

# Bioinformatics

Shakila Shams<sup>1</sup>, M. Abdus Sobhan<sup>2</sup>, Farruk Ahmed, Ph.D.<sup>3</sup>, Ali Shihab Sabbir, PhD<sup>4</sup>

<sup>1</sup>The Independent University Bangladesh, Computer Network & Communication (M.Sc.Eng.), Computer Science & Engineering (CSE)  
Plot 16, Block B, Aftabuddin Ahmed Road, Bashundhara R/A, Dhaka-1229, Bangladesh  
shamsshakila2[at]gmail.com

<sup>2</sup>Vice Chancellor of Prime University, 114, 116 Mazar Rd, Dhaka 1216, Bangladesh  
asobhan[at]secs.iub.edu.bd  
masobhan.primeuniversity[at]gmail.com

<sup>3</sup>Department of Computer Science & Engineering, Ph.D. in Electrical and Electronic Engineering, 1979,  
University of Salford, Manchester, England  
farruk[at]iub.edu.bd

<sup>4</sup>Department of Computer Science & Engineering, The Independent University Bangladesh  
Plot 16, Block B, Aftabuddin Ahmed Road, Bashundhara R/A, Dhaka-1229, Bangladesh  
asabbir[at]iub.edu.bd

**Abstract:** *The characteristics flood of data method from genome sequences and functional genomics has given rise to new field sector, bioinformatics related work, which combines elements of biology and computer science related background. Here we propose a proper definition for this new field scenario and review and follow some of the research related work that is being pursued and also issued, particularly in relation to transcriptional regulatory systems for this purpose. Our definition may be describe usually Bioinformatics is conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying "informatics" or "information" techniques (derived from disciplines such as applied mathematics, computer science related background, and statistics related computation) to understand and organize the information associated with these molecules and method, on a large-scale criteria. Another objective of bioinformatics is the enhancement of experimental data related by anticipation. A primary purpose of computational biology is the anticipation of protein structural scenario which is observed from an amino acid sequence. The spontaneous folding of proteins indicates that this must be possible matter. Progress in the improvement of methods and its application to anticipate protein folding is experiment by biennial Critical Assessment of Structure Prediction (CASP) programs, which involve blind tests of structure prediction methods and its application also.*

**Keywords:** Sequence assembly, Genome annotation, Computational evolutionary biology, Comparative genomics

## 1. Introduction

Analyses on bioinformatics predominantly focus considering three types of large datasets available in molecular biology sector area macromolecular structures related works and structures, genome sequences step by step, and the results of functional genomics experiments or examination (e.g. expression data). Extra information indicates the text of scientific papers and also its application and "relationship data" from metabolic pathways or method, taxonomy trees or construction, and protein-protein relational networks and its application are observed. Bioinformatics employees a large range of computational techniques and applications including sequential and structural alignment, database design and data mining, macromolecular geometrical phylogenetic tree construction, prediction of protein structure related background and function, gene finding, and expression data clustering is shown. The emphasis is on approaches integrating a variety of computational methods and its application and heterogeneous data sources. Finally, bioinformatics is a practical discipline and rules. We survey some of these representative applications and techniques and also methods related works, such as finding homologues, designing drugs, and performing large-scale censuses which is discussed. Additional information and applications pertinent to the review is available. To study how normal cellular activities

and its criteria are altered in different disease states and also places, the biological data must be combined to form a combined picture of these activities and its applications and methods. So, the field of bioinformatics has evolved such that the most pressing task now involves many analysis and its influences and interpretation of various types of data are also observed. Nucleotide and many elements such as amino acid sequences, protein domains, and protein structures. The Analyzing system and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include: Development and implementation of computer program enable efficient access to, use and management of, various types of information and its applications.

The Development system of new algorithms such as (mathematical formulas and methods) and statistical measurements that assess relationships among members of large data sets. Some methods to locate a gene within a sequence, to predict protein structure and/or function, and to cluster protein sequences into families of related sequences. The objective of bioinformatics is to increase the understanding of biological processes scenario area related work. It focuses biological processment system related works on developing and applying **computationally intensive techniques** to achieve this goal area level. Examples as follows: pattern recognition, data mining,

Volume 11 Issue 7, July 2022

[www.ijsr.net](http://www.ijsr.net)

Licensed Under Creative Commons Attribution CC BY

machine learning algorithms, and visualization. Main research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and some are observes protein, genome-wide association studies, the modeling of evolution and cell division/mitosis.

Bioinformatics now entails **the creation of mathematics and advancement of databases related work, algorithms, computational and statistical techniques** and its application, and theory to solve formal and practical problems arising from the management and analysis of biological data scenario.

Over the past few decades, the rapid developments and technologies in genomic and other molecular research related technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology sector for rapid development. **Bioinformatics is a type of mathematical and computing approaches computer science background which is used to glean understanding of biological processes.** Common activities and its criteria in bioinformatics include mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them, and creating and viewing 3-D models of protein.

## 2. Sequence Assembly

DNA sequencing techniques and its application produce little fragments of sequence that need to be assembled to obtain complete gene or genome sequences which is observed sequentially. The Shotgun sequencing technique (which was used, for example, by The Institute for Genomic Research (TIGR) to sequence the first bacterial genome, *Haemophilus influenzae*) generates the sequences of many thousands of little DNA fragments (ranging from 35 to 900 nucleotides long, depending on the sequencing technology). These fragments overlapping and, when aligned properly by a genome assembly program, can be used to reconstruct the complete genome. The Shotgun sequencing field sequence data quickly, but the task of assembling the fragments can be quite complicated for larger genomes. Large as the human genome, it may take many times of CPU time on large-memory, multiprocessor computers to assemble the fragments, and the resulting assembly usually contains large-memory, multiprocessor computers numerous gaps that must be filled in later.

## 3. Genome Annotation

The context of genomics, annotation is the processment of marking the genes and other biological features in a DNA sequences scenario. These processment needs to automated because most genomes are too large to annotate by hand, not to mention the desire to annotate as many genomes as possible, since the Sequencing has ceased to pose a bottleneck. Annotation is made possible by the fact of these genes have recognizable start and stop regions, although the

exact sequence found in these regions can vary between genes. A combined genome annotation system was published in 1995 by the team at The Institute for Genomic Research that performed the first complete sequencing and analytical issues of the genome of a free-lifestyle organism, the bacterium *Haemophilus influenzae*. Built a software system to identify the genes encode all proteins, transfer RNAs, ribosomal RNAs (and other sites) and to make initial functional assignments. Genome annotation systems related work similarly, but the programs available for analysis of genomic DNA, such as the Gene Mark program trained and used to find protein-coding genes in *Haemophilus influenzae*, are constantly changing and improving. The Human Genome Project left to achieve after its closure, a new project developed by the National Human Genome Research Institute in the U.S appeared. The ENCODE related project type work is a collaborative data collection of the functional elements of the human genome that uses next-generation DNA-sequencing related technologies, method and genomic tiling arrays, technologies able to automatically generate large amounts of data sequencing event occurs at a dramatically reduced per-base cost but with the same accuracy (base call error) and fidelity (assembly error).

## 4. Computational Evolutionary Biology

Evolutionary biology is the study field of the origin or primitive and descent of species, as well as their changes observed over time to time. Informatics has assisted evolutionary biologists by enabling researchers to the below characteristics:

The evolution of a large number of organisms by measuring changes in their DNA, rather than through physical taxonomy or physiological observations alone which is observed time to time, Recently compares of entire genomes, which permits the study of more complex evolutionary events, such as gene duplication, horizontal gene transfer, and the prediction of factors important in bacterial speciation, Create complex computational population genetics models to predict the outcome of the system over time to time

## 5. Comparative Genomics

The scenario of Comparative genome analysis is the establishment of the correspondence between genes (orthology analysis) or other genomic features in different organisms. These kind of intergenomic maps that make it possible to trace the evolutionary processes responsible for the divergence of two genomes. The evolutionary events acting at various organizational levels shape genome evolution. The lowest level, point mutations affect individual nucleotides. At a high range, large chromosomal segments undergo duplication, lateral transfer, inversion, transposition, deletion and insertion. Genomes are related deeply in processes of hybridization, polyploidization and endosymbiosis, often leading to rapid speciation. **The critical situation of genome evolution poses many exciting challenges to developers of mathematical models and algorithms, who have recourse to a spectrum of**

algorithmic, statistical and mathematical techniques, which computer science & engineering related works, leveling from exact, heuristics, fixed parameter and approximation algorithms for problems based on parsimony models to Markov chain Monte Carlo algorithms for Bayesian analysis of problems based on probabilistic models. Some of these studies are based on the detection of sequence homology to assign sequences to protein families which is discussed in these era.

## 6. Pan genomics

Pan genomics is a concept introduced by Tettelin and Medini which is eventually took root in bioinformatics related background. Pan genome is the complete gene repertoire of a particular taxonomic group related background: Initially applied to closely related strains of a species, it can be applied to a larger context like genus, phylum etc. The main two parts- The Core genome: Genes common to all the genomes under study (These are often housekeeping genes vital for survival) and The Dispensable/Flexible Genome: Genes not visible in all but one or some genomes under study. The Bioinformatics tool BPGA format can be used to characterize the Pan Genome of bacterial species

## 7. Conclusion

The objective of this discussion was to bring together bioinformatics related stake holders from government, academic and industry for a day of presentational scenario and dialogue. Fifteen experts identified and discussed some of the most important issues raised by the current flood of biological data scenario. May be it can be solved by following computer science & engineering related concept and of course method for establishment of bioinformatics processes and implementation in important sector. Subjects explored included the importance of database related work integration and interoperability, consistency and standards in terminological construction, mistakable prevention and correction is also observed, data provenance, ontology, the importance of maintaining privacy and security, data mining, and the need for more computer scientists with specialty training in bioinformatics field. Although formal conclusions and recommendations will not come from this particular workshop, many insights may be gleaned about the future of this background, from the context of the discussions, application and presentations described sequentially above.

## References

- [1] **Jump up** ^Lesk, A. M. (26 July 2013). "Bioinformatics". *Encyclopaedia Britannica*. Retrieved 17 April 2017.
- [2] ^ **Jump up to:** <sup>ab</sup>Sim, A. Y. L.; Minary, P.; Levitt, M. (2012). "Modeling nucleic acids". *Current Opinion in Structural Biology*. **22** (3): 273–278. doi:10.1016/j.sbi.2012.03.012. PMC 4028509. PMID 22538125.
- [3] **Jump up** ^Dawson, W. K.; Maciejczyk, M.; Jankowska, E. J.; Bujnicki, J. M. (2016). "Coarse-grained modeling of RNA 3D structure"(PDF). *Methods*. **103**: 138–156. doi:10.1016/j.ymeth.2016.04.026. PMID 27125734
- [4] Hixson, R., & Hunt-Unruh, D. Demystifying HIPAA. *Annals of the American Psychotherapy Assn*, 11(3), 10-14.
- [5] **Jump up** ^Kmieciak, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. (2016). "Coarse-Grained Protein Models and Their Applications". *Chemical Reviews*. **116** (14): 7898–936. doi:10.1021/acs.chemrev.6b00163. PMID 27333362.
- [6] **Jump up** ^Wong, K. C. (2016). *Computational Biology and Bioinformatics: Gene Regulation*. CRC Press/Taylor & Francis Group. ISBN 9781498724975.
- [7] **Jump up** ^Joyce, A. P.; Zhang, C.; Bradley, P.; Havranek, J. J. (2015). "Structure-based modeling of protein: DNA specificity". *Briefings in Functional Genomics*. **14** (1): 39–49. doi:10.1093/bfpg/elu044. PMC 4366589. PMID 25414269
- [8] **Jump up** ^Spiga, E.; Degiacomi, M. T.; Dal Peraro, M. (2014). In Karabancheva-Christova, T. *BiomolecularModelling and Simulations*. Advances in Protein Chemistry and Structural Biology. **96**. Academic Press. pp. 77–111. doi:10.1016/bs.apcsb.2014.06.008. ISBN 9780128000137.