

# Эволюционная геномика симбиотических микроорганизмов

Выполнил: Маловичко Юрий

Руководитель: Андронов Е.Е., Лаборатория микробиологического  
мониторинга и биоремедиации почв, ВНИИСХМ

# Немного об объекте работы (еще раз)

- *Sinorhizobium meliloti*  
(Bacteria:Proteobacteria:Alphaproteobacteria:  
Rhizobiaceae) – вид грамотрицательных  
почвенных бактерий
- Образует симбиотические клубеньки на  
корнях растений (e.g. представители рода  
*Medicago*)
- Возможно, в пределах вида существуют две  
генетических линии, демонстрирующие  
сцепление между аллелями генов *leu* и  
*betCB*



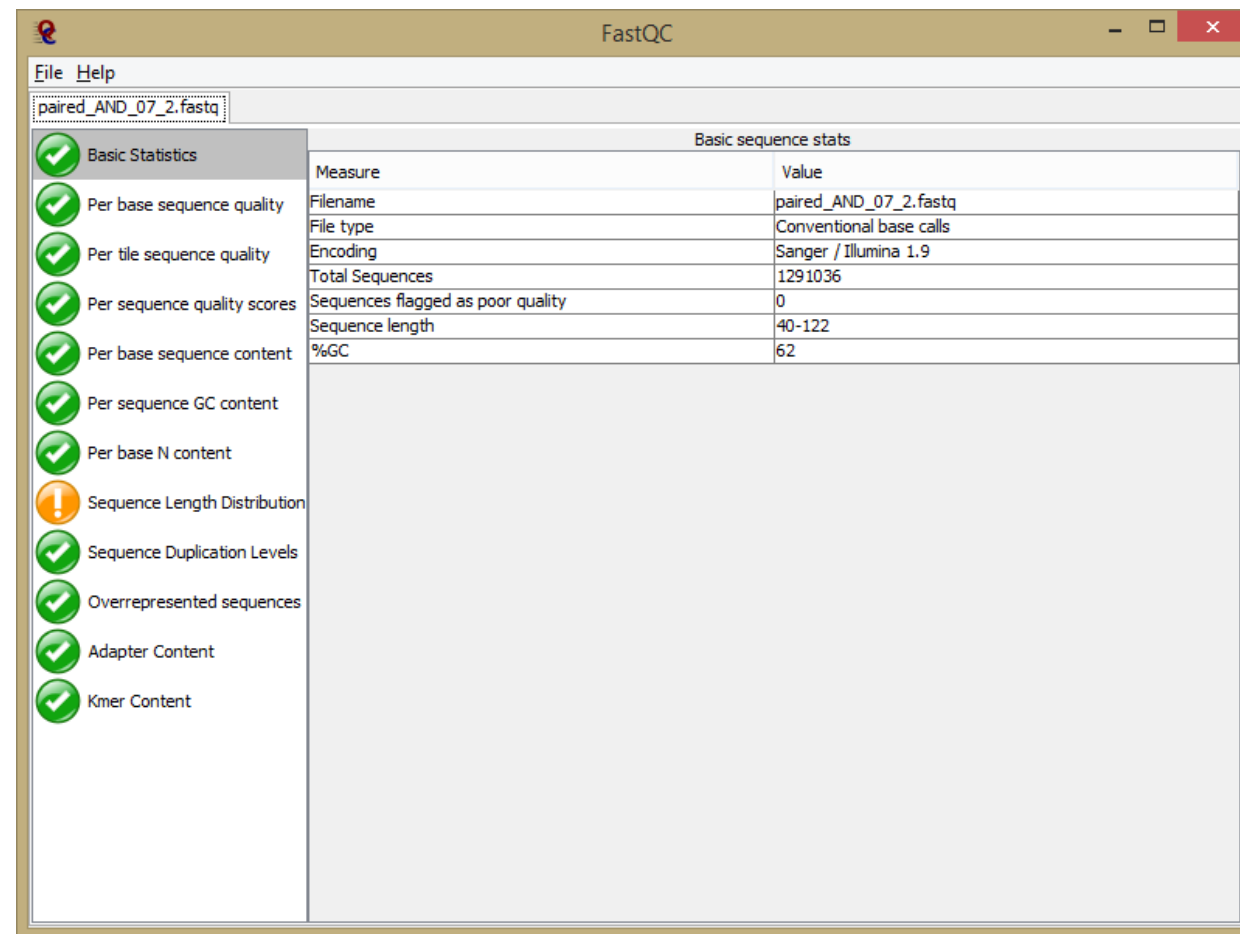
(Взято с <http://www.sinorhizobium.org/> )

# Цели и задачи работы

- Цель работы – проверить гипотезу о существовании двух линий (А и В) в пределах вида *Sinorhizobium meliloti*
- Задачи:
  - Собрать и аннотировать геномы представителей 12 изолятов *S. meliloti* (по 6 на каждую из предполагаемых линий)
  - Типировать изоляты методом MLST и построить филогенетическое дерево
  - На основании полученных данных сделать вывод о наличии/отсутствии двух выраженных линий в пределах вида

# Шаг 1: тримминг

- Использованная программа: Trimmomatic
  - На вход подавались парные прочтения, сделанные на платформе Illumina MiSeq
  - Проверка качества на выходе производилась при помощи FastQC
  - Сохранялись риды длиной до 400 нуклеотидов
  - Основные проблемы:
    - Sequence Length Distribution: неодинаковая длина последовательностей в файле на выходе
    - Kmer Content: неодинаковая представленность некоторых k-меров (следы повторов и мобильных элементов?)



Пример «хорошего» результата работы Trimmomatic в FastQC

## Шаг 2: сборка генома

- Сборка геномов de novo производилась при помощи ассемблера SPAdes (режим --careful)
- Среднее количество контигов на выходе: 489 (минимум – 324, максимум – 708)

# Шаг 3: аннотация

- Два подхода к аннотации первых 6 геномов:
  - RAST 2.0 с исходными настройками;
  - Prokka: база данных собрана из последовательностей, взятых с UniProt для f. Rhizobiaceae (1309039 последовательностей), приоритетные последовательности взяты из SwissProt для g. Sinorhizobium (1984 последовательности).
- В результате остановились на использовании Prokka

Генов обнаружено	RAST	Prokka
Белок-кодирующих	6679	6362
-аннотированных	4640	4023
-предполагаемых	2039	2339
Нетранслируемых	52	60

Средние значения результатов работы RAST и Prokka  
(данные взяты для изолятов S7-S12)

# Типирование и построение филогенетического дерева

- Используемый метод: MLST (Multilocus Sequence Typing)
- Суть: типирование бактерий производится по множественным (7+) локусам, соответствующим генам домашнего хозяйства (HKG).
- В работе 2006 года для *S. meliloti* был предложен набор из 10 генов:
  - *ssd*                      *edD*
  - *gap*                        *glnD*
  - *gnd*                        *nuoE1*
  - *ordLO2*                    *recA*
  - *sucA*                        *zwf*

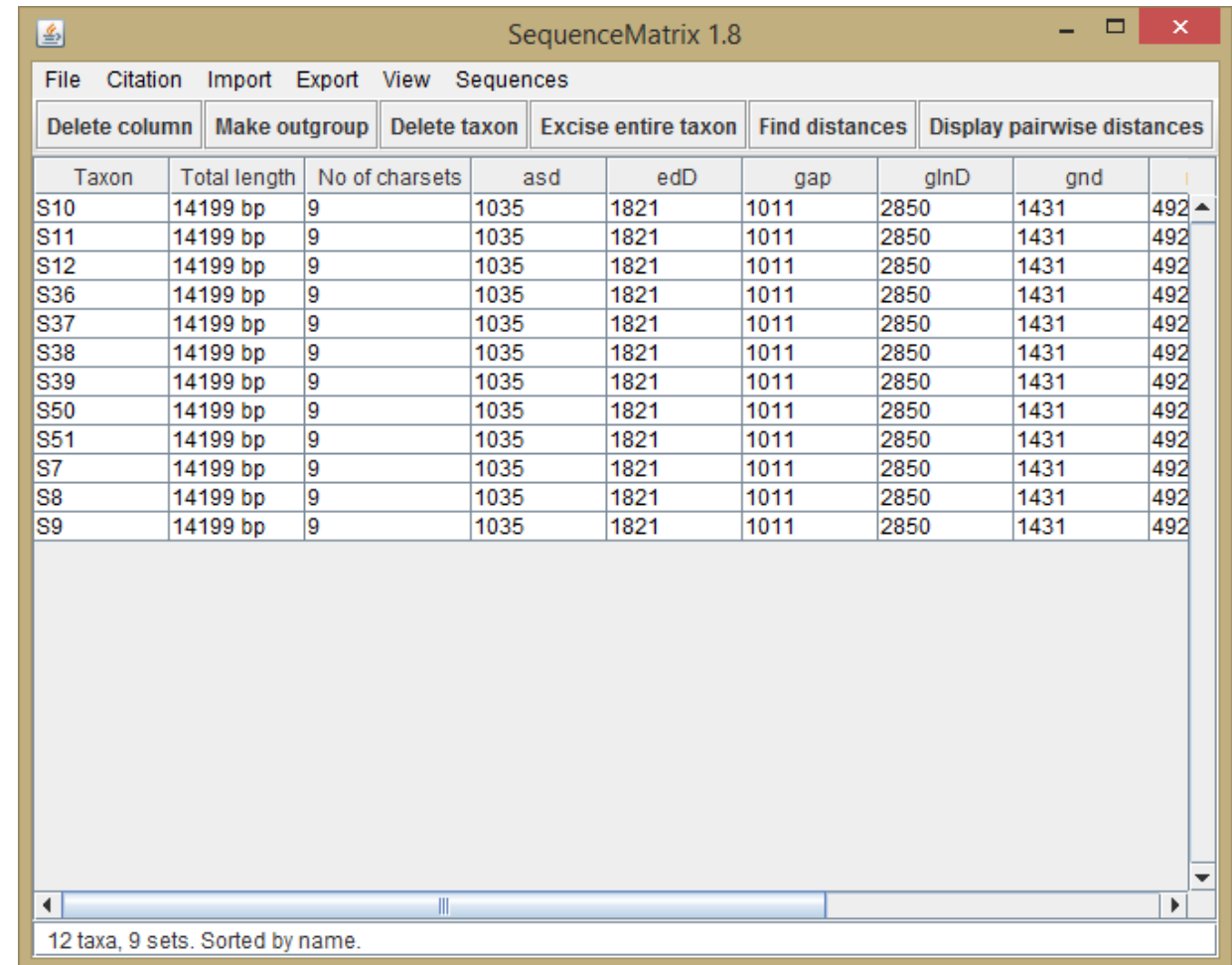
# Поиск и выравнивание последовательностей

- Необходимые гены находились в файлах с аннотациями при помощи python-скрипта
- Лocus ordL2 отсутствовал во всех 12 аннотациях:
  - Первоначально заменили ordL2 на sadH
  - В итоге отказались от локусов оксидоредуктаз в пользу модели с 9 локусами
- Выравнивание: MEGA7



# Создание конкатенированной матрицы

- Выравнивания собирались в единую конкатенированную матрицу при помощи SequenceMatrix
- Матрица экспортировалась в формате nexnaked для дальнейшей работы в MrBayes 3.2.6



The screenshot shows the SequenceMatrix 1.8 application window. The main area contains a table with the following data:

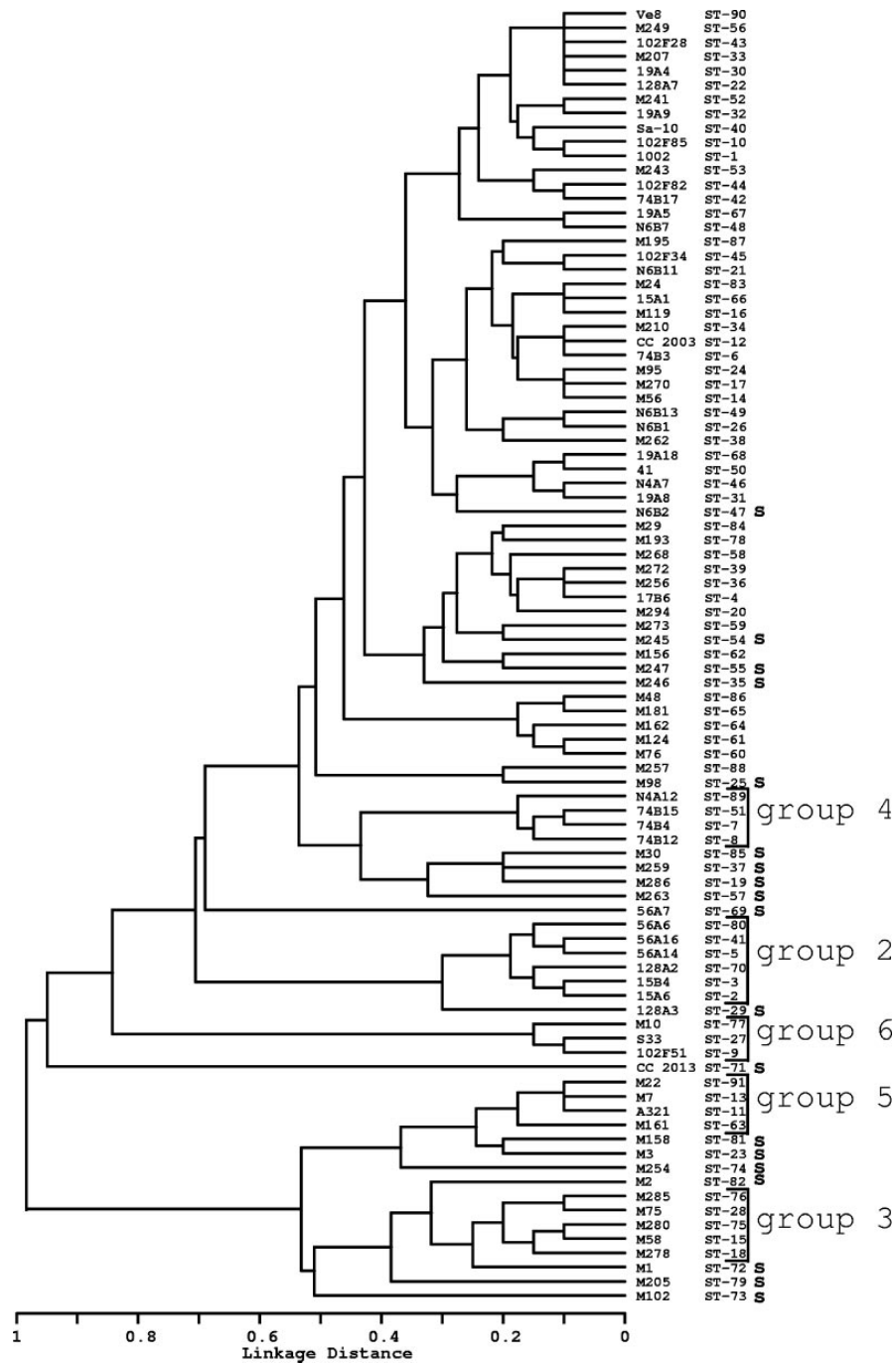
Taxon	Total length	No of charsets	asd	edD	gap	glnD	gnd	
S10	14199 bp	9	1035	1821	1011	2850	1431	492
S11	14199 bp	9	1035	1821	1011	2850	1431	492
S12	14199 bp	9	1035	1821	1011	2850	1431	492
S36	14199 bp	9	1035	1821	1011	2850	1431	492
S37	14199 bp	9	1035	1821	1011	2850	1431	492
S38	14199 bp	9	1035	1821	1011	2850	1431	492
S39	14199 bp	9	1035	1821	1011	2850	1431	492
S50	14199 bp	9	1035	1821	1011	2850	1431	492
S51	14199 bp	9	1035	1821	1011	2850	1431	492
S7	14199 bp	9	1035	1821	1011	2850	1431	492
S8	14199 bp	9	1035	1821	1011	2850	1431	492
S9	14199 bp	9	1035	1821	1011	2850	1431	492

At the bottom of the window, a status bar indicates: 12 taxa, 9 sets. Sorted by name.

# Построение дерева методом VI

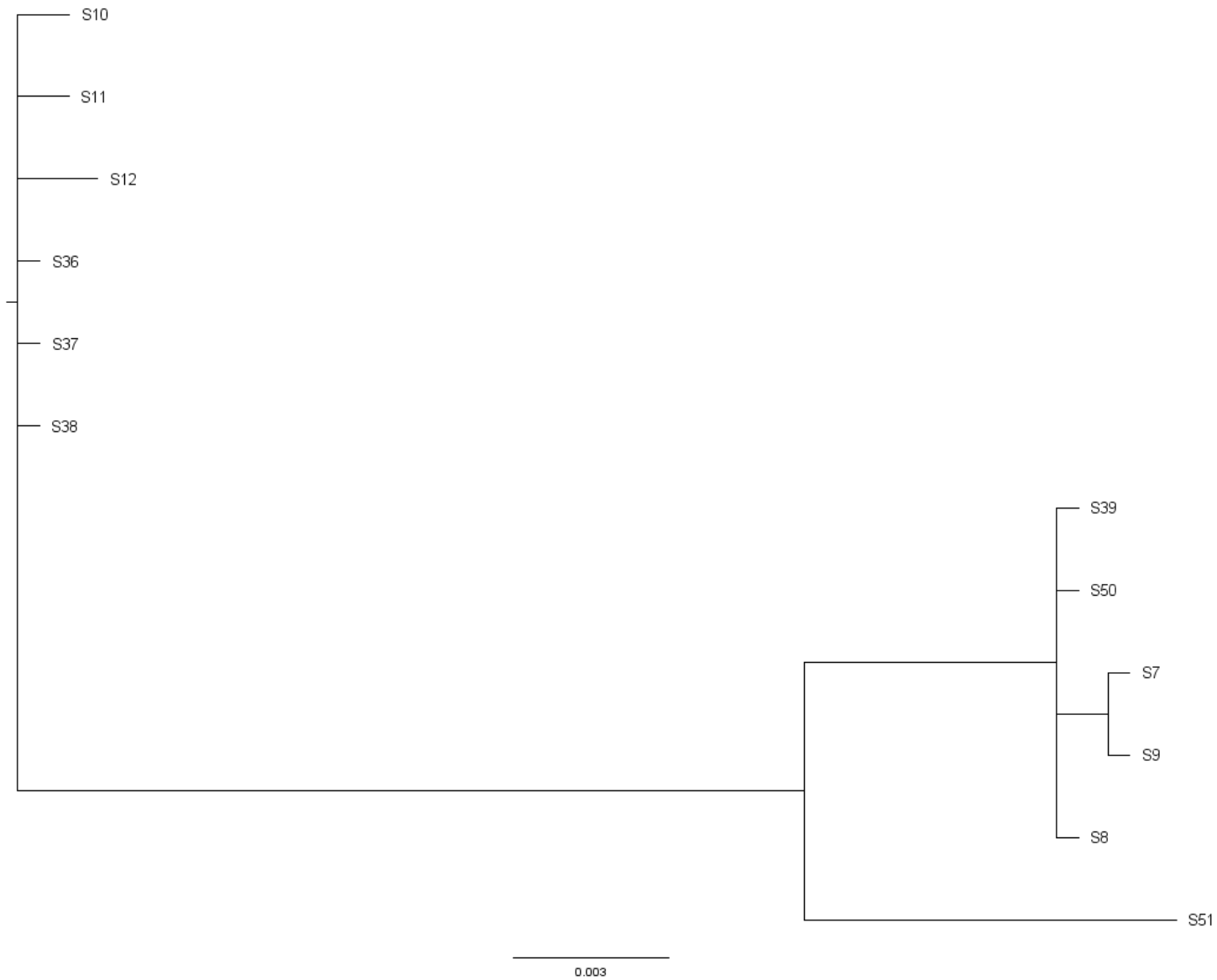
- Модели замещения подбирались отдельно для каждого гена при помощи JModelTest2
- В случае, если AIC и BIC давали разные результаты, выбор делался в пользу BIC
- В результате все локусы разделились на подпадающие под модель F81 модель HKY
- Количество итераций VI: 50 миллионов

# Чего мы ожидали



Berkum P. Van, Elia P., Eardly B.D. Multilocus sequence typing as an approach for population analysis of Medicago-nodulating rhizobia // J. Bacteriol. 2006. T. 188. № 15. C. 5570–5577.

# Что мы получили



# Выводы

- Дидактическая ценность работы:
  - Освоены инструменты для обработки и анализа геномных сиквенсов прокариот и работа с ними при помощи терминала Linux
  - Освоены методы подготовки данных для филогенетического анализа и построение филогенетических деревьев методом BI в MrBayes
- Научная ценность работы:
  - (сомнительная)
  - Возможно, предложенные в статье локусы не являются универсальными для кластеризации методом NLST или не подходят в данном случае

Спасибо за внимание!

