

Задание 1. Метагеномика

25 сентября 2025 г.

ОБЗОР

В данном задании вам нужно будет оценить влияние слабого загрязнения (**k5 – 5г керосина/кг**) и сильного загрязнения (**k25 – 25 г/кг**) на бактериальное сообщество, живущее в дерново-подзолистой почве, в сравнении с контрольными образцами (**k0 – без керосина**). Для каждой точки взято по три образца.

Задачи

1. Запустить `fastqc` на одном из образцов по выбору, описать полученные результаты
2. Статистический анализ
3. Определить таксономическую принадлежность для каждого варианта ампликона
4. Оценить разнообразие бактериального сообщества в образцах (альфа-разнообразие) и сходство образцов друг с другом (бета-разнообразие)
5. Посмотреть как распределён индекс разнообразия образцов (индекс Шеннона) в группах образцов.

Описать как меняется разнообразие бактериального сообщества со временем при разном уровне загрязнения образцов.

Задача 1. FastQC

Результатом первого этапа является обработанный файл **soil_reads.qza**. Была проведена фильтрация данных по качеству, исправление ошибок секвенирования и фильтрацию химер. Файл на выходе - **soil_ASV_table.qza** (сюда запишется итоговая таблица с вариантами ампликонов в образцах), **soil_rep_seq.qza** (тут будет файл со списком всех встреченных в образцах ампликонов), **soil_reads.dada2.stats.qza** (тут будет таблица с статистиками фильтрации в формате qza)

Задача 2. Статистический анализ

Сделаем файл с визуализацией qzv из файла qza

```
qiime metadata tabulate \  
  --m-input-file qza/soil_reads.dada2.stats.qza \  
  --o-visualization qzv/soil_reads.dada2.stats.qzv
```

Результат представлен на Рисунке 1.

sample-id	input	filtered	percentage of input passed filter	denoised	non-chimeric	percentage of input non-chimeric
SRR17307380	5128	5087	99.2	4830	4809	93.78
SRR17307593	5208	5159	99.06	4887	4866	93.43
SRR17307262	5995	5926	98.85	5754	5731	95.58
SRR17307636	6890	6812	98.87	6642	6634	96.28
SRR17307627	7624	7536	98.85	7335	7291	96.63
SRR17307643	8233	8174	99.28	7711	7705	93.59
SRR17307404	8614	8551	99.27	8158	8141	94.51
SRR17307364	8661	8605	99.35	8270	8270	95.49
SRR17307540	9247	9056	97.93	8949	8918	96.44
SRR17307436	9471	9380	99.04	9019	8993	94.95
SRR17307400	9513	9351	98.3	9127	9077	95.42
SRR17307546	9539	9471	99.29	9119	9115	95.56
SRR17307649	9577	9502	99.22	9121	9114	95.17

Рис 1. Пример работы: Визуализированная таблица со статистиками фильтрации

После всех стадий фильтрации остаётся примерно 97–99,9% исходных прочтений. Самый большой образец — SRR17307475 (~45407 прочтений), самый маленький — SRR17307380 (~5087 прочтений). Такой высокий процент сохранённых данных говорит о очень хорошем качестве секвенирования и низком уровне ошибок и химер.

Задача 3. Таксономия

Для определения таксономии использовался классификатор -

gg_2022_10_backbone.v4.nb.qza, обученный на V4 регионе всех последовательностей из базы данных GreynGenes2.

```
qiime feature-classifier classify-sklearn \  
  --i-classifier  
/projects/mipt_dbmp_biotechnology/metagenomes/soil/2022.10.backbone.v4.nb.sklearn-  
1.4.2.qza \  
  --i-reads qza/soil_rep_seq.qza \  
  --o-classification qza/soil_taxonomy.qza \  
  --p-n-jobs 16
```

Визуализация

qiime taxa barplot \

--i-table qza/soil_ASV_table.qza \

--i-taxonomy qza/soil_taxonomy.qza \

--m-metadata-file

/projects/mipt_dbmp_biotechnology/metagenomes/soil/soil_metadata_full.tsv \

--o-visualization qzv/soil_taxonomy_barplot.qzv

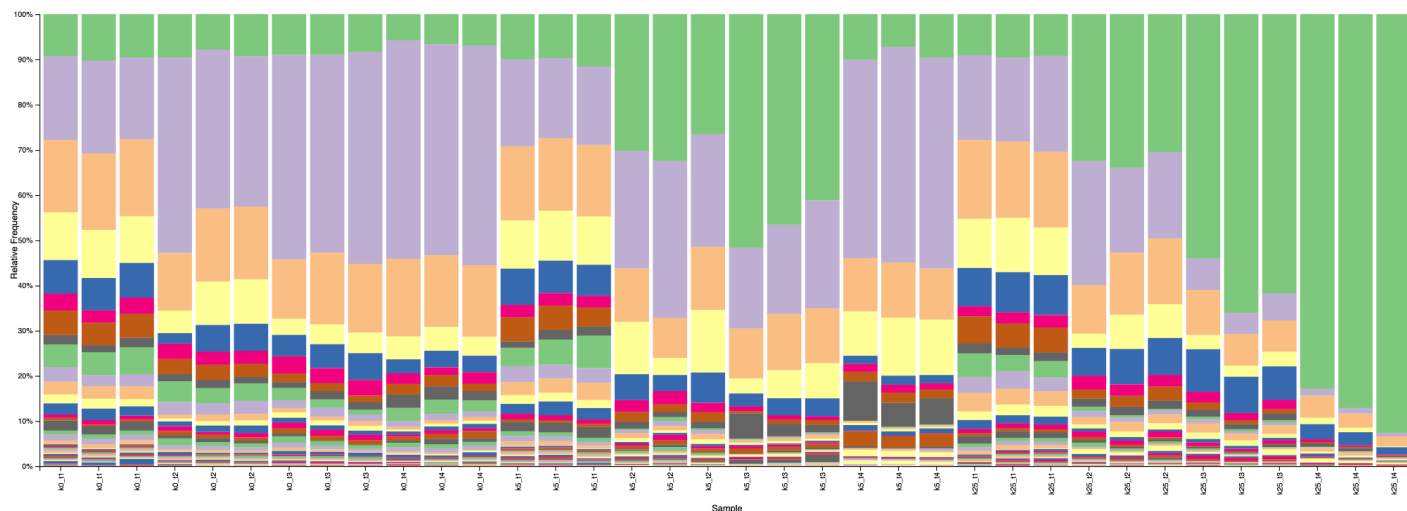


Рис 2. Распределение (level=3)

В первой временной точке микробные сообщества во всех образцах были схожи по составу и разнообразию. Со временем после загрязнения керосином наблюдалось снижение разнообразия и перераспределение таксонов. Доминирующие группы сместились от **Firmicutes** и **Bacteroidota** к **Proteobacteria**, преимущественно представителям классов **Gamma-** и **Alphaproteobacteria**.

На уровне родов в загрязнённых образцах преобладали бактерии рода **Pseudomonas**, что логично, в связи с их способностью эффективно разлагать углеводороды, включая компоненты керосина.

Задача 4. Beta разнообразие

Результаты представлены на рисунке 3. В первой временной точке образцы каждого уровня загрязнения сходны по составу микробиоты. Со временем наблюдается смещение вдоль Axis 1, отражающее изменение сообществ под воздействием керосина. При слабом загрязнении (k25) различия умеренные и к концу эксперимента заметно частичное восстановление. При сильном загрязнении (k5) изменения более выражены, а восстановление микробного сообщества не наблюдается.



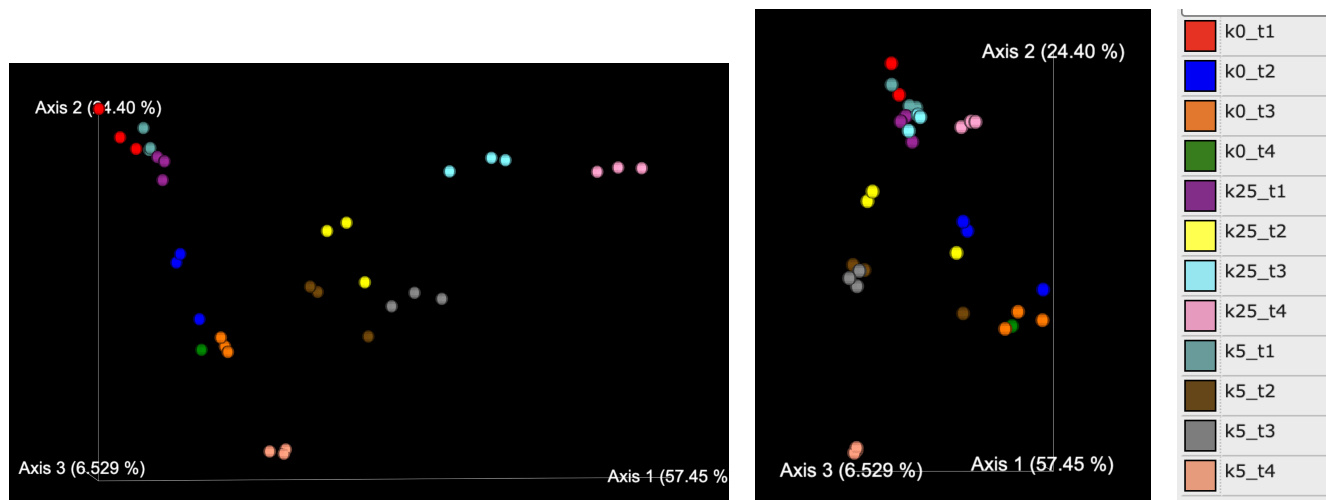


Рис 3. PCoA-анализ микробных сообществ почвы на основе UniFrac-дистанций. Каждая точка соответствует отдельному образцу; расстояния между точками отражают степень различия микробиоты. Цвет обозначает уровень загрязнения керосином (k0, k25, k5) и временную точку (t1–t4)

Задача 4. Alpha разнообразие. Индекс Шеннона

Посмотрите как распределён индекс разнообразия образцов (индекс Шеннона) в группах образцов.

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity soil_core_metrics_results/shannon_vector.qza \
  --m-metadata-file
/projects/mipt_dbmp_biotechnology/metagenomes/soil/soil_metadata_full.tsv \
  --o-visualization soil_core_metrics_results/soil_shannon_significance.qzv
```

Результат:

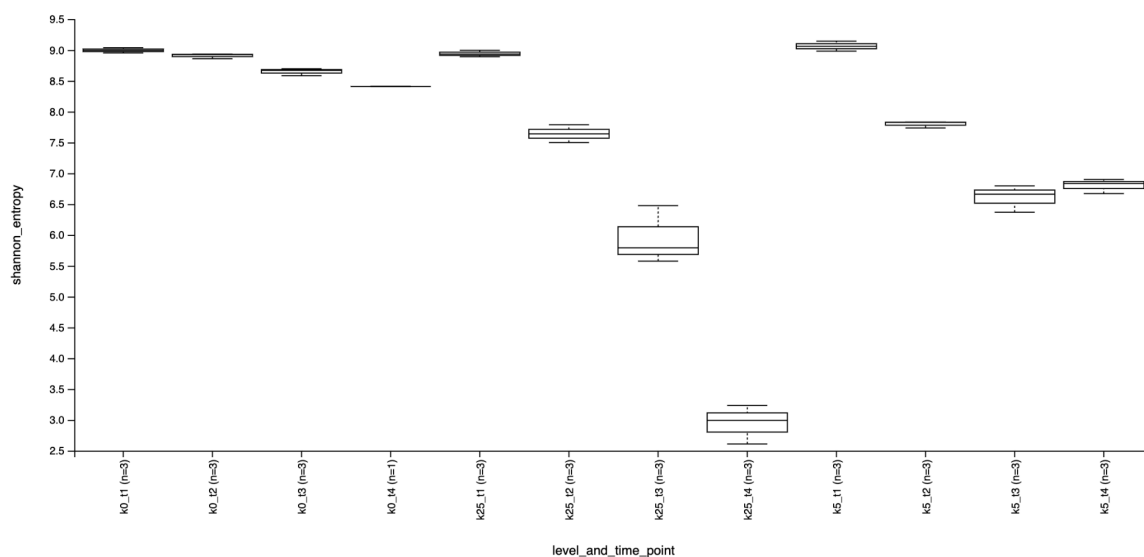


Рис 5. Показатели альфа-разнообразия микробных сообществ почвы при различной степени загрязнения керосином (k0, k25, k5) и во времени (t1–t4). По оси Y приведён уровень видового разнообразия (индекс Шеннона)

Альфа-разнообразие в контрольных образцах (k0) оказалось наибольшим, что отражает стабильное и богатое бактериальное сообщество в неизменённой почве.

При слабом загрязнении (k25) наблюдается умеренное снижение разнообразия, особенно на ранних временных точках, с последующим частичным восстановлением к концу эксперимента.

В образцах с сильным загрязнением (k5) разнообразие резко падает и остаётся низким на протяжении всего периода наблюдения, что указывает на устойчивое угнетение микробного сообщества под воздействием керосина