

# Поиск генов симбиотического комплекса фиксаторов азота

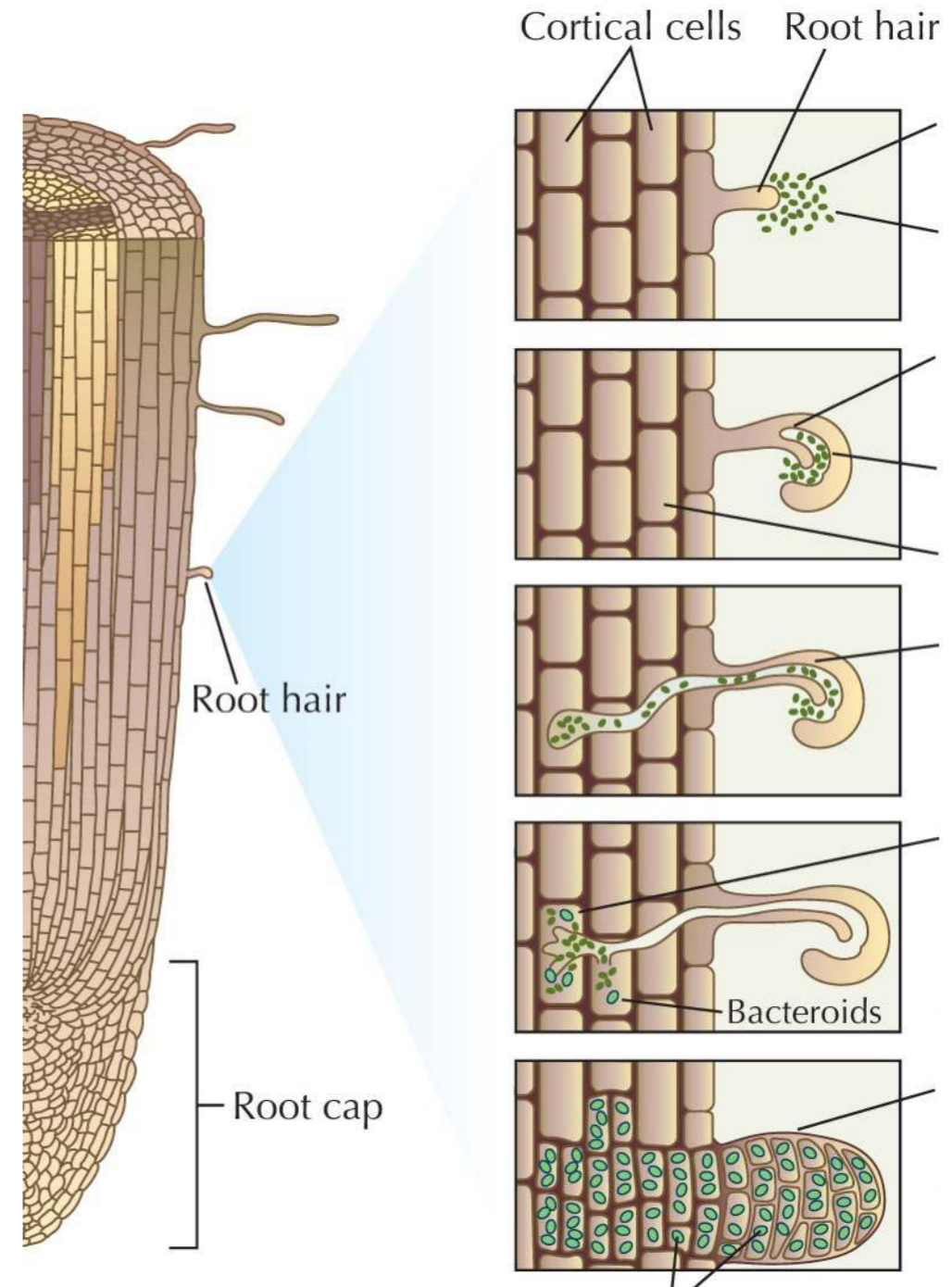
Исполнитель: Илья Корвиго

Научный руководитель: Евгений Андронов

Организация: ГНУ ВНИИСХМ РАСХН

# Самый сложный симбиоз

- Комплексы генов, ответственные за распознавание растения и перекрестного сигналинга.
- Системы регуляции нитрогеназного комплекса.
- Гены морфогенеза бактериода.
- Козэволюция с растением.



- Пути возникновения этих систем по большому счету остаются загадкой.
- Высокая дивергенция нуклеотидных последовательностей даже в пределах эволюционно близкой клады прокариот.
- Возникший однажды коровый комплекс, вероятно, неоднократно горизонтально переносился.
- Из вышесказанного естественно следует сложность поиска предковых форм генов.

# Задачи

- Разработать способ эффективного поиска далеко дивергировавших генов прокариот в собранных геномах.
- Неплохо иметь возможность искать фрагменты генов в метагеномных библиотеках.
- Реализовать это в виде удобного и интуитивного интерфейса.

# Что имеется

## **Blast**

- Основан на локальном выравнивании и хэшировании. Эвристический алгоритм - неспецифичен.
- Осложненная интерпретация получаемых на выходе данных.

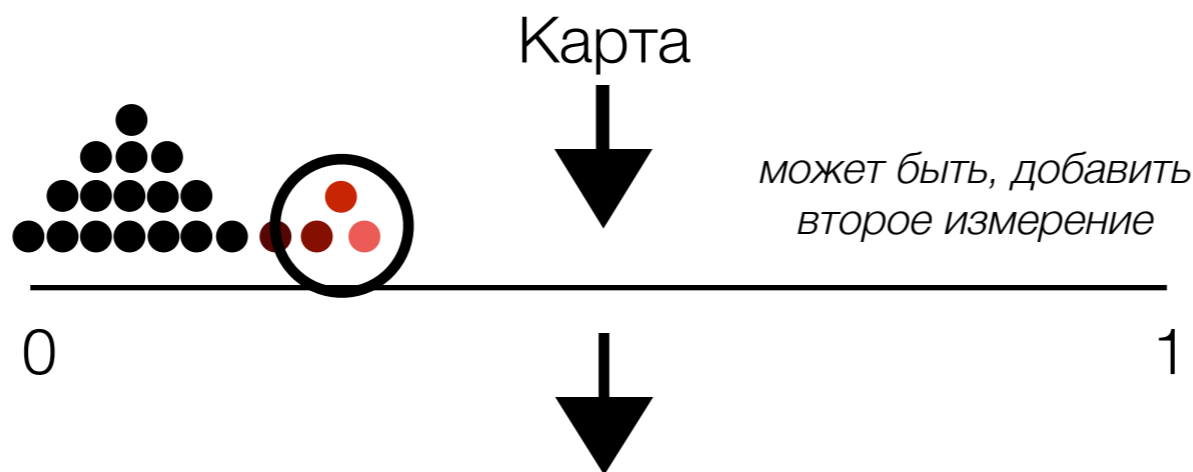
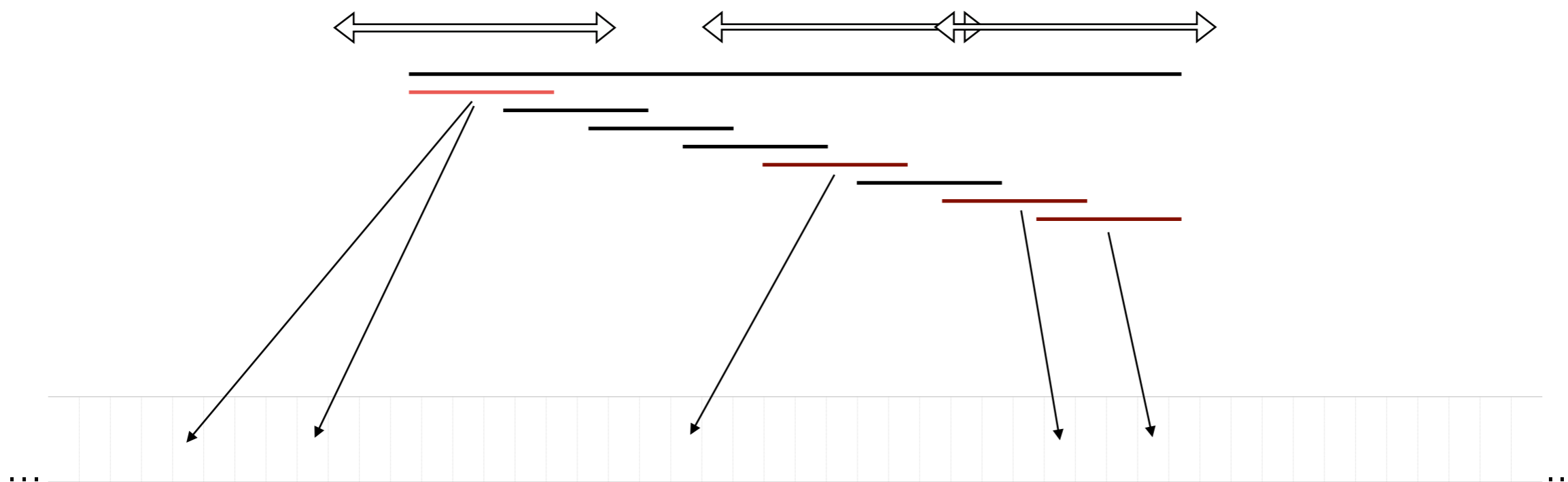
## **HMMer**

- Основан на скрытых моделях Маркова - специфичен.
- Лучше находит далеко разошедшиеся последовательности.
- Медленнее. Не имеет аналога tblastn.

# Что делается

- Полное хеширование k-меров тренировочной библиотеки на одной хэш-карте.
- Построение дополнительных k-меров на основании поиска позиций с максимальной энтропией в цепи Маркова.
- Картирование исследуемого генома (или фрагментов метагенома).
- Вычленение возможных аминокислотных последовательностей методом скользящего окна.
- Передача последовательностей HMMer'y.

# Картирование целевой последовательности



поиск точек из одной окрестности  
внутри выделенного кластера

# Куда же без проблем

- Удаление промоторов.
- Фильтрация шума
- Статистически оправданная стратегия выбора зоны поиска рамки.



# Что предстоит

- Решить проблемы (или вставить костыли).
- Распараллелить.
- Переписать узкие места на C или, на худой конец, Cython.
- Связать свой тул с HMMer'ом.
- Сделать окна или хотя бы интерактивный интерфейс командной строки - пользователь должен пользоваться, а не разбираться, почему не взлетает.

Спасибо за внимание