Metafast

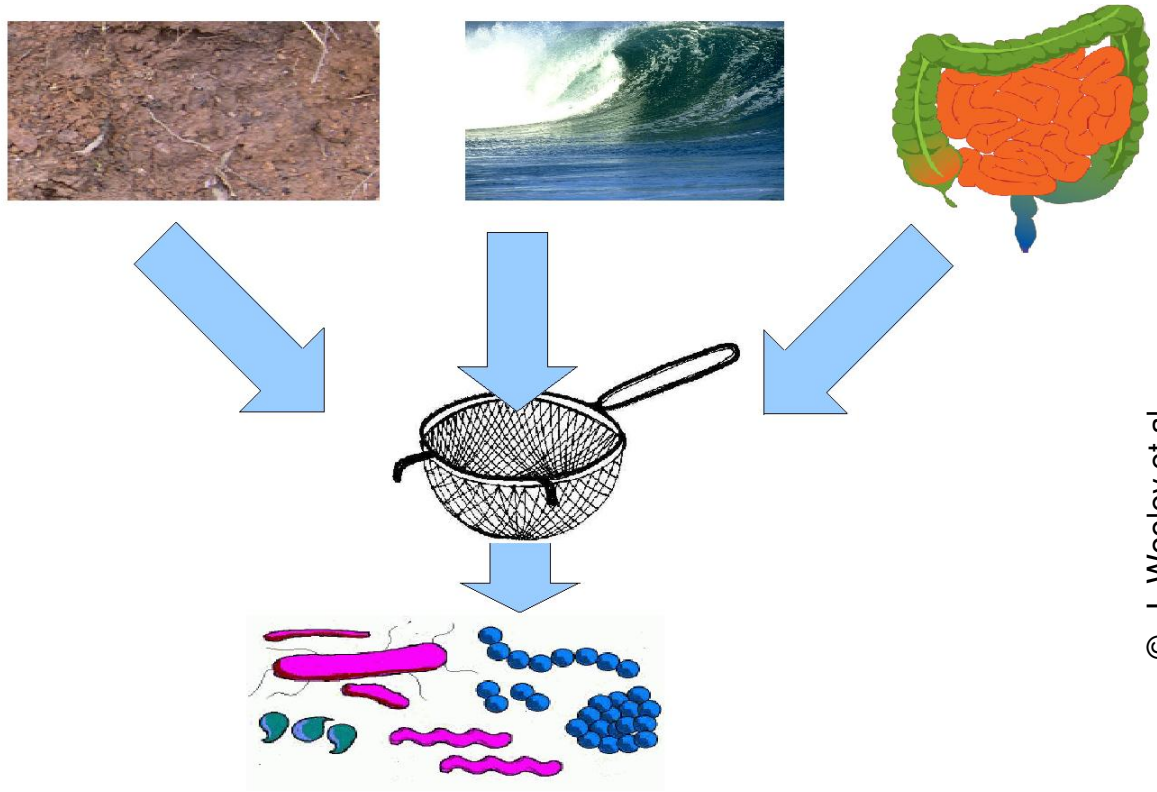
ПО для высокопроизводительного сравнительного анализа метагеномов

Сергей Казаков, Владимир Ульянцев, Вероника Дубинкина, Александр Тяхт, Дмитрий Алексеев

ITMO UNIVERSITY

ФЕДЕРАЛЬНОЕ МЕДИКО-БИОЛОГИЧЕСКОЕ АГЕНТСТВО
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
ФИЗИКО-ХИМИЧЕСКОЙ МЕДИЦИНЫ

Сравнительная метагеномика



- **Соотношение** между метагеномами:
  - Разные среды обитания
  - Разные моменты времени

Что интересно?

- Определение таксономического состава.
- Определение функционального состава.
- Разнородность сообществ(а).
- Альфа- и бета-разнообразие сообществ.

Особенности метагеномов

- Большая доля некультивированных бактерий (нет референса).
- Сложная структура сообщества.
- Большие объемы данных секвенирования.
- Насущен вопрос снижения размерности данных.

Существующие подходы

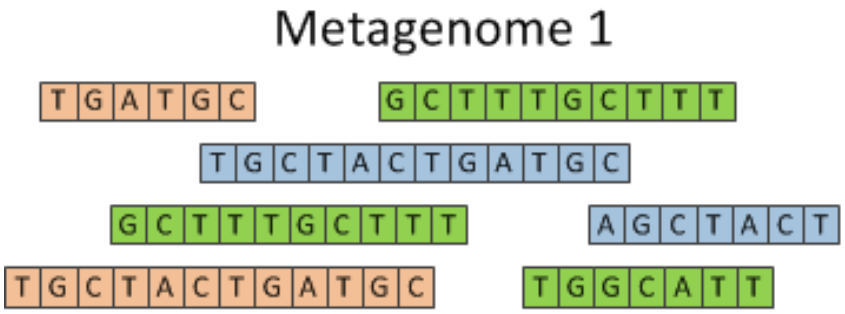
- Reference-based
  - Выравнивание на каталог известных геномов
- Assembly-based
  - Совместная сборка и дальнейший анализ
  - Metavelvet, Meta-IDBA, Ray, MetAmos, crAss.
- Composition-based
  - Анализ k-мерного спектра, нейронные сети, Марковские модели и др.
- K-mer binning
  - AbundanceBin, CompostBin, MaxBin, MetaCluster.

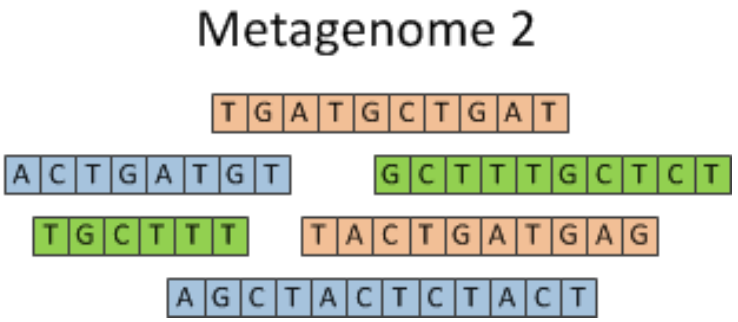
MetaFast

Метод, основанный на «упрощенной» сборке:

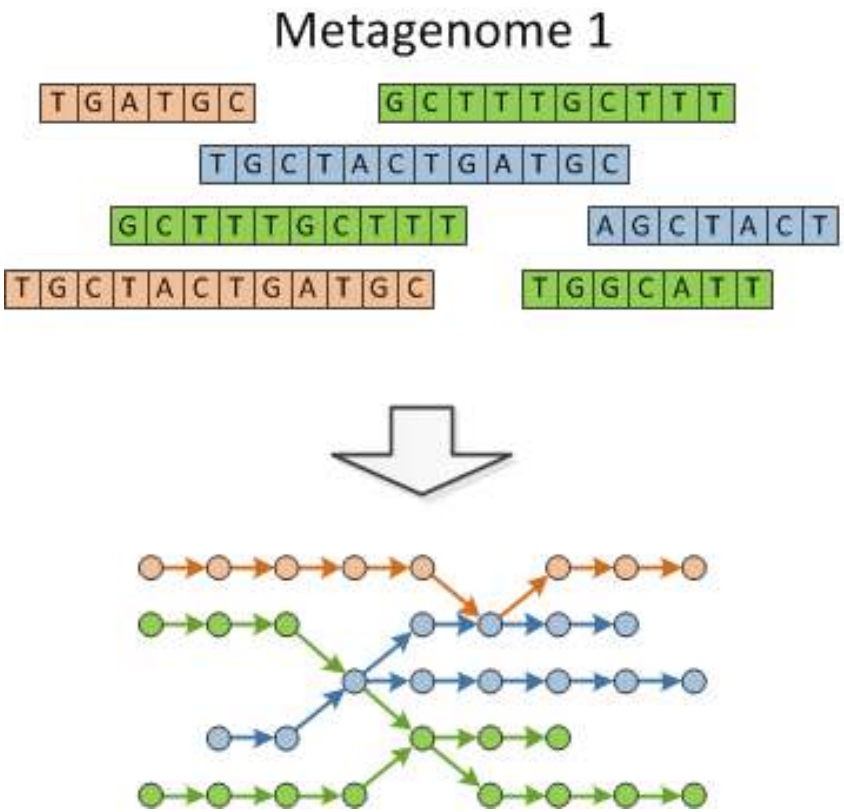
- Не требует референса.
- Высокопроизводительный.
- Выделение компонент и использование их как признаков.

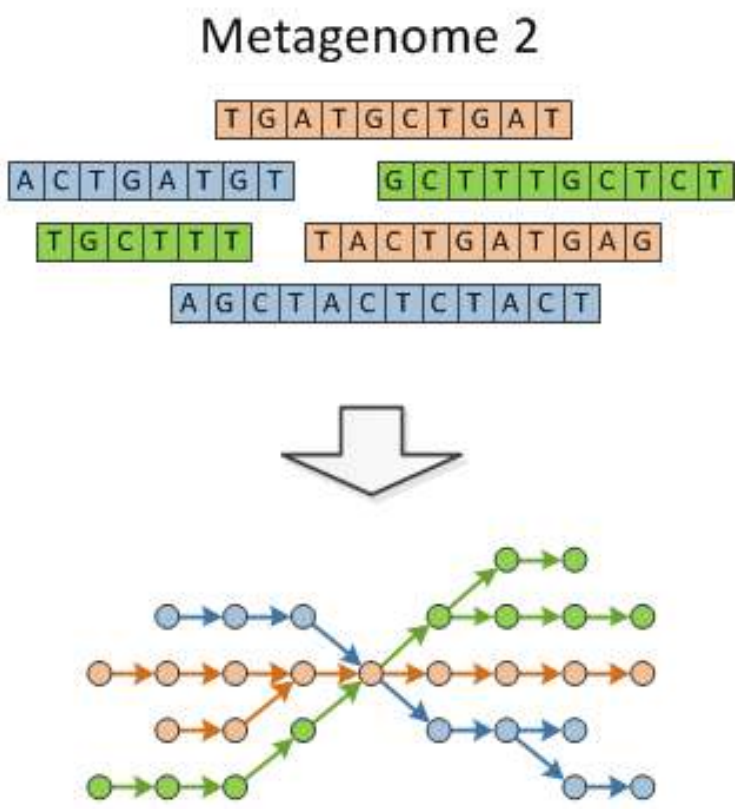
0. Исходные данные



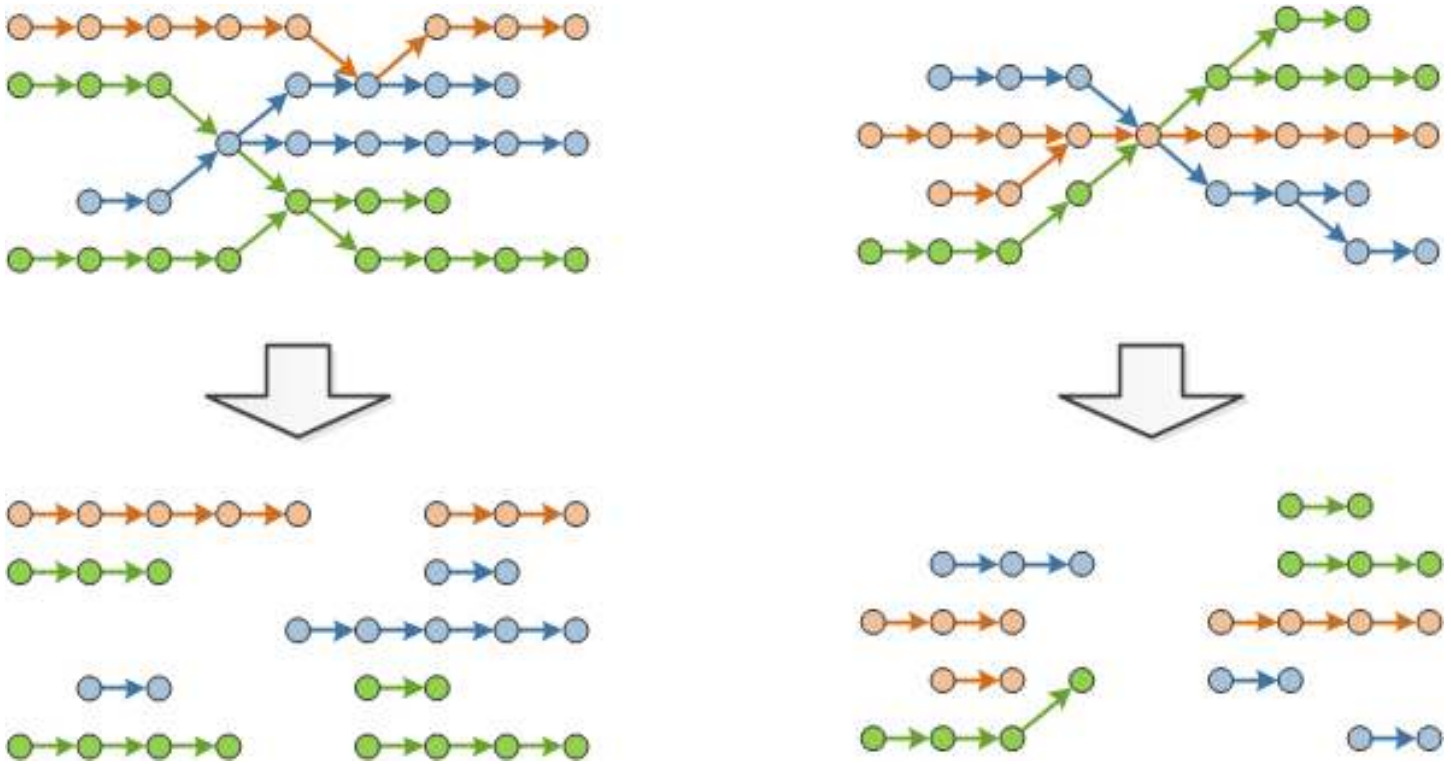


1. Построение графа де Брейна

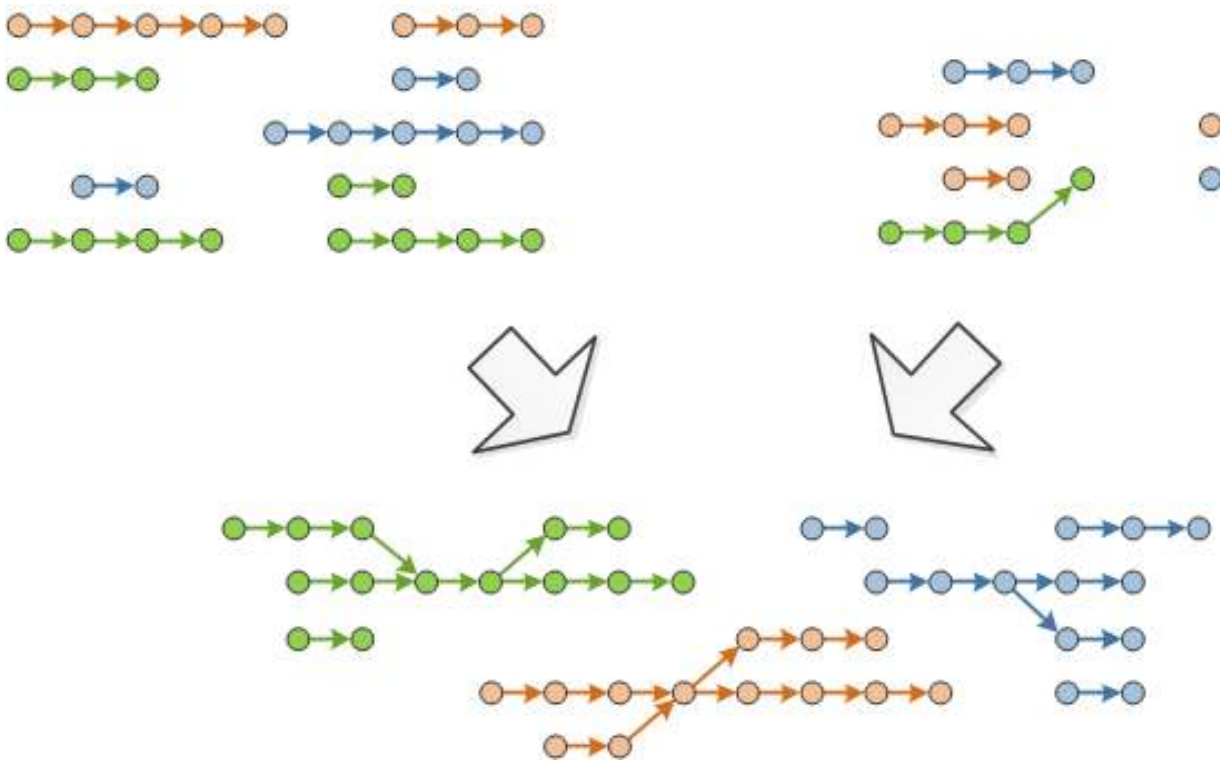




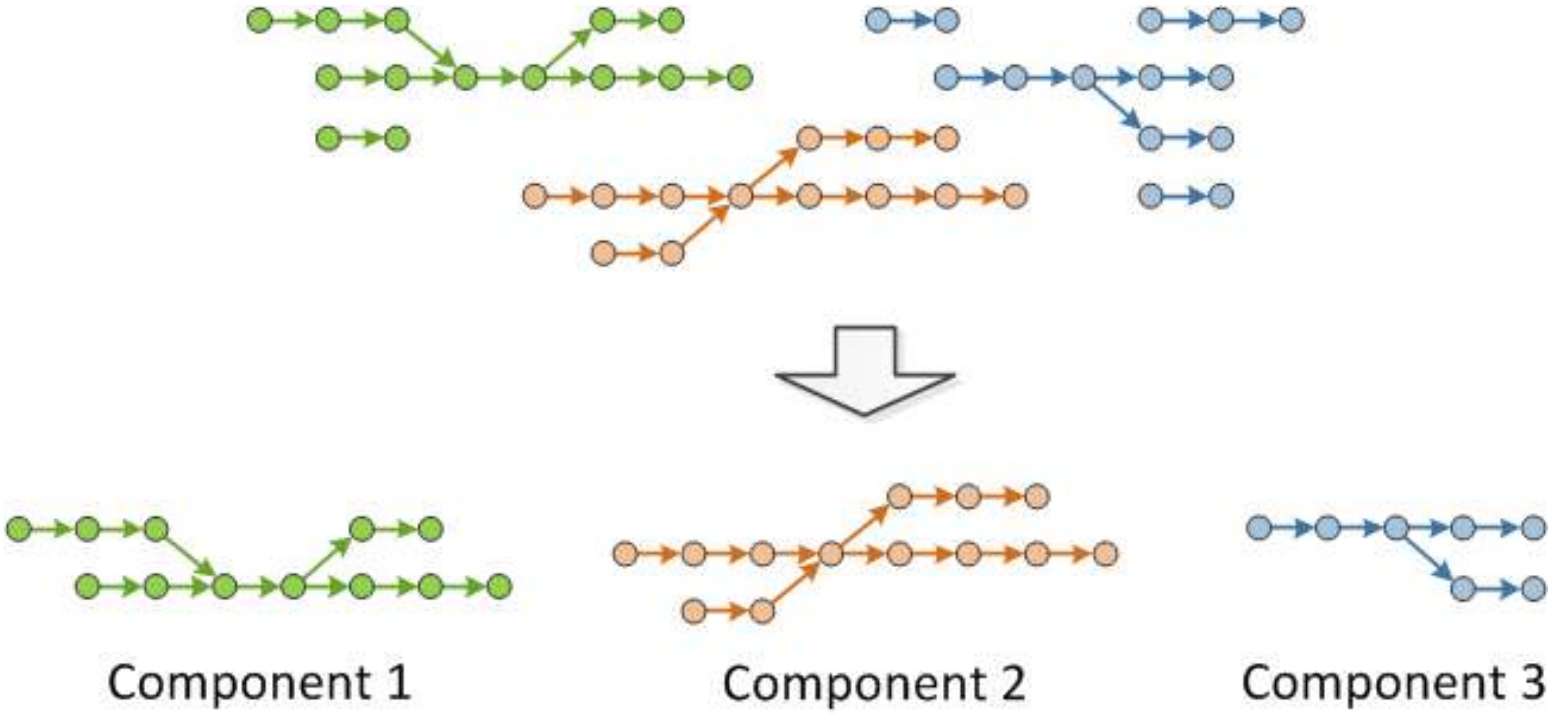
2. Выделение неветвящихся путей



3. Объединение в один граф де Брейна

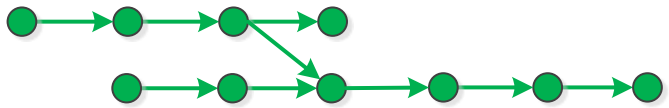


4. Выделение «подходящих» компонент



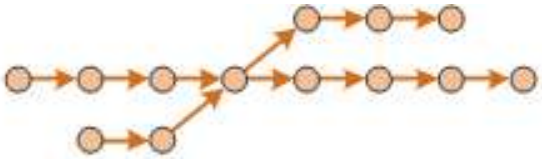
Компонента

- Набор к-меров
- B_1 <= размер <= B_2
- Большая компонента?
  - Итеративный алгоритм для выделения «основной» части

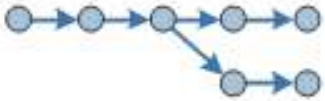


5. Вычисление характеристического вектора для каждого метагенома

Component 1



Component 2



Component 3

Metagenome 1 **(15, 0, 6)**

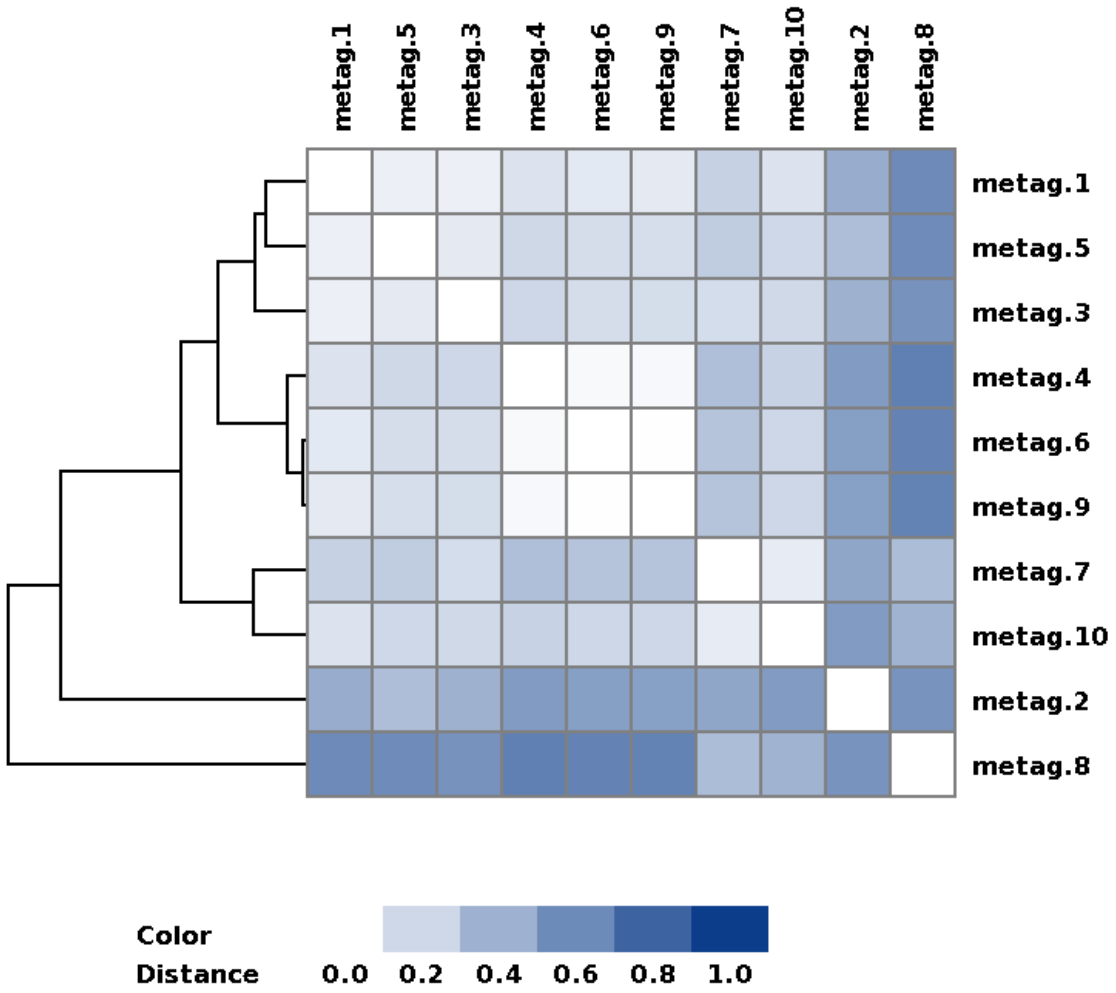
Metagenome 2 **(0, 7, 8)**

6. Вычисление расстояния между образцами

- Вычисление матрицы расстояния между образцами (расстояние Брея-Кёртиса):

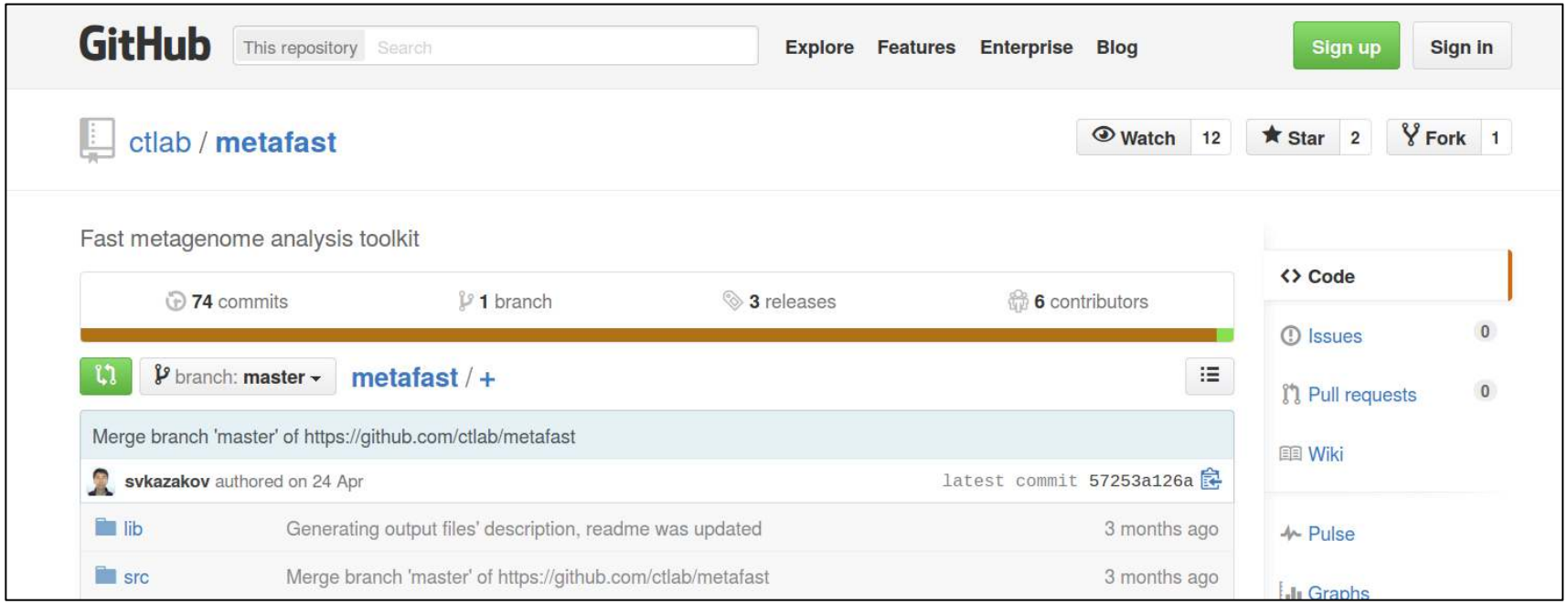
#	tinytest_A	tinytest_B
tinytest_A	0.0	0.0909
tinytest_B	0.0909	0.0

7. Построение HeatMap и дендрограмы



Реализация

- На *Java* (кросс-платформенное ПО)
- Открытое ПО (open-source)
- http://github.com/ctlab/metafast



Эксперименты

1. Искусственные метагеномы:

- Сравнение с расстоянием по таксономическому составу.
- Сравнение с совместной сборкой.

2. Метагеномы микробиоты кишечника:

- Сравнение с расстоянием по таксономическому и функциональному составу.

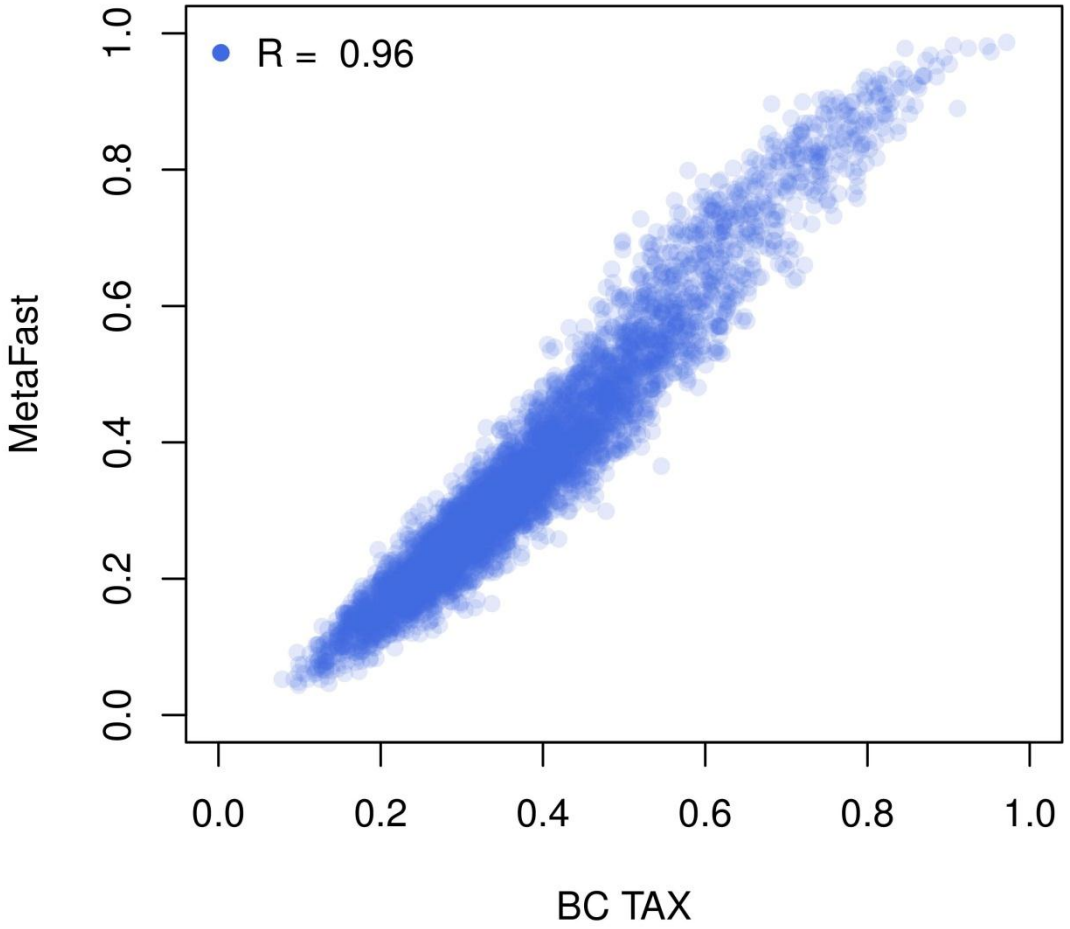
3. Метагеномы микробных сообществ в подземной шахте.

Эксперимент 1

- **100 метагеномов** получены путем нарезки в случайных пропорциях **10 бактерий**.
- Посчитана истинная матрица расстояний (по таксономическому составу).

Эксперимент 1

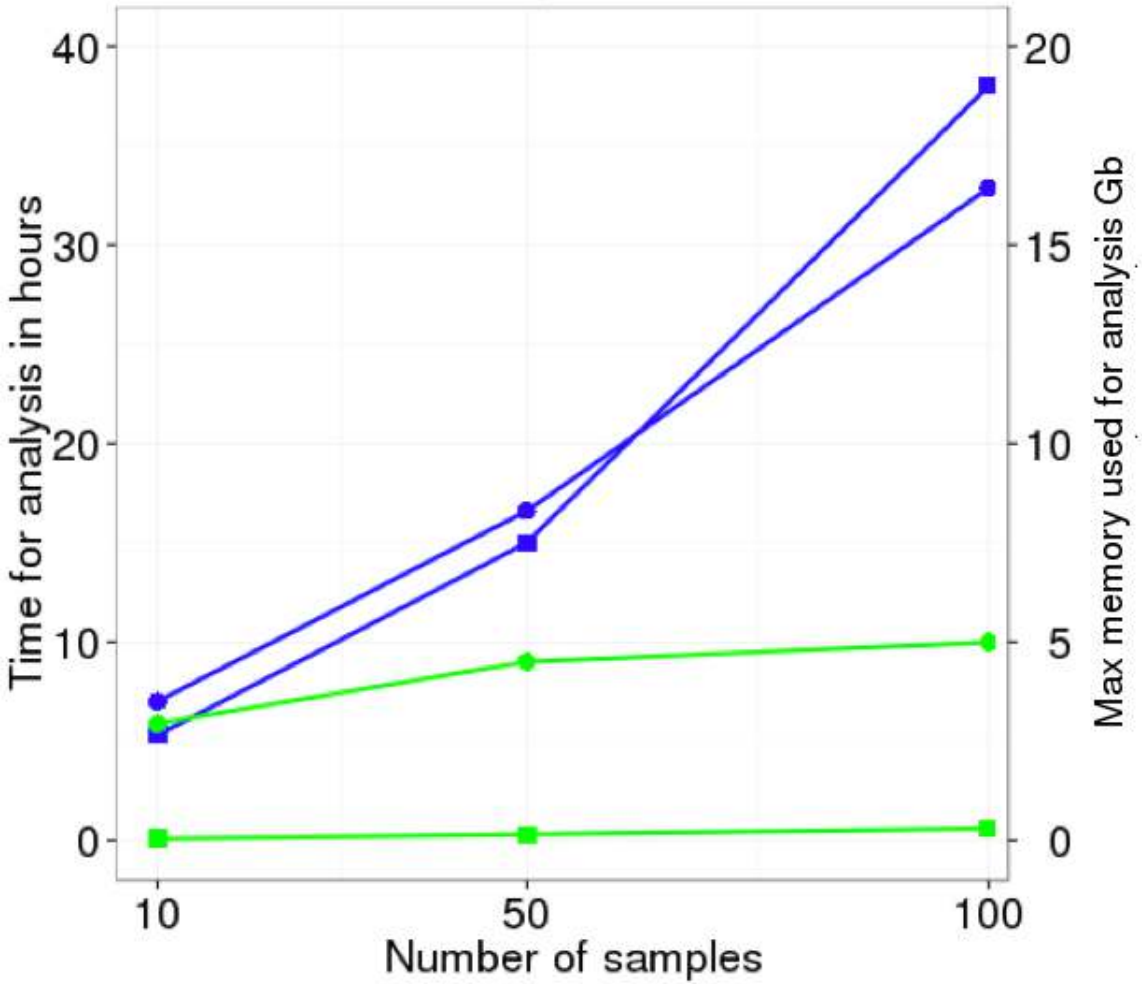
- Тест Мантеля сходства матриц: корреляция Спирмена **r=0.96**, **p=0.001**



Эксперимент 1 – сравнение с совместной сборкой

Charact. / Method	10 samples		50 samples		100 samples	
	time	memory	time	memory	time	memory
MetaFast	6 min	3.0 Gb	24 min	4.5 Gb	46 min	5.0 Gb
Combined assembly	5 h	3.5 Gb	15 h	8.3 Gb	38 h	16.4 Gb
Velvet	41 min	3.5 Gb	3.5 h	8.3 Gb	7 h	16.4 Gb
bowtie	4.5 h	1.5 Gb	10 h	1.5 Gb	26.5 h	1.5 Gb

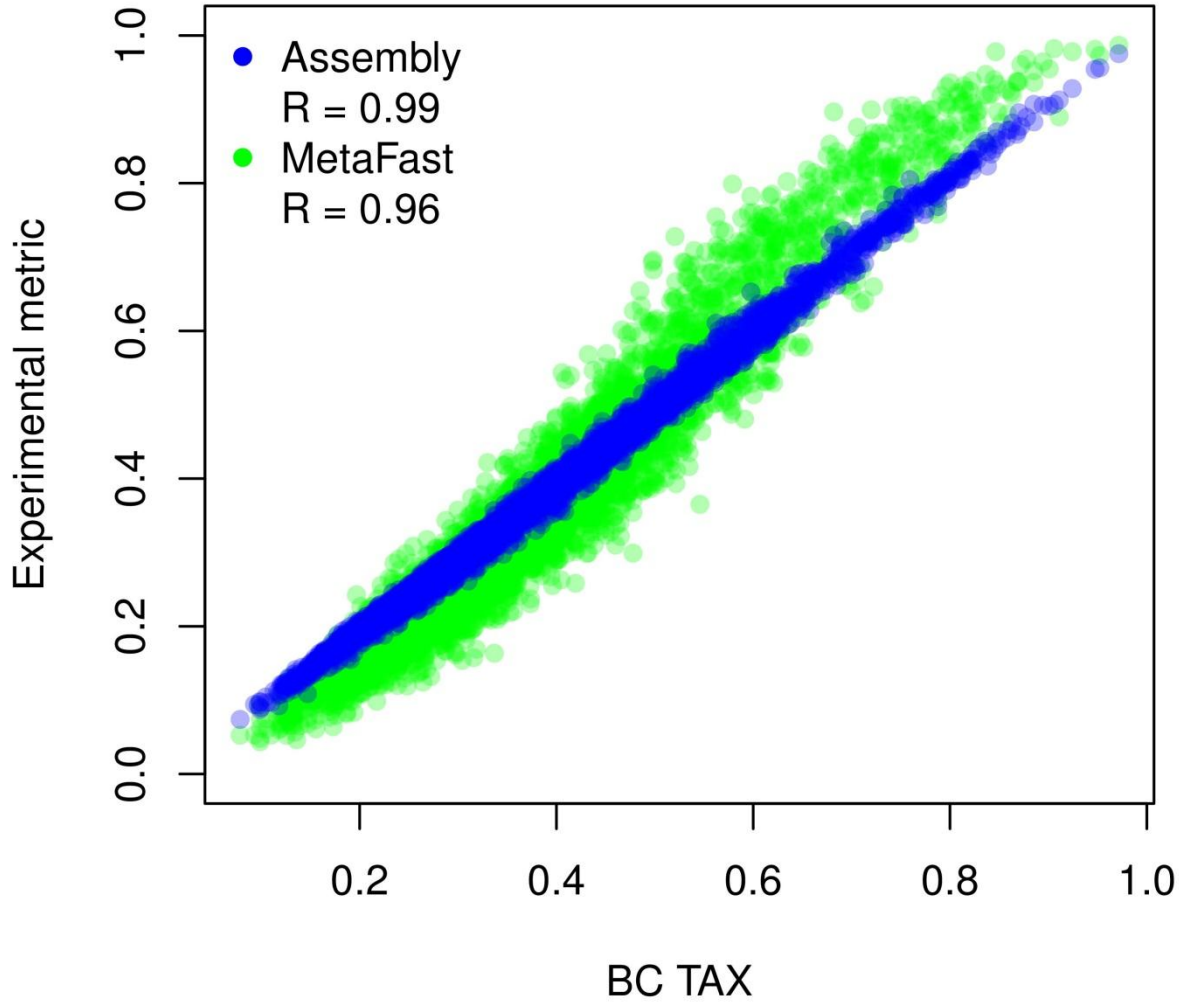
Эксперимент 1 – сравнение с совместной сборкой



- Совм. сборка
- Metafast

- Time
- Memory

Эксперимент 1 – сравнение с совместной сборкой



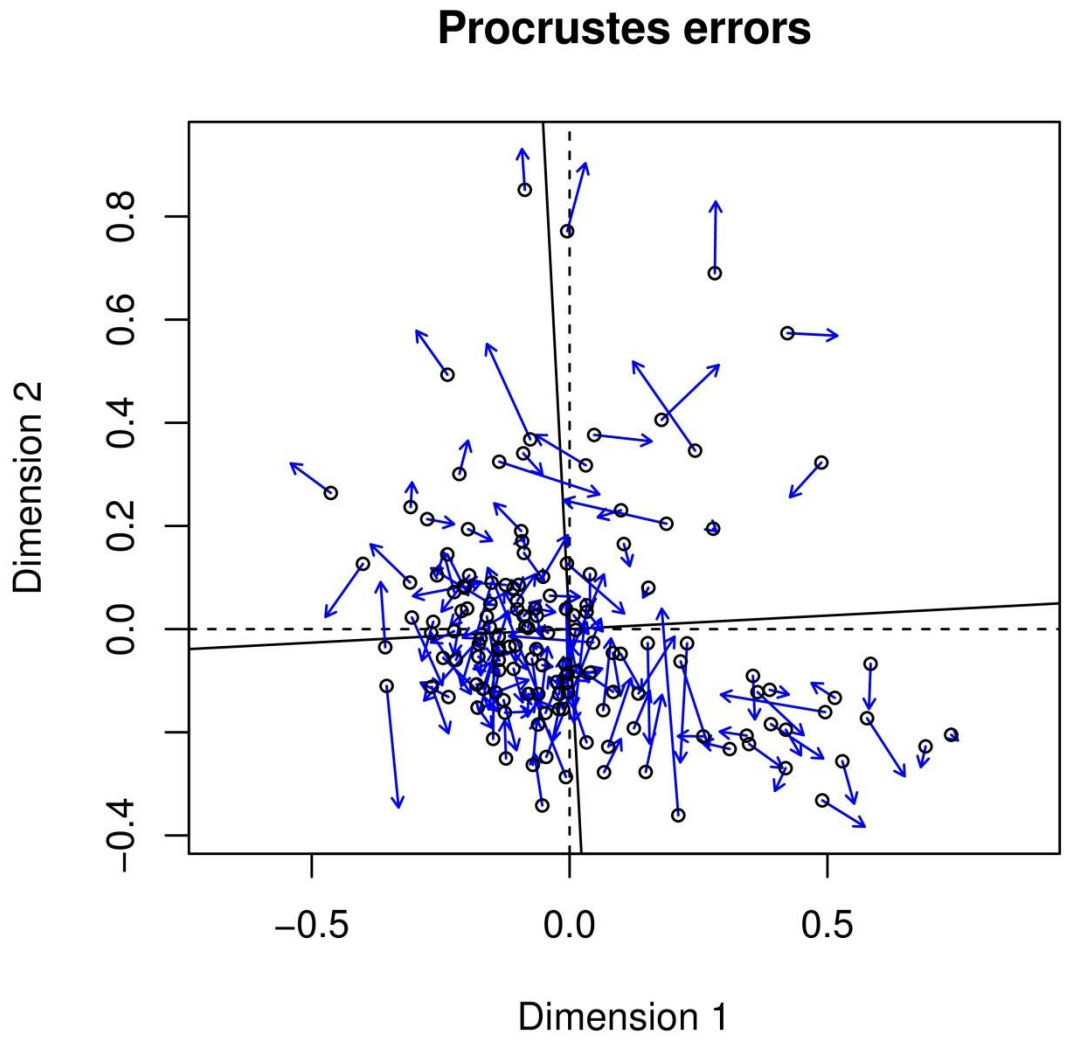
Эксперимент 2

- **157 образцов** метагеномов микробиоты кишечника человека.
- 580 Гб сжатых fastq файлов ($7.8 * 10^{12}$ чтений).
- Обсчет Metafast'а: 34 часа, используя 20 ядер процессора и 90 Гб ОЗУ.

Эксперимент 2

Comp. / Charact.	Taxonomic composition	Functional composition
r	0.91	0.78
p-value	0.001	0.001

Эксперимент 2



Направления дальнейших исследований

- Доработка экспериментов с метагеномами микробных сообществ в подземной шахте.
- Использование разных стратегий выделения компонент, анализ зависимости получаемых результатов от выбора стратегии.
- Поддержка графов де Брейна для $k > 31$.

Спасибо за внимание!

http://github.com/ctlab/metafast

С. Казаков

В. Ульянцев

В. Дубинкина

А. Тяхт

Д. Алексеев

ITMO UNIVERSITY