



文献 DOI:

10.11922/csdata.2017.13.zh

数据 DOI:

10.11922/sciencedb.480

文献分类: 生物学

收稿日期: 2017-09-14

开放同评: 2017-09-22

发表日期: 2017-11-17

摘要: 中国有毒动物数据集整合了中国陆生有毒动物物种、动物毒素活性成分、有毒动物防治等数据, 共包括数据记录 1 371 条, 物种照片、蛋白质序列和空间结构图片等 1 362 张, 数据量约 29 MB。本数据集按照生物学专家和信息技术人员共同研究确定的标准建设而成。数据收集整理由专业人员完成, 并经过了专家审核以及计算机自动校验。本数据集可以为科研机构、大学和企业等提供有毒动物研究、动物毒素资源开发利用的数据和信息支持; 有助于提高公众对有毒动物的认识、防范和自我救助的水平; 有助于环保、海关、公安等政府部门开展有毒动物鉴定, 进行违法案件查处等。

关键词: 有毒动物; 物种; 毒素; 活性成分; 防治

中国生物多样性专刊

数据库(集)基本信息简介

Dataset Profile

数据库(集)名称 Title	中国有毒动物数据集 A dataset of venomous animals in China
数据作者 Data authors	王林、李文辉、朱建国 Wang Lin, Li Wenhui, Zhu Jianguo
通讯作者 Corresponding author	朱建国 Zhu Jianguo (zhu@mail.kiz.ac.cn)
地理区域 Geographical scope	中国 China
数据量 Data volume	29 MB
数据格式 Data format	*.xls, *.rar
数据服务系统网址 Data service system	http://www.sciencedb.cn/dataSet/handle/480
基金项目 Source of funding	中国科学院信息化专项, 科技数据资源整合与共享工程(2014年), 中国有毒动物数据库(XXH12504-3-13) The Science and Technology Data Resource Integration and Sharing Project "Database of Venomous Animals in China" (2014) of the Chinese Academy of Sciences Informatization Program (XXH12504-3-13)
数据库(集)组成 Dataset composition	数据集由两部分数据组成, 其一为有毒动物物种数据, 其二是动物毒素数据。共包括2个数据文件, 分别为: animal_toxin.xls和image.rar, 其中: (1) animal_toxin.xls含2个数据表单。表tbl_species存放物种相关数据, 共50条记录。表tbl_toxin存放动物毒素相关数据, 共1 321条

	数据记录。该Excel表格数据量0.7 MB；（2）image.rar中包含了1 362张图片，其中物种照片49张，毒素蛋白质序列图片1 249张，分子结构图片64张。该压缩文件的数据量为28 MB。 The dataset consists of two files: (1) animal_toxin.xls has two worksheets: the “tbl_species” worksheet contains 50 records related to venomous animals taxonomy and the “tbl_toxin” worksheet has 1,321 records related to toxins. It has a size of 0.7 MB; (2) image.rar contains 1,362 images, including 49 photos of the species, 1,249 protein sequence pictures and 64 molecular structure images. It has a size of 28 MB.
--	---

引言

中国有着丰富的有毒物种资源。21 世纪以来，动物毒素多肽的挖掘与利用在国际生物医学研究和创新药物开发中异常活跃。动物毒素的高活力、高分子多样性、强专一性以及不同地域动物毒素功能和结构的特殊性，这些特点使其在竞争激烈的新药研发中成为不可替代的天然药物资源^[1-4]。然而，一方面由于人类活动的干扰和对有毒动物的过度开发，导致生态环境恶化，自然界有毒动物的种类和数量不断减少和下降，少数物种甚至面临灭绝的危险。另一方面，最近 20 年，我国有明确文献记载的有毒动物中毒事件超过了 46 000 例，中毒病死率为 2.52%^[5]。因此，需要对公众普及有毒动物知识，加强有毒物种识别、危害和预防，物种和生境保护的科普宣传教育工作^[6]。中国有毒动物数据集集成了物种数据、动物毒素活性成分、有毒动物防治等数据，能在有毒动物研究和保护、资源的合理开发和利用，有毒动物的识别、防范和自我救助等方面提供全面有效的内容支持。

1 数据采集和处理方法

中国有毒动物数据集规划了 3 方面的内容，并应用数据库技术建立了相关关系，如图 1 所示。数据采集和处理都围绕着这 3 部分内容展开。

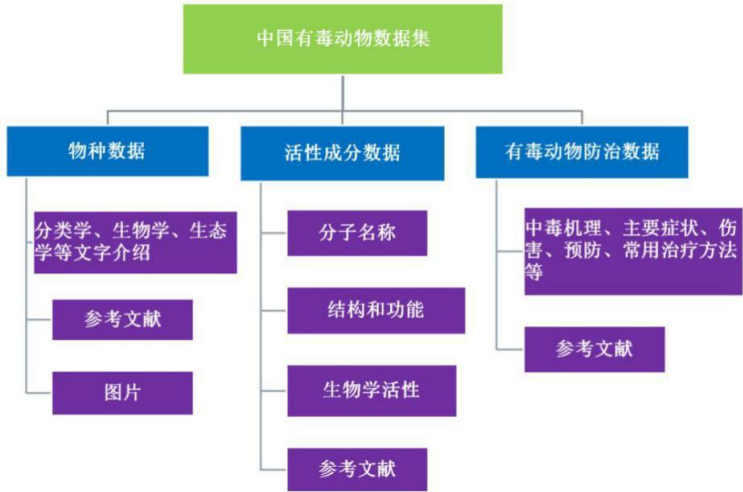


图 1 中国有毒动物数据集所含数据内容

1.1 数据来源

物种数据收集整理陆生脊椎动物和节肢动物的分类学、生态学等相关数据，数据来源有两方面：

- ① 动物志书以及国内外相关专业期刊上发表的文章；
- ② 物种照片由昆明动物所专家在野外收集数据时拍摄。

动物毒素活性成份数据包含分子名称、蛋白质序列，分子结构功能、生物学活性等内容，数据主要来自：

- ① 昆明动物所研究成果；
- ② 专业期刊上发表的文章以及权威网站（如美国生物技术信息中心，<http://www.ncbi.nlm.nih.gov>）。

有毒动物防治数据包括中毒机理、症状、伤害、预防、常用治疗方法等。同时附上了参考文献。

1.2 参考的标准规范

数据收集整理过程中参考了国家科技基础条件平台建设基础科学数据共享网项目组制定的《数据集核心元数据标准》，用于指导元数据的建立和著录^[7]；参照《专题数据库建设规范》用以指导数据库的建立，运行维护以及各类建库文件的撰写和使用^[8]；参照《数据资源加工指导规范》用于指导数据质量的控制和评估^[9]。

1.3 数据管理系统的开发

为规范管理数据，基于.net 和 SQL server 2000 开发了中国有毒动物数据管理系统，该管理系统的模块和功能如图 2 所示。

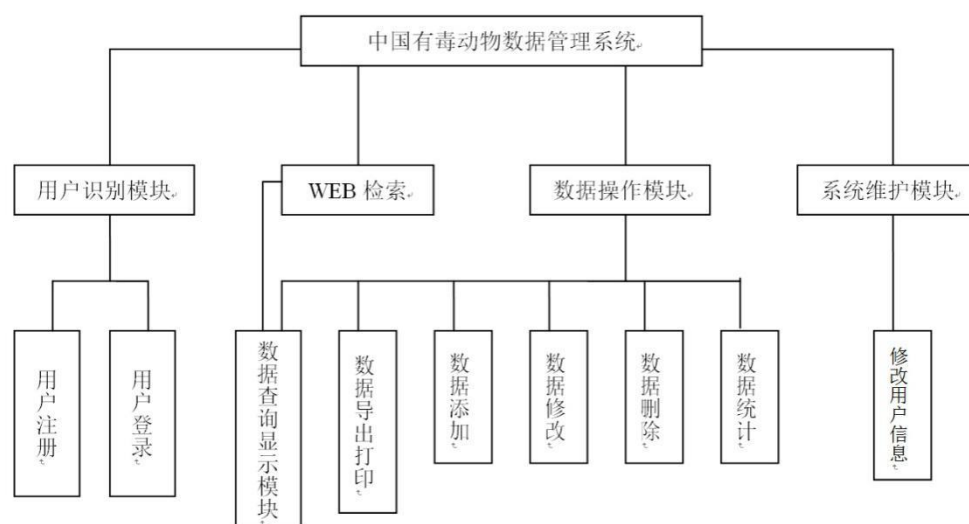


图 2 中国有毒动物数据管理系统的模块和功能

一般网络用户可以通过 WEB 检索方式查询数据（http://www.swanimal.csdb.cn/animaldb_youdu）；中国科学院昆明动物研究所所内授权用户可以对数据进行增、删、改、导出打印等操作。本文所附的中国有毒动物数据集是从已入库的数据中

导出到 Excel 表格而得。

2 数据样本描述

本数据集的结构化数据存储于 Excel 文件（animal_toxin.xls）的 2 个数据表单中：表 tbl_species 存放物种相关数据，表 tbl_toxin 存放动物毒素及有毒动物防治的相关数据。这 2 个数据表单的元数据描述如表 1 和表 2 所示。

表 1 tbl_species 数据表的元数据描述

字段名	字段类型	数据样本 ^a	关联的文件 ^b
id	短整型	35	/
物种中文名	字符型	中华蟾蜍	/
物种拉丁名	字符型	<i>Bufo gargarizans</i>	/
物种俗名	字符型	NA	/
目中文名	字符型	无尾目	/
科中文名	字符型	蟾蜍科	/
属中文名	字符型	蟾蜍属	/
鉴别特征	文本	鼓膜显著；皮肤粗糙，有不同形状的瘰粒；体背面颜色一般为棕黑色、棕褐色、黄褐色或灰褐色……体腹面浅黄棕色或黄白色。	/
形态	文本	雄性体长 67 mm，雌性 79 mm 左右。头宽大于头长……体腹面浅黄棕色或黄白色。	/
地理分布	文本	分布于宁夏、甘肃、青海、四川。	/
生态学特点	文本	生活于海拔 1 700~3 700 m 的高山溪流、沼泽草甸。产卵于静水坑。以昆虫为食。	/
参考文献	文本	费梁等，中国两栖动物图鉴，郑州，河南科学技术出版社，2000。P.136	/
物种照片	字符	2015070610226523578.jpg	image 压缩文件夹中的 2015070610226523578.jpg
有毒/泌毒部位	字符	皮肤腺	/
毒素类型	字符	混合毒素	/
中毒途径	字符	误食或使用带有分泌物的器皿进食	/
毒性描述	文本	中华蟾蜍的皮肤毒腺和耳下毒腺能分泌蟾酥，主要由生物原胺、蟾毒配基和蟾蜍毒三大类组成……蟾蜍毒和蟾毒配基均具有强烈的局部麻醉作用。此外，在蟾蜍中还发现了一些具有溶血作用的抗菌肽。	/

a: NA 表示没有被著录，在 Excel 表格中为空。

b: / 表示无与之关联的文件。

表 2 tbl_toxin 数据表的元数据描述

字段名	字段类型	数据样本 ^a	关联的文件 ^b
id	短整型	2254	/
物种中文名	字符型	中华蟾蜍指名亚种	/
物种拉丁名	字符型	<i>Bufo gargarizans gargarizans</i>	/
毒素分子中文名	字符型	水通道蛋白 1	/
毒素分子英文名	字符型	Aquaporin 1	/
蛋白质序列	字符型	Bufo gargarizans gargarizans-Aquaporin-Sequence-01.jpg	image 压缩文件夹中的 Bufo gargarizans gargarizans-Aquaporin-Sequence-01.jpg
空间结构	字符型	Bufo gargarizans gargarizans-Aquaporin-Structure-01.jpg	image 压缩文件夹中的 Bufo gargarizans gargarizans-Aquaporin-Structure-01.jpg
生物学活性	文本	能水解致病菌中的黏糖的碱性酶，主要通过破坏细胞壁中的 N-乙酰胞壁……导致细胞壁破裂而使细菌溶解。	/
中毒症状	文本	NA	/
治疗方法	文本	NA	/
蛋白质序列参考文献	文本	http://www.ncbi.nlm.nih.gov/protein/ACM51137.1	/
空间结构参考文献	文本	隋海心, 任昱. 2004. 水分子通道蛋白的结构与功能[J]. 化学进展, 16(2): 145-152.	/
生物学活性参考文献	文本	隋海心, 任昱. 2004. 水分子通道蛋白的结构与功能[J]. 化学进展, 16(2): 145-152.	/
中毒症状参考文献	文本	NA	/
治疗方法参考文献	文本	NA	/

a: NA 表示没有被著录，在 Excel 表格中为空。

b: / 表示无与之关联的文件。

物种照片，毒素蛋白质序列图片以及部分分子空间结构图片压缩保存在 image.rar 文件中。这些图片的文件名(.jpg)分别与 tbl_species 表中的“物种照片”字段，tbl_toxin 表中的“蛋白质序列”字段和“空间结构”字段相关联（表 1 和表 2）。

3 数据质量控制和评估

本数据集来源于中国有毒动物数据库^[10]。该数据库不定期更新，目前物种数据的覆盖范围仅限于中国，未来有条件时将考虑扩大覆盖范围。数据库的建设遵照中国科学院数据应用环境建设与服务项目组以及国家基础科学数据共享网项目组制定的有关标准规范完成^[7-9]。入库数据一方面来自中国科学院昆明动物研究所多年来的研究积累，另一方面来自公开发表的专著和学术论文或者权威机构的官方网站。数据入库前由专家对数据质量进行审核，发现有问题的数据将被再次组织确认，入库后的数据将被再次进行抽查，保证数据的可靠性和完整性。

为方便数据管理，我们基于 .net 和 SQL server 2000 开发了数据管理系统，部分数据字段可以通过该管理系统进行自动校验。同时，我们为中国有毒动物数据库建立了规范和完善的各类开发和说明文档，如需求说明书、软件设计概要说明书、软件设计详细说明书、数据采集与整理工作指南、建库工作规程和管理规定等。

我们还对部分非专业录入人员进行了相应的上岗培训，培训合格后才能进入数据录入工作。

最终，中国有毒动物数据库通过了领域专家组的验收，于 2015 年正式上线提供 WEB 检索服务。因此，本数据集的数据质量能够得到保证。

4 数据价值

美国国家生物技术信息中心（National Center for Biotechnology Information, NCBI）建立的蛋白质数据库，湖南师范大学建立的动物毒素数据库等收录了动物毒素的相关数据^[11-12]，但这类数据主要面向从事动物毒素研究的专业人士。少数医学、大众性网站或网页中涉及一些有毒动物（蜚）咬伤后的急救常识，但内容单一地分散在不同的非专业平台中，缺乏有毒动物系统知识和前沿领域的研究成果。本数据集集成了中国科学院昆明动物研究所乃至全国的研究成果，内容主要针对中国有毒动物物种数据，以及动物毒素活性成分以及有毒动物的防治等，全面且丰富，集科学性、权威性、实用性于一体。可以为科研机构、大学和企业提供有毒动物资源开发利用的数据支持；可帮助提高公众对有毒动物的认识、防范和自我救助的水平；有助于环保、海关、公安等政府部门开展有毒动物鉴定，进行违法案件查处，进一步保护中国的生物多样性。

5 数据使用方法和建议

本数据集结构较为简单，表 1 和 2 对数据字段进行了描述，便于用户的理解和使用。需要注意的是，毒素蛋白质序列是以图片.jpg 的格式存储，用户如需进一步编辑使用，需先从图片中提取信息。

致 谢

感谢中国科学院昆明动物所张云研究员对本数据集建设的指导和帮助。感谢李迎春先生在数据录入过程中所做的工作。

数据作者分工职责

王林（1975—），男，云南省丽江市人，硕士，助理研究员，研究方向为生物多样性信息学和景观生态学。主要承担工作：本数据集的建库和维护。

李文辉（1967—），男，云南省昆明市人，博士，副研究员，研究方向为两栖爬行动物毒素蛋白多肽分子生物多样性，结构与功能。主要承担工作：本数据集活性成分数据的收集和整理以及数据质量的把关。

朱建国（1962—），男，云南省昆明市人，学士，副研究员，研究方向为生态学和生物多样性信息学。主要承担工作：本数据集总体设计。

参考文献

- [1] Zhang Y. Why do we study animal toxins?[J]. Zoological Research, 2015, 36(4): 183–222.
- [2] 吕秋敏, 赖仞. 动物毒素及相关药物研究进展[J]. 药学进展, 2015, 39(12): 897–904.
- [3] 吕秋敏, 赖仞, 张云. 动物毒素与人类疾病——从单一成分到组学研究, 从理化性质到疾病机理, 从粗毒利用到理性药物设计[J]. 动物学研究, 2010, 31(1): 2–16.
- [4] 蓝海, 陈远聪. 中国毒蛇及蛇伤救治[M]. 上海: 上海科学技术出版社. 2008.
- [5] 何仟, 谢立璟, 马沛滨, 等. 我国有毒动物、有毒植物、毒蕈中毒现况分析[J]. 药物不良反应杂志, 2013, 15(1): 6–10.
- [6] 周静, 袁媛, 孙承业, 等. 2004–2013年全国有毒动植物中毒事件分析[J]. 疾病监测, 2015, 30(5): 403–407.
- [7] 国家科技基础条件平台建设基础科学数据共享网项目组. 数据集核心元数据标准[M]. 北京: 基础科学数据共享网, 2011.
- [8] 国家科技基础条件平台建设基础科学数据共享网项目组. 专题数据库建设规范[M]. 北京: 基础科学数据共享网, 2011.
- [9] 国家科技基础条件平台建设基础科学数据共享网项目组. 数据资源加工指导规范[M]. 北京: 基础科学数据共享网, 2011.
- [10] 中国科学院昆明动物研究所. 中国有毒动物数据库[EB/OL]. (2015–03–01) [2017–09–07]. http://www.swanimal.csdb.cn/animaldb_youdu.
- [11] National Center for Biotechnology Information. NCBI Protein Database[EB/OL]. (1993–10–10) [2017–09–07]. <https://www.ncbi.nlm.nih.gov/protein>.

[12] College of Life Sciences, Hunan Normal University. Animal Toxin Database[EB/OL]. (2009-02-25) [2017-09-07]. <http://protchem.hunnu.edu.cn/toxin>.

论文引用格式

王林, 李文辉, 朱建国. 中国有毒动物数据集[J/OL]. 中国科学数据, 2017, 2(4). (2017-11-13). DOI: 10.11922/csdata.2017.13.zh.

数据引用格式

王林, 李文辉, 朱建国. 中国有毒动物数据集[DB/OL]. Science Data Bank, 2017. (2017-09-14). DOI: 10.11922/sciencedb.480.

A dataset of venomous animals in China

Wang Lin¹, Li Wenhui¹, Zhu Jianguo^{1*}

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming
650201, P. R. China;

* Email: zhu@mail.kiz.ac.cn

Abstract: This dataset of venomous animals in China integrates the data of taxonomy and toxin active ingredients of terrestrial venomous animals in China, as well as envenomation, poisoning and treatment therapy. The dataset contains 1,371 records, 1,362 pictures of species, protein sequence and molecular structure, and the data volume is 29 MB. The structure and standard of the dataset are designed and determined by experts of zoology and IT technicians together. Biological experts collected and reviewed the data before importing them into the database. A data management system is designed and programmed to process all the collected data and to perform quality control. The dataset supplies useful data and information to the research and development of venomous animals. As the bioactive constituents from toxin are one of the important source of biomedicine, the dataset contributes to identifying and preserving venomous animals, providing knowledge on poison prevention and treatment caused by venomous animals.

Keywords: Venomous animal, Species, Toxin, Active ingredient, Treatment