Bioinformatics: Its Relevance, Today and Tomorrow

Dr. S. N. Chatterjee

Ex- Joint Director, Central Silk Board, Madiwala, Bangalore and now Visiting Faculty, Kalyani Government Engineering College

What is Bioinformatics?

It is well accepted that search for information on, and knowledge about, the subject Biology started long back during the march of human civilization. However, in the last twenty years, the immense developments in the field of computer technology and satellite communication along with the spread of these technologies to a wide range of population, especially the scientists, has made BIOINFORMATICS an independent field of R&D. In India, though the progress is slow, this subject is gaining more and more importance with every passing day and it cannot be denied that its relevance in scientific investigation of nature and its very many species would increase tremendously in the near future!!

What Bioinformatics stands for today?

Today, Bioinformatics means storing information on all aspects of biology, with special emphasis on structural and functional Genomics in universal language and making it available to public without disregarding the right of the scientific worker i.e., the intellectual property right (IPR). This obviously has provided immense scope of utilization of information generated and also the generation of new information. It also includes the methodology of analyzing structural and functional Genomics of the immense biodiversity around us. Let us examine a simple case: A student finds an insect and knows how to describe the morphological structure of it in detail. He can immediately utilize the internet-facility for knowing the true taxonomic status of it sitting in his own place without sending the material to London or elsewhere. This naturally provides him with the extra leverage to work in his chosen area of research.
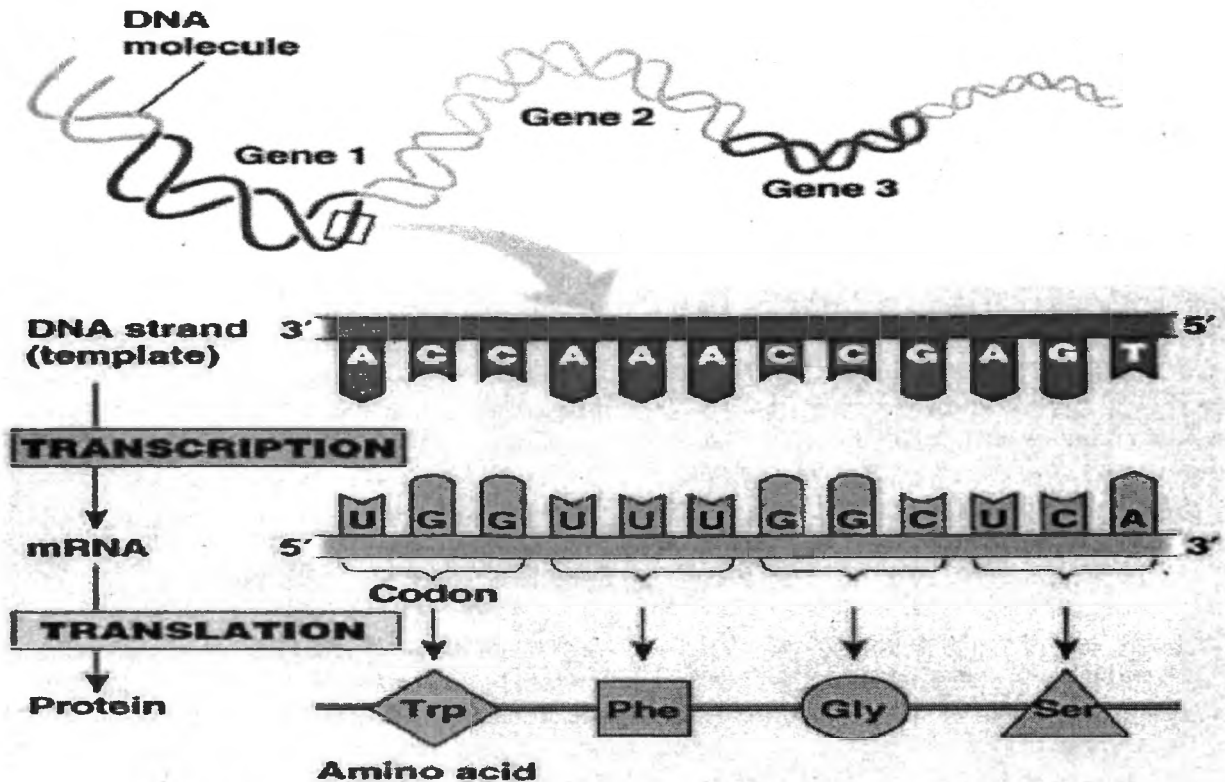
What is structural and functional Genomics ?

Genomics, as the name suggests, includes all aspects related with genome. But what is a genome? Genome is the physically and chemically well defined macromolecules, which follow certain universally applicable rules for all living matters and are inherited from parent(s) to offspring, again using certain biological processes which are universally true and can be defined in specific modalities applicable to the whole range starting from the simple organism as bacterial or a protozoa or human being or a dinosaur! However, before trying to understand "Genomics" , it would be better to approach a simpler one ,i.e., the "GENE".

What is a GENE?

We all know that it is Gregor Mendel (1822-1884) who first tried to understand the inheritance of characters in higher organism. However, it took the mankind more than 30 years to appreciate his mathematical derivation of inheritance. But even then, ideas about the physical or chemical characters of the materials that are inherited were quite imaginative to many scientists including Darwin! In 1883 Wilhelm Roux, postulated that chromosomes within the nucleus were the bearer of the hereditary factors. The experimental work of T. Boveri and WS Sutton in 1902 confirmed the contention of Roux. But it took another 7 years to introduce the term "GENE" by a Danish scientist, WL Johannsen, who was working with plants. Till then, the physical identity of gene was not known for which the world had to wait for a few more decades.

Chromosomes are made of two types of large organic molecules, proteins and nucleic acids. For quite sometime scientists were not in agreement as to which of these two is actually the GENETIC MATERIAL. During 1925-1945, some elegant experiments proved beyond doubt that it is the nucleic acids that carry the hereditary traits. Considering the components of these nucleic acids, it was found that the only variability is the four types of bases: (adenine, quanine, cytosine and thymidine (present as uracil in RNA) while two other components, phosphate and sugar molecules are the same in all the genetic materials of the living world. But how do these four bases decide the hereditary traits?

(Biology.com/AP Biology Book)

For each gene, one DNA strand functions as a template for transcription-the synthesis of a complementary mRNA molecule. The base-pairing rules for DNA synthesis also guide transcription, but uracil (U) takes the place of thymine (T) in RNA. During translation, the mRNA is read as a sequence of base triplets, called codons. Each codon specifies an amino acid to be added to the growing polypeptide chain. The mRNA is read in the 5' to 3' direction.

It was found that mother NATURE adopted a triplet coding system

GENE and Bioinformatics:

Coming back to the main frame, let us examine the connection between Bioinformatics and gene. A gene may code for a particular protein, which in turn may act as an enzyme, essential for a particular biochemical reaction. A gene can also produce a RNA molecule which can play the role of a regulator function! Further, it is to be kept in mind that certain biological functions are essentially the same in all the living organisms and so the enzyme also needs to be the same. Thus there is the issue of universality on one hand and specificity on the other. But how it is linked with BioInformatics? There are two main aspects :1/ providing bio (gene-tic) information ; 2/generating new information and enriching the existing bioinformatic database. Suppose one scientist wants to search for a gene-X in an organism-Y ! Let us see how BioInformatics helps him. Firstly, he can search for the information whether any one has worked earlier on the same through internet and his request will be processed simultaneously at USA, UK, Japan databases and he gets the confirmation either way. Let us presume, he failed to find any information for gene-X in organism-Y and then he can search information available on the gene-X in other organism close to his and using the available sequence, he can design his search tool (known as primer!) and using particular wet-lab technology the gene-X in organism-Y can be hunted out. Further, using high-precision semi-automated instrumentation he can find out the base sequence. He can thus go back to BioInformatics database to confirm the presence

of gene-X in his DNA sequence and also utilize BioInformatics to compare his gene-sequence and its expected (or predicted) products at the evolutionary level with that in other organism already worked out. Moreover, once he generates the information it becomes his solemn responsibility to deposit the information with international GENBANK, so that his endeavor is registered at the international level before publication of his result. He also can decide the date when the information would be released to others as a component of public domain database.

Further, let us presume that an evolutionary biologist likes to examine the evolutionary changes in the enzyme-Xi (RUBISCO enzyme of plant system) in the plant kingdom. He can screen the information available in BioInformatics database for various groups of plant kingdom and retrieve the sequence data of DNA or amino acids and utilize various mathematical or statistical programs to correlate the changes with the evolutionary time scale. Further, if it is possible for him to retrieve information on the habitat of the plant-taxa selected by him, he also can try to correlate the polymorphism in the DNA/amino acid sequences with specific habitat characteristics through regression statistics.

Gene and Genomics:

Genomics is the story of the total genetic make up of a particular species in the bio-world. It is known that only a small part of the total genome produces specific proteins. Does that mean that the remaining portion is JUNK? It was thought so once. But, today, we cannot ignore that part as JUNK. It is now well accepted that besides the sequences chosen to be transcribed and translated, there are other sequences, which also play a very important role. For example, there are specific DNA sequences where receptor or signal-protein gets attached. Another important component is the mobile genetic element commonly known as jumping genes. As evident from the name itself, such sequences can independently come out from a particular region, can amplify of its own and also can get inserted into another segment and either deactivate another sequence into functional state or even can inactivate a functional gene! It may be mentioned here that ONCOGENE, one of the causative agent of cancer can also be included in the group of 'jumping gene'!

Bioinformatics as a Master Course:

As indicated in the beginning, Bioinformatics is now offered as a major course subject in Graduation and Post-graduation around the world. The courses offered in various universities abroad include two major aspects: one is cell-biological aspect and the other is computational aspect. The computational aspect deals with various algorithms developed on the basis of mathematical derivation deduced from biological observations. However, it would be better to replace the term biological with molecular! Further, the computational aspect is increasingly moving towards developing model circuits of biological/physiological/molecular events taking place in most cases in a cascade fashion. The cell biological aspects include basic understanding of cell-biology, diversity of life forms, molecular basis of heredity, expression of heritable characters, various DNA technologies used in wet-lab, high-precision semi automated instrumentation and biological basis of mathematical derivations used to develop Bioinformatics tools.

What is the future of Bioinformatics ?

As we proceed further and further the load of genetic information is gradually increasing at a tremendous rate, especially through public domain. More and more biological species are being taken up for full genomic analysis. It is to be kept in mind that globalization is not being restricted at the economic front only. The under-developed countries as also the developing countries will be more and more exposed to exploitation by the developed world. Under such circumstances, two aspects will be relevant. Firstly, moral and technical responsibility of scientists of countries like India and others shall increase tremendously with regard to understanding and characterizing bio-diversities of their own countries and safe-guarding the information which can be misutilised by others. Secondly, as the outsourcing is increasing rapidly in IT and other related sectors, emergence of specialists in the field of BIOINFORMATICS would result in scopes for more profitable employment generation. However, all these require strong support from the State. As a country, India has awakened to the needs and demands of the scientists in this so very emerging a field. It is hoped that with necessary support from the government, research and studies in the field of Bioinformatics will quickly gather momentum in our country.