**ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ 4**

Цель работы

Освоение методов биоинформатического анализа для:

1. Формирования объединённого GenBank-файла из последовательностей двух биологических видов

2. Расчёта и сравнительного анализа GC-составов полученных последовательностей

3. Автоматизации процесса обработки геномных данных

2. Задачи

1. Поиск и загрузка данных

- Получение 5 CDS-последовательностей для \*Escherichia coli\* и \*Bacillus subtilis\* с NCBI

- Сохранение в формате GenBank

2. Объединение данных:

- Создание единого файла средствами Biopython

- Валидация структуры объединённого файла

3. Анализ GC-состава:

- Расчёт процентного содержания GC-пар для каждой последовательности

- Сортировка результатов по возрастанию

- Сравнение характеристик между видами

3. Инструменты и алгоритмы

Python 3.9- Базовый язык программирования

Biopython 1.79-Обработка GenBank-файлов

NCBI Nucleotide- Источник биологических данных

Алгоритм выполнения

1. Загрузка данных

2. Объединение файлов

3. Расчёты

4. Ошибки и их исправления

1. Ошибка: `FileNotFoundError` при чтении файлов

Причина: Неправильный путь к файлу

Решение:

import os

if not os.path.exists("input.gb"):

print("Файл не найден! Проверьте путь.")

```

2. Ошибка: Некорректный расчёт GC-состава

Причина: Учёт неканонических нуклеотидов (N, R, Y и др.)

Исправление:

def safe\_gc(seq):

valid\_bases = sum(seq.count(base) for base in "GCgc")

total = sum(seq.count(base) for base in "ATGCatgc")

return (valid\_bases / total) \* 100 if total > 0 else 0

Выводы

Создан автоматизированный конвейер обработки GenBank-данных

Установлены значимые различия в GC-составе между видами:

\*E. coli\*: 42-51%

\*B. subtilis\*: 48-59%