A Survey on Bio Informatics Strategies and Applications for Disease Diagnosis and Prediction

Edward Daniel Christopher.B[1] ,S.P.Victor[2]

[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelveli,
Email: edwarddanielchristopher@rediffmail.com
[2]Associate Professor, St.Xaviers College, Tirunelveli

---***---

Abstract— Bioinformatics is the application of information technology to the study of living things, usually at the molecular level. Bioinformatics involves the use of computers to collect, organize and use biological information to answer questions in fields like evolutionary biology. A fascinating application of bioinformatics or the use of programming and software development to build enormous datasets of biological information for research purposes and linking big pharmacy companies with software companies, bioinformatics is growing in demand and offers good job prospects for computer science researchers and graduates interested in biology, medical technology, pharmaceuticals and computer information science. The classic data of bioinformatics include DNA sequences of genes or full genomes; amino acid sequences of proteins; and three-dimensional structures of proteins, nucleic acids and protein–nucleic acid complexes. This paper also presents a survey on bio informatics strategies used for disease diagnosis and prediction.

Keywords—**Bio informatics, Disease,** *Diagnosis*, **Prediction, Structure**

1. INTRODUCTION

Bioinformatics is a science field that is similar to but distinct from biological computation, while it is often considered synonymous to computational biology. Biological computation uses bioengineering and biology to build biological computers, whereas bioinformatics uses computation to better understand biology. Bioinformatics and computational biology involve the analysis of biological data, particularly DNA, RNA, and protein sequences. The field of bioinformatics experienced explosive growth starting in the mid-1990s, driven largely by the Human Genome Project and by rapid advances in DNA sequencing technology.
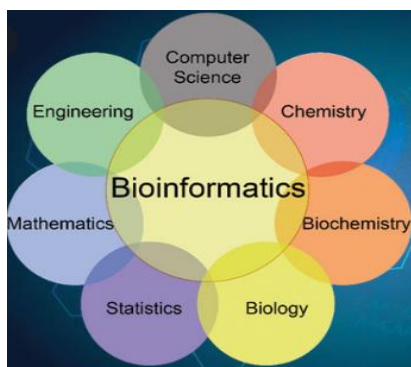

Fig-1: Bio Informatics Architecture

2. BIO INFORMATICS PATHWAYS

The figure 2 shows the web content mining process and the information retrieved in the structured format.
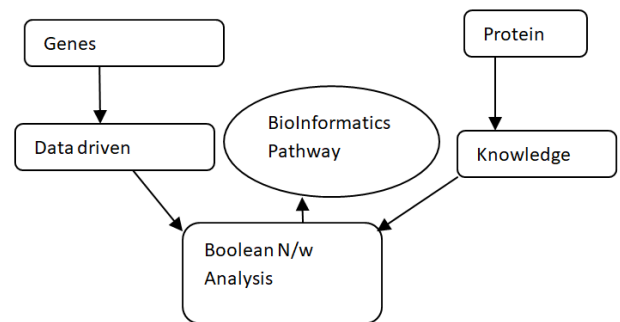

Fig-2: Progress of Web Content Mining

3. APPROACHES OF BIO INFORMATICS

The first approach to predicting the AMR is to analyze WGS (e.g., Illumina sequencing) data by identifying the presence of known ARG or gene variants. Some genes are endogenous to a given species/strain, while others reside on mobile genetic elements (plasmids) that can be shared between species.

The second approach to the AMR analysis and prediction is to study changes in gene expression of the isolate upon drug treatment. Suzuki and colleagues induced E. coli evolution in the lab under pressure of 11 AB with different concentrations to obtain strains resistant to these drugs with various degrees, measured in terms of minimum inhibitory concentrations (MIC) [4]. Subsequent gene expression analysis (based on microarray) allowed them to study dynamic compensatory mechanisms in the pathogen. They found 8 genes (acrB, ompF, cyoC, pps, tsx, oppA, folA, and pntB) whose expression levels allowed for a linear regression model achieving correlation with experimentally determined MIC with $R2$ ranging from 0.54 to 0.75 per drug. Darnell and colleagues studied transcriptional response of E. faecalis (strain JH2-2) to teixobactin [15].

The third approach to elucidating the AMR mechanisms is gene agnostic and based on global genomic comparison of multiple strains with various susceptibility to different drugs. Pamphlet compared 110 reference and 12 metagenomically detected strains of E. coli, both commensally and outbreak-related, with

respect to their genomic content and gene expression profiles [7].

The last approach, orthogonal to those described above, takes advantage of metabolic profiling. Yang and colleagues treated E. coli with 3 AB and analyzed the changes in the metabolite profiles [14]. By combining the known metabolic pathways of the organism with experimentally measured boundaries of metabolite concentrations upon drug treatment, they were able to conduct flux balance analysis (FBA).

4. LITERATURE SURVEY

The following table-1 shows the literature survey in a sketch format,

Table-1: Literature survey data

Research Articles	Disease Analysis	Disease Prediction
Typewriter by Gordon et al., 2014 [3]	*Staphylococcus aurous*	Binary SvR classification; CA
By Suzuki et al., 2014 [4]	*Escherichia coli*	Regression, MIC; EA
PhyResSE by Feuerriegel et al., 2015 [5]	*Mycobacterium tuberculosis*	Binary SvR classification; CA
Mykrobe by Bradley et al., 2015 [6]	*S. aureus, M. tuberculosis*	Binary SvR classification; CA
PanPhlAn by Scholz et al., 2016 [7]	*E. coli*	Binary SvR classification; PCU
PointFinder by Zankari et al., 2017 [8]	*E. coli, Campylobacter jejuni, Salmonella enterica*	Binary SvR classification; CA
By Zampieri et al., 2017 [9]	*E. coli*	NA; EA
By Mahe and Tournoud, 2018 [10]	*S. aureus, M. tuberculosis*	Logistic regression; PCU
By Kavvas et al., 2018 [11]	*M. tuberculosis*	NA; PCU
CRyPTIC consortium, 2018 [12]	*M. tuberculosis*	Binary SvR classification; CA
By Drouin et al., 2019 [13]	*E. coli*	Multi-class classification; PCU
By Yang et al., 2019 [14]	*Enterococcus faecalis*	NN regression; EA
By Darnell et al., 2019 [15]	*E. coli*	NA; EA

5. ISSUES IN WEB CONTENT MINING

- ✓ Protein structure prediction
- ✓ Homology searches
- ✓ Multiple alignment and phylogeny construction
- ✓ Genomic sequence analysis and gene-finding

6. IMPLEMENTATION IN DISEASE DIAGNOSIS

- ❖ Study protein-protein and protein-nucleic acid recognition and assembly
- ❖ Investigate integral functional units (dynamic form and function of large macromolecular and supramolecular complexes)
- ❖ Bridge the gap between computationally feasible and functionally relevant time scales
- ❖ Improve multiresolution structure prediction
- ❖ Combine classical molecular dynamics simulations with quantum chemical forces
- ❖ Sample larger sets of dynamical events and chemical species
- ❖ Realize interactive modeling
- ❖ Foster the development of bimolecular modeling and bioinformatics
- ❖ Train computational biologists in teraflop technologies, numerical algorithms, and physical concepts
- ❖ Bring experimental and computational groups in molecular biomedicine closer together.

7. TOOLS PERFORMANCE

The following table illustrates Bioinformatics tools with its implementation scope for different types of disease analysis.

Table-2: Tools Performance Bioinformatics

EnuMol	Enumerating Molecular Structures
MIMOX	A web tool for phage display analysis
HSEpred	predict Half-Sphere Exposure from protein sequences
PRORATE	prediction of protein folding rates
SLPFA	Sub cellular Location Prediction with Frequency and Alignment
INTREPED	an interactive DNA repair prediction server
IPPI	Inferring Protein-Protein Interactions for YEAST
PURE	A Pub med article recommendation system based on content-based filtering

8. RECENT TRENDS

- Bioinformatics: Development and prospects

- Molecular Evolution usage frequency bias

- Proteomics: A Revolutionary milestone in science
  Genomic and molecular databases

- Computational genomics, Molecular markers and plant systematic

- Basic concepts, Integration and biological importance of databases

- Introduction to Databases and sequence analysis Database Management System

- Bioinformatics as a tool for identifying and understanding human diseases

- Bioinformatics in horticulture
  Applications of Bioinformatics in vegetable breeding

- Vegetable gene list for hybridization

9. APPLICATIONS

- Varietal Information System

- Plant Genetic Resources Data Base

- Biometrical Analysis

- Storage and Retrieval of Data

- Studies on Plant Modelling

- Pedigree Analysis

- Preparation of Reports

- Updating of Information

- Diagrammatic Representation

- Planning of Breeding Program

10. CONCLUSION

The detailed research survey in the field of disease diagnosis and prediction based on bio informatics shows that the higher level of impact in the field of medical genomes with the cope up towards latest trends and systematic pathways for the cure of several disease strategies. The approaches for bioinformatics dealt with the various level of implications towards the selection strategies for the procurement and prevention of specific dangerous mutative viral components. The tools performance and applications of bioinformatics provides the several directions for the development of different methodologies to implement the bioinformatics domain in several real-time domains. In future this research will lead the direction of bioinformatics application towards the control of current pandemic situation of Covid-19.

REFERENCES

1. *Antibiotic Resistance Threats in the United States, 2019.* U.S. Department of Health and Human Services, CDC; Atlanta, GA, USA: 2019. [(Accessed on 17 February 2020)].

2. Hughes D., Andersson D.I. Environmental and genetic modulation of the phenotypic expression of antibiotic resistance. *FEMS Microbiol. Rev.* 2017; 41:374–391. doi: 10.1093/femsre/fux004

3. Gordon N.C., Price J.R., Cole K., Everitt R., Morgan M., Finney J., Kearns A.M., Pichon B., Young B., Wilson D.J., et al. Prediction of Staphylococcus aureus Antimicrobial Resistance by Whole-Genome Sequencing. *J. Clin. Microbiol.* 2014;52:1182–1191. doi: 10.1128/JCM.03117-13

4. Suzuki S., Horinouchi T., Furusawa C. Prediction of antibiotic resistance by gene expression profiles. *Nat. Commun.* 2014;5:1–12. doi: 10.1038/ncomms6792.

5. Feuerriegel S., Schleusener V., Beckert P., Kohl T.A., Miotto P., Cirillo D.M., Cabibbe A.M., Niemann S., Fellenberg K. PhyResSE: A Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. *J. Clin. Microbiol.* 2015;53:1908–1914. doi: 10.1128/JCM.00025-15.

6. Bradley P., Gordon N.C., Walker T.M., Dunn L., Heys S., Huang B., Earle S., Pankhurst L.J., Anson L., de Cesare M., et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat. Commun.* 2015;6:1–15. doi: 10.1038/ncomms10063. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

7. Scholz M., Ward D.V., Pasolli E., Tolio T., Zolfo M., Asnicar F., Truong D.T., Tett A., Morrow A.L., Segata N. Strain-level microbial epidemiology and population genomics from shotgun

metagenomics. *Nat. Methods.* 2016;13:435–438. doi: 10.1038/nmeth.3802

8. Zankari E., Allesøe R., Joensen K.G., Cavaco L.M., Lund O., Aarestrup F.M. PointFinder: A novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J. Antimicrob. Chemother.* 2017;72:2764–2768. doi: 10.1093/jac/dkx217.

9. Zampieri M., Enke T., Chubukov V., Ricci V., Piddock L., Sauer U. Metabolic constraints on the evolution of antibiotic resistance. *Mol. Syst. Biol.* 2017;13 doi: 10.15252/msb.20167028

10. Mahé P., Tournoud M. Predicting bacterial resistance from whole-genome sequences using k-mers and stability selection. *BMC Bioinformatics.* 2018;19:383. doi: 10.1186/s12859-018-2403-z.

11. Kavvas E.S., Catoiu E., Mih N., Yurkovich J.T., Seif Y., Dillon N., Heckmann D., Anand A., Yang L., Nizet V., et al. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. *Nat. Commun.* 2018;9:1–9. doi: 10.1038/s41467-018-06634-y

12. The CRyPTIC Consortium and the 100,000 Genomes Project Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing | NEJM. [(accessed on 27 November 2019)];

13. Drouin A., Letarte G., Raymond F., Marchand M., Corbeil J., Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. *Sci. Rep.* 2019;9:4071. doi: 10.1038/s41598-019-40561-2.

14. Yang J.H., Wright S.N., Hamblin M., McCloskey D., Alcantar M.A., Schrübbers L., Lopatkin A.J., Satish S., Nili A., Palsson B.O., et al. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell.* 2019;177:1649–1661.e9. doi: 10.1016/j.cell.2019.04.016.

15. Darnell R.L., Knottenbelt M.K., Todd Rose F.O., Monk I.R., Stinear T.P., Cook G.M. Genomewide Profiling of the Enterococcus faecalis Transcriptional Response to Teixobactin Reveals CroRS as an Essential Regulator of Antimicrobial Tolerance. *mSphere.* 2019;4:e00228-19. Doi: 10.1128/mSphere.00228-19.