

生命条形码新秀:DNA 微型条形码技术

程 鹏, 张爱兵
首都师范大学生命科学学院, 北京 100048

摘要: 近些年来, DNA 条形码技术为便捷的物种鉴定提供了很大的帮助, 但随着该技术的发展, 也出现了一系列的问题。微型条形码技术是作为 DNA 条形码技术的补充而出现的一项新兴技术, 具体是指通过通用引物扩增出比细胞色素 c 氧化酶 I 号基因全序列更短的一段序列, 并通过该序列进行物种鉴定、分类等研究工作。作为一项新兴技术, 其优点包括, 适用于部分降解的 DNA 样品的目的基因扩增, 能够很好地解决环境混合样品多样性的调查等。但是, 该技术所选 DNA 片段非常短, 因此标记序列包含的遗传信息有限, 在鉴定的精确度方面和 COI 全条形码存在一定的差距。本文在总结前人研究的基础上, 简要概述了微型条形码技术的优缺点, 并对其未来在害虫分子识别方面的应用做了初步探讨。

关键词: 微型条形码; 物种鉴定; DNA 降解; 环境混合样品

Notive barcode of life: DNA mini-barcode technology
Peng CHENG, Ai-bing ZHANG
College of Life Sciences, Capital Normal University, Beijing 100048, China

Abstract: In recent years, DNA barcode technology make great help in identification of species, but there are some problems with the development of the technology. Mini-barcode technology as a supplement to DNA barcoding has been proposed recently. This method amplifies a sequence which is much shorter than the cytochrome c oxidase subunit I sequence by universal primers to identify and classify species. As a new technology, its advantages include, such as amplifying the target gene for partly degenerated DNA samples, as well as a good solution to the diversity of environmental investigation in mixed samples and so on. However, the target gene of the new technology is a very short DNA fragment. Therefore, the genetic marker sequence contains limited information. So, there is a certain gap between the full barcode in the accuracy of identification. Here, based on previous studies, we make a brief summary of advantages and disadvantages of DNA mini-barcode and its potential applications, especially to identification of the invasive species.

Key words: mini-barcode; species identification; DNA degradation; environmental sample

由于资金的限制及年轻分类学家的匮乏, 分类学的发展在过去几十年里受到了严峻的考验, 曾一度有沉寂的趋势 (Mallet & Willmott, 2003), 但近年来又呈现出勃勃生机。其中主要是因分子生物学技术和因特网的介入而提出的 DNA 分类学和 DNA 条形码 (Hebert *et al.*, 2002)。DNA 条形码是指一段特殊的 DNA 序列, 既可以用于辨别物种, 还可以用于研究近缘种之间的进化关系 (Hebert *et al.*, 2003)。

在 Hebert 等人研究的基础上, 许多科学家接受了将细胞色素 c 氧化酶 I 号 (COI) 基因序列作为动物 DNA 条码的设想 (Tautz *et al.*, 2003), 这个序列也逐步成为了物种鉴定最常用的分子标记之一。

从此, 生命条形码工程便成为科学家们的一个宏伟计划, 即试图建立一个基于标准的分子序列的物种鉴定系统 (Frezal & Leblois, 2008)。但是, 随着 DNA 条形码技术研究的深入与推广, 人们发现其在实际工作中常常受到限制。如对待检标本要求高、在馆藏标本中的应用难度大等。DNA 微型条形码的出现, 在很大程度上弥补了以上传统条形码的缺点与不足, 其在物种鉴定上的使用也越来越广泛, 并取得了一系列的成果 (Miller *et al.*, 2005; 欧阳小艳等, 2007; Ward *et al.*, 2009)。本文就该技术的研究成果及优缺点做简要概述, 并对其将来可能应用的方向做初步的探讨, 以期对微型条形码技术的完善及其对外来入侵害虫的防治方面提供参考。

1 DNA 条形码的定义及微型条形码的研究成果

1.1 DNA 条形码的定义及其序列特征

DNA 条形码是指线粒体氧化酶 C 从 5'端开始的一段长度约 650 bp 的 DNA 序列 (Hebert *et al.*, 2002)。能够用于条形码的基因必须具备 2 个看似矛盾的特征:(1)相对的保守性,便于用通用引物扩增出来;(2)足够的变异性,以便将物种区别开来。综合基因序列的长度和进化速率 2 个条件,最终选定 COI 序列,即一段约 650 个碱基长度的片段作为条形编码基因。这是因为:(1)动物生命中绝大部分阶段都有 COI 基因序列;(2)大多数细胞中都有上百个线粒体,但只有一组染色体,因此等量的样品中,线粒体的 DNA 更容易被放大和使用;(3)与细胞核的 DNA 相比,线粒体 DNA 的突变速度是核 DNA 的 10 倍,意味着核 DNA 能够保留原来的变异而线粒体原来的变异会很快地丢失,使物种分离更准确;(4)线粒体的遗传方式属于母性遗传,COI 基因位于细胞线粒体中,因此只能从母体中遗传,这样基因重组的发生率低;(5)它还拥有蛋白编码基因所共有的特征,即密码子第三位碱基不受自然选择压力的影响,可以自由变异;(6)COI 基因在能够保证足够变异的同时又很容易被通用引物扩增,其 DNA 序列本身很少存在插入和缺失(即使有少数也主要分布于该基因的 3'端,对结果的分析不会造成很大的影响)。COI 基因序列现已成为鉴定动物的 DNA 条形码。

1.2 DNA 微型条形码的研究成果

DNA 微型条形码开始是针对博物馆标本提出的 (Hajibabaei *et al.*, 2006),近年来,中外学者对此做了一系列的尝试并取得了一定的成果。Hajibabaei *et al.* (2006)证明了可以从已经发生 DNA 降解的馆藏标本中扩增出 134 bp 的目的基因;同时通过对来自哥斯达黎加地区的寄生蜂进行试验,验证了 135 bp 的 COI 序列可以很好地应用于物种鉴定。尽管 COI 序列的内部切割思想在当时非常新颖并且对条形码技术的发展起到了很大作用,但其工作是基于计算机模拟,并未考虑扩增微型条形码的引物设计问题,缺乏实际证明。在此基础上,Meusnier *et al.* (2008)通过对现有 DNA 条形码数据库中的已测序列进行计算得到,随着被扩增序列长度的缩短,其引物的通用性增强,但缩短的序列鉴定物种

的能力会有一定的减弱。当 COI 序列长度大于 100 bp 时即可鉴定 90% 以上的物种;扩增 130 bp 的部分 COI 序列比扩增 650 bp 的 COI 序列容易;同时,他们将这段 130 bp 的条形码应用于鳞翅目蛾类馆藏标本的鉴定研究,结果证明了该条形码在馆藏标本鉴定中的正确性。从此,DNA 微型条形码技术逐渐走向成熟。Dubey *et al.* (2009)将复合引物 PCR 技术与微型条形码技术结合起来,直接将 PCR 的产物电泳条带 (380、265、130 bp) 用于印度濒危蛇类的鉴定。这在一定程度上利于新物种的发现,并且节约了大量成本。但是,PCR 条带始终无法直接反映 DNA 序列的内部信息,因此还需要提升测序技术。范京安等 (2009) 也将微型条形码技术用于海关检验中果实蝇的物种鉴定。其以一种生物信息学模拟与试验相结合的方法,进一步证明了微型条形码可以用于物种鉴定,并且运用统计学的方法从基因位点的统计数据上对微型条形码的应用提供了进一步的理论支持。

1.3 DNA 微型条形码的基本操作过程

DNA 微型条形码的基本操作过程主要包括:(1)样品的采集与处理;(2)基因组 DNA 的提取及电泳确认;(3)利用 COI 全条形码的传统引物扩增 COI 基因片段及电泳检验;(4)当步骤(3)确认无法扩增出完整的 COI 时,利用微型条形码的通用新型引物扩增部分 COI 基因片段及电泳确认;(5)PCR 产物的纯化与克隆;(6)测序和序列分析;(7)微型条形码数据库的构建以及在相关领域中的具体应用。

2 传统分类方法、DNA 条形码技术与微型条形码技术的比较

2.1 传统分类方法的贡献与不足

尽管传统的形态分类为分类学奠定了基础、提供了框架和依据,但其鉴定手段复杂且效率不高,对工作人员的经验和专业有相当高的要求。常规形态学鉴定方法有 4 个方面的缺陷:(1)由于物种表型的可塑性和遗传的可变性,容易导致不正确的鉴定;(2)无法鉴定许多群体中普遍存在的隐存分类单元;(3)受生物性别和发育阶段的限制,很多生物无法被鉴定;(4)虽然现代交互式鉴定系统是一个很大的进步,但它要求很高的专业技术,一旦操作不正确则很容易导致错误的鉴定结果。“形态学鉴定的局限性和不断缩减的分类学家队伍,使分类

学的发展面临巨大的挑战,亟需一种快捷方便的物种鉴定方法产生”(Hebert *et al.*, 2003)。

2.2 DNA 条形码技术的优点与不足

首先,DNA 条形码技术的运用改善了分类学多年不够活跃给生物科学带来的负面影响。如资源投入不足、合格的分类工作者较少等(Gotelli, 2004)。DNA 条形码技术既准确、快捷,又降低了对工作者经验的要求,使得一些难度较大或工作量较大的物种鉴定工作变得相对容易。其次,DNA 条形码对保护生物学、生物进化历史研究的发展起着至关重要的作用。它将完成一些传统形态学鉴定手段无法完成的工作,如鉴定生物的卵和幼体、动物或植物的寄生虫,以及解决形态学手段难以攻克的隐存种问题,或者根据对动物肠道包含物或排泄物的分析来解决食物链问题。这个计划本身的发展,要求一系列更快、更好、更廉价技术的支持,这势必会推动相关分子生物学技术的进步,从而让其他相关的生物科学受益。此外,DNA 条形码技术也可以为入侵生物的预防和控制研究提供帮助(Armstrong & Ball, 2005)。

同时,DNA 条形码技术也存在一些问题。如DNA 的易降解性大大制约了条形码技术的适用范围,而且该技术在馆藏标本以及福尔马林浸泡的标本上应用也达不到预期的效果。对于线粒体变异速率的要求,即适度的变异速率也是对DNA 条形码的制约因素之一,如植物的线粒体具有极端的进化保守性,即其DNA 异质化程度相当低,所以,COI 序列不能作为鉴定植物的条形码序列。此外,由于不同的引物对片段的扩增有偏向性(Bellemain *et al.*, 2010),DNA 条形码技术在新物种的确定以及环境混合样品的分析中也存在一些问题。如DNA 条形码技术一般要求完整的COI 序列长度为650 bp,这使得引物的通用性无法满足不同物种的要求(Meusnier *et al.*, 2008)。因此,在建立国际DNA 条形码数据库的进程中,必须去攻克这些随之而来的问题,否则,DNA 条形码生命数据系统(BOLD)永远不可能成为一个普适性的系统(Frezal & Leblois, 2008)。

2.3 微型条形码技术的优势与缺陷

随着生命条形码工程的不断发展,一系列的问题需要新的技术来完善和补充,从而使DNA 条形码能发展得更好。正如条形码的倡导者Hebert 所言“微型条形码系统显著地扩大了DNA 条形码技术的

应用范围,我们已经验证了可以从已经降解的馆藏标本中扩增出微型条形码;另外,配合高通量测序技术,微型条形码引物的通用性使得研究环境混合样品成为了可能,一种廉价而全面的研究生物多样性的方法成为了现实”(Meusnier *et al.*, 2008)。因此,微型条形码凭借其引物的通用性强、扩增能力高,且较强的物种分类能力,得到越来越广泛的应用。如Dubey *et al.* (2010)通过蛇残片微型条码的检测,来防止一些非法商人对珍稀蛇类的捕杀。

由于DNA 微型条形码技术现阶段主要还是基于COI 序列中的一段,许多COI 全条形码技术中的缺陷同样存在于微型条形码分类技术中。第一,采样不足在很多时候会造成条形码库出现鉴定盲区(Meyer & Paulay, 2005),因此,在采样过程中,应尽可能采到绝大部分已经确认存在的种并保证其足够的个体数量。这一点在实际工作中往往会成为建立DNA 条形码数据库的一个制约。第二,由于线粒体DNA 的遗传主要由母系决定,线粒体的母系遗传成为分析COI 序列过程中需要考虑的因素。已有研究证明,仅仅使用COI 序列在一些时候不能正确地鉴定物种的种类(Burns *et al.*, 2007)。因此,在微型条形码技术的使用中,同样要考虑这一点。第三,对于10 倍阈值法则的争议。在2007 年Hebert *et al.* 对北美地区260 种鸟的COI 基因进行了研究(Kerr *et al.*, 2007)。结果发现,其近源种间COI 序列差异(7.0% ~ 8.0%)约比种内(0.3% ~ 0.4%)大20 倍。因此认为,大多数北美地区的鸟类物种均可用该序列进行区分,并将10 倍的种内变异(约3.0%)作为标记物种遗传分化的“标准序列阈值”,低于此数值的物种之间即使存在生物学差异也被认为是同种。许多科学家认为,这个法则存在着很大的主观性,并且许多情况与实际不一致(Meier *et al.*, 2008)。第四,假基因在COI 基因PCR 过程中的出现也是一个值得考虑的问题(Song *et al.*, 2008)。由于微型条形码引物的通用性高,在实际工作中应尤为重视。第五,许多学者对单序列条形码的信度问题始终存在着争议,因此,许多研究选择采用多位点基因序列分析并结合其形态信息,记录更多的生态信息,综合这些信息和序列比对的结果来进行物种鉴定。此外,微型条形码较短,包含的序列信息明显低于全条形码,因此,在实际运用过程中,虽然鉴定物种的正确率符合预期要求,但与全条形码鉴定物种的正确率相比,还有微小的差距(Meusnier *et al.*, 2008)。

3 展望

微型条形码技术虽然在发展中遇到一系列的问题,但随着研究的深入,该技术将得到不断完善。国际生物条形码计划倡导全球统一行动,现在一个分子生物学实验室平均每年可得到 1000 个以上物种的目的片断序列;并且随着 DNA 测序技术的革新,配合微型条形码技术,不仅使 DNA 鉴定的成本大大降低,而且在一定程度上解决了 DNA 降解问题,条形码数据为馆藏标本的使用提供了空间,提高了生命条形码工程的可行性以及完成速度。生命条形码工程作为最直接的信息资源,将成为全球生物鉴定系统(global bioidentification system, GBS)的基础。若该数据库被全球生物多样性信息机构(Global Biodiversity Information Facility)或生物物种基金(All Species Foundation)利用,网络信息资源长期缺乏的问题将得到解决。Hebert *et al.* (2003) 预计整个计划的实施将花费 10 亿美元,远远低于人类基因组计划和国际空间站的成本。在保护生物学方面,DNA 微型条形码技术将对生物多样性研究的发展起到至关重要的作用,主要体现在可以快速有效地完成环境混合样品多样性调研。在外来物种入侵防治方面,DNA 微型条形码技术为新种和隐存种的发现提供了必不可少的佐证(Hajibabaei *et al.*, 2006),因此,运用该技术监控一个区域是否发生外来物种入侵将成为可能。

总之,DNA 微型条形码技术作为一种与传统 DNA 条形码技术结合的新兴技术,必将推动分类学乃至整个生命科学的发展。

参考文献

- 范京安,陈世界,温演庆,顾海丰,刘伟,曾晓茂,何万兴. 2009. DNA 条形码识别 VI: 基于微型 DNA 条形码的果实蝇物种鉴定. 应用与环境生物学报, 15(2): 215–219.
- 欧阳小艳,莫帮辉,余华丽,何建伟,赵明,曾晓茂. 2007. DNA 条形码识别——DNA 条形码与 DNA 芯片识别蚊媒准确性的比较. 中国国境卫生检疫杂志, 30(6): 349–352.
- Armstrong K F and Ball S L. 2005. DNA barcodes for biosecurity: invasive species identification. *Proceedings of the Royal Society*, 360: 1813–1823.
- Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P and Kausrud H. 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology*, 10: 189.
- Burns J M, Janzen D H, Hajibabaei M, Hallwachs W and Hebert P D N. 2007. DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *Journal of the Lepidopterists' Society*, 61(3): 138–153.
- Dubey B, Meganathan P R and Haque I. 2009. Multiplex PCR assay for rapid identification of three endangered snake species of India. *Conservation Genetics*, 10: 1861–1864.
- Dubey B, Meganathan P R and Haque I. 2010. DNA mini-barcoding: An approach for forensic identification of some endangered Indian snake species. *Forensic Science International Genetics*, doi:10.1016/j.fsigen.2010.03.001
- Frezal L and Leblois R. 2008. Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution*, 8: 727–736.
- Gotelli N J. 2004. A taxonomic wish-list for community ecology. *Proceedings of the Royal Society*, 359: 585–597.
- Hajibabaei M, Smith M A, Janzen D H, Rodriguez J J, Whitfield J B and Hebert P D N. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, 6: 959–964.
- Hebert P D N, Cywinska A, Shelley L and Dewaard J R. 2002. Biological identifications through DNA barcodes. *Proceedings of the Royal Society*, 27: 313–321.
- Hebert P D N, Ratnasingham S and Dewaard J R. 2003. Barcoding animal life: cytochrome c oxidize subunit 1 divergences among closely related species. *Proceedings of the Royal Society*, 270: 96–99.
- Kerr K C R, Stoeckle M Y, Dove C J, Weigt L A, Francis C M and Hebert P D N. 2007. Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, 7: 535–543.
- Mallet J and Willmott K. 2003. Taxonomy: renaissance or tower of babel? *Trends in Ecology and Evolution*, 18: 57–59.
- Meier R, Guanyang Z and Ali F. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the “Barcoding Gap” and leads to misidentification. *Systematic Biology*, 57: 809–813.
- Meusnier I, Singer G A C, Hebert P D N and Hajibabaei M. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, 9: 214.
- Meyer C P and Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, 3: 2229–2238.
- Miller S E and Schindel D E. 2005. DNA barcoding a useful tool for taxonomists. *Nature*, 435: 17.
- Song H J, Buhay J E, Whiting M F and Crandall K A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, 105: 13486–13491.
- Tautz D, Arctander P, Minelli A, Thomas R H and Vogler A P. 2003. A plea for DNA taxonomy trends. *Ecology and Evolution*, 18(2): 70–74.
- Ward R D, Hanner R and Hebert P D N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, 74: 329–356.