

Report

Bioinformatics Service, Education and Research: The EMBnet and CBI

Jingchu Luo

Centre of Bioinformatics, Peking University, Beijing 100871, P.R. China

Bioinformatics has become indispensable to all fields in the life sciences. Its importance and impact to both theoretical and applied biology cannot be overestimated. The rapid progress of genome projects has brought a vast accumulation of molecular biological information in the past decade. Millions of nucleic acid sequences with billions of bases have been deposited in EMBL, GenBank and DDBJ. Both SwissProt and PIR form the basis of annotated protein sequence database together with TrEMBL and GenPept which are computer translated sequence entries from EMBL and GenBank. Hundreds of specialist databases have been derived from the above primary sequence databases, such as TRANSFAC (gene regulation sites and transcription factors), Vector (DNA cloning vectors), EPD (eukaryotic promoters), PROSITE (protein motif sites), BLOCKS (protein alignments), PRINTS (protein fingerprints), REBASE (restriction enzymes.) Special web servers for model systems such as GoldenPath and ENSEMBL have been actively developed.

THE EUROPEAN MOLECULAR BIOLOGY NETWORK

The rapid growth of computer network makes it possible for molecular biologists to access the public database via the Internet. The European Molecular Biology Network (EMBnet) was formed in 1988 by European scientists and is an international organisation for bioinformatics service, research and education. It is organised as a foundation called "The EMBnet Stichting", registered in the Netherlands. During its 13 years of existence, EMBnet has become a world-wide organisation bringing bioinformatics professionals together to serve the expanding fields of genetics and molecular biology. It has expanded its membership, not only from European countries but also, since 1996, from other continental countries.

Two categories have been defined for EMBnet membership, the national node and the specialist node. The national nodes (Table 1) are mandated by their governmental authorities to provide biocomputing and bioinformatics services whereas the specialist nodes (Table 2) have extensive expertise and knowledge in specific areas of biocomputing, bioinformatics and biodiversity. EMBnet now has 40 nodes spread over 30 countries and supports more than 32,000 registered as well as an enormous number of web based users from all over the world. EMBnet member sites have a sum of approximately 200

Electronic publication can be found in *In Silico Biol.* **2**, <<http://www.bioinfo.de/isb/2002/02/0016/>>.

1386-6338/02/\$8.00 © 2002 – IOS Press and Bioinformation Systems e.V. All rights reserved

professionals. Computing resources of EMBnet nodes add up to around 300 hosts ranging from workstations to supercomputers.

Table 1
List of the National Nodes of the European Molecular Biology Network

Country	Node Name	URL	Institute
Argentina	IBBM	sol.biol.unlp.edu.ar	Instituto de Bioquímica y Biología Molecular
Austria	VBC	www.at.embnet.org	Vienna University Computer Center
Australia	ANGIS	www.au.embnet.org	Australia National Genomic Information System
Canada	CBR-RBC	www.cbr.nrc.ca	National Research Council of Canada, Institute for Marine Biosciences
Belgium	BEN	www.be.embnet.org	Belgian EMBnet Node
China	CBI	www.cn.embnet.org	Centre of Bioinformatics, Peking University
Cuba	CIGB	www.cu.embnet.org	Centro de Ingeniería Genética y Biotecnología
Chile	DCC	www.dcc.uchile.cl	University of Chile
Denmark	BioBase	www.dk.embnet.org	Danish Biotechnological Database
France	InfoBioGen	www.fr.embnet.org	Resource Centre Infobiogen
Finland	CSC	www.fi.embnet.org	National Center for Scientific Computing
Germany	GeniusNet	www.de.embnet.org	German Cancer Research Centre
Greece	IMBB	www.imbb.forth.gr	Institute of Molecular Biology and Biotechnology
Hungary	HEN	www.hu.embnet.org	Hungarian EMBnet Node
India	CDFD	www.in.embnet.org	Centre for DNA Fingerprinting and Diagnostics
Ireland	INCBI	www.ie.embnet.org	Irish National Centre for Bioinformatics
Israel	INN	www.il.embnet.org	Weizmann Institute of Science
Italy	CNR	www.it.embnet.org	Area di Ricerca - Bari
Mexico	CIFN	embnet.cifn.unam.mx	Nitrogen Fixation Research Center
Norway	BiO	www.no.embnet.org	Biotechnology Centre of Oslo
Poland	IBB	www.pl.embnet.org	Institute of biochemistry and biophysics
Portugal	PEN	www.pt.embnet.org	Portuguese EMBnet Node
Russia	GeneBee	www.ru.embnet.org	Belozersky Institute of Physico-Chemical Biology, Moscow State University
Slovakia	IMB-SAS	www.sk.embnet.org	Institute of Molecular Biology, Slovak Academy of Science
South Africa	SANBI	www.za.embnet.org	South African National Bioinformatics Institute, University of Western Cape
Spain	CNB	www.es.embnet.org	Centro Nacional de Biotecnología, Campus Universidad Autónoma
Sweden	LCB	www.se.embnet.org	Linnaeus Centre for Bioinformatics, Uppsala Biomedical Centre
Switzerland	SIB	www.ch.embnet.org	Swiss Institute of Bioinformatics
The Netherlands	CMBI	www.nl.embnet.org	Centre for Molecular and Biomolecular Informatics, University of Nijmegen
UK	HGMP	www.uk.embnet.org	UK MRC Human Genome Mapping Project Resource Centre

Table 2
List of the Specialist Nodes of the European Molecular Biology Network

Name	URL	Institute / Company
EBI	www.ebi.ac.uk	European Bioinformatics Institute
ETI	www.eti.uva.nl	Expert Centre for Taxonomic Identification
ICGEB	www.icgeb.trieste.it	International Centre for Genetic Engineering & Biotechnology
UMBER	www.bioinf.man.ac.uk	University of Manchester
MIPS	www.mips.biochem.mpg.de	Max-Planck-Institute fuer Biochemie
Pharmacia	www.pnu.com	Pharmacia & Upjohn
Roche	www.roche.com	F. Hoffmann-La Roche
Sanger Centre	www.sanger.ac.uk	The Sanger Centre

The Annual General Meeting (AGM), where each node has an equal vote, is the main decision making body of EMBnet. EMBnet is managed by a four-member Executive Board (EB) that is elected by the AGM. The EB effects the decisions of the AGM, deals with funding and generally takes care of the organisation between AGMs. The activities of EMBnet are run by three Project Committees: the Education and Training Committee, the Technical Management Committee and the Publicity and Public Relations Committee. Each project committee consists of four members elected at the Annual General Meeting. Collaborations within EMBnet as a whole and between individual nodes of EMBnet are ongoing. Database construction, software development, publication of journals, newsletters and quick guides, training courses, workshops and technical visits are a part of the EMBnet activities.

Although composed predominantly of academic nodes, EMBnet gains an important added dimension from its industrial members. The success of EMBnet is attracting increasing numbers of organisations to join. Membership application to EMBnet should be presented at and voted by the EMBnet AGM around September each year at one of the European nodes. The application should be submitted to the Secretary of the Executive Board three months in advance and circulated among all of the node managers by the EMBnet board mailing list. Application for national node membership of the EMBnet should be accompanied by a document from the respective national authority and indicate the capacity of the site and the expertise of the applicants.

EMBnet is also looking into possibilities of working collaboratively with other bioinformatics and biocomputing organisations such as the Asia Pacific Bioinformatics Network (APBioNet) which was initiated by a group of bioinformaticians in 1998. Gaining experience from the European colleagues, APBioNet is dedicated to the advancement of the field of bioinformatics among Asia Pacific economies, focusing on development of the bioinformatics infrastructure, exchange of data and information, the development of training programs, workshops and symposia and the stimulation of collaborations in the field of bioinformatics. Nearly twenty representative organisations and institutions from more than ten economies are currently part of this growing network.

THE CENTRE OF BIOINFORMATICS AT PEKING UNIVERSITY

The economic boom in China in the past decade has made it possible to build up a national network infrastructure. In 1994, China Education and Research Network (CERNET), promoted by the State Education Commission of China, announced its birth in Zhong Guan Cun, the scientific golden-triangle area in the North-Western area of Beijing. This is a region where dozens of research institutes and two of the largest universities in China, Tsinghua University and Peking University, are located. As one of the central nodes of CERNET, Peking University Network Centre plays an important role in network administration and service in China.

China's economic growth also offers funding opportunities for Chinese bioscientists. In addition to various projects from basic research to agricultural and medical applications in biological sciences and biotechnology, the Rice Genome Project supported by the National High-tech program has made profound progress. The China Human Genome Project, consisting of some twenty laboratories, was initiated by the National Natural Science Foundation in 1993 and is now becoming one of the biggest national projects. The North Centre and South Centre of Human Genome Research located in Beijing and Shanghai respectively started to operate in 1998. Joining in the International Human Genome Program in 1999, the Beijing Genome Institute, together with other Chinese collaborators, completed sequencing of 1% of the full human genome. BGI is working on full genome sequencing for other organisms such as rice and pig. In the biocomputing and bioinformatics field, scientists from various institutes and universities have been working on molecular simulation, structure prediction, drug design, genome analysis, molecular evolution and DNA/protein sequence analysis. The first Chinese Conference on Bioinformatics was held in April 2001 with four hundreds participants. This will be an annual event and the meeting for this year has been scheduled on 28-30 June 2002. Supported by the Sino-German Science Centre, a delegation of four scientists attended the German Conference on Bioinformatics in October 2001.

With the support from the Ministry of Education and the Ministry of Science and Technology, the National Laboratory of Protein Engineering and Plant Genetic Engineering at Peking University joined the European Molecular Biology Network in 1996. The Centre of Bioinformatics (CBI) at Peking University was then initiated in March 1997. This Centre is composed of scientists not only from the biological field, but also from the Computing Centre of Peking University. With the help of EMBnet colleagues, the EMBnet China node has gained some significant progress since its birth. CBI provides bioinformatics resource for domestic users, including databases and software. More than 90 databases, such as DNA and protein sequence and structure databases, genome and sequence-related databases, mapping, mutation, reference databases were installed in the database query system SRS. A BLAST database search engine is running locally. Various mirrors such as the Human Genome Database (GDB), biosafety regulation (BINAS), transcriptional factor database (TRANSFAC), Expert of Protein Analysis System (ExPAsy), Bacteria Genome Information (GenoList), have been set up since 1996. Other bioinformatics information, such as genome web resource, EMBnet newsletter, lists of bioinformatics books, references and catalogues, workshop abstracts and presentations were also put online. Both freeware and commercial software packages for sequence analysis and molecular design such as EMBOSS, GCG, Staden, Insight II are accessible for local users while all public databases and programs can be browsed, searched and downloaded via CBI web and ftp servers for both domestic and external users.

As an educational and academic incubator, various seminars, workshops and courses on bioinformatics and biocomputing are being organised at Peking University. With the support from domestic authorities such as the Ministry of Education, the Ministry of Science and Technology, the Natural Science Foundation and the Beijing Municipal government, as well as external sources including the EMBNet,

ICGEB, EMBL, the Swiss National Science Foundation (SNSF), CBI succeeded in organising bioinformatics courses and workshops in the past since 1998. Following the EMBnet course for which five European experts came to give lectures and hand-on practice and two workshops on molecular modelling and drug design in 1998, an ICGEB sponsored international bioinformatics course was held in April 1999 and a Sino-Swiss bioinformatics workshop was held last in 2000. A graduate course "Introduction to Bioinformatics" has been started since 2000. More than 100 students with biology and medical background obtained practical training to access the rich resource of databases and software tools. CBI scientists have also been invited to give lectures and tutorials at various domestic and international conferences and workshops as well as national wide summer schools.

In addition to service and training, CBI also carries out bioinformatics research and development. Databases of protein loop classification and protein domain assignment have been developed in collaboration with UK scientists. Cytomer, a relational database of gene regulation and expression at the tissue and cell specific level was constructed collaboratively with German bioinformaticians. Specialist database information systems in protein disulphide bridge, rice dwarf virus, as well as China's ethnic genome diversity are under construction. Sequence analysis tools and web interfaces to BLAST, FASTA and other programs were developed. Together with colleagues of mathematicians and computer scientists of Peking University and colleagues from other institutions, algorithm development, bacteria genome sequence assembly, genome sequence analysis, structure prediction, molecular design are actively going on.

It is now well recognised by both the domestic and international bioinformatics communities that EMBnet China node has been playing an important role in providing bioinformatics resource, service and education. Together with partners from the Chinese Academic Sciences in Shanghai and Beijing, CBI will start the National Bioinformatics Base under the China High-tech Programme.

Nevertheless, CBI is still in its infancy. We are keen in learning experiences from, seeking assistance of and exchanging ideas with bioinformatics and biocomputing experts from all over the world.