

中国微生物资源数据库细菌性状子库的 数据结构与检索功能设计*

马俊才 赵玉峰 王大帮 张克莉

(中国科学院微生物研究所,北京)

张荣肖 江福湘 杨立芝 张悦

张建昌 孙慈 潘玉平 王晓峰

(中国科学院计算中心,北京)

摘要 提出了构建中国微生物资源数据库 CDMR(China Database for Microbial Resources) 的方案。该库包括微生物性状库、微生物产品的商品信息库、微生物名称库以及微生物名词库四个分库。本文详细介绍了细菌性状子库的数据结构与检索功能。

关键词 数据库;微机应用

由于微生物学与经济发展和社会进步有着密切联系,微生物资源也日益为人们所重视。因而各国纷纷建立各种类型的菌株库^[1,2]。随着菌株库的建立,并为了充分发挥库内菌株在资源开发中的作用,菌株数据库也随之建立起来。欧美、日等国纷纷建立菌株数据库。然而,迄今为止我国尚没有一个为管理菌种库和微生物资源开发用的多功能的综合数据库。

为此我们提出构建中国微生物资源数据库。本数据库除收集菌株的性状及有关的管理资料外,还拟建立微生物产品的商品信息库、微生物学名词和名称库等。

(一) 性状库的数据结构

在这四个库中,我们首先开始了微生物性状库的构建。微生物数据库中存储的数据类型一般分为两类。一类是真正的微生物性状数据(如形态、生理生化等数据);另一类是与之有关的综合管理数据(如菌株来源、菌株用途、参考文献、保藏单位、菌株历史等)。我们使用 12 个表型结构来存储这两方面的数据。

下面分别说明各表的结构及存储内容。

1. 设置目的:记录综合管理数据等有关文字型信息。

2. 记录主序:菌号序。

3. 记录长度:720 个字节。

4. 数据存储规则:

A. 菌号:中国微生物菌种保藏管理委员会下属的普通微生物菌种保藏管理中心的所有菌号均以“AS”开头。在本库中对于所涉及的 AS 菌株编号,我们通过建立菌株别号表来记录这种对应关系。对于本库中尚未批准升格为 AS 型的菌株,我们用“PR-Potential Resource”作为其菌号前缀,规定其书写格式:“PRx.nnnnn”。其中“x”为菌株类别:“1”代表细菌、“2”代表放线菌、“3”代表酵母菌、“4”代表小型丝状真菌、“5”代表担子菌。“nnnnn”为菌株的顺序编号。

B. 鉴定单位、分离单位、保藏单位、提供单位均采用 20 个字符以内的缩写形式。例如:“IMAS”代表中国科学院微生物研究所、“IMASBACT”代表微生物所细菌室。

C. 保密级别:“0”代表不保密、“1”代表国

* 本项目得到中国科学院科学数据库和院生物局微机领导小组的部分资助,特此致以衷心感谢。
中国人民大学信息中心吴志宏、汪左涓同志参加部分工作,特此鸣谢。

表 1 概述表(表名 MCBTSPCF)

项号	项目名称	是否为检索项	字段长度
1	菌号	✓	CHAR(9)
2	菌名	✓	CHAR(70)
3	鉴定单位缩写	✓	CHAR(20)
4	鉴定者	✓	CHAR(40)
5	标准菌株否	✓	CHAR(1)
6	分离省市	✓	CHAR(20)
7	分离县	✓	CHAR(60)
8	分离具体地点	✓	CHAR(70)
9	分离号	✓	CHAR(30)
10	分离单位缩写	✓	CHAR(20)
11	分离者	✓	CHAR(20)
12	分离年	✓	CHAR(4)
13	分离月	✓	CHAR(2)
14	分离日	✓	CHAR(2)
15	生态环境	✓	CHAR(60)
16	栖息地水平 1	✓	CHAR(40)
17	栖息地水平 2	✓	CHAR(40)
18	栖息地水平 3	✓	CHAR(40)
19	经度	✓	FIXED(5,2)
20	纬度	✓	FIXED(5,2)
21	海拔高度	✓	FIXED(5)
22	培养基 1 号	✓	FIXED(4)
23	培养基 2 号	✓	FIXED(4)
24	培养温度	✓	CHAR(7)
25	存活	✓	CHAR(1)
26	血清型	✓	CHAR(30)
27	保藏单位缩写	✓	CHAR(20)
28	保藏者	✓	CHAR(20)
29	保藏方式	✓	CHAR(30)
30	提供单位缩写	✓	CHAR(20)
31	提供者	✓	CHAR(20)
32	保密级别	✓	CHAR(1)

内、“2”代表科学院内部、“3”特定的一部分用户、“4”代表提供者本人及其研究组。凡 0 级的记录可允许所有用户检索。

表 2 菌株别号表(表名 MCBTSYNO)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌号	✓	9
2	别号	✓	20

1. 设置目的: 建立现用 PR 菌号与其它机构菌号的对应关系。在各保藏机构间交换菌株时, 会出现同一菌株在不同的保藏机构中具

有各自不同的菌株编号的情况: 本表旨在指明这种不同菌号的对应关系。

2. 当某一菌株的别号数目不止一个时, 只需再增加若干个记录即可。

3. 记录主序: 菌号序。

4. 记录长度: 29 个字节。

表 3 菌株别名表(表名 MCBTSYNM)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌名	✓	70
2	别名	✓	70

1. 设置目的: 记录现用菌种名称的变迁历史及一些在分类上尚有争议的菌名, 随着微生物分类系统的逐步完善和分类技术的日益提高, 微生物的分类体系有时会发生改变, 甚至会有整个一个属被划入另一个属的情况, 使得菌名经常改变。同时同一种微生物有时会有好几种不同的名称。为了记录这种情况并达到菌名规范化的目的, 特设立本表。

2. 当某一菌株的别名数目不止一个时, 只需再增加若干个记录即可。

3. 记录主序: 菌名字母序。

4. 记录长度: 140 个字节。

表 4 用途表(表名 MCBTUSAG)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌号	✓	9
2	用途类别	✓	40
3	用途细则	✓	70
4	开发单位	✓	70
5	开发者	✓	40
6	开发现状	✓	50
7	用途文献	✓	100
8	保密级别	✓	1

1. 设置目的: 建立菌号与用途的对应关系, 并建立保密级别。用途是开发利用微生物的一个重要信息, 也是本库的经济效益之所在。

2. 将用途分为用途类别和用途细则两个水平, 以利于用途的检索和名称规范化。

3. 记录主序: 菌号序。

4. 记录长度: 380 个字节。

表 5 菌株性状表(表名 MCBTCHAR)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌号	✓	9
2	保密级别		1
3	性状	✓	200

1 设置目的: 为将微生物性状的自然语言描述(用文字)转换成数字描述(用数值), 故采用 RKC 代码方式, 从而将一个文字型数据库转变为数值型数据库, 以节省系统资源, 提高检索速度。假设在 500 株菌的细菌子库中(按每株菌有 150 个有效性状, 每个性状的自然语言描述用 100 个字节, 500 株菌共使用 200 个 RKC 代码, 每个代码的文字描述用 186 个字节计算), 使用 RKC 代码方式仅占系统存储容量: MB 为兆字节。 $200(\text{项}) \times 500(\text{株}) + 200(\text{项}) \times 186(\text{字节}) = 0.137\text{MB}$ 。如果不使用 RKC 代码方式:

$150(\text{项}) \times 100(\text{字节}) \times 500(\text{株}) = 7.5\text{MB}$ 。

从而可以大大节省系统资源。特别是在库中菌株数目很大时, 这种节省尤为显著。同时, 使用了 RKC 代码方式后, 由于可以使用定位的纵列查找方式, 所以使用 RKC 代码方式还可以节省检索时间。

另外由于微生物性状数目很大, 性状类型差别很大, 文字描述复杂, 很难做到性状描述的规范化。由于 RKC 代码将几乎所有的微生物性状进行了编码, 从而一定程度上有助于微生物性状规范化问题的解决。

2. 数据类型: 记录全部为菌株的 3 值型数据 (0, 1, N) 和数值型数据。

3 记录主序: 菌号序。

4. 记录长度: 随菌株量的增加而上升。500 株菌使用的 RKC 代码约为 200 个。

表 6 RKC 代码说明表(表名 MCBTRKCI)

项号	项目名称	是否为检索项	字段长度(字节)
1	RKC 代码	✓	6
2	内容叙述	✓	180

1. 设置目的: 作为 RKC 代码说明字典,

建立 RKC 代码与性状描述间的对应关系, 供输入输出时参照使用。

2. 记录主序: RKC 代码值序。

3 记录长度: 186 个字节。

表 7 RKC 代码位置表(表名 MCBTRKCM)

项号	项目名称	是否为检索项	字段长度
1	RKC 代码	✓	CHAR(6)
2	在表 5 中的位置		FIX(5,0)

1 设置目的: 此表为系统专用, 用于建立 RKC 代码与在性状表中的位置间的对应关系, 从而实现检索项目的快速定位。

2 记录主序: RKC 代码值序。

3 记录长度: 11 个字节。

表 8 培养基成分表(表名 MCBTMEDI)

项号	项目名称	是否为检索项	字段长度
1	培养基编号	✓	FIX(4,0)
2	配方内容		CHAR(450)
3	保密级别		CHAR(1)

1. 设置目的: 培养基配方多为文字型描述, 并且重复使用次数较多, 故建立该表, 当输出培养基内容时, 只打印培养基编号即可。

2. 记录主序: 培养基编号。

3 记录长度: 455 个字节。

表 9 刊物表(表名 MCBTJOUR)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌号	✓	9
2	刊物名称		70
3	卷页年		30

1. 设置目的: 记录与该菌有关的主要参考文献。

2 记录主序: 菌号序。

3 记录长度: 109 个字节。

表 10 单位缩写表(表名 MCBTINST)

项号	项目名称	是否为检索项	字段长度(字节)
1	保藏单位缩写	✓	20
2	单位英文全称		150
3	单位英文地址		100

1. 设置目的: 建立保藏单位名称缩写与全

称的对应关系,并且记录各保藏机构的地址,以方便用户与之联系。

2. 记录主序: 保藏单位缩写序。

3. 记录长度: 270 个字节。

表 11 菌株历史表(表名 MCBTHIST)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌号	√	9
2	菌株历史		150

1. 设置目的: 记录菌株保藏变迁历史。由于菌株可能经过若干个保藏单位的相互交换,用本表记录这种变迁过程。

2. 记录主序: 菌号序。

3. 记录长度: 159 个。

表 12 标准菌株说明表(表名 MCBTSSTI)

项号	项目名称	是否为检索项	字段长度(字节)
1	菌号	√	9
2	标准菌株说明		100

1. 设置目的: 记录该菌在什么范围内是标准菌株及有关说明。

2. 记录主序: 菌号序。

3. 记录长度: 109 个字节。

(二) 性状库的检索功能

基于以上数据结构,我们制定了 18 个检索功能,并按照使用频度将其分为主检索功能 8 个、辅助检索功能 10 个。

主检索功能 1: 从菌株性状检索 RKC 编码。

用户首先通过这一功能检索出与自然语言描述的菌株性状相对应的 RKC 编码。系统提供了顺序查找、关键字查找两种检索方式。

主检索功能 2: 从 RKC 编码检索菌号。

用户可以通过 RKC 编码检索到具有某些特定性状的菌株。在此,系统提供了六种检索方式,并且包括了与、或、非运算及各种逻辑关系复合组配。

主检索功能 3: 从菌名检索菌号。

用户可以检索到属于某一特定属或种的所有菌号。系统对菌名以及生态环境、栖息地、用途等数据项分别设立了字典功能。当用户不确

切地知道某些菌株的菌名、生态环境等专有名词的具体拼法时,只需打入其单词的前面一部分,就可迅速地找到其正确全称,从而极大地方便了用户检索。

主检索功能 4: 从用途检索菌号。

用户可以从用途类别、用途细则、开发单位、开发者、开发现状出发检索到其相应的菌株。

主检索功能 5: 从综合管理信息检索菌号。

用户可以用任何可检索项进行多种与、或、非等各种逻辑关系的复合组配,从而检索出符合条件的菌株。

主检索功能 6: 从菌号检索综合管理信息。

用户可以借助这一功能调出某一特定菌号的综合管理信息。

主检索功能 7: 从菌号检索用途信息。

用户可以借助这一功能调出某一特定菌号的用途信息。

主检索功能 8: 从菌号检索性状。

用户可以借助这一功能调出某一特定菌号的性状信息,并且可以自动生成数值分类用的数据工作文件。

10 个辅助检索功能分别为:

1. 从别号查菌号, 2. 从菌号查别号, 3. 从菌号查菌名, 4. 从菌名查别号, 5. 从别号查菌名, 6. 从 RKC 码查其说明, 7. 从保藏单位缩写查全称, 8. 从培养基编号查其成分, 9. 从菌号查其历史, 10. 从菌号查刊物。

(三) 软硬件环境

软件: 主要使用了三种软件语言。

操作系统: VM/CMS。

软件语言:

SQL 语言: 用于建库、库操作以及各种与综合管理数据有关的表管理。

PL/I 语言: 用于各类打印输出以及各种与性状检索有关的程序设计。

ISPF 语言: 用于屏幕显示格式的定义、显示操作以及各种屏幕输入输出管理。

文件类型: SQL/DS TABLE, ISPF
TABLE, CMS FILE

硬件:

主机: IBM 4341-M11

内存: 8MB

终端: 3180

磁盘: 3375

磁带: 3430

数据库像图书馆一样是一项社会公益事业, 它的生存需要大家的关心与帮助。本性状库目前已投入试运行。我们竭诚希望广大微生物

学工作者, 尤其是从事细菌分类、生态及资源开发工作者向本库提出检索查询要求, 并对本库的数据结构和检索功能提出宝贵意见。另外, 如您研究的菌株仍完整地保存, 欢迎您把有关数据向本库提供, 我们愿尽力为您服务。同时本库将按照数据提供者的意愿在所要求的级别上采取保密措施, 保证原始数据不外泄。

参 考 文 献

1. 菅原秀明: *Bio. Industry*, 5(8):41-51, 1988.
2. 菅原秀明, 馆野义男: *信息管理*, 26(8): 640-652, 1983.

电子计算机在厌氧菌鉴定中的应用

何道生

(军事医科学院微生物流行病学研究所)

余满松 毛远平 王亚平

(中国人民解放军 304 医院, 北京)

摘要 本文介绍了鉴定厌氧菌的微量系统配以电子计算机诊断, 提高了鉴定效果。经 356 株参考菌株和临床分离菌株的检测, 87.6% 的菌株可鉴定到种, 97.4% 菌株鉴定到属。种水平上诊断错误的占 12.6%, 属水平上诊断错误的仅占 2.6%。没有菌株不能诊断。与国外已成商品的鉴定系统相比, 无论是鉴定到种还是属, 其水平都较高。

关键词 厌氧菌鉴定; 微量系统; 计算机诊断

无芽孢厌氧菌的鉴定除根据菌落形态、细胞形态、革兰氏染色反应、药物耐受性以及代谢产物的气液相色谱分析和 DNA G+C 含量测定以外, 生化反应仍然是应用最普遍, 技术最简单, 一般实验室都能接受的方法。常规的生化试验方法已由 CDC (疾病控制中心)、VPI (弗吉尼亚多科技术研究所) 和 Wadsworth 厌氧菌实验室建立并发展。然而这些系统既费时, 又费钱, 超出许多临床实验室的承受能力^[1]。近十年来出现了许多鉴定厌氧菌的微量系统 *microsystem*, 并已制成商品。这些微量系统大致分为两类: 依赖于细菌生长的 (需时较长 48—72 小时) 和不依赖于细菌生长的 (有氧条件下孵育 4—6 小时)。前者包括 *Anaerobe-Tek (A/T)* 系统^[2]、*API-20A* 系统^[3]、*Minitek* 系统^[4] 和 *Anaerobe Combo Panel* 等^[5]; 后者包括 *RapID-ANA* 系统^[6]、*An-Ident* 系统^[7]。

我们的微量系统是属于前者配以电子计算机诊断。本文报道了用这种方法对 356 株临床分离菌的鉴定结果。

材 料 和 方 法

(一) 材料

1. 供试菌株及来源: 标准菌株 NCTC 9343 等 7 株, 临床分离菌 349 株 (口腔分离菌 212 株、阑尾分离菌 28 株、食管癌分离菌 17 株、眼疾患分离菌 12 株、实验动物分离菌 19 株, 其他来源的菌株 61 株) 共 356 株。其中含厌氧球菌 69 株, G⁺ 无芽孢厌氧杆菌 88 株, 胆汁耐受拟杆菌 62 株, 胆汁敏感无色素拟杆菌 85 株, 产色素拟杆菌 20 株, 梭杆菌 32 株。

2. 微量生化管: 本试验所用的微量生化管包括 靛基质试验、硝酸盐还原试验、七叶苷水解试验、尿素酶试验、淀粉水解试验、20% 胆汁耐