Browsing the Microverse

In this hands-on class, we will explore the microbial world through our web application the “Microverse Mapper” (https://microverse-mapper.uni-jena.de/). The project is under development, so we very much welcome any feedback through this Google form: https://docs.google.com/forms/d/e/1FAIpQLSesS7n9Q1ukTp8X8P2UjjrPQe5aO_mFpnr7B6bfQndV_nVK7w/viewform?usp=dialog. Finally, please enjoy and give us feedback, but **don’t post it online (social media or something) since it’s not yet published!**

Questions:

1. Please take a look at the Microverse Mapper website. Read the Help page.
2. In how many microbiome samples was *Roseobacter* found? Take a cutoff of 100 reads. From how many different biomes were these samples derived? In which biome is *Roseobacter* typically found? What are the 10 most abundant other genera in marine samples containing at least 100 *Roseobacter* reads? Which of these are generalists and which are specialists?

3. Which oral samples stand out? Can you figure out why? What does this tell you about the data?

4. How many data points are derived from rhizosphere samples?
5. Which are the most widespread microbial genera in the oceans?
6. How many different types of bioreactors are there?

7. Which other genus is *Geobacter* mostly found with?

8. Find the MGnify sample with identifier SRR7063047, which is derived from a certain species of *Porifera*. Are there similar microbiomes on the map? Where on earth do they come from? Are they all from the same *Porifera* species?

9. You can also place your own datasets onto the Microverse map. To do this, you need a taxonomic abundance table in BIOM format. Try it out. Do your samples fall where you expect them on the map? Select the 100 closest other samples. From how many different studies are they derived? Can you find the associated publications?

Answers:

1. If anything is unclear, please ask or let us know so we can fix it!
2. Use the “Smart Search” function to find out that there are 156 samples with >100 *Roseobacter* reads. In the “Biome” tab of the popup menu you can figure out that they are derived from 11 different biomes. *Roseobacter* is mostly found in the marine biome (54 samples), but many of the other annotated biomes are also related to the marine realm. By clicking on the Environmental:Aquatic:Marine slice of the pie diagram, you can reduce the selection to only the marine samples. Next, you can go

to the “Barchart” tab and select the “genus” rank to find the top ten genera (note that they are reverse sorted in the color legend):

1. *Desulfobacula*
2. *Amylibacter*
3. *Candidatus* Actinomarina
4. *Roseobacter*
5. *Glaciecola*
6. *Acinetobacter*
7. *Ascidiacei-habitans*
8. *Sulfitobacter*
9. *Algicola*
10. *Roseovarius*

Searching these again with the “Smart Search” function (100 read cutoff) yields the following numbers of samples in the “Biome” tab & if you download the csv file you can find the biomes (scroll down to the bottom of the pie chart color legend):

1. 41 & 4
2. 985 & 26
3. 2194 & 30
4. 156 & 11
5. 458 & 24
6. 9742 & 102
7. 573 & 18
8. 334 & 16
9. 34 & 6
10. 245 & 19

Based on these results, *Desulfobacula* is the most specialist and *Acinetobacter* the most generalist genus.

3. There is a cluster of oral samples on the right, surrounded by some wastewater and bioreactor samples. If you select them and click “Studies”, you will find that they are all derived from one study. The link to MGnify reveals that they are classified as “root:Host-associated:Human:Digestive system:Oral:tongue dorsum”, but the study description shows that they are actually from anaerobic digester sludge fed with amino acids. This highlights that you should always question metadata annotations!

4. Find the rhizosphere samples by clicking Host-Associated:Plants:Rhizosphere in the color legend (Biome GOLD coloring). There are 2,084 rhizosphere samples.
5. Clicking Environmental:Aquatic:Marine:Oceanic via the color legend (Biome GOLD coloring) gets you 1,447 samples. In the “Enrichment” tab, you will find that *Coxiella*, *Rhodopirellula*, *Marinoscilla*, *Woeseia*, and *Candidatus* Actinomarina are present in the largest numbers of ocean samples (1,287, 1,141, 998, 975, and 966, respectively).
6. Find 391 Engineered:Bioreactor samples through the color legend (Biome GOLD coloring). Click the “UMAP” tab and select your favorite taxonomic rank. Depending on the rank you will find ~5-10 clusters.

7. Smart Search for *Geobacter* and choose your cutoff. There are 4,051 samples with at least 50 *Geobacter* reads. The “Enrichment” tab shows that these samples also contain a lot of *Bryobacter* and *Haliangilum*.

8. Go to “Advanced Search” and enter “(SRR7063047)[mgnify_id]”. Zoom in on the small cluster of host-associated samples and select all of them. Zoom back out with “Reset View” and find out that these sponges were found in the Labrador Sea by selecting “Geographic” under “Select Map”. Select all 45 host-associated samples from that region and morph back to the t-SNE or UMAP. This reveals that there are two clusters, possibly from two different *Porifera* species? All derived from the same MGnify study, MGYS00002780, whose description actually mentions three deep-sea sponges: *W. bursa*, *S. fortis*, and *G. barretti*.