Clustering amplicons

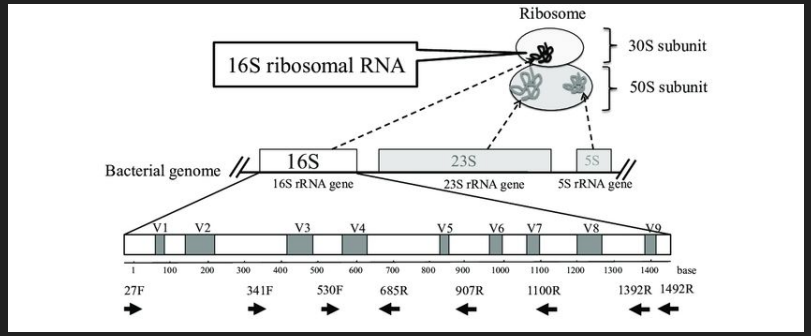
Acer VanWallendael

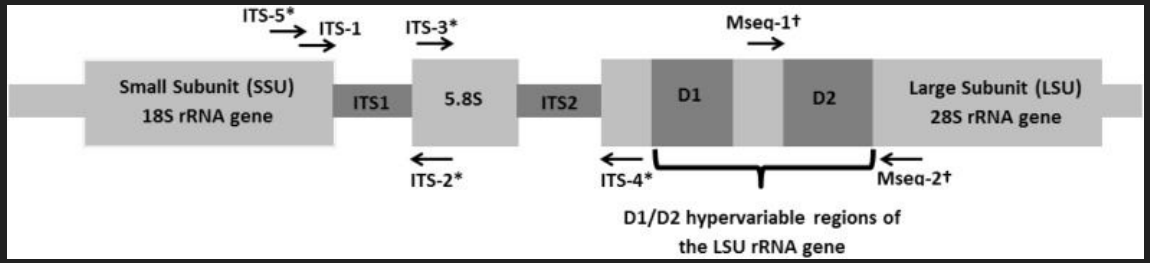
Outline

1. Amplicon sequencing background
2. Bioinformatic workflow
    a. QC
    b. Clustering
    c. Understanding outputs
3. Amplicon clustering tutorial

So you want to do a microbiome.

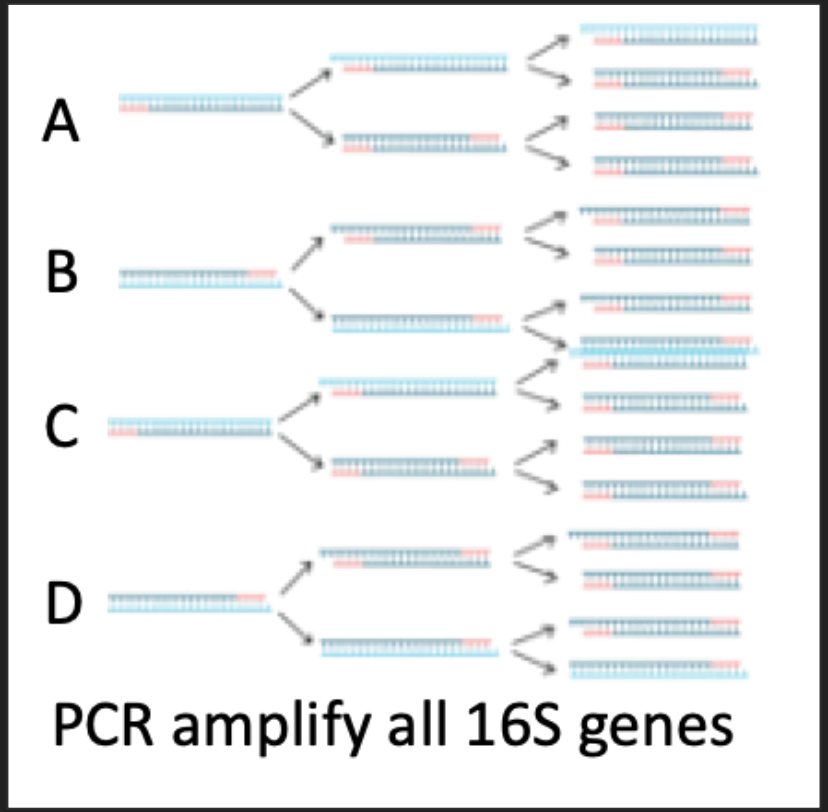
16S (bacteria) and ITS (fungi) regions of the ribosome evolve just slowly enough that primer regions are conserved, but species are different.

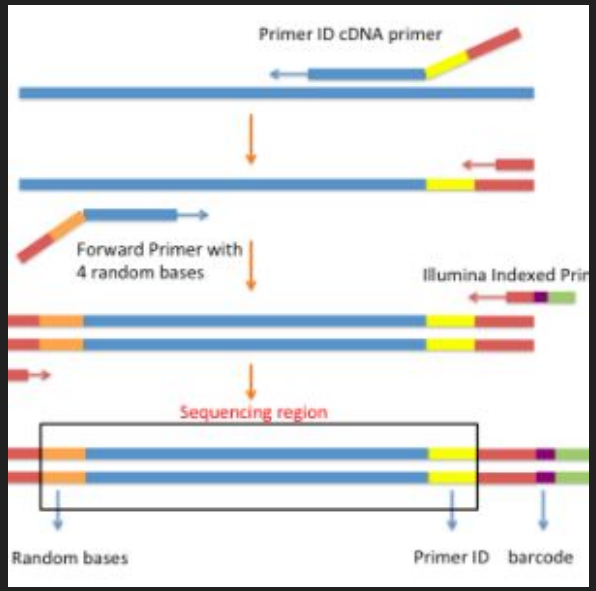




Sequencing ITS or 16S: Library prep

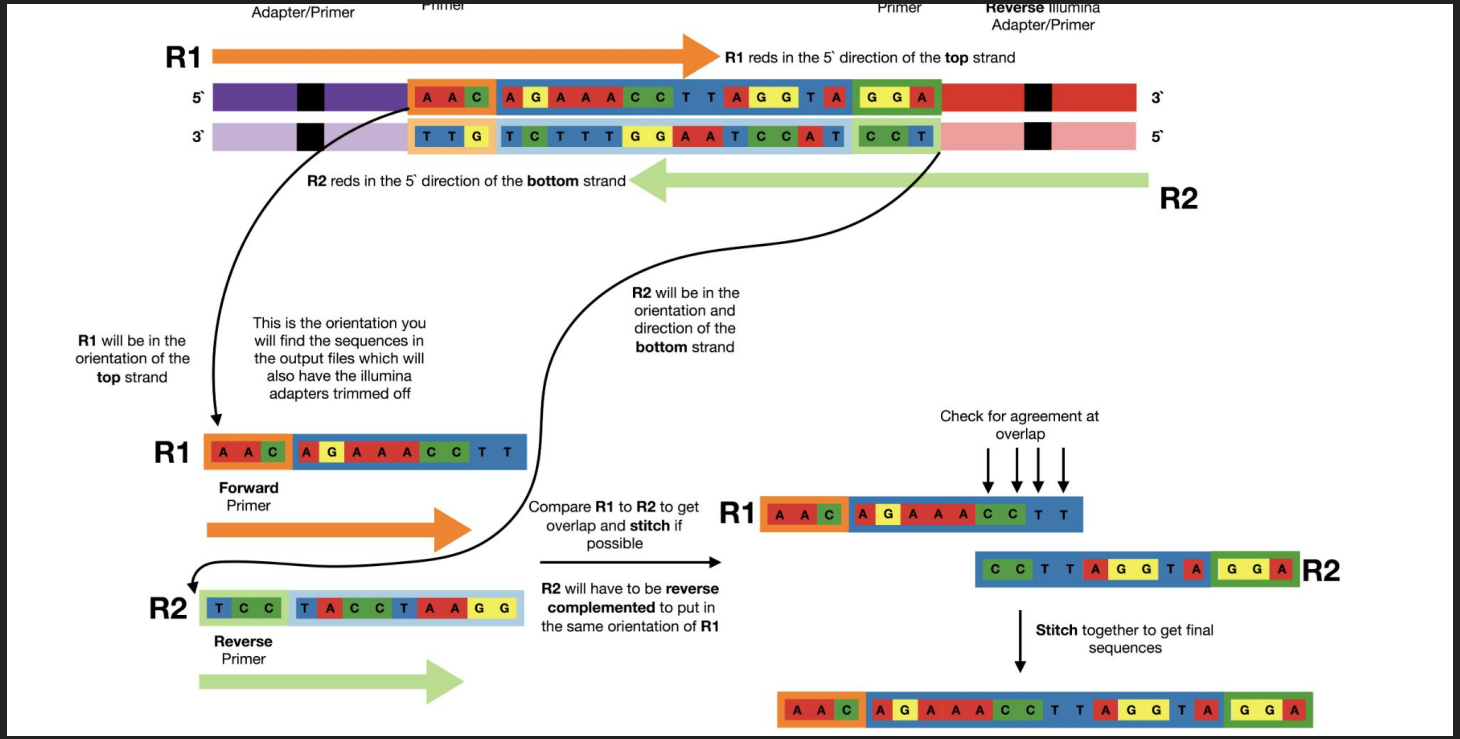
Extract, clean, and normalize DNA from sample





Ligate adapters & barcodes for multiplex sequencing

Illumina sequencing



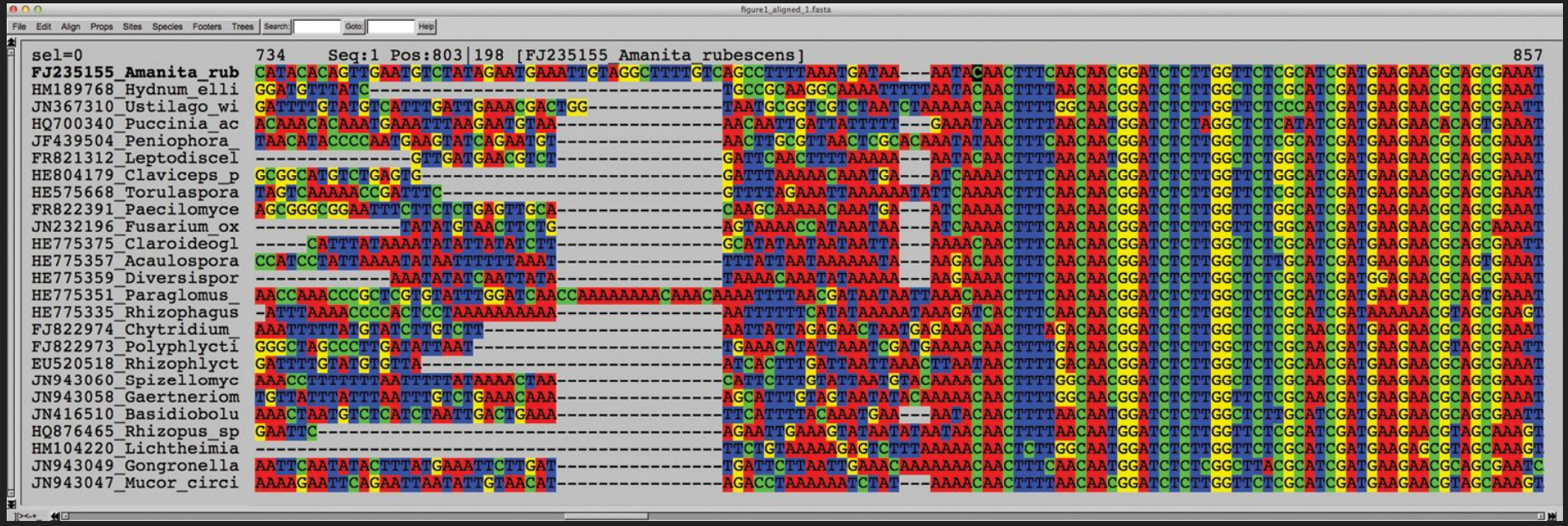
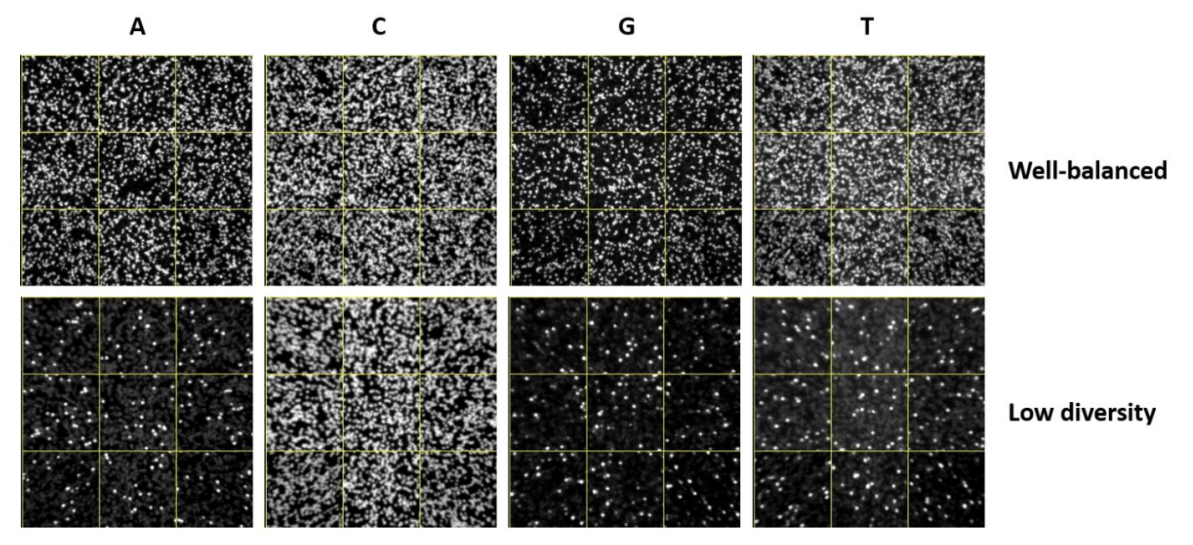


figure1_aligned_1.fasta
File Edit Align Props Sites Species Footers Trees Search: Goto: Help
sel=0 734 Seq:1 Pos:803|198 [FJ235155 Amanita rubescens] 857
FJ235155_Amanita_rub
HM189768_Hydnum_elli
JN367310_Ustilago_wi
HQ700340_Puccinia_ac
JF439504_Peniophora_
FR821312_Leptodiscel
HE804179_Claviceps_p
HE575668_Torulaspora
FR822391_Paecilomyce
JN232196_Fusarium_ox
HE775375_Claroideogl
HE775357_Acaulospora
HE775359_Diversispor
HE775351_Paraglomus_
HE775335_Rhizophagus
FJ822974_Chytridium_
FJ822973_Polyphlycti
EU520518_Rhizophlyct
JN943060_Spizellomyc
JN943058_Gaertneriom
JN416510_Basidiobolu
HQ876465_Rhizopus_sp
HM104220_Lichtheimia
JN943049_Gongronella
JN943047_Mucor_circi

Sequencing controls

Field, Extraction, Library prep controls

PhiX control library = balanced proportions of ACGTs to correct for seq errors from low-diversity samples

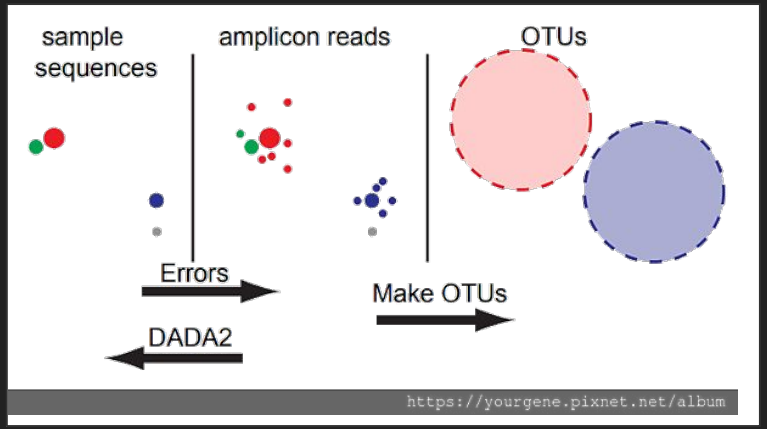


Bioinformatic workflow

1. Quality Check
2. Demultiplexing
3. Amplicon clean and QC
4. Amplicon clustering

Bioinformatic workflow

1. Quality check: were there major sequencing errors?
2. Demultiplexing: Use barcodes to distinguish samples
3. Amplicon clean and QC: Remove poor reads, primers, trim seqs, output stats
4. Amplicon clustering: Output OTUs (operational taxonomic units) or ASVs (Amplicon Sequence Variants)



Simplified Clustering Workflow

1. Remove PhiX
2. Trim primers
3. Strip variable region
4. Filter low quality
5. Identify unique reads
6. Cluster OTUs
7. Build OTU table

Simplified Clustering Workflow

1. Remove PhiX: *bowtie2*
2. Trim primers: *cutadapt*
3. Strip variable region: *usearch*
4. Filter low quality: *usearch*
5. Identify unique reads: *usearch*
6. Cluster OTUs: *usearch*
7. Build OTU table: *usearch*