

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра системного анализа и компьютерного моделирования

САРНАЦКИЙ Денис Дмитриевич

**РАЗРАБОТКА И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМОВ
МОДЕЛИРОВАНИЯ И ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА
САЙТОВ ОДНОНУКЛЕОТИДНОГО ГЕНЕТИЧЕСКОГО
ПОЛИМОРФИЗМА**

Аннотация (реферат) к дипломной работе

Научный руководитель:
канд. физ.-мат. наук,
доцент Н. Н. Яцков

Допущен к защите

«___» _____ 2025 г.

Заведующий кафедрой системного анализа
и компьютерного моделирования

кандидат физико-математических наук, доцент В.В. Скакун

РЕФЕРАТ

Дипломная работа: 74 страницы, 23 рисунка, 12 таблиц, 42 источника, 8 приложений.

ОДНОНУКЛЕОТИДНЫЙ ПОЛИМОРФИЗМ, СЕКВЕНИРОВАНИЕ, СТАТИСТИЧЕСКИЕ КРИТЕРИИ, ИМИТАЦИОННАЯ МОДЕЛЬ ДАННЫХ ГЕНОМНОГО СЕКВЕНИРОВАНИЯ, МАШИННОЕ ОБУЧЕНИЕ, ВЕБ-ПРИЛОЖЕНИЕ

Объект исследования: генетический однонуклеотидный полиморфизм в молекулах ДНК человека.

Предмет исследования: математические модели, методы, алгоритмы имитационного моделирования для идентификации однонуклеотидных полиморфизмов по данным геномного секвенирования.

Цель работы: разработка и программная реализация эффективных алгоритмов имитационного моделирования и интеллектуального анализа данных для идентификации сайтов однонуклеотидного полиморфизма в молекулах ДНК человека.

Методы проведения работы:

1. Теоретические: анализ статистических критериев, разработка математических моделей и алгоритмов.
2. Экспериментальные: имитационное моделирование и тестирование алгоритмов на данных геномного секвенирования.

Результаты работы:

1. Реализованы три метода имитационного моделирования: параметрический, на основе экспериментальных функций распределения, бутстрапирование.
2. Разработаны алгоритмы генерации признаков нуклеотидных сайтов и определены наиболее информативные признаки.
3. Проведён анализ алгоритмов на экспериментальных данных генома HG001 (хромосомы 10 и 22).
4. Создано веб-приложение для идентификации и визуализации результатов с системой защиты.

Область возможного практического применения:

Разработанные алгоритмы и инструменты применимы в медицине для идентификации сайтов однонуклеотидного полиморфизма в молекулах ДНК человека.

РЭФЕРАТ

Дыпломная работа: 74 старонкі, 23 малюнка, 12 табліц, 42 крыніцы, 8 прыкладанняў.

АДНАНУКЛЕАТЫДНЫ ПАЛІМАРФІЗМ, СЕКВЕНІРАВАННЕ, СТАТЫСТЫЧНЫЯ КРЫТЭРЫІ ВЫЗНАЧЭННЯ АДНАНУКЛЕАТЫДНЫХ ПАЛІМАРФІЗМАЎ, ІМІТАЦЫЙНАЯ МАДЭЛЬ ДАДЗЕНых ГЕНОМНАГО СЕКВЕНІРАВАННЯ, МАШЫННАЕ НАВУЧАННЕ, ВЭБ-АПЛІКАЦЫЯ

Аб'ект даследавання: генетычны аднануклеатыдны палімарфізм ў малекулах ДНК чалавека.

Прадмет даследавання: матэматычныя мадэлі, метады, алгарытмы імітацыйнага мадэлявання для ідэнтыфікацыі сайтаў аднануклеатыднага палімарфізму па дадзеных геномнага секвеніравання.

Мэта работы: распрацоўка і праграмная рэалізацыя эфектыўных алгарытмаў імітацыйнага мадэлявання і інтэлектуальнага аналізу сайтаў аднануклеатыднага генетычнага палімарфізму ў малекулах ДНК чалавека.

Метады правядзення работы:

1. Тэарэтычныя: аналіз статыстычных крытэрыяў, распрацоўка матэматычных мадэляў і алгарытмаў.
2. Эксперыментальныя: імітацыйнае мадэляванне і тэставанне алгарытмаў на дадзеных геномнага секвеніравання.

Вынікі работы:

1. Рэалізаваны тры метады імітацыйнага мадэлявання: параметрычны, на аснове эксперыментальных функцый размеркавання, бутстрапаванне.
2. Распрацаваны алгарытмы генерацыі прыкмет нуклеатыдных сайтаў і вызначаны найбольш інфарматыўныя прыкметы.
3. Праведзены аналіз алгарытмаў на эксперыментальных дадзеных геному HG001 (храмасомы 10 і 22).
4. Створана вэб-аплікацыя для ідэнтыфікацыі і візуалізацыі вынікаў з сістэмай абароны.

Вобласць магчымага практычнага выкарыстання:

Распрацаваныя алгарытмы і інструменты дастасавальныя ў медыцыне для ідэнтыфікацыі сайтаў аднануклеатыднага палімарфізму ў малекулах ДНК чалавека.

SUMMARY

Thesis: 74 pages, 23 figures, 12 tables, 42 sources, 8 appendices.

SINGLE NUCLEOTIDE POLYMORPHISM, SEQUENCING, STATISTICAL CRITERIA FOR DETERMINING SINGLE NUCLEOTIDE POLYMORPHISMS, SIMULATION MODEL OF GENOMIC SEQUENCING DATA, MACHINE LEARNING, WEB APPLICATION

Object of research: genetic single nucleotide polymorphism in human DNA molecules.

Subject of research: mathematical models, methods, and simulation algorithms for the identification of single-nucleotide polymorphisms based on genomic sequencing data.

Objective: to develop and programmatically implement effective algorithms for simulation and data mining to identify single-nucleotide polymorphism sites in human DNA molecules.

Methods of work:

1. Theoretical: analysis of statistical criteria, development of mathematical models and algorithms.
2. Experimental: simulation modeling and testing of algorithms based on genomic sequencing data.

The results of the work:

1. Three simulation modeling methods have been implemented: parametric, based on experimental distribution functions, and bootstrapping.
2. Algorithms for generating features of nucleotide sites have been developed and the most informative features have been identified.
3. Algorithms were analyzed based on experimental data from the HG001 genome (chromosomes 10 and 22).
4. A web application for identification and visualization of results with a security system was created.

Field of possible practical application:

The developed algorithms and tools are applicable in medicine for the identification of single-nucleotide polymorphism sites in human DNA molecules.