OPPORTUNITIES FOR THE PAKISTANI IT INDUSTRY WITHIN BIOINFORMATICS

MINISTRY OF INFORMATION TECHNOLOGY
GOVERNMENT OF PAKISTAN
ISLAMABAD

DECEMBER 2005

Printing
December 2005

Printed by
Roohani Art Press
Blue Area Islamabad. Ph: 051-2873274-2206849

Published by
Pakistan Software Export Board

The Funding Agency

The research study "Opportunities for the Pakistani IT Industry within Bioinformatics" is funded by the Pakistan Software Export Board (PSEB). PSEB is the entity within Government charged with the task of enhancing exports of software and IT enabled services (ITES) from Pakistan. PSEB is a guarantee limited company totally owned and funded by the Government of Pakistan. Any questions or comments about this report may be directed to PSEB Islamabad at 92-51-111-333-666 or through e-mail at research@pseb.org.pk .

Disclaimer

Contents

Executive Summary

An Overview of the Project

The present report explore the question 'whether bioinformatics can be considered as a potential business opportunity for the Pakistani IT Industry', 'what are the existing competencies of biologists that can be cashed upon in this respect' and 'what are the potential problems that will be faced by Pakistan in the process'. It provides an overview of the global bioinformatics industry, its segmentation, educational institutions, bioinformatics products and leading companies. The report also presents bioinformatics industry in India, Canada and a few other countries and explores the reasons behind emergence of India as a global bioinformatics hub in the developing world.

Finally, it concludes with a list of suggestions compiled in consultation with a group of about 50 biological scientists, IT professionals and bioinformatics education providers, on the steps that should be taken by Pakistan to develop bioinformatics potential.

Key objectives of the project were the following:

i. Review of available information and data to develop a comprehensive understanding of the status and potential of IT within bioinformatics as a discipline in Pakistan.
ii. Assessment of issues and constraints in the development of bioinformatics, both as a discipline and industry and identification of key factors for developing this field in Pakistan.
iii. Identification of opportunities for Pakistan's IT industry for collaborative ventures to enable them to obtain a share of the global bioinformatics business.
iv. Identifying the enabling measures to facilitate the development and growth of bioinformatics and IT industry to develop in this field.
v. Preparing a medium term strategy to this end.

These objectives were achieved by using a combination of primary and secondary data. The primary data was collected by conducting in-depth interviews while secondary data was collected by conducting desk research.

An initial desk research was carried out to gain background knowledge of the subject as well as to identify the stakeholders in bioinformatics business. Major part of objectives i, iii and iv were covered by conducting desk research. For this purpose published reports and statistics either electronic or in hardcopy were utilized through scientific journals, business reports, industry and technological magazines and online databases. A multidisciplinary team of researchers having background in Bioinformatics, Computer Science, Social Sciences and Biology was deployed for the purpose.

A survey of potential stakeholders of bioinformatics business was done to estimate the current state of biological research in the country, research interests of different groups, opinion on bioinformatics, use of bioinformatics software and identification of opportunities in local and international markets.

For the in-depth interviews, the sample was divided into 4 groups.

Group 1. Scientists working in biological sciences/ medicine.
Group 2. Information Technology (IT) Companies.
Group 3. Educational Institutions imparting bioinformatics education at various levels.
Group 4. Pharmaceutical companies.

The respondents were selected using snow balling sampling method. Literature search was done to identify prominent research institutes in the country. The preliminary workshop also helped to identify the stakeholders.

The head of the institute or senior member of the faculty was contacted for the interview. Initially an introductory letter was sent to the stakeholders requesting them for an in-depth interview. In case of no response the individuals was contacted three times.

The IT companies in group 2 were selected randomly from a list of companies working in three major cities i.e. Karachi, Lahore and Rawalpindi/Islamabad.

Many pharmaceutical companies were contacted from a list provided by Pakistan Pharmaceutical Association. However the response rate was low in this group. Only 3 respondents could be interviewed completely. Many refused to be interviewed on the pretext that they did not have information or that their companies do not have a research wing in the country.

The interviews were conducted in 4 cities Karachi, Islamabad, Lahore and Faisalabad. One interview from Quetta was done over the telephone and 1 questionnaire was filled from Australia. A team of interviewers having awareness of the subject matter conducted the in-depth interviews.

In addition to formal interviews, several experts in bioinformatics contributed their opinion through emails and telephonic correspondence.

A workshop was conducted towards the end of the project to bring together stakeholders from all over the country, share the report with them and to hear their opinion. The report was subsequently modified in light of the discussion generated during the workshop.

Introduction to Bioinformatics

Bioinformatics is an emerging field of science that combines biology, statistics and computer sciences to solve biological problems. It has application in drug discovery, genomic and protein analysis and particularly in biotechnology. The need for using mathematical algorithms and information management tools arises from an explosion of data in various fields of biology.

In the past one-decade, new technologies have enabled scientists to produce large databases of genomic data, of man and other species, protein sequences, expression of genes, population genetics studies, crop varieties information, etc. These projects have suddenly transformed biology into a 'data rich' science. New methods are therefore required for data capture, data mining and data visualization. These functions represent the basic tasks undertaken in bioinformatics.

Bioinformatics Market Size and Characteristics

The global bioinformatics market is estimated to be about one billion dollars. Half of this money goes into genomics, 20% into proteomics and drug discovery and the rest into other segments like pharmacogenomics and cheminformatics.

India aims to combine its strengths in IT and biotechnology to emerge as a major global hub in bioinformatics. Indian bioinformatics market was estimated to be 17 million dollars in 2004. However the government of India is hoping to grab at least a 5% share of the global bioinformatics market in a few years time.

Bioinformatics will perhaps create a moderate number of jobs, In India an estimated 1000 professionals will be needed annually for the next 5 years. Furthermore bioinformatics companies should expect revenues in a few million dollars only.

The Future of Outsourcing in Developing Countries

Outsourcing has developed as an important concept in today's business. It is cost and time efficient, can make use of the limited set of high quality professionals and expand service portfolio for companies. Inspite of a tremendous increase in demand, the US educational stream has not significantly increased the production of scientific workforce. As a result cost of research has gone higher. Many companies are realizing that they can get their research done abroad, at cheaper cost, without any compromise in quality. In clinical trials for example, GSK *(Glaxo smith Kline)* has announced that it will shift 30% of its clinical trials to countries like India and China.

As the technological difference between advanced and developing countries bridges and physical barriers become irrelevant, developing countries with a small but significant number of highly trained scientists and technicians will have a major opportunity for outsource based business. It can also help countries counter brain drain and bring back expatriates working in technical fields. Potential problems are lack of innovation and scientific infrastructure. However the prospects of Bioinformatics may however be better, considering that knowledge of data base and statistics are important in the business.

Relevance of Bioinformatics to Pakistan

Pakistan offers unique genetic resources in human population, crops and other species. In several areas such as crop varieties development, Pakistan offers a genuine promise. In certain areas such as, drug development, protein expression assays, population genetics and clinical trials, there are pockets of expertise in the country, which can be utilized to develop bioinformatics business. In other words, service provision in bioinformatics is possible, although human resources would have to be channeled towards that stream. At the moment, most researchers are not exploring their luck in knowledge based economy due to lack of awareness, interest, or absence of IT and management support system. This trend is however proving detrimental to academic research as well.

The stress should perhaps be on service provision in the beginning; product development at the moment is not feasible, due to a total lack of expertise.

Comparison with other Countries

Bioinformatics, although filled with great potential, is difficult; as it requires cutting edge biology knowledge and infrastructure coupled with advanced technical IT skills. Countries that are progressing in bioinformatics have a history that favors such development. Amongst them, America, Europe, Canada and Australia, excel in biological research and India, although not on the cutting edge of science, has made consistent efforts to develop life sciences industry in the country including

biotechnology and generic pharmaceuticals. In addition, impressive IT infrastructure has been developed to support biological research and collaboration, which includes BIOGRID INDIA, and Biotechnology Information System Network (BTISnet). Biotechnology industry in India is already a 1 Billion dollar industry.

Model of a Bioinformatics Company and the Need for Seeding Grants

Bioinformatics in a highly competitive field, there are a large number of companies offering products as well as services. After the dot-com crisis, many IT companies in India eyed bioinformatics opportunities, they were however unable to provide the desired quality. Many of them were subsequently closed. Such mistakes should not be made by Pakistan.

The development of bioinformatics business necessitates that many different government bodies, researchers and private entrepreneurs adopt a collective strategy. Awareness should be raised amongst scientists as well as IT companies through workshops and training camps and online forums.

Bioinformatics companies will consist of research scientists, IT professionals and marketing personnel. Most of the biological scientists are working in public sector research institutes and universities. These institutions need to adopt a liberal policy for sharing of remunerations from bioinformatics commercial projects. Further more, entrepreneurship should rank high on the desired qualities for scientists. Private entrepreneurship will be the forerunner in developing bioinformatics and biotechnology business in Pakistan. Public institutes may enter into partnership with private companies.

For any new company offering bioinformatics services or products, ‘proof of concept’, will be essential. This means that a nascent company will need to demonstrate its ability before getting business. The government should help in this stage by offering bioinformatics projects that would be beneficial to public research institutes. Such projects should be funded by HEC, COMSTECH or Pakistan Science Foundation, as long as the bioinformatics company has partnership with public institutions. These projects will help interested people to form a company, gather relevant expertise and knowledge capital and develop infrastructure for research. These projects could then be presented as a proof of the company's competence. It will add to the profile of these companies and help to attract outsource business. Proposed projects should be in the field of drug discovery, protein modeling, expression analysis, population genetics, database generation, etc.

Several research institutes like HEJ and Punjwani at Karachi University, NIBGE Faisalabad, Aga Khan University, and Center for Excellence in Molecular Biology, Punjab University have active research going on in various aspects of crop biotechnology, drug and vaccine development and production, genetics of disease, etc, Bioinformatics companies would start by active partnership of researchers at these and other research institutes in the country.

Bioinformatics Human Resource Development

When it comes to Bioinformatics education, Pakistan is training less than a hundred undergraduate students annually for the past couple of years, the first batch is expected to graduate this year. These programs complain of lack of appropriate interdisciplinary faculty and uncertainty about job prospects. Government support in terms of scholarship to students and collaboration with public research institutes are necessary to improve quality. Furthermore, the number of students should be increased, however a sprouting of substandard bioinformatics educational institutes should not take place.

Conclusion

It is clear that bioinformatics business will not precipitate overnight. High standard, concentrated efforts are needed, on a limited scale to develop a few segments of the field in Pakistan. Progress in the field will be gradual, so investment should be limited in the beginning, linked to progress as well as opportunities. The role of PSEB in short will be to raise awareness on the issue, bring together interested people, serve as a platform for training workshops and advocate the strategy to develop bioinformatics business proposed in this report and in consultation with other biologists and scientists. Bioinformatics is an enabler in biotechnology research and business, which will serve as the backbone of agriculture and industries of the future, hence can't be ignored by Pakistan.

1 Introduction to Bioinformatics

What is Bioinformatics?

Bioinformatics is the application of computer technology and statistics to the management of biological information. In the past decade bioinformatics has become an integral component of research and development in biological sciences. The need for bioinformatics stems from a fundamental change that is taking place in biology. Several large-scale projects like the genome projects, human diversity initiative, and new techniques like microarray analysis are turning biology into a data rich science. For the first time in history, predictions based on empirical data is possible in biology[1]. The vast amount of newly generated data requires revolutionary new methods for management and analysis of results. Thus giving birth to a new field, known as bioinformatics.

The Human Genome project

The need for bioinformatics was realized for the first time by the explosion of genomic information resulting from the Human Genome Project (HGP)[2]. The purpose of this project was to characterize the normal set of human genes and to produce a 'consensus' road map of the structure of the human genome. This includes finding the sequence of 23 pairs of human chromosomes and elucidating the functions of both genes as well as intervening sequences.

HGP data serves as a reference from which applied studies can examine the variation in special samples, such as people affected with a given disease. The HGP has produced a sequence of the entire human genome, however the exact characteristics of most genes and their functions remain unknown. In the next stage scientists are gradually trying to find correlations in the raw HGP data in order to find disease causing genes. HGP has given a boost to such research.

Utility of the genome data

The intended utility of the HGP has to do with understanding disease causation. Finding genes potentially associated with disease is only the first step. Once such a candidate gene is identified, the next step is to identify variations within that gene or its regulatory regions that is associated to a certain trait, such as susceptibility to disease. The association between a

[1] Kanehisa M, Bork P. (2003) Bioinformatics in the post-sequence era. Nat Genet;33 Suppl:305-10.

[2] Weiss, Kenneth M., *Coming to terms with Human Variations*, Annual Review of Anthropology, Vol. 27 (1998), 273-300

genetic sequence and a physical characteristic is only the first step. Biologically valid scientific explanation is however required to prove causality.

Complexity of the Genome Data

Several factors contribute to the complexity of genomic research. Some of them are outlined below:

- **Volume**

  The sheer volume of the genome data is mind-boggling. An estimated 3 billion units (base pairs) are present in the human genome. In addition an estimated 1.4 million variations exist in each individual, accounting for the bewildering human variations in physical attributes, disease susceptibility and possibly behavior. These variations are referred to as single nucleotide polymorphisms (SNPs)

  Apart from the human genome, the genomes of about 150 other species have been sequenced.

  The volume of genome data is doubling every 6 months or so.

- **Environment-Gene Interaction**

  Secondly, both genes as well as environment determine the physical *(and biochemical)* characteristics of living organisms. The environment and genes interact in complex ways to cause disease. Causal inferences, relating genes with disease are therefore much difficult to draw.

  The relationship of DNA sequences to phenotypes *(physical attributes of a living being)* is turning out to be less predictive than anticipated. As a result more detailed studies with larger sample sizes have to be conducted to fish out any subtle effects of genes. Variations in the genome also indicate the need for population specific diagnosis and therapies.

Biotechnology and Bioinformatics

Apart form biological research, bioinformatics has its application in the field of Biotechnology. Biotechnology is a set of techniques that involve manipulation or change of the genetic patrimony of living organisms. Biotechnology is a tremendously growing field with application in drug and vaccine production, medical diagnostic tests, biotechnology based foods, environmental cleaning, industrial processes and forensic science.

Bioinformatics is an enabler and a necessary tool for modern biotechnology.

2 Applications of Bioinformatics

Bioinformatics as a tool for the analysis of biological data has it's applications in all fields of biological research and development. Some of these fields are summarized below. More detailed descriptions of these application areas and the use of IT in them is provided later in the report.

Genomics - involves the use of vast databases and complex software to analyze the enormous amount of data that emerges from the sequencing or mapping of the human and other genomes. Bioinformatics offers the tools to mine the data and match the DNA information with the genes.

Proteomics - The large-scale study of proteins with the goal of understanding the functions of the millions of proteins in the human body and other important species of animals and plants. Unlike the genome which is a finite entity with a fixed number of base pairs of DNA, the proteome is "plastic", changing throughout growth and development and environmental stresses, as well as in pathological situations.

Pharmacogenomics - Pharmacogenetics is the study of how genes affect the way individuals respond to drugs. This relatively young field of medicine is poised to change the face of healthcare as we know it today, by allowing doctors to tailor medicines to an individual's genetic makeup.

Comparative Genomics - Comparative genomics is the analysis and comparison of genomes from different species.

Phylogenetics - Field of biology that studies the evolutionary relationships between organisms. It includes the discovery of these relationships, and the study of the causes behind this pattern.

Forensics - or forensic science is the application of science to questions which are of interest to the legal system. For example, forensic pathology is the study of the human body to determine cause and manner of death. Criminalistics is the application of various sciences to answer questions relating to examination and comparison of biological evidence, trace evidence, impression evidence, drugs and firearms

3 A Classification of Tasks in Bioinformatics

Bioinformatics is primarily concerned with production of databanks and analysis softwares. Users such as biotechnologists and research scientists may choose to install these components on a local computer system, or access them over the internet, using the publicly available databanks.

Bioinformatics currently deals with several main types of biological data:

- **Sequences and structure of genes and proteins.** Sequences are the simplest way to represent a macromolecule. The structure of genes that code for the sequence of amino acids in proteins is produced in this form by genome sequencing projects. Protein sequences are usually obtained via computer-based translation of genomic data.

- **3-D molecular structures.** These are obtained by physical measurements *(X-ray, Nuclear Magnetic Resonance)* combined with computer modeling.

- **Genome structure and function.** The genome of an organism is composed of its entire genetic material. Information on genome structure and function is a basic description that is continuously updated with new information including links to other databases.

- **Bibliographic data, such as abstracts of scientific articles**. The amount of scientific literature in biology has increased exponentially in recent years, especially after the onset of genome projects. This information is organized into a few large public databanks available through the internet.

- **Finding Genes in the DNA sequence of various organisms *(Gene Annotation)***

- **Developing methods to predict the structure and/or the function** of newly discovered proteins and structural RNA sequences

- **Clustering Proteins into families of related sequences** and the development of protein models

- **Aligning similar proteins and generating phylogenetic trees** to examine evolutionary relationship between different organisms.

- **Systems biology-** that looks at life as an integrated and interrelated system of genes, proteins and other chemicals instead of studying each of these components in isolation.

4 Industry Characteristics

Estimation of Bioinformatics Market

The prospects of bioinformatics are inseparable from its application areas, namely life sciences research, biotech, pharmaceutical industry and agriculture. Research developments in these fields herald an exponential growth of bioinformatics industry.

Estimation of the size of Industry

In year 2000, a large amount of equity investment in biotech, accounting to $31 billion, flowed into bioinformatics. Over all it is estimated that the pharmaceutical industry spends around 6% of its $9 billion drug discovery budget on IT, i.e. US$ 500 million per year.

A wide variety of estimates can be found on the current and projected market size of bioinformatics services and products.

The Confederation of Indian Industries (CII), predicted the global bioinformatics industry will produce an estimated turnover of $60 billion in year 2005. However this estimate seems highly exaggerated. Most other figures have put it to $840 million in 2003 and $1.8 billion by year 2007. It is estimated that the industry will grow at a rate of 17 percent per annum for the next few years.

US National Institute of Health has set $400 million for the development of bioinformatics in the next few years.

Market Estimate Based upon Sales

The difficulty to estimate the size of bioinformatics market also stems from the presence of a large number of private players. Frost and Sullivan Consultants estimated the market for bioinformatics databases, products and services at about $300 million with approximately half the annual sales coming from data suppliers and the rest from tools/ IT providers of various kinds.

Segment wise application of Bioinformatics

Segment wise application of bioinformatics is

Genomic	50%
Proteomics	20%
Cheminformatics	19%
Pharmacogenomics	11%

Further explanation of these application areas is given in chapter 2.

Bioinformatics and Drug Discovery

- **Drug Discovery**

  For the past decade, the focus of most drug discovery companies has been genomics – sequencing genome, identifying the protein-coding genes, evaluating gene functions, and transcription.

  Development in computer and hardware has facilitated the assembly and analysis of whole genomes. Predictive models have enabled the identification of coding sequences, the identification and evolutionary comparisons of genes and, in many instances, functional assignment (gene annotation).

- **Steps involved in bringing a drug to market**

  Although drugs were being discovered and marketed long before the advent of genomics, the process has been less methodical[1]. The clinical trials that tested the value of various treatments were virtually done on the pattern of 'black box testing' where the inputs were tested against the outputs without reference to the genetic makeup of the body to evaluate the impact of the drug. With the genome sequence uncovered, and most of the contents of the black box revealed, it is anticipated that drugs that target specific genes/ gene products will be discovered. The success of this proposition is highly correlated with the success in the field of functional genomics – an ability to accurately predict the functions of the genes after having analyzed its sequence.

  Recent research into this field has revealed that this path from gene to drug is much longer than anticipated.

  Identifying drug target: The first objective of drug discovery is to determine what genes are to be studied. From the vast pool of available genes, scientists need to shortlist those that are likely to have a bearing on the disease. In other words, proteins or cell structures involved in the pathogenesis of disease are to be found, these are known as 'drug targets'.

[1] The Business of Bioinformatics, 2002, available at http://www.bio-itworld.com/archive/111202/horizons_business.html

Screening for an effective chemical: The identification of drug targets opens up new opportunities for compound screening and optimization. If we have a validated target, we can predict its structure thereby decreasing the number of compounds that should be tested to identify an inhibiting compound. Several prediction algorithms are available for this purpose.

Drug Development and Clinical Trials: Once a compound has been optimized and is ready to advance to clinical trials, we should have a set of biomarkers that can be used for monitoring safety and efficacy. Clinical trials are the longest and the decisive step of drug development.

- **Role of Information Technology in Drug Development**

Each phase of the drug discovery process presents an opportunity for informatics to provide an automated solution. The scope and value of the solutions that cater to the various stages may differ, but each step in the drug discovery process can benefit from automation, by lowering cost and increasing throughput. The business opportunity is in knowing where your solution fits into the complete process.

Nearly 40% of the time of a typical pharmaceutical company is spent on information processing for effective drug discovery. IT needs are therefore growing enormously[2].

- **The Need for Bioinformatics in Pharmaceutical Research**

Human Genome Project has provided a great boost to biological research; however it has failed to yield the promise of molecular medicine, as yet. There is widespread disappointment in the pharmaceutical industry, that the progress to new drugs based on genetic knowledge has been frustratingly slow. It was promised that genetic knowledge would be rapidly converted to drugs for different disease.

Many believe that the methods of drug development are to blame which are still structured for another age. Developments in the field have yielded a large number of potential drug targets and NCEs *(Novel Chemical Entities).* However drug development is the bottleneck, delaying the translation of research into commercially beneficial products.

Many like Bill Heseltine, CEO of human genome sciences *(Rockville, MD)* believe the solution lies in out sourcing. Heseltine envisions a virtual company that out sources to groups around the world, each specializing in discovery, development, manufacturing and clinical trials. For some clinical trials, it might not be feasible to carry them out in

[2] Charles Cooney, Proceedings of International Symposium on Bioinformatics and Genomics, January 2001, India. www.educationandhra.com as accessed on May 4, 2005

the US, simply because there are not enough patients. Other hurdles include strict FDA regulations which according to some are even obstructionist[3].

In today's competitive business, the best strategy, no matter what field you are in, is preeminence: being the best at what you do. This is increasingly being accepted in drug discovery and development.

[3] Drug Development Is Virtually Dead. Technology Review, Reiss S, January 2005, p20

5 The Role of IT in Bioinformatics

The relationship between computer science and biology is a natural one for several reasons. First, the phenomenal rate of biological data being produced provides challenges: massive amounts of data have to be stored, analyzed, and made accessible. Second, the nature of the data is often such that a statistical method, and hence computation, is necessary.

Bioinformatics plays a key role in functionalities such as to gather, store, classify, analyze and distribute biological information derived from sequencing and functional analysis projects. The limitations of the wet-lab *(in vivo and in vitro)* experiments has caused a shift towards informatics centered *(in silico)* approach for data collection and analysis[1].

Summary of IT processes in Bioinformatics

The use of Information technology to facilitate biological research and development can be summarized into the following 5 processes,

1. Electronic Data Capture
2. Data Mining
3. Data Warehousing
4. Data Visualization
5. Customized Software

Electronic Data Capture

The field of bioinformatics is data driven to a large extent. One of it's main concerns are organizing vast amounts of data, finding patterns in it and converting raw data into knowledge. Biological data in the post genomic era is rich, complex & voluminous.

Pharmaceutical companies and other healthcare organizations increasingly rely on external relationships for important services, especially in data management, integration and analysis. Informatics-based solutions are sought to link databases of different types and formats, often held in different places. Innovation in the field depends upon integration of heterogeneous databases across functions and across companies.

[1] Larvol, Bruno L., L. John Wilkerson, *In Silico Drug Discovery: Tools for bridging the NCE gap,* Nature Biotechnology, Vol 16, 1998

Once the database has been created, it is expected to have the quality of being extensible and supply efficient algorithms to enable asking relevant questions in the database.

Data Mining

Data mining refers to analysis of data and the use of software techniques for finding patterns and regularities in different data sets. For this purpose, algorithms are devised, which help to find patterns by identifying the underlying rules and features in the data. The choice to apply a particular combination of techniques in a certain situation depends on both the nature of data mining task and the nature of the available data. The analysis process starts with a set of data, uses a methodology to develop an optimal representation of the structure of the data; thus transforming data into useful knowledge.

Data Warehousing

A data warehouse is a collection of data; gathered and organized so that it can easily be analyzed, extracted, synthesized, and otherwise be used for the purposes of further understanding the data.

Data Visualization

Data visualization is important in understanding the structure of proteins and other biologically important molecules[2]. The 2-D and 3-D structural representations help data to be perceived in unique ways by the researchers.

Customized Software

Many research groups would like custom made softwares, tailored to their specific needs. These may include computer programmes involved in a multitude of tasks including,

- Molecular Modeling
- Microarray gene expression analysis software
- Identification of coding sequences/genes
- Evolutionary comparisons of genes
- Functional assignment of genes
- Identifying novel biomarkers and drug targets based on computational approach
- Cutting DNA and characterizing the gene sequence
- DNA sequencing software
- Screening tools for drug discovery
- Visualization Tools

Bioinformatics Products

Bioinformatics analysis is applied to various kinds of biological data, e.g. taxonomy trees, relationship data from metabolic pathways, the text of scientific papers, and patient statistics. A large range of techniques are used, including primary sequence alignment, protein 3D

[2] Kuonen, Dr. Deigo, *Challenges in Bioinformatics for Statistical Data Miners*, Bulletin of the Swiss Statistical Society, 2003, Vol 46

structure alignment, phylogenetic tree construction, prediction and classification of protein structure, prediction of RNA structure, prediction of protein function, and expression data clustering. Algorithm development is an important part of bioinformatics. Techniques and algorithms are specifically developed for the analysis of biological data.

Bioinformatics products include individual platform software, laboratory workflow systems and complex databases. Following table summarizes the various areas in Bioinformatics that have emerged in order interpret the explosive amounts of biological data:

Data source	Data size	Bioinformatics topics
Raw DNA sequence	11.5 million sequences (12.5 billion bases)	Separating coding and non-coding regions Identification of introns and exons Gene product prediction Forensic analysis
Protein sequence	400,000 sequences (~300 amino acids each)	Sequence comparison algorithms Multiple sequence alignments algorithms Identification of conserved sequence motifs
Macromolecular structure	15,000 structures (~1,000 atomic coordinates each)	Secondary, tertiary structure prediction 3D structural alignment algorithms Protein geometry measurements Surface and volume shape calculations Intermolecular interactions Molecular simulations (force-field calculations, molecular movements, docking predictions)
Genomes	300 complete genomes (1.6 million – 3 billion bases each)	Characterisation of repeats Structural assignments to genes Phylogenetic analysis Genomic-scale censuses (characterisation of protein content, metabolic pathways) Linkage analysis relating specific genes to diseases
Gene expression	largest: ~20 time point measurements for ~6,000 genes in yeast	Correlating expression patterns Mapping expression data to sequence, structural and biochemical data
Other data		
Literature	11 million citations	Digital libraries for automated bibliographical searches Knowledge databases of data from literature
Metabolic pathways		Pathway simulations

Table 1: Sources of data used in bioinformatics, the quantity of each type of data, and bioinformatics subject areas that utilize this data

6 Application Areas in Bioinformatics and the Use of IT

The following section provides a review of different application areas and the use of IT in them. Following application areas are considered,

1. Genomics
2. Functional Genomics and Microarray Technology
3. Proteomics
4. Drug Discovery
5. Organizing biological knowledge in databases
6. Structural Bioinformatics
7. Pharmacogenomics and the Era of Personalized Medicine
8. Fields at the periphery of bioinformatics
    i. Laboratory Information Management Systems (LIMS)
    ii. Clinical Trials
    iii. Medical Informatics

Genomics

Genomics involves the storage and analysis of genotypic data, as represented by bio-sequences *(i.e. DNA, RNA and protein sequences).* The comparison of these biosequences, either between samples or between species, has been highly successful in elucidating the biological relevance of sequence variations. Various global projects have undertaken the task to sequence entire genomes, a notable project being the sequencing of the entire human genome, undertaken independently by a public as well as a commercial group.

Technical advances in Robotics and instrumentation have resulted in the ability to rapidly determine the complete genome sequence of complex organisms. Newer genomic technologies have generated experimental data of gene sequence, gene expression, protein-to-protein interactions, etc, thus allowing the possibility to look at the organism as a whole.

Bioinformatics helps geneticists tackle the important job of correlating genotypes with phenotypes. By using high throughput systems to analyze sets of genetic markers *(e.g. single nucleotide polymorphism or SNP),* researchers have been able to correlate these results with

clinical phenotypes for disease and drug action. The description of biological functions for genes and their products is called *functional genomics*.

The most pressing tasks in genomics involves the analysis of sequence information. Other significant processes include:

- Molecular sequence analysis and data mining aimed at new gene discovery *(finding genes in the DNA sequence of various organisms)*
- Comparative genomics focusing on comparing entire genomes
- Developing methods to predict the structure and/or function of newly discovered proteins and structural RNA sequences
- Clustering protein sequences into families of related sequences and the development of protein models
- Aligning similar proteins and generating phylogenetic trees to examine evolutionary relationships.

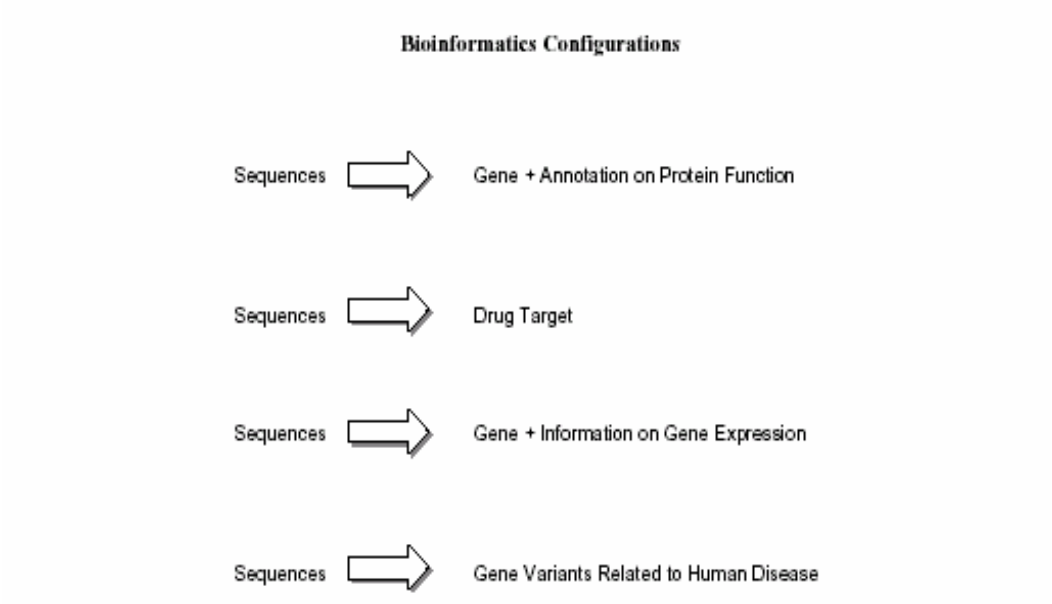

Figure 1. The figure illustrates various strategies that companies use to commercialize bioinformatics.[1]

In order to accomplish these tasks, computational techniques have been developed, which include:

- Development of new algorithms and statistics with which to assess relationships among members of large datasets
- Development of tools that enable efficient access and management of different types of

[1] *(Adapted from Jones P. The commercialization of Bioinformatics, EJB Electronic Journal of Biotechnology, Vol 3, No.2, August 2000)*

databases
- Analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains and protein structures

Products in these areas include:

1. Gene Sequencing instruments, reagents and consumables which are increasingly being used to study a variety of organisms

2. Micro-arrays including biochip systems, 'lab-on-a-chip' devices, and protein chips to study different levels of gene expression under different biological conditions

3. Analysis of single nucleotide polymorphism (SNP) to study individual genetic variation, leading to future possibilities of individualized medicine based on genetic makeup

Functional Genomics and Microarray Technology

The field of biotechnology/chip technology and molecular biology is advancing rapidly. Developments in this field will have a profound influence in the laboratory. One example of such technology is the development of microarray chip or the gene chip. In this system a small volume of sample is enough to analyze hundreds of tests, and several hundred samples can be analyzed in a very short time using a single compact instrument. With the unraveling of the human genome it will be possible to examine DNA for defects that may have adverse effects, using DNA chip technology. This technology can also be used to detect "foreign" *(viral or bacterial)* DNA.

The DNA microarray industry is comprised of companies which supply

- microarray slides;
- microarrayers *(eg. robotic spotters and photolithographic equipment);*
- scanners;
- Software for designing and analyzing microarrays

A competitive segment in the microarray industry is the software products for image analysis of microarrays, microarray database generation and management, array design and polymorphism analysis. Some of the softwares used for microarray analysis are given below,

- **Software for Microarrays:**

  almaZen System assists researchers in organizing, managing and analyzing large number of biological samples involved in a typical DNA array experiment. An important advantage of almaZen is that, it is a completely web-browser based

application, with no software installation required on the client computer. A single almaZen license serves for an unlimited number of users.

ArrayMiner® is a set of analysis tools using advanced algorithms to reveal the true structure of microarray data. Its unique graphical interface provides an intimate understanding of the microarray analysis. The software allows easy publishing of results.

ArrayStat software is a unique statistical package that has been designed for analysis of microarray gene expression data and can answer questions like: Which expression changes are statistically significant? Which spots are poorly reproduced? And, how many replicates are required to reliably detect a desired fold change.

BASE is a comprehensive database server to manage the massive amounts of data generated by microarray analysis. In short, it manages biomaterial information, raw data and images, and provides integrated and "plug-in"-able normalization, data viewing and analysis tools. Additionally, for labs that make their own in-house microarrays or for labs that wish to track probe information, the system also has microarray production LIMS features, which can be integrated with the data analysis. The organization and interface of **BASE** was designed to closely follow the natural workflow of the microarray biologist, and is compatible with most types of microarray platforms and data types *(e.g. cDNA arrays /oligo arrays spotted on any substrate, Affymetrix , CGH on arrays, etc).*

Organizing biological knowledge in databases

Biological raw data are stored in public databanks *(such as Genbank or EMBL for primary DNA sequences).* The data can be submitted and accessed via the World Wide Web. This stored data needs to be accessed in a meaningful way. Often contents of several databanks or databases have to be accessed simultaneously and correlated with each other. Special languages have been developed to facilitate this task *(such as the Sequence Retrieval System (SRS) and the Entrez system).* Different databases also differ in the way data is captured and stored. Interaction between different databases is problematic. An unsolved problem is the optimal design of inter-operating database systems.

Although, the genome data on NCBI and several other databases is freely available, many biological problems necessitate formation of other databases. Researchers, working in a certain area, need to compile their data and organize it into such a form that their colleagues may be able to use it and contribute to it. Apart from genome and protein databases, other examples are taxonomic databases of plant species, plant genome databases, biodiversity databases, clinical data repositories, etc. The department of biotechnology in India, to quote an example, has developed some 100 different databases in different areas of biology.

Pakistan Agriculture Research Council (PARC) is in the process of developing a number of databases including (1) database of medicinal plant of Pakistan that can be searched by botanical, *Tibb*, local and English names. This database is available on net. *(http://www.parc.gov.pk/data/medicinal/medsearch.asp)*, (2) Plant genetic resources, (3) compendium of published agricultural research papers on Pakistan. A complete list and descriptions on PARC databases in attached as an appendix. These and other databases already prepared by Pakistani groups, although simple, can be looked at as a model.

- **Mirror Databases/ Mirror sites:**

  Public domain servers that host major databases, suffer from congestion due to heavy traffic. This impedes effective utilization of these sites. One solution to this problem is establishment of mirror sites, of important and routinely used international databases for genomics and proteomics researches at different locations of the world. These are exact replicas of the original sites and allow for unhindered access to large databases.

  In India protein databases genome databases, EBI databases etc are being hosted by University of Pune, India institute of science *(11SC, Bangalore)*, etc.

Structural Bioinformatics

This branch of bioinformatics deals with predicting and analyzing the spatial structure of proteins and nucleic acid. Structural classification schemes elucidate the relationship between protein folds and function. Very often the primary sequence serves to uniquely specify the 3-D structure of protein. Secondary structure can be deduced from the primary sequence with statistical methods or neural networks.

Remote relationships that cannot be detected by sequence comparisons can be found out using sequence-to-structure fitness approach, whereby the 3D structure is predicted by comparing its sequence to that of other proteins and analyzing their structure.

The main driving force behind structural genomics is the desire to obtain a better understanding of the protein function. Understanding the 3-D structure is crucial for understanding the function of the protein. Similar structure is assumed to imply similar function. General structure-to-function relationships can be drawn through statistical approach.

Structural bioinformatics can also be used to screen ligants to inhibit or increase the function of a protein such an enzyme, thus playing a major role in drug discovery.

Protein-to-protein interaction predicted on the sequence level can be studied in more detail on the structural level.

Drug Discovery

The HGP gives us the enhanced ability to design cures for human diseases. Proteins are involved in disease causation either because they are faulty, hence cause disease by themselves, or are part of the normal cell machinery exploited by a disease-causing organism. In most cases, drugs work by interacting with proteins. For example, a drug might inhibit the effect of a protein essential for the transmission of a virus, or enhance the effect of a protein that helps the cells of the immune system to fight disease.

One of the key steps in drug development process is the identification and selection of the protein that will be the target for the drug. This involves a very thorough molecular understanding of the biochemical processes and molecules involved in the causes and symptoms of the disease. The drug development process is a series of sub-processes including target identification, selection, and validation to be conducted in that order.

Once one or more drug targets are selected, compounds *(that seem to be promising candidates for therapeutic intervention)* are screened to determine which ones best interact with one or more potential targets. Molecules that show a sufficient degree of affinity for the target and are selective in interacting with the target protein are selected. The process of screening often involves retrieving a small amount of material from the compound and testing it against the targets under consideration. The screening test is called an 'assay'; the compounds are called 'lead' compounds and the process of screening is known as 'lead optimization'.

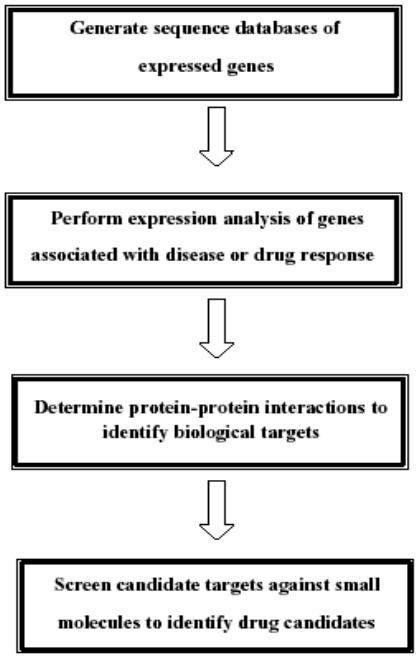

Figure 2: From genes to drugs; the process of drug discovery in the post genomic era[2]

[2] *Adapted from Jones P. The commercialization of Bioinformatics, EJB Electronic Journal of Biotechnology, Vol 3, No.2, August 2000*

Owing to the needs of the drug developers to accelerate identification and validation of potential new drugs, there is a great demand for technologies that allow companies to minimize product failures in trials and move candidates through the development process more rapidly. During the last ten years, two important experimental technologies have transformed the lead identification and optimization process. These technologies extensively use computational techniques and robotics and are known as:

- Combinatorial Chemistry, the automated synthesis of tens of thousands of prospective drug molecules using robotics, e.g. synthesizing all of the molecules that can be made by chemically bonding together chain of reagents.

- High Throughput Screening (HTS). This also uses robotics and is often combined with Combinatorial Chemistry. HTS is the automated testing of each of a number of compounds against one or more targets. Current technology can produce 10,000 to 20,000 assay results per day. Companies such as Aurora Bioscience produce equipment capable of 100,000 assay results per day, a level known as Ultra High Throughput Screening (UHTS). Of course, these data need to be systematically stored, organized for efficient retrieval and analyzed. At these rates, computational support for the HTS process is essential and ranges from data collection and management to sophisticated data mining and decision support. Another application of computational chemistry is virtual high-throughput screening (VHTS), in which libraries of compounds are characterized using computer models that predict, in some way, a measure related to binding affinity to a particular target. These techniques make heavy use of numerically intensive computation and can also be database-intensive.

Drug discovery and preclinical development itself is a burgeoning 20 billion $ area[3].

One important aspect of drug discovery is to screen traditionally used herbs for their medicinal effects. This approach has sufficient scientific credibility. Several drugs in recent past have been derived from traditional medicines, these include gingko *(for memory),* Artemisinin (Qinghaosu) for malaria, etc.

Much modern-day medicine is directly or indirectly derived from plant sources, so it would be unjust to conclude that plants offer no further potential for the treatment or cure of major diseases. Worldwide, the botanical pharmacopoeia contains tens of thousands of plants used for medicinal purposes. Hundreds, perhaps thousands, of definitive texts, monographs, and tomes on herbal remedies exist. But most of this information is outside current databases and remains unavailable to physicians, researchers, and consumers[4].

Bio diverse areas of the world also provide a suitable case for drug discovery research as naturally occurring molecules from plants and animals are extracted and tested for their

[3] Partnering challenges for startups
http://www.nature.com.gate2.inist.fr/nbt/email_response/email.taf?address=french%40entelos.com

[4] Herbal Medicine http://www.naturalhealthvillage.com/reports/rpt2oam/herb.htm

medicinal effects. Previously all of this used to take place in the lab using animal models, cell based assays etc, which is increase highly being shifted to computational studies.

J. Michael French, vice president of business development and alliances, Entelos, Menlo Park, CA, USA believes that in the presence of thousands of companies working in drug discovery and pre clinical trials area, it is a challenge for the pharmaceutical companies to find the right partner. A challenge, as great as, finding drug targets from amongst the glut of data from the genome. Different steps for drug discovery companies include (i) finding the right partner; (ii) determining what to bring to the negotiating table; and (iii) getting to the negotiating table[5].

Pakistan with an old civilization, an ancient medicinal tradition and pockets of rich biodiversity, would be an appropriate place for such research. A look at the projects funded by HEC would reveal that active research is being conducted in this area in Pakistan especially at HEJ institute. Further support might enhance bringing effective drugs to market.

The role of IT in the above mentioned areas would be that of an enabler, providing support for researchers, making their data available to them in a systematic, organized manner. This would also help develop expertise for the IT industry to venture into similar projects for foreign clients.

Proteomics

Proteins owe their function in large part to the special three-dimensional shape or 'fold' they adopt in the cell. The shapes allow proteins to fit favorably together, e.g. in protein-to-protein interaction. In order for us to understand protein function, therefore, we would like to be able to obtain complete information about the relative locations of each of the atoms in a protein molecule.

Because large number of amino acid sequences for proteins will be available as an important product of the various genome projects, and because experiments to determine full protein structures are so difficult to perform, two of the most important scientific issues being addressed are whether we can predict the 3-D structure of a protein from just its amino acid sequence, and what is the process by which proteins adopt the shape they need to perform their functions. These two problems are called the Protein Folding problems.

Proteomic software provides scientists with the ability to conduct database searches of known protein sequences utilizing batch or real-time processing. This software is capable of controlling automated hardware, i.e. robotics, as well as facilitating data transfer operations. Protein databases provide volumes of indexed biological information on proteins from model organisms enabling scientists to annotate and interpret their experimental results. Some

[5] Partnering challenges for startups
http://www.nature.com.gate2.inist.fr/nbt/email_response/email.taf?address=french%40entelos.com

challenges faced when using this technology are reliability, annotation requirements, integration capability, query formulation, cost effectiveness and ease-of-understanding.

The demand for sophisticated tools in the field of proteomics has driven new inventions that address quality, customizability, security, compatibility and accessibility. As the alternatives grow, companies involved with proteomics must understand life scientists' successes and frustrations to improve their technology and make it more powerful.

A survey of more than 450 researchers; who were using protein databases and software in 2001 showed the following opinion.

- Almost three-quarters of protein scientists currently use databases in their protein research and this number will increase by 11% over the next 12 months.

- Study respondents experience the greatest difficulty when attempting to obtain "functional data" and "tertiary structure" information from the protein databases they are using in their research.

- When selecting a particular database for use in their research, most researchers look for "comprehensiveness" and also "accessibility."
- At present, researchers are hampered by "difficult" and "time-consuming" analyses and are "overwhelmed" by the choice of database options.

- "Increased ability to identify remote sequence/structure relationships" is one of the valued improvements to proteomics software programs.

Pharmacogenomics and the Era of Personalized Medicine

Pharmacogenetics draws on the study of human variations *(predominantly SNPs).* This information is highly valuable from a commercial point of view, since studying these differences may be important to reduce drug side effects. Also many potential drugs/chemical NCEs *(Novel Chemical Entities)* fail because of inherent differences in drug metabolism and effects in a genetically heterogeneous patient population.

Pharmacogenetics has started to yield results already. NitroMed *(Lexinton, MA),* a small pharmaceutical company has marketed a drug called BiDil, which specifically targets heart failure in black population[6]. Keeping aside the moral and political implications, this proof of concept example shows that the era of personalized medicine is about to dawn soon. The integration of Medical and Bioinformatics would be instrumental to the development of personalized medicine.

[6] Race and Medicine, Technology Review, April 2005, Rotman D, http://www.technologyreview.com/articles/05/04/issue/feature_medicine.asp?p=1 as accessed on May 5, 2005

On one hand, pharmacogenetics threatens to reduce the market of a drug by reducing the target population, while on the other hand pharmacogenetics will increase drug revenues by rescuing several drugs from clinical trials, many of these drugs have been rendered useless since they were being administered to a genetically heterogeneous population.

The prospects of personalized medicine will tremendously increase research in pharmaceutical industry. At the same time, it will require physicians to obtain genetic data from their patients. Interpreting patient's genetic data and to prescribe a certain drug, will require extensive IT support.

7 Fields at the Periphery of Bioinformatics

Certain areas of biological sciences are with in the broad area of Bio-IT but not included in the definition of bioinformatics in the strict sense of the word. These include,

1. Laboratory Information Management Systems (LIMS)
2. Medical Informatics
3. Clinical Trials Management

Laboratory Information Management Systems (LIMS)

As a part of its routine tasks, a laboratory must deliver accurate, understandable results to the originator of the request for analysis, within a suitable timescale. Such an operation necessitates transferring a sample to the laboratory, analyzing the sample, checking the results, *(and if necessary re-analyzing the sample),* and issuing a report to the requester.

With no computational support to the entire scheme, there is a large scale reliability on clerical handling of the results of the analysis. Unfortunately, manual reporting systems are neither accurate nor timely. They depend on multiple transcription of results, and they are slow - studies in the author's lab showed that clerical work in typing and issuing reports took about as long as analysis. Further, manual reporting is cumbersome and labour-intensive, data retrieval becomes a hunt across multiple locations for a single piece of paper, and retrospective data analysis is virtually impossible.

The essential concept of a basic LIMS, therefore, is that of a computer system which would automate the clerical activities associated with the processing of the analytical results, improving accuracy and turnaround times to an acceptable level. LIMS is a technique independent of discipline, and has applications in any industry where laboratory analysis is important, from Healthcare through the Food & Drink industry to Pharmaceuticals and Petrochemicals. A more complete list of the attributes expected of a LIMS is shown below:

- To shorten turnaround time of lab tests
- To improve access to the results database
- To improve accuracy of analysis, by eliminating transcription steps
- To count and monitor resource utilization
- To exchange data and information both with analytical equipment and corporate mainframes
- To improve productivity

A LIMS will reduce the amount of time to manually track down samples and paper by placing that information in a centralized database that the whole laboratory accesses. A LIMS will reduce the amount of time required to enter results for samples by electronically transferring the results from instruments directly into the LIMS. Not only does this reduce entry time, it also reduces entry mistakes. Finally, the LIMS will automatically print certificates of analysis and invoices. Because the results can be automatically reported after they are electronically entered, validated and approved, the reporting time is dramatically reduced. The key to this automation is a good LIMS that is flexible, and above all, easy to use.

LIMS is an important segment of the market. A report by Silico Research states that the two major bottlenecks in clinical trial process are the recruitment of investigators and patients, and the second is the transfer of data from the investigator into the clinical data system. Over the next four years the author expects trial management tools and the electronic data capture to converge with the clinical data management and data infrastructure tools that form the basis of the market today.

The introduction of LIMS has antiquated the use of Laboratory notebooks and handwritten reports/charts that were used to track and report information. Today's advanced Information Management Systems specifically designed for the labs have a decentralized architecture, thus processing can be performed anywhere on the network. Thus, all clients and servers can operate in either capacity depending upon the data load at any particular instance

Laboratory Automation

Laboratory Automation and Robotics involves sample handling, robotic workstation and software. Much of the equipment is designed for pharmaceutical companies for High Throughput Screening. Automation in clinical laboratories is not new. There has been significant investment in automating diagnostic assays at large clinical diagnostic libraries. The larger labs are seeing Total Laboratory Automation (TLA) utilizing Robotics.

The introduction of tracking systems linked to dedicated units allow scientists to centrifuge, sort, and deliver samples to different analyzers to develop a continuous flow system. This form of automation is sometimes referred to as "microautomation". The next stage in automation is the automation of specimen processing and transport of specimens between analyzer or workstations. This is called "macroautomation" or robotics, and typically consists of an inlet unit, sorter, transport system/tracks, automated centrifuge, level detector, bar code reader, decapper, aliquotter, recapper, and so on. Following analysis, samples can be automatically stored in a refrigerated unit, from which specimens can be automatically recalled for further analyses.

There are many possible reasons to implement a robotic system in a laboratory.

Economic reasons: The cost of health care is rising universally owing to increased expectations, new technology, and the advancing age of the population. In many industrialized countries, hospitals have been closed or merged with neighboring hospitals in

order to reduce cost and to improve efficiency. Many hospital laboratories are also closed and replaced by centralized laboratory services. Macroautomation/ robotics is a possible way to improve efficiency and to cope with the high workload in these centralized laboratories.

To reduce errors: Errors in laboratories are a continuous source of problems and they are mainly due to human factors. It is estimated that the error rate in clinical laboratories may be 12%, although many errors go unnoticed. Some of these errors have clinical consequences. There are no data on the cost of these laboratory errors. One estimate puts this at 10% of all errors in hospital care, equivalent to more that $1.5 billion annually in the USA. Most of these errors are at the pre-analytical stages of the sample processing. An important incentive for implementation of total laboratory automation is the opportunity it provides to reduce human errors which occur during labeling of specimen, aliquoting, order entry, and so on.

Improved quality: One factor determining the quality of a laboratory service is the timeliness of results. If the results can be produced at the right time and place, it will have a significant implication on health care cost. With increasing pressure on bed occupancy and attempts to reduce inpatient care, earlier availability of results is now expected. Macroautomation/robotics has the potential to significantly reduce the turnaround time of tests. It has been suggested that the introduction of robotics will improve the overall turnaround time such that an emergency laboratory becomes redundant.

Technology has advanced to such an extent that it is now possible to do many tests at the bedside. These near patient testing/point of care testing devices are disposable devices, hand held devices, or bench top analyzers, and they use samples such as whole blood, urine, or saliva. The operating system of the devices is such that they can be operated successfully by non-laboratory staff.

In an effort to reduce the overall laboratory costs and improve laboratory efficiencies at all of network hospitals, many laboratories have turned to automation. When properly implemented, automation systems can reduce overall laboratory expenses, enhance patient services, and address the overall concerns facing the laboratory today: job satisfaction, decreased length of stay, and safety. The financial savings realized are primarily a result of labor reductions.

Emerging areas of interest in this industry include microchip-based detection systems and the efforts associated with genomics and proteomics. In the past few years many companies have emerged in this field committed to providing computing infrastructural consulting services in the areas of secure internetworking, seamless data storage and retrieval and parallel computing to the industry and academia.

Problems of high cost and building issues may make TLA impractical initially, and laboratories may choose to introduce partial automation as a first step to total automation. Automation may be conveniently divided into preanalytical, analytical, and postanalytical. Manufacturers of TLA are able to sell partial systems, so that TLA can be introduced in a

stepwise manner. Although this may allow some automation in the face of limited funds, it may not be economic with regard to phased alterations to the laboratory.

Medical Informatics and Bioinformatics

- **The Gap between Medical Informatics and Bioinformatics Is Bridging**

  Traditionally, medical informatics has dealt with managing patient records and hospital data. However, in the post genomic era, the difference between Bioinformatics and Medical Informatics is rapidly decreasing. IBM Life Sciences *(Rochester, MN)* have designed an electronic patient records system for Mayo Clinic Rochester, which integrates clinical symptoms with genetic data from microarrays. This will allow physicians to choose medication for a patient depending upon the genetic profile of the patient. Information about disease may it be from patients or from labs is likely to converge in coming years, making the distinction between medical and bioinformatics irrelevant. Companies venturing into this field would need competence in both of these areas[1]. Other groups working to integrate genetic and clinical data are Duke University School of Medicine *(Durham, NC)*, Hadassah Hospital *(Jerusalem, Israel)*, iCapture Research Centre, University of British Columbia *(Vancouver, British Columbia)* and Kobe General Hospital *(Kobe, Japan).*

  The information revolution will enable the use of molecular biology in medicine in two ways. First, rapid diagnostics will likely be done using computers and information technologies. Computers will be used for complex diagnoses, and the Internet will be used to conduct remote consultations between physicians. Second, information about the patient's full medical history will be instantly available through a storage system, either on a chip embedded in the patient's body or accessible from a central database. These advancements in recording, manipulating, and transmitting biomedical data—made possible by the Information Revolution—will offer improvements in health care over the next 20 years in ways that are only vaguely dreamed of now.

- **Niche market for patient care medical informatics products**

  Innovative gadgets helpful in home management of chronic illnesses are in demand in the developed world. Vikram Kumar of Brighams and Women Hospital, founded a company, Dimagi *(Boston, MA)* while he was a medical student. His company specializes in making PDA based computer programs that help patients with managing their conditions. Encouraging people to adhere to treatment regimens is one of the biggest challenges in medicine today. With the increase in chronic disease it would be essential to treat disease at home, get blood test done which could be periodically transferred to the doctor's office. Similarly computer software could help modify drug

[1] Diagnosing with Data, Technology Review, December 2003, Huang G, http://www.technologyreview.com/articles/03/12/innovation11203.asp?p=1 as assessed on May 5, 2005

regimen, suggest appropriate diet or exercise according to the patient's daily blood tests e.g. in the case of diabetes[2].

Hospital based information systems are being integrated with genetic expression data from micro arrays. Several information streams are thus integrating to the common goal of better patient care.

- **Data Mining in Patient Data**

Several companies like, Cereplex *(Gaitherburg, MD),* Theradoc *(Salt Lake City, UT),* Vecna Technologies *(College Park, MD),* Med Mined *(Birmingham, AL)* specialize in data analysis services to track hospital infections. These companies mine patient data for trends in infections, suggest courses of action for particular patients. This is likely to decrease the incidence of hospital acquired infections[3].

- **Opportunities in Paperless medicine**

Adverse drug effects are a major problem in modern medicine. Of the more than 700,000 adverse drug events estimated to occur each year in U.S. hospitals, nearly 28% are attributed to a preventable medication error, with most occurring during drug ordering[4]. Others occur at the time of pharmacist processing, drug administration, and monitoring. A study by the Institute of Medicine, USA stated in 1999 that 7,400 people died in a year from receiving a wrong medicine, many of these due to receiving a wrong prescription. One basic change-using computers to order prescriptions-has reduced medication errors by as much as 80 percent in some hospitals. Motivated by these statistics, a growing number of doctors and hospitals are weaning themselves from paper, using computers not just to order prescriptions and lab tests but also to track patients' conditions, medications, allergies, and test results.

Typically in the US, the primary care physician keeps one set of records, hospitals another, and each specialist yet another. And all of these medical histories are logged in old-fashioned paper "charts." Any information a patient forgets to tell one of his or her doctors-about a severe allergy to a medication, say,-simply doesn't appear in that doctor's record. The largest HMO in the US, Kaiser Permanente has recently shifted to electronic medical records. Many more in the US and around the world, may be planning to computerize their records in the near future, representing a major opportunity for IT industry[5].

[2] Biotech and Medicine, Technology Review, October 2004, Jonietz E, P 68-74.
[3] Fighting Infections with Data, Technology Review, October 2004, Lok C, P 24

[4] Bates DW, Cullen DJ, Laird N et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA. 1995; 274:29–34.

[5] Paperless Medicine, Technology Review, April 2003, Jonietz E, P 58-65

- **An important business for technologically competent developing countries.**

  Galaxo Smith Kline (GSK) CEO J.P. Garnier has announced that GSK will shift 30% of its clinical trials to countries like India and China in order to slash its clinical research budgets and stretch its current spending to more and larger trials.

  India aims to achieve 20% of global clinical trials by 2010 which accounts to 1.5-2 Billion $.

  An interesting example of a company working in Clinical Trials Business is Quintiles India.

- **Case Study: Quintiles India**

  Key Projects

  The company offers clinical trials, data management services and ECG Services. Quintiles have worldwide participating groups especially in India and South Africa.

  Over a hundred different projects have been conducted by the Quintiles India so far. A list of these is reproduced below.

Therapeutic Area	Total Studies	Total Sites	Total Patients
Oncology	20	103	1,691
Psychiatry	26	188	3,644
Neurology	14	113	2,418
Anti Infective	15	118	1,779
Endocrinology	12	110	2,068
Gastroenterology	4	45	360
Ophthalmology	4	25	755
Cardiology	10	108	3,453
Others	10	54	1,772
Total	**115**	**864**	**17,940**

- **Advantages to developing countries**

  The ability to recruit patients with desired criteria / disease rapidly and in appropriate numbers, owing to a huge disease burden, offers great advantage to a company such as Quintiles to operate in developing countries. Other important advantages are,

  → With modest excess to health care patients are much more willing to participate.

  → Other benefits are low investigator fees and low infrastructure cost.

  → Quintiles India boosts of US and UK trained English speaking, low cost investigators, a growing, local pharmaceutical, favorable government policies and collaborating teaching and tertiary care hospitals.

→ Large patient populations with broad range of disease profiles, faster study start up, quicker recruitment.

→ Cheaper trials than US with no compromise of quality.

→ Double disease burden *(diseases of tropics as well as developed countries)*

- **Workforce Requirement**

  A company specializing in clinical trials would require data handling system, IT experts, Physicians, Collaborating health care facilities and Management team.

  A summary of the services is given below.

  → Project Management
  → Clinical Trial Monitoring
  → Regulatory Affairs
  → Drug Safety
  → Protocol Development
  → Site Management
  → Quality Assurance
  → Clinical Trial Supplies
  → Data Management
  → Training
  → Innovex *(Commercialization)*
  → ECG *(Centralized ECG)*

- **Quantities Biostatistics Services:**

  Another service offered by quintiles is biostatistics. This wing claims to have epidemiologists, biostatisticians, statistical programmers and support staff, and helps to develop innovative study design, clinical development programs, analysis methods, data displays and interpretations. The biostatisticians become involved with projects at the earliest protocol development stage and advice on the design, size and conduct of clinical studies.

  Once the data is gathered and stored, Quintiles, biostatisticians put their experience to work analyzing and interpreting the results.

- **Difficulties faced by India in Clinical Trials Business:**

  1. Regulatory uncertainties.
  2. Involvement of Multiple agencies for approval of biotech products.
  3. Lack of extensive documentation of patient records.
  4. Bad Media publicity.
  5. Low level of patient literacy.

Knowledge of Good clinical practices and establishment of Ethics committees in health care facilities will be an important stepping stone for this business. Medical practitioners will need to be aware of regulations, ethics, Good Clinical Practices, and skills for clinical trail management.

- **Is Clinical Trials Business Possible in Pakistan?**

  Most of the advantages outlined above are not specific only to India. It is likely that similar companies can be setup in Pakistan. There might be certain areas in the field of clinical trials, where Pakistan would be at an advantage to provide such services.

  In 2000, worldwide pharmaceutical market totaled to 317 billion according to IMS health and it could reach 3 trillion $ by 2020. This growth is however possible only if the industry adds more drug targets to its current list of around 500 targets. Hence there are substantial investments in development of new drugs and in clinical trials. Clinical Trials is an important area that should not be neglected by Pakistan.

  An estimated 10% of research and development expenditure is currently spent on IT related solutions; this share is estimated to grow in the immediate future.

8 Business Models in Bioinformatics Industry

Outsourcing

- **Why would organizations want to outsource?**

  CanBiotech's[1] report on Bioinformatics states that the business models encompassing drug discovery & development are predominantly based on consolidation into the Big Pharma companies. Nevertheless, the pharmaceutical industry comprises of a large number of small medium & large companies which vary in their informatics needs as well as available resources. These firms would, inevitably, attempt to survive the sheer competition by outsourcing their needs to small & medium source providers, thus optimizing on their outsourcing partner's experiences, flexibility, standardization & componentization.

  With informatics providing enormous computing powers to the Bioscience processes, biosciences firms are readily allowing the expansion in their IT infrastructure for R & D operations. The IT industry seems willing to provide just about everything from servers to storage, database technology, visualization equipment, and a range of software programs for complex scientific tasks. But small to medium pharma and biotech companies will focus on those technologies that help them gain process efficiencies and increased throughput. The long term goal of their companies to cut down costs and bring more drugs to market at a faster pace has added to the dependency on informatics-based solutions.

  The following factors entice Pharmaceutical companies to outsource their internal functions to IT companies, as outlined by CanBiotech specialists:

  1. **Focus on Core Strength** - The opportunity cost of adding a new segment of services to an existing business is high. Companies prefer to not diverge from their focus on their core competencies, and stay competitive.

  2. **Improve Operational Efficacy and Stay Competitive** – the outsourcing partner can be relied upon to choose the best technology for automation

[1] CanBiotech, Outsourcing Research Informatics, BioMed Outsourcing Report, 2002, Vol 1, Issue 2.

3. **Leverage Top Talents from the Bioinformatics Companies –** companies can make use of the best resources engaged in IT services without having to worry about recruiting, training and retaining professionals

4. **Minimize on Total Cost of Ownership –** the rate of obsolescence is very high in the IT sector. Companies who outsource do not have to bear the replacement costs that accompany obsolescence.

5. **Improve Service Delivery to Customers** – the Company can dictate to the outsourcing partner the timelines it is required to meet.

6. **Reduce Cost and Time –** the time spent otherwise on planning and development is saved and better utilized to managing the existing resources.

7. **Expand Service Portfolio –** companies can extend their services without having to invest on building in house capabilities

Proposed Model for Informatics-enabled solutions

CanBiotech argues that with the dynamic informatics market where technologies improve and advance at a rapid pace, organizations should opt for a modular approach toward informatics-enabled solutions. The modular architecture allows the organizations to deal with the various processes in an ad-hoc way, making use of technologies that best-suit the problem in hand & later integrate these applications into its overall informatics architecture.

This incremental, ad-hoc developing of processes would ensure that organizations do not get stymied or 'locked up' into certain technologies. The model is based on the assumption that the modules or sub-systems can be integrated or linked together to allow for a smooth, seamless flow of information from one process to another.

The following methods elaborate how the various components could be linked together to bring together an informatics architecture that provides solution to an entire chain of processes

The following are examples of components that an organization may want to buy off the shelf or outsource for custom development:

- Components that allow User Interface Controls
- Analysis Packages e.g. the Statistical Analysis
- Components that implement Communication Protocols
- Visualization Components
- Algorithm Libraries

- Components for importing data
- Workflow Management
- Components that join Databases

The benefits of using a component architecture are manifold:

1. There is a clear cost advantage, as focus on core competencies can be achieved, outsourcing or buying off-the-shelf components that are required, but are not a part of the companies core competency
2. The choice of application is done on an ad-hoc basis which ensures that the application best-suited for the module is used

3. Since the entire system is built up of modules which have been independently developed, it is easy to switch pieces of architecture without having to replace the entire infrastructure.

The complete informatics architecture must posses the following qualities:

- Data Integration
- Data Access/ Data Sharing
- Data Visualization
- Usability
- Extensibility/Scalability
- Deployability
- Reliability
- Security

America still outspends all other nations when it comes to R&D. Almost 300 billion Dollars were spend in 2003 according to National Science Foundation. However there are signs that more ad more R and D is being shifted to other countries.

Many American companies like Microsoft and GSK, have opened research wings in countries like China and India.

Badruddin Syed, executive vice president at "outsourcing Firm vMoksha Technologies says "Indian companies, especially the big ones, are waking up to the patents philosophy, as they find that patents are highly profitable. The Indian software industry feels that it has to position itself as creators of high value solutions, rather than just of low cost offerings". Outsourcing is likely to happen not only in imitative, labor-intensive work but also in innovation, research and development.[2]

[2] **Outsourcing:** The next technology battlefields in Biospectrum, India

Some of the obstacles that the US is facing are:

Slower growth in the number of possible workers to recruit and train: Rand researchers predict a growth of 1.1% between 2000 and 2010 in the work force meaning the wages of qualified engineers and other technology workers in America will increase, this will further aggravate the problem leading to more outsourcing to offshore research and development centers.

"Engineers with Ph.Ds and recent college graduates alike are hearing that they are too expensive, that their job can be done more cheaply abroad" says Paul Almedia, President of the Department of Professional Employees at the Union Movement AFL-CIO.

Reverse Brain Drain

As various countries explore outsource business in high tech fields many scientists, engineers etc are returning to their home counties to explore opportunities there.

This compounded with other political issues represent a major opportunity to develop and invest in high technology and intellectual property arena in Pakistan.

According to CNET News com a survey findings shows that over 60% of IT decision makers think export of IT jobs threaten the long term technology leadership of the US.

In spite of the fact that some companies are establishing R & D wings in developing countries; as a general trend only labor intensive jobs are being moved to these places. As more low-end jobs are shifted abroad, the US will focus more on high end solutions and breakthrough research.

Countries like India enjoy a large expatriate population working in various technology areas. It is estimated that 50 of the 500 companies located in the Boston area are headed by Indian-Americans. *(India Today May 2002)* These could serve as an additional link in terms of outsourcing.

Essential Components of a Bioinformatics Company

Essential components of a bioinformatics setup would be[3]:

Scientific credibility: Does the approach work? Proof is required that the overall process *(e.g. drug discovery and development process)* makes a difference, reduces cost and time.

Revenue: to support the company's growth

The right client: It is important to tailor the technology to the client's needs. So in the question, what comes first; the technology or the partner? The answer has to be the partner.

[3] French JM, Nature 2002, Vol 20, Supplement p BE 40-BE 42.

Possible Partnerships

The field of bioinformatics is multidisciplinary. For the IT companies in Pakistan willing to venture into the field of Bioinformatics, developing partnerships is essential. This would give them the knowledge base and the direction in which development must take place. The collaboration could take one of the following forms:

1. Collaboration with the Bioinformatics Supplier

At the initial stage, it may prove profitable for the newly established Bioinformatics companies to indulge into partnerships with their counterparts in the developed countries which are reputable and well-established. This is expected to be useful in the beginning as local firms attempt to gain credibility and find a channel to introduce themselves in the international market. The local firms can acquire projects or project modules from foreign bioinformatics companies on contractual basis, and thus attain projects which are manageable in size and allow the firms to focus in the direction in which it has greater expertise. In the words of Mathukumalli Vidyasagar, vice president of service bioinformatics service company Tata Consulting Services in Hyderabad:

"You have to walk before you run. You have to do some contract research before being able to come up with products."

Thus it seems that a contractual arrangement with the already existing bioinformatics suppliers could be crucial first step for the burgeoning bioinformatics companies.

2. Collaboration with Life Science Instrumentation Supplier

Companies that traditionally supply instruments and other equipment to life science laboratories have the knowledge and access to the market; whereas the bioinformatics suppliers have the latest technology. By consolidating knowledge with technology, bioinformatics can hope to revolutionize the way experiments are done today. Collaboration between the two would result in an automation in the traditional 'wet-lab'. IT companies could find a niche in embedded system technology by designing software that is used to configure the instrument hardware. Managers at Active Motif, a firm producing reagents and assays that are used in cell biology, observe that by getting involved earlier in the research process, it will gain valuable insights into what products its customers want later on when they try producing real results and go into production with new processes.

3. Collaboration with Drug Discovery Pharmaceutical Companies

Bioinformatics is seen as primarily applied to speeding up new drug discovery. The IT-Pharma partnership is aimed at achieving just that. The enormous growth of biological data has led to a rising demand in the following areas:

i. Storage and Maintenance of databases dedicated to genomic data
ii. New Methods for analyzing these huge databases
iii. Integration of disparate databases

The partnership between IT and Pharma is an essential one; whereas Pharma companies need technology to gain accelerated processing power in drug discovery, IT companies need to enhance the credibility of their products by endorsing software output with the results obtained from Clinical trials. As Bart Challis of Active Motif puts it:

"We don't think that bioinformatics will ever replace wet-lab experiments, because scientists will always want to validate their work in silico with actual cell assays".

Thus, with the Clinical trials and wet-lab experiments retaining their significance as the most authentic source of biological knowledge, the confluence of IT and Pharmaceutical companies would position both industries to sustained high growth.

Partnership Challenges

The Pharmaceutical industry anticipates in silico approach towards drug discovery can bring a new era of enhanced drug development. They recognize the role of outsourcing in achieving this goal. However, they are also becoming increasingly cautious and selective about who to take into partnership. Recently a report of Indian Bioinformatics trends highlighted that the Hyderabad based I-Lab[4] was undergoing closure for not being able to capture the overseas market despite having developed a unique software product.

The case depicts the cut-throat competition underlying the bioscience industry that is driving biotech companies to strive for innovation rather than opt for providing 'standardized' products. Dr. Vidyasagar, executive vice president, TCS, observed, "In contrast with all the hype, bioinformatics is a niche area and will remain so in the foreseeable future."

It seems that organizations vying to occupy a portion of the global bioinformatics market shall have to work hard to formulate their business model, identify a niche for providing their products or services and consistently focus on this niche in order to be successful players in the industry. According to Dr. Vidyasagar,

"Bioinformatics as a science is still evolving. The underlying mathematical and algorithmic methods are still being developed. In this rapidly changing scenario, a company that aims to provide 'standardized' services will definitely fail. Only those companies that are contributing to the evolution of the subject itself, and generating their own intellectual property will survive in the long run."

For the informatics entrepreneurs, the important thing to keep in mind is that the bioscience industry is R & D intensive, with the domain knowledge expertise required to fully

[4] Has the bioinformatics dream soured?, India No. 1 IT Business Weekly, 2004, available at: http://www.expresscomputeronline.com/20040322/indiatrends01.shtml

understand the problems and challenges posed by this field. While the local bioinformatics companies may have the expertise and skills in IT, they lack the biological expertise and domain knowledge that comes from extensive research and development.

Due to the high confidentiality of information in this field, Pharmaceuticals would not want to outsource to countries where Intellectual Property Rights are not properly enforced[5]. In such circumstances, the Big Pharma would be tempted to hire talents from the under developed countries, instead of outsourcing.

While India has familiarized itself in the global market as an extremely reliable and profitable partner in IT outsourcing, Pakistan's IT industry has yet to see a boom. Multinationals who consider outsourcing to developing countries would be more readily outsource to the top tier companies in India, who have a good track record, than take chances in Pakistan.

In comparison to India *(which is the major competitor of Pakistan in outsourcing)*, the size of the current Biotechnology industry in Pakistan is very limited. The indigenous pharmaceutical industry of India is far more developed, and boasts of presence of the multinational companies. In addition, many pharmaceutical companies in India have been manufacturing generic medicines for a long time. Recently the government has signed many intellectual property rights agreements, threatening to put generic pharmaceuticals out of business. Many of these pharmaceuticals are turning into contract research companies specializing in drug discovery and manufacturing.

Strengths of Developing Countries

While these genome initiatives and the potential for developing new drugs and vaccines are admirable, it is notable that scientists from developing and endemic countries have had very little participation in the process of developing genome resources that are important in their fight against disease. The capacity exists within disease endemic countries to contribute to annotation and development of the analyses of the parasite and vector genomes[6].

In many instances, populations of developing countries have specific advantages in carrying out genetic research, particularly for common disorders including cardio-vascular, cancer and mental illness, that are leading causes of morbidity or even mortality world wide. In many developing countries, marriages between close relations is commonplace, the incidence of genetically-influenced diseases is thus high. Often, these families are large and live together, experiencing the same environment. These provide ideal conditions for scientists to study and analyze disease causes amongst related individuals brought up in the same environment[7]. The new tools of genomic analysis would help geneticists gain insight into many of these diseases by careful analysis of these special populations.

---

[5] French, Michael, *Partnering Challenges for Startups*, Vol 20, 2002

[6] Davila, Alberto M. R., Phelix A. O. Majiwa, *Response to Hertz-Fowler and Berriman: Continuing tsetse and Trypanosoma genome sequencing projects*, Trends in Parasitology

[7] Genomics – Your Guide to Genomics, Bioinformatics & Proteomics, 2005, available at http://www.separationsnow.com/basehtml/SepH/1,1353,6-1-1-0-0-news_detail-7321169111-270,00.html

It should be of significant interest to health authorities in these countries to know the causes of these diseases: in the short term, to be able to offer genetic counseling and pre-natal diagnosis to families, and in the longer term to find ways to ameliorate, cure, or eliminate these diseases. Moreover, these genetic "resources" have the potential to make major contributions to the study of diseases that are of worldwide significance, underlining the importance of carrying out research in these regions of the world.

Weaknesses of Developing Countries

Almost all developing countries lack in their industrial capabilities in modern biology. With their inherent lack of interest in technological development, the developing countries have lagged far behind in research and development of 'original' products. The bioscience field continues to be an R & D intensive industry, therefore, countries lacking in this area would naturally be at a disadvantage.

There has been a low rate of absorption[8] of new technology from developed countries largely because of the scarcity of financial resources and the absence of a crucial network.

The infrastructure required for development and research is limited.

Developing countries lack the human resource specialized in the field of bioscience. The very few people who have the expertise would be quickly offered better jobs and incentives *(in the form of remuneration)* in other countries.

The potential profits have driven firms from all over the world to enter this market, thus making it extremely competitive. At a national level, India, Israel and Spain are developing to become bioinformatics outsource hubs.

While some of the IT industry's capabilities can be transferred to a low-end bioinformatics business, deep knowledge and understanding of the bioscience processes is required to move up the value chain.

[8] Madhok, Anoop, Thomas Osegowitsch, The International Biotechnology Industry: A Dynamic Capabilities Perspective, Journal of International Business Studies, Vol 31, No. 2, 325-335

9 Building Bioinformatics Capacity

The research and development capabilities in developing countries are limited. However, the prospects of bioinformatics may be better, given that knowledge of database management and statistics are important in this business[1].

The second source of difficulty is a problem of building and maintaining capacity. There are few university-level bioinformatics curricula worldwide, and very few in developing countries. Moreover, graduates are often immediately absorbed by the prestigious universities, or pharmaceutical or agri-biotech multinationals. Furthermore, instead of maintaining bioinformatics research groups, many universities choose to support bioinformatics teaching in the general framework of biological curricula, using knowledgeable user-level teachers involved in other areas of biological research at the same university.

This necessitates the need for the government to invest in programs specifically addressing the bioinformatics needs of a developing country like Pakistan. Success lies in identifying the strengths of the local industry. The government needs to recognize the importance of bioscience as a multi-disciplinary area, and initiate a training program to encourage growth in the industry. Teaching programs geared towards training individuals at the graduate/undergraduate level and arousing interest in the subject are imperative. For this purpose the government would need to set up university faculties, and devise curriculum especially designed to attract individuals to this emerging industry.

The government also needs to provide a network to the existing scientists, academic institutions and research organizations operating in this important area to communicate and form linkages with each other. Effective communication and collaboration amongst all sectors is crucial for the success of bioinformatics and would push technology transfers from research to industry.

In order to convert the market potential into a profitable industry, what is required is a true understanding of the drug discovery process and the role that computational technology can play. There is a dire need to emphasize on R&D in this area. Research grants should be offered to scientists to start the project of developing gene sequences of the local population. Based on these genetic resources, development of indigenous vaccines would become a much achievable task.

[1] Pongor, S., D. Landsman, Bioinformatics and the Developing World, Biotechnology & Development Monitor, 1999, No. 40, p 10-13

One of the reasons for the lack of research initiatives is the low market demand of bioscience products in the local market. By allowing funding program to the start up companies to encourage bioinformatics entrepreneurs, the government could expand the market as well as the demand of the local bioscience industry. This would also encourage investments from the various foreign pharmaceutical companies, revive the demand conditions and attract capital resources to this industry.

While dissemination of bioscience education will reap profits in the long run, the local IT industry could play an active role in the near future by developing customized software for the global bioscience industry. Private companies could utilize their expertise in IT and Statistics to develop tools that help pharmaceuticals lower the time required to bring a new drug to market.

Building Bioinformatics Infrastructure

For bioscience research and tool development to ensue, the infrastructure has to be in place. An effective support system for Bioinformatics includes Information Technology support for Hardware, network, system administration, data tools & methodologies, academic graduate education & faculty development.

The government should build technology parks, wet-labs, provide computational resources like high-end computers,

Enforcement of Intellectual Property Rights is an essential to obtaining business from foreign companies who want to make use of the cost arbitrage.

Trends for the Future

Biotechnology is used across a wide range of applications including health care, agriculture, environmental protection, industrial applications and crime investigation. However, globally the focus of biotechnology is on health care. The industry has continued to be a highly R & D intensive with R & D expenses exceeding 50%[2] of revenues. However, there are clear signs that the industry is becoming more commercial.

Although an increasing number of potential products are reaching the final stages of clinical validation, developing biotech drugs remains an expensive proposition with the cost associated with bringing drug to market estimated to $500 million.

The sequencing of the human genome has revolutionized the way Biotechnology can be used to develop health care applications. In the long run, this is going to result in ‘individualized medicine’- where medicines would be tailor-made to suit a person's genotype. In the short term, genome sequencing may not effectively reduce he time required to develop and bring a drug to market. Drug development, therefore would continue to be R & D intensive. The success in human genome sequencing has opened up new horizons to the way information

[2] Krishnan, Rishikesha T., Anushu Gupta, Varun Matta, *Biotechnology & Bioinformatics: Can India Emulate the Software Success Story*, (Banglore) 2003

can be analyzed and utilized for developing new products of industrial significance. The demand for individuals with skill sets in information science and biology is thus on the rise.

While pure play biotech companies have carried out most of the research in the field; pharmaceutical companies continue to be of value because of their role in the commercialization of drugs. More than half of the biotech drugs in 2000 were co developed or marketed by pharmaceutical companies. The alliance between the pharmaceuticals companies and biotech companies is an important one – while biotech companies depend on the pharma companies to finance and commercialize their products, the latter seek advanced technology and tools from the former. The result is a profitable partnership between the two geared to facilitate technological development.

There has been large scale diffusion of the technology to a variety of countries around the globe. Although the core work of the human genome sequencing project was done in the U.S. and the U.K., other countries like France, Germany, Japan and China were also part of the project. The internet has facilitated access to and participation in the on going research, however, the U.S. clearly remains the locus of all bioscience-related activities because of the advantages it has in the availability of finance, the size of the market, government funding for research projects, and intellectual property rights regime.

In lieu of the rising demand conditions, many companies have attempted to provide their expertise as content supplier and IT tools provider. Large, multinational IT firms like Sun Microsystems, IBM, Oracle have invested in research to become bioinformatics solution providers. In future, there is an opportunity for these firms collaborating with other genomics companies to take a large chunk of the profits that are anticipated from advancement in this industry. Emerging companies, eager to provide their expertise in this field would have to strive to entice these companies by their distinctive new tools which are innovative, yet cost effective. Due to the emphasis on protection of intellectual property, these pharma companies would less likely be willing to outsource it to countries where the Intellectual Property Rights are not observed.

Bioinformatics cannot be disregarded by any country intending to remain up-to-date in the biomedical, biotechnological and agricultural sectors. In addition to this general trend, developing countries would also want to manage their own specific data on indigenous biological species and biodiversity programs. These tasks would necessitate that statisticians and informatics experts become advanced users of bioinformatics software and develop a capability to solve problems locally. The developing countries would also want to benefit from the cost arbitrage and provide business process outsourcing in order to grab their share of the profits. Since their R & D capabilities are limited, the coming years are thus likely to see a mushrooming of bioscience firms in developing countries, all offering low-cost, high quality IT solutions.

10 Bioinformatics Business in India and its Comparison with Pakistan

India has emerged as a major center for outsource in IT related businesses. It also supports a large generics pharmaceutical industry. The government aims to combine its strengths in IT and biotechnology to emerge as a major global hub in bioinformatics. There are already a large number of companies working in this field in India. In 2004, bioinformatics contributed $17 million to the economy, predicted to increase many folds this year. There are many factors responsible for the success of bioinformatics in India.

- **India's Edge in Bioinformatics**

  Following are the main reasons for India's edge in bioinformatics among many others.

  → Rich biodiversity and natural resources

  → Vibrant life sciences industry base

  → Large pool of skilled manpower

  → Research base of around 200 established research institutes

  → Availability of biotech parks, industry friendly state biotechnology policies and government's priority on bioinformatics.

  → 2nd largest English speaking scientific workforce in the world.

Government Policies on Bioinformatics

In recognition of its importance, the Department of Biotechnology, Government of India has identified bioinformatics as an area of high priority during the tenth plan period. A comprehensive Bioinformatics Policy of India (BPI -2004) has been announced by the Department of Biotechnology, Government of India. The principal aim of the policy is to ensure that India emerges as a key international player in the field of bioinformatics; enabling a greater access to information wealth created during the post-genomic era and catalyzing the country's attainment of lead position in medical, agricultural, animal and environmental biotechnology. The government envisions that India should make a niche in Bioinformatics industry to create bioinformatics industry with US$10 billion by the end of 10th Plan period. The following strategy has been adopted to progress towards this goal.

- Develop the programme as an array of distributed resource repositories in areas of specialization pertinent to the needs of India's economic development.

- Coordination of the network through an Apex Secretariat
- System design and implementation in terms of computing and communication infrastructure, bioinformatics utilities etc.

- Facilitate and enhance application of bioinformatics

- Support and promote organization of long-term, short-term and continued training/education in Bioinformatics. National level testing for quality assurance on human resource on Bioinformatics shall also be conducted through reputed universities.

- Establish linkages with international resources in biotechnology information

- An international institute on Bioinformatics shall also be established to promote various activities on bioinformatics particularly international and entrepreneurial participation in these activities.

India also harbors an impressive infrastructure for biotechnology and bioinformatics. Some of which is elaborated below.

Biotechnology Information System Network (BTISnet)

Biotechnology Information System Network (BTISnet) is unique and one of the well recognized major scientific networks in the world, dedicated to providing state of the art infrastructure, technologies, education, manpower and tools in bioinformatics. The programme has evolved to deal with four distinct problems viz. (i) handling and management of biological data, including its organization, control, linkages, analysis, and so forth. (ii) communication among people, projects, and institutions engaged in the biological research and applications. The communication, which includes e-mail, file transfer, remote login, video conferencing, electronic bulletin boards and establishment of web-based information resources; (iii) organization, access, search and retrieval of biological information, documents, and literature; and (iv) analysis and interpretation of the biological data through the computational approaches including visualization, modeling & simulation, and development of algorithms for highly parallel processing of complex biological structures.

The network has established a link among scientists. It offers a single window information resource in the country covering inter-disciplinary areas of biotechnology and molecular biology. The network programme consists of 10 Distributed Information Centres (DICs), 50 Sub-Distributed Information Centres (Sub-DICs) and an apex Biotechnology Information Centre (BTIC) that coordinates the activities of the entire network.

Six National Facilities on interactive graphics are dedicated to the promotion of molecular modeling and other related activities.

Research and Development

The Bioinformatics Centers linked to BTIS are being extensively used for intensive research by the hosts and neighboring institutions. The thrust areas provided are sequence analysis and molecular modeling. Apart from supporting intensive research, the centers are also actively engaged in research on frontier areas of bioinformatics. Intrinsic research in the bioinformatics centers includes gene analysis, protein structure prediction & engineering, modeling macromolecular assembly, evolutionary biology and mechanisms of disease.

Databases and Mirror sites

The BTIS centers have developed more than 100 databases on various aspects of biotechnology. Several major International databases for application to genomics and proteomics have been established in the form of mirrors as part of the network. A Virtual Private Network (VPN) has been established to link these databases through high speed and large bandwidth network to promote faster sharing of information. Presently 11 institutions have been linked through VPN. These sites are designed to act as knowledge pathways for discoveries in biotechnology.

Some of the representative databases are, 'Biotech Industries of the Rajasthan', 'Flora of the Indian Desert'. Dr. Y.S. Parmar University of Horticulture and Forestry, 'Medicinal and Aromatic Plants of Aravali in Rajasthan', , 'Marine Life of India', 'Marine crabs of India' and 'Medicinal and Aromatic Plant of J&K'

Biogrid India

Research in biotechnology, which is highly knowledge and capital intensive, has generated a deluge of information in this decade. To make use of this information effectively there is a need for high speed and large bandwidth network. Towards this end, the department has successfully established a high-speed and high-bandwidth network in the form of Virtual Public Network (VPN) named as BIOGRID INDIA. The network was established through HCL Infosystems and is being coordinated by National Brain Research Center (NBRC).

Eleven institutions have been networked under this project in the first phase. These nodes are actively pursuing bioinformatics activities such as human resource development and R&D in bioinformatics besides dissemination of biotechnology information to researchers in the country. The nodes are interconnected through 2mbps dedicated leased circuit line at each location and 4Mbps internet bandwidth shared from the central server by all the nodes.[1]

- **Initiatives for Transfer of Technology**

  The BTIS centers are, constantly developing electronic versions of information resources relevant to biotechnology. These are expected to function as vehicles for knowledge and technology transfer to the people. In this area, the Indian Institute of Spices Research, Calicut has developed a package called 'Spiceprop' featuring the

[1] Department of Biotechnology Annual Report, 2002-2003 (http://dbtindia.nic.in/annualreports/2002-3/chapter8.pdf)

micropropagation protocols in spices. The Regional Research Laboratory, Jammu has developed three CDs on 'Science and Practice of Mushroom Growing', 'Marvels of Himalayan Herbs' and 'Commercial Floriculture in the north west Himalayas'. The Central Sericulture Research & Training Institute, Mysore has completed work on an electronic edition entitled 'Practical Technologies for Commercial Bivoltine Silkworm Rearing'. These products are expected to be released shortly.

- **National Super Computing Facility**

  A National Facility has been established at IIT Delhi towards the development of In-silico Drug Development by using Bioinformatics applications.

India's Scientific Workforce

Various studies show that the scientific talent pool of 4 million Indians is the second largest English speaking group worldwide, after the US. The availability of a large resource of English speaking scientists gives India an edge over other countries to reach the forefront of R&D services in biosciences. According to some estimates, 15 per cent of the scientific population of pharmaceutical and biotechnology companies in the US is of Indian origin.

3 million graduates, 700,000 post-graduates and 1500 PhDs qualify in the scientific stream each year in India. Presently, 15,000 scientists are estimated to be engaged in India's biotechnology sector.[2]

India's Workforce In Bioinformatics

Human resource development has been recognized as an important area for effective sustenance of the bioinformatics programme. Towards meeting the needs for trained bioinformatics professionals, long-term courses are being run as part of Biotechnology information system network (BTISnet) in various reputed Universities. In addition to this about 50-60 short-term courses are being held every year to train the researchers and scholars in bioinformatics.

According to Puneet Mehrotra, technical director of Bioinformatics Institute of India (BII), "The total number of students enrolled in BII's regular and distant learning programs has reached up to 5,000

In 2003 there were 27 M.Sc. courses in general biotechnology, 7 M.Sc. courses in agriculture biotechnology, 6 M.Tech. courses in biochemical engineering, bioprocess technology and biotechnology, 2 M.Sc. courses in marine biotechnology, 2 M.Sc. course in medical biotechnology, 1 M.Sc. course in neurosciences, two post-M.D./M.S. certificate courses in medical biotechnology, one post-graduate diploma course in molecular and biochemical

[2] Vibrant Gujrat, Govt of Gujrat Official Website.
http://www.vibrantgujarat.com/sp/emerging_technology_bt.html

technology and one post-graduate diploma course in genetic engineering and bioprocess development. The intake of students in the post-graduate courses was around 800.[3]

Pharmaceutical and Biotech Industry

India has a vibrant generic pharmaceutical industry. Apart from this there is impressive growth in biotechnology sector in the past decade. Biotech products produced in India exceed 1 Billion dollars.

Another estimate puts it to around 700 million $ (2003-2004). According to Kiran Mazundar-Shaw, president Association of Biotechnology Led Enterprises (ABLE) India has addressed biotech segment over the last decade by leveraging its low cost scientific skill base in an imitative manner with which to build it capabilities.

Investment in Biotechnology and Bioinformatics

The World Bank has given $240 million to the Indian Council of Agricultural Research towards the National Agricultural Technology Project (NATP), a 5 year project to focus on plant and agriculture BT research and private sector development.

Spectramind, a call centre has decided to invest 100 crores in this arena and would recruit at least 100 experts who would work on internet and other s/w, not in biology lab. (Dainik Bhaskar)

- **Bioinformatics Society of India:**

  Bioinformatics society of India (Jubios) has been working since 2001 August. It has over 300 members. In has become a common Informal Platform for the younger generation to learn and contribute to this new industry.

Bioinformatics Companies in India

In the following section a review is given of some prominent Indian companies working in the field of bioinformatics.

- **Location:**

  In India most of the bioinformatics Companies have been born in clusters where there are a large number of IT biotech, and pharmaceutical companies.

  Among these places are Banglore, Pure, Hyderabad, Chenai, and Delhi with Pharma and biotech companies like Dr. Reddy's Lab, Shantha Biotech, Bharat Biotech, Biogenus, Indigene, Biogn, Avesthegen, metanelix, and Astra Zeneca.

[3] B.S. PADMANABHAN, BIOTECH REVOLUTION: Building expertise, Biospectrum, Volume 20 - Issue 17, August 16 - 29, 2003.

- **Source of Business:**

  Although most of these companies have active collaboration with local Parma and biotech companies they are not completely reliant on them. They are actively looking for market opportunities in US, Europe and Australia.

  The main job type that is being outsourced here can be contract research, services, maintenance, Molecule Design and Database related jobs.

- **Size and Growth:**

  Most companies are small and medium sized with 20 – 50 employees.

  A typical bioinformatics product would take a year to develop and another year or more to reach the market.

  Initially there were only few bioinformatics companies and they either decided to do contract research or never got about releasing their products. Many of the companies said they were bioinformatics companies and they were either training companies or services oriented companies.

- **Company Structure:**

  Bioinformatics companies generally have well-defined teams with a clear business focus. For example, in Mascon, there are three groups—a functional team, which is the driving force, consists of domain experts. The development team is the supporting arm and comprises the software professionals. Then there is a business development team. Likewise, SysArris has domain experts who are thorough in genomics, cheminformatics and other areas. These experts understand the requirements of customers, analyze it and suggest the solution. The software team then converts the requirements to a software solution.

- **Major Companies**

  Some of the major bioinformatics companies in India have been discussed in the following section.

  1. Strand Genomics
  2. SysArris
  3. SciNova Technologies
  4. Ocimum Biosolutions Inc.
  5. Wipro
  6. Infosys
  7. Cognizant Technologies
  8. Satyam Computers

9. Tata Consultancy Services (TCS)
10. IBM India
11. Intel
12. Sun Microsystems

- **Strand Genomics**[4]

It is a life sciences informatics company developing and marketing software for drug discovery and development. Strand's core competencies include data mining and analysis, visualization, knowledge management and software engineering. It provides value-added software to biotechnology and pharmaceutical industries.

Corporate Address	Managing Team	
Strand Genomics 237, Sir C.V. Raman Avenue Rajmhal Vilas, Bangalore 560 080, Kamataka, India Tel. :+91-80-23618992,23618993 Fax: :+91-80-23618996	Dr. Vijay Chandru Dr. Ramesh Hariharan Dr. Swami Manohar Dr. V. Vinay Dr. Kas Subramanian	Chairman, Co-Founder & CEO Co-Founder & CTO Co-Founder & Advisor Co-Founder & Advisor Chief Scientific Officer (CSO)

Products: Strand's marketed products include Avadis and Acuris. Strand also offers professional services to customize and optimize the performance of its products. Strand licenses a suite of products for microarray design and analysis, as well as a specialized product for high-throughput crystal image analysis and classification.

Avadis: It is a comprehensive data mining and visualization tool for data cleaning, filtering, transformation and normalization techniques. Avadis Microarray is a functional extension to the product for conducting microarray gene expression data analysis.

Acuris: It offers gene annotation to automatically gather, persist and present gene-related public information and literature. Acuris uses industry-standard 3-tier architecture.

Sarani: It automates large-scale design of optimal oligonucleotide probes for microarray experiments. Thousands of gene sequences can be analyzed together and the best available oligonucleotide probes with uniform thermodynamic properties and minimal similarity to non-specific genes can be selected.

Chitraka: It provides automatic grid layout, spot finding, comprehensive spot statistics and robust background signal correction.

[4] **Source:** www.strandgenomics.com

Sphatika: It is a crystal image classification tool for high throughput X-ray crystallography. It classifies protein crystals with more than 92% accuracy on harvestable crystals and crystal hits together with a 20% false positive rate.

- **SysArris**

It is a life sciences informatics company providing software services to clients globally, for drug discovery and drug development. SysArris has leveraged its experience in providing software for drug development for the last 10 years, combined with its experience in solutions for managing, mining and integrating data across applications, to provide software solutions for drug discovery.

Corporate Address	**Managing Team**	
SysArris Software No. 120 A, Elephant Rock Road Jayangar-III Block Banglore – 560 011, India Tel. :+91-80-26655165, Fax: :+91-80-26650374 Email: marketing@sysarris.soft.net www.SysArris.com	Ajay Simha C. V. Mukundan C. R. Seshadri C. Vinod Kumar	Director General Manger, Marketing V.P. Off-shore Development Center G.M. Off shore Development Center

SysArris has developed and delivered custom solutions in the areas of RNAi technology, nanotechnology, HTS apart from building products to design RT- PCRs, siRNA, microarray analysis and virtual screening. Products developed for drug development are in the areas of clinical trials, adverse reaction and drug registration amongst others. These products are compliant with international regulatory guidelines set by organizations like FDA, BFARM-caps, MCA, CIOMS, MOH and MHLW. SysArris specializes in offering software solutions in genomics and range of requirements of gene discovery programs, functional genomics, recombinant DNA technologies, target research, pharmacogenomics and combinatorial chemistry. From conceptualization to reality, it provides software services to help the clients to transform their business by lugging the right technology.

Products

Oligosys: It is a tool for the design of oligonucleotides for hybridization reactions, PCR and real time PCR.

- Available on Internet
- Multiple template sequences for batch processing
- Oligosys designs oligos with speed and accuracy
- Multiple templates in Genbank /FASTA formats

Pindrop: User-friendly, convenient, efficient tool to design siRNA oligos and to construct siRNA expression vectors for study of gene expression and function. Pindrop has been developed with algorithms, which design siRNAs that work toward reducing gene expression considerably. The algorithm is constantly refined, as more is understood about the criteria that affect siRNA potency and specificity.

- **SciNova Technologies**

SciNova Technologies is an Indian life science informatics company, having expertise in bioinformatics, chemo informatics and medical informatics. Present in 42 countries around the world, the company offers customized solutions using proprietary software, and focuses on the requirement of each client. SciNova designs products to enable cutting-edge drug discovery and helps researchers narrow down the huge list of probable drug targets for actual experimentation in the laboratory. Their products make it possible for biotech, pharma, diagnostics and chemical enterprises to design and deploy customized applications. SciNova provides services and products to organize, plan, develop, integrate and manage the applications and software specific to environment to maximize the return on investment. It also undertakes simulation and modeling studies for 'biological and chemical data /processes that are dynamical in nature.

Products

Chemlab: It is knowledge management software for laboratories. It has been designed by lab managers, scientists and technicians and coded by software experts. It captures all the parameters of an experiment in an intelligent manner. It keeps a track of yields obtained and the quality of the end product and also helps manage inventories by listing the products about to expire or have reached reorder levels. A customized report can be generated in a matter of seconds. ChemLab provides powerful querying and reporting tools so that the information can be had at the click of a button. It is as easy to use as a word processor and has the power of a relational database. Chemlab can be customized for chemical, pharma or bio-technology research.

Prometheus: It provides a modular, cutting edge yet extremely user-friendly data mining product for life science research. It allows researchers to use their domain expertise and not bother with the underlying complexity of data mining techniques.

Life science data from the molecular to the clinical level is typically multivariate and inherently non-linear. Prometheus has algorithms like SVMs, neural networks and decision trees that can build accurate predictive models to hasten research. Prometheus is the integration of the data mining components like SVMs, Decision trees, neural networks and some multivariate statistical techniques. Prometheus uses a visual pipeline approach for data mining.

Rx is a toolkit for calibration, annotation, characterization and predictions based on ECG signals. It can handle a large number of signals at the same time and can even be

configured to sit on a database of signals. Eventually it should be able to take real time data from the instrument.

Silent features of Prometheus

Bioinformatics: Prometheus brings the power of QSAR analysis to bioinformatics. Physico-chemical, information theoretic and time series descriptors are calculated from sequences with different classification. Applications include Promoter prediction, fold identification, subcellular location prediction, GPCR classification, identification of drug targets *(e.g. essential bacterial genes)*

Chemo Informatics: Prometheus is to design new molecules with "chemical rules" generated from known activity data. Properties of the small molecule structures to their biological activity are non-intuitive for a medicinal chemist.

Medical Informatics: Analysis of ECG data to annotate, analyze, diagnose and detect cardiac anomalies by using machine learning and non-linear methods.

Structure Elucidation of Natural Compounds: Isolated from plants, fungi, bacteria or other organisms s a common problem in natural product chemistry. 1H-NMR and 13C-NMR spectroscopy methods are important, as hydrogen and carbon are most abundant in natural products.

- **Ocimum Biosolutions Inc.**

It is a life sciences contract research and development company with competencies in ; Bioinformatics, LIMS, Genomics, Proteomics and custom contract research services, with operations in USA and India. Its team comprises experts from life sciences/computer sciences backgrounds.

Corporate Address	**Managing Team**	
Ocimum Biosolutions 6th Floor, Reliance Classic Road NO 1, Banjara Hills Hyderabad – 500 034, A.P, India Tel: +91-40-5562 7200 Fax: +91-40-5562 7205 Email: hyd@ocimumbio.com	Ms. Anuradha Acharya, Dr. P. Sujata, Mr. Subash Lingareddy Mr. Sunil Lingreddy	Founder & CEO Founder & CSO Found & CFO Director

Ocimum provides key, ready-to-use reliable, cost-effective software solutions for the I biotech/pharma industry with a suite of products such as Biotracker, Toxchek, Genchek, IRNAchek, Genowiz, OptGene and Nutrabase and an array of custom services including data mining, algorithm development, gene identification, multiple- platform software development, database creation and manipulation, and tools for image analysis.

The custom contract research services are provided to pharma and agricultural organizations, and fermentation services for biologicals and non-biologicals. Its labs utilize state-of-the-art systems for sample tracking, nucleic acid and protein analysis, and provide highest standards of custom tailored services reliably and cost- effectively.

Products

iRNAchek: It is one of the most powerful molecular biology tools with a wide variety of applications. RNA interference is silencing of a gene caused by the introduction of a homologous dsRNA. The success of RNA interference experiment is heavily dependent on the design of short interfering RNA, for efficiency and specificity. iRNAchek provides a comprehensive and intuitively designed environment for organization of sequence data, design of siRNAs, and tracking and analysis of successful templates.

Genchek: The product is a comprehensive, multi-platform, sequence analysis software package. It facilitates .analysis of Expressed Sequence Tags (ESTs), Complete Genome, and SNP *(Single Nucleotide Polymorphism)* data. Genchek is a research information and experiment management tool that integrates public and proprietary data through a discovery workspace that provides contextual access to sequence analysis tools, content and services. Genchek has an integral database system that can be used to access, store, organize and retrieve DNA, RNA, and protein sequences in an intuitive environment that offers editing, management and annotation of sequences.

Genowiz: It is a powerful gene expression analysis program. It includes a suite of advanced analysis methods and offers choice for selecting analysis that is appropriate to a researcher's dataset. Genowiz provides numerous visualization options to track down intricate correlation in microarray data. Genowiz allows researchers to organize experimental information (MIAME), quickly and easily import data files, preprocess, normalize, plot, perform cluster analysis, classify, visualize patterns, review gene information, and link analysis results to external tools.

OptGene: It is a novel gene optimizing tool that optimizes naturally occurring genes to achieve higher productivity, giving higher flexibility for protein design. The tool optimizes the genes using only the sequence information and the choice of expression system. Optgene allows the researcher to adapt genes and their products precisely to their specific requirements.

Nutrabase: It is a database archival and accessing system for Flavonoids. This database includes easy to use interfaces and analysis modules for chemical, biological properties, genetics, classification and geographical distribution of source plants of flavonoids.

- **Wipro**

Wipro Healthcare & Life Science division has an ambitious goal: to be among the top 3 solution providers in this segment globally. For the quarter ended March 2003, Wipro's

revenue from the health sciences sector is approximately $2.5 million. The company which almost single handedly lifted India's bio-IT solutions business in the last two years hopes to cross $500 million in the next ten 10 years.

In the life science segment, Wipro's focus is to offer pharma companies IT solutions that will reduce drug discovery and approval time. In the drug discovery stage, for example, it will work as a technology partner for a consortium of life science equipment vendors like Beckam Coulter, GE Medical Systems, Agilent, Fujinon and Cemer who supply equipment that aid genetic drug discovery process.

- **Infosys**

The software bellweather's foray into the field of bioinformatics and life sciences was revealed in August 2002, when Infosys opened its life sciences division. Although the company has been keeping mum about its progress in the life sciences and pharmaceutical markets, it is recording client wins by the quarter. In 2003, the company announced client wins of a global leader in contract research, a European biotechnology leader and an emerging India-based global pharmaceutical company. Further, Infosys is working with the US division of a global pharmaceutical leader on a performance management dashboard for their senior executives.

- **Cognizant Technologies**

The US software services major, Cognizant, is a major player servicing the healthcare sector across the world. Today, the company boasts of possessing a dedicated team of over 800 professionals globally for the healthcare sector.

The company's healthcare practice has contributed approximately 22 percent of its $229 million revenue. In the pharma space, Cognizant works with large pharma and biopharma companies providing Sales Force Automation and CRM integration solutions, enabling 21CFR11 compliance *(a US pharma regulatory standard),* pre-clinical and clinical trial management software and internalization solutions. Its customers include marquee names such as United Healthcare, John Deere Healthcare, Blue Cross of North Eastern Pennsylvania, Sierra Health Services, IMS Health and Fletcher Allen Healthcare.

"The importance of information and technology has never been greater in the life sciences industry than today. The rapid development of biotechnology over the past decade has opened up the opportunity for developing drugs that could be targeted at specific diseases and human population, thereby leading to a potential explosion in new drug discovery", says Mohan Narayanan, VP, Healthcare and Life Sciences, Cognizant. "Cognizant is at the forefront of this bio-information revolution and is playing a key role in drug development and commercialization informatics."

Cognizant's healthcare and life sciences solutions combine deep industry knowledge, methodologies and proprietary tools for IT-intensive healthcare and bioinformatics organizations.

- **Satyam Computers**

In 2001, Satyam threw its hat in to the bioinformatics ring and entered in to a five-year agreement with the Center for Cellular and Molecular Biology (CCMB), Hyderabad, to develop software tools to sift and search through large volumes of genetic material for vital DNA fragments. Satyam believes that the convergence of two technological revolutions; information technology and biotechnology helps in "time compression".

Satyam has a strategic business unit which will draw the competencies of CCMB, one of the world's leading genetics research centers, in the global market. The solutions offered by Satyam include IT manpower services *(to biotech companies),* database value-added services and pharma-cum-database value added services.

- **Tata Consultancy Services (TCS)**

Another major player in the Indian bioinformatics market is the country's largest IT company, Tata Consultancy Services. Based at the company's Advanced Technology Center (ATC) in Hyderabad, the bioinformatics practice is working to provide services such as automated genome analysis, protein structure prediction and high throughput molecular modeling, rational drug design and creation and integration of relational databases from proprietary, unstructured pharmaceutical and clinical data.

Currently, TCS is developing an end-to-end bioinformatics software package "Biosuit" on the Linux platform with domain knowledge from 20 leading Indian institutions. The Biosuit project, in collaboration with the Council for Scientific and Industrial Research (CSIR) and supported by the Department of Biotechnology (DBT) will roll out the alpha version soon with the full product expected to hit the market in 2004.

TCS also has a research agreement with the Center for DNA Fingerprinting and Diagnostics (CDFD) of DBT. The TCS Life Sciences Practice aims to offer end-to-end solutions, services and products to life sciences and healthcare institutions worldwide. The services range from application development management, outsourcing, systems integration and embedded systems.

- **IBM India**

IBM has invested $40 million on Blue Gene, a supercomputer that can simulate protein structure. It has also spent another $100 million to build its Life Sciences unit in August 2000. The Big Blue recently touted the progress that its Life Science unit had made: more than 30 analytical instrument and tools companies are basing their products on IBM middleware, server platforms and services. The company says its life sciences division is its fastest-growing unit. IBM's strategy is to team up with companies selling to the life sciences marketplace.

Buoyed by the rapid strides, IBM has now more than doubled the investments in the division and has over 1,000 employees in the unit. The IBM Research Lab in India has come up with a biochip technology that enables the simultaneous study of expressions of thousands of genes or proteins by using a single experiment in the laboratory. Such experiments generate a huge volume of data which is expected to help understand the genetic basis of diseases. The IBM Research Lab is also developing a solution to facilitate browsing, querying and analysis of data.

- **Intel**

Intel India's research work in the biotech field is aimed at revolutionizing the future with "chips that can tell someone they are at the earliest stages of a life-threatening disease, such as cancer". Intel's CTO (chief technology officer) Pat Gelsinger says, "we envision a future in which every piece of silicon will include computing technologies but also connect to multiple wireless networks and roam between them. Your socks could tell if you are going to get a blister... you could have your mirror looking back at you and tell you if you have the onset of skin cancer."

Intel's biotechnology head, Andy Berlin, dreams of creating "a chip that tells someone they are at the earliest stages of a life-threatening disease, such as cancer".

The growing popularity of clusters is part of a worldwide trend which has seen a dramatic increase in the number of Intel-based systems being used for high performance computing deployments.

A number of organizations including the Indian Institutes of Technology (IITs) and the Council for Scientific and Industrial Research laboratories use Intel-based servers and clusters for research work in the life sciences area. Some of the research efforts are based on clusters that link multiple desktop or server computers to harness their collective computing power. "Intel is looking at building life sciences applications in diagnostic instrumentation and disease detection", says Ketan Sampat, President, Intel India.

Intel's research and development labs around the globe are developing silicon radios and "context aware" computing. Intel is pursuing the development of radios based on the company's low-power CMOS silicon manufacturing process. Intel says it is closer to realizing its goal of developing "reconfigurable radios" that would automatically identify and connect to a number of wireless networks - including the 802.11 standard, Bluetooth and Ultra Wideband - enabling any device powered by one of these chips.

- **Sun Microsystems**

Sun Microsystems is actively involved in bioinformatics research and has set up a group to support the development of Java and XML tools for the Life Sciences industry. Sun says. "Bioinformatics can undo the damage caused by the loss of the dotcom revenue," is its standard line.[5]

➔ **Some Bioinformatics Research Institutes In India[6]**

- ❖ Centre for Biotechnical technology (CBT), New Delhi.
- ❖ Department of Biotechnology, JNU, Delhi
- ❖ Centre for Cellular and Molecular Biology (CCMB), Hyderabad.
- ❖ Microbiology Department of Biotech centre, M.S University of Baroda, Vadodara.
- ❖ Central Drug Research Institute, Lucknow.
- ❖ Indian Toxicology Research Centre, Lucknow.
- ❖ Department of Biological Science, Tata Institute of Fundamental Research, Mumbai.
- ❖ National Centre for Biological Science, TIFR Bangalore.
- ❖ Indian Institute of Technology, Kanpur.

[5] Source: Bioinformatics in India: Can it take a quantum leap?
http://biospectrumindia.com/content/search/showarticle.asp?arid=46373&way=search

[6] Source: Bioinformatics Institute of India (www.bioinformaticscentre.org)

11 Bioinformatics Education Institutes in India

India has over 15 Years Experience in Bioinformatics Education

The first bioinformatics education program, the Bioinformatics Centre (BIC) at Jawaharlal Nehru University, New Delhi was stated in India in 1989 funded under the Biotechnology Information Systems Programme (BTIS) of DBT.

Various bodies in India are offering courses in Bioinformatics.

Following are the educational institutes imparting education in bioinformatics and biotechnology.

BIC at Jawaharlal Nehru University, New Delhi

The Bioinformatics Centre (BIC) at Jawaharlal Nehru University, New Delhi is among the pioneers of bioinformatics education in the nation. BIC was set up in 1989 funded under the Biotechnology Information Systems Programme (BTIS) of DBT. BIC-JNU is offering one-year post M.Sc Diploma in Bioinformatics and PhD program in bioinformatics. Computational genomics is the thrust area of BIC.

Bioinformatics Institute of India, Noida UP

It offers three PG Diploma courses via distance learning program to graduates, biologists, software professionals and professionals in other related fields. These courses are provided through local study centers in cheminformatics, bioinformatics and biomedical informatics all over the country. The total number of students enrolled in BII's regular and distant learning programs has reached up to 5,000.

CBT at Anna University, Chennai

The Center for Biotechnology (CBT) at Anna University, Chennai was started in 1987 with funds from the University Grants Commission, Department of Biotechnology and Anna University, In 1992, the center started a B Tech program in industrial biotechnology. It is supported by a large consortium of industries who have set up a corpus fund to the run the program. Recently, CBT got a boost with more funds donated by a well-known biotech company, Amersham Biosciences. CBT also offers a two-year M Tech (biotech) and PhD programs in biotechnology. Anna University is the largest Technology University in the country with 222 engineering colleges in Tamil Nadu affiliated to it.

Center for DNA Fingerprinting and Diagnostics

This is an autonomous institution supported by the Department of Biotechnology (DBT). It is India's premier center providing services in the areas of DNA fingerprinting, molecular diagnostics, and bioinformatics. CDFD is now poised to initiate basic research in fields relevant to its objectives.

CDFD is also the bioinformatics national node for the European Molecular Biology network (EMBnet). The node currently provides bioinformatics services in the form of browsing biomolecular sequence databanks, macromolecular structure databank, genome and other useful databases.

Indraprastha University, Delhi

Guru Gobind Singh Indraprastha University (GGSIU) is offering BTech course in bioinformatics. It has some unique features in its course curriculum. This special academic program is a result of successful alliance between School of Information Technology and School of Biotechnology at GGSIU. The four-year B Tech program in Bioinformatics was first started in 1999 with 30 students. The university is also offering an integrated five-and-half year M Tech program in biotechnology.

University of Pune (MBA Biotechnology)

University of Pune has taken a path-breaking initiative to catalyze the biotechnology revolution. It has launched a two-year full-time MBA program in biotechnology in the academic year 2002-03. The course has been envisaged to create managers with domain specific knowledge for the biotech industry. The syllabus is based on the requirements of the industry.

Sixty percent of the syllabus covers the managerial aspects and 40 percent biotechnology. The institute received overwhelming response for the course during the first year. The program admits 60 students each year. There will be a provision to admit international students also.

VG Vaze College, Mumbai

Kelkar Education Trust's VG Vaze College offers degree-level courses in biotechnology. The college has a state-of-the-art biotechnology lab and a scientific research center. The center under takes research in areas relating to medical, aromatics, cosmetics and environmental biotechnology. DST and University of Mumbai have bestowed the center as a recognized institution for research. Students will be admitted at the center for M.Sc (by research) and PhD in life sciences. The center is focusing on the development and mass propagation of medicinal and aromatic plants. The college, in collaboration with the research center and industry, has plans to offer certificate and diploma courses and training in perfumery, cosmetics, horticulture and tissue culture techniques, agricultural applications, etc. These courses are expected to commence from this academic year 2003-04. They will be offered to the science stream students. These diploma courses will be simultaneously offered to the students pursuing other degree courses too.

Rai University, Bangalore

Founded by Rai Foundation, this group has spread its educational arms across the nation with campuses in Bangalore, Behror, Bhopal, Dehradun, Gurgaon, Hyderabad, Kolkatta, Kosi, Lucknow, Mumbai, New Delhi, Pathankot and Pune. It is offering B.Sc in biotechnology, B.Sc (Hons) in environment management studies, Graduate Diploma (Hons) in biotechnology and patent laws and GPD (Hon) in forensic sciences. It also offers post graduate diploma in bioinformatics and in industrial microbiology.

RGCB, Kerala

Rajiv Gandhi Centre for Biotechnology (RGCB) is an autonomous research institute that functions under the umbrella of Science, Technology and Environment department of the Government of Kerala. Established in its present form in 1994, RGCB has come to occupy an important position among the research institutes of the country devoted to research in the cutting edge areas of modern biology and biotechnology. It occupied a new building in January 2002. With the support of the DBT it has established research facilities required for carrying out modern biological research. Currently, the research activities are being carried out in seven major divisions: infectious diseases, plant molecular biology, environmental biotechnology, molecular human genetics, neurobiology, cancer biology and molecular endocrinology.

East West Institute of Technology

East West offers Bachelor of Engineering courses in biotechnology engineering. This is an eight-semester course, affiliated to Vishweshwaraiah Technological University. It also offers graduate and postgraduate courses in biotechnology.

Edubiotech

Edubiotech is a technology driven organization, providing world-class biotech technical support to educational institutions and corporate sectors. Edubiotech accelerates innovative ideas by providing good quality R&D and lab facilities. Each one has a place and is given the opportunity to excel in their field of interest.

Oxford Group, Bangalore

The Oxford Group of Educational Institutions is a group that yearns at strengthening the concept of value based education. There are 22 educational institutions in all the branches of this institution. The Group has a separate College of Engineering as well as a college of Science in Bangalore. Today the Group is offering engineering as well as graduate and postgraduate programs in biotechnology. The Oxford College of Engineering is affiliated to Vishweshwaraiah Technological University and approved by AICTE, New Delhi and offers four-year BE in biotechnology engineering. The Oxford College of Science offers B.Sc in biotechnology and related fields such as microbiology, genetics and biochemistry. It also offers postgraduate courses in microbiology, biotechnology, and biochemistry.

Padmashree Institute, Bangalore

Padmashree Institute of Information Science (PIIS), affiliated to Bangalore University, is offering courses in Biotechnology at both undergraduate and postgraduate levels. The biotech program at the institute has been designed to bridge the gap between industry requirements and the growing demand for skilled manpower in the sector. The institute also offers services to the industry. The areas include protocol development for micro-propagation of ornamental and medical plants, maintenance of mother cultures, supply of planting material for cultivation, testing of microbial activity, testing microbial load in food cosmetics, drugs and extracts and also it has the provision of services for research fellows in the areas of plant tissue culture, biochemistry, microbiology and diagnostics.

RK Institute of Management and Computer Science

RK Institute of Management and Computer Science, recognized by the Government of Karnataka, affiliated to Bangalore University and approved by AICTE is offering courses in biotechnology at both under graduate and post graduate levels. The institute has a highly sophisticated research lab. The lab meets the requirements of the Bangalore University curriculum. The labs are well equipped with air conditioned tissue culture chamber and advanced models of laminar flow bench, UV-Vis Spectro photometer, incubators, autoclave, microscopes, electrophoretic units, optical instruments charts, models and other useful instrument to impart and enhance the practical aspects. Such progressive instruments and equipment contribute to the richness and variety of the courses offered by the institute.

STG

Software Technology Group (STG) is a technology driven company established in 1993 in San Jose, California, providing world class software consulting and training solutions to organizations and professionals in several countries. STG is a leading provider of advanced software education and has strategic alliance with IBM, Microsoft, Red Hat and Sun Microsystems amongst others. STG trains thousands of professionals each year on leading edge technologies. Bioinformatics becoming an increasingly important competitive differentiator for public and private life science companies, STG entered into this emerging field to provide extensive and expert training. Bioinformatics at STG is structured and designed to meet the industry's requirements and has collaborated with several institutions and research organizations to value add to its students. Live projects are provided and there will be an opportunity for the students to work in the horticulture department laboratory under the guidance of professors and research scientists at UAS.

Wageningen, Bangalore

Europe's premier institution, Wageningen University, recently launched its educational module for biotechnology in Bangalore for the first time. Wageningen University in the Netherlands was established in 1918 and had over 10,000 students emerging from its portals annually. As one of the leading centers of education and research in plant, animal, environmental, agro-technological, food and social sciences, Wageningen had been offering quality programs in Asia. The mission statement of the university is to develop and disseminate scientific knowledge needed to sustain and supply society's demands for sufficient healthy food and a good environment for humans, animals, and plants. Wageningen

had tied up with two service providers in India—Bioinfraa and C-Lift, both Bangalore-based companies.

SS Infotech, Bangalore

SS Infotech started in the year 1999 as a non-profitable organization to provide high quality education to less fortunate students at an economical cost. Started with M.Sc (IT), B.Sc (IT), MBA and BBA courses, presently it is offering one year PG diploma in bioinformatics through Bioinformatics Institute of India, Noida. The objective of the course is to provide adequate knowledge and training to the student with hands on experience in the field of bio-informatics. The institute has developed a course keeping in mind the requirements of the industries and to equip the students with the latest advances in the field of bioinformatics. The institute also offers correspondence courses in biomedical-informatics, bioinformatics and cheminformatics.

Biotech Education Services &Training (BEST), Bangalore

BEST was started as a company incorporated to undertake activities in imparting quality research and training in biotechnology. The objective of BEST is to promote quality hands on training program in biotech education. Besides training the company has successfully forged into contract research services, protocol development, water analysis and waster water management. BEST is equipped with state of art facilities, competent and dedicated technical manpower. Its network has helps the students in placements and career counseling. It offers one-year PG diploma, short-term course of 10-50 days and advanced diploma courses of three months duration in biotechnology.[1]

[1] Bioinformatics red hot, Biospectrum, Monday, July 21, 2003, http://www.biospectrumindia.com/content/careers/303072101.asp

12 Bioinformatics Business and Education in Canada

Keeping in view the growing importance of bioinformatics as a technology, Canada has taken the initiative to carry out research and set up businesses with collaboration between the industry, government and the research community.

Competence

The Canadian biotechnology industry invested about $600 million in R&D in 1997 and the sales of their products exceeded $1 billion in the same year. As far as geographic concentration of biotechnology firms and research institutions is concerned, the highest number of biotechnology firms is in Quebec, followed by Ontario and British Columbia.

Canada's capacity in bioinformatics with regards to the bioinformatics programs at the undergraduate level is weak since there are very few such programs, which in turn leads to a lack of professionals in the science. However, preparation is underway to introduce specialized programs in biotechnology at a number of universities, though the problem of lack of skilled teachers in the field persists. Another problem is that very few industries in Canada are including biotechnology in their business. Most of the biotechnology and pharmaceutical firms carry out their work in the US, which has a bigger R&D industry. Availability of funds and lack of properly trained personnel needed to start biotechnology research is also a problem. Another major problem voiced by the research community and businesses is that there is not enough government support for bioinformatics. There is also a lack of communication and collaboration between the stakeholders, which is likely to lead to overlapping of work.

However, the people undergoing the development of biotechnology in Canada are amongst the best in the world. In addition to study programs being designed by universities, workshops to train participants with biotechnology skills are being organized. Moreover, the few biotechnology firms in the industry are very keen on developing adequate research capabilities in Canada. Organizations are working to establish concentrated regions of information technology industry. Though the Canadian Bioinformatics Resource needs a lot of improvement, it is a good primary project that will enable the industry and research community to access information from around the world.

Threats and Barriers

Canada faces the following threats in its quest to make its Bioinformatics capacity global:

- Canadian Biotech companies are setting up operations in the US where the biotech industry is more advanced.
- Canadian professionals are being offered high paying jobs elsewhere and there is a threat that they might leave, this coupled with the fact that already there are not enough professionals in the field will pose dire problems for the Bioinformatics industry in Canada.

Competitive Analysis

Compared to the American and the European Bioinformatics industry, the Canadian industry is more competitive as far as the quality of programs, technology and skilled individuals is concerned. However, it lags behind in the number of bioinformatics programs being undertaken, also there is a lack of networks, shortage of funds and the number of bioinformatics industries functioning in Canada. The shortage of firms in the industry is primarily due to lack of funds for start-up business. Moreover, as has already been emphasized there are not enough trained personnel in the field. Another important factor is that most of the firms have their R&D being carried out outside Canada which makes technology transfer from research to industry very slow.

Institutes, Infrastructure and Programs

- **National Research Council**

The National Research Council is the premiere biotechnology research agency of the Canadian Federal Government. The NRC Biotechnology Program was established in 1983 under the guiding principles of the National Biotechnology Strategy.

NRC's research strengths are organized around key sectors including biotechnology, information and communications technologies, measurement standards, molecular sciences, aerospace, manufacturing, construction, ocean engineering and others.

BIOTECanada deals with the issue of research in bioinformatics in Canada. It is striving to spread awareness of the importance of bioinformatics in the every day life of the people. Out of thirty nine key players selected from Bioinformatics research and industry, interviews with twenty three were carried out and 59% of the key players were from the bioinformatics industry of Canada. It is imperative that software developers, hardware developers, mathematicians and bioinformaticians work together for bioinformatics to succeed in Canada. Apart from these main players, the research community, the government and the biotech industry need to collaborate and communicate with each other to ensure a thriving biotech industry in Canada.
The Biotechnology Program is a founding member of Genome Canada and contributes to Canadian innovation in genomics through the NRC Genome and Health Initiative.

- **DNA Sequencing Facility**

NRC's Institute for Marine Biosciences boasts Canada's largest and most advanced DNA sequencing facility. Equipped with world-class sample preparation and sequencing

technologies, robotics and advanced bioinformatics tools, the facility now provides DNA sequencing for other NRC institutes and for national and international clients.
The facility greatly enhances the Institute's research capacity and support the development of biotechnology applications leading to improvements in human health, food crops, and the environment.

- **NRC Plant Biotechnology Institute (Saskatoon)**

Research is being conducted in the following areas,

- Cell Technology
- Legume Performance
- Promoter/Gene Discovery
- Gene Expression
- Molecular Pathology
- Carbohydrate Modification
- Seed Oil Biotechnology
- Growth Regulation
- DNA Technology Unit
- Spectroscopy Services
- Transgenic Plant Centre

- **Plant Transformation System**

NRC's Plant Biotechnology Institute is working to start a Canadian breeding program for nutraceutical plants. The research will enhance the uniformity of plants grown for their therapeutic attributes and will eventually lead to the production of improved lines of these high-value plants.

- **Canadian Bioinformatic Resource** (Halifax)

NRC's Canadian Bioinformatics Resource (CBR) in Halifax, gives fast access to sequence information. It also provides databases and tools for universities and research organizations across Canada. The resource provides two distinct services: CBR-I supplies high-speed access to sequence information to NRC and associate members only. CBR-II makes bioinformatics available to universities and research organizations across Canada.

- **NRC's Genomics and Health Initiative**

The Genomics and Health Initiative is a unique horizontal program that involves several NRC biotechnology institutes located across Canada. Launched in 1999, the initiative strives to bring the benefits of advances in genome sciences and health research to a variety of Canadian industrial sectors and regions. GHI is advancing fundamental and applied technical research in areas such as the diagnosis of disease, aquaculture, human pathogens, crop enhancement, environmental remediation of pollution, cancer, neurobiology and protein assembly.

- **BioMiner Software**

NRC's Institute for Information Technology and NRC's Institute for Biological Sciences are working in collaboration to address the need for data mining capability for genomics research. Funded under the NRC's Genome and Health Initiative, the project has led to the development of a prototype data analysis software package, called BioMiner.

Canadian Bioinformatics Resource (CBR) is a recognized Sun Centre of Excellence in distributed Bioinformatics. It provides collaborators access to bioinformatics and computational biology applications, databases, large-volume data storage, basic training and help desk support. CBR hosts over 120 bioinformatics information suites and 150 biological and bioinformatics databases. It has gained international recognition for its expertise in remotely managing and administering distributed informatics infrastructures in support of lead bioinformatics and biological initiatives, genome health and genome proteomics as well as Genome Canada Research Projects.

An integrated and distributed Bioinformatics Platform for Genome Canada

The Bioinformatics platform for Genome Canada builds on existing infrastructure, including the Canadian Bioinformatics Resource (CBR) and the Calgary-based Sun Center of Excellence for Visual Genomics.

Research is underway on BioMOBY which will focus on ensuring that the accessibility and usefulness of biological data is maximized through the creation of common formats for both representation and data distribution. To date, this standardization has not been developed in the biological research arena and with the increase in the amount of data available, standards and common formats will be necessary. Another scientific goal that the platform is planned to achieve is the visualization of complex genomic features.

Over the next few years the platform will install a computing GRID in which networked computers from many sites in Canada will appear to function as a single computer. At the University of Alberta, we have established a help desk and custom programming facility that began operating in April 2003 and will assist other genomics research projects in Canada.

The project also features a major training component. Two Bioinformatics workshops are planned per year. Our first workshop was held in Calgary on June 7-15th, 2003 and attracted 20 researchers from across Canada. The courses are structured to enhance the knowledge and skills of wet-lab genome researchers who have a basic understanding of computational biology and programming skills.

The goal for the platform is that all Genome Canada researchers will be able to access and benefit from the platform including utilizing the GRID, adopt the standards recommended by the BioMOBY consortium, and access the software components through our platform website. In addition, the platform will enable training of over 100 Canadian researchers in the area of Bioinformatics and provide continued support and guidance through the Help Desk

NRC-CBR is a national distributed network of collaborating institutes, universities and individuals dedicated to the provision of bioinformatics services to Canadian researchers.

NRC-CBR has approximately 1,200 registered non-profit users across the country and provides bioinformatics infrastructure to 17 research groups from Halifax to Vancouver.

Research Programs

Below is a brief description of each of the research projects that are underway as part of the Genome Canada Projects:

- **Bovine Genome Sequencing Program**

  With the completion of the human genome sequence, the opportunities available for acceleration of genomics research activities in cattle have increased dramatically. This is even more valid now that the sequencing of the bovine genome has begun. As a result of the present effort, it will be possible to determine the fine structure of transcripts by providing details as to the position and structure of genes within the bovine genome sequence.

- **Diagnostic Applications of Microarrays in Organ Transplantation**, led by Dr. Lisette Xavier, This project aligns with the objectives of Genome Canada to capture the power of new technologies to improve diagnosis and treatment of human health problems.

- **Building The Metabolomics Toolbox: Enabling Rapid Disease Diagnosis Through Metabolic Profiling**, led by Lori Querengesser, uses small molecules called metabolites to detect changes in cell behavior and organ function. It also uses these chemicals to monitor and measure the larger-scale physiological changes that occur in response to subtle changes in the environment, thus helping to improve our monitoring of adverse drug reactions and better understand individual sensitivities to prescription drugs.

- **Functional Genomics of Abiotic Stress**, led by Dr. William (Bill) Crosby of the University of Saskatchewan. This project includes researchers in Vancouver, Lethbridge, Calgary, Edmonton, Saskatoon, Winnipeg and Montreal. This project is being conducted in partnership with Genome Quebec.

- **Commercialization and Society: Its Policy and Strategic Implications**, led by Dr. Edna Einsiedel of the University of Calgary. This project includes researchers in Calgary, Edmonton and Saskatoon.

- **The Development of Enabling Technologies for Proteomic and Genomic Research**, led by Dr. Bill Davidson of MDS Sciex and co-led by Drs. Jed Harrison and Liang Li of the University of Alberta. This project includes researchers in Edmonton, Winnipeg, Toronto, Kingston and Quebec City.

- **Functional Pathogenomics of Mucosal Immunity**, led by Dr. Lorne Babiuk of the Veterinary Infectious Diseases Organization at the University of Saskatchewan and co-led by Dr. Bob Hancock of the University of British Columbia. This project includes researchers in Vancouver, Burnaby and Saskatoon. This project is being conducted in partnership with Genome BC.

- **Enhancing Canola through Genomics**, led by Dr. Wilf Keller of the NRC Plant Biotechnology Institute in Saskatoon. This project includes researchers at the NRC Plant Biotechnology Institute and the Agriculture and Agri-food laboratories in Saskatoon.

- **An Integrated and Distributed Bioinformatics Platform for Genome Canada**, led by Dr. Christoph Sensen of the University of Calgary. This technology platform includes researchers in Calgary, Edmonton, Saskatoon, Winnipeg, Toronto and Halifax.

The following chart illustrates how the Genome Canada projects are distributed across the country according to region and sector of application:

	BC	Prairie	Ontario	Quebec	Atlantic	Total
Sector						
Agriculture		3	2		1	6
Environment	1			1		2
Technological Development		1	2		2	5
Fisheries	1					1
Forestry	1		1	1	1	4
GELS	1	1	2	1		5
Health	6		9	12		26
Platforms	1	1	2	1	1	6
Total ($ M)	**100**	**91**	**190**	**172**	**31**	**584**

Bioinformatics Educational Institutes in Canada

Canadian Bioinformatics Program Listings

» British Columbia (4)
» Alberta (4)
» Saskatchewan (1)
» Ontario (13)
» Quebec (7)

British Columbia

Simon Fraser University:

Joint Major in Computing Science and Molecular Biology and Biochemistry - provides undergraduate training for a bioinformatics career; suitable for exceptional students who want a high caliber, broad based education suitable for bioinformatics, but also want to leave the door open to a wide range of other career options in the biological or computing sciences
Length: 4 Years

University of British Columbia and SFU:

Bioinformatics Training Program for Health Research - Diploma - inter-disciplinary program in bioinformatics; course of study individually tailored by each student; will allow students with backgrounds in biology, computer science, and statistics to acquire the relevant skills in each field and to work together to solve bioinformatics problems
Length: 1 year

Bioinformatics Training Program for Health Research - M.Sc. - inter-disciplinary program in bioinformatics; course of study individually tailored by each student; will allow students with backgrounds in biology, computer science, and statistics to acquire the relevant skills in each field and to work together to solve bioinformatics problems
Length: 2 years

Bioinformatics Training Program for Health Research - Ph.D. - inter-disciplinary program in bioinformatics; course of study individually tailored by each student; will allow students with backgrounds in biology, computer science, and statistics to acquire the relevant skills in each field and to work together to solve bioinformatics problems
Length: 3 - 4 years

Alberta

University of Alberta:

Bioinformatics B.Sc. Specialization in Biological Sciences - Bachelor of Science degree in biology with a specialization in bioinformatics.
Length: 4 years

Bioinformatics B.Sc. Specialization in Computing Science - Bachelor of Science degree in computing science with a specialization in bioinformatics.
Length: 4 years

University of Calgary:

Master of Biomedical Technology Program - Students acquire a broad range of skills and knowledge to prepare for a career in the biotechnology or pharmaceutical industry; graduates

will develop skills in bioinformatics and gain expertise in genomic analysis. Undergraduate degree required.
Length: 1 year

Bachelor of Health Sciences Program (BHSc): Bioinformatics major - A major in bioinformatics is available as part of the O'Brien Centre for the Bachelor of Health Sciences Program (BHSc) at the University of Calgary. The BHSc is a research-intensive interdisciplinary undergraduate program with a focus on health and health research. It is also possible to major in Biomedical Sciences or Health and Society.
Length: *4 years*

Saskatchewan

University of Saskatchewan:

Bioinformatics B.Sc. - Honours or regular 4-year degree available, with specialization in either Biochemistry or Computer Science. Additional streams in Biology and Probability & Statistics are forthcoming. This is an inter-disciplinary program providing an undergraduate education in the areas contributing to Bioinformatics *(Computer Science, Probability & Statistics, and the life sciences)*, plus a link between those areas through capstone Bioinformatics courses.
Length: *4 years*

Ontario

Carleton University:

B.Sc. (Honours) in Computational Biochemistry - In this program, you'll receive a strong background in core areas of biology and chemistry such as genetics, cell biology, organic chemistry, and analytical chemistry. You'll also take a variety of lecture and laboratory biochemistry courses, including bioinformatics. Optional courses allow you to focus on areas such as molecular genetics, pharmaceutic drug design, functional genomics, and protein structure and function.
Length: *4 years*

B.Sc. (Honours) in Computational Biology - In this program, a core of Biology courses provide a strong background in the fundamental disciplines of genetics, cell biology and biochemistry, plant and animal biology, and ecology. You can then specialize in one of two areas of computational biology: molecular bioinformatics or biodiversity.
Length: 4 years

Bachelor of Computer Science (BCS): Bioinformatics Stream - The Bioinformatics stream provides a general understanding of Biology, with a slant toward biochemistry, and specifically addresses some of the most interesting computational problems arising in Biology. The stream is geared toward those interested in a career as a computer scientist or software engineer in biotechnology, medical computing, or the life sciences in general.
Length: 4 years

Queen's University:

B.Sc. Honours - Biochemistry and Computing and Information Science - Provides the student with undergraduate training in the area of bioinformatics.
Length: 4 years

B.Sc. in Biomedical Computing - Provides an undergraduate education in computer science and life sciences; provides a link between these areas through specialized courses in medical informatics and computational biology.
Length: 4 years

B.Sc. Honours - Biology and Computing and Information Science - Degree consists of biology core program and a number of required computing courses
Length: 4 years

Seneca College:

Bioinformatics Post-Diploma Program (BIF) - Will provide specific training in the area of Bioinformatics with a focus on computer operating systems; programming and scripting; data storage, management, and analysis; computer applications used in this field; current molecular biology laboratory techniques. Three year college diploma or university degree required.
Length: *8 months*

University of Toronto:

Ph.D. Program in Proteomics and Bioinformatics - Multi-department and affiliated research institutes' academic program; will bring together researchers in biochemistry, genetics, cell biology, biophysics, chemistry, biotechnology, computational biology, and medicine. Only students in a Ph.D. program may apply.
Length: 5 years

B.Sc. (Honours) in Bioinformatics and Computational Biology - The B.Sc. (Honours) in Bioinformatics and Computational Biology is an interdepartmental Specialist Program, that is balanced between its foundational disciplines while covering advanced topics in both the theoretical and the life-sciences. The program draws on the University of Toronto's state-of-the-art facilities across the Departments of Biochemistry, Computer Science, Botany and Zoology; as well, it is firmly embedded in a comprehensive landscape of world-leading graduate and postgraduate research in one of the University's priority areas. This is a challenging new academic program that offers new opportunities to study in one of North America's leading Universities. Graduates will be well prepared to pursue graduate studies in any of the participating Departments, or to apply their skills in research-oriented industry positions.
Length: *4 years*

University of Waterloo:

Computer Science/Bioinformatics option (B.Math.) - Students primarily interested in computer science should select this route.
Length: *5 years*

Honours Bioinformatics option (B.Sc.) - Contains more biology content than the Honours Computer Science option.
Length: *5 years*

Honours Biology and Bioinformatics option (B.Sc.) - Contains more biology content than the Honours Computer Science option.
Length: *5 years*

Bioinformatics M.Sc. And Ph.D. Program - Training in bioinformatics at the graduate level.
Length:

Quebec

McGill University:

Undergraduate Minor in B.Sc. program in Computational Molecular Biology - Will provide undergraduate students in the biological sciences with the skills from Computer Science to solve computational problems arising in Molecular Biology and Genomics and to provide students with the necessary skills to build software tools from these algorithms.
Length: *24 credits to be completed during 4-year B.Sc.*

University of Montreal:

Ph.D. in Bioinformatics
Length: 4 years

M.Sc. in Bioinformatics
Length: 2 years

Baccalaureate in Bioinformatics
Length: 3 years

University of Sherbrooke:

BSc. in Bioinformatics
Length: 3 years

Challenges for the future

The following Challenges must be met if Canada needs to be competitive internationally in the field of Bioinformatics.

- There needs to be an increasing collaboration between the industry, the research and the government sector.
- More personnel need to be trained in the field.
- Sustaining and attracting key players in Canada.
- Providing assistance and support to the bioinformatics firms in Canada.

The following steps have been recommended to ensure Canada's success in Bioinformatics.

- Bioinformatics must be recognized as a vital science.
- The number of workshops to facilitate the professionals should be increased.
- Academic programs should be introduced in universities.

On the basis of these suggestions, support has been extended to the University of Montreal, University of New Brunswick, Queen's University and the University of Waterloo. In the 2000 federal budget, the Canadian government provided for a sizeable funding for the Research Chairs program, and the Canada Foundation for Innovation (CFI) program. Such initiatives will provide support to the bioinformatics programs. Moreover, collaboration between these universities and foreign universities should be encouraged.

- More students need to be encouraged to join the bioinformatics program by expanding awareness about these programs.
- Adequate funding needs to be provided to encourage new businesses in this field.
- A network, such as the proposed Canadian Bioinformatics Research Network (CBRN) that was initially declined, should be set up with the support of the government and the industry.
- Bioinformatics does not fit into any category of the Medical Research Council (MRC) and the Natural Sciences and Engineering Research Council (NSERC), who are responsible for providing funding for bioinformatics research. Industry, government and the universities need to ensure that these councils recognize bioinformatics as a vital area for more support.

It is imperative that the biotechnology industry expand to meet the growing demands for their products. Apart from three companies, the firms involved in biotechnology research are also involved in all other areas of biotechnology, including developing software for their own needs.

13 Bioinformatics Business and Education in Selected Countries

Apart form USA which is the hub of biological research as well as bioinformatics, other countries like India, Australia, Denmark and Japan are making headway in bioinformatics. India has been discussed in detail in the previous chapter. This chapter is devoted to bioinformatics infrastructure in other countries.

Australia

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 (BITS), The Computational Biology and bioinformatics Environment), Murdoch University (Centre 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.

- **Educational Institutes**

  The Government of Australia is devoting time and resources to development of its biotechnology sector. Currently, the application of bioinformatics tools is growing in two segments of biological research. One is bioinformatics related to genetic analysis and functional genomics, and the other is examination of biodiversity and analysis of species and phylogeny. While genomics bioinformatics in Australia is not widely distributed and there are only a limited number of centers with recognized skills in the discipline, biological bioinformatics *(species level)* is far more widely distributed at the commonwealth, state, university and private sector level.

Institutions such as Walter and Eliza Hall Institute for Medical Research, Institute for Molecular Bioscience, Australian National University and Monash University have a significant number of bioinformatics scientists, but in many cases these scientists are fully occupied by the work given by their employer. A number of undergraduate and postgraduate training courses have, and are, being developed at various universities, but these are rarely in the context of a critical mass of experienced scientists.

Attempts have been made to form regional consortia that assemble larger groups of scientists to undertake research and research training in bioinformatics. These include Victorian Bioinformatics Consortium and New South Wales Bioinformatics Consortium. However, there is no national framework to foster collaboration or address significant skill shortage in this discipline. Now Australia has set its eyes on a National Bioinformatics Strategy for comprehensive development of the sector.

The bioinformatics sector in Australia could continue to develop on its own without a national strategy. Some state governments have displayed a willingness to provide funding to establish bioinformatics institutes. However, a lack of coordinated intervention at a national and state level may not produce the increased capabilities and capacities needed.

Although bioinformatics is a critical biotechnological resource, Australia is affected by market failure in the bioinformatics area. Only a few bioinformatics companies operate in Australia. Very little public funding is available for bioinformatics. The Victorian Bioinformatics Consortium (VBC) conducts research in bioinformatics and assists the performance of biological research within Australia by both academic and industrial users. It plays an important role in developing Australia's skill base in bioinformatics by training postgraduate students and postdoctoral fellows. In Australia, most universities offer courses at graduate and, postgraduate level. Some universities also offer doctoral courses in bioinformatics. Some of leading universities offering bioinformatics courses are:

- University of South Australia
- Curtin University of Technology
- University of Melbourne
- University of Queensland
- National Tsing Hua University
- Latrobe University
- Flinders University
- University of Sydney
- James Cook University

Pharmaceuticals companies in Australia are using bioinformatics for identification of genes, correlation and comparison between sequences, and development of genome-based new drugs. Such companies include:

- Beckman Coulter Australia
- Biochemical Veterinary Research

- Bioinformatics
- Goodman Fielder
- Proteome Systems
- Sigma Aldrich

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 USS130 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 various public-funded research institutes, private sector companies and 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).

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, along with system integration services. A number of pharmaceutical majors also play a critical role in promoting bioinformatics in the country. They include Ajinomoto Co., Inc., Ono Pharmaceutical Co. Ltd., Taisho Pharmaceutical Co. Ltd. and Yamanouchi Pharmaceutical Co. Ltd., to name a few.

Since biotechnology-based drugs have good potential in the market, many pharmaceutical companies are busy in development of genome-based new life saving drugs and clinical trails. In new drug development at national institutes and university research centers, genomic DNA of many microorganisms has been sequenced for utility in pharmaceuticals. Japan government has recognized the need for more bioinformaticians and scaled up its genomics efforts. In 2002, Japan's Ministry of Education started upgrading bioinformatics education at national universities by creating additional staff positions and funding for both undergraduate courses and graduate-level informatics training program. A fund of USS2 million has been sanctioned for the development of bioinformatics, systems biology, protein functional analysis and software development. The following four institutions have received grants:

- Tokyo University
- Keio universities
- Nara Institute for Advanced Science and Technology
- Computational Biology Research Center at the National Institute of Advanced
- Industrial Science and Technology

Other national laboratories and universities involved in bioinformatics are:

- Kanehisa's Lab at Institute for Chemical Research

- Riken, the Institute of Physical and Chemical Research
- Bioinformatics Center, Kyoto University
- Computational Biology Research Center
- National Institute of Genetics
- Human Genome Center, University of Tokyo
- Institute for Advanced Biosciences, Keio University–DNA Data Bank of Japan (DDBJ)

Europe

Bioinformatics has attained maturity in Europe thanks to combined efforts of government, biological research institutes and private companies from Pharmaceuticals and medicine. The development of bioinformatics is more pronounced in Denmark. Currently, the Center for Biological Sequence (CBS) group is one of the largest academic centers in bioinformatics in Europe. The CBS Analysis at Technical University of Denmark was formed in 1993. In addition to a very strong research environment, the center possesses considerable teaching resources. In 2002, more than 300 students received training at CBS at doctoral and postgraduate level. This is good news for universities and start-ups starved of skilled bioinformatics manpower.

In 2001, University College London (UCL) created a separate budget for research activities in bioinformatics to benefit biological community in Europe. It collaborates with Sanger Center, UK Medical Research Council Human Genome Mapping Project Resource Center and European Molecular Biology Laboratory (EMBL). An international network of research institutes at Switzerland, it shares information between research institutes funded by EU countries. European Bioinformatics Institute (EBI) has built on this work to create a number of public domain databases covering gene and protein sequences, biological macromolecular structures and more recently, gene expression data. Presently bioinformatics is dealing with single molecules and single genes.

EMBL has sanctioned €8.4 million to EBI and intends increase this amount to €11.4 million to support expansion of research activities. Private entities such as Welcome Trust provide €2 million annually for Ensembl (the genome database) and fund macromolecular structure (MSD) database and training more people. The following outfits are actively involved in bioinformatics research and development:

- University College of London
- University of Denmark
- Southern Denmark University
- Odense University Hospital
- Department of Biochemistry and Molecular Biology
- Center for Proteome Analysis,
- Department of Mathematics & Computer Science

a) Public Institutes

- European Bioinformatics Institute
- The Sanger Center

- Human Genome Mapping Project Resource Center
- National Center for Biotechnology Information
- European Molecular Biology Laboratory
- Bloomsbury Center for Structural Biology
- Gene Ontology

b) Private Companies

- The Welcome Trust
- Inpharmatica
- Ace Biosciences
- Aros Applied Biotechnology
- Exiqon
- H. Lundbeck
- Maxygen
- MDS Proteomics
- Natimmune,
- NovoNordisk
- Novozymes
- Pride Proteomics
- Structural Bioinformatics
- Unizyme Laboratories

Bioinformatics in Denmark

Denmark was among the very few European countries who, at an early stage, invested massively in bioinformatics. The Center for Biological Sequence Analysis at the Technical University of Denmark is one of the largest academic centers in bioinformatics in Europe. In addition to a very strong research environment, the center has an extensive teaching capacity. As of 2002, more than 300 students received training at CBS at the PhD and Masters level. This means that there are good prospects for hiring people highly skilled in bioinformatics, which is a significant problem for start-up companies all over the world.

- Research in the fields of proteomics and genomics at the University of Southern Denmark relies heavily on bioinformatics.

- In 2002, an interdepartmental study program in bioinformatics was established, leading to a bachelor or masters degree in this interdisciplinary field.

- The University of Southern Denmark has received a grant from the Danish Center for Scientific Computing to establish a Beowulff-type computer cluster with approximately 500 processors.

- This computer facility is available to researchers at the University of Southern Denmark in the fields of scientific computing (chemistry, computer science, and engineering) and in biotechnology (biophysics)

→ Asia Pacific

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, clinicians, 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 5 behemoths.

Bioinformatics Education in United States

University of California Davis granted US$95 million for bioinformatics program. Virginia Polytechnic Institute is investing US$100 million in Virginia Bioinformatics Institute. University of Philadelphia, George Mason University and University of Florida also offer bioinformatics courses. They offer courses at graduate and postgraduate level. Some universities are offering Ph.D. level programs too. Some of the leading universities in US offering bioinformatics courses are:

- Rutgers University
- George Mason University
- University of New Mexico
- Washington University
- University of Pennsylvania
- Yale University

Besides universities, public research institutions and private companies are also actively involved in bioinformatics. These include:

- Schering-Plough Corp.
- Option Care
- Nektar Therapeutics
- Able Laboratories
- Akorn
- Maxygen
- Antigenics
- Caraco Pharmaceutical Laboratories
- Toxikon Corp.
- Dow Pharmaceutical Sciences

14 Bioinformatics Education in Pakistan

There are three educational institutes offering bioinformatics BS (Bachelors of Science) in Pakistan. These are:

1- COMSATS University, Islamabad
2- Mohammad Ali Jinnah University, Islamabad
3- International Islamic University, Islamabad *(Female only)*

Mohammad Ali Jinnah University (MAJU)

Mohammad Ali Jinnah University is a private sector University with programs in Information Technology, Engineering and Management Sciences in Islamabad.

MAJU has a permanent devoted faculty of 6 in Bioinformatics Department along with that there are some visiting faculty from various institutes in Islamabad and Rawalpindi.

Projects

Some of the projects carried out by MAJU students in bioinformatics are:

- Spliced Alignment through Dynamic Programming
- Multiple Sequence Alignment Clustering
- Identification of Protein Families through Hidden Markov Models
- MSA through Genetic Algorithms
- MSA through Simulated Annealing
- Clustering of Gene Expression Data

Collaborations

MAJU has been trying to forge collaborations with foreign bioinformatics companies including Shantha biotechnics Ltd Hyderabad

Faculty

- Professor Zafar Malik (Head of Dept.)
- Dr. Abdul Qadir
- Mr. Rashid Ali
- Mr. Nadeem Iftikhar

- Ms. Ayesha Fatima
- Dr. Sahar Fazal

Course Outline

Course Outline of MAJU and other Bioinformatics Programs is included as appendix 1.

COMSATS Bioinformatics BS

COMSATS Islamabad offers 4 year Bachelors in Bioinformatics. The institute is in the process of setting up a faculty of Biosciences and supports a large number of faculty members trained in biology.

Early this year COMSATS Program invited visiting faculty from George Mason University to facilitate in the improvement of its programs.

Faculty

- Dr. Shahzad Mufti
- Dr. Qamar Javed (Chairman)
- Dr. Habib Bokhari
- Dr. Rani Faryal
- Dr. Muhammad Ansar
- Dr. Mahmood A Kayani
- Dr. Muhammad Arshad Rafique
- Dr. Raheel Qamar
- Dr. Syed Sarfraz Hussain
- Dr. Asifa Ahmed
- Ms. Irum Qureshi
- Mr. Waseem Haider
- Mr. Jadoon Khan
- Mr. Syed Kumail Ali Rizvi
- Mr. Alamdar Hussain
- Mr. Zahid-ur-Rehman
- Mr. Suliaman Faisal
- Ms. Nadia Anwer

Research Interests

Comsats supports several wet laboratories within its campus. The course outline is provided in Appendix 1.

Other Institutes

In addition to BS programmes in Bioinformatics, Baluchistan University of Information Technology and Management Sciences is offering undergraduate and graduate programs in

biotechnology and Informatics. A few institutes are offering short courses in bioinformatics as part of Masters Courses in biology and associated fields. These include

- University of Arid Agriculture, Rawalpindi,
- University of Karachi

The focus of these courses is to introduce biologists to the bioinformatics tools they are likely to encounter in their research. However, a thorough knowledge of the mechanics behind softwares, trouble-shooting and development of new tools is beyond their training and expertise.

Baluchistan University of Information Technology and Management Sciences BS in Biotechnology and Informatics

The faculty of Biotechnology and Informatics was started in August 2003 with the support of HEC. Following programs are being offered,

- BS (biotechnology and Informatics)- 4 yrs program
- MS (biotechnology and Informatics)- 2 yrs program
- PGD (biotechnology and Informatics)- 1 yr program
- PhD (biotechnology and Informatics)- 3 yrs program

The course outline of the BS program is given in Appendix 1. The main aim of the program is to produce biotechnologists/ biologists with expertise in using bioinformatics tools. Thus the stress of the program is on core biology with a smattering of information technology. The institute claims collaboration with all major research institutes in the country. The program has bold goals such as developing computer models for bioproteins, collecting and organizing databases, providing consultancy to industry and developing drugs from natural compounds.

Additional information about the program can be obtained from
http://www.buitms.edu.pk/bio_program.htm

15 Possible Partners in Bioinformatics Business

Bioinformatics being an interdisciplinary field will require collaboration between biologists, IT professionals and managers. One important component of this partnership will be research scientists working in Genetics, Biotechnology, and Proteomics etc. Every major state university in Pakistan harbors a biology department, however research work in limited to a smaller number, furthermore, excellence and innovation is limited to a few individuals and institutions. Some of the prominent research institutes in Biology and Medicine might be explored as possible partners in bioinformatics business, such as,

1. HEJ Research Institute of Chemistry, Karachi
2. Panjwani Centre for Molecular Medicine and Drug Research, Karachi
3. Centre for Excellence in Molecular Biology, Lahore
4. NIBGE, Faisalabad
5. Aga Khan University, Karachi

HEJ (Hussain Ebrahim Jamal) Research Institute of Chemistry, Karachi University[1]

The H.E.J. Research Institute of Chemistry was established under the Directorship of late Prof. Salimuzzaman Siddiqui, in 1967 at University of Karachi. Emphasis of the institute has been on research and training in economically relevant fields of chemical and biochemical sciences and resource mobilization for the future growth in new disciplines of science.

- **Area of Research**

  The main areas of research and development and training of students include natural product chemistry, protein chemistry, pharmacology and plant biotechnology. The areas of research covered in the programs of the institute broadly relate to isolation, structural, synthetic and pharmacological studies on novel natural products as well as various aspects of protein chemistry.

- **Academic Programs**

  The Institute has the single largest doctoral program in the country with over 220 Ph. D. students. A number of goal-oriented projects relating to the chemistry of natural

[1] Information gathered by face to face interviews with faculty members and from the official website of HEJ Institute

products, pharmacology of herbal medicines and protein chemistry are being vigorously pursued which have led to the award of over 130 doctorate degrees, 35 M. Phil. degrees and 45 M.Sc. degrees and the publication of over 1200 research papers that have earned international recognition.

The analytical, spectroscopic and other facilities present in the institute are at par to any good Institution in the West.

- **Interaction with Industry**

  The Institute also has an Industrial Analytical Center, which is providing analytical and consultancy services to over 200 industries in Pakistan. The scientists trained in the institute are now serving the country in industry and in various R & D and academic institutes.

Panjwani Center for Molecular Medicine and Drug Research (PCMD), Karachi University

The Panjwani Center for Molecular Medicine and Drug Research (PCMD) has recently been founded through a generous support and patronage of Ms Nadira Panjwani and family. Along with HEJ Institute, it is one of the institutes affiliated with Department of Chemistry at Karachi University. It is hoped that PCMD will initiate scientific programs that lead to possible treatments of often neglected prevalent diseases of Pakistan, and help in developing scientific human resource in the country.

- **Area of Research**

  The main areas of research of the institute are the emerging fields of molecular medicine and drug development. The academicians, clinicians and pharmaceutical researchers would be brought together to translate basic scientific discoveries into new therapies, vaccines and diagnostic tests. The efforts in the center will mainly focus on developing greater and more comprehensive understanding of the pattern and causes of most common diseases of Pakistan in order to develop effective diagnostic tools and affordable treatments by using molecular medicinal, chemical and computational methods.

  The mission of the center is to foster excellence and achieve international recognition in the targeted areas of research, which are relevant to national needs. To accomplish this mission, the center will be involved in:

  a) Enhancing the quality and quantity of scientific training of graduate students in the area of molecular medicine and rational drug designing.

  b) Focusing on the development of new and effective diagnostic, and treatment techniques for prevalent diseases.

c) Providing forum for information exchange between academic disciplines and raising general awareness about the diseases and their prevention, cure and treatments.

- **Faculty at HEJ and PCMD**

Prof. Dr. Atta-ur- Rahman	Bio-organic and Natural Product Chemistry
Prof. Dr. M. Iqbal Choudhary	Bio-organic, Structural Organic Chemistry
Prof. Dr. Amin Suria	Neuropharmacologist
Prof. Dr. Sheikh Arshad Saeed	Pharmacologist
Prof. Dr. Muhammad Anwar Waqar	Cell & Molecular Biology
Prof. Ahsana Dar	Pharmacology
Dr. Kamran Azim	Biochemisty
Dr. Shabana U. Simjee	Neuropharmacology
Dr. Zaheer -ul-Haq	Organic Chemistry
Dr. Shakila Khandwala	Microbiology /Immunology
Prof. Viqar Uddin Ahmad	Organic Chemistry
Prof. Bina S. Siddiqui	Organic Chemistry
Prof. Dr. Abdul Malik	OrganicChemistry
Prof. Dr. Atiya Abbasi	Biochemistry
Prof. Dr. Shaheen Faizi	Organic Chemistry
Dr. Sabira Begum	Organic Chemistry
Dr. Muhammad Shaiq Ali	Bioorganic Chemistry
Dr. Saifullah Khan	Plant Biotechnology
Dr. Farzana Shaheen	Organic Chemistry
Dr. M. Raza Shah	Synthetic Organic Chemistry
Dr. Shabana U. Simjee	Neuropharmacology
Dr. Shazia Anjum	OrganicChemistry

- **Drug Discovery Research**

Several faculty members in HEJ and Panjwani Centre are involved in drug discovery, screening of natural compounds, discovery of new therapeutics using a combination of in silico and medicinal chemistry methods, medicinal chemistry, molecular modeling, computational chemistry, computer-aided drug design, bioinformatics related to drug design, virtual screening *(docking. scoring. 3D-QSAR; CoMFA, COMSIA),* virtual combinatorial library design, molecular dynamics (MD) simulations and abinitio calculations, enzyme inhibitors etc

HEJ and PCDM are potential foci of expertise that can be accessed to develop Drug discovery research in Pakistan on commercial scale.

NIBGE Faisalabad

NIBGE *(National Institute for Biotechnology and Genetic Engineering)* located in Faisalabad is the model institute in biotechnology that aims to focus on agricultural, industrial,

environmental and healthcare problems relevant to Pakistan. In the last 10 years since its establishment it has developed a number of technologies and products. *(list given below)*

The institute is affiliated with Pakistan Atomic Energy Commission (PAEC) and with Quaid-e-Azam University, Islamabad for its educational programs. It is presently running M.Phil and Ph.D degree programs.
The institute is divided into the following 6 divisions.

1) Plant Biotechnology Division
2) Bio-fertilizer
3) Health Biotechnology Division
4) Industrial Biotechnology
5) Bioprocess Biotechnology
6) Environment Biotechnology

NIBGE claims several collaborative ventures with industry which have facilitated the development of Practical working models in diversified areas viz, development of microbial inocula for various industries, up scaling of coal bio-desulphurization process for cement industry etc.

NIBGE also took leading role in the oil spill on Karachi beach in 2003, cleaning it through bio-mediation using oil degrading bacteria. The institute has also developed several advanced lines of virus/insect resistant cotton which will be ready for commercial use after meeting the requirements of bio-safety guidelines.

The genetic sequencing facility of the institute has successfully brought to light "The Burewala virus" that played havoc with cotton crop in certain areas in Pakistan. Current stress is on introducing genes for salt tolerance and other biotic stresses in rice, sugarcane, potato and tomato. RNAi technology is being used for this purpose.

The institute also has a group working on bio-fertilizers. Bio-fertilizer Resource centre (BIRCEN) has been established through a grant from Islamic Development Bank (IDB). In the field of health care and management, NIBGE offers PCR based diagnostic tests for various diseases. The work on development of diagnostic kits on local scale is said to be in fairly advanced stage. Other activities being carried out involve genotyping, for Thalasemia, Karyotyping, and Molecular diagnosis for genetic disorder and consultancy services. In Health sector research is also being conducted on diabetes, viral hepatitis and typhoid.

NIBGE has also established facilities for testing of genetically modified organisms (GMOS). Industrial products include production of industrial enzymes, edible acids and ethers from agro-industrial wastes.

It is claimed that a Bio-informatics cell has been established at NIBGE, however it is concerned primarily with providing software support and library assistance to scientists.

More than 100 microbial strains have been isolated from various extreme environment of Pakistan. Fifty two genes of extremophiles *(organisms living in extreme conditions)* have

been deposited to Gen Bank as a public Resource Domain. It is hoped that this information will help in the production of crops able to resist harsh conditions

- **Patents and Commercial Products**

  During the period b/w 2001-2003 2 patents have been filed by NIBGE and more than four technologies were commercialized. The patents include NIBGE-I and NIBGE-2 cotton variety of this institute. Insect resistant cotton *(IR-cotton)* varieties *(IR-FH-901, IR-CIM-448, IR-NIBGE-I and IR-CIM-443).*

- **Commercialization of Biotechnology**

  To commercialize the fruits of its research, a biotech marketing company has been established in NIBGE in 1995 called Pakistan Innovative Biotechnology Services (PIBS). The mandate of PIBS is to expedite the commercialization of technologies applied in various sectors of the industries and agriculture field, which might have a direct or indirect impact on the society.

  The main objectives of PIBS are:

  1- To commercialize the processes developed by the application of biotechnology by Genetic Engineering techniques.

  2- To provide technical services, consultancy, training, transfer of technology to industries, farmers, laboratories, institutions and all others engaged in the development of products, processes, services, related to or based on biotechnology and genetic engineering.

  3- To provide research and Development facilities or services, to the industries, laboratories, hospital, research institution or farmer to improve, discover, invent or produce the products, processes and equipment based on or related to Biotechnology and Genetic Engineering.

  In 2002-2003 PIBS reported a humble profit of 1.4 Million Rupees. With plants, Bioprocesses and Biofertilizers taking almost equal share of the profit.

Important products of PIBS are:

- Bio-power – Sold 23,000 bags in 2003
- Virus free seed of potato – 75,000 mini-tubers sold in 2002-2003
- Elite cotton seed – 2925.2 kg in 2003

Additional information can be obtained on http://www.nibge.org/

Centre for Excellence in Molecular Biology (CEMB) Punjab University, Lahore

CEMB located in Lahore, is affiliated with Punjab University. The institute has been working for the past 2 decades. One of the main aims of the institute is teaching and training to

generate a cadre of students specifically trained in molecular biology and recombinant DNA technology. M.Phil. and Ph. D programs are being offered by the institute at the moment. The institute receives funding from National Granting Agencies as well as US State Department, National Science Foundation and International Centre for Genetic Engineering and Biotechnology.

- **Areas of Research**

  The institute aims to undertake goal oriented molecular biological research on specific problems related to economic needs of the country, in agriculture, health & medicine, industrial, energy and environmental sectors. It also plans to create a repository of DNA modifying enzymes, DNA cloning vectors, novel bacterial strains and other such molecular tools for ready availability and use by various research groups at this centre and other DNA research laboratories in Pakistan.

  The institute is divided into 4 divisions

  1. Basis Molecular Biology,
  2. Plant Molecular Biology or Agriculture,
  3. Medical Molecular Biology and
  4. Industry, Energy and Environment

  The institute has produced Insect resistant rice and cotton in the past 6 years. Further research is going on the production of virus free seeds of vegetables, sugarcane, banana and gladiolas *(floral productions)*

  In the field of health the following projects are being carried out in the institute:

  - ❖ DNA typing and its utility in paternity and forensics.
  - ❖ Diagnosis of hepatitis and tuberculosis
  - ❖ Production of pharmaceutical proteins.
  - ❖ Stem cell research

- **Future Commercial Prospects**

  The institute aims to produce recombinant Interferon, vaccines for hepatitis and protein modeling and gene expression services.

- **Bioinformatics Manpower Development**

  CEMB is organizing a one week international workshop on bioinformatics with the help of International Society for Computational Biology (ISCB) on November 14, 2005, which will help educate Pakistani biologists about the field.

More information about the workshop can be obtained on
http://www.iscb.org/events/event_data.php?333

- **Senior Faculty**

  It enjoys a contingent of 45 professionals, including 10-12 Ph.d and 3 Foreign Faculty professors.

  - Dr. S. Riazuddin — Professor Director
  - Dr. Zahoor Ahmad — Associate Professor
  - Dr. Anjum Suhail — Associate Professor
  - Dr. Tayyab Husnain — Associate Professor
  - Dr. Bushra Chaudhry — Associate Professor

Additional Information can be obtained from http://www.cemb.edu.pk

Aga Khan University (AKU) Karachi

Aga Khan University was chartered in 1983 as Pakistan's first private university. AKU's objective is to promote human welfare in general, and the welfare of the people of Pakistan in particular, by disseminating knowledge and providing instruction, training, research and service in the health sciences, education and such other branches of learning.

The University consists of the Medical College, School of Nursing and Teaching Hospital, Institute for Educational Development and Institute for the Study of Muslim Civilizations.

Active research is being done in various areas of basic and applied biology, clinical and community health sciences. The university hospital is the pioneer in implementing a electronic patient records system. Currently there are 50 programmers rich Software Development Wing in the University, which is responsible for customizing hospital management systems (MAGIC EMR) for its hospitals in Karachi, Kenya and Tanzania. The group also aims at commercially marketing this product in the future.

With an eye on new technologies and extensive research and clinical faculty, the university is a potential suitable site for entrepreneurship in bioinformatics. A list of research projects being undertaken by the faculty is provided in Appendix 5.

More information about the research group at AKU can be obtained from http://www.aku.edu/medicalcollege/bbs/

16 Use of Agricultural Biotechnology for Pakistan

Market Scenario

The first major GM crop, Roundup Ready Soybeans *(glyphosphate-tolerant),* was launched by Monsanto in 1996. Since then the global market for GM/transgenic crop products grew rapidly from 1995 to 1999. Global sales were estimated at $75 million in 1995, reaching $1.6 billion in 1998, and increased to an estimated $2.1-2.3 billion in 1999, according to the International Service for the Acquisition of Agri-biotech Applications *(ISAAA).* The global market for the GM crops is projected to reach approximately $8bn in 2005 and $25bn in 2010.

In 1999, the global area of GM crops increased by 44 per cent or 12.1 million hectares from 27.8 million hectares in 1998 to 39.9 million hectares, according to ISAAA.

Prospects

The world population has topped 6 billion people and it is predicted to double in the next 50 years. Ensuring an adequate food supply for this booming population is going to be a major challenge in the years to come. GM foods promise to meet this need in a number of ways:

- **Pest resistance:** Growing GM foods such as Bt-cotton and Bt-corn can help eliminate the application of chemical pesticides and reduce the cost of bringing a crop to market.

- **Herbicide tolerance:** For some crops, it is not cost-effective to remove weeds by physical means such as tilling, so farmers will often spray large quantities of different herbicides to destroy weeds, a time-consuming and expensive process, that requires care so that the herbicide doesn't harm the crop plant or the environment. Crop plants genetically engineered require one application of weed-killer instead of multiple applications, reducing production cost and limiting the dangers of agricultural waste run-off.

- **Disease resistance:** There are many viruses, fungi and bacteria, which cause plant diseases. Plant biologists are working to create plants with genetically engineered resistance to these diseases.

- **Cold tolerance:** Unexpected frost can destroy sensitive seedlings. An antifreeze gene from cold-water fish has been introduced into plants such as tobacco, potato and strawberries.

- **Drought/salinity tolerance:** Creating plants that can withstand long periods of drought or high salt content in soil and groundwater help people to grow crops in formerly inhospitable places.

- **Phytoremediation:** Soil and groundwater pollution continues to be a problem in all parts of the world. Plants such as poplar trees have been genetically engineered to clean up heavy metal pollution from contaminated soil.

Conclusion

Biotechnology is making it possible for researchers and developers to deliver products that help farmers protect their crops; and improve the economy and environment while grow grains that improve the quality of the food we eat. Biotechnology will enhance quality of life in many ways, while helping the environment by reducing our dependence on non-renewable resources. But that's just the beginning. We have to understand the importance GM and its role and influence on our future growth, health and environment.

17 Results of the Survey of Stake Holders

The sample consists of 4 groups:

Group # 1	Information Technology Companies	(n=10)
Group # 2	Scientists working in biological sciences/ medicine	(n=20)
Group # 3	Educational Institutions imparting Bioinformatics Education at various levels	(n=7)
Group # 4	Pharmaceutical Companies	(n=3)

RESULTS

Group # 1
Survey of IT Companies

Ten people from 9 IT companies in Islamabad, Lahore and Karachi were interviewed.

Human Resource

Size of IT Companies

Size (Manpower)	Freq
Less than 50	4
50 – 100	2
101 – 300	2
> 300	2
Total	**10**

Familiarity with Bioinformatics

Are you aware of Bioinformatics

Response	Freq	Percentage
No	2	20
To some extent	3	30
Familiar	4	40
No response	1	10
Total	**10**	**100**

Projects in Bioinformatics

Q. Has your company taken up a project in bioinformatics?

In response to this question a variety of projects were quoted, none fitted exactly with the definition of a bioinformatics project.

Response	Freq
No	4
Agriculture Management System with Arid University	1
Telemedicine solutions	2
Medical Informatics	4
Biometrics	2
Radiology Storage	1
Total	**14**

* Multiple Responses

Willingness to venture into bioinformatics

Q. Would you be willing to venture into bioinformatics?

Response	Freq
No	2
Difficult field extensive expertise needed	2
Unsure	1
Would like to go into medical informatics	1
Yes	4
Total	**10**

Difficulties in entering the field

Q. Difficulties that you see in venturing into bioinformatics?

Response	Freq
Out of our area of expertise / lack of experts	8
Finances are limited	4
No R and D funds	3
Market is unknown	2
User unaware that solution exists	1
Lack of leadership vision	2
Intense competition	1
Field is in infancy	1
Total	**22**

Threats and Barriers

Q. What difficulties do you see for Pakistan in this field?

Response	Freq
Lack of R&D investment, Lack of Research	6
Lack of expertise	4
No prior example, new in the field	2
Lack of local biotech / knowledge based industries	2
Intense competition	1
Political issues / country has a low tech image	2
No focus on the field of bioinformatics	1
Total	**18**

Suggestions to start bioinformatics business in Pakistan

Response	Freq
Start bioinformatics education short courses	3
Stress on R and D	1
School education should be improved first	2
Public private partnership	2
Eliminate red tape	1
Awareness on part of IT people	1
Partnership with international groups/ companies	1
Develop biotech industry	1
Eliminate corruption, improve country's image	2
Total	**15**

General Suggestions

A variety of general suggestions were given by the IT group of the sample. These have been classified into broad categories.

- **Education:**

1. Improve IT education
2. Start BS/MS in bioinformatics
3. School education is highly substandard
4. Short courses in bioinformatics
5. Human resource development

- **Policy Issues:**

1. Private public partnership should be encouraged
2. Govt. protection should be provided like in the case of car manufactures
3. Reduce import duty
4. Tax rebates
5. Provide R and D loans
6. Severely disappointed with government
7. Eliminate red tape
8. Improve accountability

- **Bioinformatics:**

1. Bring the field to people's attention.
2. Help gain loans from banks since softwares are not considered collaterals by banks.
3. IT ministry and PSEB should be improved further.
4. Policies should be consistent and made by professionals.
5. Local software use should be promoted.
6. Invest in research.

Group # 2
BIOLOGICAL RESEARCH GROUP

A list of the respondents and their institutions is attached.

- **Major Focus of Research:**

The major focus of research of the respondents is given:

Agriculture / Environment / Life Stock	
Agriculture biotechnology	2
Biofertilizers	1
Marine animals	1
Plants and agriculture	1
Molecular Biology / Pure Science	
Biotechnology and Genetic engineering	1
Electron microscopy	1
Molecular biology	3
Protein interacting, identification of proteins	1
Medicine	
Genetic of complex diseases	1
Human Genetics	1
Health biotech	1
Hearing System Research	1
Regenerative Neurobiology	1
Pharmacology / Drug Development	
Cardiovascular Pharmacology	1
Natural products research	1
Others	
Education providers	1
No Response	1
Total	**20**

- **Funding:**

Different agencies funding projects being undertaken by the respondents were:

HEC / Govt. funding (unspecified)	8
Pakistan Agricultural Research Council	4
Aga Khan university Intramural funds	10
Karachi University / HEJ	3
International funds NIH (USA) French Embassy	6
Others	7
Pakistan Science Foundation	3

This list would not be representative since project funding is highly dependent on the institutions. *(A complete list of projects funded by HEC is given in Appendix 3)*

Software Use

Different softwares used by the respondents are:

- BLAST / Sequence Analysis
- SPSS
- Prime Design
- Gene Runner
- Matlab
- Restriction Analysis
- Modler
- ORF reader
- Thesis
- Cytoscape

(A list of some bioinformatics tools and softwares is given in appendix 6.)

Of the different softwares cited about 30 were freely available and only 4 were subscribed.

Research Interests and Use of Bioinformatics Software

Name	Research Interest	Software	Utilization
Zeeshan Ozair, AKU	Regenerative Neurobiology	P-Draw II Gene Runner	Plasmid design
Muhammad Shoaib, Ph.D Candidate, Australia	Bioinformatics Research	Matlab Weka S and R BRB SAM Cytoscape Octave and Sclab	Classification, Gene Selection and Expression Analysis Classification Statistical Analysis Classification Gene Selection Gene Networking Classification, Gene Selection and Expression Analysis
Anwar Siddiqui, AKU	Role of Proteins in Human Diseases	Pop Star Bionumeric Philip	Phylogenetic Analysis 2D gel electrophoresis and Genotyping Phylogenetic Analysis
Syed Ali, AKU	Viral Genetics	Sequentree P Draw 32 Amplify	Sequence Analysis PCR Analysis
Danish Saleheen, AKU	Human Genetics	Sequentier Thesis Chromas Gene Runner	Sequence Analysis Haplotype Analysis Sequence Analysis Primer design
Jaffar Hussain, NIBGE	Genetics	ORF-Finder Cluster W/X Restriction Summary	Protein coding regions Multiple sequence alignment Restriction Analysis
Attiya Abbasi, HEJ	Protein Inhibitors	Modler Wetlab Pro Tree View Whatif Autodock	Homology search, modeling, phylogeny Cancer Apoptosis Modeling Modeling Protein Interaction
Rumina Qazi, AKU	DNA based diagnosis	DNA Star	PCR Analysis
Waseem Ahmed, QAU	Human Genetics	Graic M Link Alegro	DNA Analysis and Simulation

The research interests of these researchers are given below:

Research interests of selected survey respondents

Name	Research Projects
Zeeshan Ozair, AKU	Isolation, telomerase immortalization and characterization of a fetal human neural stem cell line. Ex-vivo gene therapy of NPC-1
Muhammad Shoaib, Ph.D Candidate, Australia	Gene Networks, Gene Selection, Cancer Diagnosis, Imputation and Microarray data processing
Anwar Siddiqui, AKU	Hepatitis C and B Genotyping, Viral Transport in Blood Role of proteins in renal stones and gallstone formation
Syed Ali, AKU	Mutation of genes in Breast CA Molecular oncology of Human Papilloma Virus (HPV) Molecular Epidemiology of HPV and HIV Screening of antiviral drugs
Danish Saleheen, AKU	Genetics of Hypolipoproteinemia, Stroke, Hypertension, Muscular Dystrophy
Attiya Abbasi, HEJ	Isolation, Characterization of proteases, amylases, lipases from microorganisms and plants Medicinal plants-biologically active proteins for clinical purposes
Rumina Qazi, AKU	Diagnosis of pathogens in various tuberculosis Expression of REST4 in various cancers
Waseem Ahmed, QAU	Human Genetics

Satisfaction with Freely available Softwares

Four out of the 20 respondents said they were happy with their software use and did not need any improvements. 2 thought they needed customized softwares.

IT professionals in Biological Research Organizations

Only 5 of the 20 organizations / Labs had information technology support staff.

Utilization of bioinformatics professionals

In response to the question, Do you plan to incorporate more IT support in your research?

14 responded with a Yes, 4 were not sure only 1 said they were not needed.

Yes	14	70%
No	1	5%
Maybe	4	20%
No Response	1	5%

Sharing of Information / Resources

Q. Can you identify areas where sharing of information / data / services could be used in carrying out research ?

All Areas:

All areas need collaboration	5

Plants / Agriculture / Life Stock / Fishing:

Plant classification	2
Pest control	1
Fishing	1

Genetics / Molecular Biology:

Genome wide scan	1
Proteomics database	1
Genetics information database	1
Germs / plasmids database	
Online data analysis facilities	1
DNA sequences / sharing lab equipment	1
Molecular epidemiology	1
Biotechnology Knowledge	1

Medicine:

Epidemiological and clinical studies repository	3
Infectious disease incidence database in the country	1

Others:

Better access to international forums	1
Better professional / research networking	2
Toxicology	1
NR	4

Commercialization of resources / services

10 out of 20 responds thought they had services or lab equipment that can be marketed / offered to other research groups with commercial motive.

Business Opportunities for Pakistan IT / Bioinformatics Industry In Local or Global Market

	Frequency
Software Development	
Software development in genome analysis	2
Educational softwares and online materials	1
Automation of data gathering	1
Data storage softwares	1
Software development	2
Image Analysis Softwares	
Databases	
Database of all hospitals in Pakistan	1
Epidemiological and clinical studies	1
Clinical Studies	
Central database for all Pakistani hospitals	3
Epidemiological and clinical studies	1
Data Analysis of Clinical Trials	
Knowledge sharing with Industry	
Fish farming	1
Toxicology studies in food	1
Virus inactivation in agriculture	1
Services	
Statistical services	1
Proteomics/drug development	2
Image analysis	1
Drug receptor interactions	1
Translational Research	
Development of products from research	2
Vaccines development	1
Others	
All areas need IT support	1
None/ Not Feasible	
Too much competition	1
None	1
Don't know	1

Group # 3
Educational Institutions

Seven respondents identified themselves as educators in bioinformatics. Four of these were associated with bioinformatics bachelors of Science programs, while one was running one year degree program. The other two although not currently running bioinformatics programs formally were associated with bioinformatics education in some way.

Program strengths of Bioinformatics Programs

The approximate strengths of bioinformatics programs is given below:

	Strength
COMSATS BS	130 -140
MAJU BS	45
IIUI BS	90
Karachi University	26

Problems faced in the program

A number of problems were identified that are presently being faced by a bioinformatics program or the problem that are seen in starting a bioinformatics educational program in Pakistan.

Faculty / Course Designing

Response	Freq
Faculty shortage / Faculty not appropriate	4
Designing course outline according to demand	1
Lack of interdisciplinary people	3
No experts in PerL, Linux languages	1
Total	**9**

Infrastructure / Resources

Response	Freq
Lace of infrastructure / resources	2
Softwares are expensive / Difficult to obtain	1
Lack of interest / support by industry	1
Total	**4**

Job Prospects

Response	Freq
Job prospects are uncertain / not perceived by students	2
Total	**2**

Quality of Students

Response	Freq
Biology students lack mathematics background	1
No Response	1
Total	**2**

Utilization of Bioinformatics Graduates

What would be the utilization of Pakistani bioinformatics graduates?

Area	Frequency
Drug/ Vaccine Development	2
Pharmaceutical Industry	4
Proteomics	2
Genomic Analysis	2
Medical Informatics	2
Software Development and Implementation	3
Agricultural Research	1
Faculty in Bioinformatics Programs	3
Research Organizations	3
Assistance to Scientists	1
Research in bioinformatics	1
Utilization in translation of research for Industry	1
Can work in other informatics systems	1

What steps should be taken by Pakistan to develop bioinformatics business?

1. "Private sector development in software"
2. "Proper allocation of funds for education"
3. "Create jobs to attract students"
4. "Create awareness in industry"
5. "Interdisciplinary contact"
6. "Appropriate method for Bioinformatics research grants review"
7. "Increase funding for bioinformatics"
8. "Involvement of IT wing for International collaboration, advertising, promotion"
9. "Collaborate with International Companies for enzyme development/ analysis as a start".
10. "Improve research"
11. "Faculty development"
12. "Involvement of industry to device useful courses"
13. "Awareness on part of Govt. sector"
14. "International collaboration for student exchange"
15. "Introductory courses in all related institutes"
16. "Govt. support for private sector as well"
17. "Analysis of already existing biological research data"
18. "Send people abroad to conferences"
19. "Talk to biologists/offer solutions to the clients/market yourself".

Group # 4
Pharmaceutical Companies

Many pharmaceutical companies were contacted from a list provided by Pakistan Pharmaceutical Association. However the response rate was low in this group. Only 3 respondents could be interviewed completely. Many refused to be interviewed on the pretext that they did not have information or that their companies do not have a research wing in the country.

18 Opportunities and Threats

Opportunities in Bioinformatics

The opportunity in bioinformatics for Pakistan stems from its large population with a potential for skilled human resources. Also there is a fair amount of biological research infrastructure present in the country, although it is underutilized and poorly managed. Pakistan owing to its diverse and large population can serve as a good model for several disease and drug development/ clinical trials studies. Several pockets of expertise exist in several high technology areas in Pakistan, most of this is not industry oriented hence has little impact. Bioinformatics business on one hand can help local biological research to develop products and translate research into solutions and on the other hand can provide revenue from outsource business.

Threats and Barriers

The road to the above mentioned opportunities is however fraught with barriers and threats. Some of which are lack of trained manpower, need for more industry-academia partnership, lack of regulatory environment and lack of proper intellectual property protection.

Forming the right partnerships

Most of Pakistan's biological research is taking place in public sector university labs and institutes, which do not have any experience or propensity towards collaboration with industry. To kick-start such collaboration, would require substantial efforts on the part of all stakeholders in bioinformatics.

As the first example is set others will follow. It is interesting to note that the same problem would have been faced by India as it endeavored to enter into bioinformatics business. The following case study might be helpful to understand the problem and offer solution.

India's Tata Consultancy Services (TCS) has emerged as Asia's largest software company to branch out into bioinformatics. In August 2001, TCS signed an agreement with the DBT funded Center for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, for training and research in bioinformatics designed to help TCS gain expertise in bioinformatics and allow CDFD generate intellectual property through TCS-funded projects. TCS will recruit 50 to 75 engineers and post-graduate scientists over the next year. They will be trained in molecular biology and bioinformatics at the CDFD to prepare them for contract research for leading biotechnology and pharmaceutical companies worldwide.

TCS also hopes to work with CDFD to develop specialized software products in niche areas of bioinformatics. The TCS-CDFD agreement also envisages joint research by CDFD and the TCS Advanced Technology Center, Hyderabad. It is believed that the TCS has committed at least US$1.5M for the next three years on this initiative, but has indicated that it is willing to invest more if required. TCS will fund long-term R&D activities at CDFD in areas such as protein structures determination, delineation of genes and regulatory elements in genomes, identification of functionally important motifs in proteins and functional relevance of single nucleotide polymorphisms in human populations. TCS and CDFD hope to generate intellectual property within the next two to three years.

Such partnerships between public sector research institutes and national or international commercial companies would be essential for the development of bioinformatics in Pakistan.

Absence of indigenous biotech industry

Pakistan lacks a substantial biotechnology industry within the country, hence for a nascent bioinformatics company outsource might be the only business in the beginning.

Biological research is obscure, irrelevant to industry

There is a general lack of industry-oriented research in Pakistani academia. Furthermore there is little collaboration between different research groups. Research findings are often obscured by lack of proper media to propagate it. A multidisciplinary field such as bioinformatics, which has its utilization in knowledge-based industries, can only develop gradually in such an environment.

Scarcity of Human Resource

There is a severe shortage of human resource in bioinformatics. To begin with there are very few educational programs in bioinformatics, furthermore, these programs are marred by lack of proper resources and shortage of appropriate faculty. The total output of bioinformatics bachelors' degree holders in the country is less than a hundred.

Services? or product development?

Bioinformatics business can be divided into two types, product development and offering services. These two approaches are captured in the following statements.

Anuradha Acharya, CEO, Ocimum Biosolutions, says," bioinformatics is crucial for the advancement of the biotech industry since it can help to tremendously cut the time frame that is usually required to develop a product. It also helps in bringing about standardization and discipline in the field, by automating some tasks and introducing checks in the process. Since it has become standard practice in most countries, those that don't follow it will find it difficult to keep pace with developments in this area."

Kiran Mazumdar Shaw, chairperson and managing director, Biocon India Group says, "bioinformatics is a tool. People should not confuse bio-informatics as an end in itself. It is a very important tool, just like how software is a tool. It is service business. Use bio

informatics to mine a lot of knowledge. Create types of new hypothesis and to have a very interesting interpretation of databases of genomics and proteomics. Bioinformatics has a great potential in the service industry."

In the presence of freely available softwares in bioinformatics, product development is a highly competitive area. There are few pure-play bioinformatics companies in the world now. Many have merged with biotech or pharmaceutical companies and are now offering services or complete drug discovery and other solutions, from target identification to production and clinical trials. Few companies like Tata Consultancy, SysArris and Wipro are still pure play. Other areas that still allow for product development are, database generation and maintenance, data mining, etc. Phil McHale, vice president for product marketing at MDL Information Systems, does not consider it necessary to have a wet lab to carry out synthesis and screening of compounds.

Intense global competition

There is intense global competition in the field. Bioinformatics knowledge is a decisive and sensitive resource for the pharmaceutical and other industries, hence cheap cost has a benefit only if it comes with substantial quality. Bioinformatics is unlike other IT related fields in this respect. In India many educational institutes and bioinformatics companies started in early 2000s have closed due to this.

India's strong position in bioinformatics

In the highly competitive market of bioinformatics services and products, India already enjoys an edge owing to extensive scientific infrastructure and also because it's an early entrant into the field. Consider efforts would be needed to carve out niche for Pakistan in the field.

Pure play bioinformatics companies are becoming rare

Pure play companies in bioinformatics are merging with pharmaceutical and biotech companies to allow for wet labs within the company. Hence substantial infrastructure might be needed to setup a bioinformatics business. Loans and other funds would be essential to encourage such business.

Bioinformatics companies must be dynamic, rapidly changing

Bioinformatics companies must offer a variety of products to survive the fierce competition; from complete products for small companies to toolkits for large companies that build their own custom solutions.

Intellectual Property Issues

Another factor that will determine the success of bioinformatics business in the country is presence of strict laws that provide intellectual property protection. Algorithms, processes, computer programs, and functionality can all be patented. However, software tends to be

proprietary. In some cases it might be easier to keep secret than to patent, since reverse-engineering binary object code is almost impossible.

Licensing is also a major feature of academic bioinformatics programs. A stream of announcements has been made about software companies licensing their products to pharmaceutical and biotech "content" companies. This trend is sure to grow in the future as companies try to extract the most value from their intellectual capital.

Bioinformatics is more than mere programming

Bioinformatics business is fairly different from other IT related tasks. A substantial level of development skills are required to develop custom applications to knot together and integrate disparate databases *(usually from several global locations)*, simulations, molecular images, docking programs etc.

20 Recommendations

It is imperative that through this study, the need for a stepwise progress in bioinformatics is felt. Development of a highly competitive and technologically advanced field such as bioinformatics will not take place overnight. Many different government policy bodies and private institutes need to work together in this respect. Similar to other places in the world, technological development is possible as a result of a combination of industrial research, government funded academic work and commercial competition. None of them alone can provide the suitable environment for growth.

Some other questions to explore are how many jobs will be created by bioinformatics? and what profits are associated with it?

How many jobs will be created by Bioinformatics?

Although an important subclass within IT, bioinformatics should not be considered as the next big thing. It is unlikely that it creates a huge number of jobs. In 2003 an estimated 8000 jobs in the field of bioinformatics were predicted in India. According to DSQ Biotech Advisor, Mr.A.Madhvan, 5000 professionals would be required in this field for next 5 years who would mainly work on international projects, generating a revenue of at-least Rs.3,000 crores. [1]

Anuradha Acharya, CEO, Ocimum Biosolutions notes, "Bioinformatics is not a field for the masses. People who get involved should have a strong motivation and desire to remain in this field to see true results. The field is justifiable if a company is expecting revenues up to a few million dollars only,"

Joining the bioinformatics revolution is vital to Pakistan for reasons that go beyond export revenues. The country's agrarian economy requires biotechnology including bioinformatics to stay competitive in tomorrow's agricultural market. It requires these sciences to improve healthcare research and delivery. A growing population also raises concerns about maximizing food supply in a sustainable manner and preserving biodiversity. All of these questions require research and information systems. In today's world, biotechnology is required by any country that wants to become self-sufficient.

[1] Bioinformatics red hot, Biospectrum**,** Monday, July 21, 2003
http://www.biospectrumindia.com/content/careers/103072101.asp

Suggestions on how to develop bioinformatics business in Pakistan can be divided into three broad categories,

1. Raising Awareness
2. Development of Human Resource
3. Business Model Development

Awareness

- **Raising Awareness about Bioinformatics Opportunities**

  Perhaps the first and immediate measure in development of a bioinformatics industry in Pakistan is to raise awareness in the relevant fields. As apparent from the survey results, there is a general awareness about the importance of bioinformatics, however, knowledge about business opportunities and relevance of bioinformatics for Pakistan.

- **Workshop for promotion of Bioinformatics Business in Pakistan:**

  Keeping in view the importance of bioinformatics in biological sciences research and the potential for development of bioinformatics business in Pakistan, also keeping in mind that the researchers have limited knowledge of bioinformatics tools let alone business opportunities, it is imperative that a workshop should be in bioinformatics. Such a workshop should be arranged by a public university like NU-FAST or COMSATS and should consist of the following curriculum.

  1. Introduction to Biotechnology
  2. Introduction to Bioinformatics
  3. Business models for bioinformatics enterprise.

  Such a workshop should be of three weeks with one week to each of the three topics. One foreign expert and one potential local and foreign customer should be invited to the workshop. Awareness about business prospects in bioinformatics should be an essential part of all of these activities.

Business Model and Policy Initiatives

- **Bioinformatics will develop along with biotechnology**

  Bioinformatics is a tool to analyze and manage biological data. Bioinformatics is crucial for the advancement of the biotech industry since it can help to tremendously cut the time frame that is usually required to develop a product. The growth of bioinformatics will be simultaneous to the growth of biotechnology industry in Pakistan. There is growing realization that we need to and we are capable of developing indigenously/home grown solutions particularly for agriculture, which is the backbone of the country's economy.

- **First Focus on Local Market**

The global market for bioinformatics solutions looks like a messy crowded bazaar where each vender is crying out to get hold of the client. In other words there is fierce competition, a variety of products are on display, without a clear-cut proof of their functionality.

It is obvious that in such a big market "proof of concept" is essential before a client buys a product or services from a supplier. Such a "proof of concept" for a nascent Bioinformatics industry in Pakistan can be produced through two means.

a- Collaboration with local researchers and industry

Such collaboration would require that IT professionals, biologists cum entrepreneurs and bioinformatics enthusiasts form a company – focus on a certain field, meet biologists with a flare for Bioinformatics solutions and biotech products. Such a group works together to produce a new product, a solution. That can further be sold to others in the international market.

The collaboration between Tata Consultancy Services (TCS) and Center for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India, can serve as an interesting case study in this regard. (given in Threat and Barriers section)

In the question, whether the client comes first or the products the answer is the client. As it will be the client that will guide the service provider to the product. Marketing thus becomes an important part of any bioinformatics company.

b- Collaboration with the Bioinformatics Supplier

At the initial stage, it may prove profitable for the newly established Bioinformatics companies to indulge into partnerships with their counterparts in the developed countries which are reputable and well-established. This is expected to be useful in the beginning as local firms attempt to gain credibility and find a channel to introduce themselves in the international market. The local firms can acquire projects or project modules from foreign bioinformatics companies on contractual basis, and thus attain projects which are manageable in size and allow the firms to focus in the direction in which it has greater expertise. In the words of Mathukumalli Vidyasagar, vice president of a bioinformatics service provider Tata Consulting Services in Hyderabad:

"You have to walk before you run. You have to do some contract research before being able to come up with products."

Thus it seems that a contractual arrangement with the already existing bioinformatics suppliers could be the crucial first step for the nascent bioinformatics companies.

- **Support to the man power: Public – Private partnership**

Public sector scientists and teachers should be able to participate in bioinformatics business. There is merit in encouraging public sector manpower in order to activate the dormant or under-utilized physical facilities available to them. However since a large number of competent persons are working in private sector, there is a need to develop support mechanism, which might bring them together.

One proposal would be to promote, joint projects between public and private institution. These joint projects should be eligible for HEC *(Higher Education Commission)* and other grants provided that public sector partners have at least 50% of the stake. The private sector should include IT companies working in bioinformatics or closely related disciplines. The present policy of restricting HEC support to public sector only is not helpful in unleashing the other wise small reservoir manpower in the biotech and bioinformatics area. Furthermore, without introducing business companies into such partnerships, the objective of promoting bioinformatics as a commercial enterprise may not materialize.

Furthermore in the case of a few commercially beneficial products/ crops etc already developed by national institutes; marketing and production should be outsourced to private companies with due share for the research group, the parent institution as well as the marketing and production company. Such partnership is essential for better management and cost efficiency in the highly competitive market in bioinformatics and biotechnology.

In the field of bioinformatics and biotechnology, the partnership between industry and academia would take two forms:

1) the academic institution out sources translational work, development of a marketable product, any database or software product to a commercial company, the copy right is maintained by the academic institution;

2) the academic institution and the industry collaborate in the entire project and share the copy right/ patent.

Following steps have been recommended in BPI-2004 to facilitate industry-academia partnership:

- ❖ Understanding the industry environment and market environment through techno-economic market surveys
- ❖ Inventorying the transferable technologies available within the BTIS network and making them available on the public domain
- ❖ Developing an Industry-Institution Partnership programme at national, regional and international level.

- ❖ Promoting the growth of business incubators in the field of bioinformatics
- ❖ Provision for industrial and entrepreneurial consultancy services
- ❖ Encourage enterprise creation through liberalized flow of foreign capital, outsourcing, infrastructure generation etc.
- ❖ Adequate emphasis on human resource development that suits the requirement of the industries.

- **Seeding Grants to develop bioinformatics Expertise**

  In order to prepare the country for bioinformatics business, the government needs to provide seed money to local companies to develop solutions for local researchers. Such projects should be assigned to companies willing to venture in bioinformatics. Such funds will improve biological research infrastructure of the country, develop biotechnology industry on the one hand and on the other prepare local expertise in bioinformatics, such experience would be essential to grab outsource work.

  It is not likely that IT companies will jump into taking bioinformatics projects without government support in some way. An appropriate method would be to develop some potential projects and give them to local companies willing to venture into the field. Consideration should be given to the cost and time required to build the workforce and infrastructure to take up a bioinformatics project. There should also be a provision for the company to market it's product in case it is willing to do so.

- **Creation of a Model Bioinformatics Company**

  As a result of seed money or a research grant; a model companies will be created that would serve as an example and inspiration to others.

- **HEC support for bioinformatics education and startups:**

  Bioinformatics is an enabler for biological research as well as biotech industry. It should thus be identified as a priority area for the development of science and technology in the country. HEC funds millions of rupees worth of research in biology and biotechnology, a fraction of this money if utilized for development of capacity and infrastructure in bioinformatics, would help in improvement of research and increase commercial remunerations from research.

  The industrial liaison secretariat of HEC is running a program in which 80% of a certain university-industry collaboration project is funded by HEC while 20% is contributed by the industry. This should be extended to the field of bioinformatics as well.

  This has benefits both for the industry as well as academics. The academia will have not only an additional avenue for generating income through the sale of inventions but also an opportunity for students and faculty to gain practical experience, add to their profiles. It will also add to their professional satisfaction. To the industry this would provide research based insight and guidance for dealing in emerging technologies.

Foreign experts for bioinformatics programs similar to those in other fields of biological sciences should be arranged for bioinformatics.

The focus of research in Pakistan needs to be changed. Promotion and tenure positions should be decided on productivity of a researcher in terms of solving local problems and advancing knowledge on locally relevant problems, instead of publishing papers in American journals, which would alienate educational institutions form the society even further.

- **Clearing House**

  A web based clearing house, comprising the list of projects (provided) and IT tools/ softwares (also provided) should be maintained by PSEB. This would serve 2 purposes.

  1. It will allow Pakistani practitioners to form a community of their own, where they can share information with each other.

  2. It will publicize their competence to attract customers for both local and international market.

  PSEB should only take the responsibility of maintaining the clearing house mainly through updating the web site.

- **Private entrepreneurship model for Bioinformatics**

  In the American model of knowledge-based economy, a university is funded by public money as well as funds from industry. University labs and departments produce the baseline research during the course of which, some researchers start 'spin off' companies. These companies produce solutions to a certain problem with the help of venture capitalists.

  In case of Pakistan the only biotech company we have come across in our research was NIBGE affiliated public sector company called Pakistan Innovative Biotechnology Services (PIBS). Interestingly, the aims and objectives of several public sector institutes like Centre for Excellence in Molecular Biology (CEMB) Lahore, include marketing of molecular research products. A biotech company can do this far more efficiently. The rigors of the market will not allow public sector biotechnology company model to flourish.

  May it be biotechnology or bioinformatics private entrepreneurship is essential for progress in the field.

- **Bioinformatics Business: A challenging Field**

  The field of bioinformatics is a very competitive one. It requires specialized expertise, development costs and inter-disciplinary partnerships. Moreover, there are a lot of companies that are already working in different areas within this field. Technology companies find that they not only have to stand out in a sea of other companies but also have to do so in a financially attractive manner.

This becomes clear when we look at the area of drug discovery. Analysts at Lehman Brothers (New York) have observed that, "Genomics threatens to increase not only the overall associated research and development (R&D) costs but also the average cost per new chemical entity (NCE) or drug". With so many new technologies on offer, pharmaceutical companies must be selective in their collaborations, alliances, and partnerships to ensure that they do not exacerbate this already precarious situation.

The industry looks for the most gain possible from each dollar spent on outsourcing.

- **Foreign Assistance for Human Resource Development**

To improve the quality of bioinformatics education and organize business, the government can seek foreign assistance through faculty and contracts with foreign countries, also more students need to be supported for higher education in this field. Contract research would also serve to increase technology transfer through mutually beneficial accord.

In 2003 the government of Pakistan and US have signed a comprehensive science and technology cooperation agreement which includes Pak-US partial PhD support program, Optimization of large scale production of value added medical plant extract, linkages of centers for chemical sciences, etc. Similar linkages and collaboration should be considered for bioinformatics as well.[2]

- **Need of Resource Integration**

A considerable amount of resources exist within Pakistan in biological research. However there is little interaction between different groups. It is fairly common for biologists to be working in the same area and be unaware of each other's work. Same is the case between industry and academia. A look at the projects funded by various funding agency (Appendix 5) would show that various projects have commercial value such as development of tolerant crop species, discovery of drugs from traditional medicines, ecological studies and research on life stock and fisheries There are also a number of useful databases obscured by inappropriate marketing and promotion.

Keeping in view the above scenario, there is immediate need for collaboration and devising systems for inter and intra disciplinary information sharing. In this regard, HEC's role in promoting resource sharing of lab equipment is highly commendable.

- **Contract Research might be the first step**

There is a growing trend towards outsourcing in research and development. Contract research accounts for an approximate US $ 7 billion business at the moment and estimated to grow at 30% per annum for the next 5 years. Pakistani scientists and research institutions have a genuine opportunity in this regard.

[2] HEC Official website www.he.gov.pk/htmls/rsp/ils/projects.htm

Contract research opportunities as being done by India and other countries is typically basic gene sequencing work, cDNA library preparation, genetic research related to medical diagnostics, new crop varieties and early stage drug development. Large multinationals follow the strategy of sub contracting in biological research in a manner identical to sub-contracting in software development to enjoy benefits in cost, reduce time and sometimes to enjoy less strict regulations.

The key stakeholders in this segment would be pharmaceutical companies, entrepreneurs, private and public research institutions, funding agencies and educational institutes.

Contract Research Organizations can function as export oriented units. It will also facilitate technology transfer to local research groups and capacity building for local informatics companies.

Development of Human Resources

- **What sort of bioinformatics experts would be needed by Industry?**

Bioinformatics experts can be divided into 2 broad categories.

1. Biologists with bioinformatics knowledge
2. IT experts with bioinformatics expertise

1. Biologists with bioinformatics knowledge

The major research interest of these scientists is biology; however, their research requires them to use bioinformatics tools. Their education/ training requires introduction to bioinformatics tools and expertise in operating these softwares, however a thorough knowledge of the mechanics behind softwares. Trouble-shooting and development of new tools is beyond their training and expertise. A few universities in Pakistan like University of Arid Agriculture, Rawalpindi, University of Karachi and Baluchistan University of Information Technology and Management Sciences, Quetta are offering such courses.

2. IT experts with specialty in bioinformatics

Such experts have expertise in both IT as well as biology, however their major from is IT. Few programmes of this sort are being undertaken at institutes known for their strength in IT. Their curriculum includes learning computer languages and it is expected that they will have the basic training to be involved in development and maintenance of bioinformatics tools and services.

It is expected that this sort of experts in bioinformatics would be more useful for any potential bioinformatics business in the country. Hence such

programmes offering bioinformatics as an undergraduate major or a graduate specialty should be encouraged by PSEB and other stockholders in bioinformatics business.

- **Collaboration between Bioinformatics Educational Institutes and Researchers**

  There is an imminent need to improve bioinformatics education in the country through collaboration between different educational institutes and researchers. In addition to this there should be resource pooling in terms of faculty and infrastructure between different bioinformatics programmes.

  An informal national body should oversee the development of human resource in the field and serve as advocating body for the development of bioinformatics business in the country. It should also ensure that various educational institutions are imparting quality education. However it is important for such a body to be a guide, its suggestions should not have a binding legal status.

- **Funding of Bioinformatics Research**

  As a new field bioinformatics suffers from difficulty in peer review process due to lack of experts. According to HEC, no projects in bioinformatics have been funded by the agency as yet. The complaint was also raised by the bioinformatics researchers in our sample. For the development of the field, the government needs to consider bioinformatics as a high priority area in terms of research funding.

  PSEB could advocate for increasing funding for bioinformatics research. It is also necessary that these funds should also be given to private institutions as well as companies. Growth is not possible if the private sector is not encouraged.

- **Government Support for Bioinformatics Programs**

  Bioinformatics is a potential area with huge prospects. However in the absence of clearly market opportunities, there is reluctance on part of students and their parents to join bioinformatics programs. This problem becomes more pronounced at graduate level, where IT students are lured by more lucrative jobs. To overcome this situation government support is needed comparable to the support provided to medical and engineering colleges. The government needs to review its funding policies keeping in view the market demands and future prospects.

  Government funding would not only increase the number of students in the field but also provide a more competitive student body, necessary to the growth of the field. A look at FAST-NU fee structure for 4 year BS in Computer Science/ Bachelors of Business Administration shows a total expenditure of around 0.5 million rupees. Around 50 million rupees will be required to produce 100 bioinformatitions with a bachelors degree.

Fee Structure and Student Loans 4 yrs bachelors (Computer Science or Business Management) at FAST-NU

	BS / BBA	
Tuition Fee (05-06)	Rs 112,000	per annum
Tuition Fee (06-07)	Rs 118,000	per annum
Tuition Fee (07-08)	Rs 124,000	per annum
Tuition Fee (08-09)	Rs 130,000	per annum
Total expenditure	**Rs 484,000**	

Source: FAST-NU Official website http://www.nu.edu.pk/Finances.asp

An estimated 0.6 Million Rupees is spent on producing one doctor in the government funded medical colleges. This account for close to 500 million Rupees spends on medical colleges every year. In addition the government spends a fair amount of money on other public supported programmes at government universities. Keeping this in mind 100 scholarships per year for bioinformatics BS does not seen to be an irrational deal.

- **Combine Scientific Expertise with Entrepreneurial Skills**

A mix of scientific expertise and entrepreneurial skills would be helpful in the field of bioinformatics business. It is imperative that bioinformatics and biotechnology students are taught not only in core sciences but also in business management.

To enhance interaction between academia and industry, the University of Pune in India has devised a combined MBA-Biotechnology program since 2002. Each batch consists of 60 students, which are exposed to subjects like regularity matters, Intellectual Property Rights (IPR) Bio-enterprise management, BT plant management, Bio Safety issues, and emerging trends in biotechnology and short-term training in biotechnology units.

Such programs can be considered in Pakistan as part of long term planning for development in the field of bioinformatics business. An interesting case study would be to that of Entrepreneurship Development Institute of India, which aims to share research findings with academic community and enhance the boundaries of knowledge on entrepreneurship.

- **Advocacy for Improvement in School level Education**

School education in Pakistan is in need of urgent reform. There is a tendency towards early specialization. The dichotomy between 'pre-medical' and 'pre-engineering' groups during secondary and higher secondary level was a frequent concern to many respondents of our survey. Such a dichotomy should be eliminated. Furthermore, bioinformatics and biotechnology needs to be introduced at an early level to stimulate awareness and interest in the field.

The government of India has introduced biotechnology in higher secondary *(grade 12)* level.

APPENDICES

Course Outlines of Bioinformatics Educational Institutes

International Islamic University Bioinformatics Course Schedule

BS in Bioinformatics

It will be a four years/ 8semesters/135-138 credit hours degree. Students will be required to carry out course work comprising of 45 credit hours from core computer science subjects.24 credit hours will be taken from university requirements, 27 credit hours from supporting & management courses, and the rest of 33 credit hours of the program will be from Biological Sciences including Agriculture, Medicine, and Biological Research areas. 6 credit hours of the program will be from internship and project work.

Eligibility

F.Sc/A-Level or equivalent with at least 50% marks with Biology as major subjects.

The scheme of study for BS in Bio Informatics is provided in the Table 6.

Table 6

Scheme of Study for BS in Bio Informatics

1st Semester	2nd Semester	3rd Semester	4th Semester
UR103 Arabic I UR107 Functional English I UR114 Introduction to Computers AM525 Fundamentals of Mathematics CS501 Fundamentals of Algorithms	UR104 Arabic II UR108 Functional English II UR109 Islamic Studies I AM501 Calculus AM504 Statistics & Probability CS502 Programming Lang	UR110 Islamic Studies II AM506 Discrete Structures FI501 Financial management CS503 Programming Lang II BI501 Agriculture SE501 Software Engineering I	UR113 Fiqh I CE505 Digital Circuit Design CS504 Data Structures CS505 Object Oriented Programming CS507 Data Base Mgmt System BI502 Biochemistry
5th Semester	**6th Semester**	**7th Semester**	**8th Semester**
CE508 Computer Architecture BI512 Psychology SE502 Software Engineering II CS508 Visual Programming BI504 Agricultural Economics BI505 Molecular Science	CS510 Operating System Concepts CS511 Web Programming TE502 Data Communication BI507 Pharmacology and Toxicology BI508 Biomedical Sciences BI509 Phylogenetics	CS516 Computer Networks MG601 Human Resource Development BI503 Biotechnology BI510 Nucleic Acids Seq. Analysis BI511 Hospital management BI506 Genomics	UR111 Pakistan Studies BI581 Project

Muhammad Ali Jinnah University Bioinformatics Course

Bachelor of Science
Bioinformatics

Admission Requirements
A minimum of 12 years of education

- Intermediate (Pre – Engineering / Pre – Medical / ICS), A Levels or equivalent (minimum 50% marks)
- Admission Test
- Interview

Degree Requirement
Each candidate for the BS Bioinformatics degree is required to complete successfully the following undergraduate courses:

A.	Core Courses	63 sch
B.	Elective Courses	36 sch
C.	General Education Courses	30 sch
D.	Final Project	6 sch
Total		**135 sch**

Core Courses (63 sch)

BT 1003	Bio-statistics	3 sch
BI 1003	Introduction to Bioinformatics	3
BT 1013	Introduction to Biochemistry	3
BT 3033	Cell Biology	3
BT 2023	Introduction to Molecular Biology	3
BI 2013	Computational Biology	3
BT 2243	Proteome and Proteomics	3
BI 2043	Microbiology and Immunology	3
BT 3023	Molecular Phylogeny and Evolution	3
BT 2053	Genetics	3
CS 1104	Computer Programming	4
CS 2114	Data Structures	4
CS 3203	Introduction to Database System	3
CS 2104	Object Oriented Programming	4
MATH 2103	Calculus I	3
MATH 3003	Linear Algebra	3
MATH 2113	Calculus II	3
MATH 2123	Applied Differential Equations	3
MATH 1303	Discrete Mathematics	3
MATH 3123	Mathematical Modeling and Simulation	3

Elective Courses (36 sch)

BT 3143	Gene Expression and Regulation	3 sch
BT 2143	Introduction to Biotechnology	3
CS 1003	Introduction to Computer Science	3
CS 2203	Applied Graph Theory and Algorithm	3
CS 4253	Data mining	3
CS 4803	Artificial Intelligence	3
BT 2153	Applications of Biotechnology	3
BI 3023	Functional Genomics	3
BT 2443	Tissue and Cell Culture	3
BT 3303	Pharmacogenomics	3
BT 4903	Special Problem	3
BT 1023	Bio-Ethics and Biosafety Principles and Regulation	3
BI 3003	Current Topics in Bioinformatics	3
BT 2153	Virology	3
BI 2124	Mathematical Methods for DNA sequences	4
BT 2213	Genetic Engineering	3
CS 4213	Database Management Systems	3
CS 3003	Data Visualization	3
CS 4913	Neural Computing and Genetic algorithms	3
CS 4703	Software Engineering	3
CS 2123	Introduction to Object Oriented Paradigm	3
EE 4723	Image Processing	3
EE 4103	Stochastic Processes	3

Internship:
It is mandatory for every student to participate in an internship program following their 6th semester or after the completion of 90 credit hours. A formal evaluation will be carried out and Pass/Fail grade will be awarded to the student.

Duration
This is a 4 year program comprising of 8 semesters with minimum of 135 Semester Credit Hours (sch). There will be a Fall and a Spring Semester in each year. The summer session will be utilized for projects, internships or deficiency course.

COMSATS Bioinformatics BS

Bachelor of Science in Bioinformatics
Bachelor of Science in Bioinformatics offering Semesters, Course title and Credit Hours as follows:

Semester - I	Course title	Credit Hrs
HUM 1541	English Comprehension Skill	3 (3 0)
CSC 1020	Computing Fundamentals	3 (2 3)
MTH 1001	Mathematics I	3 (3 0)
BIO 1001	Cell Biology	3 (3 0)
BIO 1002	Fundamentals of Genetics	3 (3 0)
BIO 1003	Principles of Biochemistry	4 (3 3)
Semester - II		
HUM 1531	Islamic Studies	3 (3 0)
MTH 1004	Calculus	3 (3 0)
BIO 1004	Fundamentals of Microbiology	3 (3 0)
BIO 1005	Introductory Molecular Biology	3 (3 0)
HUM 1544	English Composition	3 (3 0)
BIO 1006	Statistical Methods in Biology	3 (3 0)
Semester - III		
MTH 2061	Discrete Mathematics	3 (3,0)
CSC 1123	Technical Writing	3 (3,0)
BIO 2001	Immunology	3 (3,0)
HUM 1532	Pakistan Studies	3 (3,0)
MTH 2060	Algorithms	3 (3,0)
BIO/MTH	Electives	3 (3,0)
Semester - IV		
CSC 2022	Object Oriented Programming	4 (3,3)
BIO 2002	Medical Microbiology	3 (2,3)
BIO 2003	Functional Genomics	3 (3,0)
CSC 2113	Informational Modeling	3 (3,0)
CSC 1031	Introduction to Data Structures	4 (3,3)
BIO 2004	Elective	3 (3,0)
Semester - V		
BIO 3001	Applications of Biotechnology	3 (0,0)
BIO 3002	Biochemistry of Gene Expression	3 (0,0)
BIO 3003	Bioinformatics I	4 (3,3)
CSC 4091	Computer Graphics	3 (0,0)
BIO 3004	Essential techniques in Biochemistry and Molecular Biology	3 (0,9)
BIO/CSC	Elective	3 (3,3)
Semester - VI		
BIO 3005	Protein Chemistry	3 (2,1)
BIO 3006	Molecular Genetics	3 (3,0)
BIO 3007	Bioinformatics Professional Placement (Internship)	3 (3,0)
BIO 3008	Bioinformatics II	4 (3,3)
CSC 4015	Modeling and Simulations	3 (3,0)
Semester - VII		
BIO 4001	Social, Ethical and Legal aspects of Biotechnology	3 (3,0)
BIO 4002	Cell Communication	3 (3,0)
BIO 4003	Bioinformatics III	4 (3,3)
BIO 4004	Literature survey and research in Bioinformatics (Research Proposal)	3 (3,0)
BIO/CSC	Elective	3 (3,0)
BIO 4005	Seminar and Presentations / Group Discussion	2 (2,0)

Semester - VIII		
BIO 4006	Independent projects formulations, execution and presentation or Two elective courses will be offered.	6 (6,0)
Electives		
CSC 3100	Artificial Intelligence	
CSC 3111	Advanced Database Systems	
CSC 3090	Digital Image Processing	
MTH XXX	Bioinformatics-Algorithms	
BIO XXX	Protein Classification and Structure Prediction	
BIO XXX	Metabolism	
BIO XXX	Understanding Cancer	
BIO XXX	Understanding Metabolic Disease	
BIO XXX	Genetics in Child Health	
BIO XXX	Molecular Biology of the Cell	
BIO XXX	Application of biotechnology in Agriculture	
BIO XXX	Industrial Biotechnology	
BIO XXX	Population Genetics	
BIO XXX	Introduction to Epidemiology and Clinical Trials	
MTH XXX	Bayesian Statistics	
BIO XXX	Mathematical Biology	
CSC XXX	Statistical data Mining	
MTH XXX	Advanced Mathematics	
PHY XXX	Biophysics	
CSC XXX	Database Management	
CSC XXX	Computer Programming	
BIO XXX	Developmental Biology	
BIO XXX	Pharmacology	
BIO XXX	Clinical Biochemistry	

Baluchistan University of Information Technology and Management Technology (BUIT)

Courses Outlines of BS (Biotechnology & Informatics)

Eligibility / Prerequisite Qualification:

Intermediate (Pre-Medical/Non-Medical) or Equivalent with 45%Marks or above can apply for admission Maximum Students: 50

First Semester

Course Number	Course Name	Credit Hours
BSGB-111	General Biology(for Non-Medical Students)	3+0
OR		
BSGM-111	General Mathematics (For Pre-Medical Students)	3+0
BSIG-112	Introductory Genetics	3+0
BSFE-113	Applied Functional English	3+0
BSCB-114	Introduction to Computers and bio-computing	3+1
ISL-115	Islamic Studies	2+0
BSBB-116	Introduction to Biochemistry and Biotechnology	3+1

Total credit Hours 18(16+2)

Second Semester

Course Number	Course Name	Credit Hours
BSMB-121	Introduction to Molecular Biology	2+1
PAK-122	Pak Studies	2+0
BSBM-123	Biomathematics	3+0
BSIM-124	Introductory Microbiology	2+1
BSAN-125	Introduction to Bio-analytical Techniques and Instruments	2+2
BSCS-126	Cell Structure and Function	2+1

Total Credit Hours 18(13+5)

Third Semester

Course Number	Course Name	Credit Hours
BSAG-231	Applied Genetics	2+1
BSCU-232	Cell and Tissue Culture Techniques	2+1
BSIT-233	Introduction to Toxicology and Pharmacogenomics	3+0
BSIB-234	Introductory Bioinformatics	3+0
BSBT-235	Industrial Biotechnology	2+1
BSIE-236	Introduction to Epidemiology	3+0

Total Credit Hours 18(15+3)

Fourth Semester

Course Number	Course Name	Credit Hours
BSPC-241	Protein Chemistry and Function	2+1
BSRT-364	Recombinant DNA Technology	2+1
BSAB-	Advance Biochemistry	3+1
BSIT-363	Immunology, Techniques and Applications	2+1
BSCP-246	Bioinformatics II	1+2

Total Credit Hours 16(10+6)

Fifth Semester

Course Number	Course Name	Credit Hours
BSNP-351	Natural Products and Biotechnology	3(2+1)
BSEB-352	Environmental Biotechnology	3(2+1)
BSAM-353	Applied Microbiology	3(2+1)
BSAN-125	Bio-analytical Techniques	3(2+1)
BSDS-355	Data Structures and Algorithms	2(1+1)
BSAE-356	Enzymology	2(1+1)

Total Credit Hours 16(11+5)

Sixth Semester

Course Number	Course Name	Credit Hours
BSNT-361	Nutrition and Food Biotechnology	3 (2+1)
BSVD-362	Biotechnology in Vaccine Development	2(1+1)
BSIT-233	Introduction to Toxicology and Pharmacogenomics	3(2+1)
BSRB-245	Radio Isotope in Biotechnology	2(1+1)
BSRM-365	Research Methodology	2(2+0)
BSPB-366	Computer Programming for Bioinformatics	2(0+2)
BSIE-236	Epidemiology	2(2+0)

Total Credit Hours 16(10+6)

Seventh Semester

Course Number	Course Name	Credit Hours
BSMB-471	Medical Biotechnology	3(2+1)
BSGD-472	Genetic Disorders	2(1+1)
BSBS-244	Biostatics	2(2+0)
BSBA-483	Economic and Business Application of Biotechnology	2(2+0)
BSTH-484	Thesis	3(0+3)
BSDD-475	Database Design	3(1+2)

Total Credit Hours 15(8+7)

Eighth Semester

Course Number	Course Name	Credit Hours
BSBM-481	Biomedical Informatics	3(2+1)
BSAI-482	Artificial Intelligence Applied	3(2+1)
BSTH-484	Thesis	3(0+3)
BSES-474	Ethics in Science	4(4+0)
	Biotechnology and Pest Management	3(2+1)

Total Credit Hours 16(10+6)

List of Scientists Working in Biotechnology in Pakistan[1]

Faisalabad			
1.	Ayub Agricultural Research Institute, Department of Biotechnology, Faisalabad.	Dr. Ghulam Ahmad (Director General)	Ph: 041-657281-90 Ext. 229 agribt@fsd.brain.netpk agribt@fsd.paknet.com
2.	Center for Agricultural Biotechnology University of Agriculture, Faisalabad	Dr. Iftkhar Ahmad (Chairman)	Ph: 041-9200161-70 Ext.2921 Fax: 041-647846 http://www.uaf.edu.pk/
3.	Nuclear Institute for Agriculture and Biology (NIAB). P.O.Box 128, Jhang Road, Faisalabad	Dr. Mohsin Iqbal (Chief Scientist/Director NIAB)	Tel: 041-654210 Fax: 041-654213 niab@fsd.paknet.com.pk http://www.nibge.org/
4.	National Institute for Biotechnology & Genetic Engineering (NIBGE). P. 0. Box - 577, Jhang Road, Faisalabad	Dr. Ahmad Mukhtar Khalid (Director General)	Tel: 041-651471/75 Fax: 041-651472 amkhalid@nibge.org http://www.nibge.org/
Rawalpindi / Islamabad			
1.	Agriculture Biotechnology Institute, National Agriculture Research Centre (NARC), Islamabad	Dr. Rasped Anwar (Deputy Director General)	Tel: 051-9255217 Fax: 051-9255217 abi_narc@yahoo.com
2.	Biomedical & Genetic Engineering Division, Dr. A. Q. Khan Research Laboratories, P.O. Box-2891, Islamabad	Dr. S. Qasim Mehdi (Director General)	Tel: 051-9261138 Fax: 051-9261144 E-mail: sqmehdi@comsats.net.pk
3.	Department of Biochemistry University of Arid Agriculture, Rawalpindi	Dr. Azra Khanum (Chairperson)	Tel: 9290151-2 Fax: 9290160 http://www.uaar-edu.pk/
4.	Department of Plant Pathology University of Arid Agriculture, Rawalpindi	Prof Dr. Irfan-ul-Haq (Chairman)	Tel: 9290151-2/ Ext-143 Fax: 9290160 http://www.uaar-edu.pk/
5.	Department of Biological Sciences, Quaid-i-Azam University, Islamabad	Dr. Afsari Qureshi (Chairperson)	Tel: 051-9219809 Fax: 051-9219888 :

1

(www. Pakissan.com)

HEC has also organized a list and contact details list of over 300 scientists working in biological sciences in Pakistan available on www.hec.gov.pk/htmls/life_sciences/doc_pdf/directory_staff.pdf

Karachi			
1.	Dr. Punjwani Center for Molecular Medicine and Drug Research University of Karachi, Karachi	Dr. lqbal Chaudhary (Acting Director)	Tel: 021-9243224 Fax: 021-9243190-91 zainraa@digicom.net.pk hej@khi.comsats.net.pk
2.	Plant Tissue Culture Lab, H.E.J Institute Research Institute of Chemistry, Karachi	Dr. Saifullah Khan (Assistant Professor & lncharge)	Tel: 021-9243205 Ext-148 Fax: 021-9243190/91 drsaif@super.net.pk
3.	Dr. A. Q. Khan Institute of Biotechnology & Genetic Engineering, Karachi	Dr. Mujtaba Naqvi (Director	Tel: 021-4823885 Fax: 021-4823887 kibge@cyber.net.pk
4.	Department of Biotechnology, University of Karachi	Dr. Altaf Khan (Dean Faculty of Science)	Tel: 021-9243131-42 I Fax: 021-9243161 http://www.ku.edu.pk/
5.	Center for Molecular Genetics University of Karachi, Karachi	Dr. Nuzhat Ahmad (Director Center)	Tel: 021-9243131-7 Fax: 021-4966045
6.	Dept. of Microbiology University of Karachi, Karachi	Prof Dr. Shahana Urooj Kazmi (Chairperson)	Tel: 021-9243131-7 Fax: 021-4966045
Lahore			
1.	Centre for Excellence in Molecular Biology, University of Punjab, Lahore	Dr. S. Riazuddin (Director)	Tel; 042-5421235 Fax: 042-5421316 cambl@wol.net.pk
2.	Institute of Biochemistry and Biotechnology, University of Punjab, Lahore	Prof. M. Waheed Akhter (Dean Faculty of Science)	Tel: 042-9211612 Fax: 042-9230242 mwapu@brain.net.pk http://www.pu.edu.pk/
3.	Biotechnology Laboratory Department of Botany Govt. College, University Lahore	Dr. lkram-ul-Haq (Head)	Tel: 042-9221634 Fax: 042-7243198 ikrhaq@yahoo.com
4.	School of Biological Sciences, Punjab University, P.O.Box - 54590, Lahore	Prof. M. Akhtar (Director General)	Tel: 042-5432746-47 Fax: 042-5462748
5.	Dept. of Microbiology and Molecular Genetics. University of Punjab	Prof. Shahida Hasnain (Chairperson)	Tel: 042-9231249 Fax: 042-9230481
6.	Biotechnology and Food Research Centre, PCSIR Laboratories, Lahore	Dr. Nazir Hussain Shah (Director)	Tel: 042-9230688 Fax: 042-5877433

Peshawar			
1.	Institute of Biotechnology & Genetic Engineering, N.W.F.P Agricultural University, Peshawar	Dr. Zahoor Ahmad Swati (Director)	Tel: 091-9216553 6 Fax: 091-9216553 drzaswati@yahoo.com
2.	Center for Animal Biotechnology Veterinary Research Institute, N.W.F.P, Peshawar	Dr. Mohammad Subhan Qureshir Officer-in-charge CAB	Tel: 091-9210218-9 11 Fax: 091-9210220 vrmsqureshi@yahoo.com
3.	Department of Biotechnology University of Peshawar, Peshawar	Dr. Farrukh Hussain (Director)	Tel: 091-9216701-20 Ext-3070
Multan			
1.	Central Cotton Research Institute, O1d Shujabad Road, P.O.Box - 572, Multan	Dr. Muhnmmad Islam Gill (Chief Scientific Officer/ Director)	Tel: 061-9200340-41 Fax: 061-9200342 ccri@mul.paknet.com.pk http://www.ccri.org.pk/
Quetta			
1.	Institute of Biochemistry, University of Balochistan, Quetta	Dr. Masoom Yasin Zai (Professor)	Tel: 081-9211261 Fax: 081-9211277 masoom@infolink.net.pk
Jamshoro			
1.	Institute of Biotechnology and Genetic Engineering, University of Sindh, Jamshoro.	Dr. Umar Dahot (Director)	Tel: 0221-771681-90 udehot@yahoo.com

List of projects in Biological Sciences funded by HEC (2003-2005)

To get an idea of the research interests of different scientists in the country, A list of grant awards by Higher Education Commission, Government of Pakistan is provided. HEC is the largest funding agency in Pakistan and funds more than half of all the research projects in the country. Hence this list can serve as a good indicator of the kind of research being conducted at different research institutes. Other funding agencies are, Pakistan Science Foundation, Pakistan Agriculture Research Council and research funds of different universities, like Aga Khan University, etc.

The projects list is organized by research area/ commercial utilization of research.

Medicine/ Health/ Pharmaceutical

Principal Investigator	University/Institution	Total of the Project	Approved Amount in Rs.
Dr. Waseem Ahmed Asstt Professor	Quaid-i-Azam University	Identification of Genes Involves in Hereditary hearing impairment	1,860,000
Dr.S. Raizuddin, Prof. Director, National centre of Excellence in Molecular Biology, Lahore	University of the Punjab, Lahore	Screening for Vision and Hearing Impairment in Rural Sindh and Balochistan	1,646,000
Prof. Dr. Shahidha Hasnain, Chairperson, Department of Microbiology & Genetics	University of the Punjab, Lahore	Identification, Characterization of Drug resistant E.coli and its treatment: A perspective of Novel Antibacterial Agents	1,871,000
Dr. Muhammad Umer Dahot Prof. & Director	Centre for Advance studies in Biotechnology, University of Sindh	Isolation, purficatication and Characterization of protease inhibitors from plants	1,320,000
Dr. Shaheen N.Khan National centre of Excellence in Molecular biology	University of the Punjab, Lahore	Genetic and molecular basis of recessive deafness	674,000
Dr. M. Karmran Azim, Assistant Professor, International centre of Chemical Sciences, HEJ Research Institute of Chemistry	University of Karachi	Production and Characterization of engineered proteins containing non-natural amino acids; potential tools for specific drug targeting.	1,562,000
Dr.M.Arshad Rafiq. Asstt. Prof. of Biosciences	COMSATS Institute of Information Technology, Islamabad	Characterization of Loci responsible for inherited Eye Anomalies in Families of Pakistan	4,936,800
Dr. Zafar Iqbal Chaudhary Prof. of Pathology	University of Veterinary and Animal sciences	Prevalence of Bovine Tuberculosis in Human, Animal Population and Animal Food Products	1,880,000
Dr.Shiekh Arshad Saeed, Prof. of Pharmacology	University of Karachi. Karachi,	Atherosclerosis an inflammatory disease: Roles for cyclooxygenase-2	1,539,180
Dr.Ihsan Ilahi. Professor of Biosciences	Kohat University of Science & Technology	Botincal Chemical and Pharmacological Studies of important medicinal plants by Establishment of Medicinal plant farm at Kohat university of science and technology	4,771,560

89-1	Process development for the production of enzyme invertase by Saccharomyces cerevisiae.	Government College University, Lahore	Dr. Ikram-ul-Haq	1,433,000	3 Years
65-1	Genetic and molecular basis of recessive deafness.	University of the Punjab, Lahore	Dr. Shaheen N. Khan	674,000	3 Years
11-1	Vanadium Complexes as Insulin Mimetic Agents: Coordination chemistry, Characterization and Biological Studies of Vanadium (IV) and Vanadium (V) Complexes of various Ligands Containing Carboxylate Group	Quaid-i-Azam University, Islamabad	Dr. Saqib Ali	1,212,000	3 Years
2-1	Extraction, Isolation and Investigation of some indigenous Natural Products for their Anti-leishmanial activities.	Gomal University, D.I. Khan	Dr. Gul Majid Khan	1,178,000	3 Years

21.	119-1	Isolation, structural elucidation and bioassay of the constitutions of aconitum violaceum jack.	Gomal University, D.I. Khan	Dr. Irshad Ali	940,000	2 Years
37.	54-1	Comparative study of Radiological Versus chronological age from birth to 22 years.	Liaquat University of Medical and Health Sciences, Jamshoro Sindh	Dr. Nizamuddin Memon	654,100	3 Years

74	Synthesis and Biological Screening of Combinatorial libraries of small drug-like molecules to establish compounds activity relation and drug development	Quaid-i-Azam University, Islamabad	Dr. Farzana Latif Ansari	1,693,000	3 Years
84	Synthesis and Biological studies of some anti-narcotic drugs.	University of Karachi, Karachi	Dr. Zafar Saied Saify	2,000,000	2 Years
7-1	Identification of Genes Involved in Hereditary hearing impairment	Quaid-i-Azam University, Islamabad	Dr. Wasim Ahmed	1,860,000	2 Years
120	Isolation structure elucidation and biological activity of some Medicinal Plants of Pakistan and Iran	University of Karachi, Karachi	Dr. Viqar Uddin Ahmad	1,433,000	3 Years
136	Synthetic Manipulations of Bioactive Amino surgars for the development of new antiosteoarthritis, antiviral and antitumor agents – A new approach in drug designing and SAR studies.	University of Karachi, Karachi	Dr. Shazia Anjum	1,433,000	2 Years
7	Development of new antiviral and anticancer steroids by combinatorial synthesis and high-thorughput biological screening – A new approach to drug discovery	University of Karachi, Karachi	Dr. Muhammad Iqbal Choudhary	1,990,000	3 Years
151	Peptidyl Antibiotics: structure-Function Relationship of the Antimicrobial peptides and Bacteriocins/ Lantibiotics from Indigenous Enterococci.	University of Karachi, Karachi	Dr. S.Abid Ali	1,450,500	3 Years
244	Effect of cloves, bay leaves and turmeric on blood glucose and lipid profile in diabetic individuals.	NWFP Agriculture University, Peshawar	Dr. Alam Khan	1,011,840	3 Years

92-1	Photodynamic Diagnosis System: A new method for the early recognition of bladder tumour.	Liaquat University of Medical and Health Sciences, Jamshoro Sindh	Dr. Muhammad Shahzad Laghari	1,462,500	2 Years
217	Finite element modeling of blood flow: Relevance to atherosclerosis	Mehran University of Eng. & Technology, Jamshoro	Dr. Ahsanullah Baloch	995,000	3 Years
199	Studies on the Effects of Alcohol and Antidepressants on Tryptophan Metabolism & Disposition in Stressed and Unstressed.	University of Karachi, Karachi	Dr. Samina Bano	960,400	3 years
154	Studies on the effect of Vitamin C supplementation on lead levels of blood, hair and urine in adults	The Aga Khan University, Stadium Road Karachi	Dr. Anwar-ul-Hassan Gilani	2,215,408	18 Months
24	Phytochemical Investigations on Genus abutilon with emphasis on the Isolation of some hypoglycemic factors.	University of Karachi, Karachi	Dr. Zaheer Ahmad	1,355,000	2 Years
139-1	Identification, Characterization of drug resistant Ecoli, and its treatment: A perspective of novel antibacterial agents.	University of the Punjab, Lahore	Dr. Shahida Hasnain	1,871,000	3 Years
167	Determination of essential and trace elements in biological samples of human subjects with various physiological disorders.	University of Sindh, Jamshoro	Dr. Tasneem G. Kazi	1,434,000	3 Years
16	Synthesis of Metal-based Sulfonamide Derived Novel Antibacterial, antifungal and antiviral compounds.	Bahauddin Zakariya University, Multan	Dr. Zahid Hussain Chohan	2,242,000	3 years
69	Synthesis of some natural and unnatural isocoumarins and related ompounds in search of potential chemotherapeutic agents.	University of Karachi, Karachi	Dr. Khalid Muhammad Khan	1,981,000	3 years
125-1	A histological study of human olfactory Mucosa.	National University of Sciences & Technology , Rawalpindi	Brig. Dr. Liaqat Ali Minhas	965,000	2 Years
241	Studies of nano-enzymatic glycation in Pakistani subjects suffering from Diabetes and Cataract: A Proteomic approach.	University of the Karachi	Dr. Shamshad Zarina	1,897,000	2 Years
328	Botanical, Chemical and Pharmacological Studies of Important medicinal plants by Establishment of Medicinal Plant Farm at Kohat University of Sciences & Technology.	Kohat University of Science & Technology, Kohat	Dr. Ihsan Ilahi	4,771,560	3 Years
306	Suppression and treatment of Tardive dyskinesis: Neurochemical and pharmacological studies on a rat model.	University of the Karachi, Karachi	Dr. Darakhshan Jabeen Haleem	2,576,520	3 Years
141	Search for new biological active constituents from some medicinally significant plants.	HEJ-Research Institute of Chemistry, Internatinal Center of Chemical Sc. University of Karachi	Dr. Sabira Begum	747,000	3 Years
107	Purification and Characterization of antifungal peptides/proteins from potential medicinal plants and construction ofcDNA libraries for hyper-expression.	University of Agriculture, Faisalabad	Dr. Amer Jamil	1,600,000	3 Years
129-1	Effect of Propranolal on the morphology of rat testis	National University of Science & Technology	Brig. Liaqat Ali Minhas	1,073,000	3 Years

195	Preparation and evaluation of sustained released analgesic drugs.	University of Peshawar, Peshawar	Dr. Zafar Iqbal	710,940	2 Years
342	Vaccine Development against Hepatitis – C Virus (Type 3a)	The Aga Khan University	Dr. Sohail Asif Qureshi	1,887,495	3 Years
196	Production and characterization of engineered proteins containing non-natural amino acids; Potential tools for specific drug targeting	University of Karachi, Karachi	Dr. Muhammad Kamran Azim	1,562,000	3 Years
429	Isolation of structurally novel and pharmacologically point compounds from plant cell suspension cultures	University of Karachi	Dr. Muhammad Iqbal Choudhary	3,793,000	2 Years
246	Atherosclerosis- an inflammatory disease: Role for Cyclooxygenase-2	The Aga Khan University	Dr. Sheikh Arshad Saeed	1,539,180	2 Years
172	Interaction Studies of Diltiazem and Verapamil with H2-receptor antagonist, NSAIDs, Floroquinolones and Metals Essential to Human Body.	University of Karachi	Dr. Najma Sultana	1,579,860	2 Years

Agriculture and Ecology

Principal Investigator	University/Institution	Total of the Project	Approved Amount in Rs.
Dr. Muhammad Ashraf, Prof. and Chairman of Botany	University of Agriculture, Faisalabad	Activities of Plasma membrane and Vacuolar H+ -At phase and Accumulation of antioxidants as markers of Solt tolerance in a potential oil seed crop canola.	1,750,000
Prof. Dr. Rukhsana Bajawa, Department of Botany	University of the Punjab, Lahore	Phenolic Allelochemica! Of Sunflower as Natural herbicides for weed management in wheat	1,210,000
Dr. Asghari Bano	Quaid-i-Azam University	Physiological basis of drought and high temperature tolerance in what (Triticum aestivum L)	1,657,100
Dr. Syed Ehteshamui-Haque, Dept. of Botany	University of Karachi	Utilization of plant growth promoting and nodule producing Bactria in control of root knot nematode and root infecting funi.	1,142,000
Dr. M. Arshad, Deputy Director.Cholistan Institute of Desert Studies (CIDS)	Islamia University Bahawaipur	Socio Economic improvement of Cholistani desert dwellers through exploitable genetics improvement	983,000
Dr. Faheem Aftab, Assistant Prof. of Botany,	University of the Punjab, Lahore	Biochemical Characterization of Vitro salt Tolerant cell lines and regenerated plants of potato and subsequent Establishment under Exvitro.	1,714,000
Dr. Anjam Perven. Assistant Prof. of Botany	University of Karachi	Plant Biodiversity Conservation and its sustainable use of Kirther Range	1,641,500
Prof.Dr. Qasier, Deptt. Of Botany	University of Karachi	Sustainable use of plant Wealth of Chitral and Preparation of red data list.	3,240,220
Prof. Dr.Shafiq Ch. Cholistan Institute of Desert Studies	Islamia University, Bhawalpur	A survey of germplasm resources and phytosociology, ecotype variations in morphological, anatomical biochemical characterization, and phytoremedition potential of two aromatic grasses- vetiveria zizaniodies and chmbopogon jwarancusa with special reference to Southern Punjab.	1,712,580
Dr.Bushra Mirza, Assistant Prof. of Biological Sciences	Quaid-i-Azam University	Generation of Salt resistant tomato by using HAL II genses and molecular analysis of transgentic plants	1,368,000
Dr.Moinuddin Ahmed, Professor of Botany	Federal Urdu University of Arts, Sciences and Technology, Karachi	Dendrochronological Potential of pine Trees Species of Pakistan	3,118,750
Dr. Muhammad Ashraf, Prof. and Chairman of Botany	University of Agriculture. Faisaiabad	Identification of Biochemical Indicatiors for salt tolerance and their genetic basis in potential oilseed; crops: Canola and Sunflower.	5,760,040
Dr. Safdar Hussain Sahh. Prof.	NWFP Agricultural University	Fcross Adaptation and in virtue Development of Salt Tolerant Lines of Rice	2,094,060

Principal Investigator	University/Institution	Total of the Project	Approved Amount in Rs.
Prof. Dr. Mohammad Ajmal Khan , Department of Botany.	University of Karachi, Karachi	salt-induced oxidative stress, Consequencies and Management.	5,392,607
Dr. Amanullah Jan, Assistant Prof. of Agronomy	NWFP Agricultural University, Peshawar	Impact of Agronomic practices on water use efficiency of Rainfed Wheat	1,083,240
Dr. Zabta Khan Shinwari Vice Chancellor	Kohat University of science and Technology, Kohat	Reconstruction of Chromosomal Inheritance in Pedigree of Pakistani Wheat Cultivars	4,408,000
Dr. Mudassir Asrar, Visting Prof. of Biotechnology	Balochistan University of Information and technology and Management Sciences, Quetta	Micro Propagation and cultivation of some economically important plants of Balochistan	2,512,560
Dr. Abdur Rashid, Professor of Botany	University of Peshawar	Floristic Phyto geographic and Ethno-botanical Studies of Vascular Biodiversity in Swat Kohistan (Upper Swat) Hindukush Range, Swat	1,570,000

109-1	Development of Desi Sarsoon (Brassica compestris L.) Varieties Through Conventional and Modern Techniques.	NWFP Agriculture University, Peshawar	Dr. Farhatullah	1,007,000	3 Years
49-1	Pest Risk Analysis and Integrated Pest management of Maize in NWFP.	NWFP Agriculture University, Peshawar	Dr. Farman ullah	1,256,000	3 Years

6.	59-1	Studies on biology, Ecology and Physiology of wild oats (avena fatua L.)	NWFP Agriculture University, Peshawar	Dr. Gul Hassan	1,440,000	3 Years

58-1	Studies on Epidemiology & Biology of Virus &Viroid Dieases of Citrus &their Control through Intergrated Approaches	NWFP Agricultural, Peshawar	Dr. M.Arif	998,000	3 Years
18	Efficacy of Water users Organizations and Water Management in Agriculture (A study of Water Crisis and Irrigation System in Paksitan)	Quaid-i-Azam University	Dr.Hafeez-ur-Rehman	777,000	2 Years
137-1	Farmers' capacity building through information technology in Shaikupura.	University of the Punjab, Lahore	Dr. Muhammad Zakria Zakar	474,000	1 Year
150	Integrated management of aphid in canola in D.I.Khan	Gomal University, D.I.Khan	Dr. Said Mir Khan	812,000	3 Years
115-1	Preparation and characterization of complexes of some of biological active chelating agents	University of Peshawar, Peshawar	Dr. Saeed-ur-Rehman	525,000	2 Years
114	Use of Allelopathy for reducing Herbicide imports.	University of Agriculture, Faisalabad	Dr. Zahid Ata	1,218,000	3 Years
68-1	Systematic study of the berry bugs (Heteroptera: Pentatomidae: Pentatominae: Halyini) of sindh	University of Sindh	Dr. Nasreen Memon	944,000	3 Years
119	Identification, tissue-specific expression and immunolocalization of stress proteins (dehydrins) under high temperature stress.	University of Agriculture, Faisalabad	Dr. Abdul Wahid	1,035,200	2 Years
61	Pathobiology of Shisham (Delbergia sissoo) dieback	Quaid-i-Azam University, Islamabad	Dr. Muhammad Ashraf	1,550,000	3 Years
8-1	Physiological basis of drought and high temperature tolerance in wheat (Triticum aestivum L.)	Quaid-i-Azam University, Islamabad	Dr. Asghari Bano	1,657,100	3 Years
161	Medicinal and pesticidal agents based on tagetes species.	University of Karachi, Karachi	Dr. Shaheen Faizi	1,670,015	3 Years
181	Mass rearing of Coccinellid predators on different insect pests.	Sindh Agriculture University, Tandojam	Dr. Muhammad Khan Lohar	1,623,000	3Years
118-1	Activities of plasma membrance and vaculoar H+ - ATPases, and accumulation of antioxidants as markers of salt tolerance in a potential oilseed crop canola.	University of Agriculture, Faisalabad	Dr. Muhammad Ashraf	1,750,000	3 Years

143-1	Biochemical characterization of in vitro salt-tolerant cell lines and regenerated plants of potato (Solanum tuberosum) and subsequent establishment under ex-vitro conditions.	University of the Punjab, Lahore	Dr. Faheem Aftab	1,714,000	3 Years
207	Diagnostic and research centre for Mango Orchards.	Bahauddin Zakariya University, Multan	Dr. Mushtaq Ahmed Saleem	1,366,000	3 Years
86	Bionomic Systematics & Predatory efficacy of Phytoseiid and stigmaeiid against harmful mites and small insects.	University of Agriculture, Faisalabad	Dr. Muhammad Afzal	1,035,000	3 Years
271	Bio-ecology and systematics of acridoid grasshoppers (Orthoptera) with special emphasis on Heiroglyphus nigrorepletus from AJ&K.	University of Azad Jammu & Kashmir, Muzaffarabad, Azad Kashmir	Dr. Khalid Mahmood	1,330,480	3 Years
60	Plant Biodiversity, Conservation and its sustainable use of Kirthar range.	University of Karachi, Karachi	Dr. Anjum Perveen	1,641,500	3 Years
243	Studies on low Productivty of Mango Orchards: Its causes & possible Remedies.	Bahauddin Zakariya University, Multan	Dr. Muhammad Akbar Anjum	1,760,000	3 Years
66	Phenolic Allelochemicals of sunflower as Natural herbicides for weed management in wheat.	University of the Punjab, Lahore	Dr. Rukhsana Bajwa	1,210,000	3 Years
169	Response of rice based cropping system to soil and root dip Zn application under integrated fertilization.	Gomal University, D.I. Khan	Dr. Muhammad Umar Khan	1,593,000	3 Years
187	Multiplication of Chrysoperla carnea as a biocontol-agent in Integrated Pest Management.	University of Agriculture, Faisalabad	Dr. Muhammad Ashfaq	1,390,260	3 Years
278	Sustainable use of plant wealth of Chitral and preparation of red data list.	University of Karachi, Karachi	Dr. Muhammad Qaiser	3,240,220	3 Years
296	Integrated Pest Management (IPM) of Asian Ambrosia Beetle - A vector of a Killer Disease of Mango in Sindh	Sindh Agriculture University, Tandojam	Dr. Rab Dino Khuhro	1,569,000	3 Years
305	Biosystematics of Dragonflies (Odonata) of Pakistan	University of Arid Agriculture, Murree Road, Rawalpindi	Dr. Abdul Khaliq	1,405,506	3 Years
118-1	Activities of Plasma membrane and vacuolar H+ - ATPases, and accumulation of antioxidants as markers of salt tolerance in a potional oilseed crop canola.	University of Agriculture, Faislabad.	Dr. Muhammad Ashraf	1,750,000	3 Years
180	Effect of Neem products on some economically improtant Insect Pests of Cotton, Gram and Muskmelon and their natural enemies.	Gomal University, D.I. Khan	Dr. Masood Khan Khattak	1,195,500	3 Years
349	Utilization of plant growth promoting and nodule forming rhizobacteria in the integrated control of root infecting fungi of Sunflower and Soybean.	University of Agriculture, Faisalabad	Dr. M. Inam-ul-Haq	1,436,160	3 Years
189	Inducing salt tolerance in cereals through bacterial ACC-deaminase biotechnology	University of Agriculture, Faisalabad	Dr. Muhammad Arshad	2,106,300	3 Years

98	Conservation and Cultivation of two aromatic Grasses- Vetiver & Cymbopogon-with special reference to southern Punjab.	Islamia University, Bahawalpur	Dr. Muhammd Shafiq Chaudhary	1,712,580	3 Years

364	Micropropagation and cultivation of some economically important plants of Balochistan	University of Balochistan, Quetta.	Dr. Mudassir Asrar	2,512,260	3 Years
385	Impact of agronomic practices on water use efficiency of rainfed wheat.	NWFP Agricultural University,	Dr. Amanullah Jan	1,083,240	3 Years
162	Potential species identification and evaluation as efficient resource capture and utilization for forage in remote sector of NWFP	NWFP Agricultural University, Peshawar	Dr. Muhammad Akmal	1,324,200	3 Years
393	Enhancement of performance of direct seeded rice by seed priming	University of Agriculture, Faislabad	Dr. Shahzad Maqsood Ahmed Basra	1,720,740	3 Years
341	Analytical characterization of some non-conventional oil seed crops of Pakistan for promoting their conservation, utilization and potential oil production	University of Agriculture, Faisalabad	Dr. Farooq Anwar	1,598,340	2 Years
335	Salt-induced oxidative stress, consequences and management	University of Karachi	Dr. Muhammad Ajmal Khan	5,392,607	3 Years
194	Cross Adaptation and in Vitro Development of Salt Tolerant somaclones of Rice	NWFP, Agricultural University, Peshawar	Dr. Safdar Hussain Shah	2,094,060	3 Years
190	Evalution and control of post harvest losses in colocasia	NWFP Agricultural University, Peshawar	Dr. Nawab Ali	497,960	3 Years

Fishing Industry

Principal Investigator	University/Institution	Total of the Project	Approved Amount in Rs.
Dr. Naeem Tariq Narejo, Assistant Professor of Fresh Water Biology and Fisheries	University of sindh	Development of Induceed Breeding Larval and fry rearing Techniques for Indus River Palla. Tenualosa illisha Hasmilton	560,000
Dr. Pirzada J.A. Siddique, Assistant Professor COE in Marine Biology	University Karachi, Karachi	Aqua fed formulation and assessment for Marketing	1,858,000

291	Development of Induced Breeding, Larval and Fry Rearing Techniques for Indus River Palla, Tenualosa ilisha (Hamilton).	University of Sindh, Jamshoro	Dr. Naeem Tariq Narejo	560,000	3 Years

Live Stocks

Principal Investigator	University/Institution	Total of the Project	Approved Amount in Rs.
Dr. Tasawar Hussain Khan, Institute of Pure & Applied Biology	B.Z. University	Some studies to improve meat and milk production to sheep in and around multan by gonadotrophin supplementation	1,692,000
Dr. Aqeel Ahmed, Prof. of Microbiology	University of Karachi, Karachi	Serodiagnosis tests for infectious diseases of the Poultry	1,902,000
Dr. Masood Rabbani, Assistant Professor of Microbiology	University of Veterinary and Animal Sciences, Lahore	Rapiod Diagosis and Immunoprophylaxis of acute viral diseases in Canines.	1,770,000

177	Screening Nematode Populations of sheep and Goats for Development of Anthelmintic Resistance against commonly used anthemintic on Government Livestock Farms of NWFP and Pungab	University of Agriculture, Faisalabad	Dr. Zafar Iqbal	1,350,000	3 Years
97-1	Some studies to improve meat and milk production of sheep in and around Multan by gonadotrophin supplementation.	Bahauddin Zakariya University, Multan	Dr. Tasawar Hussain Khan	1,692,000	3 Years
59	Selection and characterization of starter culture for fermented milk products.	University of Arid Agriculture, Murree Road, Rawalpindi	Dr. Tariq Masud	1,635,500	3 Years
183	Serodiagnostic tests for infectious diseases of the poultry	University of Karachi, Karachi	Dr. Aqeel Ahmed	1,902,000	3 Years

Others

130	Enantioselective and biomematic synthesis of Feralolide, A dihyhdroisocoumarin from cape aloe and its derivatives	Quaid-i-Azam University, Islamabad	Dr. Nasim Hasan Rama	885,000	3 Years
30-1	Isolation and Purification of lipase from common cereals and pulses grown in Pakistan.	Sindh Agriculture University, Tandojam	Dr. Mumtaz Ali Sethar	1,575,000	3 Years
88-1	Production of Lipases and Lipase mediated synthesis of valuable esters.	Government College University, Lahore	Dr. M Akram Kashmiri	1,064,000	2 Years
269	Synthesis of Methoxy indoles of Potential Biological interst.	Allama Iqbal Open University, Islamabad	Professor Dr. Atta-ur-Rehman	5,355,000	3 Years
211	Studies on the phycochemistry of Fresh Water algae of pakistan to facilitate its employment as a sourece of Nutrient & Medicine.	University of Karachi, Karachi	Dr. Mrs.Saleha Hasni	1,108,000	2 Years
31	An alternative approach to managing termites infesting buildings and forest trees; with emphasis on Shisham Dieback treatment.	University of the Punjab, Lahore	Dr. Muhammad Saeed Akhtar	938,000	3 Years
300	The Synthesis of Chiral Compounds of Biological and Synthetic Interest, using Anhydride of <-Tataric Acid	Quaid-i-Azam University, Islamabad	Dr. Javid Hussain Zaidi	3,803,580	3 Years
228	Study on syntheses, characterization and biological evaluation of structurally modified pyrimidine derivatives.	University of Peshawar, Peshawar	Dr. Nazar Ul Islam	1,619,760	3 Years
233	Neutraceutical Products development of industrial importance from Rice Bran	University of Karachi, Karachi	Dr. Rashida Ali	2,304,180	3 Years
212	Utilization of Rice Industrial Wastes for Oil Extraction and Value Added Products	Institute of Food Science and Technology, University of Agriculture, Faisalabad	Dr. Masood Sadiq Butt	489,000	1 Year

Databases produced by Pakistan Agriculture Research Council[2]

PARC DATABASES

RESEARCH

PAKISTAN AGRICULTURE

Database containing bibliographic information of literature published in Pakistan or elsewhere in the world about Pakistan agriculture. It is facilitating the scientists to identify, **locate, and use research literature**. From 1997 abstracts have also been added. Total records in the database are more than 32000.

MEDICINAL PLANTS OF PAKISTAN

A project of National Institute of Agricultural Biotechnology and Genetic Resources, NARC. The database contains detailed information about medicinal plants of Pakistan. It can be searched by Botanical, Local, Urdu,English or Tibb names. The database is in its infancy and a lot more data along with photographs will be provided as and when obtained. Input from visitors would be welcome.

SOCIAL SCIENCES

A database of studies conducted by the Social Sciences Division, PARC. The project has started with data from 2002 onwards. The database also contains summary of research results.

REFERENCE

UNION DATABASE OF JOURNALS

Developed by the Directorate of Scientific Information, NARC. It contains information of about 3000 journal/magazine titles with available volumes and issues in 35 libraries of Pakistan, mostly in the field of agriculture and allied subjects. It contains information of current as well as back volumes.

PLANT GENETIC RESOURCES

This database can be useful for scientists/ researchers who wish to obtain plant genetic material from PARC. Over 15000 records are there in the Database and can be searched by providing the Genus and/or species as input.

AGRICULTURAL LINKAGES PROGRAMME (ALP)

[2] **http://www.parc.gov.pk/database.html**

A searchable database of projects approved and funded by ALP.

PROJECT AND RESEARCH INFORMATION SYSTEM MODULE

RWC-PRISM is an internet-based platform for sharing and managing existing and new information on organizations, experts and projects in agricultural research and development searchable through one window. The system provides overview of projects in the region, persons and organizations associated with them.

Directorate of Scientific Information , NARC is the National Focal Point (NFP) of PRISM for Pakistan. RWC-PRISM derives its mandate from the recommendations made by the Regional Steering Committee of the Rice-Wheat Consortium comprising Bangladesh,India,Neplal and Pakistan, CG Centers(CIMMYT and IRRI) and the representative of the Donor Community (The World Bank) to strengthen the (content) management capacity of the regional NARS and other stakeholders in the Indo-Gangetic Plains by sharing and managing an information system on the agricultural research and development activities in the region. RWC, IAC (International Agricultural Center) and WIS International (Wageningen Information Services International) have catalyzed the process leading to the development of Project and Research Information System Module (PRISM) in close collaboration with WISRD, which was launched in early 2001.

List of on going projects at Aga Khan University

AGA KHAN UNIVERSITY

CURRENT RESEARCH PROJECTS AT AKU

Sl #	Project Title	Department	Prinicipal Investigator
1	Understanding Determinants of Health Seeking Behaviour and Their Relationship to Health Service Utilization using Kroeger's Framework of Health-seeking Behaviour in Northern Areas/Chitral?, Pakistan.	Department of Community Health Sciences	Babar Tasneem Shaikh
2	Tolerance Development and Effects on Anxiety and Reproductive Functions: Mechanisms of Anxiolytics Treatment	Department of Biological & Biomedical Sciences	Arif Siddiqui
3	Does Leptin Play a Role in Sex and Age Related Release of Gonadotropin from the Pituitary Gland?	Department of Biological & Biomedical Sciences	Sheikh Abdul Saeed
4	Studies on Calcium channel Blocking Activities of Indigenous Medicinal Plants	Department of Biological & Biomedical Sciences	Anwar-ul-Hassan Gilani
5	Immunohistochemical & Molecular Characterisation of T-Non-Hodgkin's Lymphoma & its Association with Epstein-Barr Virus	Department of Pathology & Microbiology	Shahid Pervez
6	Laboratory Markers of Risk for Clinical Disease in Household Contacts of Tuberculosis Patients with Active Disease	Department of Pathology & Microbiology	Rabia Hussain
7	Studies on the Lead Deoxifying Potential of Vitamin C Supplementation in Adults	Department of Biological & Biomedical Sciences	Anwar-ul-Hassan Gilani
8	Determination of Therapeutic Potential of Indigenous Medicinal Plants Which Exhibit Antihypertensive Activitiesi (Research Support Programme for Active Scientists and Technologists in Pakistan	Department of Biological & Biomedical Sciences	Anwar-ul-Hassan Gilani
9	Vaccine Development Against Hepatitis-C Virus (Type 3a)	Department of Biological & Biomedical Sciences	Sohail Asif Qureshi
10	Unraveling the Genetic Architecture of Coronary Artery Disease	Department of Biological & Biomedical Sciences	Mohammad Perwaiz iqbal
11	Comparison of 5 vs 10 days of Ceftriaxone Therapy for Bacterial Meningitis in Children	Department of Pediatrics	Zulfiqar A. Bhutta
12	Global Network Research Unit for Perinatal Infections	Department of Community Health Sciences	Anwar Islam
13	Data Management Center at AKU	Department of Pediatrics	Zulfiqar A. Bhutta
14	Socio-cultural and Behavioral Aspects of Shigellosis Disease Burden Studies in Pakistan	Department of Pediatrics	Zulfiqar A. Bhutta
15	Data Management for the IVI Related Projects	Department of Pediatrics	Zulfiqar A. Bhutta
16	Efficacy of a Combined Vi Vaccination and Health Education Program on Reducing the Burden of Typhoid during Childhood: A Demonstrarion Project in Karachi	Department of Pediatrics	Zulfiqar A. Bhutta
17	Estimating Bio-availability of Zinc, Iron, and Vitamin A from Home-availability Complementary Foods in Pakistan: A Community-based Study in Young Infants using Stable Isotopes	Department of Pediatrics	Zulfiqar A. Bhutta
18	Hypertension in the Pakistani Population	Department of Community Health Sciences	Tazeen Jafar
19	Community Based Perinatal and Newborn Care in Rural Pakistan	Department of Pediatrics	Zulfiqar A. Bhutta
20	Efficacy of a Combined VI Vaccination and Health Education Pprogram on Reducing the Burden of Typhoid during Childhood. A Demonstration Project in Bilal Colony, Karachi	Department of Pediatrics	Zulfiqar A. Bhutta

21	Pakistan Bioethics Program - Gateway to the Islamic World	Bioethics Group	Asad Jamil Raja
22	Pakistan Flocculent Health Outcome Study Protocol	Department of Community Health Sciences	Arshad Altaf
23	Population Based Strategies for Effective Control of High Blood Pressure in Pakistan	Department of Community Health Sciences	Tazeen Jafar
24	Estimating the Burden of Unnecessary Injections in Pakistan	Department of Community Health Sciences	Naveed Zafar Janjua
25	Introducing Quality of Care Dimension in HMIS to improve Patient Satisfaction of Services: A study in Malir, Pakistan	Department of Community Health Sciences	Babar Tasneem Shaikh
26	Multicentre Study of Clinical Signs Predicting Severe Illness Requiring Hospitalisation of Young Infants	Department of Pediatrics	Anita Kaniz Mehdi Zaidi
27	Community-based Perinatal and New-born Care in Rural Pakistan: A Cluster Randomized Controlled Trial	Department of Pediatrics	Zulfiqar A. Bhutta
28	Evaluation of Anthelminthics and Multivitamins for Treatment of Severe Anemia in Pregnant Women and Children 6-24 Months of Age in Pakistan	Department of Pediatrics	Zulfiqar A. Bhutta
29	Zinc Supplementation during Acute Childhood Diarrhea: A Cluster Randomised Trial in Rural Pakistan	Department of Pediatrics	Zulfiqar A. Bhutta
30	The Application of Iso topic and Nuclear Techniques in Studies Related to Intrauterine Growth Restriction (IUGR) Causes in Population from developing Countries	Department of Pediatrics	S. Qamruddin Nizami
31	A Placebo-randomized Controlled Trial of Misoprostol in the Management of the Third Stage of Labor in the Home Delivery Setting in Rural Pakistan	Department of Community Health Sciences	Juanita Hatcher
32	Relationship between Salt and Blood Pressure in South Asian Population in Pakistan	Department of Community Health Sciences	Saleem Jessani
33	Community Strategies for Control of Tobacco Use among Adoleseents in Karachi, Pakistan	Department of Community Health Sciences	Tazeen Jafar
34	Assess the Prevalance ot Tobacco & Betel Nut Use amongst the School Children of Karachi & Educate them about its Health Hazards	Department of Medicine	Javaid Ahmed Khan
35	Desinging and Implementing the Nutrition Supplement Acceptability Pilot for Pregnanat and Nursing Mothers using a Fortified Blended Food	Department of Pediatrics	Zulfiqar A. Bhutta
36	Exposure to Indoor Air Pollutants and Birth Outcomes	Department of Pediatrics	Zulfiqar A. Bhutta
37	National Survey on Bioethics Capacity	Bioethics Group	Aamir Mustafa Jafarey
38	Consultancy - Marketing Research for the Demand of Health Services	Department of Community Health Sciences	Naushaba Mobeen
39	National Study of Hepatitis B & C in Pakistan: Lahore & Gujranwala Phase	Department of Community Health Sciences	Naveed Zafar Janjua
40	Engendering the Health System: Improving the Quality of Health Services for Women	Department of Community Health Sciences	Babar Tasneem Shaikh
41	Hepatitis Survey in adults of Jam Kandah, Landhi District, Karachi, Pakistan	Department of Family Medicine	Hasan bin Hamza
42	AKU & International Partners Diabetes & Chronic Disease Management Initiative	Department of Family Medicine	Riaz H Qureshi
43	Multicenter Growth Reference Study	Department of Pediatrics	Uzma Shah
44	Hospital - based Surveillance to Estimate the burden of Severe Rotavrus Gastroenteritis among Children in Pakistan	Department of Pediatrics	Anita Kaniz Mehdi Zaidi
45	Enhanced Surveillance for Invasive Pneumococcal Disease in Children in Sindh, Southern Pakistan.	Department of Pediatrics	Anita Kaniz Mehdi Zaidi
46	Nutrition Supplementation Project for Mothers and Infants in Rural Sindh, Hala and Matiari	Department of Pediatrics	Zulfiqar A. Bhutta

47	Puberty Onset Influence of Environmental and Indigenous Regulators (PIONEER)	Department of Pediatrics	Zulfiqar A. Bhutta
48	Zinc Supplementation during Acute Childhood Diarrhea: a Cluster Randomized Trial in Rural Pakistan.	Department of Pediatrics	Zulfiqar A. Bhutta
49	Pakitan Initiative for Mothers and Newborns (PAIMAN)	Department of Pediatrics	Zulfiqar A. Bhutta
50	The Mapping of Mental Health Research in Low and Middle Income Countries.	Department of Psychiatry	Murad M. Khan
51	Teaching-Learning Issues in Large Classes in Higher Education	Centre for English Language	Nasreen Mujjahida Ahsan
52	Comparison of Physician vs Patient Controlled Analgesia for Pain-free Corporeal Shock Wave Lithotripsy	Department of Surgery	Khurram Siddiqui
53	Influence of Changes in pH on the Contractile Status of Human Tissue	Department of Biological & Biomedical Sciences	Dileep Kumar Rohra
54	Virulence Markers of Helicobacter Pylori in Pakistani Population	Department of Medicine	Javed Yakoob
55	Application of Geographic Information System to Identify High Risk Areas for Road Traffic Injuries in Karachi	Department of Medicine	Junaid Abdul Razzak
56	Assessment of Intra-operative Bleeding Using Serial Hemoglobin and Hemotocrit estimation in craniotomies	Department of Anaesthesiology	Shemila Abbasi
57	Tramadol Consumption in Patient Receiving Rectal Diclofenac Sodium in Gynaecological Surgery.	Department of Anaesthesiology	Aziza Mohammed Hussain
58	Assesing Incidence of Sensorineural Hearing Loss in Patients Receiving Interferon Therapy	Department of Surgery	Shehzad Ghaffar
59	Faculty Students Perception about the Characteristics of Clinical Faculty	School of Nursing	Seema Rehan
60	Integration of Health Promotion in Nursing Practice: A Case Study Approach	School of Nursing	Muneerah Amin Vastani
61	Comparison of Propofol and Ketamine Anaesthesia to Profol alone in Monitored Anaesthesia Care (MAC) for Upper GI Endoscopies	Department of Anaesthesiology	Abdul Monem
62	Virus-induced Cell Profileration and the Pathogenesis of Liver Damage Complicating Chronic Hepatitis C Virus Infection	Department of Medicine	Saeed S.Hamid
63	Is Sentinel Node Biospy a Predictor of Metastic Disease in the Axilla in Primary Breast Cancer?	Department of Surgery	Shaista M. Khan
64	Extent and Factors Associated with Endemic Level of Crimean Congo Haemorrihagic Fever Virus Infection amonfg Selected Risk Groups	Department of Community Health Sciences	Amna Rehana Siddiqui
65	Is Paroxysmal Atrial Fibrillaion an Important cause of Idhemic Stroke?	Department of Medicine	Saad Shafqat
66	Effects of Magnesium in Arrhythmias in Patients Undergoing Bypass Surgery	Department of Anaesthesiology	Mohammad Hamid
67	Genotype :Phenotype Corerlation of Pakistani Patients with Congenital Adrenal Hypeprlasia and Characterization of Sequence Variation in their 21-Hydroxylase Gene	Department of Pathology & Microbiology	Ayesha Habib Khan
68	Effect of Maternal Anxiety and Depression on Child Growth and Development in Two Semi-urban Communities of Karachi	Department of Community Health Sciences	Badar Sabir Ali
69	Effect of Folic Acid on Plasma Homocysteine Level and Arterial Endothelical Function among Diabetics with Acute Coronory Syndrome	Department of Medicine	Syed Muhammad Najaf Ali Nadeem
70	Risk Factors for Intestinal Metaplasia in the Gallbadder Epithelium of Patients with Gallstones	Department of Surgery	Rizwan Azami
71	Learning Dimensions Change Within a Year of Study in AKU Medical College	Department for Educational Development	Jamsheer Talati

72	Effects of length of epidural catheter in epidural space on pain relief in patients undergone total abdominal hysterectomy - randomized double blinded study	Department of Anaesthesiology	Gauhar Afshan
73	Characterisation of M.tuberculosis Beijing strains in Pakistan using molecular and immunological techniques	Department of Pathology & Microbiology	Rumina Hassan
74	Mycobacteria tuberculosis-induced host cellular activation. Effect on cytokine expression and signalling.	Department of Pathology & Microbiology	Zahra Hasan
75	Screening of the Paraoxonase Gene Cluster for Novel Mutations in Patients with Ischemic Stroke	Department of Biological & Biomedical Sciences	Philippe M. Frossard
76	Knowledge of breast cancer, and knowledge and Practice of Breast Self-examination, and the Associated Factors among Female Nurses of Karachi	Department of Community Health Sciences	Faiza Ahmed
77	Study of Depolarisation Vectors in Myocardial Ischaemia	Department of Medicine	Camer Vellani
78	Serum Chromagranin A and Prostate Specific Antigen (PSA): Predict Disease Progression Better than PSA alone in Advanced Prostate Cancer	Department of Surgery	M. Hammad Ather
79	Human Papillomavirus (HPV) Association and p53 Mutation in Oral Cavity (Squamous Cell Carcinoma) of Pakistani Patients: Its Correlation with Histologic Variables and Disease Outcome	Department of Surgery	Sohail Awan
80	CFTR Mutation Analysis: Effect of the Genotype on Phenotype and Mortality in Pakistani Patients with Cystic Fibrosis	Department of Pediatrics	Uzma Shah
81	Development of Growth Curves for Identification of Fetal Growth Restriction.	Department of Obstetrics & Gynaecology	Shama Munim
82	The ATP-binding cassette transporter gene, VMD2 gene and Macular Diseases	Department of Biological & Biomedical Sciences	Philippe M. Frossard
83	Development of Stress Scale for Pregnant Women in Pakistan	Department of Community Health Sciences	Syed Zafar Ahmed Fatmi
84	Assessing Processes and impact of Community-based early Childhood Development Interventions: Phase 2 of AKU's ECD Study 2001-02	Department of Community Health Sciences	Bilal Iqbal
85	Recrudescence and Reinfection of H.pylori after Successful Eradication	Department of Medicine	Shahab Abid
86	Analysis NPC 1 of Expression by Engrafted Neural Stem Cells and Clinical Outcome in a Niemann-Pick Type C Disease Model	Department of Surgery	Syed Ather Enam
87	Association of Genetic Diversity with Virulence and Antimicrobial Resistance in MTB Isolates at AKUH	Department of Biological & Biomedical Sciences	Asho Ali
88	Prostate Cancer: Role of Molecular Markers in Diagnosis and Prognosis	Department of Biological & Biomedical Sciences	Ahmed Yakinuddin
89	Mitochondrial DNA Haplogroups among Pakistani Infertile Males and its Association with Poor Semen Parameter	Department of Biological & Biomedical Sciences	Taranum Sultana Zahid
90	Role of the Transcription Factors Bach1 and MafK in Epileptogenesis	Department of Biological & Biomedical Sciences	Sohail Asif Qureshi
91	Screening Pakistani Patients for Polycystic Kidney Disease 1 (PKD1) Gene Mutations	Department of Biological & Biomedical Sciences	Saima Naheed Waqar
92	Role of the Neuronal Restrictive Silencing Factor Splice Variant REST4 in Neoplasia	Department of Pathology & Microbiology	Romena Qazi
93	Identification of G Protein Subunits in Healthy and Osteoarthritic Articular Chondrocytes	Department of Biological & Biomedical Sciences	Khalid Khan
94	Living with Breast Cancer: Experiences of Pakistani Women on Chemotherapy	School of Nursing	Ann Rose Castellino
95	Risk Factors, Antibiotic Susceptibility and Molecular Differences between Community Acquired and Hospital Acquired Methicillin Resistant Staphylococcus Aureus	Department of Pathology & Microbiology	Affia Zafar
96	Detection of Hepatitis C Core Antigen in Pakistani Blood Donors	Department of Pathology & Microbiology	Ghulam Nabi Kakepoto

97	Genetic Diversity of Plasmodium Falciparum Candidate Antigen Genes among Field Isolates from Karachi, Pakistan	Department of Pathology & Microbiology	Mohammad Asim Beg
98	Sentinel Node Biopsy as a Predictor of Metastatic Disease in Cervical Lymph Node from the Cancer of Oral Cavity	Department of Surgery	Mubasher Ikram
99	Risk Factors for Preeclampsia in Both Nulliparous and Parous Women in Maternity Hospitals	Department of Community Health Sciences	Uzma Shamsi
100	Cell Cycle Responses of Hepatocytes during Chronic Hepatitis C Virus Infection	Department of Biological & Biomedical Sciences	Saira Sarfraz

(Courtesy of Dr. Anwar Ali Siddiqui, Associate Dean, Research Affairs, AKU)

BIOINFORMATICS TOOLS AND SOFTWARES[3]

Bioinformatic tools are software programs that are designed for extracting the meaningful information from the mass of molecular biology / biological databases & to carry out sequence or structural analysis.

Factors that must be taken into consideration when designing bioinformatics tools, software and programmes are:

- The end user (the biologist) may not be a frequent user of computer technology
- These software tools must be made available over the internet given the global distribution of the scientific research community

Major categories of Bioinformatics Tools:

There are both standard and customized products to meet the requirements of particular projects. There are data-mining software that retrieve data from genomic sequence databases and also visualization tools to analyze and retrieve information from proteomic databases. These can be classified as homology and similarity tools, protein functional analysis tools, sequence analysis tools and miscellaneous tools.

Here is a brief description of a few of these, everyday bioinformatics is done with sequence search programs like BLAST, sequence analysis programs, like the EMBOSS and Staden packages, structure prediction programs like THREADER or PHD or molecular imaging/modelling programs like RasMol and WHATIF.

Homology and Similarity Tools:

Homologous sequences are sequences that are related by divergence from a common ancestor. Thus the degree of similarity between two sequences can be measured while their homology is a case of being either true of false. This set of tools can be used to identify similarities between novel query sequences of unknown structure and function and database sequences whose structure and function have been elucidated.

Protein Function Analysis:

This group of programs allow you to compare your protein sequence to the secondary (or derived) protein databases that contain information on motifs, signatures and protein domains. Highly significant hits against these different pattern databases allow you to approximate the biochemical function of your query protein.

[3] Suresh Kumar (2005). *Bioinformatics web.* Retrieved August 14, 2005 from: http://www.geocities.com/bioinformaticsweb/

Structural Analysis:

This set of tools allow you to compare structures with the known structure databases. The function of a protein is more directly a consequence of its structure rather than its sequence with structural homologs tending to share functions. The determination of a protein's 2D/3D structure is crucial in the study of its function.

Sequence Analysis:

This set of tools allows you to carry out further, more detailed analysis on your query sequence including evolutionary analysis, identification of mutations, hydropathy regions, CpG islands and compositional biases. The identification of these and other biological properties are all clues that aid the search to elucidate the specific function of your sequence.

Some examples of Bioinformatics Tools:

BLAST:

BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) comes under the category of homology and similarity tools. It is a set of search programs designed for the Windows platform and is used to perform fast similarity searches regardless of whether the query is for protein or DNA. Comparison of nucleotide sequences in a database can be performed. Also a protein database can be searched to find a match against the queried protein sequence. NCBI has also introduced the new queuing system to BLAST (Q BLAST) that allows users to retrieve results at their convenience and format their results multiple times with different formatting options.

Depending on the type of sequences to compare, there are different programs:

- blastp compares an amino acid query sequence against a protein sequence database
- blastn compares a nucleotide query sequence against a nucleotide sequence database
- blastx compares a nucleotide query sequence translated in all reading frames against a protein sequence database
- tblastn compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames
- tblastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

FASTA:

FAST homology search **A** ll sequences .An alignment program for protein sequences created by Pearsin and Lipman in 1988. The program is one of the many heuristic algorithms proposed to speed up sequence comparison. The basic idea is to add a fast prescreen step to locate the highly matching segments between two sequences, and then extend these matching segments to local alignments using more rigorous algorithms such as Smith-Waterman.

EMBOSS:

EMBOSS (**E**uropean **M**olecular **B**iology **O**pen **S**oftware **S**uite) is a software-analysis package. It can work with data in a range of formats and also retrieve sequence data transparently from the Web. Extensive libraries are also provided with this package, allowing other scientists to release their software as open source. It provides a set of sequence-analysis programs, and also supports all UNIX platforms.

Clustalw:

It is a fully automated sequence alignment tool for DNA and protein sequences. It returns the best match over a total length of input sequences, be it a protein or a nucleic acid.

RasMol:

It is a powerful research tool to display the structure of DNA, proteins, and smaller molecules. Protein Explorer, a derivative of RasMol, is an easier to use program.

PROSPECT:

PROSPECT (PROtein Structure Prediction and Evaluation Computer ToolKit) is a protein-structure prediction system that employs a computational technique called protein threading to construct a protein's 3-D model.

PatternHunter :

PatternHunter, based on Java, can identify all approximate repeats in a complete genome in a short time using little memory on a desktop computer. Its features are its advanced patented algorithm and data structures, and the java language used to create it. The Java language version of PatternHunter is just 40 KB, only 1% the size of Blast, while offering a large portion of its functionality.

COPIA :

COPIA (COnsensus Pattern Identification and Analysis) is a protein structure analysis tool for discovering motifs (conserved regions) in a family of protein sequences. Such motifs can be then used to determine membership to the family for new protein sequences, predict secondary and tertiary structure and function of proteins and study evolution history of the sequences.

Application of Programmes in Bioinformatics:

JAVA in Bioinformatics:

Since research centers are scattered all around the globe ranging from private to academic settings, and a range of hardware and OSs are being used, Java is emerging as a key player in

bioinformatics. Physiome Sciences' computer-based biological simulation technologies and Bioinformatics Solutions' PatternHunter are two examples of the growing adoption of Java in bioinformatics.

Perl in Bioinformatics:

String manipulation, regular expression matching, file parsing, data format interconversion etc are the common text-processing tasks performed in bioinformatics. Perl excels in such tasks and is being used by many developers. Yet, there are no standard modules designed in Perl specifically for the field of bioinformatics. However, developers have designed several of their own individual modules for the purpose, which have become quite popular and are coordinated by the BioPerl project.

Bioinformatics Projects:

BioJava:

The BioJava Project is dedicated to providing Java tools for processing biological data which includes objects for manipulating sequences, dynamic programming, file parsers, simple statistical routines, etc.

BioPerl:

The BioPerl project is an international association of developers of Perl tools for bioinformatics and provides an online resource for modules, scripts and web links for developers of Perl-based software.

BioXML:

A part of the BioPerl project, this is a resource to gather XML documentation, DTDs and XML aware tools for biology in one location.

Biocorba:

Interface objects have facilitated interoperability between bioperl and other perl packages such as Ensembl and the Annotation Workbench. However, interoperability between bioperl and packages written in other languages requires additional support software. CORBA is one such framework for interlanguage support, and the biocorba project is currently implementing a CORBA interface for bioperl. With biocorba, objects written within bioperl will be able to communicate with objects written in biopython and biojava (see the next subsection). For more information, see the biocorba project website at http://biocorba.org/ . The Bioperl BioCORBA server and client bindings are available in the bioperl-corba-server and bioperl-corba-client bioperl CVS repositories respecitively. (see http://cvs.bioperl.org/ for more information).

Ensembl :

Ensembl is an ambitious automated-genome-annotation project at EBI. Much of Ensembl\'s code is based on bioperl, and Ensembl developers, in turn, have contributed significant pieces of code to bioperl. In particular, the bioperl code for automated sequence annotation has been largely contributed by Ensembl developers. Describing Ensembl and its capabilities is far beyond the scope of this tutorial The interested reader is referred to the Ensembl website at http://www.ensembl.org/.

bioperl-db:

Bioperl-db is a relatively new project intended to transfer some of Ensembl's capability of integrating bioperl syntax with a standalone Mysql database (http://www.mysql.com) to the bioperl code-base. More details on bioperl-db can be found in the bioperl-db CVS directory at http://cvs.bioperl.org/cgi-bin/viewcvs/viewcvs.cgi/bioperl-db/?cvsroot=bioperl . It is worth mentioning that most of the bioperl objects mentioned above map directly to tables in the bioperl-db schema. Therefore object data such as sequences, their features, and annotations can be easily loaded into the databases, as in $loader->store($newid,$seqobj) Similarly one can query the database in a variety of ways and retrieve arrays of Seq objects. See biodatabases.pod, Bio::DB::SQL::SeqAdaptor, Bio::DB::SQL::QueryConstraint, and Bio::DB::SQL::BioQuery for examples.

Biopython and biojava:

Biopython and biojava are open source projects with very similar goals to bioperl. However their code is implemented in python and java, respectively. With the development of interface objects and biocorba, it is possible to write java or python objects which can be accessed by a bioperl script, or to call bioperl objects from java or python code. Since biopython and biojava are more recent projects than bioperl, most effort to date has been to port bioperl functionality to biopython and biojava rather than the other way around. However, in the future, some bioinformatics tasks may prove to be more effectively implemented in java or python in which case being able to call them from within bioperl will become more important. For more information, go to the biojava http://biojava.org/ and biopython http://biopython.org/ websites.

GLOSSARY OF IMPORTANT TERMS

A

- **autoclave:** a container for sterilizing by superheated steam under pressure.
- **amino acid:** any of a large group of organic acids containing a carboxylic group, COOH, and an amino group, NH_2, that link together into polypeptide chains to form proteins.
- **algorithms:** any systematic method of solving a certain kind of problem.
- **automation:** an electronic machine or control device equipped with a computer and designed to operate automatically in response to instructions previously fed into the computer.
- **allergy:** a hypersensitivity to a specific substance or condition which in similar amounts or degrees is harmless to most people: it is manifested in physiological disorder.
- **agriculture:** the science and art of farming; work of cultivating the soil, producing crops, and raising livestock.
- **abinito:** from the beginning.

B

- **bioinformatics:** the application of computer technology to the management of biological information.
- **biotechnology:** a set of techniques that involve manipulation or change of the genetic patrimony of living organisms.
- **biomaterial:** a synthetic or natural material used to replace a bone or tissue in a living body.
- **biostatics:** the branch of biometrics dealing with demography, especially vital statistics.
- **biodiversity:** diversity, or variety, in the living things in a particular area or region.
- **biological:** of or connected with biology; of plants and animals.
- **bioscience:** any science that deals with the functions or problems of living organisms.
- **biomarker:** a measurable phenomenon that indicates the presence of life.
- **biophysics:** the study of biological phenomena using the principles and techniques of physics.
- **black box testing:** a method of testing where the inputs are tested against the outputs without reference to the genetic makeup of the body to evaluate the impact of the drug.

C

- **chromosome:** any of the microscopic rd-shaped bodies formed by the incorporation of the chromatin in a cell nucleus during mitosis and meiosis: they carry the genes that convey hereditary characteristics, and are constant in number for each species.
- **centrifuge:** a machine using centrifugal force to separate particles of varying density, or to draw off moisture.
- **cell:** a very small, complex unit of protoplasm, usually with a nucleus, cytoplasm and an enclosing membrane: all plants and animals are made up of one or more cells that usually combine to form various tissues.
- **cell biology:** the branch of biology that deals with cells and their physiological properties.
- **cardio-vascular disorder:** a disorder of the heart and the blood vessels.
- **cancer:** any of various diseases characterized by the uncontrolled growth of cells that disrupt body tissue and metabolism, and tend to spread locally and to distant parts of the body.
- **crystallography:** the science of the form, structure, properties, and classification of crystals.
- **compound:** a substance containing two or more elements chemically combined in fixed proportions.
- **cardiac:** of, near, or affecting the heart.
- **cDNA library:** a complete, or nearly complete, set of all the mRNAs contained within a cell or organism.
- **cardiology:** the branch of medicine dealing with the heart, its functions, and its diseases.

D

- **diagnosis:** the act or process of deciding the nature of a diseased condition by examination of the symptoms.
- **data mining:** the analyses of data and the use of software techniques for finding patterns and regularities in different data sets.
- **DNA:** deoxyribonucleic acid.
- **diabetes:** any of various disease characterized by an excessive discharge of urine.
- **drug:** any substance used as a medicine or as an ingredient in a medicine which kills or inactivates germs, or affects any body function or organ.
- **disease:** a particular destructive process in an organ or organism, with a specific cause and characteristic symptoms; specifically, an illness.

E

- **enzyme:** any of various proteins, formed in plant or animal cells or made synthetically, that act as organic catalysts in initiating or speeding up specific chemical reactions.

- **electrophoresis:** A method of separating biological molecules like proteins, and analyzing molecular structure of proteins and DNA based on the rate of movement of each component in a colloidal suspension while under the influence of an electric field.
- **endocrinology:** the branch of medicine dealing with the endocrine glands and their hormones.
- **empirical data:** data relying or based solely on experiment and observation rather than theory.
- **exon:** a sequence in the genetic code that supplies the information for protein formation.
- **ecology:** the branch of biology that deals with the relations between living organisms and their environment.
- **evolution:** the development of a species, organism, or organ from its original or primitive state to its present or specialized state.
- **endemic:** constantly present in a particular region: said of a disease that is generally under control.
- **epidemiology:** the branch of medicine that investigates the causes and control of epidemics.
- **extremophiles:** organisms living in extreme conditions.
- **ECG:** electrocardiogram.

F

- **forensic science:** the application of scientific, especially medical, knowledge to legal matters, as in the investigation of crime.
- **functional genomics:** an ability to accurately predict the functions of the genes after having analyzed its sequence.
- **fermentation:** the breakdown of complex molecules in organic compounds, caused by the influence of a ferment.
- **flavanone:** a complex, colorless, crystalline ketone, $C_{15}H_{12}O_2$, derived from flavone, which is obtained from certain plants.

G

- **genotype:** the fundamental constitution of an organism in terms of its hereditary factors.
- **gene annotation:** finding genes in the DNA sequence of various organisms.
- **genome:** one complete haploid set of chromosomes of an organism.
- **gene:** any of the units occurring at specific points on the chromosomes, by which hereditary characters are transmitted and determined: each is regarded as a particular state of organization of the chromatin in the chromosome, consisting primarily of DNA and protein.
- **gladiolas:** floral productions.

- **gastroenterology:** the medial specialty that is concerned with disorders of the digestive system.

H

- **heterogenous:** of different origin; not from the same source, individual or species.
- **hyberidization:** the process of producing an offspring by crossing two individuals of unlike genetic constitution.
- **homologous:** corresponding in basic type of structure and deriving from a common primitive origin.
- **horticulture:** the art or science of growing flowers, fruits, vegetables, and shrubs, especially in gardens or orchids.
- **herbicide:** any chemical substance used to destroy plants, especially weeds, or to check their growth.
- **hypothesis:** an unproved theory tentatively accepted to explain certain facts or to provide a basis for further investigation.
- **hepatitis:** inflammation of the liver.

I

- **IT:** information technology.
- **in vivo:** occurring within the living organism.
- **in vitro:** isolated from the living organism and artificially maintained, as in a test tube.
- **in silico:** informatics centered.
- **intron:** an intervening sequence in the eukaryote genetic code, interrupting protein formation.
- **infection:** the fact or state of being infected, especially by the presence in the body of bacteria, protozoans, viruses, or other parasites.
- **immune system:** the system that protects the body from disease by producing antibodies.
- **incubator:** an apparatus for growing microbic or cell cultures in which the temperature, atmosphere, and humidity can be controlled.
- **inhibitor:** any substance that slows or prevents a chemical or organic reaction.
- **inoculation:** the injection of a disease agent into an animal or plant, usually to cause a mild form of the disease and build up immunity to it.

M

- **molecule:** the smallest particle of an element or compound that can exist in the free state and still retain the characteristics of the element or compound.
- **metabolism:** the chemical and physical process continuously going on in living organisms and cells, consisting of anabolism and catabolism

- **medicine:** the science and art of diagnosing, treating, curing, and preventing disease, relieving pain, and improving and preserving health.
- **monograph:** a treatise on a single genus, species etc of plant or animal.
- **molecular medicine:** a branch of medicine that studies the chemical and physical principles associated with the composition, properties, and activities of molecules in living cells.
- **micro automation:** an automated tracking system linked to dedicated units which allow scientists to centrifuge, sort and deliver samples to different analyzers to develop a continuous flow system.
- **macromolecule:** a very large molecule, as a protein or polymer molecule, composed of hundreds of thousands of atoms.
- **morbidity:** the rate of disease or proportion of diseased persons in a given locality or nation.
- **microbe:** a microscopic organism, especially any of the bacteria that cause disease.
- **micropropagation:** a tissue culture technique for plant propagation in which offspring are cloned from tissue taken from a single plant.
- **microbiology:** the branch of biology that deals with microorganisms.

N

- **nuclear magnetic resonance:** a phenomenon exhibited by various atomic nuclei when placed in a strong magnetic field, in which they absorb energy from specific, high frequency radio waves.
- **neuroscience:** any science dealing with the functions, abnormalities etc of the nervous system.
- **nanotechnology:** a hypothetical method or process of creating micro miniature equipment by manipulating atoms and molecules as if they were parts of a machine.
- **neurobiology:** a branch of biology that deals with the nervous system and its ability to react, learn etc.
- **nascent:** the state of an element just released from a compound and having unusual chemical activity because atoms of the element have not combined to form molecules.
- **neurology:** the branch of medicine dealing with the nervous system, its structure, and its diseases.

O

- **obsolescence:** the state of being vestigial, said of an organ.
- **oligonucleotide:** a polymeric chain of two to ten nucleotides.
- **ontology:** the branch of metaphysics dealing with the nature of being, reality, or ultimate substance.
- **oncology:** the branch of medicine dealing with neoplasms.

- **opthalmology:** the branch of medicine that deals with the anatomy, functions, pathology, and treatment of the eye.

P

- **phenotype:** physical attribute of a living being.
- **protein:** any of a large class of nitrogenous substances consisting of a complex union of amino acids and containing carbon, hydrogen, nitrogen, oxygen, frequently sulfur, and sometimes phosphorous, iron, iodine, or other elements.
- **pathology:** the branch of medicine that deals with the nature of disease, especially with the structural and functional changes caused by the disease.
- **pharmaceutical industry:** the industry carrying out the art or profession of preparing and dispensing drugs and medicines.
- **pathogenesis:** the production or development of a disease.
- **pharmacogenomics:** the study of genetic variation as revealed by various reactions to a drug.
- **primary sequence (of protein):** the types of amino acids contained in the polypeptide chain, and the sequence in which they are joined.
- **photolithography:** a process of printing from a plate, prepared by methods combining photography and lithography.
- **preclinical:** of or in the period of a disease before any of the symptoms appear.
- **petrochemical:** a chemical derived ultimately from petroleum or natural gas, as an aliphatic or aromatic hydrocarbon.
- **prenatal:** before birth or during pregnancy.
- **physico-chemical:** of or pertaining to both physical and chemical properties, changes and reactions.
- **parasite:** a plant or animal that lives on or in an organism of another species from which it derives sustenance or protection without benefit to, and usually with harmful effects on, the host.
- **phylogeny:** the lines of descent or evolutionary development of any plant or animal species.
- **photometer:** an instrument used in measuring of intensity of light, especially in determining its relative intensity from different sources.
- **pest:** any destructive or troublesome insect, small animal or weed.
- **pharmacology:** the study of the preparation, qualities and uses of drugs.
- **psychiatry:** the branch of medicine concerned with the study, treatment, and prevention of disorders of the mind.

R

- **RNA:** ribonucleic acid.
- **robotics:** the science or technology of robots, their design, manufacture, application and use.

- **reagent:** a substance used to detect or measure another substance or to convert one substance into another by means of the reaction which it causes.
- **radiology:** the science dealing with X-rays and other forms of radiant energy, especially as used in medicine for X-raying bones, organs, etc. and for diagnosing and treating disease.
- **recombinant DNA technology:** the technology by which a recombined DNA is formed in the laboratory by splicing together pieces of DNA from different species, as to create new life forms, modifying existing ones, or produce useful biological chemicals.

S

- **Single Nucleotide Polymorphisms (SNPs):** the variations, which exist in each individual, accounting for the bewildering human variations in physical attributes, disease susceptibility and possibly behavior.
- **species:** a naturally existing population of similar organisms that usually interbreed only among themselves, and are given a unique, Latinized binomial name to distinguish them from all other creatures.
- **saliva:** the thin, watery, slightly viscid fluid secreted by the salivary glands.
- **structural biology:**
- **spectroscopy:** the study of spectra by use of the spectroscope.
- **stem cell:** a cell that has the ability to continuously divide and differentiate (develop) into various other kind(s) of cells/tissues.

T

- **thermodynamic:** caused or operated by heat converted into motive power.
- **therapy:** the treatment of disease or of any physical or mental disorder by medical or physical means, usually excluding surgery.
- **taxonomy:** a system of arranging animals and plants into natural, related groups based on some factors common to each, as structure, embryology, or biochemistry.
- **therapeutics:** the branch of medicine that deals with the treatment or cure of diseases.
- **tertiary structure (of protein):** the way in which a protein coils up to form a precise three-dimensional shape.
- **typhoid:** an acute infectious disease caused by a bacillus (Salmonella typhi) and acquired by ingesting food or water contaminated by excreta.
- **tuberculosis:** an infectious disease caused by the tubercle bacillus and characterized by the formation of tubercles in various tissues of the body.

U

- **urine:** a waste product of vertebrates and many invertebrates, secreted by the kidneys or other excretory structures.

V

- **virus:** any of a kingdom (Virus) of prokaryotes, usually ultramicroscopic, that consist of nucleic acid, either RNA or DNA, within a case of protein.
- **vaccine:** any preparation of killed microorganisms, living weakened organisms, etc. introduced into the body to produce immunity to a specific disease by causing the formation of antibodies.
- **vector:** an animal, especially an insect, that transmits a disease-producing organism from a host to a non-infected animal.

W

- **whole blood:** blood for transfusion from which none of the elements have been removed.

List of Seminar Participants

(14 September 2005)

S. #	Name	Organization	Contact No.
1.	Dr. Syed Ali	Aga Khan University	
2.	Naseem Siddiqui	Aga Khan University	
3.	Rashid Gadit	Aga Khan University	
4.	Dr. Sohail Qureshi	Aga Khan University	
5.	Dr. Saqlain Naqvi	Arid Agriculture Unviersity	0333-5187253
6.	Asif Mushtaq	Arulogics, CEO	0345-5177306
7.	Zia Mohammad Khan	Axilogics	0345-5165326
8.	Rauf Malik	Bahria, Karachi	
9.	Dr. Shaukat Malik	BUITMS, Quetta	
10.	Dr. Imtiaz Taj	CARE	
11.	Adeel Yusuf (Student)	CASE	
12.	Dr. Anwar Nasim	COMSTECH	
13.	Dr. Qasim Sheikh	FAST	
14.	Dr. Qasim Sheikh	FAST	
15.	Farheem Shahid	FAST	0300-5123184
16.	Aleem Ahmed	Global Science, Karachi	
17.	Dr. Sameera Gulshad	Global Services	0300-5194227
18.	Dr. Zaheer ul Haq	HEJ	0300-2512436
19.	Rabia Sattar	HEJ	0333-2267466
20.	Dr. Shamshad Zarina	Karachi University	
21.	Ayesha Fatima	MAJU	
22.	Dr Sehar Fazal	MAJU	0300-7350918
23.	Naveed Ikram	MAJU	
24.	Students (19)	MAJU	
25.	Tahir Nijazi	MAJU	
26.	Zahid Mirza	NETSOL	051-2829972
27.	Dr. Afab Bashir	NIBGE	
28.	Jaffar Hussain	NIBGE	041-2651475
29.	M. Taha Masood	NUST	0301-5107558
30.	M. Masood Akhtar	PPS	
31.	Amir Bakht	Ultimus, Islamabad	0300-5542239
32.	Dr. Saleem Tariq	UMT, Lahore	
33.	Mamoon Khan	Univeristy of Engineering & Tech. Lahore	
34.	Mr. Humayun Ajmal	Univeristy of Engineering & Tech. Lahore	
35.	Mr. Mudassir Rasool	Univeristy of Engineering & Tech. Lahore	
36.	Bushra Saeed		0300-2783384
37.	Dr. Farhan Cyprian	Ferozsons Labs, Islamabad	0300-5065669
38.	Mrs. Azra Masood		

About the Authors and the Project Team

The present report is the culmination of a project undertaken by Gallup to research on the field of bioinformatics, its application areas and the role of IT. Key researchers include Dr. Ahmed Gilani and Ms. Sadia Zaidy.

Dr. Ahmad Ijaz Gilani completed medicine from Aga Khan University in 2004. Dr. Gilani has been a visiting research fellow at various labs at Harvard University, Boston and Mount Sinai Medical Centre, New York. His research interests include molecular medicine, public health, genetics of hypertension, diabetes and neurological diseases. His future plans include graduate studies in neuroscience and knowledge based entrepreneurship possibly in bioinformatics. He can be reached at aigilnai@yahoo.com

Sadia Zaidy did BSc (Hons) from Lahore University of Management Sciences (LUMS) in 2004 with majors in Computer Science. Currently she is Assistant Manager, Research Department at Gallup Lahore. Some of her research has been on using designs borrowed from biological systems to improve business organization models. She has done work on the new upcoming technique of detecting, rectifying errors/online frauds using the concept of Computational Immunology. Her future plans include graduate studies in Business Administration. She can be contacted at sadia.zaidy@gallup.com.pk and sadiazaidy@hotmail.com

Dr. Qasim Sheikh and Dr. Ozair-ul-Ghani have been the mentors during the project. Dr. Uzair-ul-Ghani generously provided relevant literature on Bioinformatics and Dr. Qasim Sheikh's constructive criticism has helped in bringing out a more useful report. Dr. Arif Siddiqui and Dr. Anwar Ali Siddiqui at Aga Khan University have also provided useful information on the field of Biological Sciences in the country.

Supervisor:	Dr. Ijaz Shafi Gilani
Researchers:	• Dr. Ahmed Gilani • Ms. Sadia Zaidy
Project Manager:	Ms. Fatima, Idrees
Consultant:	Professor Zafar I. Malik
Interviewers:	• Dr. Farhan Cyprin • Dr. Shehryar Hussain • Mr. Khurram Javed Mir • Mr. Usman Jawaid
Interns:	• Ms. Sana Aizad, MAJU • Ms. Mahreen Ali, MAJU • Ms. Shumaila Sayyab, MAJU • Ms. Hina Naz, MAJU • Mr. Harris Kamran, Beaconhouse, Islamabad
PSEB Team:	• Dr. Aamir Matin • Mr. Nasir Afridi • Mr. Omar Hasan • Mr. Salahuddin Yusuf • Ms. Sabahat Majeed • Ms. Kiran Afsheen