

DNA 条形码技术的研究进展及其应用

彭居俐^{1,2} 王绪桢¹ 何舜平¹

(1. 中国科学院水生生物研究所, 武汉 430072; 2. 中国科学院研究生院, 北京 100039)

THE PROGRESS AND APPLICATION OF DNA BARCODING

PENG Ju-Li^{1,2}, WANG Xu-Zhen¹ and HE Shun-Ping¹

(1. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072; 2. Graduate School of the Chinese Academy of Sciences, Beijing 100039)

关键词: DNA 条形码技术; 分类学; 物种鉴定

Key words: DNA Barcoding; Taxonomy; Species identification

中图分类号: Q-33 文献标识码: A 文章编号: 1000-3207(2008)06-0916-04

DNA 条形码技术 (DNA Barcoding) 是通过对一个标准目的基因的 DNA 序列进行分析从而进行物种鉴定的技术。这个概念的原理与零售业中对商品进行辨认的商品条形码是一样的。简单地说, DNA 条形码技术的关键就是对一个或一些相关基因进行大范围的扫描, 进而来鉴定某个未知的物种或者发现新种^[1-3]。自从提出 DNA 条形码的概念以来, 这种新兴分类学技术已经引起了越来越多的生物学家的关注。DNA 条形码技术是分类学中辅助物种鉴定的新技术, 它代表了生物分类学研究的一个新方向^[4], 因此它在生态、环境、食品等诸多领域都将会有广泛的应用^[5]。本文概括综述了 DNA 条形码技术的发展历史、原理与操作, 分析了其在生物分类中的应用及应用上的优势与限制, 对 DNA 条形码技术在鱼类学研究的意义与可行性进行了探讨。

1 DNA 条形码技术的发展历史

2003 年, Herbert 研究发现利用线粒体细胞色素 c 氧化酶亚基 (Mitochondrial cytochrome c oxidase subunit, CO₁) 基因一段长度为 648bp 的片段, 能够在 DNA 水平上成功的区分物种, 并且认为利用 CO₁ 基因从分子演化的角度, 将提供一种快速、简便、可信的分类方法^[3]。这种方法逐步发展起来并被研究者命名为 DNA 条形码技术。

2003 年于冷泉港召开的两个小型研讨会上, 分类学、分子生物学和生物信息学的专家提出利用特定 DNA 序列实现物种鉴定的目标。2004 年, 由 Alfred Sban 基金会赞助, 在美国华盛顿特区举办了一个关于 DNA 条形码的大型研讨会,

此次会议创立了“生命条形码联盟”(CBOL, the Consortium for the Barcode of Life), 大部分国家的自然历史博物馆、标本馆以及研究机构和私人机构等都加入了这个组织。

DNA 条形码技术的迅速发展吸引了研究者的广泛兴趣。2004 年秋, 美国国立生物技术信息中心 (NCBI) 与生命条形码联盟 (CBOL) 签署合作, 物种条形码的标准 DNA 序列及其相关数据将存档于 GenBank (www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html)。此前, Genbank 虽然有很多医药和其他领域的团队在扩充数据, 但是迄今为止, 除了在系统发育研究上广泛应用的 DNA 分子标记外, 从来没有与标准分类学相结合的 DNA 序列数据。

2005 年 2 月, 伦敦举办了第一届全球 DNA 条形码会议。大会对 DNA 条形码的分类概念与思想、实验技术的细节分析以及资料库建立等议题进行了讨论。目前, 全球有很多个不同项目在不同的类群中进行, 最终的目的是联合各个类群的 DNA 条形码数据库组建一个全球生物 DNA 条形码数据库, 此数据库将设置在 Genbank 中, 是一个公众可以登陆的 DNA 序列数据库。

2 DNA 条形码技术的原理及操作步骤

DNA 条形码技术有一种潜在的可能性, 能够成为地球上已被认知的 1 千万多种真核生物的有效鉴定方法^[6]。例如, 一个长度 600bp 的蛋白质编码基因的核苷酸片段在第三密码子位点含有 200 个核苷酸, 这些位点上发生的替代经常都是中性选择, 并且突变大多是通过随机漂变在种群中固

定下来。假设在一组物种中第三位点的核苷酸全部是 AT 或全部是 GC,即在其 200 多个位点上只有两种可能的核苷酸,那么就仅第三密码子位点的变化而言能够产生 2^{200} 或 10^{60} 的可能序列^[7]。如此大数目的可能序列为 DNA 条形码技术进行众多生物的物种鉴定提供了一个最基本的条件。

鉴于 DNA 条形码技术的广泛应用前景,研究者们通过归纳总结得出一个完善的 DNA 条形码标准数据库包含以下操作步骤:物种个体的保存,因 DNA 提取的需要个体应保存在浓度为 95% 的酒精中;物种个体的标签,DNA 条形码技术必须将序列与个体联合对应起来,其中还应包括个体的相关信息,如采集者、分类确认、日期和全球地理位置坐标等;提取样本 DNA;目的基因序列扩增,DNA 条形码标准基因的扩增,如在动物 DNA 条形码研究中一般选取的标准基因序列为 *CO* 基因 5 端的长度为 600bp 左右的片段;数据库结果录入,数据库所舍括的数据应包含物种个体及其标签,以及相对应的标准序列。

3 DNA 条形码技术的研究进展

自 Hebert 对 DNA 条形码技术开展研究后,越来越多的研究报告都支持 DNA 条形码技术能够准确的进行物种鉴定。为了探讨 *CO* 基因在整个动物界中辨别物种的能力,Hebert, *et al.* 对动物界中除刺胞动物门以外的所有动物门,共 11 个门 13320 个物种的 *CO* 基因序列进行分析,发现其序列间的差异能够很好的区分所有研究物种,并认为在动物界中 *CO* 基因是合适的 DNA 条形码标准基因^[3]。Vences, *et al.* 于 2005 年探讨了 DNA 条形码对两栖爬行类物种的鉴定能力,结果显示选用的 *CO* 基因能够准确的辨别各物种^[8]。此外 Vences, *et al.* 选用 16S rRNA 基因进行条形码分析,发现在两栖爬行类中 16S rRNA 基因具有较强的引物适应性,并且序列相对保守^[9]。对北美哥斯达黎加地区 1000 多种鳞翅目昆虫生物多样性的调查中,Janzen, *et al.* 利用 DNA 条形码的 *CO* 基因序列建立的一个分类学平台,能够鉴定物种以及发现这一区域存在的模糊种等,并且能准确鉴定 97% 的物种,其余不能被确认的物种是亲缘关系极近的物种或形态定义模糊的物种,因此其研究认为 DNA 条形码能够很好的协助一地区生物多样性调查^[10]。Hebert, *et al.* 利用 *CO* 基因序列对北美 260 余种鸟类进行了有效区分^[2]。利用 DNA 条形码能够对鱼类以及热带鳞翅目昆虫进行有效的物种鉴别^[11,12]。对于国际上分类混乱的两栖爬行类 DNA 条形码技术,不仅能够准确地鉴别出各物种,而且还能无误地识别出其不同地理类群的变异^[8]。对动物类群的研究表明,*CO* 基因一段长约 648bp 的序列能够准确的进行物种鉴定。陆生植物由于质体 DNA 的特殊性质,DNA 条形码研究一直难以开展。Chase, *et al.* 利用陆生植物在系统发育分析中常用的分子标记即质体 *rbcL* 基因和 *ITS* 基因进行物种鉴定分析,得到的序列变异度能够清楚的辨认物种,认为陆生植物能够利用 DNA 条形码技术可靠的进行物种鉴定^[13]。红藻门是一类难以运用传统分类手段进行物种鉴定

的类群,它们的形态结构和生理特征简单,并且有形态、生理上的趋同性和表型上高度的可变性。Saunders 选用红藻门中分类地位明确的物种检测 *CO* 基因对此类群物种的鉴别能力,研究证实 DNA 条形码能够准确的区分红藻门物种^[14]。

随着 DNA 条形码技术的推进,各类物种信息和 DNA 条形码数据在网络数据库上相继发布。最新统计显示,已经发布的 DNA 条形码序列有 168719 种,其中完整的物种 DNA 条形码序列(包含准确的物种信息和条形码序列信息)有 19163 个物种。很多国际组织自行建立的 DNA 条形码数据库构成整个 DNA 条形码项目的重要组成部分,有较大影响和规模较大组织包括 All Lep's, FISH-BOL, Canadian Fauna 和 Birds 等。All Lep's 组织主要对澳大利亚、加拿大、哥斯达黎加和美国等四国的鳞翅目共 2.5 万个物种进行 DNA 条形码数据库的构建,已经取得 7512 个物种的完整 DNA 条形码数据。FISH-BOL 组织旨在建立所有鱼类将近 3 万个种的 DNA 条形码数据库,重点对 1.5 万种海鱼的 DNA 条形码数据进行收集,目前已经获得 2538 种鱼类的 DNA 条形码数据。Canadian Fauna 组织的最终目标是建成加拿大的所有生物种的 DNA 条形码数据库,已经获得 2133 个物种的 DNA 条形码数据。Birds 组织计划在 5 年内完成所有鸟类约 1 万种的 DNA 条形码数据库的构建,1233 种鸟类的 DNA 条形码数据已经得到。

尽管 DNA 条形码技术能够准确的进行物种鉴定,但是确立 *CO* 基因在物种内变异的衡量标准还需要大量的工作。目前的研究支持利用线粒体基因的生物地理学分析给种内变异确定一个普遍的标准。因此,第一,种内差异很少有大于 2%,一般是小于 1%;第二,如果检测到高的变异时,这些个体很可能是处在不同的地理隔离种群,反映出在隔离种群形成前基因库的片段^[15]。总的说来,种内的序列变异普遍的要比种间的变异小得多。在脊椎动物中,近缘种间 *CO* 基因的差异通常要比其他线粒体基因大,例如细胞色素 *b* (Cyt *b*)。因此,利用 *CO* 基因序列的 DNA 条形码技术能够作为物种鉴定的有效手段。

地球上的 1 千多万物种仅有 15% 被分类学家描述,因此还有大量的没有被描述过的新种可以通过 DNA 条形码挖掘出来。但是这些新种的命名不能仅仅依靠 DNA 条形码的结果^[16-19],必须依靠被鉴定的标本个体、个体高清晰的图像照片和采集地等信息对物种进行标准的林奈命名^[20-22]。在动物学研究中,利用线粒体 DNA 获取的生物地理和系统分析证据经常与形态学不一致。这种差异将促使研究者对形态以及生态特征重新进行评定,甚至可能影响到分类上的修订。近期认为线粒体 DNA 序列的差异可以作为不同物种划分的首要标准^[23]。但是这个标准必须与其他生态形态等传统鉴别特征相联合^[24]。

4 DNA 条形码技术的优势

作为一种标准的物种鉴定方法,由于其快速简便的操作

过程,DNA条形码技术有巨大的潜力能够广泛的进行科学应用,在保护生物学,包括生物多样性调查等领域尤其显著。当传统的分类学受到阻碍时,这种技术就更可以发挥其优势^[25]。另外,利用DNA条形码技术为分类学家发现新种提供了一个契机^[26]。

DNA条形码技术的优势主要体现在以下几个方面:

(1)利用小块生物体组织进行鉴别。一旦生命条码数据库建立起来,条形码技术能够快速的从部分零散组织中鉴别出物种,例如通过鉴别胃里的食物小块能够帮助重建食物链循环,还有土壤层里植物根的本可被准确鉴定。(2)能够在生物的不同生长阶段鉴别物种。动植物在不同生长发育时期的不同形态都能够利用DNA条形码技术鉴别出物种。(3)准确地辨别形态相似性很高的物种。部分不同物种有高度相似的外形,还有一些具有潜在危险的物种常伪装成不具有威胁性的物种。DNA条形码技术就能够很好地区分此类物种。(4)减少鉴别的模糊性。DNA条形码技术利用碱基ACTG组成的序列使物种鉴别数字化。这种数据库使得全球范围内的现生物种能够很简单准确的鉴定出来。(5)鉴定过程更加迅捷。地球上的已经被分类专家们鉴别出的种类大概就有2百万多种^[6],但这些物种仅仅是动植物界的一小部分。在今后的鉴定工作中,分类学家依据DNA条形码数据库来鉴定物种或者发现新种,这将使分类工作变得更加迅捷。(6)鉴定过程无限制。一个标准DNA条形码数据库的建立,可使各种物种的辨认成为可能,如濒危物种或常见种、土著物种或入侵种等,也使更多的无分类学专业知识的人能够准确的辨认出某个物种。(7)在分类领域开展了用电子技术处理的新方向。条形码技术使生物鉴定与DNA测序的前沿、电子技术上的便携性以及计算机的信息储存功能相结合。(8)对生命之树新的贡献。条形码数据库提供的大量的遗传信息可以帮助构建地球上所有生物的生命之树。此外对新近发现的新种也可以通过其条形码数据来追寻它在生命之树中的系统位置。(9)展示出收藏标本的价值。通过将博物馆、标本馆、动物园、植物园以及其他机构收藏的物种个体的条形码资料进行汇编,将进一步促进这些组织机构在保护地球生物的多样性的领域的工作。(10)完善生命的百科全书。将条形码数据库与物种标本个体以及物种相关信息联合起来,将使大众更好地了解生物。

5 DNA条形码技术应用的局限性

在整个基因组中,有大量的基因序列能够满足我们一个或多个要求^[27]。作为DNA条形码技术的标准基因,一方面应该足够保守(其引物有强的保守性,能够进行大范围的物种扩增),另一方面应该有足够的变异来区分不同物种DNA序列从而进行鉴定。但是理论上不存在一种普遍适用的DNA条形码基因,因为任何一个基因都不可能在各种生物中都保守,同时又包含足够的序列变异信息来进行物种的辨别。因此在鉴定不同的物种类群时需要确定不同的目的基因。在珊瑚虫、水母和海葵中,由于存在线粒体DNA修复体

系,线粒体基因显然不适合作为条形码技术的标准基因。另外,植物的线粒体DNA因杂交和基因渗入而变异很小;真菌线粒体DNA含有内含子。对这些特殊类群的标准条形码基因的选择还有待进一步探讨^[28]。

目前,DNA条形码技术的应用主要是对物种进行鉴定。然而DNA介导的物种鉴定重点在于种内和种间遗传差异的限定。种内和种间变异范围都不明确,而且不同物种的变异范围也不相同。就线粒体DNA而言,种内差异通常小于种间差异,但是也有一些特例^[29]。不同种系间线粒体基因水平转移,近缘物种共享线粒体多态性以及同属物种间的线粒体DNA修剪机制,都将造成近期快速分化物种和通过杂交产生的新种在鉴定上的困难。此外,同属物种间较低的差异也会带来在物种分类鉴定上的疑虑。

6 DNA条形码在我国鱼类分类学的意义及可行性

由于地理环境和地区气候的巨大差异,我国的鱼类资源拥有极其丰富的生物多样性。据最新统计,淡水鱼类约有1050种^[6],它们是淡水生物资源的重要组成部分。其中,鲤科鱼类将近占一半左右,约132属532种,特有种384个,约占我国鲤科鱼类总数的74%。这些生物资源不仅是系统发育和生物地理学的特殊珍贵的研究材料,也是非常重要的生物资源。但最近几十年来,许多种类成了濒危物种,大约10%共92种中国淡水鱼类是濒危物种。因此建立一个我国鱼类的DNA条形码数据库是非常必要的,这对于保存我国鱼类尤其是濒危鱼类的基因资源和物种信息等都是至关重要的。

对于我国生物多样性极高的鱼类资源而言,研究者可以根据馆藏鱼类资源来补充采集新鲜样本,构建一个完整的包含物种标准基因序列、物种个体图片、采集地点经纬度、采集者、采集时间和分类地位等重要信息的我国鱼类DNA条形码数据库。在我国鱼类DNA条形码数据库构建的同时,研究者也应该分析所选的DNA条形码标准基因对于我国鱼类的适用性,包括是否能够在不同物种中显示出DNA序列的差别,以及这些序列差异程度能否准确清晰的辨别各个不同物种等。无论是从DNA条形码数据库的应用价值出发,还是从其科研角度来看,这种标准数据库的建立将给我国科研和生物资源商业领域带来便利。

大量的研究结果都显示COI基因能够在动物类群中很好的区分各物种,所以可以选用COI基因作为标准条形码基因构建鱼类DNA条形码数据库。我们也应该对这个基因在此生物类群中的鉴别能力先进行一番评估,制定COI基因在不同物种间序列变异度的衡量标准。通常,我们可以大概的确定不同物种间的最小基因变异在一个特定的范围。但是对于一些特殊的类群,例如快速进化的类群,划定的种间最小变异值,应该在标准值范围内有一定的波动。因此在构建不同物种类群的DNA条形码数据库时,要充分结合所研究类群的物种形成模式和进化模式,确定可行的路线和标准,使其所建立的数据库可以进行准确可靠的应用。

参考文献:

- [1] Hebert P D N, Cywinska A, Ball S L, *et al* Biological identification through DNA barcodes [J]. *Proc R Soc Lond B Biol Sci*, 2003, **270**: 313—321
- [2] Hebert P D N, Stöckle M Y, Zamlak T S, *et al* Identification of birds through DNA barcodes [J]. *PLOS Biol*, 2004, **2**(10): 1657—1663
- [3] Hebert P D N, Ratnasingham S, deWaard J R. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species [J]. *Proc R Soc Lond B Biol Sci*, 2003, **270**: 96—99
- [4] Xiao J H, Xiao H, Huang D W. DNA: barcoding: new approach of biological taxonomy [J]. *Acta Zoologica Sinica*, 2004, **50**(5): 852—855
- [5] Janzen D H. Now is the time [J]. *Phil Trans R Soc B*, 2004, **359**: 731—732
- [6] Wang X Z. Molecular phylogeny of the East Asian cyprinids (Pisces: Cypriniformes) [M]. Thesis for Doctor of Science. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 2005
- [7] Stöckle M. Taxonomy, DNA, and the barcode of life [J]. *Bio-science*, 2003, **53**(9): 2—3
- [8] Vences M, Thomas M, Bonett R M, *et al* Deciphering amphibian diversity through DNA barcoding: chances and challenges [J]. *Phil Trans R Soc B*, 2005, **360**: 1859—1868
- [9] Vences M, Thomas M, Meijden A V, *et al* Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians [J]. *Frontier in Zoology*, 2005, **2**(5): 1—12
- [10] Janzen D H, Hajibabaei M, Burns J M. Wedding biodiversity inventory of a large and complex lepidoptera fauna with DNA barcoding [J]. *Phil Trans R Soc B*, 2005, **360**: 1835—1845
- [11] Ward R D, Zamlak T S, Innes B H, *et al* DNA barcoding Australia's fish species [J]. *Phil Trans R Soc B*, 2005, **360**: 1847—1857
- [12] Hajibabaei M, Janzen D H, Burns J M, *et al* DNA barcodes distinguish species of tropical Lepidoptera [J]. *PNAS*, 2006, **103**: 968—971
- [13] Chase M W, Salamin N, Wilkinson M, *et al* Land plants and DNA barcodes: short-term and long-term goals [J]. *Phil Trans R Soc B*, 2005, **360**: 1889—1895
- [14] Saunders G W. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications [J]. *Phil Trans R Soc B*, 2005, **360**: 1879—1888
- [15] Avise J C, Walker D. Pleistocene phylogeographic effects on avian populations and the speciation process [J]. *Proc R Soc Lond B Biol Sci*, 1998, **265**: 457—463
- [16] Weckstein J D, Zink R M, Blackwell-Rago R C, *et al* Anomalous variation in mitochondrial genomes of White-crowned (*Zonotrichia leucophrys*) and Golden-crowned (*Z. atricapilla*) sparrows: Pseudogenes, hybridization, or incomplete lineage sorting [J]. *Auk*, 2001, **118**: 231—236
- [17] Lovette I J, Bermingham E. Mitochondrial perspective on the phylogenetic relationships of the *Panula* wood-warblers [J]. *Auk*, 2001, **118**: 211—215
- [18] Bensasson D, Zhang D X, Hartl D L, *et al* Mitochondrial pseudogenes: Evolution's misplaced witnesses [J]. *Trends Ecol Evol*, 2001, **16**: 314—321
- [19] Ballard J W O, Whitlock M C. The incomplete natural history of mitochondria [J]. *Mol Ecol*, 2004, **13**: 729—744
- [20] Johnson N K, Cicero C. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds [J]. *Evolution*, 2004, **58**: 1122—1130
- [21] Gregory T R. DNA barcoding does not compete with taxonomy [J]. *Nature*, 2005, **434**: 1067
- [22] Ebach C M, Holdrege C. DNA barcoding is no substitute for taxonomy [J]. *Nature*, 2005, **434**: 697
- [23] Wiens J J, Penkrot T A. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*) [J]. *Syst Biol*, 2002, **51**: 69—91
- [24] Bensasson D, Zhang D X, Hartl D L, *et al* Mitochondrial pseudogenes: Evolution's misplaced witnesses [J]. *Trends Ecol Evol*, 2001, **16**: 314—321
- [25] Witt J D S, Therloff D L, Hebert P D N. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation [J]. *Mol Ecol*, 2006, **15**: 3073—3082
- [26] Stöckle M. Taxonomy, DNA and the barcode of life [J]. *Bio-science*, 2003, **53**: 2—3
- [27] Savolainen V, Cowan R S, Vogler A P, *et al* Towards writing the encyclopedia of life: an introduction to DNA barcoding [J]. *Phil Trans R Soc B*, 2005, **360**: 1805—1811
- [28] Moritz C, Cicero C. DNA barcoding: promise and pitfalls [J]. *PLOS Biol*, 2004, **2**(10): 28—29
- [29] Avise J C. Phylogeography: The history and formation of species [M]. Cambridge (Massachusetts): Harvard University Press 2000, 447