

«Новосибирский национальный исследовательский государственный университет»

Факультет естественных наук

Кафедра информационной биологии

Выпускная квалификационная бакалаврская работа

# Метагеномный анализ микробных сообществ соленого озера №48 НСО

Выполнила:

Шипова Александра Андреевна, 4 курс

Лаборатория молекулярных биотехнологий ИЦИГ СО РАН

Научный руководитель:

к.б.н., н.с., Розанов Алексей Сергеевич

# Соленые озера

Большое Соленое озеро, США



Фото: Poldemay, 2014

Мертвое море

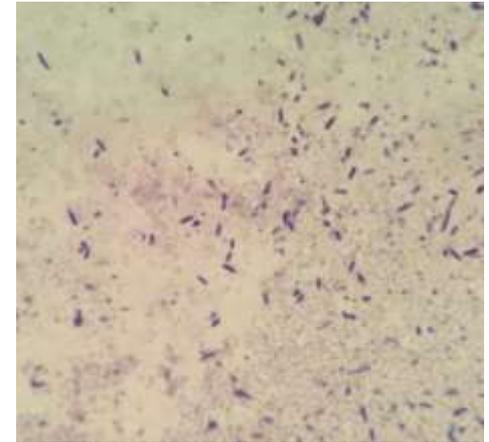


Фото: Shankbone, 2007

# Изучение микробных сообществ

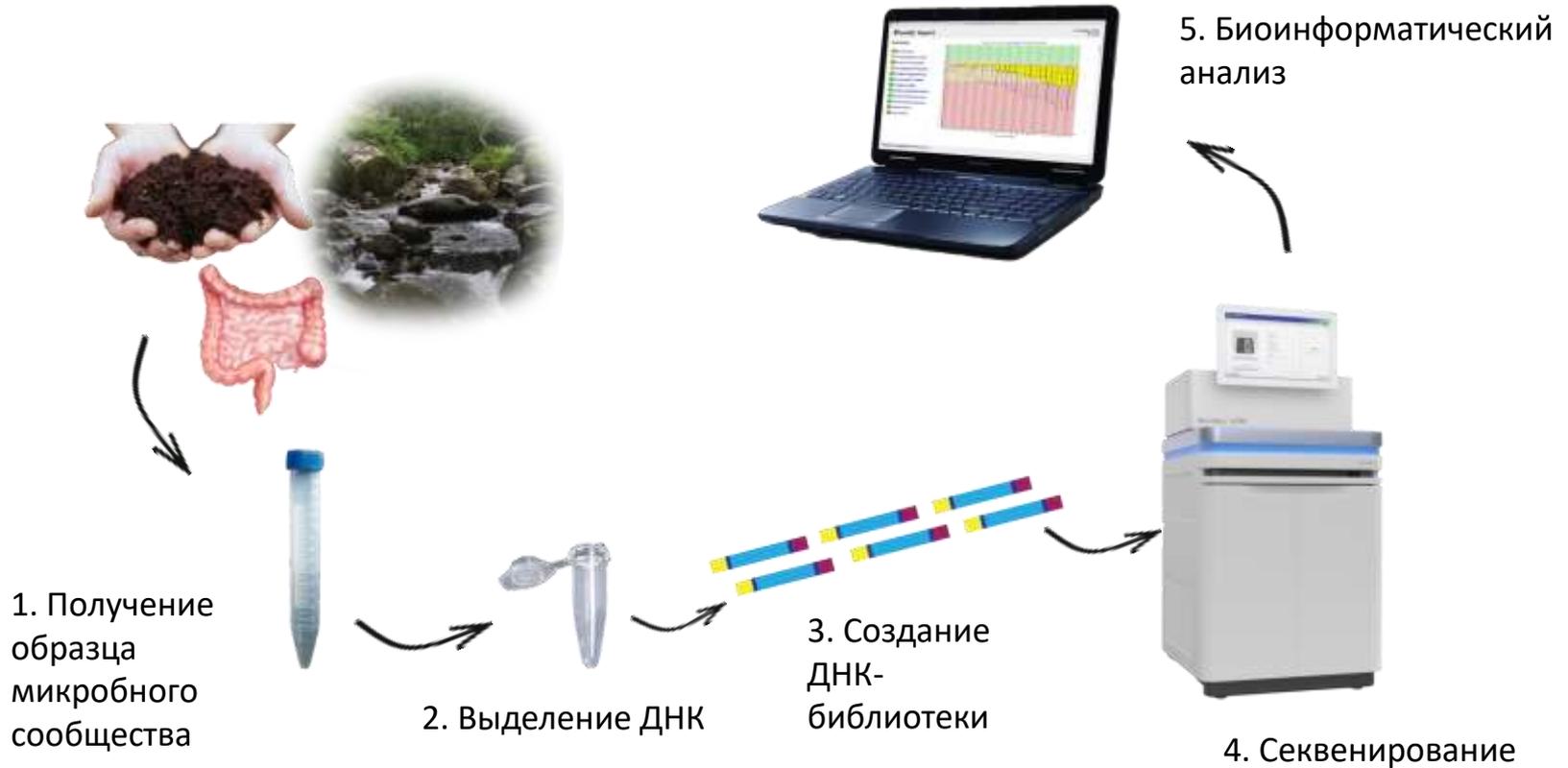


- культивирование



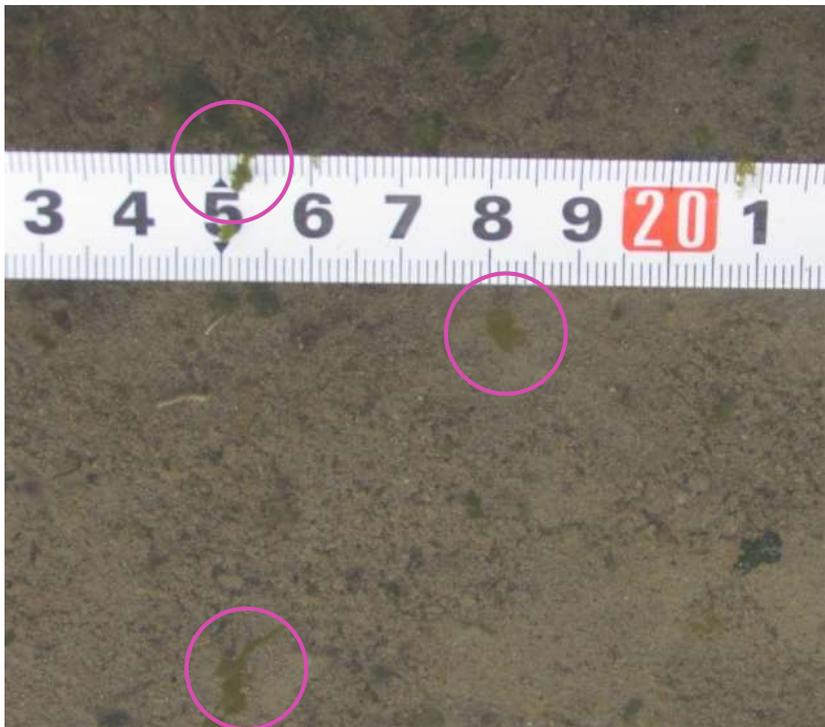
- микроскопия

# Метагеномный анализ





# Исследуемые микробные сообщества



Частицы цианобактериального мата  
(обведено)



Верхний слой донных отложений  
(обведено)

# Цель и задачи

## Цель работы:

метагеномный анализ наиболее метаболически активных микробных сообществ (цианобактериальная биомасса и верхний слой донных отложений) озера Соленого №48 НСО.

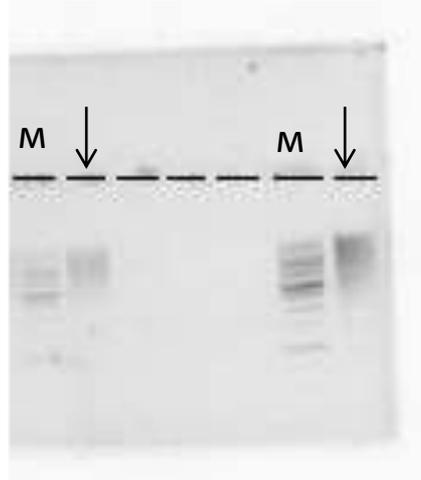
## Задачи:

1. Выделение ДНК из образцов микробных сообществ и метагеномное секвенирование.
2. Сборка скэффолдов и биннинг.
3. Филогенетический анализ микробного состава сообществ.

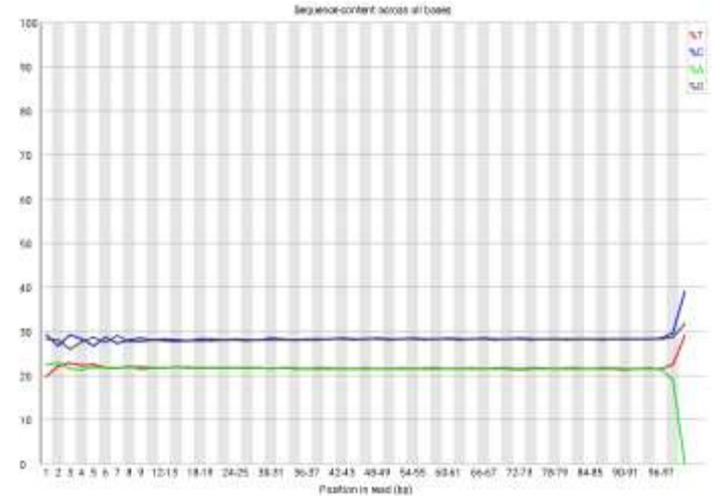
# Выделение ДНК и секвенирование



Образцы

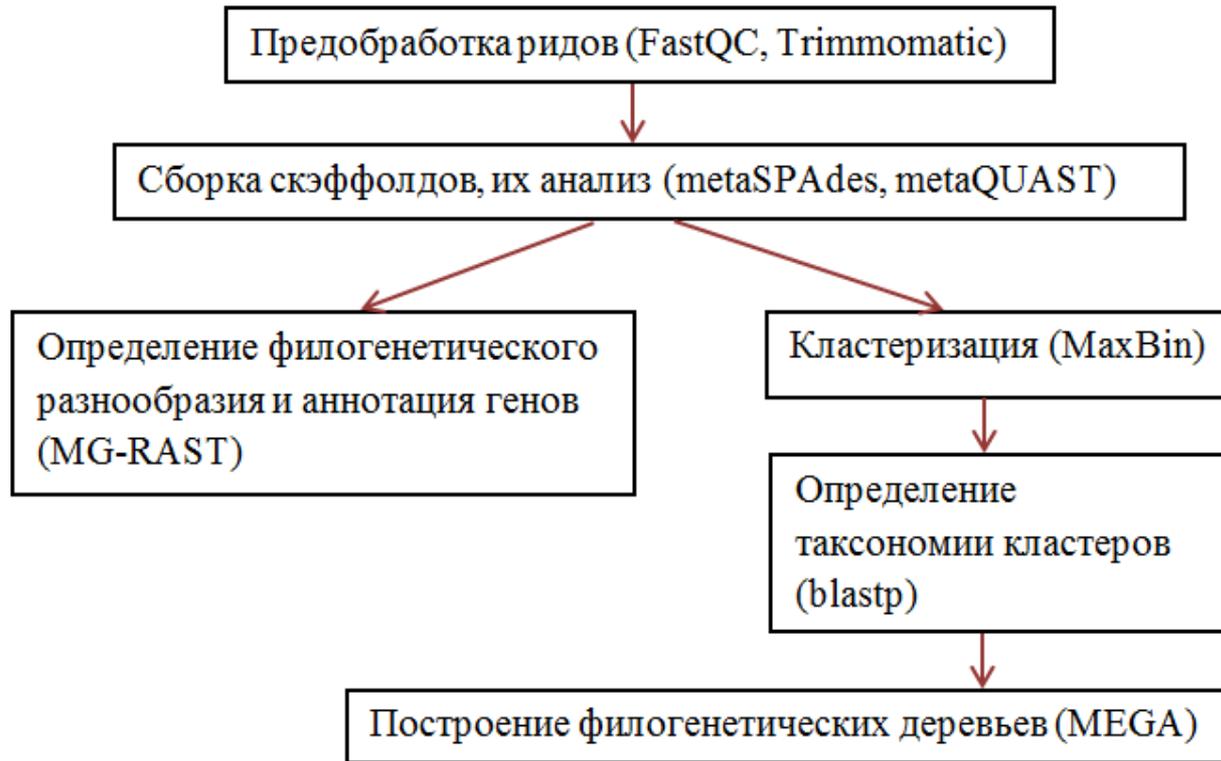


Электрофореграмма выделенной ДНК:  
ДНК образцов обозначено стрелками,  
М – маркер



Распределение нуклеотидов по длине ридов  
до их обработки (информация получена при  
помощи программы FastQC)

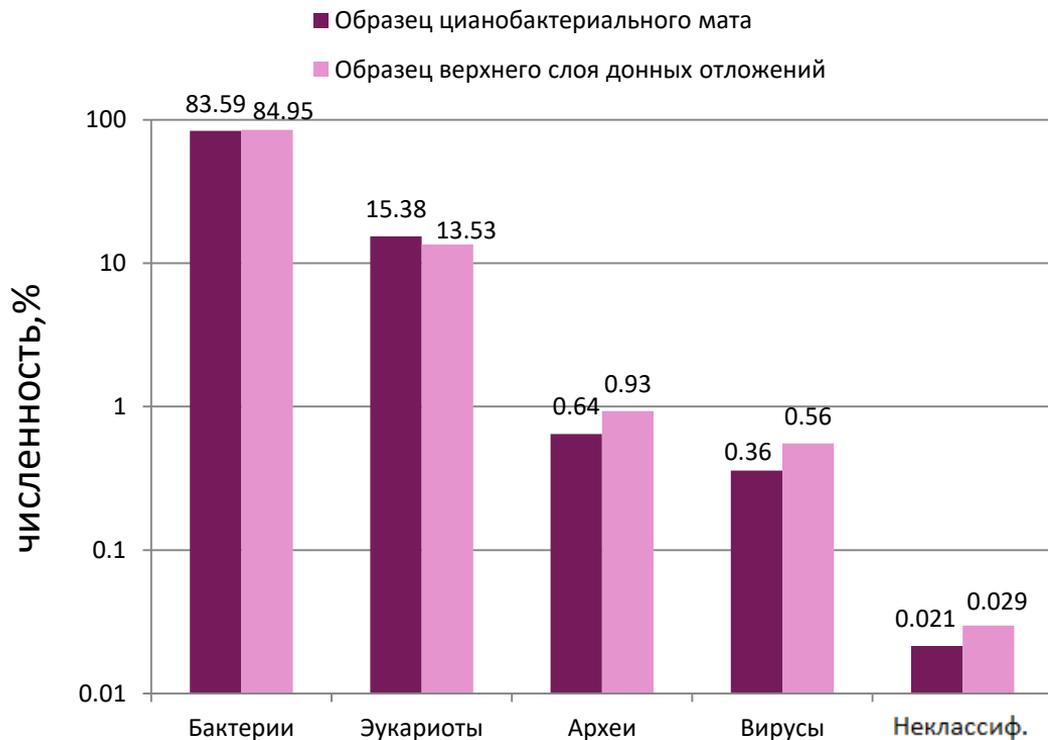
# Биоинформатический анализ



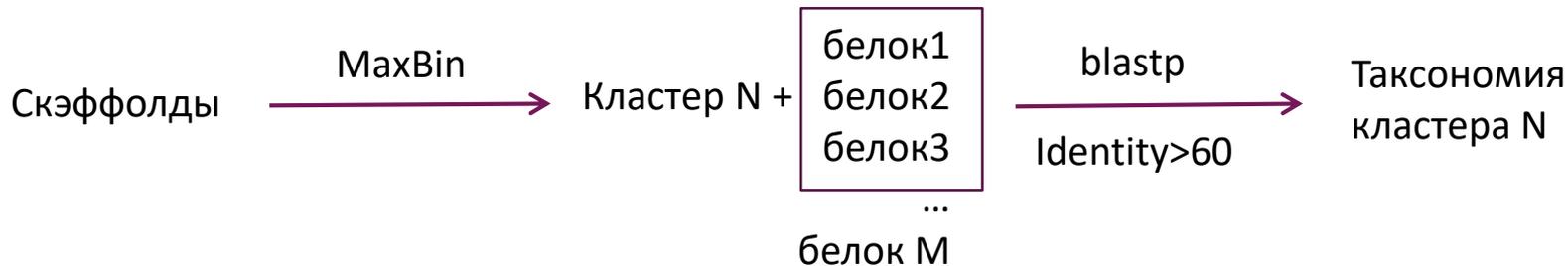
# Информация о сборке

	Образец цианобактериального мата	Образец донных отложений
Кол-во скэффолдов ( $\geq 0$ bp)	1 567 052	1 706 464
Кол-во скэффолдов ( $\geq 1000$ bp)	156 467	201 488
Самый длинный скэффолд	387 984	476 707
Общая длина ( $\geq 0$ bp)	994 597 198	1 191 582 506
Общая длина ( $\geq 1000$ bp)	632 144 908	752 234 840
GC (%)	54.02	53.88

# Филогенетическое разнообразие микробных сообществ на уровне домена, полученное при помощи MG-RAST



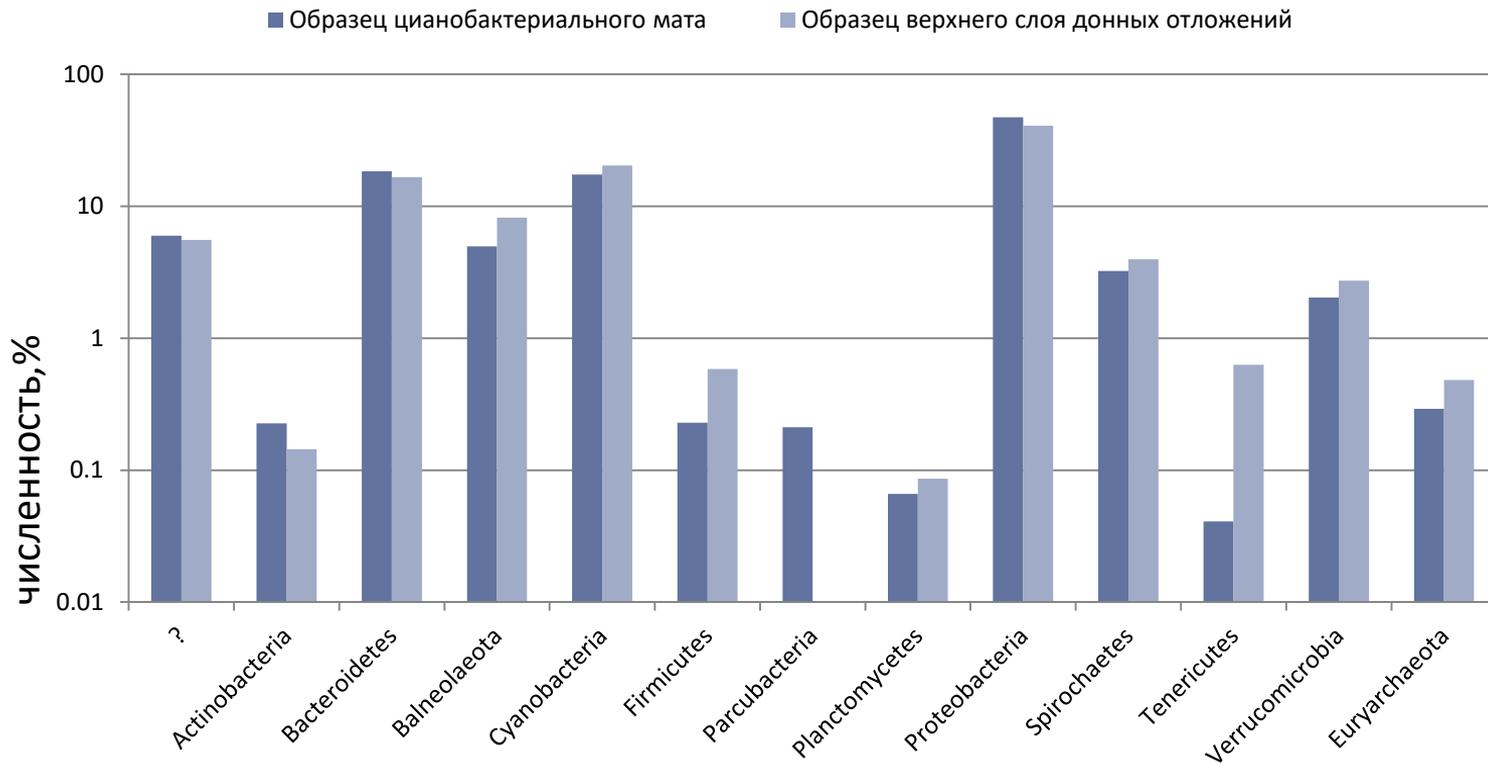
# Определение таксономии кластеров, полученных при помощи программы MaxBin



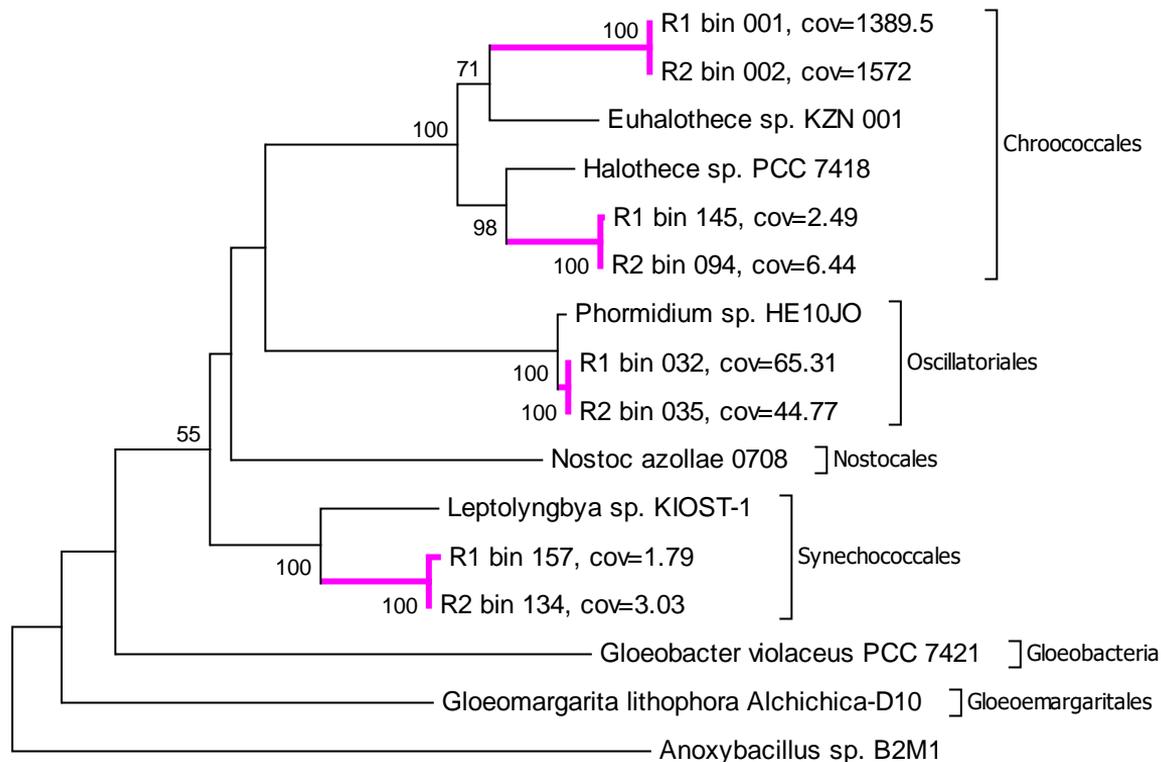
Номер бина	Покрытие	Процент маркерных генов	Размер генома	G+C контент	Тип
2	1389.5	28.9	1553506	41.9	Cyanobacteria
4	760.06	99.1	3092800	65.6	Proteobacteria
-/-					
32	65.31	95.3	4373671	52.2	Cyanobacteria
33	52.33	94.4	4423595	62.1	Proteobacteria
34	52.17	94.4	3870118	59.8	Spirochaetes
-/-					
156	1.92	60.7	2911009	44.5	Bacteroidetes
157	1.79	43	4284696	55.9	Cyanobacteria
158	1.68	32.7	2102468	63.7	Unclassified

Пример информации о некоторых кластерах для образца цианобактериального мата

# Филогенетическое разнообразие микробных сообществ на уровне типа, полученное при анализе данных кластеризации

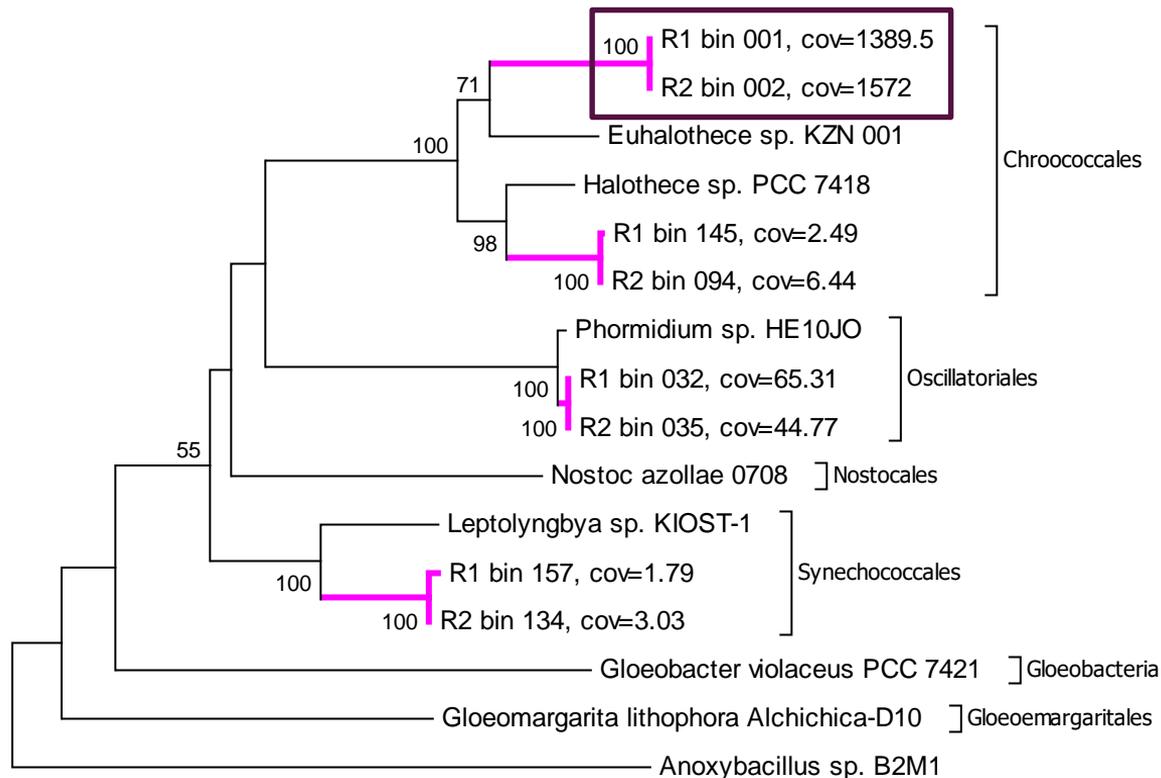


# Филогенетическое дерево цианобактерий (по белку NAD-зависимая ДНК-лигаза)



Кластеры, полученные при анализе образцов цианобактериального мата и верхнего слоя донных отложений, отмечены как «R1» и «R2» соответственно.  
Покрывание бина обозначено «cov». Отображены устойчивые узлы (>50).

# Филогенетическое дерево цианобактерий (по белку NAD-зависимая ДНК-лигаза)



Chroococcales

Oscillatoriales

Nostocales

Synechococcales

Gloeobacteria

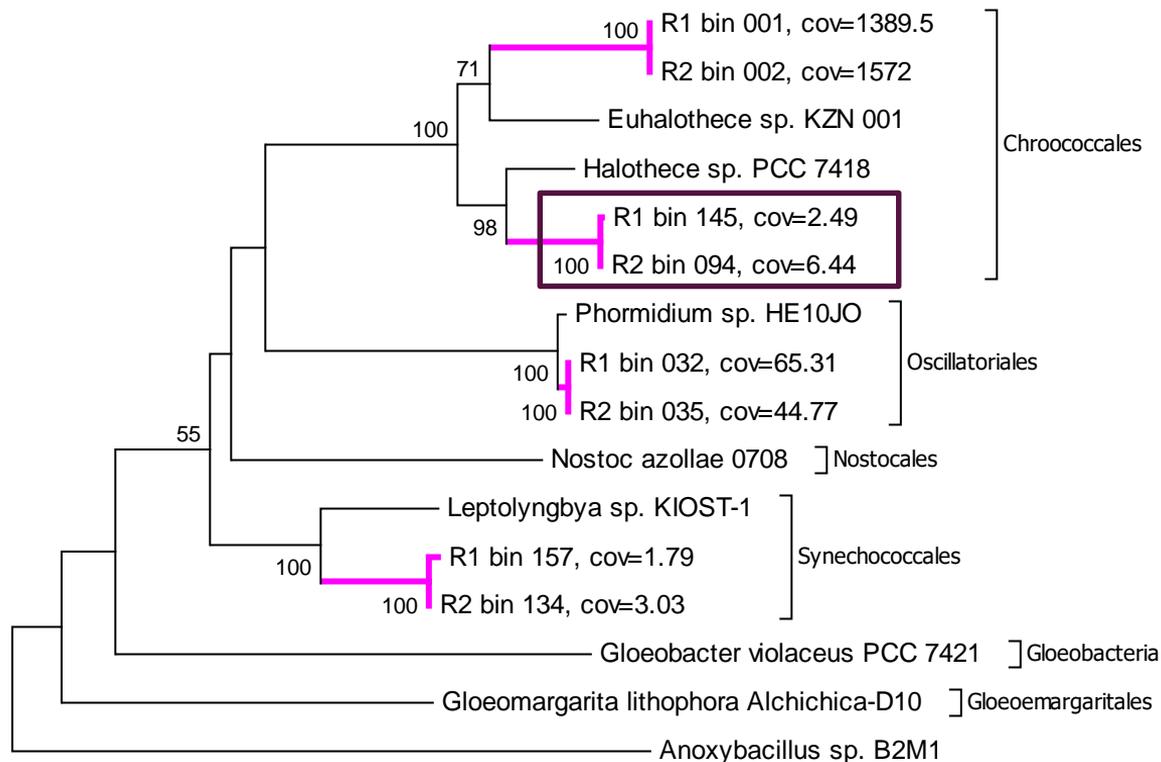
Gloeomargaritales

Кластеры, полученные при анализе образцов цианобактериального мата и верхнего слоя донных отложений, отмечены как «R1» и «R2» соответственно.

Покрытие бина обозначено «cov». Отображены устойчивые узлы (>50).

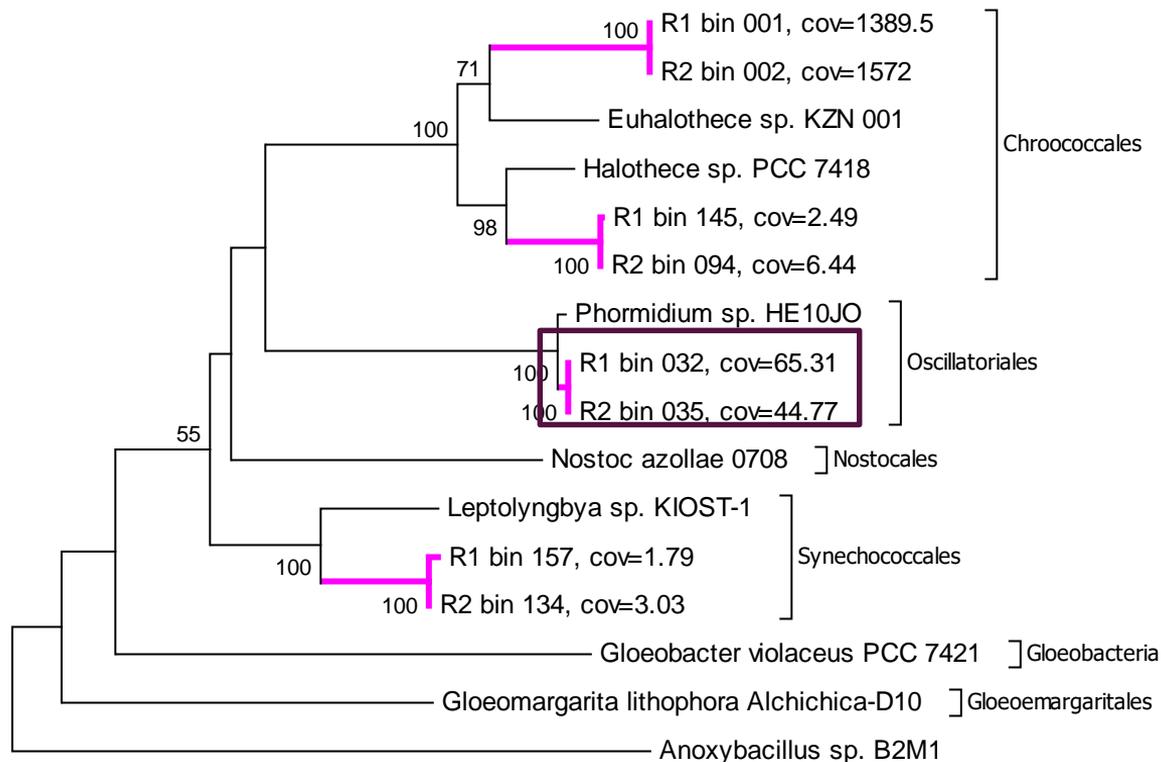
0.2

# Филогенетическое дерево цианобактерий (по белку NAD-зависимая ДНК-лигаза)



Кластеры, полученные при анализе образцов цианобактериального мата и верхнего слоя донных отложений, отмечены как «R1» и «R2» соответственно. Покрытие бина обозначено «cov». Отображены устойчивые узлы (>50).

# Филогенетическое дерево цианобактерий (по белку NAD-зависимая ДНК-лигаза)



Chroococcales

Oscillatoriales

Nostocales

Synechococcales

Gloeobacteria

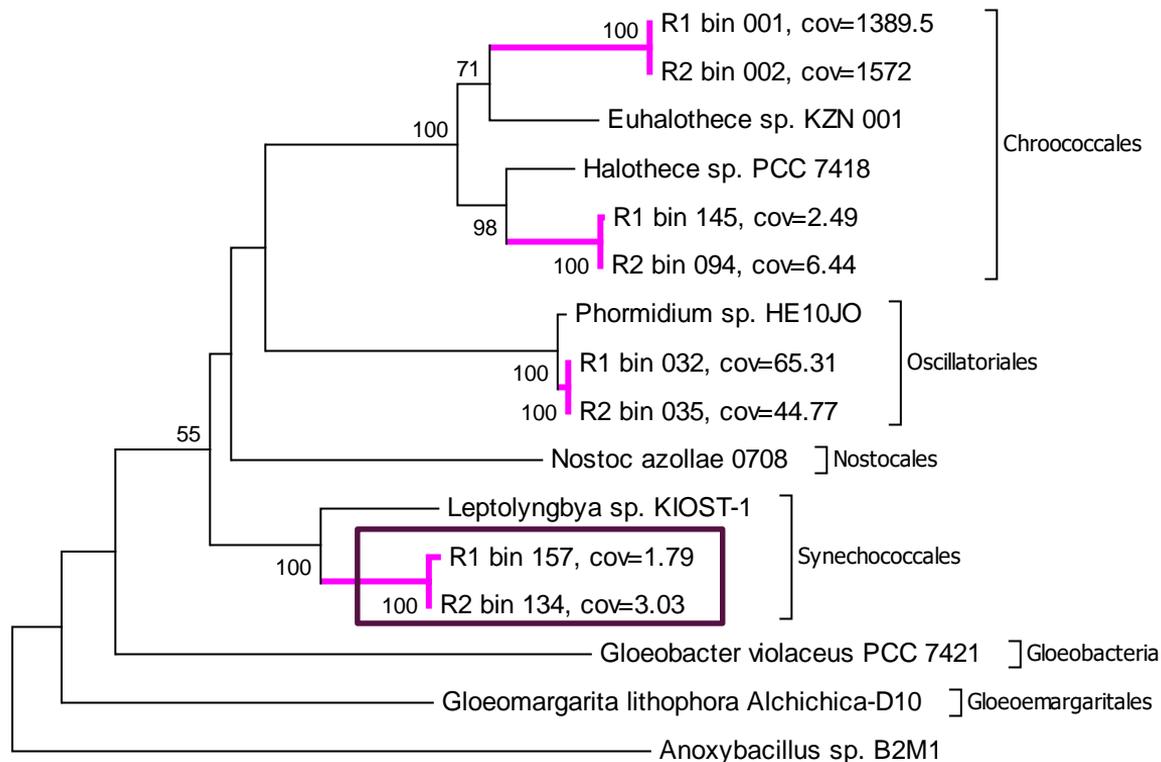
Gloeomargaritales

Кластеры, полученные при анализе образцов цианобактериального мата и верхнего слоя донных отложений, отмечены как «R1» и «R2» соответственно.

Покрытие бина обозначено «cov». Отображены устойчивые узлы (>50).

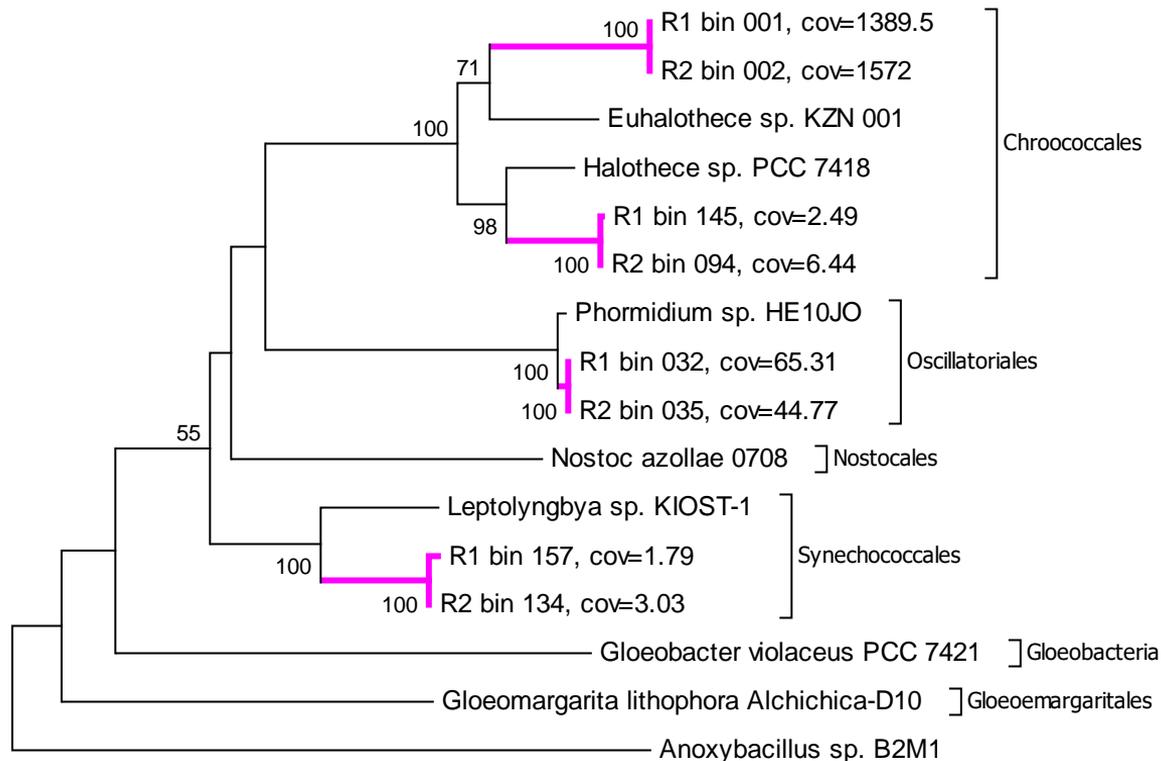
0.2

# Филогенетическое дерево цианобактерий (по белку NAD-зависимая ДНК-лигаза)



Кластеры, полученные при анализе образцов цианобактериального мата и верхнего слоя донных отложений, отмечены как «R1» и «R2» соответственно. Покрытие бина обозначено «cov». Отображены устойчивые узлы (>50).

# Филогенетическое дерево цианобактерий (по белку NAD-зависимая ДНК-лигаза)



Кластеры, полученные при анализе образцов цианобактериального мата и верхнего слоя донных отложений, отмечены как «R1» и «R2» соответственно.  
Покрытие бина обозначено «cov».  
Отображены устойчивые узлы (>50).

0.2

# Основные категории белковых функций, полученные при помощи MG-RAST

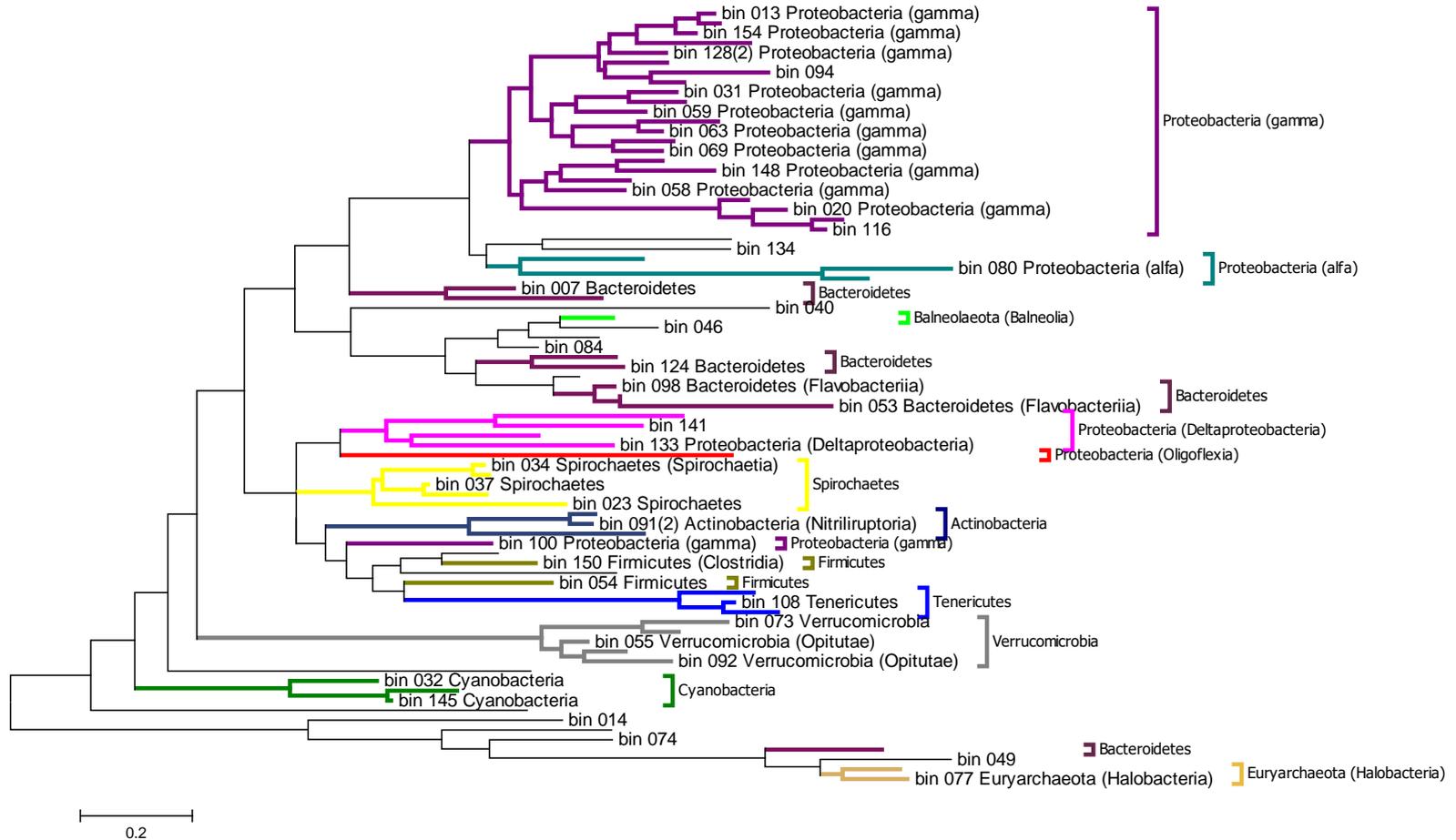
Белковая функция	Образец микроб. мата, %	Образец донных отложений, %	Прирост доли генов в образце донных отл.
Amino Acids and Derivatives	9.480205	9.692315	+2.237399
Carbohydrates	11.60548	12.09057	+4.179836
Cell Division and Cell Cycle	1.024894	1.039619	+1.436734
Cell Wall and Capsule	5.028915	4.871004	-3.14006
Clustering-based subsystems	13.84437	13.95152	+0.773961
Cofactors, Vitamins, Prosthetic Groups, Pigments	6.156593	6.164472	+0.127977
DNA Metabolism	5.116977	5.205689	+1.73368
Dormancy and Sporulation	0.149358	0.184356	+23.43229
Fatty Acids, Lipids, and Isoprenoids	3.417162	3.400773	-0.47961
Iron acquisition and metabolism	0.878882	0.835752	-4.90737
Membrane Transport	2.393606	2.397679	+0.170162
Metabolism of Aromatic Compounds	1.226715	1.160825	-5.37126
Miscellaneous	7.606006	7.624514	+0.243334
Motility and Chemotaxis	1.172111	1.204655	+2.776529

Белковая функция	Образец микроб. мата, %	Образец донных отложений, %	Прирост доли генов в образце донных отл.
Nitrogen Metabolism	0.573609	0.631692	+10.12589
Nucleosides and Nucleotides	2.340073	2.329147	-0.46691
Phages, Prophages, Transposable elements, Plasmids	5.218155	3.679806	-29.4807
Phosphorus Metabolism	0.927598	0.911877	-1.69481
Photosynthesis	0.172779	0.169169	-2.08937
Potassium metabolism	0.485279	0.493185	+1.629166
Protein Metabolism	6.520887	6.763195	+3.715875
Regulation and Cell signaling	1.500136	1.441876	-3.88365
Respiration	2.327225	2.417191	+3.865806
RNA Metabolism	5.047384	5.275567	+4.520817
Secondary Metabolism	0.275964	0.304984	+10.51586
Stress Response	2.34101	2.411232	+2.999645
Sulfur Metabolism	0.849171	0.958591	+12.88551
Virulence, Disease and Defense	2.319462	2.38874	+2.986813

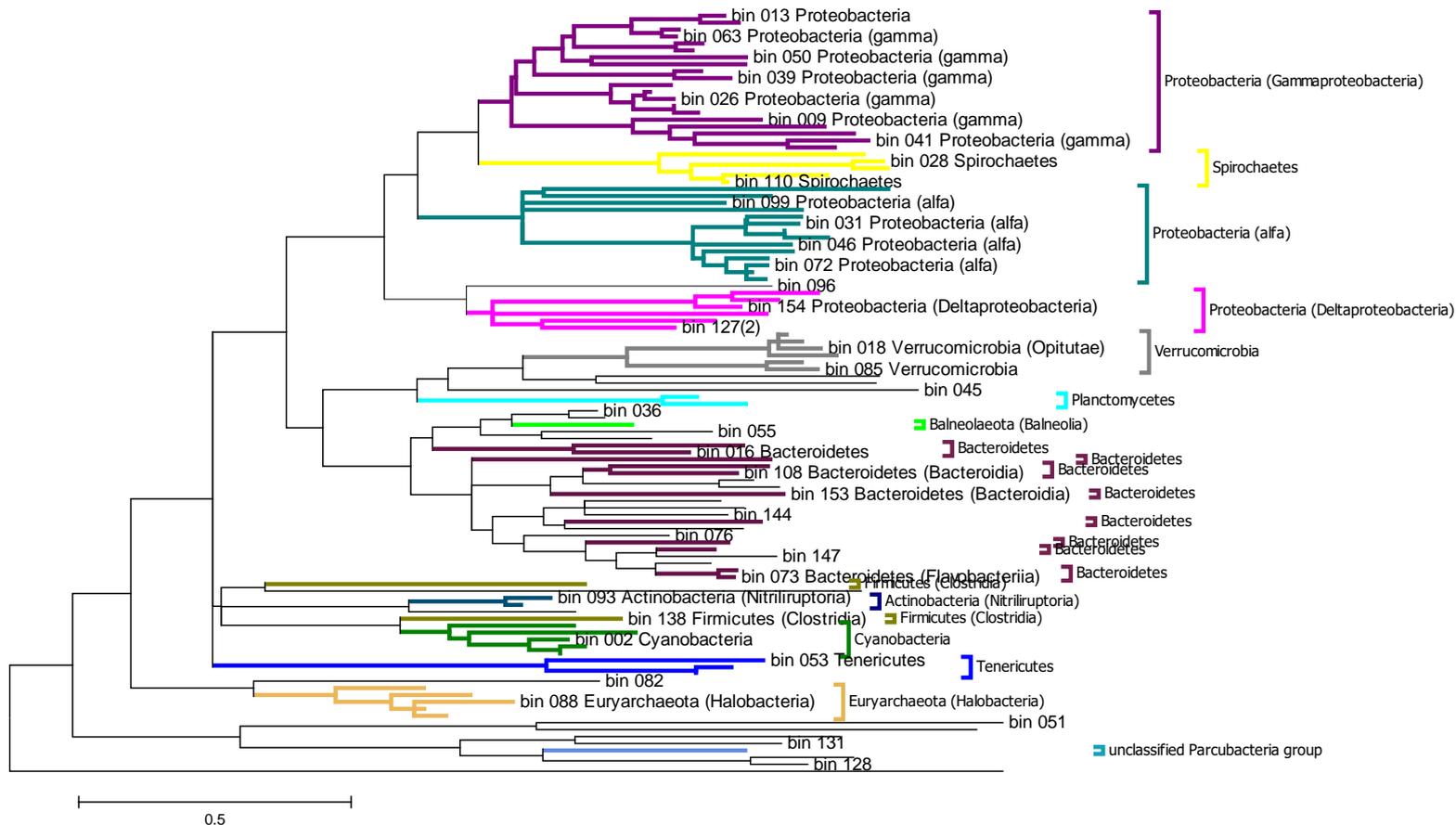
# Выводы

1. В результате секвенирования ДНК-библиотек были получены данные в объеме ~ 400 млн ридов высокого качества для каждого исследуемого образца.
2. Для образца цианобактериальной биомассы после сборки было получено 1,5 млн скэффолдов, длина метагенома составила 994,5 млн нуклеотидов. Для образца донных отложений после сборки было получено 1,7 млн скэффолдов, длина метагенома составила 1,1 млрд нуклеотидов. С помощью кластеризации удалось получить 158 кластеров для первого образца и 187 кластеров для второго.
3. Филогенетическое разнообразие микробных сообществ двух исследуемых проб оказалось практически одинаковым на уровне типа. Информация о составе сообществ, полученная с помощью сервера MG-RAST несколько отличалась от информации, которую удалось получить при определении таксономии кластеров. Общие таксоны, которые были выявлены – Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes, Planctomycetes, Proteobacteria, Spirochaetes, Verrucomicrobia, Euryarchaeota. Согласно филогенетическому анализу белковых последовательностей и их удаленности от ближайших известных гомологов можно предположить, что достаточно большое число кластеров получено из геномов ранее неизвестных видов.
4. По данным, полученным с сервера MG-RAST, последовательности, полученные для верхнего слоя донных отложений, в сравнении с последовательностями, полученными для цианобактериального мата, имеют больший процент генов, кодирующих белки метаболизма углеводов, метаболизма серы и азота, что хорошо согласуется с современными представлениями о метаболизме микробных сообществ в разных условиях аэрации. Необходимо отметить, что в случае образца верхнего слоя донных отложений доля генов, характерных для мобильных элементов, фагов и плазмид, более низкая, чем в образце цианобактериального мата.

# Филогенетическое дерево для образца цианобактериального мата (белок валин-тРНК лигаза)



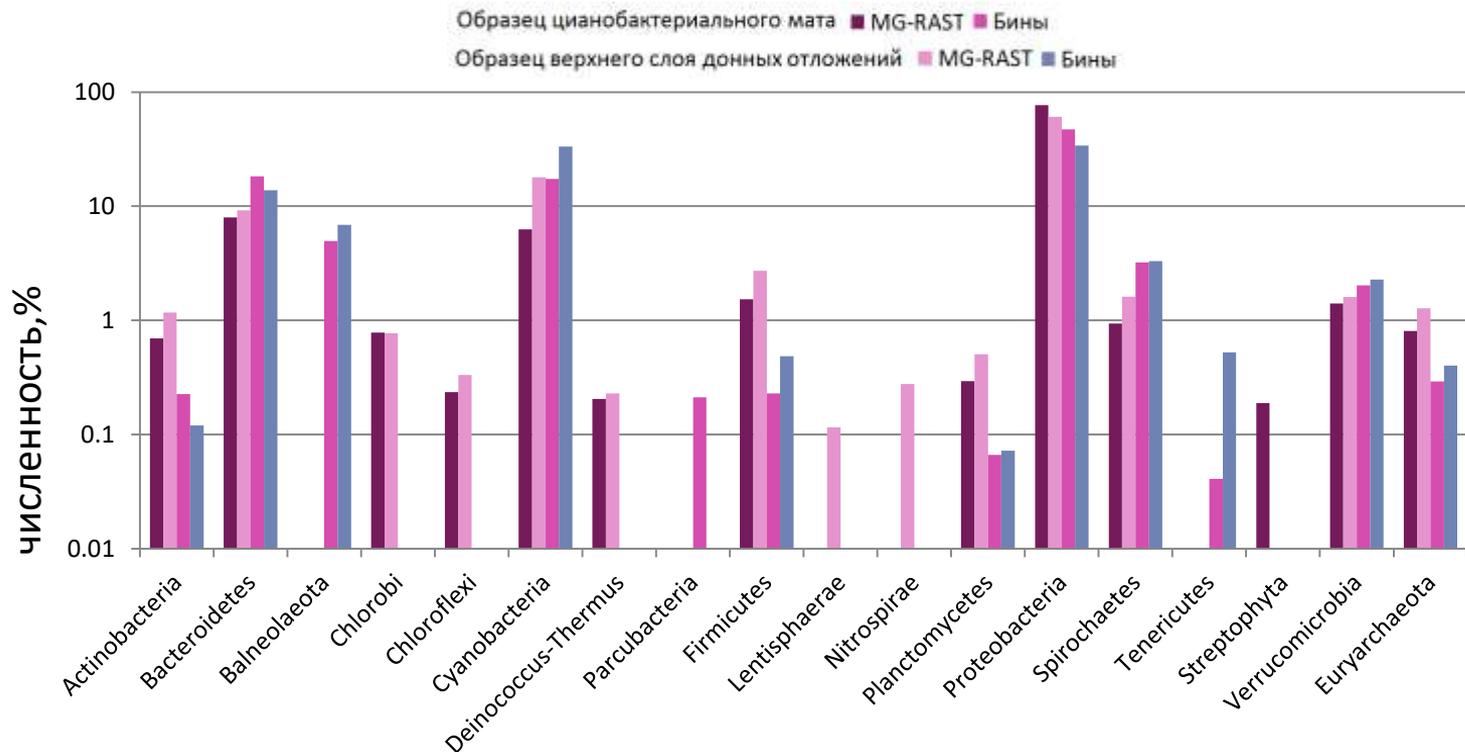
# Филогенетическое дерево для образца донных отложений (белок ЦТФ синтетаза)



# Филогенетическое разнообразие микробных сообществ на уровне типа, полученное при анализе данных кластеризации



# Сравнение данных, полученных при помощи MG-RAST и при анализе кластеризации



Общие таксоны:  
 Actinobacteria,  
 Bacteroidetes,  
 Cyanobacteria,  
 Firmicutes,  
 Planctomycetes,  
 Proteobacteria,  
 Spirochaetes,  
 Verrucomicrobia,  
 Euryarchaeota.