

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**МЕЖДУНАРОДНЫЙ ГОСУДАРСТВЕННЫЙ ЭКОЛОГИЧЕСКИЙ
ИНСТИТУТ им. А. Д. САХАРОВА**

Кафедра мониторинга окружающей среды

**РАЗРАБОТКА ПАЙПЛАЙНА ДЛЯ ОБРАБОТКИ ДАННЫХ
ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ
ЦИТОПЛАЗМАТИЧЕСКИХ ГЕНОМОВ РАСТЕНИЙ**

Дипломная работа студента IV курса

Лобкова Фёдора Леонидовича

_____ Ф. Л. Лобков

“Допустить к защите”
Зав. кафедрой информационных
технологий в экологии и медицине
к. б. н., доцент

_____ В. В. Журавков

“__” _____ 2022 г.

Научный руководитель
д.б.н., профессор

_____ С. Е. Дромашко

Минск, 2022

РЕФЕРАТ

Дипломная работа 82 страницы, 16 рисунков, 5 таблиц, 10 источников.

БИОИНФОРМАТИКА, ПАЙПЛАЙН, СЕКВЕНИРОВАНИЕ,
НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ, ВЫЗОВ ВАРИАНТОВ,
SNAKEMAKE, MAMBAFORGE, CONDA, PYTHON, АТОМ

Объект исследования:

Данные, полученные в результате секвенирования Illumina Solexa.

Цель работы:

Разработка пайплайна для обработки данных высокопроизводительного секвенирования цитоплазматических геномов растений.

Задачи:

1. Рассмотреть теоретические данные о биоинформатике, генетике.
2. Изучить используемые для получения генома методы секвенирования.
3. Изучить технологии обработки данных секвенирования с применением пайплайнов.
4. Создать схему последовательности обработки первичных данных секвенирования.
5. Проанализировать и выявить лучшую среду выполнения для разработки пайплайнов.
6. Разработать основные этапы проведения обработки данных, а также научиться использовать инструменты обработки данных.

РЭФЕРАТ

Дыпломная работа 82 старонкі, 16 малюнкаў, 5 табліц, 10 крыніц.

БІЯІНФАРМАТЫКА, ПАЙПЛАЙН, СЕКВЕНАВАННЕ, НУКЛЕАТЫДНАЯ ПАСЛЯДОЎНАСЦЬ, ВЫКЛІК ВАРЫЯНТАЎ, SNAKEMAKE, MAMBAFORGE, CONDA, PYTHON, АТОМ

Аб'ект даследавання:

Дадзеныя, атрыманыя ў выніку секвеніравання Illumina Solexa.

Мэта работы:

Распрацоўка пайплайна для апрацоўкі дадзеных высокапрадукцыйнага секвеніравання цытаплазматычных геномаў раслін.

Задачы:

1. Разгледзець тэарэтычныя дадзеныя аб біяінфарматыке, генетыке.
2. Вывучыць метады секвеніравання, якія выкарыстоўваюцца для атрымання геному.
3. Вывучыць тэхналогіі апрацоўкі дадзеных секвеніравання з ужываннем пайплайнаў.
4. Стварыць схему паслядоўнасці апрацоўкі дадзеных секвеніравання.
5. Прааналізаваць і выявіць лепшае асяроддзе выканання для распрацоўкі пайплайнаў.
6. Распрацаваць асноўныя этапы правядзення апрацоўкі даных, а таксама навучыцца працаваць з інструментамі апрацоўкі даных.

ABSTRACT

Diploma work 82 pages, 16 figures, 5 tables, 10 sources.

BIOINFORMATICS, PIPELINE, SEQUENCING, NUCLEOTIDE SEQUENCE, VARIANT CALL, SNAKEMAKE, MAMBAFORGE, CONDA, PYTHON, ATOM

Object of study:

Data from Illumina Solexa sequencing.

Purpose of work:

Developing of a pipeline for processing data from high-throughput sequencing of plant cytoplasmic genomes.

Tasks:

1. To analyze theoretical data on bioinformatics, genetics.
2. To learn the sequencing methods used to obtain the genome.
3. To learn technologies for processing sequencing data using pipelines.
4. To create a sequence diagram for processing raw sequencing data.
5. To analyze and identify the best environment for pipeline development.
6. To develop the main phases of data processing, as well as to learn how to work with data processing tools.