**Course Program of «Computer technologies, biological databases and statistics»**

**Staff Members Responsible for the course:** Assistant Prof., Dr. Gorovtsov Andrey Vladimirovich, Biochemistry and microbiology department of Southern Federal University; Associate Prof. Dr. Butenko Elena Viktorovna Genetics department of Southern Federal University. Associate Prof. Dr. Sherbina Dmitry Nikolaevich Academy of Biology and Biotechnology, Southern Federal University.

1. **The name of the educational program in which the discipline is read.**

The discipline "Biosphere and the estimation of anthropogenic impact on the environment" is read in the framework of the Master's program "Human ecology with fundamentals of biomedicine" in the direction of training 06.04.01 "Biology", graduate level.

**2. The overall complexity**

The total complexity of the discipline is 6 credit units, 216 hours. The program of the discipline provides:

First semester: 18 lectures, 54 laboratory practice, 36 student independent work, 3 ZET

Second semester: 18 lectures, 54 laboratory practice, 36 student independent work, test 3 ZET

**3. The place of discipline in the structure of the educational program.**

The discipline "Computer technologies, biological databases and statistics" refers to the basic module and is one of the disciplines of master's programs "Human ecology with fundamentals of biomedicine", "Genetics", "Ecology" and "Phytodesign." Students are engaged in this discipline in the first and second semester.

It relies on the courses “Mathematical Methods in Biology”, “Information and Communication Technologies” and provides the skills necessary for a specialist for professional activities.

**4. The purpose of studying the discipline.**

Formation of the student's competences associated with the use of modern computer and information technologies for the search, processing, analysis, presentation and publication of biological information.

**5. Requirements for the results of mastering the discipline**

In accordance with the federal state educational standard of higher education in the direction of training 06.04.01 Biology (master’s degree level) discipline "Computer technologies, biological databases and statistics" is aimed at the formation of the following competencies:

A graduate of the master program should have the ability to independently analyze available information, identify fundamental problems, set a task, carry out field and laboratory biological research in urgent specific problems using modern equipment, and computing tools, be responsible for the quality of work and scientific accuracy of the results.

A graduate of the master program should be willing to creatively apply modern computer technology in the collection, storage, processing, analysis and transmission of biological information to solve professional problems.

As a result of mastering the discipline, the student should:

**Know:**

* The principles of the search engines, used to form the query operators, the advantages and disadvantages of various search engines.
* Application areas for various types of data analysis and visualization software.
* The modern approaches to the development of software systems, modern information technologies used in the development of various applications: windows-based applications with a graphical user interface, web applications and applications for various devices, software development platforms, mobile devices, operating systems executed on them and features of software development for such devices.
* The mechanisms for the preservation of information by living systems and the implementation of programs embedded in the genomes, in ontogenesis, during differentiation and in the process of functioning of living systems; basic biological databases; Genomic browsers. GenBank.
* Have an understanding of the main information resources of biology and medicine, the peculiarities of access to these data and the possibilities of their bioinformatic analysis;
* Have an understanding of modern methods of analyzing gene expression.
* Understand the principles underlying biological sequence comparison algorithms, and have an understanding of how BLAST works.
* Understand the principles of phylogenetic analysis.
* Know the main resources in proteomics, to understand the possibilities and disadvantages of modern methods of analyzing the functions of proteins.

**Be able to:**

* Effectively find on the Internet any scientific information of interest. Generate the correct request.
* Formulate rules for assessing the correctness of data analysis results.
* To form requirements for a software product that may be required to solve professional problems.
* Use bioinformatics approaches to the analysis of genomic marker interactions.
* Use bioinformatics approaches to the analysis of associations in the study of the complete genome.
* Apply acquired theoretical knowledge in practice;
* Professionally use GenBank and UCSC Genome Browser;
* Professionally use the program BLAST;
* Build and analyze phylogenetic trees;
* To carry out statistical analysis and functional annotation of data on the analysis of gene expression.

**Have skills to:**

* Use text analysis methods to generate a query and obtain relevant results.
* Use the methods of searching for scientific articles on topics of interest and filtering the results.
* Use the search methods of graphic and video information.
* To be able to describe the results of statistical data processing in text and tabular form.
* Use the methods of visualization of the results of processing biological information, including the correct designation of types and units of measurement.
* Formulate conclusions based on digital and graphic results of biological information processing.
* Analyze sequences of genomes, genetic experiments and processing of their results using mathematical statistics methods.
* To possess the skills of independent work with web-resources on bioinformatics.
* Use routine and modern methods of genome analysis

**6. The contents of the discipline**

The content of the discipline "Computer technologies, biological databases and statistics" is built on a modular principle, with three main modules:

Module 1. WEB RESOURCES

Topic 1. Zipf laws and the mechanism of the search engines.

Topic 2. Features of Google search.

Subject 3. Search for pictures and videos

Subject 4. Search for scientific articles in Science Direct and Springer

Topic 5. Search for scientific articles in Scopus and NCBI

Topic 6. Search for information in online encyclopedias and reference books

Module 2. BIOINFORMATION DATABASES

Database classification

Genomic markers

Gene networks

Module 3. COMPUTER TECHNOLOGIES

Topic 1. The general scheme of data processing.

Topic 2. Storage and transfer of data.

Topic 3. Automation of data processing

Topic 4. Computer experiments

Theme 5. Machine learning

Topic 6. Principles of building applications

**7. Basic educational technology**

Discipline teaching provides the following forms of organization of the educational process: score-rating system of knowledge assessment during the current control, mid-term control and intermediate certification, interactive lectures, independent student work, testing, project method, presentation method

**8. Forms of control**

The discipline program provides for the following types of control: monitoring progress in the form of a test, a report with a presentation and a project assignment, mid-term monitoring of progress in the form of testing, intermediate control in the form of an exam.

**GRADING SCHEME**

|  |  |
| --- | --- |
| A | EXCELLENT - outstanding performance with only minor errors |
| B | VERY GOOD - above the average standard but with some errors |
| C | GOOD - generally sound work with a number of notable errors |
| D | SATISFACTORY- fair but with significant shortcomings |
| E | SUFFICIENT - performance meets the minimum criteria |
| FX | FAIL - some more work required before the credit can be awarded |
| F | FAIL - considerable further work is required |