

Новосибирский Государственный Университет  
Факультет естественных наук  
Кафедра информационной биологии

Выпускная квалификационная бакалаврская работа, на тему  
**«Метагеномный анализ микробных сообществ геотермальных  
источников острова Кунашир»**

Выполнил: Коржук А.В. 4 курс  
Научный руководитель: к.б.н. Розанов А.С.  
Институт Цитологии и Генетики СО РАН  
Лаборатория молекулярных биотехнологий

Новосибирск 2017

# Геотермальные источники

Большой призматический источник (YNP)



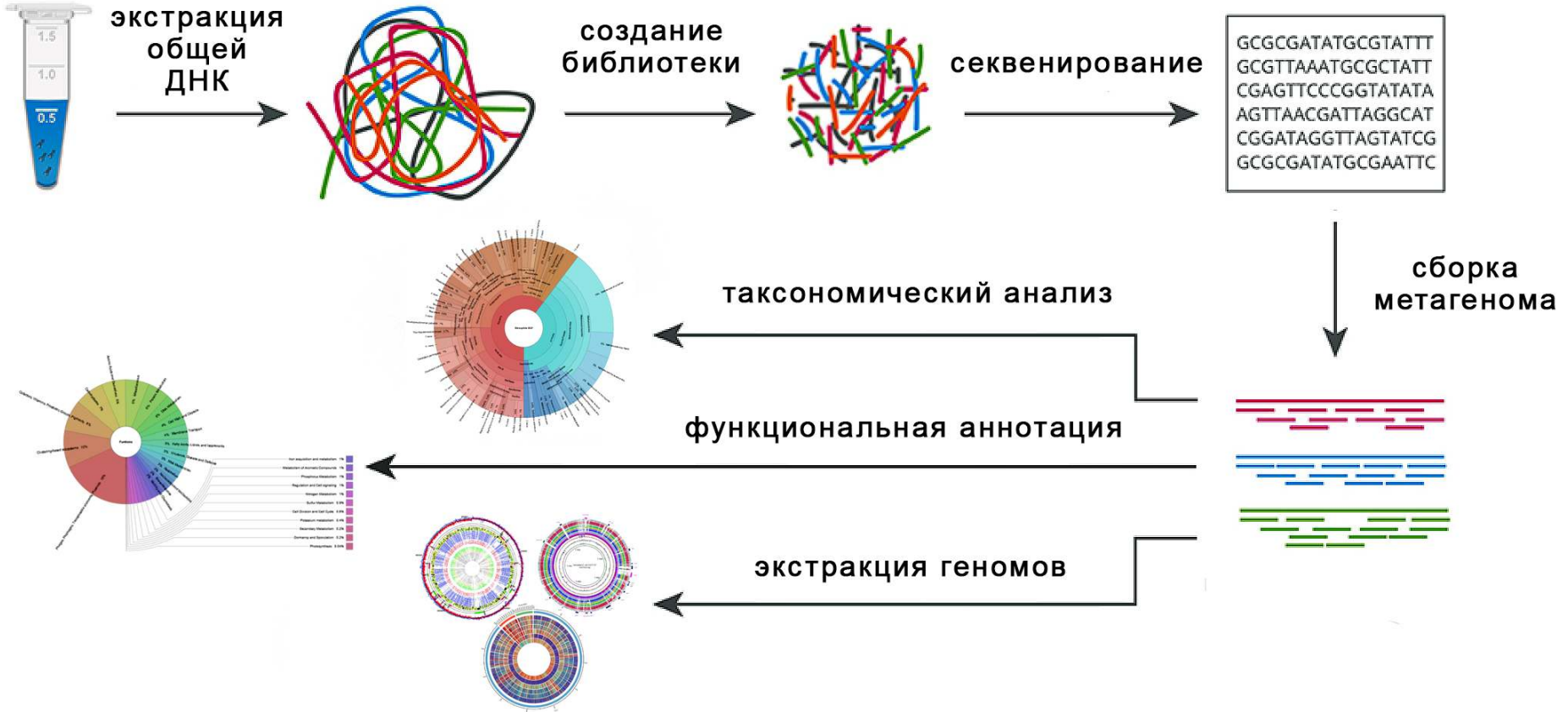
Фото: Andrea Romano, 2017

Грифон Иванова (Камчатка)



Фото: Денис Будьков, 2011

# Схема метагеномного анализа



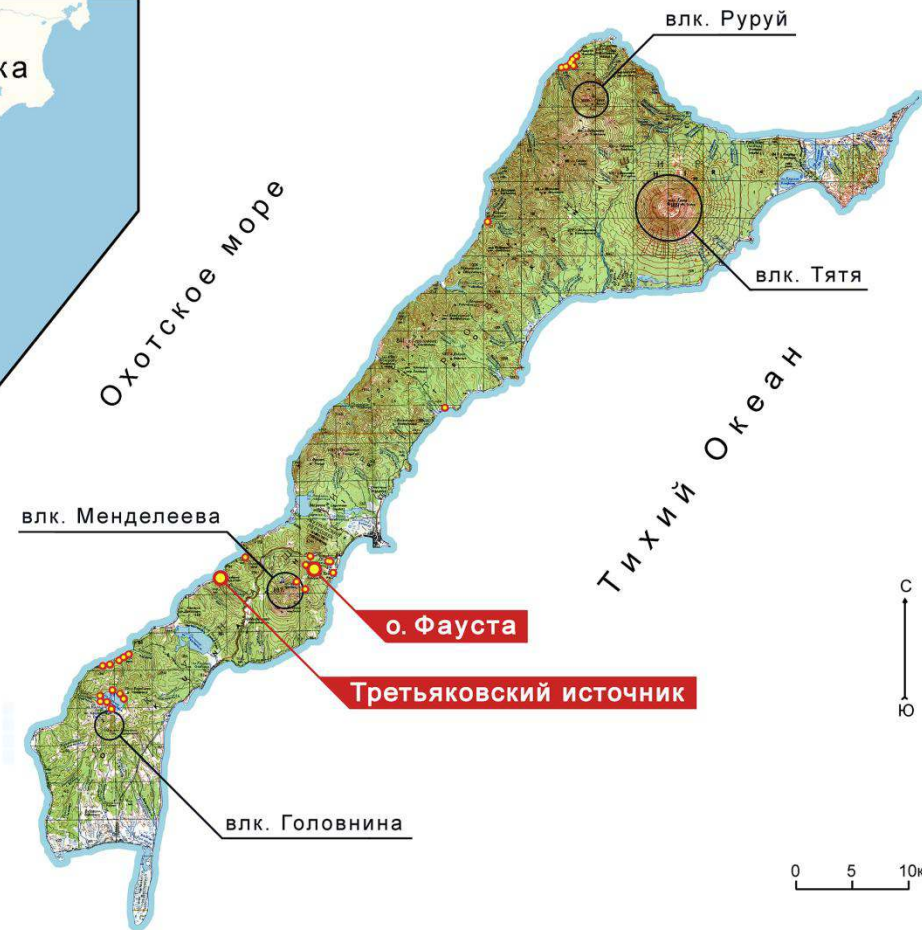
## Цели и задачи

**Цель работы:** изучение биологического разнообразия микроорганизмов геотермальных источников острова Кунашир

### **Задачи:**

- Выделение ДНК из образцов микробных сообществ геотермальных источников острова Кунашир с последующим полногеномным секвенированием;
- Анализ и фильтрация исходных данных, сборка контигов;
- Кластеризация контигов для экстракции геномных данных микроорганизмов;
- Филогенетический и геномный анализ микробного сообществ;

# Остров Кунашир



- Условные обозначения:
- геотермальный источник
  - объект исследования
  - вулкан



# Объекты исследования

## Озеро Фауста



Фото: Розанов А. С.

Координаты: 44°0'11"N, 145°46'43"E

Температура 48°C

pH ~ 2.0

## Третьяковский источник



Фото: Розанов А. С.

Координаты: 43°59'18.1"N, 145°39'05.1"E

Температура 52°C

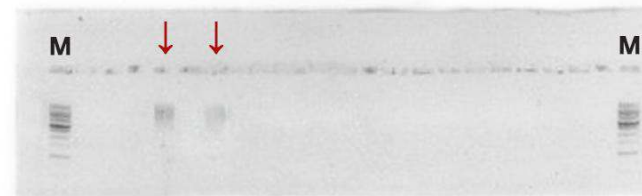
pH ~ 6.0

# Выделение ДНК и секвенирование



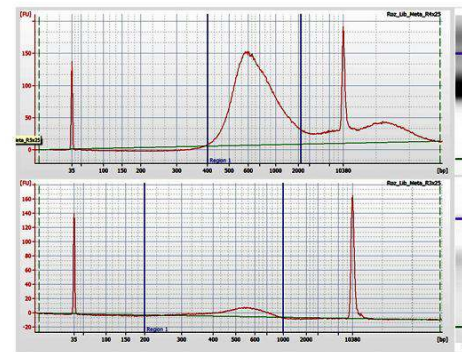
Выделение ДНК с помощью набора  
Nucleospin Soil

ГЭФ  
→



2 из 18 образцов

↓  
библиотеки



секвенирование  
на платформе  
NovaSeq 6000 Illumina



# Биоинформатический анализ





# Метагеномная сборка

Образец донных отложений оз. Фауста (R3)

Образец донных отложений  
Третьяковского источника (R4)

Сервер XL230a-Gen9: 24 ядра 192 ГБ ОЗУ  
Время сборки: ~ 38 часов

Сервер VKOP: 64 ядра 2 ТБ ОЗУ  
Время сборки: ~ 53 часа

- количество контигов длиной  $\geq 0$  bp = 271 342
- количество контигов длиной  $\geq 1000$  bp = 20 806
- общая протяженность контигов – 164 055 084 bp
- самый длинный контиг – 281 765 bp
- N50 = 6 106

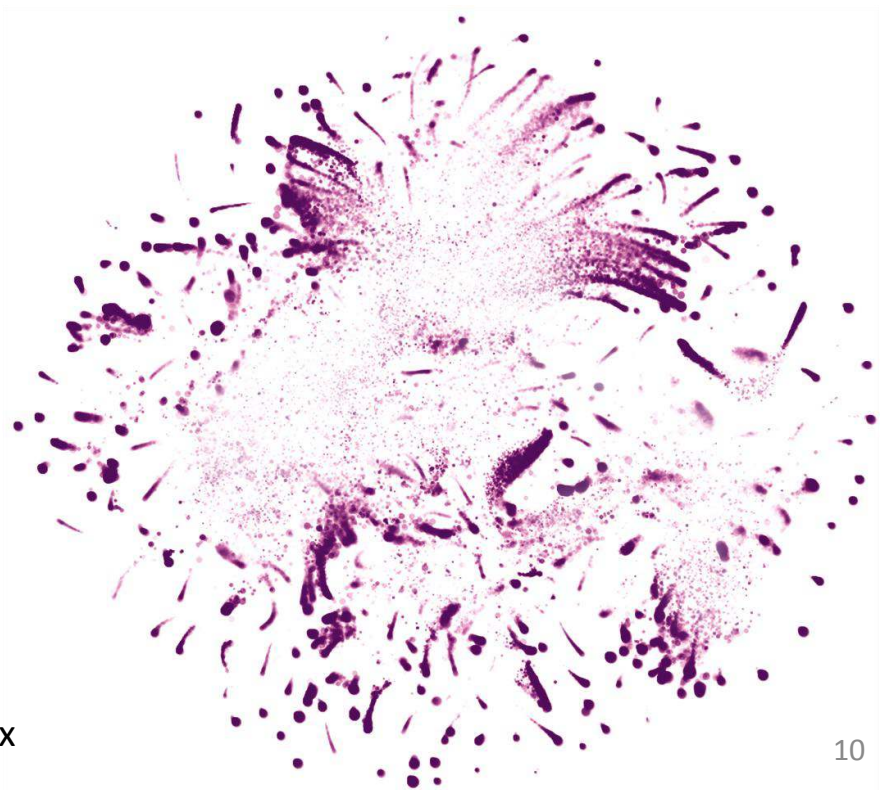
- количество контигов длиной  $\geq 0$  bp = 2 264 569
- количество контигов длиной  $\geq 1000$  bp = 279 624
- общая протяженность контигов – 2 009 229 433 bp
- самый длинный контиг – 1 623 944 bp
- N50 = 12 423

# Кластеризация

Образец донных отложений оз. Фауста (R3)  
41 бин



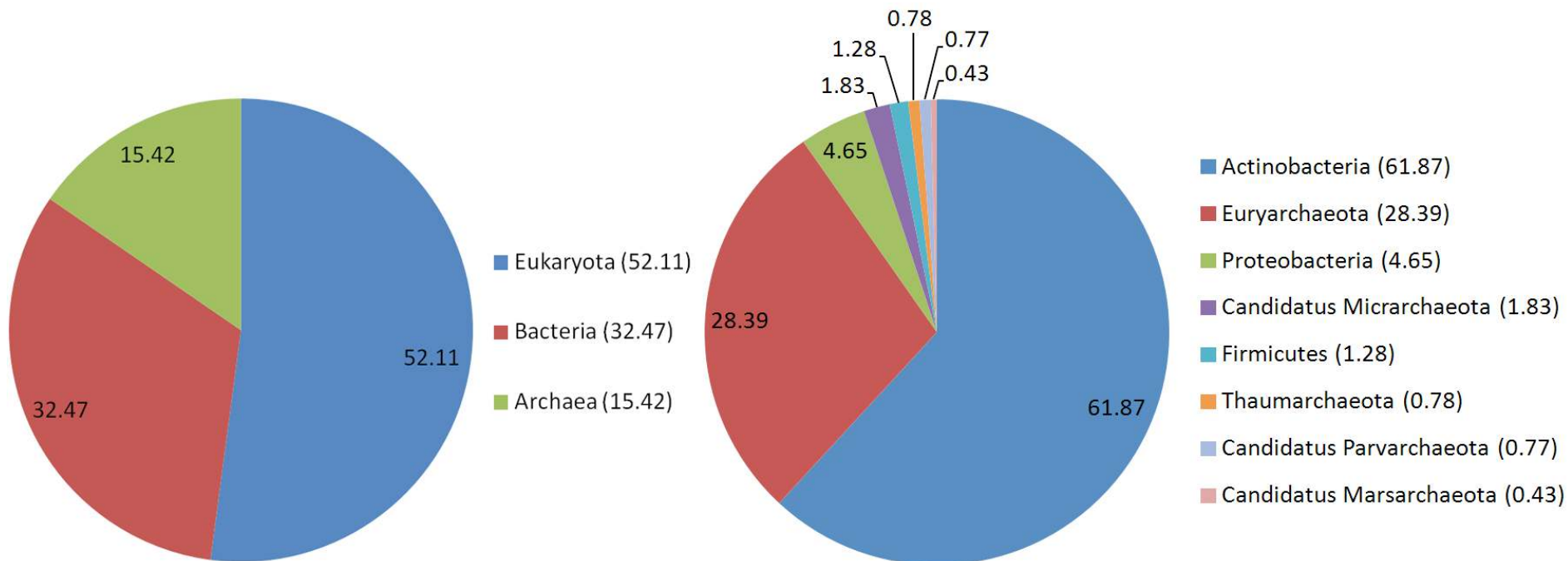
Образец донных отложений  
Третьяковского источника (R4)  
349 бинов



Визуализация кластеризации метагеномных данных

# Филогенетический анализ

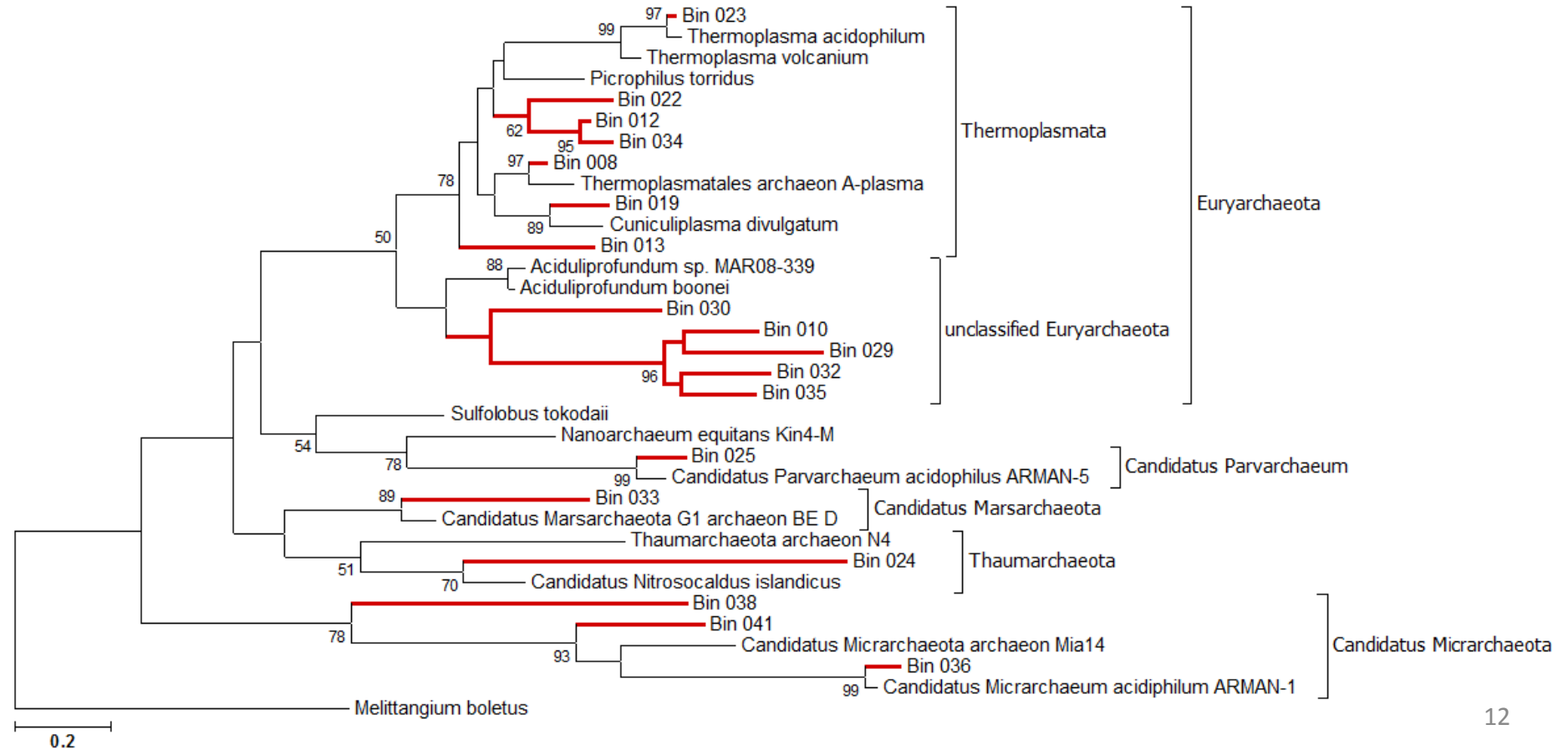
## Образец донных отложений оз. Фауста (R3)



Таксономический состав микробного сообщества по метагеномным данным R3 (%). Справа - на уровне доменов. Слева - на уровне типов, без учета представленности эукариот.

# Филогенетическое дерево бинов, относящихся к археям

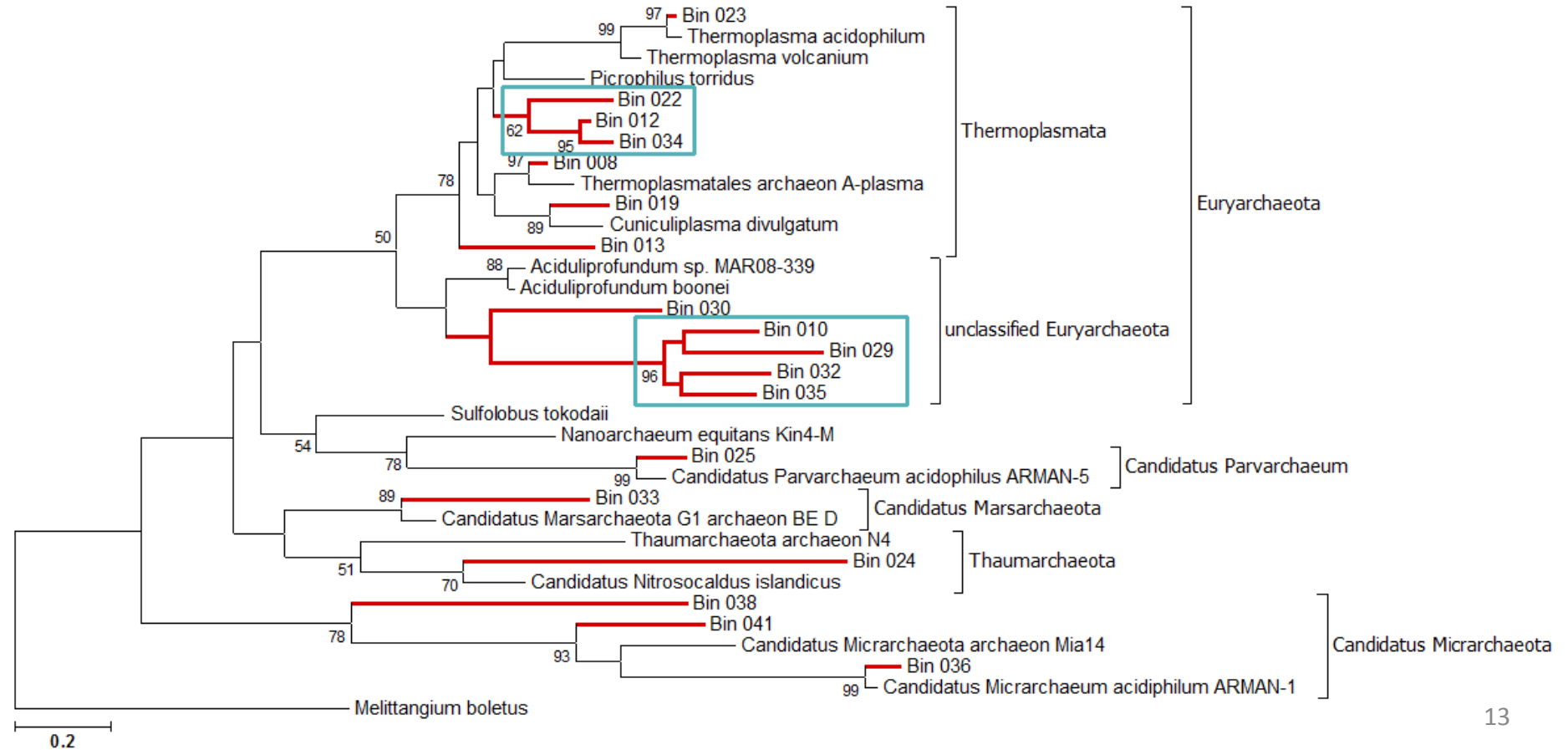
## Образец донных отложений оз. Фауста (R3)





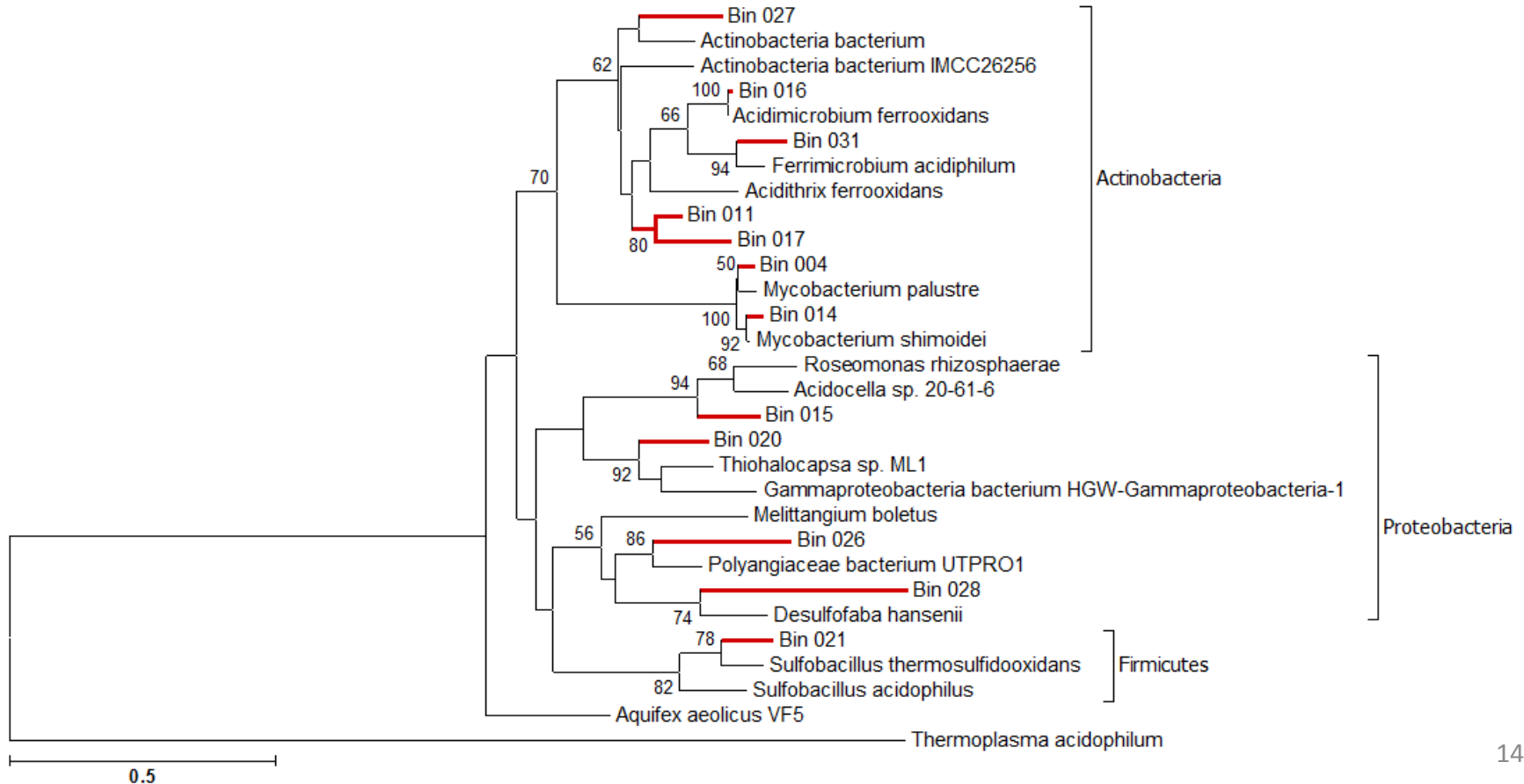
# Филогенетическое дерево бинов, относящихся к археям

## Образец донных отложений оз. Фауста (R3)



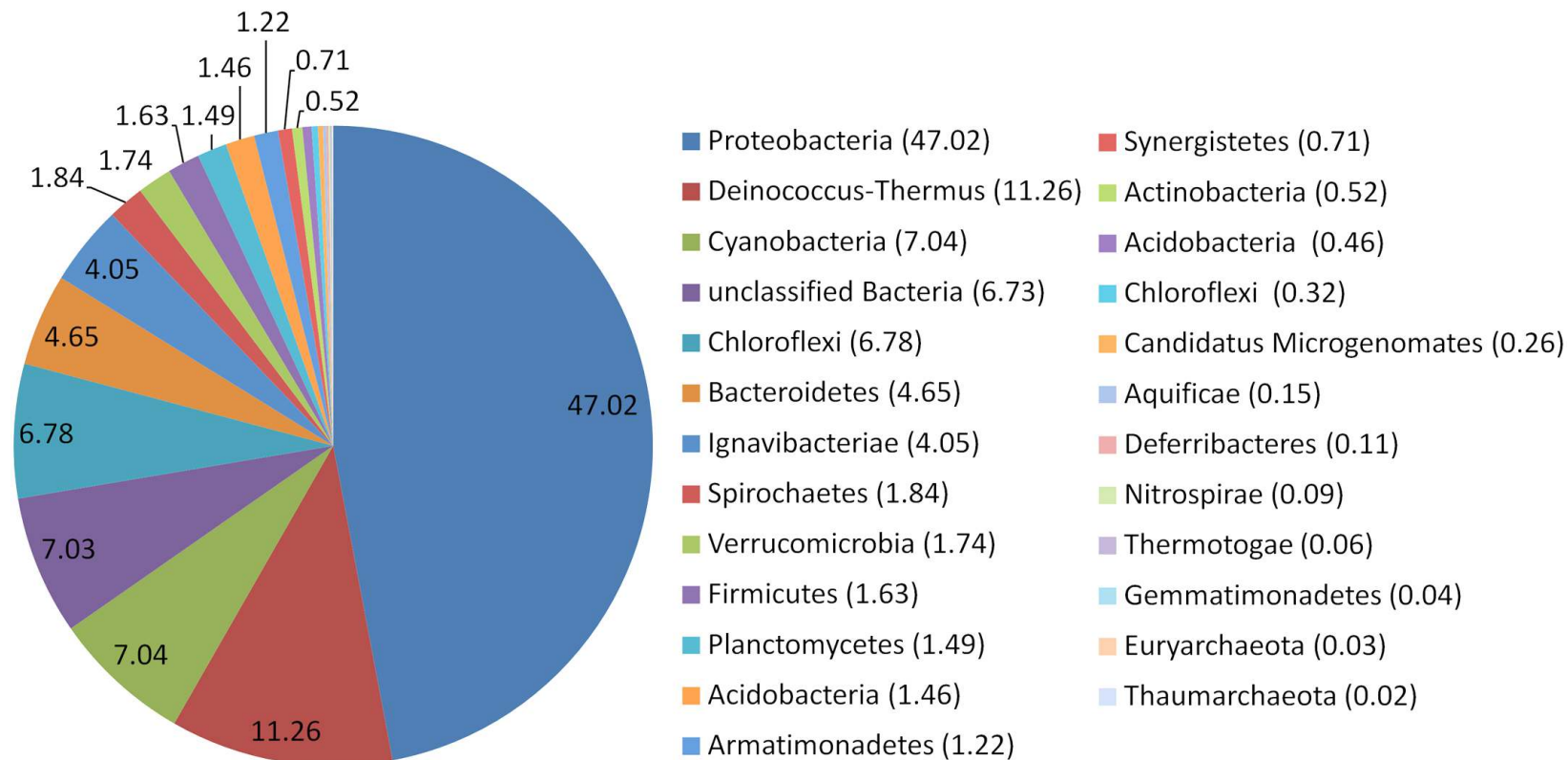
# Филогенетическое дерево бинов, относящихся к бактериям

Образец донных отложений оз. Фауста (R3)



# Филогенетический анализ

Образец донных отложений Третьяковского источника (R4)



Таксономический состав на уровне типов по метагеномным данным R4 (%)

## Экстрагированные геномные данные

Информация о бине				Информация о геноме ближайшего организма			ANI-value
М	№	Размер, мб	GC%	Вид	Размер, мб	GC%	
R3	19	1.92880	38.5	<i>Cuniculiplasma divulgatum</i>	1.90881	37.3	74.00%
	23	1.73336	46.9	<i>Thermoplasma acidophilum</i> DSM 1728	1.56491	46.0	79.68%
R4	13	3.60861	63.8	<i>Dechloromonas agitata</i> is5	3.62723	62.8	90.18%
	22	3.44674	34.2	<i>Ignavibacterium album</i> JCM 16511	3.65900	33.9	88.37%
	74	3.25294	49.4	<i>Treponema caldarium</i> DSM 7334	3.23934	45.6	79.55%
	126	4.21820	63.5	<i>Pannonibacter indicus</i>	4.17068	63.5	99.88%
	190	1.89239	57.7	<i>Thermanaerovibrio velox</i> DSM 12556	1.88084	58.8	79.64%
	197	3.99218	46.4	<i>Anaerosporomusa subterranea</i>	3.96819	47.1	77.15%
	201	2.77489	64.8	<i>Thermoanaerobaculum aquaticum</i>	2.66093	63.0	78.58%



## Выводы

- 1) Выполнено полногеномное секвенирование ДНК из образцов донных отложений, отобранных из геотермальных источников оз. Фауста и Третьяковский источник; проведена первичная обработка данных и сборка контигов. Для микробного сообщества из донных отложений оз. Фауста (R3) получено 20 806 контигов длиной более 1000 пар оснований. Для микробного сообщества из донных отложений Третьяковского источника (R4), отобранных при 52°C, получено 279 624 контигов длиной более 1000 пар оснований;
- 2) Полученные контиги были разбиты при помощи программы MaxBin на кластеры, теоретически соответствующие геномам отдельных видов микроорганизмов. Для микробного сообщества из донных отложений оз. Фауста (R3) получен 41 бин; Для микробного сообщества из донных отложений Третьяковского источника (R4), отобранных при 52°C, получено 349 бинов;
- 3) В результате анализа полученных геномных данных можно сделать вывод о том, что в микробном сообществе источника оз. Фауста преобладают микроорганизмы с очень маленьким размером генома, это характерно для всех трех доменов: эукариоты, бактерии и археи;
- 4) Большинство бинов, полученных в ходе кластеризации, содержат гены белков, филогенетически достаточно удаленные от известных, и, предположительно, относящиеся к новым видам. Данные, полученные при сравнении геномов с ближайшими родственниками с помощью программы ANI, также подтверждают филогенетическую удаленность выделенных геномов от известных таксонов.