Chapter 1
A Concise Overview of Bioinformatics

Alessandra Lima da Silva

https://orcid.org/0000-0002-2698-6889

Universidade Federal de Minas Gerais, Brazil

Diego Mariano

https://orcid.org/0000-0002-5899-2052

Universidade Federal de Minas Gerais, Brazil

Mariana Parise

Instituto de Ciências Exatas, Brazil

Angie L. A. Puelles

Instituto de Ciências Exatas, Brazil

Tatiane Senna Bialves

https://orcid.org/0000-0002-3827-2859

Instituto de Ciências Exatas, Brazil

Luana Luiza Bastos

https://orcid.org/0000-0002-6932-0438

Instituto de Ciências Exatas, Brazil

Lucas Santos

Instituto de Ciências Exatas, Brazil

Rafael Pereira Lemos

Instituto de Ciências Exatas, Brazil

ABSTRACT

Bioinformatics is a modern field of research that encompasses the use of computational techniques for analyzing biological data. Bioinformatics is directly connected to the central dogma of molecular biology, such as the analysis of DNA, RNA, and proteins. The first studies date back to the 1960s when standards were defined for the representation of proteins in files. However, it gained popularity in the early 2000s with the publication of the draft human genome. Nowadays, bioinformatics has stood out as an interdisciplinary science that uses growing computational power to help obtain new scientific discoveries. In this chapter, you will learn a little more about the fundamentals of bioinformatics.

1. WHAT IS BIOINFORMATICS

Bioinformatics is a discipline that transcends the boundaries of traditional academic disciplines, integrating and mobilizing knowledge from various areas associated with life science and computational technologies, thus demonstrating its transdisciplinarity. Moreover, bioinformatics is an interdisciplinary field mediated by biology, computer science, mathematics, and statistics. The main objectives of bioinformatics include storage and organization of biological metadata, analysis of genetic and protein sequences, genomic comparison and evolution (complete or incomplete in the form of drafts), gene expression and proteomics (microarrays, RNA sequencing, or molecule interactions), and also biological systems modeling (development of models and algorithms, virtual screening, or artificial intelligence).

2. A BRIEF HISTORY OF BIOINFORMATICS

As the processing power of computers increased after the Second World War, their use began to be considered for solving biological problems. In the 1960s, several experimental techniques were used to obtain collections of amino acid sequences. At that time, scientists were using computational methods without the need for supercomputers or computer networks to understand protein sequences. This led to a set of interesting problems that were impossible to solve without the processing power of computers (Hagen, 2000). This is partly responsible for giving rise to a fundamental area of research in biology and how information is processed in living systems, such as evolutionary biology and how genes are transmitted (Gauthier et al., 2019).

Margaret Dayhoff, Richard Eck, and Robert Ledley were pivotal figures in bioinformatics' early development, with Dayhoff's contributions being particularly notable. She is often credited as a pioneer in developing the first amino-acid substitution matrix and other foundational bioinformatics tools. Margaret Dayhoff became known as the mother of Bioinformatics. In the mid-1960s, she published the "Atlas of Protein Sequence and Structure", a collection of protein sequences known at the time, which received several updates in the following years (Dayhoff, 1972). When representing protein sequences and other polypeptide molecules, each amino acid residue was represented by a letter. For example, alanine is represented by the letter "A," and tryptophan is represented by the letter "W". The representation of a single letter facilitates comparison between different sequences (alignment). Furthermore, based on analyzing several amino acid sequences, Dayhoff and colleagues proposed the PAM (Point Accepted Mutation) substitution matrix, which estimated the possibility of mutations accepted by natural selection (Sasidharan;

Related Content

A Novel Flowsheet for the Recycling of Valuable Constituents from Waste Printed Circuit Boards
Jingfeng He, Yaqun He, Nianxin Zhou, Chenlong Duan, Shuai Wangand Hongjian Zhang (2011). *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences (pp. 296-306).*
www.irma-international.org/chapter/novel-flowsheet-recycling-valuable-constituents/48386

Genome Sequence Analysis in Distributed Computing using Spark
 Sagar Ap., Pooja Mehta, Anuradha J.and B.K. Tripathy (2015). *International Journal of Knowledge Discovery in Bioinformatics (pp. 30-42).*
www.irma-international.org/article/genome-sequence-analysis-in-distributed-computing-using-spark/167708

Green DRCT: Measuring Energy Consumption of an Enhanced Branch Coverage and Modified Condition/Decision Coverage Technique
Sangharatna Godboley, Arpita Duttaand Durga Prasad Mohapatra (2017). *International Journal of Knowledge Discovery in Bioinformatics (pp. 14-29).*
www.irma-international.org/article/green-drct/178604

Drug Optimization for Cystic Fibrosis Patients Based on Disease Pathways Crosstalk
Shuting Linand Yifei Wang (2021). *International Journal of Applied Research in Bioinformatics (pp. 1-11).*
www.irma-international.org/article/drug-optimization-for-cystic-fibrosis-patients-based-on-disease-pathways-crosstalk/267820

Quantum Phase-Hebbian Image Processing
Mitja Perušand Chu Kiong Loo (2011). *Biological and Quantum Computing for Human Vision: Holonomic Models and Applications (pp. 222-228).*
www.irma-international.org/chapter/quantum-phase-hebbian-image-processing/50509