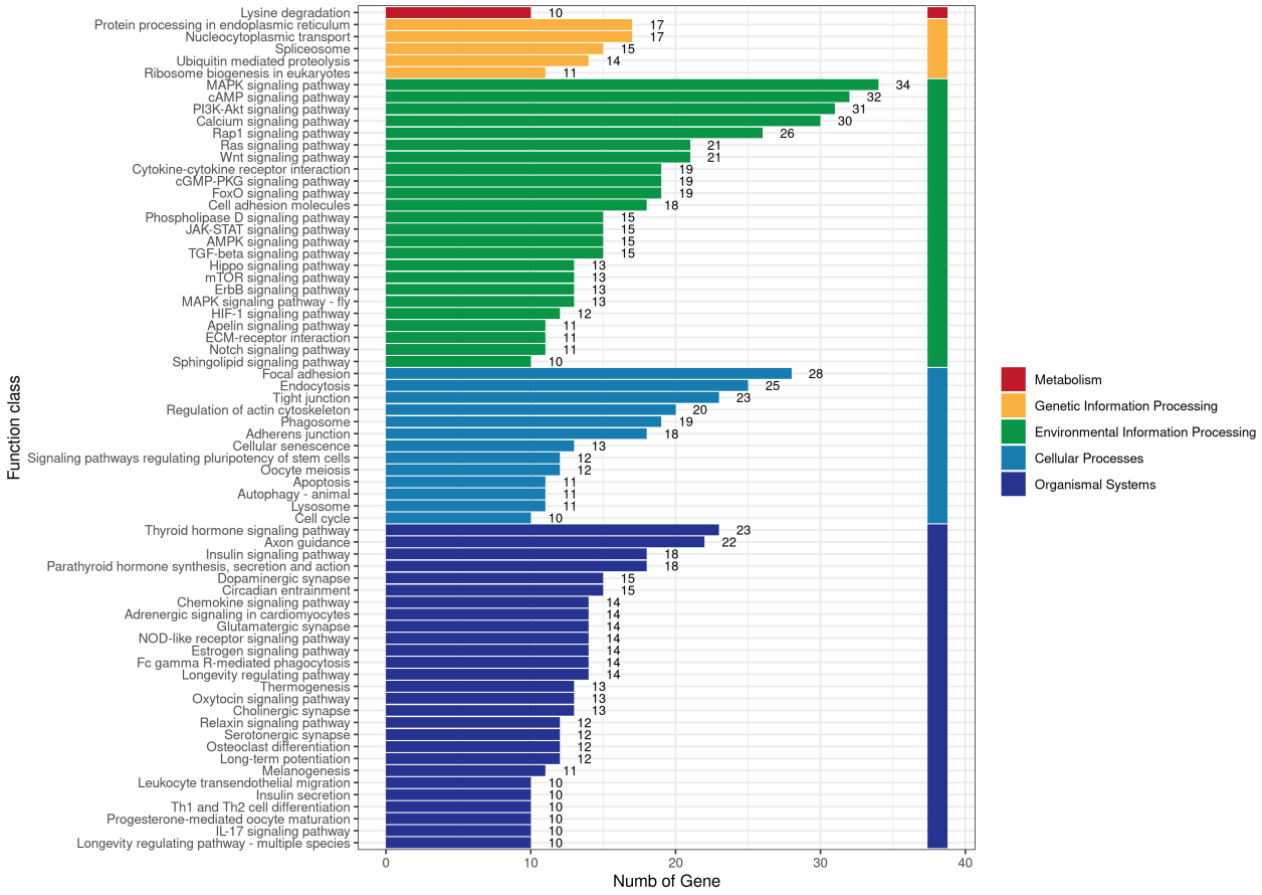


图 3 长吻鮠微卫星分布于外显子的基因 KEGG 通路富集

Figure 3 KEGG pathway enrichment of genes with microsatellites in exons in *L. longirostris*

对长吻鮠外显子微卫星相关基因开展 KEGG 通路富集分析，共富集到 198 条通路，其中黏合连接通路的富集显著性最高（*P*=0.002），而 MAPK 信号通路富集的基因数目最多，共计 34 个。所有富集通路可归为 5 大功能类别，其中有机体系统类别的通路富集数量最多（图 3）。

3 讨论与结论

本研究基于长吻鮠全基因组数据，利用 MISA 软件对其中完整型微卫星进行检索统计。在 695.9 Mb 的长吻鮠全基因组数据中筛选获得 1 180 545 个微卫星，其总长度 28.81 Mb，占全基因组总长度的

4.14%，该微卫星含量占比显著高于其同科近缘物种黄颡鱼（*Pelteobagrus fulvidraco*，1.80%）[21]和瓦氏黄颡鱼（*Pelteobagrus vachellii*，1.48%）[22]。黄颡鱼与瓦氏黄颡鱼同属，其微卫星占比差异仅为0.32%，表明物种亲缘关系越近，其基因组微卫星特征越相似[23]。而长吻鮠与黄颡鱼、瓦氏黄颡鱼分属不同属，在鲿科系统发育树中归属于不同分支且分化时间早于黄颡鱼属内物种的分化；在长时间的独立演化过程中，长吻鮠基因组微卫星位点可能在复制滑移机制驱动下发生持续变异，虽存在扩增与收缩的双向突变效应，但整体呈净扩增的累积趋势[14,18]，使其微卫星位点数量和重复长度增加，最终表现为高微卫星比例[24]。与其他水生生物如斑点叉尾鮰（*Ictalurus punctatus*，2.25%）[25]、草鱼（*Ctenopharyngodon idella*，1.43%）[26]、翘嘴鲌（*Culter alburnus*，1.35%）、团头鲂（*Megalobrama amblycephala*，1.35%）[27]、鲤鱼（*Cyprinus carpio*，0.91%）[28]以及马口鱼（*Opsariichthys bidens*，0.10%）[29]等相比，长吻鮠微卫星含量所占比例较高。不同物种微卫星数量在全基因组中的含量存在差异，可能源于不同物种间的基因组大小差异（斑点叉尾鮰 704.7 Mb，草鱼 900.51 Mb，翘嘴鲌 1.055 Gb，团头鲂 1.116 Gb，鲤鱼 1.7 Gb），也可能受物种进化差异的综合调控，该现象已被暗纹东方鲀（*Takifugu obscurus*）[30]、巨鲑（*Bagarius yarrelli*）[31]的全基因组微卫星特征相关研究所证实。

与此前研究报道的黄颡鱼[21]、瓦氏黄颡鱼[22]、斑点叉尾鮰[25]、团头鲂[27]、马口鱼[29]、中华鳉鲅（*Rhodeus sinensis*）[32]等鱼类相似，长吻鮠基因组中二碱基类型微卫星最多，有 786 139 个，占微卫星总数的 66.59%，其次为单碱基类型，占比为18.08%。而草鱼[26]、翘嘴鲌、鳡浪白鱼（*Anabarilius grahami*）、团头鲂[27]、鲤鱼[28]、中华绒螯蟹（*Eriocheir sinensis*）[33]、长江江豚（*Neophocaena asiaeorientalis asiaeorientalis*）[34]等物种的基因组中则以单碱基类型微卫星占据主导地位。不同水生生物物种中优势微卫星类型存在差异，黄纬杰等[26]推测这一现象的形成与不同物种在演化历程中形成的差异化环境适应策略密切相关。

除三碱基类型微卫星有轻微波动外，长吻鮠基因组中其他微卫星核心序列的拷贝数随重复次数增加而逐渐减少，这一趋势与黄颡鱼[21]、瓦氏黄颡鱼[22]、鲤鱼[28]等物种全基因组中微卫星核心序列拷贝数的变化规律一致。鱼类基因组中微卫星拷贝数均呈左偏分布，即低拷贝数（5～20 次）微卫星占绝大多数，高拷贝数（>50 次）微卫星极为罕见。这种分布符合“长度依赖性选择”假说：随着拷贝数增加，微卫星不稳定性升高，易引发突变和功能异常，因而受到自然选择的限制[35]。重复单元长度与平均拷贝数呈负相关[21-22]，这是鱼类微卫星最普遍且显著的特征之一。较长的重复单元（如五、六碱基）在 DNA 复制过程中更易被错配修复系统识别和纠正；同时，其形成的二级结构（如发卡结构）稳定性较低，限制了拷贝数的无限扩增。相反，较短的重复单元（如单、二碱基）复制滑移频率更高，更易积累拷贝数变异[36]。

在长吻鮠全基因组中，单碱基重复类型呈现极强的 A/T 偏向性，占比高达 98.11%，C/G 类别仅占1.89%，该特征与已发表的多数物种高度一致[21-29]。二碱基重复类型的优势类别为 AC、AG 和 AT，GC 类别占比极低，仅为 0.09%，与多数水生生物的分布特征相符[21-22,28-29,33]。此外，长吻鮠的三碱基、四碱基、五碱基和六碱基重复类型的微卫星均表现出明显的 A/T 碱基优势（表 3）。这一现象并非随机的基因组特征差异，而是由碱基理化性质与 DNA 复制修复的核心分子机制所驱动的保守演化结果：A/T 碱基对仅由两个氢键连接，相较于由三个氢键连接的 C/G 碱基对，其双链结构更易解旋，更有利于微卫星重复序列的扩增与世代保留[4]。同时，CpG 甲基化介导的胞嘧啶脱氨基反应会持续降解 C/G 重复序列，使其难以稳定积累，这一分子机制决定了真核生物单碱基重复的 A/T 偏向性在演化过程中高度保守[37]。

微卫星作为鱼类基因组中重要的分子标记，不同重复单元类型的生物学特征、分布规律及应用价值存在显著差异[34]。此前学者将微卫星用于长吻鮠遗传标记的研究表明，三碱基和四碱基重复微卫星是更精确、可靠的遗传标记，其中四碱基重复标记的实用性和稳定性更优[12-16]。长吻鮠三碱基重复类型中以 AAT 类别最多（占比 57.82%），其次为 ATC 类别和 AAG 类别，这与人类[38]以及其他水生生物物种（如瓦氏黄颡鱼[22]、马口鱼[29]）的情况相同。三碱基微卫星的异常变异会对鱼类的基因组功能、生长发育及种质特性产生负面影响：三碱基微卫星不稳定性可导致鱼类基因组稳定性下降，增加基因突变频率，进而引发肿瘤、发育异常等问题[39]；部分养殖鱼类因长期近亲繁殖，会出现三碱基微卫星异常扩增的情况，最终导致生长缓慢、抗逆性下降、繁殖力降低等种质退化现象[40]；少数研究还发现，特定基因中三碱基微卫星的扩增与神经发育缺陷相

关，如行为异常、运动失调等[41]。因此，对长吻鮠基因组中三碱基重复类型微卫星进行进一步挖掘分析，有利于长吻鮠遗传疾病的诊断和预防、种质资源的保护与利用以及分子育种工作的推进。

长吻鮠四碱基、五碱基、六碱基重复类型中，AGAT（17.42%）、AAAT（16.96%）、AAGAG（13.87%）、AATAT（11.29%）、AACCCT（20.31%）等类别微卫星占据优势，呈显著的 A/T 富集特征，符合鱼类基因组微卫星的普遍规律[21-29]。长吻鮠基因组中检索出的 AGAT、AAAT 和 ACAG 等四碱基微卫星优势类别与其他鱼类存在跨物种共性，同时呈现出物种特异性特征：AGAT 和 AAAT 类别在草鱼[26]和 5 种石斑鱼[42]全基因组中均有高频分布，而 ACAG 为长吻鮠特有四碱基优势类别，这可能与长吻鮠的物种特异性适应相关[18]。而 AACCCT 类型也在黄颡鱼[21]、草鱼[26]、鲤鱼[28]、马口鱼[29]中检出，表明该类别微卫星可能是淡水鱼类基因组中保守重复单元，具有较强进化保守性。关于四碱基、五碱基、六碱基类型微卫星的生物学意义，已有研究报道 CCTG 及其反向互补序列 CAGG 重复会提高染色体脆性发生率[43-44]；位于内含子中的五碱基重复扩增（如 ATTTC）可通过产生有毒 RNA 等方式，导致神经退行性疾病[45]；而六碱基重复在维持基因组三维空间结构中起关键作用[46]。结合本研究中长吻鮠这些微卫星的高丰度分布特征，推测其可能在长吻鮠染色体稳定性维持、基因表达调控及基因组结构塑造中发挥重要作用，为后续深入解析长吻鮠基因组功能提供新的切入点。

在明确长吻鮠全基因组微卫星分布规律的基础上，为探究微卫星在长吻鮠全基因组中的潜在功能，本研究对长吻鮠基因组编码区外显子中含有微卫星的基因进行 GO 注释，共注释得到 342 个 GO 条目；同时进行 KEGG 通路富集分析，共富集 198 条 KEGG 通路。GO 富集最为显著的条目为 RNA 结合活性和信号转导过程，分别涉及 63 个和 38 个相关基因，提示微卫星关联基因在核酸互作及细胞信号调控中的核心作用[21,28]。微卫星富集于 RNA 结合活性相关基因，其串联重复序列可改变 mRNA 二级结构、稳定性与翻译效率，重复长度变异会直接影响 RNA-蛋白结合效率，进而调控基因转录后加工、剪接与转运[35,41]。KEGG 通路富集分析结果显示，前 5 条富集最显著的通路依次为 MAPK 信号通路、cAMP 信号通路、PI3K-Akt 信号通路、钙信号通路和黏着斑信号通路，这些通路均与细胞信号传导、生长调控及应激响应密切相关[22]。MAPK、cAMP、PI3K-Akt、钙信号和黏着斑均为细胞生长、抗逆、增殖的核心通路，微卫星可通过改变基因编码区长度、氨基酸序列与蛋白构象，影响通路中关键受体、激酶与下游效应因子的功能。同时，微卫星的高突变特性可作为适应性突变热点，在环境胁迫下快速改变基因表达量与通路活性，提升长吻鮠对环境的适应能力[24]。GO 注释与 KEGG 通路富集联合分析表明，长吻鮠基因组编码区域的微卫星可能与其体内细胞信号调控、转录表达调控及应激响应过程密切相关，尤其是 RNA 结合相关基因和 MAPK/cAMP 等信号通路的富集，可作为分子标记辅助育种的候选靶标；后续可针对这些通路中的微卫星开展多态性筛选，快速定位与经济性状连锁的优势基因型，实现长吻鮠优质、速生、高抗品种的分子育种。

本研究系统解析了长吻鮠全基因组微卫星的分布特征及其与功能基因的关联模式，可为后续提升微卫星引物设计效率、筛选高多态性分子标记提供序列基础和筛选依据。研究获得的位于外显子区及信号通路相关基因中的微卫星位点，可作为候选功能标记用于长吻鮠分子辅助育种研究。本研究结果为后续开展育种实用分子标记开发、功能验证及种质创新研究提供了重要的数据支撑。

参考文献：

[1] 朱滨, 常剑波. 微卫星 DNA 及其在鱼类中的应用[J]. 水生生物学报, 1999(6): 721-728.

[2] 史雯茜, 马秋月, 陈梓馨, 等. 漾濞槭全基因组微卫星分布特征[J]. 中南林业科技大学学报, 2024, 44(12): 178-186.

[3] Ham H, Lee Y H, Moh S H, et al. Development of strain specific simple sequence repeats and assessment of genetic diversity in *Erwinia amylovora* from marker selection to phylogenetic analysis[J]. Sci Rep, 2025, 15: 30357.

[4] Vishwakarma A, Nasrin T, Hoque M, et al. Elucidating the microsatellite signature of the tri-partite genomes of the viral family Peribunyaviridae[J]. Curr Microbiol, 2025, 82(5): 191.

[5] 胡玉婷, 周华兴, 段国庆, 等. 基于 SSR 和 D-loop 序列评估安徽青鱼养殖群体的遗传多样性和遗传结构[J]. 安徽农业大学学报, 2026, 53(1): 38-47.

[6] 韩叶, 张莹, 闫春梅, 等. 增殖放流对图们江流域大麻哈鱼洄游群体遗传多样性的影响[J]. 水产学杂志, 2025, 38(3): 1-7.

[7] 赵娜, 常剑波, 陶江平, 等. 基于微卫星标记的中华鲟亲子关系判别及案例分析[J]. 水生态学杂志, 2023, 44(5): 92-99.

[8] 王会虎, 任超, 郑国栋, 等. 团头鲂选育家系亲子鉴定方法的建立[J]. 水生生物学报, 2024, 48(8): 1378-1384.

[9] Qiu G F, Xiong L W, Liu Z Q, et al. A first generation microsatellite-based linkage map of the Chinese mitten crab *Eriocheir sinensis* and its application in quantitative trait loci (QTL) detection[J]. Aquaculture, 2016, 451: 223-231.

[10] Ma H T, Yu D M, Li J, et al. Construction of first genetic linkage map based on microsatellite markers and characterization of di- and tri-nucleotide microsatellite markers for *Crassostrea* hongkongesis[J]. Aquaculture, 2022, 556: 738272.
[11] 段元亮, 黄志鹏, 张露, 等. 长吻鮠养殖技术与种质资源研究进展及发展展望[J]. 西南农业学报, 2025, 38(12): 2600-2610.
[12] 王红莹, 黄文清. 应用微卫星标记分析长江流域长吻鮠 4 个群体的遗传多样性[J]. 河南农业科学, 2011, 40(2): 146-148, 160.
[13] Zhang L, Mou C Y, Zhou J, et al. Genetic diversity of Chinese longsnout catfish (*Leiocassis longirostris*) in four farmed populations based on 20 new microsatellite DNA markers[J]. Diversity, 2022, 14(8): 654.
[14] Hou Y L, Ye H, Song X H, et al. Genetic diversity and population structure of Chinese longsnout catfish (*Leiocassis longirostris*) using microsatellite DNA markers[J]. Fishes, 2024, 9(1): 35.
[15] 于爱清, 施永海, 徐嘉波, 等. 长吻鮠 5 个养殖群体遗传多样性的微卫星分析[J]. 大连海洋大学学报, 2025, 40(4): 618-627.
[16] Jin Y J, He K, Xiang P, et al. Temporal genetic variation of the Chinese longsnout catfish (*Leiocassis longirostris*) in the Upper Yangtze River with resource decline[J]. Environ Biol Fish, 2022, 105(9): 1139-1151.
[17] Liang F F, Yang Z R, Liu W, et al. Construction of a genome-wide linkage map and QTL mapping for growth and hypoxia tolerance traits in Chinese longsnout catfish (*Leiocassis longirostris*)[J]. Genomics, 2025, 117(6): 111157.
[18] Duan Y L, Wu W Q, Huang Z P, et al. Research on longsnout catfish (*Leiocassis longirostris*) worldwide: current status and future prospects[J]. Rev Aquac, 2025, 17(4): e70092.
[19] He W P, Zhou J, Li Z, et al. Chromosome-level genome assembly of the Chinese longsnout catfish *Leiocassis longirostris*[J]. Zool Res, 2021, 42(4): 417-422.
[20] Ellegren H. Microsatellites: simple sequences with complex evolution[J]. Nat Rev Genet, 2004, 5(6): 435-445.
[21] 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[J]. 基因组学与应用生物学, 2020, 39(12): 5488-5498.
[22] 彭冶, 李杰, 王涛, 等. 瓦氏黄颡鱼全基因组微卫星的分布特征及其定位的初步研究[J]. 南方水产科学, 2022, 18(1): 90-98.
[23] 王月月, 刘雪雪, 董坤哲, 等. 7 种家养动物全基因组微卫星分布的差异研究[J]. 中国畜牧兽医, 2015, 42(9): 2418-2426.
[24] Lei Y, Zhou Y, Price M, et al. Genome-wide characterization of microsatellite DNA in fishes: survey and analysis of their abundance and frequency in genome-specific regions[J]. BMC Genom, 2021, 22: 421.
[25] 钟立强, 张世勇, 姜虎成, 等. 斑点叉尾鮰全基因组完整型微卫星生物信息学分析[J]. 基因组学与应用生物学, 2022, 41(3): 504-513.
[26] 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[J]. 水产学报, 2022, 46(2): 161-172.
[27] 刘士力, 刘一诺, 李飞, 等. 3 种鲀亚科鱼类全基因组微卫星分布特征分析[J]. 上海海洋大学学报, 2023, 32(6): 1176-1183.
[28] 梁霞, 王慧琪, 马宇璇, 等. 鲤鱼(*Cyprinus carpio*)全基因组微卫星分布特征研究[J]. 南京师大学报(自然科学版), 2021, 44(3): 103-111.
[29] 葛健辉, 关文志, 任晋东, 等. 马口鱼全基因组简单重复序列特征分析与多态性标记开发[J]. 浙江农业学报, 2023, 35(11): 2584-2593.
[30] 徐杰杰. 暗纹东方鲀全基因组微卫星标记的开发及应用[D]. 南京: 南京师范大学, 2021.
[31] 杨汶珊, 唐荣叶, 苏孟园, 等. 巨鲇(*Bagarius yarrelli*)全基因组微卫星分布特征分析[J]. 南京师范大学学报(工程技术版), 2021, 21(3): 62-68.
[32] 熊良伟, 王帅兵, 封琦, 等. 基于高通量测序的中华鳑鲏基因组微卫星特征分析及标记开发[J]. 江苏农业科学, 2018, 46(18): 164-168.
[33] 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(S2): 2422-2429.
[34] 田镇, 阙江龙, 饶榕城, 等. 长江江豚全基因组微卫星分布特征分析[J]. 海洋渔业, 2024, 46(5): 551-558.
[35] Chistiakov D A, Hellemans B, Volckaert F A M. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics[J]. Aquaculture, 2006, 255(1/2/3/4): 1-29.
[36] Ditcharoen S, De Menezes Cavalcante Sassi F, Bertollo L A C, et al. Comparative chromosomal mapping of microsatellite repeats reveals divergent patterns of accumulation in 12 Siluridae (Teleostei: Siluriformes) species[J]. Genet Mol Biol, 2020, 43(4): e20200091.
[37] Schorderet D F, Gartler S M. Analysis of CpG suppression in methylated and nonmethylated species[J]. Proc Natl Acad Sci U S A, 1992, 89(3): 957-961.
[38] 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 Biol, 2003, 4(2): R13.
[39] 瞿才惠, 瞿印权, 贺艳, 等. 许氏平鲉和褐菖鲉全基因组微卫星分布规律及其比较分析[J]. 水产学报, 2025, 49(5): 31-42.
[40] Wang J J, Li J T, Ge Q Q, et al. Effects of inbreeding on genetic characteristic, growth, survival rates, and immune responses of a new inbred line of *Exopalaemon carinicauda*[J]. Int J Genom, 2020, 2020: 5735968.
[41] Kelley K, Chang S E, Lin S L. Mechanism of repeat-associated microRNAs in fragile X syndrome[J]. Neural Plast, 2012, 2012: 104796.
[42] 王旭蕾, 高进, 齐鑫, 等. 5 种石斑鱼全基因组微卫星筛选与特征分析[J]. 渔业科学进展, 2024, 45(3): 149-158.
[43] Papp D, Hernandez L A, Mai T A, et al. Massive contractions of myotonic dystrophy type 2-associated CCTG tetranucleotide repeats occur *via* double-strand break repair with distinct requirements for DNA helicases[J]. G3-Genes Genom Genet, 2024, 14(2): jkad257.
[44] Yang Y Q, Wang Y, Yan Z Z, et al. Effects of interrupting residues onDNAdumbbell structures formed byCCTG-tetranucleotide repeats associated with myotonic dystrophy type 2[J]. FEBS Lett, 2024, 598(20): 2544-2556.
[45] Loureiro J R, Castro A F, Figueiredo A S, et al. Molecular mechanisms in pentanucleotide repeat diseases[J]. Cells, 2022, 11(2): 205.
[46] Taboada X, Rey M, Bouza C, et al. Cytogenomic analysis of several repetitive DNA elements in turbot (*Scophthalmus maximus*)[J]. Gene, 2018, 644: 4-12.