

Computational tools for the study of the genomes of filamentous fungi

L. Ellis, P. Ramos, J. Kirk, C. Floyd and J. Bender, W.M. Keck Center for Genome Informatics, Institute of Biosciences and Technology, Texas A&M University, Houston, Texas 77030

Greg May, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030

Tom Adams, Department of Biology, Texas A&M University, College Station, Texas 77843

During the past year, we have developed and deployed several new computational tools to provide new means of access to information used for research in the biology of filamentous fungi. These new tools complement and extend other resources available for access to both molecular and biological data.

bionet.mycology

A new Bionet Newsgroup, bionet.mycology, has now been established, and has been in use for most of the past year. It is interesting to note that, contrary to some of the discussion posted during the open forum preceding the vote on establishing the Newsgroup, news articles posted to date has certainly not been dominated by discussions pertinent to molecular genetics. Quite the contrary! Articles thus far have certainly represented a broad range of topics concerning the biology of numerous fungi, from taxonomy to methods to general information. If anything, molecular biologists have been rather quiet. We hope that molecular biologists will in fact become more active participants in the coming year, so that bionet.mycology will become (as we hoped in the Charter of the Newsgroup) a broad forum for discussion of all aspects of the biology of filamentous fungi.

Mosaic and the World-Wide Web (WWW)

More than any other single event during the past year, the now widespread use of NCSA's Mosaic has rapidly set a new standard for providing users with convenient hypertext links to vast distributed repositories of information on the Internet. Mosaic now is available by anonymous ftp from NCSA (<ftp://ftp.ncsa.uiuc.edu/>) for Unix, Macintosh and Windows. Navigation across the WWW is by means of Universal Resource Locators (URLs), which provide embedded links to documents rendered in HyperText Markup Language (HTML). Such documents can reside either locally, or on any remote network machine which runs a WWW daemon process (e.g., CERN's or NCSA's httpd). Furthermore, data types other than text can now also be made available, e.g., images (in gif or jpeg format), video (in mpeg) and audio. With the introduction of interactive forms in version 2.1 of Mosaic, users can now enter information and return it via email to the WWW server. We anticipate that this may provide one very useful avenue by which researchers can submit data for inclusion into filamentous fungal databases (see below). Access to the WWW server at Keck-IBT is via the following URL:

<http://keck.tamu.edu/ibt.html>

This will take the user to the logo of Keck-IBT, which when clicked leads to the first page of information. Look under the "WWW" or "What's New on the Keck-IBT WWW Server?" links, where extensive information about all of our computational activities can be found.

There are two major sites which serve as repositories for information and tools for the WWW. They are available both via the WWW, and via anonymous ftp to retrieve software.

<http://info.cern.ch/>, CERN in Geneva, where the WWW was invented by Tim Berners-Lee

<ftp://info.cern.ch/>

<http://www.ncsa.uiuc.edu/>, NCSA at the University of Illinois, home of the Mosaic client

<ftp://ftp.ncsa.uiuc.edu/>

Furthermore, the WWW is rapidly becoming the generic interface to many very useful repositories of information. Examples which we find useful for ready access to numerous databases and pointers to other WWW servers include:

<http://kufacts.cc.ukans.edu/cwis/units../index.html>, the FGSC On-line Catalog

<http://genome-www.stanford.edu/>, the *Saccharomyces* Genomic Information Resource

<http://probe.nalusda.gov:8000/>, Agricultural Genome World Wide Web Server at the National Agricultural Library, USDA

<http://www.gdb.org/hopkins.html>, the Johns Hopkins University Bioinformatics Web Server

<http://expasy.hcuge.ch/www/expasy-top.html>, The ExPASy Molecular Biology Server in Geneva

AGsDB: A Genus species DataBase

In the past 2-3 years, the ACeDB (A *C. elegans* DataBase) database engine developed by Richard Durbin and Jean Thierry-Mieg for the *C. elegans* Genome Project has been adapted for use with a wide range organisms. We have modified and extended the Class/Key structure of ACeDB to provide for the inclusion of data for multiple species, with the added functionality of queries between defined homologs of different species. As of February 1994, data in AGsDB includes *Aspergillus nidulans*, bovine and human anchor loci, cotton and, most recently, *Neurospora crassa* (the latter is with the advice of Dan Ebbole [Department of Plant Pathology, Texas A&M University]).

The starting point for both species of filamentous fungi was to enter the genetics maps and anchor loci for each species. For *A. nidulans*, the genetic map of Clutterbuck (FGN 1991) was used; for *N. crassa*, the 1993 Genetic Map of Perkins together with the information for each

locus from the 1982 Perkins compendium (Perkins et al. 1982. Microbiol. Reviews 46:426-570). Additional data types include Colleagues and References (Fungal Genetics Newsl. 40, 1993), Strains (*A. nidulans*, as of Fungal Genetics Newsl. 40, 1993) and hybridization data (when available) for Clones in the two *A. nidulans* cosmid libraries (Fungal Genetics Newsl. 40, 1993).

An important topic for discussion among interested users is how to provide an effective means by which data can be contributed, and which data are of general enough interest to include, i.e., curation of the database beyond this initial prototype that we have introduced. Furthermore, the database is designed to be able readily to include other species, as well as homolog information between species. We welcome suggestions on these topics, either via bionet.mycology or email to the authors (see below).

Once work on this initial prototype is finished, anticipated for the Spring of 1994, AGsDB will be available via anonymous ftp from:

<ftp://keck.tamu.edu/>

in the pub/AGsDB directory.

The latest versions of the ACeDB database is always available from:

<ftp://ncbi.nlm.nih.gov/>

Gopher

Given the widespread use of Gopher, the information contained in AGsDB is wais-indexed and made available via gopher at:

<gopher://keck.tamu.edu>

In addition, the following is probably the single most useful Gopher site, as it includes numerous pointers to biological resources:

<gopher://gopher.gdb.org/>

WWW-AGsDB

ACeDB is written in the C programming language, uses X windows as its graphical user interface, and was designed to run in a Unix environment (Macintosh version of ACeDB has recently been released by Frank Eckmann). An important extension to the functionality of ACeDB appeared at the end of 1993: a WWW--ACeDB server designed by Guy Decoux at INRA (decoux@moulon.inra.fr). This is an additional software module that provides an interface between user queries initiated in Mosaic, the httpd WWW daemon, and the ACeDB database itself. We feel that this is a very significant advance, given the widespread use of Mosaic, as users now do not have to maintain a local Unix of the database, but can readily access a remote one via Mosaic and the WWW.

WWW-AGsDB is available now at the Keck-IBT URL:

<http://keck.tamu.edu/ibt.html>

Much of the functionality of the native AGsDB is available through this WWW interface, particularly data browsing by moving from link to link.

At present, we are working on various strategies to provide pointers within WWW-AGsDB to other external databases. In particular, we would like to take advantage of the new WWW Server at the FGSC,

<http://kufacts.cc.ukans.edu/cwis/units../index.html>

so that now we do not have to store the fungal data contained in the FGSC catalogue, but rather point to it from WWW-AGsDB. Thus, WWW-AGsDB would provide additional functionality, especially genetic maps, as well as other information of interest that is not stored at the FGSC. The latest developments can always be found at:

<http://keck.tamu.edu/ibt.whatsnew.html>

Other sites which now have implemented WWW interfaces to ACeDBs include:

<http://inra.moulon.fr/>, where this interface was developed

<http://genome-www.stanford.edu/>, the *Saccharomyces* Genomic Information Resource

<http://probe.nalusda.gov:8300/>, WWW Interface to ACEDB Databases at the National Agricultural Library, USDA

Summary

We welcome the input from the community of researchers interested in the study of filamentous fungi. We hope that these tools will be of use, and we certainly hope to provide the requisite functionality. Suggestions are welcome either at bionet.mycology, or via email to the authors:

Leland Ellis: leland@straylight.tamu.edu

Jeff Kirk: jkirk@keck.tamu.edu

Greg May: gsmay@bcm.tmc.edu

Tom Adams: tom@isc.tamu.edu