

猫科动物 DNA 条形码及线粒体假基因对物种鉴定的影响

蔡延森¹, 李佳凌², 赵娇¹, 张亮^{3*}

(1. 西南医科大学医学细胞生物学与遗传学教研室, 四川泸州 646000; 2. 西南医科大学形态学研究室, 四川泸州 646000; 3. 成都大熊猫繁育研究基地, 四川省濒危野生动物保护生物学重点实验室, 成都 610081)

摘要: 在多种动物类群中, 基于线粒体细胞色素 c 氧化酶亚基 I (*COI*) 基因的 DNA 条形码是一种高效的物种鉴别手段, 然而猫科 Felidae 动物中广泛存在的线粒体假基因可能影响 DNA 条形码的有效性。本研究共涉及猫科动物 12 属 25 种 119 个样本。采用 3 对条形码通用引物对 6 属 11 种 29 个猫科动物样本进行了扩增及测序。结果 3 个样本扩增失败, 8 个样本得到假基因, 18 个样本获得了条形码序列。结合另外 93 条猫科动物条形码序列 (源自 BOLD Systems), 采用 Kimura 2-parameter 模型计算遗传距离, 构建 Neighbor-Joining (NJ) 树。结果显示, 遗传距离种内为 0% ~ 8.1%, 平均 0.8%; 种间为 1.4% ~ 13.1%, 平均 8.7%; 属间为 8.2% ~ 21.8%, 平均 15.1%。NJ 树显示, 除 3 个种外, 其余物种均以极高的置信度 (99%) 形成单系分支。而假基因序列有些可以单独形成分支, 有些夹杂在 *COI* 序列形成的分支中, 对物种鉴定产生干扰。

关键词: *COI*; 猫科; 线粒体假基因

中图分类号: Q959.8 文献标志码: A 文章编号: 1000-7083(2017)04-0425-06

DNA Barcoding in Felidae, and the Effects of Nuclear Mitochondrial Pseudogenes on Species Identification

CAI Yansen¹, LI Jialing², ZHAO Jiao¹, ZHANG Liang^{3*}

(1. Department of Medical Cell Biology and Genetics, Southwest Medical University, Luzhou, Sichuan Province 646000, China; 2. Department of Morphology, Southwest Medical University, Luzhou, Sichuan Province 646000, China; 3. Key Laboratory of Conservation Biology on Endangered Wildlife, Chengdu Research Base of Giant Panda Breeding, Chengdu 610081, China)

Abstract: In a variety of animal species, DNA barcoding based on mitochondrial cytochrome c oxidase subunit I (*COI*) gene is an efficient tool for species identification. However, the presence of large mitochondrial pseudogenes in felids may affect the effectiveness of DNA barcoding. In this study, 119 felid samples from 25 species of 12 genera were analyzed. 3 pairs of universal primers were used to amplify and sequence the *COI* genes of 29 felid samples that belonged to 11 species from 6 genera. The results showed that the amplification was failed in 3 samples, nuclear mitochondrial pseudogenes (numts) were obtained in 8 samples, and barcode sequences were derived in 18 samples. Combined with the other 93 felid barcode sequences (derived from BOLD Systems), Kimura 2-parameter model was used to calculate the genetic distance and to construct Neighbor-Joining (NJ) tree. The genetic distance within species were 0% - 8.1%, with an average of 0.8%; the interspecific genetic distance were 1.4% - 13.1%, with an average of 8.7%; and the genetic distance within family were 8.2% - 21.8%, with an average of about 15.1%. The NJ tree showed that most of the species formed a monophyletic clade with high confidence (99%), except 3 species; some of the numts sequences could form a clade on NJ tree, and the others were inserted in the branches of the *COI* sequences, which may compromise the species identification.

Keywords: *COI*; Felidae; nuclear mitochondrial pseudogenes

近年来, DNA 条形码作为一种分子级别的物种鉴定手段, 已经在多种昆虫、鱼类、鸟类、哺乳类等动物中显示了其准确、高效、便捷的特点 (Hebert *et al.*, 2003, 2004; 蔡延森等, 2009, 2014; Cai *et al.*, 2011,

2017; Kress *et al.*, 2015; Li *et al.*, 2017)。在动物中, 常用的 DNA 条形码是一段标准化的线粒体基因片段, 即细胞色素 c 氧化酶亚基 I (*COI*) 基因片段 (Hebert *et al.*, 2003)。该段序列带有丰富的物种信

收稿日期: 2017-03-13 接受日期: 2017-06-15

基金项目: 四川省教育厅科研项目 (12ZA242)

作者简介: 蔡延森 (1982—), 男, 硕士, 讲师, E-mail: cys_cys@aliyun.com

* 通信作者 Corresponding author, E-mail: zliang@panda.org.cn

息,适用 PCR 扩增的通用引物,极少出现插入或缺失,并且具有母系遗传、无内含子、重组受限等特点 (Brown *et al.*, 1982)。这些特点使得条形码序列的测序及序列分析方便快捷,进而准确地显示出样本的种属信息。这种快捷准确的物种鉴定技术,对于检验检疫、动物制品质量控制、生物多样性保护、圈养动物管理等各方面工作有重要意义。

线粒体假基因 (nuclear mitochondrial pseudogenes, numts) 是起源于线粒体,后来整合到核基因组中并失去功能的 DNA 序列 (Lopez *et al.*, 1994)。用条形码通用引物从总基因组 DNA 中扩增 *CO I* 基因片段时,可能得到 numts 序列而并非 mtDNA 上的目的基因片段 (周志军等, 2014)。虽然 numts 并不是普遍存在于所有物种中,但猫科 Felidae 动物核基因组中存在着大段的 numts 序列,例如家猫线粒体全基因组约 17 kb,其核基因组中就有高达 7.9 kb 的 numts (Lopez *et al.*, 1994)。这不仅会增加物种鉴定的难度,甚至可能会导致错误的鉴定结论。

本研究对 25 种猫科动物的样本进行了分析,包括国家 I 级重点保护动物——东北虎 *Panthera tigris altaica*、华南虎 *P. tigris amoyensis*、雪豹 *P. uncia*、金

钱豹 *P. pardus*, 国家 II 级保护动物亚洲金猫 *Catopuma temminckii*、兔狲 *Otocolobus manul*、猞猁 *Lynx lynx*、欧洲野猫 *Felis silvestris silvestris*、荒漠猫 *F. bieti*, 以及列入 IUCN 濒危物种红色名录的狮子 *P. leo*、美洲狮 *Puma concolor*、美洲豹 *P. onca*、猎豹 *Acinonyx jubatus* 等,并对其中出现较高频率 numts 的现象进行了分析,探究 DNA 条形码对猫科动物物种鉴定的可靠性。

1 实验材料与方法

1.1 实验材料

本研究共涉及猫科动物 12 属 25 种 119 个样本,其中 6 属 11 种 29 个样本为本实验室提供 (表 1)。所有样本均通过非损伤取样获得,采集自猫科动物的粪便,从中获取毛发后提取毛囊 DNA,然后保存于 4 °C 冰箱中备用或 95% 乙醇中 -80 °C 长期保存。其余序列下载自 The Barcode of Life Data Systems (BOLD Systems, <http://www.boldsystems.org>)。在 BOLD 猫科动物条形码目录中下载了全部 323 个序列 (截止 2017 年 3 月),删除非 *CO I* 基因及长度小于 500 bp 的 *CO I* 序列,最后剩下 93 条序列参与分析 (含 numts)。

表 1 样本信息
Table 1 Sample information

物种 Species	样品编号 Sample no.
猞猁 <i>Lynx lynx</i>	LLY34, LLY35, LLY39, LLY40, LLY41, LLY42, LLY44, LLY46, LLY50
东北虎 <i>Panthera tigris altaica</i>	jd401*, jd417*, jd437*
华南虎 <i>Panthera tigris amoyensis</i>	jdPTA31*, jdPTA01*
金钱豹 <i>Panthera pardus</i>	jdPPA826*, jdPPA827*, jdPPA829*
狮子 <i>Panthera leo</i>	jd16*
雪豹 <i>Panthera uncia</i>	PUN90
欧洲野猫 <i>Felis silvestris silvestris</i>	FSI186
荒漠猫 <i>Felis bieti</i>	jdFBI3*, jdFBI6*
亚洲豹猫 <i>Prionailurus bengalensis</i>	jdPBE153, jdPBE154
亚洲金猫 <i>Catopuma temminckii</i>	PAU5, PAU6, PAU7
兔狲 <i>Otocolobus manul</i>	jdOMA48, jdOMA54

注: * 扩增得到假基因, * 扩增失败。

Notes: * nuclear mitochondrial pseudogene was obtained, * amplification was failed.

1.2 DNA 提取、PCR 扩增及测序

DNA 提取采用 QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany)。PCR 扩增共使用了 3 对引物,首先使用脊椎动物 *CO I* 通用引物对 LCOI1490: 5'-GGTCAACAAATCATAAAGATATTGG-3', HCOI2198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3'。如果扩增

失败或测序结果为套峰,则替换引物重新扩增,使用引物对 VF1: 5'-TTCTCAACCAACCACAAAGACATTGG-3', VR1: 5'-TAGACTTCTGGGTGGCCAAAGAATCA-3'; 或 Zlf04: 5'-TCTCAACTAAYCAYAAAGAYATYGG-3', Zlr04: 5'-TAAACTTCrGGGTGACCAAArAATCA-3'。每 25 μL PCR 反应体系包含: 约 25 ng DNA 模板,

2.5 mM MgCl₂, 1 × PCR buffer (TaKaRa, 中国大连), 1.0 U Taq DNA 聚合酶 (TaKaRa, 中国大连), 上下游引物各 0.5 μM, 每种 dNTP 各 50 μM。PCR 反应参数: 94 °C 预变性 5 min; 94 °C 45 s, 51 °C 45 s, 72 °C 45 s, 共 30 个循环; 72 °C 最后延伸 10 min; 4 °C 保存。PCR 产物用 1% 的琼脂糖凝胶电泳检测, 用凝胶回收试剂盒 (E. Z. N. A. Gel Extraction Kit) 回收约 700 bp 目的片段, 直接双向测序。

1.3 序列分析

将测序所获正、反向序列导入 DNAMAN, Edit-Seq 等软件进行拼接, 使用 MEGA 6 与同源序列进行比对。由于猫科动物中 numts 出现的频率大大高于其他物种, 而 COI 假基因大多存在提前的终止密码子或带有插入/缺失导致的移码突变, 所有序列还逐一进行了手工校正并分析其蛋白序列, 根据是否存在终止密码子、移码突变进而鉴定扩增所得为 COI 基因片段或 numts。最后用 Kimura 2-parameter (K2P) 模型计算遗传距离, 并构建 Neighbor-Joining (NJ) 树。

线粒体 DNA 的核转移时间计算方法采用 Li 等 (1981) 的计算公式: $\delta = (\mu_1 + \mu_2)t$ 。其中 t 为分歧时间, δ 为序列之间遗传差异, μ_1 和 μ_2 分别为线粒体 DNA 和 numts 每年每个核苷酸替换速率, $\mu_1 = 1.7 \times 10^{-8}$, $\mu_2 = 4.7 \times 10^{-9}$ 。

2 结果

本研究共提供了 29 个猫科动物个体的新样本, 其中 3 个样本 (jdPTA01、jdPPA827、jdPPA829) 扩增失败, 更换引物后扩增依然失败或测序结果为套峰, 无法读取序列。8 个样本测序结果最初为套峰, 更换引物重新扩增测序后, 得到的序列有明显 numt 特征, 存在碱基插入/缺失或终止密码子。其余 18 个样本经过 1~3 次扩增, 得到 660 bp 左右清晰无套峰的片段, 序列用 MEGA 6 比对分析无插入/缺失、终止密码子, 通过 NCBI 中 Blast 相似性检索, 确定为 COI 序列。

2.1 遗传距离分析

DNA 条形码序列的 K2P 遗传距离在猫科动物的种内差异普遍较高, 最低为 0, 最高达 8.1%, 平均为 0.8%; 同属种间遗传距离为 1.4%~13.1%, 平均为 8.7%; 同科属间遗传距离为 8.2%~21.8%, 平均为 15.1%。种内、同属种间、同科属间的遗传距离分布频率见图 1 (numts 序列未计入)。

Hebert 等 (2004) 提出 DNA 条形码划分物种的 10 × 规则, 即种间遗传差异要高于种内遗传差异 10 倍。野猫 *Felis silvestris* 及其圈养亚种家猫 *F. s. catus* 的种内个体间差异最大达到了 10.3%, 超过平均值 (0.8%) 12.8 倍, 不符合 Hebert 的 10 × 规则。

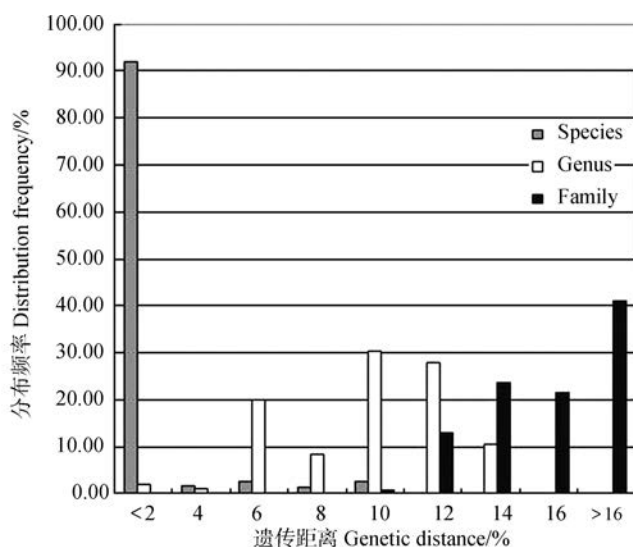


图 1 种内、同属种间、同科属间的遗传距离分布频率
Fig. 1 Distribution of genetic distances within species, genus and family of Felidae

2.2 聚类分析

本研究中有 16 个种含多个样本参与分析。MEGA 6 构建 NJ 树结果显示, 其中 13 个种以极高的

置信度 (99%) 各自形成单系分支, 表明这些种可被 DNA 条形码准确鉴定 (图 2), 包括猎豹、狞猫 *Caracal caracal*、亚洲金猫、猞猁、加拿大猞猁 *L. canadensis*

sis、短尾猫 *L. rufus*、兔狲、狮子、老虎 *P. tigris*、金钱豹、雪豹、扁头豹猫 *Prionailurus planiceps*、美洲狮。另外 3 个种没有形成单系, 分别有 2~4 个分支, 包括豹猫属 *Prionailurus* 的豹猫 *P. bengalensis*, 渔猫 *P. viverrinus* 和野猫 (*F. silvestris* 和 *F. s. catus*)。其余 9 个种只包含 1 个样本参与分析, 所有物种没有共享序列, 在 NJ 树上物种没有重叠。

在豹猫属中, 物种的种内平均差异为 1.1%~5.3%, 种间平均差异为 4.6%~6.3%, 物种之间没有共享序列。表明虽然存在较大的种内分化, 但基于聚类分析, DNA 条形码依然能够对它们进行有效的物种鉴别。

在 NJ 树上, 野猫大部分个体聚在一起形成一个 大支 (*Felis silvestris* 和 *F. s. catus*), 该群体内遗传差异平均为 0.3% (0~1.6%), 其余部分序列分别形成了 3 个小分支 (*F. s. catus* 1、*F. s. catus* 2 和 *F. s. catus* 3)。这 3 个小分支同大分支的平均遗传差异分别为 7.9%、3.6%、5.3%, 分别有 4、1、9 处碱基位点突变导致氨基酸发生改变。在 NCBI 中用 Blast 相似性检索, 这 3 个分支中的序列同家猫线粒体假基因全序列的相似度分别达到 99%、96%、95%。

2.3 假基因的分析

在狮子、老虎、金钱豹、荒漠猫 4 个物种中检测出带有终止密码子的序列, 为明显的 numts 特征。对这些 numts 序列进行分析发现: 狮子、老虎、金钱豹 3 个物种的 numts 以 87% 的置信度聚在一起, 群体内遗传差异为 0.5%~1.4%。根据 Li 等 (1981) 对 mtDNA 和 numts 的分歧时间计算方法, 可以得知狮子、老虎、金钱豹的核转移事件发生的时间分别为约 220 万年前、350 万年前、270 万年前。

荒漠猫假基因序列 (图 2, NJ 树上 *F. bieti* numts) 同野猫 2 个个体的“CO I 序列” (图 2, NJ 树上 *F. s. catus* 1) 以 100% 的置信度聚为一支, 群体内遗传差异仅为 0.5%~0.8%; 而野猫 *F. s. catus* 1 这一支中的 2 个序列同野猫其他个体之间的遗传差异高达 6.9%~10.3%, 因此推测这 2 个序列极有可能是 numts (GBMA5696-13 | *Felis catus* | CO I -5P | JX426133 和 GBMIN43334-14 | *Felis catus* | CO I -5P | KF771218)。

3 讨论

本研究中猫科动物种内和种间的平均遗传差异分别为 0.8% 和 8.7%, NJ 树上大部分物种都以极

高置信度形成单系, 这些数据表明 DNA 条形码在很大程度上能够对猫科动物实现快速、准确的鉴定。个别物种 NJ 树上没有形成单系且种内遗传差异较大, 可能有两方面的原因。

一方面, 可能是进化方面的原因: 新近分化的物种存在不完全谱系分选现象; 近源物种的种间杂交导致基因渗透; 由于地理/生殖隔离导致较大的种内遗传差异、多亚种等。这些因素都可能导致 DNA 条形码对物种的鉴别出现困难甚至失败。没有形成单系的豹猫和渔猫都是豹猫属的近源物种。其中, 豹猫至少含 8 个亚种 (Mukherjee *et al.*, 2016); 渔猫虽然没有亚种划分, 但该物种的栖息地碎片化, 长期处于地理/生殖隔离状态 (Mukherjee *et al.*, 2016)。此外, 全世界广泛分布的野猫有 22 个亚种 (Wozencraft, 2005), 在 NJ 树上有 4 个分支。可见, 正是这些进化方面的原因导致了这些物种的种内分化。

另一方面的原因是 numts 同 CO I 共扩增的现象, 会对物种鉴定造成干扰、错误, 甚至产生不存在的“新种”等。在 NJ 树上, 狮子、老虎、金钱豹 3 个物种看似有多个分支, 种内绝大部分个体形成一个大支, 另一些个体则形成几个小分支。但实际上, 这几个小分支里面所包含的都是 numts 序列。如果将 numts 误认为是 DNA 条形码对其进行分析, 则会得出错误的结论。

误把 numts 当 CO I 的情况可能是野猫鉴定失败的主要原因。在 NJ 树上, *F. s. catus* 1 这支野猫和荒漠猫 numts 以 100% 的置信度聚在一起, 二者间的遗传差异仅为 0.5%~0.8%, 并且这支中的个体同种内其他个体的差异最高达 10.3% (平均 7.9%)。虽然在 *F. s. catus* 1 中没有检测到终止密码子、插入/缺失, 但推测它极有可能是 numts; 首先, 猫科动物中不同物种的 numts 通常聚为一支 (如: 狮子、老虎、金钱豹), 而它同荒漠猫 numts 也高度相似聚在一起; 其次, 该序列在 GenBank 中 Blast, 同家猫 numts 序列相似度为 99%; 最后, 根据 Li 等 (1981) 的方法计算 mtDNA 发生核转移的时间, 发现该序列发生核转移的时间约 350 万年前, 该时间同狮子、老虎、金钱豹的 numts 发生时间相似 (220 万年~350 万年)。这些证据都从侧面暗示了它是 numts。

综上, DNA 条形码数据库中已有序列的准确性对基于该技术的物种鉴定至关重要, 而通用引物从总基因组 DNA 中扩增时, 可能得到 numts 序列而非 CO I 目的基因片段, 这一现象应该得到重视。研

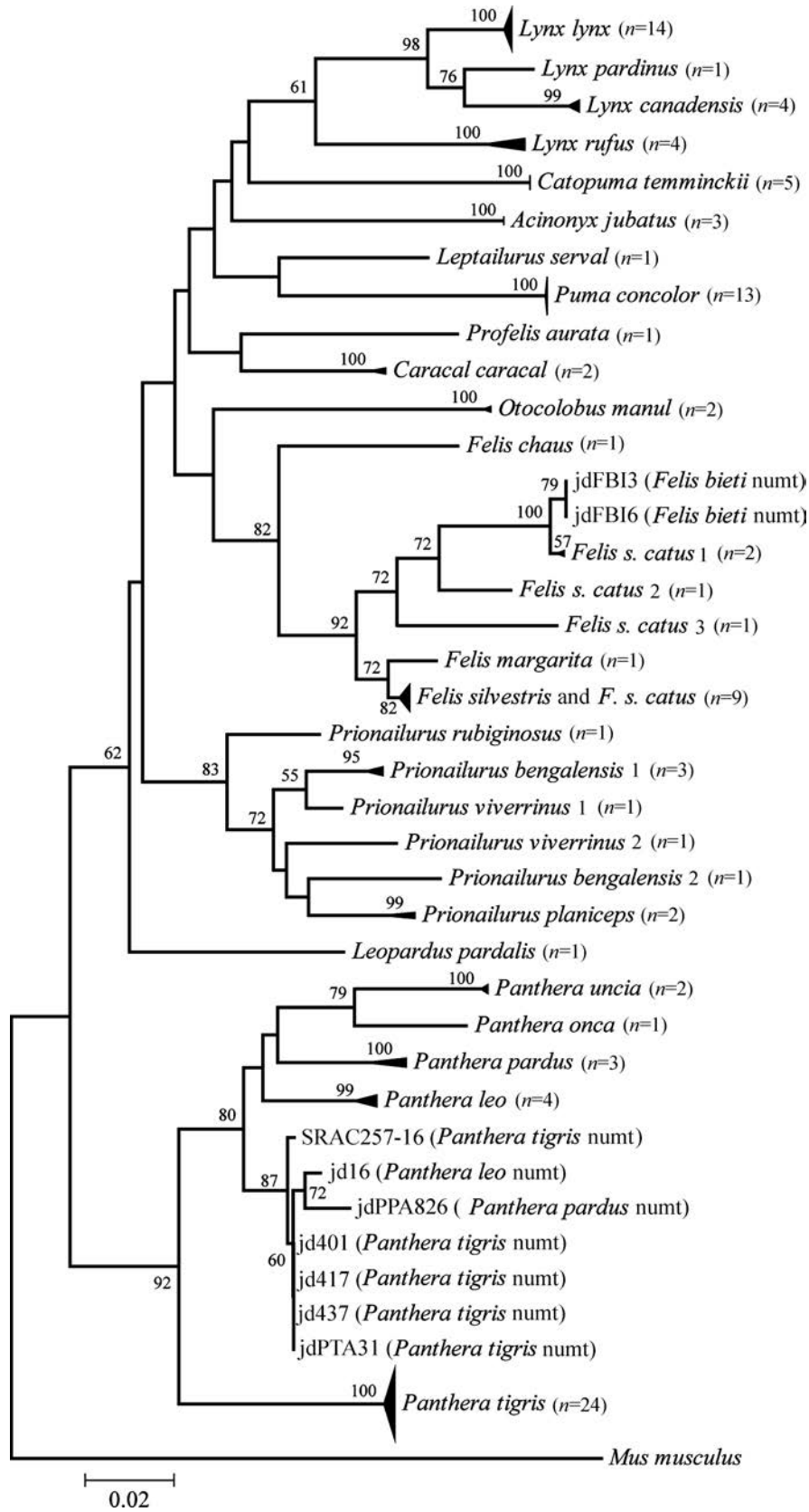


图 2 猫科动物 25 种 119 个样本的 COI 及 numts 序列所构建的 NJ 树

Fig. 2 NJ tree of 119 COI and numts sequences in 25 species of Felidae

假基因序列均在种名后标注有 numt, 仅显示高于 50% 置信度。

Pseudo sequences were labeled with "numt" after the species name; Bootstrap support values greater than 50% were shown on the branch.

究者如果不能正确区分 numts 和 mtDNA,把错误的序列关联物种信息并存入数据库,则会导致后续的各项工作中出现混乱和错误。研究表明该问题在猫科动物中普遍存在,其他种类的动物中是否也存在大量 numts,对条形码技术形成类似的干扰影响,有待进一步研究发现。

参考文献:

- 蔡延森,刘芹,王译伟,等. 2014. 四川宜宾产桃花水母的线粒体 *COI* 序列片段分析[J]. 泸州医学院学报, (1): 79-82.
- 蔡延森,张修月,岳碧松,等. 2009. 我国 8 种猛禽的 DNA 条形码技术研究[J]. 四川动物, 28(3): 334-340.
- 周志军,尚娜,常岩林,等. 2014. DNA 条形码揭示日本纺织娘 (*Mecopoda niponensis*) 个体内和个体间的序列变异[J]. 生物技术通报, (5): 129-136.
- Brown WM, Prager EM, Wang A, et al. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution [J]. Journal of Molecular Evolution, 18(4): 225-239.
- Cai YS, Zhang L, Wang Y, et al. 2016. Identification of deer species (Cervidae, Cetartiodactyla) in China using mitochondrial cytochrome c oxidase subunit I (mtDNA *COI*) [J]. Mitochondrial DNA Part A: the Journal of DNA Mapping Sequencing, and Analysis, 27(6): 4240-4243.
- Cai YS, Zhang L, Shen FJ, et al. 2011. DNA barcoding of 18 species of Bovida [J]. Chinese Science Bulletin, 56(2): 164-168.
- Hebert PDN, Cywinska A, Ball SL, et al. 2003. Biological identifications through DNA barcodes [J]. Proceedings of the Royal Society B-Biological Sciences, 270: 313-321.
- Hebert PDN, Stoeckle MY, Zemplak TS, et al. 2004. Identification of birds through DNA barcodes [J]. PLoS Biology, 2(10): e312.
- Kress WJ, Garcí'a-Robledo C, Uriarte M, et al. 2015. DNA barcodes for ecology, evolution, and conservation [J]. Trends in Ecology & Evolution, 30(1): 25-35.
- Li J, Cui Y, Jiang J, et al. 2017. Applying DNA barcoding to conservation practice: a case study of endangered birds and large mammals in China [J]. Biodiversity & Conservation, 26(3): 653-668.
- Li WH, Gojobori T, Nei M. 1981. Pseudogenes as a paradigm of neutral evolution [J]. Nature, 292(5820): 237-239.
- Lopez JV, Yuhki N, Masuda R, et al. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat [J]. Journal of Molecular Evolution, 39(2): 174-190.
- Mukherjee S, Appel A, Duckworth JW, et al. 2016. *Prionailurus viverrinus* [DB/OL]. The IUCN Red List of Threatened Species 2016: e. T18150A50662615.
- Wozencraft WC. 2005. Mammal species of the world: a taxonomic and geographic reference (3rd Ed.) [M]. Baltimore: Johns Hopkins University Press: 536-537.