**Federal State Autonomous Educational Institution of Higher Education "Moscow Institute of Physics and Technology (National Research University)"**

**Work program of the course (training module)**

| | |
|---|---|
| **course:** | Bioinformatics&System Biology/Биоинформатика и системная биология |
| **major:** | Biotechnology |
| **specialization:** | Biomedical Engineering/Биомедицинская инженерия |
| | Phystech School of Biological and Medical Physics |
| | Chair of Bioinformatics and Systems Biology |
| **term:** | 4 |
| **qualification:** | Bachelor |

Semester, form of interim assessment: 7 (fall) - Grading test

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

Independent work: 15 AH.

In total:  45 AH, credits in total: 1

Author of the program:      V.Y. Makeev, doctor of physics and mathematical sciences

The program was discussed at the Chair of Bioinformatics and Systems Biology 22.06.2023

<h1 style="text-align:center">Annotation</h1>

After completing the course, the student will understand fundamental concepts, laws, theories of post-genomic biology, the tasks of bioinformatics analysis and its relationship with other sciences, the principles of modern databases on the structure of genomes, proteins and other biological information. This course will teach students the practical skills of data analysis from proteomic and genomic experiments, create computer programs used in bioengineering and bioinformatics.

## 1. Study objective

**Purpose of the course**

    acquisition of practical skills in data analysis of proteomic and genomic experiments to build system models of biological processes.

**Tasks of the course**

• mastering by students of basic means of analysis of genomic, structural and other biological information;

• application of bioinformatics methods to obtain new knowledge in the field of living systems;

• providing advice and assistance to students in conducting their own theoretical and experimental research in the field of bioinformatics.

## 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-1 Search and identify, critically assess, and synthesize information, apply a systematic approach to problem-solving | UC-1.1 Analyze problems, highlight the stages of their solution, plan the actions required to solve them |
| | UC-1.2 Find, critically assess, and select information required for the task in hand |
| | UC-1.3 Consider various options for solving a problem, assess the advantages and disadvantages of each option |
| | UC-1.4 Make competent judgments and estimates supported by logic and reasoning |
| | UC-1.5 Identify and evaluate practical consequences of possible solutions to a problem |
| Gen.Pro.C-1 Apply knowledge of mathematical, physical, chemical, biological laws, patterns, and interrelation to study, analyze, and utilize biological objects and processes | Gen.Pro.C-1.1 Analyze the task in hand, outline the ways to complete it |
| | Gen.Pro.C-1.2 Build mathematical models, make quantitative measurements and estimates |
| | Gen.Pro.C-1.3 Determine the applicability limits of the obtained results |
| Gen.Pro.C-2 Use modern IT and software tools to perform professional tasks in compliance with information security requirements | Gen.Pro.C-2.1 Apply modern computing tools and Internet services in professional settings |
| | Gen.Pro.C-2.2 Apply numerical mathematical methods and use software applications for scientific problem-solving in professional settings |
| | Gen.Pro.C-2.3 Fulfill basic information security requirements |
| Pro.C-2 Analyze research data and make scientific conclusions | Pro.C-2.1 Adopt methods of statistical process and scientific data analysis |
| | Pro.C-2.2 Define key parameters of the studied phenomenon and make relevant numerical estimates |
| | Pro.C-2.3 Make scientific claims with supporting evidence for a professional audience in verbal and written form, state scientific problems and propose solutions |
| | Pro.C-3.1 Apply functional principles and operating ranges of scientific equipment |

| Pro.C-3 Select the necessary devices, tools, and research methods for problem-solving in a selected subject area | Pro.C-3.2 Apply theory to evaluate the accuracy of analytical calculations |
|---|---|
| | Pro.C-3.3 Estimate the accuracy of numerical methods used on a computer, learn the computational complexity of the applied algorithms and the number of required computing resources |

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

As a result of studying the course the student should:

know:

- fundamental'nyye ponyatiya, zakony, teorii postgenomnoy biologii;
- zadachi bioinformaticheskogo analiza i yego svyaz' s drugimi naukami;
- printsipy raboty sovremennykh baz dannykh po strukture genomov, belkov i drugoy biologicheskoy informatsii;
- fundamental concepts, laws, theories of post-genomic biology;
- tasks of bioinformatics analysis and its relationship with other sciences;
- operating principles of modern databases on the structure of genomes, proteins and other biological information.

be able to:

- to abstract from the irrelevant when modeling real biological processes;
- use your knowledge to solve fundamental and applied problems of post-genomic biology;
- create computer programs used in bioengineering and bioinformatics, and independently master new resources (databases and programs) and experimental methods;
- make qualitative conclusions when going to the limiting conditions in the studied problems;
- master new subject areas, theoretical approaches and experimental methods;
- to determine the relevance of goals and objectives and the practical significance of the research;
- to analyze the results and methodological experience of research in relation to the general fundamental problem in the chosen area;
- work on modern, including unique, computing equipment;
- effectively use information technology and computer technology to achieve the necessary theoretical and applied results.

master:

- skills of mastering a large amount of information;
- skills of independent work on the Internet;
- culture of modeling biological tasks;
- skills of competent processing of experience results and comparison with theoretical data;
- practice of research and solution of theoretical and applied problems of molecular medicine;
- the skills of theoretical analysis of the problems of genomics, transcriptomics, proteomics and metabolomics associated with the study of the properties of biological systems at the molecular and subcellular levels of structural organization.

## 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 | Assembly of genomes de novo | | 3 | | 2 |
| 2 | Bioinformatic analysis of mass spectrometric information in proteomics | | 3 | | 1 |
| 3 | Introduction to Bioinformatics | | 3 | | 1 |
| 4 | Visualization of experimental data in post-genomic biology | | 3 | | 1 |

| 5 | Genomics | | 3 | | 1 |
|---|---|---|---|---|---|
| 6 | Genomic mapping | | 3 | | 1 |
| 7 | Transcriptomic data processing | | 3 | | 2 |
| 8 | Presentation of genomic information | | 2 | | 1 |
| 9 | Proteomics | | 3 | | 2 |
| 10 | Statistical analysis of genomic, proteomic and transcriptome data | | 2 | | 1 |
| 11 | Technologies for reading biological texts | | 2 | | 2 |
| AH in total | | | 30 | | 15 |

| Exam preparation | 0 AH. |
|---|---|
| Total complexity | 45 AH., credits in total 1 |

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

Semester: 7 (Fall)

1. Assembly of genomes de novo

Assembler programs. Difference in assembly algorithms for short and long DNA reads.

2. Bioinformatic analysis of mass spectrometric information in proteomics

Comparison of genetic and protein sequences. Alignment methods: pair and multiple, local and global. Needleman-Wunsh global alignment algorithm. Smith-Waterman local alignment algorithm. Gibbs sampling.

3. Introduction to Bioinformatics

The history of the development of computer processing of biological data. The definition of bioinformatics. Basic concepts. General understanding of the tasks of bioinformatics analysis and its relationship with other sciences. Areas of use.

4. Visualization of experimental data in post-genomic biology

Visualization of biological experiment data. Visualization software.

5. Genomics

Definition of genomics. Structural and functional genomics. Common sequencing technologies and result formats. Software: Bowtie, samtools, MUMmer.

6. Genomic mapping

Single nucleotide polymorphisms and methods for their detection. Search for genomic translocations. Search for repetitions, complementarities and symmetries in sequences.

7. Transcriptomic data processing

Methods for recognizing promoters. Build sequences. Mapping sites of transcription start.

8. Presentation of genomic information

Sequencing Result Formats. Recognition of structural and functional motives in genetic texts. Consensus concept, weight matrix. Assessment of recognition accuracy.

9. Proteomics

Algorithms for identification of proteins by mass spectra. Proteomic analysis software packages.

10. Statistical analysis of genomic, proteomic and transcriptome data

Statistical analysis of genomic, proteomic and transcriptome data.

11. Technologies for reading biological texts

Methods for determining the DNA sequence. New generation sequencers. High throughput DNA sequencing approaches.

## 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. A. M. Campbell. Discovering Genomics, Proteomics and Bioinformatics. - 2002. - ISBN-10: 0805347224
2. M. Ridley. The genome: an autobiography of a species in 23 chapters. - Eksmo, 2008. - ISBN 978-5-699-30682-4

Additional literature
Provided at the department:
1. J.M. Claverie, C. Notredame: Bioinformatics for Dummies. - 2006. - ISBN: 978-0-470-08985-9

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

Not used

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

For some of the lessons, you will need Zoom. 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:** | Biotechnology |
| **specialization:** | Biomedical Engineering/Биомедицинская инженерия |
| | Phystech School of Biological and Medical Physics |
| | Chair of Bioinformatics and Systems Biology |
| **term:** | 4 |
| **qualification:** | Bachelor |

Semester, form of interim assessment: 7 (fall) - Grading test

**Author:**          V.Y. Makeev, doctor of physics and mathematical sciences

# 1. Competencies formed during the process of studying the course

| Code and the name of the competence | Competency indicators |
|---|---|
| UC-1 Search and identify, critically assess, and synthesize information, apply a systematic approach to problem-solving | UC-1.1 Analyze problems, highlight the stages of their solution, plan the actions required to solve them |
| | UC-1.2 Find, critically assess, and select information required for the task in hand |
| | UC-1.3 Consider various options for solving a problem, assess the advantages and disadvantages of each option |
| | UC-1.4 Make competent judgments and estimates supported by logic and reasoning |
| | UC-1.5 Identify and evaluate practical consequences of possible solutions to a problem |
| Gen.Pro.C-1 Apply knowledge of mathematical, physical, chemical, biological laws, patterns, and interrelation to study, analyze, and utilize biological objects and processes | Gen.Pro.C-1.1 Analyze the task in hand, outline the ways to complete it |
| | Gen.Pro.C-1.2 Build mathematical models, make quantitative measurements and estimates |
| | Gen.Pro.C-1.3 Determine the applicability limits of the obtained results |
| Gen.Pro.C-2 Use modern IT and software tools to perform professional tasks in compliance with information security requirements | Gen.Pro.C-2.1 Apply modern computing tools and Internet services in professional settings |
| | Gen.Pro.C-2.2 Apply numerical mathematical methods and use software applications for scientific problem-solving in professional settings |
| | Gen.Pro.C-2.3 Fulfill basic information security requirements |
| Pro.C-2 Analyze research data and make scientific conclusions | Pro.C-2.1 Adopt methods of statistical process and scientific data analysis |
| | Pro.C-2.2 Define key parameters of the studied phenomenon and make relevant numerical estimates |
| | Pro.C-2.3 Make scientific claims with supporting evidence for a professional audience in verbal and written form, state scientific problems and propose solutions |
| Pro.C-3 Select the necessary devices, tools, and research methods for problem-solving in a selected subject area | Pro.C-3.1 Apply functional principles and operating ranges of scientific equipment |
| | Pro.C-3.2 Apply theory to evaluate the accuracy of analytical calculations |
| | Pro.C-3.3 Estimate the accuracy of numerical methods used on a computer, learn the computational complexity of the applied algorithms and the number of required computing resources |

## 2. Competency assessment indicators

As a result of studying the course the student should:

**know:**
- fundamental'nyye ponyatiya, zakony, teorii postgenomnoy biologii;
- zadachi bioinformaticheskogo analiza i yego svyaz' s drugimi naukami;
- printsipy raboty sovremennykh baz dannykh po strukture genomov, belkov i drugoy biologicheskoy informatsii;
- fundamental concepts, laws, theories of post-genomic biology;
- tasks of bioinformatics analysis and its relationship with other sciences;
- operating principles of modern databases on the structure of genomes, proteins and other biological information.

**be able to:**

- to abstract from the irrelevant when modeling real biological processes;
- use your knowledge to solve fundamental and applied problems of post-genomic biology;
- create computer programs used in bioengineering and bioinformatics, and independently master new resources (databases and programs) and experimental methods;
- make qualitative conclusions when going to the limiting conditions in the studied problems;
- master new subject areas, theoretical approaches and experimental methods;
- to determine the relevance of goals and objectives and the practical significance of the research;
- to analyze the results and methodological experience of research in relation to the general fundamental problem in the chosen area;
- work on modern, including unique, computing equipment;
- effectively use information technology and computer technology to achieve the necessary theoretical and applied results.

**master:**
- skills of mastering a large amount of information;
- skills of independent work on the Internet;
- culture of modeling biological tasks;
- skills of competent processing of experience results and comparison with theoretical data;
- practice of research and solution of theoretical and applied problems of molecular medicine;
- the skills of theoretical analysis of the problems of genomics, transcriptomics, proteomics and metabolomics associated with the study of the properties of biological systems at the molecular and subcellular levels of structural organization.

## 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) The history of the development of computer processing of biological data. The definition of bioinformatics.

2) Methods for determining the DNA sequence.

3) Formats of sequencing results.

4) Single nucleotide polymorphisms and methods for their detection.

5) Programs for assembling genomes de novo.

6) Search for genomic translocations.

7) The concept of an algorithm. Computational complexity of algorithms. Comparison methods for algorithms.

8) The problem of comparing genetic and protein sequences. Alignment methods: pair and multiple, local and global. Needleman-Wunsh global alignment algorithm. Smith-Waterman local alignment algorithm. Gibbs sampling.

9) Blast package. Purpose and main features. Algorithm.

10) FASTA. Purpose and main features. Algorithm.

11) Search for repetitions, complementarities and symmetries in sequences.

12) Basics of data analysis methods. Regression analysis. Discriminant analysis. Clustering methods. Factor analysis.

13) Recognition of structural and functional motives in genetic texts. Consensus concept, weight matrix. Assessment of recognition accuracy.

14) Methods for recognizing promoters.

15) The concept of the model. Techniques and methods of modeling. The main stages of building mathematical models.

16) Statistical analysis of genomic data.

17) Visualization of biological experiment data.

During the lesson, you can conduct interactive discussions in the course chats, which is a homework assignment. It is possible to perform patent search as an independent task. Successful completion of all tasks of the course and completion of the control fragments of knowledge gives an advantage in the differential credit.

## 4. Evaluation criteria

1) The history of the development of computer processing of biological data. The definition of bioinformatics.

2) Methods for determining the DNA sequence.

3) Formats of sequencing results.

4) Single nucleotide polymorphisms and methods for their detection.

5) Programs for assembling genomes de novo.

6) Search for genomic translocations.

7) The concept of an algorithm. Computational complexity of algorithms. Comparison methods for algorithms.

8) The problem of comparing genetic and protein sequences. Alignment methods: pair and multiple, local and global. Needleman-Wunsh global alignment algorithm. Smith-Waterman local alignment algorithm. Gibbs sampling.

9) Blast package. Purpose and main features. Algorithm.

10) FASTA. Purpose and main features. Algorithm.

11) Search for repetitions, complementarities and symmetries in sequences.

12) Basics of data analysis methods. Regression analysis. Discriminant analysis. Clustering methods. Factor analysis.

13) Recognition of structural and functional motives in genetic texts. Consensus concept, weight matrix. Assessment of recognition accuracy.

14) Methods for recognizing promoters.

15) The concept of the model. Techniques and methods of modeling. The main stages of building mathematical models.

16) Statistical analysis of genomic data.

17) Visualization of biological experiment data.


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**

When conducting an oral differential test, the student is given 60 minutes to prepare. Interrogation of a student on a ticket on an oral differential test should not exceed one astronomical hour. During the oral exam, the student is given 30 minutes to prepare. The interview for a student in an oral exam must not exceed one astronomical hour.