The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2018

Volume 4, Pages 171-175

IConTES 2018: International Conference on Technology, Engineering and Science

Microbial Bioinformatics

Rukiye COLAK SASMAZER
Tat Gıda R&D Center

Mihriban KORUKLUOGLU
Uludag University

Abstract: Applying computational techniques to analyze a multitude of information about biomolecules, bioinformatics now has a solid place as a discipline in the field of molecular biology and covers a range of topics from structural biology to genomic to gene expression studies.rnBioinformatics; a technology used by computers to store, access, use and distribute information about biological macromolecules such as DNA, RNA, and proteins. Bioinformatics is basically the creation of databases for the creation and storage of biological information. Microbial bioinformatics, together with computer support, is the science of numerical analysis of information on microorganisms. The aims of bioinformatics can be examined in three parts. Regarding these data, it is easy to arrange the data so that the researchers can easily reach it and record the new data as quickly as it is produced. The second goal is to develop tools and resources that enable this data to be meaningful. The development of these resources requires expertise in computational sciences. The third goal of bioinformatics is to analyze the data using these tools and turn them into biologically meaningful information.

Keywords: Bioinformatics, Technology, Biotechnology

Introduction

Applying computational techniques to analyze a multitude of information about biomolecules, bioinformatics now has a solid place as a discipline in the field of molecular biology and covers a range of topics from structural biology to genomic to gene expression studies.

Bioinformatics; a technology used by computers to store, access, use and distribute information about biological macromolecules such as DNA, RNA, and proteins. Bioinformatics is basically the creation of databases for the creation and storage of biological information. Microbial bioinformatics, together with computer support, is the science of numerical analysis of information on microorganisms (Yu et al., 2004).

With the BLAST sequence matching program, which is an important part of bioinformatics, the DNA sequence at hand can be analyzed in detail. BLAST (Basic Local Alingment Search Tool) is a computer program that compares the searched sequence (nucleotide or amino acid) with the base sequences of the microorganisms in the database and gives the microorganism belonging to the same or nearest sequence sequence as% similarity. BLAST is a database developed by the National Center for Biotechnological Information, founded in 1988, aiming to develop computer programs for the analysis of computational and genomic data in a source of information on molecular biology. After analyzing the sequence and determining the base sequence of the searched region, this sequence is compared with the database using the program on the internet page. The result of the search gives the percentage of similarity to which microorganism the sequence of the searched sequence may belong (Sayitoğlu M. 2007)

The aims of bioinformatics can be examined in three parts. Regarding these data, it is easy to arrange the data so that the researchers can easily reach it and record the new data as quickly as it is produced. The second goal is to develop tools and resources that enable this data to be meaningful. The development of these resources requires

expertise in computational sciences. The third goal of bioinformatics is to analyze the data using these tools and turn them into biologically meaningful information (Baldi ve Brunak, 2001) .

Bioinformatics in Turkey

TUBITAK, Advanced Bioinformatics and Genome Center (İGBA I) with the name of Turkey's most advanced gene laboratory Genetic Engineering and Biotechnology (GMB) was founded on the site. This center, which is strategically critical and genetic data security for Turkey, genetic data analysis (bioinformatics) and biosensor in work areas; creating added value from research and development to ensure Turkey's technological independence, it is working for maintaining and enhancing the competitiveness. The group of multidisciplinary (computer, electronics, mathematics, bioinformatics and biotechnology) team, together with consultants and partners from home and abroad, performs their work with product and future focus (mam.tubitak.gov.tr).

The Importance of Bioinformatics

The most important task of bioinformation is; quantitative data on the genomes of all biological species including human beings, protein sequences and three-dimensional structures of proteins, metabolic pathway databases, cell line (cell line line) and hybridoma information and biodiversity-related information. In recent years, bioinformatics has played a major role in the common applications of genomic sequence projects, the successful completion of the human genome project, and biotechnology-based production and process development. This information is mostly used in bioinformatics to reduce costs and shorten the realization period because the drug design is an expensive and time-consuming process to develop. Very large bioinformatics research and development groups have been established in almost all major pharmaceutical biotechnology companies. The latest information shows that biotechnology is the fastest growing production technology. Bioinformatics has already shown significant success in the field of medical diagnosis and treatment as well as in drug design (Hogue CWV 2002).

Drug Design in Bioinformatics

Rational drug design is one of the first medical applications of bioinformatics. From nucleotide sequences the possible amino acid sequences can be determined using translation software. Subsequently, homologous regions are identified in model organisms with the aid of sequence search techniques, and human proteins can be modeled using sequence similarities. Thus, it is possible to design drugs with molecules that can be linked to the model structure by the algorithm algorithm and the biological activity on the real protein can be tested (Gatto ML 2003).

As a result of tests to be carried out prior to drug treatment, it is possible to remove the deadly side effects such as toxicity in cancer chemotherapy and to develop patient-specific treatment protocols. In addition, by identifying gene expression profiling using advanced molecular techniques, it is possible to determine which genes are active in the patient, and the individuals who can benefit from the treatment can be determined in advance (Doery M. et al., 2003).

Developments in bioinformatics allow you to make predictions about your future health status. Thus, preventive medicine will become important and unnecessary costs and treatments will be avoided. Pesticide-based treatments can be customized against patients and disease and can be achieved with minimal side effects through effective medical treatment (Polat, M, Karahan, A.G. 2009).

Open Genomic Analysis in Bioinformatics Shiga-Toxin-producing *E. coli O104: H4* Investigation

Shiga-toxin producing *Escherichia coli O104:H4;* more than 3,000 people have been infected in Germany (June 2011), a cause of H4. In this case, a genomic analysis of an isolate was made from a member of the family to identify the associated case population. In this analysis, DNA sequencing technology, open source data version and mass-based analyzes were used. In less than a week, the epidemic strain was found to produce Shiga toxin 2, which belongs to the enteroaggregative E. coli strain containing acquired genes, and to be an antibiotic resistance.

Escherichia coli is a versatile pathogen that is common in warm-blooded organisms. Enterovirulent strains of *E. coli* are a Shiga-toxin producing enterohemorrhagic and enteroaggregative strains. Interoaggregative *E. coli* strains have been found to be associated with sporadic and epidemic diarrhea and to be associated with Hep-2 cells.

Shiga-toxin producing *E. coli* has been reported to inhibit protein synthesis in eukaryotic cells. Enterohemorrhagic strains produce E. coli, Shiga toxin and a specific protein secretion system and are usually associated with hemolytic uremia and cause anemia associated with renal failure, thrombocytopenia and neurological and myocardial damage (Holger et al 2011).

Biofilm Research in Bioinformatics

Biofilms can be defined as the ensemble of microorganisms living in a gelatinous layer in their polymeric structure adhered to a surface (Leone et al., 2006)

Microbial biofilms; (Fujishige, N.A., et al., 2006), resulting in the loss of millions of dollars due to damage to the instruments, product contaminations, energy losses and infectious diseases they cause.

In addition, there is a strong relationship between infectious diseases and biofilm-forming microorganisms, and these microorganisms have been reported to cause diseases such as intestinal inflammation of the heart, periodontitis, and cystic fibriosis. Examination of biofilm in bioinformatics became important because of all these risks (Aksu et al., 2011).

In this context, the application of a laser scanning microscope image to living, hydrated, biofilms has led to a revolution in biofilm research. This has followed a number of molecular techniques, such as hybridization of fluorescence, reporter gene placement, gene expression (genes are transformed into protein constructs) et al., 1991).

In conclusion, this information has provided important information on the specific genes and proteins required for biofilm formation and development, biofilm-stage specific gene expression, division during biofilm development, and localization of gene expression (Klausen et al., 2003, Labbate et al., 2004; Lenz et al., 2008).

What to Expect from Future Microbial Bioinformatics

In 1951 Pauling and Corey's approach to accurate prediction of the secondary structure of proteins, with the difficulty of determining precisely the beginning of bioinformatics, is regarded as the starting point for bioinformatics. Pauling; is a scientist closely interested in quantum mechanics, minerology, crystallography, structural chemistry, anesthesia, immunology, medicine and evolution. Since contemporary bioinformatics requires intensive support from the computer, the publication of the first article on the drawing of computer-generated molecular graphics in Scientific American magazine in 1966 is considered the beginning for bioinformatics. The terms molecular bioinformatics, computational biology, biocomputing, where the term bioinformatics began to be used after the mid-1980s, are used in the same sense as bioinformatics (Gentleman, R.C. et al., 2004).

The National Center for Biotechnology Information (NCBI) was established in 1988 as the most effective institution for the development of new methods for the analysis and interpretation of complex data. The Human Genome Project (HGP) covers a 13-year international study and it is known that in October 1990, 30-35 thousand human genes were primarily targeted to be identified and used in biological studies. Human Genome Project studies have indicated that it is a crucial driving force in the development of bioinformatics (Collins, F. S. et al., 2003).

The closer to the end of the last decade, the more microbial genomes and metagenomes; databases and research strategies. Even in 2016, BLAST surveys in the database of NCBI (National Center for Biotechnology Information - United States) have begun to confuse a number of identical or closely related sequence results, although there is no easy way to search and download metagenomic data accumulated by mankind. This situation is expected to deteriorate in the future. For example, in 2020, even if there are not millions of genomic sequences of basic bacterial species such as Escherichia coli or Mycobacterium tuberculosis, they will have

hundreds of thousands of them. New approaches to data storage and analysis will be required for such cases (Eisenstein, 2015).

Result

With bioinformatics it is stated that significant improvements will be made in prevention of diseases. For example, the great success of the human genome project allows individual genetic screening to be performed at a very low cost. The human genome map facilitates the gene identification associated with various diseases.
It is stated that the Human Genome Project has a quality that excites people with excitement as well as some anxiety due to the expectations that they have created, new developments, question marks, triggered by the results obtained with the aim of solving the life password of other creatures such as plants, animals and microorganisms far beyond the head. One of the reasons for this is the fact that the speed of genetic research has reached a level where society is difficult to track and understand. For this reason, the informatics attracts all attention. It is thought that the results of recent advances in genetic-molecular biology research will lead us to change some of our knowledge in areas such as history, sociology and anthropology not only in fields such as medicine, biology and biotechnology in the near future. Despite the increasing pace of genetic research, the results are the basis for new questions. For example, if the number of human genes is smaller than expected, researchers lead the question of whether a gene encodes more than one protein (Polat, M, Karahan, A.G. 2009).

Some Terms Used in Bioinformatics

Gen: It is the DNA chain that encodes a polypeptide that is a function in organism. Another definition consists of nucleotide sequences at a certain point in chromosomes carrying a specific function, such as a gene, protein or RNA molecule.

Genome: The sum of the genes in an organism is called the genome.

Metagenom: All of the genetic material found in environmental samples; It consists of genomes of many (microbial) organisms in the environment.

Accession number (GenBank): This is the unique ID number for this record when it is registered to a set of GenBank. Even if the information about the entered sequence is updated, it is not changed at all.

BLAST (Basic Local Alignment Search Tool): A high-speed computer program used to compare nucleotide or protein sequences between identical or different organisms and to search for similar regions.

Blosum (block substitution matrix): An exchange matrix consisting of values derived from observations of the frequency of change of blocks obtained by comparison of proteins.

Conserved sequence: A base sequence that remains unchanged in the evolutionary process in a DNA molecule (in the amino acid sequence of a protein).

Contig: A group of different DNA fragments cloned, overlapping a chromosome.

Domain: A part of a protein that is independently considered to be able to fold and work.

E value (expectation value): Expectation value. The number of similar sequences that have a score equal to or greater than S, which is expected to result in a chance of being scanned in the database scan. A low E value indicates a large skate.

EST (expressed sequence tag): A short piece of cDNA that can be used as a gene identity. It is used for locating and mapping genes.

Homologue: The sequence is a gene that looks like another gene in a big way. These genes are thought to have a common horse and carry similar functions.

Motif: A short, conserved region in the protein sequence. Motifs are generally high-conserved regions of domains.

References

Baldi, P., and Brunak, S., 2001. Bioinformatics The Machine Learning Approach, Massachusetts Institute of Technology.

Hogue CWV. Bioinformatics for Beginners. Trends Biochem. Sci. 2002; 27, 542-543.

Klausen, M., Aaes-Jorgensen, A., Molin, S., and Tolker-Nielsen, T. (2003) Involvement of bacterial migration in the development of complex multicellular structures in Pseudomonas aeruginosa biofilms. Mol Microbiol 50: 61–68.

Lee, K.W.K., Yam, K.H.J., Periasamy, S., Mukherjee, M., Kjelleberg, S., and Rice, S.A. (2016) Interspecific diversity reduces and functionally substitutes for intraspecific variation in biofilm communities. ISME J 10: 846–857

Law, Y.Y., Kirkegaard, R.H., Cokro, A.A., Liu, X., Arumugam, K., Chao, X., et al. (2016) Integrative microbial community analysis reveals full-scale enhanced biological phosphorus removal under tropical conditions.

Loman, M.B., Moritz Hentschke, M.D., Wentong Chen, B.S., Fei Pu, B.S., Yangqing Peng, B.S. (2011) Genome Analysis Crowd-Sourcing Consortium New England Journal of Medicine

Sayitoğlu M. .Kanser Tedavisine Farmakogenetik Yaklaşım. Türkiye Klinikleri J Med Sci., 2007; 27: 434-441.

Author Information

Rukiye Colak Sasmazer
Tat Gıda R&D Center Mustafakemalpaşa, Bursa / Turkey
Contact E-mail: *rukiye.colak@tat.com.tr*

Mihriban Korukluoglu
Uludag University
Bursa, Turkey