

凤头潜鸭染色体水平基因组微卫星分布特征及其功能分析^{*}

黄 杰¹, 钟华明¹, 乔佳怡¹, 杨承忠², 杨 波³

(1. 商丘师范学院 生物与食品学院, 河南 商丘 476000; 2. 重庆师范大学 生命科学学院 动物生物学重庆市高校重点实验室, 淡水鱼类资源保护与利用重庆市重点实验室, 重庆 401331; 3. 中国大熊猫保护研究中心, 成都 610057)

摘要:为了系统揭示凤头潜鸭(*Aythya fuligula*)全基因组微卫星的分布特征及微卫星相关基因的潜在功能,本研究基于已公布的该物种染色体水平基因组数据,利用生物信息学方法对全基因组微卫星的分布特征进行了系统分析,并对位于外显子区的微卫星所在基因进行了功能注释。结果表明,在约 1 093 Mbp 的凤头潜鸭基因组序列中,共筛选出 1~6 个碱基重复的微卫星 585 346 个,总长度 13 232 974 bp,占基因组总长度的 1.17%。在 6 种重复类型中,单碱基重复的数量最多,占微卫星总数的 68.45%;其次为四碱基重复和二碱基重复,分别占微卫星总数的 13.97%和 10.48%;五碱基和六碱基重复的占比最低。微卫星碱基组成呈现明显的 A 碱基偏倚:单碱基重复以 A 拷贝类别最多,二碱基重复中 AT 拷贝类别最为丰富,四碱基重复以 AAAX 类型为主。分布数量最多的 10 种微卫星拷贝类型依次为 A、AAAC、AAAT、AT、AAT、C、AC、AAC、AAAAC 和 AAAAT,每种类型的数量均超过 9 200 个,总数为 518 077 个,占微卫星总数的 88.5%。染色体分布分析表明,微卫星数量与染色体长度存在统计学意义上的正相关($r=0.993$, $p<0.001$),Z 染色体在相对丰度和密度上均明显高于常染色体,而 W 染色体的相对丰度和密度则明显偏低。GO 注释结果显示,基因中涉及生物过程的基因数量最多。在 KEGG 通路富集分析中,有 135 条通路被注释到,其中 FoxO 信号通路、Notch 信号通路及鞘脂代谢通路等通路存在统计学意义上的富集($p<0.05$),这些通路的富集可能与凤头潜鸭的迁徙适应及神经导航能力有关。本研究首次系统揭示了凤头潜鸭全基因组微卫星的分布特征及外显子微卫星相关基因的功能富集,可为后续微卫星分子标记开发、种群遗传多样性分析和环境适应机制研究提供参考。

关键词:凤头潜鸭;全基因组;微卫星;分布特征;功能分析

中图分类号:Q953

文献标志码:A

文章编号:1672-6693(2026)04-0056-11

凤头潜鸭(*Aythya fuligula*)隶属雁形目(Anseriformes)鸭科(Anatidae)潜鸭属(*Aythya*),是全球种群数量最庞大、分布最广泛的潜水鸭之一,目前总数约 260 万~290 万只^[1],主要栖息于大型湖泊和河谷地带,大部分种群具有典型的迁徙习性^[2-3]。作为迁徙性水禽,凤头潜鸭在生命周期中面临长距离的飞行、复杂环境的导航、能量代谢的调控及对环境的适应等多种生理和生态挑战。凤头潜鸭的基因组特征,特别是高突变的微卫星(microsatellite)序列,可能蕴藏着适应广域分布和复杂迁徙行为的遗传基础。已有关于凤头潜鸭的研究多集中于种群动态、栖息地选择及城市适应机制的种群生态学和行为学层面^[4-5],而基因组层面的研究尤其是微卫星分布特征及功能解析仍属空白。

微卫星又称简单重复序列,微卫星的特征是以 1~6 个核苷酸为基本单元、经多次串联构成的 DNA 序列,在真核生物、原核生物及病毒基因组中均大量存在^[6-7]。微卫星变异是由复制滑移和 DNA 修复机制产生^[6],由于微卫星具备多态性高、易于检测、在基因组中分布广泛等诸多优点,同时它还为基因组进化提供了重要的遗传变异来源,因此它常在生命科学研究领域被用作分子遗传标记。微卫星标记在群体遗传多样性评估^[8-9]、亲子鉴定^[10]、种间基因流分析^[11]等方面得到了广泛应用。近些年来,基于全基因组分析微卫星分布特征,已成为揭示物种基因组结构、进化历史和环境适应机制的重要手段。特别是位于编码区的微卫星,能通过改变蛋白质的结构或功能来影响基因表达和表型变异,从而为物种快速适应新环境提供遗传基础^[12]。

^{*} 收稿日期:2026-04-09 修回日期:2026-06-01 网络出版时间:2026-07-23T14:10

资助项目:重庆市教育委员会科学技术研究计划重点项目(No. KJZD-202300503);河南省科技攻关项目(No. 222102320442)

第一作者简介:黄杰,女,副教授,博士,研究方向为动物分子生物学,E-mail:huangjie66666@163.com;通信作者简介:杨波,男,高级工程师, yangb_198701@163.com

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20260723.1125.002

基于 2020 年 1 月公布的染色体级别的风头潜鸭基因组数据,本研究运用生物信息学方法系统分析了凤头潜鸭全基因组微卫星的分布特征、组成规律以及染色体分布偏好。同时,针对位于外显子区的微卫星所在基因进行了 GO 功能注释和 KEGG 通路富集分析,进而探讨微卫星在凤头潜鸭基因组中可能发挥的功能。研究结果将为今后开展分子标记开发、种群遗传多样性研究以及环境适应机制解析提供基础数据。

1 材料与方法

本研究使用的凤头潜鸭基因组序列下载于 NCBI Genome 数据库(<https://www.ncbi.nlm.nih.gov>),登录号为 GCF_009819795.1,序列总长约为 1 124 Mb,该基因组序列已组装到染色体水平,将下载的染色体基因组序列以 fasta 格式保存。

利用 Krait v 2.0.6(<https://github.com/lmdu/krait2>)微卫星搜索软件^[13],从凤头潜鸭全基因组中筛选和统计不同重复类型完整型微卫星(perfect microsatellites)序列。本研究使用 Krait 软件默认的微卫星统计参数,即 1~6 碱基重复的完美型微卫星重复拷贝次数分别不少于 12、7、5、4、4 和 4。此外,本研究统计分析时,根据起始碱基顺序的不同及碱基互补配对原则,将所有可循环的序列及与该序列碱基互补的序列归为同一类型,如三碱基重复类型 AGT 与 ATC、CAT、GTA、TAG、TCA 可以合并为同一类。本研究使用的描述基因组微卫星分布特征的相关的术语,如重复类型、重复拷贝数、重复拷贝类别、相对丰度、相对密度等参考关于绿头鸭(*Anas platyrhynchos*)^[14]和白鹭(*Egretta garzetta*)^[15]基因组微卫星的研究。

使用 SPSS 26.0 软件对凤头潜鸭包括常染色体及性染色体 Z、W 在内的 36 条染色体长度与微卫星数量进行 Pearson 相关性分析。使用 R 包 clusterProfiler v. 3.12.0^[16]软件默认参数进行 GO 功能注释分析。借助于 KOBAS 2.0^[17]在线平台,对上述目标基因集进行 KEGG 富集分析。

为评估位于外显子区的微卫星对蛋白质序列的潜在影响,本研究首先使用 bedtools intersect 获得微卫星与编码序列(CDS)的重叠区域。随后,计算每个微卫星在 CDS 中的起始偏移量,用起始偏移量除以 3 所得的余数作为相位(phase)。若微卫星为非三碱基重复且相位不为 0,则预测可能引起移码突变。同时,统计了各基因编码区所含微卫星的总数,并筛选出三碱基重复微卫星。

2 结果

2.1 凤头潜鸭基因组微卫星的整体分布情况

本研究结果显示:凤头潜鸭染色体水平基因组中完整性微卫星总数为 585 346 个,总长度为 13 232 974 bp,平均长度为 22.61 bp,占全基因长度的 1.17%;基因组中微卫星的整体相对丰度为 519.38 个·Mb⁻¹,整体相对密度为 11 741.72 bp·Mb⁻¹。凤头潜鸭的 6 种重复类型微卫星中,数量最多的是单碱基类型,占基因组微卫星总数的 62.78%,四碱基类型数量次之,三碱基、二碱基、五碱基和六碱基类型微卫星的数量占比均不足 10%,且数量占比依次减少;不同类型微卫星的相对丰度的变化与该类型微卫星的数量变化一致,数量越多的微卫星类型,相对丰度也越大(表 1)。

表 1 凤头潜鸭基因组中不同重复类型微卫星分布特征
Tab.1 Distribution characteristics of microsatellites with different repeat types in genome of the *A. fuligula*

重复类型	数量/个	长度/bp	在微卫星中的 数量占比	平均长度/bp	相对丰度/ (个·Mb ⁻¹)	相对密度/ (bp·Mb ⁻¹)
单碱基	367 470	6 638 186	62.78%	18.06	326.06	5 890.11
二碱基	44 730	834 896	7.64%	18.67	39.69	740.81
三碱基	45 811	985 710	7.83%	21.52	40.65	874.63
四碱基	81 765	1 997 928	13.97%	24.44	72.55	1 772.78
五碱基	35 363	2 148 330	6.04%	60.75	31.38	1 906.23
六碱基	10 207	627 924	1.74%	61.52	9.06	557.16

2.2 凤头潜鸭基因组微卫星的优势重复拷贝类别特征

凤头潜鸭基因组中不同重复类型微卫星的优势重复拷贝类别统计结果如表 2 所示,统计范围为各重复类型内数量排名前 3 位的拷贝类别。单碱基重复类型中,A 拷贝类别(A/T)占绝对数量优势,共检测到 347 835 个,占基因组微卫星总数的 59.42%;拷贝类别 C 数量则相对很少,仅检测到 19 635 个,占基因组微卫星总数的比例仅 3.35%。二碱基重复类型中,AC 重复拷贝类别数量最多,其次是 AG 和 AT,CG 类别是在二碱基的 4 种拷贝类别中数量最少的 1 种,仅有 119 个。在三碱基重复中,AAT 出现的频率最高,其后依次为 AAC 和 AGC。四碱基重复里,数量排名前 3 的拷贝类别分别是 AAAC、AAAT 和 AAAG。五碱基重复中,AAAAC 数量最多,其次是 AAAAT 和 AAGGG。六碱基类型中,仅 AATCCC、AAAAAG 和 AAAAAT 这 3 种重复拷贝类别的微卫星数量在 1 200 个以上,其余类别数量较少。总体来看,凤头潜鸭基因组微卫星在碱基组成上呈现出明显的 A/T 偏好性。

凤头潜鸭基因组中,分布数量最多的 10 种微卫星拷贝类别按数量由多到少依次是 A、AAAC、AAAT、AT、AAT、C、AC、AAC、AAAAC 和 AAAAT,它们在基因组中的数量均超过 9 200 个,总数达 518 077 个,占凤头潜鸭基因组微卫星总数的 88.5%。

表 2 凤头潜鸭基因组中不同重复类型微卫星的优势重复拷贝类别
Tab. 2 The dominant motif categories of microsatellites with different repeat types in the *A. fuligula*

重复类型	拷贝类别	数量/个	数量占比	相对丰度/(个·Mb ⁻¹)	相对密度/(bp·Mb ⁻¹)
单碱基	A	347 835	59.42%	308.64	5 613.08
	C	19 635	3.35%	17.42	277.03
二碱基	AC	11 490	1.96%	10.20	177.95
	AG	8 011	1.37%	7.11	129.98
	AT	25 110	4.29%	22.28	431.28
三碱基	AAT	19 651	3.36%	17.44	432.29
	AAC	10 967	1.87%	9.73	172.35
	AGG	5 749	0.98%	5.10	104.54
四碱基	AAAC	32 237	5.51%	28.60	539.43
	AAAT	31 573	5.39%	28.01	583.74
	AAAG	5 406	0.92%	4.80	233.46
五碱基	AAAAC	10 297	1.76%	9.14	212.86
	AAAAT	9 282	1.59%	8.24	479.55
	AAGGG	2 482	0.42%	2.2	328.35
六碱基	AATCCC	1 625	0.28%	1.44	134.29
	AAAAAG	1 383	0.24%	1.23	86.08
	AAAAAT	1 291	0.22%	1.15	52.97

2.3 凤头潜鸭各条染色体上微卫星的分布特征

凤头潜鸭各条染色体上的微卫星分布差异十分明显,具体结果见表 3。1 号染色体的微卫星数量最多,有 124 070 个,占微卫星总数的 21.20%;2 号染色体的微卫星数量次之,共 95 507 个,占微卫星总数的 16.32%;3 号染色体有 67 765 个微卫星,占微卫星总数的 11.58%。微卫星数量最少的为 34 号染色体,仅 137 个,其次是 31 号染色体,仅 299 个。Pearso 相关性分析结果显示,微卫星数量与染色体长度存在统计学意义上的正相关($r=0.993, p<0.001$),表明染色体长度越长则该染色体上分布的微卫星数量越多。

表 3 还显示:Z 染色体上的微卫星相对丰度最高,为 664.55 个·Mb⁻¹;其次是 1 号染色体,微卫星相对丰度为 599.32 个·Mb⁻¹;2 号染色体微卫星相对丰度排第 3 位,为 598.23 个·Mb⁻¹。Z 染色体上的微卫星相对密

度同样处于最高水平,为 $17\,879.97\text{ bp}\cdot\text{Mb}^{-1}$;其次是 1 号染色体,微卫星相对密度为 $14\,282.39\text{ bp}\cdot\text{Mb}^{-1}$; 2 号染色体上的微卫星相对密度位列第 3,为 $13\,921.40\text{ bp}\cdot\text{Mb}^{-1}$ 。W 染色体的微卫星数量和相对丰度均明显偏低,微卫星数量为 3 822 个,相对丰度为 $203.98\text{ 个}\cdot\text{Mb}^{-1}$,表明性染色体之间的上述指标差异较大。

表 3 凤头潜鸭各条染色体上微卫星统计
Tab. 3 Statistics of microsatellites on each chromosome of the *A. fuligula*

染色体	染色体长/bp	微卫星数量/个	比例	相对丰度/(个 $\cdot\text{Mb}^{-1}$)	相对密度/(bp $\cdot\text{Mb}^{-1}$)
chr1	207 018 403	124 070	21.20%	599.32	14 282.39
chr2	159 650 012	95 507	16.32%	598.23	13 921.40
chr3	119 491 689	67 765	11.58%	567.11	12 930.09
chr4	76 878 203	41 563	7.10%	540.63	11 520.70
chr5	65 067 180	30 436	5.20%	467.76	9 700.37
chr6	39 514 636	18 575	3.17%	470.08	9 388.62
chr7	37 840 942	17 240	2.95%	455.59	8 931.33
chr8	32 406 295	14 603	2.49%	450.62	9 173.96
chr9	26 499 672	11 453	1.96%	432.19	8 635.43
chr10	22 162 769	9 548	1.63%	430.81	8 226.36
chr11	21 928 510	8 986	1.54%	409.79	8 292.99
chr12	21 526 058	9 540	1.63%	443.18	8 640.74
chr13	21 228 795	9 060	1.55%	426.78	8 374.66
chr14	19 924 658	8 267	1.41%	414.91	8 320.90
chr15	17 816 505	6 760	1.15%	379.42	7 464.26
chr16	16 262 610	6 446	1.10%	396.37	8 079.33
chr17	15 484 054	5 646	0.96%	364.63	7 148.58
chr18	12 903 739	4 740	0.81%	367.34	7 338.96
chr19	11 953 284	4 378	0.75%	366.26	7 575.74
chr20	11 861 279	3 920	0.67%	330.49	6 220.66
chr21	8 424 528	3 024	0.52%	358.95	7 035.29
chr22	7 582 326	2 722	0.47%	358.99	7 026.21
chr23	7 507 630	2 756	0.47%	367.09	9 755.54
chr24	6 710 917	2 873	0.49%	428.11	8 734.42
chr25	6 693 931	2 388	0.41%	356.74	8 960.36
chr26	5 835 185	2 666	0.46%	456.88	10 683.12
chr27	5 453 253	2 154	0.37%	394.99	7 504.88
chr28	3 167 070	1 662	0.28%	524.78	9 806.22
chr29	2 844 608	1 616	0.28%	568.09	10 417.96
chr30	1 560 659	411	0.07%	263.35	4 430.82
chr31	1 405 878	299	0.05%	212.68	7 072.45
chr32	1 164 621	357	0.06%	306.54	8 364.95
chr33	983 965	565	0.10%	574.21	13 684.43
chr34	723 082	137	0.02%	189.47	9 436.00
chrW	18 736 694	3 822	0.65%	203.98	4 517.50
chrZ	85 905 639	57 089	9.75%	664.55	17 879.97

对不同重复类型微卫星在凤头潜鸭各染色体上的分布进行分析,结果见图 1。单碱基重复在所有染色体上均为优势类型,尤其在 chr1、chr2 和 chr3 等长染色体上数量最多,分别为 72 835、57 741 和 42 187 个。四碱基重复在多数染色体上为第 2 优势类型,在 chr1、chr2 和 chrZ 上分别有 19 969、15 015 和 10 623 个。在部分染色体上,第 2 优势类型为三碱基重复;其中 chr18、chr19 和 chrW 的三碱基重复数量分别为 429、371 和 785 个,均高于同染色体的四碱基重复数量,体现出微卫星重复类型存在局部区域的分布偏好。五碱基和六碱基重复整体数量较少,主要分布在 chr1、chr2、chr3、chrZ 等长染色体上,在 chr30、chr31 等小型染色体上几乎缺失。

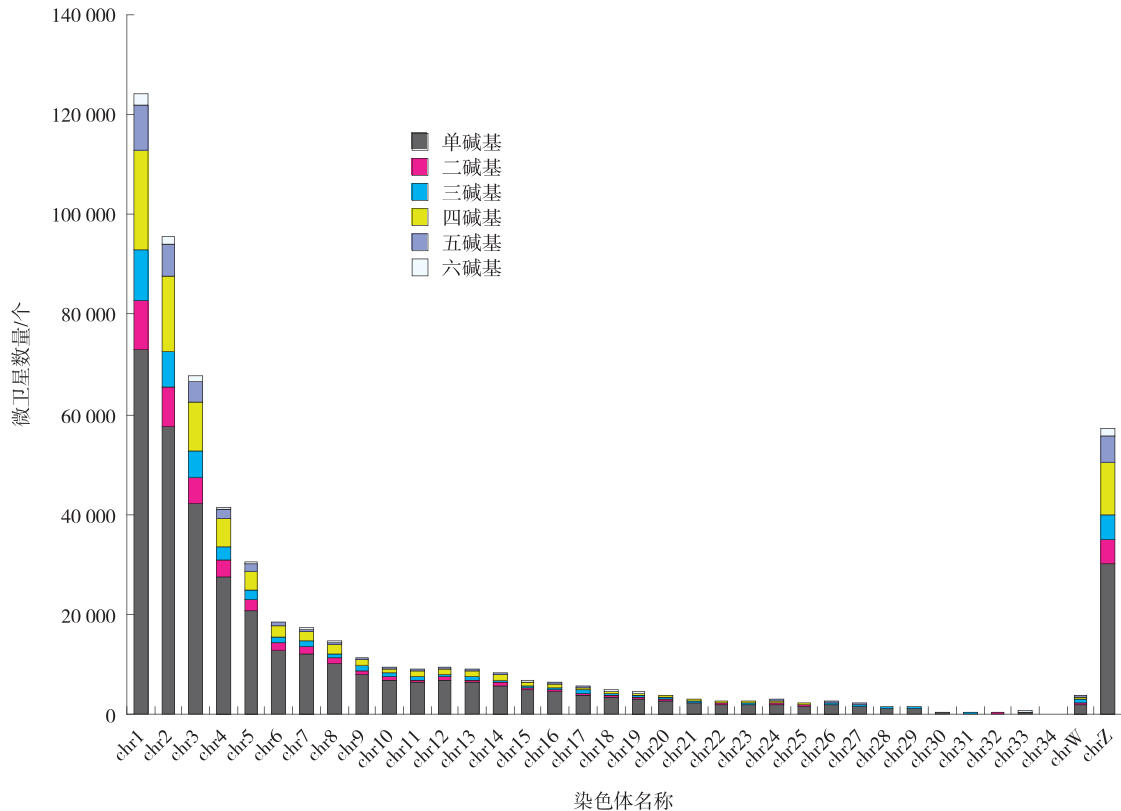


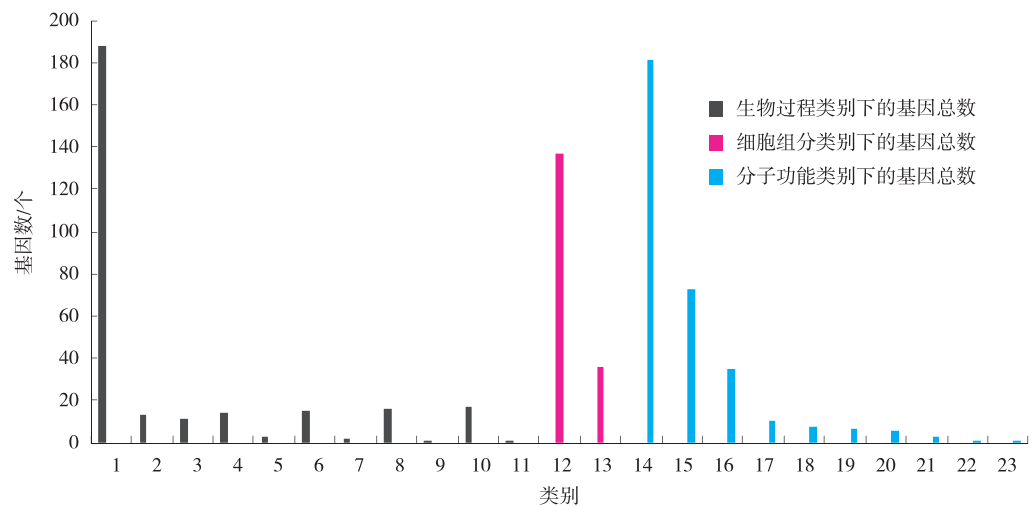
图 1 6 种碱基重复类型微卫星在凤头潜鸭各条染色体上的分布情况

Fig. 1 Distribution of six types of base repeat microsatellites across chromosomes of the *A. fuligula*

2.4 GO 功能注释和 KEGG 富集特征

研究结果显示,凤头潜鸭全基因组中共有 4 221 个微卫星分布于 1 883 个不同基因中。如图 2 所示,对这些基因进行 GO 注释,其中 727 个基因成功获得注释,共获得 1 164 个 GO 条目。其中,637 个 GO 条目归于生物过程(biological process),主要参与细胞过程、节律过程和免疫系统过程;282 个归于细胞组分(cellular component),主要参与细胞解剖结构和含蛋白质复合物;245 个归于分子功能(molecular function);主要参与结合(binding)、催化活性(catalytic activity)和转录调节活性。

对凤头潜鸭含有微卫星的 CDS 基因进行 KEGG 通路富集分析,共获得 135 条通路。按照 KEGG BRITE 功能分类体系,这些通路可以分成 6 大类。其中代谢类通路最多,有 60 条,涉及糖、脂、氨基酸、核苷酸、聚糖等代谢途径;环境信息处理类通路 35 条,主要涉及信号转导和膜运输;细胞过程类通路 20 条,涉及细胞周期、自噬、凋亡、细胞衰老及细胞运动等;生物体系统类通路 10 条,涉及免疫、内分泌和神经系统;人类疾病类通路 8 条,为与感染性疾病和癌症相关的通路;遗传信息处理类通路最少,只有 2 条,涉及核糖体生物发生和内质网蛋白质加工。进一步分析存在统计学意义富集的通路($p < 0.05$),一共 15 条,结果见表 4。其中,FoxO 信号通路(FoxO signaling pathway)、真核生物核糖体生物发生(ribosome biogenesis in eukaryotes)及 Notch 信号通路(Notch signaling pathway)富集效应在 $\alpha = 0.001$ 的显著性水平下通过统计学显著性检验。此外,鞘脂代谢(sphingolipid metabolism)、细胞周期(cell cycle)及内质网中的蛋白质加工(protein processing in endoplasmic reticulum)等通路亦存在统计学意义上的富集($p < 0.05$)。



注:1 为细胞过程,2 为多细胞生物过程,3 为定位,4 为免疫系统过程,5 为稳态过程,6 为对刺激的应答,7 为发育过程,8 为生物过程的调节,9 为生殖过程,10 为节律过程,11 为生物调节,12 为细胞解剖结构,13 为含蛋白质复合物,14 为结合,15 为催化活性,16 为转录调节活性,17 为分子功能调节活性,18 为转运活性,19 为分子传导活性,20 为分子适配器活性,21 为结构分子活性,22 为 ATP 依赖活性,23 为翻译因子活性。

图 2 凤头潜鸭外显子含微卫星的基因 GO 功能注释

Fig. 2 GO annotation of exonic microsatellite-containing genes in the *A. fuligula*

表 4 凤头潜鸭外显子含微卫星的基因 KEGG 富集结果

Tab. 4 KEGG enrichment results of exonic microsatellite-containing genes in the *A. fuligula*

功能分类	通路	KEEG ID	输入数/个	背景数/个	<i>p</i> 值
遗传信息处理	真核生物核糖体生物发生	apla03008	13	69	0.000 7
	内质网中的蛋白质加工	apla04141	16	139	0.015 5
代谢	鞘脂代谢	apla00600	10	48	0.001 5
	鞘糖脂生物合成-神经节系列	apla00604	4	14	0.016 8
	糖胺聚糖降解	apla00531	5	18	0.008 4
	N-聚糖生物合成	apla00510	8	44	0.008 7
	糖胺聚糖生物合成-硫酸软骨素/硫酸皮肤素	apla00532	5	19	0.010 2
环境信息处理	FoxO 信号通路	apla04068	20	123	0.000 2
	Notch 信号通路	apla04330	10	43	0.000 7
	ECM-受体相互作用	apla04512	11	84	0.019 0
	胰岛素信号通路	apla04910	13	109	0.021 6
	磷脂酰肌醇信号系统	apla04070	11	89	0.026 8
细胞过程	自噬-动物	apla04140	14	133	0.040 9
	细胞周期	apla04110	15	110	0.005 0
	细胞衰老	apla04218	15	134	0.022 9

2.5 外显子微卫星对编码区的潜在功能影响

为评估外显子区微卫星的潜在功能影响,本研究分析了这些微卫星在 CDS 中的阅读框相位分布。结果显示,在 4 221 个外显子微卫星中,三碱基重复共 3 914 个,占总数的 92.72%,涉及 1 726 个基因。三碱基重复的长度变异不会造成阅读框移码突变,但它的重复次数的变异可能导致氨基酸重复序列的产生,进而影响蛋白质功

能。此外,预测可能引起移码突变的非三碱基重复微卫星,共涉及 155 个基因。进一步统计发现,CDS 区微卫星数量最多的前 10 个基因如表 5 所示,其中 *KIF1A* 基因含有 35 个微卫星,*ARID1B* 基因含有 32 个微卫星,*BPTF* 基因含有 30 个微卫星。这些基因为后续研究微卫星介导的表型变异提供了候选靶点。

表 5 外显子微卫星数量排名前 10 的基因
Tab. 5 Top 10 genes ranked by the number of exonic microsatellites

基因名称	外显子微卫星数量/个	基因名称	外显子微卫星数量/个
<i>KIF1A</i>	35	<i>MAP4K4</i>	27
<i>ARID1B</i>	32	<i>KCNMA1</i>	27
<i>BPTF</i>	30	<i>KMT2C</i>	24
<i>NCOR2</i>	28	<i>WNK1</i>	20
<i>MEF2D</i>	28	<i>PAN3</i>	20

3 讨论

3.1 凤头潜鸭全基因组微卫星含量分析

本研究以已公布的凤头潜鸭全基因组序列为基础,利用生物信息学方法对该物种全基因组微卫星进行搜索,共得到 585 346 个微卫星位点,总长度为 13 232 974 bp,占全基因组总长度的 1.17%。将该结果与已报道的其他物种基因组微卫星含量进行对比后可知,凤头潜鸭基因组微卫星数量占比在鸟类中处于较高水平:绿头鸭基因组微卫星数量占比为 0.83%^[14],白鹭为 0.36%^[15],四川山鹧鸪(*Arborophila rufipectus*)为 0.43%^[8],红原鸡(*Gallus gallus*)为 0.49%^[18],普通鸬鹚(*Phalacrocorax carbo*)为 0.23%^[19],绿尾虹雉(*Lophophorus lhuysii*)为 0.54%^[20],虎皮鹦鹉(*Melopsittacus undulatus*)为 0.41%^[21];而与人(*Homo sapiens*)的 3%^[22]、绵羊(*Ovis aries*)的 4.8%^[23]、牛(*Bos taurus*)的 4.7%^[23]、小鼠(*Mus musculus*)的 2.85%^[24]等哺乳动物全基因组微卫星数量占比相比,凤头潜鸭的全基因组微卫星数量占比则相对较低。Primmer 等人^[25]预测鸟类基因组中微卫星比例低于人,本研究结果为这一观点提供了证据支持。这种差异可能源于多方面因素。首先,全基因组微卫星含量与基因组大小、重复序列组成以及进化历史密切相关^[26]。凤头潜鸭是迁徙性水禽,它的基因组可能经历过特殊的进化压力,所以微卫星序列的扩张程度比其他已报道的鸟类要高。另外,不同研究用的微卫星搜索参数如最低重复次数、碱基类型等不完全相同,也会影响检出的微卫星数量。除此之外,基因组组装的完整性和质量也会影响统计结果。本研究使用的凤头潜鸭基因组达到了染色体级别组装,因此可能检出了更多的微卫星。

3.2 凤头潜鸭不同碱基类型微卫星特征分析

凤头潜鸭全基因组微卫星中,单碱基重复类型数量最多,占微卫星总数的 62.78%,六碱基类型占比最少,这一分布规律与多数鸟类、哺乳动物及鱼类物种一致^[8,18,21,27]。与哺乳动物及鱼类以二碱基重复为第 2 优势类型微卫星不同,凤头潜鸭第 2 优势类型微卫星为四碱基重复。这一特征与已报道的绿头鸭^[14]、四川山鹧鸪^[8]、红原鸡^[18]、普通鸬鹚^[19]、绿尾虹雉^[20]等多数鸟类相同,仅发现白鹭和虎皮鹦鹉以二碱基重复为第 2 优势类型微卫星^[15,21]。上述结果表明,鸟类基因组可能普遍存在四碱基微卫星偏好,但不同物种间仍存在差异,这与 Webster 等人^[28]在黑猩猩(*Pan troglodytes*)与人基因组中发现的亲缘相近物种微卫星分布亦存在差异的结果一致。

在碱基组成上,凤头潜鸭微卫星呈现明显的 A 碱基偏倚。单碱基重复中 A 拷贝类别最多,与多数真核生物基因组中普遍存在的 ployA 现象有关^[29];二碱基重复中 AT 拷贝类别最多,GC 拷贝类别最少,后者可能与 GC 间氢键数较多、不易断裂及 CpG 甲基化后脱氨基作用有关^[30];在三碱基重复中,AAT 拷贝类别微卫星数量最多,这与上述多种鸟类,哺乳类的三碱基优势拷贝类别对比,并未发现明显规律。四碱基重复中 AAAX(X 代表除 A 以外的碱基)拷贝类别最为丰富。值得注意的是,四碱基微卫星在 PCR 扩增时不易产生滑带或影子带,基因型判读更准确,是优于二、三碱基重复的微卫星标记类型^[31]。本研究共鉴定出 81 765 个四碱基微卫星,为凤头潜鸭四碱基微卫星标记开发奠定了重要基础。

3.3 凤头潜鸭各染色体上微卫星的分布特征

凤头潜鸭的微卫星在染色体上分布与染色体长度关系密切,数量随长度增加而明显增多($r=0.993$, $p<$

0.001),这一结果和人类、鸡等已有的研究结果一致^[18,22]。Z染色体的微卫星相对丰度与相对密度均明显高于常染色体,分别为 $664.55 \text{ 个} \cdot \text{Mb}^{-1}$ 和 $17\,879.97 \text{ bp} \cdot \text{Mb}^{-1}$;而W染色体的微卫星相对丰度明显偏低,仅为 $203.98 \text{ 个} \cdot \text{Mb}^{-1}$ 。这种性染色体上的微卫星数量差异符合鸟类的进化特点——Z染色体保守性强,W染色体则趋于退化^[32]。从重复类型上看,所有凤头潜鸭染色体上都是单碱基重复最多。第2优势类型存在分化:长染色体上四碱基重复占优势,而部分小型染色体和W染色体上却是三碱基重复排第2位。这可能跟不同染色体的基因密度和染色质结构有关。五碱基、六碱基重复只出现在chr1、chr2、chr3、chrZ这些长染色体上,小型染色体里几乎没有。长重复单元更容易发生突变,可能需要更长的侧翼序列才能稳定存在;而小型染色体结构更紧凑,可能对这类重复有更强的抑制作用^[33]。这些分布特点说明微卫星在基因组里不是随机分布的,也为研究鸟类性染色体进化和不同重复单元的生物意义提供了一些线索。

3.4 凤头潜鸭外显子含微卫星的基因功能富集分析

微卫星在基因组中的分布并非随机,它在编码区、非编码区及基因间区的分布特征往往反映了不同的生物学功能^[12]。编码区微卫星的重复数变异可通过改变蛋白质结构域影响蛋白质的构象稳定性、转录活性或蛋白-蛋白相互作用,从而为物种快速适应环境变化提供遗传变异基础^[34]。此外,非三碱基重复如果发生长度变异(非3倍数),则可能导致移码突变,产生截短蛋白或功能丧失。本研究对凤头潜鸭含有微卫星的基因进行GO功能注释和KEGG通路富集分析,结果显示,含有微卫星的CDS基因共涉及135条通路,其中代谢类通路占主导,遗传信息处理类最少,仅2条。本研究富集到的通路已报道的相关微卫星研究存在一定差异。例如,在有关大蹄蝠(*Hipposideros armiger*)的研究中信号传导通路最为明显^[35],而对鳞上碘泡虫(*Myxobolus episquamalis*)的研究中代谢相关通路占主导地位^[36]。这种差异可能与物种特异性以及微卫星分布区域等有关。在含有微卫星的基因存在统计学意义富集的15条通路($p < 0.05$)中,FoxO信号通路、Notch信号通路及真核生物核糖体生物发生的富集效应最为明显。FoxO信号通路在细胞增殖、凋亡和代谢调控中起关键作用^[37],该通路的富集可能反映了凤头潜鸭对迁徙过程中能量需求的适应;Notch信号通路对细胞命运决定、细胞存活及干细胞维持等多种发育与稳态过程具有重要调控作用^[38],可能与凤头潜鸭作为迁徙性水禽在神经系统发育、能量代谢调控及环境适应中的生物学需求密切相关。遗传信息处理类通路非常少,这也符合编码区的微卫星主要影响蛋白质结构而不是基础转录翻译过程。本研究揭示了凤头潜鸭外显子微卫星相关基因可能在能量代谢、神经系统发育等过程中的潜在功能,这为理解它们如何适应迁徙生活提供了一些分子层面的线索。

值得注意的是,本研究在外显子区筛选出一批富含微卫星的基因,其中KIF1A、ARID1B和BPTF基因尤为突出,它们的外显子微卫星数量分别达到了35个、32个和30个。KIF1A编码神经元特异性驱动蛋白-3家族成员,主要参与突触囊泡前体等物质的轴突长距离运输,对神经系统的发育和功能维持至关重要^[39]。该基因外显子区丰富的微卫星序列若发生长度变异,有可能影响它编码蛋白的结构或功能,进而与凤头潜鸭的神经适应性形成关联。ARID1B和BPTF基因分别参与染色质重塑和转录调控过程^[40-41],它们所携带的编码区微卫星也可能通过影响基因表达而参与环境适应。编码区微卫星的高突变率,为凤头潜鸭在长途迁徙这一高能耗、高风险的生命活动中快速产生适应性表型变异提供了丰富的遗传底物。此外,本研究还获得155个可能发生移码突变的基因和1726个含三碱基重复的基因,这些列表为后续探究微卫星的生物学功能提供了候选资源。

综上所述,凤头潜鸭全基因组微卫星含量较为丰富,位于外显子区域的微卫星可能通过参与转录调控和神经系统发育等过程,在鸟类的生态适应和进化中发挥重要作用。本研究结果为今后开发微卫星标记、分析种群遗传多样性以及开展功能验证研究提供了基础数据。当然,后续还需要通过实验来验证某些微卫星位点的实际功能。

参考文献:

- [1] BirdLife International. IUCN Red List for birds [DS/OL]. (2026) [206-04-09].
- [2] Cramp S, Simmons K E L, Brooks D C, et al. Handbook of the birds of Europe, the Middle East and North Africa. The birds of the Western Palearctic: 3. Waders to gulls[M]. Oxford: Oxford University Press, 1983.
- [3] Scott D A, Rose P M. Atlas of Anatidae populations in Africa and western Eurasia[M]. Wageningen: Wetlands International, 1996.
- [4] Avilova K V. Dynamics of the structure and numbers of the urban population of the tufted duck (*Aythya fuligula*,

- Anseriformes, Anatidae) in Moscow[J]. *Biology Bulletin*, 2023, 50 (8): 1941-1953.
- [5] Liu Y, Keller I, Heckel G. Breeding site fidelity and winter admixture in a long-distance migrant, the tufted duck (*Aythya fuligula*) [J]. *Heredity*, 2012, 109 (2): 108-116.
- [6] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes [J]. *Nucleic Acids Research*, 1984, 12 (10): 4127-4138.
- [7] 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.
- Zhou Siqian, Jiao Weili, Peng Zhuli, et al. Analysis of microsatellite sequence distribution in ebolavirus genomes [J]. *Genomics and Applied Biology*, 2019, 38 (3): 1087-1095.
- [8] Huang J, Zhu D, Song X H, et al. High-throughput microsatellite markers discovery for the Sichuan Hill Partridge (*Arborophila rufipectus*) and assessment of genetic diversity in the Laojunshan population [J]. *Biochemical Systematics and Ecology*, 2015, 60: 266-272.
- [9] 黄皓, 范嗣刚, 王鹏飞, 等. 基于微卫星标记对 6 个花鲈群体的遗传多样性分析[J]. *南方水产科学*, 2022, 18 (1): 99-106.
- Huang Hao, Fan Sigang, Wang Pengfei, et al. Genetic diversity analysis of six geographical populations of *Lateolabrax maculatus* based on microsatellite markers [J]. *South China Fisheries Science*, 2022, 18 (1): 99-106.
- [10] 袁一铭, 张璐, 赵越, 等. 微卫星标记在非人灵长类动物亲权鉴定中的研究进展[J]. *中国实验动物学报*, 2025, 33 (12): 1817-1827.
- Yuan Yiming, Zhang Lu, Zhao Yue, et al. Research progress on microsatellite markers in paternity testing in non-human Primates [J]. *Acta Laboratorium Animalis Scientia Sinica*, 2025, 33 (12): 1817-1827.
- [11] Jaskulak M, Rorat A, Vandenbulcke F, et al. Polymorphic microsatellite markers demonstrate hybridization and interspecific gene flow between lumbricid earthworm species, *Eisenia andrei* and *E. fetida* [J]. *PLoS One*, 2022, 17 (2): e0262493.
- [12] Cheng M L, Ren J Y, Shen F J, et al. Genome-wide investigation of microsatellite polymorphism in coding region of the giant Panda (*Ailuropoda melanoleuca*) genome: a resource for study of phenotype diversity and abnormal traits [J]. *Mammal Research*, 2019, 64 (3): 353-363.
- [13] Du L M, Zhang C, Liu Q, et al. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design [J]. *Bioinformatics*, 2018, 34 (4): 681-683.
- [14] 黄杰, 徐玉荣, 付红燕, 等. 绿头鸭基因组微卫星分布规律研究[J]. *重庆师范大学学报 (自然科学版)*, 2021, 38 (5): 37-42.
- Huang Jie, Xu Yurong, Fu Hongyan, et al. Genomic microsatellite characteristic analysis for *Anas platyrhynchos* [J]. *Journal of Chongqing Normal University (Natural Science)*, 2021, 38 (5): 37-42.
- [15] 黄杰, 杨波, 贾银平, 等. 白鹭基因组微卫星分布规律研究[J]. *重庆师范大学学报 (自然科学版)*, 2019, 36 (5): 66-71.
- Huang Jie, Yang Bo, Jia Yinping, et al. The distribution regularities of microsatellites in *Egretta garzetta* genome [J]. *Journal of Chongqing Normal University (Natural Science)*, 2019, 36 (5): 66-71.
- [16] Yu G C, Wang L G, Han Y Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters [J]. *OMICS: A Journal of Integrative Biology*, 2012, 16 (5): 284-287.
- [17] 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 (Web Server issue): W316-W322.
- [18] 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.
- Huang Jie, Du Lianming, Li Yuzhi, et al. Distribution regularities of microsatellites in the *Gallus gallus* genome [J]. *Sichuan Journal of Zoology*, 2012, 31 (3): 358-363.
- [19] 黄杰, 刘磊, 杨波, 等. 普通鸬鹚基因组微卫星分布规律研究[J]. *野生动物学报*, 2020, 41 (1): 108-114.
- Huang Jie, Liu Lei, Yang Bo, et al. Distribution regularities of microsatellites in the genome of great cormorant (*Phalacrocorax carbo*) [J]. *Chinese Journal of Wildlife*, 2020, 41 (1): 108-114.
- [20] 崔凯, 岳碧松. 绿尾虹雉全基因组微卫星分布规律研究[J]. *四川动物*, 2018, 37 (5): 533-540.
- Cui Kai, Yue Bisong. Distribution patterns of microsatellites in the genome of *Lophophorus lhuysii* [J]. *Sichuan Journal of Zoology*, 2018, 37 (5): 533-540.
- [21] 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.
- Huang Jie, Yuan Baodong, Yang Chengzhong. Distribution of microsatellites in genome of budgerigar (*Melopsittacus*

- undulatus*) [J]. Chinese Journal of Wildlife, 2017, 38 (3): 422-426.
- [22] 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.
- [23] 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[J]. 畜牧兽医学报, 2013, 44 (11): 1724-1733.
- Qi Wenhua, Jiang Xuemei, Xiao Guosheng, et al. Seeking and bioinformatics analysis of microsatellite sequence in the genomes of cow and sheep[J]. Acta Veterinaria et Zootechnica Sinica, 2013, 44 (11): 1724-1733.
- [24] 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.
- Tong Xiaoling, Dai Fangyin, Li Bin, et al. Microsatellite repeats in mouse: abundance, distribution and density[J]. Acta Zoologica Sinica, 2006, 52 (1): 138-152.
- [25] Primmer C R, Raudsepp T, Chowdhary B P, et al. Low frequency of microsatellites in the avian genome[J]. Genome Research, 1997, 7 (5): 471-482.
- [26] Hancock J M. Simple sequences and the expanding genome[J]. BioEssays, 1996, 18 (5): 421-425.
- [27] 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.
- Liang Xia, Wang Huiqi, Ma Yuxuan, et al. Distribution characteristics of microsatellites in the whole genome of *Cyprinus carpio*, Linnaeus[J]. Journal of Nanjing Normal University (Natural Science Edition), 2021, 44 (3): 103-111.
- [28] Webster M T, Smith N G, Ellegren H. Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99 (13): 8748-8753.
- [29] Gallie D R. The cap and poly (A) tail function synergistically to regulate mRNA translational efficiency[J]. Genes & Development, 1991, 5 (11): 2108-2116.
- [30] 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.
- [31] Archie E A, Moss C J, Alberts S C. Characterization of tetranucleotide microsatellite loci in the African savannah elephant (*Loxodonta Africana Africana*) [J]. Molecular Ecology Notes, 2003, 3 (2): 244-246.
- [32] Stiglec R, Ezaz T, Graves J A M. A new look at the evolution of avian sex chromosomes[J]. Cytogenetic and Genome Research, 2007, 117 (1/2/3/4): 103-109.
- [33] Campregheer C, Scharl T, Nemeth M, et al. The nucleotide composition of microsatellites impacts both replication fidelity and mismatch repair in human colorectal cells[J]. Human Molecular Genetics, 2010, 19 (13): 2648-2657.
- [34] Liu S X, Hou W, Sun T L, et al. Genome-wide mining and comparative analysis of microsatellites in three macaque species[J]. Molecular Genetics and Genomics, 2017, 292 (3): 537-550.
- [35] 邵伟伟, 乔芬, 蔡玮, 等. 大蹄蝠全基因组微卫星分布特征分析[J]. 兽类学报, 2023, 43 (2): 182-192.
- Shao Weiwei, Qiao Fen, Cai Wei, et al. Characteristics of microsatellite distributions in genomes of *Hipposideros armiger* (Chiroptera) [J]. Acta Theriologica Sinica, 2023, 43 (2): 182-192.
- [36] 杨承忠, 杨雨婷, 李海涛, 等. “鲢麻风病”病原鳞上碘泡虫全基因组微卫星分布特征[J]. 水生生物学报, 2024, 48 (12): 2100-2108.
- Yang Chengzhong, Yang Yuting, Li Haitao, et al. Genomic distribution characteristics of microsatellites in *Myxobolus episquamalis* (the pathogen of mullet leprosy) [J]. Acta Hydrobiologica Sinica, 2024, 48 (12): 2100-2108.
- [37] Liu W, Meng P J, Li Z, et al. The multifaceted impact of physical exercise on FoxO signaling pathways[J]. Frontiers in Cell and Developmental Biology, 2025, 13: 1614732.
- [38] Meng Y, Bo Z H, Feng X Y, et al. The Notch signaling pathway: mechanistic insights in health and disease[J]. Engineering, 2024, 34: 212-232.
- [39] Kern J V, Zhang Y V, Kramer S, et al. The kinesin-3, unc-104 regulates dendrite morphogenesis and synaptic development in *Drosophila*[J]. Genetics, 2013, 195 (1): 59-72.
- [40] Khursheed M, Kolla J N, Kotapalli V, et al. ARID1B, a member of the human SWI/SNF chromatin remodeling complex, exhibits tumour-suppressor activities in pancreatic cancer cell lines[J]. British Journal of Cancer, 2013, 108 (10): 2056-2062.
- [41] Landry J, Sharov A A, Piao Y L, et al. Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells[J]. PLoS Genetics, 2008, 4 (10): e1000241.

Animal Sciences

Distribution Characteristics and Functional Analysis of Microsatellites in the Chromosome-Level Genome of the Tufted Duck (*Aythya fuligula*)HUANG Jie¹, ZHONG Huaming¹, QIAO Jiayi¹, YANG Chengzhong², YANG Bo³

(1. College of Biology and Food, Shangqiu Normal University, Shangqiu Henan 476000;

2. Chongqing Key Laboratory of Conservation and Utilization of Freshwater Fishes, Animal Biology Key Laboratory of Chongqing Education Commission of China, College of Life Sciences, Chongqing Normal University, Chongqing 401331;

3. China Conservation and Research Centre for the Giant Panda, Chengdu 610057, China)

Abstract: To systematically reveal the genome-wide distribution characteristics of microsatellites and the potential functions of microsatellite-related genes in the tufted duck (*Aythya fuligula*), this study based on the published chromosome-level genomic data, employed bioinformatics methods to systematically analyze the genome-wide distribution characteristics of microsatellites and performed functional annotation of genes containing microsatellites located in exonic regions. The results showed that, from the approximately 1,093 Mbp genome sequence of the tufted duck, a total of 585,346 microsatellites with motif lengths of 1–6 bp were identified, with a total length of 13,232,974 bp, accounting for 1.17% of the genome. Among the six repeat types, mononucleotide repeats were the most abundant (68.45%), followed by tetranucleotide (13.97%) and dinucleotide (10.48%) repeats, while pentanucleotide and hexanucleotide repeats constituted the smallest proportions. The base composition of microsatellites showed a clear A-base bias; among mononucleotide repeats, the A copy type was the most frequent; among dinucleotide repeats, the AT copy type was the most abundant; and tetranucleotide repeats were predominantly of the AAAX type. The top ten motifs in tufted duck genome were A, AAAC, AAAT, AT, AAT, C, AC, AAC, AAAAC and AAAAT in a descending order each with more than 9,200 copies, totaling 518,077 copies and accounting for 88.5% of all microsatellites. Chromosome-level distribution analysis revealed a highly significant positive correlation between microsatellite number and chromosome length ($r = 0.993$, $P < 0.001$). The Z chromosome exhibited significantly higher relative abundance and density than autosomes, whereas the W chromosome showed substantially lower relative abundance and density. GO annotation results indicated that genes involved in biological processes were the most common. KEGG pathway enrichment analysis identified 135 pathways, with the FoxO signaling pathway, Notch signaling pathway and sphingolipid metabolism being significantly enriched ($P < 0.05$). The enrichment of these pathways may be associated with migratory adaptation and neural navigation ability in the tufted duck. This study provides the first systematic characterization of genome-wide microsatellite distribution and functional enrichment of microsatellite-associated exonic genes in the tufted duck, laying a foundation for future development of microsatellite markers, population genetic diversity analyses, and investigations into environmental adaptation mechanisms.

Keywords: *Aythya fuligula*; genomic; microsatellite; distribution characteristics; functional analysis

(责任编辑 陈新颖)