International Journal of Education and Practice
2025 Vol. 13, No. 2, pp. 520-535
ISSN(e): 2310-3868
ISSN(p): 2311-6897
DOI: 10.18488/61.v13i2.4070

Developing principles and steps for integrating bioinformatics into teaching for pedagogical university students: A case study in Vietnam

Hang Thi Nguyen[1]
Lan Thi Ngoc Nguyen[2]
Quan Huu Nguyen[3]
Tan Quang Tu[4]
Mau Hoang Chu[5+]

[1,2,3,4,5]*Thai Nguyen University of Education, Vietnam.*
[1]*Email: hangnt@tnue.edu.vn*
[2]*Email: ntnlan.dhsptn@tnu.edu.vn*
[3]*Email: nguyenhuuquan@tnu.edu.vn*
[4]*Email: tantq@tnue.edu.vn*
[5]*Email: chuhoangmau@tnue.edu.vn*

(+ Corresponding author)

Article History
Received: 22 August 2024
Revised: 8 January 2025
Accepted: 23 January 2025
Published: 31 January 2025

Keywords
Bioinformatics and STEM
Bioinformatics integration
Biology teaching
Integration principles and steps
Pedagogical university students
Vietnamese biology education curriculum.

ABSTRACT

The main objective of this study is to develop the principles and methods for integrating bioinformatics into biology teaching. The study started with analyzing the high school biology curriculum, a survey on bioinformatics knowledge of teachers and pedagogical university students, a survey of teacher training curricula of pedagogical universities and education faculties and proposed principles and steps for integrating bioinformatics into biology teaching. The research results proposed six basic principles and a four-step process to guide pedagogical university students in integrating bioinformatics into biology teaching. The main principles include being simple, identifiable and appropriate for students, being compatible with biological knowledge content and bioinformatics, breaking down the knowledge module for analysis and empowering, personalizing students creating a model for analysis on paper before utilizing a computer and producing particularly good products related to bioinformatics. A four-step process, including selecting knowledge topics suitable for bioinformatics applications, dividing the topic into small knowledge modules, data analysis using bioinformatics, and creating bioinformatics products. The findings of this study provide valuable insights that show that integrating bioinformatics into high school biology teaching is necessary. Applying principles and methods of integrating bioinformatics into teaching promises to improve the quality of biology teaching and promote STEM education in Vietnam and developing countries.

Contribution/Originality: The study develops principles and steps for integrating bioinformatics into teaching, empowering teachers to adjust to improve teaching effectiveness and student self-learning. This unique approach addresses the difficulties and confusion of teachers and students of education when using bioinformatics to engage in biology education activities.

1. INTRODUCTION

The education system in Vietnam is open and interconnected encompassing two types of education: formal education and continuing education. This system ensures all citizens' right to access education and training nationwide. Education levels and training within Vietnam's system include preschool, school, vocational, and university education. The training products of higher education are bachelors, masters, and doctorates. Specialized universities such as universities of education and multidisciplinary universities may have faculties of education for

teacher training within Vietnam's university system. Pedagogical students are enrolled in universities or faculties of education where they learn the teaching profession. These training institutions produce teachers who specialize in subjects for high schools such as mathematics, physics, biology and others. The difference in training compared to other countries is that pedagogical students in Vietnam must acquire knowledge and learn the teaching profession, limiting their access to modern technologies to support learning. Vietnam's new general education program, Vietnam Ministry of Education and Training (2018a) emphasises the development of students' biological competence. Concurrently, it contributes to developing critical qualities in students such as love for nature and pride towards the homeland. It also contributes to forming standard capabilities, such as a capacity for autonomy and self-study, communication and cooperation, problem-solving, and creativity. Using integrated methods in teaching is an effective approach to developing students' capacity. A notable theme in the content is bioinformatics as a tool for biological research and learning in the Biology Education Curriculum in high schools in Vietnam (Vietnam Ministry of Education and Training). Integrating bioinformatics into high school biology teaching promotes STEM (science, technology, engineering, and mathematics) and education (Kelley & Knowles, 2016; Kovarik et al., 2013). In bioinformatics, the following branches of science and technology are represented: biology, chemistry, biochemistry, physics (science), biotechnology, computer science, artificial intelligence (technology), statistical techniques, and mathematics (Luscombe, Greenbaum, & Gerstein, 2001). Therefore, bioinformatics is the ideal STEM science for building innovation in biological research and teaching (Kovarik et al., 2013). Thus, bringing bioinformatics into teaching in high schools in developing countries is inevitable and in line with global trends. The rapid development of modern biology and updating knowledge about genomes, proteomes, and molecular evolution in general education programs require integrated methods in teaching biology, and bioinformatics integration is a practical direction for application. Recently, there have been many discussions about the application of bioinformatics in school teaching and the training of bioinformatics knowledge for teachers. Opinions have been exchanged on applying bioinformatics in teaching biology with particular attention paid to integrating bioinformatics in teaching. However, reports on the difficulties in meeting human resources for this application have yet to be found. Regarding human resources, high school teachers and university students need updated knowledge and skills in bioinformatics to improve teaching efficiency. This situation shows that training for pedagogical students and school teachers to identify appropriate knowledge topics and how to apply bioinformatics in teaching is very necessary. We have yet to find any studies guiding pedagogical students to approach bioinformatics integration in high school biology teaching. In this context, this study aims to develop principles and procedures for integrating bioinformatics into teaching biology for pedagogical university students. At the same time, clarify how integrating bioinformatics into teaching will promote STEM education in general biology teaching. The study's key questions are as follows: What are the principles for integrating bioinformatics into biology teaching? How is bioinformatics integrated into teaching a module of knowledge? Are pedagogical university students and school teachers the first ones who need to be trained in bioinformatics and use bioinformatics in teaching? We aim not only to resolve the contradiction between a large amount of modern knowledge and the limited teaching time but also to improve the effectiveness of teaching biology and equip learners with lifelong self-study methods by addressing these questions. The study also aims to show that school teachers and pedagogical university students trained in bioinformatics and integrating bioinformatics in teaching will benefit high school and middle school students and positively spread the integrated approach to bioinformatics in the educational community.

2. REVIEW OF LITERATURE

2.1. Bioinformatics

Knowledge is continuously updated daily and massive biological datasets have been created with the rapid development of biotechnology whose core is molecular biology with research achievements in genomics, proteomics

and molecular evolution. The abundance of biotechnological data from genomic sequencing and functional genomics has led to a new field, bioinformatics, a science combining biology and computer elements (Luscombe et al., 2001). The interdisciplinary scientific subject of bioinformatics creates techniques and software tools for interpreting biological data particularly in cases where the data sets are vast and complicated. Bioinformatics draws on a variety of disciplines, including biology, chemistry, physics, computer science, computer programming, information engineering, mathematics, and statistics to analyze and understand biological data (Suppl. 1). The interaction of science and technology fields has contributed to the formation of bioinformatics (Luscombe et al., 2001). Bioinformatics is applied to analyze genomes and the evolution of organisms from prokaryotes to eukaryotes (Casas & Saborido-Rey, 2023; Tan, Kumar, Wong, & Ling, 2022). Lewitter and Bourne (2011) argued that bioinformatics is now integral to biology, biological research and teaching. The authors discussed the application of bioinformatics with the desire to promote secondary school bioinformatics education. At the same time, the authors also provide a wealth of information about efforts to reach school students, including integrating bioinformatics into biology curricula. They wish that researchers and teachers discuss the effectiveness of integrating bioinformatics into biology teaching (Lewitter & Bourne, 2011). Bioinformatics is considered an indispensable part of modern life science. It has revolutionized and redefined how research is conducted, impacting teaching biology, biotechnology, medicine, agriculture and other fields.

2.2. The Challenge in Training Biology Teachers for Schools and Integrating Bioinformatics into Biology Teaching

The rapid growth of biological databases and a wide range of tools and services have enabled users to mine, understand and use the data. Many modern tools and knowledge areas such as bioinformatics are rarely accessed by biology pedagogy students. This poses a challenge in training biology teachers for high schools in Vietnam in the context of the 4.0 industrial revolution. The learning environment with bioinformatics will promote the construction of new knowledge structures in genetics, thus influencing the students to gain a deeper and more multidimensional understanding of this field (Gelbart & Yarden, 2006). Informatics is often taught separately from other science subjects. On the other hand, bioinformatics is not yet included in the teacher training programs of universities and faculties of education. Updating modern biological knowledge in the curriculum and textbooks, such as molecular biology, genetics, in silico docking, etc., has made it difficult for pedagogical students and high school biology teachers to receive this knowledge in learning and teaching biology. However, the resurgence of informatics education will be the basis for creating opportunities for developing bioinformatics education in schools (Gelbart & Yarden, 2006). The research on the application of bioinformatics in teaching has confirmed that bioinformatics is a tool to improve the effectiveness of integrated teaching and implement STEM education in teaching (Bain et al., 2022; El Islami, Sari, & Utari, 2023; Kovarik et al., 2013; Machluf & Yarden, 2013).

Over the last ten years, there has been discussion on integrating bioinformatics into high school education and helping teachers learn about it. The design ideas and ramifications of incorporating bioinformatics into senior high school were delineated by Machluf and Yarden (2013). Numerous writers have put forth basic guidelines for bioinformatics education in high school (Form & Lewitter, 2011; Via et al., 2011). Meanwhile, others discussed and argued that high school bioinformatics education may support jobs in science, technology, engineering, and mathematics (Kovarik et al., 2013). High school bioinformatics-based projects are crucial in developing students' literacy, curiosity and perspectives on evolution, genomics, and gene control. Bain et al. (2022) recently announced the 4273pi Bioinformatics at School study, an ongoing study in Scotland (Bain et al., 2022). Experiential workshops on bioinformatics for secondary school students with a focus on accessibility, the curriculum seeks to address this problem by creating and providing curriculum-linked. The studies generally focused on explaining the basis and affirming the necessity of applying bioinformatics to teaching biology in high school. The reports initially discussed the principles and ways of using bioinformatics tools in teaching biology and the need to foster knowledge for high

school teachers. However, the reports have not mentioned the difficulties meeting human resources and physical facilities for this change.

Regarding human resources, school teachers and pedagogical students have few opportunities to update their knowledge and skills in bioinformatics, so they often have difficulty teaching modern biological knowledge leading to low teaching efficiency. This situation shows that it is essential to foster pedagogical students and high school teachers in identifying appropriate knowledge topics and how to apply bioinformatics in teaching. We have found no research that reports on guiding pedagogical students to approach integrating bioinformatics into high school biology teaching. In that context, this study proposes a way to guide pedagogical students in identifying principles, knowledge topics, and processes for integrating bioinformatics in teaching biology accompanied by illustrative examples. At the same time, clarifying how integrating bioinformatics in teaching will promote STEM education in high school biology.

3. METHODS

3.1. The Research Process

The research process to develop principles and procedures to integrate bioinformatics into teaching includes the following steps (see Figure 1):

1) Research on theoretical and practical foundations.

2) Developing principles for integrating bioinformatics into teaching.

3) Propose steps to integrate bioinformatics into teaching.

4) Evaluation's experiment of pedagogical students after graduation.

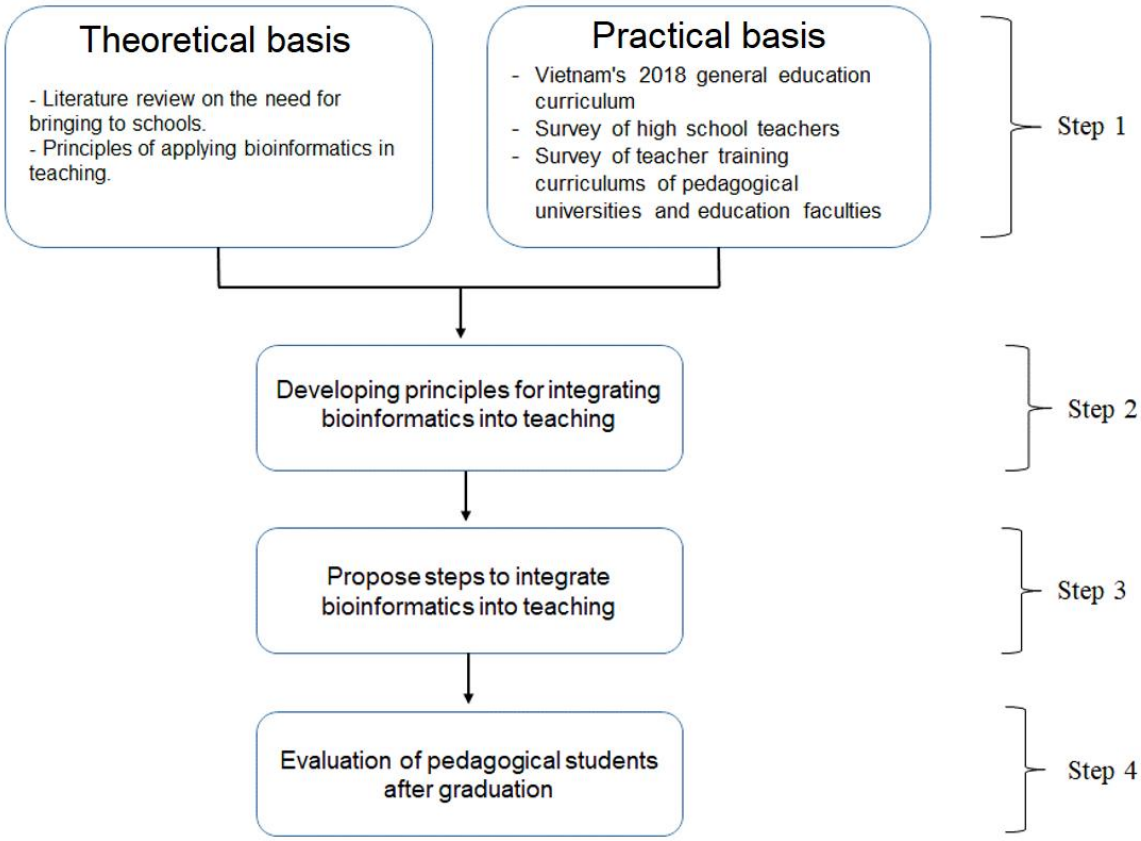

Figure 1. Summary of the research process.

3.2. Participants

The assessment and survey recruited 159 participants including 63 high school teachers and 96 recent university pedagogy graduates, all of whom were valid. Among the high school teachers, there are 28 core teachers from eight provinces in Northern Vietnam, including Lai Chau, Dien Bien, Son La, Hoa Binh, Cao Bang, Bac Kan, Lang Son, Thai Nguyen, and 35 high school teachers who are master's students and Ph.D. students of Thai Nguyen University of Education. The educational level of 63 general education teachers participating in the survey included 74.60% with a bachelor's degree, 23.81% with a master's degree and 1.59% with a doctorate. Among the 96 pedagogical students surveyed, there were 42 students in course 50, 36 in course 51 and 18 in course 52. They all had graduated from the Thai Nguyen University of Education and were preparing to go to high school to become teachers (see Table 1).

Table 1. Information on survey and evaluation participants.

Variables	Education level	N	%
Teachers (n = 63)	Bachelor	47	74.60
Core teachers from local (n=28)	Master	15	23.81
Teachers as master's students and Ph.D. students (n=35) from pedagogical university	Ph.D.	1	1.59
Students (n = 96)	Bachelor	42	43.75
51st training course			
52nd training course		36	37.50
53rd training course		18	18.75

3.3. Survey of High School Teachers

Google Forms were used to survey high school teachers to assess their perception of bioinformatics, evaluate the potential to integrate bioinformatics into teaching biology and identify the main limitations preventing teachers from implementing bioinformatics in teaching biology. The Google Form consists of 12 questions (SQ) (see Table 2), seven yes-no questions and five short answer questions. Yes-no questions (SQ1, SQ2, SQ3, SQ4, SQ5, SQ6, and SQ7) aim to determine interest, educational background and the need to foster knowledge and skills to integrate bioinformatics into teaching. The survey results of this question type were calculated as a percentage of yes or no. The goal of the short answer questions (SQ8, SQ9, SQ10, SQ11, and SQ12) is to evaluate teachers' initial knowledge capacity about bioinformatics and knowledge topics that can apply bioinformatics in teaching. The short answer questions were prepared with standard answers as a basis for comparison with the teacher's answers. The survey results of this question type were calculated as a percentage of no answer passed answer and not passed answer (see Table 2).

Table 2. Questionnaire for high school teachers on understanding bioinformatics and its application in teaching in high schools using Google form.

Questions (SQ)	Variables	Expected answers
Questions requiring a "yes or no" answer.		
SQ1	Are you interested in bioinformatics?	Yes or no
SQ2	Did you take a bioinformatics course at the university?	Yes or no
SQ3	Do you know the relationship between bioinformatics and STEM education in high school biology teaching?	Yes or no
SQ4	Does your academic background make teaching with bioinformatics tools possible?	Yes or no
SQ5	Have you attended a training course on bioinformatics?	Yes or no
SQ6	Do you wish to be fostered in bioinformatics knowledge?	Yes or no
SQ7	Do you wish to participate in an updated course on integrating bioinformatics into biology teaching in high schools?	Yes or no
Questions require short answers.		

Questions (SQ)	Variables	Expected answers
SQ8	What are the research subjects of bioinformatics?	Application of computer science and statistical techniques to the management and processing of biological information.
SQ9	Name one essential bioinformatics software that you know?	Name one of the following software: BioEdit/Or BLAST in NCBI/or MEGA.
SQ10	Name the three largest biological databases in America, Europe, and Asia?	NCBI (National center for biotechnology information, USA), EMBL-EBI (EMBL's European bioinformatics institute) and DDBJ center (Japan).
SQ11	Describe some of the major biological data in the NCBI (National center for biotechnology information, USA) database?	Nucleotide and amino acid sequences, human genome sequences and some species of organisms.
SQ12	Name some topics in the high school biology curriculum that can use bioinformatics as a teaching and learning tool?	Protein, nucleic acid, chemical composition of cells, viruses, bacterial and viral diseases.

3.4. Survey of Teacher Training Curriculums of Pedagogical Universities and Education Faculties

According to the regulations of the Ministry of Education and Training of Vietnam, teacher training curricula must be published on websites. Therefore, a survey was conducted at 14 pedagogical universities and faculties of education (Suppl. 2) to determine whether a bioinformatics course is included in the teacher training curriculum and whether it is mandatory or optional. The pedagogical universities and faculties of education surveyed for their biology teacher training programs include Thai Nguyen University of Education, Hanoi National University of Education, Hanoi Pedagogical University 2, Ho Chi Minh City University of Education, Hue University of Education, The University of Danang-University of Science and Education, An Giang University, Dong Thap University, Can Tho University, Vinh University, Quy Nhon University, Saigon University, Hong Duc University, and Hung Vuong University. The results of this question-type survey are calculated as the percentage of the curriculum with a mandatory bioinformatics course, an optional bioinformatics course, and no bioinformatics course.

3.5. Evaluation of Pedagogical Students after Graduation

After graduating from the Faculty of Biology at Thai Nguyen University of Education in Vietnam, pedagogical students' capacity to include bioinformatics into their instruction is evaluated. There are five assessment topics for 96 graduates, including 42 students from the 50th training course, 36 from the 51st training course and 18 from the 52nd training course. The assessments were carried out at two time points: before and after being fostered knowledge and skills to apply bioinformatics to teaching. The assessment results were used to calculate the mean and error of all graduates from three courses (see Table 3).

Table 3. Evaluation questions for students after graduation before and after receiving integrated training in bioinformatics.

Evaluation criteria (EC)	Variables	Expected answers
EC1	Select a topic suitable for the application of bioinformatics in teaching biology.	Proteins, nucleic acids, chemical components of cells, viruses, bacterial and viral diseases, etc.
EC2	Select bioinformatics software suitable for knowledge content.	Choose one of the appropriate software (BioEdit/or BLAST in NCBI/or MEGA).
EC3	Download data relevant to the selected topic from the biological database.	Can be nucleotide and amino acid sequences of the human genome and some species of organisms.
EC4	Use the bioinformatics software to analyse biological data.	Use one of the following software to analyze data: BioEdit/or BLAST in NCBI/or MEGA.
EC5	Create bioinformatics products.	2D or 3D model of a protein.

3.6. Analysis of Data

All the data were subjected to a one-way analysis of variance using " data analysis" in Excel and SPSS software (IBM Corp, 2021). Comparative statistical analysis between before and after being trained to integrate bioinformatics by Excel or Statistical Package for the Social Sciences (SPSS) at a 99% confidence interval of the difference.

4. RESULTS

4.1. Survey Results of High School Teachers

According to Table 4, most of the surveyed teachers are interested in bioinformatics (80.95%) but have yet to study the subject at the university (82.54%). Many teachers are not aware of the relationship between bioinformatics and STEM education (52.38%), the percentage of teachers who do not have an educational background in using bioinformatics tools accounted for 79.37%, and many have not attended the training course on bioinformatics (87.30%). The surveyed teachers aspire to be fostered in bioinformatics knowledge (88.90%) and wish to participate in updated course on integrating bioinformatics into teaching biology (90.47%).

Table 4. Survey results of teachers about the desire to foster knowledge of bioinformatics and apply bioinformatics to teaching in high schools.

Questions (SQ)	Variables	Yes (%)	No (%)	No answer (%)
SQ1	Are you interested in bioinformatics?	80.95	11.11	7.94
SQ2	Did you take a bioinformatics course at the university?	15.90	82.54	1.59
SQ3	Do you know the relationship between bioinformatics and STEM education in high school biology teaching?	23.81	52.38	23.81
SQ4	Does your academic background make teaching with bioinformatics tools possible?	14.29	79.37	6.35
SQ5	Have you attended a training course on bioinformatics?	9.50	87.30	3.17
SQ6	Do you wish to be fostered in bioinformatics knowledge?	88.90	9.52	1.59
SQ7	Do you wish to participate in a refresher course on integrating bioinformatics into biology teaching in high schools?	90.47	7.94	1.59

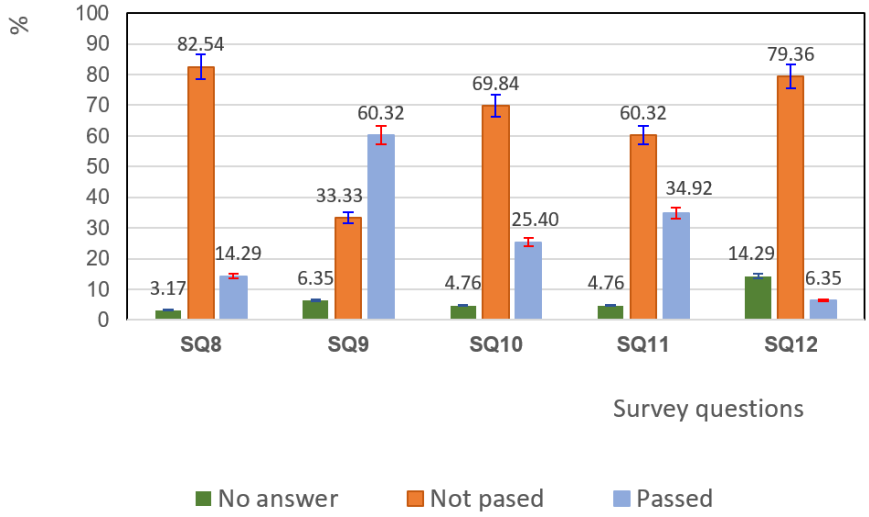

Figure 2. Survey results on teachers' understanding of bioinformatics and its application in teaching in high schools. SQ8, SQ9, SQ10, SQ11, and SQ12 are questions that require short answers (see Table 3).

The survey results for each question in Figure 2 were divided into the following three groups: the no-answer group, the passed-answer group, and the not-passed group. The non-response rates for questions SQ3, SQ4, SQ5, SQ6, and SQ7 were 3.17, 6.35, 4.76, 4.76, and 14.29 (%) respectively. The proportion of teachers with not-passed answers ranged from 33.33% to 82.54%. However, many teachers were able to name essential, commonly used bioinformatics software in biological research and learning (60.32%) and describe some of the significant biological data in the National Center for Biotechnology Information, USA (NCBI) database (34.92%).

The survey results also showed that many teachers need to learn the research object of bioinformatics and many teachers misunderstand bioinformatics as the application of information technology in teaching biology (82.54%). Most teachers surveyed did not know and had yet to use essential bioinformatics software (69.84%) especially since they could not identify knowledge topics that could integrate bioinformatics. Bioinformatics is an entirely new field for biology teachers in Vietnamese high schools. Therefore, most surveyed teachers have the desire and need to foster bioinformatics knowledge and apply bioinformatics in teaching.

4.2. Survey Results of Universities and Faculty of Biology Teacher Training for High Schools

The data in Table 5 depicts the results of a survey of teacher training curricula conducted through the websites of 14 pedagogical universities and education faculties in Vietnam (Suppl. 2). The results showed that bioinformatics had yet to be included in the training curriculum in most biology teacher training institutions. Only one pedagogical university has a compulsory bioinformatics course (7.14%), two other pedagogical universities have it as an elective (14.28%) and 11 teacher training institutions left in the curriculum need to have bioinformatics courses (78.57%).

Table 5. Survey results of teacher training curricula of 14 pedagogical universities and pedagogical faculties in Vietnam.

Variables	N	%
The curriculum has a mandatory bioinformatics course.	1	7.14
The curriculums have an optional bioinformatics course.	2	14.29
The curriculum has no bioinformatics course.	11	78.57

4.3. Proposed Principles and Steps of Bioinformatics Integration into Teaching in High Schools

Six basic principles and four steps of bioinformatics integration in teaching biology at high schools have been developed derived from the practice of teacher training programs of schools and faculties of education in Vietnam, the 2018 general education program requirements, survey results on high school teachers, and inheriting the previous research (see Table 6). These are the beginnings of applying bioinformatics to teaching suitable for high schools in Vietnam.

Table 6. Principles of bioinformatics integration in teaching biology at high schools.

Order	Principles of bioinformatics integration
1	Simple, familiar, and suitable for students.
2	Compatibility between bioinformatics and biological knowledge content.
3	The knowledge module is broken down for analysis.
4	Personalize and empower students.
5	Make an analytical model on paper before using the computer.
6	Create specific products- bioinformatic products.

Table 7. Steps to integrate bioinformatics into biology teaching.

Order	Steps to integrate bioinformatics into teaching
Step 1	Select knowledge topics suitable for bioinformatics applications.
Step 2	Divide the topic into small knowledge modules.
Step 3	Data analysis using bioinformatics.
Step 4	Create bioinformatic products (STEM products).

4.4. Examples

The knowledge topics of the biology curriculum in Vietnam's 2018 general education curriculum that can integrate bioinformatics into teaching such as topics related to DNA deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins, cell technology, gene technology, genetic and infectious diseases, evolution and phylogenetics (see Table 8).

Table 8. Some topics in the biology curriculum that can integrate bioinformatics into teaching.

Grade	Topics	Contents
10	Protein	Relationship between structure and function of protein.
10	Nucleic acid	Relationship between structure and function of DNA and RNA.
10	Chemical composition of cells	Membrane proteins, inhibitor proteins and enzymes.
10	Cell biotechnology	Stem cell culture.
10	Microbiological research methods	Identification by DNA barcoding. Establish a microbial phylogenetic tree based on DNA data.
10	Viruses	Compare gene sequences and identify virus variants. Establishing a phylogenetic tree of viral variants.
10	Bacterial and viral diseases	Infectious disease model.
11	Immunity	Vaccine technology and disease diagnosis model.
11	Transporting blood	Hemoglobin in red blood cells and transporting oxygen
12	Genes and transmission mechanisms genetic information	Structure of genes, mRNA and proteins.
12	Genome	Comparative genomics. Compare nucleotide sequences.
12	Proteome	Compare the amino acid sequence and 3D structure of the protein.
12	Inherited diseases	Genetic disease model
12	Molecular biology	Gene technology and recombinant DNA technology.
12	The modern theory of evolution	Molecular evolution and phylogenetic tree.

An example of a teaching organization model integrating bioinformatics into an infectious disease learning project in the biology 10 curriculum through the COVID-19 disease caused by the human SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) virus as described in Figure 3. Students experience using bioinformatics to search for information, download and compare nucleotide sequences and amino acid sequences, and set up and compare 3D models of antigenic proteins to see how structure matches function by teaching about "infectious diseases". Therefore, the organisation of teaching that integrates bioinformatics in disease models provides an opportunity for students to become familiar with bioinformatics tools and databases, creating a basis for students to conduct research. The creation and comparison of the spatial structure of proteins by bioinformatics can integrate the teaching and learning of biological macromolecules, genetic mutations and genetic diseases.

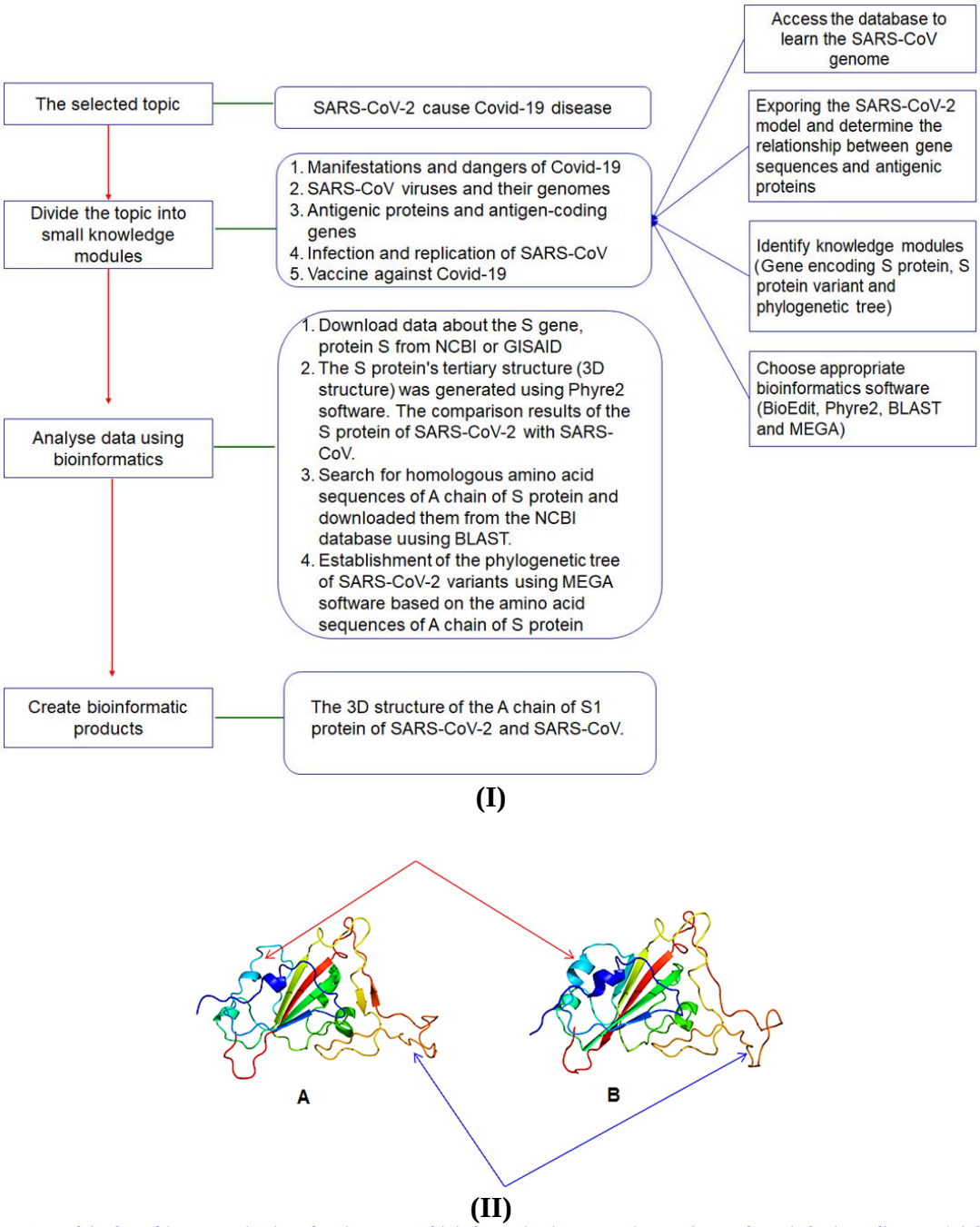

Figure 3. A model of teaching organisation that integrates bioinformatics into a project to learn about infectious diseases (Biology 10, Vietnam ministry of education and training).

Note: **(I):** The diagram of steps to integrate bioinformatics in teaching COVID19 disease.
(II): Differences in the 3D structure of the A chain of S1 protein between SARS-CoV-2 and SARS-CoV. A: SARS-CoV-2, chain A of protein S1 (194 aa); locus: 7JMW_A (https://www.ncbi.nlm.nih.gov/protein/7JMW_A); B: SARS-CoV, chain A of protein S1 (193 aa); locus: 3BGF_A; (https://www.ncbi.nlm.nih.gov/protein/3BGF_A). Arrows indicate different locations that can be observed. Phyre2 software was used to create the protein's 3D structure.

4.5. Evaluation of Pedagogical Students after Graduation

Three groups of graduate students from Thai Nguyen University of Education (Vietnam), a teacher training institution with bioinformatics in its curriculum were assessed at two points in time before and after implementing the training on integrating bioinformatics into teaching to test the application of the principles and steps of integrating bioinformatics into teaching. Table 3 of section 2.5 (methods) presents the questions and expected answers. Table 9 shows that if students only study the bioinformatics module, its application in teaching biology is minimal. The percentage of pedagogical university graduates who can apply bioinformatics to teaching in high schools is meagre which is 6.08% (EC1, EC5), 11.4% (EC4), 15.57% (EC2) and 17.46% (EC3). However, when

receiving bioinformatics integration training in teaching, the rate significantly increased by 676.64% for EC1, 280.35% for EC2, 296.22% for EC3, 445.91% for EC4 and 633.05% for EC5.

Table 9. Results of evaluating the effectiveness of integrated bioinformatics teaching of pedagogical university graduates.

Evaluation criteria (EC)	Before being trained to integrate bioinformatics (%) ($\overline{X} \pm S_{\overline{X}}$)	After being trained to integrate bioinformatics (%) ($\overline{X} \pm S_{\overline{X}}$)	Rate of increase compared to before being trained to integrate bioinformatics (%)
EC1	6.08 ± 0.52	41.14 ± 1.17	676.64
EC2	15.57 ± 1.09	43.65 ± 2.55	280.35
EC3	17.46 ± 0.79	51.72 ± 4.02	296.22
EC4	11.24 ± 1.72	50.12 ± 1.49	445.91
EC5	6.08 ± 0.52	38.49 ± 1.27	633.05

Note: The evaluation criteria designed in Table 3 of methods section 2; $\overline{X}$: Mean; $S_{\overline{X}}$: Standard error.

5. DISCUSSION

This study develops principles and procedures for integrating bioinformatics into teaching for university students of education as a preliminary study based on the analysis results of Vietnam's new general education curriculum, 2018, a survey of teachers' opinions, an investigation of the curriculum of teacher training institutions, and an assessment of pedagogical students after graduation. Advances in biology, laboratory techniques, new approaches and analyses have transformed scientific disciplines such as biology, biotechnology, agriculture, forestry, fishery, medicine and other aspects of society. However, this exciting new area of science is rarely integrated into classrooms or textbooks making high school biology education lag behind cutting-edge scientific discoveries. Establishing bioinformatics as an essential component of undergraduate life science education is in the early stages of developing a strategy for integrating bioinformatics into life science education (Dinsdale et al., 2015). The related studies include teaching genetics literacy in the post-genomic era (Stern & Kampourakis, 2017) and genomics education in the era of personal genomics (Whitley, Tueller, & Weber, 2020) identifying core bioinformatics competencies for life sciences education (Wilson Sayres, Hauser, Sierk, Robic, & Rosenwald, 2018). However, at this educational level, many barriers exist to integrating bioinformatics into life science education (Williams, Drew, Galindo-Gonzalez, Robic, & Dinsdale, 2019). Choosing an approach that integrates bioinformatics into teaching develops students' competencies helps them understand science, and arouses their interest and motivation to learn science (Machluf & Yarden, 2013). It is feasible to use bioinformatics tools in secondary school instruction with accessible online databases and tools to analyse biological data (Wefer & Sheppard, 2008). Practice shows that integrating bioinformatics into teaching in high schools is a practical approach to teaching biology. The 2018 Vietnam General Education Curriculum (Vietnam Ministry of Education and Training, 2018b) was implemented starting from the 2021-2022 school year. Bioinformatics is recommended to be applied as an effective tool in teaching biology in Vietnamese high schools. Bioinformatics is a new field of science not only for teacher training institutions for pedagogical students but also unfamiliar to teachers in high schools in Vietnam. Therefore, everything starts from surveys of pedagogical universities and faculties of education to high schools in Vietnam from pedagogical students to high school teachers. Information from the surveys is the basis for proposing principles and approaches to applying bioinformatics to teaching biology in high schools.

Bioinformatics is a multidisciplinary science that links biology, biomedicine, proteomics, genomics, engineering, mathematics and information technology (Luscombe et al., 2001) (Suppl. 1). However, bioinformatics is rarely integrated into high school teaching and learning curricula and the tool is almost unprepared for the younger generation of lifelong self-learners (Machluf & Yarden, 2013). This situation is also found in teacher training institutions in Vietnam. Our research shows that there are 11/14 teacher training institutions whose curricula do not have a bioinformatics course accounting for 78.57% (see Table 5). Thus, most pedagogical universities and

faculties of education surveyed in Vietnam do not offer bioinformatics subjects in their teacher training programs, so teachers trained in these institutions were not equipped with bioinformatics skills. However, in practice, many teachers can self-study and equip themselves with skills to apply information technology in teaching. Along with this practice, the difference in Vietnam's 2018 General Education Curriculum (Vietnam Ministry of Education and Training, 2018b) compared to previous curriculums is that for the first time, bioinformatics is included in the curriculum and required to be used as a tool in teaching and learning. During a conference in 2019, the introduction of bioinformatics into high schools was a hotly discussed topic. Martins, Fonseca, Lemos, Lencastre, and Tavares (2020) said high school teachers expressed interest and desire to learn more about their classes' bioinformatics and integration strategies. Although teachers emphasized the relevance of bioinformatics in biology education, most believe that their academic background needs to be improved to perform bioinformatics-based exercises confidently in their teaching work. The teachers suggested taking part in more training courses on bioinformatics in teaching. They acknowledged that schools need to be equipped with resources to integrate bioinformatics into teaching biology (Martins et al., 2020). 63 teachers in our study also expressed this interest. 88.90% said they intended to increase their understanding of bioinformatics and up to 90.47% said they would want to take updated course that incorporates bioinformatics into teaching biology (see Table 4). In this preliminary survey, 82.54% of respondents were ignorant of the topic of bioinformatics; 33.33% were unable to identify a key bioinformatics software; 69.84% were unable to identify the three major biological databases in North America, Europe, and Asia; 60.32% were unable to characterize some of the noteworthy biological data in the NCBI database and 79.36% of the teachers asked were unable to identify specific high school biology (see Figure 2). Thus, the issue is the necessity of promoting bioinformatics knowledge and incorporating it into the curriculum for pedagogical university students and high school teachers.

As a biology teacher-researcher, one can guide researchers-students in carrying out bioinformatics application activities according to scientific thinking approaches with different goals. It can be applied for comparing the DNA of organisms of varying degrees of complexity, searching for sequences of similarity to comparable 3D structures, and gathering genetic information about genetic and neurodegenerative diseases and health conditions such as sickle cell anaemia, Alzheimer's disease, etc. Bioinformatics significantly fosters students' literacy, interests, and attitudes towards genes, proteins and evolution (Kovarik et al., 2013) offering a promising avenue for STEM education. Presently, there are many online resources available in biological databases. However, only a tiny fraction of high school students are exposed to these materials because their integration into the science curriculum is limited. Established rules for developing the bioinformatics curriculum and teaching bioinformatics in higher secondary schools have been discussed (Form & Lewitter, 2011; Via et al., 2011). In our study, six principles and four steps for integrating bioinformatics and teaching were developed (see Tables 6 and 7). This integration process is simple and easy to apply in teacher training and for teacher training courses.

Students' knowledge selection, skill training and interests can only be achieved when they clearly define and experience their goals (Handelsman, Ebert-May, Beichner, Bruns, & Chang, 2004). Kovarik et al. (2013) noted that bioinformatics fosters students' literacy, interests and attitudes towards genes, proteins and evolution (Kovarik et al., 2013). When analyzing the content of Vietnam's 2018 general education curriculum, the biology subject, initially had 15 knowledge topics that could use bioinformatics in teaching organization, seven topics are in grade 10, two in grade 11 and six in grade 12 (see Table 8).

Recently in Vietnam, the topic of STEM education has received research attention including research on challenges for STEM education in high schools (Le, Tran, & Tran, 2021) research on developing the capacity of natural science teachers to design and organize STEM educational activities (Ngo et al., 2023) and a proposal for a STEM teaching competency framework of pre-chemistry service teachers (Nguyen, Dang, & Pham, 2020) etc. However, no research has been found on the relationship between bioinformatics and STEM education. Bioinformatics enables activities in high school classrooms in a new way in students' STEM education planning

pathway. It represents an approach to studying biological data in various fields and contexts and the linkages between biology, physics, mathematics, biomedicine and information technology based on ideas about the complexity of science. Hence, bioinformatics is a STEM science and bioinformatics integration is a new approach to promote STEM education in biology teaching. The databases and software used in bioinformatics can contribute to solving several challenges in biology education, which are i) students better understand abstract concepts such as proteins, genes, genomes and evolutionary relationships. ii) Coherence between DNA, proteins and traits, and many other topics in the biology curriculum. iii) In-depth understanding of current research methods (Machluf & Yarden, 2013). The human eye cannot see proteins and genes, so visualising these molecules requires expensive and modern equipment.

Furthermore, whether students better understand the structure, function and biological processes involved in proteins and genes remains to be seen. The computer-based approach is cheaper, more convenient, and more helpful. Using bioinformatics software to create 3D structures of proteins makes it possible to view a model of a specific protein from all angles. It is possible to zoom in, rotate the protein and effortlessly recognise the protein's function when comparing the protein of the normal gene with that of the mutant gene. In the example of a teaching organization model integrating bioinformatics into a learning project about the human disease COVID-19 as described in Figure 2. It is demonstrated that the results of bioinformatics integration have made abstract protein structures in living cells visible and intuitive. When using bioinformatics tools, the abstract concepts of genes and proteins become more tangible and are easier to understand.

For example, when defining and discussing the molecular nature of sickle cell anaemia in humans, it may be enough to look up the gene and amino acid sequence of hemoglobin (Hb) in a biological database and then show the 3D structure of the b chain in Hb. The difference in the spatial structure of the normal Hb protein and the protein synthesized by the mutant gene allows for the prediction of the functional change of the protein. It is easy for students to see that changing the spatial structure will change the protein's function. Suppl. 3 is a visual example of the difference between the 3D structure of the β-chain in Hb of a person who is not sick and a person with sickle cell disease. This model is meant to guide the students' approach to exploring knowledge of the same type. Therefore, this study has proposed six principles and four steps to integrate bioinformatics into teaching biological topics to create STEM products that are biomolecules in 3D model form (see Tables 6, 7 and Figure 3). The experiment of guiding pedagogical students in the direction of integrating bioinformatics into teaching and determining the relationship between bioinformatics and STEM education for three groups of graduates from Thai Nguyen University of Education (Vietnam) according to the principles and steps proposed above showed feasible results in all assessment criteria (see Table 9). In particular, in this study, when pedagogical students were trained in the knowledge and skills to integrate bioinformatics into teaching, the assessment criteria significantly increased by 676.64% for EC1, 280.35% for EC2, 296.22% for EC3, 445.91% for EC4, and 633.05% for EC5, respectively.

Potential challenges and limitations for biology education in Vietnam are as follows: 1) in high schools, informatics is often taught separately from other science subjects. 2) Bioinformatics has yet to be included in teacher training at the university. The importance of bioinformatics in biological teaching and research is not fully realized, leaving students unprepared for learning and studying. 3) Biology teachers in Vietnamese high schools have different bioinformatics skills and qualifications. 4) Information technology equipment and infrastructure (computers, broadband internet and smart devices) in high schools still need to be improved. Recognizing the importance of bioinformatics in education, a project for bringing bioinformatics into secondary schools was designed and implemented in Scotland (Bain et al., 2022). Recently, at a conference on research and teaching in Vietnam, our research team discussed some ideas about bringing bioinformatics into teaching in high schools. Scientists have agreed on the initial proposals and it is recommended that applications of bioinformatics be researched and widely deployed in Vietnamese high schools (Nguyen, Vu, Nguyen, Nguyen, & Pham, 2022).

6. CONCLUSION

Integrating bioinformatics into high school biology teaching has been confirmed to be necessary for Vietnam and developing countries based on theory and practice. Six essential principles and four steps for integrating bioinformatics into high school biology teaching have been developed. The six principles are as follows: simple, familiar, and suitable for students, compatibility between bioinformatics and biological knowledge content, the knowledge module is broken down for analysis, personalize and empower students, make an analytical model on paper before using the computer, create specific products- bioinformatics products. The four steps for integrating bioinformatics into teaching include selecting knowledge topics suitable for bioinformatics applications, dividing the topic into small knowledge modules, conducting data analysis using bioinformatics and creating bioinformatic products (STEM products). Fifteen knowledge topics in Vietnam's new general education 2018 curriculum, biology subjects that can integrate bioinformatics into their teaching have been proposed. A teaching organization model for integrating bioinformatics on the topic " infectious diseases" has been provided. The experimental results of evaluating pedagogical university graduates showed that when trained, the percentage of students who could apply the principles and processes of integrating bioinformatics into teaching increased significantly with 676.64% (EC1), 280.35% (EC2), 296.22% (EC3), 445.91% (EC4) and 633.05% (EC5). Bioinformatics is a STEM science, and integrating it into biology teaching is a new approach to promoting STEM education in teaching in Vietnam, contributing to the development of quality, competence and career orientation for students.

7. POLICY PROPOSALS

The principles and processes of integrating bioinformatics and teaching show potential for application in educational environments especially in biology teaching and self-study. Several policy recommendations can be proposed to improve bioinformatics' effectiveness and widespread adoption in educational policies and practices based on the results and discussions presented.

Integration into curriculum: Education policymakers should consider including bioinformatics modules in their curricula. Resources should be allocated to train and develop teachers in integrating bioinformatics in the classroom. In light of the findings of this study, teacher training institutions need to guide the design and organization of learning activities on bioinformatics and its application in teaching through teacher training and retraining courses that needs to be seriously considered.

Integration of bioinformatics in teaching relevant knowledge topics: High school teachers need to be equipped with the necessary knowledge and skills through training courses to integrate bioinformatics effectively into their teaching activities.

High schools must develop a curriculum, select topics that integrate bioinformatics into teaching, discuss teaching techniques to integrate bioinformatics, develop a lesson plan, and assess learners' capacity. Significantly developing resources, including human and material facilities, to meet the requirements of integrating bioinformatics into teaching in high schools are solutions that are essential for Vietnam and developing countries.

8. LIMITATIONS

This study has some limitations. First, the study used a relatively small sample of teachers which covered only some localities of Vietnam. Therefore, a larger group of biology teachers representing all localities of Vietnam should be included in future studies. Second, the study was only piloted for three courses with few students. In the future, it is necessary to conduct experiments with a larger group of pedagogical students and in many different training courses. Finally, the new general education program of Vietnam in 2018 has only been implemented in junior high and high school from the 2021-2022 school year and the 2024-2025 school year is the year to complete the implementation in grades 9 and 12, so future studies will conduct experiments integrating bioinformatics into teaching in all grades of junior high and high schools.

Funding: This research is supported by Thai Nguyen University of Education, Vietnam (Grant number: 3223-QD-DHSP).
Institutional Review Board Statement: The Ethical Committee of the Thai Nguyen Province, Vietnam has granted approval for this study on 15 December 2022 (Ref. No. 3141).
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

REFERENCES

Bain, S. A., Plaisier, H., Anderson, F., Cook, N., Crouch, K., Meagher, T. R., . . . Barker, D. (2022). Bringing bioinformatics to schools with the 4273pi project. *PLoS Computational Biology, 18*(1), e1009705. https://doi.org/10.1371/journal.pcbi.1009705

Casas, L., & Saborido-Rey, F. (2023). A review of genomics methods and bioinformatics tools for the analysis of close-kin mark-recapture. *Frontiers in Marine Science, 10*, 1113870. https://doi.org/10.3389/fmars.2023.1113870

Dinsdale, E., Elgin, S. C., Grandgenett, N., Morgan, W., Rosenwald, A., Tapprich, W., . . . Pauley, M. A. (2015). NIBLSE: A network for integrating bioinformatics into life sciences education. *CBE—Life Sciences Education, 14*(4), 1-4. https://doi.org/10.1187/cbe.15-06-0123

El Islami, R. A. Z., Sari, I. J., & Utari, E. (2023). Conceptualizing bioinformatics education in STEM literacy development for pre-service biology teachers. *International Journal of Advanced and Applied Sciences 10*(12), 193-202. https://doi.org/10.21833/ijaas.2023.12.021

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. *PLoS Computational Biology, 7*(10), e1002243. https://doi.org/10.1371/journal.pcbi.1002243

Gelbart, H., & Yarden, A. (2006). Learning genetics through an authentic research simulation in bioinformatics. *Journal of Biological Education, 40*(3), 107-112. https://doi.org/10.1080/00219266.2006.9656026

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., & Chang, A. (2004). Scientific teaching. *Science, 304*(5670), 521-522. https://doi.org/10.1126/science.1096022

IBM Corp. (2021). IBM SPSS statistics for windows, version 28.0. Armonk, NY: IBM Corp. Kelley, T.R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(11), 1–11.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*, 1-11. https://doi.org/10.1186/s40594-016-0046-z

Kovarik, D. N., Patterson, D. G., Cohen, C., Sanders, E. A., Peterson, K. A., Porter, S. G., & Chowning, J. T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE—Life Sciences Education, 12*(3), 441-459. https://doi.org/10.1187/cbe.12-11-0193

Le, L. T. B., Tran, T. T., & Tran, N. H. (2021). Challenges to STEM education in Vietnamese high school contexts. *Heliyon, 7*(12), e08649.

Lewitter, F., & Bourne, P. E. (2011). Teaching bioinformatics at the secondary school level. *PLoS Computational Biology, 7*(10), e1002242. https://doi.org/10.1371/journal.pcbi.1002242

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine, 40*(04), 346-358. https://doi.org/10.1055/s-0038-1634431

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: Design principles and implications. *Briefings in Bioinformatics, 14*(5), 648-660. https://doi.org/10.1093/bib/bbt030

Martins, A., Fonseca, M. J., Lemos, M., Lencastre, L., & Tavares, F. (2020). Bioinformatics-based activities in high school: Fostering students' literacy, interest, and attitudes on gene regulation, genomics, and evolution. *Frontiers in Microbiology, 11*, 578099. https://doi.org/10.3389/fmicb.2020.578099

Ngo, H. D., Vu, T. T. H., Pham, N. S., Pham, T. N., Huynh, T. P. T., & Nguyen, M. D. (2023). Developing competencies of natural science teachers for designing and organizing STEM education activities in Vietnam. *International Journal of Education and Practice, 11*(3), 500-514. https://doi.org/10.18488/61.v11i3.3430

Nguyen, T. H., Vu, T. T. T., Nguyen, H. Q., Nguyen, H. H., & Pham, T. T. N. (2022). *Integrating bioinformatics in teaching biology in vietnamese high schools: Some initial proposals.* Paper presented at the Proceeding of the 5th National Scientific Conference on Biological Research and Teaching in Vietnam, 1006-1115.

Nguyen, T. T. T., Dang, T. O., & Pham, T. B. (2020). Proposing STEM teaching competency framework of pre-chemistry service teachers. *Journal of Science Educational Science, 65*(4), 177–184.

Stern, F., & Kampourakis, K. (2017). Teaching for genetics literacy in the post-genomic Era. *Studies in Science Education, 53*(2), 193–225. https://doi.org/10.1080/03057267.2017.1392731

Tan, Y. C., Kumar, A. U., Wong, Y. P., & Ling, A. P. K. (2022). Bioinformatics approaches and applications in plant biotechnology. *Journal of Genetic Engineering and Biotechnology, 20*(1), 106. https://doi.org/10.1186/s43141-022-00394-5

Via, A., De Las Rivas, J., Attwood, T. K., Landsman, D., Brazas, M. D., Leunissen, J. A., . . . Schneider, M. V. (2011). Ten simple rules for developing a short bioinformatics training course. *PLoS Computational Biology, 7*(10), e1002245. https://doi.org/10.1371/journal.pcbi.1002245

Vietnam Ministry of Education and Training. (2018a). *General education curriculum- Biology curriculum issued together with* Circular No. 32/2018/TT-BGDĐT.

Vietnam Ministry of Education and Training. (2018b). *General education curriculum (issued together with Circular No. 32/2018/TT-BGDĐT, dated December 26, 2018).* Retrieved from https://data.moet.gov.vn/index.php/s/LETzPhj5sGGnDii#pdfviewer. [Accessed August 2022]

Wefer, S. H., & Sheppard, K. (2008). Bioinformatics in high school biology curricula: A study of state science standards. *CBE—Life Sciences Education, 7*(1), 155-162. https://doi.org/10.1187/cbe.07-05-0026

Whitley, K., Tueller, J. A., & Weber, K. S. (2020). Genomics education in the era of personal genomics: Academic, professional, and public considerations. *International Journal of Molecular Sciences, 21*(3), 768. https://doi.org/10.3390/ijms21030768

Williams, J. J., Drew, J. C., Galindo-Gonzalez, S., Robic, S., & Dinsdale, E. (2019). Barriers to integration of bioinformatics into undergraduate life sciences education: A national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLOS One, 14*(11), e0224288. https://doi.org/10.1371/journal.pone.0224288

Wilson Sayres, M. A., Hauser, C., Sierk, M., Robic, S., & Rosenwald, A. G. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLOS One, 13*(6), e0196878. https://doi.org/10.1371/journal.pone.0196878