HISTORY OF BIOINFORMATICS

Yusupov Ibragim Mirsaydaliyevich
Kokand State Pedagogical Institute, Associate Professor

Anatation. ***For today's students and researchers, it is easy to believe that modern bioinformatics appeared recently to help analyze the next generation of data. However, the beginning of bioinformatics occurred more than 50 years ago, when desktop computers were still hypothesized and DNA could not yet be sequenced. The basis of bioinformatics was laid by the use of computational methods in the analysis of protein sequences in the early 1960s. Subsequently, DNA analysis appeared due to the parallel development of (1) molecular biology methods, which made it easier to synthesize DNA, as well as its sequence and (2) the emergence of increasingly smaller and more powerful computers in Computer Science. It also needed a more appropriate new software to perform bioinformatics functions. From the 1990s to the 2000s, massive improvements in sequential technology have resulted in exponential growth of data along with cost reduction. The emergence of" Big Data "("Big Data") has come up with new challenges in terms of finding and managing data and has required more experience in the field of Computer Science. In combination with constantly increasing bioinformatics tools, biological Big Data (Big Data) has had a profound impact on the predictive power and recurrence of bioinformatics results and continues to do so. To address this problem, universities now fully integrate this science into the curriculum of biology students. Recent junior disciplines such as synthetic biology, System Biology, and whole cell modeling have emerged as a result of an ever-increasing overlap between computer science and biology.***

Key words: ***bioinformatics, origin of bioinformatics, genomics, structural bioinformatics, Big Data (Big Data), future of bioinformatics***

Computers and specialized software have become an important part of a biologist's set. To a certain extent, almost all modern research projects in biology require the use of computers in order to regularly analyze DNA or protein sequences or to analyze meaningful data from a large collection of biological data of gigabytes in size. This is especially true since the emergence of the next generation sequence (NGS-next generation sequencing), which has radically changed the direction of population genetics, quantitative genetics, molecular Systematics, microbial ecology and many other research areas.

In this regard, it is easy to believe that modern bioinformatics has appeared relatively recently for today's students and researchers and has come to the aid of the analysis of NGS (next generation sequence) data. However, the beginning of bioinformatics occurred more than 50 years ago, when desktop computers were still hypothesized and DNA could not yet be sequenced. Here we present an integrated table of key events in Bioinformatics and related fields over the past half century, as well as some information about parallel achievements in the field of Molecular Biology and informatics, and some ideas about the future of bioinformatics. We hope that this review will help the reader understand that bioinformatics has become the main driving force of biology today.

1950-1970: origin it heads with DNA analysis

At the beginning of the 1950s, much was not known about dezoksiribonucleic acid (DNA). His status as a carrier molecule of genetic information was still controversial at that time. Avery, MacLeod and McCarty (1944) showed that obtaining pure DNA from the virulent bacterial strain

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 11 Issue: 07 in July-2022
https://www.gejournal.net/index.php/IJSSIR

could give virulence to the novirulent strain, but their results were not immediately accepted by the scientific community. Many thought that proteins are carriers of genetic information. The role of DNA in the quality of genetic information encoder molekula was confirmed by Hershey and Chase in 1952 and they have no doubt proved that it is DNA, and not a protein that is accepted and transmitted by bacterial cells infected with bacteriophage.

Despite knowing its main role, much was not known about the location of the DNA molecule. That's what we knew, his pair of monomers (that is, nucleotides) were in equimolar proportions. In other words, the more adenosine, the more thymidine, the more guanidine and cytidine. Exactly in 1953 year, the double helix structure of DNA was finally solved by Watson, Creek and Franklin. Despite this achievement, it will take another 13 years to decode the genetic code and another 25 years until the first DNA sequencing methods are available. Consequently, the use of bioinformatics in DNA analysis has been almost twenty years behind the analysis of proteins whose chemical nature is better understood than DNA.

Protein analysis was the starting point

In the late 1950s, with the help of Crystallography, great progress was made in determining protein structures, in addition, insulin was discovered in the first sequence of proteins (that is, the order of the amino acid chain). This big jump solved the debate about the location of the Polyphonic chain of proteins. In addition, he encouraged the development of more effective methods of obtaining protein sequences. The Edman degradation method appeared as a simple method, which allowed the sequencing of one amino acid protein at a time, starting from the N - terminal. Together with automation, it was determined that in the next 10 years there were more than 15 different protein families in the sequence[1].

The main problem with Edman sequence was to obtain a large protein sequence. The Edman sequence works by shredding the N-terminal amino acid residues one by one with phenylisocyanate. However, the efficiency of this reaction will never be complete. Therefore, in a single Edman reaction, a theoretical maximum of 50-60 amino acids can be in a row. Larger proteins should be cut into smaller pieces, then they are separated and in a separate sequence.

The problem was not the spontaneous sequence of proteins, but the collection of the entire protein sequence from the sequence of hundreds of small Edman peptids. For large proteins, consisting of several hundreds (if not thousands) of residues, the return of the final sequence was laborious. In the early 1960s, the first known bioinformatics program was developed to solve this problem.

The main problem with Edman sequence was to obtain a large protein sequence. The Edman sequence works by shredding the N-terminal amino acid residues one by one with phenylisocyanate. However, the efficiency of this reaction will never be complete. Therefore, in a single Edman reaction, a theoretical maximum of 50-60 amino acids can be in a row. Larger proteins should be cut into smaller pieces, then they are separated and in a separate sequence.

The problem was not the spontaneous sequence of proteins, but the collection of the entire protein sequence from the sequence of hundreds of small Edman peptids. For large proteins, consisting of several hundreds (if not thousands) of residues, the return of the final sequence was laborious. In the early 1960s, the first known bioinformatics program was developed to solve this problem.

Dayhoff: the first bioinformatician

Margaret Dayhoff (1925-1983) was an American physical chemist who pioneered the use of computational methods in biochemistry[2]. Dayhoff's contribution to this area is so significant that

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 11 Issue: 07 in July-2022
https://www.gejournal.net/index.php/IJSSIR

David J. Lipman, a former director of the National Center for Biotechnology Information (NCBI), called him "the mother and father of bioinformatics" [3].

David J. Lipman is an American biologist, from 1989 to 2017 he was director of the National Biotechnology Information Center (NCI) at the National Institute of Health. NCBI is GenBank's home, the U.S. node of the international serial Database Consortium, and one of the most used sites in the world for searching and retrieving biomedical information is PubMed. Lipman is one of the original authors of the BLAST series adaptation program and is a respected figure in Bioinformatics. In 2017, he left NCBI and became the Chief Scientific Director of Impossible Foods.

Dayhoff has made extensive use of computational methods for his PhD thesis on electrochemistry and has seen the possibilities of computers in the fields of biology and medicine. In 1960 he became the director of the National Fund of biomedical resources. Dayhaff eri physicist Robert S. He started working with Ledli, who also wanted to bring computing resources to Biomedical Problems[4]. From 1958 to 1962 year ikkalasi also combined their own experiments and developed a "complete computer program for IBM 7090" Comprotein, designed to determine the basic structure of the protein using the data from the Edman peptid sequence. Fully coded on perfokartas at FORTRAN, this software is the first example of what we call de novo today.

In the program of COMPROTEIN, the input and output sequence of amino acids is expressed in three-letter abbreviations (for example, Lys for lysine, Sir for serine). In order to simplify the work with the data of the Protein sequence, Dayhoff later developed a single-letter amino acid kodini, which is still used today. This is the first time Dayhoff and Richard V for a one-letter code. In the Atlas of protein sequence and structure of Eck (1922-2006) in 1965 year, the first biological sequence was used in the database. The first edition of the Atlas contained 65 protein sequences, most of which were interspecific variants of a handful of proteins. Therefore, the first Atlas was an ideal data set for two researchers who hypothesized that the Protein sequence would reflect the evolution history of the species.

Genealogy of life with the help of a computer

Although much of the research conducted on biochemistry until the 1960s focused on mechanical modeling of enzymes, Emil Zukerkandl and Linus Pauling came out of this paradigm by studying biomolecular sequences as “information carriers”. As the specific location of words is a series of letters that denote the meaning, the molecular function of the protein (that is, the meaning) depends on how its amino acids are located to form a "word". Knowing that words and languages evolve over time by inheriting subtle changes, can the protein sequence evolve through similar mexanizm? Can these hereditary changes allow biologists to restore the history of the evolution of these proteins and restore the sequence of their "ancestors" in the same process? Tsukerkandl and Pauling introduced the term "paleogenetics" in 1963 to introduce this new branch of evolutionary biology[5,6].

Emil Zuckerkandl is originally an Austrian French biologist and is considered one of the founders of the field of Molecular Evolution. He is famous for presenting the concept of the "molecular clock", which created the neural theory of molecular evolution with Linus Pauling.

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 11 Issue: 07 in July-2022
https://www.gejournal.net/index.php/IJSSIR

References:

1. Yuldasheva, D. H. (2020). O ‘ZBEKISTONDA TURIZMNI RIVOJLANTIRISHDA TARIHIY SHAHARLARNING O ‘RNI. *Интернаука*, (21-3), 70-71.
2. Abdullayeva, N. R., & Abdullayev, A. K. (2021). The Basis is a Mobile Industrial Robot Core Characteristics and Shape of the Spatial Structure. *International Journal on Orange Technologies*, *3*(3), 253-256.
3. Usmonova, M. (2019). PROFESSIONAL COMPETENCY BUILDING FUTURE BIOLOGY TEACHER. *European Journal of Research and Reflection in Educational Sciences Vol*, *7*(12).
4. Юлдашева, Д. Х. (2015). Роль инновационных методов обучения будущих младших специалистов. *Инновационная экономика: перспективы развития и совершенствования*, (2 (7)), 331-334.
5. Qizi, Y. N. V. (2021). System of Professional Competence Development of Future Teachers in the Field of Science and Communication. *Annals of the Romanian Society for Cell Biology*, *25*(4), 14211-14215.
6. https://en.wikipedia.org/wiki/Pehr_Victor_Edman
7. https://en.wikipedia.org/wiki/Margaret_Oakley_Dayhoff
8. https://en.wikipedia.org/wiki/David_J._Lipman
9. https://en.wikipedia.org/wiki/Robert_Ledley
10. https://en.wikipedia.org/wiki/Emile_Zuckerkandl
11. https://en.wikipedia.org/wiki/Linus_Pauling
12. G’URURNI, B. A. T. M. (2021). RIVOJLANTIRISH. *Editorial board: Tone Roald, PhD Associate Professor of Psychology University of Copenhagen*, 196.
13. Маргуба, Х., & Иброхимова, М. (2018). ИЗУЧЕНИЕ ОБРАЗОВАТЕЛЬНЫХ ВОПРОСОВ В СТАРИННЫХ ПИСЬМЕННЫХ ИСТОЧНИКАХ. *Актуальные научные исследования в современном мире*, (5-6), 88-89.
14. Хайдарова, М. (2018). ВОПРОСЫ ВОСПИТАНИЯ ВО ВЗГЯДАХ ВЕЛИКОГО ПРАВИТЕЛЯ АМИРА ТЕМУРА (1336-1405). *Актуальные научные исследования в современном мире*, (12-4), 15-17.
15. Inomdjanovva, K. M. (2021). Features of formation of national etiquette skills in primary school children. *ACADEMICIA: AN INTERNATIONAL MULTIDISCIPLINARY RESEARCH JOURNAL*, *11*(1), 1593-1595.
16. Inomovna, K. M. (2021). Features of formation of national etiquette skills in primary school children: A study. *ACADEMICIA: AN INTERNATIONAL MULTIDISCIPLINARY RESEARCH JOURNAL*, *11*(2), 1646-1648.
17. Inomovna, M. (2021). Formation of National Etiquette Skills in Reading Lessons among Primary Schoolchildren-As a Social Necessity. *Middle European Scientific Bulletin*, *17*, 294-296.
18. АРТЫКОВ, С. С., ХАЛИМОВА, М. Р., & ТАШПУЛАТОВА, Д. С. (2019). ОХРАНА РЕДКИХ И ИСЧЕЗАЮЩИХ ПТИЦ. In *МОЛОДЕЖЬ И НАУКА: ШАГ К УСПЕХУ* (pp. 140-141).
19. АРТЫКОВ, С. С., ХАЛИМОВА, М. Р., & ТАШПУЛАТОВА, Д. С. (2019). КОМНАТНЫЕ РАСТЕНИЯ И ЭКОЛОГИЯ ЖИЛИЩА. In *МОЛОДЕЖЬ И НАУКА: ШАГ К УСПЕХУ* (pp. 138-140).

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 11 Issue: 07 in July-2022
https://www.gejournal.net/index.php/IJSSIR

20. ТАШПУЛАТОВА, Д. С., & ХАЛИМОВА, М. Р. (2018). ПРОБЛЕМЫ РАСТИТЕЛЬНОГО МИРА И ЕГО ОХРАНА. In *Молодежь и системная модернизация страны* (pp. 27-28).

21. ХАЛИМОВА, М. Р., & ТАШПУЛАТОВА, Д. С. (2018). ОХОТНИЧЬЕ ПРОМЫСЛОВАЯ ФАУНА ЖИВОТНЫХ ФЕРГАНСКОЙ ОБЛАСТИ И ПЕРСПЕКТИВЫ ЕЁ ИССЛЕДОВАНИЯ. In *Молодежь и системная модернизация страны* (pp. 29-30).

22. ХАЛИМОВА, М. Р., & ТАШПУЛАТОВА, Д. С. (2017). АРЕАЛ ОБИТАНИЯ МИНДАЛЯ ОБЫКНОВЕННОГО (Amygdalus communis) И ЕГО ПРИМЕНЕНИЕ В МЕДИЦИНЕ. In *Будущее науки-2017* (pp. 345-346).

23. Tudiyeva, O. M., & Ibragimova, D. A. (2019). USE OF INNOVATIVE TEACHING METHODS TO IMPROVE "REPRODUCTIVE HEALTH". *Scientific Bulletin of Namangan State University*, *1*(5), 294-299.

24. Усмонов, С. О., & Мирзарахмонов, А. А. (2021). ИЗБИРАТЕЛЬНАЯ СИСТЕМА РЕСПУБЛИКИ УЗБЕКИСТАН." КОДЕКС О ВЫБОРАХ" И ЕГО ЗНАЧЕНИЕ. *Ученый XXI века*, (10 (81)), 21-25.

25. Азимов, М. М., Урманов, Х. Н., Усмонов, С. О., & Рўзиматов, Р. Ё. (2020). КЕЙСЛАРДАН ФОЙДАЛАНИБ "НУКЛЕИН КИСЛОТАЛАР, ДНК ВА РНК МОЛЕКУЛАСИ" МОДУЛИНИ ЎҚИТИШ. *Интернаука*, (21-3), 54-55.

26. Усмонов, С. О., & Мирзарахмонов, А. А. (2021). ИЗБИРАТЕЛЬНАЯ СИСТЕМА РЕСПУБЛИКИ УЗБЕКИСТАН." КОДЕКС О ВЫБОРАХ" И ЕГО ЗНАЧЕНИЕ. *Ученый XXI века*, (10 (81)), 21-25.

27. Азимов, М. М., Урманов, Х. Н., Усмонов, С. О., & Рўзиматов, Р. Ё. (2020). КЕЙСЛАРДАН ФОЙДАЛАНИБ "НУКЛЕИН КИСЛОТАЛАР, ДНК ВА РНК МОЛЕКУЛАСИ" МОДУЛИНИ ЎҚИТИШ. *Интернаука*, (21-3), 54-55.

28. Muxayyoxon, U., & Xilolaxon, U. (2022). KASB BU-HAYOT. *Yosh Tadqiqotchi Jurnali*, *1*(5), 327-333.

29. Араббоев, К. Т. (2011). ОСОБЕННОСТИ АДАПТАЦИИ СТУДЕНТОВ ВЫСШЕЙ ФИЗКУЛЬТУРНО-ПЕДАГОГИЧЕСКОЙ ШКОЛЫ К УЧЕБНОЙ И СПОРТИВНОЙ ДЕЯТЕЛЬНОСТИ. Педагогические науки, (5), 128-131.

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 11 Issue: 07 in July-2022
https://www.gejournal.net/index.php/IJSSIR