embnet.news

Volume 2 Nr.4 (ISSN1023-4144)

December 22, 1995

Following the production of the EMBnet brochure and the activities associated with the EMBnet AGM it has been back to business as usual. So, what fine articles are in this issue? We have all the regular features including TIPS, announcements and news.

The BITS section this issue deals with mechanisms of data encryption. The Software Development section details a tcl interface to SRS. A thread which frequently crops up in the articles is the Internet. Our INTERviewNET presents a scientists-eye view of problems and potential pitfalls of this technology.

In order to facilitate discussion of articles in this, and previous, issues of embnet.news we would like to create a LETTERS TO THE EDITORS section in forthcoming releases. To get this off the ground we will, of course, need your support. Letters can discuss points raised, areas of interest or concern in bioinformatics and eventually replies to postings by other readers. Letters should be submitted to the usual editors address (emb-pub@dl.ac.uk).

Coming soon: To whet your appetites for the next issue there has been animated and productive discussion of database redundancy and accuracy by EMBnet and we will be composing an article concerning this area. In the meantime...

To keep you amused over the festive season you may like to try our quizzes. We wish you all a very happy Christmas and prosperity in the new year.

The embnet.news editorial team

Alan Bleasby	UK
Reinhard Doelz	CH
Robert Herzog	BE
Andrew Lloyd	IE
Rodrigo Lopez	NO

Contents

Editorial	1
BITS - Messaging and Privacy	1
Tips from the computer room - The csh "foreach" command	3
Node Focus - Brussels	4
INTERviewNET - Markus Sadeniemi - FUNET	5
Software Development - A Tcl based SRS	6
Node News	8
Conferences and Announcements	9
Christmas Competitions	11
The EMBnet Nodes	12
embnet.news information	13

BITS

Bioinformatics Theory Section

Reinhard Doelz and Nicole Redaschi, Swiss EMBnet node

Messaging and Privacy

Introduction

Messages may be printed, written or read in different ways. Some are meant for a group of people, some are sent by an individual to another individual, and some have to be read by as many people as possible. The Bible, for example, has been translated, painted, printed; in short communicated in all possible ways in order to spread the message as widely as possible. Refer to the electronic documentation http://keaggy.cc.cmu.edu/bible/index.html for details.

Other messages are of no interest to the public, or may even harm individuals if access to this 'secret' information becomes public. Love letters, military documents, commercial enterprise memos all need to be transmitted in a way in which the message stays confidential. Symbolic messages require intimate knowledge of your correspondent, They have been used frequently in religions and rites, such as the famous 'voodoo.' (http://www.nando.net/prof/caribe/voodoo.html.) Occasionally, writing a message with 'secret ink' will make it invisible, and only those who know how to reproduce the text by chemical or thermal treatment will be able to read its context. Religious groups were forced to set up all types of rituals in order to gain as much freedom in society as possible. As an example, refer to the story of the Jesuits as described in various historical articles (http://freethought.tamu.edu/freethought/joseph_mccabe/religious_controversy/chapter_29.html)

Alternatively, secret messages might be readable for anybody, but the transmission pathway is selected in a way that security can be guaranteed. As history shows, the assumption that such documents are really transmitted without being witnessed or exposed to the enemy is crucial and examples are known where the secret was broken purposely. Refer to the recent exposure of United Kingdom Royal Family life in the public press for the consequences of such confidentiality omissions. (http://www.spb.su/sppress/77/diana.html)

Electronic mail messages and other documents which are transmitted via public telephone lines, such as the internet,

should be considered as 'public' even if addressed to individuals, since 'sniffing' and other types of observational activities can easily reveal the contents of internet messages. This article in embnet.news focusses on methods concerning how to keep electronic messages private. We use the term 'encryption' to describe the method to make a text unreadable by third parties, but readable by all those who have a 'key' to revert the encryption process.

Classical encryption

Romances, religious needs and military applications were the driving forces for encryption in past centuries. Most elegantly, the 'key' which was used to render messages unreadable was education. Many people could not read or write, and most people could not understand Latin, the language used by the clergy. However, these methods were only applicable if groups of people were addressed. In a person-to-person communication, several approaches may be used to render a text unreadable to those who do not know the 'secret' of the text. Such a convention might be that only every third word is used (*Sir Arthur Conan Doyle, The Gloria Scott*, July-Sept 1874, http://ftp.std.com/obi/Arthur.Conan.Doyle/Memoirs/gloria.scott.txt.Z), or the introduction of apparently irrelevant symbols which resemble characters of the alphabet (*Sir Arthur Conan Doyle, The Dancing Men*, July-August 1898 http://ftp.std.com/obi/Arthur.Conan.Doyle/Return/dancing.men.txt.Z).The advantage of these methods is that no sophisticated equipment is required, and the 'key' is either a description of the system as such or a small table which explains the use of the symbols. The knowledge of how to 'decrypt' the secret message must be transmitted securely. However, linguistic and combinatorial methods allow one to 'guess' the solution of this kind of encryption. This is possible because letters in an alphabet show a typical frequency of distribution, and the analysis of the observed frequency of words or characters permits outsiders to break the key and read the message. This discipline of deciphering messages is known as 'cryptoanalysis'.

Messages have been encrypted more sophisticatedly using machines or even cmputers. The famous 'enigma' machine used in World War II worked with aset of cylinders and electrical wires in a special arrangement. The correct sequence of cylinders had to be built into the machine in order to encrypt or decrypt messages. Without these cylinders, the knowledge of the how to reassemble a message was useless. On the other hand, once the enemy had access to the physical machine, the secret of the code was soon discovered. More information on aspects of encryption are available, as an example see http://www.cityscape.co.uk/channel4/big_bytes/cybersecrecy/cyber040.html

In order to apply sophisticated encryption to biological applications we face the problem that the DNA alphabet is repetitive. It has to be avoided that the deciphering of messages by 'brute force' can be achieved more rapidly than mathematically expected because the outcome has an expected shape. More intuitively, the 'magic images' use this technology by 'encrypting' an image in dots, and the human brain is sophisticated enough to generate a three-dimensional image.

Public key cryptography

The disadvantage of the classical encryption is the use of a secret key which must be known to both parties. Electronic networks do not easily allow the exchange of these messages, and the use of specific hardware is impossible due to resource and legal restrictions. Confidential real-time traffic on the internet will, unfortunately, need to use classical methods to encrypt our messages for the sake of performance. This is achieved on World Wide Web applications as described below. To exchange private memos over the internet, however, we can use electronic mail, which need not necessarily operate in real time and also will allow the addition of modules which handle the encryption. The exchange of the key would, to be convenient, have to be achieved using insecure methods, if it can be assured that the resulting encrypted message is of no use to anybody other than the recipient. The recipient must have the sole means to decrypt the message. If the message is subversively misdirected, or broadcasted otherwise, it must not be possible for an accidental recipient to read the message, as the method to decrypt the message must be unknown. The basic theorem is to use a 'private key' (which is invisible to others, and only known to a single individual, i.e. the recipient) and a 'public key' which will serve for any sender. The sender will use the 'public key' to make the message to be sent readable only to the individual who owns the second, 'private key'. As no one but a single individual can access this 'private key', the 'public key' can be distributed widely and privacy can be assured.

The favourite program today for 'public key' encryption in electronic mail sent via the Internet is 'pgp' which stands for 'pretty good privacy', and was written by Phil Zimmermann. Be warned - this is a type of program which will not be without political and legal strings attached. The use of pgp in the US is legal, but export of the code is strictly prohibited. There are several sources, however, which offer a European counterpart of the software, mostly labeled with an 'i' for International. Francis Litterio has assembled a nice listing of resources on http://draco.centerline.com:8080/~franl/pgp/where-to-get-pgp.html which deribes legal issues and where to obtain which versions. NOTE: At the time of writing, only France does not allow the use of 'pgp' or similarly strict cryptography. The European Union, however, is considering announcing a common policy on this issue in the future.

The way how 'pgp' works is straightforward:

Public key transmission
The public key is distributed using electronic methods. You

might use electronic mail, WWW, FTP or any other method as the transmission of the 'public key' is not critical and can be unlimited in theory. Other programs might be employed as well. Try it out: use 'finger embnet@www.ch.embnet.org' to obtain the public key of the 'embnet' user on our node.

Message encryption
If you want to send a message, you must add the 'public key' obtained earlier to your configuration and encrypt your message with this key. This is a standard procedure and can be automated - see below.

Message transmission
Next, the transmission via electronic mail can be achieved just like sending any text file. Some mailer programs even allow the use of the multimedia extensions to electronic mail, as we have discussed previously in EMBnet.news, as MIME types. (http://www.ch.embnet.org/embnet.news/vol2_2/mime.html)

Message decryption
Documents are channeled through a program which will either be launched from the electronic mail program (if a MIME type was used) or which needs to be run after the mail was saved into a file. The message will be readable as before, as the 'private key' is part of the individual's mailor message transfer system configuration.

Further use of keys: signatures

Keys need not necessarily be used to encrypt messages. The authentication f messages is also very important. The problem of messages sent via public networks is their vulnerability to 'faking' and falsification - you cannot be sure that mail has really been sent from the originator unless you have a way to prove that the message is complete and authentic. To achieve this, 'pgp' can be used to create a checksum and an electronic 'signature'. The process of creation uses the 'private key' as this is the key assigned to a single individual. If messages are distributed using this method, each recipient will be able to read the message and check its integrity and origin if the 'public key' of the individual is used.

Application of cryptography and further use

Biologists do not currently use cryptography by default. The use of secure transmissions for sequence data, for example, is only possible if both sender and recipient know this technology and can seamlessly use programs which they do not need to configure themselves. The World Wide Web (WWW) uses two different mechanisms: one is the 'S-HTTP' from NCSA, and the other is called 'SSL' and available in a commercial implementation from Netscape. Both will only work if the server system the provider side and the client on the customer side have the same software. An extensive discussion list on security issues via World-Wide Web is available on http://www-ns.rutgers.edu/www-security/archives/subject.html. Future efforts will need to strengthen this issue, however, the new generation of the IP protocols has an increased degree of security built in (e.g., RFC1675 as described electronically in http://www.cis.ohio-state.edu/htbin/rfc/rfc1675.html).

Further reading on security can be found in the following bibliography: http://draco.centerline.com:8080/~franl/crypto/

TIPS FROM THE COMPUTER ROOM

The use of the C-shell (csh/tcsh) 'foreach' command.

Alan Bleasby. UK national EMBnet node

Have you ever been in the position where you want to examine many files in a directory using e.g.

```
grep –i protein *
```

i.e. search all files for the word 'protein'. For well-managed directories this will normally work but what do you do and what does it mean if you are summarily dismissed with the error message

```
"too many arguments"
```

It means that you've got too many files in the directory. csh is naturally interpreting the asterisk as a wildcard meaning all my files. What it then tries to do is to create a line of the form

```
grep –i protein fred jim bill bob ....
```

where the final arguments are your filenames. The problem is that shell command lines are limited to approximately 512 characters depending on the implementation.

You can get round this problem by using 'foreach'. An example follows:

```
%foreach i (*)
>grep –i protein $i
>end
```

(the chevrons are placed automatically).

The command is saying find all my files one at a time and place that filename in the variable 'i' ... you can call this variable anything within reason. Then, perform the grep on the filename. Note how you use the '$' symbol to refer to a variables' contents. The foreach command syntax must always finish with the line 'end'. You can have multiple lines in a foreach command. For example, if you want to

uncompress all your compressed files and print out their names as well you could type:

```
%foreach j (*.Z)
>echo $j
>uncompress $j
>end
```

NODE FOCUS

BEN, the Belgian EMBnet Node

The Belgian EMBnet Node[1] arises from the efforts of Dr. Robert Herzog and Dr. Marc Colet. For several years they have been providing the bioinformatics support for the Departments of Molecular Biology of the Université Libre de Bruxelles (ULB) and the Vrije Universiteit Brussel (VUB), who share a nice campus in Rhode Saint Genèse, not far from the capital Brussels. They learned about EMBnet in 1990 and wondered "Why not extend our efforts to the whole country?"

After some lobbying and with the support of Raymond Hamers and Nicolas Glansdorff of VUB, they were granted by, in November 1992, the Belgian Science Policy Offices the management of the Belgian EMBnet Node. The Belgian node would be created within the framework of the Brussels Free Universities Computing Centre (BFUCC).

The success of the BEN project was actually due to two major assets :

- the Brussels Free Universities Computing Centre, where the main computer of BEN was going to be located, had long standing experience in serving large groups of users. It was well equipped as far as national and international networks were concerned, running among other connections a 2 Mbit Europanet line with Amsterdam.
- BEN was presented as a joint initiative of the Dutch and the French-speaking universities of Brussels, which is certainly important in a country with two major linguistic communities.

Personnel, Hardware and Communications

The initial grant for BEN covers a period of 4 years and allows hiring two full-time staff members (Dr. Philippe Alard and Dr. Guy Bottu). It also covers the day-by-day functioning of the node. Basic hardware is provided by BFUCC.

Philippe became our support person for all French speaking users, as well as taking care of the system management, the databanks, the alerting software and part of the sequence analysis software. Guy's specialities are in the teaching area, where his in depth knowledge of GCG is highly appreciated. He manages the GCG software and baby-sits our SRS servers. His systematic approach gives him a special gift at revealing nasty bugs in any software...

In the period when BEN was founded, the Belgian Government also started the BELNET project to connect the universities and the main national scientific institutions by Internet. 64 kbaud lines are presently connecting about 25 sites, as well as high speed lines to Paris and Amsterdam. National line performance and network density of this academic network are steadily growing. In the course of its life, BEN has already moved twice to more powerful computers. Presently, it uses a DEC Alpha 3000/500[2] with the operating system OSF/1 3.2. A DEC ULTRIX 5000/200 located in Rhode Saint Genèse is still serving the informatics needs of the Departments of Molecular Biology in both universities and is used as a development and testing machine for BEN services.

BEN offers its services to all Belgian molecular biology research groups, from academic to industrial institutions. A yearly "research group" fee of 135 ECU covers the right to access all BEN facilities, with virtually no limitations except individual disc space. Besides these group accounts, BEN offers individual accounts at strongly reduced rates. Fees are slightly higher for industrial users than for their academic counterparts. BEN is being accessed both over Internet connections for the BELNET participants and over modem connections for all the other users. A park of more than 20 modems are available for these connections and a SLIP/PPP server is available for academic users.

Services

Until now, BEN has put emphasis on sequence analysis. BEN maintains a non-redundant and daily updated nucleic acid databank composed of EMBL and GenBank and a non-redundant and weekly updated protein databank composed of Swiss-Prot, PIR and GenPept. These databanks are accessible through GCG, BLAST, Gopher[3] (WAIS-indexed) and SRS[4] The updating of these databanks is performed automatically with the help of UNIX shell scripts developed by Philippe. The xNDT service of Peter Gad is running in parallel. BEN also offers several additional software packages and databanks among which Staden, Pearson-FASTA, Aidsbase[5], various ACeDB format banks[6] and IGD are included.

Our SRS server gives access to most of our databanks, and we are hosting the "Malaria Antigen Database" (MAD) which contains information about antigens identified in the malaria parasites (Plasmodium sp.). This database is maintained by Lou Dierick (The Consultory Ltd.) as part of an international project, funded by the EC, WHO and USAID, with the goal to develop a malaria vaccine. MAD can also be accessed directly by the WWW[7]. BEN is keeping two other Belgian Molecular Biology centres updated on a weekly basis with full databanks or selected records thereof.

BEN News

In order to keep in contact with its present users and to reach potential new users, BEN publishes a quarterly newsletter "BEN News". This is being sent to an ever growing list of addressees (well over 2500 at the last count). This publication significantly contributes to our perception by our national scientific community, where the need for computer tools by biologists is ever increasing. One of us (R.H.) is member of the embnet.news editorial team, where some cross printing of BEN News contributions have appeared.

Sequence monitoring agents : CSA and CSSA

BEN has developed systems that inform users about the arrival in the EMBL databank of sequences of potential interest to them. Current Sequence Awareness (CSA) is a Perl script which searches each day's new arrivals by combinations of keywords. Current Sequence Similarity Awareness (CSSA) searches for similarities with a query sequence (using blastn or blastx). These popular services relieve the users of the need to repeat the same search at regular intervals and save a lot of CPU time, since they run only on the daily retrieved sequences. CSA currently monitors about 150 keyword combinations, while CSSA presently takes care of about 200 sequences every day.

User interfaces : menugcg and WWW2GCG

GCG users used to complain about the need to learn commands by heart or to work with a huge manual on one's lap. To ease their work, Marc Colet developed a UNIX shell script named "menu", which is closely connected to the GCG environment and help files. menu provides the user with screen pages of GCG commands corresponding to the chapters in the manual. With menu, it is easy to choose the program from the chapter headers, start the program or obtain on-line help about it. All UNIX commands can still be executed in the classical way however. Therefore, menu is an invaluable assistance to all users, be they experts or beginners.

From release 8.0 onwards, GCG has contained the graphical user interface WPI. However, to use it comfortably, you need not only to run in X-Window mode, but also to have a reasonably fast connection with the main computer. This is still a luxury by today's Belgian network standards. Therefore Marc also developed WWW2GCG, a Web interface to the GCG software. This became the interface of choice to GCG for a sizeable part of BEN's users, as presently most of the programs can be run under the Web interface and graphical outputs are generated in a format directly understood by the browsers. All that is needed for a BEN registered user to run WWW2GCG on a microcomputer (Macintosh or PC) is a good Web Browser (Netscape or the most recent versions of Mosaic) and some vt100 emulation software linked to it when using the sequence editors.

Training program

Besides the basic function of keeping a large set of databanks updated and widely accessible, the Belgian EMBnet Node seems to fill a gap in the average Belgian molecular biologist's training. We are running our 3 day introductory and training programme every month and its popularity is still growing. We have access to a well equipped computer room on the ULB main campus where we can welcome up to 25 people, each of them getting access to a PC workstation[8] running X-Windows emulation. Up to now, BEN's training sessions were attended by more than 250 scientists, and recently, our program has been granted the status of a doctorate training course by the Flanders universities. We have also run our training sessions several times on remote campuses. It is amazing how many simultaneous sessions you can run over a 64 kbaud line !

The near future

For the near future, BEN will probably acquire a second AXP computer, in order to separate the user tasks from the database management and network server chores. We are looking into improvements of the user interface and work on developing client-server applications for sequence analysis. The rapid growth of the sequence databanks and the necessity to add support for genome data are our present main concerns.

Footnotes:

1 The "BEN" abbreviation was cast in 1992 by Jeroen Coppieters, presently working at the EBI, in Hinxton
2 128 Mbyte memory and more than 27 gigabyte of disk space
3 URL=http://gopher.be.embnet.org
4 URL=http://www.be.embnet.org/srs/srsc
5 The slightly messy nature of Aidsbase makes that the installation of a new release always involves some manual editing like removing doubles and avoiding UNIX incompatible file names
6 ACeDB, MycDB, AAtDB, GrainGenes, C21DB
7 URL=http://www.be.embnet.org/malaria/mad.html
8 Novell networked Pentium PC's with 16 Mbyte memory and the Micro-X terminal emulator program

---

INTERviewNET

In this issue Heikki Lehväslaiho (EMBnet.fi) interviews Markus Sadeniemi <http://www.funet.fi/funet/faces/markus_fi.html>, director of FUNET (Finnish University Network, <http://www.funet.fi/index-en.html>).

HL: Markus, you have recently gained fame in Finland by criticising the way the Internet is presented in the daily press.

MS: My concern has been mostly the printed misinformation about Internet. To me it looks like the press is trying to say that the most important things on the Internet are meowings

of the US President's cat or the ability to select tonight's movie in some experimental setup, while in fact the real stuff is elsewhere. Internet is already being used for real productive work.

HL: Just like we in EMBnet are trying to do...

MS: Yes. 5 years ago it was quite difficult to put anything onto the network. Now practically anyone can do it. The amount of information on the Internet is already so huge that it is really getting difficult to find relevant information. Automatons (Lycos, WebCrawler) are one way of doing it, but we also need people who really know what is in there to filter out information into manageable form. Not being directly involved with EMBnet, that is what to me seems to be the role of EMBnet. What worries me most is the uncontrollable commercialisation that is gaining momentum now.

HL: Do you see a conflict in this commercialisation tendency on the Internet and the science, especially the basic science, and research it is supposed to serve?

MS: Alas, I do. The old co-operative spirit of the Internet had its origin in the openness and solidarity in the research community. One example of this are public file servers. When Finland joined the Internet (in 1986) we were very grateful that we could freely use US servers. Now we are running our own server with 60000 visitors each day. Two thirds of them come from foreign countries. Should we keep this service free, if most of the users come from commercial companies? I am afraid we are heading to a future, where every WWW-page you click costs you a penny and a lot of time is used for billing even if it is automated. With this development libraries will also face problems.

HL: Why do you say that?

MS: Libraries have been able to buy a book and lend it several times without extra cost. If digital books are on the net and you have to pay by the number of pages you read, what is the difference between a library and a bookstore? I also see a conflict between growing demands for strict protection of intellectual property rights and the technical possibilities of making perfect copies. Pirate copies of computer programs is only a small sign of the problems we will see in the future.

HL: Educational systems will be in trouble?

MS: One new development in education is remote teaching. I've always been skeptical about it, but now suddenly it seems to be a real functional alternative.

HL: Birkbeck College at University of London is already announcing the second protein structure course in the Internet <http://www.cryst.bbk.ac.uk/PPS2/>.

MS: Expert subjects are one thing. Technology is suddenly available and it seems that nowadays for people to keep their jobs, they have to re-educate themselves several times during their active work life. For anyone living far from main university cities and with a family, remote studying is beginning to sound quite attractive. Think about the effect it will have on lecturers if students can decide with a click of a mouse that they'd rather listen someone else in a different university!

HL: That really should hit hard on lecturers! Finally could you tell us how do you see the future for an organisation such as EMBnet?

MS: Technically the research network, the Internet, is starting to mature. Providing network bandwidth becomes easier and even putting information on the net is technically easy. What makes all the difference for the research community is how to make relevant information available and how to make it easy for a researcher to find it. This is the great challenge of informatics in the coming years. But no theoretical work can replace human judgement on what piece of scientific knowledge is relevant and what is not.

I see EMBnet as a means for researchers to jointly organise the flow of information in a way that serves everyone in the specific field of science. This of utmost importance and I pity those brances of science that don't have an "EMBnet" of their own!

SOFTWARE DEVELOPMENT

A Tcl based SRS vs.4 Interface

Gijs Schaftenaar: CAOS/CAMM EMBnet Netherlands

An EMBnet funded project at the CAOS/CAMM Center, EU grant BIO2-CT944002.

Abstract

A new SRS (Sequence Retrieval System) User Interface has been developed for SRS vs. 4. Key features are the support of simple character oriented (ASCII, VT100) terminals, support of X Windows graphics terminals and support of an X Windows client/server environment. The interface is coded in Tcl, using some dedicated Curses calls and Tk and TDP extensions.

Introduction

The internal SRS query language, is user accessible through either a command line user interface, “getz”, or through a more user friendly character oriented window interface. SRS versions up to vs.3, supported this window interface compatible with VT100 terminals. SRS vs.4.x only supported a World Wide Web interface, depriving users with only a simple terminal/terminal emulator connection from access to SRS or forcing them to use the ASCII WWW client LYNX.

An EMBnet user poll, conducted in March 1995 showed an overwhelming interest in the availability of a character oriented SRS vs.4 window interface, preferably with the look-and-feel of its SRS vs.3.1 predecessor. This “market demand” has generated EMBnet funding for the development of such an interface. In addition, some “extensions” were coded to widen the applicability to graphics terminals and to a client/server environment. Development was initiated by Jack Leunissen and supervised and executed by Gijs Schaftenaar with the assistance from Koen Cuelenaere in close collaboration with Thure Etzold.

System Specification and Coding

The interface should meet the following specifications:

- coding with only widely available and well maintained public domain tools and libraries
- serving the large usergroup with only simple ASCII terminals/terminal emulators
- look-and-feel of SRS vs. 3 window interface
- extensions for X windows graphics terminals
- compatability with successive SRS release

These requirements were met by coding the interface in Tcl, (Ousterhout, 1994). The character oriented (ASCII) version employs some C-Curses calls and a dedicated C-Curses widget library for screen management. For the graphics the Tk (X11 Toolkit) extension to Tcl was used. For the client/server version the TDP was used. Tk adds commands for graphical attributes like windows, buttons and scrollbars. The TDP extension (Smith et.al. 1994) adds commands which enable communication between different computers in a network, using RPC’s. The necessary Curses extensions to Tcl, Table I, were taken from the literature (Kamp, 1994). The interface code has been tested on a variety of platforms.

Three interface versions are available:

- srscurs which has the look-and-feel of the previous SRS vs. 3. Fig. 1 shows an exemplary query mask. Appendix I summarises the C-Curses window interface elements
- srstk which uses the Tk extension to Tcl for screen management instead of C-Curses calls. This is the Xwindows graphics version of the interface. This version uses the extra functionality as supplied by the possibilities of a graphics

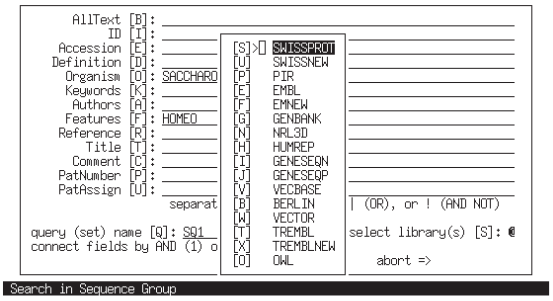

fig 1

screen. (Fig. 2)

- srsctk which is based on the CTk C-Curses version of Tk. CTk was released on Internet in May 1995 (Andrews, 1995). Due to the incompleteness of CTk this interface lacks some of the functionality of our basic ASCII version. Fig. 3 shows the screen.

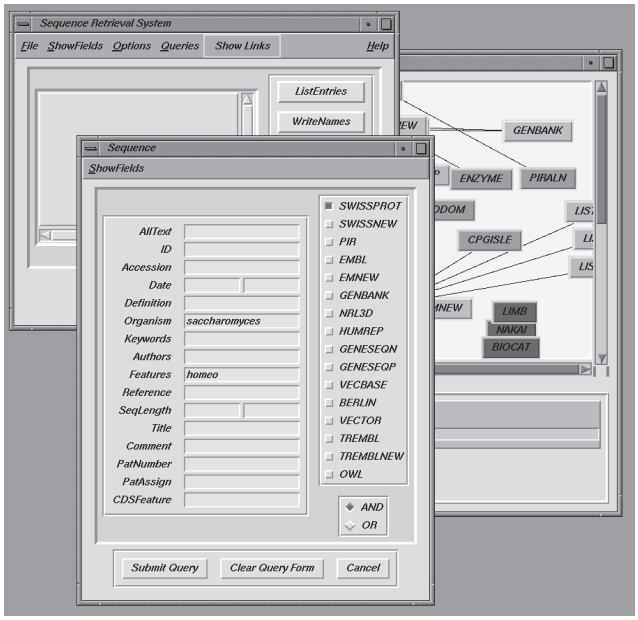

fig 2

Both srscurs/srsctk and srstk can be executed in different run-modes. These modes differ mainly in the way in which the SRS query engine is addressed. Details of these run modes are discussed in the paper submitted for publication in the

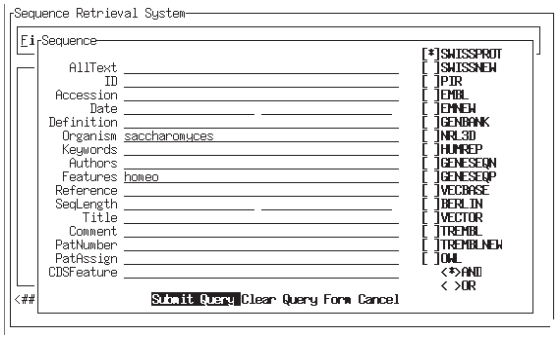

fig 3

Application Notes section of CABIOS.

References

- Andrews,M., (1995). andrew@ccfadm.eeg.ccf.org Cleveland Clinical, Cleveland, Ohio, US.
- Kamp, Poul-Hennig., (1994). Curses extensions to Tcl. phk@data.fls.dk
- Ousterhout,J.K. (1994) Tcl and the Tk Toolkit. Addison-Wesley Publishing Company; ISBN 0-201-63337-X
- Smith,B., Yen,S., and Tu,S. (1994). Tcl-DP; Distributed Programming extension to Tcl/Tk. Documentation file released with the TDP extension.

The software has been incorporated into the SRS distribution at the EMBL, ftp site . You will need release 4_07 or higher.

NODE NEWS

News from the EMBnet Nodes

1. News from the Irish EMBnet node: INCBI

To coincide with Charles Darwin's 187th birthday, INCBI is hosting its first Annual BioInformatics Lecture on the 12th February 1996. The speaker will be Dr Des Higgins, from the EBI, who will talk on "101 uses for trees and alignments"

All welcome!

---

2. News from the Italian EMBnet node

"KEYnet : the Keywords Hierarchical structure (EMBnet news, vol 1. n.2 , 1994) funded under the BRIDGE program <Promotion of EMBnet> and developed at the Italian EMBnet node, has now been implemented on a WWW browser from Matteo di Tommaso at the EBI. This is an experimental version. Further improvements of the structure will be performed in the next 3 years within the project recently funded under the BIOTECH program in collaboration with EBI. The browser can be accessed at the following adress http://www.ebi.ac.uk/htbin/bari.pl"

---

3. News from the Finnish EMBnet node: CSC

A three-day course on "Basics of Biocomputing" was held at CSC 27th - 29th November <http://www.csc.fi/suomi/Kuvaus_biocomp.html>. The course got very good ratings form the students. The lecturers were Rodrigo Lopez (EMBNet.no) and Heikki Lehväslaiho (EMBnet.fi).

The contents of CSC Biobox can now be searched with keywords <http://www.csc.fi/molbio/wais/>.

The local newsletter "CSC Bionews" is regularly sent to more than 500 local molbio users to the mailing list "gcg", The newsletter issues are now archived in html format at <http://www.csc.fi/molbio/bionews>.

New installations:
Databases:

- Vector - Cloning vector sequences
  http://cypress.csc.fi:8001/srs/srsc?-info+VECTOR
- TransTerm - Splice sites and codon usage for 93 species in GCG format
  http://www.csc.fi/molbio/progs/db/transterm.html
- TREMBL - Translated EMBL
  http://cypress.csc.fi:8001/srs/srsc?-info+TREMBL
- PDB & SRS - daily updated PDB entries con now be displayed directly from SRS.
  http://www.csc.fi/molbio/bionews/1995/8.12.html#PDB

Programs:

- EGCG - Additions and improvements to GCG package
  http://www.csc.fi/molbio/progs/egcg
- TreeAlign - Simultaneous multiple sequence alignment and parsimony tree reconstruction
  http://www.csc.fi/molbio/progs/treealign/
- totreealign - Interactive frontend to TreeAlign
  http://www.csc.fi/molbio/progs/treealign/#TIPS
- Pratt - Flexible pattern finding in unaligned protein sequences
  http://www.csc.fi/molbio/progs/pratt/
- Fastalert - Automatic, periodic FastA scanning of new nucleotide entries at EMBnet.ch.
  http://www.csc.fi/molbio/progs/fastalert/

---

4. News from the Spanish EMBnet node: CNB

BioImage Workshop

A workshop was organised in Madrid on the 9th and 10th of November on the topic of exploring the use of low-resolution volume information in structural biology. The ultimate goal was to set up a new database of this type of information that could complement existing information in other databases, most notably PDB.

A total of 30 people came to Madrid, covering topics ranging from examples on how volume information has solved key problems in biology, to database-related topics such as distribution and maintenance. EMBnet was represented by Dr. Rodrigo Lopez.

The meeting was extremely productive, providing a general overview on the topic of how volume data relates to other forms of data in structural biology that will be key for further developments of the project

The minutes of the workshop, as well as the list of participants, can be obtained at http://indy.cnb.uam/Bioimage/. A small book with examples of key applications in biology was compiled, and is available upon request.

The meeting was jointly funded by the Biotec Program and EMBnet.

New staff members at EMBnet/CNB, the Spanish EMBnet node

We are glad to say welcome to a new staff member at the Spanish EMBnet node. His name is ***Patrick Moutel*** and he has recently joined the EMBnet/CNB team.

Patrick holds a licensure degree on Chemistry, with specialization on Biochemistry, and a Master degree on Biotechnology. He comes from the Biocomputing Service of the Molecular Biology Center (CBM) of Madrid, Spain.

Jose R. Valverde also recently joined the CNB as the Manager of the Service of Bioinformatics and will be responsible of taking care of the EMBnet node.

Jose R. Valverde graduated in Medicine and Surgery by the University Autonoma of Madrid, obtained Master and Ph. D. degrees in Molecular Biology, and in 1991 started working as the Manager of the Service of Bioinformatics at the Biomedical Research Institute of Madrid. He comes to Madrid from the EMBL Outstation in Hinxton, the European Bioinformatics Institute, where he worked since its creation last year.

5. Swiss EMBnet Node News

Operational improvements (N.Redaschi)

The BioComputing Facility and services started a restructuring activity locally in order to improve and optimise service provision to the researchers. This includes ongoing collaboration with the EBI in order to access, and process, sequence database updates.

SRS developments boosted by HASSLE (M.Schmitz)

The HGETZ/XSRS1 service on top of the recently released HASSLE v5 protocol runs smoothly over wide area networks as well as locally. Collaboration with EMBL/T.Etzold proceeds to refine the product in order to incorporate it into the released SRS package. Combination with Gijs Schaftenaar's developments is possible, as well as the incoroporation of the new XSRS interface which was developed outside of Basel, and will be announced separately.

The BioCompanion is released (R.Doelz)

On the URL http://www.ch.embnet.org/jam/jam.html the most recent version of a new teaching/tutorial guide can be found. This document is an entire rewrite of the Biocomputing Tutorial/the BioComputing Survival Guide which have been released in past years. The format is configurable, as is the contents, using the recently released JAM Formatting system. The BioCompanion presents a basic introduction to BioComputing and focusses on the GCG software as well as various extensions, such as EGCG, SRS, MPSRCH and others. The BioCompanion is also available in a printed version from a publisher.

CONFERENCES AND ANNOUNCEMENTS

New Spanish node

EMBnet has admitted into its ranks a second non-academic node. The Research and Development Department of TDI (TECNOLOGIA PARA DIAGNOSTICO E INVESTIGACION S.A.), a Madrid based firm which specialises in the development of software and hardware for the analysis of one and two dimensional gels as well as phylogenetic analysis. During the period prior to the next EMBnet board meeting, scheduled to November 1996 in Helsinki, this node will operate as an observer node. TDIs Research and Development Department have a information page on the WWW and full Internet connection. For further details please see: http://www.tdi.es/indexi.htm

International Symposium on

THEORETICAL AND COMPUTATIONAL GENOME RESEARCH

25-27 March 1996
Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

Organizing Committee: M. Bishop, Ph. Dessen, R. Haux, R. Hofestaedt, W. Jaeger, J. G. Reich, O. Ritter, S. Suhai, P. Tautu, M. Vingron

TOPICS:

- BIOMOLECULAR SEQUENCE ANALYSIS

- GENOME MAPPING AND SEQUENCING
- MODELLING OF PROTEIN STRUCTURE
- INTEGRATED GENOMIC INFORMATION MANAGEMENT
- COMPLEX GENOMIC SYSTEMS

LECTURERS will include: S. Altschul, Bethesda; S. Brunak, Lyngby; J.-M. Claverie, Marseille; R. Cottingham, Houston; K. H. Fasman, Baltimore; J. Fickett, Los Alamos; H. Frauenfelder, Los Alamos; M. Geller, Warsaw; M. Kanehisa, Kyoto; P. D. Karp, Menlo Park; M. Karplus, Cambridge MA; H. Lehrach, Berlin; V. M. Markowitz, Berkeley; T. G. Marr, Cold Spring Harbor; E. W. Myers, Tucson; L. Nilsson, Stockholm; O. Ritter, Heidelberg; Ch. Sander, Cambridge; P. Schuster, Vienna; D. S. Searls, King of Prussia; T. F. Smith, Boston; G. Stormo, Boulder; J. Sussman, Brookhaven; W. R. Taylor, London; J. Thierry-Mieg, Montpellier; J. Thornton, London; D. Weeks, Oxford; S. Wodak, Bruxelles

Closing date for POSTER ABSTRACTS: 15 February 1996

REGISTRATION: DM 250 (DM 100 for students). Early registration is advisable due to space limitations.

For further information and registration please contact:
Dr. Sandor Suhai, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 280, D - 69120 Heidelberg
Phone +49/6221/422369, Fax +49/6221/422333
E-mail S.Suhai@DKFZ-Heidelberg.de

Getting The Most From Your Protein Sequence

British Biophysical Society & CCP11
11 March, 1996
The Wellcome Trust, London UK

For the biologist wishing to do the best science with the greatest ease and convenience, the choice and application of sequence analysis methods can be a bewildering process. This meeting aims to provide an overview of current techniques in sequence and structure analysis, outlining the most widely-used methods, explaining the pitfalls, and braving the biological information revolution!

For more information, and the registration form, try looking at the following WWW page:

http://www.biochem.ucl.ac.uk/bsm/dbbrowser/ccp/

or e-mail attwood@biochem.ucl.ac.uk

The Fourth International Conference on Computational Biology

Intelligent Systems for Molecular Biology '96

June 12-15, 1996
Washington University, St. Louis, Missouri, USA
http://ibc.wustl.edu/ismb96
Call For Papers

An electronic mailing list can be joined by sending a message to ismb96@ibc.wustl.edu with the word "subscribe" as the body of the message.

Key Dates :
Meeting: June 12-15, 1996
Paper Submissions
- Papers due: Feb 1, 1996
- Replies to authors: Mar 15, 1996
- Revised papers due: Apr 1, 1996

Open Poster Submissions
Abstracts due: Apr 1, 1996
Tutorial Proposals
- Full proposals due: March 18, 1996
- Replies to authors: April 1, 1996
- Draft handouts due: April 15, 1996
- Final handouts due: May 13, 1996
- Tutorials presented: June 12, 1996

The purpose of the ISMB conference is to disseminate the latest developments in computational molecular biology and biophysics and to stimulate new work on the application of intelligent computational systems to problems in molecular biology. ISMB is a multidisciplinary conference bringing together scientists from computer science, mathematics, statistics, and molecular biology. Its scope extends to any computational method or system supporting a biological task that is algorithmically, cognitively, or conceptually challenging, involves a synthesis of heterogeneous information, or exhibits the emergent properties of an "intelligent system." From a computational perspective, areas of interest include adaptive systems, intelligent experimental control, data modeling, machine learning, artificial intelligence, combinatorics, stochastic optimization, string and graph algorithms, linguistic methods, and parallel computer technologies. Biological areas of interest include molecular structure, genomics, molecular sequence analysis, evolution and phylogenetics, adaptive experimental systems, and molecular biology. Emphasis is placed on the validation of methods using real data sets and on practical application in the biological sciences.

The ISMB conference has attracted a large and enthusiastic audience comprising scientists involved in application areas such as artificial intelligence, structural biology, DNA, RNA

and protein sequence analysis and structure prediction, genome mapping, gene identification, molecular biology data and knowledge bases, and the modeling of biochemical processes. We are continuing the tradition of soliciting original papers, which will be rigorously refereed and published (by AAAI Press and the MIT Press) in proceedings available at the conference. The conference proceedings are indexed in the Medline database. The previous ISMB meetings were

- 1993: National Library of Medicine, USA
- 1994: Stanford University, USA
- 1995: Cambridge University, UK

The four-day conference will feature introductory and advanced tutorials on June 12th and presentations of original refereed papers, posters, and invited talks (June 13-15).

There will be special sessions at the conference on "Whole Genomes: Challenges and Implications," and on the "Interconnection of Molecular Biology Databases (MIMBD)." A test suite of raw data will be set aside to evaluate base-calling and gene-finding programs. A job fair and a vendor fair are also being organized.

Organizing committee
- David States (states@ibc.wustl.edu)
- Terry Gaasterland (gaasterl@mcs.anl.gov)
- Randall Smith (rsmith@imgen.bcm.tmc.edu)

Keynote Speakers
- Robert Waterston, Washington Univ., St. Louis
- David Haussler, Univ. of California, Santa Cruz
- Russell Doolittle, Univ. of California, San Diego
- Chris Sander, EMBL, Heidelberg

Contact address
ISMB '96
Institute for Biomedical Computing
Washington University School of Medicine
700 South Euclid Avenue
St. Louis, MO 63110-1012
USA
Phone: (314) 362-2134
FAX: (314) 362-0234

CHRISTMAS COMPETITIONS

Following tradition, here are your Christmas puzzles...

1. The Mysterious Case of WEIRDnet

EMBnet consists of many nodes. The latest (fictional) node is called WEIRDnet. It is an altogether strange node even by EMBnet standards. Like any other node there are Users, Node Managers & Sysops. On WEIRDnet the sysops always lie, the managers always tell the truth; the users sometimes tell the truth and sometimes lie (so its probably not too strange a node after all!)

A minor disaster recently happened on WEIRDnet when, either through an operating system loophole found by a user or accidentally by a manager/sysop, all data on the node was erased. Automatic procedures recorded over the network that immediately before the disaster only three people were logged on. It was also known that one of these people was responsible for the crash. Interestingly it was also known that one of these people was a user, one a sysop and one a node manager. One of them did the dirty deed. It was known that the sysop didn't do it though... she had a watertight alibi.

To preserve anonymity we'll call these people A, B & C. Suspicion fell on A. On questioning they made the following statements:

A: I did not wipe the system
B: 'A' is indeed not guilty
C: Rubbish, 'A' is guilty

Well, the WEIRDnet staff analysed these statements and couldn't come to any conclusion as to guilt or innocence. They therefore took the obvious route, the EMBnet Virtual Technical Manager was contacted (a hologram). He was intrigued by this problem and wanted to find out not only who was guilty but also which was the user, sysop & manager. Furthermore, to avoid excess network traffic he asked just enough questions to solve the problem specifying yes/no answers.

So, first of all he emailed 'C' asking "Are you the offender?" He got an answer and thought for a while.

Next he emailed 'A' and asked: "Did C wipe the system?"

After 'A' replied our virtual technical manager got the full picture. So, who was guilty and who was the sysop, manager & user?

2. And for those of you who are more pictorially minded here's a Winter Solstice Gif Quiz

Just look at the URL
http://www.ie.embnet.org/embnet.news/vol2_4/quiz.html

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 (rodrigol@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

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 contributions 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 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/vol2_3/contents.html***
- ***http://www.be.embnet.org/embnet.news/vol2_3contents.html***
- ***http://www.ch.embnet.org/embnet.news/vol2_3contents.html***
- ***http://www.no.embnet.org/embnet.news/vol2_3/contents.html***
- ***http://www.ie.embnet.org/embnet.news/vol2_3/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/***
- ***ftp.ch.embnet.org in the directory pub/embnet.news/***

Back issues.

Online:
- ***http://www.ch.embnet.org/embnet.news/info.html***

Postscript by ftp:
- ***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/***
- ***ftp.ch.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:

Alan Bleasby, Daresbury Laboratory, UK
(bleasby@daresbury.ac.uk)
FAX +44-1-925-603100
Tel +44-1-925-603351

Reinhard Doelz, Basel University, CH
(doelz@comp.bioz.unibas.ch)
FAX +41-61-2672078
Tel +41-61-2672247

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, BiO, University of Oslo, NO
(rodrigol@biotek.uio.no)
FAX +47-22694130
Tel +47-22958756

embnet.news
Vol.2, No.4, 1995
22 December 1995 ISSN 1023-4144