WFCC-MIRCEN: world data centre for microorganisms

Hideaki Sugawara

Director
World Data Centre for Microorganisms
Center for Information Biology
National Institute of Genetics
1111 Yata, Mishima,
Shizuoka 411-8540, Japan
Tel: (81) 559 6895
Fax: (81) 559 6896
E-mail: wdcm@wdcm.nig.ac.jp

Introduction

In the 1960s, microbiology and culture collection experts met several times to develop a strategic plan to foster microbial culture collections. They quickly recognised that statistics on the activities of culture collections worldwide were not available and that there was a need for a 'world directory of collections of cultures of microorganisms' that accumulated information on culture collections. It was obvious also that a world data centre for microorganisms was required to develop and maintain the directory.

The production of the first edition of the World Directory in 1972 by the late Professor VBD Skerman of University of Queensland marked the start of the World Data Center for Microorganisms (WDCM). The WDCM subsequently moved to the Institute of Physical and Chemical Research (RIKEN) in Japan, under the guidance of the World Federation for Culture Collections (WFCC), then to the Center for Information Biology and DNA Data Bank of Japan in the National Institute of Genetics (NIG), Japan.

The WDCM has adopted the most up to date information technology: online databases connected to the Packet Switching System, the Internet, gopher, the world wide web (WWW) and eXtensible Markup Language (XML). The WDCM is now heavily dependent on XML technology, although it is not visible to users.

The World Directory is accessible online in the form of two databases – CCINFO and STRAIN. The CCINFO database contains information of acronym, full name, address, staff, services and categories of holdings. The STRAIN database contains lists of scientific names of microbial strains and names of cell lines held by the culture collection registered in CCINFO.

Online registration and update to the databases of CCINFO and STRAIN

WDCM implemented an online system for registration of culture collections and the update of their information in 2001 on the WDCM homepage at http://www.wdc.org/ (Figure 1). Clicking the menu of CCINFO in the box at the middle of Figure 1 takes the user to the menu shown in Figure 2. Completing the registration form (Figure 3) allows WDCM to assign a user-id and password to a culture collection. The WDCM system identifies valid update of a culture collection by recognising its user-id and password. Culture collections can update their holding by downloading the list(s) in the STRAIN database, editing locally and uploading the new list(s) to the WDCM server.

WDCM is able to prepare statistics on the activities of culture collections when required using the CCINFO and STRAIN databases. The number of culture collections currently registered with WDCM is 489 in 65 countries. WDCM has assigned more than 700 registration numbers to culture collections; however, a number of culture collections have since disappeared. This fact emphasises the need for the concept of biological resource centres (BRCs) discussed in the BRC taskforce of OECD Working Party for Biotechnology. Some statistics on culture collections worldwide are shown in Table 1.

e-Workbench

In addition to the mandatory task of maintaining the World Directory, WDCM has developed an e-Workbench that supports 'databasing', classification and identification of microbes. The functions of the e-Workbench include:

- Fully open system using JAVA, CORBA and XML.
- Integration of remote/local databases and tools.

Table 1. Some statistics on culture collections registered with WDCM. Note that only 175 culture collections publish catalogues.

	No. of collections
Supported by government	165
Semi-governmental	33
Supported by university	140
Supported by industry	7
Private collections	19
Collections producing a catalogue of holdings	175
No. people working in collections	2701

Figure 1. HomePage of WFCC-MIRCEN world data centre for microorganisms.

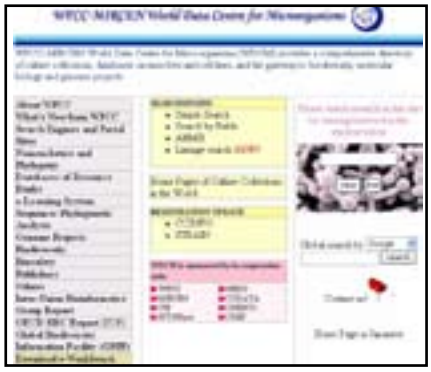


Figure 2. The web page to start the registration, update and delete of the information on a culture collection in the CCINFO database.

Figure 3. A part of the form for the interactive registration and update of the information on a culture collection.

- Import and export XML documents and CSV files.
- Numerical taxonomy (principle component analysis and dendrogram).
- Phylogenetic analysis (multiple alignment and phylogenetic tree).
- Finding phenotypic character(s) to differentiate clusters either in a dendrogram or in a phylogenetic tree.
- Probabilistic identification including construction of a standard data matrix.
- Identification by a decision tree including construction, import and export of a decision tree.
- Homology search (Blast).
- Design of primers.
- Graphical representation of results of data analysis.
- Seamless operation from data retrieval to analysis and back.
- Downloadable from the WFCM site and also available in CD-ROM.

e-Workbench can be accessed by downloading InforBIO from the WDCM site (Figure 4.). A help menu is available (Figure 5). Figure 6 shows an example of data analysis, in this case clustering based on phenotypic characters and the resultant dendrogram. Each colour of the branches in the dendrogram reflects the grouping of each strain. The colouring is also valid to see consistency between the dendrogram and the phylogenetic tree of the same set of strains. Selecting a cluster in the phylogenetic tree by clicking a node in the tree changes the colour of the members (namely, strains) of the cluster chosen and the colour of the members/ strains in the dendrogram also change.

The architecture of InforBIO is flexible and expandable. G-InforBIO (Figure 7) for comparative genomics has been developed based on the same framework as InforBIO. The results of comparing homologs in two chromosomes are shown in Figure 8.

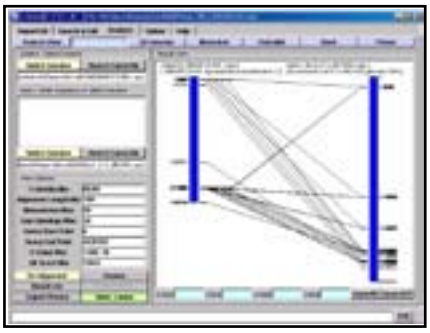
Figure 4. The top panel of InforBIO displayed in your laptop computer.

Figure 5. The menu of the help file of InforBIO.

Figure 6. The phenotypic data in the left window were used for the clustering of strains and the result is displayed in the right window as a dendrogram.

Figure 7. The top panel of G-InforBIO displayed in your laptop computer.

Figure 8. Homologs in two genomes are graphically linked by lines.



Conclusion

WDCM assigns a globally unique identifier to each culture collection. If this identifier is combined with a type of collection and the accession number of a strain, the strain will have a globally unique identifier too. This identifier will be very useful for a global information network of culture collections. Thus the mechanism for the registration of culture collections with WDCM should remain stable. However, WDCM is maintained voluntarily by the author so is not necessarily stable in the long-term. Should a global network of BRCs (GBRCN), as discussed in OECD, materialise, then WDCM should find a home in the GBRCN.