embnet.news

Volume 2 Nr.1 (ISSN1023-4144)

March 21st, 1995

Welcome, again, to embnet.news. This edition contains articles on multiple sequence alignment programs, a method for improving the speed of your sequence similarity searches, and a tool for examining the efficiency of European networks. Regular features include Tips from the Computer Room about searching for 'lost' files and Node Focus which this issue is BMC, the Swedish EMBnet node at the Uppsala Biomedical Centre. We also describe, both in this editorial and in INTERviewNET, the way in which collaboration between EMBnet and the European Bioinformatics Institute (EBI) will benefit European and Worldwide science.

EMBnet welcomes the establishment of the EBI. An obvious question concerns the relationship and differences between these two vital organisations. The answer is that they are symbiotic and above all friendly and collaborative. Representatives of the EMBnet Executive Board from Norway and the United Kingdom recently visited the EBI to sign a Letter of Understanding which explains our complementary roles.

EMBnet, currently consisting of 26 establishments (nodes) in 19 European countries, provides nationwide computing services which include database and software services and export via fast local networks, a pan-European knowledge base, training and help lines in native languages, hands-on experience of local needs and also research. EBI, a member of EMBnet, is a dedicated primary data provider via the data library (EMBL plus other databases and software), a research site and a focal point for European information which is disseminated by EMBnet.

The two organisations are mutually supportive and also collaborate on research projects. The matter was put very succinctly at our joint meeting when the statement was made that "If EMBnet didn't exist the EBI would have to invent it". The obverse is equally true. Europe is therefore in the fortunate position that both organisations exist and can act as a team in the provision of worldwide bioinformatics resources.

Alan Bleasby (UK)
Reinhard Doelz (Switzerland)
Robert Herzog (Belgium)
Andrew Lloyd (Ireland)
Rodrigo Lopez (Norway)

The embnet.news editorial group.

BITS
BioInformatics Theory Section

ALGORITHMS FOR MULTIPLE SEQUENCE ALIGNMENTS

Guy Bottu, BEN, The Belgian EMBnet node

In a previous issue of embnet.news, we considered the alignment of pairs of sequences and the search for similar sequences in databanks. We now turn our attention to multiple sequence alignments.

If you have several similar nucleic acid or protein sequences it is often useful to align corresponding bases or amino acids in columns. For instance, you might wish to group bases or amino acids that occupy similar positions in the three-dimensional structure which exercise similar functions or that have evolved by substitution from the same base or amino acid in an ancestral sequence. In the latter case you might also like to construct a phylogenetic tree.

1. Global alignments.

The Needleman and Wunsch algorithm for finding the best global alignment of two sequences can readily be extended to multiple sequences. The problem is that the time the computer needs for such a job is roughly proportional to the product of the sequence lengths. So, if aligning two sequences of 300 positions takes 1 second, aligning 3 sequences takes 300 seconds and aligning 10 sequences would take 300^8 seconds, which is longer than the lifetime of the universe!

Since searching for a best global alignment using a rigorous algorithm is not realistic for more than three sequences, a

Contents

Editorial	1
BITS - Algorithms for Multiple Sequence Alignments	1
Node Focus - BMC in Sweden	3
INTERviewNET - Mike Ashburner, EBI	4
Hardware development - Bioccelerator: a solution for faster searches	5
Network development - Performance monitoring in EMBnet	6
Tips from the computer room	8
Node News	9
Announcements	12
EMBnet nodes	14
embnet.news information	15
Answers from the Christmas leisure page	15

number of strategies have been developed to carry out a multiple global alignment in a reasonable amount of time with a reasonable chance of finding the best alignment. The GCG program pileup first aligns all possible pairs of sequences according to Needleman and Wunsch (for n sequences, this makes n*(n-1)/2 alignments). Then it uses the pairwise similarity scores to construct a tree using the UPGMA method (see below). Finally, this tree serves as a guide for a progressive multiple alignment starting from the tips. Once two sequences have been aligned, their relative alignment is no longer changed. Clusters of previously aligned sequences are treated as a linearly weighted profile when they are subsequently aligned with another sequence or another cluster.

Other approaches include:

- The very popular CLUSTALW program differs only from pileup in that it performs the initial pairwise alignments using the fast algorithm of Wilbur and Lipman. CABIOS 8:189 (1992). Clustalw is obtainable by ftp from ftp.ebi.ac.uk /pub/software/unix and /pub/software/vax
- Starting with a search for words of n bases or amino acids that are common between the sequences. An example is Martin Vingron's program MALI. CABIOS 5:115 (1989). MALI is not distributed freely but may be obtained from its author Martin Vingron (vingron@embl-heidelberg.de)
- PIMA uses pattern-matching, rather then profile matching, while making the progressive alignment. PNAS 87:118 (1990) PIMA can be obtained via ftp, for UNIX systems, from mbcrr.harvard.edu /MBCRR-Package
- Building a phylogenetic tree, using a more elaborate algorithm, as the sequences are progressively aligned. An example is Jotun Hein's program TreeAlign. Meth.Enzymol. 18:626(1990) TreeAlign can be obtained for UNIX and VMS from the same address as given for Clustalw (see above)
- Making the best multiple alignment in a limited area of alignment space. This can only realistically be performed with eight to ten sequences.

2. Local alignments

There are cases where sequences share a similar region but are otherwise completely different. Take, for example, the amino acids in the active site of an enzyme or transcription factor binding sites in a DNA sequence. To handle these cases local multiple alignment algorithms have been developed. Usually they only look for ungapped alignments thereby avoiding the problem of choosing the optimal gap penalty. Two such programs have been developed at the NCBI :

MACAW by Schuler, Altschul and Lipman first tries to find high scoring segment pairs (HSPs) for each possible pair of sequences using the BLAST algorithm (with the sensitivity set high). It then assembles overlapping HSPs into blocks. An interesting feature of MACAW is that it does not try to align all sequences, but can pick out only those that share similar regions. Proteins 9:180 (1991). There are versions of MACAW obtainable for the PC under Windows and for the Mac: cnbi.nlm.nih.gov /pub/macaw The MACAW distribution also contains Gibbs (see below) and a pattern searcher.

The **Gibbs** sampler algorithm involves iteratively making a profile with stretches of n bases or amino acids, selected from the sequences, and then searches this profile against one of the sequences. The result of the search is used to weight the selection of the stretches at the next run. A drawback is that the user must choose the width n and the number of elements in each sequence and thus must have a certain idea of the outcome, or run the program several times. An interesting feature is that the Gibbs sampler algorithm avoids the choice of an externally added scoring scheme since it derives the highest scoring profile, in a self-consistent manner, from the data. Science 262:208 (1992). Gibbs for is available by ftp for UNIX from ncbi.nlm.nih.gov /pub/neuwald/gibbs

3. blast3

It is also worth mentioning the program blast3. This searches a protein against a protein databank using the BLAST algorithm (with the sensitivity set high) and then makes threefold alignments between the query sequence and each possible pair of databank sequences that have been found. Only the statistically significant threefold alignments which are made from three nonsignificant pairwise alignments are retained. blast3 is useful in finding proteins that share a region of only weak similarity. Occasionally it can show that a query sequence makes the bridge between two databank sequences whose relationship had not yet been suspected.
It is possible to access a BLAST (including blast3) server at the NCBI, either through WWW or with a specific blast Internet client that you can install on your computer.

4. phylogenetic trees

Ideally a researcher would like to have a black box in which to throw sequences and get out a fully annotated phylogenetic tree. This is, however, not possible for two reasons. First, an algorithm that considers all possible multiple sequence alignments and then, for each alignment, all possible phylogenetic trees and picks out the best one, would take too much time. That is why most phylogenetic programs work on previously aligned sequences. Second, the result is always strongly influenced by the criteria that are used to define the best tree. Phylogenetic analysis will be the subject of a separate column in a later issue of embnet.news. However, a few remarks seem appropriate here. There are three main kinds of tree building methods: distance matrix, maximum likelihood and parsimony.

Distance matrix methods first estimate the pairwise distances between the sequences (which means that the information in the alignment of two sequences is reduced to one number) while the other methods construct many trees from all the information in the multiple alignment and decide which is best.

The simplest distance based method is UPGMA (unweighted pair-group method using arithmetic averages) which involves iteratively taking together the two sequences that have the shortest distance from each other, placing them at the end of branches on a node of the tree, and replacing their distances from the other sequences by an average value.

The guide tree used by pileup and CLUSTALW should never be used to infer phylogeny! It has been derived from the distances between pairwise aligned sequences and these distances are not necessarily the same as the distances between sequence pairs taken from the multiple sequence alignment.

Node focus

EMBnet Sweden
BMC, Uppsala Biomedical Centre,
Box 570, S-751 23 Uppsala, Sweden

Background:

The Biomedical Centre houses more than 35 departments belonging to any of three faculties of Uppsala University, or to the State Veterinary Institute (SVA) or the Swedish University of Agriculture (SLU). There are more than 1400 employees and about the same number of students at all levels of their academic education.

History:

When minicomputers appeared on the scene, the BMC set up a small computing department running two Nord-10 computers. That was in 1976 and those two computers served us for ten years. When they needed replacing, it turned out that our customers were changing from being "statisticians" to "sequencers". It was thus decided that the next computer should be a VAX and the major software package should be GCG. That marked the beginning of the bioinformatics era at the BMC. At the same time a VAX 8200 was obtained, also a 1.6 km Ethernet network was installed throughout our buildings which brought everyone within 50m of an Ethernet connection. With this purchase we also saw the beginning of another era - when you buy a new computer it is out of date when it is finally installed! Quite soon the 8200 had to make friends with a 3100 VAXstation in a LAVCluster, to gain some more CPU power.

Time rolled on to 1988 and there was a call for establishing a Swedish national EMBnet node. The BMC was (s)elected in 1989 from the Swedish universities to fulfill that role. A new VAX 6410 computer was installed as the central resource and fully automated redistribution of EMBL nucleotide data to all universities in Sweden was set up from day one of the service. Soon it became clear that Unix was coming into the academic world, even in Swedish biomedicine, so we bought a number of SUN workstations; a SparcServer 2 serving two SLC's with operating system software and general discspace.

Present situation:

Today we run, apart from those three SUNs, three ELCs, one IPX, eight SparcStation 10s, one 630-2xCPU Server and a 4 CPU SparcServer 20 with all in all more than 67 Gbyte of disc. The SUN environment is a homogeneous "cluster" using Solaris 2.x and NIS+ (with two exceptions - machines that are still running SUN-OS 4.1.3 to support third-party software). Moving CPU power from central dragons to everyone's desktop has been going on for quite some time now and we see, on almost a daily basis, a growing number of very capable MacIntosh and DOS machines being connected to the Ethernet. This once more shifts the focus of things. To meet the needs of Mac users, CAPservices were set up on our SUN servers thus enabling AppleShare connections to our systems. For DOS/Windows users, we use PC-NFS, connecting PCs to the native SUN NFS service. Today we give not only login service over telnet but also library literature database services like Medline, Chemical Citation Index, Intl. Pharmaceutical Abstracts and soon Biosis. Most work done on our system falls within the area of sequencing. To meet the demand from roughly 700 users, we currently run GCG, Staden, Phylip and programs for linkage analysis. The EMBnet node is also closely related to the GDB system, giving access to GDB and OMIM services for the nordic countries. For information services, an experimental gopher server was installed early in 1993 but was never really expanded to a full service. Likewise, an experimental WWW server was started during summer 1994; this seems to be more of a success than gopher ever was. Shortly before Christmas 1994, an SRS-WWW server was installed.

A major effort of the Swedish EMBnet node has, by a central decentralisation decision, been devoted to the transferring of EMBL nucleotide data to all Swedish universities. To enhance the transfer of data the NDT protocol was constructed and implemented. It is now used for all EMBL nucleotide data redistributions within Sweden.

The activities at the Swedish EMBnet node have been maintained by a staff of 1.2 people (two working 50% and one at 20% part-time) but has just recently been extended with another 50% position dedicated to education.

All work is funded by the Swedish research councils.

INTERviewNET

In this issue Alan Bleasby interviews Professor Michael Ashburner, Coordinator of Research at the European Bioinformatics Institute.

AB: How do you view the relationship between EBI & EMBnet?

MA: I describe it with that overused word: synergy. The EBI has two central missions, to collect, annotate and distribute data and to be a focus for bioinformatics research in Europe. Neither mission can be accomplished alone. A close interaction with the community of wet biologists and those who live staring at a screen is essential for the success of the EBI. The EMBnet provides an essential component for the success of these missions. The National Nodes provide their users with data, software and training at a level that would be extremely difficult, if not impossible, for the EBI to do alone. For this reason, if no other, were the EMBnet not to exist, then the EBI would need to invent it!

The two organisations are complementary despite some similarity of their activities. Take training as an example. The EMBnet nodes provide training at several different levels, from the most elementary introductions to computer use to more advanced. They offer support with helpdesks which can answer questions at all levels and in a national language. The EBI is also planning courses and workshops but these will be targeted at advanced users. We will have a well equipped course room for this purpose (I hope that this will be ready by late 1996). I very much welcome input from the EMBnet nodes as to what courses might be useful. We are happy to host courses formally organised by organisations such as EMBnet, EMBO and FEBS.

To encourage interactions between researchers in Europe the EBI has recently announced a Visitors Programme. We will provide facilities, and limited funds, for people to come and work at Hinxton for short periods, not necessarily in collaboration with a member of the EBI's staff. Anyone needing further details of this programme should write to me at the EBI (or email ashburner@ebi.ac.uk).

AB: As Director of research at EBI what do you do?

MA: First, my title! I am "Research Coordinator" not "Director of Research"! Second, some background. The EBI is an Outstation of the European Molecular Biology Laboratory. It is located at Hinxton, some 16 km south of Cambridge along with the Sanger Centre and the MRC's Human Genome Mapping Project Resource Centre. In addition, the Wellcome Trust, who own and are developing the Hinxton site, will build a new Conference Centre at Hinxton together with a training facility and on-site accomodation. All of this is set in the middle of a nice park just 20 minutes from Stansted airport! Do look at our WWW page (http://ebi.ac.uk/hinxton/hinxton.html) for a picture of the site and how to reach us. The EBI, as an EMBL Outstation, provides a home for the "traditional" activities of the EMBL Data Library within the context of an organisation whose mandate includes both "R&D" and "research". Clearly, any organisation producing products as complex as those of the EBI must engage in R&D if only to develop and improve its products. This is done by a small R&D team working with Graham Cameron. In addition the Data Library services could not continue to evolve unless it is within an environment of active research, for it is this research which will determine future developments. Although there was (and is) strong bioinformatics research at the EMBL, bringing together, at the EBI, the database service and R&D and bioinformatics research will be of mutual benefit.

My brief is to coordinate these research activities. My position is parallel to Graham Cameron's, whose job it is to run the service and service related R&D activities of the EBI. I am now both thinking about the general areas in which we wish to encourage research at the EBI and scouting for individuals who might be encouraged to apply for positions of research group leaders.

I am spending quite a lot of time visiting and talking to people! I announced at Chester last September that I intend to visit all of the EMBnet National Nodes, both to tell them about developments at the EBI and to learn what goes on locally. OK, I have yet to make it north to Cheshire, but I have already visited Italy, Sweden, Norway, The Netherlands, Ireland and Israel. I also had lunch with Reinhard Doelz in Basel (which was very nice) on my way to visit Hoffman La Roche, and have a visit to MIPS planned for later this month.

AB: What size is the current research effort and what is projected?

MA: Two appointments on the 'research side' of the EBI have already been made. Chris Sander is moving from EMBL-Heidelberg to the EBI and Shoshana Wodak has a half-time appointment. The plans for the EBI call for a size of about 70 people by the end of 1995. The breakdown is roughly 50 on the 'service' and service related R&D side, 13 on the research side and the balance for general support. However I expect that the research side will become larger by including people funded from outside sources. The new building has been designed to allow significant expansion and I hope that the research activities of the EBI will attract funds to allow us to fill it!

AB: What research efforts are currently underway and what is planned?

MA: Neither Chris Sander nor Shoshana Wodak have yet

moved to Hinxton: we are now rather cramped in temporary huts! It is clear from these two appointments that protein structure will be a major field of research at the EBI. We have plans for a strong involvement in macromolecular structure databases. A strong European 'presence' in this area is very important. There are important local resources with the Cambridge Crystallographic Data Centre just down the road, and the MRC Laboratory for Molecular Biology and the University's Department of Biochemistry both analysing macromolecular structures .

We would like the spectrum of research activities of the EBI to reflect the cutting edge of bioinformatics research. Apart from excellent quality, we want to have a group of researchers at the EBI who can interact both among themselves and with the service side. It is not the mission of the EBI to grow into an Institute for Theoretical Biology!

AB: What computing facilities are available at the EBI?

MA: I am not the person to answer this (_you_ should have known that!) but I know a chap who is, so what follows is from Roy Omond and not me:

[1] VMS: DSSI-cluster 2 x DEC 4000-720 (each with 256 Mbytes mem and 2 procs) 32 Gbytes disk.
[2] OSF/1: 4 x DEC 3000-600 (each with 128 Mbytes) 1 x DEC 3000-800 Bunch of disks etc.
[3] Cisco 7000 router.
[4] DEC GIGAswitch (full-duplex FDDI switch i.e. 200 Mbits/sec point-to-point) currently with 4 machines connected, soon to be more. (This is interesting because of the way computing is expected to evolve at the EBI, in particular allowing parallel processing on a loosely connected but nevertheless high-speed, very-low- latency network of workstation class machines).
[5] Bunch of Macintoshes (mostly now Power Macs)
[6] Bunch of MS-DOS PC's.
[7] 1 x DEC AlphaStation 200 model 4/166 (166 MHz, 64 Mbytes) running Windows NT.
[8] Bunch of NCD X-terminals.
[9] Network connection to the outside world is via shared (with Sanger and HGMP) 8-Mbit line to SuperJanet backbone in Cambridge.

I am sure that you can decode this !

AB: How is FlyBase doing?

MA: Great! Actually FlyBase is my other half, that employed by the University of Cambridge and not by the EMBL, although there is certainly a lot of common interest. FlyBase is funded by the US National Institutes of Health and I have one of the three groups involved (the other two are in the other Cambridge and in Bloomington, Indiana). One of the ideas that we have pioneered in FlyBase is links between genetic and sequence databases. At the time these were first established we did not know just how useful these would be. With the success of WWW browsers such as Mosaic, the real power of the links we have established with both nucleic acid and protein sequence databases has become clear. I am keen to establish similar links between different single organism databases, and will be using my EBI hat to foster these. There are other problems general to such databases. One of these is the classification of gene products. For enzymes we have a system, but for other gene products (protein and RNA) there is nothing available. Even for enzymes the classification, by type of catalytic reaction, is wholly inadequate. Other attributes which could be a basis for classification include sequence, structure, 'function' and cellular localisation. By collaboration between these databases and bodies like the EBI we can also build systems that will allow biologists to view an organism's sequence data not as a collection of unrelated records but integrated and closely related to data such as genetic maps.

AB: Thanks Mike. I'm tempted to say that this interview has gone quickly but will resist saying that "Time flies like an arrow but fruit flies like a banana", or maybe I won't.

Hardware Development

BIOCCELERATOR: A CURRENTLY AVAILABLE SOLUTION FOR FAST PROFILE AND SMITH-WATERMAN SEARCHES.

Leon Esterman,Israeli National Node (INN)

Background.

DNA and Protein databases and sequence analysis programs have become critical tools in biological, medical and genetic research. The load on INN's server for frequently requested services like database searches, FASTA, TFASTA, Profilesearch and multiple sequence alignment is rapidly increasing with the ever expanding size of the databases. This is a situation probably occurring at every EMBnet site.

The amount of data available is growing exponentially and many computational resources are needed in order to find meaning in this information. Biologists are aware of the need for more processing power and are examining solutions based on RISC workstations and on massively parallel computing[1]. Workstations are relatively cheap but will quickly be overwhelmed by the data that is being generated. Massively parallel computers will be able to support future needs but they are expensive.

The BIOCCELERATOR, on the other hand, is an accelerator

board designed to support the specific functions needed to run a DNA or protein sequence search extremely fast. Developed and built by Compugen Ltd. of Petach-Tikva, Israel, and tested extensively by the Biological Computing Division of the Weizmann Institute, the BIOCCELERATOR is the simplest and most cost effective solution available today for rigorous sequence analysis. A high end BIOCCELERATOR completes, in seconds, runs that can take hours or days on fast workstations. The BIOCCELERATOR is just as fast, or faster, than the massively parallel solutions available but costs less than one tenth of these solutions and is more convenient to use.

Algorithms supported by the BIOCCELERATOR include the Smith-Waterman algorithm [2] and PROFILESEARCH [3]. The BIOCCELERATOR is the only currently available solution for fast profile searches. Smith-Waterman has been shown by various studies [4] to be the most sensitive algorithm for searching protein databases. It is the only algorithm that allows an unlimited number of gaps in the alignment. This fact has been known by biologists for a long time but the algorithm was used infrequently because of the large amount of time each run required. With the BIOCCELERATOR, it has become feasible to run such searches on a regular basis and thus increase the chances of detecting significant homologies. PROFILESEARCH is an extension of the Smith-Waterman algorithm. Instead of using a single sequence as a query, a profile is constructed from a multiple alignment of related proteins. This profile reflects information known about a family of proteins, not just one member of the family. PROFILESEARCH is one of the most sensitive methods for finding new members of protein families.

The BIOCCELERATOR was designed to be completely compatible with the widely-used Genetics Computer Groups Wisconsin Sequence Analysis Package. For users of the GCG package, the BIOCCELERATOR is transparent, employing the same file and database formats. For numerous sites which are GCG-literate, the BIOCCELERATOR is a simple solution that doesn't require re-educating users.

The Weizmann Institute was the natural beta site for the BIOCCELERATOR. The Institute's Biological Computing Division [BCD] hosts the Israeli National Node [INN] of EMBnet and provides computing services to over 1000 users in Israel. BCD provides an international e-mail service based on the BIOCCELERATOR for performing rigorous database searches. For information send a blank e-mail message to bicserv@sgbcd.weizmann.ac.il .

With the phenomenal growth in cDNA sequencing use of the Smith-Waterman algorithm has also grown. Because of its ability to handle gaps, the Smith-Waterman algorithm can accurately retrieve sequences when cDNA queries are used. cDNA sequence fragments require sensitive algorithms because of the need to retrieve the parent genomic sequences in which the cDNA fragments are surrounded by non-coding regions.

BIOCCELERATOR, along with NCBI and EMBL, was used last year to perform the database comparison that was instrumental in finding the gene BRCA1 [5]. For academic institutions the BIOCCELERATOR provides a comprehensive solution for speeding up the most time consuming database searches and enables researchers to work almost interactively instead of waiting hours for results.

In 1993, the EMBnet BRIDGE project supported the introduction of this technology.

References:

1. Guy Bottu (1994) Fundamentals of database similarity searching methods. embnet.news 1(1): 2-3
2. Temple F. Smith and Michael S. Waterman (1981) Comparison in biosequences. Advanced in Applied Mathematics 2: 482-489
3. Michael Gribskov, Andrew D. McLachlan, and David Eisenberg (1987) Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA 84: 4355-4358.
4. Saul B. Needleman and Christian D. Wunsch (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453
5. Mark H.Skolnick et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66-71.

Network Development

PERFORMANCE MONITORING IN EMBNET

Hans Engelkamp and Jan H. Noordijk,CAOS/CAMM Center, Nijmegen, The Netherlands

Introduction

To monitor the performance, and hopefully the rapid improvement, of European network connections all EMBnet nodes are examined on a daily basis from three different locations; The Netherlands, Spain and Norway. Data are collected using the PING utility. Processed data are presented below.

Ping is a tool for network testing, measurement and management. It utilises the ICMP protocol's ECHO_REQUEST datagram to elicit an ICMP ECHO_RESPONSE from a host or gateway. ECHO_REQUEST datagrams ("pings") have an IP and

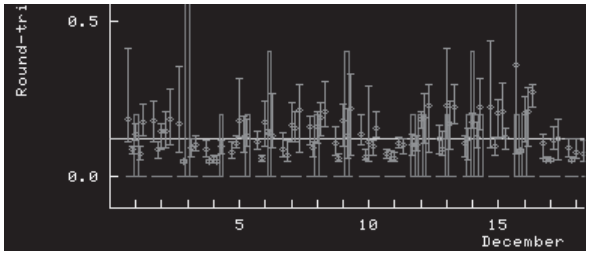
ICMP header, followed by an 8-byte timestamp, and then an arbitrary number of "pad" bytes used to fill out the packet. For the current monitoring facility packet sizes of 64 are used. These mimic average traffic, which is usually a mix of Telnet and FTP, rather well. It has been shown that data collected in this way semi-quantitatively represent data transfer and interactive work to and from remote sites. The Dutch Network organisation SURFnet uses, as a "quality-of-service" level for interactive on-line work, an upper RTT (round-trip times) limit (with 0% packet loss) of 125 msec. Using this parameter with PING indicates that, for the majority of EMBnet Nodes, the current EMBnet services require far better connectivity than actually available.

Implementation and examples

The monitor program, as implemented, uses the ping command in the following way. Five times a day (01:00, 09:00, 12:00, 15:00 and 18:00 hrs.) a fixed number of requests (5) is sent and round-trip times (RTT) and packet loss statistics are computed. These data are collected and displayed graphically using gnuplot as in picture # 1.

One should be aware that RTT's, owing to both network congestion and destination occupancy, can be strongly dependent on the time of the day as is exemplified in picture # 2.

This example shows measurements to a German destination from the Netherlands. These daily fluctuations are also visible in the full Node results, although in less detail because of the sampling.

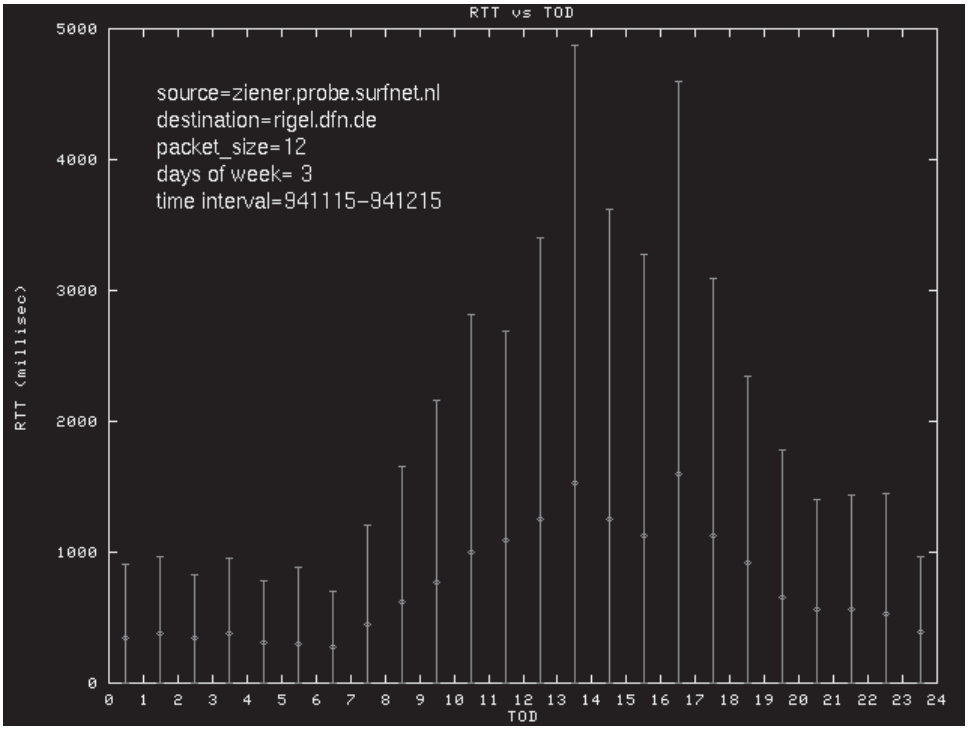


Picture 1 - Fragment of a gnuplot

Minimum, average and maximum round-trip time. Round-trip times are given in seconds, ranging from 0.0 to 2.0, even though some round-trip times can be up to ten seconds.

Relative packet loss. It ranges from 0.0 to 1.0, and is the number of 'lost' packets divided by the number of packets sent. A packet loss of 1.0 means that all packets were lost (which often indicates that the remote machine was switched off). In the latter case obviously no round-trip times are reported.

Average round-trip time measured over the whole month, again in seconds.



Picture 2 - Plot of round-trip times versus the time of the day

A few more notes with respect to the interpretation of the result: As presented the measurements show the combined result of network congestion and destination occupancy. The measurements at 1 o'clock at night are considered to represent a "zero" level, or the "best one can get" situation.

As a subsidiary result the Node availability can be determined from the packet-loss statistics. Blue bars of 100% indicate that the Node was down at the time of the measurement.

The data do not give an indication of the specific bottlenecks along the data paths. This project only tries to collect and present factual data on the reachability of the different EMBnet nodes. This determines if the network problems experienced do indeed hamper node services. By performing these measurements over longer periods (some years) improvements for specific nodes can be identified easily. Incidental bottleneck tracing might be performed by "traceroute" measurements. For continuous measurements of this kind network and routing topology is required and would result in a project far beyond the scope of the current one. It might be advisable to contact DANTE to start up a project of that size.

The (daily updated) results are available on the WWW-server at the CAOS/CAMM center, via URL "http://www.caos.kun.nl/Ping/".

Tips from the Computer Room

SEARCHING IN AND FOR FILES

Reinhard Doelz, BioZentrum Basel.

Have you ever had the problem of knowing that you have saved a sequence in a file but forgotten what the file was called? Have you been in the position that you saved an electronic mail message with the tag 'IMPORTANT' and can't find it anymore?

There are ways to solve this problem :

(1) Consider a preselection of possible files by specifying the last part of the filename (the 'extension'). Usually you will save sequence files as *.seq or *.ref files. Mail messages, as in the second example, are a bit more tricky, but possibly these are in distinct mail folders rather than anywhere in your disc space. Then use a program to scan all filenames selected.

In UNIX, you use the 'find' command as explained in a previous computip corner of embnet.news, on VMS you can use the file names directly (see below). In both examples you will search all the directories below your current position in the tree.

Example:

```
VMS: $ DIR [...]*.seq
UNIX: % find . -name "”*.seq"” -print
```

(2) Consider a significant word which will allow to you to pick out the desired file. If you use the GCG program package, all sequences will have the word 'Length' next to the sequence name. An electronic mail message might have the sender of the message as the significant word. The command to do this type of search in a text file is 'grep' in UNIX and 'search' in VMS. To make the latter case-sensitive you must use the /EXACT qualifier. The commands below will only search in the current directory.

Example:

```
VMS:     $ SEARCH/EXACT *.seq ”Length”
UNIX: % grep Length *.seq
```

(3) Combine the searching _of_ files with searching _in_ files. This is achieved by using the commands mentioned in step (1) and (2) on the same command line. The following examples show the output of the commands as well. Note that searching the word 'Length' in sequence files gives you the entire line of the file. On VMS systems the SEARCH command will report the name of the file which has the keyword in it. In UNIX, the command as typed below will give you all files found plus the key line below the name of the file. As GCG program package sequence files have the date and other parameters attached to this line as well it is possible to get even more information for this file other than the keyword itself e.g. the type and length of the sequence.

Example, VMS:

```
$ SEARCH/EXACT [...]*.seq ”Length”
```

Example, UNIX (as a one-liner):

```
% find . -name ”*.seq” -print -exec
    grep Length {} \;
```

Node.news

Swiss EMBnet node (BioComputing Basel)

Hardware/Operating systems:
The disk pool was increased by another 10 GByte in order to cope with the data flood. OS/2 is under evaluation as the programming platform for application software.

Databases:
Thure Etzold's SRS program package can compile a protein database from DNA by utilizing the 'trembl' program. We provide TREMBL as both translated EMBL database and as TREMBLNEW made from the updates. Access to the database is possible via the SRS-WWW system.

Software:
The 'HGETZ' software has been developed. Based on a HASSLE client, the program talks to a SRS server which is interlinked to the HASSLE provider. Its interface is based on plain text. Dr. Nicole Redaschi joined the Basel team and will participate in this specific development. EMBnet funding has been allocated to support the final packaging, distribution and release of this program.

In order to meet the challenge of the steadily growing databases, the tools which were developed as EMBnet Switzerland over the years were harmonized and assembled into the List Update Processing (LUP) package which is currently being tested. It is anticipated that a HASSLE-based version of LUP is released in summer 1995 to facilitate the integration of EMBL database updates into the GCG software and more simple packages.

German EMBnet node (GENIUSnet)

New databases:
The Encyclopedia of the Mouse Genome
FlyBase: The Drosophila genetic database
AAtDB: Arabidopsis thaliana database
ACeDB: Caenorhabditis elegans database
EcD: Escherichia coli database

New Software:
A menu system using Lynx or Mosaic
RasMol: Graphics Program for Visualisation of Molecules
Mapping programs:
Linkage: Lathrop & Lalouel's Package
Mapmaker: Eric Lander's Package
Crimap: Phil Green's Package
Map: Multipoint Map from Two-Point Lod-Scores
Fastmap: Approximate Multipoint Lod-Score
LIPED: Two-Point Linkage Analysis
Simlink: Linkage Simulation Program
LDB: Morton's Location Database
SIGMA: System for Integrated Genome Map Assembly
HUSAR:
RMAP: computes restriction maps from single+double digest fragment length data
FOLDSPLIT: computes minimum energy of RNA/DNA segments
SEG: divides sequences in low and high complexity regions
HUSAR/GCG interface to the email server PredictProtein

Spanish EMBnet node (CNB)

WWW servers:
Official CNB home page is at http://www.cnb.uam.es/, Database of three-dimensional density distributions is available at http://indy.cnb.uam.es/Base.

New software:
(Available on CNB's anonymous FTP server ftp.cnb.uam.es)

Exhaustive sequence similarity program that searches using Smith-Waterman and Needleman-Wunsch algorithms on message passing parallel computers. The programming language is C with PVM. The approach uses a modified self scheduling algorithm for better efficiency and is able to accomodate heterogeneous systems. The programs have been tested on systems ranging from workstation farms to the Cray T3D parallel computer. For a fuller description seeTrelles-Salazar, Zapata and Carazo (1994) CABIOS 10(5):509-511.

ORF (Optimal Region Finder): a program for finding the shortest regions from which reliable phylogenetic relationships can be obtained within a set of aligned sequences. The aim of the program is to optimize sequencing experiments in which a lot of samples need to be analysed by reducing the number of bp necessary to obtain a correct classification. The program runs under VMS or OpenVMS operating systems. Details of the algorithm can be obtained in "Martin, M.J., Gonzalez-Candelas, F., Sobrino, F. and Dopazo (1995) A method for defining the position and size of optimal sequence regions for phylogenetic analysis. J. Mol. Evol.(in press)" and also in the documentation provided with the program.

Finnish EMBnet node (CSC)

Server:
SRS 4.0 WWW server installed:
href="http://cypress.csc.fi:8001/srs/srsc"

Databases:
(All new local databases are available through SRS)

SWISS_NEW: updates to SWISS-PROT (GCG)
NRL_3D: sequences from PDB (GCG)

PROSITE: sites and patterns in proteins
ECDC: E. coli sequence database
OMIM: on-line Mendelian Inheritance in Man
ENZYME: enzyme nomenclature
CPISLE: CpG isle
LiMB: Listing of Molecular Biology Databases
TFSITE: Transcription factor sites
TFFACTOR: Transcription factors

Software:
SAPS: Statistical Analysis of Protein Sequences
fastDNAml: maximum likelihood method for large DNA data sets

HGMP-RC

People:
The HGMP-RC now has more than 3160 registered users.

New Hardware:
Two IBM RS6000s with 0.5Gb of memory for running Linkage analysis programs. A Silicon Graphics Power Challenge with 1Gb of memory for running sequence database searches. A SparcServer 20 which has enabled us to move the Mouse Backcross Database onto a dedicated machine.

System upgrade:
We are upgrading our SUN machines to Solaris 2.4

New Software:
HOMOZ: Very rapid multipoint mapping of disease genes in nuclear pedigrees, including homozygosity mapping
PROMOTER SCAN: Find putative eukaryotic Pol II promoter sequences in primary sequence data
FGENEH: Predict gene structure in complete Human DNA
FEXH: Predict exons in partially sequenced Human DNA
HEXON: Predict internal exons in Human DNA
HSPL: Splice site prediction in Human DNA
GAS: for reading, writing, sectioning and performing statistical analyses on phenotypic and genotypic data
HANDLINK: Tabulates lodscores (z) for any given set of recombination fractions (R) and sums them (Z) over a number of families
DOLINK: Outputs the results of genetic analyses in a variety of formats

New Newsletter:
The latest edition of the HGMP newsletter 'Genome News' is available by applying to 'admin@hgmp.mrc.ac.uk'.

Hungarian EMBnet node (ABC)

Server:
SRS 4.02 SRS-WWW server (http://hubi.abc.hu/srs/srsc)

Databases:
LIMB: Listing of Molecular Biology Databases
EPD

Software:
GCG 8.0 + WPI, + remote blast searching capabilities
Nentrez + nclever clients for browsing Entrez databases
DCSE 3.3.1 for X-Windows

Israeli EMBnet node (INN)

Hardware:
Upgraded Bioccelerator computer: 320 Million Matrics/sec, 16 Unit processors

Servers:
SRS 4.0 WWW server
BIOCCELERATOR E-mail server.
To use bicserv, send a formatted e-mail message to:
bicserv@sgbcd.weizmann.ac.il
An answer will be automatically sent back to you. Sending an empty file will retrieve a help document.
If you encounter any problems please send mail to:
lsestern@weizmann.weizmann.ac.il

Databases:
Full PDB mirror

Netherlands EMBnet node (CAOS/CAMM Center)

Server:
The SRS WWW-server has been upgraded to version 4.02 of SRS. The URL of the (experimental) server is "http://cammsg3.caos.kun.nl:8000/srs/srsc".

Databases:

HOVERGEN: database of homologous vertebrate genes by Duret, Mouchiroud and Gouy has been installed. A version accessible under SRS is under preparation.

PDB: prerelease files are currently updated via mirroring. It is envisaged that with the next full release of the PDB, we will switch to full mirroring of the entire release.

The catalogue of molecular biology programs, by Patricia Rodriguez-Tome, is now available under SRS.

BERLIN: database of 5S rRNA sequences (by Specht, Wolters, and Erdmann) has been put under SRS as well.

ECDC replaces ECD.

Software:
HOVERGEN: see above

tbob: the Blast Output Browser scans a directory for BLAST output files, and makes it easier to browse through the resulting hits.
MOLPHY: for MOLecular PHYlogenetics by Adachi and Hasegawa
It mainly features ML methods for protein and nucleic acid sequences. If only there was some proper documentation coming with it...
ASSP: a program to predict the accuracy of secondary structure prediction by Geoff Barton
AMPS: multiple alignment program by Geoff Barton
ALSCRIPT: format multiple alignments to postscript by Geoff Barton
SMITE: (Attwood, Beck, and Bleasby) enables querying the PRINTS database of protein motif fingerprints.

Norwegian EMBnet node (BiO)

Server:
SRS updated to version 4.03.

Software:
GCG 8.0 installed with html documentation.
EGCG 8.0 (beta testing) New versions of most EGCG programs including the Quick suite of programs, New exchangers, own genlib and applib, (see http://www.no.embnet.org/ and
ftp://www.sanger.ac.uk/~pmr/egcg8/release.notes)
WiseTools: package from Ewan Birney (opensearch, searchise,
pairwise)
Linkage analysis programs i (gs, mapmaker, fred, fastlink dolink, crimap and map)
Molecular Visualization Laboratory officially opened. Services include: Gert Vriend's WhatIf, M. Carson's Ribbons, P J Kraulis MolScript, D. McRee's xtalview, CAME's Prosa, Andrej Sali's Modeller and MSC xmol.

Polish EMBnet node (IBB)

People:
We have above 250 users from different molecular biology institutes in Poland.

Staff:
Piotr Zielenkiewicz - node manager.
Katarzyna Kruszewska
Andrzej Kierzek

Hardware:
SGI Challenge (2 R4400 processors, 64 MB RAM) and about 10 GB of
disk space.

WWW server:
The WWW server is installed and the pages are under development.
(home page http://info.ibb.waw.pl/)

Databases:
Non redundant nucleic acids database from both EMBL and GenBank.
New EMBL sequences are automatically fetched by NDT protocol.
PIR and SwissProt protein sequence databases are in use. The genome databases of yeast and arabidopsis managed by ACeDB system have been just installed. PDB structure database is available and the mirror of the PDB from Brookhaven is under development. The subset of PDB containing representative structures is also available.
Remote searches: BLAST and ENTREZ clients at NCBI.

Software:
GCG 8.0 and WPI.
ClustalV: program for multiple sequence analysis
Phylip: package for phylogenetic analysis
Staden: public domain version of the package

UK (SEQNET) EMBnet node

Server:
SRS-WWW 4.03

Software:
GCG 8.0
AMPS, ASSP & ALSCRIPT
MasPar and MPsrch software available via email server (send the word 'help' in the body of an email message to mpsrch@dl.ac.uk). This service is also available via local GUIs.

Teaching:
The PPS (Principles of Protein Structure) course is available with the URL http://www.dl.ac.uk/PPS/index.html

Research:
A three year position has been filled at Daresbury for the development of an intuitive GUI for molecular biologists. We are integrating all common molbiol packages using a drag-and-drop icon based system called disGUise.

Announcements

The Third International Conference on Intelligent Systems for Molecular Biology (ISMB-95) will take place at Robinson College, Cambridge, England during July 16-19th, 1995.

THE DEADLINE FOR SUBMISSION OF PAPERS TO THIS CONFERENCE HAS NOW PASSED

The ISMB conference is intended to bring together scientists who are addressing problems in molecular biology using advanced computational methods including data modelling, machine learning, artificial intelligence, cognitive science, robotics, combinatorial/stochastic optimisation, adaptive computing, string & graph algorithms, linguistic methods and parallel computer technologies. The 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 in some other way exhibits the emergent properties of an "intelligent system." In general these methods will have been validated on real data sets or have clear practical applications.

The four-day conference will feature introductory and advanced tutorials (on July 16th), and presentations of original refereed papers, posters and invited talks (on July 17th-19th).

FURTHER INFORMATION

For more information concerning the conference, tutorial submissions and travel fellowships see the ISMB'95 World Wide Web page:ftp://ftp.icnet.uk/icrf-public/ismb/ismb95.html
email: ismb95@biu.icnet.uk
Fax : +44 (0) 171 269 3067
Organising Committee:
Christopher Rawlings, ICRF, UK
Dominic Clark, ICRF, UK
Russ Altman, Stanford U, USA
Lawrence Hunter, NLM, Bethesda, USA
Thomas Lengauer, GMD-SCAI, Germany
Shoshana Wodak, UCMB, ULB, Belgium

INTERNATIONAL CONFERENCE ON MOLECULAR STRUCTURAL BIOLOGY Organized by the Austrian Chemical Society Vienna (Austria), September 17-20, 1995

TOPICS:

1) The Impact of Molecular Biology on Structural Biology
2) Biomolecular Structure Determination
 a) X-Ray Diffraction
 b) NMR Spectroscopy
3) Dynamics and Function of Biomolecules
4) Computational Methods
5) Protein Engineering and Design

Speakers include:

Tom Blundell(UK), Hans-J. Boehm(D), Charles Cantor(USA),
Christopher Dobson(UK), Manfred Eigen(D), Alan Fersht(UK),
Robert Huber(D), Oleg Jardetzky(USA), Martin Karplus(USA),
Joseph Lakowicz(USA), Paul Roesch(D), Ilme Schlichting(D),
Jeffrey Skolnick(USA), Malcom Walkinshaw(CH)

Posters are invited on any of the conference topics
Outstanding posters will be selected for 20 minute oral presentations

Abstract deadline: MAY 31, 1995

Head of Scientific Committee: P.Schuster

Participation Fees: Regular Participant 4000 ATS
 GOeCH Member 3500 ATS
 Student 2000 ATS

For further information and preregistration, contact:

A.Kungl
Gesellschaft Oesterreichischer Chemiker
AG Biophysikalische Chemie
Nibelungengasse 11
A-1010 Wien, Austria
Tel.: (43) 1 587249
FAX.: (43) 1 587966
e-mail: msb95@helix.mdy.univie.ac.at

THIRD INTERNATIONAL WORLD-WIDE WEB CONFERENCE
Technology, Tools, and Applications
April 10-14, 1995, Darmstadt, Germany

Organizer:
Fraunhofer Institute for Computer Graphics in Darmstadt, Germany. More information: http://www.igd.fhg.de/www95.html

SPECIAL WORKSHOPS ARE SCHEDULED, INCLUDING ONE FOR BIOLOGY:

WORKSHOP E is to be held in the Audimax area at Monday, April 10, 1995, from 2 to 5:30 pm.

We will start with users, go to analysis of genes and sequences, proceed with a program package interface and end with high-tech and 3D future. All talks are expectedly 20 minutes and 10 minutes discussion. More details on talks at http://www.ch.embnet.org/bio-www/www95.html and http://www.igd.fhg.de/www/www95/workshops/work-e.html for most recent schedule changes.

14:00 Introduction; overview on submitted contributions as suggested by colleagues not present at the conference
R.Doelz, Workshop chair, BioComputing Basel, Switzerland

14:30 Experience and Expectation A WWW User Report
J.Suehnel, Inst.f.Molekulare Biotechnologie, Jena, Germany

15:00 A WWW Server for Linkage Analysis, A.Audic, G.Zanetti, Center for Advanced Studies, Cagliari, Italy

15:30 World-Wide Molecular Biology Data Browsing
T.Etzold, EMBL Heidelberg, Germany

16:00 Sequence Analysis via WWW using a command-line application package, M.Colet, BEN, Brussels, Belgium

16:30 VMRL for the combination with 3D scenarios of biomolecules, H.Vollhardt, G.Moeckel, C.Henn*, M.Teschner*, J.Brickmann, Inst f. Physikal. Chemie, Darmstadt, Germany, and *SGI, Riehen, Switzerland

17:00 Round-table (all participants)
What would be specific developments in Biology which are required for WWW development?

Please note that all attendees *MUST* register for the workshop - see http://www.igd.fhg.de/www95.html for details.

SECOND EMBNET/CNB COURSE ON SEQUENCE ANALYSIS

EMBnet/CNB, the Spanish EMBnet node, is very pleased to annouce that a second EMBnet/CNB course on sequence analysis will take place in Madrid in June this year. The course is mainly aimed at the Spanish scientific community at an introductory level. A maximum of 20 students are expected to come to the CNB. Speakers include Drs. Carazo, Dopazo, Ficket, Guigo, Harper and Pezzi.

The course is partly supported by an EMBnet Training Grant.

Contact:
Jose Maria Carazo
BioComputing Unit
Centro Nacional de Biotecnologia-CSIC
Universidad Autonoma
28049 Madrid

Tf.: + 341 585 4543 or + 341 585 4510
Fax: + 341 585 4506

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] 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@ehpmic.mips.biochem.mpg.de)
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

[FI] CSC (Heikki.Lehvaslaiho@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)
Human Genome Mapping Project Resource Centre,
Hinxton, Cambridge, United Kingdom

[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 (attimonelli@mvx36.csata.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] SEQNET (bleasby@daresbury.ac.uk)
DRAL Daresbury Laboratory,
Daresbury, United Kingdom

Answers from the Christmas issue (embnet.news 1.2) Leisure Page:

The cryptic (complementary if ungrammatical) message was:

EMBNETWISHESMERRYCHRISTMASANDAHAPPYNEWYEAR

The response to the Xmas Xword was overwhelming: at least two people attempted it.
First prize (a weekend up a mountain with Alan Bleasby): Alan Bleasby,
Second prize (two weekends up a mountain with Alan Bleasby): Rodrigo Lopez.

Answer:

```
SUPREME#PASTEUR
I#O#I##S#S##U#A
TOLERANT#PATRON
T#S#E#BARI##O#K
I#K##V#R#R#OPAL
NEANDERTHAL#E#E
G##O#R#S#T#RA#S
#PHENYL#SILENT#
C#AL#B#I#O#A##C
H#R#PRIMENUMBER
EDDY#I#P#S##E#A
E#W##TOOL#S#A#C
SCAMPI#RAGNAROK
E#R##S#T##O#D#E
SLEIGHS#NOWISER
```

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.dl.ac.uk/embnet.news/vol2_1/contents.html
http://ben.vub.ac.be/embnet.news/vol2_1/contents.html
http://www.ch.embnet.org/embnet.news/vol2_1/contents.html
http://biomaster.uio.no/embnet.news/vol2_1/contents.html

A Postscript version (ISSN 1023-4144)is also available. You can get it by anonymous ftp from:
s-ind2.dl.ac.uk in the directory pub/embnet.news/
dbmdec5.ulb.ac.be in the directory pub/embnet.news/
ftp.no.embnet.org in the directory pub/embnet.news/

Back issues.
Online:
http://www.ch.embnet.org/embnet.news/info.html
Postscript by ftp:
s-ind2.dl.ac.uk in the directory pub/embnet.news/
dbmdec5.ulb.ac.be in the directory pub/embnet.news/
ftp.no.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, DRAL Daresbury, 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

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

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

Rodrigo Lopez, BiO, University of Oslo, NO
(rodrigol@biotek.uio.no)
FAX +47-2-2694130
Tel +47-2-2958756

embnet.news
Vol.2, No.1, 1995
21st March 1995
ISSN 1023-4144