

# DNA 分类概述\*

窦向梅<sup>1,2</sup> 肖 晖<sup>1</sup> 黄大卫<sup>1</sup>

(1 中国科学院动物研究所 北京 100101 2 北京市十一学校 北京 100039)

**摘要** 介绍了利用 DNA 进行分类的原理,强调了 DNA 分类的优越性,并着重介绍了 COI(Cytochrome Oxidase Subunit I)基因作为 DNA 分类的目的基因序列的可操作性。对 DNA 分类仍有许多争论。在实际分类工作中,建议将形态学数据与 DNA 序列数据结合起来分析。

**关键词** 分类学 DNA 分类 COI 基因

中国图书分类号:Q75 文献标识码:A

分类学是生物学研究的基础,但发展相对较慢,据估计地球上约有 1 000 万~1 亿种生物,经过 250 多年的形态分类工作,已正式描述的仅约 1 700 万种。分类学家的短缺和研究资金的匮乏使分类学面临严重的危机,分类学一度成为一个被忽视的角落。随着生物技术的发展,尤其是 DNA 测序技术的普及,Tautz 等(2002, 2003)首先提出 DNA 分类的概念(DNA taxonomy),把 DNA 序列作为生物分类系统的主要平台,即利用保守的分子序列,来定义和判断与传统意义上的物种有相似范围的分子操作分类单元(molecular operational taxonomic unit, MOTU)。在病毒、细菌和原生动物的形态上难于区分的类群中,利用 DNA 序列进行区分得到了广泛的应用,并逐渐应用于很多高等生物。DNA 分类的提出使人们看到了分类学的新希望。

## 1 DNA 分类的原理和优点

DNA 序列在每一个物种内都是特定的。一段长度为 15 bp 的核苷酸序列,每一个位置上有 A、T、C、G 4 种选择,由此可产生  $4^{15}$  约 10 亿种变换方式,是估计动物种数的 100 倍。长期的自然选择使不同物种的核苷酸在某些特定的位置是固定的,使得组合数目减少一些。蛋白编码基因可以突破这一局限,因为在蛋白编码基因序列内由于遗传密码的简并性,密码子第 3 位是可变的,一段 45 bp 的蛋白编码基因可以识别近 10 亿个物种。随着现代测序技术的发展,很容易获得几百个碱基长度的序列,目的基因在种间的差异性可能会随着目的片段核苷酸数目的增多而增加,因此可选择长一些的目的基因片段。目的基因片段的特点是由物种的年龄和演化速率决定的,要准确识别物种没有限定它的长度。

DNA 分类的思路比较简单,一般程序为常规分子生物学方法提取 DNA、PCR 扩增目的片段、纯化 PCR 产物和测序。进行序列分析后,将序列所含有的信息做为某一物种的鉴定标签(identification tag),这条序列就可以做为未来研究的参考标准,相关的模式种和 DNA 提取物则存放在博物馆里。理想状况下,可以通过此方

法获得所有已知物种的 DNA 序列,辅以各物种的描述和其他信息,以及它的分类地位和所参考的依据,据此可建立一个包括很多物种的序列数据库。对于任意一头未知标本,只需要把它的序列与数据库中已有的序列进行对比,就可以快速确定物种的分类地位(尤其是外形差别较大的处于不同生活阶段的幼虫),判断其是否是一个新种或隐存种,还可用于鉴定濒危物种和由贸易往来而引入的外来种。即使有些物种不能根据序列数据进行精确定位,至少可以找到与其密切联系的种类。依据此方法建立的系统称为基于 DNA 的物种识别系统。

DNA 序列信息是数字式的,易于管理,不受主观判断的影响,没有地域限制,具有可重复性。一旦建立起一个基于 DNA 的物种识别系统,借助于互联网技术,将会为物种鉴定信息的积累和获取提供一个更可行而通用的平台,可以加快交流的速度,更易获得物种的鉴定和描述信息。这一系统更易于管理和调用,提高分类工作的效率。假设一个分类学家能够准确地鉴定 1 000 个物种(事实上几乎没有),要鉴别地球上现存的 1 000 万~1 亿种生物就需要 1 万~10 万个分类学家持续地工作,而对于严重缩减的分类学队伍来说,根本是不可能的。如果基于 DNA 的物种识别系统能够正常使用,就可以使这种困窘的局面得到改观。

基于 DNA 的物种识别系统不但有可能解决物种鉴定问题,还可以利用数据库中大量的序列进行相关的系统发育分析,从而推进生物演化历史研究。分类学的目的是进行物种鉴定并把它置于较高的分类阶元,后者常与系统发育假设相关,生物分类应反映它们的演化历史,而系统发育关系可以从 DNA 序列分析中直接获得。分子数据尤其适于再分析(re-analysis)和后分析(meta-analysis),来自于多个独立研究的数据可以在新的参数和假设下进行比较和研究。用于 DNA 分类的序列数据库主要用于提供物种鉴定信息,也会提供许多的系统发育信息。利用这些序列至少可以在种级水平上进行初步的系统发育分析。DNA 分类学数据

\* 基金项目:国家自然科学基金项目(No.30370188, No.30330090)资助,部分由国家基础科学人才培养基金项目(No. J0030092)资助

库可以提供构建生命树(Tree of Life)的框架,还可以为各分支提供足够可靠的“树叶”。

基于DNA的物种识别系统将会为现有物种鉴定提供一种可靠而可行的方法,同时也使我们从新的视角来看待生物的多样性。这种方法可以克服传统形态学方法的缺点,鉴定范围更广,包括生物的卵和幼体、内脏内容物、人类疾病的媒介物、动植物的寄生虫等,并有可能发现新种和隐存种。Brown等(2003)根据DNA序列和形态特点发现巴布亚岛新几内亚鳞翅目昆虫一新种 *Xenothictis gnetivora*。Hebert等(2004)对哥斯达黎加西北部一种广泛分布的鳞翅目昆虫 *Astraptus fulgurator* 进行了研究,根据对博物馆标本的DNA序列分析并结合形态学知识,发现这个地区存在的 *A. fulgurator* 至少是10个种的集合。这个结果使我们认识到自然界中还有许多的隐存种有待去确定,势必会促进生物多样性的研究。现在人们已认识到保护生物多样性的重要性和紧迫性,我国已经成立了中国生物多样性信息系统(Chinese Biodiversity Information System, CBIF),建立了专门的数据库来提供生物多样性的科学数据,基于DNA的物种识别系统可以促进基础数据库中物种编目数据、濒危和保护物种数据等的完善,促进我国生物多样性保护和持续利用事业的发展。

## 2 目的基因的确 定

基于DNA的物种识别系统的核心问题是确定适于进行物种识别的目的DNA序列。目的DNA序列最好能满足3个条件:1)在所研究的物种中均能获得;2)易于分析;3)信息量足够识别所有物种。目前用于分类研究使用最多的基因是核的核糖体小亚单位(18S rRNA基因)和线粒体的核糖体小亚单位(12S rRNA基因),但这2个基因都相当保守,不能用于区分关系密切的种。一种可行的方法是使用核糖体DNA(rDNA)序列的D-loop区(divergence loops),尤其是核糖体大亚单位的RNA基因,它们的演化速率较快,位于较保守的区域,可以使用通用引物从少量或已部分降解的样品中进行扩增。核糖体RNA基因在细胞中含量丰富,可以直接作为DNA微阵列(DNA-microarray)探针来鉴别物种。核糖体基因序列比对中存在很多问题,而且与不同程度的协同演化(concerted evolution)有关,从长远来看以核糖体基因作为目的基因是不合适的。

线粒体基因组不含间隔区和内含子,无重复序列,很少发生不等交换,在遗传过程中不易发生基因重组、倒位、易位等突变,更有望成为目的基因。一个快速变异且含有大量信息的基因是线粒体的控制区基因,但它的演化速率太快,不能成功区分高级分类阶元的种类。相对来说,线粒体基因中13个蛋白编码基因更为适用,这些蛋白编码基因中很少有插入或缺失发生,大多数的插入或缺失会引起阅读框的移动。线粒体 *Cytb* (细胞色素b)基因在脊椎动物中得到了广泛的应用,利

用它可以解决关系密切的种间关系,也可以重建高级阶元的系统发育关系,因为 *Cytb* 基因密码子第3位点的演化速率非常快,而蛋白序列的演化相对较慢。*COI* (细胞色素c氧化酶I)基因演化速率较慢,相对 *Cytb* 基因有2个优点:首先,此基因的通用引物适用范围很广,可以获得绝大多数动物的5'端序列;其次,*COI* 比 *Cytb* 基因在分类上具有更大的识别能力,前者由于氨基酸的变化速率更慢而能区分分异时间更远的种类。在系统学研究中 *COI* 基因比线粒体其他基因应用更广泛。

Hebert等(2003)对GenBank中2002年11月到2003年2月的所有动物同属物种的 *COI* 序列进行分析,结果表明同属物种的 *COI* 基因序列间有足够的变异,通常情况下能够区分除刺丝胞动物门(Cnidaria)的所有动物门的近缘种,刺丝胞动物(Cnidarian)中的变异程度较低,与这种生物的线粒体演化速率很慢有关,至少部分是由于其线粒体内存在切除修复系统所致。他们提出利用线粒体 *COI* 基因作为全球动物类物种识别系统的核心基因,期望建立以DNA序列为物种条码(DNA barcoding)的系统,并期待给所有生物进行编码。他们将动物界7个主要的门、昆虫纲8个主要的目和鳞翅目200个近缘种的代表样本的 *COI* 序列进行比对,分析每一阶元内 *COI* 序列的多样性。再以此为基础,根据 *COI* 序列的相似性作为门、科或种级的鉴定标准。根据来自于较高分类单元取样的 *COI* 基因的特点,可以将新分析的物种归入合适的阶元,其中鳞翅目的种级分析完全正确,由于鳞翅目是动物界中分类多样性最丰富的目之一,而且序列差异性较低,因此可望在其他类群里面也可以成功分类。

## 3 争 论

国际生物条形码计划倡导全球统一行动,在未来20年时间里建立完整的DNA Barcoding系统。Bessansky等(2003)强调DNA条码系统只是为了加速发现物种的进程、缩短调查和分析物种多样性的时间,但针对这项计划及相关问题引发了一些争议。

一个颇具争议的问题是,一段DNA序列(如 *COI* 基因)能否解决所有物种鉴定的问题,尤其是关系非常近的物种问题?当一对物种具有较近的分异时间时, DNA序列的鉴别能力是有限的。如刚分离的姐妹种由于基因流或较近的共同祖先仍具有等位基因,这种情况下仅根据一条或几条序列不足以把某一物种准确地归入一个特定的类群。还有一个复杂的问题,对于细胞器DNA来说(如线粒体或叶绿体),偶尔会发生基因转移,至少在关系较近的种间是存在的,使用核基因和线粒体基因可能得到不同的分析结果。即使单纯的一段 *COI* 基因,也不可能适用于所有的分类阶元,如 *COI* 基因不适用于导管植物和某些小型动物。有3个因素会限制 *COI* 基因识别物种,即:杂交;年轻的物种

(young species)和较慢的分子演化。线粒体 *COI* 基因不可能解决全部问题,如有些物种的界限由于杂交或基因渗入被打破,这就需要一种或多种核基因作为补充。同样地,对于某多倍体物种,还需研究基因组的大小。研究显示大多数动物类群中发生杂交、多倍体形成和线粒体基因渗入的个体不足1%。假设序列的变异速率为2%每100万年,生殖隔离后100万年的2个物种间600 bp的 *COI* 序列就可以发现12 bp的差异。如果物种生殖隔离后演化的年代较短,物种识别就比较困难。然而化石材料显示大多数物种形成年代超过100万年,所以大多数物种还是很容易识别的。不同的动物谱系(lineage)分子演化速率可以相差100倍,识别演化速率慢的类群中的年轻物种更加困难。总之,这3个因素不会影响属的划分,只会偶尔影响种的识别。Tautz等建议使用多个基因区段来进行物种的分类鉴定工作,对于需要深入研究的可能的杂交种会提供一些有用的信息。

第2个争议的问题是阶元划分标准,到底序列差异到何种程度属于种间差别,何种程度又分别是属间、科间差别? Hebert等对GenBank中所有动物的同属物种的 *COI* 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* 序列差异程度很少超过2%,绝大部分低于1%。如果都以此差异度为标准,那么利用 *COI* 序列的差异程度就很容易区分不同物种,事实上并非如此。Sperling(2003)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的物种不容易根据序列差异来区分。即使现在所有的数据都来自于各属代表性种类,并都能明确鉴定,也会出现一些例外,如刺丝胞动物(Cnidarian)同属物种对 *COI* 序列表现出很低的变异,94.1%的物种对种间变异小于2%,远小于同属各种 *COI* 序列的平均差异程度(11.3%)。还可能会出现种内变异可以忽略或者至少比种间变异低的情况。随着分析数据与类群的增多,区分标准还会不断变化。

另一个问题是基于DNA序列的分类数据库如何为已命名物种提供一条特定的DNA序列? 双名法命名的物种给我们提供了分类关系(taxonomy affinities)、形态、分布和可能的生态功能等信息,生物学家可以将该物种与相关知识联系起来,基于DNA的物种鉴定系统不能提供这些功能,需要一个独立的命名系统将物种序列数据跟以前的种名实行链接。为已命名的模式种提供特定的DNA序列时,最好是由这个模式种获得,但很多模式标本已经不能提取DNA,找到的新模式标本跟原来的是否一致也很难确知。大动物和绝大部分植物及真菌利用标本的一部分提取DNA就足够了,但对昆虫标本来说,为了获得足够的DNA可能要毁坏整头标本。后来的研究者希望看到整体或尽可能多的保存标本,虽然在毁坏标本前可以使用相机把标

本的情况照下来,但毕竟有许多细微结构已无法观察。

建立一个整合分类学信息的通用序列数据库时存在不同的意见。Tautz等认为现在的DNA数据库如NCBI(National Center for Biotechnology Information)和EBI(European Bioinformatics Institute),虽然已经拥有10万多条真核生物的序列信息,但不能保证所有序列的物种名字都是正确的,这些数据库没有包括形态的、生物地理的、生态的信息和文献资料情况,在递交序列时没有必须填写的分类标准。现有DNA数据库的建立者认为不必将这些DNA数据库转变成分类为主的数据库,而是希望建立一个整合了分类学信息的通用的分类学序列数据库。现在最重要的就是建立一个DNA分类通用的注册系统,还有一个类似于NCBI和EBI的序列号,分类注册序列号不用于鉴定特定的序列,而是说明与特定序列相关的DNA存放和标本的情况。单个的序列仍然可以递交到NCBI或EBI,但应明确标出DNA分类注册号。目前一个前导系统已经成立了,网址是:<http://www.zsm.mwn.de/DNATAX/>。假如现存DNA数据库中有鉴定错误的物种,只需对参考标本(voucher specimen)进行检视就可以解决问题。递交序列时由谁来决定分类标准存在争议,物种的界限(circumscription)界限会随着研究的不断深入而变化。

一旦DNA分类得到普遍认可,建立相应的DNA数据库需要大量的资金。新技术的发展会降低测序费用,但DNA测序相对传统分类技术仍然是一种复杂且昂贵的技术。这将会增加南北国家分类学发展的差距,将许多未接触测序技术的分类学家排除在外。Seberg等(2003)认为将DNA序列强制性引入分类中似乎是一种倒退。

#### 4 结语

不可否认,DNA序列数据会加速物种鉴定的进程,还会促进生物多样性研究,辅助其他研究如整合分类信息系统(Integrated Taxonomic Information System, IT IS)、物种2000(Species 2000);生命树互联网计划(The Tree of Life Web Project)等。但它只是提供了一套不同于形态性状的适用范围更广的数据,从另一个角度来看待物种鉴定的问题,可以用它来辅助分类工作,解决传统分类中难以解决的问题,没有必要用DNA分类系统替代现有的分类系统。必须明白,DNA序列并非万能药(panacea),非同源相似、序列比对和不同的分析方法都会增加现有的序列数据分析的不可靠性。DNA序列不可能解决形态分类中的所有难题,分类学发展的200多年时间中,积累了大量宝贵的形态学物种鉴定资料,不能轻易地否定并抛弃它。在实际应用中,我们认为以形态学数据为主,DNA序列数据为辅,将传统分类系统的优势和分子生物学的测序技术结合起来,推进分类学的研究。

#### 主要参考文献

# 神奇的热带特异资源——卡瓦胡椒\*

施江<sup>1,2</sup> 辛莉<sup>1</sup> 谭琳<sup>2</sup> 郑学勤<sup>2</sup>

(1 河南科技大学农学院 河南洛阳 471003 2 海口热带作物生物技术国家重点实验室 海南海口 571101)

**摘要** 介绍了卡瓦胡椒的发现、起源、化学成分及药理作用。近年来卡瓦胡椒在欧洲和美国开始为人们所接受,1990年,德国联邦卫生局批准卡瓦胡椒制剂 WS1490 用于治疗焦虑症,卡瓦胡椒成为德国销售量最大的 20 味草药之一。1992 年卡瓦胡椒根被美国食品、药品检验局(FDA)认证为“精神健康”食品。1998 年,卡瓦胡椒被美国人称为“植物药之星”。1999 年卡瓦胡椒又在美国创下年销售额 50 亿美元的奇迹。2001 年,卡瓦胡椒的引种、适应性和生物技术的研究项目被列为我国“十五”科技攻关项目。

**关键词** 卡瓦胡椒 卡瓦内酯 药理作用 临床研究

中国图书分类号:Q949.95 文献标识码:A

卡瓦胡椒(*Piper methysticum*),俗称卡瓦(Kava or Kawa),是瓦努阿图、斐济、汤加、巴布亚—新几内亚及所罗门群岛等南太平洋诸岛野生的一种多年生胡椒科(Piperaceae)胡椒属(*Piper*)灌木类植物。植株一般高 3~5 m,叶片心脏形、全缘、较大而薄,茎干非藤本,直径 3~5 cm,与栽培胡椒一样属肉穗花序,黄绿色小花,但只开花不结果,无种子形成。南太平洋岛国的居民喜欢用卡瓦胡椒的根或根茎制备能放松肌体和情绪、恢复体力、减轻痛苦的饮料,尤其是在宗教活动、重大节日的庆祝活动、成人仪式及宴请宾客时,常常专门为宾客设立卡瓦宴,是礼仪和社会活动中必备的饮料,饮后肌体神经先兴奋,而后抑制,进入极其良好的睡眠,可使神经系统保持平衡,不影响智力,又无成瘾性。

## 1 卡瓦胡椒的发现

1769 年英国探险家詹姆斯·库克率船队登上波利尼西亚岛探险途中,亲眼目睹许多离奇怪诞的事情,其中尤以当地少女咀嚼卡瓦胡椒根最引人注目和令人驻足。随同库克探险的艺术家 Parkinson 在社会岛上第 1 次画出了这种植物的图,这幅画现被保存在英国伦敦大英博物馆内。1772—1775 年随同库克第 2 次远航探

险的植物学家 Forster 第 1 次从分类学的角度描述了这种植物。1886 年 Lewin 第 1 次写出了一本有关卡瓦胡椒的学术专著,并声称自己是继 Forster 之后从植物学角度研究卡瓦胡椒的第 1 个人。自从 18 世纪卡瓦胡椒与欧洲人的首次有意义的接触,通过各种媒体的宣传,卡瓦胡椒这种中南太平洋岛屿上特有的药用植物才逐渐为外面的世界所熟悉。特别是近半个世纪以来,人们对卡瓦胡椒的认识才更加广泛和深入。1960 年 Steinmetz 出版了第 2 本有关卡瓦的学术性专著。1992 年 Vincent Lebot, Mark Mertin 和 Lamont Lindstrom 出版了最新的第 3 本有关卡瓦的专著,书中结合自己的研究工作并在总结前人研究成果的基础上对卡瓦胡椒进行了广泛系统阐述,并列有大量作者参考的文献资料,所有这些使我们对卡瓦胡椒的认识进一步加深和全面。

## 2 卡瓦胡椒的起源

最近的遗传研究提供了额外的证据进一步表明卡瓦胡椒是起源于 *P. wichmannii*。Vincent Lebot 等人 1992 年采用电泳技术和细胞学技术研究了卡瓦胡椒和 *P. wichmannii* 的遗传背景。细胞学研究表明卡瓦胡椒、*P. wichmannii* 和 *P. gibbilimum* 3 种植物染色体数

- 1 Tautz, D., Arctander, P., Minelli A. et al. DNA points the way ahead in taxonomy. *Nature*, 2002, 418: 479.
- 2 Tautz D., Arctander P., Minelli A. et al. A plea for DNA taxonomy. *Trends Ecol. & Evol.*, 2003, 18: 70—74.
- 3 Brown J.W., Miller S.E. and Horak M.. Studies on New Guinea moths. 2. Description of a new species of *Xenochictis* Meyrick (Lepidoptera: Tortricidae: Archipini). *Proc. Entomol. Soc. Wash.*, 2003, 105: 1043—1050.
- 4 Hebert P.D.N., Penton E.H. and Burns J.M. et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Nat. Acad. Sci. USA.*, 2004, 101: 14812—14817.
- 5 Hebert P.D.N., Ratnasingham S. and deWaard J.R.. Barcoding animal

life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B (Suppl.)*, 2003, 270: S96—S99.

- 6 Besansky N.J., Severson D.W. and Ferdig M.T. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends in parasitology*, 2003, 19: 545—546.
- 7 Sperling F.. DNA Barcoding: Deus ex Machina. *Newsletter of the Biological Survey of Canada (Terrestrial Arthropods)*, 2003, 22: Opinion Page.
- 8 Seberg O., Humphries C.J., Knapp S. et al. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.*, 2003, 18: 63. (E-mail: douxm2810@sohu.com)

(BH)

\* 基金项目: 国家十五重点攻关课题“热带特异资源卡瓦胡椒引种繁殖的研究”(编号: 2001BA707B)项目资助, 洛阳市科技计划项目“牡丹转基因育种技术研究(0602042A)”资助, 洛阳市科技计划项目“牡丹新品种选育及产业化开发专项(0602042A)”资助。