Bioinformatics Research in Australia

Yi-Ping Phoebe Chen
Center for Information Technology Innovation, Queensland University of Technology, Brisbane, Australia.
p.chen@qut.edu.au

Geoff McLachlan
Department of Mathematics, University of Queensland, Australia.
gjm@maths.uq.edu.au

Abstract

Bioinformatics is the intersection of computer science, statistics, molecular biology and genetics. It is one of the most important emerging research areas of the 21st century and has already attracted worldwide interest. It is clear that major initiatives are being undertaken which will establish Australia both as a vital link in the international bioinformatics community for research and development and also as an Asia-Pacific service for bioinformatics. This article briefly notes some groups carrying out bioinformatics research in Australia.

Keywords: Bioinformatics; Australia.

1 Introduction

An excellent account of bioinformatics in Australia has been provided in the editorial of Littlejohn (2000). Since the writing of this editorial, interest in bioinformatics in Australia has increased enormously. The volume of biological data is growing exponentially, partly due to the deluge of information that has stemmed from the working draft of the human genome (released in February 2001) and other similar sequencing projects. Also, the increase in advanced computer power has allowed groups to generate more data at a faster rate. Consequently, there are many challenging and novel problems on the interface of modern biology and computational statistics that are in need of solutions. For example, there needs to be statistical methodology and database technologies that can analyze and draw reliable inferences from these huge and complex data sets, where the biological signal is often very difficult to distinguish from the noise which can arise from many sources. There is also the need to adapt existing statistical and computational approaches to accommodate the often novel constraints or assumptions of bioinformatics problems.

The bioinformatics research community in Australia has been fortunate in having one of the leading researchers on the international stage in the field, Prof. Terry Speed (University of California Berkeley), now spending six months each year at the Walter & Eliza Hall Institute (WEHI) in Melbourne. Research in bioinformatics at WEHI is focussed on statistical genetics, design and analysis of gene expression studies, biological sequence analysis, and proteomics.

At the University of Queensland (UQ), research in bioinformatics is centered in the Computational Biology and Bioinformatics Division of the Institute of Molecular Bioscience (IMB). The IMB is directed by Prof. John Mattick whose research interests in this field include the development of an intron and alternative splicing database. The Computational Biology and Bioinformatics Division at the IMB is directed by Prof. Mark Ragan. It is actively building critical mass in data mining, genomic and expression databases, protein structure/function, gene identification, molecular modeling, comparative genomics, chemical informatics and high-performance computing. It collaborates closely with other initiatives at UQ: the Advanced Computational Modeling Center, under the direction of Prof. Kevin Burrage; the Computational Biology and Bioinformatics group funded by the Commonwealth Department of Education, Training and Youth Affairs (DETYA); the Queensland Parallel Supercomputing Foundation (QPSF) in collaboration with the Australian Partnership in Advanced Computing (APAC); and the Center for Statistics directed by Prof. Geoff McLachlan. There are also close collaborative ties between the
Division and Dr. Phoebe Chen’s group at the Queensland University of Technology (QUT).

Research in bioinformatics at QUT is focused in the Center for Information Technology Innovation led by Dr. Chen. As part of her research, Dr. Chen is undertaking ARC funded investigations with UQ collaborators on two projects: “Integrating Database Technologies and Visual Analysis in Bioinformatics for Genome Data” and “Computational Infrastructure for High-Throughput Genome Bioinformatics.” QUT has identified bioinformatics as a strategic area for research and education. Its Faculty of Information Technology has created a research concentration in bioinformatics. Research projects have been collaborated among the Faculty of Information Technology, the Faculty of Science, Department of Maths and the Center for Molecular Biotechnology.

Among other groups researching bioinformatics in Brisbane, there is the UQ node of the Australian Center for Plant Functional Genomics, headed by Prof. Kaye Basford, the Queensland Institute of Medical Research (QIMR), and the CSIRO Divisions of Livestock Industries (Dr. Brian Dalrymple) and Plant Industry (Drs. Rosanne Casu and Scott Chapman). Bioinformatics research is being carried out also in some other CSIRO divisions either in Brisbane or elsewhere in Australia, including Entomology and Mathematical & Information Sciences.

QUT has identified bioinformatics as a strategic area for research and education. Its Faculty of Information Technology has created a research concentration in bioinformatics.

The Center for Bioinformation Science at the Australian National University (ANU), co-directed by Prof. Sue Wilson and Prof. Simon Easteel, collaborates extensively with groups within the ANU and outside. The former include the Human Genetics group at the John Curtin School of Medical Research and the Advanced Data Mining Group (Dr. Marcus Hegland). The latter include the Garvan Institute of Medical Research, Prof. Speed’s group at the University of California Berkeley, and the Ludwig Institute for Cancer Research.

Concerning bioinformatics in Tasmania, the state government has recognized the growing commercial and academic research potential of this field with financial support for the setting up of the Tasmanian Bioinformatics Center of Excellence (with Dr. Michelle Allan as its CEO).

Australia’s national bioinformatics service provider (ANGIS, the Australian National Genomics Information Service) is based at the University of Sydney. It is headed by Dr. Michael Poidinger. Over the years, ANGIS has developed a series of integrated bioinformatics workflows. Recently, the University of Sydney has announced a new postgraduate course in bioinformatics. In other recent initiatives in Sydney, the Garvan Institute of Medical Research has established a bioinformatics center.

The bioinformatics research community in Australia has been fortunate in having one of the leading researchers on the international stage in the field, Prof. Terry Speed (University of California Berkeley), now spending six months each year at the Walter & Eliza Hall Institute (WEHI) in Melbourne.

The Victorian Bioinformatics Consortium (VBC) is directed by Prof. Ross Copple. VBC members include: Monash University, the central institution with several departments and centers involved; Plant Biotechnology Center of Agriculture Victoria at La Trobe University; CSIRO Mathematical and Information Sciences; the Peter MacCallum Cancer Institute (directed by Prof. David Bowtell, a pioneer of array technology); the Howard Florey Institute; and other commercial companies. Prof. Copple’s research interests include the molecular biology of malaria antigens and malaria vaccine development. Bioinformatics research in Melbourne is being carried out also at the Victorian Partnership for Advanced Computing (VPAC) and the Ludwig Institute for Cancer Research.

The University of Adelaide hosts the Microarray Analysis Group (MAG) directed by Dr. Patty Solomon. Its primary aims are to provide advice on the statistical design and analysis of microarray experiments specific to the local context; to provide access to state-of-the-art image processing and data analysis software; and to collaborate with national and international groups working on microarray informatics.

Concerning bioinformatics in Tasmania, the state government has recognized the growing commercial and academic research potential of this field with financial support for the setting up of the Tasmanian Bioinformatics Center of Excellence (with Dr. Michelle Allan as its CEO).
Bioinformatics research in Australia is also taking place in many of the Cooperative Research Centers (CRCs), particularly those that come under the umbrella of the Biotechnology Education and training program. For example, there is the Cooperative Research Center for the Discovery of Genes for Common Human Diseases, which actively pursues bioinformatics research.

The Center for Bioinformation Science at the Australian National University (ANU), co-directed by Prof. Sue Wilson and Prof. Simon Easteal, collaborates extensively with groups within the ANU and outside.

Mention should also be made of (i) The Australian Genome Research Facility (AGRF) directed by Prof. Mattick at UQ (with a Melbourne branch located at WEHI), which provides an engine for large scale genetic discovery by Australian research organizations and industry; (ii) BioLateral Pty. Ltd. which has provided training programs across the spectrum of bioinformatic problems (molecular modeling, microarray analysis, data management, programming); and (iii) The Australian Biotechnology Association (AusBioTech), which supports regular bioinformatics related events at the various branches throughout Australia, including the annual AusBioTech conference, which brings together scientists from industry and academia.

This year two international conferences on bioinformatics are to be held. Firstly, there is the First Asia-Pacific Bioinformatics Conference, to be held at Adelaide within Australia Computer Science Week (February 4 – 7, 2003). The Asia-Pacific Bioinformatics Conference series, of which Dr. Phoebe Chen is a founder, is to be an annual forum for exploring research, development and novel applications of bioinformatics. The second conference is the 11th International Conference on Intelligent Systems for Molecular Biology (ISMB 2003) to be held in Brisbane (June 29 – July 3). This conference is sponsored by the International Society for Computational Biology and is one of the major conferences in bioinformatics.

References


Acknowledgments

The authors would like to thank Ms. Jess Mar (European Bioinformatics Institute (formerly of UQ)) for her assistance in the preparation of this article.

Some of Australia Bioinformaticians

Tim Littlejohn
Michael Poidinger
Mark Ragan
Yi-Ping Phoebe Chen
Kevin Burrage
Ross Coppel
Georg Weiller
Goff McLanchian
Contents: Exploring RNA Intermediate Conformations with the Massively Parallel Genetic Algorithm; Introduction to Self-Assembling DNA Nanostructures for Computation and Nanofabrication; Mapping Sequence to Rice FPC; Graph Theoretic Sequence Clustering Algorithms and their Applications to Genome Comparison; The Protein Information Resource for Functional Genomics and Proteomics; High-Grade Ore for Data Mining in 3D Structures; Protein Classification: A Geometric Hashing Approach; Interrelated

COMPUTATIONAL BIOLOGY AND GENOME INFORMATICS

edited by Jason T L Wang (New Jersey Institute of Technology, USA),

This book contains articles written by experts on a wide range of topics that are associated with the analysis and management of biological information at the molecular level. It contains chapters on RNA and protein structure analysis, DNA computing, sequence mapping, genome comparison, gene expression data mining, metabolic network modeling, and phyloinformatics.

The important work of some representative researchers in bioinformatics is brought together for the first time in one volume. The topic is treated in depth and is related to, where applicable, other emerging technologies such as data mining and visualization. The goal of the book is to introduce readers to the principle techniques of bioinformatics in the hope that they will build on them to make new discoveries of their own.

Readership: Molecular biologists who rely on computers and mathematical scientists with interests in biology.

266pp • Feb 2003 • 981-238-257-7 • US$38 • £26
The Status of Formalized Bioinformatics

Abstract

This review describes the current state of bioinformatics education in Australia, outlining the award and non-award courses offered by Australian Universities and centers.

Keywords: Bioinformatic degree; Australia.

1 Introduction

Bioinformatics is a new, under developed and under resourced field in Australia. Whilst there are a number of centers with a focus on bioinformatics research, including the Victorian Bioinformatics Consortium at the University of Melbourne and Monash University, the Institute for Molecular Biosciences (IMB) at the University of Queensland, The Bioinformatic Center of Excellence in Tasmania, the Center for BioInformation Science (CBIS) at the Australian National University and the Sydney University Biological Informatics and Technology Center (SUBIT), the amount of research and number of qualified bioinformaticians in Australia remains small, compared with other countries.

However, in the past five years there has been an increase in the number of bioinformatics degrees appearing in Australia. Compared to the US markets, there is a serious lack of postgraduate degrees in Australia with only two universities (University of Sydney and Curtin University) providing postgraduate degrees. As of yet there have been no graduations from these as of December 2002.

Note that in the past ten years most universities in Australia have incorporated some practical bioinformatics into their molecular biology and biotechnology teaching, many using the resources of the Australian National Genomic Information Center (ANGIS, see later). In addition, several other universities also teach bioinformatics within their current Bachelor of Science/Bachelor of Biotechnology degree structure. They allow students to major in bioinformatics but do not offer Bioinformatics as a degree in its own right. These courses will not be covered in this article.

2 Australian Degrees in Bioinformatics

2.1 Undergraduate Degrees

In 2003 there will be nine Australian Universities that will be providing specified Bioinformatics degrees. As of 2002 only two (University of Sydney and La Trobe University) have produced graduates. Of the remainder three will be taking applications for the first time in 2003 (Macquarie University, University of Southern Queensland and University of Western Australia).

The content of these degrees differs from course to course but there are a few underlying similarities (see Table 1). The courses can be roughly divided into two streams; a degree that produces bioinformatics programmers and a stream that produces bioinformatics technicians. The former includes the degrees that have the major emphasis on computer science, mathematics and statistics while the latter is more biological in orientation.

Bioinformatics, with its emphasis on computer science, mathematics and statistics, had not been considered a suitable candidate for distance education due to the molecular biology component. Also the universities that had so far produced the Bioinformatics degrees tend not to have strength in the area of distance education. However, in 2003 the University of Southern Queensland will launch is Bachelor of Bioinformatics degree that can be taken by distance education with compulsory residential schools in the semester break.


2.2 Postgraduate Degrees

Despite the number of undergraduate degrees on offer, only two of the nine universities have established a post graduate degree. The University of Sydney has set up a Graduate Certificate, Graduate Diploma, Coursework Masters degree combination that has been running since August 2002. These post graduate coursework degrees are directed at both working scientists who need retraining in bioinformatics and computer scientists who wish to be training in biology to work as bioinformaticians. A Graduate Certificate in bioinformatics has been available at Curtin University for several years, however, there have been no graduates as yet.

2.3 Award Courses

2.3.1 ANGIS

ANGIS has been training working scientists in the application of currently available bioinformatics programs for 12 years. Courses are given in database searching, applied bioinformatics and the use of the ANGIS bioinformatics interfaces. The aim of these courses is to train applied bioinformaticians and not computer programmers. Courses are held at least once a year in Sydney, Melbourne, Brisbane and Adelaide.

2.3.2 Biolateral

Biolateral is a Sydney based company that offers intensive courses in bioinformatics application and theory. These courses include training in computer programming, and the analysis of data from new technologies such as microarrays.

References

Additional information about these degrees can be found at http://www.angis.org.au/new/education/study.html along with links to the individual university course information sites.

Table 1: Undergraduate Degrees in Australia (2003)

<table>
<thead>
<tr>
<th>University</th>
<th>Degree Offered</th>
<th>Emphasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Sydney</td>
<td>Bachelor of Science (Bioinformatics)</td>
<td>Molecular Biology, Computer science</td>
</tr>
<tr>
<td>La Tobe University</td>
<td>Bachelor of Bioinformatics</td>
<td>Mathematics, Molecular Biology</td>
</tr>
<tr>
<td>Flinders University of SA</td>
<td>Bachelor of Science (Bioinformatics)</td>
<td>Mathematics</td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>Bachelor of Engineering (Bioinformatics)</td>
<td>Programming</td>
</tr>
<tr>
<td>University of Western Sydney</td>
<td>Bachelor of Science (Bioinformatics/Chemical Informatics)</td>
<td>Molecular Biology and Mathematics or Computer Science</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Bachelor of Science (Bioinformatics)</td>
<td>Molecular Biology *</td>
</tr>
<tr>
<td>University of Wollongong</td>
<td>Bachelor of Computer (Bioinformatics)</td>
<td>Biochemistry, Mathematics</td>
</tr>
<tr>
<td>University of Southern Queensland</td>
<td>Bachelor of Bioinformatics</td>
<td>Molecular Biology, Computer Science</td>
</tr>
<tr>
<td>University of Western Australia</td>
<td>Bachelor of Science (Informatics)</td>
<td>Computer Science *</td>
</tr>
</tbody>
</table>
1 Summary

Singapore seeks to be an international center for the biomedical sciences and its related industries. Bioinformatics is seen as an important ingredient in this ambition. We provide in this short report a brief overview of bioinformatics in Singapore. We cover aspects such as training (Section 3), research (Section 4), and commercialization (Section 5). We also introduce some of the main centers of activities, as well as some of the bioinformaticists in these centers.

2 Background

As the Singapore government is the main driver to develop the biomedical sector into a new pillar of the Singapore economy, most of the main organizations that are currently involved in bioinformatics research and development in Singapore are government-funded institutions. Before we get into the details, it is useful to lay out these main organizations.

There are two large universities in Singapore, namely the National University of Singapore (NUS) and the Nanyang Technological University (NTU). The NUS has a longer history in bioinformatics and biomedical training and research, while the NTU did not have a life science school until recently. Both universities are in the process of stepping up their bioinformatics training and research programs.

3 Training Programs

There are four major university-level bioinformatics training programs in Singapore. The first and oldest program is the S* Life Science Informatics Alliance, an organization involving Karolinska Institutet, NUS, Stanford University, UC San Diego, University of Sydney, University of Uppsala, and University of Western Cape. This is a global alliance to provide a global, unified bioinformatics learning environment
made up of modular courses in the disciplines of genomics, bioinformatics, and medical informatics. The modules are assembled from selected lectures from the seven universities involved and are offered through distance learning. The main driver of this Alliance in Singapore is Shoba Ranganathan and Tin-Wee Tan in the NUS. Shoba also organizes a bioinformatics module in the NUS Department of Biochemistry that is aimed at biology and biochemistry majors.

The second program is a full-time MSc-level course organized by the BII in partnership with the NUS. This is a two-year course. It was started in 2002 and has a small but selective enrolment. A distinctive quality of the BII program is that it has a very active literature review and journal club series. The program is lead by Gunaretnam Rajagopal, director of the BII.

The third program is a part-time MSc-level course to be launched in 2003 in the NTU. It is also a two-year course. The course co-ordinator is Chee-Keong Kwoh. The renowned bioinformaticist Ming Li, of the University of Waterloo, is the external examiner of this course.

The last program is not yet a full program. It currently comprises a couple of very good bioinformatics modules, viz. combinatorial methods in bioinformatics and computation foundation in bioinformatics, offered to undergraduates at the School of Computing in the NUS. These two modules were mounted in 2002 after the arrival of Ken Sung at the school.

4 Research Projects

The NUS currently has several pockets of bioinformatics activities. In the Department of Biochemistry, Shoba Ranganathan and Tin-Wee Tan are active in Asia-Pacific Bioinformatics Network, a non-profit and non-governmental organization that promotes bioinformatics awareness and education in the Asia-Pacific region. Their group has done curation and research in alternate splicing, databases for structural biology, assignment of function based on 3D structural model (Ranganathan 2001), etc. In the Department of Computational Science, Yuzong Chen has been working on computer-aided drug design and dynamics of biomolecules. In the Department of Mathematics, there are also several researchers who have made interesting contributions to bioinformatics. Particularly notable is Kwok-Pui Choi and Louxin Zhang’s recent theoretical work on spaced seeds, which is an idea that can significantly improve the efficiency and sensitivity of sequence homology search.

The NUS School of Computing did not have any activity in bioinformatics until this year. Besides Ken Sung, there are now several other researchers in the school with an interest in this field: Gabriel Ciobanu is interested in the logical modeling of protein interaction; Roland Yap, Tzu-Yun Leong, Wynne Hsu, and Mong-Li Lee are interested in medical data analysis; and Beng-Chin Ooi and Kian-Lee Tan are interested in indexing of genomic and protein sequences (Ooi, Pang, Wang, Wong & Yu 2002).

The NTU has started its Bioinformatics Research Center (BIRC) in 2002. The key members of the BIRC include Jagath C. Rajapakse — who is an experienced researcher in neural networks and medical imaging, Chee-Keong Kwoh — who is an expert in probabilistic inference and medical imaging, and K. C. Lun — who is well-known in the medical informatics world. The research interests of BIRC covers a very broad spectrum, such as computational genomics, functional genomics, structural genomics, molecular imaging, and medical informatics.

The NTU School of Mechanical & Production Engineering also has a small team of bioinformatics researchers. The most experienced amongst them is Meena Sakharkar, who earned the first bioinformatics PhD awarded by the NUS. She has done excellent curation work on introns and exons (Sakharkar, Kangueane, Woon, Tan, Kolatkar, Long & de Souza 2000, Sakharkar, Tan & de Souza 2001). Besides this work, her group are also exploring peptide vaccine design and protein-protein interaction.

The BII was set up in 2001 to encourage, develop, and support trained expertise in bioinformatics to advance biomedical research and development in Singapore. It currently has skeletal research groups in systems biology, structural & functional genomics, computational genomics, medical genomics, and distributed computing. The key researchers are Gunaretnam Rajagopal — who has significant experience in supercomputing techniques for modeling complex systems, Pawan Dhar, Kuo-Bin Li, and Adrian Mondry. The BII scientific advisory board has many luminaries in the bioinformatics world, including Michael S. Waterman, Leroy E. Hood, and John Wooley.

The GIS is Singapore’s flagship in genomic sciences. The bioinformatics works conducted in GIS are largely in support of various biological research projects in GIS. The key bioinformatics researchers in GIS are Philip Long — who is well-known in the machine learning...
field, Prasanna Kolatkar — who is an expert in protein crystal structures, and Lin Kui — who is an excellent bioinformaticist. It is also worth noting that, David J. Lipman, a luminary of the bioinformatics field, sits on the GIS scientific advisory board. An important recent bioinformatics project at the GIS was the design and development of a *S. pombe* genome-wide ORF microarray. This is one of the few *S. pombe* microarrays that are being used in the studies of global gene expression patterns and their regulation. Another recent project was a study of cellular responses to estrogens. Here, bioinformatics was used to identify direct targets of estrogen receptors. This project has successfully produced over 100 early estrogen-responsive genes with putative estrogen response elements, which was a five-fold increase over the 20 or so known estrogen-responsive genes with estrogen response elements. A third project was the study of gene expression of hepatocellular carcinoma. Bioinformatics was used to examine the variation in gene expression patterns in paired samples of hepatocellular carcinoma tumour tissues and non-tumour counterparts. A significant on-going bioinformatics effort at the GIS is the development of a protein-protein interactions database to help better understand the intricate relationships between the individual proteins within different genomes.

The IMCB’s recent main bioinformatics activities were centered on its Fugu project. Its informatics team, lead by Elia Stupka, was the work horse that annotated and analyzed the draft Fugu genome sequence (Aparicio et al. 2002).

The IHPC does not have a bioinformatics group. However, its computational chemistry group, lead by Wu Ping, does have some bioinformatics capabilities, especially in the protein structure and modeling aspects. An interesting project of this group in 1998 was a study of prostate cancer where prognosis was made based on chemical data from hair samples of patients.

I²R currently has the most active bioinformatics researchers in Singapore — after all, it was the first organization to conduct bioinformatics research in a significant way in Singapore. The key bioinformatics researchers in I²R are Vladimir B. Bajic — who has created an extremely accurate human transcription start site prediction system, Vladimir Brusic — who is a pioneer in immunoinformatics research, Jinyan Li — who has made innovative contributions to datamining technology research with his idea of emerging patterns, See-Kiong Ng — who wrote the key genotyping automation software that was used by DeCODE to genotype the Icelandic population, and Limsoon Wong — who conducted the first piece of bioinformatics research in Singapore back in 1994. The group’s research also benefits from the participation of Louxin Zhang of the NUS Department of Mathematics — who is an expert in sequence analysis and phylogeny reconstruction algorithms, and Christian Schoenbach of RIKEN Japan — who is a member of the FANTOM02 core team.

As mentioned earlier, the first piece of bioinformatics research in Singapore was conducted in 1994 in I²R. It was the development of the Kleisli query system for the integration and manipulation of data sources that are highly heterogeneous, highly complex, and geographically dispersed. Amongst the first problems the system was applied to was the so-called “impossible” biological data integration problems identified by the US Department of Energy at that time (DOE 1993). This was one of the earliest and most general data integration system for biology (Buneman, Davidson, Hart, Overton & Wong 1995, Wong 2002). The system was subsequently licensed in 1998 to geneticXchange of Menlo Park for commercialization, see http://www.geneticXchange.com.

This project was followed by the FIMM project in 1996 which focused on curation of information pertaining to human functional immunology (Schoenbach, Koh, Sheng, Wong & Brusic 2000), and on development of methods for recognizing immunogenic peptides from antigens (Brusic & Zeleznikow 1999, Brusic, Petrovsky, Zhang & Bajic 2002). This project was a pioneer of the field that has now come to be known as immunoinformatics (Petrovsky & Brusic 2002). The group has made successful epitope predictions in several disease areas such as IDDM, melanoma, HIV, Malaria, renal transplantation, just to name a few.

By around end of 1998, the I²R group became sufficiently experienced and launched several major projects in the subsequent years aiming at post-genome knowledge discovery problems. The earliest amongst these later projects was the PIES project which was started in 1998 to use natural language processing techniques to extract protein interaction information from scientific literature (Ng & Wong 1999). The group also started the literature datamining track in the Pacific Symposium on Biocomputing series and organized the track for three years until 2002, when this topic was finally recognized as an important challenge (Hirschman, Park, Tsujii, Wong & Wu 2002). A prototype
of PIES was licensed in 2001 to Molecular Connections in Bangalore as the backbone of its curation system, see http://www.molecularconnections.com.

Following PIES, the Dragon project was started in 2000 to analyze and predict gene features in genomic DNAs. By late 2001, the project has delivered initial success — its human transcription start site recognizer is currently the most accurate in the world by a significant margin (Bajic, Chong, Seah & Brusic 2002). It has recently been licensed to BioBase of Germany to be integrated into their TRANSPLORER product in 2003, see http://www.biobase.de.

Most recently, the I^2R group has also embarked on a project to develop technologies for analyzing clinical data, including gene expression and proteomic data. A driving interest of this project is the research on new datamining and classification algorithms that are both highly accurate and highly comprehensible. Most of the initial ideas are based on the concept of emerging patterns (Li & Wong 2002a) and have delivered some initial successes (Li & Wong 2002b, Yeoh et al. 2002).

5 Commercial Sector Activities

In the commercial sector, the most notable is the National Cancer Center, where Patrick Tan is leading the OmniArray Research Group. This group investigates the influence of genetic differences on the development of certain cancer diseases. The group has made good progress in the use of cDNA microarray technology and in the management and analysis of microarray data (Ooi & Tan 2003).

Another organization is the Lilly Systems Biology Research Center managed by Santosh Mishra, a very seasoned bioinformaticist. The Lilly Systems Biology Research Center is a subsidiary of Eli Lilly and Company. The center is to deliver innovative computational tools to help identify novel drug targets and elucidate drug mechanisms.

On a smaller scale are start-up companies such as KooPrime, Helixense, Agenica, and several others. KooPrime is a spin-out from the NUS and specializes in laboratory information management and automation. Helixense is marketing a data integration platform called gRNA that they developed in-house. Agenica is a joint venture between Mitsui and the National Cancer Center to develop a molecular gene expression database for human breast cancer.

6 Closing Remarks

We would like to close this report by a few brief remarks on the history and progress of bioinformatics in Singapore.

As early as 1991 or 1992, there was already a limited amount of bioinformatics activities in Singapore that were championed by Tin-Wee Tan. However, these activities were mostly about mirroring of data collections and standard sequence searching applications.

The nature of bioinformatics activities in Singapore began to take on a more research-oriented character when Limsoon Wong started work on the Kleisli query system in 1994. This work generated significant interests from several large international pharmaceutical companies.

This helped convinced the Singapore Economic Development Board to provide the funding in 1996 to start a Bioinformatics Center in the NUS as a joint collaboration between the support services of Tin-Wee at the NUS and the research of Limsoon at the I^2R (then known as the Institute of Systems Science). The resources of this Bioinformatics Center allowed several talented
researchers to be recruited and nurtured into full-fledged bioinformaticists. Some of them have already been mentioned earlier: Louxin Zhang, Prasanna Kolatkar, Meena Sakharkar, Vladimir Brusic, Kui Lin, Shoba Ranganathan, and several others.

By 2000, the potential of bioinformatics in modern biomedical research and high-throughput biological investigations has been fully recognized. Therefore, A*STAR initiated significant new funding to encourage and support research and development in this area. The GIS was established as the flagship for high-throughput genomic research in Singapore. A year later, the BII was established to drive bioinformatics training and research. Both the NUS and the NTU also responded by establishing proper degree programs in bioinformatics in 2002 and 2003.

Today, as described in the main sections of the report, we see considerable bioinformatics activities in Singapore. We can therefore expect increased vibrance in this field in Singapore in the near future.

References


Bioinformatics Service, Research and Education at CBI

Jingchu Luo
College of Life Sciences, National Laboratory of Protein Engineering and Plant Genetic Engineering
Center of Bioinformatics, Peking University, Beijing 100871, China.
luojc@pku.edu.cn

Abstract
Bioinformatics plays an important role for in the research and development of the life science and biotechnology. This paper intends to give an overview of the activities of bioinformatics service, research and education at the Center of Bioinformatics, Peking University; the national node of the European Molecular Biology Network and the Asia Pacific Bioinformatics Network.

Keywords: Bioinformatics; EMBnet; APBioNet.

1 Introduction
It has been only 12 years since the term "bioinformatics" started to appear in biomedical journals, according to the text search against the PubMed database (Benson, et al. 1990). Only three papers related to this topic were found in 1993, but this number increased to 988 in 2002. A simple Google search (1 Jan 2003) using the keyword "bioinformatics" returned 999,000 entries of which 486,000 pages were in English. Needless to say the rapid growth of bioinformatics is due to the rapid growth of the internet, which in turn, is a reflection of the rapid growth of information science, during the last decade of the 20th century. The genome projects, especially the model organisms including human and mouse, pour sequence data into databases accessible over the internet. DNA chip technology and proteome research make the biological resource even more complex and comprehensive.

Computational approaches in biological science can be traced back to the early 1970s as indicated by Mount (2000) in his book. The work by pioneers of biocomputing, such as the PAM scoring matrix constructed by Dayhoff, the sequence alignment algorithms developed by Needleman-Wunsch and Smith-Waterman, the DNA sequence database initiated by Goad and the database search tools contributed by Lipman and Pearson have been the basis of, and are still major resources for the current research, development and application in bioinformatics.

The database of 3D structures of proteins (PDB) started to collect atomic co-ordinates of protein molecules 25 years ago (Bernstein et al. 1977). The Chou-Fasman and GOR methods of protein secondary structure prediction, and the efforts to analyse and classify, the attempt to predict and design protein 3D structures can also be dated to late 1970s (Sternberg 1978, Lišon 1979).

Driven by the huge amount of data and the availability of the information highway, a new paradigm shift has been emerged in the life sciences (Gilbert 1991). However, the diversity and complexity of the biological data make it difficult for biologists to access the information (Stein 2002). Integration of software tools and databases is in great demand in current bioinformatics research and development (Chichurel 2002).

The diversity and complexity of the biological data make it difficult for biologists to access the information (Stein 2002). Integration of software tools and databases is in great demand in current bioinformatics research and development (Chichurel 2002).
In April 1999, 41 senior Chinese scientists were gathering at the First Forum of the Core Sciences in the 21st Century organized by the Natural Science Foundation of China (NSFC). As implied by the title of this Forum, “Information science in the life science”, both Chinese scientists and funding agencies had recognized the importance of bioinformatics for biological research and biotechnology development. In April 2000 the first Chinese bioinformatics conference was held in Beijing, following a workshop “Life Science in the Internet Times”. Research in computational biology in China can be tracked to late 1980s (Zhang 1989). Scientists in several universities and academic institutions started to work on computer analysis of DNA and protein sequences, as well as protein modeling and design. Computer programs such as the sequence analysis package GoldKey were also developed. Application of computational tools to molecular biology has been used to help with experimental data analysis.

Research in computational biology in China can be tracked to late 1980s (Zhang 1989). Scientists in several universities and academic institutions started to work on computer analysis of DNA and protein sequences, as well as protein modeling and design. Application of computational tools to molecular biology has been used to help with experimental data analysis.

2 Activities

The European Molecular Biology Network (EMBnet, http://www.embnet.org) was formed in 1988 by European scientists and is an international organization for bioinformatics service, research and education. In 1996 the Center of Bioinformatics at Peking University (CBI, http://www.cbi.pku.edu.cn) joined EMBnet and became the national node of China. Together with colleagues in the Asia Pacific rim, CBI has also been actively working in the Asia Pacific Bioinformatics Network (APBioNet, http://www.apbionet.org). With support from the Ministry of Science and Technology, the Ministry of Education, the China Education and Research Network (CERNET, http://www.cernet.edu.cn) and the various help from EMBnet colleagues, CBI has been playing a significant role in bioinformatics service, research and education.

2.1 Service

CBI provides bioinformatics resources for domestic users, which include both databases and software tools. More than 140 databases such as the DNA and protein sequences and structures, genome and sequence-related mapping, mutation and literature reference databases have been installed in the database query system SRS (http://srs.pku.edu.cn). Mirrors of dozens of sites such as ExPASy, GoldenPath, ENSEMBL, GDB, RGD (Rat Genome Database), S-Star (Bioinformatics online education), have been set up to make access easy for domestic users. Other bioinformatics information, such as description of software tools, user manuals, documents and workshop presentations were also put online either in English or in Chinese. A local BLAST server and the sequence analysis GUI JEMBOSS are open to the public through the internet. A molecular design platform is accessible for local users. All the public databases and programs can be downloaded via the CBI FTP server.

2.2 Research

In addition to the bioinformatics service, CBI carries out bioinformatics research and development. The human proteome annotation system has been set up and the protein annotation knowledge database was constructed (http://pak.cbi.pku.edu.cn). Prediction of secreted proteins from the human proteome and construction of the secreted protein database have been completed. Specialist database information systems for the protein neurotoxin database, rice dwarf virus and China’s ethnic genome diversity have been constructed. 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 (Wingendar et al. 2001). A prokaryotic genome assembly assistant system (Yu et al. 2002) and a microbe genome annotation platform were developed. Jupdate, a Java interface for the
automatic updating and maintaining of biological databases is being tested. Based on the idea that “half day on the web saves you half month in the lab”, an integrated bioinformatics platform WebLab is being developed. Projects of algorithm development, genome sequence analysis, structure prediction, molecular design and drug discovery are all active fields (Wang et al. 2002).

2.3 Education

CBI is playing a leading role in bioinformatics education and training in China. Both Ph.D. and MSc programs for bioinformatics and biocomputing have been started to meet the need of this fast growing field (Pearson 2001). A practical course “Application of bioinformatics to molecular biology” has been running for graduate students of biology, medicine and agriculture at the College of Life Sciences and the Medical School of Peking University, as well as China Agricultural University and China Academy of Agricultural Sciences. Various seminars, workshops and courses on bioinformatics and biocomputing have been organized during the past five years. Bioinformatics experts were invited to give lectures and to help with hands-on practice for participants not only from mainland China, but also from Hong Kong, Taiwan and Thailand. Text books were edited or translated (Luo et al. 2002).

In addition to the bioinformatics service, CBI carries out bioinformatics research and development. Specialist database information systems for the protein neurotoxin database, rice dwarf virus and China’s ethnic genome diversity have been constructed.

3 Future Direction

It is now well recognized by both the domestic and international bioinformatics communities that CBI has been playing an important role in providing bioinformatics services, carrying on research and training bioinformatics talents in China. Nevertheless, CBI is still in its infancy. We are keen to learn experiences of, seek assistance from and exchange ideas with bioinformatics and biocomputing experts from all over the world.

References


automatic updating and maintaining of biological databases is being tested. Based on the idea that “half day on the web saves you half month in the lab”, an integrated bioinformatics platform WebLab is being developed. Projects of algorithm development, genome sequence analysis, structure prediction, molecular design and drug discovery are all active fields (Wang et al. 2002).

2.3 Education

CBI is playing a leading role in bioinformatics education and training in China. Both Ph.D. and MSc programs for bioinformatics and biocomputing have been started to meet the need of this fast growing field (Pearson 2001). A practical course “Application of bioinformatics to molecular biology” has been running for graduate students of biology, medicine and agriculture at the College of Life Sciences and the Medical School of Peking University, as well as China Agricultural University and China Academy of Agricultural Sciences. Various seminars, workshops and courses on bioinformatics and biocomputing have been organized during the past five years. Bioinformatics experts were invited to give lectures and to help with hands-on practice for participants not only from mainland China, but also from Hong Kong, Taiwan and Thailand. Text books were edited or translated (Luo et al. 2002).

In addition to the bioinformatics service, CBI carries out bioinformatics research and development. Specialist database information systems for the protein neurotoxin database, rice dwarf virus and China’s ethnic genome diversity have been constructed.

3 Future Direction

It is now well recognized by both the domestic and international bioinformatics communities that CBI has been playing an important role in providing bioinformatics services, carrying on research and training bioinformatics talents in China. Nevertheless, CBI is still in its infancy. We are keen to learn experiences of, seek assistance from and exchange ideas with bioinformatics and biocomputing experts from all over the world.

References


