Group Project: Pundemic

Description of challenge:
In the summer of 2024, a mysterious disease dubbed the “pundemic” by the media began cropping up worldwide. Even if they had never previously engaged in wordplay, those afflicted simply cannot refrain from making puns at every opportunity, much to the chagrin of their colleagues, friends and family. The medical community scrambles to find an explanation for this highly pun-usual condition.

Once strong evidence emerged of a link between the pundemic and changes in the gut microbiome, a doctor at the University Hospital Zurich set up a clinical trial to fecal microbiota transplants (FMT) as a possible treatment. The trial involved collection of fecal microbiome samples from pundemic patients before and after the trial, from both treatment and placebo groups. Pundemic severity in patients was quantified in terms of puns per hour. Fecal samples were collected from the FMT donors as well.

Because the bacterial and fungal gut microbiome are both of interest, the USZ team collected both 16S rRNA gene and ITS data from the study cohort. You have been tasked with analyzing the ITS data in order to further explore the connection between pundemic symptoms and an altered gut mycobiome composition, as well as the potential of FMT as a pandemic treatment option. You have received DNA sequences as well as metadata allowing you to distinguish pundemic from healthy samples.

The medical community is eagerly anticipating your results, and patients' families are hoping you can help end months of PUNishment for them.

Some specific questions*:

1. Look at your metadata. What are the main groups and timepoints being compared in the study? What can you say about the demographic distribution of your cohort?
    - Hint: Explore the metadata and create visualizations to describe your patient cohort.
2. Look at the taxonomic composition of the gut mycobiome in the pun group prior to treatment as well as in the donors. Which phyla are most abundant and/or prevalent? Are these phyla the same or different in the pun and donor groups?
3. Now consider the treatment outcomes. What differences in mycobiome composition can you see between the FMT response, FMT no response and placebo groups? Which taxa are differentially abundant in the FMT response group pre- and post-treatment?

Information on provided datasets:

1. **Sequences**: The sequences originate from the ITS-1 region and were sequenced on an Illumina MiSeq machine. The sequence files contain demultiplexed single-end sequences and are provided as a QIIME 2 artifact (semantic type: `SampleData[SquencesWithQuality]`).
2. **Sample metadata**: The metadata table contains additional information on each collected sample. Each row represents a unique sample that can be identified with its unique ID. The remaining columns depict health data of the patient.

Where to get this data:
(Use `wget` to download these links in your analysis notebooks for fully executable analyses!)

1. Sequences: https://polybox.ethz.ch/index.php/s/o8HqHJqvuf9e2on
2. Sample metadata: https://polybox.ethz.ch/index.php/s/7LxWSbaw2q37yof

Questions? Your course instructors are here to help. Meet with instructors during in-class group work sessions or ask questions via email.

Contact TA: Milo Schärer, milo.schaerer@hest.ethz.ch

** **Note on questions:** This is not a list of the only questions to answer about this dataset. All group semester projects are expected to incorporate most analysis topics covered in the course when possible (see Group Semester Projects assignment document). This is merely a list of specific questions that should be answered/ attempted in addition to the "standard" analyses covered. Not everything will work or give a clear result! But that does not mean that you should not try. Your goal as scientists is to figure out why.*