

•保护论坛•

# 地球生物基因组计划与生物分类学

朱朝东<sup>1\*</sup> 罗世孝<sup>2</sup> 周欣<sup>3</sup> 陈华燕<sup>4</sup> 张冠阳<sup>5</sup>

1 (中国科学院动物研究所动物进化与系统学院重点实验室, 北京 100101)

2 (中国科学院华南植物园保护与可持续利用重点实验室, 广州 510650)

3 (中国农业大学植物保护学院昆虫学系, 北京 100193)

4 (Department of Entomology, The Ohio State University, Columbus 43212, USA)

5 (Florida Museum of Natural History, University of Florida, Gainesville, Florida 32607, USA)

**摘要:** 结合第19届国际植物学大会发表的“植物科学深圳宣言”, 我们从四个方面对“地球生物基因组计划”(Earth BioGenome Project, ESP)与生物分类学之间的内在联系进行了讨论。ESP立足组学, 旨在推动并加速物种多样性信息数字化和可计算性。生物分类学仍然是目前生物多样性研究的根基, 也会为ESP提供不可或缺的重要信息。ESP既需要和生物分类学现有知识和队伍有机整合, 也需要与中国现有生物多样性重大研究计划和战略规划相结合, 这样才能真正实现学科突破创新, 发挥引领作用。

**关键词:** 生物分类学; 基因组计划; 学科交叉; 学科整合

## Earth BioGenome Project and Taxonomy

Chaodong Zhu<sup>1\*</sup>, Shixiao Luo<sup>2</sup>, Xin Zhou<sup>3</sup>, Huayan Chen<sup>4</sup>, Guanyang Zhang<sup>5</sup>

1 Key Laboratory of Zoological Systematics and Evolution, Chinese Academy of Sciences, Beijing 100101, China

2 Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

3 Department of Entomology, College of Plant Protection, China Agricultural University, Beijing 100193, China

4 Department of Entomology, The Ohio State University, Columbus 43212, USA

5 Florida Museum of Natural History, University of Florida, Gainesville, Florida 32607, USA

**Abstract:** Reflecting on the recent Shenzhen Declaration on Plant Sciences at XIX International Botanical Congress, we discuss connections between the Earth BioGenome Project (ESP) and Taxonomy. Against a backdrop of rapid species extinction, the ESP aims to speed up species conservation and preservation. It holds promise in promoting biodiversity informatics. Taxonomy underpins biodiversity research, and it will contribute important information to the ESP. For the ESP to make breakthroughs and innovations in biological sciences, it requires not only organic integration with the taxonomic community, but also needs to align with current major research programs and national strategies on biodiversity in China.

**Key words:** taxonomy; genome project; interdisciplinary research

2017年, 中国生物多样性研究领域发生了两大重要事件: 一是2月份美国史密森生物多样性基因组计划(Smithsonian Initiative on Biodiversity Genomics)与中国深圳华大基因研究院共同启动了比人类基因组计划更宏大的“地球生物基因组计划”(Earth BioGenome Project, ESP) (<http://www.sciencemag.org/news/2017/02/biologists-propose-sequence-dna-al>

l-life-earth); 二是7月份第19届国际植物学大会在深圳隆重召开, 并发表“植物科学深圳宣言”(<http://www.abc2017.cn/CN/Declaration/>)。ESP和深圳宣言都是研究者面对生物多样性发生急剧变化的现状而提出的, 二者分别从不同的角度对当前发生的生物物种灭绝和生物多样性丧失表达了深切关注, 并提出了一些需要优先行动的领域。同时, 随着国力

收稿日期: 2017-10-24; 接受日期: 2017-11-20

基金项目: 国家杰出青年科学基金(31625024)

\* 通讯作者 Author for correspondence. E-mail: zhucd@ioz.ac.cn

和研究能力的增强,中国科学家也向世界展现出保护生物多样性的行动能力和勇于承担关键角色的自信心。

## 1 ESP的近期和远期目标

和人类基因组计划一样,ESP自提出之初,便是生命科学研究的热点话题之一。该计划关注生物多样性,试图在基因组水平解码物种的遗传特征。其远期目标为破译地球上所有生命的基因组;近期目标首先集中在科级水平,完成约9,330个真核生物代表性物种的基因组测定,获得与人类参考基因组相当或比人类参考基因组更好的参考基因组数据集。由于其内在的数字化和可计算性,该计划产生的大型基因组数据集也必将对生命科学的很多分支学科,特别是生物多样性科学、系统与进化生物学、形态与发育、生态学、生理学和遗传学等领域产生深远的影响。

从目前学术界推动的基因组项目可以看出一些共性的地方:(1)动物基因组测序计划,如鸟类(万种鸟类的基因组生命之树的数字化项目, <http://b10k.genomics.cn>)、脊椎动物(万种脊椎动物基因组计划项目, <https://genome10k.soe.ucsc.edu>)等,大多具备非常好的分类学基础,均有国内外知名分类学者的参与。(2)试图探索昆虫演化之谜的IKITE项目,旨在通过对1,000种昆虫的转录组测序构建昆虫高级阶元系统发育关系(<http://1kite.org>),从项目设计、样本采集、数据获取、数据整合、数据分析、数据展示等每一个环节都活跃着非常强大的国际分类学队伍。此外,孑遗植物无油樟(*Amborella trichopoda*)的全基因组测序也已完成(*Amborella Genome Project*, 2013); 2009年清华大学深圳研究生院与深圳市兰科植物保护研究中心(全国兰科植物种质资源保护中心)启动了“兰花基因组计划”。这些都是分类学背景清楚也相对容易识别的类群。总之,无一例外,上述研究都涉及了生物分类学学者长期关注的类群系统发育和奇妙的进化生物学问题。

## 2 生物分类学是ESP的基础

生物分类学的目的和独特的使命是建立和完善一个反映生物进化历史、用以沟通和存储生物多样性信息并比较生物学知识和数据的参考体系。ESP将在生物分类学现有框架内选择代表物种,测定基因组。两者都有助于建立更加完善的生命之树

框架,有助于探索、认知和描述地球上的生物,以更好地保护和利用它们。两者的区别只是解决问题的方式不同。作为生物分类学工作者,笔者欣喜地看到该计划的组织者们拟按照系统发育关系和生物分类阶元来逐步实施该计划的目标。如加利福尼亚大学戴维斯分校的进化基因组学家Harris Lewin提到:该计划将需要大量的专家和专业背景知识,特别是分类学工作者的参与。在代表科属中选择一个代表物种进行基因组测序,需要分类学者的准确鉴定,而真正的瓶颈在于某些类群比较罕见或缺乏从事研究的专家。当类群包含大量尚未被描述的物种时,如果没有分类学者的野外工作与分类鉴定,基因组解码这些物种将无从谈起。不难想象,物种分类学障碍必然会在ESP中迅速显现。该问题在生物多样性发现与保护、宏观生态学研究以及DNA条形码技术发展过程中已反复出现。

## 3 ESP能否避免分子系统学和DNA条形码的尴尬

2005年,由生命条形码联盟(CBOL)和英国自然历史博物馆主办召开了第一届国际DNA条形码大会。多位大力推动该项技术的学者到会并做会议报告。其中,Alfried Vogler教授领导的英国自然历史博物馆昆虫系分子系统学实验室大力推动DNA分类学的思路和方法(Tautz et al, 2003); DNA条形码之父Paul Hebert教授等期待用一台手机大小的便携测序设备,将野外采集到的生物样本即时鉴定到物种。之后不久,Alfried Vogler教授和Quentin Wheeler教授组织了两人之间的专场学术辩论会,讨论实体标本体系的存废,建立全新的基于DNA信息的生物系统学的必要性等。近10余年来,这些新思路一方面对包括笔者在内的学者和团队成长起到了引领和推动作用;另一方面使我们更加意识到生物分类学的重要性。DNA条形码技术和DNA分类学的兴起解决了很多物种问题,大大加速了物种多样性的发现进程。但大量的研究表明:现有基因数据库中的数据缺乏对其客观载体——物种的分类学信息。以昆虫为例(Chesters & Zhu, 2014),在已经公开的基因序列中,一半以上缺乏分类学阶元信息。实际上,缺乏分类鉴定的问题不仅存在于基因或者组学数据中(Page, 2016),大量的博物馆收藏标本同样存在分类鉴定的问题。根据初步调研和分析,在其他类群中也存在大致相同的情况。从DNA条形码技术设

计的理念和具体的实施流程来看, 分类学知识是其中不可或缺的一个关键组成部分。对于个别较小的类群, 在实验室层面就可以把分类学人才和知识整合进来。但是, 对于一些很大的类群则需要巨大的分类学人才队伍和知识储备。以《中国植物志》和《中国生物物种名录》为例, 完成这些浩大工程需要大量分类学工作者的参与和付出。显然, 无论基于形态学、DNA序列还是基因组, 全球生物测序都需要充分的样本采集与鉴定等过程。可以说, 如果没有生物分类学家长期野外工作形成的“专业感觉”, 很多物种乃至科属连采集都很难。

与分子系统学发展历程中普遍存在的问题一样: 常见的类群容易获得样品并分类, 而系统性取样则必然涉及分类学知识匮乏的类群。在生物界, 很多类群都缺乏能够鉴定到物种水平的专家。很多科甚至更高级别的阶元都缺乏可以参考的文献等专业知识, 能鉴定的专家就更少了。从分类学研究比较透彻、分类系统比较稳定的科入手, 属的问题相对比较好解决。但很多类群属级阶元的变动都比较普遍。分子系统学的发展验证了大部分基于形态学得出的分属建议, 但也推翻了不少前人的假说而提出了新的属级阶元变动(Waugh, 2007)。显然, 多基因乃至组学的数据会有助于固定一部分变动频繁的属级阶元, 但分类阶元的变动应该得到分类学工作者的论证和支持。

物种是客观存在的进化单元, 其分类鉴定难度要远远超过属以上水平。一般生物在科级以上水平比较容易区分, 但大多数的属和物种则需要非常专业的分类学知识和能力才能鉴别(Giangrande, 2003)。因此, 某种意义上讲, 分类学人才的培养就在于对他们的属和物种鉴别能力的培育。

#### 4 ESP等新思路对分类学提出的新挑战

生物分类学是一门高度综合的学科。从事分类研究的学者首先必须是一位生物学家, 特别是遗传学、发育生物学的知识背景, 要深刻理解生物的变异和进化(洪德元, 2016)。然而, 从事分子生物学和基因组学的学者未必都具备分类学功底。

物种划分是分类学的核心内容, 采用多学科的手段、不断吸收新的学科和技术的过程, 也是分类学发展的过程。从数据的角度看, 快速积累的DNA信息和形态学信息的整合尤为必要。从肉眼观察、解剖镜观察、电子显微镜观察到基因、基因组测序、

DNA和形态学数据为理解物种界定和准确分类提供了多维度的证据。其中, DNA等分子数据实现了生命信息的快速、自动数字化。无论基因、基因组还是其他新的技术所产生的数据, 都是对客观主体——物种的观察和信息提取与整合。最近快速发展的三维成像技术实现了生物内部解剖和外部形态结构的数据自动提取, 并产生了大量前所未有的形态学数据; 红外相机技术实现了对生物物候、发生动态等信息的海量数据获取。利用手机和数码相机, 普通公众如鸟类爱好者、植物爱好者也可以实现对物种的观察和数据的积累, 并通过互联网参与科学研究。相信未来会有更多新的仪器和技术, 可以帮助我们实现对个体数据的自动、快速提取, 人工智能、机器学习、深度学习等领域的进展也会给生物分类带来新的认知思路和实现方案。

基于图像库的生物识别系统的建立为物种识别提供了便利, 如“形色”(<https://itunes.apple.com/cn/app/形色-会上瘾的识花神器/id1018747351?l=en&mt=8>)、“智能花卉识别系统”([http://www.chinadaily.com.cn/interface/yidian/1138561/2016-08-10/cd\\_26422376.html](http://www.chinadaily.com.cn/interface/yidian/1138561/2016-08-10/cd_26422376.html))等。而昆虫尽管物种多, 图片也多, 但是还没有整合, 更缺乏较好的终端识别和分析工具。如果实现了目、科甚至属的大类区分, 经典分类学工作者就可以把最宝贵的经验和技能集中在物种鉴定环节上。这样不仅可以大大提高物种鉴定效率和准确率, 还可使分类学者带着物种问题来应用DNA条形码、宏基因组学等新技术开展研究。这样的模式将不仅解决分类学障碍问题, 还将在获得大量数据的同时保存凭证标本。

生物分类学从来都不是孤立发展的一门学科。新一代测序技术、互联网技术、超高清相机、图片数据库、模式识别、机器人等技术和工具如何和经典分类学有机整合? 以昆虫分类学为例, 如何开发一项市场认可的昆虫识别技术? 这应该有助于提高该学科的社会认可度。在大科学、大数据工程中要不断吸纳已有的生物分类学科学知识和人力资源, 形成有机整合; 同时关注并加强对源头创新队伍中青年人才的长期培育。

如果分类学者参与ESP, 如何定位自己的角色并发挥积极作用呢? 显然, 分类学者的作用远远不止贡献标本, 而在于类群科学问题的提出、物种分类鉴定、生境适应、关键性状的解析和生物学精彩故事的讲述。缺乏分类学基础的分子生物学研究将

可能导致新发现与人类数百年积累的生物学知识脱节。分类学工作者则可以考虑如何参与并加速物种多样性的研究进程。关注新技术并超前规划引用新技术,兼收并蓄,加速物种多样性发现甚至描述过程、理解生物多样性的格局及生命奥妙的本质。

## 5 中国学者如何引领ESP和开拓分类学新时代

深圳宣言特别提到:“产生大量数据的新技术往往受到当前基础设施和信息管理能力的限制;对实验室研究的日益重视,让人们忽略了在迅速消失的生境中优先开展野外考察的重要性;纯理论研究和应用研究失去了应有的平衡。”植物学家们在宣言中提出了7个优先领域制定行动战略。其中,生命的编目研究成为十分迫切的任务。“虽然我们已命名了很多植物,但是对其中大多数知之甚少,很多植物仍然有待发现。我们可以保护和保存那些已发现的植物,而对于未发现的物种,在其灭绝前发现并了解它们显得尤为迫切。开展这些工作所需的具有一定规模的合作和整合机制尚未建立。拯救植物需要了解植物。时间已经十分紧迫。”

深圳华大基因研究院拥有国际最大规模的综合基因组分析平台。美国史密森学会是世界上最好的标本保藏场所之一,拥有一流的设施和分类学者以及丰富的馆藏标本。ESP由他们联合倡议,形成了基因组测序和生物分类学的有机结合与合作。生物分类学对资金、仪器等资源依赖程度低,虽然历经波折至今还能持续发展。它的学科历史实际上就是立足物种,对新理论和新技术不断兼收并蓄,实现发展的进程。笔者认为要为该计划点赞,以开放、包容和融合的心态,期待该计划能够真正吸收世界生物分类学长期以来的专家知识库(包括提出的问题),推动包括生物分类学在内的分支学科有新的突破。那么,突破来自哪里?从数据的角度来看,从小数据集到大数据集,数据获取、整合、分析、展示等每个环节都需要学科交叉和技术进步。但是,我们仍需要回到物种本身来探讨生物学问题。而大多数国内同行对于物种相关的理论和方法研究尚处于跟随阶段。但是理论和方法的研究一旦有框架性的突破,则会使某一学科有翻天覆地的变化。这会让很多生物系统学学者联想到Hennig系统发育论引起的学科大讨论和后续学科的大发展,而这一点在中国目前仍然极为薄弱。因此,生物学工作

者需要利用包括形态学和基因组学在内的各类数据,进行整合、分析、展示,深入理解物种本身的生物学问题,解析物种形成及其对所处生态系统的作用和适应机制。

ESP的实施将革新我们对生物学的理解,有望极大改善物种保护工作,并为农业、医药和生态系统服务创造新的基因资源。中国生物多样性非常高,并拥有世界第三极这种特殊的生物多样性区系,中国分布的特有类群以及中国学者发表的新科或新属模式种的数量也非常多。在ESP首先要集中完成的约9,330个真核生物科级代表性物种中,中国分类学者需要考虑并列优先进入测序的物种名单。2017年8月19日,中国科学院在拉萨启动了第二次青藏高原综合科学考察,集中了不同学科领域的科学家参与,相关的研究所均派出科研队伍组成了综合考察队。如何将这种具有战略性和前瞻性的基础研究 with ESP有机结合?“泛第三极生物基因组计划”能否成为ESP的引领项目之一?这些项目能否有助于包括生物分类学相关的学科培养、锻炼和储备年轻人才队伍?这些都需要中国科学家认真思考与谋划。

## 参考文献

- Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science*, 342, 124089.
- Chesters D, Zhu CD (2014) A protocol for species delineation of public DNA databases, applied to the Insecta. *Systematic Biology*, 63, 712–725.
- Giangrande A (2003) Biodiversity, conservation, and the ‘Taxonomic impediment’. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13, 451–459.
- Hennig W (1950) Grundzüge einer Theorie der phylogenetischen Systematik. Deutscher Zentralverlag, Berlin.
- Hong DY (2016) Opinion of raising rationality in species delimitation. *Biodiversity Science*, 24, 360–361. (in Chinese)
- [洪德元 (2016) 关于提高物种划分合理性的意见. 24, 360–361.]
- Page R (2016) DNA barcoding and taxonomy: dark taxa and dark texts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 317, DOI: 10.1098/rstb.2015.0334.
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler A (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution*, 18, 70–74.
- Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, 29, 188–197.