APPROVED
Head of the Phystech School of Biological and Medical Physics
D.V. Kuzmin

Work program of the course (training module)

course:	Analysis of OMICS Data/Анализ омиксных данных
major:	Applied Mathematics and Physics
specialization:	Applied Bioinformatics/Прикладная биоинформатика
	Phystech School of Biological and Medical Physics
	Chair of Bioinformatics and Systems Biology
term:	1
qualification:	Master

Semester, form of interim assessment: 2 (spring) - Exam

Academic hours: 30 AH in total, including:
    lectures: 0 AH.
    seminars: 30 AH.
    laboratory practical: 0 AH.

Independent work: 30 AH.

Exam preparation: 30 AH.

In total: 90 AH, credits in total: 2

Authors of the program:

    I.V. Antonov, phd (candidate of biological sciences)
    Y.A. Medvedeva, candidate of biological sciences

The program was discussed at the Chair of Bioinformatics and Systems Biology 04.06.2020

Annotation

The aim of the course is hands on experience in applying bioinformatics methods for analyzing and interpreting biological data.

1. Study objective

Purpose of the course

Hands on experience in applying bioinformatics methods for analyzing and interpreting biological data.

Tasks of the course
- give an idea of the basic mechanisms of gene transcription regulation;
- make students familiar with modern understanding of regulatory transcriptome and epigenomics;
- teach how to use the main bioinformatics tools, programming languages and databases;
- introduce basic algorithms and data formats.

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
UC-2 Able to manage the project through all stages of implementation	UC-2.1 Set an objective within a defined scientific problem; formulate the agenda, relevance, significance (scientific, practical, methodological, or other depending on the project type), forecast the expected results and possible areas of their application
	UC-2.2 Forecast the project outcomes, plan necessary steps to achieve the outcomes, chart the project schedule and monitoring plan
	UC-2.3 Organize and coordinate the work of project stakeholders, provide the team with necessary resources
	UC-2.4 Publicly present the project results (or results of its stages) via reports, articles, presentations at scientific conferences, seminars, and similar events
UC-3 Organize and manage a team, and develop the team strategy to achieve the objectives	UC-3.1 Organize and coordinate the work of the project stakeholders and help resolve disputes and conflicts
	UC-3.2 Consider the interests, specific behavior, and diversity of opinions of team members/colleagues/counterparties
	UC-3.3 Foresee the results (consequences) of both individual and collective actions
	UC-3.4 Plan teamwork, distribute tasks to team members, hold discussions of different ideas and opinions
UC-4 Use modern communication tools in the academic and professional fields, including those in a foreign language	UC-4.1 Exchange business information in oral and written forms in Russian and at least one foreign language
	UC-4.2 Use the acquired skills to write, translate, and edit various academic texts (abstracts, essays, reviews, articles, etc.)
	UC-4.3 Present the results of academic and professional activities at various academic events, including international conferences
	UC-4.4 Use modern ICT tools for academic and professional collaboration
UC-6 Determine priorities and ways to improve performance through self assessment	UC-6.1 Achieve personal growth and professional development, determine priorities and ways to improve performance

performance through self-assessment	UC-6.2 Evaluate performance results in correlation with the set objectives and applied methods
Pro.C-3 Use research and testing equipment (devices and installations, specialized software) in a selected subject field	Pro.C-3.1 Understand the operating principles of the equipment and specialized software
	Pro.C-3.3 Evaluate the accuracy of the experimental (numerical) results
	Pro.C-3.2 Conduct an experiment (simulation) using research equipment (software)

3. List of the planned results of the course (training module)

As a result of studying the course the student should:

know:
- physical and chemical properties of nucleic acids and chromatin proteins;
- processes involving nucleic acids and chromatin proteins;
- research methods of the genome and evolution of living organisms;
- the main mechanisms of regulatory genomics, transcriptome and epigenetics;
- basic bioinformatics algorithms, databases, file formats used in transcriptome and epigenomics;
- modern molecular biological methods for studying cell processes;

be able to:
- use the main databases of regulatory genomics;
- run bioinformatics tools in UNIX environment and program in python/R;
- analyze large scale datasets.

master:
- of working with large volumes of biological data;
- of planning and implementation of multi-stage bioinformatic analysis.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Introduction to bioinformatics tools		2		2
2	Biopython and R		4		2
3	Bacterial genome annotation		2		2
4	Systems biology		2		2
5	Molecular evolution		2		2
6	Non-coding genome		2		4
7	RNA-seq		4		4
8	DNA methylation		4		4
9	ChIP-seq		4		4
10	DNA accessibility		4		4
AH in total			30		30
Exam preparation		30 AH.			
Total complexity		90 AH., credits in total 2			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 2 (Spring)

1. Introduction to bioinformatics tools

The main UNIX commands, installation of tools, approaches to genome assembly from NGS data

2. Biopython and R

GenBank file format, Download sequences from NCBI, parse genome annotation in GenBank format

3. Bacterial genome annotation

Predict the protein coding genes with GeneMark, compare the de novo prediction with the GenBank annotation

4. Systems biology

KEGG database. Predict bacterial phenotype from its genes.

5. Molecular evolution

Comparative genomics - Build 16S rRNA gene tree

6. Non-coding genome

RepeatMasker, Composition of the genomes of different organisms (Bacteria, Archaea, Yeast, Drosophila, Arabidopsis, mouse, human)
Regulatory elements in the genomes of higher organisms -- promoter, enhancer, insulator, TADs and Hi-C experiments.

7. RNA-seq

De novo transcript assembly, splicing. Differentially expressed genes, DESeq2 R package.

8. DNA methylation

DNA de-methylation during embryogenesis, differentially methylated regions (DMRs)

9. ChIP-seq

ChIP-seq for histone marks: NGS.plot & ChromHMM, ChIP-seq for transcription factors, sequence logo.

10. DNA accessibility

DNAse-seq and ATAC-seq

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

Equipment needed for lectures and seminars: computer and multimedia eqipment (projector, sound system)

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature

Provided at the department

1. Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge university press., ISBN: 978-0521629713

2. Borodovsky, M., & Ekisheva, S. (2006). Problems and solutions in biological sequence analysis. Cambridge University Press. ISBN: 978-0521612302

Additional literature

Provided at the department

Рэндал Л. Шварц, Том Феникс, Брайан Д Фой. Изучаем Perl. Символ-Плюс, 2009

Р. Дурбин, Ш. Эдди, А. Крог, Г. Митчисон. Анализ биологических последовательностей. Регулярная и хаотическая динамика, 2006. Марк Бородовский, Светлана Екишева. Задачи и решения по анализу биологических последовательностей. Регулярная и хаотическая динамика, 2008

7. List of web resources that are necessary for the course (training module) mastering

Scientific bibliographic and patent databases in the field of physico-chemical biology, available on the Internet in free mode - Science Citation Index (Web of Science), Medline (PubMed), Scientific Electronic Library (NEB), Russian Patent DB of FGU FIPS and American USPAFULL patent database; email addresses of major scientific publishers who provide access to the full text of current and archival issues of these journals.

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

Zoom is required for some of the lessons. Google Drive to access course materials. The presence of smartphones / laptops during classes is encouraged to participate in interactive exercises.

9. Guidelines for students to master the course

A student who studies discipline must, on the one hand, master a general conceptual apparatus, and on the other hand, must learn to apply theoretical knowledge in practice.
As a result of studying the discipline, the student should know the basic definitions of the discipline, be able to apply this knowledge to solve various problems.

Successful learning requires:
- visits to all classes provided by the curriculum for the discipline;
- conducting the abstract of occupations;
- intense independent work of the student.

Independent work includes:
- reading recommended literature;
- study of educational material, preparation of answers to questions intended for self-study;
- solving problems offered to students in the classroom;
- preparation for performance of tasks of the current and intermediate certification.

An indicator of possession of the material is the ability to answer questions on discipline topics without an outline.

It is important to achieve an understanding of the material being studied, and not its mechanical memorization. If it is difficult to study individual
topics, questions, you should seek advice from the teacher.

Intermediate control of students' knowledge in the form of problem solving in accordance with the subject of classes is possible

Assessment funds for course (training module)

major:	Applied Mathematics and Physics
specialization:	Applied Bioinformatics/Прикладная биоинформатика
	Phystech School of Biological and Medical Physics
	Chair of Bioinformatics and Systems Biology
term:	1
qualification:	Master

Semester, form of interim assessment: 2 (spring) - Exam

Authors:

I.V. Antonov, phd (candidate of biological sciences)

Y.A. Medvedeva, candidate of biological sciences

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
UC-2 Able to manage the project through all stages of implementation	UC-2.1 Set an objective within a defined scientific problem; formulate the agenda, relevance, significance (scientific, practical, methodological, or other depending on the project type), forecast the expected results and possible areas of their application
	UC-2.2 Forecast the project outcomes, plan necessary steps to achieve the outcomes, chart the project schedule and monitoring plan
	UC-2.3 Organize and coordinate the work of project stakeholders, provide the team with necessary resources
	UC-2.4 Publicly present the project results (or results of its stages) via reports, articles, presentations at scientific conferences, seminars, and similar events
UC-3 Organize and manage a team, and develop the team strategy to achieve the objectives	UC-3.1 Organize and coordinate the work of the project stakeholders and help resolve disputes and conflicts
	UC-3.2 Consider the interests, specific behavior, and diversity of opinions of team members/colleagues/counterparties
	UC-3.3 Foresee the results (consequences) of both individual and collective actions
	UC-3.4 Plan teamwork, distribute tasks to team members, hold discussions of different ideas and opinions
UC-4 Use modern communication tools in the academic and professional fields, including those in a foreign language	UC-4.1 Exchange business information in oral and written forms in Russian and at least one foreign language
	UC-4.2 Use the acquired skills to write, translate, and edit various academic texts (abstracts, essays, reviews, articles, etc.)
	UC-4.3 Present the results of academic and professional activities at various academic events, including international conferences
	UC-4.4 Use modern ICT tools for academic and professional collaboration
UC-6 Determine priorities and ways to improve performance through self-assessment	UC-6.1 Achieve personal growth and professional development, determine priorities and ways to improve performance
	UC-6.2 Evaluate performance results in correlation with the set objectives and applied methods
Pro.C-3 Use research and testing equipment (devices and installations, specialized software) in a selected subject field	Pro.C-3.1 Understand the operating principles of the equipment and specialized software
	Pro.C-3.3 Evaluate the accuracy of the experimental (numerical) results
	Pro.C-3.2 Conduct an experiment (simulation) using research equipment (software)

2. Competency assessment indicators

As a result of studying the course the student should:

know:

- physical and chemical properties of nucleic acids and chromatin proteins;
- processes involving nucleic acids and chromatin proteins;
- research methods of the genome and evolution of living organisms;
- the main mechanisms of regulatory genomics, transcriptome and epigenetics;
- basic bioinformatics algorithms, databases, file formats used in transcriptome and epigenomics;
- modern molecular biological methods for studying cell processes;

be able to:
- use the main databases of regulatory genomics;
- run bioinformatics tools in UNIX environment and program in python/R;
- analyze large scale datasets.

master:
- of working with large volumes of biological data;
- of planning and implementation of multi-stage bioinformatic analysis.

3. List of typical control tasks used to evaluate knowledge and skills

During the current control, the student should be able to answer the following questions:
1. Basic UNIX Operating System Commands
4. Basic concepts of molecular biology
5. Smith-Waterman alignment algorithm
6. Needleman-Wunsch alignment algorithm
7. Approaches to the prediction of genes in the genome. The difference between eukaryotic and prokaryotic genes.
8. Methods of phylogenetic analysis.
9. Bioinformatics databases, UCSC Genome Browser
10. Methods for studying the genome and gene expression.
11. Methods for determining the primary structure of nucleic acids.
12. Polymerase chain reaction (PCR).
13. Replication, transcription, translation.
2. Basics of the Python programming language
3. Regular expressions
14. Levels of regulation of gene expression. Systems biology.
15. Regulation of translation.
16. The genetic code. The codon frequencies.
17. Approaches to the analysis of large biological data.

During the class, interactive discussions can take place in the course chats, which will be homework. It is possible to perform patent search as an independent task. Successful completion of all tasks in the course and the completion of knowledge control slices gives an advantage in the exam.

4. Evaluation criteria

1. Basic UNIX Operating System Commands
4. Basic concepts of molecular biology
5. Smith-Waterman alignment algorithm
6. Needleman-Wunsch alignment algorithm
7. Approaches to the prediction of genes in the genome. The difference between eukaryotic and prokaryotic genes.
8. Methods of phylogenetic analysis.
9. Bioinformatics databases, UCSC Genome Browser
10. Methods for studying the genome and gene expression.
11. Methods for determining the primary structure of nucleic acids.
12. Polymerase chain reaction (PCR).
13. Replication, transcription, translation.
2. Basics of the Python programming language

3. Regular expressions
14. Levels of regulation of gene expression. Systems biology.
15. Regulation of translation.
16. The genetic code. The codon frequencies.
17. Approaches to the analysis of large biological data.

Examples of exam tasks:
1. Methods of phylogenetic analysis.
2. Bioinformatics databases, UCSC Genome Browser
3. Methods for studying the genome and gene expression.
4. Methods for determining the primary structure of nucleic acids.
5. Polymerase chain reaction (PCR).

Can be left unchanged

The mark is excellent (10 points) - it is given to a student who has shown comprehensive, systematic, deep knowledge of the curriculum of the discipline, who has an interest in this subject area, has demonstrated the ability to confidently and creatively put them into practice in solving specific problems, and a free and proper substantiation of decisions.

The mark is excellent (9 points) - it is given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently put them into practice in solving specific problems, free and proper substantiation of the decisions made.

The mark is excellent (8 points) - given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently apply them in practice in solving specific problems, correct justification of decisions made, with some shortcomings.

A mark is good (7 points) - it is put up for a student, if he knows the material firmly, sets it up competently and in essence, knows how to apply the knowledge gained in practice, but does not competently substantiate the results obtained.

Evaluation is good (6 points) - it is put up to a student, if he knows the material firmly, sets it up correctly and in essence, knows how to apply this knowledge in practice, but admits some inaccuracies in the answer or in solving problems.

A mark is good (5 points) - it is given to a student, if he basically knows the material, correctly and essentially sets it out, knows how to apply this knowledge in practice, but allows a sufficiently large number of inaccuracies to answer or solve problems.

Grade satisfactorily (4 points) is given to a student who has shown the fragmented, fragmented nature of knowledge, insufficiently correct formulations of basic concepts, violations of the logical sequence in the presentation of program material, but at the same time he has mastered the main sections of the curriculum necessary for further education and can apply knowledge is modeled in a standard situation.

Grade satisfactorily (3 points) - given to a student who showed the fragmented, scattered nature of knowledge, making mistakes in formulating basic concepts, disrupting the logical sequence in presenting program material, poorly masters the main sections of the curriculum required for further education and even applies the knowledge gained in a standard situation.

The rating is unsatisfactory (2 points) - is given to a student who does not know most of the main content of the curriculum of the discipline, makes gross mistakes in the wording of the basic principles and does not know how to use this knowledge when solving typical tasks.

Unsatisfactory mark (1 point) - is given to a student who does not know the main content of the discipline's curriculum, makes gross errors in the wording of the basic concepts of the discipline and does not have any skills to solve typical practical problems.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

During the examination proccedure (or another type of control), the student is given 30 minutes to prepare. The interview for a student in an oral exam must not exceed one astronomical hour.