Microbiome 590-12
Microbiome Analysis

ITS Practicum

Christine Hawkes

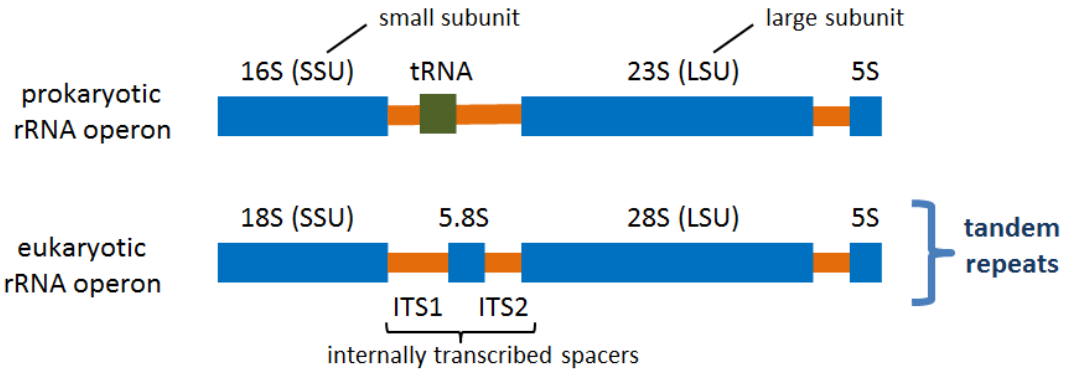
NC STATE UNIVERSITY

Upcoming deadlines

- **Feb 9** – proposal for paper selection due
  - Make “Final Project Proposal” folder in your class GitHub repo
  - Proposal is 1-2 page pdf on
    - Why this paper is a good choice
    - Your plans for re-analyzing the data and how that differs from the original
    - Confirmation that the sequences and metadata are available for you to carry out the re-analysis
    - Confirmation regarding R code
- **Feb 22** – (after proposal approval) download data from SRA
- **Feb 22** – completed HPC training (certificate sent to me)

ITS Sequences for fungal identification

- Practical species concept
- Previously we used 16S rRNA sequences to identify prokaryotes
- rRNA sequences are also used in amplicon sequencing to identify fungi



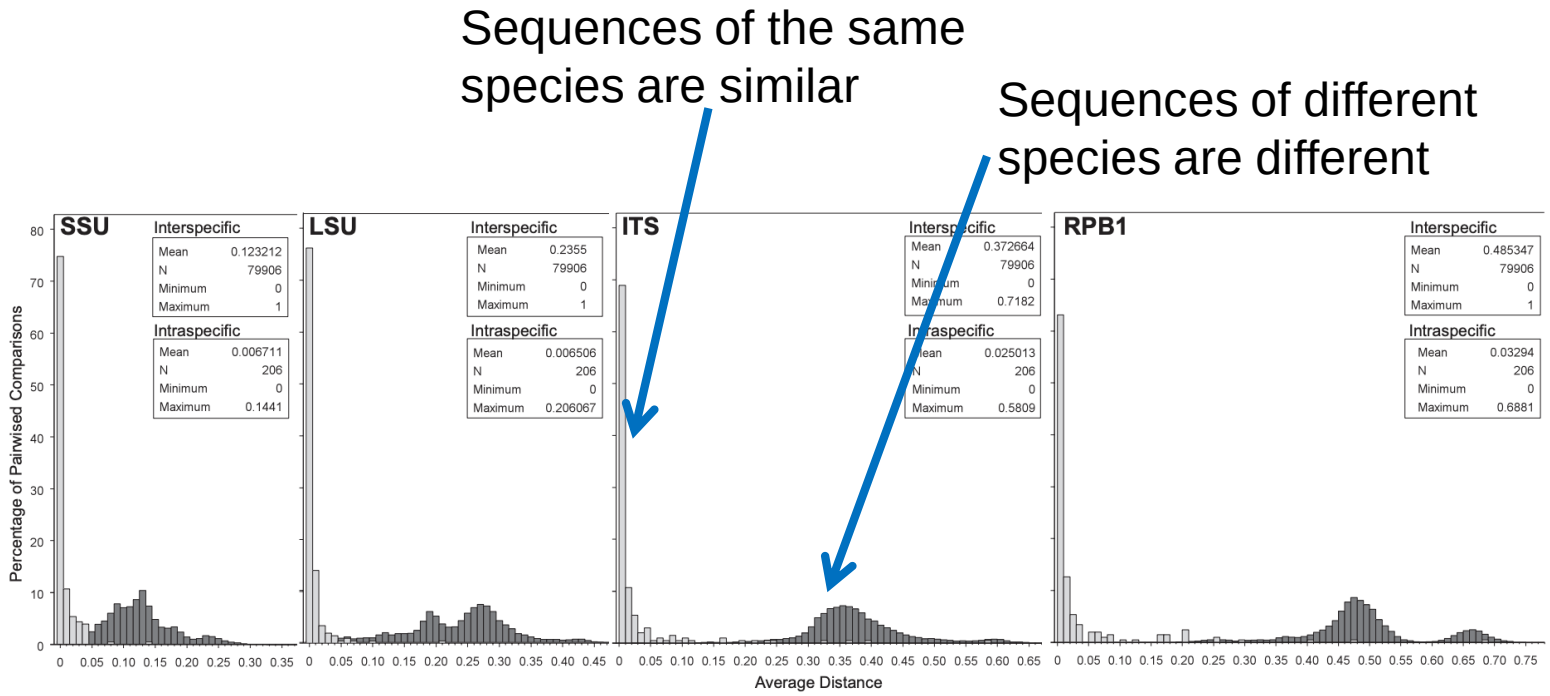
Type	LSU	SSU
prokaryotic	5S - 120 bp 23S - 2906 bp	16S - 1542 bp
eukaryotic	5S - 121 bp 5.8S - 156 bp 28S - 5070 bp	18S - 1869 bp

http://www.sixthresearcher.com/amplicon-sequencing-and-high-throughput-genotyping-metagenomics/

Choosing a marker gene for fungi

- A good marker gene should be similar within species and different between species

- ITS has a clearly defined “barcode gap” – distance between the intra- and interspecific peaks
- LSU can also be used for fungi (and good for phylogenetics)
- Do not use SSU for fungi



SSU: The 18S nuclear ribosomal small subunit rRNA gene (16S homolog)
LSU: The 28S nuclear ribosomal large subunit rRNA gene
ITS: internal transcribed spacer region
RPB1: largest subunit of RNA polymerase II

Schoch (2012) PNAS doi.org/10.1073/pnas.1117018109

Bioinformatic processing

- Of 360 tested bioinformatic approaches on a mock community, DADA2 results appeared most sensitive and similar
- LULU curation increased precision compared to other filtering methods

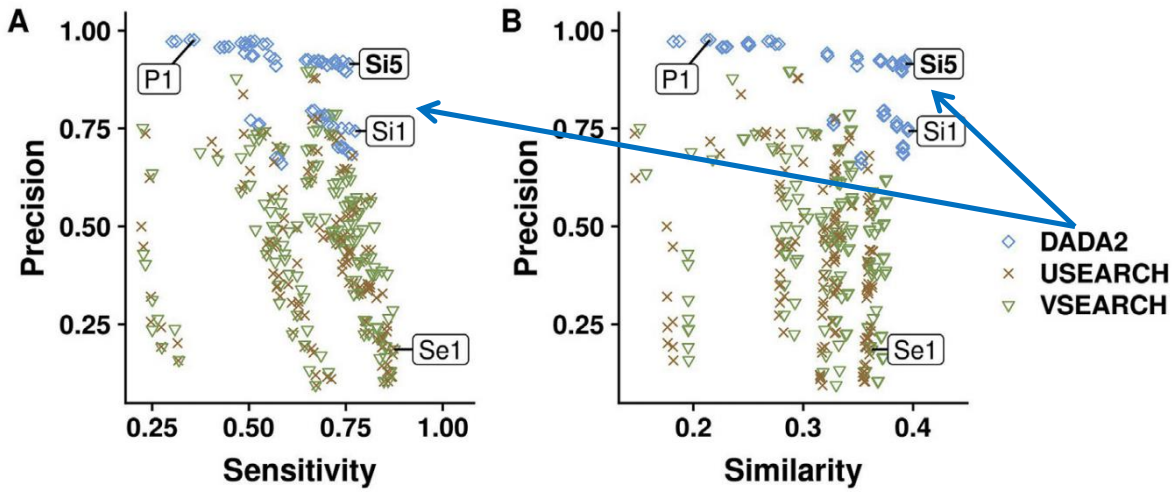


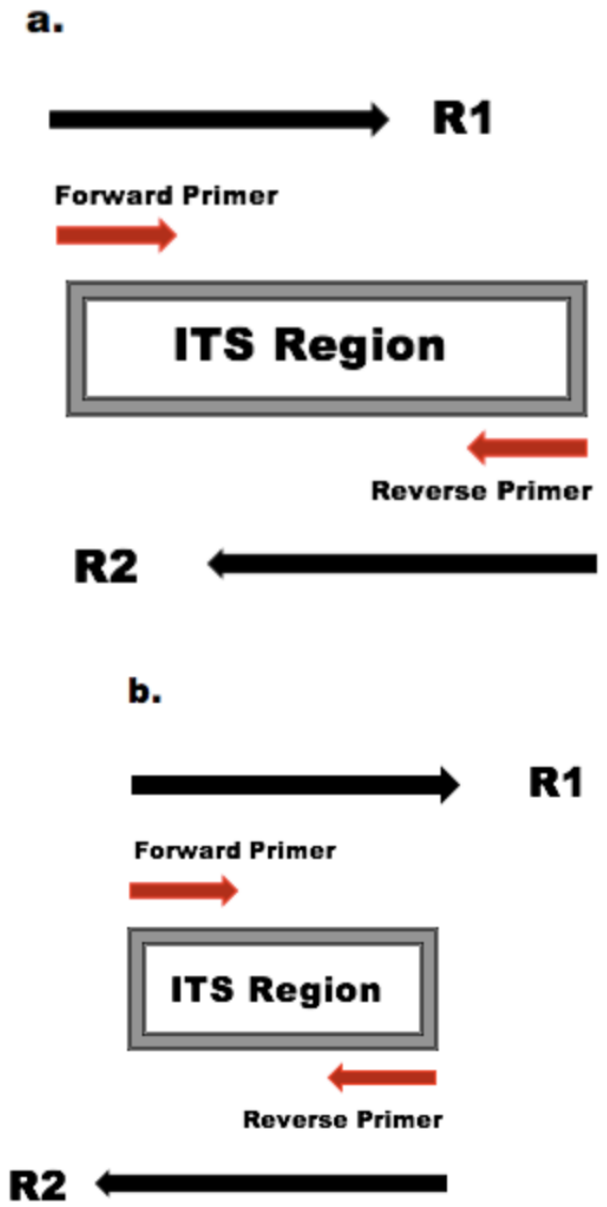
Table 2
List of the 10 most precise approaches. Sensitivity, precision and similarity values were averaged over the three replicates for each bioinformatic approach. Richness is defined as the mean number of ASVs. LULU was applied with default settings.

Approach	Assembly	Extraction	Variation	Chimeras	Filtering	Richness	Sensitivity	Precision	Similarity
P1	PEAR_150	No	DADA2	Retained	LULU	69	0.358	0.976	0.215
P2	PEAR_150	No	DADA2	Removed	LULU	67	0.347	0.976	0.212
P3	CUTADAPT_MERGED	No	DADA2	Retained	LULU	100	0.515	0.973	0.271
P4	CUTADAPT_MERGED	No	DADA2	Removed	LULU	98	0.504	0.973	0.268
P5	FASTQJOIN_150	No	DADA2	Retained	LULU	61	0.312	0.973	0.187
P6	FASTQJOIN_150	No	DADA2	Removed	LULU	59	0.302	0.972	0.182
P7	PEAR_100	No	DADA2	Retained	LULU	96	0.49	0.969	0.251
P8	PEAR_100	No	DADA2	Removed	LULU	94	0.48	0.968	0.249
P9	QUALITY_R1	No	DADA2	Retained	LULU	107	0.547	0.966	0.278
P10	QUALITY_R1	No	DADA2	Removed	LULU	105	0.536	0.965	0.275

Pauvert et al (2019) *Fungal Ecology*
doi.org/10.1016/j.funeco.2019.03.005

DADA2 ITS Tutorial:

- https://benjjneb.github.io/dada2/ITS_workflow.html
- Variable length poses a problem:
  - Read through past the opposite primer is possible
  - Need to remove primers in a way that accounts for this
  - Look for reverse complement of forward read in reverse reads
  - Look for reverse complement of reverse read in forward reads



DADA2 ITS Tutorial:

- Data: use the link provided in the Rmd (not the tutorial)

when you get to the Remove Primers step

- We will use FastqCleaner to remove primers (not cutadapt)
  - cutadapt is challenging to install on Windows
  - We will use a modified version of FastqCleaner today to avoid compatibility issues
  - But FastqCleaner is buggy, so use software like cutadapt, BBduk, or trimmomatic for your own data analysis,

DADA2 ITS Tutorial: Taxonomy Assignment

- Not necessary to run this today but if you do:
  - Use the UNITE ITS database
  - Download and expand the file
  - Check your file paths

The taxonomy assignment command is computationally expensive. In case R freezes, SAVE YOUR FINAL ASV TABLE as a .csv file before you attempt this!

Example:

```
write_csv(as.data.frame(seqtab.nochim), file = “~/ASV_table.csv”)
```

Perform LULU Curation

- Save OTU table from DADA2 output
- Save ASV sequences in a fasta file (example code in R markdown)

- Code also provided for
  - Converting ASV sequences to names
  - Pruning existing ps object with LULU results

Coding Exercise – Decontam tutorial

- Paper: Davis et al. 2018 BMC Microbiome
  - https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-018-0605-2
- Tutorial
  - https://benjjneb.github.io/decontam/vignettes/decontam_intro.html

- Identified contaminants based on:
  - Frequency relative to input DNA concentration
  - Prevalence in samples relative to negative controls
  - Combinations of the above