

Bioinformatics: a promising field for Medical Biochemists

Manojit Majumder

Dept of Biochemistry, Sylhet MAG Osmani Medical College, Sylhet

Medical informatics is the field concerned with the optimal use of information, often aided by technology, to improve individual health, healthcare, public health and biomedical research. Categories of *medical informatics* are i) *Medical or Clinical Informatics* - application of informatics to individuals ii) *Public Health Informatics* - application of informatics in public health iii) *Bioinformatics*- application of informatics in cellular and molecular biology, often with focus on genomics.

The term *bioinformatics* was first used by Pauline Hogeweg in 1978 to refer to 'the study of information process in biotic system'. Human DNA organizes into 46 chromosomes. Set of chromosomes that determine an organism is known as its genome - encoding all the functions required to go from a single cell to complete human being. Concept of genomics has developed over a long period of time. Structure of DNA was a key breakthrough in the history of genomics followed by Molecular cloning, DNA sequencing and RFLP mapping technique. In mid 80's, some important technical breakthrough happened that led to the ability to embark on Human Genome Project: i) Pulse-field gel electrophoresis to separate out large segment of genomic DNA, ii) PCR that enable chemical synthesis of DNA, iii) Global physical clone maps developed in model organism, iv) Yeast artificial chromosome were developed in 1987, and v) Automated DNA sequencer came in late 80's. Combination of all these different technologies led to concrete proposal to map and sequence the human genome by National Research Council in 1988. The draft sequence of human genome was published in 2001 and the complete sequence was published in 2004. Invention in 2005, of Next-Generation Sequencing technique with cheaper and high throughput

sequencing method, enabled expanding scope of genome analysis and DNA sequencing for all kind of biological experiments.

The field of bioinformatics involves the analysis and interpretation of various types of biological data - nucleotide and amino acid sequences, protein domain and protein structures. Database built up to store sequenced data in Genbank and UniPort. The PAM matrices- the default matrices for sequence alignment since 1978 was replaced by BLOSUM in 1992. In the 1990s, Smith-Waterman local alignment algorithm published to find the most similar pair of sub-sequences. BLAST algorithm developed to find out whether similar sequences exist in sequence database. Also, computer program developed for finding gene from the genome and algorithms were published to test the similarity among the genomes. Bioinformatics and Human Genome Project supported each other. The Human Genome Project provides for bioinformatics a lot of data produced by highthroughput technologies and bioinformatics is indispensable for developing the methods for analysis of such data. With the arrival of Next-Generation Sequencing technologies in 2005, the amount of data are increasing exponentially. Sequence Read Archive (SRA) was created to store the read sequences. Today, in addition to the human genome, there are also over 160,000 other species whose genomes or genes have been sequenced. Bioinformaticians are needed to understand and to make use of this information.

Bioinformatics is the field of science in which *biology, computer science, mathematics and information technology* merges in a single discipline. Sub-disciplines within bioinformatics

include: i) *computational biology* - process of analyzing and interpreting data, ii) *algorithms* and *statistics* - to assess relationships among members of large data sets, and iii) *Information technology* - to develop and implement tools that enable efficient access to and management of information.

Researchers are using bioinformatics to identify genes, establish their structure and functions and to develop gene-based strategies for preventing, diagnosing and treating diseases. Major research efforts in the field include sequence alignment, gene finding, genome assembly, protein structures alignment, protein structure prediction, drug design, drug discovery and prediction of gene expression, genome-wide association studies and the modeling of evolution. The primary goal of bioinformatics is to increase the understanding of biological processes. Human Genome Project costs \$3 billion and took twelve years. Today, one can get his/her genome sequenced at less than \$3,000 and within a month.

It is realistic to envisage that within next 10 years or so, everyone's genome could be sequenced, if they needed it, at the cost of \$100 per genome and in a much shorter period of time. Bioinformatics, at the moment is a hot, attractive and fascinating field because there are too few people adequately trained in both biology and computer science. Being a teacher of molecular biology, it is a promising field for the young Medical Biochemists. It is expected that within a short period of time, public and private Universities in Bangladesh will open *Bioinformatics* as a separate discipline and they will definitely search for competent faculty member, in this field. It is the prime time for Medical Biochemists to prepare themselves to avail the golden opportunity in near future.