embnet.news

Volume 1 Nr.1 (ISSN1023-4144) July 1st, 1994

Welcome to the first edition of embnet.news! This news-letter is intended for all researchers who are interested in, or have a need to, work with molecular biology data on computers. This news letter naturally focuses on, but does not limit its scope to, reports and new achievements from the European Molecular Biology network (EMBnet). EMBnet is not parochial and its members, like all scientists, collaborate worldwide. But what is EMBnet? Here is some background.

EMBnet, established in 1988, is a group of European sites (nodes) which provide computational molecular biology services to both national and international researchers. Each country has a `national' node supplying internal bioinformatics services but there are also `special' nodes providing specialist functions, genome mapping projects are one example.

EMBnet sites are typically mandated by their national governments to operate as centres of excellence and are dedicated to the management and supply of the molecular biology data and software to their local, European and global communities. EMBnet is also very active in training researchers, pure research and developing methods for dissemination of bioinformatics by the most effective means e.g. the production of user-interfaces and network protocols. EMBnet flowered under EC BRIDGE funding and has recently adopted the structure of a non-profit, international association registered under Dutch law as a "Stichting". EMBnet has worked so well it is now an organisation and continues to grow.

Contents

Editorial - Welcome to embnet.news	1
BITS - Fundamentals of Database Searching Methods	2
Node Focus - The ICEGB in Trieste	3
Interview-Net	4
Databases & Programs - The Volume Database at the CNB	5
Tips from the computer room	6
Node News	7
Conferences	8
EMBnet nodes	8
Dear Reader	9

What are some achievements of EMBnet? The setup of a communications network whereby all national EMBnet researchers know their site is the contact point for international advice. We have set up a fault-tolerant NetNews infrastructure and fully functional network information servers with a common hierarchy within the GOPHER system. Future issues of embnet.news will report on some of the more advanced networking achievements, such as sequence updating and remote sequence searching.

Research is a strength of many EMBnet nodes and this issue describes a new database produced by the Spanish EMBnet node. Training is particularly important; the Belgian EMBnet node (BEN) has contributed an article on sequence searching. Each issue of embnet.news will highlight an existing EMBnet member site, the ICGEB at Trieste is featured in this publication.

A metric of the success of EMBnet is that the membership continues to grow. An interview with Ireland is presented, a node applying to join EMBnet. Other regular features include conference and training announcements and, as no one has yet programmed the perfect interface, there is a corner for 'tips from the computer room' to help biologists get to grips with computers.

Embnet.news is available electronically on 'WWW servers', and as postscript files which you can print locally. It is a platform of information and reports, hopefully both useful and interesting, for researchers using computers in molecular biology. We welcome feedback on this, our first issue, and also submissions for inclusion in future publications (e-mail address emb-pub@dl.ac.uk). Let's make communication a success!

The embnet.news editorial board

Alan Bleasby (SEQNET, UK)
Reinhard Doelz (Biozentrum, Switzerland)
Robert Herzog (BEN, Belgium)
Rodrigo Lopez (BiO, Norway)

BITS
BioInformatics Theory Section
FUNDAMENTALS OF DATABASE SIMILARITY SEARCHING METHODS

Dr. Guy Bottu: Belgian Node (BEN)

If you've ever wondered, for example, how the FASTA database similarity searching program works, or even whether it is the best program for your problem, then this article will help. Knowing how a program works is beneficial in determining whether it is suitable for a particular task. This article, though not comprehensive, covers the essentials of databank searching, multiple alignment and threading.

A simple heuristic is that any program which finds sequences in a databank that are similar to a query sequence usually computes a similarity score between them. The program then picks out the hits that yield the highest score. For databank searching the two main methods are

a) Global alignment: an attempt to align every residue.
b) Local alignment: report only sections of good alignment

Each of these methods can be either 'rigorous' or 'fast'. The rigorous methods generally require special computers and are usually accessed via network services. The fast methods, which are available on all EMBnet nodes, are suitable for the vast majority of applications. With all of these methods it must be borne in mind that:
- the results are very much dependent on the scoring scheme. The choice of gap penalty is a particularly difficult problem.
- an ordinary sequence alignment may not be appropriate. If you have additional, for example structural, information then use it (see later).

Network services

As sequences are of different lengths, trying to do a global alignment between the query sequence and the sequences from the databank can be of limited utility. Therefore most databank searching programs use local alignment methods. The Smith and Waterman best local alignment algorithm (SW) is a very good general-purpose method for databank searching.

The Blitz E-mail Server in Heidelberg offers an SW search of a protein against the Swissprot database. The program runs on a MasPar parallel computer with 4096 processors. A weak point of Blitz is that it gives a constant penalty for each gap (a 'b.n' approach), whereas a gap-opening plus gap-extension approach ('a+b.n') is usually preferable.

An experimental E-mail Server is running at the Weizmann Institute in Rehovot (the EMBnet node in Israel). It supplies an SW search (and also a GCG profilesearch, see below) against various databanks using a 'bioccelerator'; a special purpose computer which can be programmed to facilitate sequence comparisons.

Local services for fast databank searching

Fast databank searching programs have been developed. They all begin by making a 'hash table' which contains the position in the query sequence of all possible 'words' of 'n' residues; the same is done for each database sequence. It is therefore easy for the programs to find matching 'words'. What happens next depends on the program:
- The algorithm of Wilbur and Lipman makes best local alignments whilst restricting the computation to the neighborhood of where most matching words have been found.
- The GCG program wordsearch counts matching words within a chosen range of frames of alignment. An optional 'mask' can be supplied too for 'words' containing positions which are allowed to mismatch e.g. for searching coding DNA against a databank, the mask '++-++-++' can be useful to ignore the wobble base. Wordsearch is of limited applicability unless sequences are gapless and very similar.
- The FASTA program, as supplied in the GCG package, assembles words in contiguous stretches that are scored with a symbol comparison table. Using a joining penalty it assembles stretches that are compatible with an alignment. The FASTA program supplied by the original author (William Pearson) optimizes alignments; a restricted best local alignment is made against the whole databank, not just for the sequences which had the highest score after joining.
- The BLAST suite developed by Altschul, Gish et al. extends words in order to obtain 'high scoring segment pairs'. Its strength lies in that, at least for proteins (blast3), it matches the words of the query sequence with 'words' in at least two other database sequences thereby providing more certainty of a hit against well-characterized proteins.

Multiple alignments and database searching

Once you have several sequences you know to belong to the same family, you can make a multiple sequence alignment (for example using the GCG program Pileup) and derive from it a 'profile'. This is essentially a sequence represented by tables for each residue position. Each table defines match/mismatch scores and gap penalties. Again, the computation of the optimal gap penalty is a problem. In order to identify new members of the family, a profile can be searched against a sequence databank or a sequence can be searched against a profile databank (using an extension of the Smith and Waterman algorithm). The advantage of a profile is that the parts of the multiple sequence alignment

where the sequences are dissimilar have a low score in the profile and thus contribute only marginally to the SW alignment score. Profilesearch (GCG package) is an example of this type of program.

Using protein structural information

Related protein structures frequently share a common 'core'. The core can consist of, for example, connected stretches of alpha-helix or beta-pleated sheet structures. The number of amino acids in the core can be remarkably conserved, even if the sequences themselves have diverged beyond recognition. The core sequence segments may be connected by 'loops' with a very variable length and structure. A modest way to make use of such knowledge is to search a protein with known structure against a databank and use a greater gap penalty for the core regions than for the loop regions. Once the structure of a number of proteins belonging to the same family has been determined, and closely related proteins have been aligned with them, it is possible to define a 'core model'.

Aligning a sequence with a core model is called 'threading'. You can imagine the core as a set of tubes and the sequence as a wire that is passed through the tubes. A core model has, for each position, a score for each amino acid and, for each pair of positions, a score for each pair of amino acids (e.g. for some position pairs there is a very high score for pairs of amino acids with opposite charge and a low score for pairs of amino acids with identical charge). A core model can then be searched against a databank of sequences and a sequence can be searched against a databank of core models. Programs using this method will soon be widely available.

NODE FOCUS

ICGEBnet: A United Nations computer resource for biotechnology

International Centre for Genetic Engineering and Biotechnology
Padriciano 99, 34012 Trieste, Italy

ICGEB: The International Centre for Genetic Engineering and Biotechnology

Dr. Sandor Pongor, Valeria Bevilacqua and Zsolt Hatsagi

ICGEB is an international organisation of the United Nations, established in order to promote the safe use of biotechnology with special regard to the needs of the developing world. ICGEB aims to be a research and training Centre of Excellence for its Member Countries. ICGEB has two main laboratories, one in Trieste, Italy (where the Direction of the Centre is also located), and one in New Delhi, India. In addition, ICGEB coordinates a network of 17 Affiliated Centres (national laboratories in Member Countries) whose research activity is partially funded by ICGEB.

ICGEB member countries are: Afghanistan, Algeria, Argentina, Bhutan, Bolivia, Brazil, Bulgaria, Chile, China, Colombia, Congo, Costa Rica, Croatia, Cuba, Ecuador, Egypt, Greece, Hungary, India, Indonesia, Iran, Iraq, Italy, Kuwait, Mauritania, Mauritius, Mexico, Morocco, Nigeria, Pakistan, Panama, Peru, Poland, Russia, Senegal, Sri Lanka, Sudan, Syria, Thailand, Trinidad & Tobago, Tunisia, Turkey, Venezuela, Vietnam, Yugoslavia and Zaire.

ICGEB is a node in EMBnet providing specialized services.

ICGEBnet

ICGEBnet is a central biocomputing resource located in Trieste that currently provides login facilities to over 600 users world-wide via INTERNET and X.25 connections. ICGEBnet provides a computer environment that allows molecular biologists to analyze nucleotide and protein sequences. ICGEBnet provides access to a large variety of databases (biosafety, genetics, biodiversity, etc.) and to various tools of electronic communication (bulletin boards, electronic mail and wide area information services). Access to ICGEBnet is available free of charge to all ICGEB Member Country scientists; however, preference is given to those scientists whose research is directly related to the research goals of ICGEB.

Molecular biology services

ICGEB hosts copies of the most important biological sequence databanks, including GenBank, EMBL, PIR, Swiss-Prot and Prosite. In addition, ICGEB provides access to virtually all molecular biology databases available in the EMBnet and world-wide communities via the INTERNET.

Analysis software includes three major program packages for biological sequence analysis (GCG, Staden, IG) and a large variety of specialized software. In addition, ICGEBnet provides on-line access to all computer program descriptions as well as a complete user manual. The programs of ICGEB courses, fellowships and application forms are also available on-line.

Information services, biosafety

ICGEBnet provides on-line access to a large number of databases pertinent to genetics, molecular biology and biotechnology. Special emphasis is given to biosafety and the release into the environment of genetically modified organisms. ICGEBnet collects documents and guidelines on biological and chemical laboratory safety, lists of experts, biosafety committees etc. Using the gopher wide area information server, access is provided to

specialized data collections on biodiversity, microbiology, cell culture collections etc. The aim of this information service, the first such service worldwide, is to assist national authorities, academic institutions and industry in ICGEB Member Countries in the safe evaluation of the environmental effects of genetically modified organisms, and to help in the creation of national regulations.

Electronic communication and other services

ICGEBnet provides electronic mail services (mm, elm), access to bulletin boards (nn), and wide area information servers. While being increasingly popular in industrialized countries, these communication tools are rarely available in the developing world. Presently ICGEBnet appears to be one of the main INTERNET gateways for biologists in ICGEB Member Countries.

Research and training activities

Research at ICGEBnet concentrates on computer methods for detecting distant protein homologies and the maintenance of SBASE, a comprehensive and annotated collection of protein domain sequences. SBASE is available through gopher, anonymous ftp (ftp.icgeb.trieste.it) and www (http://www.icgeb.treiste.it/). BLAST searches of SBASE are available through the automated e-mail servers sbase@icgeb.trieste.it and domain@hubi.abc.hu. UNIX biocomputing utility programmes (menu interfaces, database update, sequence retrieval, etc.) are developed in-house and are available through the anonymous ftp facility.

ICGEBnet also serves as a training facility with 2-3 yearly computer courses. The practical course 'Computer Methods in Molecular Biology' is held in July each year. This course provides an introduction to bioinformatics and biological sequence analysis. EMBnet has provided support to several of ICGEB's computer courses.

User support

User consultation is available via telephone and electronic mail in addition to on-line and hard copy documentation of the major programs. User training is provided through the computer courses organised at ICGEB.

ICGEBnet Biosafety Archives*

1. Laboratory Chemical and Biological Safety (general)
   - Material Safety Data Sheets
   - US Institutional Biosafety Committees
   - Collection of documents, institutional regulations
2. Environmental Biosafety
   - Genetically Modified Organisms
   - US Regulations
   - US GMO Permit Applications Forms
   - US Field Test Permit Approvals
   - ICGEB Documents
   - European Documents
   - Useful Infos and Contacts (Legislation, Experts)
3. Biodiversity Issues
   - UNCED Documents
   - Biodiversity gopher servers

*available through the gopher server and also partly through e-mail(docserver@icgeb.trieste.it).

For Further info contact: Valeria Bevilacqua,
Systems Manager, ICGEB (valeria@icgeb.trieste.it)
Tel: +39-40-3757330
FAX: +39-40-226555
Sandor Pongor Ph.D. Head, Protein Structure and Function Group and Computer Services, (pongor@icgeb.trieste.it)

INTERviewNET

An EMBnet telephone interview

Alan Bleasby talks to Andrew T. Lloyd

In this issue we have contacted Ireland, a country soon destined to become an EMBnet node. This interview demonstrates the collaborative nature of the European molecular biology community and why European countries want to be an EMBnet node.

Q: Hello, who am I speaking to?
A: Andrew T. Lloyd, Director of the Irish National Centre for BioInformatics (INCBI)

Q: Who supports INCBI?
A: The Irish National Centre for BioInformatics (INCBI) has been operating informally since 1986 to provide a service to Irish molecular biologists. Since April 1994, INCBI has been funded by FORBAIRT, the Irish Science and Technology Agency. INCBI has applied to be the Irish EMBnet node. I have houseroom on the top floor of the Genetics Department at Trinity College Dublin.

Q: What kind of equipment do you have?
A: Our central processor is a DEC AXP 3000 600 workstation with 96MB ram and 12Gb of disk, Internet address: acer.tcd.gen.ie. I sit, bathed in the aura of its 19" screen, for much of the day. I have a museum of terminals, including a new (today!) LC475, a suitcase sized IBM-PS2 (it's actually bigger than our DEC alpha), a PC-386 and a couple of PC-ATs.

Q: How many users do you have?
A: For the last 7 years, some 60 user accounts have been open on an ageing VAX. For reasons beyond my control, accounts external to Trinity College are multi-user which is a security nightmare. We would expect to have perhaps 200 user accounts (from Graduate students through to Heads of Department) set up in 1994. These will come primarily

from the seven Universities, but we also have interest from hospitals, regional technical colleges and industry.

Q: Can you give examples of typical projects?
A: My base, the Genetics Department in Trinity College, shows something of the range of projects which INCBI is having to service: Genome (yeast, arabidopsis) sequencing projects. Mapping disease loci and beginning to move into gene therapy. Phylogenetic trees from mammalian cytokines, mitochondrial D-loops, microsatellites and prokaryotic structural genes. Classical, structure and function, molecular biology of prokaryotes, plants and mammals. Molecular evolution studies are strongly represented. Other Irish users come from pretty much every sort of biological discipline: medicine, food science, biochemistry, microbiology. INCBI provides software, collaboration and advice.

Q: As a prospective EMBnet node, what do you hope for?
A: The benefits look to be enormous; particularly from the viewpoint of a one-man-band who must learn UNIX, become a system administrator, get up to date with and install the latest BioInformatics software, write documentation for users, offer workshops, seminars and lectures on demand and be at the end of telephone and email help lines. It is nice to know that there are folk out there who have been through all this before and are willing to help out. The prospect of getting funding from EMBnet to facilitate communication, travel and setting-up of software is icing on the support cake. It will also be of interest to our users to know that there are many EMBnet-supported training programs. I am impressed with the freedom with which highly sophisticated software has been developed and made FREELY available throughout the community (i.e. not only through EMBnet).

Q: What can you offer to EMBnet?
A: While it may seem, at least in the early days, that we are all take and no give, Ireland has a positive contribution to make in the world of BioInformatics. We would hope to play a part in teaching and have plans to offer an MSc course in the next academic year. We might claim to be a centre of excellence in codon usage analysis. A lot of our work requires the development of non-redundant databases of, say, coding sequences for particular species. My personal contribution to BioInformatics software now sits on EMBnet file servers, (where it is FREELY available as is right and proper!): its called CODONS. As the Alpha is still a minority machine in the molecular biological world, we have a strong interest in promoting portability to that environment.

Q. Well, many thanks Andrew, I guess we'll see you at the Genes, Proteins and Computers conference in Chester?
A. It's the one conference that I've been to since the beginning, wouldn't miss it!

DATABASE AND PROGRAM DEVELOPMENT

The Microscope Volume Data Base at EMBnet/CNB, Spain

Dr. Jose-Maria Carazo

The need for a 3D structural data base:

We address in this development-note the problem of information access that arises in the field of three dimensional structural determination by means of image processing of data obtained by microscopy.

A substantial number of newly determined structures in different resolution ranges are starting to appear in the literature, and it is essential to allow the structural biology community full access to these data. This situation is not much different from the one that existed before in the fields of sequence analysis and X-ray crystallography, which led to the establishment of several scientific publicly accessible databases. In this work a prototype of a distributed database containing three dimensional structural information is presented. This type of structural information is cross linked, when applicable, to other sources of information, such as sequence data, atomic coordinates and bibliography.

Data Organisation

The first point to be considered when discussing possible ways of organizing information is the precise informational content of the data. Although the final result of all the analytical approaches mentioned above is a volume, there are other important pieces of descriptive information that pertain to the experiment. We are referring to a detailed description of the conditions under which the volume was obtained, as well as any pointers that allow the 3D volume to be linked to other data sources. In our case, that complementary information can be expressed in a textual form, and we have therefore decided to split each data set into two files, one containing the textual information and the other the volume data.

As to the textual information, we have organised it into keyed fields. In this way it is possible to formulate complex queries to the database by relating the information in the different fields. Additionally, and inan effort to integrate information from different sources, explicit links to other databases such as EMBL or SWISSPROT have been provided.

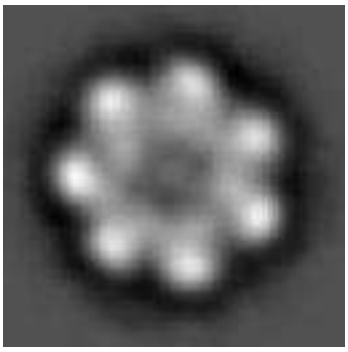
Data Distribution

Considering the large number of 3D volume results obtained by the different kinds of microscopy in so many different areas, we have set means to allow for a distributed organisation of the data. Our standpoint was that the data would finally be organised by a relatively small number of stable "collection centres" with expertise in different areas of application, which might be located anywhere in the world, but linked together via Internet.

The data is accessed under a client/server architecture, using the WWWprotocol. All the programs needed by the client laboratories (browsers, viewers, search utilities, etc.) are public domain.

An Example Session

During a query session, the sequence of events is as follows:-The "client" laboratory runs a X windows based WWW browser, such as Mosaic.
- A keyed query is assembled and passed to a searching engine.
- The search is performed over the textual part of the data base and a document containing links to images, volumes, anonymous sites (etc) is created on the fly and visualized on the screen.



CH60 ECOLI Chaperonin obtained from the httpd server at the CNB in Spain.

A prototype configuration of two servers has been set up between Madrid (Centro Nacional de Biotecnologia, Spanish EMBnet node) and Albany (Wadsworth Center for Laboratories and Research, New York). Please feel free to connect to our server at:
http://indy.cnb.uam.es/Base/home_page.html

We hope that this prototype EMBnet project will be the embryo of a new biological data base of volume information obtained from the microscopy. Certainly these new data would complement the detailed biological information already contained in other data bases such as EMBL or SWISSPROT. Comments and suggestions are most welcome.

Acknowledgments

Supported, in part, by the Spanish DGICYT Plan General de Promocion del Conocimiento grant number PB91-0910 and by the European Molecular Biology Network (EMBnet, European Union BRIDGE project number PL890235/A2A).

CompuTips

Helpful advice from the computer room

This month's tips demonstrate some uses of the UNIX '**find**' command. This command allows you to search for files and, once found, perform specific actions. It also allows searching with different criteria.

1. The exact name of the file for which you want to search is known. To find the file 'data.dat' in the partition 'docs', use the command:

```
find /docs -name data.dat -print
```

Note that the expression -print is needed. If it is omitted, find will indeed detect the file, but won't bother telling you where it is!

2. You are looking for files whose names contain a well known string of characters. To find all files in the partition '/docs' which contain the string 'data' as an extension of their name. Use the command

```
find /docs -name "*data" -print
```

Note that the double quotes are essential for find to properly interpret your intention.

3. You want to take a specific action on a set of files found by the find command. Assume you want to delete all data files having the ".dat" extension from your current '.' directory and all your subdirectories. Use a command of the form

```
find . -name '*.dat' -exec rm {} \;
```

Note that you have to "escape" the final semicolon with the \ character for the expression to be properly interpreted. Also, be careful with 'rm'!

4. To delete all files from your area which have not been modified for 7 days use

```
find . -mtime +7 -exec rm {} \;
```

A final note concerning the find command: Use it only in your own file area. It is very computationally expensive when used to search very large file systems, like those found on most EMBnet computers.

NodeNews

This section reports some of the recent highlights at a cross-section of EMBnet nodes. Software, database, hardware, network and research developments are included.

Austrian EMBnet node
Hardware: 4 Gbyte diskstorage on DEC system 5900 added.
Software: Egcg, Xgcg (CAOS/CAMM Extensions), SRS.

Belgian EMBnet node (BEN)
Software: SRS installed on both SUN & DEC platforms. `Current Sequence Awareness' installed; our automatic system to make users aware of update entries.

Finnish EMBnet node (CSC)
WorldWideWeb: Conversion of the EMBL Software files listing to HTML:
http://shamrock.csc.fi/embldos.html
http://shamrock.csc.fi/emblmac.html
http://shamrock.csc.fi/emblvax.html
http://shamrock.csc.fi/emblunix.html
allowing reading and downloading brief descriptions of the files.
Converted Una Smith's Bio Faq to hypermail.
Converted embnet.general archive to hypermail.
Available from CSC BioBox main menu:
http://shamrock.csc.fi/brochure.html

ICGEBnet EMBnet node
Network: e-mail server sbase@icgeb.trieste.it: BLAST searches of the SBASE protein domain library e-mail server docserver@icgeb.trieste.it: Documents of the ICGEB Biosafety archives (Biological safety, genetically modified organisms)

WorldWideWeb: showing basic information about ICGEB, access to EMBnet and bionet groups, access to the SBASE protein domain library. Retrieval of entries by fields and links to locally mounted databases (e.g. BLOCKS7.01, PRODOM24 & PRINTS4.0) and remote databases (e.g. EMBL, MIM, MEDLINE, Swiss-Prot & MEDLARS). Submission of a protein sequence for BLAST searches against SBASE is incorporated.
http://base.icgeb.trieste.it/ or
http://www.icgeb.trieste.it/

Netherlands EMBnet node (CAOS/CAMM)
Hardware: A Bioccelerator, 4 processor and 32Mb DRAM, to boost database search speeds. This is big enough to hold the full SWISSPIR or OWL in memory.
Software: The EMBnet GMS (general menu system) has now reached such a mature state that it has been released to our users. It seems like they like it!
The switch from the VMS environment to UNIX (IRIX) is currently in progress.
WorldWideWeb: A CAOS/CAMM WWW-server is available with the URL: http://www.caos.kun.nl/

Norwegian EMBnet node (BiO)
Hardware: New R4600 Indigo and Upgrade of the Challenge to R4400.
Software: Sequence Analysis: Testing GCG 8.0, DCSE, alscript installed.
Molecular Visualisation lab: Setor, Whatif, xtalview and tops installed.
Phylogeny Lab: PAUP, Phylip and ODEN installed.

Databases: PDB, HSSP and DSSP installed.

Spanish EMBnet node (CNB)
Databases: A prototype 3D Imaging Volume Database which is now operational (presented as an article in this issue of embnet.news).
Software: A package for the classification of Kohonen Neural Networks as an addition to our X-based Microscopy Image Processing package (XMIPP) running on SGI, IBM and Sun platforms. A package designed to work within the framework provided by the PHYLIP programs. This allows testing of evolutionary models. A self-extracting executable, ABLE-DOS.EXE is available. New population & ecology software for calculating the intricsic rate of natural increase (rm) and its error and confidence intervals using a friendly pop-up menu interface. All this software can be anonymously ftp'd from: ftp.cnb.uam.es

ABLE-DOS.EXE	DOS	-
/SOFTWARE/MOLEVOL	Self-extracting file.	
RM-DOS.EXE	DOS	-
/SOFTWARE/OTHER	Self-extracting file.	
xmipp_ko_tar.Z	UNIX	-
/SOFTWARE/IMAGE_PROCESSING/XMIPP		

Swiss EMBnet node (Biozentrum)
Software: The remote sequence searching software package 'Hierarchical Access System for Sequence Libraries in Europe' (HASSLE) was released in June 1994 (version 4.1). Network communication: A virtual reality conference at the BioMoo located at the Weizmann Institute, Rehovot (Israel) introduced the working principle of HASSLE. Fifteen attendees commented on various aspects of accounting, security and implementation. The meeting place was the 'EMBnet Cafe'. It was introduced by R.Doelz at the Moo with the help of the node staff there.

United Kingdom EMBnet node (SEQNET)
Hardware: Two DEC 3000 AXP, running OSF.
Software: Egcg, babel, tops, dcse.
Databases: Yeast chromosome 11.

Conferences

Genes, Proteins and Computers III: An international conference on bioinformatics, networking and computing in molecular biology. Chester 7-9th September 1994.

This conference is the third of a biannual series of International conferences. Topics covered in the sessions include large scale mapping, protein threading, membranes, database accuracy/connectivity and accessing remote services.

Contributions are invited for poster presentation. The best four submissions will be invited for oral presentation. There is a conference fee of 100 pounds. A limited number of bursaries are available for non-UK European students.The conference schedule summary appears below.

Wednesday 7th September.
Chair: Peter Murray-Rust [Glaxo]

10:00 - 12:30 Registration
12:30 - 14:00 Lunch
14:00 - 14:45 Tom Blundell [London]
Protein structure
14:45 - 15:30 Director of EBI [Cambridge]
15:30 - 16:15 Patricia Rodriguez-Tome [Genethon]
Mapping
16:15 - 17:00 Steve Oliver [UMIST]
Chromosomes
Evening: assembly of posters, mixer+commercial exhibition

Thursday 8th September. Theme: structure
Chair: Janet Thornton [UCL]

09:00 - 09:45 Steve Benner [ETH]
Prediction
09:45 - 10:30 David Jones [UCL]
Threading
10:30 - 11:00 Coffee/Tea
11:00 - 11:45 Willie Taylor [NIMR]
Membranes
11:45 - 12:30 Jeff Skolnik [SCRIPPS]
Lattice
12:30 - 14:00 Lunch
14:00 - 15:30 4 20-minute poster contribution talks
15:30 - 16:00 Coffee/tea
16:00 - 17:30 Poster session (manned)

Evening: conference meal & after dinner speaker

Friday 9th September. Theme: Connectivity
Chair: Alan Bleasby (SEQNET)

09:00 - 09:45 Peter Rice [EMBL]
DB accuracy
09:45 - 10:30 Rob Harper [Finland]
Remote services
10:30 - 11:00 Coffee/Tea
11:00 - 11:45 Thure Eztold [EMBL]
DB links & SRS
11:45 - 12:30 Joel Sussman [PDB]
PDB

Details of registration can be found via anonymous ftp (s-ind2.dl.ac.uk pub/events/gpcIII), via World-Wide Web (http://www.dl.ac.uk) or by contacting the SAS conference office, DRAL Daresbury Laboratory, Warrington WA4 4AD, UK.

THE EMBnet NODES

AT EMBnet (grabner@embdec.bcc.univie.ac.at)
Bio Computing Centre
University of Vienna, Vienna, Austria

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

CH EMBnet (embnet@comp.bioz.unibas.ch)
Biozentrum der Universitaet Basel, Basel,
Switzerland

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

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

DE EMBL (datalib@EMBL-Heidelberg.de)
European Molecular Biology Laboratory,
Heidelberg, Germany

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

DE MIPS (mewes@ehpmic.mips.biochem.mpg.de)
Max-Plank-Institut fur Biochemie,
Martinsried, Germany

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

ES CNB (carazo@cnbvx3.cnb.uam.es)
Centro national de Biotechnologia
CSIC, Madrid, Spain

FI CSC (harper@convex.csc.fi)
Centre for Scientific Computing,
Espoo, Finland

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

FR CNRS-INSERM (dessen@genome.vjf.inserm.fr)
INSERM, Villejuif, France

GB HGMP (mbishop@crc.ac.uk)
HGMP Resource Centre,
Harrow, UnitedKingdom

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 (lsestern@weizmann.weizmann.ac.il)
Weizmann Institute of Science,
Rehovoth, Israel

IT CNR (attimonelli@mvx36.csata.it)
Consiglio Nationale delle Ricerche,
Bari, Italy

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

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

NO BiO (rodrigol@biotek.uio.no)
Biotechnology Centre of Oslo,
Oslo, Norway

PR EMBnet (pfern@gulbekian.pt)
Instituto Gulbenkian de Ciencia,
Oeiras, Portugal

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

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

Dear Reader,

If you have any comments or suggestions regarding this newsletter we would be very glad to hear from you. If you have a tip you feel we can print in the 'Tips from the computer room' section, please let us know. Submissions for the BITS section are most welcome, but please remember that we cannot extend space beyond two pages per article. Please send your contribution to one of the editors. You may also submit material by Internet E-mail to:

emb-pub@dl.ac.uk

If you had difficulty getting hold of this newletter, 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:

http://www.embnet.unibas.ch/embnet.news/info.html
http://biomaster.uio.no/embnet.news/info.html
ftp://s-ind2.dl.ac.uk/pub/embnet/embnet.news/
ftp://dbmdec5.ulb.ac.be/pub/embnet.news/

or via *anonymous* ftp from:

s-ind2.dl.ac.uk	**directory pub/embnet.news**
dbmdec5.ulb.ac.be	**directory pub/embnet.news**

Acknowledgements:
This publication was made possible with funds from the EU BRIDGE contract.

Title: embnet.news
Publisher: EMBnet
Administration Office c/o J. Franklin, ASFRA BV,
Voorhaven 33, 1135 BL EDAM, The Netherlands
Published: Quarterly, by the EMBnet nodes
Editorial Board:
Alan Bleasby (SERC Daresbury, UK)
(bleasby@daresbury.ac.uk)
Fax: +44-925-603100
Tel: +44-925-603351
Reinhard Doelz (Basel University, CH)
(doelz@urz.unibas.ch)
Fax: +41-61-2672078
Tel: +41-61-2672247
Robert Herzog (Free University of Bruxelles, BE)
(rherzog@ulb.ac.be)
Fax: +32-2-6509767
Tel: +32-2-6509762
Rodrigo Lopez (BiO, N)
(rodrigol@biotek.uio.no)
Tel: +47-22958756
Fax: +47-22694130

embnet.news
ISSN 1023-4144