Scientists that had spent 40 years looking at proteins through a microscope. And today with the merger of IT and the scientific community, you can literally--what was done before in 30 years can be done in years or months and, in some cases, in days.
Mick Gallagher, Manager of Life Sciences, Oracle

In recent years, our advances in technology have allowed us direct that computing power inward in an attempt to learn more about ourselves. The merger of biology and technology has generated enormous amounts of data on how we humans are constructed. Bioinformatics is termed coined for the science of culling useful information from all this data.

What is Bioinformatics?

Bioinformatics is a combination of Computer Science, Information Technology and Genetics to determine and analyze genetic information. Bioinformatics deals with the issues created by the massive amounts of new types of data obtained through novel biological experiments. The problem and the promise in bioinformatics are best understood in the context of the Human Genome Project. The project resulted in 3.3 billion base-pair DNA sequence of the entire human genome, as pre-released by an international consortium of 16 institutes on 26 June 2000.

DNA sequences are valuable because they provide the most detailed anatomy possible for any organism - the blueprint of life with the complete set of instructions for how each working part should be assembled and operate. Yet while identifying the order of letters in our genetic alphabet is an important and necessary first step, a much more complicated task lie ahead. This is to figure out what those letters mean, what they do, and what can be done if the messages they spell out are in error - that is, laying bare the genetic triggers for hundreds of diseases of genetic predisposition

Three levels of bioinformatics:

1. Analysis of a single gene sequence. For example:
   - Similarity with other known genes
   - Evolutionary relationships -- Phylogenetic trees
   - Identification of well-defined domains in the sequence
   - Sequence features (physical properties, binding sites, modification sites...)

2. Analysis of complete genomes. For example:
   - Linking gene families, identifying missing ones
   - Gene location on the chromosomes, correlation with function or evolution
   - Large-scale events in the evolution of organisms

3. Analysis of genes and genomes with respect to functional data. For example:
    - Identification of essential genes, or genes involved in specific processes
    - Deletion or mutant genotypes vs. phenotypes

Industry relevance

Bioinformatics is of direct relevance for the pharmaceutical, agricultural & bioinformatics company. In the pharmaceutical industry for example, d*ata analysis* is an essential and defining constituent. For the Pharmaceutical company the goal is to:

- Identify patterns in this information that can be used to develop more effective therapeutics--drugs that work more quickly, are safer, are less toxic, and have better bioavailability.
- Manage the rapidly expanding information available in genomic and other databases
- Screen immense libraries of natural and chemical compounds for desirable biological and therapeutic activity.

Bioinformatics companies acquire clinical and late pre-clinical stage compounds with defined pathways to the clinic and the market and develop products and leading-edge technologies with large, unsatisfied markets or broad applications.

Research firms indicate the market for bioinformatics software and services are growing at 17% annually. Frost & Sullivan, a research and consulting firm, has predicted the growth potential to be as much as $5 billion in the next five years. In comparison, the global pharmaceutical industry is worth more than $150 billion per year.

The appeal of this technology has drawn the attention (and dollars) of software giants like Microsoft, Oracle, and IBM. IBM, for example, has invested more than $100 million in its new life-sciences division, and has 100 bioinformatics researchers who customize and test drug discovery products for biotech companies.

Challenges

The fundamental challenges of bioinformatics lies in how we describe, analyze, simulate and predict the dynamics of life processes. The body contains more than 1 million proteins, which regulate the structure of cells and tissues and trigger or deactivate genetically linked diseases such as Alzheimer's, cancer, and diabetes. Knowledge of protein structures is the key to understanding these diseases. Proteins, not genes, are the endpoints of life sciences investigations, since proteins ultimately regulate metabolism and disease in the body. Drug makers believe that an understanding of proteins will lead to new therapies that will revolutionize the way disease is diagnosed and treated

The need is for a common language to standardize disparate pieces of data and a set of computational tools that will help interpret these data. Although some current technology

2601 Floyd Road M/S SM23
PO BOX 830688
Richardson, Texas 75083-0688
School of Management
The University of Texas at Dallas
citm@utdallas.edu
http://citm.utdallas.edu
(972) 883-4755 (v) 4754 (f)

developments like XML do help, the fragmentation is more fundamental than that. It is due to different ways of viewing the data, and this is a genuine scientific problem.

One of the most important challenges is management of data. The mapping of human DNA produced reams of data that is more than a match for even today's most sophisticated computers. The human genome represents around 300 terabytes of data. The complete Library of Congress can be stored in about 3 terabytes. In the future as the study on proteins progresses, the data volumes are going to expand exponentially. According to industry experts, "If dealing with the genome is a terabyte problem, dealing with the proteome (the protein "map") will be a hundreds-of-terabytes problem".

The Future

The societal impact of this technology will be immense. One day, we can expect doctors to use a computer to see your genetic profile, predict what diseases you are likely to get later in life and give you personalized drugs to prevent them.

Future experts in biotechnology predict, the industry shift from one that is dependent on chemistry to one that will be centered towards information technology

Bioinformatics is the future of the life sciences industry. There are of course other issues to be concerned with, such as scale up from research to production environments and the economies and accuracy of scale which are required. However, this technology also comes with serious ethical issues to be resolved by society before the technology is applied.

Also some side effects of this modern science such as patenting a part-human animal, creating a child with possibly eight parents, or being tested secretly for genetic disease need to be addressed. These possibilities will raise some inevitable questions; Do advances in biotechnology confuse or clarify what it means to be human, to be a parent, to inhabit a body?

Mihir Jani
mihirjani@hotvoice.com

More Information on Bioinformatics can be obtained from the following links

http://www.cshl.org/public/genome.html --Cold Spring Harbor laboratory
http://genome.mcgill.ca/~pdlee/bioinformatics.html --Montreal genome centre
http://www.ornl.gov/hgmis/TRCAL.HTML -- *Human Genome Project*
http://www.ebi.ac.uk/ --- European Bioinformatics Institute