embnet.news

Volume 3 Nr 3 (ISSN1023-4144)

December 21, 1996

Editorial

It looks like circumstances overwhelmed us this year. We lost two members of the editorial board and also one issue of the newsletter. However, we have co-opted two new volunteers as editors - Rob Harper from the EBI and Peter Rice from the Sanger Centre - and trust that they will help keep embnet.news on course nad on time. Thanks are due to retired editors Reinhard Doelz and Alan Bleasby who have been associated with the newsletter since its foundation.

This, year-end, edition has a number of contributions of interest to molecular biologists. Terri Attwood explains her multiple sequence editor, Cinema and Nicole Redaschi descibes the SynCron tools for updating the EMBL database. There is a report on the 10th EMBnet AGM in Helsinki last month and an interview with Bill Pearson who was an invited speaker at the scientific meeting as, indeed, was Terri Attwood.

Prof Pearson emphasised the need for more, and more effective, training in the world of bioinformatics and we have a description of one such course from Joelle Thonnard. EMBnet is moving next year into a new grant from the EU. A major part of this money will be devoted to new training courses. The EMBnet WWW bioinformatics training course has now been officially launched - see the announcements section. It is hoped that all this effort will not be wasted. It is certainly heartening to see people running programs correctly and interpreting the output accurately rather than merely running programs. There is no doubt that properly those trained in bioinformatics are a desirable commodity in the jobs marketplace.

We would like to wish you Happy Computing in a Happy New Year.

The embnet.news editorial board.

Contents

Editorial	1
The 10th EMBnet AGM in Helsinki	1
Syncron: tools for maintaining databases	2
CINEMA: an interactive multiple sequence alignement editor	4
TIPS from the Computer Room	5
INTERviewNET : Bill Pearson	5
The VSNS-BCD biocomputing course	7
Book Review section	8
Announcements	8
Node news	9
The EMBnet Nodes	11
embnet.news information	12

The 10th EMBnet AGM in Helsinki

Rodrigo Lopez, EBI.

Going to Helsinki during the month of November can be quite an experience even for seasoned travelers. Last year the temperature in November was -15 degrees Celsius but then, to be fair, it was blowing a bit!

This year it turned out Helsinki was much warmer that the rest of Europe and I didn't get away with my descriptions of going snow-rolling or ice-dipping after a sauna. All I got was wet from easterly showers and temperatures above 5 degrees Celsius.

The weather was merciful and our three special guests were spared the contrasts of the Finnish Sauna. Steven Altschul, Bill Pearson and Terri Atwood did the honours during the scientific part of the 10th EMBnet AGM. You will find an interview with Bill Pearson as well as an article on CINEMA by Terry Atwood in this issue of embnet.news.

The EMBnet organisation seems to be settling well within the molecular biology community and more nodes continue to be added. This time applications from Russia, China, Australia and Pharmacia-Upjohn in Sweden were accepted. The world turns around and with it changes takes place within the old membership: The transfer of the national node mandate for Switzerland from Basel to Lausanne was accepted. Thus a new node is established for the Swiss community. The node reports displayed interesting activity in the fields of software development, with Spain embarking in Java molbio application development, and education being added to the Swedish portfolio. On the hardware side, some nodes have gone through major computer upgrades such as Israel and news of new models of Bioccelerators (Second Generation) was greeted with enthusiasm.

Alan Bleasby from the UK Seqnet node announced the impending availability of AJAX, a C language function library which will be the core of the new EMBOSS (European Molecular Biology Open Software Suite). EMBOSS is the next step for the EGCG project and aims primarily at it gaining independance from the Genetics Computer Group function libraries. Quite a number of nodes are already involved in various ways in the project and interested parties may contact Peter Rice, from the Sanger Centre node, for further details.
As in previous AGMs, the election or re-election of officials took place. A new Executive Board was elected which

consists of: Sandor Pongor (Chairman), Andrew Lloyd (Secretary), Rodrigo Lopez (Treasurer)and Peter Rice (vara member). I'd like to take this opportunity to thank Dr. Alan Bleasby and Dr. B. Savakis on behalf of EMBnet for their services as EB members!

The project commitees consist today of:

R&D TechMan:	
Peter Gad	Chair
Gijs Schaftenaar	Treasurer
Thure Etzold	Secretary
Leon	Esterman
E&T:	
Martin Bishop	Chair
Jack Leunissen	Treasurer
Martin Grabner	Secretary
Takis Benos	
P&PR:	
Pedro Fernandes	Chair
Heikki Lehvaslaiho	Treasurer
Linda Langholm	
J-R Valverde	
embnet.news:	
Rob Harper	
Robert Herzog	
Rodrigo Lopez	
Andrew Lloyd	
Peter Rice	

Dr. Jack Franklin, who has been the Manager of the EMBnet Stichting for the past 4 years completed his contract. We would like to thank him very much for his services and advice.

Finally, a descision to hold the next 1997 AGM due south was taken. EMBnet will go to sun-drenched Italy. The Bari node has agreed to host this meeting and promises a stimulating scientific meeting on the topic of molecular evolution.

SynCron

tools for maintaining synchronised copies of the EMBL Nucleotide Sequence Database

Nicole Redaschi (1), Leif Landeman (1), Matteo diTommaso (1) and Reinhard Doelz (2)

Introduction

The EMBL nucleotide database doubles in size approximately every twelve months. A release of the database is made every 3 months and distributed via ftp and CD-ROM. Between releases, updates are available in two forms on the EBI ftp server: a cumulative set, containing all updates since the last full release to the current day, and incremental sets issued at least once a day and compiled into weekly sets. The incremental update files range between 100Kbytes and 25Mbytes with the typical file being less than 2Mbytes. The cumulative set at the end of release 48 has exceeded 150Mbytes.

Using the cumulative file to update a local database copy offers the advantage that there is no need for further data processing and, hence, no risk of errors arising through the manipulation of incremental update files. At many sites the available network bandwidth does not make it reasonable to transfer the cumulative file by ftp from the database provider site daily or even weekly. Downloading incremental files requires much less bandwith, but higher local maintenance effort, since the incremental files have to be integrated into a cumulative file to present the data as a single file for conversion into GCG format or indexing by SRS. To facilitate this processing step, and to provide a reliable mechanism to regenerate the cumulative file locally from incremental updates, we have developed the SynCron tools.

Systems and methods

The SynCron package contains a set of file manipulation utilities written in the C language and scripts to run the requested updating job on Unix (csh shell) and VMS (DCL) operating systems. SynCron is available by anonymous ftp from:

UNIX Version:
ftp://ftp.ebi.ac.uk/pub/software/unix/listtools/SynCron_XXX.tar.gz

VMS Version (backup/gzip):
ftp://ftp.ebi.ac.uk/pub/software/vms/listtools/SynCron_XXX.bck-gz

VMS Version (tar/compress):
ftp://ftp.ebi.ac.uk/pub/software/vms/listtools/SynCron_XXX.tar_Z

The ~XXX version number (currently, 005) will change as the programs are updated.

1 Database Programming Group, EMBL Outstation, The European Bioinformatics Institute E-mail: ditommaso@ebi.ac.uk, landeman@ebi.ac.uk, redaschi@ebi.ac.uk
2 Present address: Sandoz Pharmaceuticals, Basel, Switzerland EMail: reinhard.doelz@sandoz.com

SynCron makes use of transaction listings which have been made avaiable by the EBI for the nucleotide database since January 1996. The EBI supplies a listing for each of its update files that describes the update, insert and delete operations to the database represented in the flat-file updates. The core program of SynCron merges updates into the cumulative file following the instructions given in the transaction listings. Two additional utilities are provided to verify that the resulting cumulative file contains the correct entries. Currently, accession number, entry name, version, divison and a datestamp of an entry are validated. In the future, we may include the NID of an entry (a unique identifier for the sequence), or a checksum for the sequence to prove its identity.

The programs are launched by a configurable script. The script keeps track of the incremental files that have already been merged into the cumulative file by writing their names to a logfile. If there is more than one new incremental file, the files are processed in sequential fashion. The script can easily be customized to exclude classes of data (e.g., EST, GSS, etc.) from the updates by preprocessing the transaction listings.

The SynCron package also includes tools to assist file transfer. If you have already a working configuration for obtaining the update files, there is no need to change this, except that you will need to add a procedure that fetches the list files in addition to the data files.

List files are available from ftp://ftp.ebi.ac.uk/pub/databases/embl/new/list. The naming scheme is the same as for incremental and cumulative update files with the extension «.lis». All customization for local file names and paths can be done in a general configuration file, as explained in the installation instructions for the package.

Discussion

Surveys of the EBI ftp server show that incremental update files are already used by some sites. We noticed, however, that the data available on public servers (such as the EMBnet SRS servers) often differ by a number of entries.

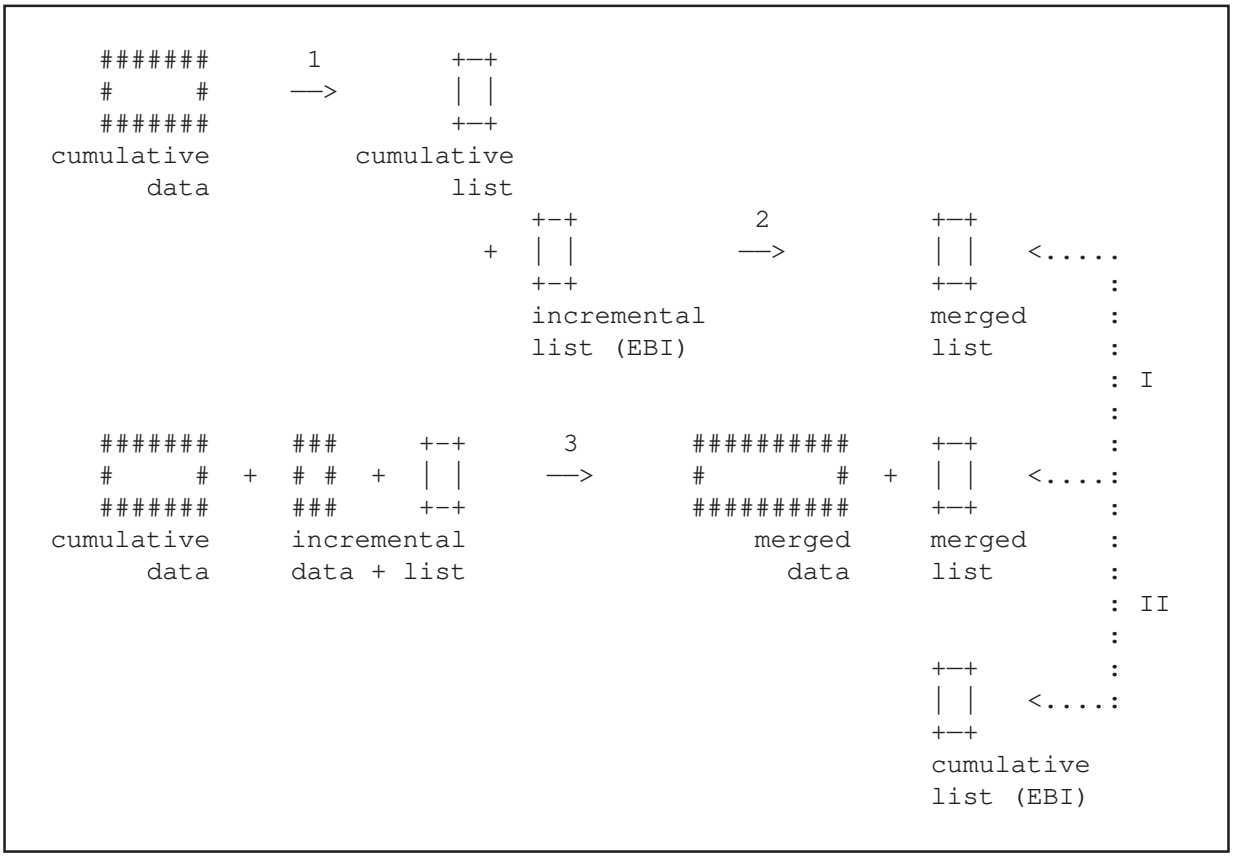

Figure: A cumulative list is derived (1) from the local cumulative data file and merged (2) with the new incremental list to produce a merged list that will serve to verify the actual data merging process. The new incremental data file is then merged into the cumulative data file according to the instructions given in the incremental list (3), generating the new cumulative data file, as well as a merged list file. The latter is compared to the merged list produced in the previous steps (1+2), to verify the contents of the merged data file (I). If using daily update files there is an additional, optional verification step, a comparison to the cumulative list file that is generated daily at the EBI (II).

Using SynCron it should be possible to keep a copy of the EMBL Nucleotide Sequence Database that exactly matches the contents of the database in operation at the EBI for external services, with manual intervention required only in the event of network failure, etc. We hope, thereby, to help users to improve and guarantee the quality of the EMBL Nucleotide Sequence Database updates obtained by electronic transfer.

CINEMA

A Colour INteractive Editor for Multiple Alignments -

Attwood, T.K.[1], Payne, A.W.R.[2], Michie, A.D.[1] and Parry-Smith, D.J.[3]

Introduction

The World Wide Web has become the most popular vehicle for distributing bio-information. The 'information superhighway' was initially used primarily to ship data, but we are now seeing a shift away from data dissemination per se to its use in transmitting concepts. This is true, for example, in the pharmaceutical industry, where extraction of information about potential structural or functional sites from sequence data is an essential component of drug-discovery protocols. Thus, while the provision of, and links between, databases has revolutionised the way we access data, visualisation and interactive manipulation of data are now key goals in allowing users to get the most from their bio-information.

For the sequence analyst, a vital tool is an alignment editor. Numerous programs are now available, either in a stand-alone form or as components of larger packages. The programs range from fully-manual to fully-automatic, but results from automatic procedures almost invariably require manual editing. This often presents problems, as there is currently no standard format for output, storage and distribution of alignments.

Java, an object-oriented network programming language that allows interactive use of software over the Web, begins to address some of these problems. Java-capable browsers may run applets on a variety of platforms (applets are small applications launched from a server via HTML pages). To an extent, this obviates the need to distribute code, as software is loaded on-the-fly from the server, and cached for that session by the browser. Executable code will thus run on virtually all desktop platforms without modification: when modifications are made and the source recompiled, the program should run everywhere.

To address the torrent of genome data, new-generation tools are required to deliver up-to-date information to the community via user-friendly, interactive interfaces. To this end, we have developed CINEMA, a tool both for local alignment construction and modification, and for visualisation and manipulation of alignments resident at different sites on the Internet.

The program

CINEMA is embedded in a comprehensive help file, so instructions on its use are immediately to hand. The applet is demonstrated with an alignment of lysozyme sequences, which are easily purged from the display to allow input of user-specified files: these may be loaded directly from local sequence or alignment databases, from a temporary directory, or they may be input by the user (via cut-and-paste or file-upload facilities). The program thus provides the flexibility to extend and enhance pre-existing alignments, and to generate alignments from scratch.

Navigation around the display is effected by means of scroll bars, and gaps are inserted/deleted using a click-and-drag mouse action. Other facilities include group editing (to allow simultaneous gap insertion into sets of sequences); sequence re-ordering/removal; variation of font sizes and colours; etc.. By default, alignments are coloured according to residue property groups consistent with those of physical modelling components and graphics packages, vis: acidic=red; basic=blue; polar neutral=green; hydrophobic aliphatic=white; hydrophobic aromatic=purple; P,G=brown; C=yellow. Output options allow files to be saved to a temporary directory for future program input, or alignments may be mailed-back (in text or PostScript formats). Alternatively, results may be output in gif format for display within the browser.

Conclusion and future directions

Interactive editors are essential where alignments from automatic programs require manual adjustment. The use of colour then allows rapid interpretation of the results, allowing different properties to be depicted in an immediately informative way, no matter how large the alignment. Such tools thus offer a rapid and informed means of selecting residues suitable for mutagenesis studies by revealing regions

1. Dept. of Biochemistry and Molecular Biology, University College London, London WC1E 6BT, UK
2. Cyberdynamics (UK) Ltd., 38 Lordsbury Field, Wallington, Surrey SM6 9PE, UK
3. Dept. of Molecular Sciences, Pfizer Central Research, Sandwich, Kent CT13 9NJ, UK

crucial to the structure and/or function of a protein: e.g., critically conserved residues can be seen at a glance; unusual mutations may stand proud against a smooth backdrop of conservation; and mutational hotspots are readily pin-pointed.

CINEMA is the first component of a modular network-oriented analysis package. A structure display module is now well-advanced; this links the sequence to features of biological interest, by allowing visualisation of conserved motifs in a 3D context. Other emerging tools include hydropathy plots, diagonal plots, etc.. Our intention is to extend the applet flexibly, by dynamically loading new classes to 'plug in' additional functionality. With this open approach, our aim is to permit multiple centres to develop and deploy extensions to the applet through construction of custom 'pluglets', allowing the package to grow rapidly. Such cooperation avoids costly duplication of effort, it encourages global collaboration, and allows convergence to a set of standards.

The different features of this program are not new or remarkable in themselves. What is striking is that alignment manipulation is able to happen in real time, that users may swap data with the applet, and that the applet may be enhanced and expanded without the need to distribute code.

Availability

CINEMA is accessible via UCL's Bioinformatics server at http://www.biochem.ucl.ac.uk/bsm/dbbrowser/CINEMA/. For security reasons, use of the Internet is not a universal solution. We are therefore making the applet available for installation on organisations' Intranets.

TIPS from the computer room.

Looking for help with the details of bioinformatics ? No manual available ? You might find the assistance you need from the WWW bioinformatic turorials which can be found at two EMBnet nodes:

1. HGMP-RC
http://www.hgmp.mrc.ac.uk/Embnetut/Universl/embnettu.html

2. BioBase
http://biobase.dk/Embnetut/Universl/embnettu.html

Here you should find a wealth of information about bioinformatics, unix and much much more. Your comments and feedback are welcome.

INTERviewNET

Robert Herzog and Andrew Lloyd interview Bill Pearson

Helsinki 22nd November 1996

RH. It is a great honour having you around here at our annual EMBnet meeting ! It was certainly an excellent idea from our organiser to invite you to give us a talk about the current FASTA software. Is this your first contact with EMBnet as such?

BP. I think so. I knew it existed. Maybe I have used a node in the past ? I certainly use SwissProt.

RH. EMBnet is primarily trying to serve the community of molecular biologists and associated people with services in biocomputing. Is it a different situation from what exists in the United States ?

BP. Yes. Few institutions in the U.S. have a formal support mechanism. However, at the University of Virginia, we have a computational biologist who teaches classes and consults with researchers. Places where I give talks may not have a local copy of databases, they may not have local copies of GCG, they may only have access to databases over Web servers. So their perspective on what you can do with sequences is quite different from when you have somebody at your university who actually keeps the databases up to date and makes sure the programs work properly.

RH. So in most major institutes, there are some personnel acting ...

BP. «Most» is certainly not true. Perhaps 20 - 25% of the institutes have molecular biology computing support personnel - usually only one person. I don't know of any place except the genome centres where you have more than one person doing these kind of things, because at most institutions, the number of people that are actually using the sequences is not large.

RH. You are the author of one of the major tools for exploiting the databases and all the diversity they contain. You have been developing this software for many years. How do you feel the integration of that software into other packages would best take place ?

(BP: William Pearson, AL: Andrew Lloyd, RH: Robert Herzog)

BP. The software is available on the Internet and the major companies have licences for integrating it in their packages. GCG has it, Intelligenetics has it People at the EBI, people working of egcg or the SRS system can build connections between their software and my software very easily.

I am not a very big operation (laugh). I've done most of the programming for this myself or by getting code from other people. The easiest way to get into packages from my perspective is basically to make my program code available to everyone and try not to be restrictive in the way it can be redistributed. I think it has been pretty successful.

AL. With a software set that has become one of the de-facto standards, and so popular, are you tempted to go commercial ?

BP. No! The university of Virginia does licence FASTA to companies that redistribute it. I had some experience at doing this long before I wrote FASTA, and the result of going commercial was that nobody got to use the program. My business is being a professor, not selling software. My first priority is to have people use my programs.

The thing that people should pay for is the support, and that is not something that I want to sell. I certainly don't mind people asking questions about my software, but I think that if I charged people for the software, they would feel I should have more responsibility.

RH. Do you think that this is an area where EMBnet should have a role ?

BP. Certainly the more teaching you do, the better. When I give talks and courses, frequently people get surprised to learn that there are these things that they wanted to do, were actually easy to do, but they did not know how to do them. Maybe the Web will make that easier in the future. It will be easier to show all the different options and parameters, and so people will wonder about their use.

RH. Have you felt in the past that the software produced by Steve Altschul from the NCBI with BLAST was in some kind of rivalry with your approach ?

BP. Well, their package is very popular, and it is a very good package. I think that for some things, FASTA does a better job. I think the reason their package is used so much is because it is available on a server with databases that get updated every night. If EBI or EMBnet or somebody else would do that for FASTA, that would be great.

AL. But there is an unfortunate discrepancy here. For example, in the GCG documentation, they say «for doing nucleic acids homology searches, don't use BLAST, use FASTA». It would be nice if there was an integrated service so that, when you say you want to do a nucleic acids homology search, it would tell you «for this job, you might rather use FASTA.»

BP. I try to discourage people from doing do DNA sequence comparisons. There usually isn't very much information in the DNA that is not in the protein. I guess I don't mind that people don't have better access to FASTA for doing DNA sequence comparisons, because it's so much better to do protein comparisons. It is unfortunate that so much DNA comparison gets done. I am told that DNA searching is used at least 10 times as much as protein searching. And the databases are much larger, so you are likely to spend maybe hundred times more computer time doing DNA comparisons. That is very inefficient.

RH. This is certainly an area where teaching has a role to play. How is the teaching of bioinformatics in the United States organised ?

BP. There are several courses. I organise one at Cold Spring Harbor with Randy Smith. There is a course at the Pittsburg Supercomputing Centre. These are one week courses.

RH. None of them have been integrated in the curriculum of faculty students ?

BP. Different universities have courses and I have participated in some of those. One problem is that we don't have a textbook. And we don't have many faculty that have training in sequence analysis. I teach a course at my university as any computational biologist with a faculty position does. But there aren't very many of us (laugh). Unfortunately, computational biology is new. There are not very many students that know about it at all. Most computation biologists in the United States don't have many graduate students. Students tend to choose their graduate courses from the undergraduate courses they followed. Maybe in five years from now, all of them will have had computational biology.

RH Thank you very much for this interview, Professor Pearson.

The VSNS-BCD biocomputing course

A contribution by Joëlle Thonnard (UCL, Laboratoire de Biologie Moleculaire), user of BEN, who assisted to the on-line interactive VSNS-BCD biocomputing course, organized by the University of Bielefeld.

As you know, molecular biologists rely more and more often on computer aided sequence analysis tools. For example, they need computer programs to compare DNA and protein sequences, to search for coding regions in DNA sequences, to predict the secondary and tertiary structure of DNA, RNA and proteins.

Once, programs for sequence analysis were mainly running on UNIX systems and required cryptic commands. Currently, more user friendly programs are available. Therefore, when biologists want to use tools for computer aided sequence analysis, they do not necessarily need to be expert in mathematics and computer sciences. However, they still face a difficult task because the number of tools is large and even exploding : there are a large number of databases (DNA sequences, gene maps, protein sequences, protein 3-D structure, specific databases,...), a lot of programs and various ways to access the tools (one can run a terminal session on a UNIX machine, install programs on a personal computer, access servers via the Web).

Thus it is necessary for a biologist to acquire reliable knowledge of biocomputing. In other words, to learn the assumptions and the principles of the methods on which the programs are based, in order to use them correctly and efficiently. But in the universities and research centers, there are not yet a lot of specialists in biocomputing and courses in this field are rare. Fortunately, it is now possible to find, on the Web, introductory material about biocomputing. Moreover, it is also possible to access interactive biocomputing courses.

For the last two years, the VSNS-BCD (Virtual School of Natural Sciences -BioComputing Division) has delivered a biocomputing course over the internet. The main components of the course were the Hypertext Coursebook, on-line sessions using the Electronic Conferencing system BioMOO, and the hands-on visualization of some important concepts such as sequence alignments.

An updated Hypertext Coursebook is accessible via the Web. Hence the students can proceed at their own pace. The Coursebook comprises the following topics:

Introduction
Networking
Pairwise sequence alignment
Weight matrices for sequence similarity scoring
Fasta/Blast exercises
Multiple alignment
Mathematical basis of molecular phylogenetics
Genetic algorithms and protein folding

The MOO in BioMOO stands for MUD, Object-Oriented. MUD stands for Multi User Dungeon, (or more recently and pretentiously, Multi User Dimension). The MUDs are on-line interactive communication tools. They are based on a client/server architecture and TCP/IP connections. They were firstly developed for «dungeon and dragon» role-playing games but, currently, they are also used as virtual places on the network where people can meet and collaborate on various projects. Indeed, the MUD server represents a virtual environment composed of different rooms where small groups of people can enter and interact. The people connect to the server using a simple terminal connection (telnet) or a MUD client. The BioMOO is such an environment where the course sessions were held. Thus, over a 2 month period, 6 groups of about 6 students and an instructor met once a week and discussed the chapter and the exercise assignments. An interesting aspect of this is that, during the session, the instructor encourages the discussion rather than gives a professoral course. Thus, each one learns from each other and the teacher/student boundary is gradually erased.

Java Applets begin to be used in the course, for example, to visualize sequence alignment. Indeed, an alignment of two sequences can be thought of as a unique path through a two dimensional coordinate system. In the case of three sequences, the alignment path can be visualized within a three dimensional cube. With the Java tool Visu3d, the cube can be rotated in all three dimensions using the mouse to point into the desired direction.

In conclusion,

all the participants agree that the course allowed them to acquire some skill in biocomputing but also meet interesting people and, moreover, have a lot of fun together. Every interested person is invited to visit the VSNS Biocomputing Web site. Here one can get a better idea about the course: the coursebook but also the transcripts of the sessions are available ; there are also some statistics which show the range of people who have participated, how many disciplines and how many nationalities they come from. I hope that new courses will materialize !

You can access the home page of the VSNS course at: http://www.techfak.uni-bielefeld.de/bcd/welcome.html. You will find contributions of the students (among which the author herself) at: http://www.techfak.uni-bielefeld.de/bcd/ForAll/welcome.html.

Book Review section

Internet for the molecular biologist

Eds Simon Swindell, R. Russell Miller, Garry S.A. Myers.

The Internet is growing! For molecular biologists, every month brings new resources, new web-sites, new servers. It

is hard to keep up with it and still hold down a regular job at the bench.
One consequence of the insidious creep of commercialism into the internet is that Web pages get more «professional» and take longer to download.

There are several more or less famous on-line jump-sites or resource lists such «Pedro's BioMolecular Research Tools», but for the average biologist surfing in a corner of the lab with a 13" screen it can be useful to have «The Manual» open beside them on the desk. This is certainly true for me and I suspect for many of the print generation - which is most people born before, say, 1965. So where *is* that bioinformatic manual ? Where is the Maniatis of the Internet ? Three years ago, you could ask the same question of the Internet itself and be given a choice of two: Ed Krol's The Whole Internet and Zen and the Art of the Internet by Brendan Kehoe. Now the bookstore shelves groan under yards of Internet guides.

With Internet for the Molecular Biologist we have the first such guide, professionally produced and commercially available as hard copy; albeit between soft covers. It has a comfortable, rather dated, feel because it is acting on the hypothesis that real biologists still use e-mail servers. To many the idea of constructing an e-mail message such as:

```
Program blastn
Datalib nr
Expect 0.75
Begin
>myclone
atcgcgtgcgtgtgctgatgacgtgcgtaaatgatcag
END
```

and mailing it off to NCBI will seem quaint but it does force you to think a bit about the mechanics of what you are doing. This command-line approach is so pervasive that there is only one picture of a netscape screen in the whole book. The publishers have kindly put chapter abstracts up on their web page (http://www.almac.co.uk/a/hspabsv3) and arranged for the first editor to have many of the resources brought together on the web at (http://www.ccc.nottingham.ac.uk/~mbzsrs/IFTMB.HTML). This will ameliorate the inevitable problem of such books being out of date before they reach the printer.

The first several chapters introduce the internet itself to emphasise that there is more to it than Netscape browsers. There is a good chapter listing servers for sequence retrieval and analysis and describing the nuts and bolts of connecting to them, followed by a chapter on resources for computational gene identification. The soft centre of the book consists of chapters on BioSci newsgroups and BioMOO. The last few chapters are a bald list and short description of resources (e-mail, gopher, ftp and WWW) for the molecular genetics of human/mouse, fungi, invertebrates, plants and microbes.

Is all this useful ? I think, in the main, it is, not only in the content but also to have in a medium that you can have open beside you and fill up with bookmarks and annotation. But perhaps you should wait to see what the Christmas publishing deluge brings to the market as an alternative.

Andrew T. Lloyd
EMBnet Ireland
11/10/96

ANNOUNCEMENTS

CLEANUP v 1.8

The lastest version of Cleanup is now available.

CLEANUP, using an algorithm based on an «approximate string matching» procedure, is able to determine the overall degree of similarity between each pair of sequences contained in a nucleotide sequence database and to generate automatically nucleotide sequence collections from which redundancies have been purified.

Cleanup was fully described in the last issue of embnet.news :
http://www.embnet.org/embnet.news/vol3_2/software.html
and was also published in CABIOS:
Reference : Giorgio Grillo, Marcella Attimonelli, Sabino Liuni and Graziano Pesole (1996) CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. CABIOS 12, 1-8

CLEANUP 1.8 works both on a GCG environment and on a GCG independent platform. It has been compiled and tested using DEC Alpha AXP C compiler, Borland C++ 4.0 compiler, and, moreover, it is portable over any machine with an ANSI C standard compiler.

CLEANUP 1.8 is available on anonymous ftp
ftp://area.ba.cnr.it/pub/embnet/software/Cleanup

For support and any information contact:
Giorgio Grillo : giorgio@area.ba.cnr.it
Sandra Brunetta: areasb16@area.ba.cnr.it
Sabino Liuni: sabino@area.ba.cnr.it

CCP11 Web site

The UK Collaborative Computational Project for Biosequence and Structure Analysis (CCP11) now has a

website and can be found at the URL http://www.dl.ac.uk/CCP/CCP11

Its goals are to implement 'flagship' code development projects, to maintain and distribute code libraries, to organise training courses, to hold meetings and workshops, to encourage overseas researchers to visit the UK for conferences and colloborative visits and to issue a regular newsletter.The first of these newsletters is to be found on our website.

Genes, Proteins and Computers IV

19-21 February 1997
CLRC Daresbury Laboratory

This is the fourth in a successful series of international conferences on bioinformatics, networking and computing in molecular biology. This is a conference for both inexperienced and experienced users of computational molecular biology tools. It focuses on the biology and on how novel computer applications can aid the researcher. Subjects covered in this conference include comparative modelling, protein docking, ab initio prediction, genome informatics, sequence and structural phylogeny and hidden Markov models.

The speakers include Michael Ashburner, Amos Bairoch, Peer Bork, John Collins, Richard Durbin, Sean Eddy, Nick Goldman, Tim Hubbard, David Jones, Andrew Lyall, John Moult, Christine Orengo, Paul Sharp and Mike Sternberg. Further details of the conference, including registration information, can be found on the CCP11 homepage http://www.dl.ac.uk/CCP/CCP11

NODE NEWS

Austria

SOFTWARE

To determine whether a protein contains a PEST region, a program called PEST-FIND was developed by Martin C. Rechsteiner and Scott W.Rogers in 1986. Recently the significance of PEST sequences and the PEST-FIND algorithm was reviewed in TIBS:

Rechsteiner M. and Rogers S.W.; PEST sequences and regulation by proteolysis; TIBS 21, 267-271 (1996)

We translated the core algorithm of the original PEST-FIND QBasic code kindly provided by Martin C. Rechsteiner into PERL and offer a possibility to analyse protein sequences on our Website (http://www.at.embnet.org/embnet/tools/bio/PESTfind/)

HARDWARE

The Vienna University Computer Center added a multimedia workstation SGI Indy 5500 (96MB RAM) to our equipment.

Because of a damage caused by a circuit breakdown our old EMBnet Server a DEC System 5900 has to be retired. Cost benefits consideration did not allow us to repair the server.

The computer equipment for the teaching room was replaced. Now courses are held on PCs (P5 120 MHz, 16 MB RAM) with Windows 95 as OS.

Announcement of the FEBS Advanced Lecture Course «ATP Binding Cassette (ABC) Transporters: From Multidrug Resistance to Genetic Disease»

Supported by the Austrian Federal Ministry of Science, Transport and the Arts

Organized by Karl Kuchler and Angela Simeon
ABC transporters comprise the largest protein family known to date, as more than 150 different ABC proteins have been identified operating from bacteria to man. The fact that the ABC protein family is the fastest growing protein family overall, and the medical importance of many ABC transporters, prompted us to organize an advanced lecture
To find out more about the scientific programme, registration information and the location, please look up the following URL: http://www.at.embnet.org/molg/abc-meeting/

Netherlands

Free GDB access at the CAOS/CAMM Center

K. Cuelenaere

Besides «generic» biocomputing services, seven EMBnet Nodes also offer genetic services by providing access to «mirrors» of the Baltimore Human Genome Datbase. The CAOS/CAMM Center is of of these seven. With the introduction of <B>GDB 6.0</B>, the Center's access policy to GDB has changed.

In previous versions of GDB, one actually had to have a login on the computer running the GDB ASCII interface and at the CAOS/CAMM Center these login's are not free of charge. Effectively this put an access restriction on our GDB service. With <B>GDB 6.0</B>, the ASCII interface

was dropped and all queries have to be done via the Web through run-time generated HTML forms. This interface is uniform at all GDB nodes. With this access mechanism computer login's have disappeared and (almost) all GDB nodes did not even protect their Web interface with a password. The result was free GDB access on a node of choice.

At our Center, we had replaced the computer login by a Web password. However we soon realized that, while other GDB nodes did not restrict GDB access to their local users, it would not be «healthy» to maintain our restriction. Dutch users, unauthorized at the CAOS/CAMM Center GDB, simply went to other GDB nodes for their information.A nice illustration of the well known phenomenon that you can't sell things that are given for free by others.

On the other hand, free is not for free. To keep up the GDB idea of distributed nodes serving their local community, these nodes have to be financed either by grants (the tax payer) or by the user. Up to recently, the CAOS/CAMM Center financing model for GDB services was a mix of the two with a limited user fee. In the upcoming Dutch GDB node grant renewal discussion, a decision recognizing the fact that users do not pay for a service that is «free» just one click away seems to be inevitable. Even if such a decision implies the fact that it might seriously undermine local knowledge infra structure.In any case, now the GDB Service at the CAOS/CAMM Center is unrestricted too. It can be reached at http://www-gdb.caos.kun.nl/gdb

The first forms available for GDB 6.0> were too complex to use and users sometimes were unable to retrieve available GDB information. This resulted in a drop of GDB usage, not only at the CAOS/CAMM Center, but also at other GDB nodes. However, since the removal of our GDB WWW authorization procedure, a gradual increase of CAOS/CAMM GDB usage can be seen again (see Reports and Statistics at the GDB homepage).

Nevertheless it might still take some time before users, to whom access has previously been rejected, find out that our GDB access is no longer restricted. A mechanism by which users are told where to find the closest GDB mirror, will soon be introduced at all nodes.

The EMBnet Nodes

[AT] EMBNet (martin.grabner@cc.univie.ac.at)
VIENNA BIOCENTRE
University of Vienna, Vienna, Austria

[BE] BEN (rherzog@ulb.ac.be)
Brussels Free Universities,
Rhode-St-Genese, Belgium

[CH] Biocomputing Basel (info@ch.embnet.org)
Biozentrum der Universitaet,
Basel, Switzerland

[CH] ROCHE (doran@embl-heidelberg.de)
Hoffmann-La Roche,
Basel, Switzerland

[CH] SWISSPROT (bairoch@cmu.unige.ch)
Med. Biochem. Dept. CMU, University of Geneva
Geneva, Switzerland

[DE] EMBL (datalib@EMBL-Heidelberg.de)
European MolecularBiologyLaboratory,
Heidelberg, Germany

[DE] GENIUS
(dok419@genius.embnet.dkfz-heidelberg.de)
DKFZ, Heidelberg, Germany

[DE] MIPS (mewes@mips.embnet.org)
Max-Plank-Institut fur Biochemie,
Martinsried, Germany

[DK] BIOBASE (hum@biobase.aau.dk)
BioBase,
Aarhus, Denmark

[ES] CNB (carazo@samba.cnb.uam.es)
Centro national de Biotechnologia
CSIC, Madrid, Spain

[ES] TDI (dopazo@tdi.es)
Technologica para Diagnostico e Investigation
Madrid, Spain

[FI] CSC (Heikki.Lehvaslaiho@csc.fi)
Centre for Scientific Computing,
Espoo, Finland

[FR] CEPH (claude@genethon.fr)
GENETHON,
Evry, France

[FR] INFOBIOGEN (dessen@infobiogen.fr)
INSERM,
Villejuif, France

[GR] EMBnet Node (savakis@myia.imbb.forth.gr)
Institute of Molecular Biology and Biotechnology,
Heraklion, Greece

[HU] EMBnet (remenyi@abc.hu)
Agricultural Biotechnology Centre,
Godollo,Hungary

[IL] INN (isestern@weizmann.weizmann.ac.il)
Weizmann Institute of Science,
Rehovoth, Israel

[IT] CNR (marcella@area.ba.cnr.it)
Consiglio Nationale delle Ricerche,
Bari, Italy

[IT] ICGEB (pongor@genes.icgeb.trieste.it)
International Centre for Genetic Engineering,
Trieste, Italy

[IE] INCBI (atlloyd@acer.gen.tcd.ie)
Irish National Centre for BioInformatics,
Dublin, Ireland

[NL] CAOS (jackl@caos.caos.kun.nl)
Katholieke Universiteit,
Nijmegen, Netherlands

[NO] BiO (linda.akselberg@biotek.uio.no)
Biotechnology Centre of Oslo,
Oslo, Norway

[PL] IBB (piotr@ibbrain.ibb.waw.pl)
Institute of Biochemistry and Biophysics,
Polish Academy of Sciences, Warsaw, Poland

[PR] EMBnet (pfern@gulbenkian.pt)
Instituto Gulbenkian de Ciencia,
Oeiras, Portugal

[SE] EMBnet.se (gad@perrier.embnet.se)
Computing Department, Biomedical Centre,
Uppsala, Sweden

[UK] HGMP (mbishop@hgmp.mrc.ac.uk)
Human Genome Mapping Project Resource Centre,
Hinxton, Cambridge, United Kingdom

[UK] SEQNET (bleasby@daresbury.ac.uk)
Daresbury Laboratory,
Daresbury, United Kingdom

[UK] Sanger Centre (pmr@sanger.ac.uk)
Hinxton Hall
Cambridge, United Kingdom

Vol.3, No.3, 21st December 1996

Dear Reader,
If you have any comments or suggestions regarding this newsletter we wouldbe very glad to hear from you. If you have a tip you feel we can print inthe Tips from the computer room section, please let us know. Submissions forthe BITS section are most welcome, but please remember that we cannot extendspace beyond two pages per article. Please send your contributions to one ofthe editors. You may also submit material by Internet E-mail.

You are invited to contribute to the LETTERS TO THE EDITOR section.

If you had difficulty getting hold of this newsletter, please let us know. We would be only too happy to add your name to our mailing list. This newsletter is also available on-line using any WWW client via the following URLs:

The Online version(ISSN 1023-4152) :

- ***http://www.uk.embnet.org/embnet.news/vol3_2/contents.html***
- ***http://www.be.embnet.org/embnet.news/vol3_2/contents.html***
- ***http://www.no.embnet.org/embnet.news/vol3_2/contents.html***
- ***http://www.ie.embnet.org/embnet.news/vol3_2/contents.html***

A Postscript version (ISSN 1023-4144) is also available. You can get it by anonymous ftp from:

- ***ftp.uk.embnet.org in the directory pub/embnet.news/***
- ***ftp.be.embnet.org in the directory pub/embnet.news/***
- ***ftp.no.embnet.org in the directory pub/embnet.news/***
- ***ftp.ie.embnet.org in the directory pub/embnet.news/***

Publisher:

EMBnet Administration Office.
c/o J.Franklin,
ASFRA BV,
Voorhaven 33,
1135 BLEDAM.
The Netherlands

Editorial Board:

Rob Harper, EBI, Hinxton Hall, UK
(harper@ebi.ac.uk)
FAX +44 (0)1223 494 468
Tel +44 (0)1223 494 429

Robert Herzog, BEN, Free University Bruxelles, BE
(rherzog@ulb.ac.be)
FAX +32-2-6509767
Tel +32-2-6509762

Andrew Lloyd, INCBI, Trinity College Dublin, IE
(atlloyd@acer.gen.tcd.ie)
FAX +353-1-679-8558
Tel +353-1-608-1969

Rodrigo Lopez, EBI, Hinxton Hall, UK
(Rodrigo.Lopez@ebi.ac.uk)
FAX +44 1223 494468
Tel ++44 (0)1223 494423

Peter Rice, The Sanger Centre,Hinxton Hall, UK
pmr@sanger.ac.uk
FAX +44 1223 494919
Tel +44 1223 494967

Back issues online:

UK EMBnet node
Belgian EMBnet node
Irish EMBnet node
Norwegian EMBnet.node

Postscript by ftp:

UK EMBnet node
Belgian EMBnet node
Irish EMBnet node
Norwegian EMBnet.node

embnet.news
Vol.3, No.2, 1996
21 December 1996 ISSN 1023-4144