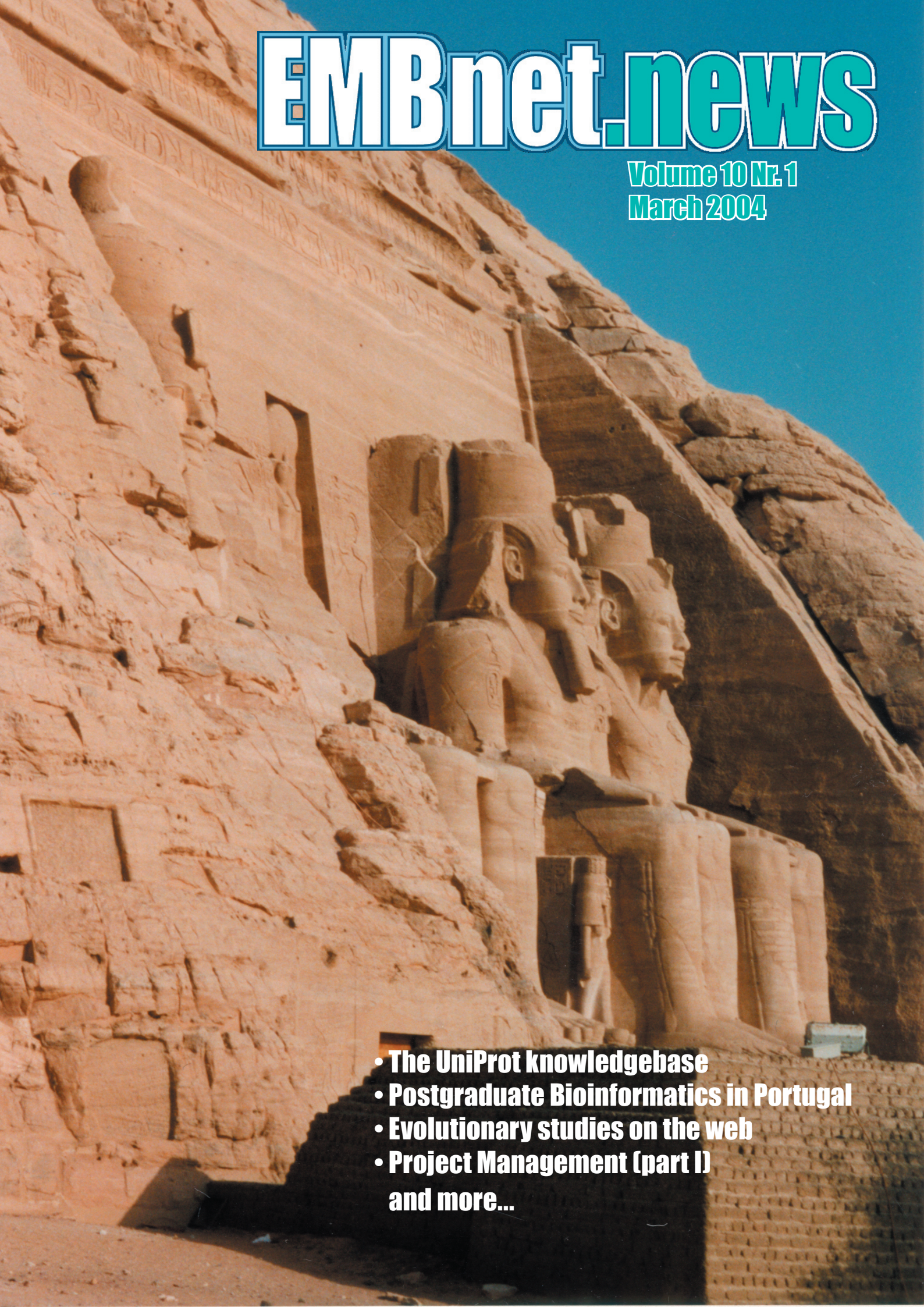
EMBnet.news

Volume 10 Nr. 1
March 2004



- The UniProt knowledgebase
- Postgraduate Bioinformatics in Portugal
- Evolutionary studies on the web
- Project Management (part I)

and more...

Editorial

The creation of the UniProt knowledgebase in December 2003 illustrates a unification of forces. Just as in ancient Egypt the strength and skills of many workers were combined to build one of the most impressive human creations, researchers in today's proteomics field are working together to achieve a common goal.
While UniProt may not last a millennium (we must stay humble), other joint projects already proved to be successful (e.g., InterPro, and of course, EMBnet!) the impact of these collaborative projects might incite other to follow their example and join forces. What about a UniRNA, UniMicroarray, Uni*?

In this issue you will also discover how to use the PHYLIP package on the web, how to manage a project, and a report of the Portuguese Post graduate programme in Bioinformatics.

The editorial board: Erik Bongcam-Rudloff, Laurent Falquet, Pedro Fernandes, Oscar Grau, Gonçalo Guimaraes Pereira

Protein Spotlight

Protein Spotlight (ISSN 1424-4721) is a periodical electronic review from the SWISS-PROT group of the Swiss Institute of Bioinformatics (SIB). It is published on a monthly basis and consists of articles focused on particular proteins of interest. Each issue is available, free of charge, in HTML or PDF format at
`http://www.expasy.org/spotlight`

We provide the EMBnet community with a printed version of issue 44. Please let us know if you like this inclusion.

Publisher:

EMBnet Administration Office
c/o Jack Leunissen
Dept. of Genome Informatics
Wageningen UR, Dreijenlaan 3
6703 HA Wageningen
The Netherlands
Email: Jack.Leunissen@wur.nl
Tel: +31 317 484074

Contents

Cover picture: Two statues of Ramsès II of 20 m high at the entrance of Abu Simbel temple, September 1997 [® Monique Zahn & Marta Puente]

A great achievement within the realm of Proteomics: UniProt, the universal protein knowledgebase

Vivienne Baillie-Gerritsen
Science Communication
Swiss Institute of Bioinformatics
Swiss-Prot Group
CMU - 1, rue M.-Servet
CH-1211 Geneva 4
Switzerland

In December 2003 a press release, which informed the world at large of the birth of yet another database in the Life Sciences, was enthusiastically launched. The news may have sounded tedious to some but to a great number of life scientists, whose research is based on proteins and the fascinating world these macromolecules revolve in, the news was paramount. Here is a database – no better still a knowledgebase – which is a central hub not only for all existing public information on protein sequences but also a number of tools for their analysis and a colossal amount of cross-references for all kinds of additional knowledge. And to which researchers can refer for free.

UniProt
the Universal protein resource

PIR-PSD, the Protein International Resource protein sequence database, was the very first protein sequence database to exist. It was developed by the late Margaret O. Dayhoff (1925-1983), a pioneer in bioinformatics, and existed in book form between 1965 and 1978 as the 'Atlas of Protein Sequence and Structure' before continuing its existence within the entrails of a computer. Dayhoff has actually been credited as one of the founders of the field of Bioinformatics. The Atlas' first edition held information on barely 65 different protein sequences.

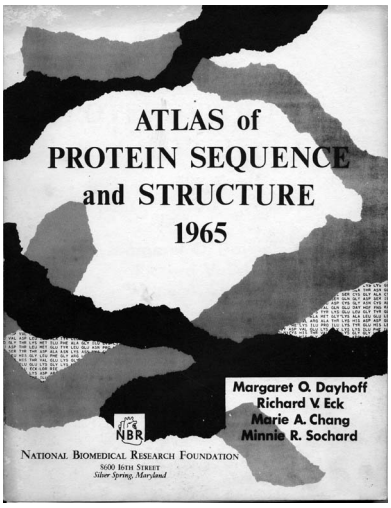


Figure 1 Cover of the Atlas of Protein Sequence and Structure 1965

Then came Swiss-Prot in 1986, a protein sequence database developed by Dr Amos Bairoch in Geneva. Why start a second one? Whilst working on his thesis, Bairoch felt the need not only to gather protein sequences but also to gather information related to each protein sequence – something he felt lacked in Dayhoff's database. Annotation was born… The first Swiss-Prot release boasted a little less than 4000 annotated protein sequence entries. Naturally, Swiss-Prot – like all databases – bounced along with the evolution of computer facilities; first stored on tapes, then disks and CD-ROMs , Swiss-Prot became available over the internet with the World Wide Web facilities. Today the Swiss-Prot knowledgebase prides itself with almost 150'000 entries and is developed by the Swiss-Prot group of the Swiss Institute of Bioinformatics in Geneva and the Swiss-Prot group of the European Bioinformatics Institute in Hinxton, England.

Swiss-Prot is renowned for its manual annotation. However, where something manual is involved, so are humans and humans do things at a human speed. In the past decade, the amount of biological information at hand has simply soared with

growing improvements in biotechnology. Gene sequences, and hence protein sequences, are churned out faster than you can swear. Manual annotation simply cannot keep up with the pace. That's why an additional database which complements Swiss-Prot, TrEMBL, saw the light of day. TrEMBL is a database of protein sequences which are automatically annotated and await further manual annotation by the Swiss-Prot team. It is an offshoot of the EMBL-Bank, a database of gene sequences, developed by the European Molecular Biology Laboratory whose main office is in Heidelberg, Germany. Currently, TrEMBL sports over one million entries.

So, here we have three large and distinct protein sequence databases: PIR-PDS, Swiss-Prot and TrEMBL, in addition to the multitude of more specialised protein sequence databases. Besides TrEMBL, which is the anteroom for protein sequences awaiting manual annotation, PIR-PDS and Swiss-Prot fulfilled the same kind of role. And it soon became apparent that it would be a wise step to merge the two in order to offer an even more comprehensive and non-redundant database of protein sequences to the scientific community. In 2002, the National Institutes of Health donated $15 million, over a period of three years, to do just that. In December 2003, the new merged database, UniProt – the Universal Protein knowledgebase – was completed and went on-line. What happened is that all the suitable PIR-PDS entries which were missing in Swiss-Prot + TrEMBL were incorporated into the latter, and cross-references are made to the original PIR entries for easy tracking back to PIR-PDS. Hence, UniProt is a bloated version of Swiss-Prot + TrEMBL. It sports two databases: one which is manually annotated and one which is only automatically annotated. The two continue to be called Swiss-Prot and TrEMBL, respectively.

Besides the UniProt knowledgebase, two other resources are an integral part of the Universal Protein Resource: UniProt Archive (or UniParc) and UniProt NREF (or UniRef). UniParc is the most comprehensive, publicly accessible pool of protein sequences. Which would make Swiss-Prot + TrEMBL quite redundant then you are thinking. Well no. The same protein sequence can be found in many different databases, not to mention more than once in one database. What UniParc does is assign a single identifier to an identical protein sequence found in a number of different sources and provides cross-references to the source databases, so that a UniProt user is not only aware of

Figuer 2 The UniProt Knowledgebase home page

the Swiss-Prot + TrEMBL protein sequence but also its existence within other databases. So, in effect, a UniParc entry is in no way similar to a UniProt entry: there is no annotation. It is a catalogue which informs you of all the various places to go for a given protein sequence. The third resource, UniRef, is itself made up of three sublayers: UniRef100, UniRef90 and UniRef50. The UniRef databases use recently developed procedures to combine closely related sequences into a single record. UniRef100 is a non-redundant version of all the sequences in the knowledgebase; UniRef90 collapses all the sequences that are at least 90% identical, or more, into a single record and UniRef50 collapses sequences that are at least 50% identical. UniRef50 not only speeds up searching significantly but also does not reduce the effectiveness of homology searching. Hence, the three UniRef sections give the user the opportunity to choose between either a fast search or a truly comprehensive one.

What are the aims of UniProt? UniProt provides the scientific community with a comprehensive, fully classified, richly and accurately annotated protein sequence knowledgebase, with extensive cross-references and query interfaces. What is more, the data is publicly available. As mentioned before, a UniProt entry is identical to the familiar Swiss-Prot entry. Besides the protein sequence, there are a number of comments made on a protein's function, post-translation modifications, subcellular location etc. as well as information on specific amino acids or domains within the sequence itself. Numerous cross-references are made to the literature from where information was extracted as well as to other databases and indeed to analysis tools which contribute to an ever-widening knowledge of a given protein sequence. So, we have certainly come a very far way from Margaret Dayhoff's 'Atlas of Protein Sequence and Structure' but without it protein sequence knowledgebases would not be where they are today.

Links

http://www.expasy.uniprot.org/
http://www.ebi.ac.uk/uniprot/

Bibiliography

Apweiler R., Bairoch A., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovan C., Redaschi N., Yeh L.S.
UniProt: the Universal Protein knowledgebase
Nucleic Acids Res. 32:D115-D119(2004)
PMID : 14681372

Planned Activities for 2004

Workshops

April 19-20: Videoconferencing systems, CSC Espoo FI.
eija.korpelainen@csc.fi

May 6-7: Regulatory sequence motif discovery, LCB Uppsala SE.
Erik.Bongcam@lcb.uu.se

August 30-31: BioMinT - Biological Text Mining Summer School, Geneva CH.
attwood@bioinf.man.ac.uk

September 16: Collaborative EMBnet workshop day. In connection with the AGM. Brussels, BE.
rherzog@dbm.ulb.ac.be

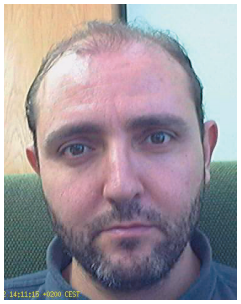
date to be announced
Federating the SRS servers within the EMBnet infrastructure, BE.
rherzog@dbm.ulb.ac.be

Meetings

September 17-19: Annual General Meeting (AGM), Brussels, BE.
rherzog@dbm.ulb.ac.be

For more detailed information, consult our web site: http://www.embnet.org/TM/workshops.php

Evolutionary studies on the Web



José R. Valverde
EMBnet/CNB,
Centro Nacional de Biotecnología, CSIC
Campus Univ. Autónoma
Cantoblanco, Madrid
28049, Spain

The analysis of evolution is more than just an interesting subject or a curious oddity. Certainly, knowing when dinosaurs died is appealing to the public at large, though it has little repercussion on everyday life. However, the study of evolution encompasses much more than that and actually has an impact on everyday life when used appropriately: from allowing us to identify relevant genes to tracing the origin of an epidemics and how it changes, or just tracing a donor for a transplant.

It is therefore of the utmost importance, for those who must deal with evolutionary techniques that applying them and understanding their results becomes as easy and efficient as possible. Furthermore, as its applications extend, new collectives must deal with technologies that seem the most alien to them since they were developed for a different collective with different ideas, needs and culture in mind. Ease of use is a must if we want the society (and ourselves) to benefit from the advantages of evolutionary studies.
In this article we have a look at Web interfaces for PHYLIP, a package for comprehensive phylogenetic analysis developed by Joe Felsenstein (see `http://evolution.genetics.washington.edu/phylip.html`).

PHYLIP

PHYLIP is a free package of programs for evolutionary studies that has become a standard used in many institutions for a number of reasons:
First, **PHYLIP** is free. Compared to other classical phylogenetic analysis packages (like e.g. PAUP) which are commercially distributed at very expensive costs, PHYLIP cost is certainly unbeatable; specially at academic institutions where funds are always at a stake, this supposes a serious advantage.

PHYLIP is open source: anybody may see how it works, amend and correct it if necessary and, what is more important, use it to develop new tools, methods and approaches. And so there are additional programs or enhanced versions of PHYLIP programs (like e.g. FastDNAML or PROTML) developed at other sites.
It is a comprehensive package including methods for most of the common phylogenetic analysis, and it certainly covers successfully most needs of the average and advanced user. While it is true that some methods are better represented than others, or that the philosophical tendencies of the authors in the interpretation of phylogeny are reflected in the package, one would need to go for a very expensive commercial package to get a more comprehensive integrated environment.

Programs in **PHYLIP** are well proven and tried. Most of them have been around for many years (well over one or two decades) and being open source have been scrutinized and corrected at large. Surely, new programs may have hidden bugs, but the wide professional user community and its very openness reduce them to a minimum.

On the minus side, **PHYLIP** lacks the niceties of a modern, graphical user interface, which draws aback many potential users, and might be somewhat slower than some of its commercial counterparts which -with current computers- is seldom significant.

WebPHYLIP

This is a very useful web interface developed exclusively for PHYLIP. Although taste and preferences widely vary, it still has some niceties missing in other interfaces, providing a really practical user environment.
WebPHYLIP (see `http://sdmc.lit.org.sg:8080/~lxzhang/phylip/`) was developed

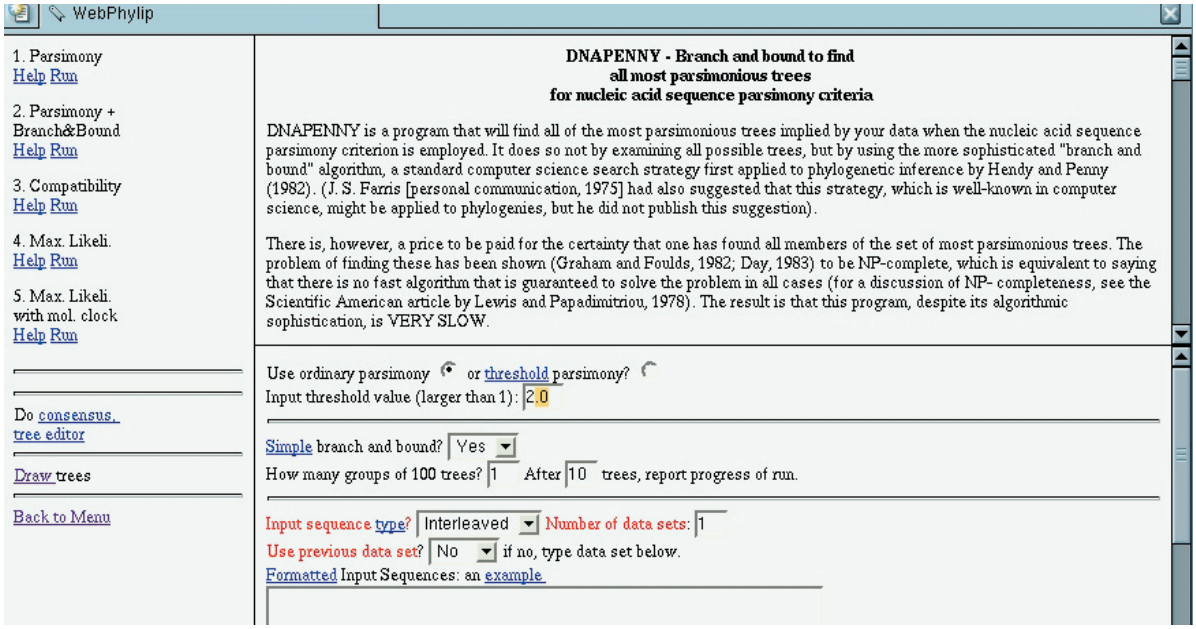
in 1999 by A. Lim and L. Zhang. It is a simple yet easy to use and powerful interface to most PHYLIP programs. The only ones missing are the interactive programs (which are unnecessary on a Web environment), a few programs of very narrow application (DNAINVAR, CONTRAST and FACTOR) and the unsupported programs MAKEINF and PROTML. Shortly put, you get access to many more programs than you do with EMBOSS and hence get extended phylogenetic productivity.

To use **WebPHYLIP** all you need is to install PHYLIP (from Joe's site) and the web interface (from WebPHYLIP site). Installation of WebPHYLIP is very easy and is described in the accompanying README file: it involves a configuration step using Perl and a building step with make and a C compiler. All in all, it is a lot simpler than the many-step installation required by EMBOSS and thus it is better if all you are interested in are PHYLIP phylogeny applications.

Once installed, you may proceed to use it. WebPHYLIP divides the screen in three parts: a hierarchically organized menu on the left, and two sections on the right: one above with the help needed to use any selected program, and one below with a form to fill in your experimental data (Figure 1).
The menu on the left has been organized so it is easy to locate the appropriate programs for your data and analysis type. It starts with a section on conversion tools so you may import data from other programs (like Clustal, Paup or GCG), and goes on to provide options for distance computations, data sampling and phylogeny methods for DNA, Proteins, Restriction sites, Gene frequencies, 0-1 data and distance matrix, as well as tools for analysing and plotting trees.
As you navigate options towards programs you may request at any time additional help (to be shown on the right, upside) or run the program (with the form on the bottom right). This allows for a very comfortable work method, where you may have both the program and its description and exhaustive help at the same time in your screen. You may even resize the different portions to suit your needs.

This kind of layout, when you are dealing with complex analysis, is a really useful life-saver. Sure enough you may get help with other web interfaces, on a separate pop-up window, but for the average user with not-so-big screens this is often more of an annoyance than a help and results in people avoiding documentation (and making mistakes). WebPHYLIP layout makes it easy to

Figure 1. WebPhylip example



keep everything at sight, which most people find more convenient and practical.

PHYLIP and EMBOSS

EMBOSS, the European Molecular Biology Open Software Suite (see `http://www.emboss.org`), developed within EMBnet to provide an integrated environment for molecular biological studies, has facilities for a wide number of sequence analysis tasks, and being free and open source as well, has become widely popular too. EMBOSS does not cover everything, most notably it lacks in molecular evolution tools.

Not to worry though: EMBOSS provides a mechanism for other application to be integrated within, and since there is already a good, well proven, free and comprehensive package for evolution, it made more sense to integrate it than to build a new one. In this way, PHYLIP programs have been integrated within EMBOSS as an EMBASSY application (an external application integrated with EMBOSS).

Thus, to be able to use PHYLIP within EMBOSS you must get and install all, PHYLIP (from Joe Felsenstein's site), EMBOSS and the "EMBASSY" interface between both (from EMBOSS site). In addition, you need to install the WWW user interface of your choice (e.g. wEMBOSS, W2H, SRS...).
Once integrated, any application (and PHYLIP programs are not an exception) is deal with as another EMBOSS application, and therefore, all EMBOSS user interfaces (Jemboss, wEMBOSS, W2H, SRS) will provide a seamlessly integrated access to PHYLIP applications (Figure 2).

The advantages of using PHYLIP within EMBOSS are many, but mainly they consist of taking advantage of an integrated environment where sharing sequences, analysis and results between applications is fully transparent.

PISE

We can't ignore the "interface of interfaces", **PISE**. Developed by Catherine Lethondal

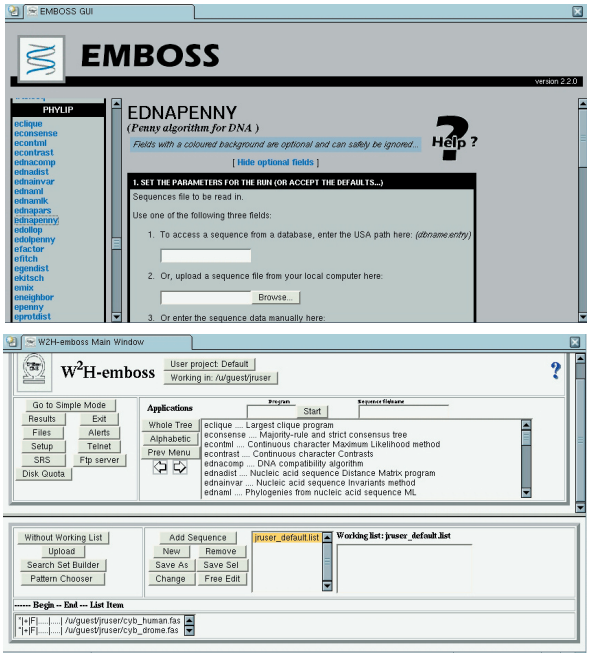


Figure 2 Examples of EMBOSS interfaces

at the Pasteur Institute (see `http://www.pasteur.fr/recherche/unites/sis/Pise/`) this is a generic framework for building web interfaces to command line applications

Almost any application maybe used with PISE after it has been appropriately described to the system using an XML schema. There are already schemas provided for many applications (EMBOSS and PHYLIP among them) and therefore normally you do not need to take care of these details yourself, just use the provided definitions. In a sense, you may include PISE among the above mentioned EMBOSS interfaces, although you don't need EMBOSS to run PHYLIP from PISE.
Using PISE you get a uniform user interface for all the supported programs. Actually this is not totally true: you get two interfaces for each: a basic interface for simple, quick use of the programs and an advanced form for more elaborate use where you can tweak options to your heart's content.

Installing PISE is not such an easy task however: although all you need are the programs and the interface generator (PISE) which is written in Perl, it is in itself a highly complex system that relies on many additional Perl libraries,

some of which in turn depend on additional libraries to various depth levels, but once you are done with all of them, you get a really comprehensive web environment.

When you access your PISE server you get a listing of supported applications, with links to both, the basic and advanced forms and to the program documentation, which open as independent web pages. This, as we have mentioned, often becomes a hindrance for the average user with a medium-sized screen, who is forced to switch among windows constantly to navigate.
Being automatically generated, PISE forms tend to be terse and less appealing than other, more elaborate or hand-crafter interfaces, but everything needed is there, and they usually come with basic descriptions of relevant fields (Figure 3).

Summary

When dealing with phylogenetic analysis, one would like to have the easiest interface possible while keeping access to advanced features. Several Web interfaces are available for PHYLIP programs, among which we have reviewed EMBOSS, WebPHYLIP and PISE.

WebPHYLIP provides a pleasant, helpful interface to almost all of PHYLIP programs, and is probably to be preferred when you are mainly interested in PHYLIP programs. It has been hand-crafted specifically for PHYLIP and that shows off. It is easier to install, easier to use, well thought out and its ability to integrate help in the main window allows for an easier introduction of the novice user and for an easier way to explore advanced features as well.

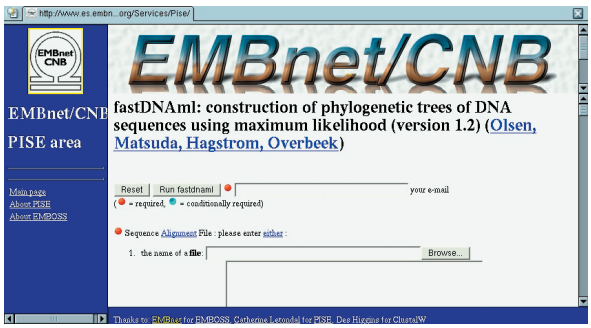


Figure 3 Example of PISE interface

If you (or your users) need to deal with other applications, integrating PHYLIP with then may prove a definitive advantage and you should consider a more generic interface for your work.

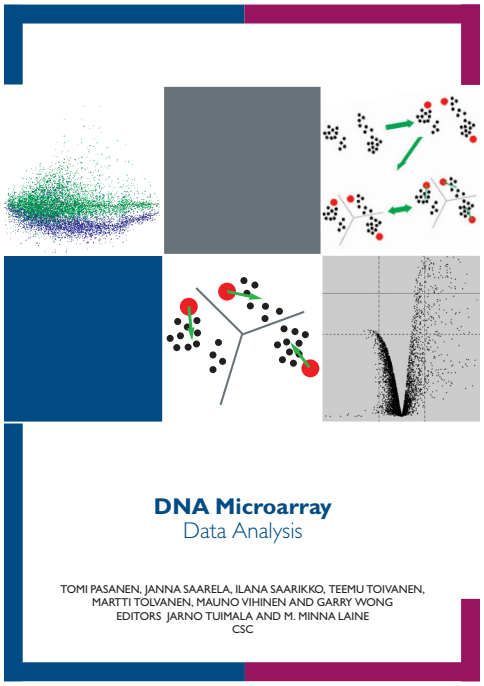
EMBOSS provides access to most common molecular biological applications, and as long as it covers all your needs, your best bet is to use it and integrate PHYLIP as an «EMBASSY» external application. Then any EMBOSS interface will do (Jemboss, wEMBOSS, W2H, SRS, PISE...). Just have a look and see which of them suits better your needs (a review of EMBOSS interfaces is beyond the scope of this article).
However, if EMBOSS is not enough for all your needs, and you still want to use one single interface for all your work, then PISE is certainly the way to go, despite the extra effort needed to install it. It will surely be terse and less appealing, but it allows you to integrate much more applications and process data with them.

Generally speaking, the best option is often starting with the more specific interfaces (WebPHYLIP for all of PHYLIP, an EMBOSS-specific one for PHYLIP+EMBOSS or PISE for a wider range of applications). As you go from generic to more specific, interfaces usually go from terseness to more tailored and full featured (albeit exceptions always exist). In case of doubt, install and try various options and decide which is the one that best suits you.

In any case, when speaking of phylogenies, you will always be interested in having **a good tree drawing program** that goes beyond the basic diagrams drawn by PHYLIP. Don Gilbert's Phylodendron TreePrint is one of the nicest web based tools and you should consider adding it to your toolbox (see `http://iubio.bio.indiana.edu/soft/molbio/java/apps/trees/`).

A guidebook for DNA Microarray Data Analysis

Rob Harper
Center for Scientific Computing
Tekniikantie 15 a D
02100 Espoo, Finland



CSC has produced a guidebook on DNA Microarray data analysis. It also has a series of PowerPoint presentations dealing with the subject, and stored in the media archive at Funet TV there are a set of videos.

This guidebook on DNA Microarray data Analysis was a collaboration between several Finnish researchers from different universities and research institutions.

The purpose of this book is to serve as course and teaching material to introduce the basic concepts of microarray data analysis. We hope that researchers who are starting microarray data analysis can benefit from the book.
The first edition of the guidebook was written by M. Minna Laine (chapters 4, 8 and 14), Tomi Pasanen (chapter 11), Janna Saarela (chapters 2 and 3), Ilana Saarikko (chapter 8), Teemu Toivanen (chapter 14), Martti Tolvanen (chapter 12), Jarno Tuimala (chapters 4, 6, 7, 8, 9, 10, 13 and 15), Mauno Vihinen (chapters 10, 11 and 12), and Garry Wong (chapters 1 and 5).

Each chapter has a section on suggested reading, which introduces some of the relevant literature. Several chapters also include data analysis examples using GeneSpring software.
We are very interested in receiving feedback about this publication. Please email your comments about this book to Jarno Tuimala: `jtuimala@csc.fi`

How to read the pdf-version of the book?

First, if you don't have it already, download Adobe Acrobat Reader (5.0 or later) software from Adobe (`http://www.adobe.com/`).

The whole document (13.2 MB) can be downloaded from
`http://www.csc.fi/oppaat/siru/siruwww.pdf`

Or four parts of the book separated into individual pdf files:
Part I Introduction (3.9 MB)
`http://www.csc.fi/oppaat/siru/sirupartI.pdf`

Part II Analysis (7.7 MB)
http://www.csc.fi/oppaat/siru/sirupartII.pdf
Part III Data mining (0.6 MB)
http://www.csc.fi/oppaat/siru/sirupartIII.pdf
Part IV Tools & data management (0.7 MB)
http://www.csc.fi/oppaat/siru/sirupartIV.pdf

PowerPoint presentations concerning Microarray analysis

A whole series of PowerPoint presentations on the subject of microarray analysis can be found at
http://www.csc.fi/molbio/opetus/sirukokous/sirupresentations.html
The presentations from the symposium called "Microarray data analysis developers' days" include the following topics:

1. The functionality of BASE, and how to extend it. **Carl Troein**, Complex Systems Division, Department of Theoretical Physics, Lund University, Sweden

2. J-Express - an integrated system for microarray data analysis **Inge Jonassen**, Department of Informatics, University of Bergen, Norway

3. Expression Profiler **Jaak Vilo**, EGeen Inc., Tartu, Estonia

4. What do Biologists want from Microarrays. What do computer scientists need to know **Garry Wong**, A.I. Virtanen Institute, University of Kuopio

5. Data mining of microarray data for inference of biological insight **Mauno Vihinen**, Institute of Medical Technology, University of Tampere

6. Microarray bioinformatics at TUCS/CBT Bioinformatics Laboratory **Tapio Salakoski**, University of Turku and Turku Centre for Computer Science

7. Analysis of microarray data: current research and future challenges **Jaakko Hollmén**, Laboratory of Computer and Information Science, Helsinki University of Technology

8. ArrayExpress and more **Misha Kapushesky**, European Bioinformatics Institute, Cambridge, UK

Microarray video files on the FUNET media archive

A symposium called "Microarrays and data mining" dealing with microarrays was recorded and stored to the videoserver of CSC in Mpeg1 and RealVideo-format. To be able to watch the files, your need to have an appropriate video player installed. The videos are available from
http://www.csc.fi/molbio/workshop.html

Listed below are the names of the lecturers and the topics under discussion:

Victor Jongeneel
Using EST data to produce a catalog of genes: the challenges of clustering, assembly and probe choice for cDNA arrays
Heikki Mannila
Global analysis of expression data sets
Esko Ukkonen
Computational Problems in Structural Biology
Garry Wong
Analysis of gene expression data using self organizing maps
Yike Guo
Kensington Discovery Edition: Towards a Web Platform for e-Science
Jaak Vilo
Extracting information from microarray data
Mauno Vihinen
Microarray analysis: development of B and T cells
Benedict Arnold
Supporting scientific discovery through the integration of heterogeneous data sources and applications
Michael Zhang
Super-Paramagnetic Clustering (SPC): A New Tool for Large-Scale Gene Expression Data Analysis

PGBIOINF – a Portuguese Post Graduate Programme in Bioinformatics

Pedro Fernandes
Head, Bioinformatics Unit of the IGC
Member of the PGBIOINF Executive Committee

Teaching Bioinformatics

Bioinformatics' usefulness comes from the fact that it addresses biological problems and finds solutions for them in the information world when they exist.

The Life Scientist of the early 21st century is continuously referencing the objects in his research to the existing knowledge through the vast amounts of information that are made available in the public and private domains. Bioinformatics is many times inevitable and the need to train people for using it in their daily life has become evident.

The Instituto Gulbenkian de Ciência (IGC) hosts the Portuguese node of the EMBnet since 1991. In an effort to address the needs of the local community, in particular – but not only – the ones that use the facilities provided by the national node, the IGC has stepped-in as a provider for such training skills and, in the past five years, has organized training courses in which more than 600 scientists have had first level or specialist education in Bioinformatics. A continuous operation was created, named the Gulbenkian Training Programme in Bioinformatics (GTPB). Training courses are held for 20 attendees, lasting for 4 to 5 days. Numerous teachers, many from the EMBnet community, were invited to run the courses. The training courses provide the training needed to use the tools, not the knowledge to build or modify them. The entry-level ones are, naturally, too broad in scope to allow for any in-depth knowledge. The specialized ones are focused on specific subjects such as Molecular Evolution or Population Genetics but their design is such that acquiring the skill to use Bioinformatics tools is still the main focus.

User level education is insufficient for the professional practice of Bioinformatics, in other words, for the research and the development that is needed to create and continuously improve the methodology and the tools. The PGBIOINF is an educational step that aims at the creation of young, productive professional Bioinformaticians, therefore it focuses on the methods themselves and the research around them, keeping in mind the broad knowledge that a professional Bioinformatician must have in order to be able to borrow experience across the various subjects that he encounters in the profession itself. Bridging between these areas is therefore an essential ingredient in the design of PGBIOINF.

Professional Bioinformaticians need to acquire considerable knowledge both in Biology and in Computer Science in order to understand biological problems from the information side, conceive new solutions, build programs to query large databases and screen the results with reliability (quality) criteria.

Origins of the Programme

The PGBIOINF was created in 2002, as a partnership, by the Instituto Gulbenkian de Ciência (IGC) and the Faculdade de Ciências da Universidade de Lisboa (FCUL). In the IGC, providing user level training extensively (120 course attendees per year for the last 5 years) had revealed that a deepening of the level of such courses was possible, extending them into the research side. Many of the teachers involved in providing the training courses in Oeiras (GTPB) are active researchers in Bioinformatics. Only a few of the course attendees would be able and willing to follow Bioinformatics professionally as a research activity, which was normal, but receiving applications from very capable undergraduates was becoming

more frequent. Such undergraduates were finishing their first degree in Biology, Biochemistry, Medicine, Engineering, Mathematics, etc. and had discovered that Biology and its problems could be addressed with Computer Science methods. Naturally the source of such students was mainly the Portuguese Universities but, as the GTPB receives foreign students quite frequently, we were aware that some demand from foreign countries could be expected.

Several Portuguese Universities were contacted to establish a partnership that could conciliate the necessary competences and the ability to provide degrees. The idea was to start with Post-Graduate education and reach the possibility of offering a PhD within a year or two. The FCUL was building similar plans and looking for complementary teaching resources, namely the means to provide advanced thematic seminars, so a program was set-up in only a few months and was approved at the Universidade de Lisboa. The proposal within FCUL was initially headed by the departments of "Informatics" and "Chemistry and Biochemistry" but, within a month or two the "Statistics" and the two departments of Biology were in the team.

Scope and objectives of PGBIOINF

PGBIOINF aims clearly at creating new competences in the field, starting mainly from freshly graduated students. Bioinformatics is to be taken by them as a profession, not an additional skill. To do this, PGBIOINF is providing them with a coherently designed framework of knowledge acquisition opportunities, in close contact with the community of users and a selected set of active Bioinformaticians that teach the advanced topics with a strong emphasis on the current research in the area.

"Biology and medicine are moving from bench-based to computer-based sciences as models replace some experiments and complement others" as stated by charter of the Biomedical Information Science and Technology Initiative (BISTI) created at the NIH. Biologists can hardly do without, but their focus is and must be on the Biology itself not on the ways that need to be professionally created to handle these problems and to tackle them from the information side. And as the information and knowledge repositories grow, the need for smarter ways to use them grow even more, and the need for people that can be professionally dedicated to this activity becomes increasingly evident.

Creating such professionals is a positive move towards this set of objectives. It is also an affirmative acknowledgement of the need to provide well-equipped intellects into a field where traditionally only self-made skills have been used for years. It is our firm conviction that a Bioinformatician that works in the prediction of protein function has nothing to loose if he also masters Phylogenetic analysis, for example, and that all Bioinformaticians can only work much better is their knowledge in Biostatistics is fresh and articulated.

Future perspectives for PGBIOINF students

PGBIOINF students are able to pursue their studies to a Master's (in Research) by writing a Master's Thesis and to step directly into a PhD project. This can be done at FCUL, in which these degrees have been created in early 2003, or elsewhere. This is also easy because PGBIOINF students are graded according to the methods of the European Credit Transfer System (ECTS) that allows for the easy interchange of students between schools that have implemented the EU directives, following the decisions of the Bologna convention and the "Lisbon strategy" (see reference URLs at the end of this paper).

PGBIOINF students are ready to step into a Master or PhD project once they finish, but they also have the aptitudes that allow them to take a professional job in several existing and foreseeable areas of employment such as:

- Biotechnological industries
- Pharmaceutical companies
- Food industry
- Research Laboratories
- Nature conservation
- Non-Governmental organizations

Moreover, many of these students will be able to respond to new opportunities in business that naturally arise from making their aptitudes available in a market that extends easily beyond national borders. They can also decide to create and manage Bioinformatics support units and actively deliver services to the research community in a completely autonomous way.

Programme structure

An executive committee with members of FCUL and IGC manages the PGBIOINF. The committee defines major orientations for the courses, chooses the teaching staff, selects the student candidates and monitors the progress of the Programme.

The PGBIOINF is organized in two main parts:

Part A follows a more classic university model. It has a workload 30 ECTS credits, consisting of classes and laboratory sessions at the FCUL, complemented with weekly trips to IGC to ensure a direct contact with the local research labs, the local scientists (many of them Bioinformatics users) and the Library is taken. Students obtain competencies in the various subjects that provide the basic science supporting Bioinformatics: on one hand, Chemistry, Biology and Biochemistry; on the other hand, Data Analysis, and Systems Design and Programming.
The choice of courses allows each student to acquire the skills that are needed to complement his/her graduate profile. For example, a Software Engineer picks-up Biochemistry and Genetics, while a Biologist picks-up Programming and Database Design.
The executive committee, according to the student's profile, initially takes this choice, and each student receives individual follow-up to ensure that not only the choice but also the learning rates are appropriate throughout the Programme.

Part B is structured after a model of thematic seminars and intensive courses conducted by well-recognized experts, also totalling 30 ECTS units, to be held in the IGC. At this time, albeit

Table 1. Part A Courses

Title	*Type*	*Credits (ECTS)*
Introduction to Bioinformatics	3	6
Biostatistics	3	6
Biologically Inspired Algorithms	3	4
Biodiversity	3	4
Introduction to Programming	2	4
Algorithms and Data Structures	2	4
Introduction to Databases	2	6
Molecular Genetics	1	4
Molecular Structures, Metabolism and Diversity	1	4
Genetic Engineering	1	4
Chemistry and Biochemistry Laboratory	1	2

Type 1: for Engineering, Maths and Computer Science students (Info Profile)
Type 2: for Biology, Biochemistry and Medicine Students (Bio Profile)
Type 3: for all students

Table 2. Part B Courses

Title	*Type*	*Credits (ECTS)*
Protein Structure and Function	3	2
Genetic Analysis	3	2
Phylogenetics and Molecular Evolution	3	4
Population Genetics	3	2
Functional and Comparative Genomics	3	4
Genetic Expression and Microarrays	3	2
Data Warehousing and Data Mining	3	2
Limits and Expectations in Bioinformatics	3	2
Gene Prediction and Identification	3	4
Proteomics, Transcriptomics and Metabolomics	3	2
Gene Ontology	3	2
Population Dynamics and Epidemiology	3	2

Type 3: for all students

their different origin and interests, students are sufficiently synchronized in aptitudes to be able to attend these seminars, in which major specialized Bioinformatics topics are covered in considerable depth. The main methodologies are explained in detail, not only in their theoretical roots but also in terms of the details of the several implementations that have been made available so far. Bringing recent publications to classroom discussions, in the Journal Club style, reveals the current status of research in each of these topics and allows each student to better judge his/her own interests. It also gives them exposure in the sense of preparing them to address audiences, a practice that is not sufficiently stimulated in undergraduate courses and is an essential ingredient the life of any scientist.

The first edition

The first edition of the PGBIOINF started late, relative to the normal academic year. The announcement of the Programme could only be out in the beginning of the summer, thus conditioning the rest of the calendar. There were 24 candidates, 6 from abroad, of which 9 were admitted. The first lecture was held on Jan 4th 2003. Part A ended 3 weeks beyond schedule in consequence of adjustments to avoid superimpositions and to the need to ensure a deeper training in fundamental subjects such as Biostatistics (mandatory for all students). Part B started in Oeiras in the planned format and students gradually adapted to the seminar form of the courses. Most of the students have chosen a subject to follow PGBIOINF with a Master's

thesis. The only exceptions are the ones that were already in PhD programmes and had interrupted them to follow the PGBIOINF.
The appraisal from the teaching staff was rather good both in the course's organizational features and in the overall quality of the end product: the student's skills.
IBM chose the student that obtained the highest grades for the "Student Achievement Prize" of 2003, that was awarded to two post-graduate students per country.

The second edition

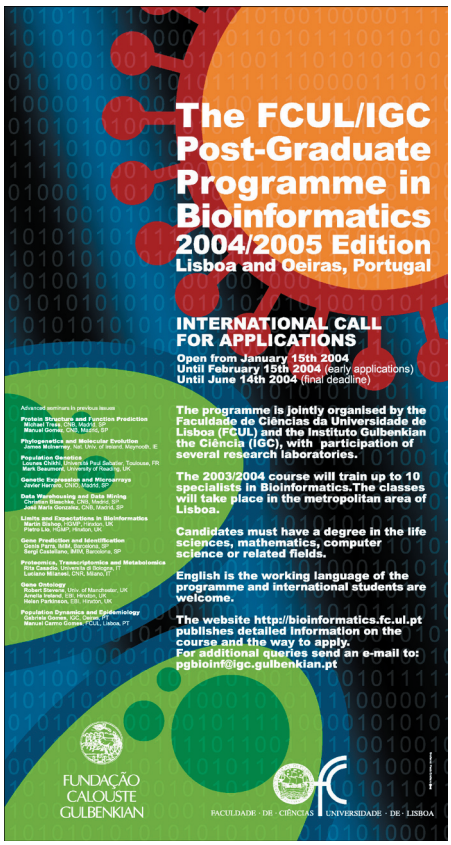
The 2003-2004 edition was announced in January, thus allowing for a longer period for the applications. 36 applicants responded, many of them with excellent first-degree classifications. This time a considerable amount of candidates had a much better idea of where this Programme can take them in terms of their professional careers. Part A began on Sept 8th 2003, synchronized with most academic activities, which makes the planning and logistics far easier to manage. Part B is about to begin (March 22nd 2004). We now have 8 students, 6 of them are Portuguese.
Considerable amounts of experience from the first edition were used in order to make fine adjustments to this edition. For instance, more time was given to Biostatistics as we found that progress is slower than we thought, and teachers of the Computer Science courses were encouraged to use working examples from Biology in place of more abstract ones.

Present Status

At this moment, the call for applications for the third edition is open. If possible, due to the demand, we will admit more than 10 applicants. We are eager to admit foreign students, so please consider recommending it to members of your local community.

At this point I consider that the key characteristics that make the PGBIOINF an attractive option for a fresh graduate interested in practising Bioinformatics professionally are:

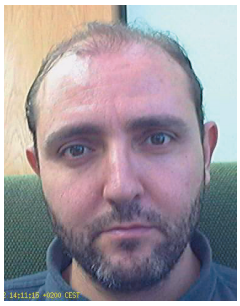
- Mixture of student origins and skills
- Taught and fully documented in English
- Graded in ECTS (student mobility)
- Advanced seminars in 12 subjects with research emphasis
- Individual attention to the student (small number of students, time to talk and discuss)
- Problem-driven classes, deep into the biological subjects



For further reference

Course website
http://bioinformatics.fc.ul.pt
European Credit Transfer System (ECTS
http://europa.eu.int/comm/education/programmes/socrates/ects_en.html
European directives towards mobility and educational strategies (Bologna, "Lisbon strategy")
http://europa.eu.int/eur-lex/en/com/rpt/2004/com2004_0021en01.pdf

Project Management (part I): a gentle introduction



José R. Valverde
EMBnet/CNB,
Centro Nacional de Biotecnología, CSIC
Campus Univ. Autónoma
Cantoblanco, Madrid
28049, Spain

Basic concepts

In this article we introduce some basic concepts about Project Management. Here you will find the basic concepts about what a project is, the respective roles of project members and a general overview of the Project Life Cycle (PLC).

Introduction

Project Management is mainly about *common sense*[1]. However, it covers such a wide variety of tasks and abilities that it is easy to forget some if you rely on some unconscious, vague confidence on your experience. Having a structured view of the overall process and its most relevant details will help you to avoid forgetting some odd important step.

It is highly advisable to read about the discipline. There are many excellent books, and you will find that many of them concentrate on exposing the acquired experience (both good and bad) in previous projects which, when well written, makes in addition for an entertaining reading. You will find references to books and pointers in the Project Management section of the EMBnet/CNB portal:

`http://portal.cnb.uam.es/tiki/tiki-directory_browse.php?parent=46`

Whether you are a project manger, you plan to become one in due time or you are simply just a worker in a project, learning what a Project is all about will help you better understand the process and make it work to your advantage.

Actually, it is arguable that you are not a Project Manager yourself. As we shall see, managing a project involves delivering a product to some interested stakeholders. Whether you are Research Policy Coordinator for the EU, a Group Leader at EMBL or an intern is moot: you still have someone (your client, a funding agency, or your boss) that is interested in the timely delivery of some results that you are responsible for, and thus you must manage your own projects (even of it is with a team of only one –**you**!).

What is a project?

A project, as defined in the "PMBoK[2]" is *a temporary endeavor undertaken to create a unique product or service*.

This has several implications:

It is **temporary**, something with a definite start and end times, not a continuing labour like, e.g. ongoing maintenance or provision of a service. There are those witty ones who say that you know when you start a project but not when it will end. This is a typical error: you must be able to realize when a project goal cannot be achieved and stop there.

It is an **endeavour**, a complex process that requires a number of resources to be put into place in the appropriate order to succeed. Usually a project is executed by a team, whose existence is tied to the project itself, but it may as well be a one-person task. In addition it is a complex enterprise, composed of one or more tasks and associated resources that must be orchestrated to complete the goal.

It is **unique**, meaning that it purports to create something unique. Its goal may be to create thousands of the same unique product, but still it attempts to create something new or in a new way. A repetition of something that has already been done does not need the entire project apparatus, it just needs to go by the book.

It has a **definite goal**, which is the production of a **product or service**. This goal determines the creation of a new project, defines when the project ends and the whole characteristics of the project.

1 The least common of all senses
2 The Project Management Body of Knowledge, compiled, maintained and published by the PMI, the Project Management Institute, one of the main reference bodies for PM in the World.

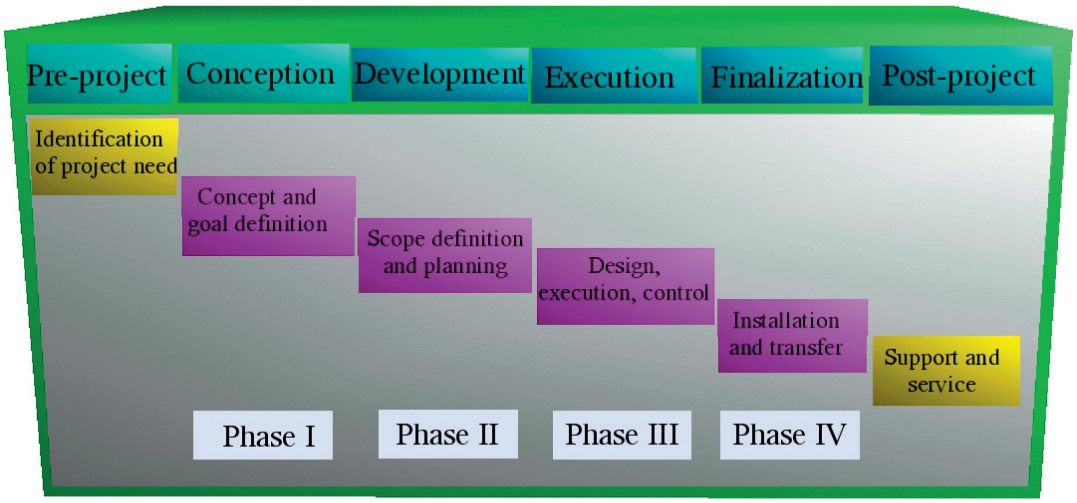


Figure 1 The Project Life Cycle

A Project Life Cycle

A project goes through several phases in what is termed a Project Life Cycle (Figure 1).
There is some general consensus on the division, although not on the names. But basically, it is obvious that before a project comes to live, something must happen: an idea must pop up about a need to be fulfilled. This idea, when accepted will result in the onset of a process whose goal will be to fulfil the need addressed.

The conception phase

Once the need has been accepted, work starts towards satisfying it. While you will probably find at most sites that corporate culture is "let's go straight for it without costly delays", common sense dictates that it is never a good idea undertaking a new enterprise without giving it a second thought and some planning. So, before rushing up, stop and take time to think.
The first step in every project should be the definition of the project goals. What do we want to achieve? The goal needs to be refined until everybody clearly understands what it is that we want to achieve. Otherwise, different people will start with incompatible conceptions and chaos will ensue, often resulting in a scandalous failure.

The development phase

Once we know what we want, it is time to make decisions. Ideally we would like to have everything, now, perfect and at no cost. Practically we can't. The old saying "Price, Quality, Time, pick two of three" is just as true: given enough time anything can be built to the most exquisite quality standards at marginal costs. Given enough money you may always build a quality product in very little time. If quality is not a problem, you can get anything now at minimum cost. But you can't get everything except in very rare occasions (Figure 2).

Therefore, before starting, it is important to make it clear how it is exactly that we will achieve the goal proposed: we need to make compromises, establish the project scope, its reach, the resources available to accomplish the goal, deadlines, cost and deliverables. It is compulsory to make a breakdown of the project into smaller subtasks/subprojects, and these assigned

Figure 2 The trade off.

the resources needed and to the people that will be responsible for them.
This can only be accomplished successfully through in-depth communication among all the parties involved, and that in turn means involvement of everyone, the client, project managers, developers, etc.

The execution phase

Once the project has been refined, planned and approved, it is time to get the work done. The people involved needs to be assembled, usually building a team and resolving all group dynamics problems, getting it to work until production level is reached, maintaining productivity until the goal is achieved and taking corrective action when something may get out of control.

The final phase

Once the product is finished, we are not done yet: it must be delivered to the client, and we need to make sure everything works as expected. The product must be deployed, installed, and possibly the client trained into using it. As the saying goes, the show isn't over until the fat lady signs, and it is always a pity to lose a lot of work because the final product hasn't bee finally delivered.

Project Post-mortem

Excellent! Your project has finished and you are now able to pursue other, more exciting endeavours. But this is not the end of the story. Rather, it links to the beginning of the next. Whether your project was an astonishing success or an utter failure, there is a lot to learn from it, what you did well, what you didn't, what you may expect of yourself and the people involved... If you did your homework while the project was alive, there will be a wealth of information on how everything went, and you can milk this information to help you better plan and manage your next projects. Ignoring this information will lead you to repeat all over again the same mistakes and, don't be naïve: if your project was a success you may not be so lucky next time.
Collecting project statistics and experiences once it is finally dead is the single and only way to learn from the past for the future. You may delay it to the next proposal, but we contend that at least some degree of self-examination should take place right after the project has end.

The Project Manager

The role of the project manager is to be responsible for the successful completion of the project. This entangles a number of tasks during all the project life cycle.
More specifically, the project manager will have to coordinate all the labours in the process, build and cater for the team that will carry out the tasks, ensure that all resources needed are available in a timely fashion, provide a hub for communication among all parties involved (the team, upper management, the client), and monitor and control the project well-being, taking corrective action when anything deviates from the project plan.
An important aspect in Project Management is the attitude you take to it. While the approach you use may change in the different phases, there is something that is of the utmost importance for success: whether you act proactively or not.
The very inception of project management is proactive. That's why planning is the start of all the process. You should therefore strive to preview and plan in advance for all possible contingencies, trying to have a contingency plan to save the day if need comes.
Waiting and hoping for the best is easier. Sometimes you will be lucky, and if no problem occurs, you save a lot of headaches. And, if you and your team are skilled enough you may get out of difficulties fast enough to reduce impact on the project schedule, quality or costs. But whenever something happens, the only way out will be through additional, unexpected stress on the team, and this will quickly lead to burnout and a snowball of additional problems.
So, for the Goodness sake, always try to be proactive and preview risks before they happen, working out alternative solutions ahead of time.

The project team

It is a resource like any other... or is it not? Most other resources are reliable and well

known, but the team deals with people and its intrinsic unpredictability.
Most of the project life cycle is usually spent in the execution phase, and carried out by the project team. It is therefore paramount that it works as a well-greased engine. Understanding how teams coalesce and build up, how people interact within them, and providing for successful interactions is at the core of project success.
You can never consider team members as interchangeable pieces in your puzzle. You need to know each of them, their needs and expectations and their abilities. People need to be matched to the appropriate tasks, and interest and morals kept up all the way through.
While the project manager has the ultimate responsibility, team members will do the actual work, and without their backing, no project can survive. It is therefore important to get team members involved from the earliest phases, getting their experience and agreement on the plan and schedule, work subdivision and resource allocation. This actually implies making each team member a project manager for the tasks under his responsibility, and helping them succeed as managers of them.
In other words, don't just think of team members as workers, but as skilled people and fellow project managers under your wings. Care for them and shield them as much as you may from external interference: your success depends on easing their way.

The client

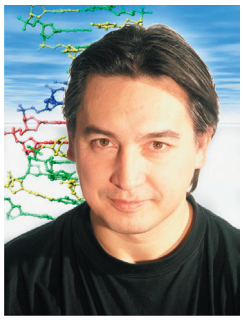
It is him, or her, or them, who has the need that will -hopefully- be satisfied by the successful completion of the project. It may be an actual client from outside the company, or another department, or upper management, or even yourself. Anyhow, the very existence of the project is to satisfy his/her needs.
An important part of the preparation of any project is the successful identification of all project stakeholders, that is, everybody with a direct or indirect interest in the project and influence in its development and outcome. Failure to do so results in ignoring key people and influences and usually leads to disaster due to unforeseen influences and conflicts of interest.
Just like the team, the client must be an active part in all the process. It seldom works to follow the old method of carefully laying out the needs and then enclosing yourself in an ivory tower to make the product, which you will magically produce to the client's astonishment and awe at the delivery session. If you are a client, demand to always have a say (and a *see*) on how the project develops: it is done in your interest, and it is in your interest to supervise its evolution.
The client original expectations may change during the lifetime of the project, often as a result of the project itself, and these shifting expectations must be met to guarantee client satisfaction and hence production of a successful deliverable.
Things may go wrong, out of schedule, problems appear. Every decision that may affect the final outcome, its costs, quality, or delivery date must be negotiated with the client. Every sensible change must be acknowledged and agreed. Otherwise you will end up releasing something other than what was initially agreed, and the client will have a lawful right to consider him/herself sacked.

Want to know more?

You may now take the "What is a project" quiz at the EMBnet/CNB portal and test your understanding of the topics covered.
`http://portal.cnb.uam.es/tiki/tiki-take_quiz.php?quizId=3`
To learn more about how projects get started, see the upcoming article on Project Conception or get yourself some good references like those included in the portal directory.

This article will be publicly available at the EMBnet/CNB portal, which we encourage you to visit for future updates, extensions, etc. If you disagree, have experiences to share or feel something is wanted or needed, please, visit the EMBnet/CNB portal and feel free to contribute and help build a significant help resource that may become a reference for the community.

MacOSX and Bioinformatics: a short report



Erik Bongcam-Rudloff
Assistant Professor
Anders Nister
Project Student
Swedish University of Agricultural Sciences
The Linnaeus Centre for Bioinformatics

New versions

We created new versions of the packages and we offer them for download at the usual place: `www.ebioinformatics.org`.

eBioinformatics
MacOSX Software

Two different packages are compiled for G4 or G5 processors, be sure to download the correct one.

New in the packages

All packages are now installed in `/usr/ebiotools/`, this new location makes it easier to upgrade and/or remove packages/programs. All licenses and documentations are now installed in `/usr/ebiotools/share`.

New as well:
Staden new version 1.4, a new **shlibs-package** with tcl/tk, itcl, iwidgets, f2c and tablelist, **Emboss 2.8** plus emboss-db (containing PRINTS, PROSITE, CUTG, AAINDEX, REBASE, and TRANSFAC), **Qt** updated from 3.2.0 to 3.2.3, and **EMBOSS.kaptn** updated to 0.95.
A program from TIGR for gene finding in microbial DNA, **glimmer**, has been added (see also: `http://www.tigr.org/software/glimmer/`)
Plus many bug fixes and changes to make the packages easier to upgrade or to update components.

Troubleshooting

Many users report having problems starting emboss programs, emboss-launcher.kaptn. or Staden programs. Please be sure to have the latest Apple X11 or X-Darwin version installed, most of the problems are related to old libs on previous X11 releases. Be also sure to have your `.cshrc` file with the line: `setenv DISPLAY :0.0` or if you are using bash then ensure that your `.bashrc` file has the line `DISPLAY=:0.0; export DISPLAY` as described in previous EMBnet.news issues.

Success story

We are also very proud to report having users in many different countries: the USA, Germany, Canada, New Zeeland, Australia, Sweden, and Denmark among others.

Example of comments from people around the world:
From USA
«*I really appreciate what you've done in putting these packages together AND writing a tutorial. I tried downloading from biotools and NCBI, and that was a waste of time.*»

From Canada
«*Thank you very much for your ebiotools release. It certainly makes installation much easier to have so many packages in one installation*»

We appreciate comments and reports on successful Installations. We are also open to comments for improvements and if you have suggestions on programs that should be included. Please mail to: `Erik.Bongcam@lcb.uu.se` or `Anders.Nister@lcb.uu.se`.

More details and a new rewritten tutorial in next EMBnet.news issue.

WHAT MOSQUITOES SNIFF

By Vivienne Baillie Gerritsen

In these parts of the world, it is hardly the time to talk of mosquitoes when the cold winter winds are still blowing. In other parts of the globe however, mosquitoes are out and about and causing millions of deaths yearly through their ability to transmit diseases, such as encephalitis, dengue, yellow fever and of course malaria. According to the World Health Organization, malaria alone is the cause of over two million deaths in Africa, one million of which are children under the age of five. A number of different mosquito species transmit malaria; these are the anopheline mosquitoes.

The female member of *Anopheles gambiae* is the mosquito which transmits malaria to humans. The question is why does it choose humans? The answer is it may well have to do with body odor. And where odor is involved, odorant receptors are also. It has recently been discovered that female *A.gambiae* do indeed have odorant receptors on their olfactive tissues which are specific for certain chemicals found in human sweat. And this is why they make a mosquito-line, so to speak, for certain humans. A greater understanding of the molecular processes, which underlie vector-host interaction, will help develop mosquito traps or repellents in the quest to beat malaria – or indeed other types of parasitic or viral diseases.

What causes malaria? It is neither a virus nor a bacteria but a plasmodium, a single-celled animal distantly related to an amoeba. *Plasmodium falciparum* and *Plasmodium vivax* are the two main culprits for causing malaria, the former being the more wicked of the two. When a mosquito sucks in human blood from a person suffering from malaria, it also sucks in the creatures causing the disease. The latter take advantage of this by reproducing inside the mosquito, waiting for the mosquito to dig its teeth into another host and seizing the opportunity to be released into the victim's blood thus causing a new infection. So what *A.gambiae* does – quite innocently in effect – is not only act as a taxi, which ships the disease from one host to another, but also as a cosy and quiet corner for the plasmodium to reproduce. And whilst *A.gambiae* acts as a taxi, human blood acts as a highway for the plasmodium.

Fig.1 *Anopheles gambiae* at work

One way then to eradicate malaria altogether would be to empty all humans of their blood, wouldn't it? However, that is most obviously not an option. Interestingly, research on ways to fight off the disease has come round in a circle in the past 50 years. By the second half of the 20th century, entomologists were aware that it was the female mosquito which stung. A far cry from what was believed in the times of Julius Caesar. In those days, it had been noted that people who lived close to swamps and marshes were more susceptible to be struck down by malaria. Bereft of the knowledge of biology we possess today, the people assumed that the actual stench was the perpetrator of the disease. And this is where the word 'malaria' originates from: 'mal' meaning 'bad' and 'aria', air… And while we are flirting with the science of etymology, the word 'mosquito' is derived from the Spanish 'little fly'.

So, by the 1950s, scientists all over the world were imagining all sorts of experiments to understand not only why it was the female mosquitoes that were attracted but also how they were attracted. One of the experiments involved a human steel robot which was invariably warmed to body temperature, imbibed in human sweat or even made to exhale CO2. Research was blossoming until

DDT made its appearance in the 1960s and was so effective as an insecticide that scientists' eagerness in understanding the vector-host interaction was somewhat dulled. However today, resistance to DDT coupled with second thoughts on the uncontrolled use of insecticides has flared an interest in alternative modes of fighting off mosquitoes.

The *Anopheles gambiae* genome has been completely sequenced. On it are found around 100 odorant receptor genes. These different odorant receptors are dispersed all over the mosquito's olfactive tissues. One of them, odorant receptor Or1 is found solely on female mosquito antennae and is particularly attracted to one of the 300 chemical compounds found in human sweat: 4-methylphenol. Or1 is a transmembrane G-protein coupled receptor, typically around 400 amino acids long. Its role, like all odorant receptors, is to bind its specific odorant. As a result, a number of downstream effector enzymes induce second messengers, which in turn stimulate odorant neurons. And the whiff of human sweat is transmitted to the brain thus giving the mosquito the drive to sting. What is more, not only does Or1 seek out 4-methylphenol for the mosquito's blood-feed but it also seems to have a role in turning off the process once the mosquito has had its fill. Indeed, once the insect is replete, the odorant receptor seems to be inhibited, perhaps following belly distension.

The existence of Or1 only in female mosquitoes – and hence its indirect role in its involvement in the transmission of malaria – as well as the discovery of its ligand should lead to interesting biotechnological applications in the quest for mosquito traps or indeed repellents. Substitute ligands could be synthesized which would trick the odorant receptors, and hence the mosquitoes, by disorienting them and leading them into a trap where they would be left to buzz aimlessly. Alternatively, ligands could be thought up which would inhibit the receptors and act as insect repellents. The discovery is promising. More research must be done to grasp in greater molecular detail the intricacies of vector-host interaction and then a cheap solution must be found to help fight off a disease which is one of the scourges of developing countries. Unfortunately, we cannot change how we smell – that is not the smell the mosquitoes are out for – but biotechnology could invent a way for mosquitoes to smell differently. And who would refuse a hot, humid summer night without unrelenting, whining mosquitoes?

Cross-references to Swiss-Prot

Q8WTE7: *Anopheles gambiae* (African malaria mosquito) odorant receptor Or1

References

1. Hallem E.A., Fox A.N., Zwiebel L.J., Carlson J.R.
Mosquito receptor for human-sweat odorant
Nature 427:212-213(2004)
PMID: 14724626

2. Kanzok S.M., Zheng L.
The mosquito genome – a turning point?
Trends Parasitol. 19:329-331(2003)
PMID: 12901929

3. Enserink M.
What mosquitoes want: secrets of host attraction
Science 298:90-92(2002)
PMID: 12364778

Protein Spotlight (ISSN 1424-4721) is published by the Swiss-Prot group at the Swiss Institute of Bioinformatics (SIB).

National Nodes

Argentina
Oscar Grau
IBBM, Facultad de Cs. Exactas, Universidad Nacional de La Plata
Email: grau@biol.unlp.edu.ar
Tel: +54-221-4259223 Fax: +54-221-4259223
http://www.ar.embnet.org

Australia
Sonia Cattley
RMC Gunn Building B19, University of Sydney,NSW, 2006
Email: scattley@angis.org.au
Tel: +61-2-9531 2948
http://www.au.embnet.org

Austria
Martin Grabner
Vienna Bio Center, University of Vienna
Email: martin.grabner@univie.ac.at
Tel: +43-1-4277/14141
http://www.at.embnet.org

Belgium
Robert Herzog, Marc Colet
BEN ULB Campus Plaine CP 257
Email: rherzog@ulb.ac.be mcolet@ulb.ac.be
Tel: +32 2 6505146 Fax: +32 2 6505124
http://www.be.embnet.org

Brasil
Gonçalo Guimaraes Pereira
Laboratório de Genômica e Expressão - IB
UNICAMP-CP 6109
13083-970 Campinas-SP, BRASIL
Tel: 0055-19-37886237/6238
Fax: 0055-19-37886235
Email: goncalo@unicamp.br
http://www.br.embnet.org

Canada
Canadian Bioinformatics Resource, National Research Council Canada, Institute for Marine Biosciences,
Email: manager@cbr.nrc.ca
Tel: +1-902-426 7310 Fax: +1-902-426 9413
http://www.ca.embnet.org

Chile
Dr. Ricardo Baeza-Yates
Dept. of Computer Science, Santiago,
Email: rbaeza@dcc.uchile.cl
Tel: N/A
http://www.embnet.cl

China
Jingchu Luo
Room 303, Exchange Centre, Peking University
Email: luojc@cbi.pku.edu.cn
Tel: +86-10-6275 9001
http://www.cbi.pku.edu.cn

Colombia
Emiliano Barreto Hernández
Instituto de Biotecnología
Universidad Nacional de Colombia
Edificio Manuel Ancizar
Bogota - Colombia
Tel: +571 3165027 Fax: +571 3165415
Email : ebarreto@ibun.unal.edu.co
http://bioinf.ibun.unal.edu.co

Cuba
Ricardo Bringas
Centro de Ingeniería Genética y Biotecnolgía,
La Habana, Cuba
Email: bringas@cigb.edu.cu
Tel: +53 7 218200
http://www.cu.embnet.org

Denmark
Hans Ullitz-Moeller
BioBase, University of Aarhus
Email: hum@biobase.dk
Tel: +45-86-13 9788
http://www.dk.embnet.org

Finland
Eija Korpelainen
CSC, Espoo
Email: eija.korpelainen@csc.fi
Tel: +358 9 457 2030
http://www.fi.embnet.org

France
Jean-Marc Plaza
INFOBIOGEN, Evry
Email: plaza@infobiogen.fr
Tel: +33 1 60 87 37 11 Fax: +33 1 60 87 37 96
http://www.fr.embnet.org

Germany
Sandor Suhai
EMBnet node at the German Cancer Research Center
Department of Molecular Biophysics (H0200)
Email: genome@dkfz.de
Tel: +49-6221-422 342 Fax: +49-6221-422 333
http://www.de.embnet.org

Greece
Babis Savakis
Institute of Molecular Biology and Biotechnology
Heraklion, Crete
Email: savakis@nefeli.imbb.forth.gr
Tel: +30-81-391 114 Fax: +30-81-391 104
http://www.imbb.forth.gr

Hungary
Endre Barta
Agricultural Biotechnology Center
Szent-Gyorgyi A. ut 4. Godollo,
Email: barta@abc.hu
Tel: +36 30-2101795
http://www.hu.embnet.org

India
H.A.Nagarajaram
Laboratory of Computational Biology & Bioinformatics facility, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad
Email: han@www.cdfd.org.in
Tel: +91 40 7155607 / 7151344 ext:1206
Fax : +9140 7155479
http://www.in.embnet.org

Israel
Leon Esterman
INN (Israeli National Node) Weizmann Institute of Science
Department of Biological Services, Biological Computing Unit, Rehovot
Email: Leon.Esterman@weizmann.ac.il
Tel: +972- 8-934 3456
http://www.il.embnet.org

Italy
Cecilia Saccone
CNR - Institute of Biomedical Technologies
Bioinformatics and Genomic Group
Via Amendola 168/5 - 70126 Bari (Italy)
Email: saccone@area.ba.cnr.it
Tel. +39-80-5482100 - Fax. +39-80-5482607
http://www.it.embnet.org

Mexico
Cesar Bonavides
Nodo Nacional EMBnet, Centro de Investigación sobre Fijación de Nitrógeno, Cuernavaca, Morelos
Email: embnetmx@cifn.unam.mx
Tel: +52 (7) 3 132063
http://embnet.cifn.unam.mx

The Netherlands
Jack A.M. Leunissen
Dept. of Genome Informatics
Wageningen UR, Dreijenlaan 3
6703 HA Wageningen, NL
Email: Jack.Leunissen@wur.nl
Tel: +31 317 484074
http://www.nl.embnet.org

Norway
George Magklaras
The Norwegian EMBnet Node
The Biotechnology Centre of Oslo
Email: admin@embnet.uio.no
Tel: +47 22 84 0535
http://www.no.embnet.org

Poland
Piotr Zielenkiwicz
Institute of Biochemistry and Biophysics
Polish Academy of Sciences Warszawa
Email: piotr@pl.embnet.org
Tel: +48-22 86584703
http://www.pl.embnet.org

Portugal
Pedro Fernandes
Instituto Gulbenkian de Ciencia
Unidade de Bioinformatica
2781-901 OEIRAS
Email: pfern@igc.gulbenkian.pt
Tel: +351 214407912 Fax: +351 2144079070
http://www.pt.embnet.org

Russia
Sergei Spirin
Biocomputing Group, Belozersky Institute Moscow
Email: sas@belozersky.msu.ru
Tel: +7-095-9395414
http://www.genebee.msu.ru

Slovakia
Lubos Klucar
Institute of Molecular Biology SAS Bratislava
Email: klucar@embnet.sk
Tel: +421 7 5941 2284
http://www.sk.embnet.org

South Africa
Ruediger Braeuning
SANBI, University of the Western Cape, Bellville
Email: ruediger@sanbi.ac.za
Tel: +27 (0)21 9593645
http://www.za.embnet.org

Spain
José M. Carazo, José R. Valverde
EMBnet/CNB, Centro Nacional de Biotecnología, Madrid
Email: carazo@es.embnet.org, jrvalverde@es.embnet.org
Tel: +34 915 854 505 Fax: +34 915 854 506
http://www.es.embnet.org

Sweden
Nils-Einar Eriksson, Erik Bongcam-Rudloff
Uppsala Biomedical Centre, Computing Department, Uppsala, Sweden
Email: nils-einar.eriksson@bmc.uu.se
erik.bongcam@bmc.uu.se
Tel: +46-(0)18-4714017, +46-(0)18-4714525
http://www.embnet.se

Switzerland
Laurent Falquet
Swiss Institute of Bioinformatics, CH-1066 Epalinges
Email: Laurent.Falquet@isb-sib.ch
Tel: +41 (21) 692 5954 Fax: +41 (21) 692 5945
http://www.ch.embnet.org

United Kingdom
Alan Bleasby
UK MRC HGMP Resource Centre, Hinxton, Cambridge
Email: ableasby@embnet.org
Tel: +44 (0) 1223 494535
http://www.uk.embnet.org

Specialist Nodes

EBI
Rodrigo López
EBI Embl Outstation, Wellcome trust Genome Campus, Hinxton Hall, Hinxton, Cambridge, United Kingdom
Email: rls@ebi.ac.uk
Phone: +44 (0)1223 494423
http://www.ebi.ac.uk

ETI
P.O. Box 94766
NL-1090 GT Amsterdam, The Netherlands
Email: wouter@eti.uva.nl
Phone: +31-20-5257239
Fax: +31-20-5257238
http://www.eti.uva.nl

EU
Frederick Marcus
DG Research - European Commission
Brussels BELGIUM
Email: Frederick.Marcus@cec.eu.int

ICGEB
Sándor Pongor
International Centre for Genetic Engineering and Biotechnology
AREA Science Park, Trieste, ITALY
Email: pongor@icgeb.trieste.it
Phone: +39 040 3757300
http://www.icgeb.trieste.it

LION Bioscience
Thure Etzold
LION Bioscience AG, Heidelberg, Germany
Email: Thure.Etzold@uk.lionbioscience.com
Phone: +44 1223 224700
http://www.lionbioscience.com

MIPS
H. Werner Mewes
Email: mewes@mips.embnet.org
Phone: +49-89-8578 2656
Fax: +49-89-8578 2655
http://www.mips.biochem.mpg.de

UMBER
Terri Attwood
School of Biological Sciences, The University of Manchester, Oxford Road, Manchester M13 9PT, UK
Email: attwood@bioinf.man.ac.uk
Phone: +44 (0)61 275 5766
Fax: +44 (0) 61 275 5082
http://www.bioinf.man.ac.uk/dbbrowser

TECH-MGR
Email: tech-mgr@embnet.org
The team gives support to EMBnet nodes and helps them with maintenance and troubleshooting.
The team is formed of experienced system administrators and programmers who ensure the availability of local services for all EMBnet users.

EMBnet.news

ISSN 1023-4144

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 then please let us know. Please send your contributions to one of the editors. You may also submit material by e-mail.

Past issues of embnet.news are available as PostScript or PDF files (ISSN 1023-4144). You can get them by anonymous ftp from:
the EMBnet organisation Web site
`http://www.embnet.org/download/embnetnews`
the Belgian EMBnet node
`ftp://ftp.be.embnet.org/pub/embnet.news`
the UK EMBnet node
`ftp://ftp.uk.embnet.org/pub/embnet.news`
the EBI EMBnet node
`ftp://ftp.ebi.ac.uk/pub/embnet.news`

Submission deadline for next issues:

May 31, 2004
August 15, 2004
October 31, 2004
February 29, 2005