🇺🇦

tp-prylutskyi23

Understanding your results

You have been given a download link to receive the following data:

Raw sequence data

The sequencing files, in the form of paired-end fastq files generated as standard by the Illumina sequencing platform, can be found in the ‘ITS2’ and ‘SSU’ folders. ‘ITS2’ contains the data generated by sequencing an amplicon produced using primers that amplify the ITS2 region of the rDNA (SPUN uses ITS3/ITS4). ‘SSU’ contains the data generated by sequencing an amplicon produced using primers that amplify the SSU region of the rDNA (SPUN uses WANDA/AML2).

Use this data for your own bioinformatics analyses if you do not want to use the Phyloseq object generated by SPUN’s bioinformatics pipeline.

Quality metrics of raw sequence data

Inside the ‘ITS2’ and ‘SSU’ folders is a folder named ‘MultiQC_report’. This contains a folder with a code name, and inside is a file called ‘multiqc_report.html’.

- Use this report to understand the quality of the raw sequence reads. The most important information is the ‘Sequence Quality Histogram’ that tells you the average quality score across each read in a sample. If you have failed samples, these may have been removed from the analysis, and if so, this will be listed in this document. The ‘Sequence Duplication Levels’ graph can be ignored, as this always fails in metabarcoding analyses.

Phyloseq objects

You have been provided with phyloseq objects, these contain all information about the samples and their associated identified taxa. This was generated as a result of cleaning and filtering the sequence data, followed by clustering into OTUs (for ITS2 EcM data) or forming ASVs (for SSU AMF data). Use these to carry out analytics into the community composition of your samples within R.

There are multiple phyloseq objects that have been provided to you:

- phyloseq_decontam.R - A Phyloseq object that has had all contaminating OTUs or ASVs removed (using the R package decontam.R and your negative controls). You will have a phyloseq_decontam.R file inside each of your results folders, 'lotus2_ITS2' and 'lotus2_SSU_ASVs'. This phyloseq object contains all taxa (except contaminants) identified in your samples, including metazoan taxa, non-mycorrhizal fungal taxa, and more. *Use this if you are interested in analysing more than just mycorrhizal data, and would also like to look at saprotrophs and other fungi in addition to mycorrhizal fungi.* Decontamination could not be carried out due to a lack of negative control samples. The phyloseq.R object you have been provided has had all OTUs represented by a low read count (1%) filtered out. This phyloseq object is not included in your analysis.
- ecm_physeq.R - A Phyloseq object that contains only EcM taxa. This was produced by using a modified version of the FungalTraits database to remove any taxa from the ITS2 Phyloseq object that are not designated with ectomycorrhizal as their primary lifestyle. This Phyloseq object can be found inside 'lotus2_ITS2' inside your results folder. The script and database used to generate this can be found on the SPUN Github. *Use this if you are only interested in analysing ectomycorrhizal fungi from your ITS data.*
- phyloseq.R - This is the initial phyloseq object produced by Lotus2 prior to the removal of likely contaminant OTUs or ASVs, it contains all OTUs, not just EcM.

Bioinformatics processing

The software Lotus2 was used to designate OTUs from ITS2 sequences and ASVs from SSU sequences. Exact commands including parameters are listed below. Following generation of a phyloseq object by Lotus2, OTUs or ASVs made up of a

very low read count are filtered out, and the decontam program is run to remove contaminating OTUs or ASVs based on the included presence of which taxa are identified in the control samples. Full code and scripts can be found on SPUN's Github page.

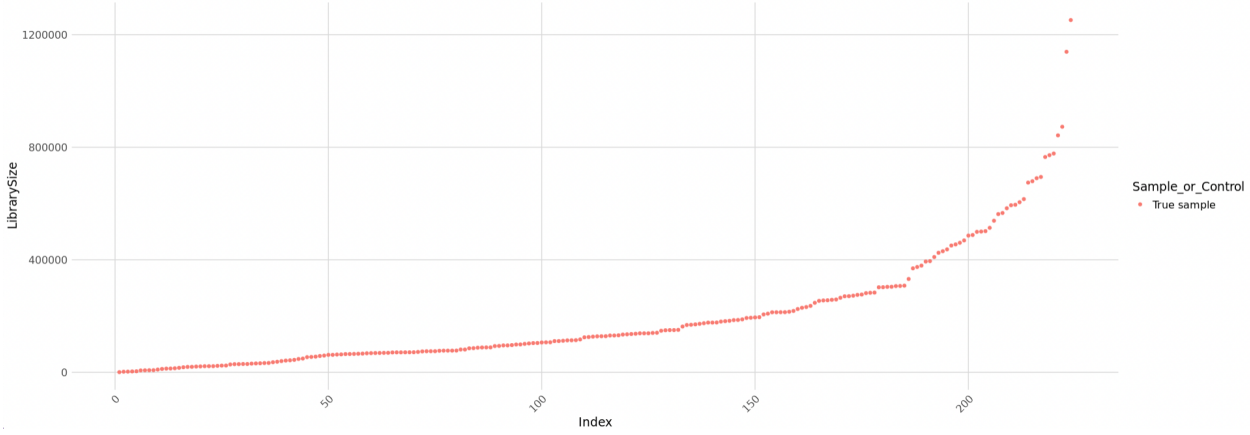
ITS2 data

Missing ITS2 samples

Aa Sample name	Reason	Run info	Resequencing progress
aku_m11	Low read count	Seq run	

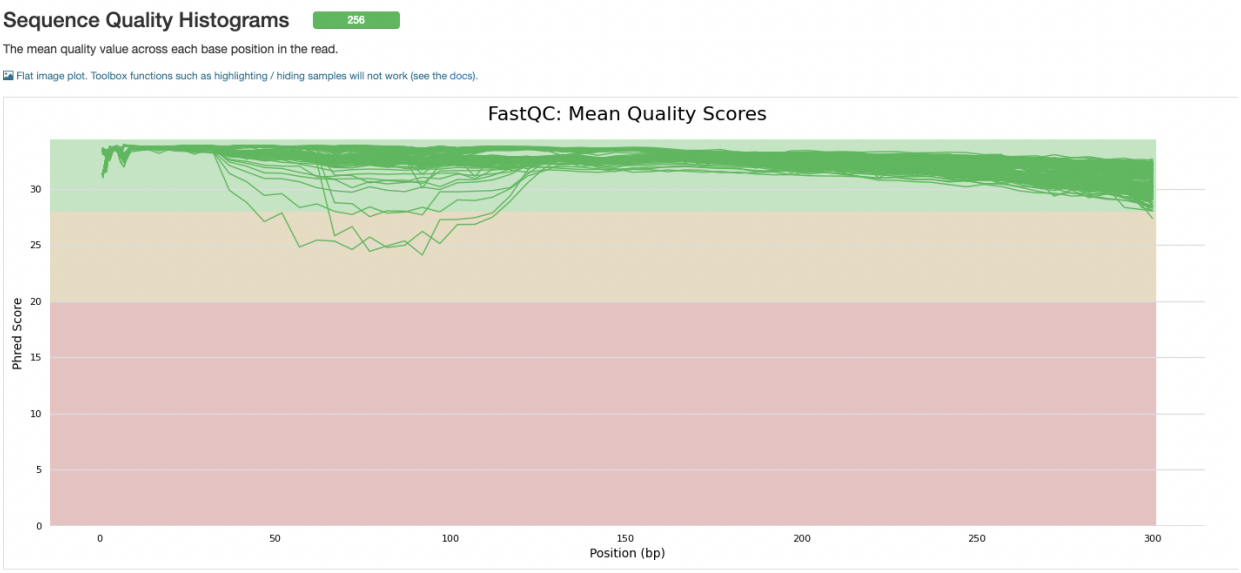
Library size

Figure below displays read count of True Samples vs Control Samples.



QC report

A Quality Control report describing the sequence quality of each sample file is generated by MulitQC. This is also included in a folder inside the folder ‘ITS2/MultiQC_reports’



OTU table

OTUs were produced using VSEARCH within the Lotus2 pipeline:

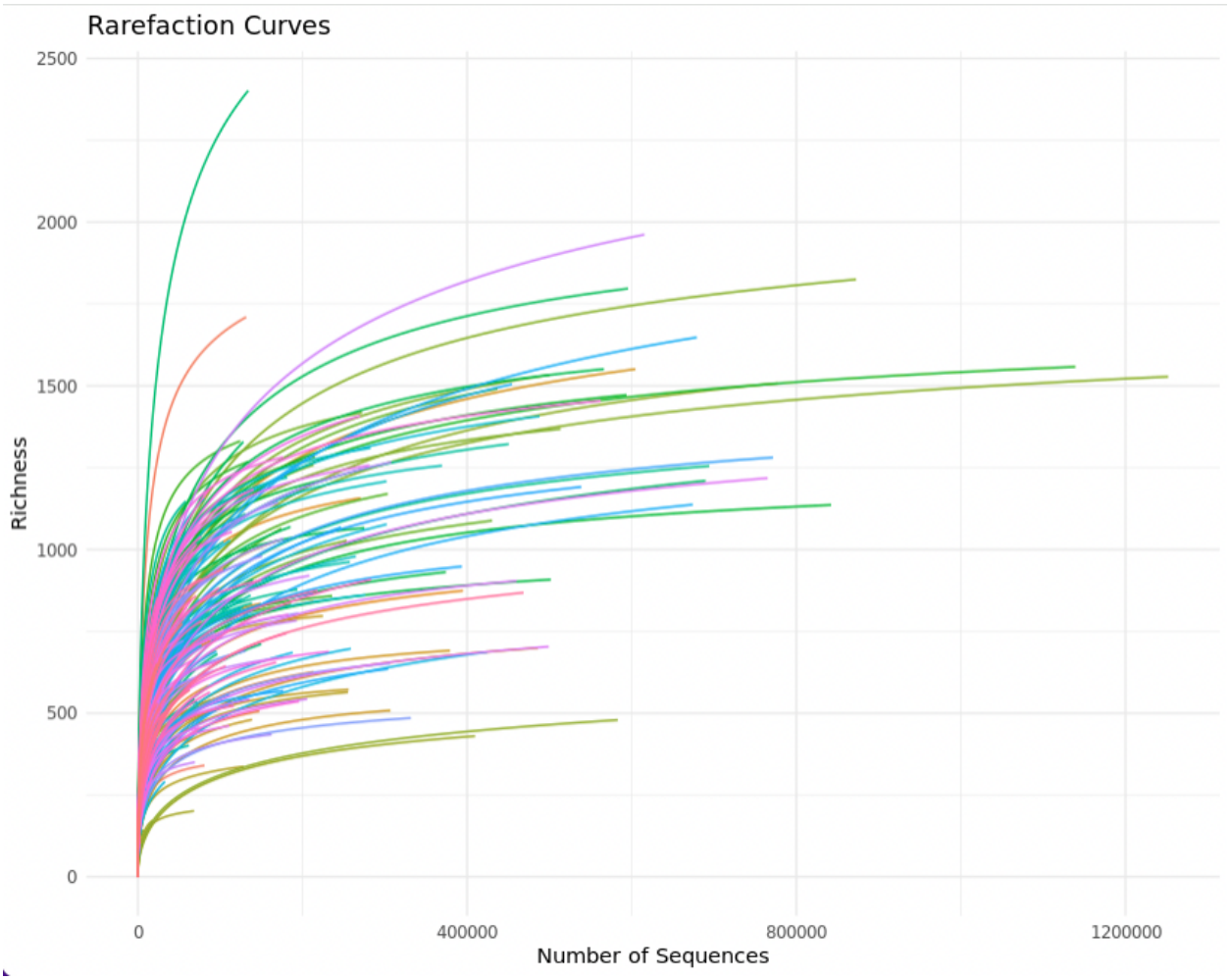
```
lotus2 -i Project_name/ITS2 \
-o lotus2_ITS2_OTUs \
-m Project_name_metadata.txt  \
-refDB UNITE \
-amplicon_type ITS2 \
-LCA_idthresh 97,95,93,91,88,78,0 \
-tax_group fungi \
-taxAligner blast \
-forwardPrimer GCATCGATGAAGAACGCAGC  \
-reversePrimer TCCTCCGCTTATTGATATGC \
-clustering vsearch \
-derepMin 10:1,5:2,3:3 \
-id 0.97
```

Decontamination

No decontamination took place because a negative control sample was not provided and sequenced.

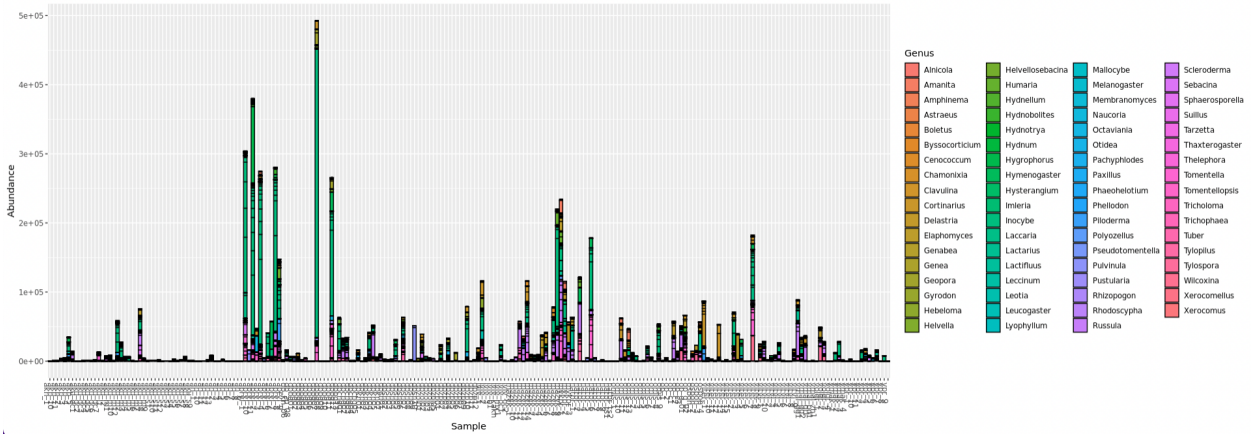
Rarefaction curves

Below is a rarefaction curve displaying the richness vs sequencing depth of each of your samples. This can be used to infer whether the sequence depth for each sample was sufficient to capture the fungal diversity in the sample. If not all of your sample diversity is plateauing, you might want to subsample (or rarefy) the reads to a fixed number, or sequence depth. This is to ensure that diversity isn't inflated in one sample compared to another just because the sequencing depths are different.



Your data at a glance - EcM only

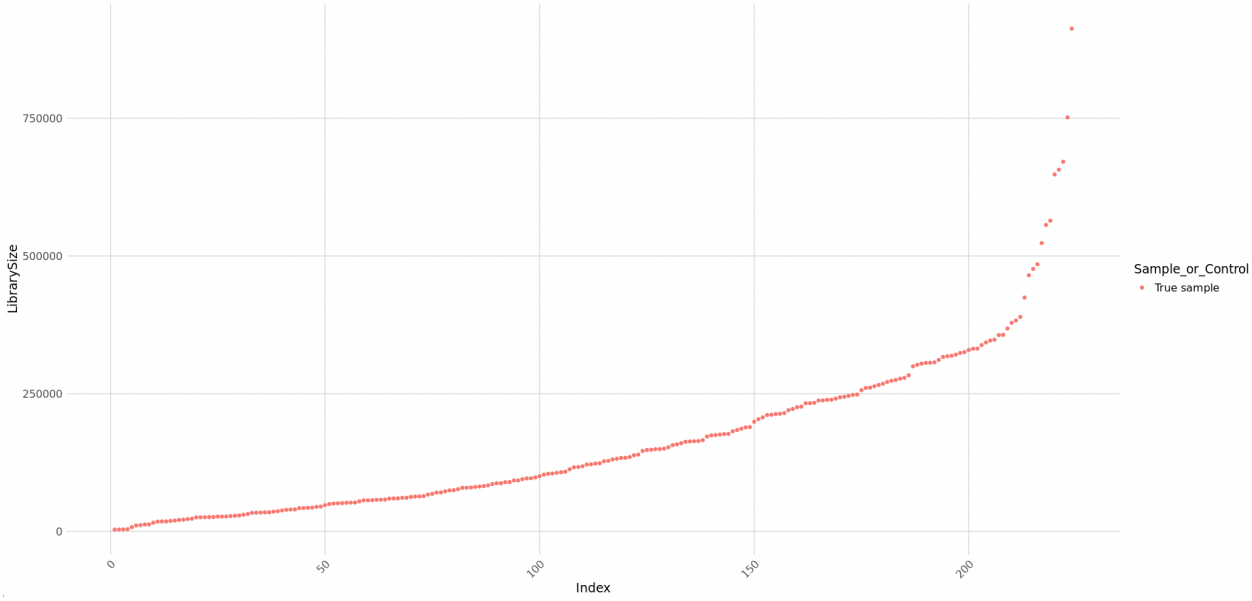
Below is a bar plot displaying the EcM genera in your ITS2 samples from the 'ecm_physeq.R' object provided in your 'Results/lotus2_ITS2' folder. This is provided for you to check that the only taxa shown are ectomycorrhizal and that you have no contaminants remaining in your control samples.



SSU data

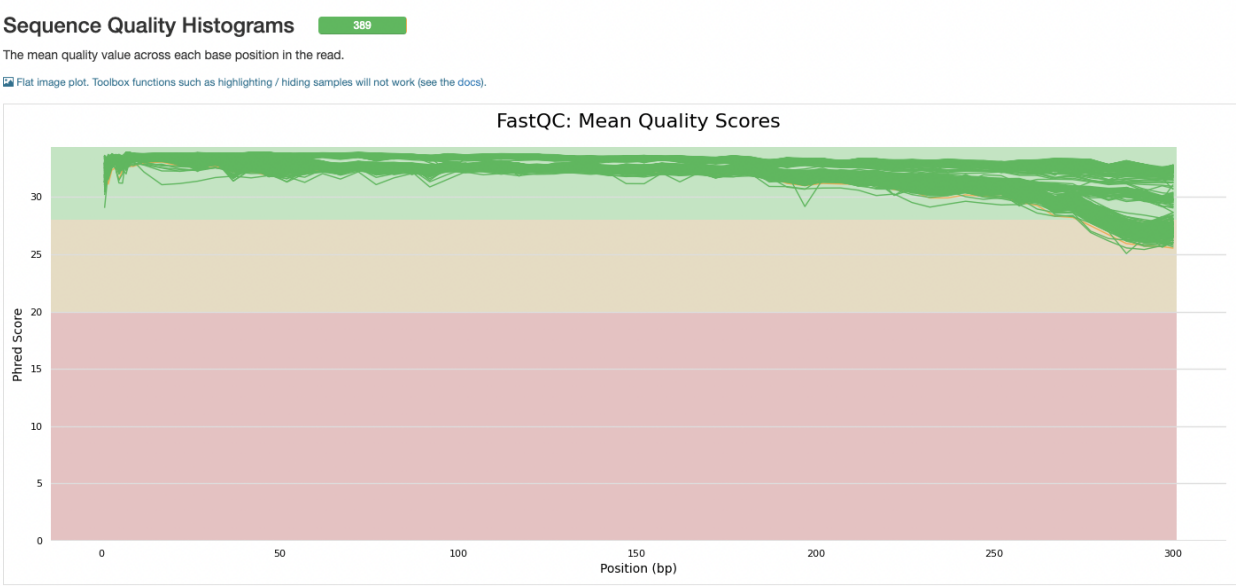
Library size

Figure below displays read count of True Samples vs Control Samples.



QC report

QC report generated by MulitQC located in file SSU/MultiQC_report



ASV table

ASVs were produced using DADA2 within the Lotus2 pipeline:

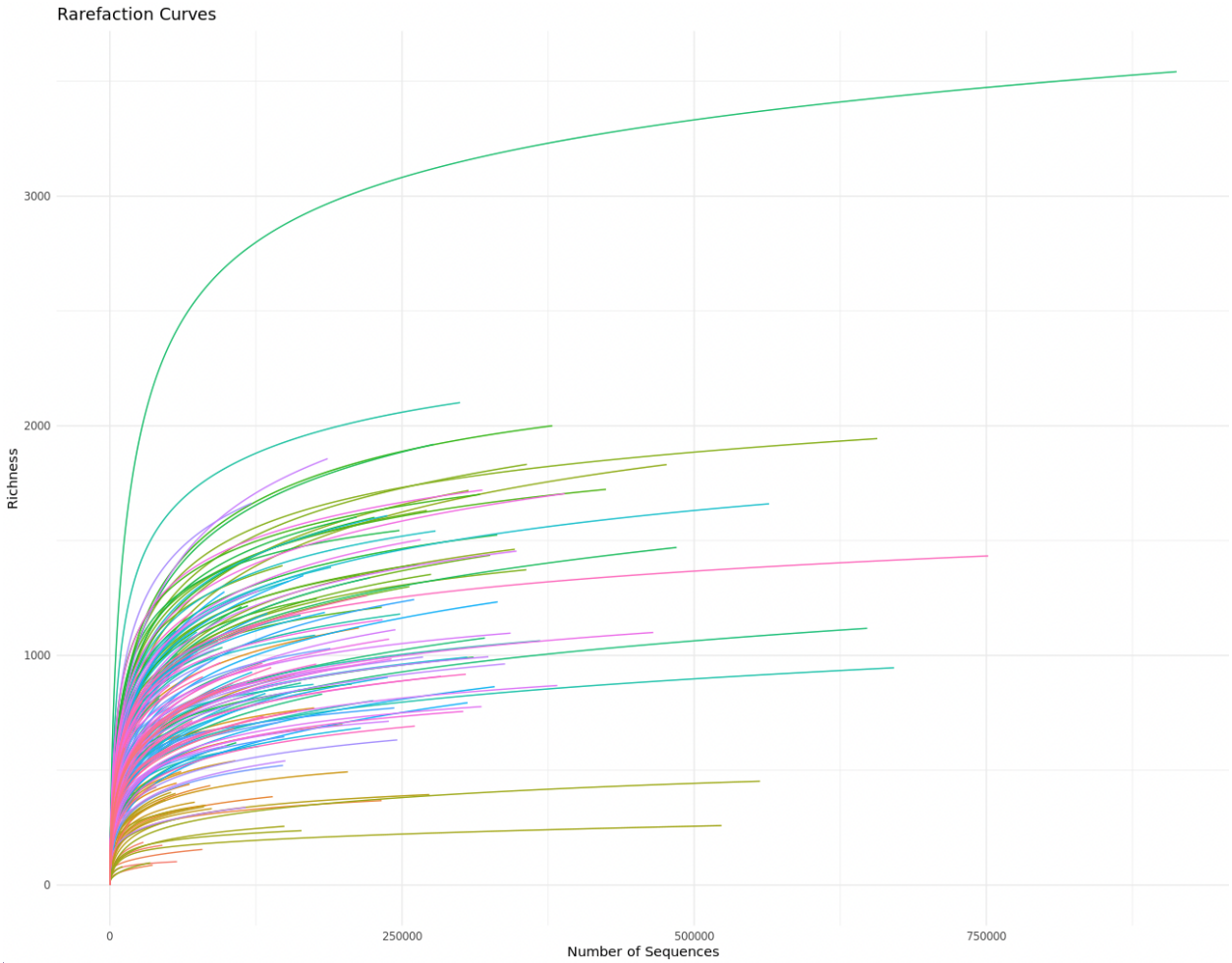
```
lotus2 -i ~/seqdata/ukraine/SSU \
-o ~/seqdata/ukraine/lotus2_SSU_ASVs  \
-m ~/seqdata/ukraine/SSU_Mapping.txt \
-refDB ~/maarjam_db_SSU_GF/vt_types_fasta_from_05-06-2019.qiime
-tax4refDB ~/maarjam_db_SSU_GF/vt_types_GF.txt \
-amplicon_type SSU \
-LCA_idthresh 97,95,93,91,88,78,0  \
-tax_group fungi \
-taxAligner blast \
-clustering dada2 \
-LCA_cover 0.97 \
-derepMin 10:1,5:2,3:3 \
-sdmopt ~/miniconda3/envs/lotus2/share/lotus2-2.34.1-0/configs/s
```

Decontamination

No decontamination took place because a negative control sample was not provided and sequenced.

Rarefaction curves

Below is a rarefaction curve displaying the richness vs sequencing depth of each of your samples. This can be used to infer whether the sequence depth for each sample was sufficient to capture the fungal diversity in the sample. If not all of your sample diversity is plateauing, you might want to subsample (or rarefy) the reads to a fixed number, or sequence depth. This is to ensure that diversity isn't inflated in one sample compared to another just because the sequencing depths are different.



Your data at a glance - AMF only

Below is a bar plot displaying the AMF genera in your SSU samples from the 'amf_physeq.R' object provided in your 'Results/lotus2_SSU/ASVs' folder. This is provided for you to check that the only taxa shown are arbuscular mycorrhizal fungi and that you have no contaminants remaining in your control samples.