ISSN Print: 2394-7489
ISSN Online: 2394-7497
IJADS 2020; 6(3): 610-614
© 2020 IJADS
www.oraljournal.com
Received: 12-07-2020
Accepted: 25-08-2020

Dr. Kiran Ghatole
Reader, Al-Badar Dental College and Hospital, Department of Conservative Dentistry and Endodontics, Kalburagi, Karnataka, India

Corresponding Author:
Dr. Kiran Ghatole
Reader, Al-Badar Dental College and Hospital, Department of Conservative Dentistry and Endodontics, Kalburagi, Karnataka, India

Bioinformatics: A new arsenal in dental sciences

Dr. Kiran Ghatole

DOI: https://doi.org/10.22271/oral.2020.v6.i3i.1015

Abstract
Bioinformatics is an emerging field that focuses on using information science to solve biological problems. It is mainly concerned with collecting, storing, retrieving and analysing data from databases. In the last decade it has catalysed the research in the field of healthcare to a great extent. Bioinformatics can aid the research in dentistry by understanding the underlying pathways and mechanisms in certain oral diseases. It can also help in early prediction and personalized treatment of cancer. Application of Bioinformatics in drug design has been a boon to the pharmaceutical companies reducing the time required to develop a novel drug. It can also assist in developing patient care databases, image Processing of X-rays and CT can supplement the diagnosis. In this era of Machine Learning and Artificial Intelligence, incorporation of these technologies into dental sciences can improve the research, diagnosis and treatment to a great extent.

Keywords: Bioinformatics, healthcare, microarray, technology

Introduction
The Human genome project and development of high throughput technologies led to generate large quantity of data. The increasing demand to understand, analyse and interpret this data gave rise to a new field of science: Bioinformatics. Bioinformatics can be defined as an emerging discipline combining mathematics, computer science, and biology to solve biological problems. It is an interdisciplinary field requiring in-depth understanding of biological systems, algorithms and statistics. It is now becoming essential for management of data in modern biology and medicine. Bioinformatics now includes concepts of machine learning, artificial intelligence and neural networks to analyse biological data and give useful results and conclusions.

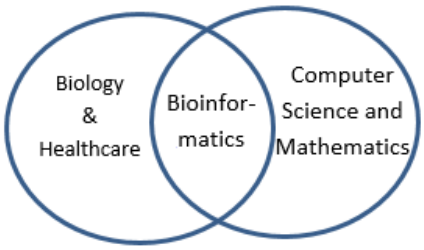

Fig 1: Show the different management of data in modern biology and medicine

Tools and techniques in bioinformatics
Alignment
Sequence alignment is the comparison of residues between sequences. This is done by aligning the unknown sequence with one or more known sequences to predict the common portions. Similar residues have similar functional and structural role tend to be preserved by natural selection in the course of evolution the optimum alignment arranges two or more sequences in such a way that a maximum number of identical or similar residues are matched. The sequences may be nucleotide sequences like DNA or RNA or amino acid sequences.

There are two types of sequence alignment, pairwise sequence alignment (PSA) and multiple sequence alignment (MSA). PSA considers two sequences at a time whereas MSA aligns more than two related sequences. MSA is more advantageous than PSA as it considers multiple members of a sequence family and thus provides more biological information. Proteins are key biological molecules that carry structural and functional information therefore sequence alignment at amino acid level is more relevant. Global alignment is performed on sequences with similar length and alignment has to be performed across the length. Local alignment is performed to know the local similar regions among the sequences. When there is a large difference in the lengths of the sequences to be compared, local alignment is generally performed [1]. A major etiological pathogen of dental caries is Glycosyltransferase enzyme from *Streptococcus mutans* [2]. The results obtained by Multiple Sequence alignment of Glycosyltransferase from *Streptococcus mutans* has are included for better understanding.

```
CLUSTAL O(1.2.4) multiple sequence alignment

tr|A0A2J9QD85|A0A2J9QD85_STRMG      MNEREALRTFGTGENFHAQHYFGFHETEKNGVKGYIFRVWAPN-AE----DLHLIGD---  52
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ----------------------MLTTTIIAGIISFILTILAMPFFIRFYQLKKINGQQM-  37
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      --MKDIFEKLAAQENLTEAEITAVAEKIFKGELSE-------AQIAAFLMALKIKGETPE  51
                                                  .  .   *  .                  :: *:

tr|A0A2J9QD85|A0A2J9QD85_STRMG      -FTGWF----DN--------PLQMDKNEAGVW--EVFTD-LPKEGHIYKYLVKRQGGQIV  96
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ---------------HEDVKQHLAKAGTPTMGGTVFLLVAALVTFICAFVLHITGGPAF  81
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      EILGLVKTIKKNAAVIPSQTSDAMDNCGTG--GDKSF---SFNISTTCSFVLAAGGIHIA  106
                                                : :  :       *            :::   *

tr|A0A2J9QD85|A0A2J9QD85_STRMG      EKMDPFAIYLEERPGTGSLIRTIPEKKWKDGLWLGRRKRWGFFKRPVNIY---EVHAGSW  153
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      GA-------------TLAILFIVLIY----------GTIGFLDDFLKIFKKINQGLTAW  117
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      KH-------------GN--RSVSSKS----------GSADVLEELG-----INVVLAPE  135
                                                :                 ..:.        :

tr|A0A2J9QD85|A0A2J9QD85_STRMG      KQNEDGSPYSFEQLKDELVPYLVKMNYTHVEFMPLMAHPLGMS--------------WGY  199
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      QKMT-------LQLIGGLVFYLVH-----------------------------------  134
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      KLAQ-------VLDETGIVFLFAQKMHPAMRYISPARQALGIPTVMNLIGPLTHPMDLET  188
                                    :               :*  :.:

tr|A0A2J9QD85|A0A2J9QD85_STRMG      QLMGFFAFEHTYGTPEQFQDFVEACHLNNIGVIVDWVPGHFTQNDDALAYFDGTPTFEYQ  259
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ------------------------------------VKPSGTDSLNLF-GFPL----  150
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      QLLGLYQADLQDD-------LAQVLK--LLGRKRAVII-TGPDNMDEAALY-GKNH----  233
                                                                          .. *    : *

tr|A0A2J9QD85|A0A2J9QD85_STRMG      DHDRAHNYRWGALNFDLGKNQVQSFLISSAKFWIDFYHIDGIRVD--------AVSN--M  309
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      -------------H------LGVFYIIFILFWVVGFSN-AVNLT--DGIDGLASISVVI  187
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      -------------YTLLD--NGQIS--QGSFTFEDFDMPKVTLEDIKGGDAKENAQILV  275
                                                .        :     * .  :    : :           .  :

tr|A0A2J9QD85|A0A2J9QD85_STRMG      LYLDYDEGPWQPNI-------EGNNRNLEGYYFLQRLNTVLKLAHPDVMMIAEESTATTK  362
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      SLLTYSVIAYMQNQFDVLLIISMVGALLGFFVYNRKPAKVFMG--DVGSLALGAMLA--  243
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      SVLKNEPSPYLE---------TTLLNAGLGFYANGK-----------VDSIKEGIDLA--  313
                                      *  .   :                  *::   :           *  :     :

tr|A0A2J9QD85|A0A2J9QD85_STRMG      ITGRREEGGLGFDYKWNMGWMNDILKFYEEDPIYRKYDFNLVTFSFMYLFSENFILPFSH  422
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ------AISITLRQEWTL----------------------LIIGIVYVLETASV-----  269
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      ------REVI-------------------------------ASGKAYA----------  324
                                           :                                  .  *

tr|A0A2J9QD85|A0A2J9QD85_STRMG      DEVVHGKKSLMHKMWGDRYNQFAGLRNLYTYQICHPGKKLLFMGSEFGQFLEWKYDHALE  482
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ----------M-------------------------------------LQ---VSYFK  277
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      ------------------------------------------------------------K  325
                                                                                                    :

tr|A0A2J9QD85|A0A2J9QD85_STRMG      WTNLEEEDGLNLKMQDFTSQLNQFYKDHKVLWQLDTSYDGLEIIDADNVDESVLSFIRKN  542
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      WTKKRKGEG-------------------------------------------------  286
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      L-------------------------------------------------------  326


tr|A0A2J9QD85|A0A2J9QD85_STRMG      EKGDLLVCVFNMVPVERKGFTIGVPVAGIYEEVWNTELEEFGGVWKEHNMTVKTQKNLWK  602
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ------QRIFRMTPFHHHL-----EL-------GGLRLRESGKKWSEWQVDF----FLWS  324
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      ----------------------------------------KALQEAQVD---------  335
                                                                                .* ::

tr|A0A2J9QD85|A0A2J9QD85_STRMG      EYENTLSFTLPALGASIWKIKRRLRK   628
tr|A0A2J9QGH7|A0A2J9QGH7_STRMG      ---------IGLLGSLLILTILYL--   339
tr|A0A2J9QGC2|A0A2J9QGC2_STRMG      -------------------------   335
```

Fig 2: Multiple Sequence Alignment performed in Clustal Omega (Uniprot IDs: A0A2J9QGH7, A0A2J9QD85, A0A2J9QGC2|)

Phylogenetic analysis

A phylogenetic tree depicts relationships and provides a visual representation of the estimated branching order of the taxonomical groups. It can be used to determine relationships among a collection of viruses, bacteria, animal or species, or other operational taxonomic units. They also play a role in identifying viral subtype and new recombinant subtypes arising from combinations of known subtypes. Probabilistic methods and matrices are used to determine the branching, distance between branches and branch length. Factors such as population subdivision, migration, and changing population size impact the tree. Forces such as recombination, mutation, and selection impact the genetic data of a given tree [3].

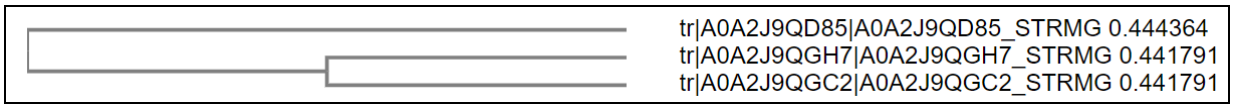

Fig 3: Phylogenetic performed in Clustal Omega (Uniprot IDs: A0A2J9QGH7, A0A2J9QD85, A0A2J9QGC2)

Protein structure and function prediction

Protein structure prediction using bioinformatics includes different methods like sequence similarity searches, MSA, identification of domains, secondary structure prediction, solvent accessibility prediction, fold recognition, constructing 3D models and model validation [4] Structure prediction is often divided into three areas: ab-initio prediction, fold recognition, and homology modelling. Ab-initio method predicts the structure of a protein based on the laws of physics and chemistry. The term ab-initio or de novo is also applied to the prediction of the structure of proteins for which there is no similar structure in the Protein Database (PDB) but where local sequence and structural relationships involving short protein fragments, as well as secondary structure prediction, are incorporated into the prediction process. Prediction of the three-dimensional structure of a protein from the known structures of one or more related proteins is called Homology Modelling. From a given library of known structures, we can determine which of them share a folding pattern with a query protein of known sequence but unknown structure is known as fold recognition. In fold recognition one or more structures similar to a given unkown sequence exist in the PDB but are not easily identified [5]. The function of the protein depends on the structure. Function prediction from structure can be achieved by detecting homology or using of structural templates derived from the active sites of enzymes. It is also possible to explore the protein surface for sequence-conserved patches, clefts and electrostatic potentials [6].

Fig 4: Protein structure predicted using SWISS-MODEL using its amino acid sequence (Uniprot IDs: A0A2J9QGH7)

Metabolic Pathways

A metabolite is a molecule that undergoes transformation in a biological system, either under the action of enzymatic catalysis or by spontaneous reaction. Metabolic pathways are the road maps defining the possible transformations of metabolites. They form a network, representable as a graph. Working out the individual reactions using classical methods such as following radioactive tracers is time consuming. It is much easier to sequence the genome, infer the amino acid sequences of the enzymes, look for sequences similar to enzymes of known function, and assemble the metabolic networks from the assignable enzymatic functions [7]. Many enzymes bind substrates in crevices not necessarily between domains. These binding crevices are known as active sites. Bioinformatics can predict the binding sites. The docking process involves two basic steps: prediction of the ligand conformation as well as its position and orientation within these sites and assessment of the binding affinity.

Importance of Biological Databases

The huge amounts of data generated are stored in online databases. A biological database is a computerized software or website designed to update, query, and retrieve information. (Bioinformatics Databases) Databases can be classified into Primary and secondary databases. Primary databases include experimental data like nucleotide sequence, protein sequence or macromolecular structure. E.g.: Protein Databank. Secondary databases include results derived from primary data. E.g.: Ensemble. Some of the commonly used databases are GenBank, SwissProt, Protein Data Bank. Each database focuses on a specific function. These function have been tabulated in Table [8].

Table 1: Commonly used domain specific databases

S No.	Database	Function
1	GenBank	Nucleotide sequence
2	Protein Data Bank	Three-dimensional macromolecular structures
3	Array Express Archive	Functional genomics data
4	UniProt	Sequence and functional information on proteins
5	EMBL	DNA database

6	Gene Expression Omnibus (GEO)	Functional genomics data repository
7	National Cancer Institute	Cancer Gene expression profiles
8	Human Protein Atlas (HPA)	Expression profiles of human protein coding genes
9	SwissProt	Protein knowledgebase
10	Swiss-model	Protein structure models
11	Reactome	Human biological pathways
12	Kyoto Encyclopedia of Genes and Genomes	Biological pathways and metabolism

Integration of Bioinformatics in Dentistry

Application of computer and information science in Dental Sciences can help improve dental practice, research, education, and management.

Microarray Analysis

Microarray technologies have become a common tool in the research for measuring the gene expression profiles. This method can be used to examine thousands of genes simultaneously [9]. The DNA microarray is a tool used to determine whether the DNA from a particular individual contains a mutation in genes. The chip consists of a small glass plate encased in plastic. A microarray chip contains thousands of short, synthetic, single-stranded nucleotide sequence complimentary to the normal gene or variants of that gene [10]. The most popular platforms for microarray analyses is Bioconductor, an open source and open development software project based on the R programming language. Arrays can be used to classify and identify molecular profiles of genes related to a number of cancers, including oral cancer. This type of genetic approach will aid in the understanding disease progression, thus improving diagnosis and treatment for patients [11]. It can aid in identifying and classifying genes associated with cancers, periodontal diseases and caries based on their expression profile. Bioinformatics analysis plays an important part of processing the information obtained from microarray analysis

Oral Cancer

Oral cancer is one of the most common type of cancer affecting large number of people. The diagnosis, therapies and prognoses of the disease is mainly due to the variation of severities, durations, sensitivity and resistance against drugs, cell differentiation and origin, and understanding of pathogenesis. Most of the cases, detection of cancer is in later stages decreasing the chances of cure or improvement [12]. With the advancement in Artificial Neural networks and image analysis of biopsy samples help in early detection. The brain consists of billions of interconnected neurons that have the ability to deal with complex and computationally demanding tasks. Artificial Neural Networks (ANN) was developed to artificially acquire the learning ability of the biological neurons system. ANN has learning ability and computational power and can be used in predictive or diagnostic models [13] The increasing evidence that the interaction between genes and proteins play an important role in investigation of cancer molecular mechanisms, hence it is necessary to study cancer at molecular level. With the healthcare industry moving towards personalized medicines, role of bioinformatics in understanding the mechanism and effects of treatment is crucial. Information about a patient's genetic, enzymatic and metabolic profile is used to develop a customised medical care to that individual's needs and assess a patient's risk factors to develop personalised medicine [14].

Insight on viral and fungal Infections

Oral disorders like periodontal diseases, caries are caused by various microorganisms. Many of these oral pathogens have been sequenced, and their genomic information has been stored in the databases. This data can provide insight into underlying mechanisms, the disease processes, molecular identification, genotypic, antimicrobial resistance analysis, detection of unknown disease-associated pathogens and effects on the host. They can also assist in for therapeutic development [15]. Next-generation sequencing (NGS) technology supported by bioinformatics can be used to determine the phylogenetic, and patho-genomics analyses and help in the identification of the causative agent. PathogenFinder is an important bioinformatics tool to test the pathogenicity of a newly discovered bacterial pathogen. It is a webserver used for the prediction of bacterial pathogenicity utilizing proteomic, genomic, or raw reads. The presence of antibiotic resistance loci in newly isolated bacterial pathogens can then be investigated using specialized search tools and services. Antibiotic Resistance Gene Search can be used for rapid, accurate detection and understanding of resistance factors and mechanisms. Several Bioinformatics tools are available for analysing data for combating and control of infectious. Incorporation of bioinformatics tools and applications in clinical microbiology and infectious diseases can be used to develop simple yet robust user-friendly bioinformatics pipelines.

Computer Aided Drug Design

The process of drug development is known to be very challenging, expensive and time consuming. It has been accelerated due to development of computational tools and methods. Over the last few years, computer aided drug design (CADD) also known as in silico screening has become a powerful technique. In-Silico screening of chemical moieties can help in short-listing the potential drug molecules. In silico drug designing include: Homology modelling ab-inition used for protein structure prediction, molecular docking to determine the drug receptor interaction, multi-target searching and design to search drugs against multiple targets, pharmacophore development to determine the three-dimensional arrangement of chemical functional groups for increased biological function, and quantitative structure activity relationship (QSAR) to determine the relationship between chemical structure and biological activity and ADMET properties to study the absorption, distribution, metabolism elimination and toxicology profile of the drug [16].

Patient care databases

Patient care databases are online storage including data related to the patient's diagnosis, procedures, drug prescription etc. Information on rare cases can be stored and can be easily accessed. Machine learning algorithms applied on the obtained data that can be used to give appropriate and effective treatment at minimal cost. A personalised well-designed hospital database can be used to collect up-to-date information from the visiting patient. This makes the relevant information available with a click-of-a-button. If a patient needs the services of healthcare providers in different

hospitals, exchange of information is possible. It is an important tool to monitor and improve healthcare services. It can also assist with documentation and billing. Reduction of paperwork and clerical staff can reduce the medical facility running costs.

Image Analysis

Image processing includes modifying, enhancing and detecting any particular image with high accuracy. Medical imaging research is increasingly dependent on computer-aided diagnosis (cad) where the details that can be missed out during visual inspection can be effectively avoided. Computer-aided diagnosis can be used for detection of bone fracture images obtained from different variation like X-ray, CT, MRI. It uses automated segmentation of medical images and various images algorithms. It can highlight changes from previous images or quantifying the size, shape, and textures. Computer-aided diagnosis is getting popular among medical practitioners and researchers. It provides very accurate, time & efforts saving, and less expensive diagnosis which can be used in the medical imaging profession. This research is used to interpret and to diagnose ailment from medical images with less aid from experts [17].

Conclusion

Research in dental sciences can be transformed by the evolution of high-throughput techniques and generation of large amount of data. The detailed study of the molecular sequences and structures can aid in understanding the genotypes and their clinical phenotypes. Informatics accelerates the process from research level to incorporation in the clinical applications. In the future, we anticipate that dental informatics the incorporation of clinical data into the analysis of genomic information will increase our understanding of the mechanisms underlying the biological challenges in dentistry. In this approach, we can eventually change the current practice of dentistry, including diagnostics, therapeutics, and prognostics of common oral diseases and disorders. This new approach to dental medicine will be both molecularly informed and informatically empowered.

References

1. Chowdhury B, Garai G. A review on multiple sequence alignment from the perspective of genetic algorithm. Genomics [Internet]. 2017; 109(5-6):419-31. Available from: https://doi.org/10.1016/j.ygeno.2017.06.007
2. Ren Z, Chen L, Li J, Li Y. Inhibition of Streptococcus mutans polysaccharide synthesis by molecules targeting glycosyltransferase activity [Internet]. Vol. 8, Journal of Oral Microbiology. Co-Action Publishing; 2016. [cited 2020 Sep 6]. Available from: /pmc/articles/PMC4841093/?report=abstract
3. Burr T. Phylogenetic Trees in Bioinformatics. Curr Bioinform. 2010; 5(1):40-52.
4. Edwards YJK, Cottage A. Bioinformatics Methods to Predict Protein Structure and Function A Practical Approach, 2003, 23.
5. Petrey D, Honig B. Protein Structure Prediction: Inroads to Biology. 2005; 20:811-9.
6. Manoj Bhasin GPSR. Protein Structure Prediction - an overview. In: Applied Mycology and Biotechnology [Internet]. 2006 [cited 2020 Sep 6]. Available from: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/protein-structure-prediction
7. Lesk A. Introduction to Bioinformatics, 2014.
8. Aryal S. Bioinformatics Databases | List of High Impact Articles | PPts | Journals | Videos [Internet]. [cited 2020 Sep 4]. Available from: https://www.longdom.org/scholarly/bioinformatics-databases-journals-articles-ppts-list-2876.html
9. Zhang Y, Szustakowski J, Schinke M. Bioinformatics analysis of microarray data. Vol. 573, Methods in molecular biology (Clifton, N.J.), 2009, 259-84.
10. DNA Microarray Technology Fact Sheet [Internet]. [cited 2020 Sep 5]. Available from: https://www.genome.gov/about-genomics/fact-sheets/DNA-Microarray-Technology
11. Kuo WP, Whipple ME, Jenssen TK, Todd R, Epstein JB, Ohno-Machado L *et al*. Microarrays and clinical dentistry. J Am Dent Assoc [Internet]. 2003. 2020; 134(4):456-62. Available from: https://pubmed.ncbi.nlm.nih.gov/12733779/
12. Wu D, Rice CM, Wang X. Cancer bioinformatics: A new approach to systems clinical medicine [Internet]. Vol. 13, BMC Bioinformatics. BioMed Central; 2012 [cited 2020 Sep 5]. p. 71. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3424139/
13. Jiang Y, Yang C, Na J, Li G, Li Y, Zhong J. A brief review of neural networks based learning and control and their applications for robots. Vol. 2017, Complexity. Hindawi Limited, 2017.
14. Verma M. Personalized medicine and cancer [Internet]. Vol. 2, Journal of Personalized Medicine. MDPI AG; 2012 [cited 2020 Sep 5]. p. 1–14. Available from: /pmc/articles/PMC4251363/?report=abstract
15. Informatics B, Avenue L. Over view of Bio informatics and its Application to Oral Genomics, 2003.
16. Bisht N, Singh BK. Role of Computer Aided Drug Design in Drug Development and Drug Discovery. Int J Pharm Sci Res. 2018; 9(4):1405-15.
17. Kajla V, Gupta A, Khatak A. Analysis of x-ray images with image processing techniques: A review. In: 2018 4th International Conference on Computing Communication and Automation, ICCCA 2018. Institute of Electrical and Electronics Engineers Inc, 2018.