

生物多样性信息学研究进展

王利松¹ 陈彬¹ 纪力强² 马克平^{1*}

1 (中国科学院植物研究所, 北京 100093)

2 (中国科学院动物研究所, 北京 100101)

摘要: 生物多样性信息学是一门蓬勃发展的新学科。它将现代的信息技术带入生物多样性及其相关学科的研究领域。它在生物多样性基础数据的数字化、模型工具和各种工具软件的开发、数据整合, 以及全球、地区和国家尺度生物多样性信息网络等多个方面的发展, 向我们展示了未来在全球范围内自由、免费共享生物多样性数据和信息, 以及人们行动起来共同关注、调查与监测野外生物多样性的前景。目前, 已有大量数字化的物种编目、标本馆标本、多媒体影像、研究文献等生物多样性基础信息可以通过互联网检索和利用。其中, 最值得关注的是一些成功的国际性研究项目, 如物种2000、全球生物多样性信息网络、生命条形码以及网络生命大百科全书。这些项目的成功不仅体现在对大量基础信息和数据的发布, 而且它们通过与生物多样性信息标准TDWG(Biodiversity Information Standards: TDWG)的合作, 推动了达尔文核心标准(Darwin Core)等一些重要的生物多样性信息标准的应用, 以及地区和国家性生物多样性信息节点的建立, 这些都为将来全球范围生物多样性信息的共享和数据交换奠定了重要基础。在数字化信息的基础上, 研究人员也开发了一些在特定研究领域应用的数据挖掘和模型工具, 例如基于数字化标本的地理分布预测工具MAXENT, 分类学专家知识管理的LifeDesk。公民科学理念的发展则向我们展示了公众和科学爱好者广泛参与以互联网为基础的生物多样性信息学研究活动。因此, 生物多样性信息学的发展前景广阔, 它将为我国实现全球保护战略目标, 应对生物多样性危机, 解决全球气候变化条件下生物多样性资源管理和利用建立坚实的信息基础。

关键词: 生物多样性, 信息学, 数据库, 物种, 标本, 地理信息系统

Progress in Biodiversity Informatics

Lisong Wang¹, Bin Chen¹, Liqiang Ji², Keping Ma^{1*}

1 Institute of Botany, Chinese Academy of Sciences, Beijing 100093

2 Institute of Zoology, Chinese Academy of Sciences, Beijing 100101

Abstract: Biodiversity Informatics is a young and rapidly growing field that brings information science and technologies to bear on the data and information generated by the study of biodiversity and related subjects. Recent years, biodiversity informatics community has made an extraordinary effort to digitize primary biodiversity data, and develop modelling tools, data integration, and county/regional/global biodiversity networks. In doing so, the community is creating an unprecedented global sharing of information and data produced by biodiversity science, and encouraging people to consider, survey and monitor natural biodiversity. Due to success of several international biodiversity informatics projects, such as Species 2000, Global Biodiversity Information Facility, Barcoding of Life and Encyclopedia of Life, digitized information on species inventories, herbarium specimens, multimedia and literature is available through internet. These projects not only make great contributions to sharing digitized biodiversity data, but also in prompting the implementation of important biodiversity information standards, such as Darwin Core, and in the establishment of regional and national biodiversity networks. These efforts will facilitate the future establishment of a strong information infrastructure for data sharing and exchange at a global scale. Besides focusing on browsing and searching digitized data, scientists should also work on building data mining and modeling, such as MAX-

收稿日期: 2010-04-19; 接受日期: 2010-05-30

基金项目: 中国科学院信息化专项项目(INFO-115-C01-SDB1-02)和国家自然科学基金(No. 30800057)

* 通讯作者 Author for correspondence. E-mail: kpma@ibcas.ac.cn

ENT for Ecological Niche Modelling and LifeDesk for taxonomist's knowledge management. At the same time, the idea of citizen sciences gains popularity showing us the benefit of the public working closely with the scientific community in completing internet-based biodiversity informatics activities. Therefore, biodiversity informatics has broad prospects, and is helping to build strong facilities that will aid in implementing the goals set by Global Plant Conservation Strategy and related international treaties, resolving biodiversity crises and the management of biodiversity resources in global climate change scenarios.

Key words: biodiversity, informatics, database, species, specimen, GIS

信息技术的发展对科学研究的影响是巨大的。从基础数据的采集、信息存储和分析到科学假设的提出和验证,所有科学研究的相关活动在计算机和网络技术的帮助下正逐渐实现自动化。数字地球、数字图书馆、数字地图等不仅仅是时髦的词汇,而是现代知识创新、知识转移和知识传播在信息时代的形象表达。

在生命科学领域中,已有的信息量几乎是无限的,并且依然保持着指数式的增长。以植物为例,通过几个世纪的研究积累,全世界现已认知的有花植物约22–40万种(Tangley, 1997; Govaerts, 2001, 2003; Groombridge & Jenkins, 2002; Thorne, 2002; Scotland & Wortley, 2003; Kier *et al.*, 2009);我国已知的高等植物有34,665种(The Biodiversity Committee of Chinese Academy of Sciences, 2009),占世界已知物种的10%左右。

植物物种是最基本和最重要的自然资源。物种及其存在凭证的重要原始材料是保藏在标本馆中的植物标本、植物园栽培的活植物、自然保护区和原始林区等自然植被中存在的物种。据粗略估计我国数百个标本馆共保存植物标本约1,700万份(傅立国, 1993),涵盖中国全部植物物种和世界物种约5万种;数百个植物园引种栽培我国的野生植物约2万种,引种国外物种约0.5万种。近年来国际上把基于高科技方法保存的植物器官(种子)、营养体、蛋白质和DNA大分子材料等,也都归为自然资源保护、研究和利用的重点之一。

目前,自然资源的研究管理已经进入迅猛发展的年代,随着宏观和微观生物学实验手段的改进,生物数据和信息呈爆炸式增长。传统的信息管理方式已显得力不从心。引入信息学及计算机科学的最新技术和成果,成为生物学家驾驭庞大数据信息的唯一选择,这一过程又带来工作模式的革新与观念的变化。信息技术的发展,例如大容量电子存

储媒介、因特网、分布式数据库技术和开放性数据共享平台(包括大范围的数字化工作、数据对公共普通用户的开放性使用),为生物多样性信息的产生、维持、传播和使用带来了革命性的发展(Bisby, 2000; Edwards *et al.*, 2000; Krishtalka & Humphrey, 2000; Oliver *et al.*, 2000; Smith *et al.*, 2000; Bisby *et al.*, 2002; Krishtalka *et al.*, 2002; Causey *et al.*, 2004; Brent & Bruck, 2006; Kagan, 2009; Paton, 2009)。正是在这样的背景下,生物多样性信息学(Biodiversity Informatics)蓬勃发展,并逐渐成长为一个信息技术和生物多样性研究相结合的新增长点(Canhos *et al.*, 2004; Graham *et al.*, 2004a; Godfray, 2007; Godfray *et al.*, 2007; Johnson, 2007; Sarkar, 2007, 2009; Page, 2008; Guralnick & Hill, 2009; Kagan, 2009)。

1 概念

生物多样性信息学指利用信息技术,对生物有机体基础数据(primary data)的管理和算法的探究、分析和解释,尤其是在物种水平上的应用(Soberon & Peterson, 2004; Berendsohn & Geoffroy, 2007; Johnson, 2007)。它涉及到对信息的收集、存储、提交、检索和分析,主要关注单个有机体、居群和分类群以及它们的相互关系,其内容覆盖了系统学、进化生物学、居群生物学、行为科学,以及从传粉生物学到寄生病学和植物群落学等领域。

作为信息学的一个分支,生物多样性信息学依赖计算机技术和信息学的普遍原则和技巧来处理异质性基础数据,同时也需新创一些方法来解决特殊性问题。例如:如何在信息系统中建立物种名称与分类群概念间的联系,从而完整记录分类学家多年来对某个物种认识的历史和变化过程;如何对早期的标本采集信息进行地标化(Georeferencing),从而可以在大的空间尺度上进行有效地空间分析和

比较; 如何为不同的数据和信息对象如物种、标本、图像和文献信息建立全球信息唯一标识符, 从而实现信息间的流畅共享和管理; 如何能有效促进人群间的互动合作, 促进人类对生物多样性的认知与针对性的活动, 扩大生物多样性信息获取的源头等。

生物多样性信息学和生物信息学(Bioinformatics)相互联系但又有区别。前者主要关注物种到生态系统水平(Sarkar, 2007, 2009; Paton, 2009)的生命现象, 而后者主要关注分子水平上的问题。也有研究者认为生物多样性信息学是生物信息学的一个研究方向(钟扬和张亮, 2000)。不管怎样, 理解生物多样性信息学研究及其与相关领域关系的问题在于: 随着数字化编目、数字化标本馆、数字化植物园等大量研究活动的涌现, 我们需要一个整体的知识框架来融合生物学家和技术专家的知识, 规范相关的研究活动, 一起来讨论、规划这一领域未来的发展蓝图。这也正是本文的主要目的, 即能够推动生物多样性信息学在国内的应用和发展。

目前, 生物多样性信息学最重要的研究目标就是为生物多样性信息的管理和一些重要议题如全球变化研究提供一个坚实的下层建筑(Infrastructure)。下层建筑在这里是一个宽泛的概念, 它不仅指全球、地区或国家性生物多样性信息系统的建立, 分布式共享网络体系的构建, 相关软件和工具的研制, 信息共享标准、协议和政策的制定; 也包括研究组织、机构和人才队伍的建设, 对生物多样性信息可持续增长、维持与利用的技术与平台提供支持。

2 发展历程

生物多样性信息学相关的实践活动可以追溯到20世纪70年代早期第一个计算机分类学数据库的建立。最早有文献记录的分类学数据库可能是美国弗吉尼亚海洋科学研究所开发的Biota of Chesapeake Bay数据库和澳大利亚联邦科学与工业研究组织(Commonwealth Scientific and Industrial Research Organisation, 简称CSIRO)开发的DELTA(Description Language for Taxonomy)系统(Dallwitz, 1974)。澳大利亚是生物多样性信息学研究的先行者之一, 70年代中期开始了标本馆标本的数字化工作; 1989年又发起了Environmental Resources Informa-

tion Network(ERIN), 来积累不同地理区域的环境数据。同时他们也发布了最早的标本馆数据共享协议和标准HISPID(Herbarium Information Standards and Protocols for Interchange of Data) (<http://plantnet.rbgsyd.nsw.gov.au/HISCOM/HISPID/HISPID-3/hispidright.html>)。到了80-90年代, 一些大型生物信息网络系统和数据库建立, 例如墨西哥的CONA-BIO(1992年)、哥斯达黎加的INBio(1989年)、生物多样性信息标准TDWG (Biodiversity Information Standards: TDWG, 简称TDWG) (1985年)、整合分类学信息系统(Integrated Taxonomic Information System, 简称ITIS)(1996年), 以及后来的物种2000项目(1996年)。进入21世纪, 生物学数据库和信息系系统出现爆发式增长。根据生物多样性信息学标准(TDWG)网站上的统计就有超过600多个生物多样性信息学研究项目(<http://www.tdwg.org/biodiv-projects/projects-database/>) (截至2010年3月)。数据和信息服务的内容也从单一的分类学信息转向覆盖分类学、系统学、生态学、进化生物学和生物地理学等生物多样性相关的研究领域。

2000年以后, “生物多样性信息学”这个词语及其涉及的相关研究活动开始更为正式地出现在科学研究群体。这一方面得益于科学家通过一些世界顶级学术刊物的专题栏目、增刊或者是学术会议专辑等连续报道。例如: *Science*在2000年289卷的一个专栏中以生物多样性的数字化、分类学的复兴、网络上的化石数据库、安静的革命—生物多样性信息学、互联网与生物多样性数据库的互操作性—生物多样性信息在我们的计算机桌面等5篇文章讨论了这个领域的相关问题(钟扬和张亮, 2000)。 *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*在2004年的“21世纪的分类学”专题中用19篇文章讨论了与生物多样性信息学相关的议题。 *BMC: Bioinformatics*则于2009年专门在“生物多样性信息学”专题下组织了10篇文章, 特别是DNA条码技术及其相关的研究进展丰富了生物多样性信息学的内容。另一方面, 近几年发展和运作非常成功的一些全球性生物多样性信息学研究项目(见表1)也让科学家们看到了这一领域发展的广阔前景。

2009年6月在英国举行的世界生物多样性信息

表1 一些国际上著名的生物多样信息学项目

Table 1 Important Biodiversity Informatics Projects in the world

信息系统名称/简称 System name (Abbreviation)	网址 Web address	核心数据 Core data
Catalogue of Life-Species 2000(COL-Species2000)	http://www.sp2000.org/	物种编目 Checklist
Global Biodiversity Information Facility (GBIF)	http://www.gbif.org/	标本 Herbarium specimen
Tree of Life (TOL)	http://tolweb.org/tree/	系统发育 Phylogenetic data
Barcoding of Life (CBOL)	http://www.barcodinglife.org/	DNA条形码 DNA barcode
Encyclopedia of Life (EOL)	http://www.eol.org/	综合物种信息 Integrated species information
Atlas of Living Australia (ALA)	http://www.ala.org.au/	综合物种信息 Integrated species information
International Plant Name Index (IPNI)	http://www.ipni.org/	植物名称 Plant name
USDA PLANTS	http://plants.usda.gov/	综合性植物物种信息 Integrated species information
UBio	http://www.ubio.org/	生物名称 Scientific name
Biodiversity Heritage Library (BHL)	http://www.biodiversitylibrary.org/	生物多样性历史文献 Biodiversity literature
ARKive	http://www.arkive.org/	珍稀濒危物种图像和多媒体信息 Multimedia of endangered species
MorphoBank	http://www.morphobank.org/	形态性状的系统发育分析 Morphological phylogeny

学国际会议 e-Biosphere'09(<http://www.e-biosphere09.org/>)则是生物多样性信息学领域的第一次国际性盛会。有来自69个国家的500多名代表参加, 齐聚了世界生物多样性信息学领域最著名的机构和科学家, 比如: 美国伍兹霍尔海洋生物学实验室(Marine Laboratory Woods Hole, USA)和美国斯密森研究院(Smithsonian Institution)。除了在一些主流的生物学或计算机刊物上介绍和探讨该领域的相关活动, 科学家们还创立了一些在线电子刊物, 例如: *Biodiversity Informatics*(2004年创刊), 来专门刊载这个领域的重要进展。

我国植物科学工作者在20世纪80年代初期, 敏锐地捕捉到国际植物自然科技资源领域数字化研究趋势, 并开始与国际同步进行生物物种和标本信息数字化的工作。代表性的工作包括: (1)中国科学院生物多样性信息系统(Chinese Biodiversity Information System, CBIS)(<http://cbis.brim.ac.cn/>), 由动物学、植物学、微生物学、内陆水体生物学和海洋生物学五个分部构成, 涉及各主要生物类群的物种编目、分类代码、名称、志书文献、濒危等级评估与保护策略等方面的信息; (2)中国动物信息网(<http://www.animal.net.cn/>), 主要包含动物分类系统、物种名称与编目、标本查询等资料; (3)中国微生物与病毒主题数据库(<http://www.micro.csdb.cn/>), 主要包括微生物与病毒编目、图像、菌种保藏等方面的信息; (4)在中国科技部平台项目“国家标本资源共享平台”的支持下, 近年相继建立了中国数字

植物标本馆(Chinese Virtual Herbarium, 简称CVH)、国家数字动物博物馆(<http://museum.ioz.ac.cn/>)、教学标本标准化整合与资源共享平台(<http://mnh.scu.edu.cn/>)、中国自然保护区资源平台(<http://www.papc.cn/>)、国家岩矿化石标本资源信息网(<http://www.nimrf.net.cn/>)、极地标本资源共享平台(<http://birds.chinare.org.cn/>)等6个全国性的多样性信息平台。

此外, 中科院植物研究所还初步建成了中国植物图像库(Plant Photo Bank of China, 简称PPBC)、中国自然标本馆(Chinese Field Herbarium: 简称CFH)、物种2000-中国节点高等植物信息系统(Catalogue of Life: Higher Plants in China, 简称CNPC)和中国数字植物园(Chinese Virtual Botanical Garden, 简称CVBG)等一批覆盖全国植物标本、图像、物种和引种植物的信息系统, 相关研究力量也通过“中国科学院植物研究所生物多样性信息学重点实验室”的组织形式加以整合, 形成目前国内规模最大的生物多样性信息学研究团队。

3 研究内容

如同生物多样性研究一样, 生物多样性信息学涉及到的内容宽广而复杂。从信息学角度说, 它涵盖了从基础数据的采集、存储、整合、管理、发布和分析等多个环节, 还涉及到各种数据共享政策和标准的制定。从传统研究领域说, 它涉及到分类学、系统学、进化生物学、生态学、保护生物学和生物

地理学等多个学科和研究领域。综合信息学的原则和所涉及生物学科的特点, 我们将从基础信息的数字化、信息化工具、标准和协议、模型工具以及国家和地区生物多样性网络这5个方面来介绍生物多样性信息学的主要内容。它们分别代表了从基础数据的收集整理、辅助性研究工具支持、信息交换和共享到数据挖掘和知识发现等信息化研究的几个重要环节, 也是生物多样性信息学研究最值得关注的几个方面。

3.1 基础信息的数字化

全球植物保护战略(Global Strategy for Plant Conservation, 简称GSPC)包括16个全球植物保护目标。为了有效地执行和实现这些目标, 需要获得从小地区到全球空间尺度, 涉及个体、居群、生态系统、遗传和有机体各个层次的信息和数据。Paton(2009)结合这些目标列举了潜在和已有的各种信息化资源, 指出要实现这些目标需要的信息不仅数量巨大而广泛, 并且需要这些信息相互联系, 从而可以对它们进行综合分析和评估。

因此对基础信息的数字化我们需要优先考虑涉及生物多样性基本问题最关键的信息成分, 例如物种及其地理分布。这两个方面的信息对回答如中国有多少植物物种、它们的地理分布格局如何、国家或地区尺度上应该采取什么样的保护策略等问题是非常关键的, 也是保护生物学研究经常涉及的问题, 是我们认知生物多样性最基本的信息元素(Sarkar, 2007)。这两方面的数据经常被国外学者称为生物多样性基础数据(primary data)(Soberon & Peterson, 2004; Chapman, 2005b; Chavan & Ingwersen, 2009)。此外, 我们认为生物多样性基础数据还应该包括图像和文献信息。因此, 物种、标本、图像和文献基本上构成了生物多样性信息学研究中最主要的信息组分。

3.1.1 物种信息

生命科学中最基本的科学问题是物种问题, 生物物种的研究决定着从宏观的全球生态系统至微观的生命DNA分子构造等全部内容, 更与国家目标中的农业、能源、信息、环境、人口与健康以及可持续发展等问题的研究和解决直接相关。物种信息这里使用较为宽泛的概念, 包括用来描述物种的学名、异名、俗名、形态特征、地理分布、珍稀濒危状况、经济利用价值、保护利用情况等基本信息。

物种信息中最为关键的是编目, 它是物种其他描述信息集的基础。据科学家们粗略估计, 地球上的37万多种高等植物大概涉及到90万个名称(Paton *et al.*, 2008), 如果使用一些过时的分类学信息可能会引起混乱。例如: 以前所认定的不少珍稀濒危物种是狭域分布, 后来发现, 它们实际上和一些原来的广布种是同种。IUCN在1997年的红皮书中曾列举了胡桃科的22个珍稀濒危物种, 其中9个现在已经确认实际上是一些广布种的异名。在科学文献中, 只有不到5%的植物学名在使用时给出了明确的分类学定义的引证和来源(Patterson *et al.*, 2006; Paton *et al.*, 2008)。因此, 不仅需要把应用到同一个物种的不同名字(同物异名)关联起来, 而且要把相同名字应用到不同植物(异物同名)的情况区别出来(Patterson *et al.*, 2006)。如果我们不对引用物种名字的分类学含义给出明确的依据和来源, 会给信息检索、整合和分析造成极大的困难。据估计, 异物同名的情况在植物中占据了大约4%。

全球植物保护战略的目标之一就是提供一个目前得到广泛认可的物种名录, 并将异名关联起来, 向那些应用了不同分类处理的信息源提供一个参考标准, 从而促进不同来源信息的整合。目前公认的世界植物物种名录已经完成了大约60%, 到2010年可能达到85%(Paton *et al.*, 2008)。物种2000是目前应用最为广泛的物种接受名和异名的信息来源, 已经成为检索物种信息的一个基本的索引框架。虽然目前已经建立了如上所述的全球性、地区性和类群为主的编目数据库, 但是《生物多样性公约》缔约方大会认为还应该在国家水平上增强这方面的投入, 加强分类学队伍建设。生物分类学倡议及其履行进展在这方面起到了重要作用(<http://www.cbd.int>)。

物种编目信息的数字化有非常悠久的历史, 从早期的邱园索引(Index Kewensis)到由许多国际组织联合建立的国际植物学名索引(International Plant Name Index, IPNI)、整合分类学信息系统(Integrated Taxonomic Information System, ITIS)、美国密苏里植物园的TROPICOS、全球生物物种名录-物种2000(Catalogue of Life-Species 2000)等全球性电子名录, 以及以类群为主的国际豆科植物信息系统(International Legume Database & Information Service, 简称ILDIS)、全球菊科植物名录(Global Compositae

Checklist)、世界茄属植物资源(World *Solanum* Resource)、鱼类数据库(FishBase)和以地区为主的非洲植物名录数据库项目(African Plant Checklist and Database Project, APCD)(Klopper *et al.*, 2007)、澳大利亚植物名称索引(Australia Plant Names Index, APNI)、北美植物志整合系统(Synthesis of the North American Flora, SNAF)等。

生物物种编目-物种2000项目是物种编目信息最有影响力的系统,它由国际生物科学联盟(The International Union of Biological Sciences, IUBS)发起,并联合科学技术数据委员会(The Committee on Data for Science and Technology, CODATA)、国际微生物联盟(The International Union of Microbiological Societies, IUMS)于1994年9月共同建立。它是一个与用户、分类学家和赞助单位保持紧密合作的联邦式数据库系统(Federal Database)。成立的目的是建立一个包含全球主要生物类群的物种工作名录(working list),并通过互联网和光盘向全世界发布,提供免费下载和使用。目前该数据系统已有77个加盟的专家数据库,包括了1,257,735个生物物种(<http://www.sp2000.org/>)。ITIS和世界生物多样性数据库(ETI-WBD)都是物种2000的重要合作伙伴。它还向全球生物多样性信息网络(GBIF)和网络生命大百科(EOL)等全球重要的生物多样性信息机构提供核心的物种名录数据。物种2000项目的具体目标是:针对目前已知生物物种,提供一个具有一致性、经认可且可成为实际使用工具的名录索引。

最近,由世界生物多样性信息学最为成功的GBIF发起了Global Name Architecture, GNA(<http://www.gbif.org/informatics/name-services/global-name-architecture/>)项目,目的是要建立一个多层次的名称数据体系,为将来更为广泛的生物学信息整合奠定基础。

3.1.2 标本信息

标本馆标本是表明某个物种在某个时间和地点存在的第一手证据,它提供的不仅仅是某个物种已知的分布信息,而且包括了大量历史性信息(Chapman & Busby, 1994)。它不仅是分类学家进行物种修订的凭据,也是开展生物多样性研究的原始材料(Canhos *et al.*, 2004; Graham *et al.*, 2004a; Soberon & Peterson, 2004; Chapman, 2005a)。据粗略估计,保藏在世界各大标本馆和研究机构的标本达

25-30亿份(Duckworth, 1993; Soberon, 1999),估计有5-10%已被数字化(Soberon, 1999)。互联网的发展,使大量数字化的标本信息可以被访问(Krishtalka & Humphrey, 2000; Causey *et al.*, 2004; Graham *et al.*, 2004a)。

数字化使这些标本得到了更加广泛的应用,如传统分类学修订,名录和鉴定手册的编辑,动、植物志的编纂,地理分布图的绘制,物种多样性空间格局和动态的分析,生活史和物候学的研究,珍稀濒危物种的保护生物学,外来入侵物种的管理等等(Funk & Richardson, 2002; Chapman, 2005b)。

数字化标本信息潜在价值的挖掘要归功于地理信息系统(GIS)技术、空间分析手段和大量免费获取的电子化环境数据的组合应用(Hijmans *et al.*, 2005; Swenson, 2008)。一方面,通过给大量历史标本进行地标化处理,可以实现点对点地绘制物种分布的详细信息图,相对于传统手工绘制点图和轮廓图方式来说,提高了制图的准确性。基于这种GIS化的分布图,还可以在统一的地理空间分析框架下展开大空间尺度上的比较生物地理学研究(Colwell & Coddington, 1994; Rahbek, 2005; Miller & Knouft, 2006; Kidd & Liu, 2008; Kozak *et al.*, 2008; Martinez-Freiria *et al.*, 2008)。另一方面,还可以利用这些地标化的标本信息,结合环境数据、气候变化、土地利用、植被变化等信息对物种的潜在地理分布进行预测,从而对入侵物种和珍稀濒危物种的管理,特别是在全球气候变化情景下的管理,作出更为科学的决策(Peterson *et al.*, 2002a; Peterson, 2003; Graham *et al.*, 2004b; Gaubert *et al.*, 2006; Irfan-Ullah *et al.*, 2006; Stigall & Lieberman, 2006; Rissler & Apodaca, 2007; Sergio *et al.*, 2007; Steiner *et al.*, 2008)。

要实现植物物种分布的空间分析和预测,首先需要对这些历史标本进行地标化,也就是给它们的采集地点配上经纬度坐标。对于历史采集信息,由于很多地点描述不规范,甚至信息完全缺失,以及早期的很多地名已经发生了变化等原因,使地标化成为一项非常复杂和艰巨的任务。多数情况下需要手工逐条核查并更正。现在已有一些工具来帮助处理地标化问题,例如BioGeomancer(Murphey *et al.*, 2004; Guralnick *et al.*, 2006)。

尽管我们已经大量数字化的标本信息可以

利用,但是由于有些标本在地理空间上的代表性不够,有些标本缺乏权威和准确的鉴定,再加上地标化的坐标信息不精确甚至错误,以及标本鉴定信息没有及时更新等原因,极大地限制了数字化标本信息的应用(Soberon & Peterson, 2004; Yesson, 2007)。

GBIF是目前世界上最大的生物标本信息门户,自2002年建立以来,已经有50多个国家,40多个国际组织参与建设,有超过1.7亿份标本信息,其中有经纬度记录的有花植物标本信息达2,700万份。纪力强等(2005)对GBIF作过详细的介绍。

3.1.3 图像和文献信息

图像信息在生物多样性信息学研究中有着非常特殊的价值。它不仅包括野外调查和采集过程中拍摄的活生物的数码图像,也包括历史标本的后期影像处理,以及数字化植物志书等资料中的科学绘画等内容。现代信息技术和数码摄像手段的发展已经赋予了生物图片新的内涵。它不仅展示生物物种在自然环境中的整体外貌,还能够不受存储空间限制详尽地记录物种形态特征的各个细节信息,例如:物种器官的颜色、质地,以及栖息地和行为等。这些信息经常是形态学描述、标本和科学绘画无法表现的。

国际上对图像信息非常重视,比如Discovery Life(<http://www.discoverylife.org>)收集了超过100万种生物的图片,帮助读者辨识物种。康奈尔大学的鸟类学实验室(<http://www.macaulaylibrary.org>)收集了12万条脊椎动物(主要是鸟类)的音频和4万条视频数据,供科学研究和科学普及使用。牛津大学的The Virtual Field Herbarium项目(<http://herbaria.plants.ox.ac.uk/vfh/about/>)提供了万余张植物关键鉴定特征的图像。The Linnaeus' Collections (<http://www.linnean-online.org/>)项目专门致力于Linnean早期研究的模式标本、手稿、通信等信息的图像采集,目前已经包括了1.4万余种植物,168种鱼类,3,198种昆虫,以及3,000多条信件和手稿的图像信息。近几年来,中国植物图像库(Plant Photo Bank of China, PPBC)在中国植物图像信息的收集和整理方面也取得了可喜的进展。目前已经收录了植物图片41万余幅,初步鉴定的有24万余张,分属301科2,523属,共计11,000多种(截至2010年1月)。

当然,我们也不应该忽视早期的许多植物科学绘画的价值。植物科学绘画是在科学研究的范畴之

内,运用绘画技法,科学、客观、艺术、真实地表达科学内容的创作过程,它是表现植物、认识植物的一个重要手段(孙英宝等,2008)。植物科学绘画承载了大量早期植物学研究历史的信息,是科学研究历史中非常有意义的文化遗产。对于不少物种来说,已经很少有机会在野外找到它们的踪迹,而经常是通过文献记录中的科学绘画来认知。综合了野外活植物和科学绘画等多种图片载体的影像信息集将来或许能够成为记录物种多样性图像博物馆。

大量生物多样性研究的基础信息还深藏在各种文献之中。近年来,已有不少相关的文献信息得到数字化(Davidson *et al.*, 1997)。目前与生物多样性信息学研究最为密切的是生物多样性遗产图书馆(BHL)(<http://www.biodiversitylibrary.org/>)项目。该项目最初由American Museum of Natural History、The Field Museum、Harvard University Botany Libraries、Missouri Botanical Garden、The New York Botanical Garden、Royal Botanic Gardens和Smithsonian Institution等10家单位共同发起。BHL项目的开始阶段主要将那些不涉及版权的老文献数字化。随着时间推移,不涉及版权的文献会逐步增加,即所谓的墙体移动策略(moving wall strategy)。通过与Internet Archive(<http://www.archive.org/>)的合作,BHL目前已经数字化了71,000多册书,并建立了方便的学名、关键词、页码检索方式。这些早期的每一篇历史文献都可能包括着一个甚至多个生物有机体相关的信息。将这些信息与目前最新的分类学研究结果建立关联,我们就有可能完整记录和追溯特定研究对象的历史研究过程和动态。

3.1.4 野外生物多样性实体信息

传统上保存于标本、图像、文献等实物信息媒介中的生物多样性基础信息,需要进行数字化、地标化之后才能在各种生物多样性信息系统加以利用,生物多样性采集与利用效率受到很大限制。相对于这个蓝色地球上现存的各层次的生物多样性,人类已经采集的生物多样性基础信息仅是沧海一粟,有更多自然生存着的生物多样性有待被记录与发现。

相对于林耐与达尔文时代,当今发达的交通系统、成熟的全球定位系统、多媒体信息采集技术以及以数据库与互联网为代表的信息技术,使人们可以将整个地球看作活的生物多样性博物馆,直接将

野外个体、居群、群落等各种水平的生物实体及其变化过程作为研究对象, 将它们的形态、生态信息与地理分布坐标数字化, 形成数字标本。数字标本以数码照片为主体, 同时包含GPS坐标、调查路线的GPS轨迹、采集与鉴定相关信息、其他信息的录音记录, 是综合的数据集, 具有信息量大、效率高的特点。通过广泛的公民科学协作, 可以发动大量志愿者通过采集数字标本的方式参与到生物多样性的调查与监测之中, 极大提高生物多样性信息的采集与加工效率, 将人类的认知活动与野外生物多样性直接关联, 同时满足人们的科普教育、科研基础数据采集、保育活动的前期调研与成效评估等多方面的需求。作为这些理念的实践, 中国自然标本馆(Chinese Field Herbarium, CFH)受到了用户的欢迎, 在两年时间积累了大量的资料, 包括86.5万张数码照片, 其中约30万张具有精确的GPS坐标, 已鉴定植物种类超过1.2万种。自2009年6月份以来, CFH日均新增照片2,000多张, 新增鉴定物种30多种, 成为中国增长最快的生物多样性原始数据源。

3.2 信息化工具

传统上我们对物种的鉴定主要依赖标本馆标本、图像和植物志书上的检索表或请专家咨询。科学家们也开发出一些电子化的检索工具, 它们早期都被称为“专家系统”。过去40多年来, 已有许多这样的工具被开发出来(表2)。DELTA可能是最早和应用最为广泛的电子检索工具(Dallwitz, 1993; 李健均, 1996; 陈翔和陈训, 2008; 张明理, 2009), Lucid是后来形成的一个商业化产品, 主要在昆虫学上有比较多的应用。由于早期计算机技术的限制,

这些工具都采用一些专有的数据存储格式, 并且只能在个人计算机上使用。后来的研究者开始开发一些能够在互联网上使用, 便于不同系统交换数据的工具。比如eFlora的交互式检索表(Brach & Song, 2006)。这些工具很多都采用了TDWG推荐的结构化描述性数据(Structured Descriptive Data, SDD)标准, 并且能够和早期的DELTA数据兼容。

物种鉴定的本质是人们对物种实体资料的准确认知。除了将检索表做成各种类型的鉴定工具外, 建立信息化的互动平台实现物种调查与认知能力的共享具有重大现实意义, 通过参与者在资源调查与物种鉴定上拥有的不同能力的互补合作, 能更好地满足好奇心、提高认知水平、促进新物种资料的采集与整理鉴定。仿照传统标本馆的鉴定流程实现的网上互动合作, 将是物种鉴定的重要形式。

近年来, 网络信息技术的发展使我们不仅仅是开发一些这样的辅助性工具, 而是更多地考虑如何将研究实践的各个环节都纳入信息化管理。尤其是对分类学这样一个面临大量基础数据积累并不断需要更新信息的研究来说更是如此。因此, 研究人员开始考虑信息环境下分类学信息化能力建设的问题(Wheeler, 2004; Wheeler *et al.*, 2004; Raven, 2004; Scoble, 2004; Miller, 2005; Patterson *et al.*, 2006; Tillier & Roberts, 2006; Alves *et al.*, 2007; Godfray *et al.*, 2007; Johnson, 2007; Walter & Winterton, 2007; Franz *et al.*, 2008; Mayo, 2008), 提出了如Cybertaxonomy的理念(Scoble, 2004; Berendsohn & Geoffroy, 2007; Mayo, 2008), 并开展了不少实践活动。例如: 由英国自然环境研究委员

表2 电子检索工具

Table 2 Tools of Electronic Taxonomy Key

工具名称 Name of the tool	网址 Web address
ETI Linnaeus II/III	http://www.eti.uva.nl/products/linnaeus/download.php
Lucid	http://www.lucidcentral.com/
XPer2	http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2
DeltaAccess	http://www.diversityworkbench.net/OldModels/Descriptions/index.html
DELTA	http://delta-intkey.com/
EFG (Electronic field guide)	http://efg.cs.umb.edu/efg/
navikey	http://www.navikey.net/
SLIKS	http://www.stingersplace.com/SLIKS/
X:ID	http://names.ubio.org/xid/
eFlora/Actkey	http://www.efloras.org/index.aspx

会(Natural Environment Research Council, NERC)资助, 大英自然历史博物馆(The Natural History Museum, London)、牛津大学(University of Oxford)和邱园(Royal Botanic Gardens Kew)联合发起的CATE (Creating a Taxonomic E-Science, <http://www.cate-project.org/>)项目、世界茄科植物数据库(<http://www.nhm.ac.uk/research-curation/research/projects/solanaceasource/>)项目、世界禾草数据库GrassBase (<http://www.kew.org/data/grasses-db.Ht-ml>)项目和欧洲分类学研究中心(European Distributed Institute of Taxonomy, EDIT)等。

在这些项目的实践中也产生了一些专门为分类学家提供服务的信息整合和管理工具, 比较著名并得到广泛应用的是EDIT开发的Scratchpad(Smith *et al.*, 2009)(<http://scratchpads.eu/scratchpad-taxonomy>)和EOL支持开发的LifeDesk(<http://www.Lifedesks.org/>)。这两个工具都是基于开源的Drupal内容管理系统进行的改造。CATE项目也正致力于开发出一个所有分类学家都可以使用的网络在线分类学修订管理系统的模板。这些工具与早期鉴定工具的不同在于, 它们从一体化分类学研究工作程序出发, 在工具的设计和使用上充分考虑到分类学家对名录、文献、标本、图像等基础数据资源的利用方式, 及研究成果的在线发布。应该说这些为专家们量身定做的工具已经为我们进入电子分类学研究时代奠定了技术基础。

除这些鉴定和专家使用的分类学数据整合工具之外, 生物多样性信息遍及人类生产、生活的诸多方面, 由人们在各种野外调查探索与研究认知等各种互动合作性的活动过程中产生。如果将这个过程通过信息学解决方案让野外数据采集、物种鉴定、数据组织等相关的活动尽量实现信息化, 则使数据能更快地积累, 更好地组织和有效地利用, 以便更好地服务于相关知识的普及以及分类、进化、生态、保育、驯化等学科的研究, 让喜欢、研究与利用生物多样性的人都能从高效率的信息化工作流程, 有效积累和组织的数据库, 以及互惠共享的交流氛围中获益。因此, 覆盖生物多样性信息野外采集、自动化组织、互动合作整理等以生物多样性信息业务流程信息化为目标的技术体系与支撑系统, 是建立更高效获取与利用生物多样性信息的必然要求。中国自然标本馆(CFH)采用包含数字化

地标准化野外生物多样性信息采集技术体系、网上互动鉴定、自动化数据组织与用户互动空间等内容的信息化平台, 展示了广阔的发展与应用前景。

3.3 信息标准

与其他信息学相关的学科一样, 生物多样性信息学也需要采用相关的数据和共享协议标准, 实现数据库间数据传输和信息互操作, 加速信息交换和共享。这个过程中一般涉及到三类标准: 一是数据标准(data standards), 它处理的问题是我们需要共享什么样的数据; 二是协议(protocols), 即这些信息如何进行共享; 三是元数据标准(metadata standards), 即向用户提供一个信息摘要, 告知用户信息系统存储的是什么样的信息, 可以提供哪些信息。数据标准是首要和最基础的, 它是实现协议和元数据标准的前提。

TDWG在这方面做出了非常卓越的贡献。这个组织早期主要致力于分类学数据库的标准建设, 随着生物多样性数字化的发展, 也开始研究生物多样性信息方面的标准。这个组织到目前已经发布了涉及标本采集信息的达尔文核心标准(Darwin Core, DwC)和ABCD(Access to Biological Collection Data), 涉及植物学名和描述信息的TCS (Taxonomic Concept Transfer Schema)和SDD, 涉及植物地理分布记录的WGSRPD(World Geographical Scheme for Recording Plant Distributions)标准, 已经整合数据和协议的SPM(Species Profile Model)、DiGIR、TAPIR等多个标准。

近年来, 科学家越来越重视相关标准的实践问题。国内已经成立了专门处理生物学信息化标准的组织, 即全国生物信息标准化技术委员会, 由国家标准化管理委员会领导和管理, 来帮助相关项目和研究处理生物学信息标准及其相关的问题, 以促进和加强生物信息资源跨国界、跨部门的整合、共享、服务和应用为目的。

为了使信息化数据在不同应用环境和研究领域发挥最大的价值, 数据提供者、管理者和使用者都需要考虑采用统一的语言来描述基础信息和数据。

3.4 模型工具

一般来说, 现存的生物多样性基础数据本身并不能直接为生物多样性资源的管理和评估提供有效的决策依据。我们还需要经过筛选、统计、分析、

建模以及相应的推理步骤才能变为有用的信息,从而为我们评估现存生物多样性知识的缺失,制定将来的研究计划,评估保护的优先性等提供参考。物种地理分布的生态位模型(Ecological Niche Model, EMs或ENM)是一个在生态学和生物多样性信息学中快速成长的研究方向(Peterson *et al.*, 1999; Peterson *et al.*, 2002c)。

生态位模型的基本原理是利用地标化的标本分布信息,结合这些物种已知分布点的环境变量信息,来模拟物种的生态需求(生态位)。尽管生态位模型方法还处在发展的过程中,但科学家们对模型方法的不断改善和广泛应用(Peterson, 2003; Martinez-Meyer & Peterson, 2006; Wang & Wang, 2006; Sergio *et al.*, 2007; Steiner *et al.*, 2008; Kozak *et al.*, 2008; Martinez-Freiria *et al.*, 2008; Nosil & Sandoval, 2008; Engler & Guisan, 2009),已经向我们展示了它在研究全球气候变化对陆地和海洋生物多样性影响(Peterson *et al.*, 2002a, b; Thuiller *et al.*, 2006; Sergio *et al.*, 2007),外来入侵物种的管理(Steiner *et al.*, 2008),物种的界定(Raxworthy *et al.*, 2007; Rissler & Apodaca, 2007),新物种的发现(Peterson *et al.*, 2002a),以及进化生物学(Carstens & Richards, 2007)等诸多研究方向上的巨大潜力(Peterson *et al.*, 2002a; Martinez-Meyer & Peterson, 2006; Stigall & Lieberman, 2006; Sergio *et al.*, 2007; Steiner *et al.*, 2008)。

生态位模型有许多可以利用的算法和技巧,比如 BIOCLIM(Nix, 1986), GLM (Generalized linear models)(Austin *et al.*, 1994), GAM(Generalized additive models) (Yee & Mitchell, 1991), CART (Regression and classification tree analyses)(Breiman

et al., 1984), Genetic algorithms (Stockwell & Peters, 1999)和ANN (Artificial neural networks)(Olden & Jackson, 2002)。针对不同的数据类型和算法,不少开发人员也设计了许多工具(表3)。我们还需要设计出可以方便使用不同模型,并且对不同模型结果进行比较分析的界面,从而使研究人员能够更多关注对结果的分析 and 解释。未来的发展是将数据输入和投影图层的生成放在一个连续的工作流程和框架下实现,并且可以通过远程分布式计算和网格技术加快大量数据的处理(Canhos *et al.*, 2004)。

3.5 全球、地区和国家生物多样性网络

在欧洲委员会的支持下,欧洲建立了几个比较大的生物多样性网络。CETAF(Consortium of European Taxonomic Facilities) (<http://www.cetaf.org/>)是联合了欧洲自然历史博物馆、植物园和其他生物学采集保藏机构的最大分类信息平台。它包括大量动物、植物、古生物和地质学采集,向多学科的研究人员提供服务。它致力于建立欧洲动物志的语义网络体系,有400多位动物学专家参与。欧洲动物志包括了115,000种非海洋动物物种及其地理分布等信息。这些信息和ERMS(European Register of Marine Species)相互补充,完善了欧洲动物物种的基础信息。植物方面是欧洲植物志及欧洲-地中海植物多样性数据库(European Flora and the Euro+Med Plant Database)。欧洲自然历史标本信息网络 ENHSIN(European Natural History Specimen Information Network)项目的目标则是在欧洲研究机构间建立起标本信息共享、互操作的信息基础。BioCASE(Biological Collection Access Service for Europe)(<http://www.biocase.org/>)联合了欧洲30个国家的35个研究机构通过网络提供生物学采集信息

表3 常用的生态位模型工具

Table 3 Ecological niche modelling tools

软件名字 Name of the tools	网址 Web address
Maxent	http://homepages.inf.ed.ac.uk/lzhang10/maxent.html
GARP	http://www.nhm.ku.edu/desktopgarp/
Biomapper	http://www2.unil.ch/biomapper/
BIOCLIM	http://fennerschool.anu.edu.au/publications/software/anuclim/doc/bioclim.html
openModeller	http://openmodeller.sourceforge.net/
Lifemapper	http://www.lifemapper.org/
AquaMaps	http://www.aquamaps.org/
DIVA-GIS	http://www.diva-gis.org/

的服务。欧洲生物多样性信息网络ENBI (European Network for Biodiversity Information)(<http://www.enbi.info/forums/enbi/index.php>)项目发起于2003年,由欧盟给予经费支持,通过3年的工作完成语义网络建设项目。它实际上也是GBIF在欧洲最大的数据提供者。目前ENBI已经包括了欧洲24个国家的66个机构成员,其中包括GBIF国家节点成员。

美洲各国间生物多样性信息网络IABIN (Inter-American Biodiversity Information Network)(<http://www.iabin.net/>)是由世界银行和全球环境基金等多个机构和组织资助的覆盖西半球的生物多样性基础数据信息网络。REMIB(World Biodiversity Information Network)(http://www.conabio.gob.mx/remib_ingles/doctos/acerca_remib_ing.html)是一个覆盖墨西哥的生物采集信息网络,目前已覆盖140多个国家,拥有超过600万条的生物标本信息。澳大利亚虚拟标本馆AVH (Australian Virtual Herbarium)(<http://www.anbg.gov.au/avh/>)是澳大利亚植物学研究群体组成的在线网络标本馆体系。除了通过GIS应用程序已向用户提供650万份植物标本信息外,它的最终目标是形成一个电子化的植物区系研究网络,向政府、公众提供决策和信息服务。AVH是一个真正意义上的分布式系统,在这个平台上查询的标本数据实际上都由各自标本馆来管理和发布。

在国家科技部、环保部、中国科学院等多个部门的组织和支持下,经过多年努力,也建立了动植物生物多样性信息节点(<http://www.biodiv.gov.cn/>)、中国生物多样性信息系统(<http://cbis.brim.ac.cn/>)、中国森林生物多样性监测网络(<http://www.cfbi.org/>)、中国生态系统研究网络(<http://www.cern.ac.cn/0index/index.asp>)、中国自然保护区(<http://www.nre.cn/>)、中国数字植物标本馆(Chinese Virtual Herbarium, CVH)、中国植物图像库、中国自然标本馆、中国高等植物信息系统和中国数字植物园(Chinese Virtual Botanical Garden, CVBG)、中国植物物种信息网(<http://www.plants.csdb.cn/>)等一大批生物多样性基础信息系统。

但是我们应该看到在全球、地区和国家水平上的生物多样性网络体系所面临的一些挑战。为了有效地执行各种生物多样性保护行动计划,研究人员需要获得从小地区到全球空间尺度,包含时间序列,涉及个体、居群、生态系统、遗传和有机体各

个层次的信息和数据,并且需要对多个来源不同层次的数据进行评估和分析。在全球范围来看,这些信息目前还没有被有效地组织起来发挥它们应有的作用(Paton, 2009)。首先我们需要建立有效的数据共享机制和标准来容纳来自不同研究领域和学科以及不同研究层次和异质性的数据源;其次要建立高效率的工作平台与技术体系去更好地获取最新的基础信息;最后我们还需要开发出一些普遍适用的工具和系统。这对生物学家和计算机人员都是一个巨大的挑战。

近年来,由世界40多个机构和组织共同合作和参与的网络生命大百科全书(EOL)项目向我们展示一个可以在全球、地区和国家水平上使用统一的工具、标准来整合分散的生物多样性基础信息的机会。EOL项目的发起来自有生物多样性研究之父之称的Wilson(2003)的一个简洁而明确的想法。那就是为地球上已知的每一个生物物种建立一个网页,这个网页将展现该物种从分类、形态、地理分布、生态、进化等宏观研究,到DNA条形码、分子、遗传等微观研究,以及利用和保护等百科全书式的信息集合,并且通过文本、图像、多媒体、GIS等多种技术手段来展现丰富的信息。除了专业研究领域,一个不容忽视的群体是社会公众。公众通过公民科学的形式参加生物多样性信息的采集已经有几百年的历史(Schnoor, 2007, Bonney *et al.*, 2009, Silvertown, 2009)。世界各国社会公众关注与参与生物多样性调查的热情逐渐提高,已形成巨大的生物多样性信息采集与加工的潜在力量,专门为公民科学提供的野外调查工具(Luther *et al.*, 2009)与网络数据积累平台也逐渐出现。在兼顾专业研究和公众参与方面,EOL是值得肯定的。它是一个专家监管和公众参与的开放式系统。通过有效的专家体系保证数据的科学性和可靠性,同时也为对生物多样性保护和利用有热情的公众和爱好者敞开了参与经院式研究的大门。

无论是从宏观上还是微观上研究生物有机体,最根本的总是与具体的研究对象相联系,这个具体对象就是生物物种。对于浩瀚的生物多样性信息来说,物种是联系所有信息的一个纽带。因此,EOL项目找到了一条整合多个学科、研究领域、研究层次和不同来源信息的机制和方法。而且EOL及其相关的软件工具都是采用开源形式,它所移植的分布式

共享技术、Web2.0元素、MVC (Model-View-Controller)开发模式等使它成为一个灵活、可扩展的门户系统。尽管EOL项目目前还在发展的初始阶段,但它在研究思路和技术条件上的优越性向我们展示了一个在全球范围使用统一的标准和协议共享生物多样性信息的美好前景。

4 结束语

信息技术的发展给我们很多新的方法来共享生物多样性的基础知识。一些世界性的研究项目,比如网络生命大百科全书(EOL)、生物条形码协会(CBOL/BOL)、全球生物物种名录(COL)、生命之树(TOL)、全球生物多样性网络(GBIF)在数字化生物多样性信息基础方面已经建立了比较好的框架。这些项目在世界范围的广泛成功展示出信息化时代生物多样性信息采集、共享和利用的新蓝图,并且将对科学研究群体和社会有深远的影响。这些影响包括了新物种的发现、珍稀濒危物种保护策略的制定、新药用植物的研发、人类在复杂生命网络体系中的重要角色的认识等等。

生物多样性信息学是一个快速生长的领域。它把信息科学和相关的技术带到了生物学领域,促进了基础生物多样性知识更广泛应用。我们将看到一个史无前例的,在全球范围通过广泛的合作实现对自然生物多样性的信息采集、调查与监测,自由获取生物在基因、有机体、物种、居群、生态过程、生态系统以及人类在生物多样性系统中的作用等广泛的信息的局面。

然而,我们也要认识到生物多样性信息学研究是一个长期而艰巨的任务,还有许多未知问题值得我们去探究和解决。例如:我们需要建立一种长期机制,保证从个人和公共数据库定期收集和更新数据;通过同行评议(peer review)和自动校验方式来实现对数据质量的控制,建立对新增加信息的索引、链接和自动提交的处理流程,向用户提供不仅是数据浏览和查询,而且还包括深层次的数据操作和分析的界面等。

我们这一代比较完整地理解了我们所面临的生物多样性危机。同时,我们也可能是最后有机会去探究和记录我们这个地球上物种多样性的一代。我们这一时代最大的挑战之一是用数字化的方法和手段来记录我们所栖居星球的生物多样性及其

相关的知识遗产。

参考文献

- Alves RJV, Filho V, Machado MD (2007) Is classical taxonomy obsolete? *Taxon*, **56**, 287–288.
- Austin MP, Nicholls AO, Doherty MD, Meyers JA (1994) Determining species response functions to an environmental gradient by means of a beta-function. *Journal of Vegetation Science*, **5**, 215–228.
- Berendsohn WG, Geoffroy M (2007) Networking taxonomic concepts- uniting without “unitaryism”. In: *Biodiversity Databases: From Cottage Industry to Industrial Networks. Systematics Association Special* (eds Curry G, Humphries C), pp. 13–22. Taylor & Francis, Boca Raton.
- Bisby FA (2000) The quiet revolution: biodiversity informatics and the internet. *Science*, **289**, 2309–2312.
- Bisby FA, Roskov YR, Orrell TM, Nicolson D, Paglinawan LE, Bailly N, Kirk PM, Bourgoin T, Baillargeon G (2009) *Species 2000 & ITIS Catalogue of Life: 2009 Annual 2000 Checklist*. www.catalogueoflife.org/annual-checklist/2009/. (Accessed April 2009)
- Bisby FA, Shimura J, Ruggiero M, Edwards J, Haeuser C (2002) Taxonomy, at the click of a mouse. *Nature*, **418**, 367.
- Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, **59**, 977–984.
- Brach AR, Song H (2006) eFloras: new directions for online floras exemplified by the Flora of China Project. *Taxon*, **55**, 188–192.
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) *Classification and Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- Brent R, Bruck J (2006) 2020 Computing: can computers help to explain biology? *Nature*, **440**, 416–417.
- Canhos VP, Souza S, Giovanni R, Canhos DAL (2004) Global biodiversity informatics: setting the scene for a ‘new world’ of ecological forecasting. *Biodiversity Informatics*, **1**, 1–13.
- Carstens BC, Richards CL (2007) Integration coalescent and ecological niche modelling in comparative phylogeography. *Evolution*, **61**, 1439–1454.
- Causey D, Janzen DH, Peterson AT, Vieglais D, Krishtalka L, Beach JH, Wiley EO (2004) Museum collections and taxonomy. *Science*, **305**, 1106–1107.
- Chapman AD (2005a) Uses of primary species-occurrence data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- Chapman AD (2005b) Uses of primary species-occurrence data. In: *Global Biodiversity Information Facility*, Copenhagen, pp. 106.
- Chapman AD, Busby JR (1994) Linking plant species information to climate modelling for continental biodiversity inventory and monitoring. In: *Mapping the Diversity of Nature* (ed. Miller RI), pp. 177–195. Chapman & Hall, London.

- Chavan V, Ingwersen P (2009) Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics*, **10**, S2.
- Chen X (陈翔), Chen X (陈训) (2008) Application of new DELTA system in plant taxonomy: study on *Festuca L.* as an example. *Guihaia* (广西植物), **28**, 759–763. (in Chinese with English abstract)
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B*, **345**, 101–118.
- Dallwitz MJ (1974) A flexible computer program for generating identification keys. *Systematic Zoology*, **23**, 50–57.
- Dallwitz MJ (1993) DELTA and INTKEY. In: *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision* (ed. Fortuner R), pp. 287–298. Johns Hopkins University Press, Baltimore, Maryland.
- Davidson SB, Overton C, Tannen V, Wong L (1997) BioKleisli: a digital library for biomedical researchers. *International Journal on Digital Libraries*, **1**, 36–53.
- Duckworth WD (1993) *Preserving Natural Science Collections: Chronicle of Our Environmental Heritage*. National Institute for the Conservation of Cultural Property, Washington, DC.
- Edwards JL, Lane MA, Nielsen ES (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, **289**, 2312–2314.
- Engler R, Guisan A (2009) MigClim: predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**, 590–601.
- Franz NM, Peet RK (2009) Towards a language for mapping relationships among taxonomic concepts. *Systematics and Biodiversity*, **7**, 5–20.
- Franz NM, Peet RK, Weakley AS (2008) On the use of taxonomic concepts in support of biodiversity research and taxonomy. In: *The New Taxonomy. Systematics Association Special* (ed. Wheeler QD), pp. 63–86. Taylor & Francis, Boca Raton.
- Fu LG (傅立国)(1993) Index Herbarium Sinicorum (中国植物标本馆索引). China Science and Technology Press, Beijing. (in Chinese)
- Funk VA, Richardson KS (2002) Systematic data in biodiversity studies: use it or lose it. *Systems Biology*, **51**, 303–316.
- Gaubert P, Papes M, Peterson AT (2006) Natural history collections and the conservation of poorly known taxa: ecological niche modeling in central African rainforest genets (*Genetta* spp.). *Biological Conservation*, **130**, 106–117.
- Godfray HCJ (2007) Linnaeus in the information age. *Nature*, **446**, 259.
- Godfray HCJ, Clark BR, Kitching IJ, Mayo SJ, Scoble MJ (2007) The web and the structure of taxonomy. *Systematic Biology*, **56**, 943–955.
- Govaerts R (2001) How many species of seed plants are there? *Taxon*, **50**, 1085.
- Govaerts R (2003) How many species of seed plants are there? A response. *Taxon*, **52**, 583–584.
- Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004a) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**, 497–503.
- Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004b) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781–1793.
- Groombridge B, Jenkins MD (2002) *World Atlas of Biodiversity*. University of California Press, Berkeley.
- Guralnick R, Hill A (2009) Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics*, **25**, 421–428.
- Guralnick RP, Wieczorek J, Beaman R, Hijmans RJ, the BioGeomancer Working Group. (2006) BioGeomancer: automated georeferencing to map the world's biodiversity data. *PLoS Biology*, **4**, e381.
- Hijmans RJ, Cameron SE, Parra JL (2005) Very high resolution interpolated global terrestrial climate surfaces. *International Journal of Climatology*, **25**, 1965–1978.
- Irfan-Ullah M, Giriraj A, Murthy MSR, Peterson AT (2006) Mapping the geographic distribution of *Aglaia bourdillonii* Gamble (Meliaceae), an endemic and threatened plant, using ecological niche modeling. *Biodiversity and Conservation*, **16**, 1917–1925.
- Ji LQ (纪力强), Qiao HJ (乔慧捷), Xie BG (谢本贵), Zhang SW (张尚武), Lin B (林斌), Zhu H (朱慧), Deng H (邓浩), Li N (李诺), Han Y (韩艳) (2005) Global biodiversity information facility: organization, activities, projects and information service. In: *Advances in Biodiversity Conservation and Research in China* (中国生物多样性保护与研究进展VI) (ed. Ma KP (马克平)). China Meteorological Press, Beijing. (in Chinese with English abstract)
- Johnson NF (2007) Biodiversity Informatics. *Annual Review of Entomology*, **52**, 421–438.
- Jones MB, Schildhauer MP, Reichman OJ, Bowers S (2006) The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 519–544.
- Kagan JS (2009) Biodiversity informatics: challenges and opportunities for applying biodiversity information to management and conservation. *Northwestern Naturalist*, **87**, 80–85.
- Kidd DM, Liu X (2008) geophylobuilder 1.0: an arcgis extension for creating 'geophylogenies'. *Molecular Ecology Resources*, **8**, 88–91.
- Kier G, Kreft H, Lee TM, Jetz W, Ibsch PL, Nowicki C, Mutke J, Barthlott W (2009) A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences, USA*, **106**, 9322–9327.

- Klopper RR, Gautier L, Chatelain C, Smith GF, Spichiger R (2007) Floristics of the angiosperm flora of Sub-Saharan Africa: an analysis of the African Plant Checklist and Database. *Taxon*, **56**, 201–208.
- Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology and Evolution*, **23**, 141–148.
- Krishtalka L, Humphrey PS (2000) Can natural history museums capture the future? *Bioscience*, **50**, 611–617.
- Krishtalka L, Peterson AT, Vieglais DA, Beach JH, Wiley EO (2002) The Green Internet: a tool for conservation science. In: *Conservation in the Internet Age: Strategic Threats and Opportunities* (ed. Levitt JN), pp. 143–164. Island Press, Washington, DC.
- Li JJ (李健均) (1996) Delta system: an international standard for processing plant taxonomic description. *Acta Phytotaxonomica Sinica* (植物分类学报), **34**, 447–452. (in Chinese with English abstract)
- Ludascher B, Lin K, Bowers S, Jaeger-frank E, Brodaric B, Baru C (2006) Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience, Special Issue on Scientific Workflows*, **18**, 1039–1065.
- Luther K, Counts S, Stecher KB, Hoff A, Johns P (2009) Pathfinder: an online collaboration environment for citizen scientists. In: *27th Annual CHI Conference on Human Factors in Computing Systems*, pp. 239–248. Association for Computing Machinery, Boston, MA.
- Martinez-Freiria F, Sillero N, Lizana M, Brito JC (2008) GIS-based niche models identify environmental correlates sustaining a contact zone between three species of European vipers. *Diversity and Distributions*, **14**, 452–461.
- Martinez-Meyer E, Peterson AT (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography*, **33**, 1779–1789.
- Mayo SJ (2008) Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis. *Kew Bulletin*, **63**, 1.
- Miller AJ, Knouft JH (2006) GIS-based characterization of the geographic distributions of wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* (Anacardiaceae). *American Journal of Botany*, **93**, 1757–1767.
- Miller G (2005) TAXONOMY: Linnaeus's legacy carries on. *Science*, **307**, 1038–1039.
- Muggleton SH (2006) 2020 Computing: exceeding human limits. *Nature*, **440**, 409–410.
- Murphey PC, Guralnick RP, Glaubitz R, Neufeld D, Ryan JA (2004) Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database-Informatics Initiative (Mapstedi). *Phyloinformatics*, **3**, 1–29.
- Nix HA (1986) A biogeographic analysis of Australian elapid snakes. In: *Atlas of Elapid Snakes of Australia Australian Flora and Fauna Series Number 7* (ed. Longmore R), pp. 5–15. Australian Government of Publishing Service.
- Nosil P, Sandoval CP (2008) Ecological niche dimensionality and the evolutionary diversification of stick insects. *PLoS ONE*, **3**, e1907.
- Olden JD, Jackson DA (2002) Illuminating the 'black box': A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, **154**, 135–150.
- Oliver I, Pik A, Britton D, Dangerfield JM, Colwell RK, Andrew JB (2000) Virtual biodiversity assessment systems. *BioScience*, **50**, 441–450.
- Page RDM (2008) Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics*, **9**, 345–354.
- Paton A (2009) Biodiversity informatics and the plant conservation baseline. *Trends in Plant Science*, **14**, 629–637.
- Paton AJ, Brummitt N, Govaerts R, Harman K, Hinchcliffe S, Allkin B, Lughadha EN (2008) Towards target 1 of the global strategy for plant conservation: a working list of all known plant species progress and prospects. *Taxon*, **57**, 602–611.
- Patterson DJ, Remsen D, Marino WA, Norton C (2006) Taxonomic indexing: extending the role of taxonomy. *Systematic Biology*, **55**, 367–373.
- Peterson AT (2003) Predictability of the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, **78**, 419–433.
- Peterson AT, Ball LG, Cohoon KP (2002a) Predicting distributions of Mexican birds using niche modelling methods. *Ibis*, **144**, e27–e32.
- Peterson AT, Ortega-Huerta MA, Bartley J, Sanchez-Cordero V, Soberon J, Buddemeier RH, Stockwell DRB (2002b) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.
- Peterson AT, Soberón J, Sánchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.
- Peterson AT, Stockwell DRB, Kluza DA (2002c) Distributional prediction based on ecological niche modeling of primary occurrence data. In: *Predicting Species Occurrences: Issues of Scale and Accuracy* (eds Scott JM, Heglund PJ, Morrison ML, Raven PH), pp. 617–623. Island Press, Washington, DC.
- Rahbek C (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, **8**, 224–239.
- Raven PH (2004) Taxonomy: where are we now? *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, **359**, 729–730.
- Raxworthy CJ, Ingram CM, Rabibisoa N, Pearson RG (2007) Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos *Phelsuma* from Madagascar. *Systematic Biology*, **56**, 907–923.

- Rissler LJ, Apodaca JJ (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology*, **56**, 924–942.
- Sarkar IN (2007) Biodiversity informatics: organizing and linking information across the spectrum of life. *Briefings in Bioinformatics*, **8**, 347–357.
- Sarkar IN (2009) Biodiversity Informatics: the emergence of a field. *BMC Bioinformatics*, **10**, S1.
- Schnoor JL (2007) Citizen science. *Environmental Science and Technology*, **41**, 5923.
- Scoble M (2004) Unitary or unified taxonomy? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 699–710.
- Scotland RW, Wortley AH (2003) How many species of seed plants are there? *Taxon*, **52**, 101–104.
- Sergio C, Figueira R, Draper D, Menezes R, Sousa AJ (2007) Modelling bryophyte distribution based on ecological information for extent of occurrence assessment. *Biological Conservation*, **135**, 341–351.
- Silvertown J (2009) A new dawn for citizen science. *Trends in Ecology and Evolution*, **24**, 467–471.
- Smith AT, Boitani L, Bibby C, Brackett D, Corsi F, da Fonseca GAB, Gascon C, Dixon MG, Hilton-Taylor C, Mace G, Mittermeier RA, Rabinovich J, Richardson BJ, Rylands A, Stein B, Stuart S, Thomsen J, Wilson C, Edwards JL, Lane MA, Nielsen SE, Bisby FA (2000) Databases tailored for biodiversity conservation. *Science*, **290**, 2073–2074.
- Smith V, Rycroft S, Harman K, Scott B, Roberts D (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. *BMC Bioinformatics*, **10**, S6.
- Soberon J (1999) Linking biodiversity information sources. *Trends in Ecology and Evolution*, **14**, 291.
- Soberon J, Peterson T (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 689–698.
- Steiner FM, Schlick-Steiner BC, Van Der Wal J, Reuther KD, Christian E, Stauffer C, Suarez AV, Williams SE, Crozier RH (2008) Combined modelling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. *Diversity and Distributions*, **14**, 538–545.
- Stigall AL, Lieberman BS (2006) Quantitative palaeobiogeography: GIS, phylogenetic biogeographical analysis, and conservation insights. *Journal of Biogeography*, **33**, 2051–2060.
- Stockwell DRB, Peters D (1999) The GARP Modeling System: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143–158.
- Sun YB (孙英宝), Ma LY (马履一), Qin HN (覃海宁) (2008) A brief history of botanical scientific illustration in China. *Acta Phytotaxonomica Sinica* (植物分类学报), **46**, 772–784. (in Chinese with English abstract)
- Swenson NG (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology*, **21**, 421–434.
- Tangley L (1997) How many species are there? In: US News and World Report Aug. 18, 1997. http://www.usnews.com/usnews/culture/articles/970818/archive_007681.htm. (Accessed March 2010)
- The Biodiversity Committee of Chinese Academy of Sciences (2009) Catalogue of Life China: 2009 Annual Checklist China. CD-ROM; Species 2000 China Node, Beijing, China.
- Thorne RF (2002) How many species of seed plants are there? *Taxon*, **51**, 511–512.
- Thuiller W, Lavorel S, Sykes MT, Araujo MB (2006) Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions*, **12**, 49–60.
- Tillier S, Roberts D (2006) Taxonomy on the fly in a European web project. *Nature*, **440**, 24.
- Walter DE, Winterton S (2007) Keys and the crisis in taxonomy: extinction or reinvention? *Annual Review of Entomology*, **52**, 193–208.
- Wang R, Wang YZ (2006) Invasion dynamics and potential spread of the invasive alien plant species *Ageratina adenophora* (Asteraceae) in China. *Diversity and Distributions*, **12**, 397–408.
- Wheeler QD (2004) Taxonomic triage and the poverty of phylogeny. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, **359**, 571–583.
- Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: impediment or expedient? *Science*, **303**, 285.
- Wilson EO (2003) The encyclopedia of life. *Trends in Ecology and Evolution*, **18**, 77–80.
- Yang QE (杨亲二), Zhu GH (朱光华), Hong DY (洪德元), Wu ZY (吴征镒), Raven PH (2005) World's largest flora completed. *Science*, **309**, 2163b.
- Yee TW, Mitchell ND (1991) Generalized additive models in plant ecology. *Journal of Vegetable Science*, **2**, 587–602.
- Yesson C (2007) How global is the global biodiversity information facility? *PLoS ONE*, **2**, e1124.
- Zhang ML (张明理) (2009) DELTA system, a recommendable information processing tool for taxonomic description. *Journal of Plant Resources and Environment* (植物资源与环境学报), **18**, 87–90. (in Chinese with English abstract)
- Zhong Y (钟扬), Zhang L (张亮) (2000) Biodiversity Informatics: a new direction of bioinformatics and biodiversity science and related key techniques. *Chinese Biodiversity* (生物多样性), **8**, 397–404. (in Chinese with English abstract)

(责任编辑: 周玉荣)