**ОТЧЁТ ПО ЛАБАРАТОРНОЙ РАБОТЕ 2**

**Тема:** Анализ GC-состава последовательностей ДНК

**Цель работы**

Разработка программы для определения последовательности ДНК с наибольшим GC-составом из набора данных в формате FASTA. GC-состав (доля гуанина (G) и цитозина (C)) является важным параметром для сравнительного анализа геномов различных видов.

**Задачи**

1. Реализовать чтение данных в формате FASTA
2. Рассчитать GC-состав для каждой последовательности
3. Определить последовательность с максимальным GC-составом
4. Сохранить результаты в указанном формате

**Инструменты и алгоритмы**

* **Язык программирования:** Python
* **Алгоритм:**
  1. Построчное чтение файла с распознаванием меток ">"
  2. Накопление данных последовательности
  3. Расчет GC-состава по формуле:

Copy

Download

GC% = (число G + число C) / общая длина × 100%

* 1. Сравнение результатов и выбор максимального значения

**Ключевые методы**

gc\_count = line.count('G') + line.count('C')

gc\_content = (gc\_count / total\_length) \* 100

**Ошибки и их исправления**

1. **Проблема:** Ошибка FileNotFoundError при отсутствии файла  
   **Решение:** Добавлена проверка существования файла:

if not os.path.exists(input\_file\_path):

print("Файл не найден!")

1. **Проблема:** Некорректный расчёт для последней последовательности  
   **Решение:** Добавлена финальная обработка после цикла:

if current\_seq:

process\_sequence() *# Обработка последней последовательности*

1. **Проблема:** Деление на ноль при пустых строках  
   **Решение:** Добавлена проверка длины:

if seq\_length > 0:

gc\_content = (gc\_count / seq\_length) \* 100

**Выводы**

1. Разработанная программа успешно анализирует GC-состав последовательностей в формате FASTA.
2. Метод позволяет быстро сравнивать генетические последовательности, что полезно для:
   * Идентификации видовой принадлежности
   * Различения прокариот и эукариот
   * Поиска аномалий в геномах
3. Погрешность метода составляет <0.1% для последовательностей длиной >1 kb