FROM MENDEL PEAS TO BIOINFORMATICS: AND THE ETHICAL PRINCIPLES?

THIAGO FERREIRA DE TOLEDO

Master in Electrical and Computer Engineering
E-mail: thiagoftoledo@live.com

Abstract – The fusion of biological sciences with computer science has given rise to bioinformatics. As it is an area that deals directly with life, where computers perform a central role, ethical issues become even more complex. In this sense, this work begins by tracing a parallel between the evolution of genetics with that of computers and, then, an essay is proposed, inviting reflection on how responsible actions by professionals in the field can help them conduct their work on an ethical basis.

Keywords – Applied Ethics, Bioinformatics, Computer Ethics, Ethics, ICT.

I. INTRODUCTION

Computing has been used as a means to achieve progress in various areas of knowledge, assisting, for example, in complex calculations, storage, search and data processing, among other infinities of possibilities that may be required by human needs to resolve any issue.

On the other hand, the introduction of computers brought to light questions, until then, unpublished, as well as the correct way to use them, without abuse of any kind. So, to solve situations concerning the good use that should be made of computers, ethical factors applied to this area need to be considered.

For example, bioethics emerged, which deals with ethical issues related to biomedical sciences and genetic engineering, as the manipulation of genes [1]. And, computer ethics, which provides conceptual models to guide computer professionals to perform their function with good conduct, considering what human beings value and what they do not [2].

Bioethics and computer ethics are two branches of what can be called applied ethics, whose objective is to provide the public with ethical judgments on relevant contemporary issues, providing vocabulary, theories and argumentative methods to deal with problems. Therefore, the ethical issues related to bioinformatics refers to the use of computers to manipulate genetic data [3].

The genetic data are obtained, for example, with the genomic sequence, which is an encrypted code that contains the raw information for the formation of living organisms and how they function. The study of the content of this genomic information is called bioinformatics [4]. Bioinformatics, therefore, is a hybrid science, because it encompasses both biology and computing, where computational systems are used to manage and analyze biological data. The data are analyzed using statistical techniques and artificial intelligence [5]. In this way, bioinformatics encompasses the use of hardware, software and also mathematics [6].

In this sense, this work aims to shed light on the ethical issue applied to bioinformatics, and, for that, initially historical points are investigated where discoveries in the biological area and inventions in the computational field occurred.

As bioinformatics is a science where there is the junction of biology and computing, the evolution of both areas was traced in parallel. And then, having verified the close connection between these sciences, a way of using bioinformatics responsibly is analyzed and suggested.

This article is organized as follows: in the next section, some points related to the evolution of genetics in parallel with the evolution of computers are addressed; in addition, concepts that can help the reader to better understand are also explained. Section III contains an analysis of how responsibility can be a vital factor to be considered in making good use of bioinformatics. In Section IV a brief analysis is reported, containing key points of discoveries and inventions achieved in the areas of genetics and computing, and, finally, the last considerations are made.

II. EVOLUTION OF GENETICS IN PARALLEL WITH COMPUTING

In the monastery, located in the city of Brüm, Austria (currently Brno, in the Czech Republic), the monk, Gregor Mendel, seeking to understand how the transmission of the characteristics of the genitors to their descendants occurs, he grew peas in his garden between 1856 and 1863. During this time, Mendel conducted experiments of cross-pollination between different varieties of peas; one of the experiments involved crossing purple and white flowers, and, all the first hybrid generation resulted in purple flowers. Of these flowers, Mendel made the self-pollination, giving rise to the second generation, in which, of the 929 flowers of the progeny, 705 were born purple and 224 white, therefore, analyzing the results, observed that genes perform a fundamental function in the transmission of characteristics. Mendel also proposed

the occurrence of two genetic variants, called alleles; one allele conditioning to purple and the other to white. In the experiment, there was a predominance of the purple allele, in which a gene encodes an enzyme producing the purple pigment in the cells. In his investigations, Mendel noted the law of inheritance [4].

Over the years of his studies, Mendel understood the importance of statistical research and the application of mathematics to a biological problem, serving as the basis for your most important work, entitled "Experiments in Plant Hybridization", which was read at the meetings of Natural History Society of Brünn, in 1865. And, the following year, it was printed in the Proceedings of the Natural History Society [7].

Several other discoveries were made in the area of genetics, for example, that proteins are large mass organic molecules formed by one or more chains of amino acids, that is, they are the final products of the gene expression, whose molecules perform the function of assigning properties to living systems. Therefore, the diversity of proteins and biological functions that gives the enormous diversity of life; originate when proteins are constituted with the combinations of 20 different amino acids [6].

Before the appearance of the first computers, in 1937, the British Turing [8], projected a machine capable of computing mathematical functions. Then, the following year, the German Konrad Zuse, built the first programmable binary computer, called Z1. And, in 1945, the American Vannevar Bush, in a text titled "As We May Think", influenced, to some extent, what would later become the World Wide Web, because he proposed a device, called MEMEX (Memory Index), to store books, records and communications, which could be consulted quickly [9]. Bush's text speculated about how science and technology could change the way of thinking and organizing knowledge [10].

At the end of the 1930's and during the following decade, the first computers appeared, for example, in chronological order: Bombe, ABC (Atanasoff-Berry Computer), Complex Number Computer, Claude Shannon, RDA2 (Rockfeller Differential Analyzer II), Colossus Mark I, IBM ASCC (Automatic Sequence Controlled Calculator), ENIAC (Electronic Numerical Integrator and Computer) [9].

Between 1941 and 1944 experiments were conducted to identify the role performed by genes within cells that enabled the control of different states and characteristics, for example, like the color of a flower. So, it has been demonstrated that the genes encode the enzymes, that is, they perform a metabolic function within the cells, since enzymes perform biochemical functions in the cell. The evidence that genes are made in DNA (deoxyribonucleic acid), was demonstrated by extracting DNA from a virulent strain of bacteria, whose genetic information could transform a non-virulent strain into a virulent strain. But the question arose: How genetic information can be stored in a DNA molecule? The answer came in 1953, when Watson and Crick [11] determined that the molecular structure of the DNA had the form of double helix, that is, two strands of DNA interlaced in a spiral. Thereby, the sequence adenine (A), thymine (T), guanine (G) and cytosine (C), represents the information that is encoded by the DNA molecule, in which the adenine in a filament is always paired with the thymine, and the guanine with the cytosine. That same year, Sanger and Thompson [12], [13], reported, in two scientific publications, the first sequence of a protein, insulin [4].

The genetic information must present four characteristics: replication, information storage, information expression and variation by mutation. Through cell division, replication of genetic material occurs; the molecule stores genetic information that may, or may not, manifest in the cell, depending, for example, on climatic conditions. It is important that genetic material encodes the genetic products that are found in life forms, so that they can transmit it to the cells and organisms of the offspring; the expression of the information, which comprises the transcription of the DNA, in which three main types of RNA molecules (ribonucleic acid) are synthesized: messenger RNA (mRNA), transporter RNA (tRNA) and ribosomal RNA (rRNA). The mRNAs are translated into proteins, mediated by tRNAs and rRNAs, with each one representing the product of a specific gene, leading to the synthesis of a protein. In translation, mRNA guides the construction of a chain of amino acids, called polypeptide, which then, folds into a protein. At the end of this process, what is known as the central dogma of molecular genetics is formed [14], where "the DNA produces the RNA, which produces the proteins"; how genetic material is a source of variability among organisms, through mutation processes, when a change in DNA chemistry occurs, the change in transcription and translation affects the protein [6].

During the development of a cell, they need a mechanism to turn genes on or off to form specific cell and tissue types and on specific occasions [4]. Conceptually, this was reported by Jacob and Monod (1961) [15], when they demonstrated that the genes have regulatory elements, gene on or off, called gene expression. That is, the regulatory elements are DNA sequences that receive the binding of a protein performing the function of activator or repressor of the gene expression [4].

In 1961, Crick *et al.* [16], discovered the basic unit of a protein. So, it was possible to deduce as a sequence of nucleotides in the DNA, with each one having a base, A, T, C or G, codifies the set of 20 amino acids to form the proteins. Since then, the genetic code has been "unveiled", making it possible to "read" it [4]. This result was published by Crick (1968) [17], where the general characteristics of the genetic code

were described. In 1965, nucleotides were sequenced by a tRNA, from the amino acid called alanine [18].

In the area of computing, the 1950's marked the beginning of commercial computing, accessible to companies and other organizations. The first operating systems appeared, the magnetic disk, the first programming languages and the transistors, because, until then, computers operated with valves [9].

In the following decade the integrated circuit was invented, which allowed the construction of a whole set of transistors from a single block, without the need for welds and wires. Packet switching also arose, which allowed the transmission of information between computers through several different paths, in which the message was divided into information packets, giving rise to the principle of the computer network, designed by ARPANET (Advanced Research Projects Agency Network).

Other examples of inventions from this decade include the touch screen, virtual and augmented reality, notebook, microprocessors and computer graphics [9]. At the beginning of this decade, Dayhoff and Ledley (1962) [19], developed a computational system to help determine the primary protein structure. After four years, a molecular model was created to be visualized by computer [20].

In the early 1970s, Ramachandran and Lakshminarayanan (1971) [21] developed a technique for the mathematical process of reconstruction of an object in three dimensions. Four years later, a computational simulation was developed to demonstrate the protein entanglement, that is, it can be visualized in three-dimensional form [22].

The advance in computer processing capacity has allowed the creation of more complex systems. For example, Needleman and Wunsch (1970) [23], developed a computer adaptive method to find similarities in two protein sequences. To measure the similarity of sequences, the Needleman-Wunsch algorithm, by means of dynamic programming, finds the optimal alignment of sequence pairs, dividing the complete sequence into smaller segments. The similarities in the sequences receive a score that are inserted in a matrix, and thus, the algorithm detects gaps in the sequence alignment [24]. Years later, Murata, Richardson and Sussman (1985) [25], extended the Needleman-Wunsch algorithm, in order to allow the comparison of three biological sequences, however, the processing time for its execution made it unfeasible. Then, two years later, Feng and Doolittle (1987) [26], developed another algorithm, also based on Needleman-Wunsch, which made it possible to compare the sequence of three proteins.

In 1972, there was the first sequence of the Bacteriophage MS2, a biological virus [27]. Four years later, the complete MS2 virus genome was sequenced [28]. The DNA sequencing techniques continued to evolve, for example, in the researches reported by Sanger and Coulson (1975) [29], Maxam and Gilbert (1977) [30] and Sanger, Nicklen and Coulson (1977) [31]. All efforts made in this direction have made possible increasingly adequate methods to determine the nucleotide sequence of DNA molecules, for example, the creation of the computer program developed by Staden (1979) [32].

In the early 1970s the first computer virus was created. In this decade computers were also introduced for personal use and the modem, which is an equipment that allows communication between computers through a telephone network. ARPANET started connection tests with other organs, through TCP/IP (Transmission Control Protocol / Internet Protocol) and the relational databases emerged [9]. The emergence of the databases allowed the storage of biological data, for example, DNA sequences, whose structures are complex, because they contain numerous relationships between data elements and overlapping data fragments. However, the application of the computational technique called data mining, which involves statistical analysis and artificial intelligence, made it possible to analyze the data [5]. Statistics were also used to obtain the DNA sequence, applying the technique called maximum likelihood [33].

In 1984, other computer programs of genetic manipulation appeared, for example, Malthiery *et al.* [34] e Johnsen [35]. Another program was developed by a team from the University of Wisconsin Genetics Computer Group, whose set of computational tools was named GCG [36].

In 1962, a group of scientists met to discuss the possibility of creating an international molecular biology laboratory, giving rise, the following year, to EMBO (European Molecular Biology Organization), which, later, in 1974, became a legal entity, originating the EMBL (European Molecular Biology Laboratory). And, in 1981, the world's first nucleotide sequence data repository was released [37]. The following year, the EMBL begins working in cooperation with GenBank [38], belonging to the NCBI (National Center for Biotechnology Information) [39] and request assistance from the Japanese database, which later in 1987, is officially inaugurated, receiving the name of DDBJ (DNA Data Bank of Japan); however, they had already collaborated with the project since the early 1980s [40]. In 2005, DDBJ, EMBL and NCBI, in common accord, accept to call the mutual collaboration of INSDC (International Nucleotide Sequence Database Collaboration) [41].

In 1988, employees of EMBnet (European Molecular Biology Network) have teamed up with collaborators from other locations to develop an extension package for the GCG program; the new system was named EGCG, distributing it for free, however, when the GCG stopped providing its program code, it was no longer possible to distribute the EGCC [42]. In this context, EMBOSS (European Molecular Biology

Open Software Suite) emerged in 1996, whose objective was to develop tools for molecular sequence analysis and provide them free of charge [43].

In 1985, the first journal focused on the publication of works related to biological computing was established, called CABIOS (Computer Applications in the Biosciences) [44]. Currently, the journal is entitled Bioinformatics [45].

In the area of information technology, the 1980's were marked by the popularization of personal computers and the emergence of the Internet, whose milestone was reached when ARPANET connected to other networks around the world, mainly government agencies and universities. In this decade were also initiated the first studies and theoretical developments related to quantum computing [9].

In 1995, researchers at TIGR (The Institute for Genomic Research), Johns Hopkins University School of Medicine and State University of New York, were the first to achieve the complete genome sequence of a bacterial life species, *Haemophilus influenzae* [46]. The 20th century ended with the complete sequencing of a eukaryote, *Saccharomyces cerevisiae* [47]; of an animal, *Caenorhabditis elegans* [48]; and a plant, *Arabidopsis thaliana* [49].

In the 1990s, the World Wide Web established itself as the leading Internet service and wireless network standard, Wi-Fi, was created, inspired by the cable network standard, Ethernet [9].

In February 2001, the first results of the sequencing of the human genome were published, in which, in Nature magazine, included analysis of initial sequences generated by the Human Genome Project[1] [50], with public sponsorship; in the publication of Science, the focus was on reporting a draft of the sequence [51], obtained by the private company Celera Genomics [52]. In 2003, the Human Genome Project completed the sequencing of all genes of a human being [53].

In 2005, researchers from Life Science Corp., The Rockefeller University and The Rothberg Institute, reported a sequencing system, whose method, according to the authors, surpassed the best systems, being therefore the best of its time [54]. The Genomic Standards Consortium[2] was also founded that year, with the objective of being an open community for development, implementation and standardization in the genomic field [55].

In the first decade of the 21st century, computers were connected in a worldwide network, the devices began to be miniaturized and, in 2007, the first quantum computer was presented [9].

Within this context, after the explanation of some concepts and historical events, a more complete definition of what is bioinformatics can be given by Araújo *et al.* [56]: Bioinformatics includes the use of computer programs to treat biological data and identify gene sequences, predict the three-dimensional configuration of proteins, identify enzymatic inhibitors, promote protein grouping, establish phylogenetic trees (graphic representation to demonstrate the evolutionary relations between species) and analyze experiments in gene expression. In this sense, bioinformatics is responsible for ordering and grouping the results generated by the analysis of gene sequencing, produced from the compositions of DNA, RNA and proteins. Therefore, due to the large amount of biological data, molecular biologists started to use statistical methods capable of analyzing them, and thus, deducting functions from genes and demonstrating relationships between genes and proteins.

In view of this, having a more complete definition of the term bioinformatics, in the next section it will be suggested a way to consider ethical principles to use it in order to maximize the benefits that this science can provide to human beings.

III. ETHICS IN BIOINFORMATICS

To begin this essay on how responsibility is a fundamental factor to guide human actions, the words of the Argentine thinker and humanist, Carlos Bernardo González Pecotche (1901-1963), founder of Logosophical Science, will guide the analysis: "Freedom, which is the essential foundation of life, forms the vertex of the triangle whose base rests on duty and the right. In the face of this ternary that shapes the synthesis of human responsibility, it will be necessary to raise the conscience of men and to make it manifest itself in all its splendor and in its maximum potency."[3](p. 199; Author translation) [57].

Freedom, supported on one side by the duty, and the other, the right, leads to the reflection that for the individual to have freedom, it is necessary to ponder between your duty, as an individual, towards other people, and only then, check your right. That is, how you can act without hurting, under no circumstances, the freedom or right of another person.

So, the freedom of an individual to act, weighted by the duty and the right, will summarize the essence of responsibility, because the acting of the individual will be dosed with what he wants to do with what he can effectively do, to thereby, maintain their freedom and harmony in society.

Since freedom is the action of the individual, aligned with your predilections and your will, according to what is your right, balanced with your duty, will culminate in your responsibility to the like. That way,

1 https://www.genome.gov/human-genome-project

2 https://gensc.org/

3 In Portuguese language: "A liberdade, que é fundamento essencial da vida, forma o vértice do triângulo cuja base repousa no dever e no direito. Perante este ternário que plasma a síntese da responsabilidade humana, será preciso erguer a consciência dos homens e fazer com que ela se manifeste em todo o seu esplendor e na sua potência máxima."

the sense of responsibility can be a principle to be considered when making use of bioinformatics. In other words, the result achieved with bioinformatics, under no circumstances, may hurt another person. In this sense, we start with responsibility, knowing that, in this concept, freedom is contained, balanced by duty and right.

Therefore, the central question in focus is: How to make use of bioinformatics responsibly?
Doing a reflection exercise, expanding the borders for a more comprehensive panorama, for example, of the human being consider the responsibility towards all other forms of life on the Planet, and then, will be to consider the responsibility as a fundamental starting point in research, with bioinformatics, involving genetic alteration, because it implies in hurting the right of the species of life to continue being what Nature has determined, therefore, the use of bioinformatics produced a result that hurt another living being, and, by consequence, there was no responsibility in the actions.

In view of how the application of ethical principles can guide the use of technologies, the following are the final considerations, which include points of relationship between the biological and computational sciences.

IV. FINAL CONSIDERATIONS

In this work a brief evolution of genetics and computers was reported, tracing a parallel between both. And after that, was suggested how to consider ethical principles to guide bioinformatics professionals to use it in the best way.

When Mendel discovered the law of inheritance – in its physical manifestation –, until the construction of the first computers, approximately 70 years have passed. However, in his experiments, Mendel had already used statistical and mathematical techniques, basing what would become, a century ahead, the so-called bioinformatics.

From the appearance of the Turing machine to computers for personal use, approximately 40 years passed; at the same time that computers became popular, research on quantum computers was started, which would be built around 30 years later.

In fact, there has been an accelerated growth in the area of computer science. In this sense, it becomes even more important for there to be a greater dialogue between professionals in the various areas of knowledge with professionals in Information and Communication Technology, so that they can mutually find appropriate solutions related to ethical issues, maturing the dialogue to deal with the different adverse situations that may arise.

Regarding the parallel traced between progress in the areas of genetics and computing, can highlight moments when an area, to some extent, has been related, more closely, with the other.
For example, in the 1970s, there was the complete sequencing of a biological virus. Analogous to its sequencing, in 1971, the first computer virus was created, that, similar to a biological virus, replicates and infects other computers.

In the 1980's, computer programs for genetic sequencing have emerged. Through the World Wide Web, such programs could make use of biological data stored in central repositories, as that of the EMBL database.

The 20th century ended with the complete sequencing of a plant and an animal, and the beginning of the next century, the complete sequence of the human genome was reported. Therefore, it can be inferred that the introduction of computers made it possible to obtain the integral sequencing of the human genome.
Finally, all science must be practiced to improve the living conditions of human beings, considering it as a central focus. Then, the following question will remain open for further work: What is the role of applied ethics, be it bioethics, computer ethics, both, or a third branch, can perform to contribute in providing good ideas or indications to make the right use of bioinformatics?

REFERENCES

[1] CROCETTI, Zeno. **Ethics and Citizenship** [*Ética e Cidadania*]. 1. ed. Curitiba: IESDE Brasil, 2012.

[2] MOOR, James H.. What is computer ethics?. **Metaphilosophy**, v. 16, n. 4, p. 266-275, 4 Oct. 1985.

[3] HONGLADAROM, Soraj. Ethics of Bioinformatics: A Convergence between Bioethics and Computer Ethics. **Asian Biotechnology and Development Review**, v. 9, n. 1, p. 37-44, Nov. 2006.

[4] GRIFFITHS, Anthony J. F. *et al.***Introduction to Genetic Analysis**. Translated by Sylvia Werdmüller von Elgg Roberto [*Introdução à Genética*]. 11. ed. Rio de Janeiro: Guanabara Koogan, 2016.

[5] RAMAKRISHNAN, Raghu; GEHRKE, Cornell. **Database Management Systems**. Translated by Célia Taniwake; João Eduardo Nóbrega Tortello [*Sistemas de Gerenciamento de Bancos de Dados*]. 3. ed. Porto Alegre: AMGH, 2011.

[6] KLUG, William S. *et al.***Concepts of genetics**. Translated by Maria Regina Borges-Osório Rivo Fischer. [*Conceitos de Genética*]. 9. ed. Porto Alegre: Artmed, 2010.

[7] Villanova University. **Gregor Mendel's "Experiments in Plant Hybridization"**. Villanova University. n.d.. Available in: https://www1.villanova.edu/villanova/president/university_events/mendelmedal/aboutmendel/experiments.html. Access in: Aug. 31, 2020.

[8] TURING, Alan M.. On Computable Numbers, with an Application to the Entscheidungsproblem. **Proceedings of the London Mathematical Society**, v. s2-42, n. 1, p. 230-265, Jan. 1, 1937.

[9] WAZLAWICK, Raul Sidnei. **History of Computing** [*História da Computação*]. 1. ed. Rio de Janeiro: Elsevier, 2016.

[10] CRUZ, Carlos Henrique Brito. Vannevar Bush: An Introduction [*Vannevar Bush: Uma Apresentação*]. **Revista Latinoamericana de Psicopatologia Fundamental**, São Paulo, v. 14, n. 1, Mar. 2011. Available in: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47142011000100001. Access in: Sep. 13, 2020.

[11] WATSON, James D.; CRICK, Francis H. C.. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. **Nature**, v. 171, p. 737-738, Apr. 25, 1953.

[12] SANGER, F.; THOMPSON, E. O. P.. The Amino-acid Sequence in the Glycyl Chain of Insulin. 1. The Identification of Lower Peptides from Partial Hydrolysates. **Biochemical Journal**, v. 53, n. 3, p. 353-366, Feb. 1953.

[13] SANGER, F.; THOMPSON, E. O. P.. The Amino-acid Sequence in the Glycyl Chain of Insulin. 2. The Identification of Lower Peptides from Partial Hydrolysates. **Biochemical Journal**, v. 53, n. 3, p. 366-374, Feb. 1953.

[14] CRICK, Francis. Central Dogma of Molecular Biology. **Nature**, v. 227, p. 561-563, Aug. 8, 1970.

[15] JACOB, François; MONOD, Jacques. Genetic Regulatory Mechanisms in the Synthesis of Proteins. **Journal of Molecular Biology**, v. 3, n. 3, p. 318-356, Jun. 1961.

[16] CRICK, Francis H. C. *et al.* General Nature of the Genetic Code for Proteins. **Nature**, v. 192, p. 1227-1232, Dec. 30, 1961.

[17] CRICK, Francis H. C.. The Origin of the Genetic Code. **Journal of Molecular Biology**, v. 38, n. 3, p. 367-379, Dec. 28, 1968.

[18] HOLLEY, R. W. *et al.* Structure of a Ribonucleic Acid. **Science**, v. 147, n. 3664, p. 1462-1465, Mar. 19, 1965.

[19] DAYHOFF, Margaret Oakley; LEDLEY, Robert S.. Comprotein: A Computer Program to Aid Primary Protein Structure Determination. *In:* ASSOCIATION FOR COMPUTING MACHINERY. 1962. **Proceedings [...]** New York, 1962. 262-274 p.

[20] LEVINTHAL, C.. Molecular Model-building by Computer. **Scientific American**, v. 214, n. 6, p. 42-52, Jun. 1966.

[21] RAMACHANDRAN, G. N.; LAKSHMINARAYANAN, A. V.. Three-dimensional Reconstruction from Radiographs and Electron Micrographs: Application of Convolutions instead of Fourier Transforms. **Proceedings of the National Academy of Sciences**, v. 68, n. 9, p. 2236-2240, Sep. 1, 1971.

[22] LEVITT, Michael; WARSHEL, Arieh. Computer Simulation of Protein Folding. **Nature**, v. 253, p. 694-698, Feb. 1, 1975.

[23] NEEDLEMAN, Saul B.; WUNSCH, Christian D.. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of two Proteins. **Journal of Molecular Biology**, v. 48, n. 3, p. 443-453, Mar. 28, 1970.

[24] LESK, Arthur M.. **Bioinformatics**. Britannica. n.d.. Available in: https://www.britannica.com/science/bioinformatics. Access in: Sep. 11, 2020.

[25] MURATA, M.; RICHARDSON, J. S.; SUSSMAN, J. L.. Simultaneous Comparison of Three Protein Sequences. **Proceedings of the National Academy of Sciences of the United States of America**, v. 82, n. 10, p. 3073-3077, May. 1985.

[26] FENG, Da-Fei; DOOLITTLE, Russell F.. Progressive Sequence Alignment as a Prerequisitetto Correct Phylogenetic Trees. **Journal of Molecular Evolution**, v. 25, p. 351-360, 1987.

[27] JOU, W. *et al.* Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein. **Nature**, v. 237, p. 82-88, May. 12, 1972.

[28] FIERS, W. *et al.* Complete Nucleotide Sequence of Bacteriophage MS2 RNA: Primary and Secondary Structure of the Replicase Gene. **Nature**, v. 260, n. 5551, p. 500-507, Apr. 8, 1976.

[29] SANGER, F.; COULSON, A. R.. A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. **Journal of Molecular Biology**, v. 94, n. 3, p. 441-448, May. 25, 1975.

[30] MAXAM, Allan M.; GILBERT, Walter. A New Method for Sequencing DNA. **Proceedings of the National Academy of Sciences of the United States of America**, p. 560-564, Feb. 1977.

[31] SANGER, F.; NICKLEN, S.; COULSON, A. R.. DNA Sequencing with Chain-terminating Inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 12, p. 5463-5467, Dec. 1977.

[32] STADEN, R.. A Strategy of DNA Sequencing Employing Computer Programs. **Nucleic Acids Research**, v. 6, n. 7, p. 2601-2610, Jun. 11, 1979.

[33] FELSENSTEIN, Joseph. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. **Journal of Molecular Evolution**, v. 17, p. 368-376, Nov. 1981.

[34] MALTHIERY, Bruno *et al.* Apple II Pascal Programs for Molecular Biologists. **Nucleic Acids Research**, v. 12, n. Issue 1Part2, p. 569-579, Jan. 11, 1984.

[35] JOHNSEN, Morten. JINN, an Integrated Software Package for Molecular Geneticists. **Nucleic Acids Research**, v. 12, n. Issue 1Part2, p. 657-664, Jan. 11, 1984.

[36] DEVEREUX, John; HAEBERLI, Paul ; SMITHIES, Oliver. A Comprehensive Set of Sequence Analysis Programs for the VAX. **Nucleic Acids Research**, v. 12, n. Issue 1Part1, p. 387–395, Jan. 11, 1984.

[37] EMBL. **EMBL History**. EMBL. n.d.. Available in: https://www.embl.de/aboutus/general_information/history/index.html. Access in: Sep. 13, 2020.

[38] NCBI. **GenBank Overview**. NCBI. n.d.. Available in: https://www.ncbi.nlm.nih.gov/genbank/. Access in: Sep. 13, 2020.

[39] NCBI. **Programs & Activities**. NCBI. n.d.. Available in: https://www.ncbi.nlm.nih.gov/home/about/programs/. Access in: Sep. 13, 2020.

[40] DDBJ. **About DDBJ Center**. DDBJ. n.d.. Available in: https://www.ddbj.nig.ac.jp/aboutus-e.html. Access in: Sep. 13, 2020.

[41] INSDC. **International Nucleotide Sequence Database Collaboration**. INSDC. n.d.. Available in: http://www.insdc.org/. Access in: Sep. 13, 2020.

[42] RICE, Peter *et al.***History: Background to EMBOSS**. EMBOSS. 2009. Available in: http://emboss.open-bio.org/html/use/ch01s01.html. Access in: Sep. 13, 2020.

[43] RICE, Peter *et al.***Introduction to EMBOSS**. EMBOSS. 2009. Available in: http://emboss.open-bio.org/html/use/pr02s01.html. Access in: Sep. 13, 2020.

[44] BEYNON, R. J.. CABIOS Editorial. **Bioinformatics**, v. 1, n. 1, p. 1, Jan. 1985.

[45] Oxford Academic. **About the Journal**.Academic. n.d.. Available in: https://academic.oup.com/bioinformatics/pages/About. Access in: Sep. 13, 2020.

[46] FLEISCHMANN, R. D. *et al.* Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd. **Science**, v. 269, p. 496-512, Jul. 28, 1995.

[47] GOFFEAU, A. *et al.* Life with 6000 Genes. **Science**, v. 274, n. 5287, p. 546-567, Oct. 25, 1996.

[48] C. Elegans Sequencing Consortium. Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology. **Science**, v. 282, n. 5396, p. 2012-2018, Dec. 11, 1998.

[49] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. **Nature**, v. 408, p. 796-815, Dec. 14, 2000.

[50] LANDER, E. *et al.* Initial Sequencing and Analysis of the Human Genome. **Nature**, v. 409, p. 860-921, Feb. 15, 2001.

[51] VENTER, J. C. *et al.* The Sequence of the Human Genome. **Science**, v. 291, n. 5507, p. 1304-1351, Feb. 16, 2001.

[52] LOPES, Guilherme S.. Human Genome Project. In: ZEIGLER-HILL, Virgil. **Encyclopedia of Personality and Individual Differences**. 1. ed. Springer International Publishing, 2017, p. 1-4. Available in: https://link.springer.com/referenceworkentry/10.1007/978-3-319-28099-8_724-1. Access in: Sep. 13, 2020.

[53] National Human Genome Research Institute. **Human Genome Project Timeline of Events**. Genome. 2018. Available in: https://www.genome.gov/human-genome-project/Timeline-of-Events. Access in: Sep. 17, 2020.

[54] MARGULIES, M. *et al.* Genome Sequencing in Microfabricated High-density Picolitre Reactors. **Nature**, v. 437, p. 376-380, Jul. 31, 2005.

[55] FIELD, Dawn *et al.* Genomic Standards Consortium Projects. **Standards in Genomic Sciences**, v. 9, n. 3, p. 599-601, 15 Jun. 2014.

[56] ARAÚJO, Nilberto Dias de *et al.* The Age of Bioinformatics: Its Potential and Implications for the Health Sciences [*A Era da Bioinformática: Seu Potencial e suas Implicações para as Ciências da Saúde*]. **Estudos de Biologia**, v. 30, n. 70/71/72, p. 143-148, 2008.

[57] PECOTCHE, Carlos Bernardo González. **Logosophy Magazine Collection: Tomo III** [*Coletânea da Revista Logosofia: Tomo III*]. Translated by José Dalmy Silva Gama. 1. ed. São Paulo: Editora Logosófica, 2010. Translation: *Coleccíon de la Revista Logosofia*.

★★★