Bioinformatics Industry in Australia

1 Introduction

In the last couple of years, the role of bioinformatics in Australia has been recognized as a key ingredient in the development of high tech industries in the country (Littlejohn 2000, Littlejohn 2002a). This recognition is based on a strong and growing biotechnology sector (Littlejohn 2002b, Thorburn and Hopper 2002), which is underpinned, by vibrant life science research and ICT (information and communication technology) R&D.

Recently an industry analysis of bioinformatics in Australia was conducted and the report (Littlejohn 2002a) recommended that four major areas be focused on:

a. Education and skills development;

b. Research;

c. Infrastructure;

d. Commercialization.

The taskforce identified that skills development in bioinformatics is the most pressing issue for industry development. But what skills? This very much depends on the definition of bioinformatics.

2 Defining Bioinformatics

Bioinformatics is one of the most frequently used terms in the modern molecular life sciences and bioindustries, and perhaps one of the least well defined.

Common usage of bioinformatics defines it broadly as the intersection point of the life sciences and technologies (e.g. genomics) with the information sciences and technologies (e.g. computing). Further refinement of the term, however, typically leads to disagreements: Fig. 1 illustrates the level of semantic confusion that surrounds the term “bioinformatics”.

There are four layers to bioinformatics (Littlejohn 1997) that have been used in various ways over the last five years. All are valid uses of the term bioinformatics but all are very distinct and their definitions are frequently the subject of heated debate. These layers are:

- **Users = biologists.** This refers to the application of bioinformatics technologies to the life science discovery process. However, much as molecular biology was a discipline and later became a technique and its practitioners were no longer considered molecular biologists (Morange 1999), those scientists who use bioinformatics tools should no longer considered bioinformatics practitioners.

- **Product and service builders and deliverers = BioIT.** The newest term used to describe bioinformatics, BioIT should be associated with the “IT” (technical) aspects of bioinformatics. Examples of this might be provision of an online service such as a WWW interface to a databank.

- **Engineering = BioIT.** Software, database and computing systems engineering should also now be considered BioIT. Examples include development of new software tools for biological data management.

- **Research = bioinformation science.** The term “bioinformation science” has recently been applied to describe the research activities in bioinformatics. In the past the term “computational biology” has also been used to describe “pure” bioinformatics research. Research also overlaps with biostatistics, especially in areas such as bioinformatics problems in functional genomic data such as microarray datasets. Examples of research in this area include gene detection algorithms in novel genome sequences.

3 Bioinformatics Consumers

In Australia, the bioinformatics “industry” is best carved into two slices: “consumers” bioinformatics of products and services and “suppliers” of these products and services. Owing to the ambiguity in definition of the term “bioinformatics”, these two need to be addressed separately.
A large proportion of publicly funded life science research in Australia in organizations such as universities, hospitals and the CSIRO are big consumers of bioinformatics. There are too many of these to consider separately here.

On the commercial side, the biotechnology industry in Australia is a growing consumer of bioinformatics technologies. In fact, the peak biotechnology industry body, AusBiotech, has recently launched a Bioinformatics Special Interest Group (ABSIG) in recognition of the growing importance of bioinformatics to biotechnology industrial growth in Australia (http://www.ausbiotech.org).

Biotechnology firms that heavily use bioinformatics (some even having their own bioinformatics departments) include Bionomics in Adelaide (http://www.bionomics.com.au), GeneTraks in Brisbane (http://www.genetraks.com) and the Brain Resource Company in Sydney (http://www.brainresource.com). Instrument companies such as Proteome Systems Limited (PSL) in Sydney (http://www.proteomesystems.com) are also voracious consumers of bioinformatics and producers of bioinformatics tools.

Several publicly funded or semi-commercial organizations are also consumers of bioinformatics, including the Innovative Dairy Products Cooperative Research Center in Sydney (http://www.dairycrc.com) and the Intelligent Island Center Of Excellence in Bioinformatics in Hobart (http://www.intelligentadvantage.net.au). In addition, Australia has three national facilities that are consumers of bioinformatics — ANGIS in Sydney is Australia’s National Bioinformatics facility (http://www.angis.org.au) and also a supplier of bioinformatics to many research and educational organizations in Australia. In addition, APAF in Sydney as the National Proteomics Facility (http://www.proteome.org.au) and the AGRF in Melbourne and Brisbane, as the National genomics facility (http://www.agrf.org.au), are also heavy bioinformatics consumers.

4 Bioinformatics Suppliers

Australia has a number of pure bioinformatics companies offering products and services not just to the consumers mentioned above but also to organisations outside Australia. These companies include Sydney based BioLateral (http://www.biolateral.com.au) and Desert Scientific (http://www.desertsci.com), Canberra based GeneticXchange (http://www.geneticxchange.com/v3/index.php), Brisbane based GeneXbase (http://www.genxbase.com) and GeneticSolutions (http://www.geneticsolutions.com.au) and Perth based Biogene-E-Us (http://biogene-e-us.com).

In addition to these “pure” bioinformatics companies, some instrument companies are also suppliers of bioinformatics products. These companies include Axon Instruments (http://www.axon.com) and PSL (http://www.proteomesystems.com). Australia has a number of clinical trials software companies too, including Sydney based iLife (http://www.iLife.com.au) and Phase Forward (http://www.phaseforward.com).

The life science sector is seen as one of the fastest growing markets for the IT industry, and as a result, a number of IT firms are actively selling equipment, software and services into the sector. These firms include IBM (http://www-3.ibm.com/solutions/lifesciences), Apple (http://www.apple.com.au), Cray (http://www.cray.com), Sun (http://www.sun.com) and HP (http://www.hp.com).

Australia has a large number of publicly supported research groups with active research programs in bioinformatics (see accompanying article). These groups produce bioinformatics innovations that are in turn consumed by the life science research communities mentioned above. Bioinformatics research groups include Canberra based CBiS (Center for Bioinformation Sciences) at the ANU (http://cbis.anu.edu.au) and the University of Canberra Medical Informatics Center (http://nhsc.org.au), Melbourne based WEHI (http://www.wehi.edu.au/bioweb/index.html) and Victorian Bioinformatics Consortium (http://www.vicbioinformatics.com), Sydney based PeterWills Bioinformatics Center at the Garvan Institute (http://www.garvan.org.au) and the Centenary Institute.
5 Future Industry Directions

The future of bioinformatics in Australia will very much be shaped by how rapidly the life science research sector gains the skills in bioinformatics. As a consequence of this recognized demand for skills, Australia has seen rapid growth in the range of bioinformatics education opportunities available for both life scientists and information scientists. BioLateral, a Sydney based bioinformatics company, offers professional training in a range of bioinformatics topics (http://www.biolateral.com.au), and many universities now offer bioinformatics training at the undergraduate levels. Undergraduate courses are available at universities around the country (summarized at http://www.angis.org.au/new/education/study.html) and post-graduate courses are available from the University of Sydney (http://www.scifac.usyd.edu.au/future/pg/pgc_bioinform.html) and Curtin University (http://www.curtin.edu.au/curtin/handbook/courses/30/30 1135.html). It will be interesting to watch the growth of the industry as skills levels rise over the coming years.

The life science sector is seen as one of the fastest growing markets for the IT industry, and as a result, a number of IT firms are actively selling equipment, software and services into the sector.

References


Abstract

The Oracle relational database management system, with object-oriented extensions and numerous application-driven enhancements, plays a critical role worldwide in managing the exploding volumes of bioinformatics data. There are many features of the Oracle product which support the bioinformatics community directly already and there are several features which could be exploited more thoroughly by users, service vendors, and Oracle itself to extend that level of support. This paper will present an overview of Oracle features which support storage of bioinformatics data and will discuss extensibility features which give the product room to grow. Some attention will be given to Oracle’s own efforts to use that extensibility to exploit emerging standardization of many of the complex data and computation requirements of the life sciences.

Keywords: Oracle; database; bioinformatics; extensibility.

1 Introduction

Huge volumes of bioinformatics data are emerging from DNA and protein sequencing efforts, gene expression assays, X-ray and NMR protein structure determination, and many other activities. Almost every important effort to understand life processes or to develop therapeutic drugs is a producer or consumer of large data volumes. Database management systems have a very important role to play in the storage and serving of this explosion of data. Relational technology has emerged as the standard for databases in many fields including bioinformatics. Oracle’s relational database management system (RDBMS) is now used for about 85 percent of the bioinformatics database market. The
Oracle RDBMS has evolved from a simple relational data store to one that enables complex data structures to be explicitly represented in the database, and for those data structures to be provided with diverse functional behavior within the server. There are emerging standards for the content of bioinformatics data structures and for their associated functional behavior. It is Oracle’s plan to provide implementations of many of these standards so that users of bioinformatics data can concentrate on science and not on data management. This paper will discuss features of the Oracle RDBMS that support bioinformatics and will talk about plans for future extensions.

2 The Oracle Object-Relational Database

Object-relational database management systems (ORDBMS) incorporate object-oriented capabilities in a database environment to better support specialized applications. Starting with the 8i release, the Oracle database has been a fully functional object relational database which extends the relational database in three main areas: (1) definition and storage of new datatypes; (2) incorporation of object behavior in the query language; and (3) extensible indexing and the query optimizer.

In the Oracle database, a user can define new data types that closely represent the objects used in the application. With this approach, applications can focus on object manipulation logic (the object behavior) without worrying about the object storage and retrieval logic. The query language used in all the commercial relational database systems is SQL, and it can be extended to integrate the new operations defined by applications with standard SQL queries. This has the advantage that all the operations on the database objects are handled using a standard query model resulting in simplified application logic.

Indexes are optional structures associated with relational tables. Indexes speed up the execution of SQL statements in the database by providing a faster access path to the data, primarily by reducing disk I/O. Traditional databases provide an indexing mechanism which can work with simple data but this is not suitable for use with application-defined complex data types. With Oracle Extensible Indexing, the application can define and manage the structure and semantic content of the index, and can base the index on properties of complex abstract data types that are only understood by the application. The database system then interacts with the application to build, maintain, and employ the index. The main advantage of this extensible indexing framework is that the index is always in synchronization with the data; once the index is built, all the updates on the base table will automatically result in updates in the index data. Thus the users are not required to worry about the data integrity and index currency issues. Once the index is built, it is treated like a regular B-tree index. The database server knows the existence of the index and manages all the index-related work using user defined functions. In this extensible indexing model, the query optimizer also understands the application’s index structure and operators. When the user-defined operators are evaluated, the corresponding index can be used to efficiently process the SQL query. This is similar to using a B-tree index (built in to Oracle) while evaluating an equality operator. The application can store the index data either inside the Oracle database (in the form of tables) or outside the Oracle database (in the form of files) although it is highly desirable for the database to handle the physical storage of the index. By doing this, the index data will have the same security and reliability as the rest of the data in the database.

3 Current Applications of Oracle Extensibility in Bioinformatics

Many independent service vendors (ISVs) and Oracle customers have taken advantage of the database features for object definition and storage to create database abstractions for sequences, structures, annotations, and other bio data. Furthermore Oracle’s object-oriented extensibility permits the behaviors of the abstractions to be programmed as well. This powerful feature enables ISVs and in-house teams to offer domain products based on Oracle that have functional capabilities that are tightly integrated with the data. The externally programmed functional capabilities then become accessible to users and other applications via SQL or Java APIs.

Oracle extensibility also includes extensible indexing, a very useful feature for querying over complex scientific data types in large stores. Several life sciences vendors, including MDL and Daylight, have built chemoinformatics cartridges which use extensible indexing to do fast searching for structural subgraphs in large tables of small molecules. These proprietary indexes, whether based on hash keys (Daylight) or digital fingerprints (MDL), are fully registered and made active within the Oracle database via extensible indexing so that standard SQL queries can implicitly make use of them.
4 Future Oracle Plans to Extend the Database for Bioinformatics

When developing a major application-driven extension to the DB server the issue of standardization is very important. This applies to the data dictionaries or ontologies of structured data as well as to the specifications for standard computations. The burgeoning complexity and quantity of life sciences scientific data present such an immense challenge that uniform data standards (ontologies) for sequences, structures, gene-expression profiles, annotations, etc. are absent, with some exceptions. This is not the same as saying there is no format for the data. There are lots of formats, some with an old and awkward history. The same is true for standard computations.

In making its entry into this rich area, Oracle is focused initially in two areas: sequence searching and alignment, which together we will call sequence matching; and macromolecular structures.

In the case of sequence matching, it is the computational requirements that have become more or less standard. Although there is no single standard for the underlying data type, the linear DNA or polypeptide sequence, this is a fairly simple data type as things go with biological data. Oracle will be delivering a server-based implementation of the Basic Local Alignment and Search Tool (BLAST) together with its variants BLASTP, etc. in a near release of the product. This sequence alignment capability will operate on sequences stored in the database as character large objects (CLOBs). It will be simple to move the services later to a proper ontology-based representation of sequences and annotations in the database, if suitable standards emerge for sequence objects. In the meanwhile Oracle is investigating data compression techniques for sequences so that we may one day do an efficient storage implementation of an emerging standard. Further on the horizon Oracle is looking into fast ways of solving the general pattern matching problem for sequences, so that rapid searches could be done in the database for things like flexible promoter sequences, intron/exon boundaries, and statistical secondary structure predictors in the case of protein sequences. There have been no standard computational requirements to emerge for this generalized matching (although there are lots of software implementations for particular require ments), but this does not preclude R&D on powerful, general methods for matching sequence data that will be useful and efficient in the database for a lot of typical problems.

In the case of macromolecular structures, it is the data dictionary for the solved structure that has shown signs of standardization. The Protein Data Bank (PDB) is now supporting a specification for a structured data type for macromolecular structures based on a thorough scientific ontology called the Macromolecular Crystallographic Information File (mmCIF) (Bourne et al. 1997) This specification has received the blessing of the Object Management Group, Life Sciences committee, as a standard and has been served to users as an option by the PDB since August 2002. Oracle intends to do an efficient schema and object data type implementation of mmCIF and make it part of the DB server offering. Initially Oracle will provide a SQL and Java interface to the macromolecular structures in the database consistent with the low level accessor/depositor functions of the OMG standard. Over time Oracle will layer computations over the data structures as standard computational requirements emerge. Already Oracle has identified many common operators that are believed to be of wide and uniform applicability to molecular biology problems. In other words, Oracle will implement standards about the things researchers commonly want to ask about large molecules, whether explicitly captured in the data or derived by computation. With such a smart database implementation of mmCIF, researchers will be able to load and utilize not just public molecular structures but also proprietary structures and those derived via homology, or ab initio protein folding, if the latter problem is ever solved.

In the Oracle database, a user can define new data types that closely represent the objects used in the application. With this approach, applications can focus on object manipulation logic (the object behavior) without worrying about the object storage and retrieval logic.

References
BioITWorld Conference & Expo - HK 2003 will highlight the following two catalogues:

**Medical Informatics:**
- Database Storage
- Data-mining/KM/Software
- System Integration
- Digital Imaging
- IT infrastructure
- Integrated ERP Systems... etc.

**Bioinformatics:**
- Biosynthetic Technology
- Biotechnology Discoveries
- Therapeutics and Diagnostics
- Basic or commercial research
- Microarrays / Biochips
- New drugs or protein discovery... etc.

**Opening Remarks:**
Professor Lap-Chee Tsui, Vice-Chancellor,
The University of Hong Kong
The Better technologies to study proteins, tools to handle Bioinformatics, and methods to identify the causes of common diseases.

**Advisors of BioITWorld Conference & Expo - HK2003**

- Dr. Kevin Davies, Editor-In-Chief, Bio-IT World
- Dr. Edward Sellers, Member of the World Health Organization Expert Advisory Panel on Drug Dependence.
- Dr. Haffner ME, MD, M.P.H. Director, Office of Orphan Products Development, The Food and Drug Administration

Reply Slip: Please fax to (852) 2529 9956 or email to paul_szeto@idgexpo.com.hk
Enquiry: Please contact Mr. Paul Szeto at (852) 2233 9347

Name
Title
Company
Address
Tel
Fax
E-mail

I would like to exhibit at BioITWorld HK 2003. Please contact me.
I would like to sponsor at BioITWorld HK 2003. Please contact me.
I would like to attend BioITWorld HK 2003. Please send us further information.
I would like to be Speaker at BioITWorld HK2003. Please contact me.
Bowling Ball
‘Bioinformatics’ hits Ninepins of Asia Pacific

by
Balaji K.
(Healthcare Practice, Asia) FROST & SULLIVAN
jgovan@frost.com

For many burgeoning start-ups, bioinformatics is considered to be “a pot of gold at the end of the rainbow”. The key questions are: Is there a real opportunity for them? And has the bowling ball called bioinformatics truly hit the ninepins of Asia Pacific — Japan, Australia, India, Singapore, South Korea, China, Taiwan, and Indonesia?

1 Introduction

“Information is power” has been one of the oft-repeated clichés in corporate boardrooms. With information originating from various processes, biotechnology has reworked the old cliché. “Information capture, storage, management, and analysis is power”. To put it in a nutshell, biotechnology is looking toward advances in information technology, like for instance, bioinformatics, to get the job done faster and quicker.

To define, “Bioinformatics is the capture, management, analysis, and dissemination of information related to the emerging drug-discovery paradigm using genomics, combinatorial chemistry, proteomics, and high throughput screening”.

With pharmaceutical and biotech companies increasingly relying on infotech to provide them with solutions and tools to manage their growing data, bioinformatics is the most important part of the biotech wave to hit the Asia Pacific region.

2 The Global Snapshot

2.1 The opportunity

Drug research is data rich, but information poor. Genomics or gene-sequencing projects, high throughput screening, combinatorial chemical synthesis, gene-expression investigations, pharmacogenomics, and proteomics studies have created massive volumes and multiple sources of biological and chemical data. This data is threatening to create a bottleneck that might hamper drug discovery and development. The primary goal of bioinformatics is to link and convert this complex data into useful information and knowledge. As computing and biology have converged, software tools for data capture, management, analysis, mining, and dissemination have also emerged. The convergence of biotech and infotech has become inevitable.

It has been estimated that about 20 percent of the current novel discovery programs are based on genomics, and this is fueling the growth of bioinformatics. It is predicted that virtually all new discovery programs will be genomics-based in the near future. Currently, there is an increased pressure to develop breakthrough drugs and shorten the drug discovery time and costs involved. This presents an opportunity to bioinformatics companies as data
capture, management, analysis, and dissemination could play a vital role for drug discovery companies in containing both cost and time.

2.2 The Market

The total market for bioinformatics is estimated to be US$837.7 million in 2002 growing at a compound annual growth rate (CAGR) of 8.64 percent.

The global bioinformatics market is directly proportional to the investments taking place in the human healthcare R&D for new drug development. Pharmaceutical companies are under pressure to develop blockbuster drugs. With costs to take a drug from “bench to bottle” ranging from US$400 to US$900 million, and the percentage of drugs that make it to market, being low, the pharmaceutical companies are increasingly looking at biotechnology to deliver results. With informatics being the backbone of biotechnology, informatics market today is piggybacking the investments made in drug discovery and development and encompasses every stage of the pharmaceutical and biotech R&D process. Most of the data regarding the previously discovered gene sequences is available in the public domain, bringing down the cost of having to invest in costly Intellectual Property Rights (IPRs), and this has become a crucial driver for the growth of bioinformatics globally.

Globally, the markets in US, Japan, Western Europe, Taiwan, and Singapore are in a high growth phase, while those in the remaining parts of Asia and Latin America are in the medium growth phase. In Asia — Japan, Australia, India, Singapore, Taiwan, and South Korea have shown good potential as supplier of Bioinformatics tools. Software for data management and data analysis are on the higher end of value chain and contribute more revenues than other segments. Some of the important market players include Celera Genomics, Lion Biosciences, Informax, Accelrys, AlphaGene, Base4 Informatics, Gene Codes Corp., DNASTAR, Inc., Genomics Crop., Incyte Genomics, Molecular Informatics, Inc., Cognia Corp., CuraGen and Textco, Inc.

2.3 Regional Attractiveness Index

The following chart gives the regional attractiveness for bioinformatics. The index takes into account macro-economic variables such as GDP; market related variables such as market age, government initiatives, growth rate, IT expenditure, and degree of competition; and micro variables such as technical expertise and R&D expenditure of biopharma companies.

2.4 The Market Segments

The global bioinformatics market can be segmented into two — the supply side and the demand side.

The supply side consists of vendors in the market that provide data-generation software, which is used to control gene-sequencing hardware, data storage and management that allows the raw sequence data to be stored so that it can be analyzed and sequenced later, data-analysis software applications which are used to
mine data from available databases, and then perform various forms of analysis such as gene-sequence manipulation, comparison, alignment and the data dissemination, which makes available the information on discovered sequences to all concerned parties inside, and frequently outside, the organization.

The demand side consists of gene-based informatics, which comprises all the uses of bioinformatics techniques for the analysis of genetic elements. This includes the discovery phase of genomics (structural genomics), the target validation phase (functional genomics), and the stratification of patient populations based on genotype for diagnostics and therapeutics (pharmacogenomics). The other component of demand side is the Cheminformatics, which is the characterization and optimization of combinatorial libraries with the intent of using these libraries in the drug discovery process.

2.5 Trends

The global marketplace is currently witnessing the emergence of increased numbers of validated technologies with increasing levels of integration. Open data formats such as Java and XML are driving the market. The market is also witnessing increasing market consolidation and use of legacy systems. The decoding of the human genome sequence has accelerated the need for high throughput target identification and validation, driving the need for bioinformatics as a tool in both biotechnology and drug discovery. Today, IT capabilities cover the entire R&D chain, which includes gene identification, target identification and validation, and patient monitoring, further fueling the growth of bioinformatics. Companies are increasing the application of biotechnology in areas such as therapeutics, agriculture, industry, food, and so on. This increasing application of biotechnology is driving the growth of bioinformatics. An increase in the number of teaching institutes, which provide systematic training in bioinformatics, has ensured that adequate skilled manpower is available to take advantage of the opportunities presented. Huge volumes of data available in public and commercial databases along with the data generated by pharmaceutical companies poses formidable challenges for data management and also aids the growth of bioinformatics. The global market is also witnessing increased interest in bioinformatics from pharmaceutical, biotechnology, and IT companies, and this has resulted in a spate of mergers, acquisitions, and joint ventures, both among them and with bioinformatics start-ups as well.

3 Asia Pacific: Bioinformatics Bowling the Ninepins

Bioinformatics in Asia Pacific is driven predominantly by the government-funded initiatives. This is all set to change in the near future. More and more private informatics start-ups are being formed. Today, the informatics initiative is being driven by the large pharmaceutical companies in the region as well as the IT companies entering the arena with their technical skills. The Asia Pacific market is witnessing a shift to a large number of private sectors holding considerable sequence information in the proprietary domain. The biggest driver of bioinformatics continues to be the availability of significant amount of data in the public domain. This continues to play a significant role in terms of training a whole new generation of scientists, environmentalists, healthcare professionals, clinicans, computational biologists, bioethicist, and so on. The key countries in the region are Japan, Australia, India, Singapore, and South Korea. Taiwan, China, and Indonesia continue to lag behind these top five behemoths.
3.1 Japan

The chief promoter of Japanese bioinformatics industry continues to be Japan Biological Informatics Consortium (JBiC). JBiC is an organization of large chemical and pharmaceutical companies that aims to improve the international competitiveness of the Japanese biotechnology industry by using bioinformatics to the speed of R&D in all sectors of biotechnology. JBiC has spent approximately US$130 million till date on various bioinformatics projects. JBiC focuses on SNP, protein analysis, and establishment of e-commerce systems for the biotech industry. It coordinates with the various ministries associated with Japanese biotechnology such as Ministry of Education, Culture, Sports, Science and Technology (MECSST), the Ministry of Agriculture, Forestry, and Fisheries (MAFF), Ministry of Health, Labor, and Welfare (MHLW), and Ministry of Economy, Trade, and Industry (METI) with the various public-funded research institutes and the private sector companies.

One of the important private players in the Japanese bioinformatics arena is Hitachi Life Science. Hitachi offers a variety of research services, right from DNA sequencing and SNP discovery to genetic analysis, protein structure modeling, and establishment of e-commerce systems for the biotech industry. It coordinates with the various ministries associated with Japanese biotechnology such as Ministry of Education, Culture, Sports, Science and Technology (MECSST), the Ministry of Agriculture, Forestry, and Fisheries (MAFF), Ministry of Health, Labor, and Welfare (MHLW), and Ministry of Economy, Trade, and Industry (METI) with the various public-funded research institutes and the private sector companies.

3.2 Australia

The surge in bioinformatics field is fuelled by the advances in genomics and proteomics generating increasing volumes of biological data that need to be managed and interpreted. The Australian bioinformatics industry is still in its infancy. A number of Australian companies, research institutes, Cooperative Research Centers (CRCs), universities, and other organizations are active in bioinformatics and the development of bioinformatics technologies. A large volume of bioinformatics work happens in the universities and government research institutes. Developments within Australia’s bioinformatics capability include unique databases and libraries, innovative screening and analysis technologies, and various bioinformatics programs.

The important centers of action are The Australian National University, Monash University, The University of Queensland (Biological Information Theory group (BITIS), The Computational Biology and bioinformatics Environment), Murdoch University (Center for Bioinformatics and Biological Computing), Sydney University (Australian National Genomic Information Service (ANGIS)), CSIRO (Bioinformatics Research Group), and Macquarie University (Australian Proteome Analysis Facility (APAF)). The Walter and Eliza Hall Institute and the Garvan Institute of Medical Research (through the Peter Wills Center for Bioinformatics) both have dedicated bioinformatics research programs. Tasmania has announced an Intelligent Island Program, which supports the Tasmanian Center of Excellence for Bioinformatics.

3.3 India

India’s Department of Biotechnology (DBT) initiated a bioinformatics program to create a network-based infrastructure that now extends across 57 universities and public-funded institutions. This initiative comprises research groups involved in database creation, molecular modeling, and algorithm development. The DBT has pledged US$65 million for genomics research over the next five years, and has announced plans to enhance the infrastructure for bioinformatics research in public-funded institutions. The Indian bioinformatics market was worth about US$25 million in 2001, contributing to about 51 percent of the overall Indian biotechnology market.

The most crucial advantage for India lies in the low cost of development and the high success rates enjoyed by bioinformatics companies in the country.
The Indian advantage also stems from strong government support for the biotech sector and the availability of technical expertise, which gives an impetus to the research process and the demand for bioinformatics tools. The low number of competitors makes India attractive to new entrants. India has already made a mark in the IT sector and has a strong IT infrastructure, as well as skilled manpower, which would help it to become a strong player in the bioinformatics market.

Growing volumes of genomics data and increasing numbers of participants contracting work to Indian companies have encouraged many pharmaceutical, IT, and biotechnology companies to enter the bioinformatics sector. Indian IT companies such as Tata Consultancy Services (TCS), Cognizant Technologies, Infosys, Wipro, and Satyam have already set up their bioinformatics divisions. Not far behind are pharmaceutical companies such as Dr. Reddy’s Laboratories, Ranbaxy, Biological E, and Nicholas Piramal that have either opted for the joint venture route or have tied up with start-up companies. India is also witnessing the emergence of pure-play bioinformatics companies such as Strand Genomics. Government and private institutes, as well as participants such as the National Institute of Information Technology (NIIT), are fostering the growth of man power skilled in bioinformatics.

3.4 Singapore

Bioinformatics is an area where Singapore can leverage its position as a country with a thriving IT industry to its advantage. Unlike biological sciences, bioinformatics expertise in Singapore can be easier to develop as Singapore has an “IT culture”. However, bioinformatics is not just IT; the “bio” component is as important as the latter. Singapore expects its efforts in bioinformatics education to pay off in the form of a talented pool of bioinformaticians.

Singapore houses a number of key bioinformatics institutions. The Bioinformatics Institute (BII), a division within the Biomedical Research Council (BMRC), is one such institution. The BII, in partnership with the National University of Singapore (NUS), has set up a bioinformatics graduate program. This includes a two-year Masters program in Bioinformatics (by course work). The Singapore BioMedical Computation Resource (SBCR) is a national resource that will allow public (academic, research) institutions and organizations access genomic data as well as run biomedical and bioinformatics applications (BLAST, FASTA, and CLUSTALW). The Bioinformatics Center (BIC) at the NUS was set up in 1996 with an initial funding from the Economic Development Board (EDB). Now, it receives funding from A*STAR, GlaxoSmithKlineBeecham, and NUS. BIC aims to pursue leading-edge research in the field of bioinformatics and life sciences, promote and develop state-of-the-art services in bioinformatics, and to offer informatics solutions to complex biological problems for Singapore and international scientific users with possible commercialization.

BIC has research groups at Kent Ridge Digital Labs (now, I2R), the Center for Natural Product Research (CNPR), and the Institute of Molecular and Cell Biology (IMCB). New groups are emerging in other research centers and life sciences departments. BIC is among the first in the region to undertake research in bioinformatics.

The key bioinformatics initiatives of Singapore include the BioMed Grid (BMG), another development in this area. It comprises a network of distributed computing resources in Singapore and collaborates with other R&D centers in the US, Europe, and Japan. The other key initiative is the 5* Life Science Informatics Alliance. This alliance is an attempt by six top institutions from five continents to share their resources to establish global course modules in bioinformatics. The six institutions, led by the NUS Bioinformatics Center as the Secretariat, include Stanford University, Karolinska Institutet and Uppsala University in Sweden, University of Sydney, and the South African National Bioinformatics Institute (SANBI) at the University of Western Cape, South Africa.

3.5 South Korea

The South Korean Ministry of Science and Technology (MOST) declared 2001 as the Year of Biotechnology, and is setting out to promote developments in this industry. The ministry plans to nurture 600 biotech companies by the end of 2002 to make South Korea the seventh most developed country from its current 14th place in terms of competitiveness in the industry. Some of the key factors responsible for the country’s attractiveness include domestic potential, government policies and favorableness to multinational companies, investment climate, infrastructure, attitude of private sector toward cutting-edge technology, and South Korea’s position as an international hub. Bioinformatics plays a crucial role in Korean biotech plans.
An important advantage of Korea in bioinformatics is its mature IT industry. The bioinformatics onslaught in Korea is led by the chaebols such as LG, SK, and Samsung. These companies have established new bioinformatics teams. The bioinformatics industry also finds support from pharmaceutical companies, such as Chongkeundang and Green Cross. Some of the start-up bioinformatics companies in Korea are Bioinfomatix, Inc., Pax GENETICA, Inc, Macrogen, IDRTech, Badasoft Co., SmallSoft Co., BITEK CHEMS, Inc., IStech Co., IDGENE, Inc, RNA, Inc., Genoprot Co., and Proteogen.

Some of the public institutions involved in bioinformatics research include Korean Society of Bioinformatics (KSBI), Biological Research Information Center (BRIC), National Institutes of Health (NIH), Korean Institute of Science and Technology Information (KISTI), Korea Research Institute of Bioscience and Biotechnology (KRIIBB) and the 21c Frontier the Center for Functional Analysis of Human Genome.

Some of the public institutions involved in bioinformatics research include Korean Society of Bioinformatics (KSBI), Biological Research Information Center (BRIC), National Institutes of Health (NIH), Korean Institute of Science and Technology Information (KISTI), Korea Research Institute of Bioscience and Biotechnology (KRIIBB) and the 21c Frontier the Center for Functional Analysis of Human Genome.

4 Conclusion

Asia Pacific definitely enjoys a distinct advantage in bioinformatics. The strength of its IT sector along with the presence of a large base of man power skilled in IT would enable a number of countries in the region to grab opportunities in bioinformatics. This large pool of technically-skilled personnel can be used to meet the demand for bioinformatics in other regions as well. Growth in the volume of genomic data along with technological advances and increased automation of the R&D process present bright growth prospects for bioinformatics worldwide and Asia Pacific is well placed to harness the same.

If there are a couple of constraints, it would be Asia Pacific’s ability to hard and soft sell its skill sets to get a chunk of the global pie as well as the low expenditure on biopharma R&D locally. Only time can reveal whether countries in the region have the resourcefulness needed to play a winning game in bioinformatics.

The primary goal of bioinformatics is to link and convert this complex data into useful information and knowledge. The convergence of biotech and infotech has become inevitable. The total market for bioinformatics is estimated to be US$837.7 million in 2002 growing at a compound annual growth rate (CAGR) of 8.64 percent.

Abstract

This paper considers the problem of inferring an original sequence from a number of erroneous copies. The problem arises in DNA sequencing, particularly in the context of emerging technologies that provide high throughput or other advantages at the cost of an increased number of errors. We describe and compare two approaches that have recently been developed by the authors. The first approach searches for a sequence known as a Steiner string; the second searches for the most probable original sequence with respect to a simple Bayesian model of sequencing errors. We present the results of extensive tests in which erroneous copies of real DNA sequences were simulated and the algorithms were used to infer the original sequences. The results are used to compare the two approaches to each other and to a third, more conventional, approach based on multiple sequence alignment. We find that the Bayesian approach is superior to the Steiner approach, which in turn is superior to the alignment approach.