



大熊猫粪便中线粒体基因库特点研究

李晓达^{1#}, 任晓彤^{1,2#}, 高雨薇¹, 王文慧¹,
张琦¹, 褚佳宁¹, 林立钧³, 徐艳春¹, 杨淑慧^{1*}

1. 东北林业大学野生动物与自然保护地学院, 哈尔滨, 150040;
2. 菏泽市疾病预防控制中心, 菏泽, 274000;
3. 东北林业大学奥林学院, 哈尔滨, 150040)

稿件运行过程

收稿日期: 2024-03-22

修回日期: 2024-05-13



关键词: 大熊猫;

粪便 DNA;

异质性;

线粒体基因库;

年龄

Keywords: Giant panda (*Ailuropoda*

melanoleuca);

Fecal DNA;

Heteroplasmy;

Mitogene pool;

Age

中图分类号: Q34

文献标志码: A

文章编号:

2310-1490(2025)-02-0269-12

DOI: 10.12375/ysdwxb.20250204

摘要

为了解大熊猫(*Ailuropoda melanoleuca*)肠道脱落细胞中线粒体 DNA(mtDNA)的异质性特点及其与年龄的关系,以 92 只不同年龄的大熊猫粪便为材料,通过扩增子高通量测序研究非编码的 D-loop 区、编码 rRNA 的 16S rRNA 区以及编码蛋白质的 ND1 和 ND2 区的异质性。结果显示:D-loop 区存在 3 个异质性热点区段,突变碱基的数量远远高于其他 3 个编码区;D-loop 区 T_2/T_1 小于 16S rRNA 和 ND 两个编码区,表明后两者较 D-loop 区受到更强烈的纯化选择;ND 区第 3 位密码子的突变频率显著高于前两位,选择压力最低。此外,16SA3、ND1A6 和 ND2A7 三个编码区的异质性可能与年龄相关。上述结果表明,大熊猫粪便 DNA 的异质性突变构成了一个多样性较高的线粒体基因库,突变在组织细胞水平上受到自然选择的作用。该线粒体基因库为揭示动物适应性进化机制、个体对环境的适应方式、个体适合度和动物与肠道微生物及环境的关系等提供一个全新的角度,也可能成为有潜力的大熊猫个体年龄预测标记。

Study on the Characterization of the Fecal Mitogene Pool of Giant Panda

基金项目: 国家自然科学基金项目(32170517); 大熊猫国家公园珍稀动物保护生物学国家林业和草原局重点实验室开放课题项目(KLSFGAGP2020.017)

第一作者简介: 李晓达(1999—),男,硕士研究生;主要从事保护遗传学研究。E-mail: 925194078@qq.com

* 共同第一作者: 李晓达; 任晓彤

* 通信作者: 杨淑慧, E-mail: 4745682512@qq.com

LI Xiaoda^{1#}, REN Xiaotong^{1,2#}, GAO Yuwei¹, WANG Wenhui¹,
ZHANG Qi¹, CHU Jianing¹, LIN Lijun³, XU Yanchun¹, YANG Shuhui^{1*}

(1. College of Wildlife and Protected Areas, Northeast Forestry University, Harbin, 150040, China;

2. Heze Center for Disease Control and Prevention, Heze, 274000, China;

3. Olin College, Northeast Forestry University, Harbin, 150040, China)

Abstract: In order to obtain the heteroplasmy characteristics of mitochondrial DNA of intestinal exfoliated cells in the feces and its relationship with age in giant panda (*Ailuropoda melanoleuca*), fecal samples of 92 giant pandas were collected, and analyzed heteroplasmy of partial D-loop, 16S rRNA and ND region, which represent the non-coding, rRNA coding and protein coding region by using amplicon massive parallel sequencing. The results showed that heteroplasmic mutations took place across all regions studied, where D-loop carried the greatest number of heteroplasmic sites in three hot regions comparing to the other three regions. T_g/T_v ratio was smaller in D-loop than in 16S rRNA and ND region, indicating the later regions were under stronger purifying selection than D-loop. The mutation frequency of the third codon in ND region was significantly higher than the first two codon, and the selection pressure was the lowest. Meanwhile, the heterogeneity of the three coding regions 16SA3, ND1A6 and ND2A7 may be related to age. These results suggest the fecal DNA of giant panda contain a mitogene pool with high diversity, and heteroplasmic mutations are selected at cell and tissue level. The mitogene pool may provide novel insights into the adaptive evolution, individual fitness, response fashion to environments and interactions between animals and gut microbiota, and even become a potential novel marker to prediction age.

线粒体是真核细胞内重要的细胞器之一,由原始真核细胞内共生的好氧菌进化而来,对细胞内能量代谢、Ca²⁺缓冲和细胞的程序性死亡等具有至关重要的作用^[1]。它是一个由内外双层膜围成的细胞器,拥有独立于核基因组的环状基因组,即线粒体基因组(mitochondrial genome, mtGenome),也称线粒体DNA(mitochondrial DNA, mtDNA)^[2-3]。在哺乳动物中,mtDNA长约16.5 kb,包括1个控制区(D-loop区)和37个基因,编码22种tRNAs、2种rRNA和13个电子传递链的蛋白亚基和ATP合成酶的2个亚基。虽然这些基因编码的蛋白质大约只占线粒体蛋白质总量的1%,但对线粒体氧化磷酸化和细胞能量转化是必不可少的^[4]。

每个细胞中线粒体的数量与其代谢水平密切相关,代谢水平越高,线粒体数越多^[5];而每个线粒体中也含有多个mtDNA拷贝,且拷贝数也与细胞状态有关^[6]。因而,每个细胞中都含有大量的mtDNA拷贝。随着损坏的线粒体被不断清除,新线粒体被不断复制出来,mtDNA也在不断增殖和更新。在这一过程中,合成新mtDNA的DNA聚合酶是Pol γ ,由于

该酶缺乏纠错能力,在链复制过程中经常出错,从而易产生新的突变(*de novo mutation*)^[7]。同时,在细胞代谢过程中,细胞自身产生的氧化物质(ROS等)和从外界摄入的基因毒素(genotoxins)会损伤mtDNA,而受损mtDNA的修复也由Pol γ 来完成,修复过程也会产生新突变^[8-9]。因而,一个细胞、组织、器官和个体的mtDNA都包含大量突变体,这些突变体被称为mtDNA的异质性(heteroplasmy)^[10-15],它们的总和可看作一个线粒体基因库(mitogene pool),其组成、数量会对机体的适合度产生显著影响,具有适应意义^[16]。以往的研究通过利用mtDNA独特的进化方式进行系统发育关系的重建^[17]、遗传多样性的评估^[18]、种群遗传结构和种群间遗传关系的分析、物种及杂交个体的鉴定^[19-20]等,但关于线粒体基因库的研究则十分少见。

大熊猫(*Ailuropoda melanoleuca*)是举世瞩目的明星物种,经过半个多世纪的努力,种群得到显著恢复,已从濒危(EN)转为易危(VU)^[21]。但是,围绕大熊猫这一生物学特性非常明显的物种开展的科学研究继续向着纵深发展,如组装了大熊猫高质量的参

考基因组^[22-23],揭示了其演化与亚种分化历史^[24]、低代谢率的适应机制^[25]、特殊食性的进化机制^[26-28],以及代谢组学和脂肪组学特性所反映的适应机制等^[29-31]。线粒体基因库又提供了一个新的角度来探究大熊猫生理、发育、衰老和适应背后的基因组学机理,为进一步保护大熊猫提供新的科学依据。

大熊猫粪便是最易得的材料之一,可通过粪便获得大熊猫肠道上皮脱落的细胞DNA^[32]。这些肠道上皮细胞主要来自大肠^[33-34],包括大肠上皮中吸收性上皮细胞、分泌性杯状细胞和肠内分泌细胞等^[35]。肠上皮细胞的更新速度极快,3~5 d就会完全更新一次^[36]。海量的细胞增殖也伴随着线粒体及其基因组的海量增殖,所产生的突变均累积在子代上皮细胞中^[37]。这些子代细胞完成使命后脱落到肠腔,随粪便排出体外。有研究表明,小鼠每24 h脱落的肠上皮细胞数可达 2×10^{11} 个^[38]。所以,利用粪便中数量庞大的肠道上皮细胞,足以精准检测和研究mtDNA的异质性。

本研究利用92只大熊猫的粪便DNA,通过PCR和扩增子高通量测序方法,研究其肠道上皮组织中mtDNA D-loop区异质性的空间分布特点,初步揭示这一区段突变和选择的特点,为深化对线粒体基因库的认识提供资料。

1 材料与方法

1.1 样本采集

从中国大熊猫保护研究中心都江堰、核桃坪和耿达3个基地内采集92只大熊猫的粪便样本,大熊猫为22~357月龄,包括雌性57只,雄性35只。粪便在排出后1~3 h内收集,装在自封口样品袋中,于冰箱中-20℃冷冻保存,然后放于干冰中冷链运输到实验室,继续在冰箱中-20℃保存。

1.2 粪便DNA的提取

用无菌手术刀或剪刀,从每个粪便样本前端、中段和后端3个不同部位切取表层样品,用CP513型电子天平(OHAUS,美国)称取180~220 mg放于2 mL无菌离心管内。粪便总DNA的提取采用本团队创建的方法(专利号:ZL 202010037073.3),主要步骤包括:(1)向装有粪便样本的2 mL离心管中加入500 μ L 10 mmol/L的PBS溶液,使样本充分浸泡约10 min,随后加入125 μ L 5%的SDS,震荡使粪便

与溶液快速混合均匀,瞬时离心后于室温下静置20 min,4 000 r/min离心10 min,吸取全部上清液到新的1.5 mL离心管内。(2)用AxyPrep Genomic DNA Mini Kit(Axygen,美国)从上清液中提取总DNA,完全按照试剂盒说明书操作。提取后的DNA用Nano-Drop 2000超微量紫外分光光度计(Thermal Fisher Scientific,美国)测定DNA浓度。

1.3 mtDNA异质性分析

1.3.1 大熊猫线粒体基因组上部分区段的PCR扩增

课题组在此前的研究(尚未发表)中发现,在线粒体基因组上,D-loop、16S rRNA、ND1和ND2四个区域的异质性频率明显高于其他区域,且在不同个体间能够表现出明显的差异性,因此本研究针对这4区在大熊猫物种上进行具体研究。4个区域分别代表非编码区(D-loop区)、rRNA编码区(16S rRNA区)和蛋白质编码区(ND1和ND2区),设计PCR引物分别扩增5个片段(表1)。把引物序列通过NCBI-BLAST排除扩增线粒体DNA核拷贝(Numt)的可能。引物由吉林省库美生物科技有限公司合成。每个片段的扩增均在50 μ L体系中进行,包括:Phanta Max Super-Fidelity DNA Polymerase(南京诺唯赞生物科技有限公司)1 μ L、2 \times Phanta Max Buffer 25 μ L、dNTP Mix (10^{-3} mol/L each)1 μ L、上下游引物(10^{-6} mol/L)各1 μ L、DNA模板4 μ L(~80 ng总DNA)及去离子水17 μ L。用9700型PCR扩增仪(Applied Biosystems,美国)扩增,程序:95℃预变性3 min;95℃变性15 s, T_m ℃退火15 s,72℃延伸30 s,30个循环;循环结束后,72℃延伸5 min。扩增结束后,取4 μ L扩增产物在1.5%琼脂糖凝胶上100 V下电泳50 min,SYBR Safe染料染色(Thermal Fisher, USA),以DL2000(TaKaRa,大连)为分子质量marker,在紫外光下观察扩增产物的大小。

此外,mtDNA的异质性主要是单个拷贝水平的变异分析,因为高通量测序会引入一定的误差^[39],因此,需要对这一误差进行估算并将其从序列中消除才能保证异质性数据的真实性。因此,以大熊猫mtDNA基因组为模板设计3对引物:MG3-2、MG4-1和MG5(表1),分别扩增4 434、7 283、6 062 bp片段。扩增条件与前述5对引物相同,但每个循环的延伸时间为4 min。

表1 大熊猫线粒体基因组中8个片段的PCR引物

Table 1 PCR primers amplifying eight fragments of the mitogenome of the giant panda

引物名称 Primer name	扩增区段 Target region	引物序列(5'→3') Primer sequence(5'→3')	片段长度 /bp Fragment length	退火温度 /°C Annealing temperature
D-loopA1	D-loop	F-TCCACCTCTCATTTCATTCACTTCA R-GGAGCGAGAAGAGGTACACG	330	57.0
D-loopA2	D-loop	F-ACTATAGATCACCCGGACTGG R-GTTTGGCAGGACATCCATAAGT	450	59.0
16SA3	16S rRNA	F-CAGCAACGGATAACCACTGATAGTT R-TTAACTGATGTCACCCGGGCA	220	59.0
ND1A6	ND1	F-CAGCCCCTACATACCAGAATTATAC R-TGCTTGCTGTAATAATAGGTAGAGT	200	57.0
ND2A7	ND2	F-GCCATTACAGCCCTACTAAACC R-ATTTTCACTTACTTGAGGCTTTGA	250	57.0
MG3-2	16S rRNA—ATP8	F-TACACGTAGGGCGAGGCTTA R-CCCTCGTGTGGCCATTTCATA	4 434	57.5
MG4-1	COX2— <i>Cyt b</i>	F-ACAATAGACGCACAAGAAGTGG R-TATAAGCCTCGCCCTACGTG	7 283	57.5
MG5	<i>Cyt b</i> —16S rRNA	F-AACCCCGCCTGTTTACAAA R-GGGGTACTAAAAGAGGCGAACA	6 062	57.5

1.3.2 PCR产物的高通量测序

对每个个体D-loop、16S rRNA、ND1和ND2四个区域的扩增产物切胶回收(Axygen, USA)。测序文库构建采用MGIEasy酶切DNA文库制备试剂套装(华大智造, 中国)。用文库制备试剂盒中的末端修复组分(ERAT Buffer和ERAT Enzyme Mix)对DNA进行末端修复和添加dA尾。用文库制备试剂盒中的Ligation Buffer和DNA Ligase在DNA片段两段连上接头, 将连接产物进行磁珠纯化回收。将回收产物用PCR Enzyme Mix和PCR Primer Mix进行扩增, 反应体系为50 μ L, 按照推荐的反应程序扩增8个循环。然后对PCR产物进行再度纯化、定量, 取1.0 pmol PCR产物95 °C变性3 min, 立即置于冰上, 用MGIEasy环化模块中的Splint Buffer和DNA Rapid Ligase在37 °C条件下对单链环化30 min, 将环化产物用Digestion Buffer和Digestion Enzyme进行酶切30 min。将环化产物进行磁珠纯化回收, 用Qubit® ssDNA Assay Kit对酶切消化纯化后的产物定量, 当酶切消化产物产量/PCR投入量(ssDNA/dsDNA) \geq 7%时, 视为合格文库。通过质检的文库由MGI Seq2000测序平台(华大智造, 中国)进行PE250测序。下机数据经过质检, 达到 $Q_{30} > 85$ 后, 通过Geneious Prime v2020.1.2软件(<https://www.geneious.com>)去除接头污染和N碱基比例 $> 5\%$ 的reads, 将得到clean数据去掉接头, 并将同一拷贝的双端测序拼接为一条read, 保存备用。

1.3.3 线粒体序列测序误差的估算

对引物MG3-2、MG4-1和MG5的扩增产物切胶回收(Axygen, USA)后, 用pMD18-T载体进行连接、克隆(TaKaRa, 大连)。克隆经过1.3.1中同样体系的PCR鉴定后, 用ABI 3730x1测序仪(Thermal Fisher Scientific, USA)进行Sanger测序, 同时采用1.3.2中的方法在MGI Seq2000测序平台(华大智造, 中国)进行高通量测序。所不同的是, MG3-2、MG4-1和MG5的扩增产物采用MGIEasy酶切DNA文库制备试剂盒中的酶切组分(Frag Buffer II和Frag Enzyme II)将DNA打碎, 用MGIEasy DNA纯化磁珠试剂盒进行纯化回收, 在得到主带约为330 bp的产物后, 再进行后续的建库和测序。

用Geneious Prime v2020.1.2软件把高通量测序数据进行质控处理后, 生成一个一致性序列(consensus sequence)。将同一个体经Sanger和高通量测序所获得的序列, 用DNASTar Lasergene软件包(DNASTAR Inc. USA)中的MegAlign软件与大熊猫线粒体基因组(GenBank accession No. NC_009492.1)序列进行比对, 确认扩增的正确性; 同时, 比较二者的碱基差异, 观察一致性(identity)。使用Geneious Prime v2020.1.2软件, 以Sanger序列为参照, 以Variant Frequency $> 0.1\%$ 、覆盖度 ≥ 2000 的标准筛选多态性碱基位点(即SNP), 将得到的数据导入到Excel 2006(Microsoft, USA)。因Sanger测序和高通量测序为同一个克隆的DNA片段, 可以认为高通量

测序中所有 SNP 均来自测序错误。由此,计算所有 SNP 频率及其 95% 的置信区间,并以上限作为可信异质性突变的过滤阈值。

1.3.4 mtDNA 异质性的分析

将所有个体的高通量测序数据导入 Geneious Prime v2020.1.2 软件,筛除 Average Quality < 30、Coverage < 2 000 的数据行,再对每个个体生成一个一致性序列作为参照,从高通量测序数据中比对、筛查 SNP。在 1.3.3 中得到的高通量测序误差 95% 的置信区间为 [0.002%, 0.207%],参考其上限值筛除 Variant Frequency ≤ 0.2% 的位点,从而获得每个个体正式的异质性数据。然后,对每个 SNP 识别替换

类型,即转换 (T_S) 和颠换 (T_V),计算每种类型碱基的频率。通过皮尔逊相关分析法及 Wilcoxon 秩和检验来对比 D-loopA1、D-loopA2、16S rRNA、ND1 和 ND2 五个区域整体的异质性碱基总数,观察其中转换和颠换的数量有无相关关系。

2 结果

2.1 粪便 mtDNA 异质性位点分布

从 92 只大熊猫粪便中均成功提取了总 DNA,表 1 中的 5 个短片段和 3 个长片段引物的扩增产物在琼脂糖凝胶上均显示为单一而明亮的条带,大小符合设计预期(图 1)。

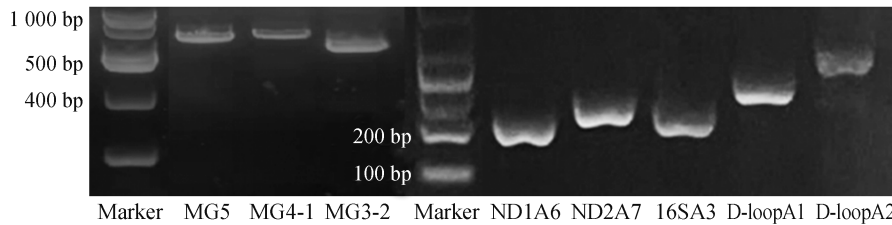


图 1 大熊猫线粒体基因组片段引物的扩增产物电泳图

Figure 1 Electrophorogram of amplification products of the primers for the mitochondrial genome fragment of the giant panda

高通量测序与 Sanger 测序的比较表明, MG3-2、MG4-1 和 MG5 扩增产物出现变异位点频率的 95% 置信区间为 [0.002%, 0.207%],表明高通量测序引入误差的最高限为 0.207%,即低于这一界限的可被认为是测序误差,因此,将 0.200% 作为过滤测序误差的标准。将每个位点上的 SNP 数量与碱基覆盖度进行回归分析,发现两者之间无显著相关性 ($P = 0.441$),表明数据筛选条件可靠。

从整个群体的线粒体基因组来看,非编码区的

D-loopA1 和 D-loopA2 区段每个个体平均异质性位点数分别为 (149.0 ± 43.1) 个和 (92.9 ± 60.0) 个,远高于其他 3 个编码区段 $[(3.2 \pm 3.8) \sim (5.2 \pm 7.2)$ 个]。D-loopA1 和 D-loopA2 区段每个个体平均异质性频率分别为 (0.015 ± 0.006) 和 (0.071 ± 0.030) ,也高于其他 3 个区段(表 2)。D-loopA1 的异质性位点分布于第 50 ~ 330 位碱基,而 D-loopA2 明显分为 2 个热点区段,一是第 700 ~ 850 位碱基,二是第 860 ~ 1 100 位碱基位,前者异质性频率较低,后者频率较高(图 2)。

表 2 大熊猫线粒体基因组 5 个区段的个体异质性特征

Table 2 Characteristics of heteroplasmy (HP) in five regions of giant panda mitogenome

区段 Regions	基因组范围 Range on mitogenome	异质性位点 HP site				转换 (T_S)、颠换 (T_V) 频率 Frequency of T_S and T_V		
		最大值/个 Maximum	最小值/个 Minimum	位点数量/个 No. of sites	频率 Mean frequency	T_S 平均频率 Mean frequency of T_S	T_V 平均频率 Mean frequency of T_V	T_S/T_V
D-loopA1	11 th ~ 340 th	210	18	149.0 ± 43.1	0.015 ± 0.006	0.013 ± 0.005	0.021 ± 0.009	0.706 ± 0.098
D-loopA2	669 th ~ 1 118 th	273	7	92.9 ± 60.0	0.071 ± 0.030	0.060 ± 0.024	0.058 ± 0.031	0.979 ± 0.277
16SA3	3 142 th ~ 3 361 st	23	1	5.2 ± 3.7	0.005 ± 0.003	0.005 ± 0.003	0.006 ± 0.005	3.638 ± 4.288
ND1A6	4 856 th ~ 5 055 th	23	0	3.2 ± 3.8	0.008 ± 0.007	0.005 ± 0.005	0.005 ± 0.005	2.819 ± 2.300
ND2A7	6 127 th ~ 6 376 th	35	0	5.2 ± 7.2	0.005 ± 0.003	0.006 ± 0.004	0.009 ± 0.005	0.962 ± 0.977
整体 Overall		420	29	273.8 ± 70.7	0.042 ± 0.014	0.030 ± 0.013	0.031 ± 0.010	0.844 ± 0.123

注:表中除个体异质性位点数的最大值外,其余数据为平均值 ± 标准差。

Note: Data in the table are in form of mean ± SD except for the maximum and minimum values.

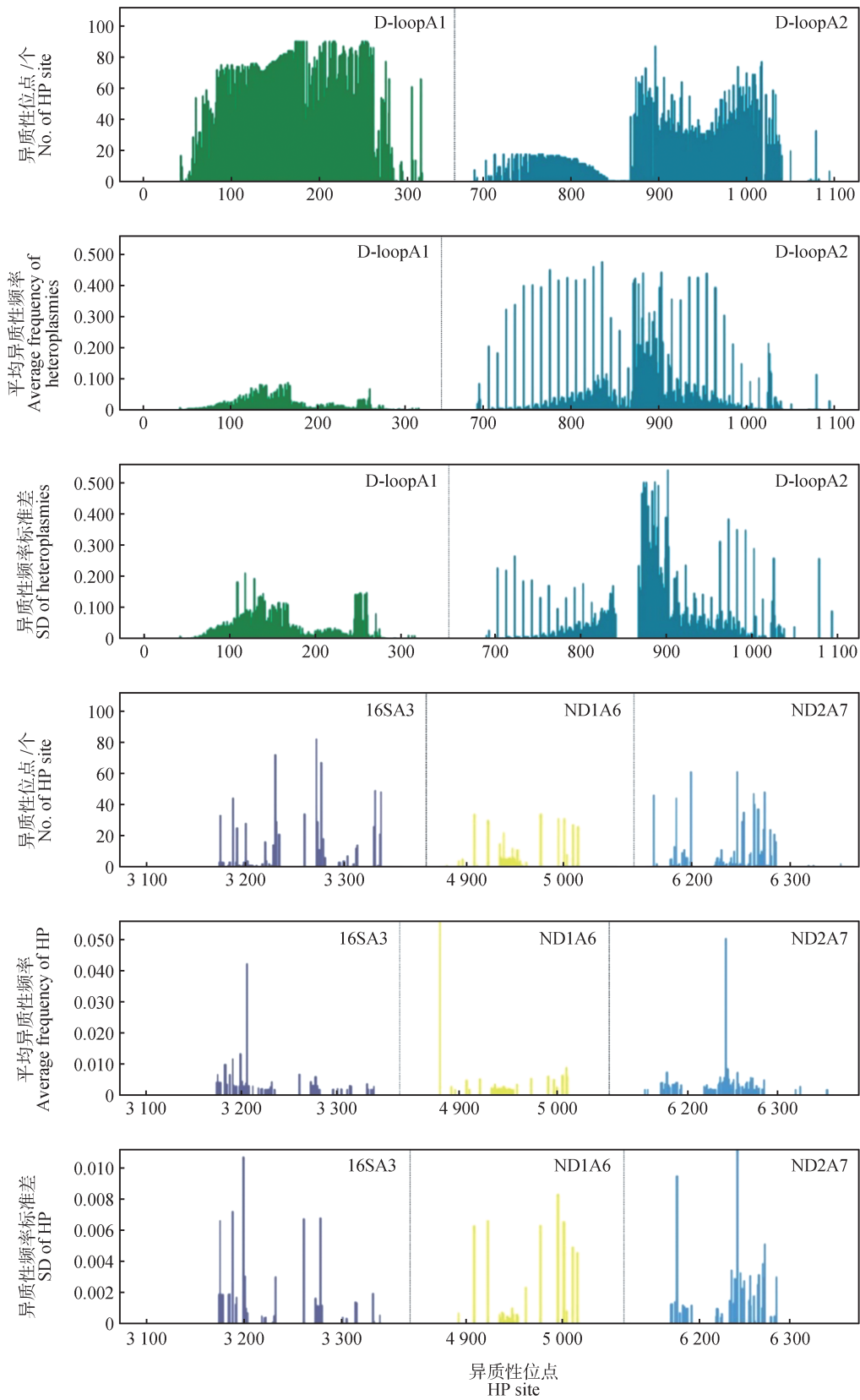


图2 大熊猫粪便 mtDNA 5 个区段异质性(HP)的空间分布格局

Figure 2 Spatial distribution pattern of heteroplasmy (HP) in five regions of fecal mtDNA of giant pandas

此外,大熊猫线粒体异质性表现出极其显著的个体差异。如D-loopA1区一个个体最多检测到210个异质性突变,而最低只有18个;D-loopA2区最多检测到273个,而最低只有7个。编码区也表现出很大的个体差异,每只个体的最大数为23~35个,最低为0。因此,个体平均异质性位点数的标准差也较大,变异系数(标准差/均值)为0.289(D-loopA1)~1.385(ND2A7),平均可达0.844。

2.2 mtDNA 不同区段异质性的碱基替换特点

从5个区段的整体来看,个体上检测到的转换

(T_S)为8~425个,平均为(164.7 ± 74.1)个;颠换(T_V)为20~384个,平均为(202.6 ± 70.7)个; T_S/T_V 为0.400~1.107,平均为(0.786 ± 0.136)。各区段上的 T_S/T_V 变化较大,如D-loopA1区 T_S/T_V 为0.267~0.897,平均值为0.706,而D-loopA2区 T_S/T_V 为0.385~2.250,平均值为0.979,说明D-loop区受到正选择和中性选择。相比而言,其他编码区更倾向于纯化选择(负选择),如16SA3区域 T_S/T_V 为0.333~19.000,平均值为3.638,ND1与ND2区域 T_S/T_V 为0~9.000,平均值为1.671(图3)。

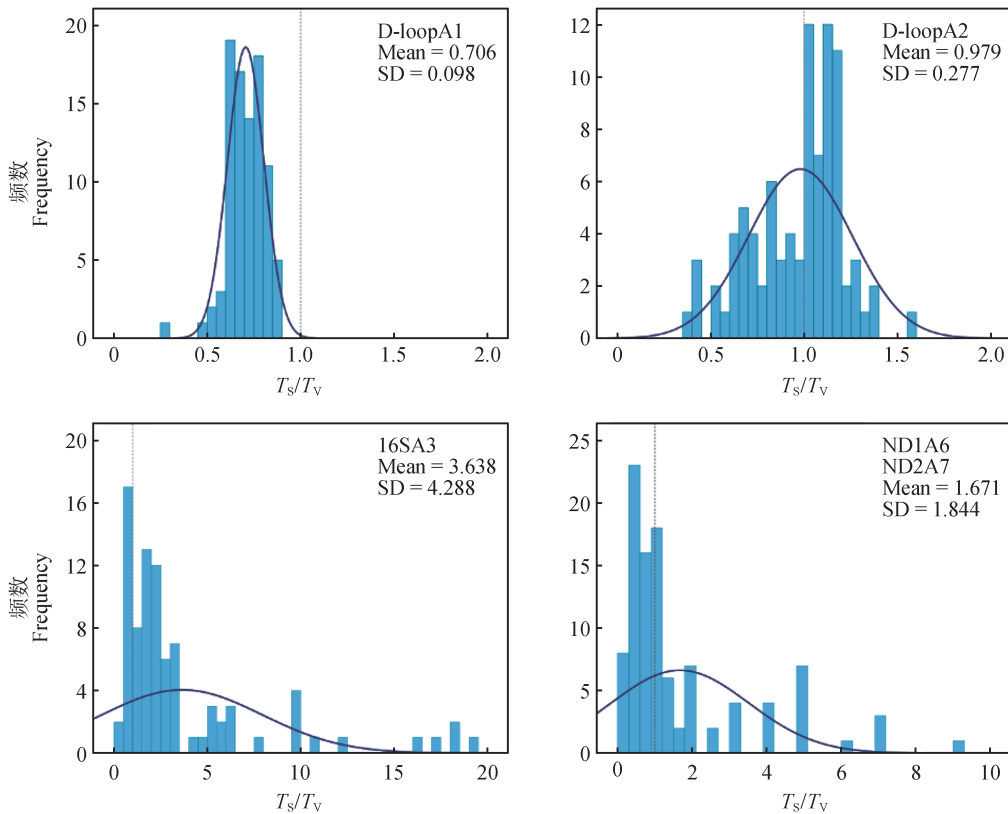


图3 大熊猫 mtDNA 四个区域异质性突变转换与颠换比(T_S/T_V)的频率分布

Figure 3 Histogram of heteroplasmic transitions-transversion ratio (T_S/T_V) in four regions of giant panda mitogenome

将ND1和ND2两个蛋白质编码区域按照密码子C1、C2和C3分别计算异质性突变频率并进行比较,结果显示,ND1A6区域上3个密码子之间都存在极显著差异(C1 vs. C2: $P < 0.001$; C1 vs. C3: $P = 0.008$; C2 vs. C3: $P < 0.001$);同样,ND2A7区域也显示相同的差异性,3对比较的 P 值均小于0.001(图4)。

2.3 不同年龄大熊猫 mtDNA 的异质性

将每只个体的异质性碱基总数与其月龄作图5,结果显示,把5个区域作为整体(total)来看,点的分布没有显示出显著的倾向性,而是以围绕质心向四

周发散。把异质性碱基中的转换和颠换分别提取出来,也同样没有显著的倾向性分布。单独计算各个区段,发现D-loop区的2个区域没有倾向性分布,但编码区16SA3、ND1A6和ND2A7三个区域的异质性碱基数量分布出现一定的倾向性,预示其与年龄可能存在一定的关系。

皮尔逊相关性分析显示,月龄与5个区域整体的异质性碱基总数,或其中的 T_S 和 T_V 数量没有显著相关关系($P = 0.058 \sim 0.139$);在单独考量各个区段时,D-loopA1和D-loopA2区均未显示出显著相关

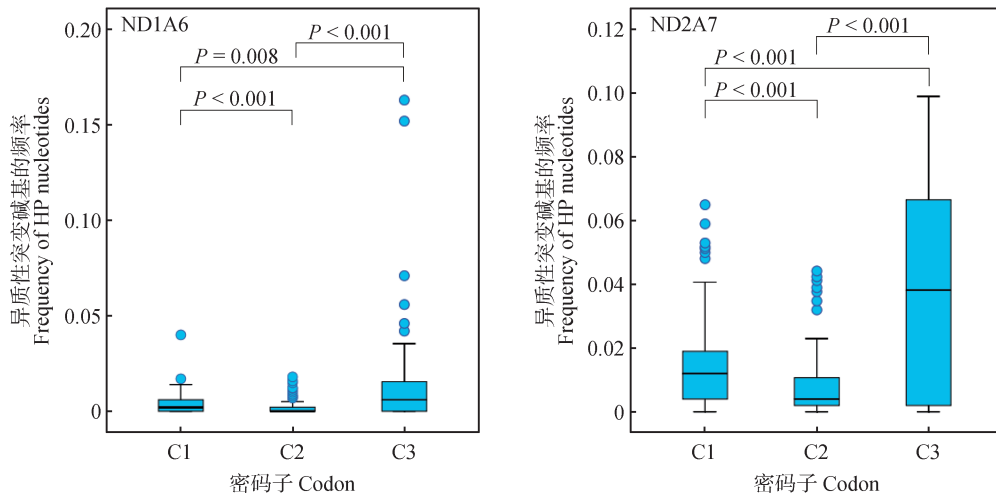
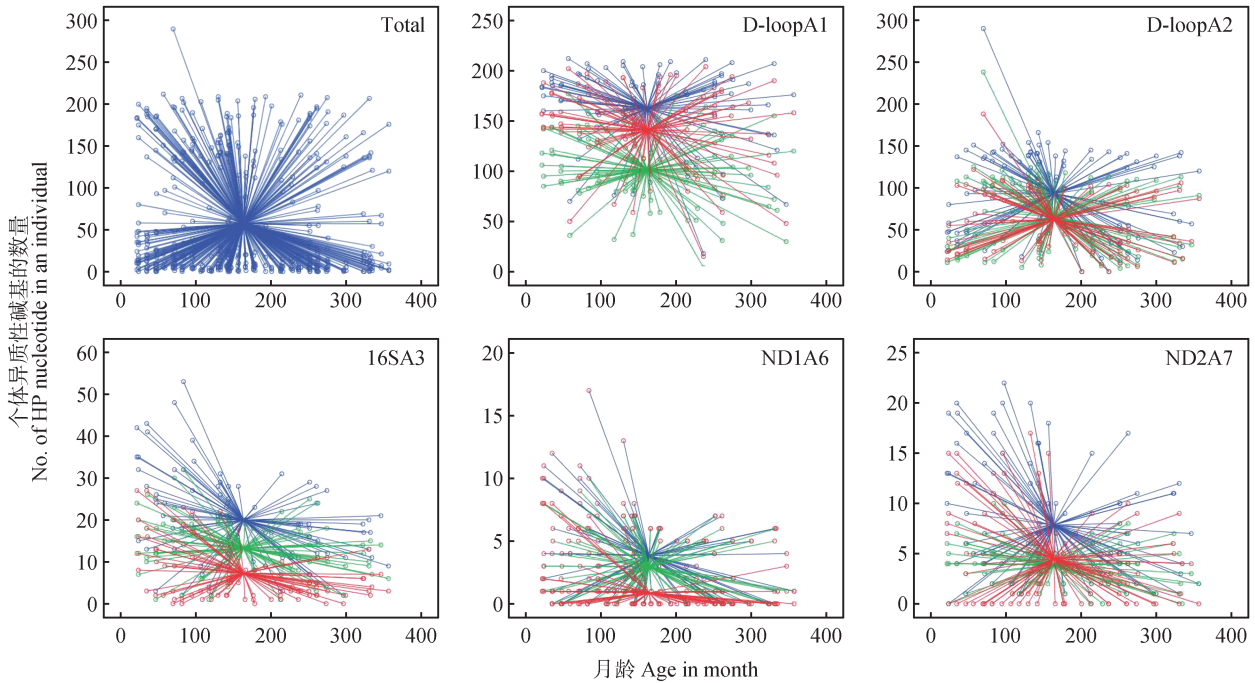


图4 ND1和ND2蛋白质编码区不同密码子异质性突变频率的分布

Figure 4 Frequency distribution of total heteroplasmic mutations at different codon positions in ND1 and ND2 protein coding regions



蓝色表示该区域异质性碱基的总数;绿色表示转换(T_s)碱基的数量;红色表示颠换(T_v)碱基的数量。

Blue circles and lines are the total number of heteroplasmic nucleotide; Green indicates the number of transitions (T_s); Red indicates the number of transversions (T_v).

图5 大熊猫粪便mtDNA异质性随着月龄的分布

Figure 5 Distribution of heteroplasmy of giant panda fecal mtDNA along the age classes in months

性($P = 0.058 \sim 0.291$),但其他3个区域则显示出显著相关性:月龄与16SA3区的相关系数 r 为 -0.426 , $P < 0.001$;与ND1A6的 r 为 -0.369 , $P < 0.001$;与ND2A7的 r 为 -0.295 , $P = 0.002$ 。通过多元线性回归建立3个区域异质性碱基数和月龄的模型:

$$A = 235.725 - 2.564x_1 - 4.013x_2 - 0.671x_3。$$

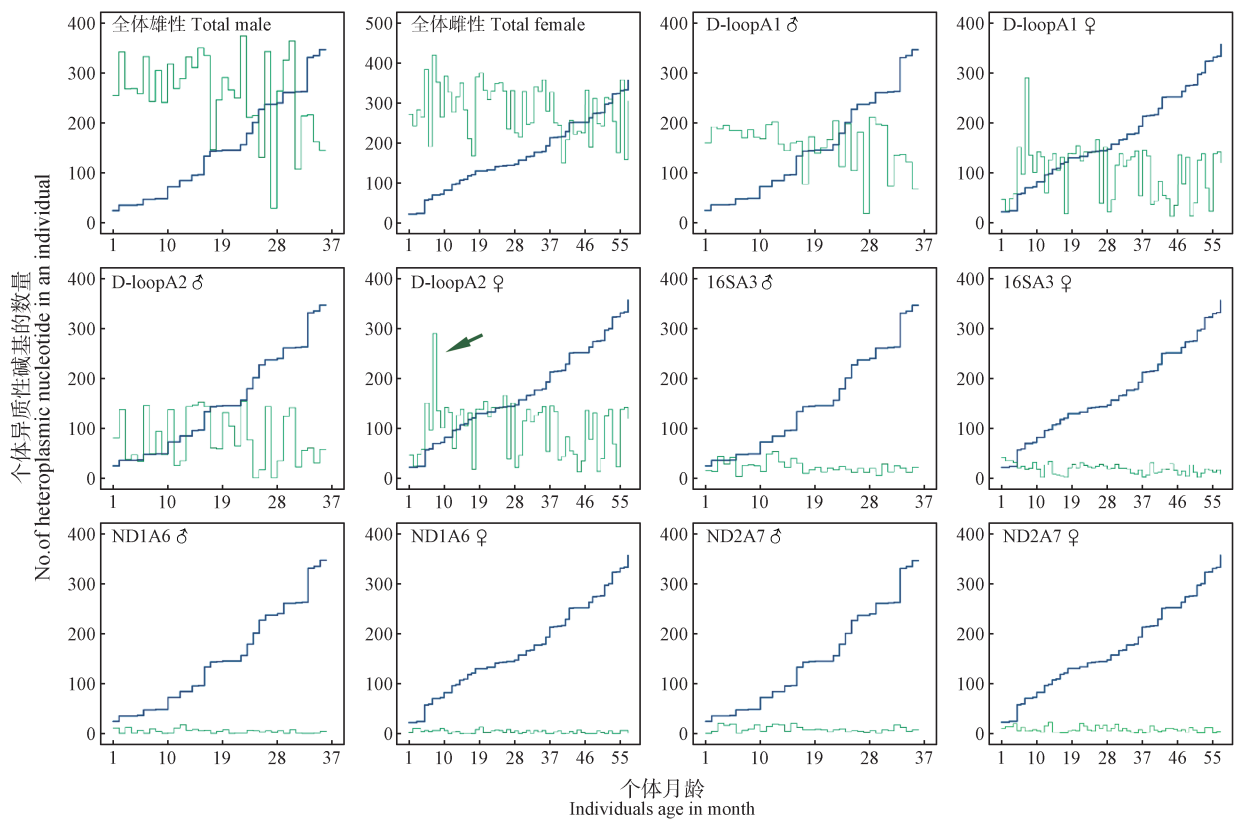
其中, A 为年龄, x_1 、 x_2 和 x_3 分别为16SA3、ND1A6和ND2A7的异质性碱基数。然而,虽然这一模型是显著的($F = 6.924$, $P < 0.001$),但相关系数的平方(R^2)只有0.191,说明异质性碱基数与年龄的关系较弱。

2.4 mtDNA异质性与性别的关系

通过相关样本Wilcoxon秩和检验进行比较,5个

区域总的异质性碱基数在雌性之间存在显著差异 ($P = 0.020$)。在分区域比较时, D-loopA1、16SA3、ND1A6 和 ND2A7 均无显著差异 (P 分别为 0.318、0.647、0.510、0.765), 只在 D-loopA2 存在显著差异 ($P = 0.024$)。按性别分组, 把每组样本再按年龄从小到大排列制作折线图, 观察异质性碱基数的分布

情况。发现5个区域的整体和 D-loopA2 的性别差异可能受极端值的影响所致(图6绿色箭头所示)。把此值去除后, P 值分别增加到 0.040, 差异的显著性明显下降。因此, 总体上可以认为异质性在性别之间没有显著差异。



蓝色线表示月龄;绿色线表示异质性碱基的数量。

Blue lines are individual age of giant panda, and green lines are the total number of heteroplasmic nucleotides.

图6 大熊猫粪便 mtDNA 异质性与月龄的关系

Figure 6 Correlation between heteroplasmy of fecal mtDNA and the age (in month) of giant panda

3 讨论

3.1 mtDNA 异质性的特点与纯化选择

本研究选择大熊猫线粒体基因组上 5 个区段来探索其异质性, 分别代表非编码区、rRNA 编码区和蛋白质编码区。试验中为了消除 PCR 扩增、测序掺入的假阳性突变, 课题组以单克隆模板进行扩增和高通量测序, 来检测 PCR 扩增以及二代测序过程中碱基错配导致的突变率, 以突变率的上限 0.200% 作为过滤假阳性数据的阈值, 保证试验数据的可靠性。

考察到 5 个区域都存在明显的异质性, 以 D-loop 区最高, 16S rRNA 区次之, 而蛋白质编码区的 ND 区

最低, 这意味着异质性突变可能对线粒体的机能产生影响而受到细胞的选择, 非编码区的突变影响较小而得以保留, 编码区的负面影响较大而被清除。这一推测得到 3 方面的证据支持。一是突变位点碱基替换频率的标准差在编码区低于非编码区(图 2), 意味着编码区的异质性突变比非编码区更为保守。二是本课题组检测到 D-loop 区的两个区段的 T_s/T_v 值分别为 0.706 和 0.979, 而 16S rRNA 和 ND 两个编码区的 T_s/T_v 平均值为 3.638 和 1.671。从化学过程的难易程度来讲, T_s 比 T_v 实现的难度小于 50%。在完全没有选择的条件下, 任何碱基替换都不会被清除, 于是 $T_s/T_v = 0.5$ ^[40]。实际上 T_v 比 T_s 更能引起 DNA 空

间构象的变化,在转录因子的结合区等重要功能区域发生 T_V ,往往会对其结合能力产生很大的影响,引发比 T_S 更为明显的功能改变^[41],故此在细胞器和细胞水平会受到更大的选择压力^[42]。本研究结果表明,D-loop区保留着较多的 T_V 突变,而有编码功能的16S RNA区和ND区有大量 T_V 被淘汰,受到更强烈的纯化选择。三是ND1和ND2两个蛋白质编码区3个密码子的突变频率渐次提高(图4),这完全符合密码子简并性,即C1和C2的突变更可能改变编码的氨基酸而受到较强的选择压力,而C3具有较高的简并性,选择压力较轻^[43]。

3.2 mtDNA的异质性与年龄的关系

异质性的来源包括mtDNA复制过程和损伤修复过程中出现的错误^[7-9]。随着动物体年龄增加,肠隐窝的干细胞因不断分生和线粒体复制而积累突变^[44],同时衰老会引起较多的DNA损伤和错误修复^[45]。因此普遍认为随着年龄的增长,异质性突变会随之增加^[4,46]。本研究发现,除了D-loop区外,16SA3、ND1A6和ND2A7三个编码区的异质性碱基数量都可能与年龄相关。但是,相关系数较低(-0.426 ~ -0.295),多元线性回归建立三者与年龄的关系模型 R^2 只有0.191,说明与年龄的关系较弱。本研究使用的是异质性突变碱基数作为变量。在从高通量测序数据中检测异质性突变时,本课题组使用0.200%作为过滤条件。这一条件相比其他研究低了1个量级^[4,47-48],所得到的信息包含两种突变来源,一是分化的上皮细胞中发生的新突变,二是隐窝干细胞中发生的突变。在分化细胞中的突变无法扩增而始终处于低频状态,并在不断产生的同时,也会随着上皮细胞的脱落而被不断消除。故此,其所反映的异质性信息不具有随时间的累加性。在构建异质性与年龄关系模型时,此类突变信息会成为模型的噪音。相反,隐窝干细胞中产生的突变可通过子代细胞的增殖而发生克隆性扩增,使频率增高。随着时间的推移,干细胞中会发生越来越多的突变,并被克隆性放大。因而,这类异质性突变具有随时间的累加性。如果提高过滤阈值来滤除第一类突变信息而保留第二类信息,可能在一定程度上会提高模型的精确性。此外,如果再考虑选择压力,通过给密码子不同位置和转换、颠换赋予不同的权重,也可能进一步丰富异质性的时间属性,进而提高模型的精确性。

综上所述,大熊猫粪便DNA中存在多样性很高的线粒体基因库,异质性突变是基因库发展的基本源泉。异质性突变在组织、细胞水平上受到自然选择作用,表现为蛋白质编码区的纯化选择压力大于rRNA编码区,更大于非编码区;同时,蛋白质编码区的第3位密码子的选择压力显著低于前两位。这一现象与种群水平的mtDNA多样性完全相同。在宏观世界中,mtDNA的自然选择是通过动物体的死亡或失去繁殖机会来实现,而机体内的选择则是通过mtDNA、细胞器和细胞等的降解(死亡)来实现^[42]。因此,从个体角度出发,线粒体基因库的发展、变化始终会影响细胞的生命过程,进而影响机体的适合度^[46]。粪便DNA主要来自大肠上皮的脱落细胞^[32],而这个位置也是肠道微生物与宿主的主要互作界面^[49],其线粒体基因库对揭示适应性进化机制、个体对环境的适应方式、个体适合度、动物与肠道微生物及环境(食物条件)的关系等具有重要价值,也可能成为动物研究和保护的新工具。

参考文献:

- [1] 巫小倩,张顺华,朱砾. 线粒体DNA异质性[J]. 中国生物化学与分子生物学报, 2017, 33(1): 11-16.
WU X Q, ZHANG S H, ZHU L. Mitochondrial DNA heteroplasmy [J]. Chinese Journal of Biochemistry and Molecular Biology, 2017, 33(1): 11-16.
- [2] AMORIM A, FERNANDES T, TAVEIRA N. Mitochondrial DNA in human identification: a review [J]. PeerJ, 2019, 7: e7314.
- [3] SOBENIN I A, MITROFANOV K Y, ZHELANKIN A V, *et al.* Quantitative assessment of heteroplasmy of mitochondrial genome: perspectives in diagnostics and methodological pitfalls [J]. BioMed Research International, 2014, 2014: 292017.
- [4] ZHANG R Y, WANG Y Q, YE K X, *et al.* Independent impacts of aging on mitochondrial DNA quantity and quality in humans [J]. BMC Genomics, 2017, 18(1): 890.
- [5] VAKIFAHMETOGLU-NORBERG H, OUCHIDA A T, NORBERG E. The role of mitochondria in metabolism and cell death [J]. Biochemical and Biophysical Research Communications, 2017, 482(3): 426-431.
- [6] CASTELLANI C A, LONGCHAMPS R J, SUN J, *et al.* Thinking outside the nucleus: mitochondrial DNA copy number in health and disease [J]. Mitochondrion, 2020, 53: 214-223.
- [7] ZHENG W M, KHRAPKO K, COLLER H A, *et al.* Origins of human mitochondrial point mutations as DNA polymerase γ -mediated errors [J]. Mutation Research, 2006, 599(1/2): 11-20.
- [8] SHOKOLENKO I, VENEDIKTOVA N, BOCHKAREVA A, *et al.* Oxidative stress induces degradation of mitochondrial DNA [J]. Nucleic Acids Research, 2009, 37(8): 2539-2548.

- [9] FONTANA G A, GAHLON H L. Mechanisms of replication and repair in mitochondrial DNA deletion formation[J]. *Nucleic Acids Research*, 2020, 48(20): 11244–11258.
- [10] MASTRANTONIO V, LATROFA M S, PORRETTA D, *et al.* Paternal leakage and mtDNA heteroplasmy in *Rhipicephalus* spp. ticks[J]. *Scientific Reports*, 2019, 9(1): 1460.
- [11] WEN M, PENG L Y, HU X J, *et al.* Transcriptional quiescence of paternal mtDNA in cyprinid fish embryos[J]. *Scientific Reports*, 2016, 6(1): 28571.
- [12] UJVARI B, DOWTON M, MADSEN T. Mitochondrial DNA recombination in a free-ranging Australian lizard[J]. *Biology Letters*, 2007, 3(2): 189–192.
- [13] STEWART J B, CHINNERY P F. Extreme heterogeneity of human mitochondrial DNA from organelles to populations[J]. *Nature Reviews Genetics*, 2021, 22(2): 106–118.
- [14] PARAKATSELAKI M E, LADOUKAKIS E D. mtDNA heteroplasmy: origin, detection, significance, and evolutionary consequences[J]. *Life*, 2021, 11(7): 633.
- [15] SOLIGNAC M, MONNEROT M, MOUNOLOU J C. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1983, 80(22): 6942–6946.
- [16] WANG W H, LIN L J, ZHANG Q, *et al.* Heteroplasmy and individual mitogene pools: characteristics and potential roles in ecological studies[J]. *Biology*, 2023, 12(11): 1452.
- [17] LI B, LU J Y, MONAKHOV V, *et al.* Phylogeography of subspecies of the sable (*Martes zibellina* L.) based on mitochondrial genomes: implications for evolutionary history[J]. *Mammalian Biology*, 2021, 101(1): 105–120.
- [18] NABHOLZ B, MAUFFREY J F, BAZIN E, *et al.* Determination of mitochondrial genetic diversity in mammals[J]. *Genetics*, 2008, 178(1): 351–361.
- [19] DU H Y, ZHANG X Y, DINH T D, *et al.* Identification of hybrid green peafowl using mitochondrial and nuclear markers[J]. *Conservation Genetics Resources*, 2020, 12(4): 669–683.
- [20] DINH T D, NGATIA J N, CUI L Y, *et al.* Influence of pairwise genetic distance computation and reference sample size on the reliability of species identification using *Cyt b* and *COI* gene fragments in a group of native passerines[J]. *Forensic Science International: Genetics*, 2019, 40: 85–95.
- [21] XU W H, VIÑA A, KONG L Q, *et al.* Reassessing the conservation status of the giant panda using remote sensing[J]. *Nature Ecology & Evolution*, 2017, 1(11): 1635–1638.
- [22] WEI F W, FAN H Z, HU Y B. *Ailuropoda melanoleuca* (giant panda)[J]. *Trends in Genetics*, 2020, 36(1): 68–69.
- [23] LI Y, XU W, WANG Y, *et al.* An improved, chromosome-level genome of the giant panda (*Ailuropoda melanoleuca*)[J]. *Genomics*, 2022, 114(6): 110501.
- [24] GUANG X M, LAN T M, WAN Q H, *et al.* Chromosome-scale genomes provide new insights into subspecies divergence and evolutionary characteristics of the giant panda[J]. *Science Bulletin*, 2021, 66(19): 2002–2013.
- [25] YANG S C, LAN T M, WEI R P, *et al.* Single-nucleus transcriptome inventory of giant panda reveals cellular basis for fitness optimization under low metabolism[J]. *BMC Biology*, 2023, 21(1): 222.
- [26] HU Y B, WU Q, MA S, *et al.* Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(5): 1081–1086.
- [27] MA J N, ZHANG L, SHEN F J, *et al.* Gene expressions between obligate bamboo-eating pandas and non-herbivorous mammals reveal converged specialized bamboo diet adaptation[J]. *BMC Genomics*, 2023, 24(1): 23.
- [28] ZHOU C, LIU Y, ZHAO G Q, *et al.* Comparative analysis of olfactory receptor repertoires sheds light on the diet adaptation of the bamboo-eating giant panda based on the chromosome-level genome[J]. *Animals*, 2023, 13(6): 979.
- [29] ZHU C L, LAGHI L, ZHANG Z Z, *et al.* First steps toward the giant panda metabolome database: untargeted metabolomics of feces, urine, serum, and saliva by ¹H NMR[J]. *Journal of Proteome Research*, 2020, 19(3): 1052–1059.
- [30] ZHU C L, PAN X, LI G, *et al.* Lipidomics for determining giant panda responses in serum and feces following exposure to different amount of bamboo shoot consumption: a first step towards lipidomic atlas of bamboo, giant panda serum and feces by means of GC-MS and UHPLC-HRMS/MS[J]. *International Journal of Molecular Sciences*, 2022, 23(19): 11544.
- [31] YAN Z, XU Q, YAO Y, *et al.* Fecal metabolomics reveals the foraging strategies of giant pandas for different parts of bamboo[J]. *Animals*, 2023, 13(8): 1278.
- [32] PERRY G H, MARIONI J C, MELSTED P, *et al.* Genomic-scale capture and sequencing of endogenous DNA from feces[J]. *Molecular Ecology*, 2010, 19(24): 5332–5344.
- [33] BANDALETOVA T, BAILEY N, BINGHAM S A, *et al.* Isolation of exfoliated colonocytes from human stool as a new technique for colonic cytology[J]. *APMIS*, 2002, 110(3): 239–246.
- [34] WILLIAMS J M, DUCKWORTH C A, BURKITT M D, *et al.* Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip[J]. *Veterinary Pathology*, 2015, 52(3): 445–455.
- [35] PARIKH K, ANTANAVICIUTE A, FAWKNER-CORBETT D, *et al.* Colonic epithelial cell diversity in health and inflammatory bowel disease[J]. *Nature*, 2019, 567(7746): 49–55.
- [36] KAUNITZ J D, AKIBA Y. Control of intestinal epithelial proliferation and differentiation: the microbiome, enteroendocrine L cells, telocytes, enteric nerves, and GLP, too[J]. *Digestive Diseases and Sciences*, 2019, 64(10): 2709–2716.
- [37] ZHANG H B, MENZIES K J, AUWERX J. The role of mitochondria in stem cell fate and aging[J]. *Development*, 2018, 145(8): dev143420.

- [38] POTTEN C S. A comprehensive study of the radiobiological response of the murine (BDF1) small intestine [J]. *International Journal of Radiation Biology*, 1990, 58(6): 925–973.
- [39] JEON S A, PARK J L, PARK S J, *et al.* Comparison between MGI and Illumina sequencing platforms for whole genome sequencing [J]. *Genes & Genomics*, 2021, 43(7): 713–724.
- [40] YANG Z H, NIELSEN R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage [J]. *Molecular Biology and Evolution*, 2008, 25(3): 568–579.
- [41] GUO C, MCDOWELL I C, NODZENSKI M, *et al.* Transversions have larger regulatory effects than transitions [J]. *BMC Genomics*, 2017, 18(1): 394.
- [42] STEWART J B, FREYER C, ELSON J L, *et al.* Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease [J]. *Nature Reviews Genetics*, 2008, 9(9): 657–662.
- [43] MCCLELLAN D A, WHITING D G, CHRISTENSEN R, *et al.* Genetic codes as evolutionary filters: subtle differences in the structure of genetic codes result in significant differences in patterns of nucleotide substitution [J]. *Journal of Theoretical Biology*, 2004, 226(4): 393–400.
- [44] TAYLOR R W, BARRON M J, BORTHWICK G M, *et al.* Mitochondrial DNA mutations in human colonic crypt stem cells [J]. *Journal of Clinical Investigation*, 2003, 112(9): 1351–1360.
- [45] GORBUNOVA V, SELUANOV A, MAO Z Y, *et al.* Changes in DNA repair during aging [J]. *Nucleic Acids Research*, 2007, 35(22): 7466–7474.
- [46] BURGSTALLER J P, KOLBE T, HAVLICEK V, *et al.* Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations [J]. *Nature Communications*, 2018, 9(1): 2488.
- [47] DING J, SIDORE C, BUTLER T J, *et al.* Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2,000 Sardinians using tailored sequencing analysis tools [J]. *PLoS Genetics*, 2015, 11(7): e1005306.
- [48] HÜBNER A, WACHSMUTH M, SCHRÖDER R, *et al.* Sharing of heteroplasmies between human liver lobes varies across the mtDNA genome [J]. *Scientific Reports*, 2019, 9(1): 11219.
- [49] ASHIDA H, OGAWA M, KIM M, *et al.* Bacteria and host interactions in the gut epithelial barrier [J]. *Nature Chemical Biology*, 2012, 8(1): 36–45.