

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

**«Биоинформатический анализ геномов человека для установления
популяционной»**

Гуль Анна Павловна

Научный руководитель – доктор химических наук,
Андрианов А. М.

Минск, 2020

Реферат

Магистерская диссертация, 48 страниц, 27 рисунков, 15 источников, 1 таблица, 1 приложение.

БИОИНФОРМАТИКА, ГЕНЕТИЧЕСКИЙ МАРКЕР, ДЕРЕВО ГАПЛОГРУПП, Y-ХРОМОСОМА, ДЕРЕВО ПРИНЯТИЯ РЕШЕНИЙ, ГАПЛОГРУППА, СЕКВЕНИРОВАНИЕ, РЕФЕРЕНСНЫЙ ГЕНОМ, МАШИННОЕ ОБУЧЕНИЕ

Объект исследования – последовательность ДНК Y-хромосомы индивидов. В частности, были исследованы SNP и STR маркеры и возможность их применения для установления популяционной принадлежности.

Цель работы – разработка программного обеспечения для установления популяционной принадлежности неизвестного индивида по его последовательности ДНК Y-хромосомы. Кроме основной функции приложение также должно поддерживать визуализацию вероятностей принадлежности к популяции на территории Республики Беларусь.

В ходе работы был произведен анализ и разработка методов определения гаплогруппы, указывающей на популяционную принадлежность неизвестного индивида. Выбраны маркеры для исследования, по которым с помощью дерева гаплогрупп устанавливается гаплогруппа индивида. Также в рамках решения задачи были реализованы методы обработки больших объемов информации и методы машинного обучения.

Результатом работы является приложение, реализованное на языке python с использованием фреймворка Flask. Данное приложение позволяет по vcf файлу секвенированного образца определить гаплогруппу. Также реализованный модуль даёт возможность для заданной гаплогруппы определить вероятность принадлежности популяции и отобразить распределение на карте Республики Беларусь.

Область применения – определение популяционной принадлежности неизвестного образца в криминалистике.

Abstract

Master thesis, 48 pages, 27 figures, 15 references, 1 table, 1 attachment.

BIOINFORMATICS, GENETIC MARKER, HAPLOGROUP TREE, Y-CHROMOSOME, DECISION TREE, HAPLOGROUP, SEQUENCING, REFERENCE GENOME, MACHINE LEARNING

The object of the research is a DNA sequence of Y-chromosome individuals, also SNP and STR markers were studied and the possibility of using them for the prediction of human population.

The objective of the thesis is to develop a software for identifying the human population for an unknown individual based on the sequence of DNA for Y-chromosome. In addition, the application also must support visualization the probability of belonging to a human population in the territory of the Republic of Belarus.

During the study, analysis and development of the methods for identifying the haplogroup indicating the population belonging to an unknown individual were performed. Also, the markers, by which the haplogroup of an individual is determined with the use of the haplogroup tree, were selected. In addition, algorithms for processing large volumes of data and some machine learning methods were implemented to solve the problem.

The result of the research is the proposed approach for solving the problem of determining the haplogroup and human population of an unknown individual. The application in python and the Flask framework was implemented, which allows to identify a haplogroup from a vcf - file of a sequenced sample. In addition, the module allows to determine the probability of a human population for a selected haplogroup and visualize the distribution on the map of the Republic of Belarus.

The field of application is a determination of the human population for an unknown individual in criminalistics.