

# 植物DNA条形码与生物多样性数据共享平台构建

曾春霞<sup>1#</sup> 王亚楠<sup>2#</sup> 王雨华<sup>2,3\*</sup> 王 红<sup>1,4\*</sup>

1 (中国科学院昆明植物研究所中国西南野生生物种质资源库, 昆明 650201)

2 (中国科学院昆明植物研究所科技信息中心, 昆明 650201)

3 (中国科学院昆明植物研究所资源植物与生物技术重点实验室, 昆明 650201)

4 (中国科学院昆明植物研究所生物多样性与生物地理学重点实验室, 昆明 650201)

**摘要:** DNA条形码基于较短的DNA序列实现物种的快速、准确鉴定, 不仅加快了全球生物物种的鉴定和分类步伐, 也为生物多样性的管理、保护和可持续利用提供了新思路和研究方法。植物DNA条形码标准数据库的不断完善, 将使植物多样性信息的快速获取成为可能; 将不同类型数据资源整合、共享和利用, 构建植物DNA条形码数据共享平台, 是满足公众对物种准确鉴定和快速认知的重要支撑。本文介绍了近年来植物DNA条形码的研究进展; 植物DNA条形码参考数据库的研发现状和存在的问题。结合上述问题, 围绕“大数据”时代背景, 对如何管理和使用好海量的植物信息, 如何构建数据共享平台提出了一些设想: (1)数据共享平台的元数据应尽可能翔实、丰富、准确和多关联; (2)数据标准应统一规范; (3)查询入口方便、迅速、多样, 易于管理, 便于实现更大程度的数据共享和全球化的合作交流。

**关键词:** 植物DNA条形码, iFlora, 生物多样性, 共享平台

## Plant DNA barcoding and framework for biodiversity data sharing platform

Chunxia Zeng<sup>1#</sup>, Yanan Wang<sup>2#</sup>, Yuhua Wang<sup>2,3\*</sup>, Hong Wang<sup>1,4\*</sup>

1 China Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201

2 Science and Technology Information Center, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201

3 Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201

4 Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201

**Abstract:** DNA barcoding technology provides an opportunity for rapid, accurate, and standardized species-level identification using short DNA sequences. This method speeds up species identification and classification, and presents a new tool for the management, conservation and sustainable development of biodiversity at a global level. Due to improvements in plant barcode database availability and functionality, it is becoming feasible to meet increasing demands for biodiversity information. A framework is needed for a barcoding server platform that utilizes, integrates, and shares among different data types. Such a platform would be an important step towards enabling the public to rapidly identify species and acquire species-related digital information. In this paper, we review current progress on plant DNA barcoding. Secondly, we summarize the current status of, and identify bottlenecks for, plant DNA barcode reference libraries specifically. Thirdly, in the Big Data era, it is indispensable to manage and make good use of massive amounts of plant information. We provide the following suggestions for the framework of server platform: (1) metadata should be substantial, accurate and correlative; (2) data should be normalized; (3) query entrance is convenient, efficient, easy to manage, and available for large-scale data sharing and global communication.

收稿日期: 2013-12-30; 接受日期: 2014-04-14

基金项目: 科技部科技基础性工作专项项目(2013FY112600)、科技部国家高技术研究发展计划(863 计划)主题项目(2012AA021801)和中国科学院大科学装置项目(2009-LSFGBOWS-01)

# 共同第一作者 Equally contributed to this paper

\* 共同通讯作者 Co-authors for correspondence. E-mail: wangyuhua@mail.kib.ac.cn; wanghong@mail.kib.ac.cn

**Key words:** plant DNA barcoding, iFlora, biodiversity, server platform

DNA条形码(DNA barcoding)是一种基于DNA序列进行生物物种鉴定的技术,即利用标准化的一个或者几个DNA片段进行序列分析,根据核苷酸序列差异,对物种进行快速和准确的鉴定(Hebert *et al.*, 2003; Kress & Erickson, 2007; Hollingsworth, 2011)。DNA条形码的选择标准,一是片段长度较短,片段两端连接相对保守的区域,通用引物容易扩增;二是片段要有足够的变异,根据种内特异性和种间多样性将物种区分开来。DNA条形码与广义形态性状相结合,共同应用于物种鉴定,是对传统鉴定和分类的有效补充(DeSalle *et al.*, 2005; Will *et al.*, 2005; Hajibabaei *et al.*, 2007; Vogler & Monaghan, 2007),同时也为生物多样性的管理、保护和可持续利用提供新的思路和研究工具(Smith *et al.*, 2005; Agnarsson & Kuntner, 2007; Chase & Fay, 2009)。

2010年,国际生命条形码计划(International Barcode of Life, iBOL)(<http://ibol.org/>)启动,这是世界上迄今最大的生物多样性研究计划,预计在2015年前完成50万种真核生物500万份标本的DNA条形码参考数据库,目前已经完成约21万个物种290万份标本的DNA条形码测定([http://www.barcodinglife.com/index.php/TaxBrowser\\_Home](http://www.barcodinglife.com/index.php/TaxBrowser_Home))。该计划现阶段最重要的目标是建立全球共享的生命条形码标准数据库。

作为iBOL计划的4个中心节点之一,中国植物DNA条形码团队通过大数据分析,推荐将ITS作为种子植物核心条形码之一,得到国际同行的广泛认可(Hollingsworth, 2011),中国也是提交植物DNA条形码数量最多的国家(参见<http://boldsystems.org/>)。中国多次建议与生命条形码数据库系统(Barcode of Life Data Systems, BOLD)网站建立实时更新的镜像站点,但几番努力仍未成功(陈之端和李德铎, 2013)。因此,构建中国植物DNA条形码参考数据库以及生物多样性数据共享平台就成为当前首当其冲的任务。

## 1 植物DNA条形码

植物DNA条形码研究稍滞后于动物学相关的

工作。编码细胞色素C氧化酶I的线粒体基因 $COI$ (或 $cox1$ )被选作动物DNA条形码,在一些昆虫、鱼类和鸟类的研究中取得了较好结果(Hebert *et al.*, 2004; Smith *et al.*, 2005; Hajibabaei *et al.*, 2006; Yoo *et al.*, 2006)。而对于陆地植物,由于线粒体基因的进化速率相对较慢, $COI$ 并不适合。因此,从叶绿体基因组和核基因组中寻找通用性高、序列质量好、物种分辨率高的DNA条形码候选片段是问题的关键。2009年,国际生命条形码联盟植物工作组(Consortium for the Barcode of Life, Plant Working Group, 2009)根据已往研究和该研究组数据的分析结果,建议将 $rbcL + matK$ 组合作为陆地植物的核心DNA条形码,用于构建植物物种鉴定的统一框图。在第三届国际生命条形码大会上,ITS和 $trnH-psbA$ 被作为植物的辅助条形码(Hollingsworth *et al.*, 2011)。2011年,中国植物条形码研究团队建议将ITS(或ITS2)作为种子植物的核心条形码之一(Li *et al.*, 2011)。

植物核心条形码的确定,实现了植物标本鉴定过程的自动化和标准化,突破了对经验的过度依赖,并可在较短时间内建立易于利用的应用系统。因此,植物DNA条形码技术作为传统分类的有效补充,在生物多样性研究领域显示出强大的生命力。利用DNA条形码可以解决缺乏有效形态学证据的已知或未知物种的鉴定,有助于新种或者隐存种的发现。例如,Pei等(2011)以及裴男才(2012)从叶片或树皮中提取DNA,对热带雨林或亚热带常绿阔叶林中一些林冠层物种快速鉴定到科、属甚至种。Liu等(2011)对欧亚分布的红豆杉属(*Taxus*)植物的DNA条形码研究中,发现4个物种(类群)可能是新种或隐存种,并得到了居群遗传学(Liu *et al.*, 2013)和形态学(Möller *et al.*, 2013)证据的支持。

DNA条形码在群落生态学研究作为一种快速识别物种的方法,还可以用于群落系统发育关系的构建,成为生物多样性或者物种丰富度快速评估的新工具。例如,Kress等(2009)在对巴拿马Barro Colorado Island (BCI)样地的群落系统发育结构的研究中,利用 $rbcL + matK + trnH-psbA$ 片段组合,使物种鉴定的准确率提高至98%。

生物多样性的形成和维持机制已有较多假说,

但仍然没有形成广为接受的理论。DNA条形码在群落物种间网络(尤其是植食网络和传粉网络)关系等的研究中,得到了较为成功的应用。例如借助DNA条形码技术分析动物消化道中食物残渣,鉴定其取食的植物,进而研究野生动物的取食习性(Passmore *et al.*, 2006; Huler *et al.*, 2007; Bourlat *et al.*, 2008)。

García-Robledo等(2013)在构建美国中部热带雨林中植物与昆虫之间的网络关系时,证实DNA条形码是一种行之有效的办法。Ramírez等(2011)通过DNA条形码探讨特化的兰花类群与其传粉昆虫之间的拓扑网状结构,以及它们的分化时间和多样化式样,表明兰花的多样化与传粉昆虫的分化相一致,同时与新热带森林中化学环境的变化密切相关。此外,Roy和Lawson(2012)借助DNA条形码,研究了寄主与寄生网络的关系。可见,DNA条形码为生物多样性的形成和维持机制的研究开辟了新的途径。

随着第二代测序技术的出现,开发出通过高通量技术获得多物种(样品)的DNA条形码序列的方法,即DNA metabarcoding技术(Pompanon *et al.*, 2011; Riaz *et al.*, 2011)。该方法结合生物信息学手段,可以获取复杂混合样品甚至是大尺度范围样品的DNA条形码序列,自动识别多个物种,具有快速、可重复、高效及综合性的特点,可以对当前生物多样性和古生物多样性进行评估。Yoccoz等(2012)从环境土壤提取了混合DNA,获得其中叶绿体*trnL*(UAA)内含子P6环序列,利用metabarcoding技术进行物种鉴定,结果能够较好地反映地上植物的分布。Hiiesalu等(2012)利用metabarcoding技术不仅分析了地上植物物种的分布及丰度,还发现对地下土壤的取样可以获得传统分析中容易忽视的信息,如物种共存分布等,能更全面地反映植被生态及其影响因素。借助于metabarcoding技术也可以更好地重现古植物多样性。例如Murray等(2012)分析了植食性动物的粪堆遗存,发现了之前没有报道的动植物物种。此外,利用metabarcoding技术对冻土DNA样品进行分析,DNA鉴定结果可以与花粉鉴定和大化石鉴定结果相互印证,已成为传统古植物学研究的辅助手段(Jørgensen *et al.*, 2012)。

## 2 植物DNA条形码参考数据库

为了实现对物种的快速鉴定,构建了DNA条形码参考数据库(Reference Library),包括实物库(植

物标本库、植物分子材料库和植物总DNA库)、植物DNA条形码数据库、物种信息库和计算分析共享系统(曾春霞等, 2012)。这些数据库需要统一规范化的工作流程,才能够在数据展示系统(应用平台)中,实现对物种的简单快捷的鉴定并获取相关信息。生命条形码联盟(CBOL)建立了生命条形码数据库系统BOLD,是专门收集和分析DNA条形码数据的平台。与目前常用的核苷酸数据库(例如GenBank、EMBL和DDBJ)不同,BOLD数据库除了要求提交DNA条形码序列外,还要求使用者提交以下信息:物种名称、标本图片、凭证标本信息(目录号和馆藏号)、采集号(采集人、采集日期和GPS定位地点)、标本鉴定人、用于PCR扩增的引物、测序原始峰图。这样,研究人员在对核苷酸序列进行比对的同时,还能够结合其他信息,完成对物种的鉴定。据BOLD数据库最新统计,约21万个物种已经完成DNA条形码信息的描绘,同时具DNA条形码序列信息的标本也有290万份([http://www.barcodinglife.com/index.php/TaxBrowser\\_Home](http://www.barcodinglife.com/index.php/TaxBrowser_Home))。该数据库以鱼类、鸟类和昆虫类为主,而真菌和植物DNA条形码信息相对较少。

由于植物物种中存在广泛的多倍化、杂交或基因渐渗等事件,同时近缘(姐妹)或者近期分化类群普遍存在,这就给植物DNA条形码参考数据库的构建提出了更高的要求。为了能够实现种级水平的快速鉴定,如何统一物种界定标准和居群取样标准等是当务之急(陈之端和李德铎, 2013)。然而要构建一个全面的、准确的植物DNA条形码参考数据库,需要保证数据库中的DNA序列均来自经分类学家可靠鉴定的标本、这些凭证标本具有详尽的注释和可靠的保藏、DNA序列正确,以保证分子鉴定的可靠性。可见,植物DNA条形码标准数据库的构建将是长期而艰巨的任务。

近年来,中国科学院昆明植物研究所联合国内相关科研院校开展了中国重要植物类群的采集、DNA条形码的测定与分析等工作,在此基础上,进一步融入现代植物学、新一代测序技术、地理信息数据和计算机信息技术等新元素,提出了新一代植物志(iFlora)的研究计划(李德铎等, 2012)。iFlora不但包含传统植物志的检索表、物种描述等信息,还包括物种详细分布信息、图片或图像、DNA条形码数据库和系统发育等信息,同时建立了高效的电子

化比对和搜索工具。通过系列关键技术的集成和研发, iFlora将构建一个便捷、准确识别植物和掌握相关数字化信息的智能植物志(或智能装备)(王红等, 2013)。目前, iFlora研究计划已经积累了中国维管束植物约230科1,665属6,800种约67,000条DNA条形码序列, 同时构建的植物DNA条形码参考数据库已初具雏形。

### 3 植物DNA条形码数据共享平台构建

随着数据的爆炸性增长, 人们正进入一个“大数据”的时代。大数据可概括为4V特征, 即数据量大(volume)、类型繁多(variety)、价值密度低(value)和速度快时效高(velocity)。在这样的背景下, 如何管理和使用好海量的信息, 是一个值得探讨的问题。

现今关于植物DNA条形码数据不再是单一的、一个维度上的数据, 而是丰富的、多维度的数据。构建中国植物DNA条形码数据共享平台, 需要充分考虑到大数据的特征, 整合和管理好海量的数据资源, 并思考如何使用植物学大数据。以往的DNA条形码数据共享平台多是通过序列比对算法(如BLAST)来鉴定物种, 可当大数据的概念引入之后, 通过加入GPS信息与以往的标本分布进行比对作为佐证, 以及植物图片识别作为辅助手段, 物种将得到更准确的鉴定。这种使用多维度的植物学数据“交叉复现”准确鉴定物种, 将使得植物学信息“全息可见”。下面就中国植物DNA条形码数据共享平台的构建提出一些思考。

#### 3.1 植物DNA条形码数据

##### 3.1.1 数据的内容

在大数据时代, 植物DNA条形码数据不再限于其获取过程中产生的信息, 如样本采集、标本鉴定、标本凭证、标本图片、测序图谱、处理文件、序列等信息, 还应加入更加丰富的内容。如BOLD SYSTEMS(<http://www.boldsystems.org/>)的页面加入了来源、出版信息等内容; Tree of Life Web Project(<http://tolweb.org/tree/>)的详细页面加入了物种名称(包括图片、生命之树)、简介、特征、亲缘关系、异名、参考文献、互联网上的信息、标题插图等内容; iFlora系列的“国家重点保护野生植物鉴定信息平台”(<http://www.iflora.cn/>)的详细页面结合了昆明植物研究所的优势资源, 加入了《中国植物志》关于物种的信息、种质资源、民族植物学、植物照片

等内容。

以下有价值的DNA条形码相关植物学数据可供整合参考:

(1)《中国植物志》、*Flora of China*以及地方植物志书等关于物种的描述。

(2)《中华本草》、《中国民族药志要》、《中国中药资源志要》中关于物种药用价值的介绍。

(3)“物种2000”(<http://www.sp2000.cn/joaen/>)的异俗名数据。

(4)中国植物图像库(<http://www.plantphoto.cn/>)的海量植物照片。

(5)国家标本资源共享平台(<http://www.nsii.org.cn/>)的标本分布信息。

(6)中国西南野生生物种质资源库(<http://www.genobank.org/>)的种质资源信息等。

(7)生物技术信息中心(<http://www.ncbi.nlm.nih.gov/>)的GenBank条形码数据等。

##### 3.1.2 数据的标准化

在平台建设之前, 不同单位和部门的数据库对于检索系统设计的技术方案、技术与信息组织规划选择、相关技术标准的采用, 基本上是各行其是, 数据标准的不统一常常给使用者带来一些不必要的麻烦, 造成数据共享的壁垒。因此有必要建立一套统一的DNA条形码数据标准。基于这样的需求, 基础科学数据共享网重要生物类群DNA条码数据库(<http://www.pbl.csdb.cn/page/index.vpage>)、BOLD Mirror(<http://www.boldmirror.net/>)等网站在建站时就提出了植物DNA条形码研究技术规范(高连明等, 2012)。中国植物DNA条形码数据共享平台在建立时应进一步完善此类规范, 供以后数据库建设者及科研工作者参考。

#### 3.2 数据共享平台的界面设计

数据共享平台是目前国际上普遍采用的数据共享方式, 通过集中数据储存, 建立数据服务机制, 实现数据的共享。以下从界面设计入手, 提出“中国植物DNA条形码数据共享平台”的建设思路。

##### 3.2.1 门户模块

现有的一些知名DNA条形码网站可以给我们提供参考。BOLD SYSTEMS(<http://www.boldsystems.org/>)使用了数据门户、条形码集群数据库、教育门户、数据收集工作站这4个模块来实现数据的共享。中国生命条形码数据门户(<http://www.barcodeoflife.org/>)

cn/)主要提供新闻、资源的文字信息;中国生命条形码信息管理系统和BOLD Mirror两个模块实现管理及数据共享的任务。其中中国生命条形码信息管理系统(<http://data.barcodeoflife.cn/index.php>)必须注册才能进行数据的提交及管理;BOLD Mirror模块(<http://www.boldmirror.net/>)主要采用BLAST查询方式实现DNA条形码数据的共享;iFlora系列——国家重点保护野生植物鉴定信息平台(<http://www.iflora.cn/>)采用了多种智能化查询方式获得DNA条形码数据。

因此,中国植物DNA条形码数据共享平台至少包括3个门户:

(1)数据门户:采用多种方式的智能化查询并提供一定的下载方式。

(2)交流门户:除了基本的个人信息管理,用户还可以进行学术交流,提交DNA条形码相关的数据信息,实现更大程度的共享。

(3)管理门户:主要针对管理员级别的可视化数据管理,包括数据查询、导入、增删改、用户角色管理以及日志维护等操作。管理员还兼具对于交流门户提交的信息审核的责任,将通过审核的数据归入数据门户的表中。

### 3.2.2 数据门户的鉴定入口

数据门户能让用户通过智能表检索、DNA条形码鉴定、图像识别三个检索入口获取物种详细的信息。其中,智能表检索是指在优化数据结构的基础上进行常规检索,如名录索引、分等级检索、志书式检索、模糊搜索等方式。DNA条形码鉴定是通过本地BLAST(local BLAST)集成在网页中,实现序列比对的功能,鉴定得出可能的物种,从而获得鉴定物种的详细信息。图像识别是利用计算机对图像进行处理和解析,以识别各种不同模式的目标。对于植物来说,就是用植物的某部位如叶片、花、果实、整体形貌等的照片,通过一定的算法与图像库内的图片进行比对,鉴定出可能的物种列表,从而获得相应物种的详细信息。现在“百度识图”以及“Google识图”已发展到一定水平,越来越多的科研工作者也开始关注这种方式。2013年9月在上海辰山植物园召开的“第三届全国生物多样性信息学研讨会”上,很多学者报告了图像识别技术并进行了深入的讨论,中国科学院植物研究所也正与“百度识图”合作开发进行“拍花识植物”的工作。“图像识

图”正成为一种新的热点(庄会富和王雨华,2012)。

### 3.2.3 平台框架的设计

通过以上分析,可以得出以下平台框架架构,即基于一定的数据标准规范体系以及安全保障体系,采用“瘦”客户端、“胖”服务器端的浏览器/服务器(B/S)结构,以表现层(View)、业务逻辑层(Business Logic Layer, BLL)、数据访问层(Data Access Layer, DAL)、数据库(Database)多层架构的设计模式进行开发。中国植物DNA条形码数据共享平台应该提供一个智能化的物种鉴定引擎,高效化的合作交流平台以及人性化的管理门户。具体的架构如图1所示。

中国植物DNA条形码数据共享平台应具备以下的特点:

(1)方便、迅速、多元的数据查询入口。用户可以通过多种查询入口,方便快捷地找到并下载所需数据。

(2)详实、丰富、准确、多关联的数据页面及元数据。元数据不仅仅包括样本的采集信息和鉴定信息以及DNA条形码的序列信息等,还应包括如物种信息、图片信息、药用植物信息和标本分布信息等,同时保证数据的详细、准确、权威。多关联性表现在能展示物种更多的相关信息。

(3)便于合作、交流。为了实现更好的数据共享以及全球化的合作交流,需要提供便利的交流合作平台,不仅提供日常的学术交流及讨论,还能提供数据的提交乃至定制服务。

(4)易于管理。有强大的可视化后台管理界面,实现数据的查询、导入、管理、用户角色管理、日志维护等功能,对已有数据实现增、删、改和检索的操作,并对提交的数据进行审核和更新。

## 4 展望

近年来,随着网络化和信息化的飞速发展,各种移动互联网设备迅速普及,“云计算”、“云服务”等概念已形成一种趋势。中国植物DNA条形码数据共享平台最初的设计可能只是集成BLAST序列比对进行物种鉴定,随着工作的深入,将会集成越来越多的生物信息学软件,可能涉及到一些复杂的算法,计算量较大,加之遗传信息的海量性和生物信息的多样性,采用“瘦”客户端、“胖”服务器端的结构就要求高效的计算服务环境。基于科学计算需求

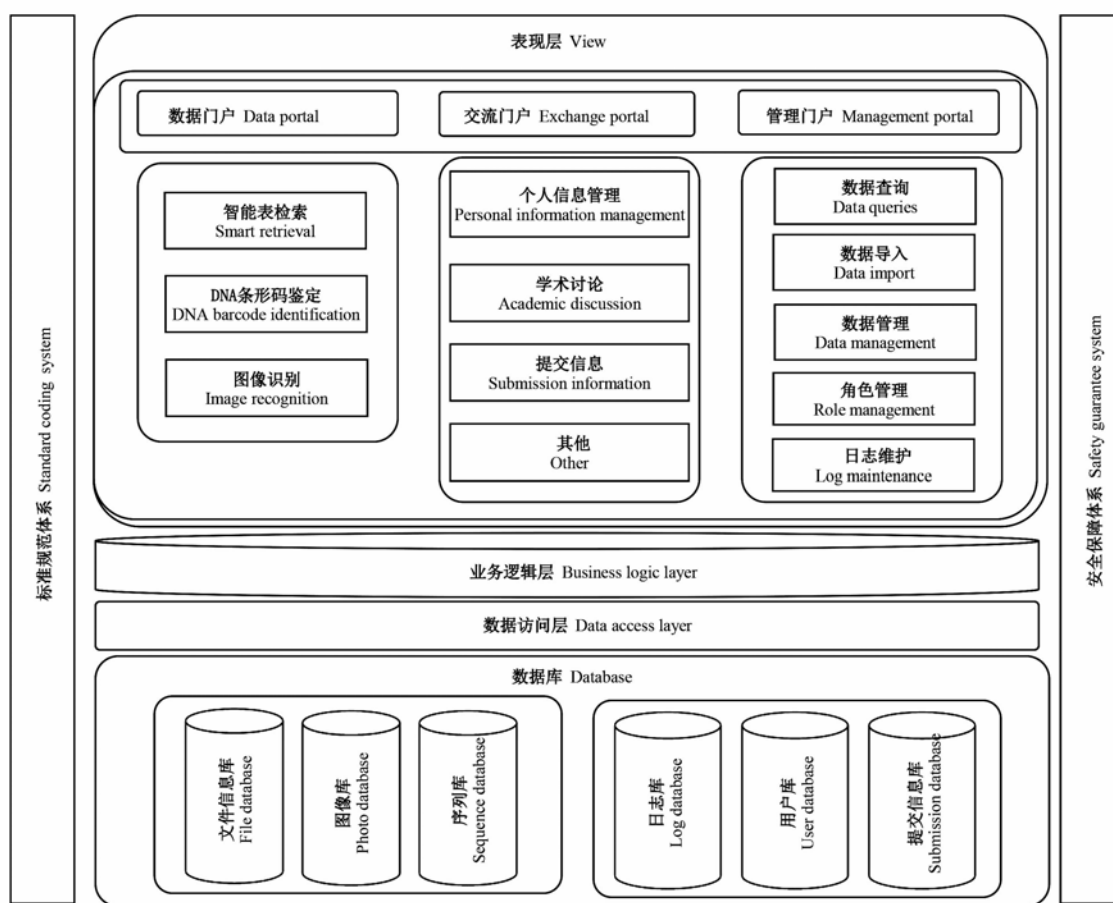


图1 中国植物DNA条形码数据共享平台架构

Fig. 1 Proposal framework of plant DNA barcoding server platform

的共享平台的设计势必是一个满足多平台客户端、立足科研工作者核心用户群、采用“客户端-云服务端”构架的云服务平台。

Tree of Life Web Project (<http://tolweb.org/tree/>) 正致力于地球上所有生命起源的追溯, 构建世界的生命进化树, 解释现存生物的演化关系, 并以生命系统发育树的形式表现出来。我们正在构建的中国植物DNA数据共享平台, 在实现DNA条形码数据共享的同时, 将为进一步构建中国植物生命之树提供重要基础, 最终建立一个服务于科研、管理和公众的数据共享平台。

#### 参考文献

- Agnarsson I, Kuntner M (2007) Taxonomy in a changing world: seeking solutions for a science in crisis. *Systematic Biology*, **56**, 531–539.
- Bourlat SJ, Nakano H, Akerman M, Telford MJ, Thorndyke

- MC, Obst M (2008) Feeding ecology of *Xenoturbella bocki* (phylum Xenoturbellida) revealed by genetic barcoding. *Molecular Ecology Resources*, **8**, 18–22.
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA*, **106**, 12794–12797.
- Chase MW, Fay F (2009) Barcoding of plants and fungi. *Science*, **325**, 682–683.
- Chen ZD (陈之端), Li DZ (李德铎) (2013) On barcode of life and tree of life. *Plant Diversity and Resources* (植物分类与资源学报), **35**, 675–681. (in Chinese with English abstract)
- DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London: Series B: Biological Sciences*, **360**, 1905–1916.
- Gao LM (高连明), Liu J (刘杰), Cai J (蔡杰), Yang JB (杨俊波), Zhang T (张挺), Li DZ (李德铎) (2012) A synopsis of technical notes on the standards for plant DNA barcoding. *Plant Diversity and Resources* (植物分类与资源学报), **34**, 592–606. (in Chinese with English abstract)

- García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ (2013) Tropical plant–herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS ONE*, **8**, e52967.
- Hajibabaei M, Hanzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences, USA*, **103**, 968–971.
- Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**, 167–172.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London: Series B*, **270**, 313–321.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.
- Hiiesalu I, Opik M, Metsis M, Lilje L, Davison J, Vasar M, Moora M, Zobel M, Wilson SD, Pärtel M (2012) Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Molecular Ecology*, **21**, 2004–2016.
- Hollingsworth PM (2011) Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA*, **180**, 19451–19452.
- Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA Barcode. *PLoS ONE*, **6**, e19254.
- Huler J, Mogia M, Isua B, Novotny V (2007) Host specificity of ambrosia and bark beetles (Col., Curculionidae: Scolytinae and Platypodinae) in a New Guinea rain forest. *Ecological Entomology*, **32**, 762–772.
- Jørgensen T, Kjaer KH, Haile J, Rasmussen M, Boessenkool S, Andersen K, Coissac E, Taberlet P, Brochmann C, Orlando L, Gilbert MTP, Willerslev E (2012) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA meta-barcoding. *Molecular Ecology*, **21**, 1980–1988.
- Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE*, **2**, e508.
- Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjurjo O, Bermingham E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences, USA*, **106**, 18621–18626.
- Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA*, **108**, 19641–19646.
- Li DZ (李德铨), Wang YH (王雨华), Yi TS (伊廷双), Wang H (王红), Gao LM (高连明), Yang JB (杨俊波) (2012) The next-generation flora: iFlora. *Plant Diversity and Resources* (植物分类与资源学报), **34**, 525–531. (in Chinese with English abstract)
- Liu J, Möller M, Gao LM, Zhang DQ, Li DZ (2011) DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources*, **11**, 89–100.
- Liu J, Möller M, Provan J, Gao LM, Poudel RC, Li DZ (2013) Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, **199**, 1093–1108.
- Möller M, Gao LM, Robert RM, Liu J, Zhang DQ, Poudel RC, Li DZ (2013) A multidisciplinary approach reveals hidden diversity in the taxonomically challenging *Taxus walliichiana* complex. *Taxon*, **62**, 1161–1177.
- Murray DC, Pearson SG, Fulagar R, Chase BM, Houston J, Atchison J, White NE, Bellgard MI, Clarke E, Macphail M, Gilbert MTP, Haile J, Bunce M (2012) High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews*, **58**, 135–145.
- Passmore AJ, Passmore AJ, Jarman S, Kawaguchi S, McMinn A, Nicol S (2006) DNA as a dietary biomarker in Antarctic krill, *Euphausia superba*. *Marine Biotechnology*, **8**, 686–696.
- Pei NC (裴男才) (2012) Identification of plant species based on DNA barcode technology. *Chinese Journal of Applied Ecology* (应用生态学报), **23**, 1240–1246. (in Chinese with English abstract)
- Pei NC, Lian JY, Erickson DL, Swenson NG, Kress WJ, Ye WH, Ge XJ (2011) Exploring tree habitat associations in a Chinese subtropical forest plot using a molecular phylogeny generated from DNA barcode loci. *PLoS ONE*, **6**, e21273.
- Pompanon F, Coissac E, Taberlet P (2011) Metabarcoding a new way to analyze biodiversity. *Biofutur*, **3**, 30–32.
- Ramírez SR, Eltz T, Fujiwara MK, Gerlach G, Goldman-Huertas B, Tsutsui ND, Pierce NE (2011) Asynchronous diversification in a specialized plant-pollinator mutualism. *Science*, **333**, 1742–1746.
- Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145.
- Roy HE, Lawson Handley LG, Dunn A (2012) Networking: a community approach to invaders and their parasites. *Functional Ecology*, **26**, 1238–1248.
- Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1825–1834.
- Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolu-*

- tionary Research*, **45**, 1–10.
- Wang H (王红), Wang YH (王雨华), Yang ZL (杨祝良), Li DZ (李德铎) (2013) iFlora: innovation and development. *Plant Diversity and Resources* (植物分类与资源学报), **35**, I–III. (in Chinese with English abstract)
- Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.
- Yoccoz NG, Brathen KA, Gielly L, Haile J, Edwards ME, Goslar T, Von Stedingk H, Brysting AK, Coissac E, Pompanon F, Sørensen JH, Miquel C, Valentini A, De Bello F, Chave J, Thuiller W, Wincker P, Cruaud C, Gavory F, Rasmussen M, Gilbert MT, Orlando L, Brochmann C, Willerslev E, Taberlet P (2012) DNA from soil mirrors plant taxonomic and growth from diversity. *Molecular Ecology*, **21**, 3647–3655.
- Yoo HS, Eah JY, Kim JS, Kim YJ, Min MS, Paek WK, Lee H, Kim CB (2006) DNA barcoding Korean birds. *Molecules and Cells*, **22**, 323–327.
- Zeng CX (曾春霞), Yang JB (杨俊波), Yang J (杨静), Yi TS (伊廷双), Lin CY (林春艳) (2012) A proposed framework for iFlora. *Plant Diversity and Resources* (植物分类与资源学报), **34**, 555–561. (in Chinese with English abstract)
- Zhuang HF (庄会富), Wang YH (王雨华) (2012) A framework of scientific-cloud for iFlora. *Plant Diversity and Resources* (植物分类与资源学报), **34**, 623–630. (in Chinese with English abstract)

(责任编辑: 马克平 责任编辑: 时意专)