

研究报告

根瘤菌资源数据库(RRDB)的建立

李颖 陈文新

(北京农业大学生物学院, 北京 100094)

摘要 在总结我室多年对根瘤菌资源调查与分类研究的基础上,建立了根瘤菌资源数据库(RRDB)。此库依据根瘤菌研究特点而设计,包括:基本信息、采集信息、保藏信息、回接信息、参考文献、性状信息、菌株说明等七个子库。目前收录了来自国内21个省或地区及国外部分研究单位提供的,并经全面性状分析与分类研究的286个菌株的信息,每个菌株设有寄主来源、固氮酶活性、碳源、氮源的利用等321个数据项。该库具有数据维护、查询检索、数据统计、输出打印等功能,引入并连接了聚类分析软件包-MINTS系统,可以满足当前研究的需要。

关键词 根瘤菌资源,数据库技术,计算机应用

据美国 Wisconsin 大学 O. N. Allen 和 E. K. Allen^[1]估计,目前世界上有748属近20,000种豆科植物,其中做过结瘤情况调查的不过3,000多种,对其共生体系经过研究的仅占0.5%,且多局限在一些常见的农作物、牧草、绿肥等植物。我国地大物博,资源丰富,约有15,000种豆科植物,为充分开发祖国的资源,自70年代末,我室相继开展了对新疆、海南全境以及四川、西藏、山西、黑龙江、河北、湖北(神农架林区)等部分地区的根瘤菌资源调查,分离了近千株不同寄主、不同生态环境的栽培及野生豆科植物的根瘤菌菌株,发现其中一百多种植物的结瘤情况在 Allen's 所著《THE LEGUMINOSAE A Source Book of Characteristics, Uses, and Nodulation》一书中未曾记载,对分离纯化的部分菌株进行数值分类^[2-4]、全细胞蛋白电泳^[5]、核酸分析^[6]等研究,证实我国不同地区的根瘤菌有很大的异质性,一些菌株抗逆性强,耐盐碱,是我国特有的种质资源。目前,根瘤菌分类工作中研究对象太局限,尚有大量的已知菌未给予分类命名,已有的分类系统与系统发育关系尚不一致。为此,“八五”期间我们

将有计划、有步骤地继续对我国三北(东北、西北、华北)地区进行广泛的根瘤菌资源调查与分类研究,并逐步推向全国。

为总结、保存以往的研究数据和今后研究的发展需要,构建了我国第一个根瘤菌资源数据库。此库的建立,将为材料的存贮管理、数据的分析预测、信息的查询统计等方面带来便利,并为建立我国根瘤菌分类中心及撰写《中国根瘤菌资源》专著奠定基础。

1 数据的收集、整理与规范化处理

数据的收集、筛选、归类及数据项的确定是构建本库的基础。将我室十几年来积累的实验数据及野外采集记录,对应每个菌株分别编号登卡。将所得资料按建库需要分为七类,即:“基本信息”、“采集信息”、“保藏信息”、“回接信息”、“参考文献”、“性状信息”、“菌株说明”等(信息的划分及各子库的建库原则见库结构设计)。因以往数值分类分别进行性状指标比较分散,经统计近600个,为使今后数据产生逐步走向标准化、规范化,我们确定具有鉴别特征的生

理生化及其它表型性状 248 项并分别规定了测定标准,加之其它来源的信息 73 项,每个收录的菌株确定 321 个数据项。

2 硬件及软件的配备

2.1 硬件

长城 286B 型微机、80 兆硬盘、1 兆内存
CEGA 高精度彩色显示器
EPSON LQ 1600K 打印机

2.2 软件

CCDOS 3.3 操作系统
XE 中文编辑软件
CW 中文录入软件
FoxBASE 2.1 关系型数据库软件

3 根瘤菌资源数据库结构及功能设计

3.1 库结构设计

3.1.1 概况:依数据来源和研究工作需要,将每类信息各建一个子库。各子库均以菌株编号相沟通。字段总长度为 1992 个字节,当前库容为 408.2 千字节。

3.1.2 各子库数据结构:

3.1.2.1 基本信息子库:本子库旨在说明每个菌株的基本情况,共设定字段 23 个,字段总长度 480 个字节。字段包括:编号、菌号、菌名、定名者、鉴定时间、鉴定者、鉴定单位、别名、别号、国别、提供单位、提供者、标准菌否、培养基、培养温度、血清型、分离号、分离时间、分离者、分离单位、分离基物、保存方式、DNA 碱基比等。

3.1.2.2 采集信息子库:此子库的建立,目的是保存菌株的寄主信息及其所处的生境信息,共设字段 31 个,字段总长度为 344 个字节。字段包括:编号、采集日期、采集人、采集省县、采集地点、寄主拉丁名、寄主中文名、当地名称、地形、经度、纬度、海拔、植生型、年降雨量、降雨季节、土壤条件、表土色、排水情况、土壤 pH、土地使用、经营情况、植物形态、根特征、根瘤着生、根瘤颜色、根瘤形状、根瘤大小、自然酶活、植物生长期(指生长年限)、生长阶段、胶卷编号(指采集时植物照像编号)等。

本子库记录了某一自然状态下,根瘤菌与

豆科植物共生的环境条件。由于长期自然选择,根瘤菌在一些特殊的环境中生存,会出现相应的生理生化特征,如耐盐碱、耐高温或低温等。其种群分类可能与生态条件发生联系,详细记录菌株所处生态环境的信息,以便我们分析研究生境与生理生化性状关系及变化规律,为选择植物-菌株优良组配和抗逆性强的种质资源提供依据。

3.1.2.3 保藏信息子库:保藏信息子库共设字段 4 个,包括:编号、保藏方式、保存数量、制作日期等,字段总长度为 27 个字节。

我室现已收集、保存了大量的根瘤菌菌株,并作为一个较重要的保藏单位列入国际细菌库目录,为进一步与世界各国交流信息、交流菌株特设此库。

根据前人经验及我们试验的结果,将根瘤菌细胞用 20—30% 的甘油混合后,置 -30℃ 低温保存效果较好,冷冻干燥的方法也适用于长期保存。因此,入库的菌株均采用三种保存方式,即:“冷冻干燥”、“-30℃ 低温”、“油管常温”,前两种用于较长时间的保存,后一种时间稍短,适用于日常研究使用。

目前保藏库共收藏近千个根瘤菌菌株,保藏时间最长的已达 10 年。

3.1.2.4 回接信息子库:回接信息子库共设字段 7 个,包括:编号、回接植物、回接基质、结瘤天数、结瘤温度、光照强度、酶活测定等,字段总长度为 101 个字节。

豆科植物与根瘤菌共生固氮,虽早已被人们所认识,但其结瘤机制尚待搞清,根瘤菌对寄主的专一性和有效性是其在自然界长期选择的结果,至今尚未完全揭开二者辨认结合的机理。采集并分纯的菌株,必须先做回接实验确定其能使原寄主结瘤并有固氮酶活性,才符合柯赫原则,确认其寄主来源,在根瘤菌的分类鉴定中,此项工作必不可少,目前还没有简便可靠的实验手段来替代回接实验。回接信息子库是根瘤菌与植物共生的特殊要求。此外,根瘤菌与豆科植物的交叉结瘤与有效性的研究也将为生产

提供科学依据。详尽记录回接与互接信息,掌握豆科植物与根瘤菌菌株组配结瘤固氮的最佳条件,寻找一定的规律,将会对农业生产起促进作用。

3.1.2.5 参考文献子库:为避免重复输入参考文献,节约存贮空间,在结构设计上将参考文献信息分为两部分。其一为参考文献子库,共设8个字段,字段总长度为402个字节,此子库主要记录参考文献的来源、作者、文章类型、文章语种等信息;其二为文献摘要子库,共设三个字段,字段总长度为74个字节,用以记载文献摘要。

3.1.2.6 性状信息子库:菌株的生理生化及表型性状,一直作为数值分类的依据。如何选择性状指标,目前尚无固定的模式可循,研究者们仍在探索和试验。本室曾分别选取136—241个生理生化及其它表型性状进行分析。根据我室对根瘤菌进行数值分类的经验,参照Moffett等^[7]和Graham等人^[8]的实验,我们选定248项具有鉴别特征的性状作为近期内对未知菌株进行数值分类的依据,此子库是在这个基础上设计的,包括:“唯一碳源的利用”68项;“唯一氮源的利用”33项;“对抗生素的抗性”72项;“对维生素的需求”11项;“对染料的耐受性”16项;“对化学药物的耐受性”2项;“耐盐性”9项;“生长pH”9项;“生长温度”6项;“酶活性”5项;“其它性状”17项。每类信息各构成一个数据项(字段),字段长度则代表此类性状信息所需测定的项目数。性状选取的原则是尽可能广泛而均匀地遍布于所研究的机体上,以上性状的选定是否合理,还有待于在今后的研究工作中进一步验证。我们将在大量收集菌株进行数值分类的基础上,逐步摸索出一套既简便可靠,又相对标准化、规范化的实验手段。

3.1.2.7 菌株说明子库:菌株说明子库共设3个字段,包括:编号、菌株历史、典型菌说明等,字段总长度为173个字节。

本子库设立的数据项并非每一株菌都具有。“菌株历史”主要记录菌株的特殊来源、曾用

名、或在各其它保藏单位的菌号等信息;“典型菌说明”主要记录模式菌株的特殊说明。

3.2 根瘤菌资源数据库功能

根瘤菌资源数据库具有“数据维护”、“查询检索”、“输出打印”、“数据统计”、“聚类分析”等功能。全部命令文件(除聚类分析软件包以外)总计326.2KB(千字节),在实现各个功能的设计上,力求遵循操作简便、快速、准确的原则。

3.2.1 数据维护:此模块包括数据输入、信息追加、数据删改、文献维护等内容,由数据库专职管理人员操作,“数据输入”用于菌株信息的录入及追加;“信息追加”用于保藏信息、回接信息、文献编号等一菌多记录信息的添加;“数据删改”用于菌株各类信息的修改或删除;“文献维护”用于参考文献信息的录入、添加、修改及删除。由于每株菌的信息较多,各类信息连续录入,每录入一屏后,都留有充分的复核时间检查或修改。此功能力求做到合理安排屏幕格式,提示信息清晰,用户界面友好,使用安全方便。

3.2.2 查询检索:查询检索模块包括浏览查询、指定查询和条件查询三部分。用户可根据实际情况通过不同的模块达到查询信息的目的。

a. 浏览查询

此模块分为两部分,其一为“顺序浏览”,可通过库中的任意编号进入到某菌的基本信息做前、后顺序浏览,并能查询选定菌株的其它信息,适用于全部菌株信息的查询。其二是“分类浏览”,可对菌株的采集、保藏、回接等信息做分别浏览,适用于单独查询菌株的某类信息。

通过“浏览查询”,可对库内信息综合了解。从键入编号到屏幕上出现信息,只需要1—2秒种的时间。

b. 指定查询

进入此功能后,可分别通过菌株的编号、菌号、分离号查询指定菌株的全部信息。

c. 条件查询

通过此模块,可任意设置检索条件进行菌株信息的查询。

息中挑选出可能用到的检索项目(字段)共计26个,加上性状信息248项,用户可将这些信息任意组合进行查询。所有用到的条件均可在屏幕上显示供选择,不必自行输入条件组合的表达式,只需根据提示,用光标进行条件的组配。

例如:查询条件为 菌名:“*Rhizobium huakuii*”;鉴定者:“祁幼林”;定名者:“陈文新等”;提供者:“曹燕珍”;保藏方式:“冷冻干燥”;葡萄糖利用:“阳性”;蔗糖利用:“阳性”,以上条件选择后,只需十秒钟即可显示查询结果。按上述条件,屏幕上先显示符合条件记录的编号,还可根据查询者需要显示记录的详细内容。此功能的设计优点是:对于不熟悉计算机的查询者来说,采用此方法十分直观、方便,由于不必输入汉字和表达式,不易出错,从而提高了查询效率。

3.2.3 输出打印:用户的查询结果,均可在此模块的驱动下,按信息的类别打印成不同的表格。表格根据每类信息的特点设计,可一次分页打印某菌株的全部信息,也可根据需要打印菌株的某类信息,所有操作按选择菜单方式完成,由于执行此功能,需要连接打印机,因此在程序设计上注意增加了必要的提示和相应的保护措施。

3.2.4 数据统计:统计功能的实现,可及时掌握数据库中各类信息当前的状态,如发现错误,可随时纠正。

随机统计的项目包括以下内容:

a. 数据概况统计

屏幕可首先向用户展示当前数据库存贮信息的概况,包括菌株总数、菌株保存总数、原寄主类型总数、回接植物种类总数、生理生化性状测定总数等。

b. 基本信息统计

对基本信息子库中的数据项进行统计,包括菌株属、种;菌株分离者、鉴定者、提供者及其单位等。

c. 采集信息统计

采集信息子库的统计内容包括菌株的采集人;采集时间;采集省份和具体地点;采集地所处的经度、纬度及海拔高度等。

d. 保藏及回接信息统计

对保藏信息子库进行的统计内容有不同保藏方式及菌株的保存数量;对回接信息子库的统计包括被回接植物的种类及数量;回接菌株的种类及数量等。

e. 参考文献信息统计

对参考文献子库进行的统计包括文章类型、文章语种、刊物名称等。

以上统计结果,均可按表格形式随机打印。

统计功能的实现,能及时了解掌握数据库的存贮信息,使以往几小时甚至几天才能算清的数据,瞬间即可完成,既节约时间又准确、可靠。

3.2.5 聚类分析软件包的连接及应用:根瘤菌资源数据库引进了中国科学院微生物研究所马俊才、赵玉峰开发的专用于细菌数值分类的软件包-MINTS^[9]。此系统包括13种计算菌株相似性的方法和八种聚类方式,可自动绘出树状图。在本室的根瘤菌分类研究中曾多次运用该软件。实践证明,MINTS软件基本满足了目前根瘤菌数值分类的研究,并按国内外软件应用惯例已合法引进。我们还为此设计了连接程序,将MINTS软件与本库相沟通。在进入数据库“聚类分析”功能后,首先通过人机对话的方式向用户提问,将用户根据提示键入的信息自动整理归类,生成进行聚类分析必备的菌名、菌株数、分析项目数及性状代码等四个文件后,即可运行MINTS。

由于目前根瘤菌数值分类的研究工作还没有做到标准化、统一化,性状分析的结果,除了与所选择的性状有关外,测定数据受人为因素干扰较大,不同研究者的操作方式、不同时间、药品的纯度差异或药品的批号以及菌株培养条件的差异,会给结果带来影响,特别是靠肉眼观察判定的颜色反应鉴别很难一致。因此,用于聚类分析的材料,最好是选择同一人同期内的实

验数据进行比较才较可靠。一般情况下,菌株只有在经聚类分析完成后,才将有关信息送入数据库中保存。此功能是根据目前工作要求设计,还需要逐步建立一套规范化的测试手段,以满足从数据库中随意提取数据进行聚类的功能。

4 结束语

根瘤菌资源数据库的建成,是运用较成熟的数据库技术为本学科研究服务的有益尝试,是在总结、整理我室十几年来对根瘤菌研究的基础上,参照国内外同行的相近研究构建的。它的结构依据根瘤菌研究内容和特点而设定,作为国内第一个根瘤菌资源数据库,它的数据来源准确可靠,收集的菌株信息覆盖面较大,数据项比较完整,在程序的设计中,力求做到使用简单,易于操作,用户界面友好,查询统计快速,它将对根瘤菌信息的管理与交流、菌株资源的调查与鉴定、信息的归类与分析、资源的开发与利用等方面带来方便。特别是为“八五”、“九五”期间推向全国的根瘤菌资源调查与分类工作创造了信息管理的必要条件。

目前研究过的结瘤豆科植物中,多为农作物,而绝大多数野生豆科植物的根瘤菌还有待于开发和利用。随着研究工作的深入开展,信息、数据的积累不断增多,本库还将提供豆科植物根瘤菌最适生长条件的选择;植物-菌株最佳组配的判定;不同生态地区有效菌株的使用等信息,并结合农业生产的实际需要,为根瘤菌资源的开发和利用提供更广泛的信息。

近20年来,国际上采用现代分类技术对根瘤菌分类做了很多研究,废弃了以“互接种族”的观念定属、种的体系,重新制定了比较合理的根瘤菌分类系统。1991年国际系统细菌学杂志(IJSB)上公布了描述根-茎瘤菌新属、种的最低标准^[10],强调细菌命名必须尽最大可能反应基因关系。该标准以遗传特征为重点,同时结合表型特征与共生特征。定种时需要测试的项目包括与寄主的共生表现;培养和形态特征;DNA-

DNA 同源性;rRNA;DNA 杂交或 16s rRNA 序列分析;DNA 限制酶切片多型性及多位点酶电泳等。

依据此标准,我室将逐步扩大研究项目。根瘤菌资源数据库的存贮信息和软件功能正在进一步扩大和完善,除存贮现有的信息外,还将收集存贮 DNA-DNA 同源性、核酸序列分析结果、多位点酶切图谱等信息。在逐步增加测试项目的基础上,将信息以多媒体图文形式入库,使图文最新信息管理技术在本库中实现。

目前本库内录入的菌株包括我组自行分离鉴定并分类过的来自国内 21 个省、市的 258 个以及从国外引进的部分菌株,共计 286 个。随着工作的继续深入,争取在“八五”计划完成时,此库收集的菌株量达到 1,500—2,000 个。

致谢: 中科院微生物所赵玉峰高工、马俊才工程师对本工作提出建议并给予帮助;在工作的过程中得到本专业杨苏声教授、吴柏和副教授、汪恩涛副教授的帮助,在此一并致谢。

参 考 文 献

- [1] O N Allen, E K Allen. The leguminosae. A Source Book of Characteristics, Uses, and nodulation. The University of Wisconsin Press. 1991.
- [2] 牛天贵, 陈文新, 李季伦, 等. 北京农业大学学报, 1986, 12(1): 57—64.
- [3] 陈文新, 祁幼林, 李季伦, 等. 微生物学报, 1988, 28(2): 102—108.
- [4] 孙建光, 章凡, 王昌平, 等. 微生物学报, 1993, 33(2): 135—143.
- [5] 汪恩涛, 陈文新, 李季伦, 等. 微生物学通报, 1987, 14(2): 86—88.
- [6] 李广善, 陈文新. 微生物学报, 1992, 32(2): 151—154.
- [7] M L Moffett, R R Colwell. J Gen Microbiol, 1968, 51: 245—266.
- [8] 马俊才, 赵玉峰. 微生物学通报, 1986, 13(5): 225—228.
- [9] P H Graham. J Gen Microbiol. 1964, 35: 511—517.
- [10] P H Graham, M J Sadowsky, H H Keyser, et al. Int. J Syst Bacteriol, 1991, 41: 582—587.

ESTABLISHMENT OF THE RHIZOBIAL RESOURCES DATABASE

Li Ying Chen Wenxin

(*Beijing Agricultural University, Beijing 100094*)

Abstract Based on the former studies and serving for the further researches, the Rhizobial Resources Database (RRDB) was constructed. According to the specificities of rhizobia, the RRDB was composed of seven sub-database including the informations of general description, collection, storage, inoculation, reference, characteristics and remarks, Confident records of 321 data items containing the hosts, nitrogenase activities, utilization of carbon and nitrogen sources and enzyme reactions etc. covered the above mentioned seven subdatabase for 286 strains have been collected in this RRDB.

The main functions of this RRDB were data management, searching, statistics and stamp. For clustering and identifying strains, the numerical taxonomic software package named MINTS was introduced from Institute of Microbiology Academia Sinica, and linked to RRDB.

Key words Rhizobial resources, Database technique, Application of computer