Utility and features of a customized PC Windows-based relational database for managing microbial strain data

L Sigler and AL Flis

University of Alberta Microfungus Collection and Herbarium, Devonian Botanic Garden, Edmonton, Alberta, Canada

In addition to storing microbes, culture collections in industry, government or universities manage a vast and continuously expanding library of information on strain history and properties. An efficient and cost-effective computer database system is required for entering, analyzing and searching these data. This report describes the utility and features of a comprehensive database which consists of a commercially available relational database system combined with customized screens for data entry, viewing and report generation. The application was developed using Microsoft Access and Visual Basic to operate in the Windows environment on a local area network.

Keywords: relational database; microbial strain data

Storing and retrieving microbial strain data

Microbial collections conserve microorganisms of scientific, industrial or heritage value and distribute them to users in industry, universities and government. In addition to storing microbes, collections accumulate large amounts of information on taxonomy, strain history, properties, metabolites or applications. An efficient and cost-effective mechanism for archiving and retrieving these data is required by collection managers and by users who want to rapidly access critical information about collection holdings. While computer databases have been in use for handling these types of data for over 20 years, no single system meets all requirements. Collection databases have largely developed independently and only in recent years has there been any concerted effort to harmonize systems within coordinated networks [3–5,8–11]. By the time that integrated or stand-alone database systems are fully functional, they may be rendered obsolete by changes in information technology. Parameters affecting the design and development of collection databases include the rapid evolution of both computer hardware and software, a general shift from large mainframes to stand-alone or networked personal computers, a change from hierarchical to relational databases, and major changes in telecommunications linkages including rapid growth of internet use.

Our experience illustrates the impact of such changes. In 1984, the first publicly accessible interactive Canadian microbial database was established on a mainframe to maintain accession records for the University of Alberta Microfungus Collection and Herbarium (UAMH) [1,6,7]. In 1993 the existing database was made redundant by a University decision to abandon the mainframe in favor of a distributed computing environment. In developing a new collection database management application, several needs were considered. These were to select an alternate platform which allowed data transfer with relative ease, to reduce dependence on outside technical programming support as much as possible, and to design efficient mechanisms for data entry, viewing and report writing. After review of various options, a customized information system was developed by adapting and modifying Microsoft applications (Microsoft Corporation, Redmond, WA, USA). Because of commercial availability, this database application offers flexibility and utility to any researcher managing microbial strain data. The customized screens described here demonstrate one approach for data entry, reviewing and reporting of fungal data which are divided into modular units (ie related tables).

Database design

Different constraints occur in the modification of an existing database than are present in the design of a new one. Operating under the Stanford Public Information Retrieval System (SPIRES, Stanford University, Stanford, CA, USA), the original UAMH database consisted of two linked subfiles: SPECIES contained taxonomic information such as species name, authority, year of publication, synonyms and related names, and taxonomic comments, while STRAINS included information on strain history (isolation and source), key characteristics or general properties [6,7]. Since the data within SPIRES were already partially in modular units, a relational database management system (RDMS) appeared to be suitable for transfer of the data. In a RDMS, facts about different subjects are stored in different tables, and can be manipulated by defining the relationships between the tables. The new database was developed in Access, a RDMS designed for Windows and the rest of the system was designed using Office and Visual Basic [2]. Because Office is a group of applications (including Word, Power Point and Excel) that look alike and work in the same way, it offers considerable flexibility in report generation since documents can be created from more than one package.

Data were first organized into manageable related units using the data organization concepts in the original data-

base. Thirteen Access tables, containing a total of 60 fields, were established (Figure 1). The SPECIES and STRAINS subfiles became separate tables, as did each multiply-occurring field such as species SYNONYM (ie other names for the same organism), strain CHARACTER (ie property or application) or COMPOUND (ie metabolite). The powerful ability of a RDMS to quickly access the data stored in the separate tables is accomplished by defining relationships between the tables. Each table must contain a common key field (shown in lower case letters within each box in Figure 1) that uniquely identifies the record and allows the data to be retrieved and updated based on data stored in the other table.

Additionally, customized screens were developed to protect the integrity of the data stored in the tables, to simplify data entry and to facilitate viewing individual records. In a RDMS, information about an individual record may be stored in several tables. Because of the complexities in reviewing these data, the viewing screens were designed to pull together information from different tables in an efficient manner. The screens were designed in Visual Basic, a programming tool compatible with Office. Screens for data entry, review or search/report functions are shown in Figures 2 and 3.

All organisms are assigned a species name before they are given a UAMH accession number. In instances where identification is incomplete, provisional names and numbers can be used. To enter a new strain, the first few letters of the name are typed into the search bar on the Search screen (Figure 2b), the correct species name is selected from the menu, and the Species screen is then displayed (Figure 2c). Nomenclatural information (ie authority, publication reference and year) is divided into separate fields for easy access when generating reports. Other names (ie synonyms or morphs) for the species are displayed but to access or update the complete information on them, another screen must be opened (Other Names screen, not shown). The screen also displays the accession numbers of all strains belonging to this species (Figure 2c).

Reporting features are accessed from the Main Menu screen (Figure 2a). Strain records may be exported to report templates by stacking accession numbers or by searching. To automate a series of tasks for retrieving information, a query routine (Partial SQL) based on Structured Query Language was developed. Essentially all data can be queried. Figure 2d illustrates the results of a search for all strains in which cyclosporin is listed as a compound. The queued records can be exported to be used as the data source for Word mail merge function. In Word, a report template is created in which codes for the selected field names are inserted. When the exported data are merged with the report template, Word replaces the codes with the contents of each field allowing for the design and generation of various reports including documents for shipping of cultures or more complex reports such as comprehensive catalogues.

New strains are entered or information on accessioned strains is reviewed using similar screens (Figure 3). The New screen automatically assigns the next available accession number and displays the field names along with blank boxes to allow for data entry. Figure 3a illustrates how the strain data have been divided into the most useful fields for use in searching or collection management. Fields pertain to the history of the strain, including substrate or habitat, location, isolator or collector, sender, status as ex-type, confidential deposit, etc, or track preservation methods and location of stored strains. Additional information (if available) such as accession numbers assigned to the strain by other collections or herbaria, strain properties, applications or metabolites, can be viewed in related data screens (Figure 3b,c). These screens allow the entering, editing and viewing of information from several related
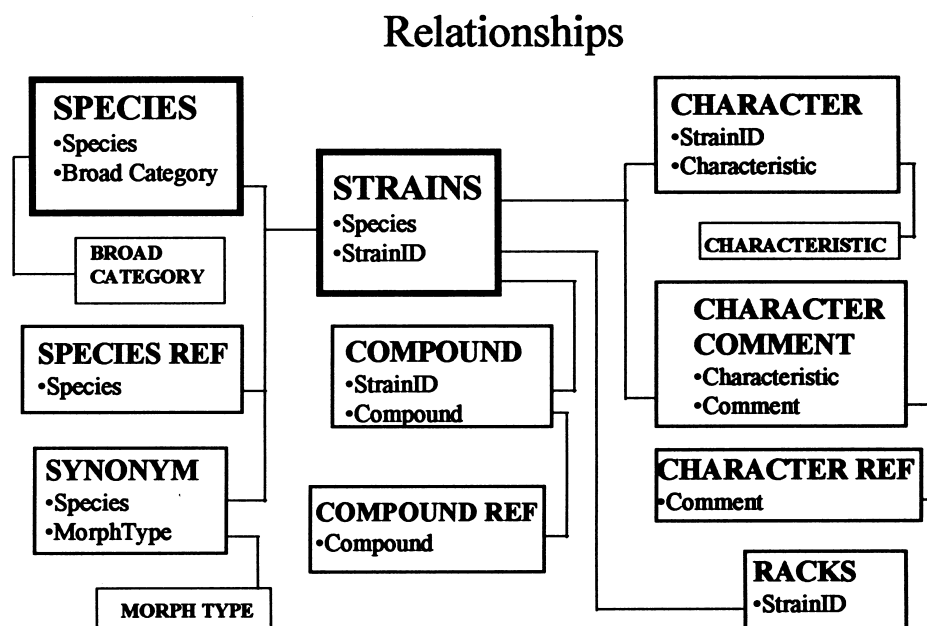
Relationships

Figure 1 Overview of database design showing organization of data into 13 related Access tables. Individual tables are shown by boxes with names in upper case. Relationships between tables are shown by lines. Related tables contain a common key field (names in lower case).
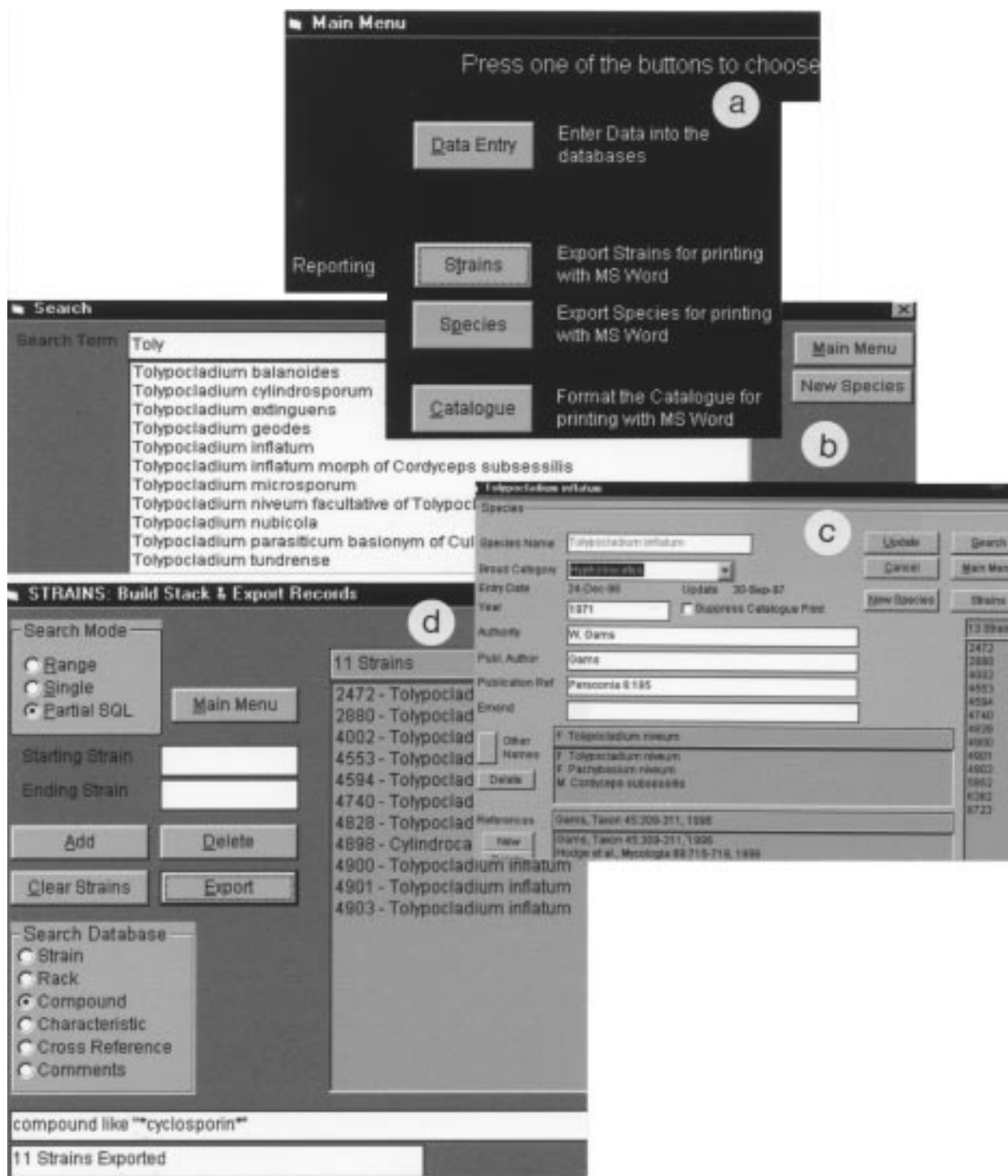
Figure 2 (a) The Main Menu screen offers data entry, searching or reporting functions. (b) The search screen allows for searching by species or strain or for entry of a new species record if New is selected. Selection of a species name from the drop down menu (eg within the genus *Tolypocladium*) brings up the Species screen. (c) The reviewing window for the Species screen displays species nomenclatural and taxonomic information, and the strains held for a species (eg *Tolypocladium inflatum*). Information on strains can be viewed by selecting Strains. (d) The Stack and Export screen allows for export to report templates by stacking accession numbers or by searching using a query routine (Partial SQL). In this example, a search for the compound cyclosporin yielded 11 strains. These data can be exported for use in a Word mail merge function to produce customized reports.

tables at one time, rather than having to move from table to table.

By necessity, databases undergo continuous evolution and development. The need for change has great impact on

staff and financial resources of smaller, specialized culture collections. Many factors impact the selection of an appropriate database platform. Our model may have utility to others seeking options for management of collection data
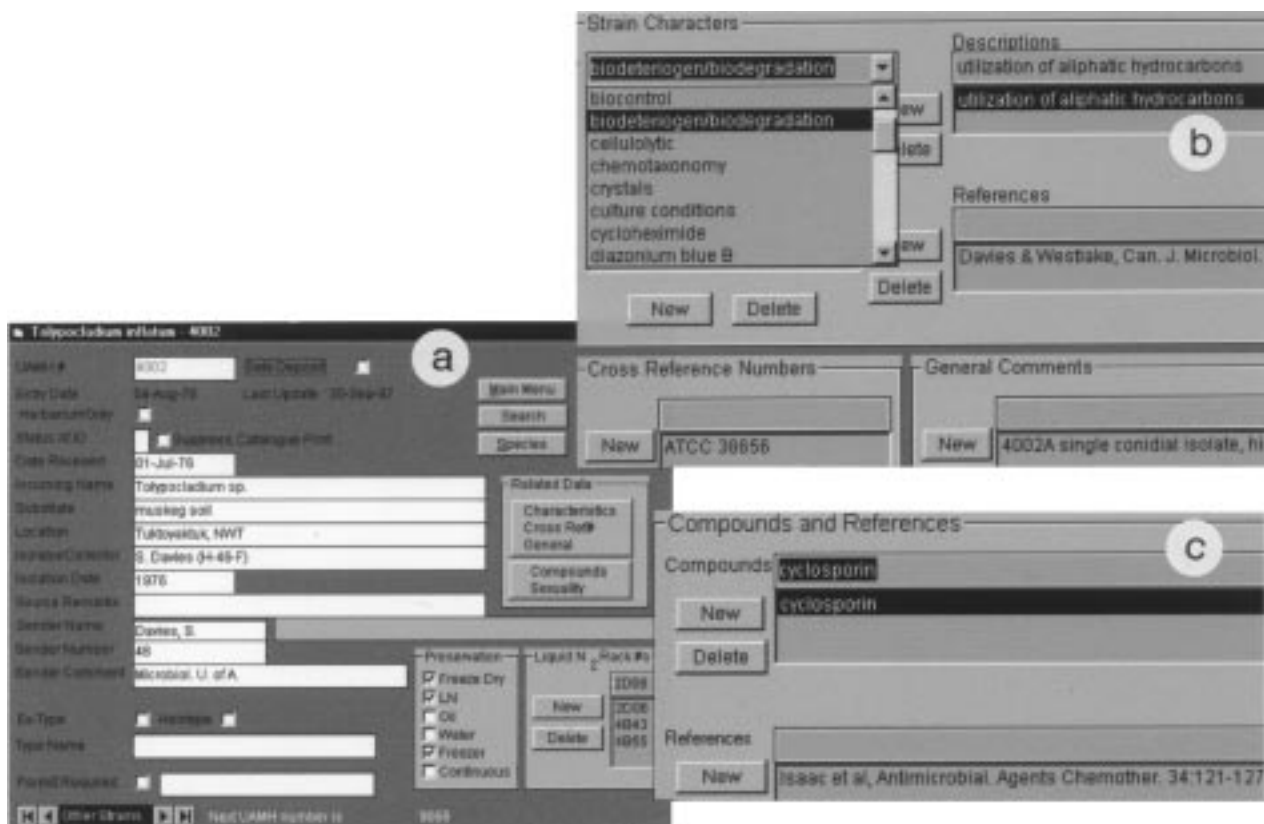
Figure 3 Composite screens allow for entering, editing and viewing of strain data from several related tables. (a) The main screen contains the fields common to most strains. Records of other strains belonging to the species may be browsed by using the Other Strains button. (b) Strain Characters can be selected from a drop down box (eg degradative capacity as shown here) and then modified with descriptions and references if available. (c) Related information including compounds produced by a fungal isolate are entered or reviewed on this related data screen.

since it shows how commercially available products may be modified to produce a customized application.

Acknowledgements

References

1 Currah RS and L Sigler. 1986. Curating and displaying collections of microfungi. In: Life Sciences Miscellaneous Publications (Waddington J and DM Rudkin, eds), pp 85–88, Royal Ontario Museum, Toronto, ON.

2 Flis A and L Sigler. 1996. Development and functionality of a PC Windows-based database for a fungal collection. In: Culture Collections to Improve the Quality of Life. Proceedings of the International Congress of Culture Collections (Samson RA, JA Stalpers, D van der Mei and AH Stouthamer, eds), pp 264–266, Centraalbureau voor Schimmelcultures, Baarn, NE.

3 Gams W, JA Stalpers, GJ Stegehuis and J Smith. 1990. Computerization of strain data in the Microbial Information Network Europe (MINE). Sydowia 42: 218–230.

4 Manfio GP and S de Souza. 1996. Brazilian catalogue on internet and other services via WWW. In: Culture Collections to Improve the Quality of Life. Proceedings of the International Congress of Culture Collections (Samson RA, JA Stalpers, D van der Mei, and AH Stouthamer, eds), pp 247–250, Centraalbureau voor Schimmelcultures, Baarn, NE.

5 Pullan M. 1996. The design and implementation of culture collection databases. In: Culture Collections to Improve the Quality of Life. Proceedings of the International Congress of Culture Collections (Samson RA, JA Stalpers, D van der Mei and AH Stouthamer, eds), pp 251–259, Centraalbureau voor Schimmelcultures, Baarn, NE.

6 Sigler L. 1986. Catalogue of the University of Alberta Microfungus Collection and Herbarium, 166 pp (mimeogr+supplement). Devonian Botanic Garden, Edmonton, Canada.

7 Sigler L. 1994. Maintaining fungal diversity—integration of a herbarium and living collection. J Ind Microbiol 13: 191–192.

8 Stegehuis G. 1996. European biological resources on the internet. In: Culture Collections to Improve the Quality of Life. Proceedings of the International Congress of Culture Collections (Samson RA, JA Stalpers, D van der Mei and AH Stouthamer, eds), pp 266–267, Centraalbureau voor Schimmelcultures, Baarn, NE.

9 Sugawara H. 1996. WDCM and worldwide information. In: Culture Collections to Improve the Quality of Life. Proceedings of the International Congress of Culture Collections (Samson RA, JA Stalpers, D van der Mei and AH Stouthamer, eds), pp 155–157, Centraalbureau voor Schimmelcultures, Baarn, NE.

10 Sugawara H, S Miyazaki, J Shimura and Y Ichiyanagi. 1996. Bioinformatics tools for the study of microbial diversity. J Ind Microbiol 17: 490–497.

11 Tiedje J, J Urbance, N Larsen, T Schmidt, O Strunk, S Pramanik, R Overbeek, R Martin and J Holt. 1996. Towards an integrated microbial database. In: Culture Collections to Improve the Quality of Life. Proceedings of the International Congress of Culture Collections (Samson RA, JA Stalpers, D van der Mei and AH Stouthamer, eds), pp 63–68, Centraalbureau voor Schimmelcultures, Baarn, NE.