Bioinformatics for undergraduate Biological Sciences students using active methodology: a case study

Everton Silva Mota[1], José Alves dos Santos Neto[2], Bruno Lassmar Bueno Valadares[3], Marcus Vinicius de Aragão Batista[4]

Abstract

With the advancement of Molecular Biology, several datasets have emerged and together with them, there was a need to systematize and analyze these data for the generation of useful information. To perform such a function arises a new area called Bioinformatics, that encompasses the Biological Sciences, Mathematics and Computer Science. Few Brazilian universities offer courses that deal with aspects related to Bioinformatics at the undergraduate level, thus, many biologists leave universities without having any knowledge about Bioinformatics. Because it is an area closely linked to Molecular Biology, Bioinformatics proves to be an important tool to aid in some disciplines teaching, such as Genetics. In this context, the present study aimed to introduce the discipline of Bioinformatics for undergraduate Biological Sciences students, using an active teaching methodology, aiming to evaluate the relevance and importance of this discipline for undergraduates, as well as the effectiveness of this teaching methodology. It was verified that the students themselves feel the need for a more in-depth approach to Bioinformatics in the academic curriculum, besides, it was observed that active methodologies assist in the teaching-learning process of these students, facilitating a meaningful learning.

Keywords: Biology teaching; Molecular genetics; Active methodology; Bioinformatics education.

Recebido em: 21-01-2025; Aceito em: 10-10-2025

https://doi.org/10.5335/ctpf2x24

ISSN: 2595-7376

---

[1] Licenciado em Ciências Biológicas, Mestre em Biotecnologia pela Universidade Federal de Sergipe. E-mail: evertonmotabio@gmail.com

[2] Licenciado em Ciências Biológicas, Mestrando em Biotecnologia na Universidade Federal de Sergipe. E-mail: joseeneto33@gmail.com

[3] Professor Associado do Departamento de Biologia da Universidade Federal de Sergipe. E-mail: brunovaladares@academico.ufs.br

[4] Professor Associado do Departamento de Biologia da Universidade Federal de Sergipe. E-mail: mbatista@academico.ufs.br

Introduction

Since DNA discovery, studies about this molecule have brought important data Farah (2007), with technology's advance were also an increasing amount of data, with the number of these data growing, then the need to organize and analyze them arises, effectively generating information. For solve this problem, it appears the Bioinformatics (Farah, 2007; GenBank, 2017; Genomes Online Database [GOLD], 2017).

In Brazil, Biological Sciences undergraduate courses hardly offer knowledge about Bioinformatics, nor do they use these tools to facilitate the teaching-learning process in undergraduate classes (Ribeiro et al., 2012). Therefore, the need to generate human resources capable of acting in this area is noticeable, since the increase in molecular data generated by scientific studies establishes a significant demand. However, many undergraduate students leave universities without even knowing what Bioinformatics is about. Some authors (Ribeiro et al.; Ribeiro, 2010) have showed this fact when they say that the insertion of Bioinformatics is very rare and still occurs generally in postgraduate courses, with only a few universities introducing Bioinformatics knowledge in undergraduate courses.

It is known that, in the academic and social environment, the technique is strongly associated with the theory and the opposite is also valid (Ribeiro & Jesus, 2015). Therefore, the professor might always seek to associate theories with the practices to help the learning process.

As Silva (2012) says, the professors should provide some characteristics to the student, which would be the acquisition of creative, participative and questioning awareness; present theoretical references for analysis and interpretation of reality; and to be able to link theory and practice.

The choice of methodologies that facilitate this process is of great value for the teacher's reflection, since these reflections help students to follow the teacher's goals (Silva, 2012).

The active teaching methodology is a viable alternative to promote such characteristics, and those that enable the student to build their knowledge, while becoming the center of the discussions related to the theme (Melo & Sant'Ana, 2013).

Bioinformatics brings the opportunity of a differentiated pedagogical approach through problem solving (Wightman & Hark, 2012). Use tools to stimulate the student in the search for solutions is an interesting method to develop skills (Farias et al., 2012).

It is necessary to mention that Bioinformatics is a growing area and very important for biologists. However, there is a gap in the Bioinformatics teaching in undergraduate courses. Therefore, this study aimed to evaluate how the introduction of the Bioinformatics interferes in the quality of the Biological Sciences students training.

Several authors have already used Bioinformatics to improve the training of undergraduates. In Nunes et al. (2015), students used programming language to understand processes related to Molecular Biology such as: Transcription, Translation, Complementarity of the DNA strand, and others. The study brought a differentiated approach to Molecular Biology teaching.

For Badotti et al. (2014), Bioinformatics has shown to be an important tool for proteomics teaching, so that even with the aid of tutorials, students are more active in the process and in fact think in an analytical way.

According to Mertz e Streu (2015), the active methodology and Bioinformatics can be used to improve students writing and reasoning skills. In their study, the authors propose to students to write a project based on previously provided biochemistry courses, but in these courses the contents are mediated by Bioinformatics tools under active methodology, facilitating the understanding of the processes approached. Thus, the students can then think of ways to use the studied tools for the creation of other research projects.

Other studies also used Bioinformatics to assist in the teaching process of several disciplines, such as: Mathematics (Chapman, 2010; Wightman & Hark, 2012), Biochemistry (Badotti et al., 2014; Inlow et al., 2007; Junior, 2010; Mertz & Streu, 2015), Programming, Chemistry, Physics (Furge et al., 2009), Physiology (Nunes et al., 2015; Zhang, 2009), among others.

Therefore, under the analysis of the current situation of Bioinformatics, the need to introduce Bioinformatics in the academic curriculum of the undergraduate Biological Sciences courses is questioned. It is still valid to inquire that the use of Bioinformatics tools in an active way contributes significantly with the learning of undergraduate students and consequently in the training of these future professionals.

Thus, this study aimed to evaluate the introduction of Bioinformatics to Biological Sciences undergraduate students through an active pedagogical methodology, where the students' sought solutions for biological problems.

Material and Methods

To achieve those goals, this study was carried out at the Federal University of Sergipe, where the sample were undergraduate students who accepted to participate in this study. In total, 27 Biological Sciences students participated in the study. This study was approved by the Research Ethics Committee Involving Human Beings of the Federal University of Sergipe (protocol number: 80239717.3.0000.5546). All the participants, individually, signed the free informed consent form and filled the survey.

Lessons were planned to introduce the main aspects of Bioinformatics as well as to present the tools to be used, as shown in Table 1. Theoretical and practical lectures were elaborated for all the classes that involved the use of computational tools, where students understood the theory of each tool before to put this knowledge into practice. For the accomplishment of the practical procedures, the students used tutorials and scripts elaborated specifically for this purpose. Thus, the students had the opportunity to learn more about each tool that would be used.

All classes were planned with the purpose of creating the best possible scenario so that the students could then explore the tools in an autonomous way, discussing in groups and acting in an active way on the problems presented to them.

In all, four tutorials and one script were elaborated, the tutorials were used for the sequence analysis classes, BLAST alignment, multiple alignments, and molecular phylogeny, and the script was prepared for application in the database class as shown in Table 1.

Table 1: Themes of the classes, and hours dedicated to each theme.

ACTIVITY	HOURS
Introduction to Bioinformatics	2
Chromatograms analysis	2
Database	2
Sequence alignment	4
Molecular Phylogeny	4
Questions	4
Presentation of problem solutions	2

In addition to the theoretical-practical classes, four classes were available for students to solve problems related to the operation of the software. The theoretical classes aimed to address the main characteristics of the subject in question and demonstrate the functioning of the tools that would be used.

After the classes, students solved a problem of low complexity in the classroom with the aid of a tutorial and the professor, in this way the students became familiar with the programs used. After all the theoretical and practical classes, the students formed five groups, each one with their own themes and biological problems to be solved, which was randomly defined.

Each problem would have some questions that should be answered, and the students should prepare a presentation where they would demonstrate their problems, their results and the methodologies used. Each group was responsible for finding solutions presented to them in the way they believed to be the best possible.

The purpose in establishing problems is to stimulate the student in seeking the possible solutions for such in an autonomous way, aiming at amplifying the critical thinking, the reasoning and investigative abilities of the students.

Biology students had a week to analyze their problems and make a presentation to show the class how they solve their problems, each group had the theme of one of the classes, and their respective problems, as shown in Table 2.

Table 2: Distribution of the themes to the formed groups.

Group	Theme
Group 01	Chromatograms analysis
Group 02	Database
Group 03	Sequence alignment with BLAST
Group 04	Multiple alignment
Group 05	Phylogeny

The Problems of each group were as follows:

Problem 1: Sequencing Analysis

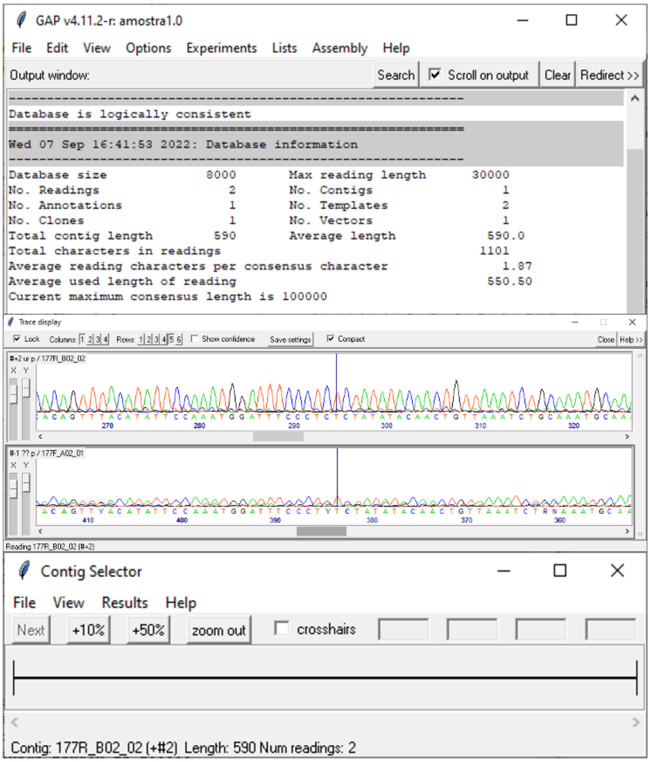
Background

For these analyses, the students received files of four sequences, two files for sample 01 (lower quality) and the others referring to sample 02 (higher quality) (figure 01). These characteristics were selected to enable students to understand the main points to be observed during sequencing analysis.

In the first sample, using a Phred value of 15 it was possible to assemble a sequence with a length of 590bp using the two sequences provided. In sample 02, with the same Phred value, it was only possible to assemble a consequence using only a sequence of files provided, and the total compliance with the agreement was 176pb.

Figure 1: Contigs formed by the chromatogram files provided to the students using the Staden Package program. This graphic output shows the numerical information about the reads and the contigs formed, the chromatogram trace display. A: sample 01 B: sample 02.

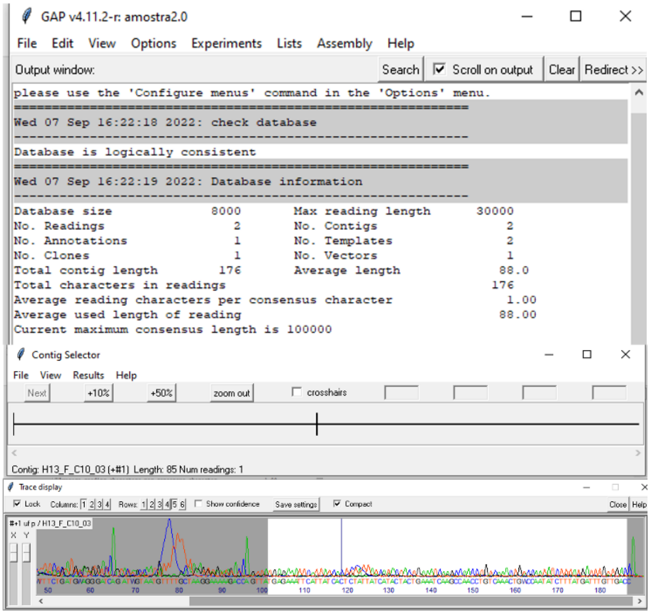
A



```
-------------------------------------------------------------
Database is logically consistent
=============================================================
Wed 07 Sep 16:41:53 2022: Database information
-------------------------------------------------------------
Database size                8000       Max reading length      30000
No. Readings                    2       No. Contigs                 1
No. Annotations                 1       No. Templates               2
No. Clones                      1       No. Vectors                 1
Total contig length           590       Average length          590.0
Total characters in readings                                     1101
Average reading characters per consensus character                 1.87
Average used length of reading                                   550.50
Current maximum consensus length is 100000
```

Contig: 177R_B02_02 (+#2) Length: 590 Num readings: 2

B



```
please use the 'Configure menus' command in the 'Options' menu.
=============================================================
Wed 07 Sep 16:22:18 2022: check database
-------------------------------------------------------------
Database is logically consistent
=============================================================
Wed 07 Sep 16:22:19 2022: Database information
-------------------------------------------------------------
Database size                8000       Max reading length      30000
No. Readings                    2       No. Contigs                 2
No. Annotations                 1       No. Templates               2
No. Clones                      1       No. Vectors                 1
Total contig length           176       Average length           88.0
Total characters in readings                                      176
Average reading characters per consensus character                 1.00
Average used length of reading                                    88.00
Current maximum consensus length is 100000
```

Contig: H13_F_C10_03 (+#1) Length: 85 Num readings: 1

Problem

“You are part of a Bioinformatics lab team, and you have received from another lab some results from a Sanger sequencing that was performed. The laboratory requests your services to answer the following questions:”

a)Did all the samples show a good sequencing? Justify.

b)What is the cut-off value of PHRED used in this analysis? Why?

c)Are the contigs formed correctly? What is the size obtained in the consensus sequences?

Problem 2: Biological Databases

Background

For this problem the answers will depend on the databases used by the students, during the bioinformatics introduction classes some databases were presented to them such as Uniprot, GenBank, PubMed and PDB, however, students could bring information from other databases

Problem

Your professor is requesting a set of information regarding the Human Papillomavirus E6 (HPV16) gene and requests the following responses:

a)Which database will you use to access the information of a single gene? Why?

b)What is the size of the gene, the location of the gene in the genome of the organism and how many amino acids does the protein have?

c)What is the function/functions of the gene in question?

Problem 3: Alignment with BLAST

Background

For this problem, 4 different sequences were given to students, the samples 01, 03 and 04 had nucleotides sequences, however the sample 02, contained a sequence of amino acids.

Different types of sequences have been strategically selected, so students should understand the differences between BLAST types and which to choose for each sample.

These sequences where selected aiming creates some difficulties for the students to solve (see table 03). As in the sample 04 for example that had only a fragment of nucleotide sequence obtained from Genbank, so the students would have only part of data and could verify e-value's differences.

Table 3: Information about the sequences provided to the students initially.

Sample	Gene	Organism	Length	Biomolecule	Accession ID
Sample 01	COI	*Homo sapiens*	386 bp	Nucleotide	KC750830.1
Sample 02	RuBisCO (Large subunit)	*Arabdopsis thaliana*	479 a.a.	Amino acids	NP_051067.1
Sample 03	DRD4 (mRNA)	*Homo sapiens*	1260 bp	Nucleotide	NM_000797.4
Sample 04	matK	*Cocos nucifera*	64 bp	Nucleotide	HG969989.1

Problem

A research group divided the tasks for a job execution, you and your team were responsible for identifying which organisms belong to a set of sequences, finding species genetically like the samples that were sent to you through a local alignment search tool. They asked for the following answers:

a)There are 5 ways to perform searches according to the sequences you get, knowing this, what method will you use (BLASTn, BLASTp, tBLASTn ...)? Justify.

b)Are the results found in your search statistically significant? Why?

Problem 4: Multiple Sequence Alignment

Background

For this problem 12 different sequences were given to the students each sequence belongs to a different primate species', the objective in this problem is to analyze the similarities and polymorphisms between the sequences besides define types of mutations, and how some mutations could change the protein characteristics.

Problem

Upon receiving a set of sequences, your advisor asks you to perform an alignment and do a simple analysis on these, and asks you to check:

a) Which method would you choose to perform the alignment? ClustalW or MUSCLE? Why?

b) After the alignment, it was possible to verify that at the 4th site there is a substitution of bases, does this substitution change the corresponding amino acid? What kind of mutation is this?

c) At site 634 is it possible to observe that a Gap was inserted into the Tarsius syrichta species. Did that Gap modify the sequence reading matrix? If yes, how can you minimize the effects of the insertion of this Gap taking into consideration the reading matrix?

Problem 5: Molecular Phylogeny

Background

The sequences analyzed by students were the same 12 as given in the previous problem, but here the objectives were to understand phylogenetic tree construction methods' and evaluate the accuracy of tree performed and empiric knowledge about phylogenetic relations between those species.

Problem

Among the tasks of your research group, you and your team were given the function of performing the phylogenetic analysis of some aligned genes. To create the phylogenetic tree, your boss asked you the following:

a) Among the methods of phylogenetic tree construction, which one will you use and for what reason (Neighbor-joining, maximum likelihood, maximum parsimony)?

b) Is this tree reliable? Can it explain the phylogenetic relationships between the species analyzed? Why?

Students perspectives

This data was obtained by using a survey with 11 questions, where 08 (eight) were open ended questions, aiming to obtain greater clarity in the answers, and 03 (three) were close ended questions, facilitating the categorization of the data. The data obtained were tabulated (for close ended questions) and analyzed individually (for open ended questions).

Results and Discussion

During the search for solutions, it was possible to notice that some problems were more complex than others, students have some difficulty in solving and answering problems about sequencing analysis, database, and multiple sequence alignment. However, when solving the problems about molecular phylogeny, students reported greater difficulties.

The main difficulty reported by the students was the understanding of the phylogenetic tree construction methods, and, therefore, being able to justify whether such trees would be “reliable” or not to discuss their results.

Problem 1: Sequencing Analysis

The students presented their results demonstrating some changes made in relation to the phred value and how this interfered in the analyses.

Regarding the questions, the students answered:

a)Did all the samples show a good sequencing? Justify.

No, the second sample resulted in a consensus sequence of smaller size and quality, even reducing the Phred value to 5, that is, with an accuracy of approximately 68% for each base, one of the sequencing reads was not used.

b)What is the cut-off value of PHRED used in this analysis? Why?

The cut-off value is 20, as this value indicates that the probability of a base being incorrect is 1 in 1000, that is, the analyzes would be 99.9% accurate.

c)Are the contigs formed correctly? What is the size obtained in the consensus sequences?

No, the consensus sequence formed were 590bp and 176bp, however when changing the Phred value to 10, the contig sizes were 652bp and 196bp, but the accuracy for base call output was 90%.

During the questions asked by the teachers, the students demonstrated a certain mastery about content, and were able to discuss their results. The students were able to understand that by changing the

Phred value it was possible to obtain more information from the sequencing, however the accuracy of these data was reduced. When asked which Phred value they would use, the group of students chose 20 as the Phred value, as the data would be “Safer” for later analyses. And when asked what should be done with sample 02, they replied that extraction and sequencing should be done more carefully to obtain better quality.

Problem 2: Biological Databases

a)Which database will you use to access the information of a single gene? Why?

“We will use GenBank to find out the size of the gene and the name of the protein it generates and some other basics information. Because this database is widely used by researchers around the world, so they have a lot of data in there”.

b)What is the size of the gene, the location of the gene in the genome of the organism and how many amino acids does the protein have?

“The gene has 476 nucleotides and is located after the L1 gene and before the E7 gene, its protein has 158 amino acids”.

c) What is the function/functions of the gene in question?

For this question the students also used the Uniprot database, and their answer was: “It acts in cellular transformation and also as an oncoprotein, stimulating the destruction of many regulatory proteins of the host cell.” The students cited the database as a source of information.

It was notable that the students in this group did not have much difficulty in carrying out the task. Perhaps in future applications it would be necessary to create more questions that seek other information about the gene and its products, such as whether the gene has intros and exons, basic characterization of the protein structure, among others.

In some graduate programs it is possible to observe some difficulties of graduate students in the search for information such as those listed above Lewis e Bartlett (2013), which encourages this type of approach, because then students would be able to obtain more information. insights into the genes of interest and possibly acquire some independence in researching this information for future projects

Problem 3: Alignment with BLAST

a) There are 5 ways to perform searches according to the sequences you get, knowing this, what method will you use (BLASTn, BLASTp, tBLASTn ...)? Justify.

For samples 1, 3 and 4, BLASTn was used, as this type of blast compares input nucleotides with nucleotides deposited in the database. For sample 2, which has amino acids as input data, BLASTp was used.

It is important to point out that during the presentation the students also explained in a superficial way the method of each type of BLAST, which reinforces that there was research before solving the problems imposed on them.

b) Are the results found in your search statistically significant? Why?

For this question, the students explained the method used by BLAST to score the alignments, but as some hits have the same score, the alternative found by the students was to use the e-value as a filter for choosing the “ideal hit” (results in Table 04).

It is important to point out that the students showed some fear in this questioning, as they had not understood that some sequences are “redundant” because they present gene fragments, isolated regions, among others, that is, in some situations 3 or more hits had a score and e-values.

identical, however, when analyzing it was possible to notice that some hits were the same gene, however isolated or just a fragment of it.

Table 4: Information about the sequences selected by students after the BLAST search.

Sample	Gene	Organism	Score	Query cover	e-value	Identity	Acession ID
Sample 01	COI	*Homo sapiens*	713	100%	0.0	100%	KC750830.1
Sample 02	RuBisCO large subunit	*Arabidopsis thaliana*	962	100%	0.0	100%	NP_051067.1
Sample 03	DRD4 (mRNA)	*Homo sapiens*	2327	100%	0.0	100%	NM_000797.4
Sample 04	Matk	Some different species	***	***	***	***	***

For samples 01, 02 and 03 the results were the same as the initial ones, that is, the method used by the students worked, however for sample 04 the gene was selected previously, however several species presented a 100% coverage and the same score.

When asked about sample 04, the students reinforced that the amount of information would be insufficient for a result with less bias and related this argument to the e-value.

Another interesting point that was not raised by the students is that the gene in question is relatively conserved, since it is linked to several important functions for plants. This could be another question raised by the problem, encouraging students to relate genes to their products and their functions and not just the sequence of nucleotides in isolation from the biological context.

Problem 4: Multiple Sequence Alignment

a) Which method would you choose to perform the alignment?

ClustalW or MUSCLE? Why?

ClustalW implements an iterative algorithm so that errors produced in the previous step are unlikely to be corrected in the later step, while MUSCLE implements a progressive algorithm that allows new optimizations throughout the process.

Despite being correct, the students did not show much confidence when answering this question, which leads to the assumption that other methodologies or questions could be used to improve the understanding of how the methods work, better unraveling their algorithms.

b)After the alignment, it was possible to verify that at the 4th site there is a substitution of bases, does this substitution change the corresponding amino acid? What kind of mutation is this?

“Yes, there was a transition type substitution (Cytosine to Thymine) and changes the amino acid, where the codon encoded the amino acid Glutamate and with the mutation it encoded a Lysine”.

c) At site 634 is it possible to observe that a Gap was inserted into the Tarsius syrichta species. Did that Gap modify the sequence reading matrix? If yes, how can you minimize the effects of the insertion of this Gap taking into consideration the reading matrix?

“Yes, it did. In order to reduce the effect of this gap, we could extend the gap (adding another two gaps shortly afterwards)”.

At this point, the objective was to show students that “gaps” do not exist biologically, but are artifices used to maintain coherent alignment without modifying the codon reading matrix. Based on the students' response to the questions made during the presentation, this idea was well understood by them.

Problem 5: Molecular Phylogeny

a) Among the methods of phylogenetic tree construction, which one will you use and for what reason (Neighbor-joining, maximum likelihood, maximum parsimony)?

Neighbor-joining is a simpler clustering algorithm that groups haplotypes based on genetic distance, that is, it would not be interesting to use this method for the purpose of the problem.

Maximum parsimony is a method that tries to minimize the lengths of the branches by minimizing the distance (minimum evolution) or minimizing the number of mutations (maximum parsimony) that explain the input data. However, the problem with this method is that it does not consider many factors of sequence evolution (e.g., reversals, convergence, and homoplasy).

Maximum likelihood can be considered an interesting model for building a phylogeny using sequence data. This method is widely used in publications like the Bayesian method; however, they need more powerful computers.

This problem was the one that the students had the most difficulties, despite the lack of explanations about the Bayesian method, the students realized that it is a method used in publications, this is another point that could be added to this problem.

Another important point is that the students seemed to have memorized what each method does, however, they did not seem to really understand the advantages and disadvantages of each method or how to associate the method with the objective of the problem.

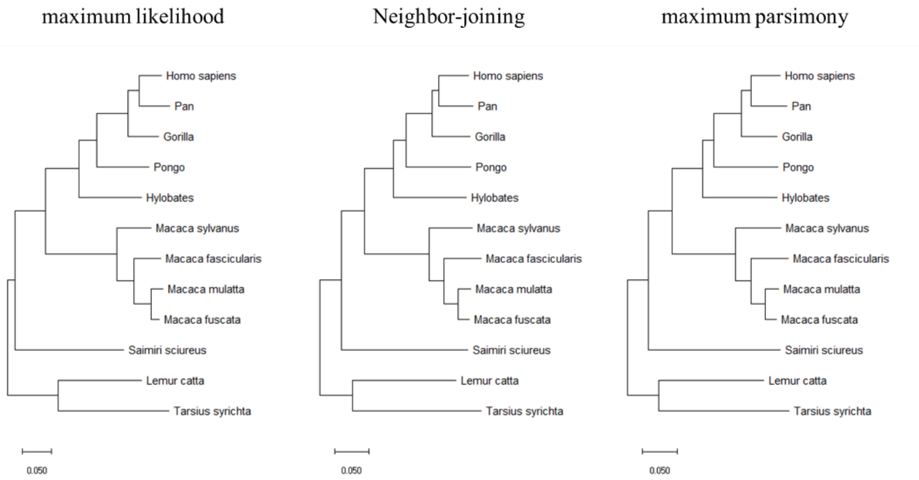
This may be due to several reasons, one of which may be that the approach to phylogeny must be improved, or the problem better constructed. Another point may be the students' background on the topic "biological evolution" and "phylogenetics", as it is observed that some students may have difficulties with these topics (Chapman et al., 2010).

b)Is this tree reliable? Can it explain the phylogenetic relationships between the species analyzed? Why?

The students presented three distinct phylogenetic trees (Figure 2) The trees were made by different methods; however, they presented similar topologies, the students used this argument to defend that the trees would be reliable, another point used by them was other evidence found in the literature that reinforced the phylogenetic relationship between some species such as *Pan troglodytes* com *Homo sapiens* e *Macaca mulatta* com *Macaca fuscata*.

However, from the problems presented, this was the one that appeared the most doubts, the students appeared not to understand much about the relationships between the evolutionary models and the methods used. As seen in Chapman, 2010, most of the students seem not to have mastered the concepts related to phylogeny and evolution.

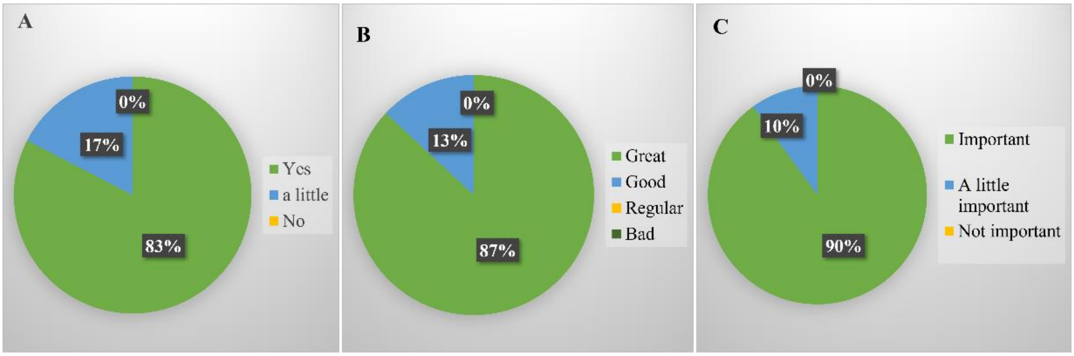
Figure 2: Phylogenetic trees made by the three different methods: maximum likelihood, neighbor-joining and maximum parsimony.



Students perspectives:

When analyzing the close ended questions, it was possible to observe a satisfactory evaluation about the teaching methodology used. When asked about the importance of Bioinformatics in their academic training or if Bioinformatics increased their investigative abilities, the most part of the students considered important and helpful, as it is observed in Figure 3.

Figure 3: Graphics representing the answers of the students about the following questions: A: Do you believe that Bioinformatics classes have expanded your investigative skills? B: How would you evaluate the insertion of Bioinformatics in your class? C: About the importance of Bioinformatics in your academic training, how would you classify it?



When asked about the introduction of Bioinformatics to Biological Sciences undergraduate students, 47% (11 participants) answered the first question in a similar way, highlighting the importance of introducing Bioinformatics in undergraduate courses, as well as the lack of information about this topic, which can be observed in some of the answers provided: ‘’Student 1: Important. Because we have little contact with this kind of subject, even less the practices.”; ‘’Student 2: It is interesting because it is something that we do not have in the curriculum, and it is a very used tool in the area of research and laboratory.”; ‘’Student 3: I found it important because it introduced knowledge beyond the curriculum, opening up new perspectives for other forms of employment in the field of biology.”; ‘’Student 4: I found it very interesting, because it is an area that is not widely spoken in undergraduate courses and is currently widely used in postgraduate studies [...]”; ‘’Student 5: A great advantage, because we do not have this kind of lectures in the curriculum of our course.”

Based on these answers, it is notable that students realize that there is a need to use Bioinformatics even in the undergraduate curriculum. Thus, the professor can contribute with aspects that aid in the learning of these students, basic knowledge about this subject is likely to be fundamental for the postgraduate research of many of these students. According to Wightman e Hark (2012), the introduction of Bioinformatics concepts to students early in their careers is effective and results in significant learning gains. This was clearly demonstrated in students own speeches.

The second question asked if Bioinformatics classes would expand their investigative skills. In total, 82% (19 participants) said they believed that their investigative abilities were indeed broadened according to the approached content. The student’s evaluation of the course in general was very positive, where 73% (17 participants) evaluated as great and the other 27% (6 participants) as good.

These two questions can demonstrate that the introduction of this discipline has contributed with important aspects to the professional future of the students. Establishing critical and analytical thinking has not been an easy task in teaching and the use of these tools under this approach can contribute to this purpose.

Regarding the importance and relevance of Bioinformatics in the

Biological Sciences undergraduate curriculum, 22 participants considered it important for their training. In addition, all the students agreed that, in fact, this subject is relevant to the curriculum. However, some students attributed greater importance of this discipline, while others (56%) said that Bioinformatics lectures would be able to increase the number of Bioinformatics trainees, and to prepare them for a job in the field.

Another important point is that the modern job market requires students to be competent in the use of computational tools to generate and analyze data McDonald et al., (2022), and it was possible to notice that the students initially had difficulties in understanding the programs and realizing how the changes they made interfered with the results. However, in a short time of use, the students were already able to better interact with the algorithm and get better results. This was also observed in McDonald et al., (2022), where the author asserts the importance of using computational tools to incorporate algorithmic thinking.

One of the fundamental points for this study was the active methodology used to approach the chosen themes, since it used praxis in the search of assigning meaning to the fundamental concepts of Bioinformatics. The participants of the study were also invited to evaluate this question through questions 6 and 7. Therefore, 100% of the participants considered this approach to be more efficient than the expository traditional method, some even justify the reason why this methodology was well evaluated by them, as can be observed in the answers provided. ‘’Student 1: It is more efficient, since it makes the student "protagonist ", they who will seek and build the knowledge within the subjects that are exposed”; ‘’Student 4: Very efficient, because it stimulates the student.”; ‘’Student 6: More efficient, therefore, made us exercise what we learned in theory and thus know how to identify in what kind of situation we can apply it in practice.”; ‘’Student 7: Much more effective, because if we only worked the exhibition without the active search, we would see only closed problems that would not represent real problems that we could face in our careers.”

Research stimulates creativity and curiosity and increases the critical attitude of students towards knowledge and become more motivated to learn (Munford & Lima, 2013; Rodrigues & Silva, 2009; Zompero & Laburú, 2010). In relation to the suggestions for approaching the subject, some students brought interesting characteristics that should be considered, for example: "to follow a step by step from the collection of the data to the analysis", or "every student to think about problems and find ways to solve it”. These statements connote how students become enthusiastic when feeling part of the teaching-learning process, as well as feel more protagonists when actively participating in the construction of their knowledge.

As for the exclusive approach of the subject, the class was divided. Some students (56%) believe that with an exclusive approach there would be the possibility of further deepening of the contents, besides more time for the accomplishment of the activities. Others (44%) believe that interdisciplinarity may be the key to an efficient approach since Bioinformatics is used in many areas.

Regarding previous knowledge on Bioinformatics, 12 participants (52%) claimed not to know the discipline before this approach. The others (48%) said they knew the subject superficially. Some authors as (McDonald et al., 2022; Balasubramanian & Chatterjee, 2022) used a specific approach to bring Bioinformatics tools to students. However, an interdisciplinary and more general approach can show students the possibilities in the use of Bioinformatics, avoiding a limited understanding of the application of the tools.

A possible approach would be to introduce Bioinformatics in an interdisciplinary way, and then use more specific approaches with the aim of training qualified human resources for more specific analyses, as suggested by (Justino et al., 2021).

Regardless of whether the Bioinformatics approaches should be exclusive or associated with another area of knowledge (Biochemistry, Genetics, or Evolution), it is suggested to use problems applied to students so that they find the solutions. The teacher can make use of questions that will guide the students, avoiding possible ambiguous approaches or outside the teacher objective. These questions also may serve as a concrete means for the instructor to assess what students have learned (Inlow et al., 2007).

Similar results have been shown by (Junior, 2010; Aguiar et al., 2022) with the introduction of bioinformatics in undergraduate courses in Biological Sciences, reinforcing the idea discussed here, expanding the importance of bioinformatics also for the training of basic education teachers (Coutinho et al., 2022).

Throughout history, changes in the curriculum of undergraduate courses in Biological Sciences have occurred due to the evolution of scientific knowledge, breaking with traditional conceptions of training and opening up to new study contents Lucas e Ferreira (2017) as well as to innovative methods, such as active methodologies and information, and communication technologies (Colares & Oliveira, 2018; Santos et al., 2020).

Final Considerations

In the case reported in this study, it is notable that few Biological Sciences undergraduate students know Bioinformatics, and consequently they do not know how to use or apply its tools. Thus, the importance of introducing Bioinformatics tools in the Biological Sciences curriculum is notorious, since it has proved to be a facilitator of the teaching-learning process of the students, besides expanding skills considered fundamental for these professionals and, above all, stimulating the critical thoughts about the subjects explored.

The problems used in this study aimed to improve some skills and, from the students' perspective, this objective was achieved. In addition, it was possible to identify some changes that could contribute to the improvement of the introduction of bioinformatics. Perhaps asking more questions with increasing complexity seems to be useful to lead students to a better and deeper understanding of biological concepts and bioinformatics tools.

Another topic that could be improved is on evolution and phylogenetic concepts since students had more difficulties in this type of questions. Considering the students' performance in the maids' presentation, it is possible to affirm that the methodology was used to understand the biological concepts.

Study participants reported being able to identify that this discipline helped them to perceive possible solutions to a biological problem. In addition, bringing Bioinformatics to undergraduate courses is important for the training of undergraduates, generating qualified human resources, not only for the immediate job market, but also for postgraduate training.

Thus, it is possible to infer that the application of activities such as those proposed in this study adds knowledge and amplifies important skills for the professional future, which demonstrates how these methodologies improve the quality of the biologist training.

Acknowledgements

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico [grant number 307128/2022-9]; Fundação de Apoio à Pesquisa e a Inovação Tecnológica do Estado de Sergipe [grant number 02/2020]; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Finance Code 001].

References

Aguiar, C. C., Casotti, M. C., Souza, K. C., & Alencar, I. C. C. (2022). Bioinformática como ferramenta de aprendizagem no ensino de biologia celular e molecular pela percepção do estudante. A incorporação das ferramentas digitais como um modelo transformador no processo de ensino na área da saúde (pp. 30–45). https://doi.org/10.22533/at.ed.2612216026.

Badotti, F., Barbosa, A. S., Reis, A. L. M., Do Valle, I. F., Ambrosio, L., & Bitar, M. (2014). Comparative modeling of proteins: A method for engaging students' interest in bioinformatics tools. Biochemistry and Molecular Biology Education, 42(1), 68–78. https://doi.org/10.1002/bmb.20721.

Balasubramanian, A., & Chatterjee, J. (2022). Bioinformatics approach used in undergraduate research to predict siRNA as ZIKV therapeutics. Biochemistry and Molecular Biology Education, 50(2), 237–245. https://doi.org/10.1002/bmb.21605.

Colares, K. T. P., & Oliveira, W. (2018). Metodologias ativas na formação de profissionais de saúde: Uma revisão. Sustinere, 6(2), 300–320. https://doi.org/10.12957/sustinere.2018.36910.

Coutinho, T. J. D., Sousa, F. B., & Vasconcelos, R. A. (2022). Bioinformática para a educação básica: Capacitando docentes para o uso de ferramentas computacionais em sala de aula. Revista de Ensino de Ciências e Matemática, 13(4), 1–16. https://doi.org/10.26843/rencima.v13n3a17.

Farah, S. B. (2007). DNA: Segredos e mistérios. Sarvier.

Farias, A. Q. P., Chacon, P. F. S., & Silva, N. R. R. (2012). A bioinformática como ferramenta de formação de recursos humanos no IFRN. Holos, 6, 113–123.

Furge, L. L., Strevens-Truss, R., Moore, D. B., & Langeland, J. A. (2009). Vertical and horizontal integration of bioinformatics education: A modular, interdisciplinary approach. Biochemistry and Molecular Biology Education, 37(1), 26–36. https://doi.org/10.1002/bmb.20249

GenBank. (2017). GenBank and WGS statistics. NCBI.

https://www.ncbi.nlm.nih.gov/genbank/statistics

GOLD. (2017). Statistics: Complete and permanent draft genome totals in GOLD. https://gold.jgi.doe.gov/statistics

Inlow, J. K., Miller, P., & Pittman, B. (2007). Introductory bioinformatics exercises utilizing hemoglobin and chymotrypsin to reinforce the protein sequence-structure-function relationships. Biochemistry and Molecular Biology Education, 35(2), 119–124. https://doi.org/10.1002/bambed.30

Justino, G. C., Nascimento, C. P., & Justino, M. C. (2021). Molecular dynamics simulations and analysis for bioinformatics undergraduate students. Biochemistry and Molecular Biology Education, 49(4), 570–582. https://doi.org/10.1002/bmb.21512

Lewis, L., & Bartlett, A. (2013). Inscribing a discipline: Tensions in the field of bioinformatics. New Genetics and Society, 32(3), 243–263. https://doi.org/10.1080/14636778.2013.773172

Lucas, M. C., & Ferreira, M. S. (2017). História do currículo da formação de professores de ciências e biologia (1960/70). Educação em Foco, 22(2), 145–166.

McDonald, A. R., Roberts, R., Koeppe, J. R., & Hall, B. L. (2022). Undergraduate structural biology education: A shift from users to developers of computation and simulation tools. Current Opinion in Structural Biology, 72, 39–45. https://doi.org/10.1016/j.sbi.2021.07.012

Melo, B., & Sant'Ana, G. A. (2013). A prática da metodologia ativa: Compreensão dos discentes enquanto autores do processo ensino-aprendizagem. ESCS Educação, 23(4), 327–339.

Mertz, P., & Streu, C. (2015). Writing throughout the biochemistry curriculum: Synergistic inquiry-based writing projects for biochemistry students. Biochemistry and Molecular Biology Education, 43(6), 408–416. https://doi.org/10.1002/bmb.20914

Munford, D., & Lima, M. E. C. (2013). Ensinar. Journal of Chemical Information and Modeling, 53(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Nunes, R., Barbosa de Almeida Júnior, E., Pessoa Pinto de Menezes, I., & Malafaia, G. (2015). Learning nucleic acids solving by bioinformatics problems. Biochemistry and Molecular Biology Education, 43(5), 377–383. https://doi.org/10.1002/bmb.20886

Ribeiro Junior, H. L. (2010). Atividades práticas de bioinformática como recurso pedagógico para a formação de professores do curso de ciências biológicas na Universidade Estadual do Ceará - UECE Fortaleza - Ceará [Trabalho de conclusão de curso, Universidade Estadual do Ceará].

Ribeiro Junior, H. L., de Oliveira, R. T. G., & Ceccatto, V. M. (2012). Bioinformática como recurso pedagógico para o curso de ciências biológicas na Universidade Estadual do Ceará – UECE – Fortaleza, Estado do Ceará. Acta Scientiarum. Education, 34(1), 129–140.

https://doi.org/10.4025/actascieduc.v34i1.1458

Ribeiro, D. da S., & de Jesus, T. F. (2015). A práxis pedagógica no ensino superior: Um relato baseado na necessidade de ações inovadoras de avaliação. Cairu em Revista, 4(5), 124–142.

Rodrigues, F. A., & Silva. (2009). O ensino por investigação e as práticas epistêmicas: Referências para a análise da dinâmica discursiva da disciplina “Projetos em Bioquímica.” VII Encontro de Pesquisa em Ensino de Física. Encontro Nacional de Pesquisa em Educação em Ciências.

Santos, R. O. B., Cabette, R. E. S., & Luis, R. F. (2020). Novas tecnologias aplicadas ao ensino: Utilização da gamificação como estratégia didática. Revista Brasileira de Ensino de Ciências e Matemática, 3(6), 69–78.

Silva, A. C. R. (2012). “Educação por competência,”.

Wightman, B., & Hark, A. T. (2012). Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. Biochemistry and Molecular Biology Education, 40(5), 310–319. https://doi.org/10.1002/bmb.2063

Zhang, X. (2009). Using Arabidopsis genetic sequences to teach bioinformatics. Biochemistry and Molecular Biology Education, 37(1), 16–23. https://doi.org/10.1002/bmb.20250

Zompero, A. D. F., & Laburú, C. E. (2010). As relações entre aprendizagem significativa e representações multimodais. Ensino e Pesquisa em Ensino de Ciências, 12(3), 31–40.