

长江江豚和东亚江豚染色体水平基因组微卫星分布特征分析

李寒姝^{1,2} 杨连东¹ 孟明辉³ 许攀³ 范飞^{1,4} 王丁^{1,4} 郑劲松^{1,4}

(1. 中国科学院水生生物研究所, 中国科学院水生生物多样性与保护重点实验室, 武汉 430072; 2. 中国科学院大学, 北京 100049; 3. 迪格思(武汉)生物科技有限责任公司, 武汉 430070; 4. 国家水生生物种质资源库, 武汉 430072)

摘要: 研究采用生物信息学方法, 首次对长江江豚(*Neophocaena asiaeorientalis asiaeorientalis*)和东亚江豚(*N. a. sunameri*)的染色体水平基因组(大小分别为2.62和2.50 Gb)的微卫星分布特征进行了分析和比较。结果表明, 在长江江豚和东亚江豚基因组中分别发现888424和867792个完美型微卫星, 总长度分别为15340771和14724482 bp, 均占基因组序列总长度的0.59%。在长江江豚和东亚江豚基因组中, 六种重复类型微卫星的数量分布模式基本一致: 单碱基微卫星最为丰富, 其次是二碱基、六碱基、四碱基、三碱基和五碱基重复微卫星; 数目最多的6种微卫星重复单元分别是A、AC、AT、CCCTAA、AGGGTT和C。研究还着重分析了长江江豚和东亚江豚Y染色体微卫星分布特征, 分别在两个亚种Y染色体基因序列中发现6195和5024个完美型微卫星。在Y染色体6种重复类型微卫星中, 单碱基是最丰富的碱基重复类型, 然后依次是二碱基、四碱基、三碱基、五碱基和六碱基重复。在Y染色体不同类别的完美型微卫星中, 以A、AC、AAT等为重复单元的微卫星最为丰富。研究分析了长江江豚和东亚江豚染色体水平基因组微卫星分布特征和差异, 为后续微卫星遗传标记的开发与利用奠定了重要基础。

关键词: 长江江豚; 东亚江豚; 基因组; Y染色体; 微卫星

中图分类号: Q347 **文献标识码:** A **文章编号:** 1000-3207(2025)05-052508-09



自从白鱀豚(*Lipotes vexillifer*)2007年被宣布功能性灭绝^[1]之后, 长江江豚(*Neophocaena asiaeorientalis asiaeorientalis*)成为长江中唯一的鲸类物种, 仅分布于长江中下游干流以及鄱阳湖和洞庭湖中^[2]。在过去的30多年中, 由于人类活动干扰的加剧, 长江江豚种群急剧下降, 2013年被国际自然保护联盟(International Union for conservation of Nature and Nature Resources, IUCN)升级为极度濒危种群^[3]。截至2022年长江江豚自然种群仅剩约1249头, 亟待加强研究和保护^[4]。东亚江豚(*N. a. sunameri*)主要分布于中国和朝鲜沿岸的黄海、渤海和东海, 以及日本的日本海和太平洋温带沿岸海域。在我国, 东亚江豚被列为二级保护野生动物, 关于其种群数量、分布规律等生态学研究极少^[5]。2002—2013年

间, 我国学者应用线粒体DNA (MtDNA)、核微卫星(Microsatellite)及单核苷酸多态性位点(Single nucleotide polymorphisms, SNPs)等多种遗传标记, 对中国水域分布的江豚不同地理种群开展了遗传多样性评估、种群结构分析等系列研究。研究结果表明, 长江江豚与包括东亚江豚在内的海洋江豚之间存在显著的遗传结构, 并且长江江豚与海洋江豚之间的基因交流水平非常低^[6-9]。2018年, Zhou等^[10]基于基因组和种群基因组学的研究表明, 长江江豚已完全适应淡水环境, 与东亚江豚等海洋江豚之间存在显著并稳定的遗传分化, 已形成独立的进化支系并出现生殖隔离, 支持长江江豚应该具有独立的物种地位。然而, 目前长江江豚与东亚江豚依然被划分为窄脊江豚(*Neophocaena asiaeorientalis*)

收稿日期: 2024-11-22; 修订日期: 2024-12-18

基金项目: 国家重点研发计划(2022YFF1301600); 国家自然科学基金(31870372); 中国科学院科技服务网络计划(KFJ-STQ-QYZD-2021-27-001)资助 [Supported by National Key Programme of Research and Development (2022YFF1301600); the National Natural Science Foundation of China (31870372); Science and Technology Service Network Initiative, Chinese Academic of Sciences (KFJ-STQ-QYZD-2021-27-001)]

作者简介: 李寒姝, 女, 硕士研究生; 主要从事鲸类保护遗传学研究。E-mail: lihanshu@ihb.ac.cn

通信作者: 郑劲松, 男, 博士; 主要从事鲸类保护遗传学与分子生态学研究。E-mail: zhengjinsong@ihb.ac.cn

的两个亚种^[11]。

微卫星是一种广泛存在于真核生物、原核生物和病毒基因组中的短串联重复(Short Tandem Repeats, STR)序列, 亦称为简单序列重复(Simple Sequences Repeats, SSR)。微卫星序列由1—6个核苷酸组成的核心序列及其侧翼保守序列构成。根据微卫星的结构, 微卫星序列被分为完美型、不完美型、复合型三种类别。其中, 完美型微卫星是由不间断的单一重复单位构成的微卫星。微卫星核心序列长度在不同个体中具有差异, 因而具有较高的多态性。由于微卫星标记具有多态性高、广泛分布、选择中性、共显性遗传及易于分析操作等特点, 被广泛应用于遗传多样性评估、种群遗传结构分析、亲子鉴定、遗传谱系构建等诸多领域。此外, 微卫星是重要的基因组特征, 能够为物种鉴定、比较基因组和保护基因组等研究提供宝贵资源^[12]。

2022年, Yin等^[13]组装得到一头雄性东亚江豚的染色体水平基因组。最近, 作者所在的研究团队成功解析了一头雄性长江江豚的T2T (Telomere to Telomere)基因组。这两个高质量染色体水平基因组为从全基因组水平准确分析微卫星种类、数量、丰度等特征提供了基础, 特别是为分析Y染色体微卫星组成特征提供了便利和可能。本研究利用微卫星搜索软件, 扫描长江江豚和东亚江豚染色体水平基因组, 对它们的微卫星序列组成和分布特征进行分析与比较。该研究有助于了解长江江豚和东亚江豚这两个亚种在基因组构成方面的差异, 同时可为后续大批量开发有效微卫星遗传标记奠定基础。

1 材料与方法

1.1 基因组数据来源

近期, 研究者所在课题组采用ONT、HiFi、Hi-C等测序技术, 组装得到一头雄性长江江豚的T2T基因组, 大小为2.62 Gb, 相关数据已上传数据库(内部数据, 待发表)。本研究用于比较分析的雄性东亚江豚染色体水平基因组大小为2.50 Gb^[13], 数据从Figshare下载获得(<https://figshare.com/ndownloader/files/36523131>)。基于这两个染色体水平基因组, 本研究对长江江豚和东亚江豚进行全基因组微卫星分布特征分析和比较, 并且着重分析了Y染色体微卫星分布特征。

1.2 微卫星的鉴定、分类及数据统计

采用MISA软件(<https://webblast.ipk-gatersleben.de/misa/>)搜索长江江豚和东亚江豚全基因组, 定位

1—6个碱基重复完整型微卫星, 并进行后续统计分析。分析统计中标准参数设置如下: 单碱基重复次数 ≥ 10 个拷贝, 二碱基重复次数 ≥ 6 个拷贝, 三碱基重复次数 ≥ 5 个拷贝, 四碱基重复次数 ≥ 5 个拷贝, 五碱基重复次数 ≥ 5 个拷贝, 六碱基重复次数 ≥ 5 个拷贝。本研究采用自行编写的R脚本结合Origin软件进行数据统计及绘图。

2 结果

2.1 长江江豚和东亚江豚基因组各重复类型微卫星分布特征

本研究分别在长江江豚和东亚江豚基因组中鉴定出完美型微卫星888424和867792个, 总长度分别为15340771和14724482 bp, 均占基因组序列总长度的0.59%。长江江豚基因组中微卫星相对丰度为339.7个/Mb, 相对密度为5866.5 bp/Mb; 东亚江豚基因组中微卫星相对丰度为347.5个/Mb, 相对密度为5896.9 bp/Mb。在长江江豚基因组中, 占比最多的是单碱基类型, 有613752个, 占微卫星总数的69.08%, 相对丰度为234.7个/Mb; 其次是二碱基类型, 然后依次是六碱基、四碱基、三碱基和五碱基(表1)。东亚江豚与长江江豚的碱基类型数量排序基本一致, 占比最多是单碱基类型, 有611301个, 占微卫星总数的70.44%, 相对丰度为244.81个/Mb; 其次是二碱基类型, 然后依次是六碱基、四碱基、三碱基和五碱基(表1)。

2.2 不同碱基类型微卫星的重复次数比较

长江江豚基因组中数目最多的是单碱基重复10次的微卫星, 占总数比例15.69%, 然后依次是单碱基重复11、12和13次的微卫星, 分别占总数比例10.25%、7.84%和6.36%。东亚江豚与长江江豚极为相似, 数目最多的前七项类型完全一致, 但第八项出现不同, 长江江豚的第八项为重复5次六碱基类型, 而东亚江豚的为重复16次单碱基类型(表2)。此外, 在不同微卫星类型中, 随着微卫星核心序列拷贝数的增加, 微卫星数量呈现递减趋势。在长江江豚和东亚江豚基因组单碱基重复类型中重复拷贝数主要集中在10—24次; 在二碱基类型中, 重复拷贝数主要集中在6—25次; 在三、四、五、六碱基类型中, 重复拷贝数主要集中在5—9次(图1)。

2.3 长江江豚和东亚江豚重复拷贝类别

长江江豚与东亚江豚重复拷贝类别分布极为相似, 出现次数排名前三的类别均依次为A、AC、AT。C/G类别微卫星占比在两个亚种间存在差异, 在长江江豚和东亚江豚基因组中分别排名第六和第四(表3)。

表 1 长江江豚和东亚江豚染色体水平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 chromosomal genomes of the Yangtze finless porpoise and East Asian finless porpoise

SSR重复类型 SSR Motif	长江江豚 <i>N. a. asiaeorientalis</i>			东亚江豚 <i>N. a. sunameri</i>		
	SSR数目	SSR分布频率 (个/Mb)	占总数的比例 (%)	SSR数目	SSR分布频率 (个/Mb)	占总数的比例 (%)
单碱基Mononucleotide	613752	234.70	69.08	611301	244.81	70.44
二碱基Dinucleotide	163414	62.49	18.39	162279	64.99	18.70
三碱基Trinucleotide	22376	8.56	2.52	22243	8.91	2.56
四碱基Tetranucleotide	33502	12.81	3.77	33529	13.43	3.86
五碱基Pentanucleotide	2961	1.13	0.33	2985	1.20	0.34
六碱基Hexanucleotide	52419	20.05	5.90	35455	14.20	4.09
总计Total		888424			867792	
总长度Total length (bp)		15340771			14724482	
相对丰度Relative frequency (loci/Mb)		339.7			347.5	
相对密度Relative density (bp/Mb)		5866.5			5896.9	
基因组大小Genome size (bp)		2615550326			2497825093	
总长度占比Proportion (%)		0.59			0.59	

表 2 长江江豚和东亚江豚染色体水平基因组中数量最多的 11种微卫星

Tab. 2 The 11 most abundant microsatellites in chromosomal genome of the Yangtze finless porpoise and East Asian finless porpoise

排名 Ranking	长江江豚 <i>N. a. asiaeorientalis</i>				东亚江豚 <i>N. a. sunameri</i>			
	SSR 数目	Motif 长度	重复 次数	占总数 比例 (%)	SSR 数目	Motif 长度	重复 次数	占总数 比例 (%)
1	139412	1	10	15.69	138676	1	10	15.98
2	91099	1	11	10.25	91493	1	11	10.54
3	69621	1	12	7.84	69447	1	12	8.00
4	56543	1	13	6.36	56145	1	13	6.47
5	51348	2	6	5.78	50586	2	6	5.83
6	45965	1	14	5.17	45352	1	14	5.23
7	37919	1	15	4.27	37681	1	15	4.34
8	34456	6	5	3.88	30311	1	16	3.49
9	30095	1	16	3.39	24526	1	17	2.83
10	24741	1	17	2.78	24487	2	7	2.82
11	24526	2	7	2.76	23415	6	5	2.70

2.4 长江江豚和东亚江豚不同重复类型中优势重复类别

长江江豚和东亚江豚各碱基类型中优势碱基类别相同。二碱基重复类型中占比最多的是AC, 分别占长江江豚和东亚江豚微卫星总数的55.08%和55.36%, 其次依次是AT和AG。三碱基类型中的前三种优势碱基类别分别依次是AAT、AAC和ACC。在四碱基重复类型中, 占比最多的类别是AAAT, 其次依次是CCTC和AAAC。五碱基重复类型中以AAAAC碱基序列为最多。六碱基重复类型

中仅排名前二的AGGGTT和CCCTAA两个类别就分别占总数98.98%和98.61%, 是长江江豚和东亚江豚六碱基重复类型中的主要重复类别(表 4)。

2.5 长江江豚和东亚江豚Y染色体不同重复类型微卫星分布特征

长江江豚和东亚江豚Y染色体中共分别检测出 6195和5024个完美型微卫星, 总长度分别为99613和81069 bp, 相对丰度分别为413.0和456.7个/Mb, 相对密度分别为6640.9和7369.9 bp/Mb (表 5), 长江江豚与东亚江豚Y染色体中分别含有约0.63%和0.74%的微卫星序列。其中, 单碱基重复是两个物种Y染色体中最丰富的微卫星类型, 共分别检测到3663和2891个, 占有所有 SSRs的 59.13%和 57.54%。然后依次是二碱基、四碱基、三碱基、五碱基和六碱基。六碱基重复数目最少, 仅分别在长江江豚和东亚江豚Y染色体中检测到1和2个, 分别占Y染色体微卫星总数的0.02%和0.04%。

2.6 长江江豚和东亚江豚Y染色体不同重复类型中的优势重复类别

长江江豚和东亚江豚Y染色体中不同重复类型的优势碱基类别也极为相似。单碱基重复类型中A是主要的重复类别, 分别在长江江豚和东亚江豚中鉴定到3432个和2706个, 占单碱基总数的93.69%和93.60%。二碱基重复类型中占比最多的是AC, 分别占总数的75.4%和78.41%, 其次是AT, 排名第三的是AG。长江江豚和东亚江豚在三碱基类型中的前三种优势碱基类别略有不同, 在两个亚种中, 均是AAT最多, 分别占三碱基重复类型总数的38.52%和41.67%。在长江江豚中, AAC (18.03%)

是排名第二的三碱基重复类别,而在东亚江豚中,AAG (12.50%)是排名第二的三碱基重复类别。此外,相对于长江江豚,三碱基重复类型中排名第三的类别除CCG外,东亚江豚Y染色体基因组中还多出了一种ACC。

而对于四碱基重复类型,长江江豚与东亚江豚的优势重复类别出现较大差异。在长江江豚中占比最多的是CCTC (17.53%),而CCTC在东亚江豚中仅排名第二,占比14.05%。东亚江豚中排名第一的四碱基类别AAAT (16.76%)仅在长江江豚中排名第三。

两个亚种的五碱基重复类型中均仅含有两个重复类别,各重复类别均只有1个,无法进行排名。长江江豚Y染色体基因组中仅含1个六碱基重复类

别CCCCCA,且仅为1个。而东亚江豚含有两个,分别为ACCCTA和CCCCCA,个数均为1个(表6)。

3 讨论

3.1 长江江豚和东亚江豚染色体水平基因组微卫星分布特征

本研究以长江江豚和东亚江豚染色体水平基因组序列数据为基础,利用MISA软件对两个亚种全基因组中的完美型微卫星进行定位、统计与分类,并比较两者基因组中的微卫星分布特征差异。结果表明,长江江豚和东亚江豚基因组中分别鉴定到完美型微卫星888424和867792个,总长度分别为15340771和14724482 bp,均占全基因组的0.59%。长江江豚和东亚江豚的微卫星占比与其他哺乳动

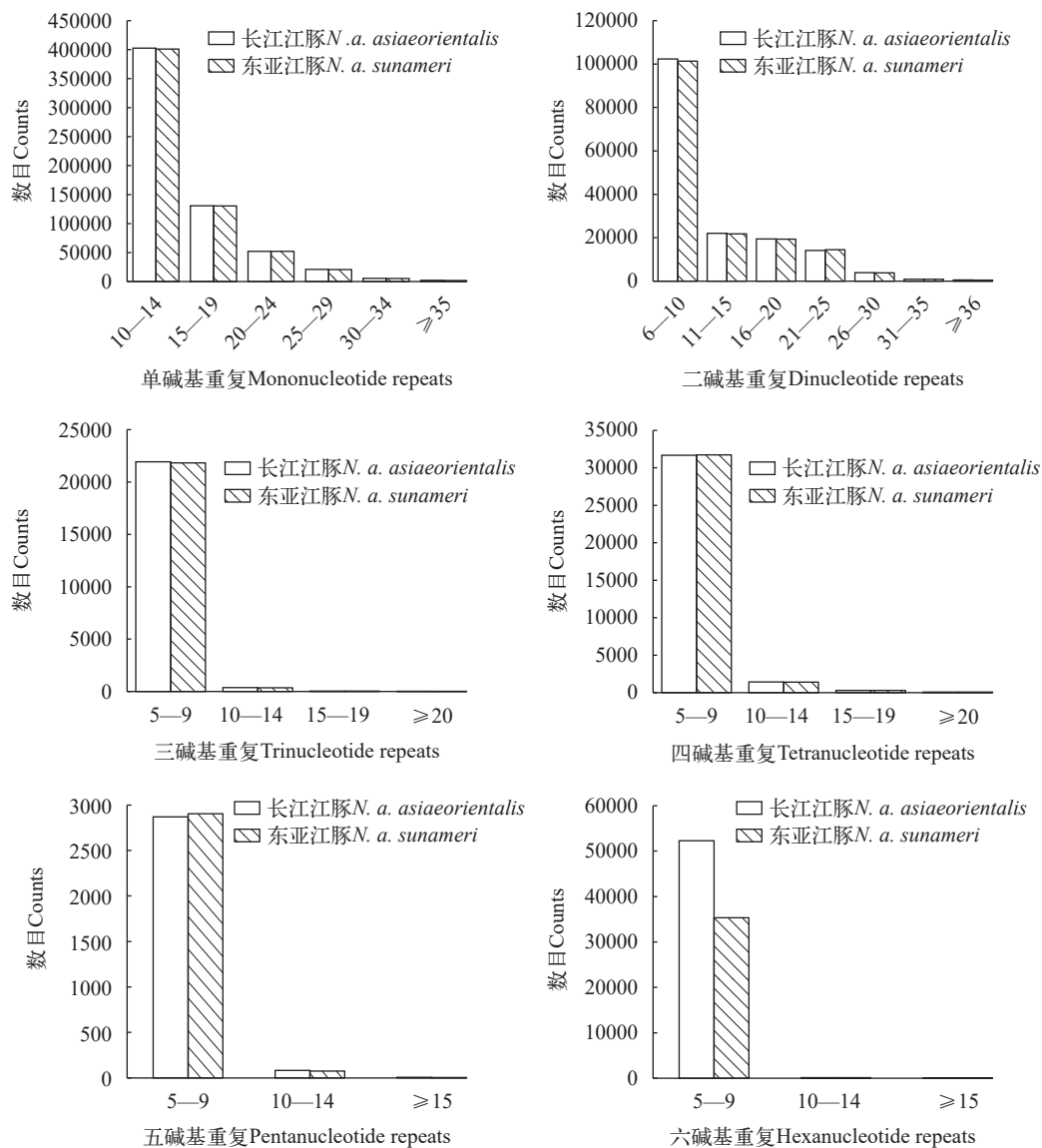


图1 长江江豚与东亚江豚基因组中6种重复类型微卫星的重复次数及分布情况

Fig. 1 Characterization of abundance and distribution patterns of six repetitive microsatellite types in genomes of the Yangtze finless porpoise and East Asian Yangtze finless porpoise

物相似, 如大熊猫*Ailuropoda melanoleuca* (0.64%)、北极熊*Ursus maritimus* (0.79%)、山羊*Capra hircus* (0.56%)、藏羚羊*Pantholops hodgsonii* (0.54%)、牦牛*Bos grunniens* (0.51%)等^[14-16], 但与人类*Homo sapiens* (3.00%)、小鼠*Mus musculus* (2.85%)、犬

Canis lupus familiaris (1.23%)等相比^[17-19], 微卫星序列在长江江豚和东亚江豚基因组中分布较少。该结果表明, 不同物种间的微卫星在基因组中的占比差异较大, 可能与物种本身的基因组大小、重复序列占基因组的比例有关, 但也不排除基因组质量、检测软件、筛选标准等造成的影响。

在长江江豚和东亚江豚基因组中, 不同微卫星重复类型丰度差异较大。长江江豚和东亚江豚基因组中均以单碱基重复类型占主要优势, 其次依次是二碱基、六碱基、四碱基、三碱基和五碱基, 与多数已发表动物基因组相似, 例如鸟类虎皮鹦鹉(*Melopsittacus undulatus*)、鱼类翘嘴鲇(*Culter alburnus*)、哺乳类高原鼠兔(*Ochotona curzoniae*)等^[20-22]。在长江江豚和东亚江豚基因组中, 大部分微卫星重复类型占总数的比例极为相近, 但长江江豚基因组中的六碱基重复类型占比(5.90%)要显著高于东亚江豚(4.09%)。这可能与本研究用于分析的长江江

表 3 长江江豚和东亚江豚基因组中出现次数最多的6种重复拷贝

Tab. 3 The six most frequent repetitive copies in genomes of the Yangtze finless porpoise and East Asian finless porpoise

长江江豚 <i>N. a. asiaeorientalis</i>		东亚江豚 <i>N. a. sunameri</i>	
重复类别	出现次数	重复类别	出现次数
A/T	585009	A/T	583422
AC	90007	AC	89835
AT	46126	AT	45834
CCCTAA	15964	C/G	27879
AGGGTT	14714	CCCTAA	10784
C/G	28743	AGGGTT	10201

表 4 长江江豚和东亚江豚基因组中6种主要重复类型排名前3的优势重复单元

Tab. 4 The top three predominant repetitive units of six major repeat motifs in genomes of the Yangtze finless porpoise and East Asian finless porpoise

重复类型Motif	长江江豚 <i>N. a. asiaeorientalis</i>			东亚江豚 <i>N. a. sunameri</i>		
单碱基 Mononucleotide	A (585009) 95.32%	C (28743) 4.68%		A (583422) 95.44%	C (27879) 4.56%	
二碱基 Dinucleotide	AC (90007) 55.08%	AT (46126) 28.23%	AG (26171) 16.02%	AC (89835) 55.36%	AT (45834) 28.24%	AG (26047) 16.05%
三碱基 Trinucleotide	AAT (6755) 30.19%	AAC (3735) 16.69%	ACC (3014) 13.47%	AAT (6748) 30.34%	AAC (3757) 16.89%	ACC (3003) 13.50%
四碱基 Tetranucleotide	AAAT (11724) 34.99%	CCTC (3485) 10.40%	AAAC (3239) 9.67%	AAAT (11631) 34.69%	CCTC (3382) 10.09%	AAAC (3352) 10.00%
五碱基 Pentanucleotide	AAAAC (1203) 40.63%	AAAAT (591) 19.96%	CCCCT (247) 8.34%	AAAAC (1212) 40.60%	AAAAT (581) 19.46%	CCCCT (233) 7.81%
六碱基 Hexanucleotide	AGGGTT (28054) 53.52%	CCCTAA (23829) 45.46%	AAAAAC (68) 0.13%	AGGGTT (20017) 56.46%	CCCTAA (14944) 42.15%	AAAAAC (78) 0.22%

表 5 长江江豚和东亚江豚Y染色体不同类型微卫星的分布特征

Tab. 5 Statistics of microsatellites with different repeat types in the Y chromosome of the Yangtze finless porpoise and East Asian finless porpoise

Y染色体中SSR重复类型 Motif in ChrY	长江江豚 <i>N. a. asiaeorientalis</i>			东亚江豚 <i>N. a. sunameri</i>		
	数目	相对丰度 (个/Mb)	占总数的比例 (%)	数目	相对丰度 (个/Mb)	占总数的比例 (%)
单碱基Mononucleotide	3663	244.20	59.13	2891	262.82	57.54
二碱基Dinucleotide	2142	142.80	34.58	1862	169.27	37.06
三碱基Trinucleotide	122	8.13	1.97	72	6.55	1.43
四碱基Tetranucleotide	251	16.73	4.05	185	16.82	3.68
五碱基Pentanucleotide	16	1.07	0.26	12	1.09	0.24
六碱基Hexanucleotide	1	0.07	0.02	2	0.18	0.04
总计Total		6195			5024	
总长度Total length (bp)		99613			81069	
相对丰度Relative frequency (loci/Mb)		413.0			456.7	
相对密度Relative density (bp/Mb)		6640.9			7369.9	
Y染色体大小ChrY size (bp)		15723222			11022723	
总长度占比Proportion (%)		0.63			0.74	

豚和东亚江豚基因组的组装方式、基因组组装长度以及gap数更少有关。研究表明,不同物种的优势微卫星重复类型存在差异。即使是与长江江豚和东亚江豚同属于海豚总科的中华白海豚(*Sousa chinensis*)的优势重复类型也存在差异,中华白海豚基因组中排名第三的优势重复类型是四碱基,然后依次是五碱基、六碱基和三碱基^[12]。在其他哺乳动物中,大熊猫和北极熊基因组中,除单碱基和二碱基重复类型外,四碱基重复类型为优势重复类型^[14];在山羊和藏羚羊中,三碱基重复类型为优势重复类型^[15]。而六碱基重复类型占优势的情况在以往对其他物种微卫星分布的研究中几乎没有发现,长江江豚和东亚江豚是目前发现六碱基重复类型占优势的两个亚种,这可能是长江江豚和东亚江豚的物种特征,也可能与长江江豚和东亚江豚基因组均达到染色体级别有关,因为利用三代测序数据能更容易组装出基因组中较为复杂的重复序列。此外,在以往发表的文章中,长江江豚除单碱基外的优势碱基类型排序为二碱基、四碱基、五碱基、六碱基和三碱基,本研究发现的长江江豚的优势碱基类型排序与以往发表的文章有较大差异^[12]。排除微卫星参数设置影响外,推测是基因组组装质量造成的差异。以往发表文章中采用的是以二代测序拼接的Scaffold级别的基因组序列片段,而本研究分析的长江江豚和东亚江豚基因组均达到染色体级别,基于三代测序数据更易组装出复杂的重复序列。要验证这一推测需要基于更多物种的染色体水平基因组来探索微卫星分布特征。

有研究表明,在所有脊椎动物中,AC是最常见的二核苷酸重复基序^[23]。本研究发现,在长江江豚和东亚江豚二碱基重复中,也以AC为基序的重复序列占比最多,分别为55.08%和55.36%;其次分别是AT (28.23%和28.24%)和AG (16.02%和16.05%),这一趋势与中华白海豚及其他一些鲸类物种,如宽

吻海豚(*Tursiops truncatus*)、虎鲸(*Orcinus orca*)、白鲸(*Delphinapterus leucas*)、白暨豚等的全基因组微卫星特征基本一致^[12]。在不同物种中,这三类主要优势碱基类别所占的比例具有一定程度的差异,例如长江江豚和东亚江豚以AC为基序的重复序列占比要显著低于中华白海豚(73.03%),而以AT为基序的重复序列占比要显著高于中华白海豚(18.57%);白暨豚基因组中以AC为基序的重复序列占比64.97%,以AT为基序的重复序列占比26.95%;小鳁鲸(*Balaenoptera acutorostrata*)基因组中以AC为基序的重复序列占比79.51%,以AT为基序的重复序列占比12.54%^[12]。

此外,在长江江豚和东亚江豚基因组中,不同微卫星类型数量随着核心序列拷贝数的增加均呈现递减趋势,这与大部分已发表的物种(如牦牛、绿尾虹雉、团头鲂、翘嘴鲃等)基因组微卫星分布特征一致^[24],验证了微卫星的稳定性会随着长度的增加而下降的结论^[25-27]。

3.2 长江江豚和东亚江豚Y染色体微卫星分布特征

基于两个染色体水平的高质量基因组,本研究还首次对长江江豚和东亚江豚Y染色体中微卫星分布特征进行了比较分析,两个亚种Y染色体中分别鉴定到6195和5024个完美型微卫星。本研究发现,在Y染色体的六种重复类型中,单碱基重复类型仍占主要优势(分别占有所有重复类型总数的59.13%和57.54%),后续依次为二碱基、四碱基、三碱基、五碱基和六碱基重复类型。然而,在长江江豚和东亚江豚全基因组中除单碱基和二碱基类型外,六碱基重复类型占主要优势,说明基因组中六碱基类型微卫星主要分布在Y染色体以外的染色体上。此外,在其他哺乳动物,例如恒河猴、山羊、食蟹猴基因组中^[28],全基因组优势微卫星类型排序与各染色体基因组优势微卫星类型排序也存在部分差异,说明不同微卫星类型的分布规律在不同染色体上

表6 长江江豚和东亚江豚Y染色体6种主要重复类型排名前3的优势重复类别

Tab. 6 The top three predominant repetitive units of six major repeat motifs in the Y Chromosome of the Yangtze finless porpoise and East Asian finless porpoise

重复类型Motif	长江江豚 <i>N. a. asiaorientalis</i>			东亚江豚 <i>N. a. sunameri</i>		
单碱基 Mononucleotide	A (3432) 93.69%	C (231) 6.31%		A (2706) 93.60%	C (185) 6.40%	
二碱基 Dinucleotide	AC (1615) 75.40%	AT (333) 15.55%	AG (191) 8.92%	AC (1460) 78.41%	AT (265) 14.23%	AG (134) 7.20%
三碱基 Trinucleotide	AAT (47) 38.52%	AAC (22) 18.03%	CCG (12) 9.84%	AAT (30) 41.67%	AAG (9) 12.50%	ACC/CCG (8) 11.11%
四碱基 Tetranucleotide	CCTC (44) 17.53%	TTCT (38) 15.14%	AAAT (35) 13.94%	AAAT (31) 16.76%	CCTC (26) 14.05%	TCTA (25) 13.51%
五碱基 Pentanucleotide	AAAAC (1) 6.25%	TTTAT (1) 6.25%		AGAGG (1) 8.33%	AGGGG (1) 8.33%	
六碱基 Hexanucleotide	CCCCCA (1) 100%			ACCCTA (1) 50%	CCCCCA (1) 50%	

并不一致。

此外, 本研究发现长江江豚和东亚江豚Y染色体中不同重复类型中的优势类别存在较大差异。长江江豚Y染色体三碱基重复类型中的优势类别为AAT (38.52%)、AAC (18.03%)、CCG (9.84%), 而东亚江豚Y染色体三碱基重复类型中的优势类别为AAT(41.67%)、AAG(12.50%)、ACC/CCG(11.11%)。此外, 长江江豚Y染色体上六碱基重复序列仅一个CCCCCA, 而东亚江豚Y染色体上存在2个六碱基重复序列, 分别为ACCCTA和CCCCCA; 该结果意味着今后可能开发基于六碱基重复的微卫星标记应用于亚种鉴定。对2个亚种基因组中Y染色体上的六碱基重复序列进行注释, 发现这些序列并不位于外显子中。对位于Y染色体的三碱基重复序列进行注释, 发现长江江豚基因组中有一个以CCG为重复单元的微卫星序列位于外显子中, 并且该外显子对应的基因所编码的蛋白家族TBL1 (Transducin Beta Like 1)与遗传性听力缺陷、致癌信号的调节等有关^[29, 30]。东亚江豚基因组中有3个以CCG为重复单元的微卫星位于外显子中, 2个以CCG为重复单元的微卫星位于5' UTR区域。研究表明, (CCG)_n、(GAA)_n等三碱基重复类型微卫星可以通过环-折叠结构形成复杂的DNA结构, 这些结构对基因表达起到调控作用^[31]。推测长江江豚与东亚江豚三碱基微卫星类别差异及六碱基微卫星优势类别的组成差异可能与这两个江豚亚种某些基因调控功能有关, 然而该猜想需要后续进一步对编码序列进行深入分析来验证。

由于Y染色体具有较高的遗传稳定性, 并且具有父系遗传特性。因此, Y染色体微卫星成了研究雄性谱系和群体遗传学的重要工具, 广泛应用于性别鉴定、遗传谱系构建、种群结构及社群行为研究等领域。随着高通量测序技术的快速发展, 2023年人类的Y染色体被成功破译^[32]。然而, 鲸类Y染色体测序及序列分析相关研究鲜见报道。2022年, Ming等^[12]首次对中华白海豚的基因组进行了微卫星分布分析和标记开发, 并利用已发表的宽吻海豚未定位的Y染色体基因组框架序列, 从白海豚基因组中比对并筛选出1349个雄性白海豚候选微卫星标记, 这些候选标记经过更多样品测试之后有望应用于性别鉴定等研究。基于染色体水平基因组, 本研究分析并比较了长江江豚和东亚江豚的Y染色体微卫星分布特征及其差异, 为后续物种特异性Y染色体微卫星标记的开发与应用奠定了重要基础。

微卫星是重要的基因组特征, 能够为物种鉴

定、遗传多样性评估及比较和保护基因组等研究提供宝贵资源。基于长江江豚和东亚江豚的染色体水平基因组, 本研究首次对窄脊江豚两个亚种的全基因组完美型微卫星种类、数量、丰度等特征进行分析和比较, 并且着重分析了两者Y染色体微卫星组成特征和差异。该研究有助于了解长江江豚和东亚江豚这两个亚种在基因组构成方面的差异, 同时可为后续大批量物种特异性和多态性微卫星遗传标记的开发奠定基础。

(作者声明本文符合出版伦理要求)

参考文献:

- [1] Turvey S T, Pitman R L, Taylor B L, *et al.* First human-caused extinction of a cetacean species [J]? *Biology Letters*, 2007, **3**(5): 537-540.
- [2] Zhao X, Barlow J, Taylor B L, *et al.* Abundance and conservation status of the Yangtze finless porpoise in the Yangtze River, China [J]. *Biological Conservation*, 2008, **141**(12): 3006-3018.
- [3] Wang D, Turvey S T, Zhao X, *et al.* *Neophocaena asiaeorientalis* ssp. *asiaeorientalis* [R]. IUCN Red List of Threatened Species Version 2013.1.
- [4] Hao Y J, Tang B, Mei Z G, *et al.* Further suggestions on conservation of the Yangtze finless porpoise based on retrospective analysis of the current progress [J]. *Acta Hydrobiologica Sinica*, 2024, **48**(6): 1065-1072. [郝玉江, 唐斌, 梅志刚, 等. 长江江豚保护进展的回顾性分析及进一步保护建议 [J]. 水生生物学报, 2024, **48**(6): 1065-1072.]
- [5] Cheng Z L, Li Y T, Zuo T, *et al.* Threats and conservation strategies of the East Asian finless porpoises in China [J]. *Journal of Applied Oceanography*, 2024, **43**(3): 597-606. [程兆龙, 李永涛, 左涛, 等. 我国东亚江豚的研究现状、面临的威胁与保护建议 [J]. 应用海洋学学报, 2024, **43**(3): 597-606.]
- [6] Yang G, Ren W, Zhou K, *et al.* Population genetic structure of finless porpoises, *Neophocaena phocaenoides*, in Chinese waters [J]. *Marine Mammal Science*, 2002, **18**(2): 336-347.
- [7] Zheng J S, Xia J H, He S P, *et al.* Population genetic structure of the Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*): implications for management and conservation [J]. *Biochemical Genetics*, 2005, **43**(5/6): 307-320.
- [8] Chen L, Bruford M W, Xu S, *et al.* Microsatellite variation and significant population genetic structure of endangered finless porpoises (*Neophocaena phocaenoides*) in Chinese coastal waters and the Yangtze River [J]. *Marine Biology*, 2010, **157**(7): 1453-1462.
- [9] Li S, Xu S, Wan H, *et al.* Genome-wide SNP and population divergence of finless porpoises [J]. *Genome Biology*

- and Evolution*, 2013, **5**(4): 758-768.
- [10] Zhou X, Guang X, Sun D, *et al.* Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater [J]. *Nature Communications*, 2018, **9**(1): 1276.
- [11] Committee on Taxonomy. 2024. List of marine mammal species and subspecies. Society for Marine Mammalogy, www.marinemammalscience.org, consulted on 10th September 2024.
- [12] Ming Y, Yu X, Liu W, *et al.* The landscape of genome-wide and gender-specific microsatellites in Indo-Pacific humpback dolphin and potential applications in cetacean resource investigation [J]. *Journal of Marine Science and Engineering*, 2022, **10**(6): 834.
- [13] Yin D, Chen C, Lin D, *et al.* Gapless genome assembly of East Asian finless porpoise [J]. *Scientific Data*, 2022(9): 765.
- [14] 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. [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.]
- [15] Qi W H, Yan C C, Xiao G S, *et al.* Distribution regularities and bioinformatics analysis of microsatellite in the whole genomes of goat and Tibetan antelope [J]. *Journal of Sichuan University (Natural Science Edition)*, 2016, **53**(4): 937-944. [戚文华, 严超超, 肖国生, 等. 山羊和藏羚羊全基因组微卫星分布规律及其生物信息学分析 [J]. 四川大学学报(自然科学版), 2016, **53**(4): 937-944.]
- [16] Xu Y H, Ma Z J, Guo W X, *et al.* Analysis on the distribution features of perfect microsatellite in the whole genome of yak [J]. *Journal of China Agricultural University*, 2024, **29**(3): 134-142. [徐宇辉, 马志杰, 郭卫兴, 等. 牦牛全基因组单纯型微卫星分布特征分析 [J]. 中国农业大学学报, 2024, **29**(3): 134-142.]
- [17] 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.
- [18] Tong X L, Dai F Y, Li B, *et al.* Microsatellite repeats in mouse: Abundance distribution and density [J]. *Current Zoology*, 2006, **52**(1): 138-152. [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.]
- [19] Wang Y Y, Liu X X, Dong K Z, *et al.* Distribution difference of microsatellite in 7 domestic animals genomes [J]. *China Animal Husbandry & Veterinary Medicine*, 2015, **42**(9): 2418-2426. [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.]
- [20] Huang J, Yuan B D, Yang C Z. Distribution of microsatellites in genome of Budgerigar (*Melopsittacus undulatus*) [J]. *Chinese Journal of Wildlife*, 2017, **38**(3): 422-426. [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.]
- [21] Liu S L, Liu Y N, Li F, *et al.* Distribution characteristics of microsatellites in the whole genome of three Culterinae species [J]. *Journal of Shanghai Ocean University*, 2023, **32**(6): 1176-1183. [刘士力, 刘一诺, 李飞, 等. 3种鲃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.]
- [22] He Y J, Hu W J, Du Y R, *et al.* Development of microsatellite primers of *Ochotona curzoniae* with RAD-seq data [J]. *Chinese Journal of Wildlife*, 2021, **42**(3): 730-736. [贺玉姣, 胡万军, 都玉蓉, 等. 基于高原鼠兔简化基因组数据的微卫星引物开发 [J]. 野生动物学报, 2021, **42**(3): 730-736.]
- [23] Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis [J]. *Genome Research*, 2000, **10**(7): 967-981.
- [24] Cui K, Yue B S. Distribution patterns of microsatellites in the genome of *Lophophorus lhuysii* [J]. *Sichuan Journal of Zoology*, 2018, **37**(5): 533-540. [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.]
- [25] 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.
- [26] Ellegren H. Heterogeneous mutation processes in human microsatellite DNA sequences [J]. *Nature Genetics*, 2000, **24**(4): 400-402.
- [27] Leopoldino A M, Pena S D J. The mutational spectrum of human autosomal tetranucleotide microsatellites [J]. *Human Mutation*, 2003, **21**(1): 71-79.
- [28] Webster M T, Smith N G C, 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] Di S M, Collesi C, Vozzi D, *et al.* *TBLIY*: a new gene involved in syndromic hearing loss [J]. *European Journal of Human Genetics*, 2019, **27**(3): 466-474.
- [30] Pray B A, Youssef Y, Alinari L. *TBLIX*: At the crossroads of transcriptional and posttranscriptional regulation [J]. *Experimental Hematology*, 2022(116): 18-25.
- [31] Fabregat I, Koch K S, Aoki T, *et al.* Functional pleiotropy of an intramolecular triplex-forming fragment from the 3'-UTR of the rat Pigr gene [J]. *Physiological Genomics*, 2001, **5**(2): 53-65.
- [32] Hallast P, Ebert P, Loftus M, *et al.* Assembly of 43 human Y chromosomes reveals extensive complexity and variation [J]. *Nature*, 2023, **621**(7978): 355-364.

MICROSATELLITES DISTRIBUTION CHARACTERIZATION IN THE CHROMOSOME-LEVEL GENOMES OF THE YANGTZE FINLESS PORPOISE AND THE EAST ASIAN FINLESS PORPOISE

LI Han-Shu^{1,2}, YANG Lian-Dong¹, MENG Ming-Hui³, XU Pan³, FAN Fei^{1,4}, WANG Ding^{1,4} and ZHENG Jin-Song^{1,4}

(1. Key Laboratory of Aquatic Biodiversity and Conservation of the Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China; 2. University of Chinese Academy of Sciences, Beijing 100049; 3. Diggers (Wuhan) Biotechnology Co., Ltd, Wuhan 430070, China; 4. National Aquatic Biological Resource Center, Wuhan 430072, China)

Abstract: In this study, bioinformatic methods were used to analyze and compare the genome-wide characterization of microsatellites in the chromosome-level genomes (2.62 and 2.50 Gb respectively) of the Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*; YFP) and the East Asian finless porpoise (*N. a. sunameri*; EAFP). The results showed that a total of 888424 and 867792 perfect microsatellites were identified in the whole genomes of the YFP and the EAFP, with a total length of 1534077 and 1472448 bp respectively, accounting for about 0.59% of the whole genome sequences. In both the YFP and EAFP genomes, the numbers of the six major repeat types of perfect microsatellites were similar, with mononucleotide repeats being the most abundant, followed by dinucleotide, hexanucleotide, tetranucleotide, trinucleotide, and pentanucleotide repeats. The most abundant repeat motifs in the two genomes were A, AC, AT, CCCTAA, and AGGGTT. Additionally, we also analyzed the microsatellite distribution characteristics in the Y chromosome of the YFP and EAFP. In total, 6195 and 5024 perfect microsatellites were detected in the Y chromosomes respectively. Among the six repeat types of perfect microsatellites on the Y chromosome, mononucleotide repeat was also the most abundant type, followed by dinucleotide, tetranucleotide, trinucleotide, pentanucleotide, and hexanucleotide. In addition, A, AC, and AAT repeat motif categories were the most dominant. In this study, we examined the distribution characteristics and disparities of microsatellites in the chromosomal-level genomes of the YFP and the EAFP, thereby establishing a solid foundation for future research on the development and utilization of species specific microsatellite markers.

Key words: the Yangtze finless porpoise; the East Asian finless porpoise; Genome; Y chromosome; Microsatellite