

Алгоритмы сравнительного анализа серий метагеномных образцов с использованием графов де Брейна для библиотек метагеномных чтений

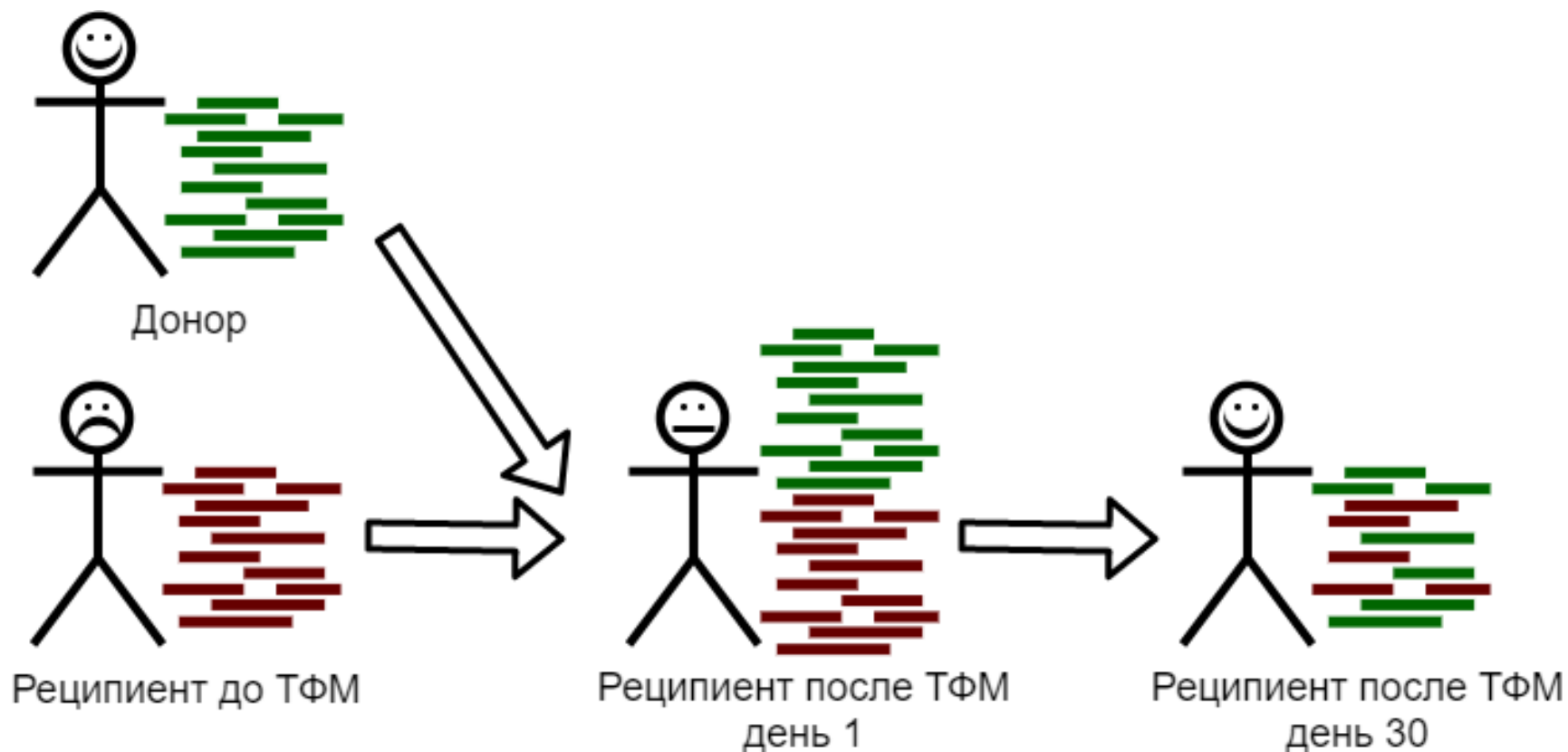
Артем Иванов, студент 4 курса ФИТиП ИТМО

руководитель Ульяновцев В.И., доцент ФИТиП ИТМО
консультант Олехнович Е.И., научный сотрудник
лаборатории биоинформатики ФНКЦ ФХМ России

Секвенирование нового поколения



Процесс трансплантации



Известные факты¹

- Донорские бактерии приживаются и находятся у реципиента спустя 3 месяца
- Приживаются бактерии близкие к уже живущим у реципиента
- Бактерии от одного донора по-разному приживаются у реципиентов

1. Durable coexistence of donor and recipient strains after fecal microbiota transplantation / S. S. Li [et al.] // Science. — 2016. — Vol. 352, no. 6285. — P. 586–589. 4

Существующие решения

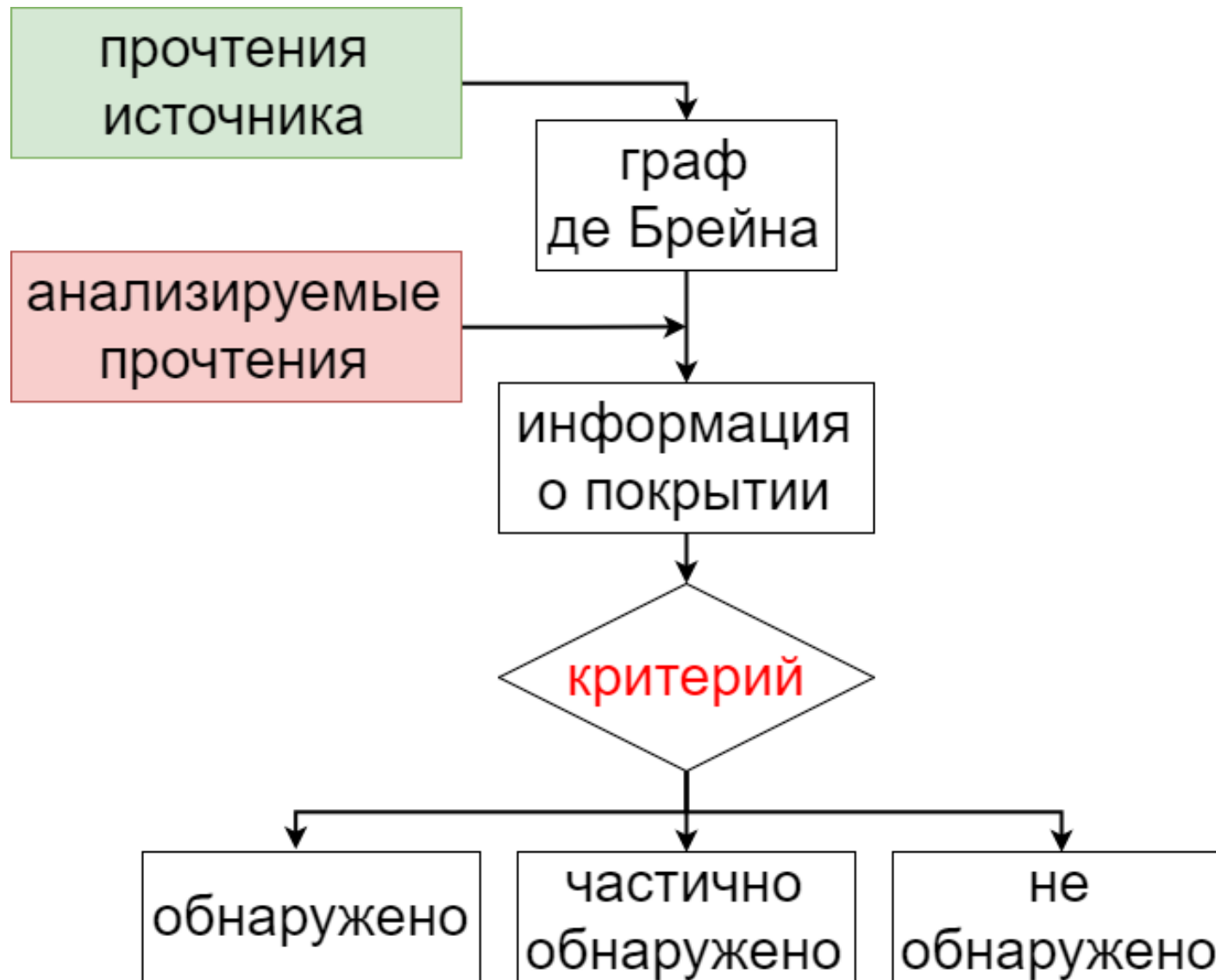
- Таксономическая аннотация
 - сложно (Kraken – 32 %, MIDAS – 49 %)
- Сборка отдельных геномов из метагенома¹
(англ. Metagenome-Assembled Genomes)
 - зашумленность и химерные связи

1. Tracking microbial colonization in fecal microbiota transplantation experiments via genome-resolved metagenomics / S. T. Lee [et al.] // Microbiome. — 2017. — Vol. 5, no. 1. — P. 50.

Задачи

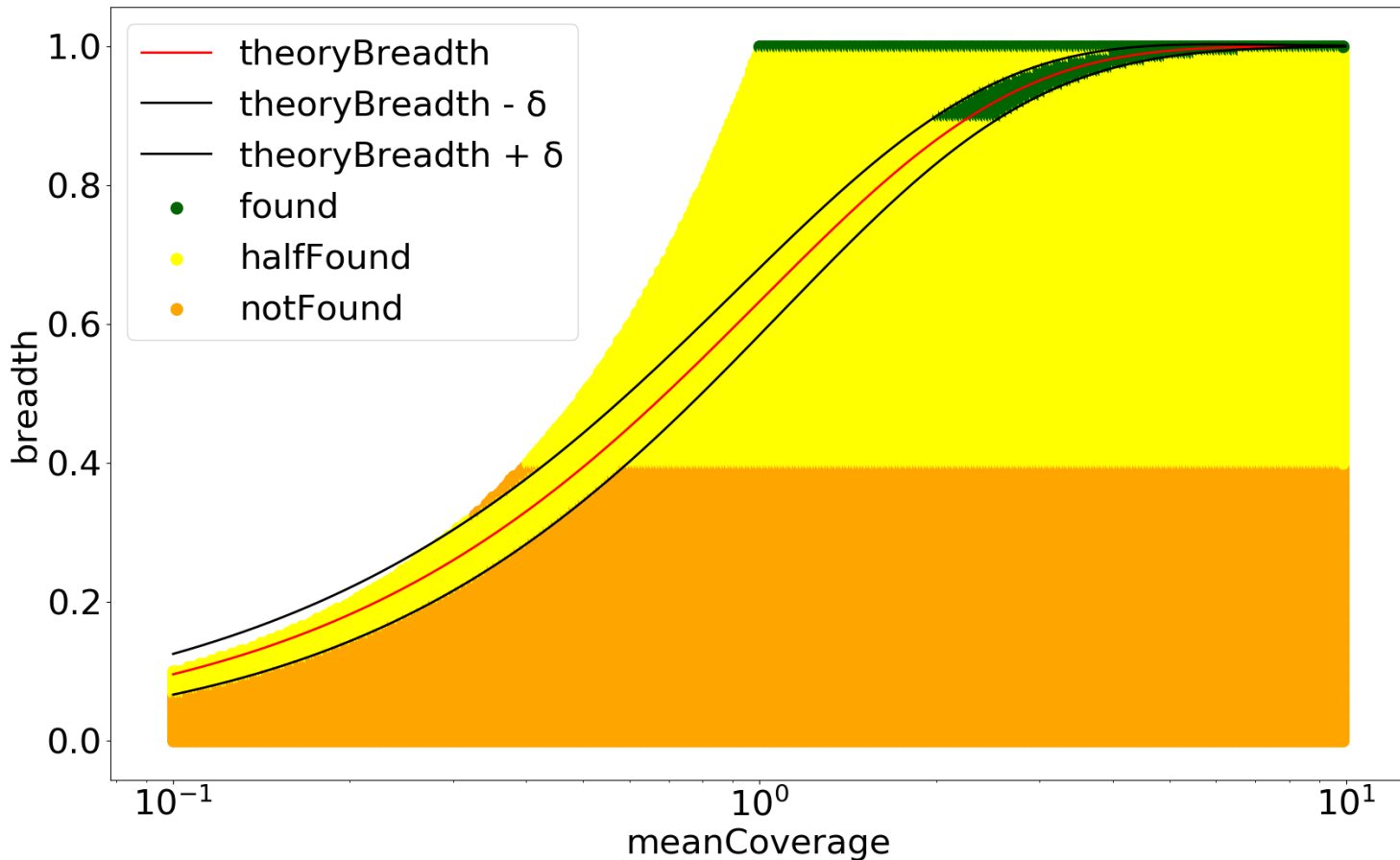
- Разработка алгоритма для сравнительного анализа серий метагеномных образцов
- Разбиение прочтений из входных метагеномов на категории
- Создание тестового набора метагеномов и тестирование
- Запуск на реальных метагеномных данных

Алгоритм поиска прочтений



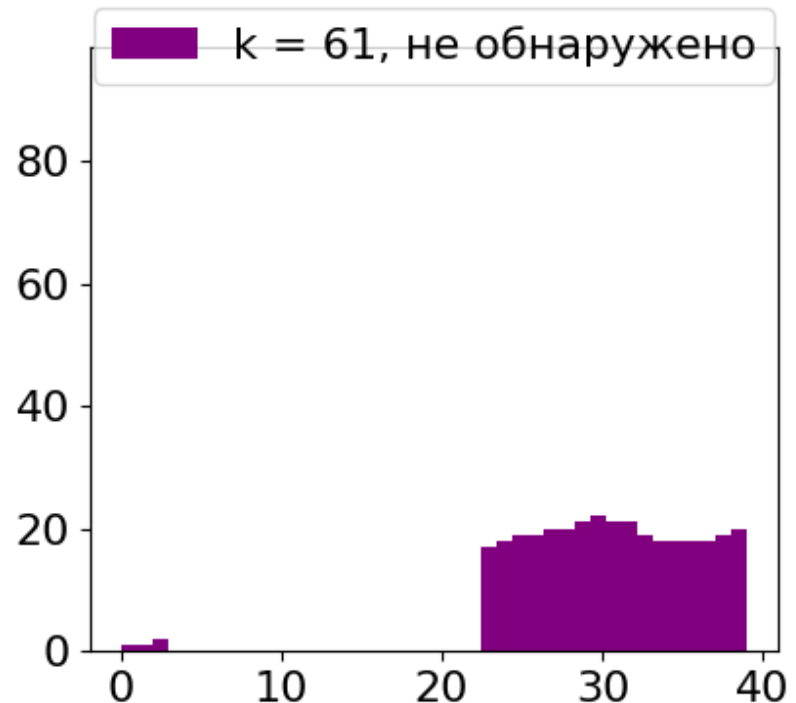
Критерии обнаружения

- $\text{theoryBreadth} = 1 - \exp^{-\text{meanCoverage}}$
- ЭВРИСТИКИ



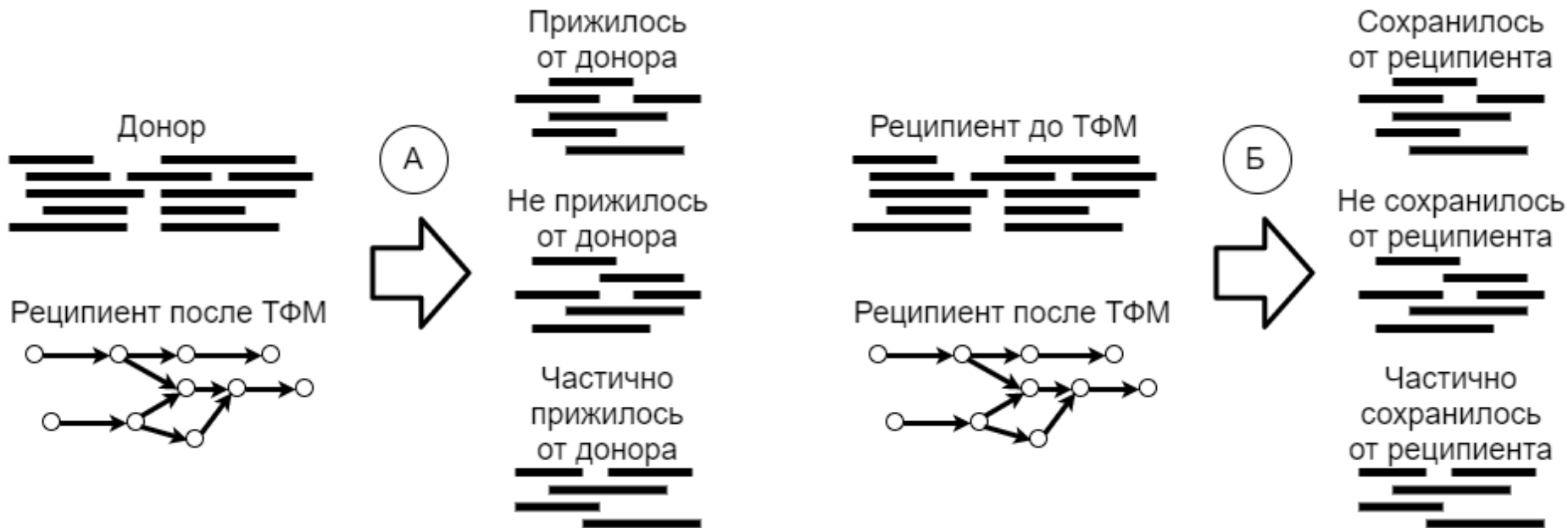
Критерии обнаружения

- Обнаружение прочтения при различных k



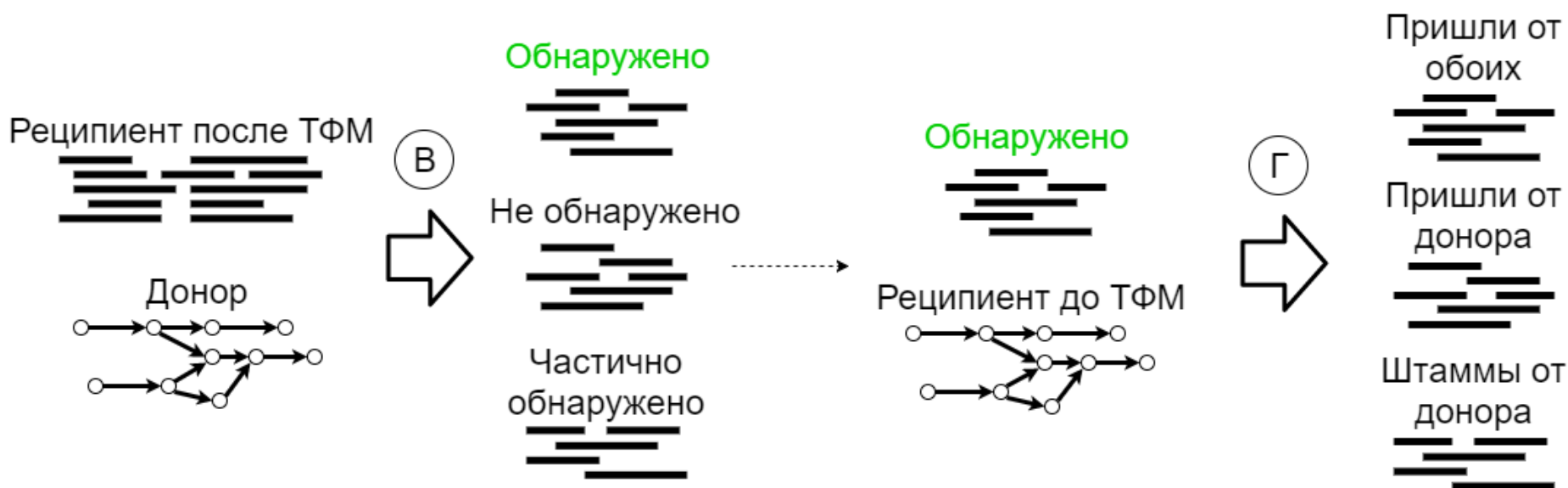
Классификатор прочтений

- Прочтения донора и реципиента до ТФМ



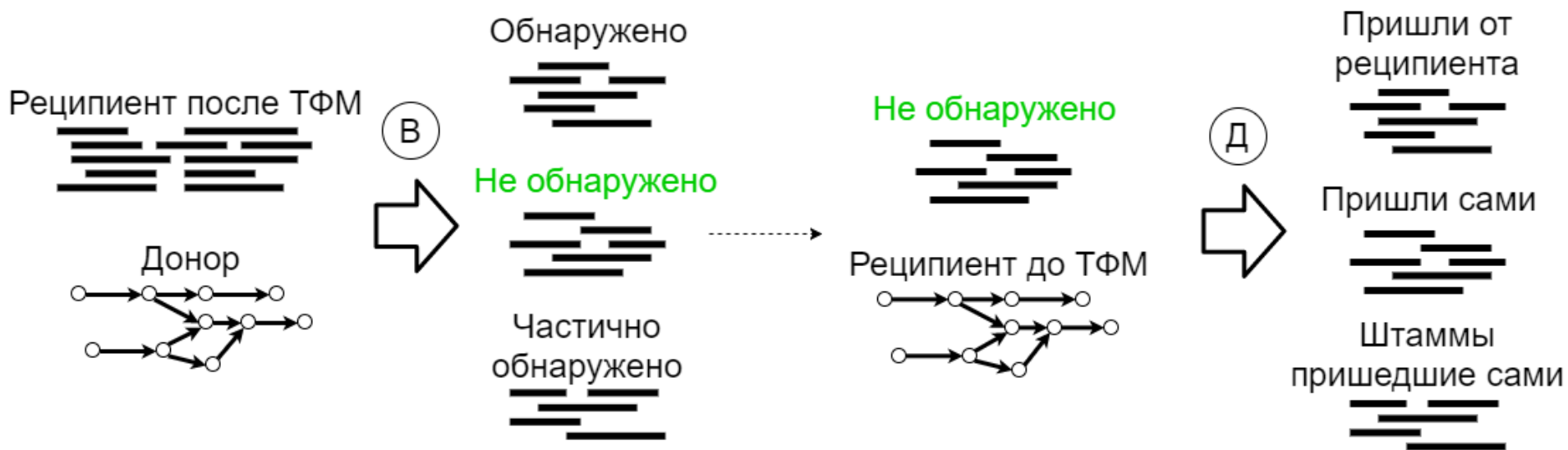
Классификатор прочтений

- Прочтения реципиента после ТФМ



Классификатор прочтений

- Прочтения реципиента после ТФМ



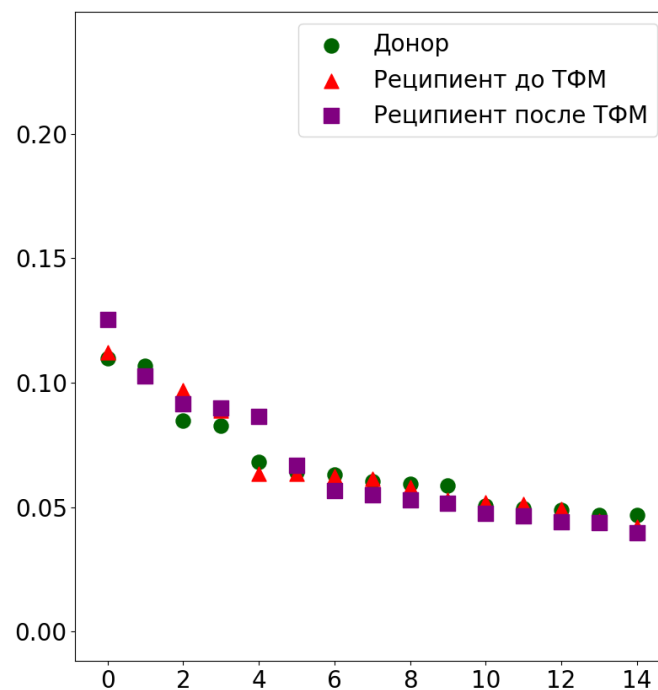
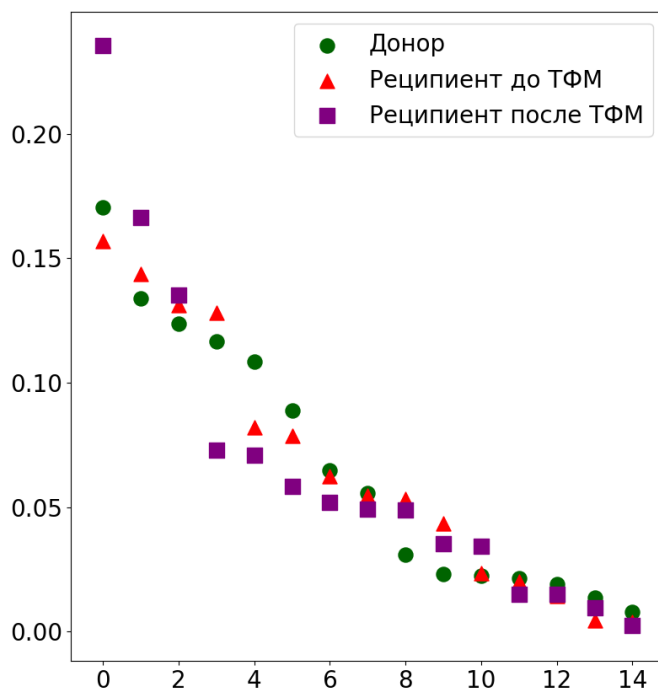
Классификатор прочтений

- Прочтения реципиента после ТФМ



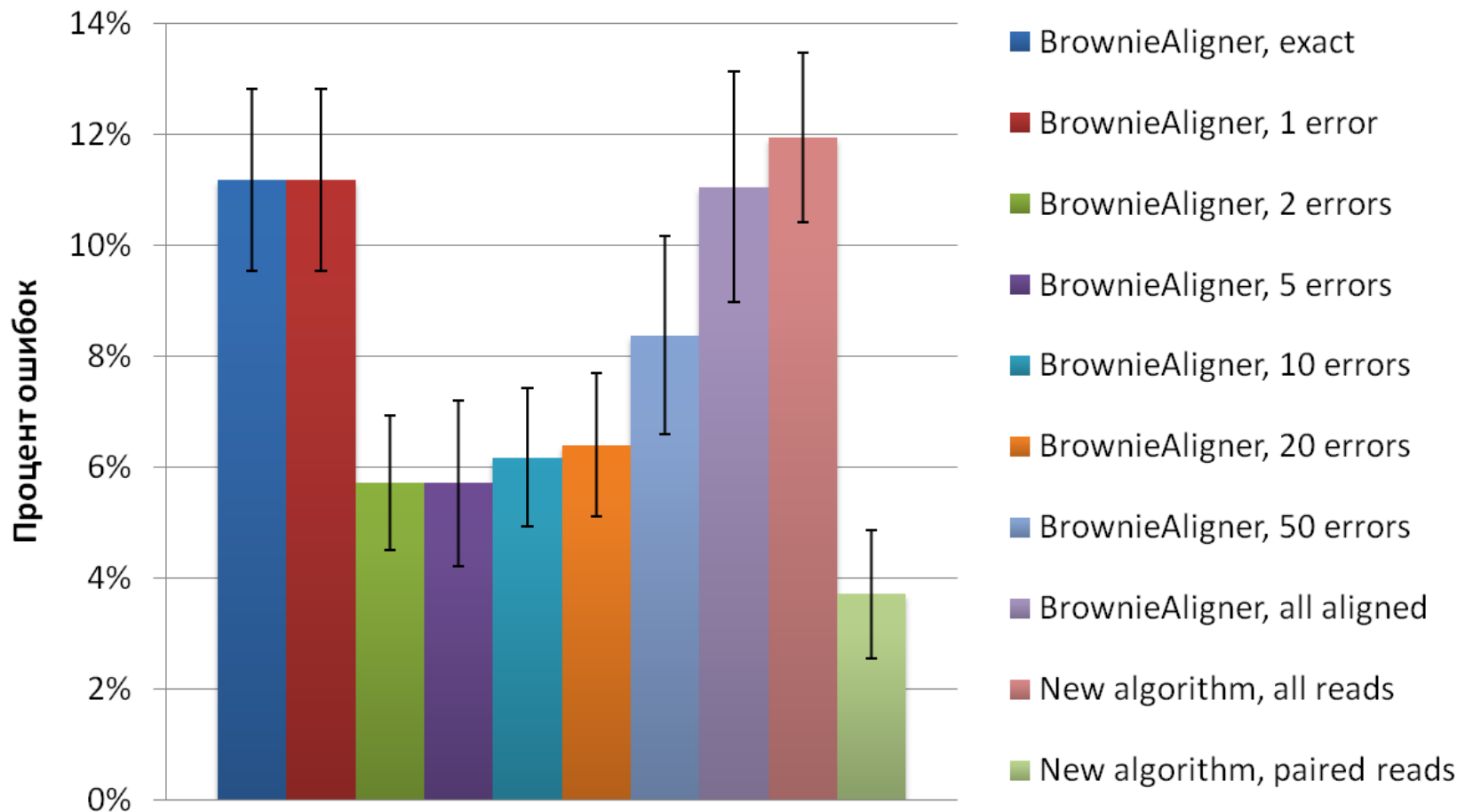
Тестирование – идея

- Сгенерированные метагеномы из видов или штаммов (по 15 в образце)
- С равномерным или показательным распределением

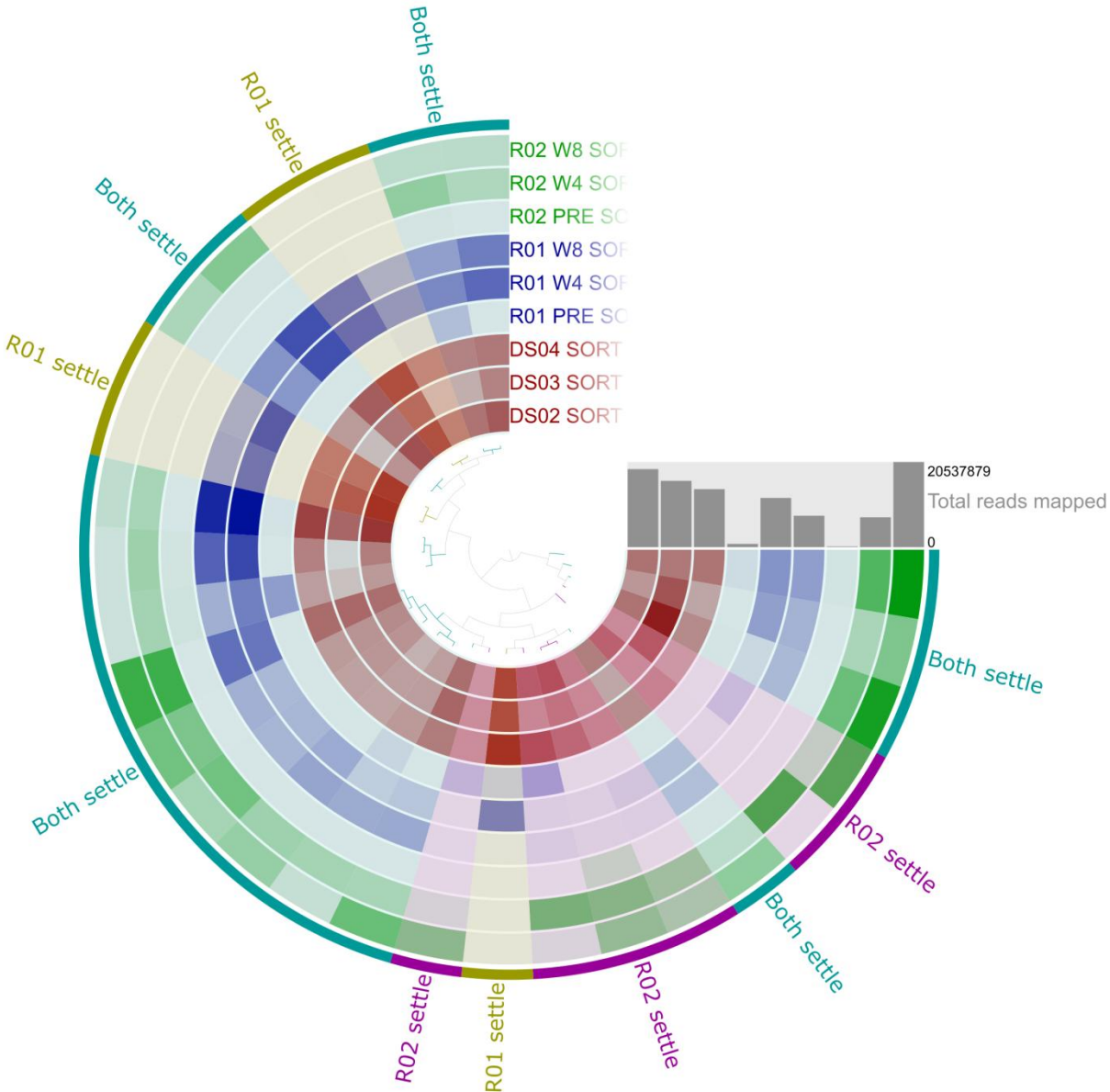


Сравнение с BrownieAligner

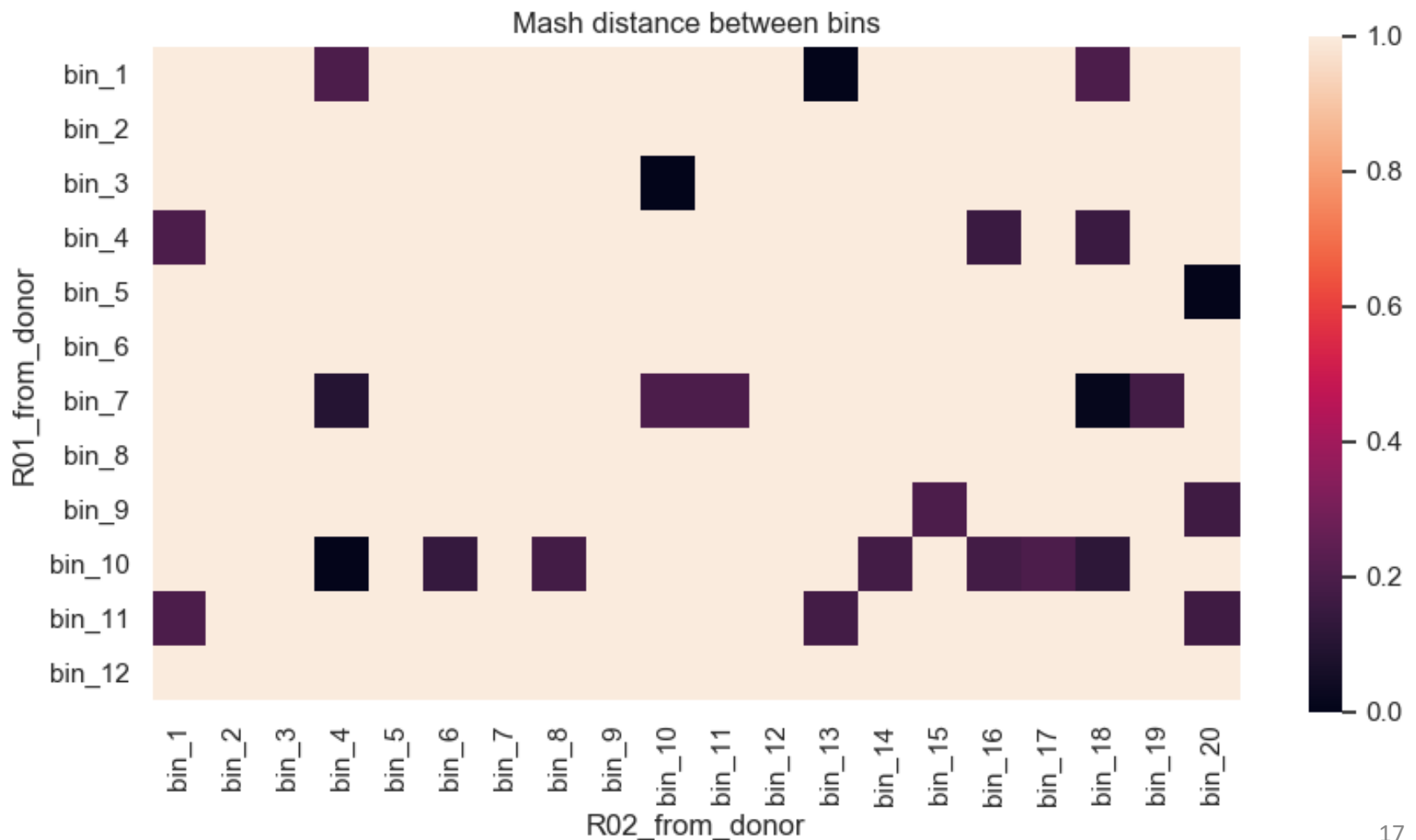
- Виды с экспоненциальным распределением с ошибками 1 на 1000 в прочтениях. Порог = 0,001



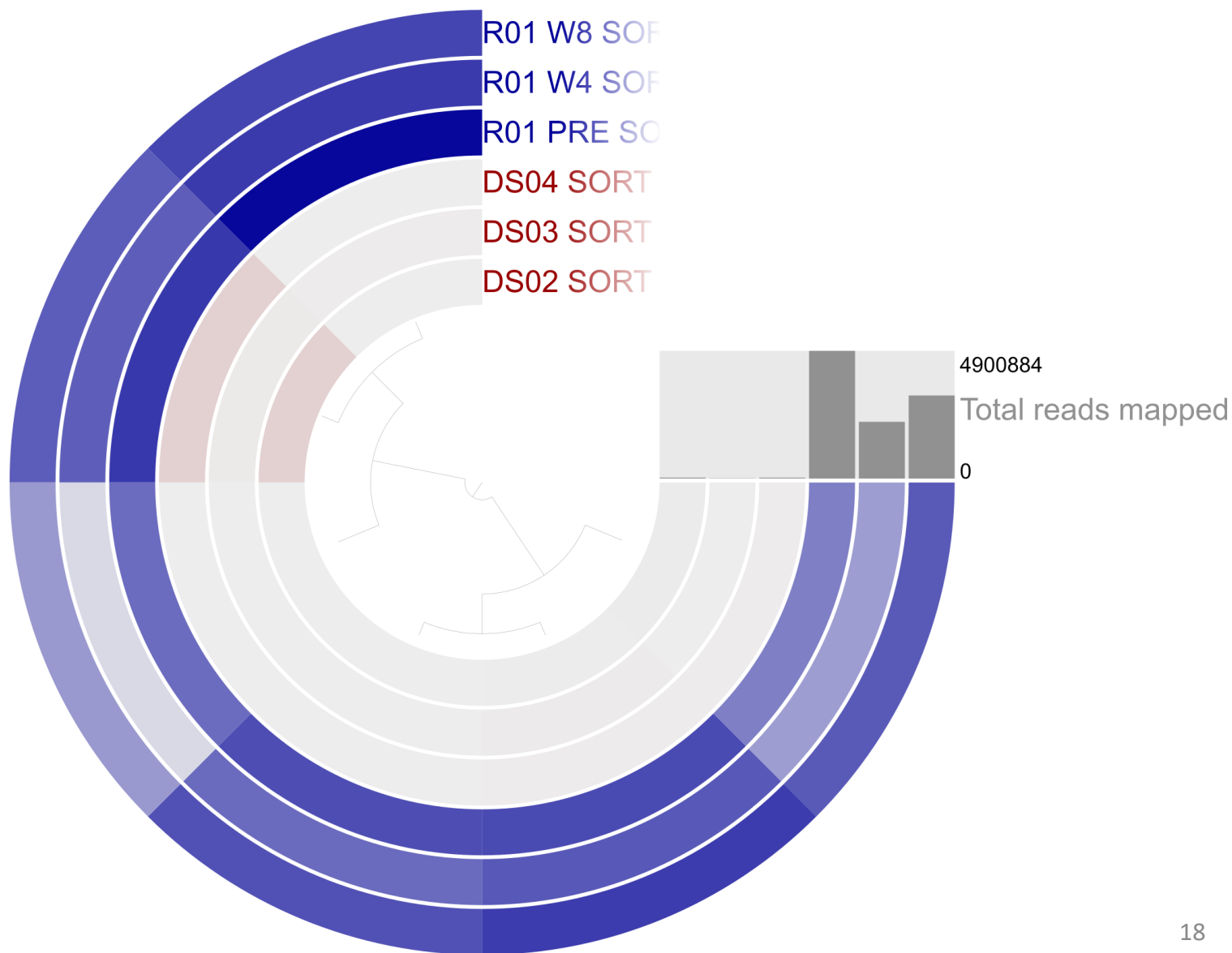
Прижилось от донора



Прижилось от донора – сравнение корзин



Сохранилось у реципиента



Выводы

- Разработан и реализован алгоритм классификации прочтений на категории
- Построена программа обработки серий метагеномов для исследования ТФМ
- Решение применяется для анализа реальных практических результатов в ФНКЦ ФХМ
- Победитель конкурса «За лучший доклад» на КМУ ИТМО 2019

СПАСИБО ЗА ВНИМАНИЕ!