

Геном в геноме

Выделение геномной последовательности потенциального бактериального эндосимбионта губки *Halisarca dujardini*.

Дарья Абузова
Александр Черкасов



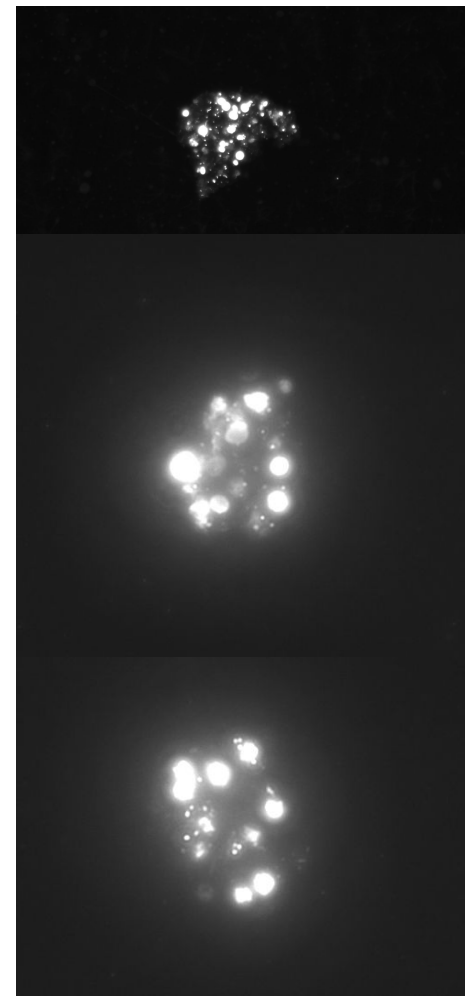
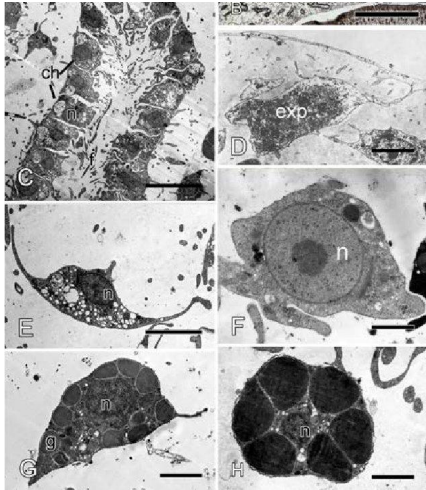
Background



Губки являются одними из наиболее древних и просто устроенных многоклеточных животных. *Halisarca dujardini* морская холодноводная губка, населяющая побережья северных морей России. Как и большинство других губок *Halisarca dujardini*, населена разнообразным бактериальным сообществом, многие представители которого не только неизвестны, но и трудны в культивировании. Это делает непростой задачу изучения видового состава и других аспектов биологии подобных сообществ.

Методы изучения организации хроматина, такие как Hi-C, помимо данных характеризующих укладку ДНК, позволяют также выявлять и выделять геномные последовательности симбиотических организмов и использовать их для исследования микробных сообществ.

В данной работе нашей задачей стало выделение и характеристика геномной последовательности симбионтов губки *Halisarca dujardini*.



Данные и рабочий процесс

Данные:

1. Геномная (метагеномная) сборка губки *Halisarca dujardini* (Nanopore)
2. Короткие прочтения Illumina геномной библиотеки (не использовались для сборки)
3. Hi-C датасет “Весна” (2 библиотеки)
4. Hi-C датасет “Осень” (5 библиотек, различные стадии регенерации губки)

Поиск и характеристика
рибосомальных РНК
**Barrnap + NCBI rRNA
database**

Определение
присутствия
бактериальной ДНК в
разных наборах
данных
Bowtie2+Jucier

Геномная
(метагеномная)
сборка губки
H.dujardini

Выделение и сборка
скаффолдов
бактериального
происхождения **Juicebox
Assembly Tools**

Функциональная
аннотация **RAST**

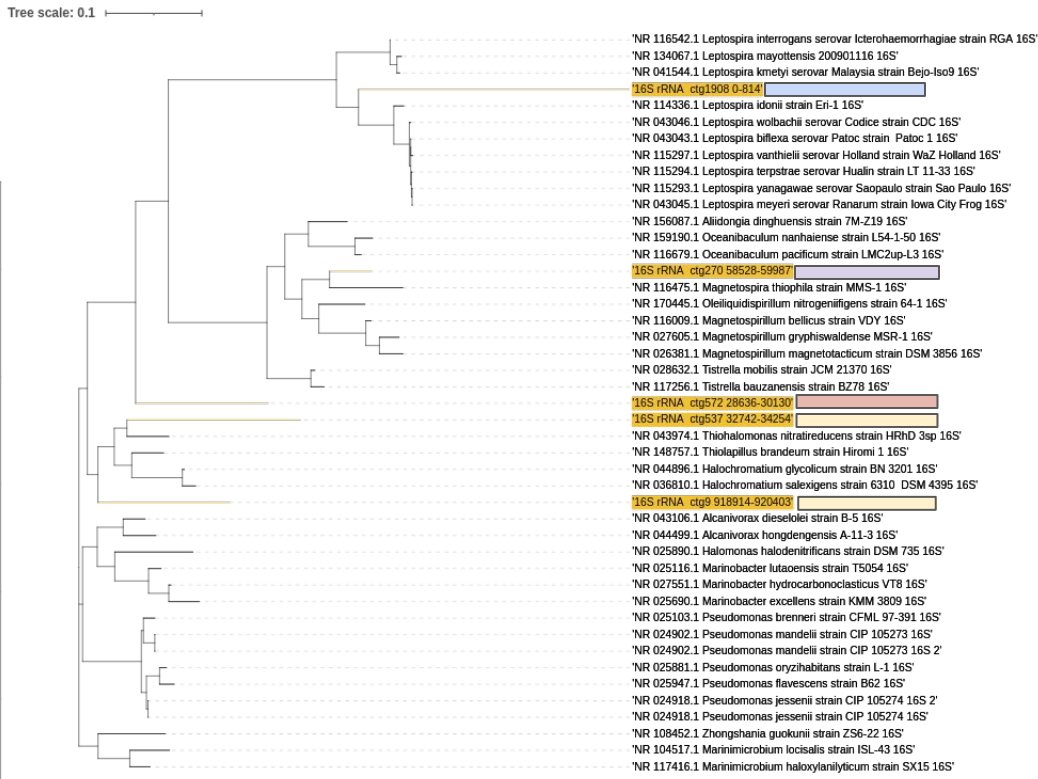
Таксономическая
классификация
GTDBTk

Оценка полноты полученных
сборок **CheckM**

Поиск рРНК

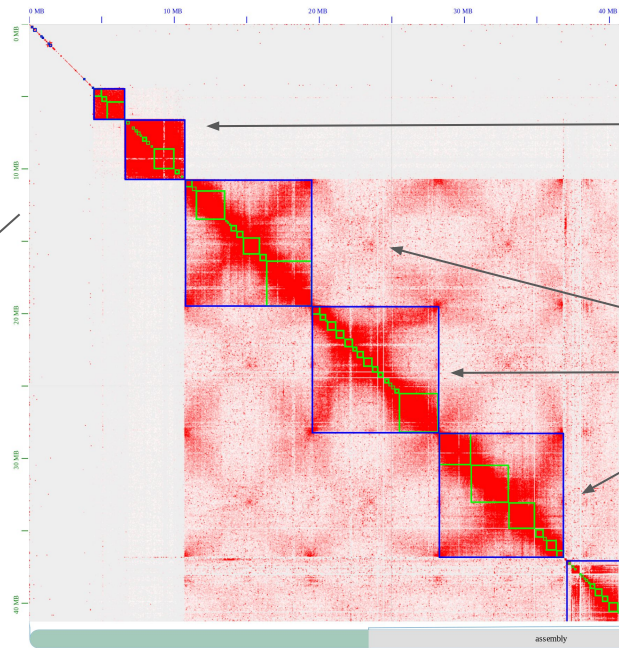
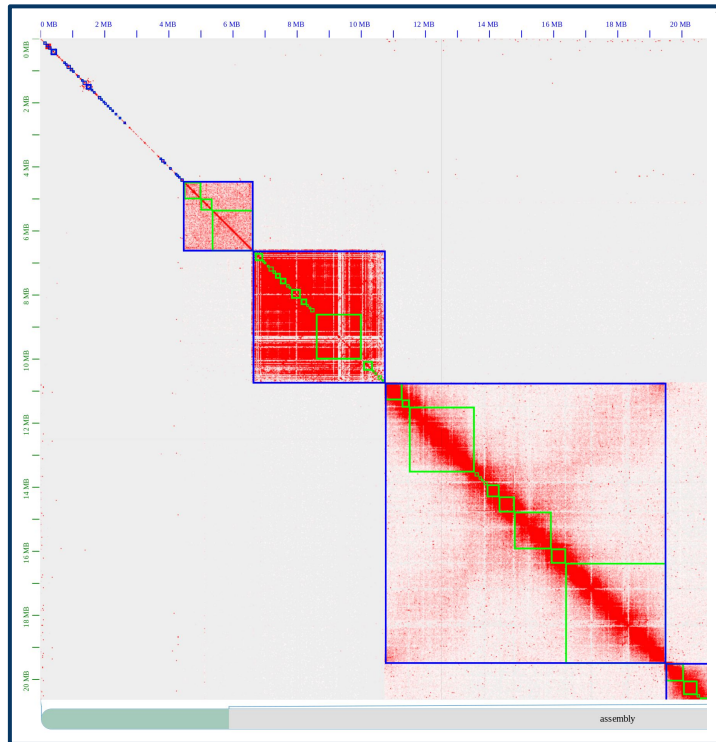
Варгнар Использовался для идентификации рРНК в геномной сборке. Поиск по базе данных NCBI rRNA database позволил выявить 5 последовательностей не включающих Eukarya с query cover >30 %

рРНК	Домен	Тип	Класс	Порядок	Семейство	Род
16S_rRNA::ctg270:585-28-5987(+)	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	Rhodospirillales [76%]	Thalassospiraceae [41%]	Magnetospira [41%]
16S_rRNA::ctg572:286-36-30130(+)	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	Nitrosomonadales [99%]	Methylophilaceae [98%]	Methylothera [96%]
16S_rRNA::ctg537:327-42-34254(+)	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	Cellvibrionales [36%]	Cellvibrionaceae [26%]	Marinagarivorus [13%]
16S_rRNA::ctg9:91891-4-920403(-)	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	Chromatiales [50%]	Thioalkalspiraceae [20%]	Endothiovibrio [20%]
16S_rRNA::ctg1908:0-814(-)	Bacteria[92%]	Aminicenantae[5%]	Saccharicenantae[5%]	Saccharicenantae[5%]		



Сборка метагенома

Juicebox Assembly Tools использовался для выявления контигов относящихся к бактериальным и последующей коррекции метагеномной сборки



Скаффолды
бактериального
происхождения

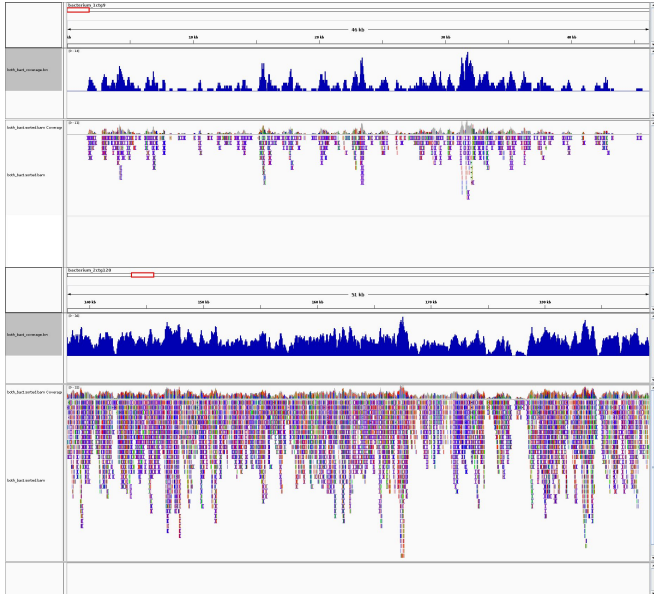
Хромосомы
губки
H. dujardini

Контиги бактериального происхождения характеризуются отсутствием контактов с геномной ДНК губки и значительным количеством контактов внутри себя и с другими контигами, принадлежащими этой бактерии. Использование метода Hi-C позволяет достоверно различать контиги, принадлежащие разным бактериям.

Наличие бактериальной ДНК в разных наборах данных

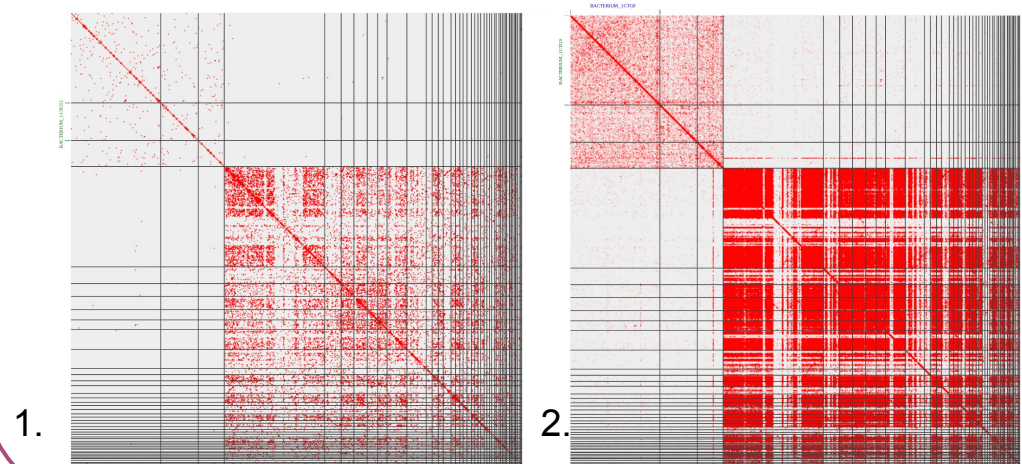
Геномные библиотеки

1. Напорге (геномная сборка)
2. Короткие прочтения Illumina



Hi-C библиотеки

1. “Весна”
2. “Осень”

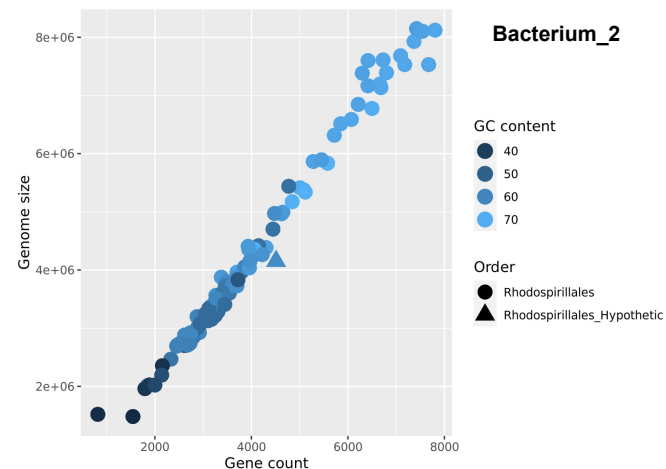
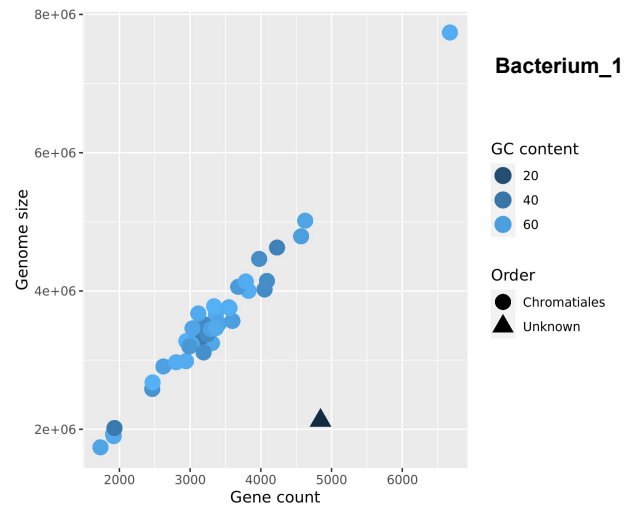


Картирование разных типов библиотек, подготовленных из материала собранного в разное время, демонстрирует присутствие бактериальной ДНК во всех образцах.

QC и классификация

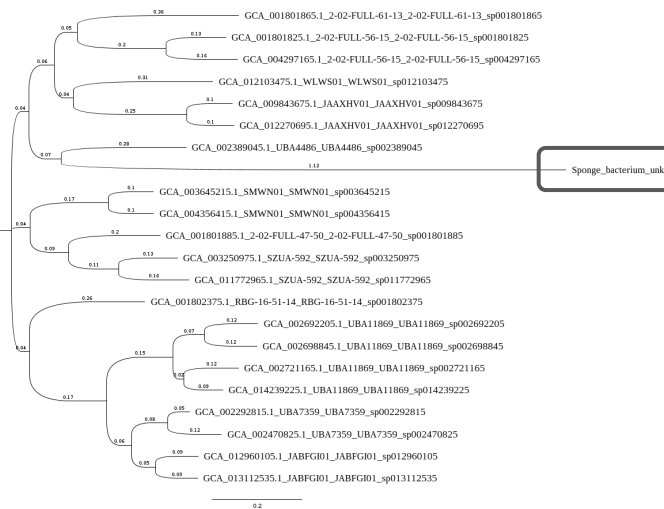
Bin Id	Bacterium_2	Bacterium_1
Набор маркеров для:	o__Rhodospirillales (UID3754) (порядок)	c__Gammaproteobacteria (UID4267) (класс)
# геномов	63	119
# генов-маркеров	336	544
Выявлены 0 раз	216	390
1 раз	106	152
2 раз	13	2
3 раз	1	0
4 раз	0	0
5 раз	0	0
Полнота, %	29.76	24.39
Контаминация, %	3.72	0.35
Гетрогенность штамма, %	18.75	0

CheckM использовался для оценки полноты сборки по наличию специфичных для таксона генов-маркеров. Дополнительно оценивалось то насколько характеристики геномов (размер, содержание GC и количество генов) согласуются с данными характеристиками для определенного таксона.



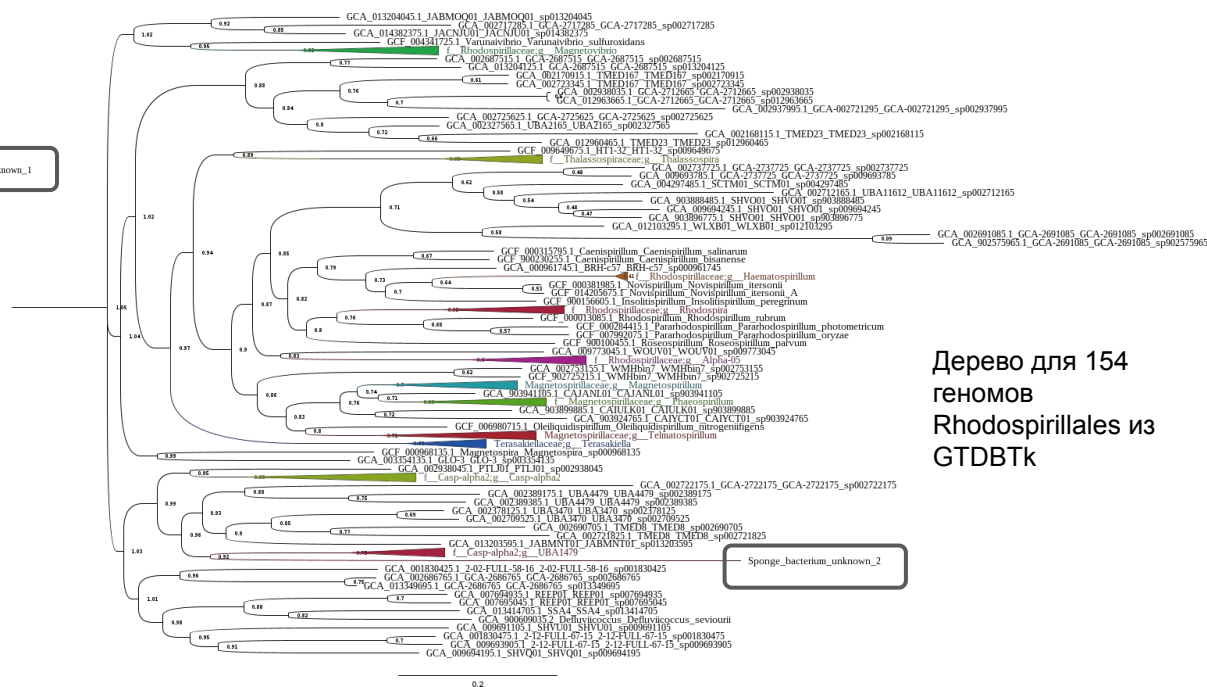
Классификация - GTDBTk

Геном	bacterium_1	bacterium_2
Классификация	d_Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA4486;f__UBA4486;g__;s__	d_Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Casp-alp ha2;g__;s__
Классификация ближайшего соседа на филогенетическом древе	N/A	d_Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Casp-alp ha2;g__GCA-2722175;s__GCA-2722175 sp002722175



Дерево для 22 геномов порядка UBA4486 из GTDBTk

Построено по 119 маркерам класса Proteobacteria



Дерево для 154 геномов Rhodospirillales из GTDBTk

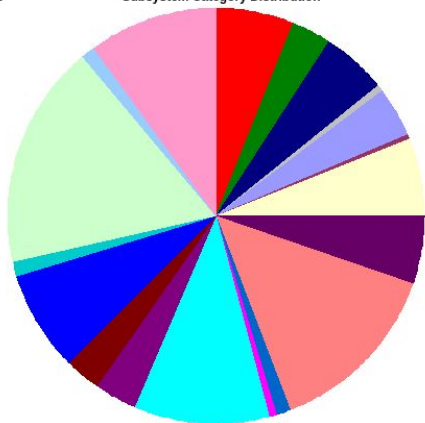
Функциональная аннотация

bacterium_1

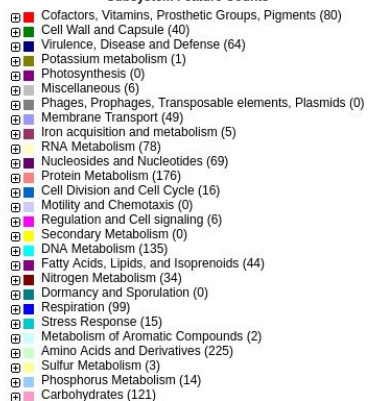
Subsystem Coverage



Subsystem Category Distribution



Subsystem Feature Counts

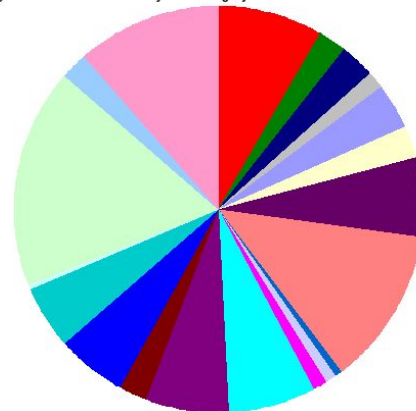


bacterium_2

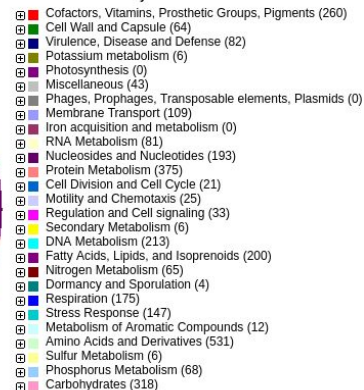
Subsystem Coverage



Subsystem Category Distribution



Subsystem Feature Counts

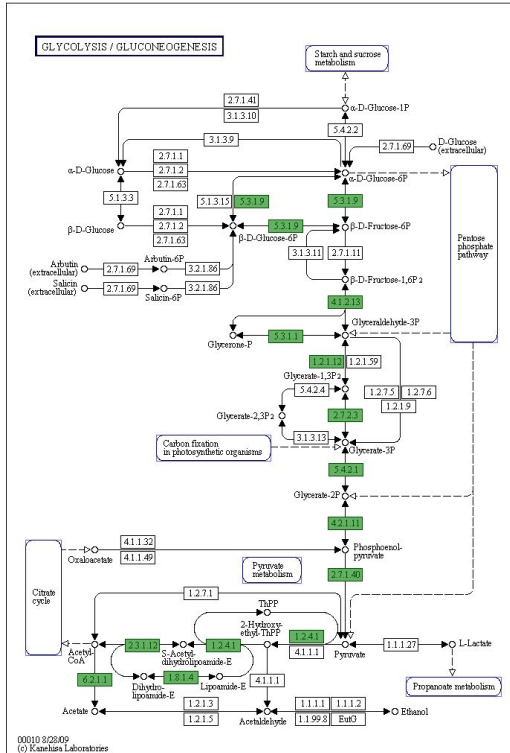


К наиболее представленным функциональным подсистемам относятся:

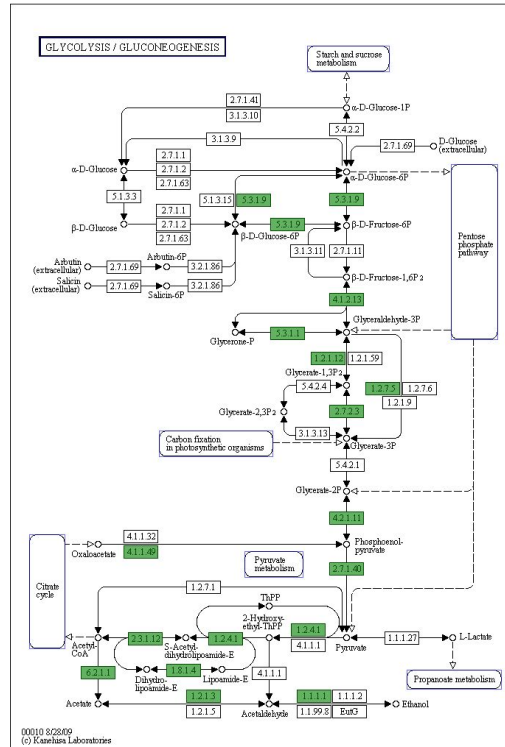
- Аминокислоты и их производные
- Метаболизм белков
- Метаболизм ДНК
- Метаболизм углеводов

Функциональная аннотация

Наиболее представленные метаболические пути



bacterium_2



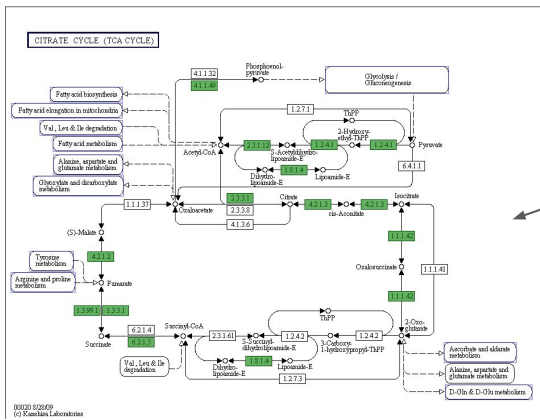
bacterium_1

Энергетический обмен -
Гликолиз и глюконеогенез

Представители обоих видов
демонстрируют схожий набор
ферментов энергетического
обмена .

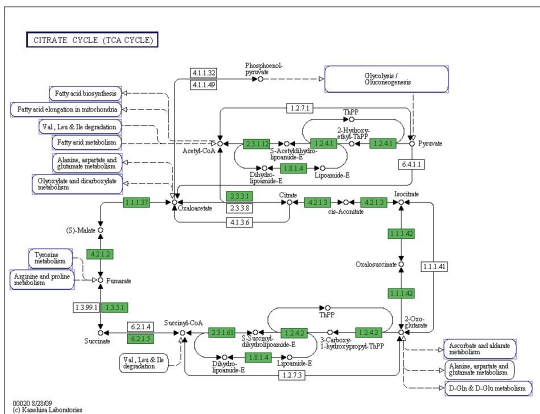
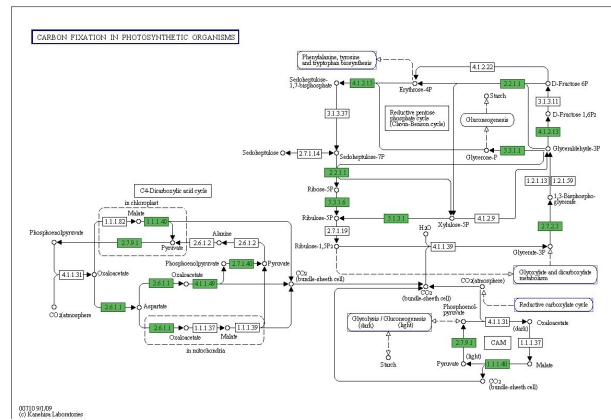
Функциональная аннотация

Наиболее представленные метаболические пути

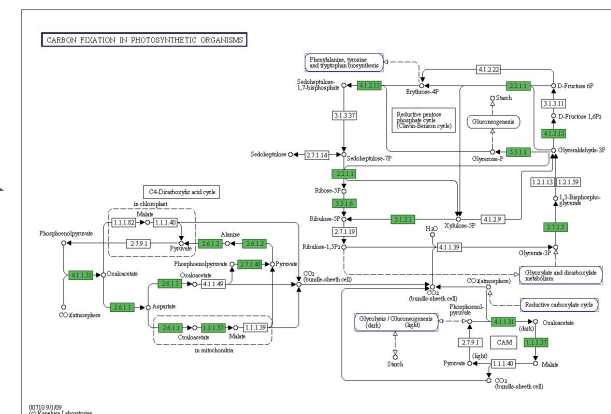


Энергетический обмен -
Цикл трикарбоновых кислот

bacterium_2



bacterium_1

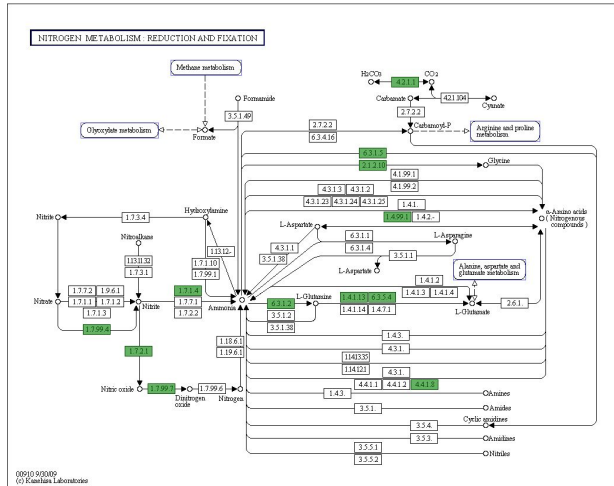


Фиксация углерода

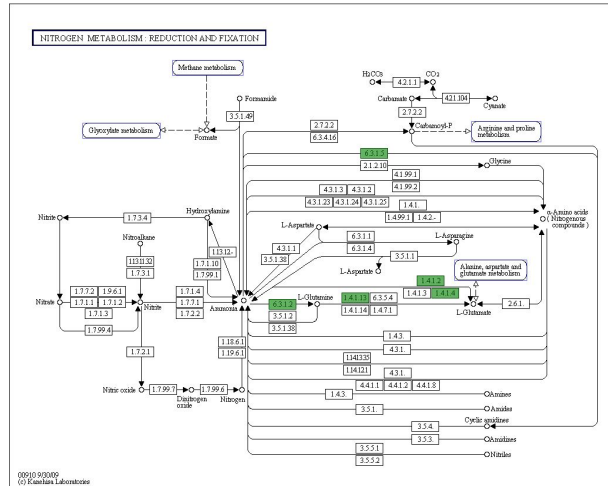
Функциональная аннотация

Наиболее представленные метаболические пути

Обмен азота



bacterium_2



bacterium_1

Спасибо за внимание!

