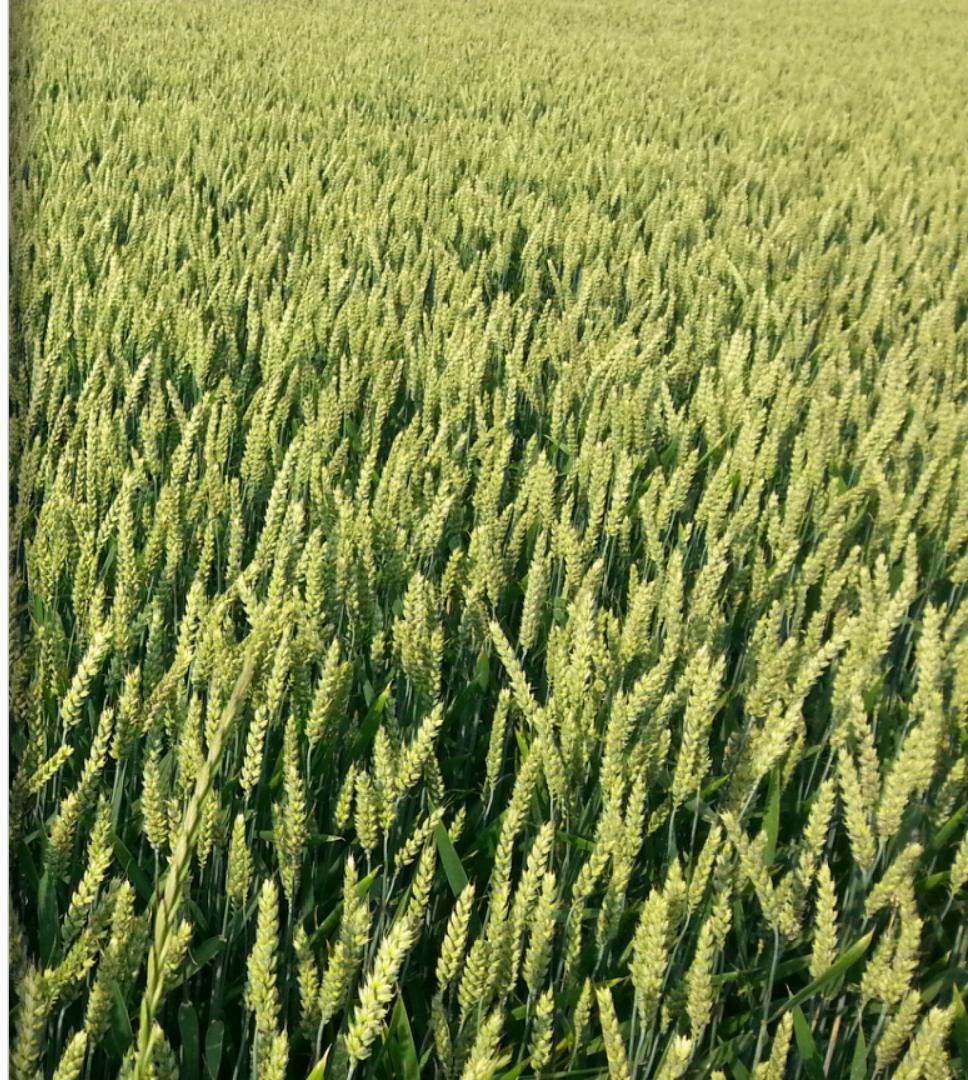




Amplicon sequencing

USDA ARS Microbiome Workshop
August 29, 2017
Adam R. Rivers



Hello!

The Genomics and Bioinformatics Unit

Brian Scheffler – Research leader and CSIO

Stoneville, Mississippi

Brian.Scheffler@ars.usda.gov

Adam Rivers – SY Microbiomes and microbial genomes

Gainesville, Florida

Adam.Rivers@ars.usda.gov

Justin Vaughn – SY Crop genetics

Athens, Georgia

Justin.Vaughn@ars.usda.gov

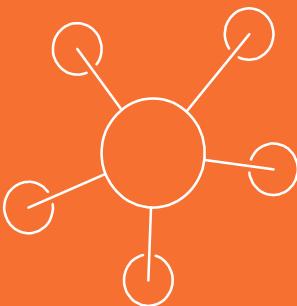
Amanda Hulse-Kemp – SY complex and polyploid Genomes

Raleigh, North Carolina

Amanda.Hulse-Kemp@ars.usda.gov



Learning Objectives



- By the end of the training you should
 - 1. What amplicon sequencing is and the questions it can address
 - 2. Know how amplicon samples are processed
 - 3. Understand a standard bioinformatics workflow
 - 4. Know the types of statistical analyses that are possible with amplicon data

1.

What is
amplicon
sequencing?



Amplicon sequencing

“Amplicon sequencing is the amplification of a particular gene locus from a mixed group of organisms followed by the random sequencing of those targeted amplicons.”

Which gene?

Functional

NifH
Cox

Taxonomic

16S
18S
ITS

Which part?

Full length

Variable
region

Which primer?

Designing
good
“universal”
primers is
hard.

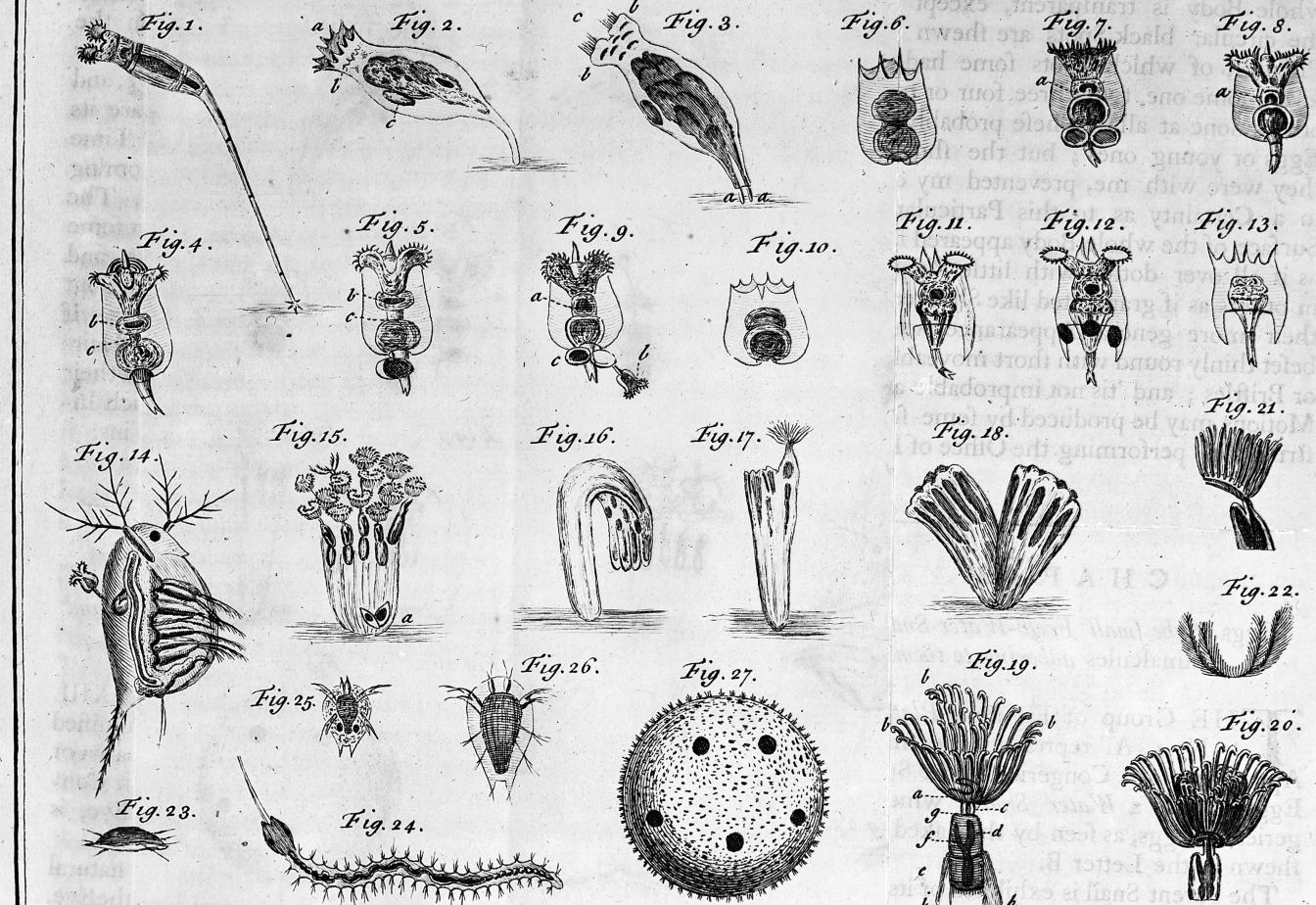
2.

What can
amplicon
sequencing
answer?

*Composition, relative
abundance, dynamics*

Plate XII pa. 324

Animalcules

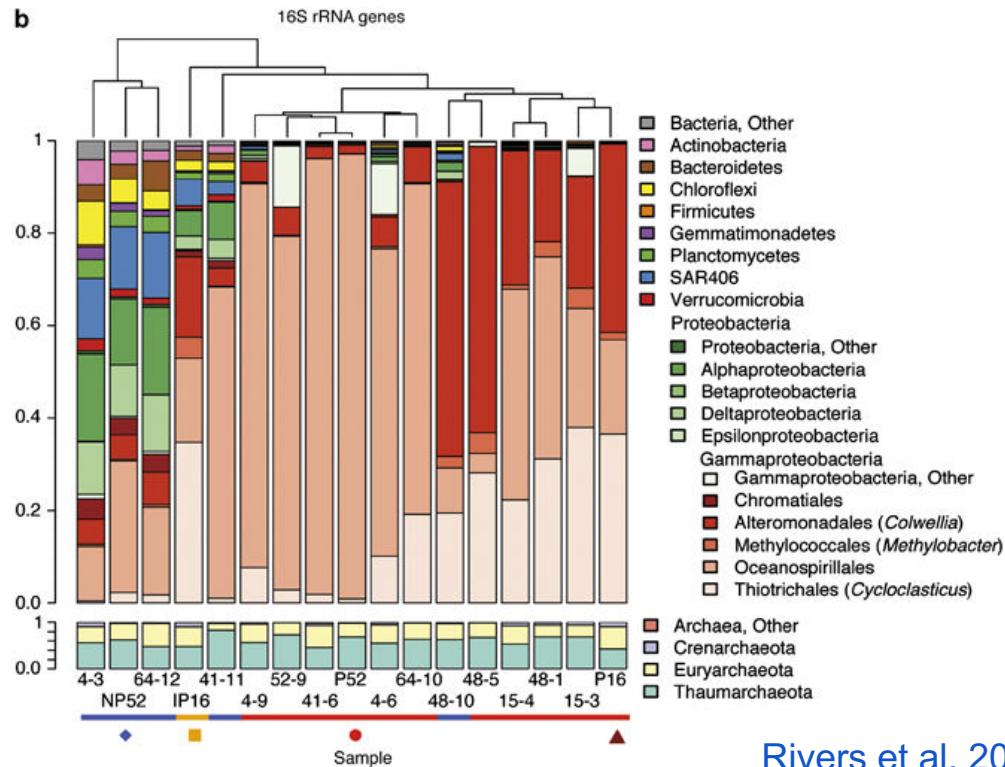


Composition

- Resolution can be a challenge
- Different primers can't be compared
- Linking environmental data is hard

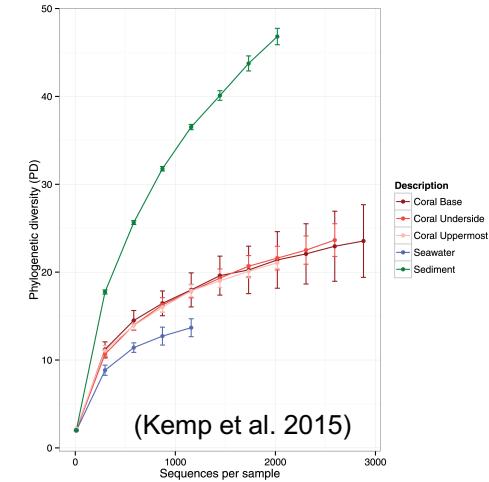
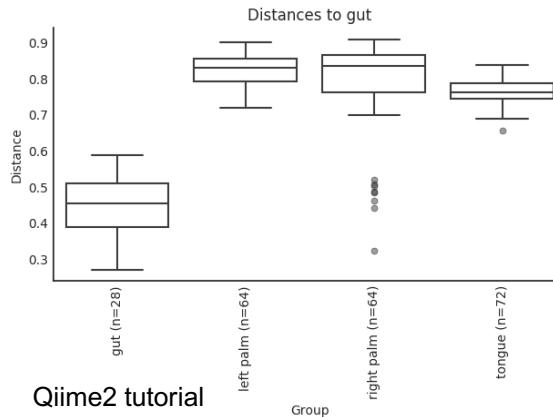
Often 16S data is used to select samples of metagenomics

At its most basic level amplicon sequencing allows for the taxonomic profiling of communities



Diversity

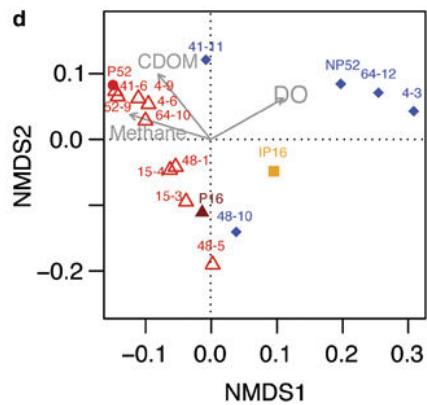
- Alpha, Beta and Gamma diversity ([R. H. Whittaker, 1972](#)) are used to describe biodiversity spatially.
- Diversity – a combination of richness and evenness
- Alpha – within community diversity
- Beta – between community diversity ([Anderson et al. 2011](#))
- Gamma – total diversity across a study area



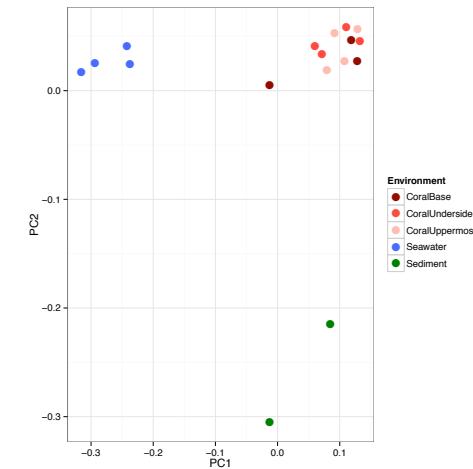
Looking for patterns in communities

Ordination is a dimensionality reduction technique

- Start with Taxa vs sample table
- Calculate a dissimilarity matrix
- Perform ordination, e.g. NMDS, PCA, PCoA,
- Fit environmental variables

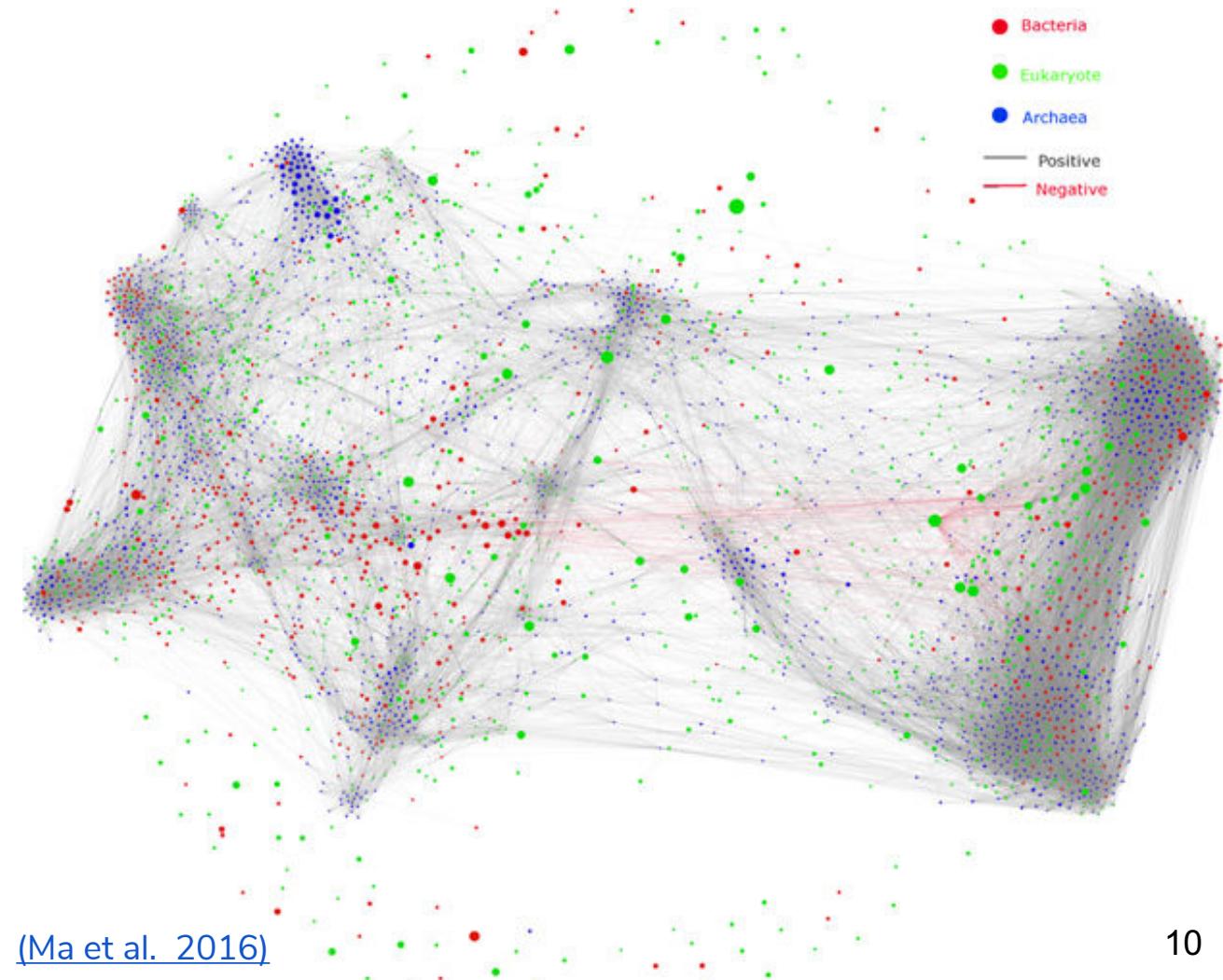


NMDS from [Rivers et al. \(2013\)](#)



PCA from [Kemp et al. \(2015\)](#)

Looking for
co-
occurrence

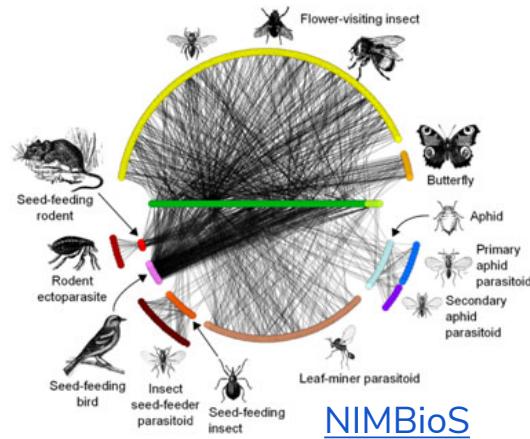


What does co-occurrence mean?

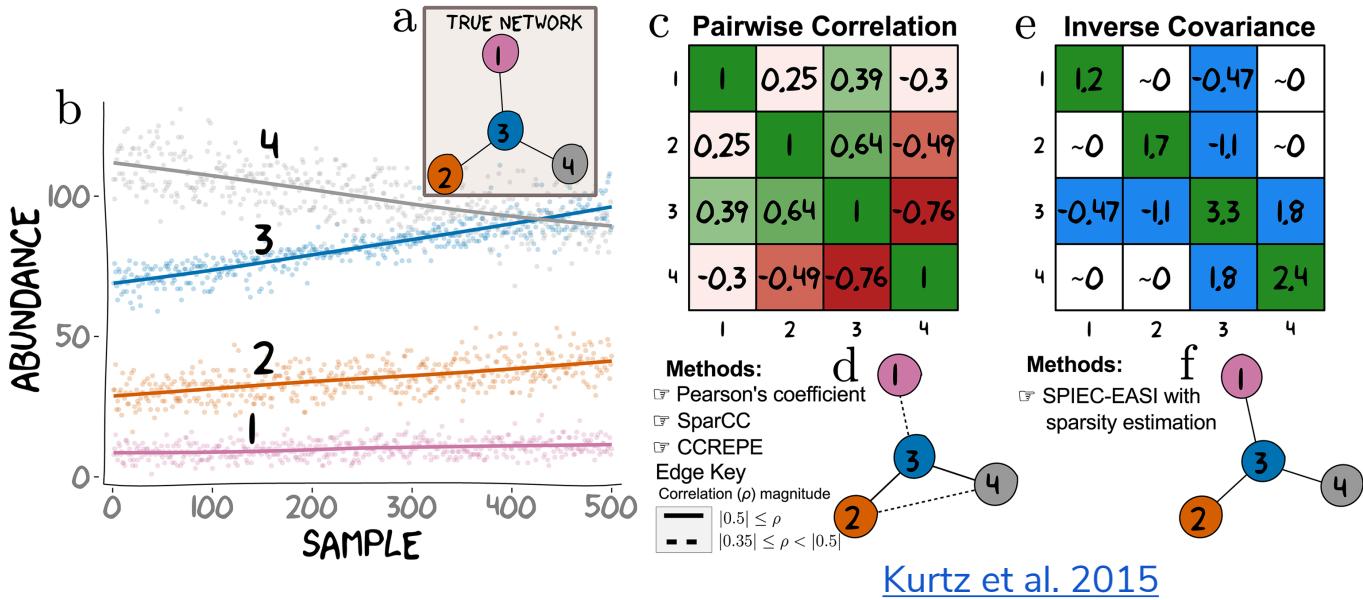
Ideally it maps to an ecological networks

But microbial communities are different...

- Direct vs. indirect interaction
- Simplex measurements
- Vastly different sampling efforts
- Artefactual Co-variance

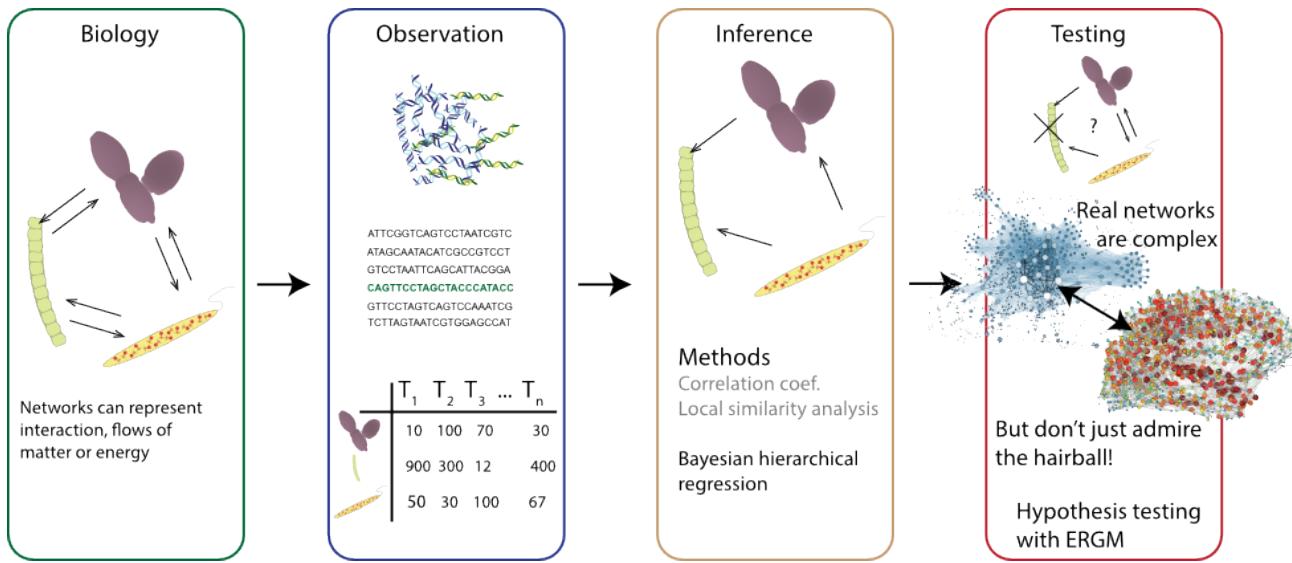


The Covariance problem



Methods to address this:
SPIEC-EASI
Gniess
Bayesian network methods

Ecological networks with predictive capability



Types of Microbiome studies

- | | |
|---|---|
| <p>Observational</p> <p>Time series</p> <p>Spatial</p> <p>Experimental treatments</p> | <ul style="list-style-type: none">• Who's there• Diversity• [C] ~ environmental parameters• Co-varying OTUs and network structure• OTUs with significant relationships to gradient <ul style="list-style-type: none">• Dynamic networks, seasonal succession• Few studies have the resolution to use ARIMA, etc.• Repeated measures <ul style="list-style-type: none">• Spatial scaling latitudinal diversity has been studied• Many GIS tools and methods, Kriging not widely used <ul style="list-style-type: none">• Experimental manipulation is becoming more complex• Pairwise tests are common, some GLM frameworks• Dealing with normalization, the Simplex and the count distribution are active research areas• Internal standards are sometimes used |
|---|---|

Rarefying, normalizing, oh my!

DOI 10.1186/s40168-017-0237-y

Microbiome

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

CrossMark

RESEARCH

Open Access

Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible

Paul J. McMurdie, Susan Holmes*

Statistics Department, Stanford University, Stanford, California, United States of America

Normalization and microbial differential abundance strategies depend upon data characteristics

Sophie Weiss¹, Zhenjiang Zech Xu², Shyamal Peddada³, Amnon Amir², Kyle Bittinger⁴, Antonio Gonzalez², Catherine Lozupone⁵, Jesse R. Zaneveld⁶, Yoshiki Vázquez-Baeza⁷, Amanda Birmingham⁸, Embrette R. Hyde² and Rob Knight^{2,7,9*}

Rarefaction subsamples without replacement reads from each sample often to the of the smallest library in an experiment sometimes 10-100x.

Alternatively data can be scaled using a normalization method like the Trimmed Mean of Means transformation and modeled using a negative binomial model.

3.

Field to sequencer – the nuts and bolts.

*Sample preparation
and sequencing*



Collect DNA

Amplicon DNA sampling is much more forgiving than RNA sampling.



4°C Short-term
-20°C Long-term
[\(Rubin et al. 2013\)](#)



Collect and store
within 1-2 hours

Consider collecting RNA and storing it in RNAlater
(Saturated Ammonium Sulfate) at 4°C to sequence the
active fraction.

Collect Metadata

“There is no such thing as metadata,
everything is data.”
- Susan Holmes

- Sample collection is the time to record environmental data
- The GSC has created environmental and sequence data standards, [MIxS](#). Use them as a guide for your collections.
- Store environmental data in NCBI or ENA Biosamples databases or Gold database. Do it now, while you still remember what you did. You can link sequence data later.

DNA extraction

Amplicon sequencing is more forgiving than metagenome sequencing.



Target amount of DNA for sequencing:

50-100ng at 3-50 ng/ul in 10-50 ul

About 10ng is needed, but who wants to have just the bare minimum of Flair-DNA?

Quality:

Length is less important than amplification.

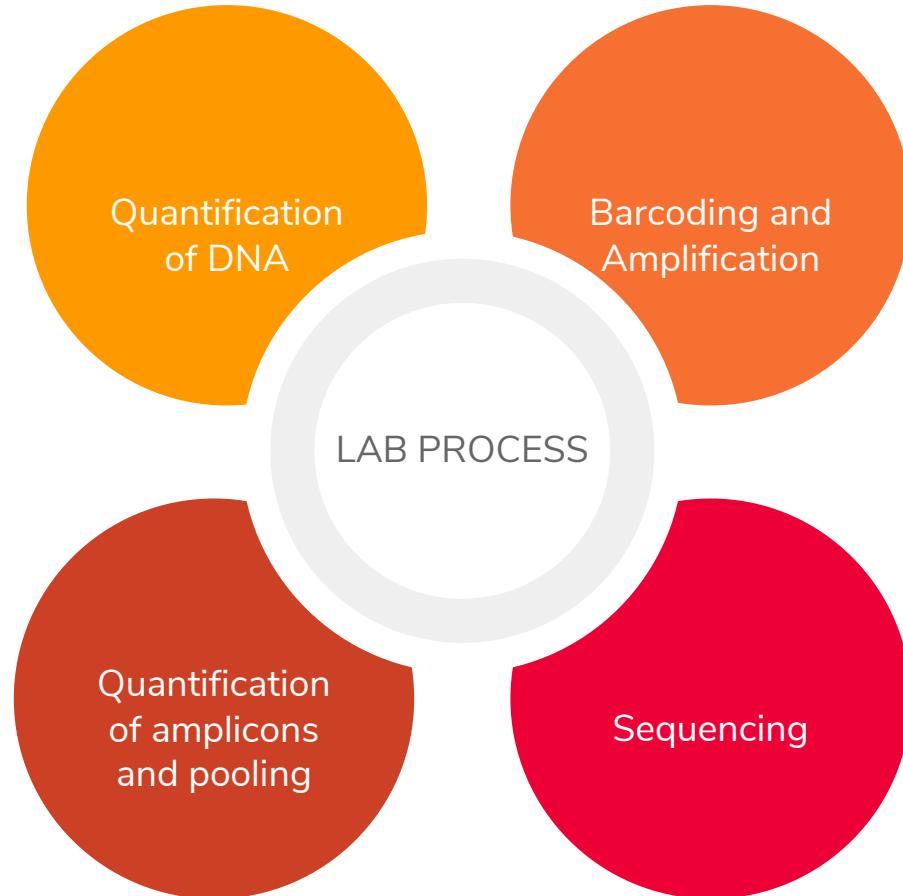
PCR test with universal part of sequencing primers.

PCR inhibitors like humic acids can be most disruptive.

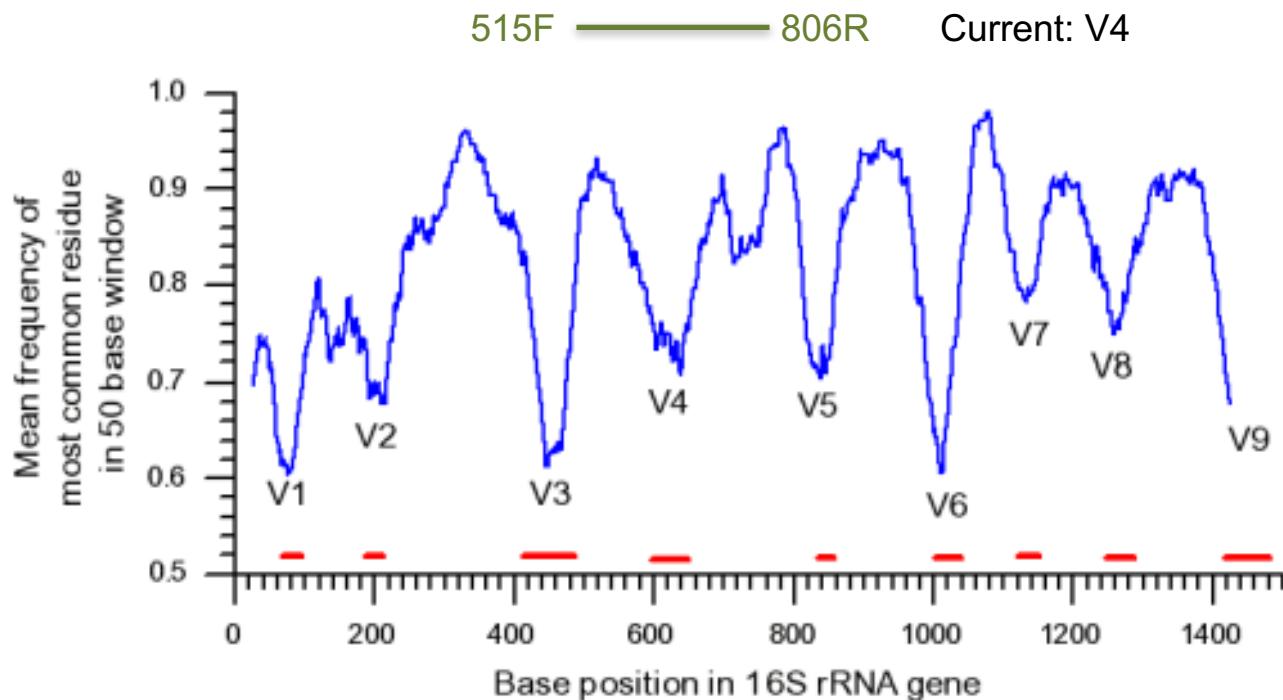
Internal standards:

For quantitative work control DNA is sometimes added during extraction ([Moran et al. 2013](#)).

Amplicon Processing

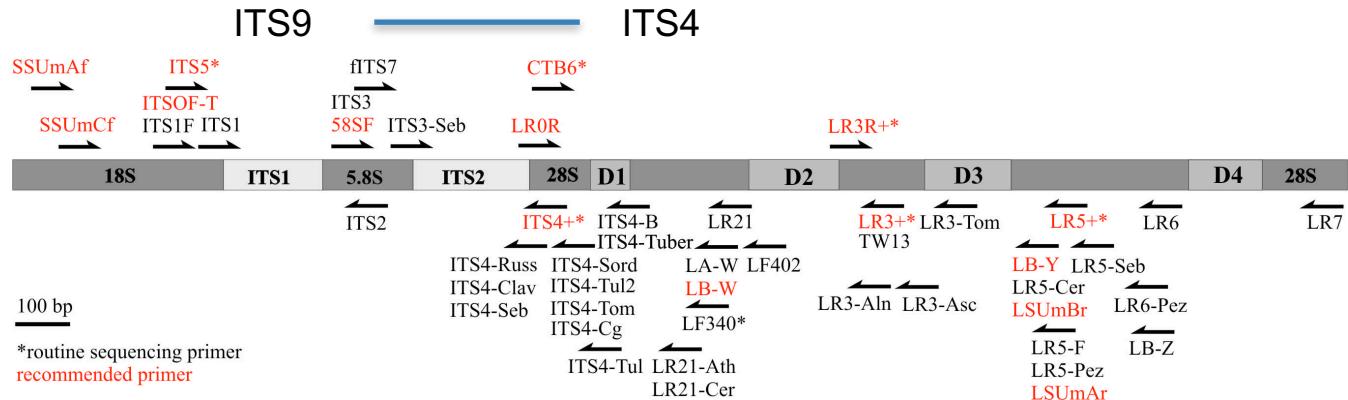


Sequencing primers



Sequencing primers

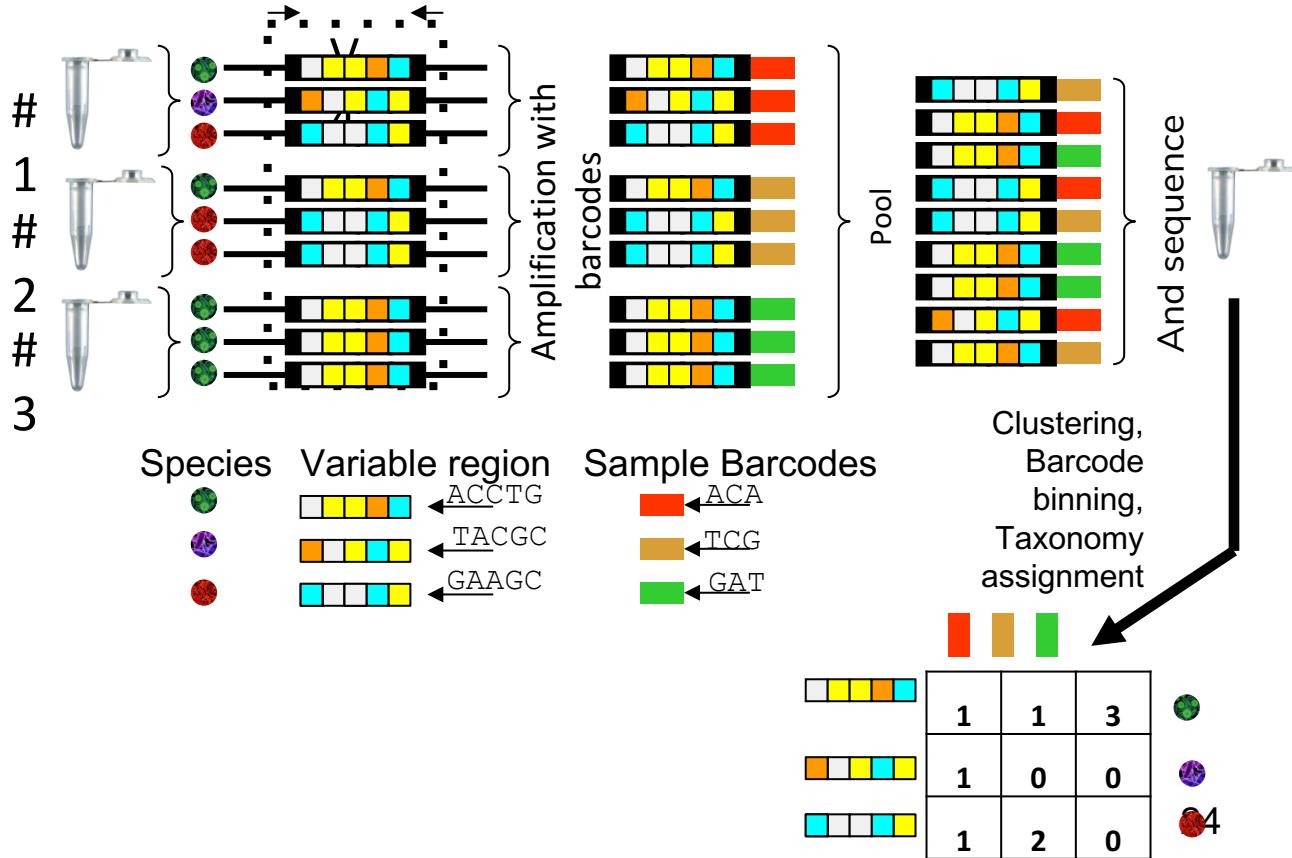
Sequenced Fungal ITS2 Region



Sequencing primers

16S		
V4		
515FB GTGYCAGCMGCCGCGGTAA	Caporaso et al. 2012 Updated EMP	
806RB GGACTACNVGGGTWTCTAAT	Caporaso et al. 2012 Updated EMP	
18S		
V4		
565F CCAGCACSCYGCCTTAATTCC	Stoeck et al. 2010	
948R ACTTTCGTTCTTGATYRA	Stoeck et al. 2010	
18S		
ITS9F GAACGCAGCRAAIIGYGA	5.8S	ITS2
ITS4R TCCTCCGCTTATTGATATGC	Menkis et al. 2012 White et al. 1990	
28S		

Barcoding



Sequencing



2 96 well plates



PCR with 16 forward primer
24 distinct reverse indexes

- 2x300bp Paired Reads
- 44-50M reads
- ~360,000 tags per sample
- 36 hours

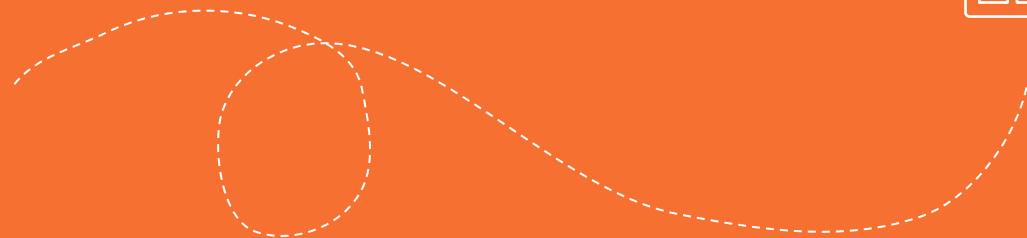
illumina®
MiSeq

4.

Analysis of amplicon data

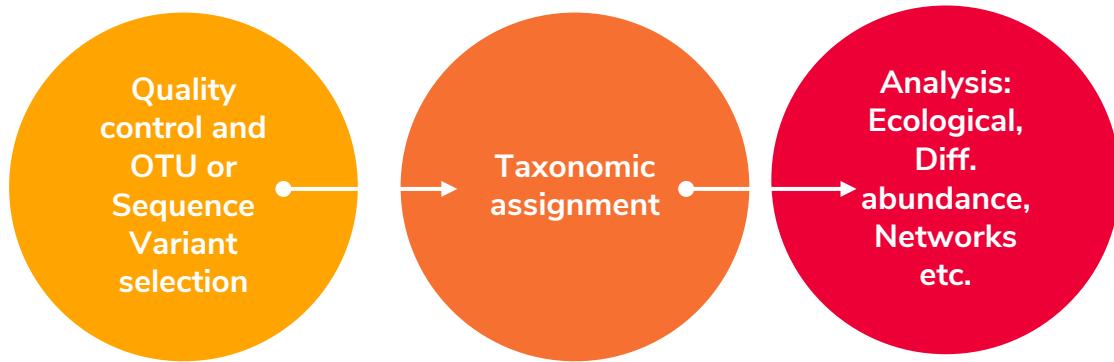
*What to do with all
those fastq.gz files.*

GATC
CATC
CCGA

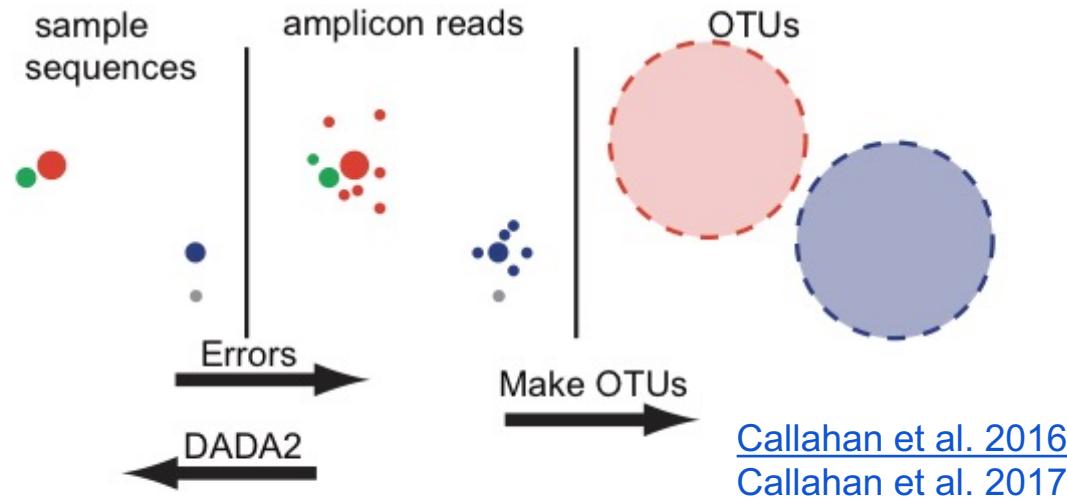


Amplicon sequencing

Amplicon analysis follows a basic workflow with many possibilities for custom analysis



Why OTU's were used and why Sequence variants are replacing them.



The field is moving this way:

Dada2 - [Callahan et al. 2016](#) (Holmes Lab)

Denoise - [Amir et al. 2017](#) (Knight Lab)

Unoise2 – [Edgar 2016](#)

Taxonomic assignment

Databases

Database	Description	License
Greengenes	A curated database of archaea and bacteria - static since 2013	CC BY-SA 3.0
Silva	The most up-to-date and extensive database of prokaryotes and eukaryotes, several versions	Free academic / Paid commercial license
The RDP database	A large collection of archaeal bacterial and fungal sequences	CC BY-SA 3.0
UNITE	The primary database for fungal ITS and 28S data	Not stated

Classifiers

RDP [Classifier](#) – The go-to NB classifier for most people

[Sintax](#) – Edgar's short Kmer classifier

Qiime2's - NB classifier based on Scikit learn

Analysis

The range of analysis preformed after 16S is wide:

- Taxonomic profiling
- Differential abundance analysis
- Diversity measurement
- Network analysis
- Hypothesis testing
- Identifying responsive SV's
- Correlating taxa with environmental conditions
- Understanding how related taxa are.

Is best to jump in and try these
yourself in the [Amplicon Tutorial](#)



Photo credits

- 1: DNA - Made By [MadeByOliver](#) from Flaticon.com
- 1: Wheat - Photo by [Kai Pilger](#) on [Unsplash](#)
- 4: glacier Photo by [Patrícia Cassol Pereira](#) on [Unsplash](#)