

# 生物分类学的新动向——DNA 条形码\*

肖金花<sup>1,2</sup> 肖 晖<sup>1</sup> 黄大卫<sup>1,3\*\*</sup>

1. 中国科学院动物研究所, 北京 100080
2. 中国科学院研究生院, 北京 100039
3. 山东农业大学植物保护学院, 山东 泰安 271018

**摘要** 过去的一年中, DNA 条形码 (DNA Barcoding) 成为生物分类学中引人注目的新方向。DNA 条形码, 根据对一个统一的目标基因 DNA 序列的分析, 达到物种鉴定的目的, 它操作的简便性和高效性将以我们无法想象的速度加快物种鉴定和进化历史研究的步伐, 但国际上对此的争论也不少。本文综述了 DNA 条形码的原理、操作过程及最新进展, 讨论了其可能存在的问题 [动物学报 50 (5): 852 - 855, 2004]。

**关键词** 细胞色素 c 氧化酶亚单位 (COI) DNA 条形码 DNA 分类学

## DNA barcoding: new approach of biological taxonomy\*

XIAO Jir-Hua<sup>1,2</sup>, XIAO Hui<sup>1</sup>, HUANG Da-Wei<sup>1,3\*\*</sup>

1. Institute of Zoology, The Chinese Academy of Sciences, Beijing 100080, China
2. Graduate School of the Chinese Academy of Sciences, Beijing 100039, China
3. College of Plant Protection, Shandong Agricultural University, Tai'an 271018, Shandong, China

**Abstract** DNA Barcoding is rising to be a charming approach in taxonomy last year. With the sequence analysis of a uniform target gene, it can enable species identification. Being simple and highly efficient, it can accelerate the process of species identity and the research on evolutionary history at a speed that we can not imagine. But it is still in dispute internationally. We here summarize its principles, manipulation and recent advances, with a discussion of the potential problems [*Acta Zoologica Sinica* 50 (5): 852 - 855, 2004].

**Key words** Cytochrome c oxidase subunit (COI), DNA Barcoding, DNA Taxonomy

由于资金不足以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 and Willmott, 2003), 但在最近一年多时间里突然又呈现出勃勃生机, 让人看到了新的希望。其原因之一就是由于分子生物学技术和因特网的介入而提出的 DNA 分类学 (DNA Taxonomy) 和 DNA 条形码 (DNA Barcoding) (Hebert et al., 2003)。

Tautz 等人首先提出要用 DNA 序列作为生物分类系统的主要平台 (即 DNA Taxonomy) (Tautz et al., 2002; Tautz et al., 2003), 随后 Paul Hebert

等连续发表两篇文章, 他们考虑得更深入, 提出利用线粒体 COI 这一特定基因的特定区段来做 DNA 条形码的基础, 期待给所有生物种进行编码 (Hebert et al., 2003a, 2003b)。继而, 仅去年一年时间里, 在著名刊物上发表一系列相关的文章, 对此专题展开了热烈的讨论 (Blaxter, 2003; Mallet and Willmott, 2003; Pennisi, 2003)。

2003 年 3 月, 20 多位分类专家、分子生物学家和生物信息学家会聚美国冷泉港, 召开了题为 “TAXONOMY AND DNA” 的会议, 提出对全球所有生物种的某个特定基因进行大规模测序, 以期

2004-03-25 收稿, 2004-05-07 接受

\* 国家自然科学基金 (No. 30330090, 30370188)、中国科学院院长基金 (No. 1730730400019) 和国家基础科学人才培养基金 (No. NSFC-J0030092) 资助 [This research was funded by the grants from National Natural Science Foundation of China (No. 30330090), Presidential Foundation of Chinese Academy of Sciences (No. 1730730400019), and National Science Fund for Fostering Talents in Basic Research (No. NSFC-J0030092)]

\*\* 通讯作者 (Corresponding author). E-mail: huangdw@panda.ioz.ac.cn

实现物种鉴定的目标，进而推进生物进化历史的研究。5 月，海洋生物普查研究首次将 DNA 条形码技术用于其标本编码。6 月，美国软体动物学会 (American Malacological Society) 年会上提出“所有蜗牛行动” (Allsnails Initiative)，预计在未来 5 年时间里，利用线粒体 COI 和 16S 两个基因对所有现存腹足动物进行条形码编码，从而推动这一世界第二大纲动物的生物鉴定及生物学其它方面的研究。7 月，生物条形码网站 (<http://www.barcodinglife.com/>) 开通，为全球所有研究者提供有关生物条形码研究的信息。9 月，在冷泉港再次召开题为“Taxonomy, DNA and the Barcode of Life”的会议，对 DNA 条形码所有真核生物的科学性、社会利益有了更深入的讨论和确定，并且还提出了组织策略及国际生物条形码计划 (International Barcode of Life Project) 的发展蓝图 (Stoeckle, 2003)。同年 12 月，英国 MICROGEN 公司代表到 Paul Hebert 所在的 Guelph 大学进行实地考察，这对 DNA 条形码计划的推进起了重要的作用。同时，GenBank 提供的 COI 序列数迅速增长，尤其表现在除脊索动物之外类群里 COI 数量的剧增，目前脊索动物的分类基本上都已经完成，大量未知的生物都集中在体型较小的生物类群里 (Blaxter, 2003)。这些类群 COI 序列数的迅速增长将大大促进 DNA Barcoding 工作的开展。

## 1 DNA 条形码编码的意义

条形码技术 (Bar code techniques) 在零售业的发展过程中起到了举足轻重作用，它大大节省了交易时间，提高了销售效率。类似地，在分类学上，根据对一个统一的目标基因 DNA 序列的分析来完成物种鉴定的过程被称为 DNA 条形码编码过程。加拿大生物学家 Paul Hebert 首先倡导将条形码编码技术应用到比零售业更复杂的生物物种鉴定之中 (Hebert et al., 2003a, 2003b)。

由于依靠形态学手段进行物种鉴定本身的复杂性和低效性，传统分类学家即使持续不断地工作也很难在几个世纪之内把整个地球上的生物完全鉴定出来，可见物种鉴定是一项很艰巨的任务。常规形态学鉴定方法尚有 4 个很大的缺陷：1) 表型可塑性 (Phenotypic plasticity) 和遗传可变性 (Genetic variability) 容易导致不正确的鉴定；2) 形态学方法无法鉴定许多群体中普遍存在的隐存分类单元；3) 形态学鉴定受生物性别和发育阶段的限制，因

此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。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., 2003a; Hebert et al., 2003b)。能够充分利用现有分子生物学和因特网技术的 DNA Barcoding 技术似乎能够满足这种要求。

DNA Barcoding 一旦被证明是一个行之有效的生物鉴定手段，它将对保护生物学，生物多样性研究的发展起着至关重要的作用。其主要作用包括可以完成物种的区别和鉴定，发现新种和隐存种，重建物种和高级阶元的演化关系。它将完成一些传统形态学鉴定手段无法完成的工作，比如可以鉴定生物的卵和幼体、动物或植物的寄生物，还能很快鉴定新种，并有可能解决形态学手段难以攻克的隐存种问题，或者根据动物肠道包含物或排泄物分析来解决食物链问题。另外，这个计划本身的发展，要求一系列更快、更好、更廉价技术的产生，这势必会推动相关的分子生物学技术的进步，从而让其它相关的生物科学受益。再者，一旦解决了鉴定的问题，必将给构建系统发育树提供足够可靠的“树叶”，从而推进生物进化历史研究 (Tautz et al., 2003)。

## 2 DNA 条形码编码的原理和操作

就像零售业条形码一样，每个物种的 DNA 序列都是惟一的。因为在 DNA 序列上，每个位点都有 A、T、G、C 四种选择，从理论上讲，15 个碱基位点就有 415 种编码，是现存动物种数的 100 倍。由于自然选择的原因，某些位点上的碱基是固定的，从而导致可能的编码组合数减少。这可以通过只考虑蛋白编码基因来解决，因为在蛋白编码基因里，由于密码子的简并性，其第三位碱基通常都不受自然选择作用，是自由变化的，因此只要考虑在蛋白编码基因上的一条 45 个碱基序列就可以获得将近 10 亿种可选择的编码。更何况，由于分子生物学技术的发展，在实际研究过程中，要获得一段几百个碱基长度的序列已经比较容易，所以根本就没有必要考虑仅仅 45 个碱基长度的序列。DNA 条形码编码工作可以建立在一段长度为几百个碱基的基因序列信息的基础之上，从理论上讲完全可以包括所有物种。

作为能够用作条形编码的基因,必须具备两个看似矛盾的特征:1) 必须具有相对的保守性,便于用通用引物扩增出来;2) 要有足够的变异能够将物种区别开来。核内基因变化速率通常要低于线粒体基因好几倍,太过保守,要解决物种级的问题有困难。目前线粒体 12S 和 16S 基因被广泛用作系统发育研究的标志基因,但这些核糖体基因中存在大量的插入和缺失(即 Indels)现象,从而使序列比对(Alignment)受到障碍,不便操作,而且还容易造成错误的比对。但线粒体中存在的 13 个蛋白编码基因却很少存在插入和缺失,在这 13 个候选基因中,综合基因序列的长度和进化速率两个条件,Paul Hebert 最终选定了 COI 中一段约 645 个碱基长度的片断。因为 COI 在能够保证足够变异的同时又很容易被通用引物扩增,而且目前研究表明,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)。同时,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。

DNA Barcoding 的基本操作过程简单,包括以下几个步骤:提取 DNA、利用通用引物 PCR 扩增目的片断(位于线粒体轻链 1 490 和重链 2 198 之间)、纯化 PCR 产物、测序以及序列分析。序列分析的思路很简单,只要将所有序列进行两两比较并计算其差异值,然后根据差异值来确定物种之间的关系即可。

### 3 DNA 条形编码的发展趋势

Diethard Tautz 等甚至提出了一个以发达国家自然历史博物馆为基地的 DNA Taxonomy 发展计划,其中涉及 DNA 测序系统、DNA 保存系统的建立,还包括全球 DNA Taxonomy 管理机构的组建以及数据库的管理、维护和软件的开发等,如此的统一管理既有利于发展中国家的参与,增加研究队伍的力量,又有利于得到高质量的序列,避免混乱(Tautz et al., 2003)。

国际生物条形编码计划倡导全球统一行动,在未来 20 年时间里建立起完整的 DNA Barcoding 系统。现在一个分子生物学实验室平均每年可得到 1 000 个以上物种的目的片断序列,而且随着 DNA 测序技术的革新, DNA 鉴定(DNA Identification)的成本还会大大降低。最终建成的 COI 序列数据库将包含所有生物 COI 基因的全部信息,是分类

学最直接的信息资源,将成为全球生物鉴定系统(Global Bioidentification System, CBS)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,将会解决网络信息资源长期缺乏的问题。Paul Hebert 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(Hebert et al., 2003a; Stoeckle, 2003)。

### 4 有关 DNA 条形编码的争论

这是一个很大胆的设想,在理论上也很有力,可以给物种鉴定的工作提供一个统一的标准,从而使这项工作可以在全球范围联合起来井然有序地进行。但是,国际上对此项计划的争论也相当大,其中最主要的质疑就是,单靠一段 600 多碱基序列能否解决所有物种鉴定的问题,尤其是能否解决关系非常近的物种问题。Felix Sperling 根据他所在实验室一系列昆虫 COI 序列的数据,认为至少有 1/4 的物种是不容易用 DNA Barcoding 的方法来区分的(Sperling, 2003)。再者,所有这些关系远近问题的解决都是基于两两比较的差异度,那么,这些差异到何种程度才可以被称为传统意义上的一个物种呢? Paul Hebert 利用鳞翅目的序列数据分析结果表明,同属各种 COI 序列的平均差异程度是 11.3%,而种内的 COI 序列差异程度通常都很低,低于 2%,因此不妨碍物种鉴定(Hebert et al., 2003b)。

目前所引数据似乎证明 COI 能够完成生物条形编码,但其前提是现在所有的数据都来自于各属非常有代表性的种类,并都能很明确地鉴定,如果要深入到以后更难鉴定的物种中去,估计还会碰到很多问题。而且随着数据的增多,我们还不得不再去考虑一个阶元划分标准的问题,到底序列差异到何种程度是一个种间差别,到何种程度又分别是属间、科间差别。

自林奈双名法以来,分类学发展了 200 多年时间,积累了大量宝贵的形态学物种鉴定数据,分类学家已经给相当多的物种命名并确定了其分类地位。现在如果仅仅依靠一小段 DNA 片断来给所有物种重新贴标签,势必存在一个如何将 DNA 标签和根据林奈命名系统所定的名字两两对应联系起来的问题。再说很多模式标本已经不能提取 DNA 并制作分子标签,必须找到一个新模式标本,而找到

的新模式标本跟原来的到底是否一致也很难确知。即便超市给商品贴条形码, 也要首先基于实物比较, 故 DNA Barcoding 仅靠单个基因片断可能会显得力度不够, 也要跟形态学数据结合起来, 不能完全抛弃形态学手段 (Lipscomb et al., 2003; Tautz et al., 2003)。

线粒体 COI 基因用作 Barcoding 的目的基因尚存在某些不足。目前 COI 序列主要适用于动物界, 对于植物界和其它生命界, 尚需选择其它基因序列作为标记; 而且线粒体基因中可能会存在杂交或者基因渗透现象, 不能解决某些复杂生物的关系问题, 因此就需要联合几个核内基因来增加标记数, 以求解决问题; 同时线粒体基因也许不能解决年轻物种 (Young species) 的问题, 我们假定基因序列变化的速率是每百万年 2%, 相对于 600 bp 长度的 COI 来说, 对那些 100 万年前分化的物种将有 12 个以上可供分析的变化位点, 但是随着物种生殖隔离时间的缩短, 可供分析的变化位点就会变少, 这些物种的关系问题就很难解决, 好在化石记录表明大部分物种的进化时间都超过了 100 万年。

应该扩大基因选取的范围。毫无疑问, DNA 条形码是大势所趋, 不可阻挡。但我们必须突破单分子标记的局限性, 要尝试多个分子标记的结合使用, 而且在选取除 COI 之外的基因时, 也决不能仅仅囿于现行通用的分子标记, 要大胆尝试一些新的有潜力解决问题的基因, 比如物种形成基因可能是区分近缘物种的重要分子标记。

该计划期望所有的 DNA 提取物、所有的序列数据在全球范围内集中统一管理并全球共享。通过互联网, 序列数据完全可以实现统一管理并全球共享, 作为生物多样性评估和物种鉴定的基础 (Blaxter, 2003; Tautz et al., 2003)。另一方面, 期望将所有的 DNA 提取物集中保存, 合适的时候取出一部分用作科研或者商业用途, 这看似是个很好的规划, 但是由于 DNA 序列本身易被降解, 其保存期限是我们必须考虑的问题; 而且很多标本由于用于提取 DNA 而被毁坏, 从此也许不再存在

(Seberg et al., 2003; Tautz et al., 2003), 这对于传统分类学模式标本的保存来说是个很大的损失。

## 5 关于在我国开展 DNA 条形码编码研究的建议

目前国内还没有相关工作进展的报道。尽管许多实验室已经在基因测序方面积累了大量经验与数据, 但将 DNA 序列用于物种鉴定的报道还不多, 更没有 DNA 条形码的系统报道。我们在冷静思考其可行性的前提下, 应该积极参与到这项工作中, 充分利用我国丰富的生物资源, 促进 DNA 条形码编码工作的进行, 为构建一个强大的 DNA 分类学平台贡献我们应有的力量。我们建议国家有关机构关注这方面的国际动态, 支持我国科学家开展相关工作。我国科学家们也应该互相联合起来, 齐心协力, 共同推进这项全球性工作的顺利进展。

### 参考文献 (References)

- Blaxter M, 2003. Molecular systematics: counting angels with DNA. *Nature* 421: 122 - 124.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR, 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B.* 270 (1 512): 313 - 321.
- Hebert PDN, Ratnasingham S, deWaard JR, 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B (Suppl.)*. 270: S96 - S99.
- Lipscomb D, Platnick N, Wheeler Q, 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology & Evolution* 18 (2): 65 - 66.
- Mallet J, Willmott K, 2003. Taxonomy: renaissance or Tower of Babel? *Trends in Ecology & Evolution*. 18 (2): 57 - 59.
- Pennisi E, 2003. Modernizing the tree of life. *Science* 300 (5 626): 1 692 - 1 697.
- Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM, 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution* 18 (2): 64 - 65.
- Sperling F, 2003. DNA Barcoding: deus ex machina. *Newsletter of the Biological Survey of Canada (Terrestrial Arthropods)* 22 (1): Opinion Page.
- Stoeckle M, 2003. Taxonomy, DNA, and the bar code of life. *BioScience* 53 (9): 2 - 3.
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP, 2002. DNA points the way ahead in taxonomy. *Nature* 418: 479.
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP, 2003. A plea for DNA taxonomy. *Trends in Ecology & Evolution* 18 (2): 70 - 74.