

# 一个植物分类数据库<sup>\*</sup>

汪 庆 凌萍萍

(江苏省中国科学院植物研究所, 南京 210014)

**摘 要** 本文介绍了一个苔藓植物分类数据库的设计及其结构, 是使用 Visual Basic 4.0, 以中国苔藓植物的标本, 学名, 分布, 文献, 分类性状等为基本资料, 以打印标签、名录、文献索引, 构成分类性状的原始数据矩阵, 并与数量分类程序, 以及 DELTA, PAUP 和 CLUSTAN 等应用程序的连结为目标进行开发的。其基本构想是在这样一个框架之下, 逐渐积累中国苔藓植物的基本资料, 能够完成大部分的标本室和分类学工作, 并在使用过程中进一步扩展和完善其结构和功能, 从而为中国苔藓植物的研究提供极大的便利。

**关键词** 植物分类数据库; 苔藓植物

## A taxonomic database for Chinese Bryophyta

Wang Qing Ling Pingping

(Institute of Botany, Jiangsu Province & Academia Sinica, Nanjing 210014)

**Abstract** A plant taxonomy database for Chinese Bryophyta was introduced for the design and construction, databases are included for the nomenclature, bibliography, curatorial, phyto geography and the character descriptions, DELTA format was used in construction of the taxonomic morphological description of Bryophyta, it is try to occupy a range of taxonomic activities in Bryophyta research, containing the outputs for the label, checklist, bibliography, character description texts through filter and original data matrix which can be used by other applied programs, such as DELTA systems, PAUP or CLUSTAN *et al.* It is an experimental taxonomic database, which can be justified only when the information stored is to be use frequently by many people. We created the database by using the Visual Basic 4.0.

**Key words** Plant taxonomic database; Bryophyta

植物分类学作为植物学科中最基础和最综合性的一门分支学科, 包含了大量的信息资料, 借助于数据库技术无疑能更有效的管理, 分析, 评价和利用这些信息资料, 因此植物分类学领域内出现了各种数据库, 包括单一的资料性数据库, 标本管理数据库, 特征描述数据库, 专著研究数据库, 以及大型的植物志数据库如 FV, FNA, ESFEDS 等。植物分类数据库朝着包含植物信息愈来愈全面, 所能执

1997-06-06 收稿

第一作者简介: 汪 庆, 男, 1965 年出生, 硕士, 主要从事华东地区苔藓植物的研究。

\*江苏省自然科学基金资助项目的部分工作

行的分类学任务愈来愈多的方向发展，即一个包含大量分类学信息的数据库系统和与之相连的可以执行广泛的分类学任务的应用程序组。这样一个植物分类数据库中，信息资源可以反复使用，为多种分类学目的服务，为多个用户所共享，同时让用户更方便地使用这些信息资源。显然建立一个内容广泛的植物分类数据库是很有意义的。

苔藓植物是除被子植物之外最大的绿色植物群，但由于体形小，经济上的重要性不大而很少受到支持，由于财力和人员的缺乏，苔藓植物的研究大大的滞后于其他分类群，因此更有必要在苔藓植物研究中采用植物分类数据库以期加快其研究进程。

我们使用 Visual Basic 4.0，在中国苔藓植物中，以标本，分布，学名（模式，异名等），文献，分类学性状（形态学，解剖学，细胞学，和化学等）等分类学资料构成基本数据库；通过筛选能够有选择的打印标签，名录，参考文献；通过累积程序构成分类性状的原始数据矩阵，进而与有关数量分类程序连接；通过格式转换程序与 DELTA，PAUP 和 CLUSTAN 等应用程序连结；从而实行多任务的目标。

### 1 中国苔藓植物分类信息

**1.1 标本库 specim** 标本号，国家，省，县，地点，生境，习性，采集人，采集日期，植物名称，鉴定人，鉴定日期，收藏标本馆，借出（入），借阅日期，附记。该库目前包括江苏苔藓标本 3000 号，和中国苔藓志 1—2 卷引用标本 2000 号。

**1.2 文献库 bibilia** 文献号，作者，第二作者，发表日期，题目，刊号或书名，卷数，页码，关键词，关键词 2。包括中国苔藓植物研究文献和科属研究文献 1000 余份。

**1.3 分类群库 taxon** 分类群代码，科名，属名，种加词，种下等级，异名，定名人，原始描述，参考文献，中国分布（中国各省分布和参考文献）。该库主要依据《Moss Flora of China》建立的。

**1.4 分类性状** 包括 3 个库，一标本特征库（charact），用于存放每一经过研究的标本的分类特征记录；另两个是各级分类单元的分类特征库。用以构成特征的原始数据矩阵，是与 DELTA system 相对应的，描述库（descrip）对应于 CHARS，描述库-x（descrip-x）对应于 ITEMS，目前描述库包含了 118 个分类特征，可根据所研究的具体对象对其进行修改，而后描述-x 库自动修改；和描述-x 库的每个域的输出为下拉选择项。还可以通过累积程序来实现从标本特征库到描述-x 库。

**1.5 标本室 (herbar)** IAPT 代号，标本室，收藏标本数目，联系地址，负责人，附记。收入中国苔藓标本收藏的 25 个中国标本馆和世界上收藏中国苔藓标本的 10 个标本馆。

### 2 分类性状的 DELTA 格式

分类群分类性状的记录是整个数据库的关键，通过该库的建立才能够实现多任务的目标，但

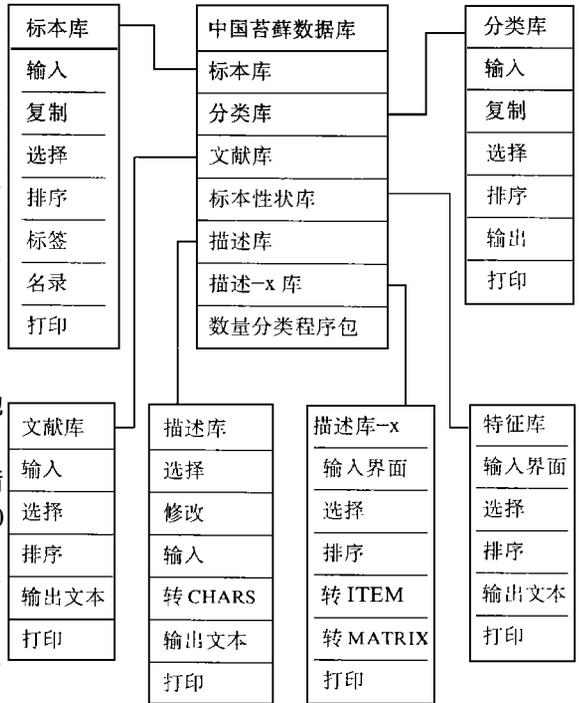


图 1 数据库界面示意

由于分类性状数目多, 类型各异, 分类学的趋势更是利用的分类学资料愈多愈好, 且在不同的分类学任务中所需要的性状是不同的, 同一性状在不同的分类群中其分类学价值也不同, 并存在不同的习惯表达形式, 如何尽可能在整个数据库中保证分类学特征整体性和均一性是非常重要的。

#1 生境/	#31 叶缘〈锯齿〉/	#61 雄苞〈配丝〉/	#91 齿片〈形状〉/
#2 习性/	#32 叶缘〈分化〉/	#62 雌苞〈位置〉/	#92 齿片〈颜色〉/
#3 颜色/	#33 分化边缘/	#63 雌苞〈配丝〉/	#93 齿片〈具疣〉/
#4 光泽/	#34 叶基部〈形状〉/	#64 鳞叶〈外雌苞叶〉/	#94 齿片〈外面〉/〈横脊〉/
#5 原丝体/	#35 叶面皱摺/	#65 内雌苞叶〈长度〉/	#95 齿片〈内面〉/〈横隔〉/
#6 主茎〈排列〉/	#36 中肋〈形态〉/	#66 内雌苞叶〈形状〉/	#96 内齿层〈颜色〉/
#7 茎枝〈形状〉/	#37 中肋〈颜色〉/	#67 内雌苞叶基部/	#97 基膜〈与齿片〉/
#8 茎〈长度〉/ cm/	#38 中肋长度/ mm/	#68 内雌苞叶上部/	#98 基膜〈具疣〉/
#9 支茎〈排列〉/	#39 中肋宽度/ mm/	#69 内雌苞叶边缘/	#99 齿条与齿片/
#10 假根〈类型〉/	#40 中肋〈细胞层数〉/	#70 内雌苞叶细胞/	#100 齿条〈形状〉/
#11 假根〈程度〉/	#41 中肋背面/	#71 内雌苞叶中肋/	#101 齿条〈隆起〉/
#12 假根〈位置〉/	#42 中肋腹面/	#72 蒴柄〈长度〉/〈cm〉/	#102 齿条〈宽度〉/ mm/
#13 中轴〈分化〉/	#43 栉片〈位置〉/	#73 蒴柄〈具疣〉/	#103 齿条〈穿孔〉/
#14 茎横切面〈形状〉/	#44 栉片细胞/	#74 蒴柄〈状态〉/	#104 齿毛/
#15 分枝/	#45 栉片顶部细胞〈具疣〉/	#75 蒴柄〈颜色〉/	#105 蒴盖〈形状〉/
#16 分枝〈排列〉/	#46 栉片顶部细胞〈形状〉/	#76 孢蒴〈状态〉/	#106 蒴盖〈具喙〉/
#17 鳞毛〈程度〉/	#47 叶上部细胞〈形状〉/	#77 孢蒴〈长度〉/ mm/	#107 蒴盖/
#18 鳞毛〈形状〉/	#48 叶中部细胞〈长度〉/	#78 孢蒴〈宽度〉/ mm/	#108 蒴帽/〈形状〉/
#19 茎叶〈与枝叶〉/	#49 叶中部细胞〈宽度〉/	#79 孢蒴〈位置〉/	#109 蒴帽基部/
#20 叶〈长度〉/ mm/	#50 叶中部细胞〈形状〉/	#80 孢蒴〈形状〉/	#110 蒴帽纵摺/
#21 叶片〈宽度〉/ mm/	#51 叶缘细胞〈形状〉/	#81 孢蒴口部〈缢缩〉/	#111 蒴帽〈覆盖孢蒴〉/
#22 叶片〈细胞层数〉/	#52 叶基部细胞〈形状〉/	#82 孢蒴外壁/	#112 蒴帽〈开裂〉/
#23 叶对称性/	#53 叶细胞〈层数〉/	#83 孢蒴外壁〈气孔〉/	#113 孢子〈形状〉/
#24 叶列/	#54 叶细胞〈排列〉/	#84 环带〈分化〉/	#114 孢子〈颜色〉/
#25 叶/〈干〉/〈排列〉/	#55 叶细胞〈具疣〉/	#85 环带〈宿存〉/	#115 孢子〈形态〉/
#26 叶/〈湿〉/〈排列〉/	#56 叶基近中肋细胞/	#86 蒴齿〈层数〉/	#116 孢子〈直径〉/ μm/
#27 枝叶/〈干〉/〈排列〉/	#57 叶角细胞〈分化〉/	#87 蒴齿〈数目〉/	#117 染色体〈数目〉/
#28 叶〈形状〉/	#58 叶角细胞〈形状〉/	#88 蒴齿〈位置〉/	#118 分布区类型/
#29 叶尖〈形状〉/	#59 两性分化/	#89 齿片〈排列〉/	
#30 叶缘〈形状〉/	#60 雄苞/	#90 齿片〈分裂〉/	

DELTA (DEscriptive Language for TAXonomy) 是一个对所有植物特征进行编码的系统。由 Dallwitz (1980) 发表, 被 International Working Group on Taxonomic Databases for plant sciences (TDWG) 作为植物特征编码的标准格式。该系统对分类单位无特殊要求, 对于特征类型也无特殊要求, 所能够进行的分类学工作比较全面, 并能转换为其他分类应用程序如 PAUP 所要求的数据格式。使用 DELTA 系统需要建立 3 个文件 CHARS、ITEMS 和 SPECS。描述库和描述库一 x

分别对应于 CHARS 和 ITEMS, SPECS 则根据具体的研究对象编写。

2.1 苔藓植物分类特征 (CHARS) 描述库 (descrip) 对应于 CHARS, 选择了 118 个基本特征作为特征输入的起点, 在以后使用过程中增删。可根据所研究的具体对象对描述库进行修改, 之后描述-x 库自动进行修改。

2.2 分类群分类特征的描述 (ITEMS) 描述库-x (descrip-x) 对应于 ITEMS; 为其另行设计了一个输入界面, 每个域的输入为下拉选择项, 大大简化了输入的工作量。描述库-x 的结构, 即域的多少及其定义是由描述库决定的。一个植物特征表达为 c, v, c 为特征的序号, 即域的名; v 为该特征状态的取值, 即域值。

例如: 钟帽藓 *Venturiella sinensis* (Vent) C. Muell. 植物体深绿色, 纤细而交织的树生藓类。茎匍匐, 不规则密羽状分枝, 随处有稀疏假根; 分枝直立等长, 叶密集, 叶干时覆瓦状排列, 湿时倾立, 阔卵形或长卵形, 内凹, 有毛状短尖, 尖部有时无色透明; 叶边平直, 全缘, 仅近尖端有细齿; 无中肋; 叶细胞卵形或六边形, 无疣, 叶边细胞四方形, 叶角细胞由多列方形或扁方形构成。雌雄异株。雌苞生于分枝顶端。雌苞叶较大, 直立, 具长毛尖。蒴柄短, 直立, 黄色。孢蒴半隐于苞叶内, 成熟后始伸出, 卵长形。环带永存。蒴齿单层, 靠近蒴口着生; 齿片成对排列, 披针形, 棕红色, 有密疣及横隔。蒴盖基部宽, 有纵摺, 基部分瓣, 孢子有密疣。

#345 11, 1 12, 1 13, 2 14, 2 16, 3 23, 6/8 20, 1 21, 2 25, 1 32, 1 36, 1 37, 2 38, 1/2 39, 1 40, 1 41, 1 59, 1 60, 1 61, 1 62, 1 65, 2 70, 1 71, 2 73, 1 74, 6 79, 1 80, 1 82, 1 83, 1 85, 1 86, 1 87, 2 90, 1 101, 3 103, 1 104, 2 106, 4 110, 3

### 3 分类过程与累积程序

分类性状一是源于现成的植物描述, 例如原始描述, 专著, 订正, 植物志等, 二是标本。前者是前人研究的结论, 使用比较方便, 但特征描述不全面, 格式不统一, 也不能反映该分类群的

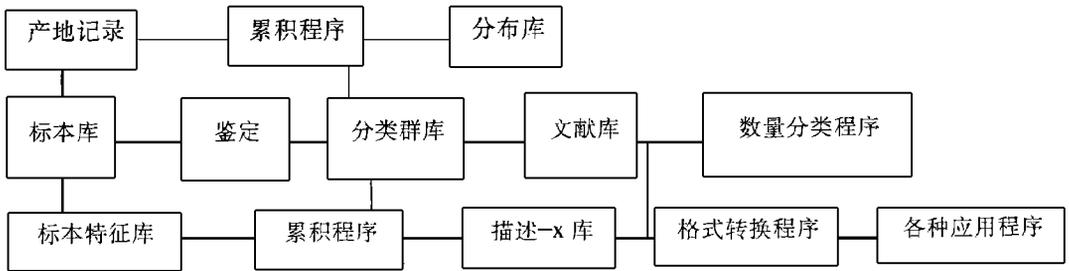


图2 数据库的流程

变异情况; 植物分类学研究的过程, 也就是从标本获取分类信息, 然后进行分析评价的过程, 如果将两者结合起来, 并能够在两者之间进行选择则更为方便和全面。据此建立一个从标本上获取分类信息直至分类群的 DELTA 的 ITEMS 的累积程序, 在进行分类学研究时使用。

3.1 标本特征录入 标本特征输入界面被用来将标本的分类特征存放在标本记录库中, 该库中的一个记录为一个标本的分类特征数据。第一个域是标本代码, 紧接着是每一个特征的记录。不允许使用广幅的数量特征或关联的或交替的定性特征, 但可以用文件中的几个域来表达。

3.2 特征累积 分烟特征累积程序以重复的方式, 依据标本鉴定所确定的归属, 自动产生每一

分类群的特征，这一过程参考 Mated 等 (1992) 的算法。对于一个给定的分类群算法的外循环是标本循环，特征循环相互嵌套使程序从标本 1 特征 1 开始，然后进行标本 1 特征 2，直到最后一个标本的最后一个特征。当一个分类群的累积过程完成，计算每一定量特征的平均数。每一标本的每一特征状态的次数（包括特征 0=缺失特征）给予每一分类群，缺失数据的标本数目也予以记录，和数量特征的最小和最大值和所有记录值的平均。在对这个临时数据文件进行必要的修改之后进行下一步。

**3.3 格式转换** 格式转换程序以与累积程序相同的方式工作。该程序从每一分类群的分类特征库中获取所需要的数据，以特征的序号排，然后拷贝每一分类群的分类特征为一个 DELTA items file，包括了每一分类群在累积过程中积累的数据，可直接使用于 DELTA 相关程序。

**3.4 地理分布资料的累积** 植物分布资料的累积过程同分类性状的累积一样，程序通过标本鉴定结果来确定标本的归属，并计算每一分类群在每一地理单元的标本数目。在这里每一标本的产地作为定性特征，与同一分类群的其他标本进行累积而获得分类群的分布。该程序可以从小到大的地理单元上进行累积，同样也可用于不同分类等级的分类单元，如从种到组、系、属、科等。

## 4 讨 论

Visual Basic 是采用可视化工具、面向对象、事件驱动编程的高级程序设计语言。VB 提供的可视化设计工具，使得图形用户界面编程变得异常的简单，应用了面向对象的编程思想，更多的考虑对象之间的逻辑关系，没有明显的主程序，只是由一些图形对象和对象的事件程序组成。执行时，也是操作这些对象的事件而实现这些功能。操作具有随机性。这些特点为设计一个较开发的数据库系统提供了很多方便。

面对日益浩繁的标本和文献，植物分类数据库无疑是一个很好的选择。以上是一个苔藓植物分类数据库的基本结构，其基本构想是在标本室的日常事物性工作中，出就是在标本积累过程中，逐步积累和完善每一分类群的分类特征和分布资料等；通过累积程序进而形成原始数据矩阵，直接用为数量分类程序的数据文件；还可以通过格式转换程序，用作如 DELTA 系统的 items file；从而达成多任务的目标，如特征描述、各种诊断检索表、匹配鉴定、表型分析的 distance 矩阵，上线鉴定的数据组和可以转换进入其他程序的数据组如 PAUP or CLUSTAN。

## 参 考 文 献

- 1 Allkin R, Bisby F A. The structure of monographic database. *Taxon*, 1988, **37**: 756~763
- 2 Allkin R, Moreno N P, Gama Campillo L, Mejia T. Multiple uses for computer-stored taxonomic description: keys for Veracruz. *Taxon*, 1992, **41**: 413~435
- 3 Bisby F A. Automated taxonomic information systems. 1984 Pp. 301~322 in: Heywood, V. H. & Moore, D. M. (ed.). Current concepts in plant taxonomy. London & Orlando
- 4 Cutbill J L. New methods for handling biological information. *Bot J. Linn. Soc*, 1971, **3**: 253~260
- 5 Dallwitz M J. A general system for coding taxonomic descriptions. *Taxon*, 1980, **29**: 41~46
- 6 Maxted N. Towards defining a taxonomic revision methodology. *Taxon*, 1992, **41**: 653~660
- 7 Maxted N, White R J, Allkin R. The automatic synthesis of descriptive data using the taxonomic hierarchy. *Taxon*, 1993, **42**: 51~62
- 8 Pankhurst R J. A method for data capture. *Taxon*, 1972, **21**: 549~558
- 9 Pankhurst R J. Database design for monographs and floras. *Taxon*, 1988, 733~746