

中国西南野生生物种质资源数据库

李拓径¹, 李洪涛¹, 李德铎^{1*}

1. 中国科学院昆明植物研究所, 昆明 650201

摘要: 中国西南野生生物种质资源库是国家重大科学工程, 在野生种质资源的收集保藏工作中扮演着重要角色, 目的是保障国家重要战略生物资源。本数据库的构建开发, 结合了野生生物种质资源收集和保藏的工作规范, 主要包括采集、登记、检测、鉴定、库存和再生等方面信息, 并按照相关管理办法对外开展实物和数据的共享, 为国内外研究机构、高等学校、政府职能部门和公众提供服务。本数据库已收录植物种子、植物DNA样品和植物离体材料等数据信息11.6万余条, 用户可以根据自己的需要对数据进行检索。本数据库的建设将有利于我国种质资源的收集、保藏和生物多样性保护, 对生态保护有积极的示范意义。

关键词: 野生生物; 野生种质资源; 战略生物资源; 数据库

文献 DOI:

10.11922/csdata.2017.19.zh

数据 DOI:

10.11922/sciencedb.490

文献分类: 生物学

收稿日期: 2017-09-29

开放同评: 2017-12-11

发表日期: 2018-03-07

数据库(集)基本信息简介

数据库(集)名称	中国西南野生生物种质资源数据库
数据作者	中国西南野生生物种质资源库
数据通信作者	李德铎 (dzl@mail.kib.ac.cn)
数据时间范围	2005~2017年
地理区域	中国境内
数据格式	SqlServer
数据服务系统网址	http://www.genobank.org/ http://www.sciencedb.cn/dataSet/handle/490
基金项目	中国西南野生生物种质资源库国家大科学工程
数据库(集)组成	本数据库由野生植物种子数据71 200多条, 植物离体材料数据2 000多条, 植物DNA样品数据43 300多条等组成。

引言

中国西南野生生物种质资源库(简称种质库)是国家大科学工程。其总体科学目标是: 立足西南, 面向全国, 建成具有国际影响力、亚洲一流的野生生物种质资源保护设施和科学体系^[1]。其保藏的野生种质资源是国家重要的战略生物资源, 具有重要的科学和经济价值, 对社会经济可持续发展和生态文明建设具有重要意义^[2]。在参考英国皇家植物园邱园“千年种子库”保藏模式的基础上, 基于种质库的资源保藏目标和科研工作流程, 按照“科学规划、逐步推进”的原则, 以工作业务流程为基础, 设计数据库模型, 采用模块化技术将种质资源采集、处理、

* 论文通信作者

李德铎: dzl@mail.kib.ac.cn

保藏、功能基因挖掘等科研活动数据的存储、管理、集成、检索、维护和分析统计融为一体，建成基于 B/S 架构（Browser/Server 架构，即浏览器和服务器架构）的种质库信息管理系统（Germplasm Bank Information System，简称 GBIS 系统），从而构建种质资源采集保藏研究的国家战略种质资源保藏信息系统^[3]。在信息系统管理的基础上，通过各个业务模块数据的合成、关联，有效形成野生种质资源数据库。

中国西南野生生物种质资源数据库收录了野生植物种子、植物 DNA 样品、植物离体材料以及动物材料和微生物资源等的各类保藏相关信息，包括登记管理信息、监测信息、库存信息、实验及使用信息等。其中野生植物材料数据超过 11.6 万条，全部对外公开查询，是本文阐述的主体；另有动物材料数据 53 800 余条和微生物资源数据 2 200 余条暂未对外公开展示，因而在本文中未详细描述。各类信息以关系型数据库的方式进行存储，并且可通过数据库实现自动备份、恢复。

除此之外，为了进一步规范种质资源的分类信息管理，其基础数据库建立了以 Flora of China^[4]（以下简称 FOC）为主、The Plant list^[5]（以下简称 TPL）为辅的植物分类学名称体系，以实现植物分类学名称数据库系统内的统一审核、规范等功能。并且系统整理 FOC 相关性状描述信息和图片信息，形成种质资源分类学基础名录数据库。基础名录数据库成功地与 GBIS 系统结合，为采集策划、物种鉴定和分析统计等提供数据依据，确保野生种质资源采集、保藏、研究过程中数据的合理性、规范性。目前，本数据库已经开放共享，并支持申领可分发实物资源。

1 数据采集和处理方法

作为国家的重大科技基础设施，种质资源的保藏与创新有着一套严格、标准的工作流程，基本分成 4 个关键环节和 9 个组成部分^[3]。4 个关键环节为：采集计划、监测评价、实验分析和研究创新。这 4 个环节可分解为 9 个有序的部分：采集计划、野外采集、种子清理、质量检测、入库保存、活力检测实验（TZ 实验，TZ 即氯化三苯基四氮唑 triphenyltetrazolium chloride）染色法、创新研究、再生扩繁、开发利用^[3]。这些工作分别由不同的部门承担，在每一个活动部门都有严格的操作流程和标准规范，在每一个关键环节都需要依赖原有的科学积累和分析实验设备，并产生大量的数据。因此，这 4 个环节之间如何有效、科学地协作，并达到数据积累和分析的目的，是种质资源保藏与创新成功的关键。

1.1 数据采集方法

种质资源数据主要来源于自 2005 年 11 月组建种质保藏团队至今的野生种质资源野外调查及保藏管理、研究等各项数据。数据采集建立了严格规范的标准流程（图 1），本数据库首先由管理人员根据采集材料递交情况进行 GBIS 系统登记，然后采集人员通过离线或在线的方式补充野外采集信息，标本鉴定人员通过在线方式审核完善鉴定信息，管理人员通过实验检测生成检测数据信息，实验人员通过信息系统在线管理和完成实验信息的录入。由此产生以种质资源实物登记为中心的中国西南野生生物种质资源数据库。数据库通过关系数据库关联的模式，实现各个模块分块工作、有效关联等功能，实现数据的标准化，系统有效地形成数据关联，为种质资源统计分析和应用奠定基础。

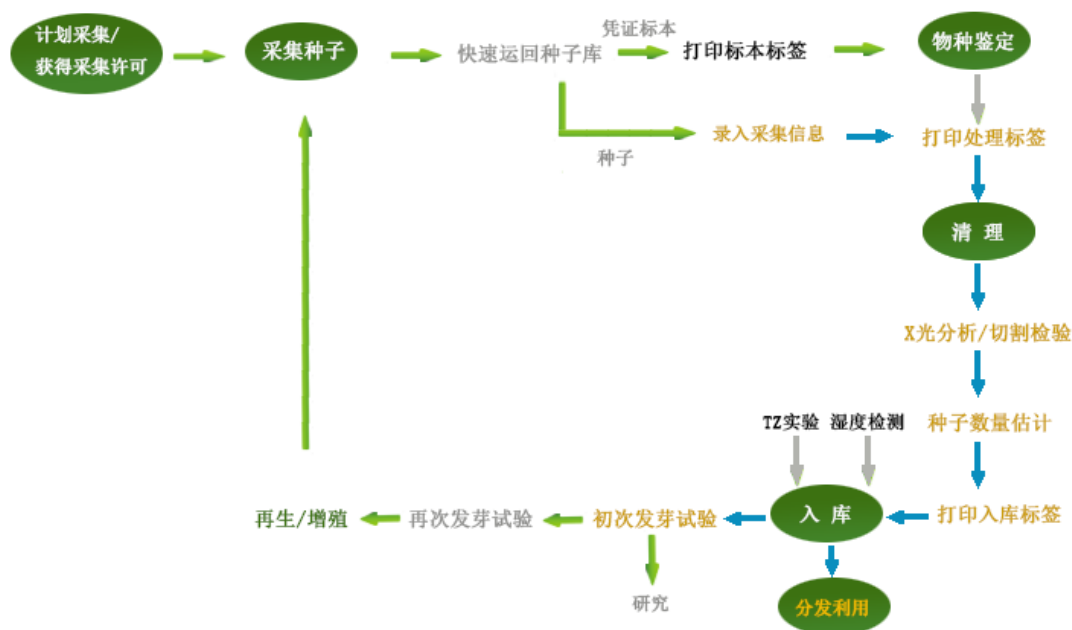


图1 野生植物种子采集保藏流程图

1.2 数据的加工与处理

GBIS 系统根据野生种质资源采集保藏的特点建立了标准化工作模式（图2），数据信息主要通过 GBIS 系统进行标准化录入。在借鉴英国“千年种子库”工作经验的基础上，种质库的种质采集保藏工作主要由种质资源的采集、登记、处理、保藏等前期工作，以及萌发、TZ 染色实验、水分测试等后期工作组成。为了能够全面准确地信息化种质资源的保藏活动，GBIS 系统以模块化设计和开发的形式完全融合了保藏种质资源的全部工作流程。每个模块的数据录入、校验、存储均由系统统一控制，规范数据的格式和内容，确保数据录入的准确性。在关键数据部分，采用二次鉴定和方差分析的方式实现人工或者自动化加工处理，确保数据的准确性。

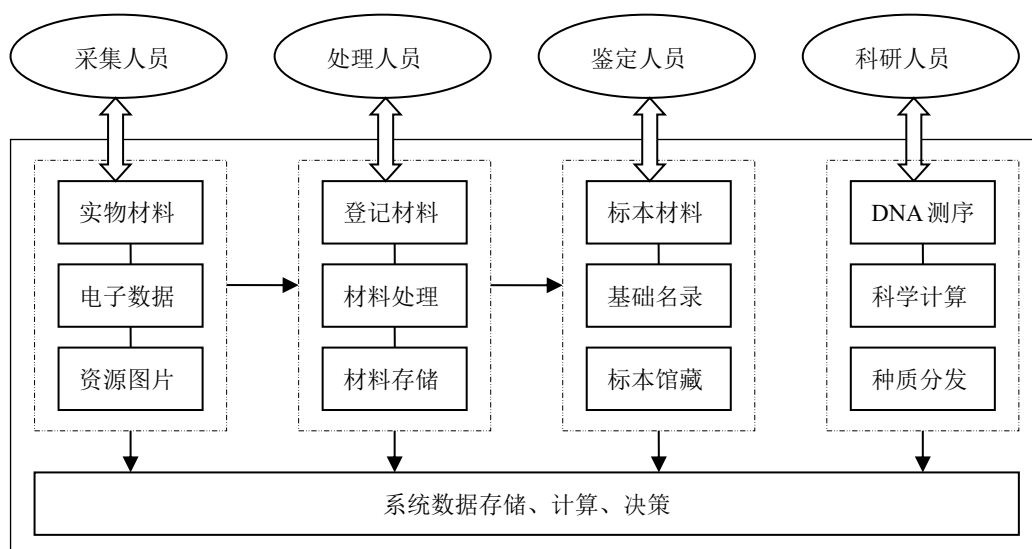


图2 GBIS 系统工作模式

2 数据样本描述

本数据库目前收录了超过 11.6 万条对外公开查询的信息数据，分别记录了：野生植物种子数据 71 200 余条，涉及 220 科 1940 属 9 484 种；植物离体材料数据 2 000 余条，涉及 104 科 377 属 1 850 种；植物 DNA 样品数据 43 300 余条，涉及 189 科 1 462 属 4 864 种。

数据样本以野生植物种子数据为例。其记录的信息包括序列号、采集编号、采集单位、负责人、采集年份、国家、省、市、区县、科中文名、科拉丁名、属中文名、属拉丁名、种中文名、种拉丁名、鉴定人、鉴定日期、图像、千粒重（克）、剩余种子数（表 1）。

表 1 毛柄杜鹃采集保藏数据展示

字段名称	示例
序列号	868710142722
采集编号	YDDXS0690
采集单位	云南永德大雪山国家级自然保护区管理局
负责人	李永亮
采集年份	2011
国家	中国
省	云南省
市	临沧地区
区县	永德县
科中文名	杜鹃花科
科拉丁名	Ericaceae
属中文名	杜鹃属
属拉丁名	<i>Rhododendron</i>
种中文名	毛柄杜鹃
种拉丁名	<i>Rhododendron valentinianum</i> Forrest ex. Hutchinson
鉴定人	李锡文（Li X. W.）
鉴定日期	2011-06-14
图像	IMAGE
千粒重（克）	0.059 600 0
剩余种子数（粒）	62 945

3 数据质量控制和评估

本数据库的数据主要来源于种质库采集保藏工作的各个环节。首先制定了针对种质资源采集保藏的标准规范，如由采集部门制定的种子采集技术规范，系统地规范了采集过程中产生的各个资源项，如采集号编写、采集地点、经纬度、生境等诸多信息^[6]。其次，在标准规范的基础上，依赖 GBIS 系统构建的关系型数据库，数据质量得到有效的控制。在应用数

数据库设计的同时,对常用项进行数据字典规范化,有效把控录入数据的质量。数据字典是 GBIS 系统中各类数据描述的集合,是详细描述各项种质资源数据收集和统计分析所获得的主要成果。例如在物种名称字典中,我们将 FOC 和 TPL 进行结合,得到标准的植物名称字典数据。在实际应用中,用户在录入时候通过选择标准植物名称,以达到规范植物名称规范输入的目的。

通过 GBIS 系统的有效录入和严格审核,实现了数据的有效管理,数据质量得到较大的提升,有效支持了中国西南野生生物种质资源库的各项研究工作。以野生植物种子保存为例,数据经历登记、质量检测、计数审核、标本鉴定和采集校验等多个环节。每个环节的数据均通过人工审定确认,以将错误率控制在 1% 以下。

4 数据价值

本数据库系统反映中国西南野生生物种质资源库的运行状态,并且涵盖了种质库的各项种质资源的保藏研究工作。例如数据库能够系统统计出保藏的科、属、种情况以及资源的利用状况。通过数据库集中数据,可以挖掘、评估种质资源的潜在价值,为进一步利用种质资源提供依据。本数据库收集了 2005 年至今 10 余年的野生种质资源保藏状况,并建立了方便的查询平台,对植物分类学、系统学、种子生物学、植物基因组学研究乃至整个植物学研究提供了有效的支撑。

5 数据使用方法和建议

本数据库主要通过 <http://www.genobank.org/> 提供 web 数据服务,在 Science Data Bank <http://www.sciencedb.cn/dataSet/handle/490> 也有访问入口。用户可以在输入框中输入物种名或者编号进行检索,或输入物种性状进行全文搜索。同时,注册用户经过电子邮箱验证后,可以进行种名和区域名称的精确查询。如果用户需要对某一类群资源或者定制某些特定字段数据进行批量数据获取,可以通过联系本数据库相关管理人员获取。

致 谢

本数据库得到中国科学院重大科技基础设施运行费的支持,数据主要来源于中国科学院昆明植物研究所所属种质资源保藏中心、分子生物学实验中心和标本馆等相关部门,在此表示衷心感谢!

数据作者分工职责

李拓径(1982—),男,广西河池人,硕士,工程师,研究方向为数据库技术。主要承担工作:数据库建设与管理。

李洪涛(1971—),男,湖南冷水江人,博士,高级工程师,研究方向为生物大数据。主要承担工作:数据库建设。

李德铨(1963—),男,江西南康人,博士,研究员,研究方向为植物分类、系统发育与

生物地理学。主要承担工作：项目组织实施。

参考文献

- [1] 中国西南野生生物种质资源库 [DB/OL]. (2015-01-13) [2017-09-15]. <http://news.genobank.org/bmjs/zykjj/>.
- [2] LI D Z, PRITCHARD H W. The science and economics of *ex situ* plant conservation [J]. Trends in Plant Science. 2009, 14(11): 614–621.
- [3] 王雨华, 李德铎. 国家战略种质资源保藏与创新系统的 e-Science 环境[J]. 科研信息化技术与应用, 2009 (02): 32–39.
- [4] 吴征镒. 中国植物志 (英文版). Flora of China [M]. 北京: 科学出版社, 2013.
- [5] The Plant list [DB/OL]. [2017-09-15]. <http://www.theplantlist.org/>.
- [6] 蔡杰, 张挺, 刘成, 等. 野生植物种子采集技术规范[J]. 植物分类与资源学报, 2013, 35(3): 221–233.

论文引用格式

李拓径, 李洪涛, 李德铎. 中国西南野生生物种质资源数据库[J/OL]. 中国科学数据, 2018, 3(1). (2018-02-23). DOI: 10.11922/csdata.2017.19.zh.

数据引用格式

中国西南野生生物种质资源库. 中国西南野生生物种质资源数据库[DB/OL]. Science Data Bank, 2017. (2017-09-29). DOI: 10.11922/sciencedb.490.

Germplasm bank of wild species in southwest China

Li Tuojing¹, Li Hongtao¹, Li Dezhu^{1*}

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

*Email: dzl@mail.kib.ac.cn

Abstract: Germplasm Bank of Wild Species is one of the large-scale scientific facilities in China, whose preserved wild germplasm constitutes an important national strategic biological resource. Data produced by the Germplasm Bank of Wild Species play a key role in the collection and preservation of wild germplasm. This database is built with modern computer information technologies based on standard specifications for the collection and preservation of wild germplasm. It stores and manages information on the collection, registration, examination, verification, curation, and regeneration of wild germplasm according to standard protocols, serving domestic and international research institutions, colleges and universities, government departments and the public. The database houses more than 116,000 pieces of data on such aspects as plant seeds, plant genomic DNAs, and plant in vitro materials. Users can retrieve data on the website of this database. The database will be conducive to the preservation and conservation of wild germplasm in China and has a positive demonstration significance for ecological protection and environmental management.

Keywords: wild species; wild germplasm; strategic biological resources; database

Database Profile

Title	Germplasm bank of wild species in southwest China
Data author(s)	Germplasm Bank of Wild Species
Data corresponding author	Li Dezhu (dzl@mail.kib.ac.cn)
Time range	2005–2017
Geographical scope	China
Data format	SqlServer
Data service system	http://www.genobank.org/ http://www.sciencedb.cn/dataSet/handle/490
Source of funding	Large-scale Scientific Facility of China on Germplasm Bank of Wild Species in Southwest China
Database composition	The database consists of more than 71,200 pieces of data on wild plant seeds, over 2,000 pieces on plants in vitro materials, and over 43,300 pieces on plant genomic DNAs.