Metabarcoding analysis

- Please log into the VM using the given credentials.
  - The hostname is `quadram.seq.space`.
  - The username is `bio0X`, where *X* will be a number provided via Skype
  - The password is `Norwich0X`, where *X* will be a number provided via Skype
- You can install tools using Miniconda, and (if needed) you can write small scripts using your favourite editor.
- The session is monitored and logged
- Should you wish to skip a step, you'll find the pre-computed output in `/media/data/precomputed/16S/`.

16S Dataset

This test will ask you to perform some computations related to 16S data, using a simplified and unrealistic protocol.

Explore your data

In the shared directory `/media/data/16S/` you will find a metadata file and a set of reads.

- Examine the metadata file and check what is the link between the metadata and the read filenames in `/media/data/16S/`.
- Create a directory in your home called `~/16S/reads`.
- For each stool sample, take 5,000 reads and place them in `~/16S/reads`.
- Create a subdirectory `merged` inside your 16S folder (*i.e.* `~/16S/merged`).
- The read pairs for each sample overlap, so please merge them using a tool such as *FLASH* or *USEARCH*, saving the output to `~/16S/merged`.

Pre-computed steps

Some steps were made for you, You will find the output of these steps in `~/output/`.

1. Create a single file with all the reads from all of the samples produced in the last step (call it `~/16/merged/all.fastq`), relabeling each *read name* to begin with the sample name (filename prefix) and a progressive number. You can use the dot as separator.
2. Create an OTU table by running the command below:

```
merged_to_otus ~/16/merged/all.fastq <output_directory>
```

Now:

- Extract the sequences named *OTU1*, *OTU2*, *OTU3*, *OTU4*, *OTU5*, from the output (otus.fa) of the previous step and save it as `~/16S/five.fasta`.

Numerical analysis

Using R or Python (Pandas):

1. Calculate the sum of counts of each sample (column) of the `otutab.txt` and `otutab.raw` files.
2. Create a stacked bar chart of the OTU composition of each sample from `otutab.txt`.
3. Create a `otutable.sorted.txt` file, sorting the table by total otu abundance (sum the counts in all samples)
4. Extract the top 5 most abundant OTUs from the OTU database and save them as `~/16S/top5.fa`