Integrating Genomics and Proteomics - A Bioinformatics based pedagogical approach to elucidate Gene-Protein Relationship.

Sharanappa Achappa[1] & S.V. Desai[2]

[1 & 2]Department of Biotechnology, KLE Technological University, Hubballi- 580 031, India.

[1]sharanappaa@kletech.ac.in, [2]desaisv@kletech.ac.in

Abstract: The central dogma of life beginning with DNA leading to RNA and subsequently to protein is the unifying principle and corner stone to the understanding of biology. With the advent of sophisticated and high throughput sequencing, enormous data is being generated. This has made the bio-informatics tools indispensable to decipher the data into meaningful conclusions. The objective of the present study was to comprehend the understanding of Genomics and Proteomics with the use of Bioinformatics tools. The course project activity was performed for Genomics & Proteomics course of VII semester undergraduate students of Engineering in Biotechnology. The students were grouped into team of four and assigned with one pathogenic bacterium. The activity involved identifying a pathogenesis-related protein, retrieving the DNA sequences of the protein, finding conserved domains, building phylogenetic relationship amongst the homologous organisms, genome annotation, predicting secondary and tertiary structure of proteins and deducing its stability. In the process, the students were engaged in active learning by the use of various Bioinformatics tools and repositories. The students were thus able to predict the secondary and tertiary structure of protein beginning with DNA sequences. The activity was instrumental in addressing the Graduate attributes. Rubrics-based assessment was performed to measure the attainment level on the scale of 10. The attainment ranged from 7.28 to 9.25 for different performance indicator (PI) of the graduate attributes. A formal anonymous feedback from the students revealed that the activity helped in appreciating the intricate relationship between Genomics and Proteomics by integrating the related concepts.

Keywords: Genomics, Proteomics, Genes, Proteins.

_____

Corresponding Author

Shivalingasarj V. Desai, Associate Professor, Department of Biotechnology, KLE Technological University, Hubballi- 580 031. India.

E-mail: desaisv@kletech.ac.in

1. Introduction

"The human genome is a script, waiting for the amino actors, the protein players to strut and fret their hour."

-Johnny Rich

Genomics is the study of all the genes in a haploid cell of an organism and Proteomics is the study of the entire set of proteins expressed by a cell (Horgan RP, Kenny LC, 2011). The scope of Genomics and Proteomics encompasses the stages of central dogma of life. Genomics lead to proteomics via transcriptomics as a sequential step (Berrar D., Granzow M., Dubitzky W., 2007). The two complement the study of each other and holds various applications in terms of understanding pathogenesis, disease prognosis and drug discovery (Rastogi *et al*, 2008). Bioinformatics has enabled the use of several disease associated molecular markers with huge number of gene sequences to be successfully annotated (Ostrowski and Wyrwicz, 2009., Bednarski *et al*, 2005).

Genomics as a course requires Microbiology, Cell Biology, Genetics, Molecular Biology and Bioinformatics as prerequisites for a good understanding. Genomics & Proteomics as a course relies heavily upon the repository of gene and amino acid sequences as the basic raw materials. It involves the annotation (deciphering the gene) and identifying the resulting protein it codes for. It is highly challenging and demanding to study the concepts and tangibly demonstrate the relationship between Genomics & Proteomics due to its dependence on high-throughput data and sophisticated techniques (Claudia Manzoni, *et al* 2018). In this context, an in-silico-based approach using Bioinformatics tools to correlate Genomics & Proteomics has been made in the present study.

Bioinformatics is an interdisciplinary field involving biology, computer science, mathematics & statistics concerned with the development & application of computer hardware& software to acquire, storage, analysis & visualization of biological information to draw a meaningful conclusion (Yun-Can Ai, 2002). This interdisciplinary nature makes bioinformatics an ideal framework to experience students the interplay between different scientific areas, while touching on societal aspects mainly on health and environment.

The objective of the pedagogical activity-Course project was to integrate Genomics & Proteomics to comprehend the relationship between Genomics & Proteomics using Bioinformatics open-access databases and software tools. Course project is an independent study project carried out during an academic year (course, semester) by students of higher technical schools and technicums und

er the supevision of their professors and instructors. The purpose of the course project is to deepen and consolidate the theoretical and practical knowledge gained by the student in the process of studying. Implementation of the course project allows you to determine and assess the degree of understanding of theoretical material and the ability to apply theory in solving specific practical situations (Agung W. Setiawan, 2020; Pohorielova, 2020)

The activity was implemented for VII Semester undergraduate students of engineering in Biotechnology. An open-ended problem on different diseases was floated to the groups (group comprising up to 4 students). Cascade of laboratory experiments were performed to understand the molecular aspect of protein involved in pathogenesis. The activity was instrumental in addressing graduate attributes namely: problem identification & solving, identification of proper tools, data analysis and interpretation, communication skill and ability to work in teams were the major outcomes of this course.

2. Methods

A. Course Description:

Genomics is the study of all the genes in a haploid cell of an organism and Proteomics is the study of the entire set of proteins expressed by a cell. The scope of Genomics and Proteomics encompasses the stages of central dogma of life. Genomics lead to proteomics via transcriptomics as a sequential step.

This course mainly focusses on genomes of prokaryotes and eukaryotes, genome sequencing techniques, molecular markers and application of genome data in medicine. Proteomics focuses on structural functional relationship of proteins, different analytical techniques in purification and characterization of proteins and finally bioinformatics tools for genomics and proteomics study.

The main aim of this course is to train students extensively to develop hands-on skills in the use of bioinformatics tools and databases for analysis of genomics and proteomics data helpful in treating diseases. Expected outcomes include:

- Describe the scope & nature of genomes and its approaches.
- Explain different tools and methods of genome analysis.
- Discuss the advancements and specific applications of genomics studies.
- Explain the scope & nature of proteomics and its approaches.
- Discuss the different approaches used for proteome separation and purification.
- Explain the various applications of proteomics.
- Explain the tools and methods of bioinformatics used in genomics and proteomics.

- Demonstrate the use of modern engineering tools, teamwork and effective written communication.

B. Design of course project:

Genomic & Proteomics course project was designed considering the prior knowledge of bioinformatics course. In course projects, groups of students will work on identified pathogen to understand gene proteins relationship.

The activity was group-based, comprising four members in each group with a total of 12 groups. The activity was implemented for VII semester students for the academic year 2016-17 and 2017-18. Each group was informed to select 3 bacterial pathogen through literature survey. Groups are assigned with an activity to collect information related to the identified pathogen with respect to molecular aspect of pathogenesis, available diagnosis method, traetment and proteins involved in pathogenesis.Based on the collected information, each group was informed to select one pathogen and assigned with a title of a disease.

The titles chosen for the activity comprised commonly encountered bacterial pathogens (Mycobacterium, Cholera, Clostridium etc.) by humans. The task for the exercise was to identify few genes from the sequence repository, perform multiple sequence alignment, develop a phylogenetic tree, predict the secondary structure, search for the possible protein patterns and determine the various physico-chemical properties of the protein.

C. Execution of course project:

The present study was conducted for VII semester undergraduate students of engineering in Biotechnology for Genomics & Proteomics course for the academic year 2016-17 and 2017-18. An open-ended problem on different diseases (restricted to bacteria) was floated to the groups. Cascade of laboratory experiments were performed to understand the gene -protein relationship and molecular aspect of protein involved in pathogenesis.

The activity was influential in addressing various graduate attributes. The course project was executed in two modules as represented in fig.1.
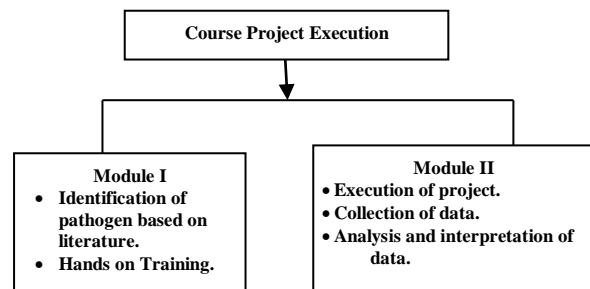
```
┌─────────────────────────────┐
│  Course Project Execution   │
└─────────────────────────────┘
```

Module I	Module II
• Identification of pathogen based on literature. • Hands on Training.	• Execution of project. • Collection of data. • Analysis and interpretation of data.

Fig.1 Modules of Course Project.

In module-I, individual students groups identified three different pathogens (restricted to bacterial pathogen) based on the literature survey. Training on different bioinformatics software tools was given. Practical session involved a step-by-step display of procedure of different bioinformatics software tools with an example was performed. The session was kept interactive so as to enable the students to clarify their doubts, if any.

In module-II, project execution was done by performing series of experiment for the identified protein of the identified pathogen, which involved the student's groups engaging themselves in executing the exercise and arriving at proteins structure and their characteristics as outputs. The students were entertained to use online repository of gene sequences and amino acids and various open access tools for the activity.

Analysis and interpretation of the results was completed for the protein to understand its relationship for the pathogenesis of the microorganism.

D. Assessment of the Course Project:

Rubrics-based assessment was practiced for the course project activity as shown in table 1.

The assessment parameters considered were Conduct of experiment (PI Code: 4.1.3), Data Collection & representation (PI Code: 4.3.1), Analysis of data (PI Code: 4.3.2), Interpretation of data (PI Code: 4.1.4), Identification of problem (PI Code: 4.1.1), Identification of tools (PI Code:4.1.2), Selection of procedure (PI Code:4.1.2), Selection of modern tools (PI Code:5.2.2), Team Work (PI Code:9.3.1) Write up (PI Code: 10.1.2) and Presentation (PI Code: 10.2.2).

Table 1. Rubrics for course project

Rubrics parameter	Inadequate (up to 25%)	Average (up to 50%)	Admirable /Good (up to 75%)	Outstanding (up to 100%)
Conduct of experiment	Inadequate quality of experimental work.	Experiments were conducted.	Adequate conduct of experiment	Proper conduct of experiment
Data Collection & representation	Raw data, including units, are not recorded in a way that is appropriate and clear.	Raw data, including units, are recorded although not as clearly or appropriately as they might be.	Raw data, including units, are recorded although not as clearly or appropriately as they might be.	Raw data, including units, are recorded in a way that is appropriate and clear.
Analysis of data	Data are presented in ways (charts, tables, graphs) that are very unclear.	Data are presented in ways (charts, tables, graphs) that can be understood and interpreted, although not as clearly as they might be.	Data are presented in ways (charts, tables, graphs) that can be understood and interpreted, although not as clearly as they might be.	Data are presented in ways (charts, tables, graphs) that best facilitate understanding and interpretation.
Interpretation of data	The results are not interpreted in a logical way.	The results are interpreted but not clear.	The results are interpreted, but not as fully as they might be.	The results are fully interpreted.
Write up/Communication	Report contains many distracting mistakes, making it generally difficult to follow and poorly organized.	Report is generally clear, but distracting errors and flow make it difficult to follow at times and organization of report is weak.	Report is logical and easy to read, and may contain a few errors causing minimal reader distraction and organized strongly.	Report is virtually error-free, and contains few if any reader distractions and clearly organized with excellent transitions.
Identification of Problem	Identified no parameters with no reference to the literature	Parameters identified but not clearly related to the experiments with very little reference to the literature.	Parameter identified but not clearly correlates to the experimental setup with reference to literature.	Parameters identified according to the experimental setup with clear reference to literature.
Selection of Bioinformatics tools	Demonstrated little or no ability to conduct experiments. Did not collect Meaningful data.	Demonstrated some ability to conduct experiments. Collected some Meaningful data.	Demonstrated adequate ability to conduct Experiments. Collected most of the needed data.	Demonstrated superior ability to conduct experiments. Collected all the appropriate data.
Team Work	Contributed little to the group effort during the project	Finished individual task but did not assist group/partner during the project.	Assisted group/partner in the finished project.	All team members contributed equally for completion of project

Modern tools	Fails to apply any modern tools to solve engineering problems	Attempts to apply modern tools but has difficulties to solve engineering problems.	Shows ability to apply fundamental procedures in using modern tools when solving engineering problems	Shows ability to apply the most appropriate and effective modern tools to solve engineering problems

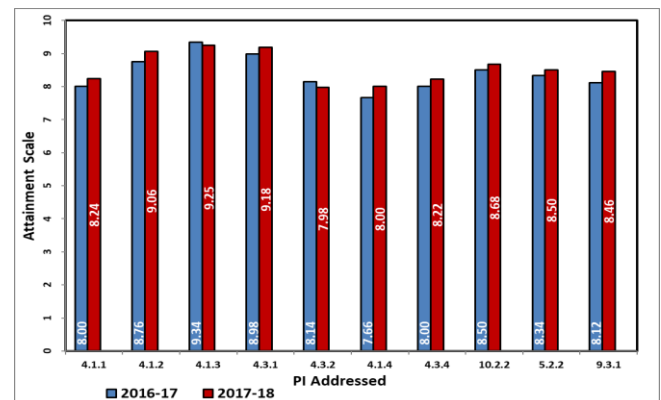| Sl. No.	Graduate Attribute	Competence	Performance Indicator PI Code)
8	Data Collection & representation	Demonstrate an ability to critically analyze data to reach a valid conclusion	Use appropriate procedures, tools and techniques to collect and analyze data (4.3.1)
9	Analysis of data		Critically analyze data for trends and correlations, stating possible errors and limitations (4.3.2)
10	Conclusion		Synthesize information and knowledge about the problem from the raw data to reach appropriate conclusions (4.3.4)

3 Results

Implementation of course project in Genomics & Proteomics aimed at establishment of the basic skill-sets of individual students. Performance of the students during the course project indicated a thorough understanding of the concept learnt through experiential learning. Also, the continuous evaluation process shows the involvement of the students in the group activity, and also technical contribution towards the project. The activity was instrumental in attaining the various Graduate attributes, Competencies and Performance indicators as shown in Table 2.

Table 2. Attainment of Graduate Attributes, Competence and Performance Indicator

Sl. No.	Graduate Attribute	Competence	Performance Indicator PI Code)
1	Modern tool usage	Demonstrate an ability to select and apply discipline specific tools and resources	Demonstrate proficiency in biochemical, microbial, molecular and tools (5.2.2)
2	Individual and team work	Demonstrate success in a team-based project	Present results as a team, with integration of contributions from all individuals (9.3.2)
3	Communication	Demonstrate an ability to comprehend technical literature and document.	Produce clear, well-constructed, and well-supported written engineering documents (10.2.2)
4	Identification of Problem/ parameter	Demonstrate their ability to conduct investigations of technical issues consistent with their level of knowledge and understanding	Define a problem to carry-out investigation with its scope and importance. (4.1.1)
5	Selection of appropriate procedure		Identify and apply relevant experimental procedure /bioinformatics tools /databases for a defined problem (4.1.2)
6	Conduct of experiment		Use appropriate analytical instruments /software tools to carry-out the experiments (4.1.3)
7	Interpretation of data		Correlate the experimental outcomes with underlying theoretical concepts and principles (4.1.4)

The student's attainment of Program Outcomes (PO) was evaluated by mapping the rubrics parameter with Performance Indicator (PI) as shown in the table 2. PI at different stages of course project was assessed based on the rubrics. Based on the mapping of rubric parameters with PI, % attainment of various PI was measured and represented in Fig.2.

Fig.2: PI attainment for the academic year 2016-17 & 2017-18

The attainment of different PI on the scale of 10 was evaluated for the academic year 2016-17 & 2017-18. The attainment measured showed highest for PI 4.1.3 & 4.3.1 of 9.25 and 9.00 respectively, which showed that the students are good at performing experiments and collection of data, while attainment was lowest for PI 4.3.2 and 4.1.1 of 7.98 & 8.24 respectively, which indicated the scope for enhancement in analysis and guiding them in identification of problem statement.

One of the case-study of *Clostridium tetani* as organism was considered to show the implementation of course project. The bacteria *C. tetani* was accession number NZ_JWIX01000088 and Elongation Factor Tu (EF-Tu) GTP-binding protein was selected as target protein with a sequence length of 382. The sequences were subjected to Blastp and found that the conserved domains existed in 10 different organisms. The phylogenetic tree developed showed 3 clusters, amongst which two had further

ramifications and one was stand-alone. The secondary structure predicted using SOPMA resulted in 125 sequences lying in Alpha helix and 124 sequences in random coil region, 91 sequences in extended strand and 42 in Beta turn. The physic-chemical characters of the sequence as determined by ProtParam showed the Molecular weight as 42 K.Da, theoretical pI of 5.1. The major amino acid contributors were Glutamine, Valine, Alanine, Aspargine, and Isoleucine,in that order. The ratio of positively charged amino-acid residues to negatively charged was 42:63. The instability index of the protein was found to be 32 which classified it as stable.

The screen-shots of some selected output of the using different bioinformatics tools is shown in Fig. No.3.
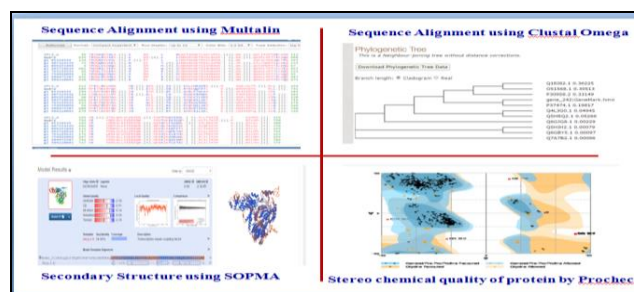

Fig 3. Screen-shot of results from Bioinformatics Tools

A formal anonymous feedback was taken from students through Type form tool. The questionnaire was framed to assess the learning curve of the students and scope for further improvement of the activity. The respondents were instructed to rate the questions on a scale of 10 with 0 denoting strongly disagree and 10 denoting strongly agree. The following table (Table No. 3) represents the questions asked as part feedback survey.

Table 4. Questions for Feedback Survey of the Activity

Sl. No.	Questionnaire
1	The course activity gave an exposure to the use of different Bioinformatics software tools related to Genomics and Proteomics
2	The activity was helpful in coherent understanding of Genomics and Proteomics.
3	The activity helped in understanding the gene-disease association.
4	The activity helped in understanding the structure-function relationship of proteins.
5	The activity gave a first-hand experience of determining secondary structure of pathogenesis related proteins.
6	The activity demanded team co-ordination for planning and execution.
7	The compilation of the report enhanced the technical writing skills.
8	Any suggestions/comments for improving the teaching-learning experience regarding the activity.

The results of the feedback survey indicated majority of the respondents agreed that the activity engaged them in active learning and helped in correlating the relationship between a typical gene and protein. Few opined that scope existed for further in-depth analysis of outputs obtained for deeper understanding of the concepts involved.

4. Conclusion:

Based on the experience of implementing course project in Genomics and Proteomics course, we conclude that the students successful in demonstrating their ability in identification of problem & selection of appropriate tools, data analysis and interpretation, communication skills and ability to work in teams. Attainment of PI showed scope for improvement in area like problem identification, analysis of data, and interpretation of data.

It was also inferred from the activity that the student participants could engage in active learning and able to comprehensively establish and appreciate the intricate relationship that exists between genomes and Proteomes. They had an opportunity to undergo experiential learning by using different bioinformatics tools and their intended applications which helped them in deciphering the raw Genomic sequences, translating it into protein structure and elucidating its physic-chemical characteristics and arriving at a logical conclusion.

.

References:

Yun-Can Ai, "Teaching bioinformatics: A student-centred and problem-based approach," *The China Papers' Issue 1, October 2002*, (2002).

Bednarski, A.E., Elgin., S.C.R., Pakrasi, H.B.(2005). An Inquiry into Protein Structure and Genetic Disease: Introducing Undergraduates to Bioinformatics in a Large Introductory Course. Cell Biology Education.4(3), 207-220

Ostrowski, J., Lucjan S Wyrwicz, L.S. (2009). Integrating genomics, proteomics and bioinformatics in translational studies of molecular medicine.Expert Review in Molecular Diagnostics. 9 (6), 623-630.

Rastogi,S.C., Rastogi, P., Namita Mendiratta. (2008). Bioinformatics Methods And Applications: Genomics Proteomics And Drug Discovery, 3rd Edition. PHI Learning Pvt. Ltd.New Delhi.

Berrar D., Granzow M., Dubitzky W. (2007) Introduction to Genomic and Proteomic Data Analysis. In: Dubitzky W., Granzow M., Berrar D. (eds) Fundamentals of Data Mining in Genomics and Proteomics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-47509-7_1

Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist 2011; 13:189–195

Claudia Manzoni, Demis A Kia, Jana Vandrovcova, John Hardy, Nicholas W Wood, Patrick A Lewis, Raffaele Ferrari, Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences, *Briefings in Bioinformatics*, Volume 19, Issue 2, March 2018, Pages 286–302, https://doi.org/10.1093/bib/bbw114.

W. Setiawan, "Implementation of Internet of Things in Biomedical Measurement and Instrumentation Course Project," *2020 IEEE Global Engineering Education Conference (EDUCON)*, Porto, Portugal, 2020, pp. 1657-1661, doi: 10.1109/EDUCON45650.2020.9125197.

Guidelines for the implementation of the course project on the discipline "Human resource management": for students of specialty 073 "Management" / comp. T. Pohorielova; National Technical University "Kharkiv Polytechnic Institute". – Kharkiv : Modelist, 2020. – 22 p.

Sharanappa Achappa, L. R. Patil, V. S. Hombalimath, Anil R. Shet (2020); DOI: 10.16920 /jeet /2020 /v33i0 /150154

Sharanappa Achappa, L. R. Patil, V. S. Hombalimath, Anil R. Shet (2018); DOI: 10.16920/jeet/2018/v0i0/120904.