



The Role on Bioinformatics in DNA Sequencing and Assembling

Dara Zain*

Biology and Genetics of Reproduction Unit, Multidisciplinary Unit for Biomedical Research, Portugal

*Corresponding Author's E-mail: zaindara@rediff.com

Received: 01-Apr-2022, Manuscript No. IRJBB-22-84061; **Editor assigned:** 04-Apr-2022, PreQC No. IRJBB-22-84061 (PQ); **Reviewed:** 18-Apr-2022, QC No. IRJBB-22-84061; **Revised:** 22-Apr-2022, Manuscript No. IRJBB-22-84061 (R); **Published:** 29-Apr-2022, DOI: 10.14303/2250-9941.2022.10

Abstract

Bioinformatics is an interdisciplinary field that develops Methods and software tools for understanding biological data, particularly when the data sets are large and complex, are developed in the interdisciplinary field of bioinformatics. Bioinformatics is an interdisciplinary field of science that analyzes and interprets biological data by combining biology, chemistry, physics, computer science, information engineering, mathematics, and statistics. Computational and statistical methods have been utilized in in silico analyses of biological queries by means of bioinformatics (Hamilton et al., 2002).

Keywords: Bioinformatics, DNA Sequencing, Genomics

INTRODUCTION

The term "bioinformatics" refers to both specific analysis "pipelines" that are frequently utilized, particularly in the field of genomics, and biological studies that employ computer programming as a component of their methodology. The identification of potential genes and single nucleotide polymorphisms (SNPs) are two common applications of bioinformatics. Such an identification is frequently made with the intention of gaining a deeper comprehension of the genetic basis of disease, distinctive adaptations, and desirable characteristics (e.g. in species used for agriculture) or differences between populations. A less formal subfield of bioinformatics is proteomics, which aims to comprehend the underlying organizational principles of protein and nucleic acid sequences (Craveiro et al., 2018).

The processing of images and signals makes it possible to get useful results from a lot of raw data. It aids in the sequencing and annotation of genomes and their observed mutations in the field of genetics. It assumes a part in the text mining of natural writing and the improvement of natural and quality ontologies to sort out and question organic information. The study of gene and protein expression and regulation also relies on it. Tools developed in bioinformatics aid in the

comparison, analysis, and interpretation of genomic and genetic data, as well as, more generally, in comprehending evolutionary aspects of molecular biology. It aids in the analysis and cataloguing of the biological pathways and networks that are a crucial component of systems biology at a more integrative level. It aids in the simulation and modeling of DNA, RNA, proteins, and biomolecular interactions in structural biology (Connor et al., 2007).

Since the completion of the Human Genome Project, both speed and cost have decreased significantly. Some laboratories are now able to sequence more than 100,000 billion bases annually, and a full genome can be sequenced for less than \$1,000. Following Frederick Sanger's discovery of insulin's sequence in the early 1950s, the availability of protein sequences necessitated the use of computers in molecular biology. It turned out to be impossible to manually compare multiple sequences. Margaret Oakley Dayhoff was an early pioneer in the field. She pioneered methods for sequence alignment and molecular evolution and created one of the first protein sequence databases, which were initially published as books. Elvin A. Kabat, who pioneered biological sequence analysis in 1970 with his comprehensive volumes of antibody sequences published with Tai Te Wu between 1980 and 1991, was another pioneer in bioinformatics (Ibba 2002). In the 1970s, new methods

for sequencing DNA were applied to bacteriophage MS2 and X174, and informational and statistical algorithms were used to parse the extended nucleotide sequences. These studies demonstrated that straightforward statistical analyses can reveal well-known features like the triplet code and coding segments, demonstrating the value of bioinformatics (Andreini et al., 2012).

DISCUSSION

The improvement of our comprehension of biological processes is the primary objective of bioinformatics. What separates it from different methodologies, be that as it may, is its emphasis on creating and applying computationally serious procedures to accomplish this objective. Some examples are: machine learning algorithms, pattern recognition, data mining, and visualization. Sequence alignment, gene discovery, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, genome-wide association studies, and the modeling of evolution and cell division/mitosis are among the major areas of research in this field (Feig et al., 2002).

For the purpose of resolving formal and practical issues that arise from the management and analysis of biological data, bioinformatics now entails the creation and development of databases, algorithms, computational and statistical methods, and theoretical frameworks.

Information technology advancements and rapid advancements in genomic and other molecular research technologies have produced a tremendous amount of molecular biology-related data over the past few decades. The mathematical and computing methods used to gain an understanding of biological processes are referred to as bioinformatics. Mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them, and creating and viewing 3-D models of protein structures are all common bioinformatics tasks.

While it is frequently regarded as synonymous with computational biology, bioinformatics is a scientific discipline that shares some similarities with but is distinct from biological computation. Bioinformatics uses computation to gain a deeper comprehension of biology, whereas biological computation builds biological computers by combining biology and bioengineering. The analysis of biological data, particularly DNA, RNA, and protein sequences, is a component of both bioinformatics and computational biology. Beginning in the middle of the 1990s, the Human Genome Project and rapid advancements in DNA sequencing technology fueled the explosive growth of bioinformatics. Writing and running software programs that make use of algorithms from graph theory, artificial intelligence, soft computing, data mining, image processing, and computer simulation are all necessary for producing meaningful information from biological data. Theoretical foundations like discrete mathematics, control

theory, system theory, information theory, and statistics are also used by the algorithms.

In order to obtain complete genome or gene sequences, the majority of DNA sequencing methods produce brief sequence fragments that must be assembled. The sequences of many thousands of small DNA fragments (ranging from 35 to 900 nucleotides long, depending on the sequencing technology) are produced by the so-called shotgun sequencing technique, which was utilized, for instance, by The Institute for Genomic Research (TIGR) to sequence the first bacterial genome, *Haemophilus influenzae*. If a genome assembly program correctly aligns the ends of these fragments, they can be used to reconstruct the entire genome. Although shotgun sequencing provides quick sequence data, the process of putting the fragments together for larger genomes can be quite challenging. On large-memory, multiprocessor computers, it may take many days of CPU time to assemble the fragments for a genome as large as the human genome. The assembled genome typically contains numerous gaps that must be filled in later. Almost all genomes that have been sequenced so far [when?] have used shotgun sequencing as their method of choice. Additionally, an essential area of bioinformatics research is genome assembly algorithms.

Other properties of sequences can be used to predict the function of genes, despite the fact that genome annotation is primarily based on sequence similarity (and, consequently, homology). In point of fact, the majority of methods for predicting gene function concentrate on protein sequences because these are more feature-rich and informative. For instance, transmembrane segments in proteins are predicted by the distribution of hydrophobic amino acids. However, external data like protein structure, protein-protein interactions, or data on gene (or protein) expression can also be used in protein function prediction.

CONCLUSION

Establishing the correspondence between genes (orthology analysis) or other genomic features in various organisms is the core of comparative genome analysis. The evolutionary processes that led to the divergence of two genomes can be traced using these intergenomic maps. A large number of developmental occasions acting at different hierarchical levels shape genome advancement. Point mutations affect individual nucleotides at the simplest level. Large chromosomal segments go through processes like duplication, lateral transfer, inversion, transposition, deletion, and insertion at a higher level.[28] In the end, whole genomes go through processes like hybridization, polyploidization, and endosymbiosis, which frequently result in rapid speciation. Developers of mathematical models and algorithms have recourse to a variety of algorithmic, statistical, and mathematical techniques to deal with the complexity of genome evolution. These techniques range from exact, heuristic, 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.

REFERENCES

1. Hamilton JA, Kamp F, Guo W (2002). Mechanism of cellular uptake of long-chain fatty acids: Do we need cellular proteins? *Mol cell biochem.* 239: 17-23.
2. Craveiro Sarmiento AS, de Azevedo Medeiros LB, Agnez-Lima LF, Lima JG, de Melo Campos JTA (2018). Exploring seipin: from biochemistry to bioinformatics predictions. *Int J Cell Biol.*
3. Connor RF, Roper RL (2007). Unique SARS-CoV protein nsp1: bioinformatics, biochemistry and potential effects on virulence. *Trends microbial.* 15: 51-53.
4. Ibba M (2002). Biochemistry and bioinformatics: when worlds collide. *Trends Biochem Sci.* 27: 64.
5. Andreini C, Bertini I (2012). A bioinformatics view of zinc enzymes. *J inorg biochem.* 111: 150-156.
6. Feig AL, Jabri E (2002). Incorporation of bioinformatics exercises into the undergraduate biochemistry curriculum. *Biochem Molecul Biol Edu.* 30: 224-231.