

Rosalind Armory

Тимофей Бондарев

Руководители:
Михаил Райко, Николай Вяххи

Задача

Развить локацию **Bioinformatics Armory**, предназначенную для обучения способам решения биоинформатических задач с использованием

- биоинформатических сервисов,
- языка python и библиотеки BioPython

Результаты

Реализованы задачи:

- GROO - преобразование подформатов FASTQ
- GCTR - выявление и удаление праймера по GC-content
- TSAM - введение в sam-формат файлов выравнивания

GROO

Изначально: по входному файлу в формате FASTQ неизвестного подформата преобразовать его в формат FASTQ, поддерживаемый sanger.

Задача в общем случае неразрешимая => переработка условия.

GROO

Итог: задача преобразования известного подформата FASTQ в подформат sanger FASTQ.

Определение подформата FASTQ по хранящимся данным перенесено в дополнительные задачи.

GCTR

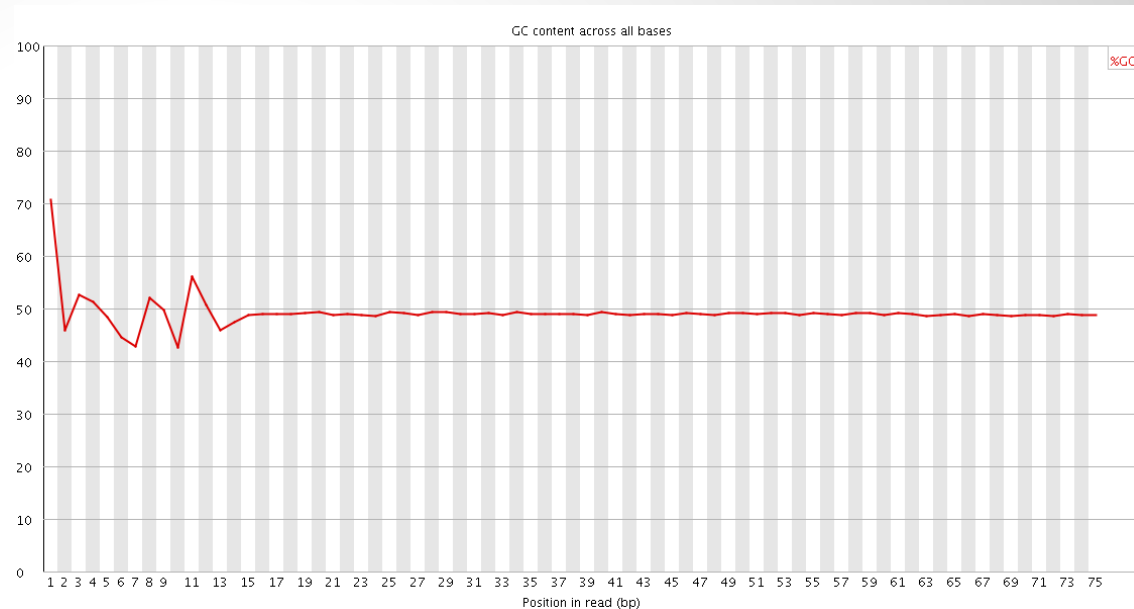
Дано: набор строк одинаковой длины, в начале каждой входит произвольный суффикс праймера.

Решение: использование GC-content.
Вычисление зависимости GC-content от позиции в ряде.

GCTR

Выявление
аномального
участка
поведения
зависимости

GC-content от позиции в риде, удаление
участка ридов, соответствующего праймеру.



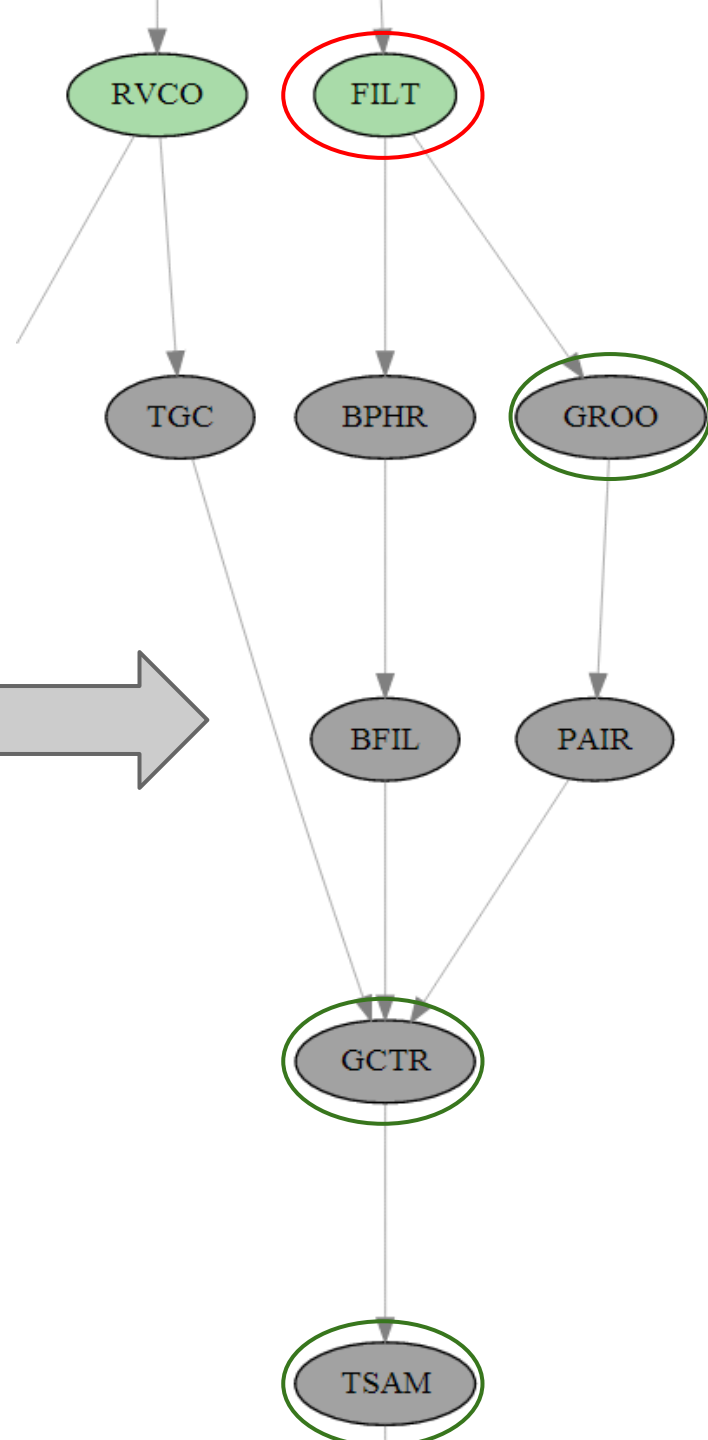
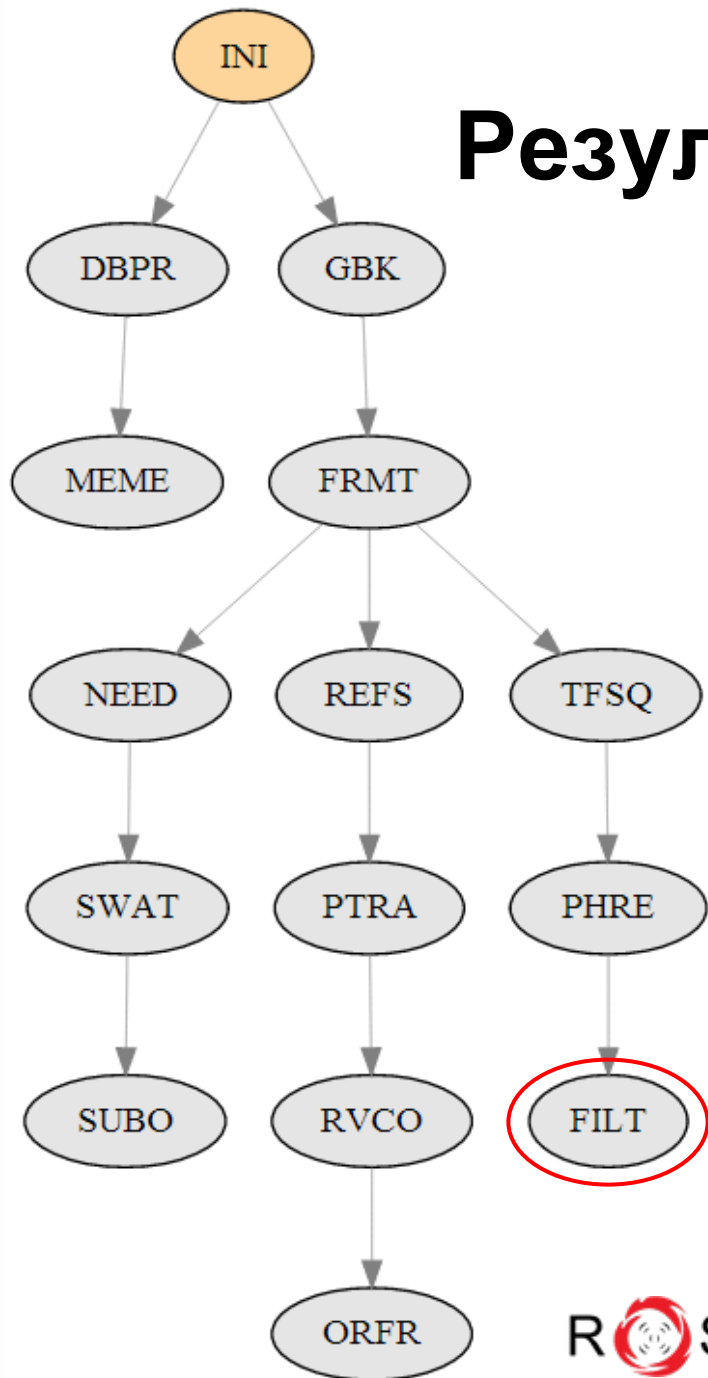
TSAM

Дано: sam-файл, номер записи

Найти: CIGAR-string соответствующей строки, общая длина выравнивания, позиция начала выравнивания

Решается через разбор sam файла и подсчет длин участков, закодированных в CIGAR-string.

Результат



Вопросы?