

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.  
ЧЕРНЫШЕВСКОГО»  
(СГУ)

Кафедра дискретной математики и  
информационных технологий

**РАЗРАБОТКА МОДЕЛЕЙ И МЕТОДОВ ДЛЯ РАСПОЗНАВАНИЯ  
СВОЙСТВ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНЕТИЧЕСКИХ КОДОВ В  
РЕШЕНИИ ПРИКЛАДНЫХ ЗАДАЧ**

АВТОРЕФЕРАТ

НАУЧНО-КВАЛИФИКАЦИОННОЙ РАБОТЫ

аспиранта 4 курса  
направления 09.06.01 Информатика и вычислительная техника  
направленности «Математическое моделирование, численные методы и  
комплексы программ»

**КАРЯКИНА ДЕНИСА АЛЕКСЕЕВИЧА**

Научный руководитель  
к. ф.-м.н., доцент

\_\_\_\_\_

Тяпаев Ливат Борисович

Зав. Кафедрой

\_\_\_\_\_

Тяпаев Ливат Борисович

Саратов 2019

## Общая характеристика работы

**Актуальность проблемы.** При анализе структур и функций генетического материала в решениях проблем наследственности и патологий живых организмов, а также в профилактике и лечении наследственных патологий явно или неявно используются классификации и задачи распознавания генетических последовательностей. После развития технологий секвенирования следующего поколения (NGS) у исследователей появился доступ к огромному количеству данных. Возможность реализации различных анализов на больших объемах генетических данных зависит от используемого математического аппарата.

В квалификационной работе взят за основу математический аппарат, разработанный Твердохлебовым В.А. С использованием которого определяются свойства последовательностей, по которым оказывается возможным распознавать последовательности, определять принадлежность последовательности конкретному классу последовательностей и различать классы последовательностей. На основе предложенных им способов рекуррентного и  $Z$ -рекуррентного определения последовательностей в квалификационной работе реализованы программная система и проведены вычислительные эксперименты.

**Цели работы.** Разработка методов для распознавания по свойствам генетических кодов представленных  $Z$  – рекуррентными определениями свойств кодов как самих кодов, так и классов кодов. Разработка алгоритма, составление и отладка программы определения числовых показателей рекуррентного определения последовательностей. Разработка библиотеки программ для вычисления числовых показателей рекуррентного и  $Z$ -рекуррентного определения последовательностей генетических кодов.

Работа составленных программ показана на основе проведения вычислительных экспериментов, в которых проиллюстрированы решения следующих задач:

- задача распознавания последовательностей по соответствующим им показателям  $Z$  – рекуррентных определений последовательности,
- задача определения принадлежности последовательности к заданному классу последовательностей, охарактеризованному множеством выполняющихся для класса  $Z$  – рекуррентных определений,
- определение показателей сложности для выбранных генетических последовательностей с использованием рекуррентного определения последовательностей,
- классификация выбранных генетических последовательностей с использованием рекуррентного и  $Z$  - рекуррентного определения последовательностей,
- определение филогенетического сходства двух генетических последовательностей.

**Научная новизна.** В работе получены следующие новые научные результаты:

- разработан алгоритм для вычисления рекуррентных и  $Z$  - рекуррентных определений последовательностей имеющих содержательную интерпретацию в приложениях к генетике,
- разработан набор программ для автоматической загрузки и анализа генетических последовательностей,
- при использовании разработанной программной системы было показано, что возможно создавать классификации представленных для

рассмотрения генетических последовательностей, а также определять их филогенетическое сходство.

**Методы исследований.** В работе используются методы дискретной математики и методы распознавания последовательностей по выполняющимся для них  $Z$  – рекуррентными определениями.

**Апробация результатов работы.** Основные результаты докладывались на международной конференции “Компьютерные науки и информационные технологии”, Саратов, 2016.

**Публикации.** Основные результаты по теме диссертации подготовлены к публикации в российский журнал из списка рекомендованных ВАК Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика. 2019, выпуск 3.

### Содержание работы

В первой главе вводятся основные понятия предметной области. Рассматривается понятие генетической последовательности, ДНК, свойства генетического кода. Рассмотрены основные банки генетических данных, дана краткая характеристика.

Вторая глава посвящена описанию рекуррентному и  $Z$  – рекуррентному определению последовательностей. Рассмотрено применение нового (что подтверждено публикациями и выступлениями на конференциях) аппарата  $Z$  — рекуррентного определения свойств последовательностей для решения задач распознавания последовательностей, определения принадлежности последовательности заданному классу и распознавание классов последовательностей.

Третья глава содержит описание метода распознавания генетических последовательностей по числовым показателям рекуррентных определений. Описываются алгоритмы вычисления множеств определений выполняющихся для последовательности  $Z$  - рекуррентных определений. Производится выбор генетических последовательностей и массивов генетических последовательностей для проведения вычислительных экспериментов. Производится анализ результатов вычислительных экспериментов.

### **Заключение**

В научно-квалификационной работе получены следующие основные результаты:

- разработан алгоритм для вычисления рекуррентных и  $Z$  - рекуррентных определений последовательностей имеющих содержательную интерпретацию в приложениях к генетике,
- разработан набор программ для автоматической загрузки и анализа генетических последовательностей,
- при использовании разработанной программной системы было показано, что возможно создавать классификации представленных для рассмотрения генетических последовательностей, а также определять их филогенетическое сходство.