

# 鲤鱼 (*Cyprinus carpio*) 全基因组微卫星分布特征研究

梁霞, 王慧琪, 马宇璇, 宋磊, 吴超, 李亮徽, 张国松

(菏泽学院生理生化及应用实验室, 山东 菏泽 274000)

**[摘要]** 本研究利用微卫星搜索软件 MISA, 对鲤鱼全基因组中的完整型微卫星进行了搜索, 并对其分布规律进行了分析, 同时对微卫星定位在编码区域的基因进行了 GO 注释、富集和 KEGG 富集分析。结果显示, 在约 1.7Gb 鲤鱼全基因组中, 共搜索出 837 004 个完整型微卫星, 相对丰度为 488 个/Mb。微卫星总长度为 15 513 551 bp, 占鲤鱼全基因组大小的 0.91%, 相对密度为 9 051 bp/Mb。在 6 种完整型微卫星中, 单碱基微卫星数目最多, 共有 495 421 个, 在 6 种微卫星中占比 59.19%; 其他依次是二碱基 (26.79%)、三碱基 (8.48%)、四碱基 (4.21%)、五碱基 (1.25%) 和六碱基 (0.07%) 类型微卫星。其中, A、AC、AT、AAT、AG、C、AAAT、AAC、AGAT、AAG 依次为鲤鱼全基因组中出现次数最多的前十种微卫星类别, 表现出明显的 A/T 碱基优势。通过对全基因组微卫星进行定位, 共 24 222 个微卫星定位在基因外显子上, 并分布在 3 853 个基因上。GO 功能注释的 1 758 个 GO 条目中, 注释到生物学过程的条目数和基因数均最多, GO 功能富集最为显著的条目是代谢过程的调节。KEGG 富集表明, 机体系统分支富集到的基因总数最多。环境信息处理分支中的鞘脂信号通路基因富集最为显著。分布到代谢分支的通路种类最为广泛。分析结果推测, 定位在基因编码区域的微卫星可能在细胞间信息交流、信号转导、细胞新陈代谢以及生物合成调控等方面发挥作用。本研究为鲤鱼后续的种群遗传信息评估、微卫星引物的开发及功能鉴定等提供了数据支持。

**[关键词]** 鲤鱼, 全基因组, 微卫星, 分布特征

**[中图分类号]** Q959.46+8 **[文献标志码]** A **[文章编号]** 1001-4616(2021)03-0103-09

## Distribution Characteristics of Microsatellites in the Whole Genome of *Cyprinus carpio*, Linnaeus

Liang Xia, Wang Huiqi, Ma Yuxuan, Song Lei, Wu Chao, Li Lianghui, Zhang Guosong

(Physiological, Biochemical and Applied Laboratory, Heze University, Heze 274000, China)

**Abstract:** In this study, the microsatellite search software MISA was used to search for the complete microsatellite in the whole genome of *C. carpio*, and its distribution pattern was analyzed. Meanwhile, GO annotation enrichment and KEGG enrichment analysis were conducted for the genes located in the coding region of the microsatellite. The results were as follows, 837,004 complete microsatellites were searched in the whole genome of *C. carpio* with a relative abundance of 488/Mb and a total length of 15 513 551 bp, accounting for 0.91% of the whole genome size of *C. carpio*, with a relative density of 9 051 bp/Mb. Among the 6 complete microsatellites, the number of Mononucleotide was the largest (59.19%), followed by the Dinucleotide (26.79%), Trinucleotide (8.48%), Tetranucleotide (4.21%), Pentanucleotide (1.25%) and Hexanucleotide (0.07%). Among them, A, AC, AT, AAT, AG, C, AAAT, AAC, AGAT, AAG were the top ten microsatellite categories that appeared most frequently in the whole genome of *C. carpio*, showing obvious A/T base advantage. Through genome-wide microsatellite localization, a total of 24 222 microsatellites were located on gene exons and distributed on 3 853 genes. Among the 1 758 GO entries annotated by GO function, the number of entries and genes annotated by biological process were the most, and the most significant entry enriched by GO function was the regulation of metabolic process. KEGG enrichment showed that the total number of genes enriched in the branches of the organism system was the largest, the accumulation of sphingolipid signaling pathway genes was the most significant in the branches of environmental information processing, and the types of pathways distributed in the metabolic branches were the most extensive. The results suggest that microsatellites located in the gene coding region may play a role in intercellular information communication, signal trans-

收稿日期: 2020-09-07.

基金项目: 山东省自然科学基金项目 (ZR2019QC002)、山东省高等学校青创科技支持计划和菏泽学院博士基金 (XY18BS08)。

通讯作者: 张国松, 博士, 副教授, 研究方向: 鱼类遗传育种。E-mail: hzyzgs@163.com

duction, cell metabolism and biosynthesis regulation. This study provides data support for the subsequent population genetic information assessment, development of microsatellite primers and functional identification of *C. carpio*.

**Key words:** *Cyprinus carpio*, genome, microsatellite, characteristics

微卫星 DNA, 又称为简单重复序列 (simple sequence repeats, SSRs), 是广泛存在于真核、原核以及病毒基因组中<sup>[1-2]</sup>的 1~6 个碱基串联重复, 随机分布于基因间区、基因的内含子区和编码区等区域。微卫星由高突变性的核心序列和较为保守的侧翼序列两部分组成, 具有杂合率高、分布均匀和共显性遗传等特点, 研究人员通常在微卫星侧翼序列设计引物对微卫星序列进行 PCR 扩增, 以探究物种的遗传多样性和筛选功能标记等。作为优良的第二代分子标记技术, 微卫星已在遗传图谱构建<sup>[3]</sup>、种群遗传多样性评估<sup>[4-5]</sup>以及分子标记辅助育种<sup>[6-7]</sup>等领域中广泛应用。

鲤鱼 (*Cyprinus carpio*) 属于鲤形目、鲤科、鲤亚科, 广泛分布于除澳洲和南美洲外的全世界淡水水域, 也是我国淡水鱼类中品种最多、分布最广、养殖历史最悠久、产量最高者之一。2014 年, 中国水产科学院联合多个单位完成了鲤鱼全基因组注释和图谱绘制, 成为了国际上首个完成全面解析的异缘四倍体硬骨鱼类基因组图谱。研究表明, 其基因组大小约为 1.69G, 含 52 610 个功能基因, 约为已完成基因组测序的硬骨鱼类基因数目的两倍。重复元件分析表明鲤鱼基因组包含了超过 31% 的转座元件, 显著超过了目前完成基因组测序的模式鱼类 (仅少于斑马鱼), 揭示了鲤鱼较高的基因组复杂度。细胞遗传学证据表明, 鲤鱼在减数分裂过程中其 100 条染色体形成 50 个二价体而非 25 个四价体, 进一步证实了鲤鱼基因组的异缘四倍体起源特征。为深入研究鲤鱼基因组四倍化特征, 科研人员开展了鲤鱼 ( $2n=100$ ) 和近源模式鱼类斑马鱼 ( $2n=50$ ) 的比较基因组研究。结果表明, 二者染色体呈现典型的 2:1 同源共线关系, 与斑马鱼相比, 鲤鱼基因组获得了加倍。该研究首次在全基因组水平证实了鲤鱼基因组的四倍化特征和其独特的全基因组复制事件。鲤鱼全基因组测序和图谱的完成, 标志着鲤科鱼类重要经济性状的遗传解析和遗传选育研究正式进入了基因组时代<sup>[8]</sup>。

目前国内外关于鲤鱼微卫星方面的研究已有较多报道: 如孙效文等<sup>[9]</sup>利用微卫星等多种分子标记对柏氏鲤与黑龙江鲤的杂交子二代的单倍体样品做基因型分析, 构建了初步的鲤鱼遗传连锁图谱; 赵兰<sup>[10]</sup>利用鲤鱼微卫星标记和 SNP 标记构建了鲤鱼较高密度的遗传连锁图谱以及遗传-物理整合图谱; 郑先虎等<sup>[11]</sup>利用 4 个鲤鱼群体构建了鲤鱼的整合图谱, 并根据图谱比较分析了不同群体生长性状 QTL 的分布及变异规律; 吴明林等<sup>[12]</sup>利用 10 对微卫星标记对长江野鲤和两种养殖鲤群体进行了遗传多样性评估。

传统开发 SSR 标记的方法主要是使用重复探针筛选基因组文库和小片段阳性克隆测序等来开发 SSR 引物<sup>[13]</sup>。目前关于鲤鱼微卫星标记的开发方法主要集中在磁珠富集法、探针筛库法及 ISSR 片段扩增法等<sup>[14-16]</sup>, 成功开发的微卫星标记主要应用于鲤鱼不同地理群体的遗传特征评估、分子辅助育种研究及种质资源鉴定等。然而, 这些传统筛选方法操作困难、耗时低效。近年来, 随着二代测序技术的迅速发展, 许多物种的基因组相继公布并得到了极大地开发和利用, 这为在全基因组水平上搜索微卫星奠定了基础。目前越来越多生物<sup>[1, 17-19]</sup>的全基因组微卫星研究已被报道。MISA 工具 (MicroSatellite identification tool) 是近些年用于筛选物种全基因组微卫星的一种较为普遍、高效、快捷的微卫星搜索软件, 适用于 EST 序列、文库核酸序列及基因组测序序列等多种类型材料<sup>[20]</sup>; 且具有运行指令简单、运行耗时短及不需联网等优点, 已成为更多物种基因组微卫星搜索的首选工具<sup>[21-23]</sup>。本研究利用 MISA 微卫星筛选软件对鲤鱼基因组数据库 (NCBI 已公布) 检索, 针对鲤鱼全基因组中的微卫星进行搜索并分析其分布规律。同时对外显子区含有微卫星的基因进行 GO 注释、富集和 KEGG 富集分析, 来进一步研究微卫星在黄颡鱼基因序列中的分布特征以及功能定位, 这对于今后分析鲤鱼全基因组特征、评估不同地理群体遗传特征以及开发鲤鱼大量高质量微卫星标记等具有重要的参考价值。

## 1 数据与方法

### 1.1 数据来源

本文中使用的 1.7Gb 鲤鱼全基因组下载自 NCBI Genome 数据库 (<https://www.ncbi.nlm.nih.gov/>), 登录号为 GCF\_000951615.1。

## 1.2 SSR 搜索方法

本研究利用 MISA 微卫星搜索软件,从鲤鱼全基因组中筛选 1~6 种完整型微卫星序列. 搜索标准参考 MISA 软件默认参数,即 1~6 种微卫星核心序列最少重复次数分别为 10、6、5、5、5 和 5. 本研究统计原则将所有可循环的序列及其碱基互补序列归为同一类别,如 ACT 重复拷贝类别,可以与之合并为一类的重复拷贝类别是 CTA、TAC、TGA、GAT 和 ATG.

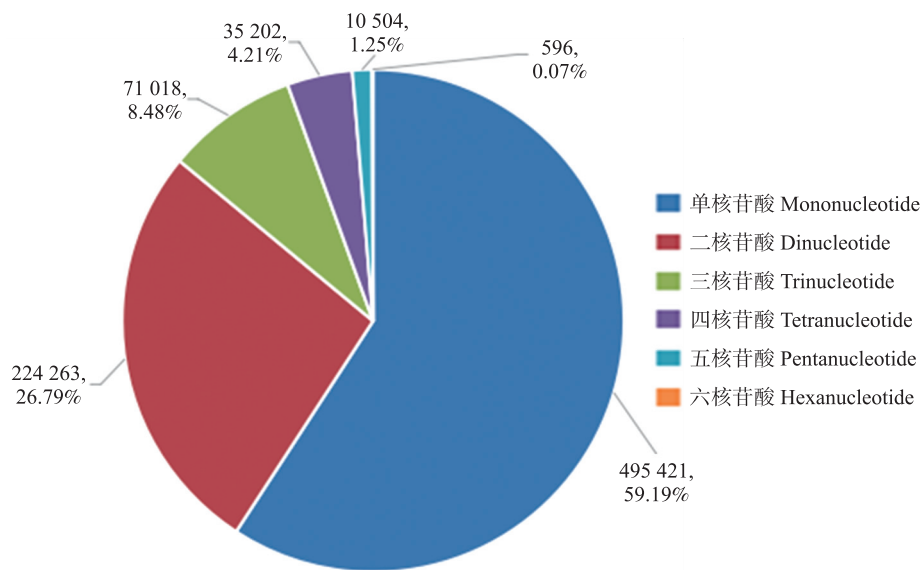
## 1.3 微卫星所在外显子上的基因 GO 注释和 KEGG 富集分析

利用鲤鱼基因组 GFF 注释文件和运行 MISA 软件得到的微卫星位置信息(misa 文件),对全基因组中所有微卫星进行定位. 结合本实验室编写的 Python 脚本及使用 Blast2 GO 软件<sup>[24]</sup>进行 GO 注释,利用 KOBAS 2.0 在线软件进行 GO 富集和 KEGG 富集分析<sup>[25]</sup>.

# 2 结果与讨论

## 2.1 鲤鱼全基因组微卫星总体分布规律

利用 MISA 微卫星筛选软件在 1.7Gb 鲤鱼全基因组中共搜索到 837 004 个完整型微卫星,相对丰度为 488 个/Mb,微卫星总长度为 15 513 551 bp,占鲤鱼全基因组大小的 0.91%,相对密度为 9 051 bp/Mb. 1~6 碱基重复类型的微卫星数目差异较大,其中单碱基重复数目最多,占微卫星总数的 67.11%;其次分别是二碱基、三碱基、四碱基和五碱基. 其中六碱基类型出现频率最低,仅有 596 个微卫星,在所有微卫星中占比 0.07%(图 1 和表 1).



图中数字代表该碱基类型在全基因组中的数量,百分比代表该碱基类型在所有微卫星中所占比例

图 1 鲤鱼全基因组中 6 种微卫星类型数量分布

Fig. 1 Quantitative distribution of six types of microsatellites in the *Cyprinus carpio* whole genome

表 1 鲤鱼全基因组中 6 种微卫星重复类型统计分析

Table 1 Different types of microsatellite sequences in the *Cyprinus carpio* genome

微卫星类型	总长度/bp	占比/%	平均长度/bp	频率/(个/Mb)	密度/(bp/Mb)
单核苷酸	6 427 459	41.43%	12.97	289.04	3 749.98
二核苷酸	5 446 318	35.11%	24.29	130.84	3 177.55
三核苷酸	1 697 370	10.94%	23.90	41.43	990.30
四核苷酸	1 281 444	8.26%	36.40	20.54	747.63
五核苷酸	655 890	4.23%	62.44	6.13	382.67
六核苷酸	5 070	0.03%	8.51	0.35	2.96

## 2.2 鲤鱼全基因组中 6 种微卫星类型不同类别分布特征

除了 6 种碱基类型微卫星在基因组中分布频率差异显著外,每种微卫星类型中不同碱基类别的分布规律也有很大差别(表 2 和表 3).

表 2 鲤鱼微卫星中出现频率最高的 10 种重复拷贝类别

Table 2 The 10 types of repeated copies with the highest frequency of microsatellites in the *Cyprinus carpio*

SSR 类型	SSR 数目	占各微卫星类型比例	SSR 类型	SSR 数目	占各微卫星类型比例
A	478 415	96.57%	C	17 006	3.43%
AC	122 398	54.58%	AAAT	15 608	44.34%
AT	62 844	28.02%	AAC	8 627	12.15%
AAT	48 895	68.85%	AGAT	7 831	22.25%
AG	38 646	17.23%	AAG	4 220	5.94%

表 3 不同碱基类型微卫星前 3 种优势类别在鲤鱼基因组中的分布

Table 3 Distribution of the first three dominant categories of microsatellites of different base types in the *Cyprinus carpio* genome

SSR 类型	类别	数目	占各微卫星类型比例	SSR 类型	类别	数目	占各微卫星类型比例
单碱基	A	478 415	96.57%	四碱基	AAAT	15 608	44.34%
	C	17 006	3.43%		ATCT	7 831	22.25%
二碱基	AC	122 398	54.58%	五碱基	AAAC	3 505	9.96%
	AT	62 844	28.02%		AATAT	3 240	30.85%
	AG	38 646	17.23%		AAAAT	2 086	19.86%
三碱基	AAT	48 895	68.85%	六碱基	AAATT	703	6.69%
	AAC	8 627	12.15%		AACCCT	74	12.42%
	AAG	4 220	5.94%		AAAAAT	66	11.07%
					AATCAG	51	8.56%

在单碱基类别中,A 类别(A、T)占有绝对的碱基优势,共搜索到 478 415 个,占单碱基总数目的 96.57%;相比较而言,C 类别(C、G)出现频率较低,在基因组中仅有 17 006 个,在单碱基中占比 3.43%。

在二碱基四种类别中,AC 类别数目最多,共出现 122 398 个,占二碱基总数目的 54.58%;其次分别是 AT 和 AG 类别,在二碱基中分别占比 28.02%和 17.23%;CG 类别数目最少,仅占比 0.17%。

在三碱基、四碱基和五碱基类别中,也均表现出明显的 A、T 碱基优势。在三碱基中,AAT 类别出现频率最高,共出现 48 895 次,占比 68.85%。其次分别是 AAC 和 AAG,在三碱基中分别占比 12.15%和 5.94%。CCG 类别出现频率最低,仅有 112 个,占比 0.16%。在四碱基中,AAAT 为出现频率最高的类别,占比 44.34%,其次分别为 ATCT 和 AAAC。在五碱基中,AATAT、AAAAT 和 AAATT 为数目最多的前三种类别。

由于六碱基微卫星在基因组中出现频率较低,绝大部分六碱基类别出现次数较少,部分类别甚至缺失。AACCCT 为六碱基中数目最多的类别,共有 74 个,占比 12.42%。其他类别占比均较低。

2.3 鲤鱼全基因组微卫星重复拷贝数分布

尽管微卫星在鲤鱼 6 种碱基类型中的数目分布和各碱基不同类别数目差异较大,但每种微卫星的核心序列拷贝数变化趋势是一致的。即 1~6 种碱基均随着核心拷贝数的增加,其微卫星数目逐渐递减(图 2)。

鲤鱼基因组中每种微卫星核心序列的拷贝数变化范围较大,从 5~4 029 拷贝数不等。但 6 种微卫星核心序列拷贝数均集中在 5~40 次,占微卫星总数的 99.44%。单碱基核心拷贝数主要集中在 10~33 次,占单碱基微卫星总数的 99.84%;二碱基核心拷贝数在 6 种微卫星中最为分散,主要集中在 6~40 次,在二碱基

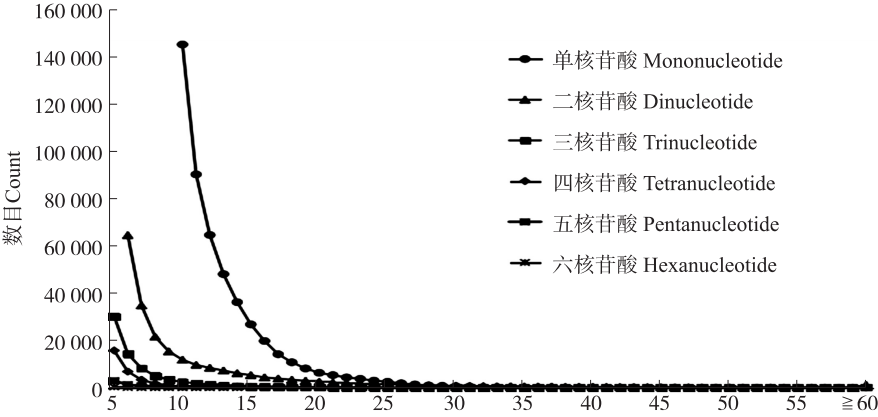


图 2 鲤鱼各碱基类型微卫星不同拷贝数分布

Fig. 2 Distribution of different microsatellite repetitions in *Cyprinus carpio*

中占比 98.22%. 其中最高拷贝数为 4 029 次,在 6 种微卫星拷贝范围中拷贝数最大;三碱基、四碱基和五碱基核心拷贝数范围较窄,分别集中在 5~21 次、5~25 次和 5~20 次,分别占比 99.37%、97.96% 和 95.70%. 六碱基核心拷贝数更为集中,主要集中在 5~9 范围内,占比 88.59%.

2.4 GO 功能注释、富集和 KEGG 富集特征

通过对鲤鱼全基因组中搜索的微卫星进行定位,并对微卫星定位在编码区的基因进行功能注释. 结果表明,包含 60 139 个基因注释的鲤鱼基因组中,17 678 个基因的外显子与微卫星有重叠. 基因的 GO 功能注释主要分为生物学过程、细胞组分和分子功能三大类. 通过对外显子上包含微卫星的 17 678 个编码基因进行 GO 功能注释,共得 1 758 个 GO 条目. 注释到生物学过程的条目数和编码基因数均最多,其中注释条目数共 1 120 个,占比 63.7%,涉及的生物学过程主要包括细胞过程 (cellular process)、单组织过程 (single-organism process) 及代谢过程 (metabolic process) 等;注释到细胞组分的条目有 416 个,占比 23.7%,主要包括膜 (membrane)、细胞 (cell) 和细胞成分 (cell part) 等;222 个条目 (占比 12.6%) 注释到分子功能 (molecular function) 方面,主要集中在连接 (binding)、催化活性 (catalytic activity) 以及信号传感器活动 (signal transducer activity) 等方面,其中注释到结合活性 (binding) 功能的基因在所有注释功能中最多 (图 3). GO 功能富集最为显著的条目是代谢过程的调节 ( $P = 8.47\text{E}-15$ )、大分子代谢过程的调控 ( $P = 1.24\text{E}-14$ ) 以及细胞生物合成过程的调控 ( $P = 5.55\text{E}-14$ ) (表 4).

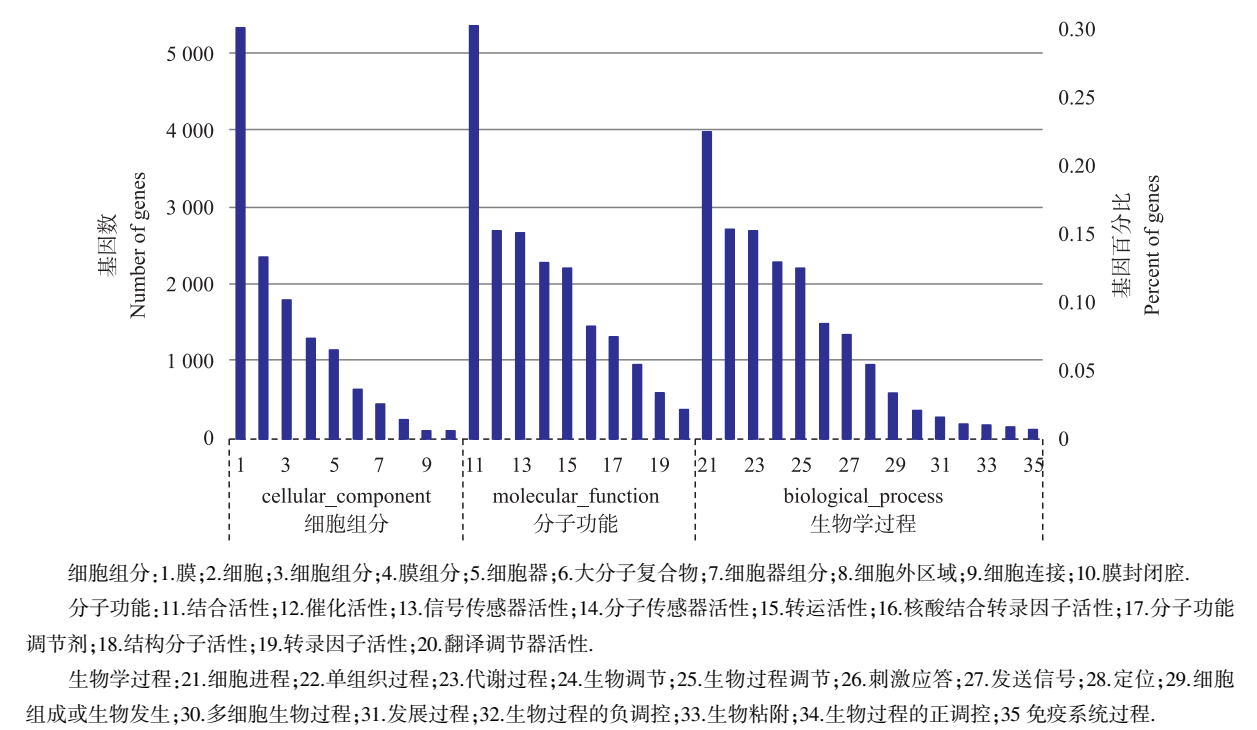


图 3 鲤鱼微卫星分布于外显子的基因 GO 功能注释  
Fig. 3 The GO function annotation of exon microsatellites in *Cyprinus carpio*

表 4 鲤鱼微卫星分布于外显子的基因 GO 富集

条目	GO ID	输入数	背景数	P 值
代谢过程调节	GO:0019222	787	2 234	8.47E-15
大分子代谢过程的调控	GO:0060255	770	2 183	1.24E-14
细胞生物合成过程的调控	GO:0031326	697	1 963	5.55E-14
生物合成过程的调控	GO:0009889	697	1 964	6.24E-14
高分子生物合成过程的调控	GO:0010556	690	1 946	1.03E-14
基因表达调控	GO:0010468	715	2 026	1.07E-13
初级代谢过程的调节	GO:0080090	745	2 122	1.07E-13
氮化合物代谢过程的调控	GO:0051171	743	2 116	1.12E-13
细胞代谢过程的调节	GO:0031323	751	2 145	1.64E-13
细胞大分子生物合成过程的调控	GO:2000112	687	1 941	1.76E-13



对微卫星分布于外显子上的基因进行 KEGG 富集分析,共富集到细胞过程( cellular processes)、环境信息处理( environmental information processing)、遗传信息处理( genetic information processing)、代谢( metabolism)和有机系统( organismal systems)五大分支中,包括 280 条通路. 其中机体系统分支富集到的基因总数最多,共 6 520 个. 环境信息处理分支中的鞘脂信号通路( sphingolipid signaling pathway)基因富集最为显著( $P=2.36E-10$ ),PI3K-Akt 信号通路( PI3K-Akt signaling pathway)上富集的基因最多,共有 319 个. 分布到代谢分支的通路种类最为广泛,包含 113 条通路,共富集到 1 609 个相关基因. 其次是有机系统( 81 条通路),分布到遗传信息处理的通路种类和基因总数均最低( 22 条通路,881 个基因)( 表 5).

表 5 鲤鱼微卫星分布于外显子的基因 KEGG 富集  
Table 5 The KEGG enrichment of exon microsatellites in *Cyprinus carpio*

功能分类	通路	输入数	背景数	P 值
机体系统	胰岛素信号通路	162	410	4.51E-07
	长寿调节途径-多种物种	92	210	6.56E-07
	血小板活化	158	402	8.94E-07
	神经营养素信号通路	146	370	1.68E-06
	趋化因子信号通路	179	470	2.09E-06
代谢	醚脂质代谢	42	104	3.16E-03
	角质、角质和蜡的生物合成	3	4	6.53E-03
	赖氨酸退化	57	154	8.14E-03
	糖胺聚糖降解	15	37	3.83E-02
遗传信息处理	RNA 降解	74	197	2.13E-03
	信使核糖核酸监测途径	80	232	1.82E-02
环境信息处理	鞘脂信号通路	168	391	2.36E-10
	cGMP-PKG 信号通路	202	533	7.58E-07
	FoxO 信号通路	157	395	4.32E-07
	Rap1 信号通路	260	713	1.03E-06
	MAPK 信号通路-果蝇	89	219	3.70E-05
	VEGF 信号通路	76	184	6.36E-05
	MAPK 信号通路	276	801	7.03E-05
细胞过程	调控干细胞多能性的信号通路	136	374	3.33E-04
	局部粘连	223	656	7.18E-04
	黏着连接	107	294	1.19E-03
	肌动蛋白细胞骨架的调节	222	660	1.36E-03

2.5 鲤鱼全基因组微卫星总体特征分析

本研究以鲤鱼全基因组为基础,利用生物信息学方法对全基因组中的完整型微卫星进行了搜索、统计并分析. 在 1.7Gb 全基因组中共搜索出 837 004 个完整型微卫星,占鲤鱼全基因组大小的 0.91%,其比例低于已公布的人 *Homo sapiens* (3%)<sup>[18]</sup>、小鼠 *Mus musculus* (2.85%)<sup>[26]</sup> 和大鼠 *Rattus norvegicus* (1.41%)<sup>[27]</sup>,但与红鳍东方鲀 *Takifugu rubripes* (0.73%)、双斑东方鲀 *Takifugu bimaculatus* (0.84%)、菊黄东方鲀 *Takifugu flavidus* (0.84%)<sup>[28]</sup> 及红原鸡 *Gallus gallus* (0.49%)<sup>[29]</sup> 相比所占比例较高. 可见在鲤鱼全基因组中微卫星的含量较为丰富. 究其物种间含量差异原因,可能是由于不同物种间基因组大小和基因组内碱基组成及排列方式的差异导致的.

在鲤鱼 6 种完整型微卫星中,单碱基微卫星分布数目最多,占完整型微卫星的 59.19%,总长度也最大,在基因组中占比 41.43%. 这一现象与已报道的大熊猫 *Ailuropoda melanoleuca*、北极熊 *Ursus maritimus*<sup>[30]</sup>、食蟹猴 *Macaca fascicularis*<sup>[31]</sup> 等哺乳动物和红鳍东方鲀、双斑东方鲀、菊黄东方鲀、黑青斑河鲀 *Tetraodon nigroviridis*<sup>[28]</sup> 等水生动物的优势碱基类型相同. 而在中国对虾 *Fenneropenaeus chinensis*<sup>[32]</sup>、三疣梭子蟹 *Portunus trituberculatus*<sup>[33]</sup> 和中华鲟 *Rhodeus sinensis*<sup>[34]</sup> 等水生动物中均为二碱基占优势. 由此可知,不同物种的优势微卫星类型也各不相同. 因而,不同物种的微卫星序列既存在基因组间的进化现象,也可能存在着一定的保守现象.

2.6 鲤鱼 6 种微卫星中多种类别特征分析

在鲤鱼全基因组 1~6 种微卫星类型中,均表现出明显的 A、T 碱基优势. 在单碱基类别中,A 类别 (A/T) 微卫星数目占比 96.57%,跟 C 类别 (3.43%) 相比具有明显的优势. 这种现象同样存在于人、果蝇 *Drosophila melanogaster*<sup>[35]</sup>、食蟹猴<sup>[31]</sup>、大熊猫和北极熊<sup>[30]</sup> 等绝大部分物种中.

在二碱基类别中,AC 类别占有较大的比例,其次是 AT 和 AG. 而在黄颡鱼<sup>[36]</sup>、金钱鱼<sup>[21]</sup>、斑鳢<sup>[23]</sup>和 4 种河鲀<sup>[28]</sup>等水生动物中二碱基前三类别排序均为 AC、AG 和 AT. 尽管 AG 和 AT 在不同物种二碱基中排序不同,但均有 AC 类别占比较大,AG、AT 类别占比相差较小的特征.

AAT、AAC 和 AAG 分别为三碱基中的前三类别,这表明 AAN(N 代表除 A 以外的任何碱基)类别在三碱基中具有绝对的碱基优势. 该现象与人<sup>[18]</sup>等三碱基分布相一致. 研究表明,在 6 种完整型微卫星中,三碱基微卫星不同于其他碱基类型,其与生物体的遗传疾病具有一定的相关性<sup>[37]</sup>. 而 Toth<sup>[35]</sup>报道在外显子中,三核苷酸重复在 6 种碱基类型中总是最丰富的. 因此猜测,伴随着微卫星核心序列拷贝数的随机波动,有可能会改变鲤鱼功能基因的结构,从而会影响其相关性状的表达,引起鲤鱼部分遗传疾病的发生. 因此对鲤鱼三碱基微卫星进行定位和研究分析,将有利于对其相关遗传疾病做出早期的预防与判断.

在四碱基、五碱基和六碱基类别中,AAAN、AAAAN 和 AAAAAAN(N 代表除 A 以外的任何碱基)的数量较为丰富,这与灵长类和啮齿类等物种中微卫星的研究相一致<sup>[35]</sup>. AACCCCT 类别在六碱基中数目最多,在黄颡鱼<sup>[36]</sup>、红鳍东方鲀等 4 种河鲀<sup>[28]</sup>中同样很丰富,但在其他物种中分布较少. 推测该类别可能在水生生物遗传进化及环境适应方面发挥一定的作用.

## 2.7 鲤鱼 6 种微卫星核心拷贝数分布特征分析

鲤鱼全基因组微卫星分析表明,6 种完整型微卫星均随着核心序列拷贝数的增加其微卫星的数目逐渐递减. 这种现象与人<sup>[18]</sup>、虎皮鹦鹉 *Melopsittacus undulatus*<sup>[38]</sup>、虾夷扇贝 *Patinopecten yessoensis*<sup>[39]</sup>、金钱鱼<sup>[21]</sup>以及 4 种埃博拉病毒<sup>[2]</sup>等绝大部分物种全基因组中微卫星的拷贝变化规律相一致. 有关研究表明,当 6 种微卫星核心序列拷贝数达到产生滑动突变的最低阈值时,DNA 复制就会出现滑动突变<sup>[40]</sup>. 此外,对于各类别微卫星出现频率随核心序列拷贝数变化的规律性,Wierdl 等<sup>[41]</sup>认为与微卫星的稳定性和突变率有关. 即随着微卫星长度的增加,其稳定性会越低,同时突变率会越高,这使得越长的微卫星序列其数目越少.

## 2.8 鲤鱼微卫星相关基因 GO 注释、富集和 KEGG 富集分析

微卫星在基因中的分布对于基因功能的表达非常重要. 在生物体内,不同基因相互协调行使其生物学功能,通过 GO 注释、富集能够分析目的基因集在基因功能上的集中体现,而 KEGG 通路显著性富集能够确定目标基因集参与的最主要生化代谢途径和信号转导途径. 本研究通过对外显子上包含微卫星的基因进行 GO 功能注释、富集和 KEGG 富集分析,来进一步探究微卫星在基因功能方面发挥的作用. GO 注释分析表明,注释到细胞组分中的相关基因主要定位于膜、膜组分和细胞组分等;分子功能主要集中在结合活性、催化活性和信号传感器活性等方面;参与的生物学过程主要与细胞过程、单组织过程和代谢过程等相关. GO 富集前十条目均与代谢调节、生物合成调控以及基因表达调控三个功能方面显著相关,其中以代谢过程调节、大分子代谢过程调控和细胞生物合成过程调控富集最为显著. KEGG 共富集到 280 条通路中,其中环境信息处理类别中的鞘脂信号通路在所有通路中富集最为显著,这表明微卫星富集的较多基因与信号转导紧密相关. 综合 KEGG 5 个分支的通路富集情况可知,富集的目的基因主要参与了信号转导通路、代谢通路以及生长相关通路等通路的调节过程.

基于 GO 注释、富集和 KEGG 富集分析结果推测,定位在基因编码区域的微卫星功能可能集中表现在参与细胞间信息交流、信号转导、细胞新陈代谢以及生物合成调控几个方面. 通过与黄颡鱼<sup>[36]</sup>、绿尾虹雉<sup>[42]</sup>2 个物种基因组中分布在外显子上的微卫星 GO 注释、KEGG 富集比较发现,3 个物种中 GO 注释的基因均与膜组分、信号转导、结合活性以及代谢过程功能紧密相关,KEGG 富集的通路也都集中在信号转导通路和代谢通路方面. 因此猜测不同物种中定位在基因编码区域上的微卫星可能在参与信号转导、细胞代谢等过程方面发挥一定的功能.

## 3 结论

尽管关于鲤鱼微卫星开发及应用的报道已有很多,但可有效利用的微卫星标记仍相对较少,目前还不能满足鲤鱼以及鲤科鱼类遗传育种相关的大量研究. 本研究利用生物信息学软件,在鲤鱼全基因组内对微卫星进行搜索分析,并对处于基因编码区域的微卫星进行了功能探究,这为鲤鱼后续的基因组分析、种群遗传信息评估、品系亲缘关系鉴定以及种群良种选育等提供了数据支持.

## [参考文献]

- [1] TAUTZ D, RENZ M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes[J]. Nucleic acids research, 1984, 12(10): 4127-4138.
- [2] 周思倩, 焦伟丽, 彭珠黎, 等. 埃博拉病毒基因组中微卫星序列的分布分析[J]. 基因组学与应用生物学, 2019, 38(3): 1087-1095.
- [3] SHEN X Y, YANG G P, LIU Y J, et al. Construction of genetic linkage maps of guppy (*Poecilia reticulata*) based on AFLP and microsatellite DNA markers[J]. Aquaculture, 2007, 271(1-4): 178-187.
- [4] NARASIMHAMOORTHY B, SAHA M C, SWALLER T, et al. Genetic diversity in switchgrass collections assessed by EST-SSR markers[J]. Bioenergy research, 2008, 1(2): 136-146.
- [5] HULAK M, KASPAR V, KOHLMANN K, et al. Microsatellite-based genetic diversity and differentiation of foreign common carp (*Cyprinus carpio*) strains farmed in the Czech Republic[J]. Aquaculture, 2010, 298(3/4): 194-201.
- [6] SERBEZOV D, BERNATCHEZ L, OLSEN E M, et al. Mating patterns and determinants of individual reproductive success in brown trout (*Salmo trutta*) revealed by parentage analysis of an entire stream living population[J]. Molecular ecology, 2010, 19(15): 3193-3205.
- [7] XIA J H, LIU F, ZHU Z Y, et al. A consensus linkage map of the grass carp (*Ctenopharyngodon idella*) based on microsatellites and SNPs[J]. BMC genomics, 2010, 11(1): 135-150.
- [8] XU P Z, XIAO F, WANG X M, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*[J]. Nature genetics, 2014, 46(11): 1212-1219.
- [9] 孙效文, 梁利群. 鲤鱼的遗传连锁图谱[J]. 中国水产科学, 2000, 7(1): 1-5.
- [10] 赵兰. 鲤遗传\_物理整合图谱的构建及鲤与斑马鱼的比较作图[D]. 大连: 大连海洋大学, 2013.
- [11] 郑先虎, 匡友谊, 吕伟华, 等. 基于整合图谱的鲤生长相关性状 QTL 的分布及变异规律[J]. 中国科学: 生命科学, 2013, 43(2): 159-167.
- [12] 吴明林, 侯冠军, 李海洋, 等. 长江野鲤 (*Cyprinus carpio*) 及两种养殖鲤群体遗传多样性评估[J]. 基因组学与应用生物学, 2020, 39(1): 70-78.
- [13] XIONG L W, WANG Q, QIU G F. Large-scale isolation of microsatellites from Chinese mitten crab *Eriocheir sinensis* via a Solexa Genomic Survey[J]. International journal of molecular sciences, 2012, 13(12): 16333-16345.
- [14] 魏东旺, 楼允东, 孙效文, 等. 鲤鱼微卫星分子标记的筛选[J]. 动物学研究, 2001, 22(3): 238-241.
- [15] 全迎春, 孙效文, 梁利群. 应用微卫星多态分析四个鲤鱼群体的遗传多样性[J]. 动物学研究, 2005, 26(6): 595-602.
- [16] 孙效文, 贾智英, 魏东旺, 等. 磁珠富集法与小片段克隆法筛选鲤微卫星的比较研究[J]. 中国水产科学, 2005, 12(2): 126-132.
- [17] Van BELKUM A, SCHERER S, Van ALPHEN L, et al. Short-sequence DNA repeats in prokaryotic genomes[J]. Microbiology and molecular biology reviews, 1998, 62(2): 275-293.
- [18] 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): 1-10.
- [19] KARAOGLU H, LEE C M Y, MEYER W. Survey of simple sequence repeats in completed fungal genomes[J]. Molecular biology and evolution, 2004, 22(3): 639-649.
- [20] 王希, 陈丽, 赵春雷. 利用 MISA 工具对不同序列进行 SSR 标记位点挖掘的探讨[J]. 中国农学通报, 2016, 32(10): 150-156.
- [21] 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.
- [22] 郭新颖, 张健, 李梦柔, 等. 藏鸡基因组微卫星特征分析[J]. 中国家禽, 2020, 42(1): 116-120.
- [23] 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[J]. 南方水产科学, 2020, 16(3): 47-60.
- [24] CONESA A, GOTZ S, GARCIA G J M, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research[J]. Bioinformatics, 2005, 21(18): 3674-3676.
- [25] XIE C, MAO X Z, HUANG J J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases[J]. Nucleic acids research, 2011, 39(suppl): W316-W322.
- [26] 童晓玲, 代方银, 李斌, 等. 小鼠基因组中的微卫星重复序列的数量、分布和密度[J]. 动物学报, 2006, 52(1): 138-152.



- [27] 涂飞云,刘晓华,杜联明,等. 大鼠全基因组微卫星分布特征研究[J]. 江西农业大学学报,2015,37(4):708-711.
- [28] 徐杰杰,郑翔,张鑫宇,等. 2019,4种河鲀全基因组微卫星分布特征分析研究[J/OL]. 基因组学与应用生物学,2019. <http://kns.cnki.net/kcms/detail/45.1369.q.20191126.1019.002.html>.
- [29] 黄杰,杜联明,李玉芝,等. 红原鸡全基因组中微卫星分布规律研究[J]. 四川动物,2012,31(3):358-363.
- [30] 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.
- [31] 涂飞云,刘俊,韩卫杰,等. 食蟹猴全基因组微卫星分布特征分析[J]. 野生动物学报,2018,39(2):400-404.
- [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,2004(5):424-431.
- [33] 宋来鹏,刘萍,李健,等. 三疣梭子蟹基因组微卫星特征分析[J]. 中国水产科学,2008,15(5):738-744.
- [34] 熊良伟,王帅兵,封琦,等. 基于高通量测序的中华鲟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.
- [35] TOTH G. Microsatellites in different eukaryotic genomes:survey and analysis[J]. Genome research,2000,10(7):967-981.
- [36] 徐杰杰,郑翔,李杰,等. 黄颡鱼(*Pelteobagrus fulvidraco*)全基因组微卫星分布特征研究[J/OL]. 基因组学与应用生物学,2019. <http://kns.cnki.net/kcms/detail/45.1369.Q.20191204.1055.002.html>.
- [37] PEARSON C E,SINDEN R R. Trinucleotide repeat DNA structures:dynamic mutations from dynamic DNA[J]. Current opinion in structural biology,1998,8(3):321-330.
- [38] 黄杰,原宝东,杨承忠. 虎皮鹦鹉全基因组中微卫星分布规律研究[J]. 野生动物学报,2017,38(3):422-426.
- [39] 倪守胜,杨钰,柳淑芳,等. 基于高通量测序的虾夷扇贝基因组微卫星特征分析[J]. 渔业科学进展,2018,39(1):107-113.
- [40] LAI Y,SUN F. The relationship between microsatellite slippage mutation rate and the number of repeat units[J]. Molecular biology and evolution,2003,20(12):2123-2131.
- [41] WIERDL M,DOMINSKA M,PETES T D. Microsatellite instability in yeast:dependence on the length of the microsatellite[J]. Genetics,1997,146(3):769-779.
- [42] CUI K,YUE B S. Distribution patterns of microsatellites in the genome of *lophophorus lhuysii*[J]. Sichuan journal of zoology, 2018,37(5):59-66.

[责任编辑:黄 敏]