

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.021:616

Савик
Олег Владимирович

АЛГОРИТМЫ АНАЛИЗА ГЕНЕТИЧЕСКОГО КОДА И ПРОГНОЗИРОВАНИЯ
БОЛЕЗНЕЙ ЧЕЛОВЕКА

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1 - 40 80 02 Системный анализ, управление и обработка
информации (по отраслям)

Научный руководитель

А.П. Курулев,
кандидат технических наук,
профессор

Минск 2020

ВВЕДЕНИЕ

Конец XX века ознаменовался значительными достижениями в биологической науке и в частности в области молекулярной биологии. Одним из важных достижений по праву можно считать создание роботизированных систем для определения последовательностей оснований ДНК, позволяющих за приемлемое время расшифровывать отдельные участки ДНК и даже полные геномные последовательности различных видов организмов. В последние 10-15 лет усилия научного сообщества были направлены на накопление последовательностей, и вершиной этой деятельности стала расшифровка генома человека, состоящего из 3×10^9 нуклеотидных пар. В настоящее время объем накопленной генетической информации в основных банках данных последовательностей ДНК составляет более 19 Гб и объем вновь расшифрованных последовательностей стремительно растет. Невозможно переоценить значение полученной информации для науки, медицины и других областей жизни человечества. Однако, для успешного использования этой информации необходимо прежде всего понять ее биологический смысл, «прочитать», что закодировано в последовательностях ДНК и аминокислот

Цель: разработка алгоритма анализа символьной генетической последовательности на поиск периодичности и дисперсных повторов последовательностей.

Объект исследования: символьная генетическая последовательность.

Предмет: анализ символьной генетической последовательности.

Задачи:

- 1 Обзор существующих подходов анализа генетических текстов.
- 2 Разработка нового алгоритма анализа генетических текстов на основе метода символьного разложения последовательности для поиска периодичности и дисперсных повторов последовательностей.
- 3 Исследование разработанного метода, сравнение с ныне существующими.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель исследования

Целью диссертационной работы является исследование существующих подходов к анализу генетического текста, разработка методов анализа генетической последовательности, которые позволят получить более детальные данные о закономерностях выстраивания генетической последовательности, а также предположения о возможных мутациях.

Задачи исследования

- 1 Обзор существующих подходов к анализу генетического текста.
- 2 Разработка нового алгоритма анализа генетических текстов, поиск периодичности и дисперсных повторов в символьной последовательности.
- 3 Исследование разработанного метода, сравнение с ныне существующими.

Личный вклад соискателя

Соискателем выполнены все изложенные в работе разработки и исследования. Постановка задач и обсуждение результатов проводились совместно с научным руководителем, доцентом кафедры систем управления Белорусского государственного университета информатики и радиоэлектроники. Обработка, интерпретация данных, а также выводы сделаны автором самостоятельно.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались на следующих научных конференциях:

– 55-я юбилейная научная конференция аспирантов, магистрантов и студентов" учреждения образования "Белорусский государственный университет информатики и радиоэлектроники"

СОДЕРЖАНИЕ РАБОТЫ

Детальное изучение и анализ генетических последовательностей можно провести экспериментальным путем. Для этого в рамках молекулярной биологии были разработаны следующие методы: электрофорез, специфическая химическая модификация азотистых оснований в составе молекул ДНК, способы радиоактивного и флуоресцентного мечения, клонирования и др. Проблема заключается в том, что применить вышеупомянутые методы для изучения огромного количества доступных в настоящее время последовательностей ДНК и аминокислот просто невозможно из-за их большой трудоемкости и значительной стоимости. Поэтому первоочередная задача состоит в привлечении к изучению последовательностей мощной компьютерной технологии. Соответственно возникает необходимость в разработке математических алгоритмов анализа генетических последовательностей и компьютерных программ, реализующих эти алгоритмы.

В первой главе диссертационной работы производится обзор существующих подходов анализа генетических последовательностей на поиск периодичности в них.

Приводится обзор математических методов и алгоритмов на поиск периодичности в последовательностях на основании преобразования Фурье, динамического программирования, нейронных сетей. Рассматриваются математические алгоритмы для обнаружения дисперсных повторяющихся последовательностей в генетических текстах, а также их преимущества и недостатки.

Во второй главе диссертационной работы производится разработка новых алгоритмов для поиска периодичности и дисперсных повторов в символьных последовательностях.

В третьей главе диссертационной работы производится анализ полученных результатов, а также приводится сравнительная характеристика существующих алгоритмов с разработанным, который может быть применен к любым символьным последовательностям другого рода.

ЗАКЛЮЧЕНИЕ

В процессе диссертации была разработаны и исследованы алгоритмы анализа генетических последовательностей на поиск периодичности и дисперсных повторов символьных последовательностей. Результаты анализа разработанного алгоритма показывают, что он может быть применен к символьным последовательностям другого рода.

Был проведен краткий обзор структуры генетической последовательности. Рассмотрены математические алгоритмы анализа символьной последовательности, применимые к генетической. Представлена их сравнительная характеристика, а также преимущества и недостатки.

Разработан и описан метод анализа символьной генетической последовательности на поиск периодичностей и дисперсных символьных повторов символьных последовательностей.

Произведено исследование разработанного алгоритма, представлены его преимущества и недостатки, произведено сравнение с ныне существующими алгоритмами, а также предложения по его модернизации.

В работе описана пошаговая реализация алгоритма. Его применение позволит получить более детальное представление о структуре символьной генетической последовательности, предсказанию возможных мутаций в генетической последовательности, а также предсказании возможных последствий для исследуемого организма. Также существует возможность анализа символьных последовательностей другого рода на поиск периодичности, например, анализ языковых структур, неизвестных языков.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Савик О.В. Алгоритмы для анализа данных ДНК / О.В. Савик, А.П. Курулев// Материалы научной конференции, 55-ая юбилейная научная конференция аспирантов, магистрантов и студентов учреждения образования «Белорусский государственный университет информатики и радиоэлектроники, Минск, 22–26 апреля 2019 года/ – с. 37.

Библиотека БГУИР