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 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?

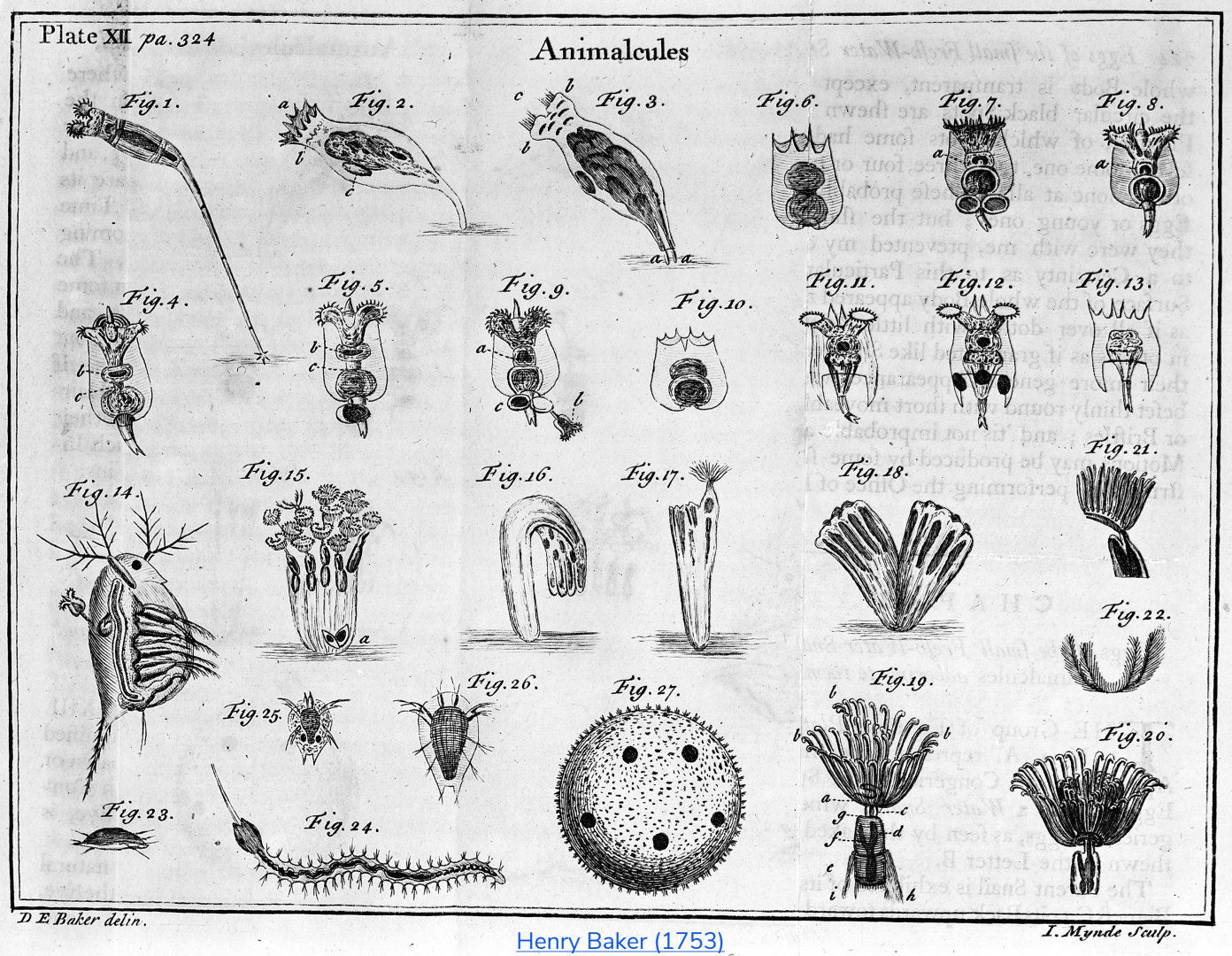
Amplicon sequencing

"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."

2.

What can amplicon sequencing answer?

Composition, relative abundance, dynamics



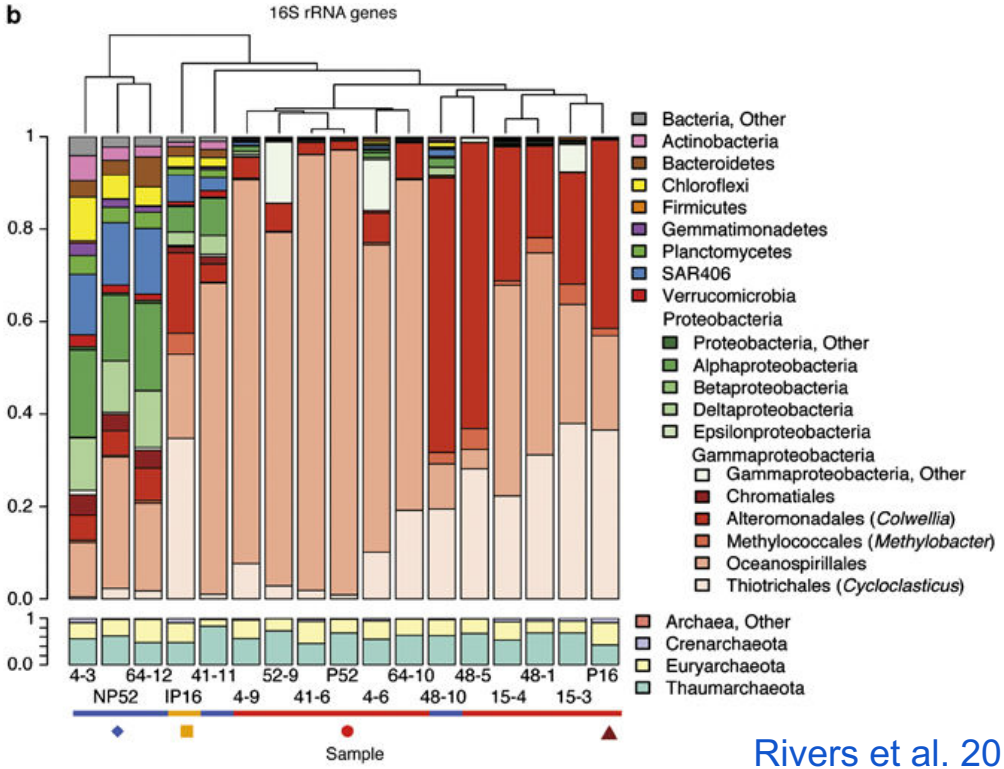
Henry Baker (1753)

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



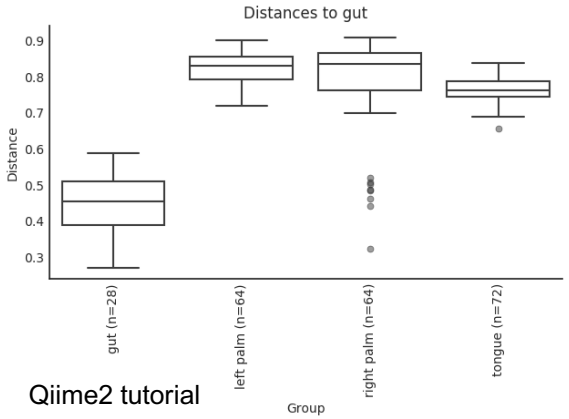
Rivers et al. 2013

Diversity

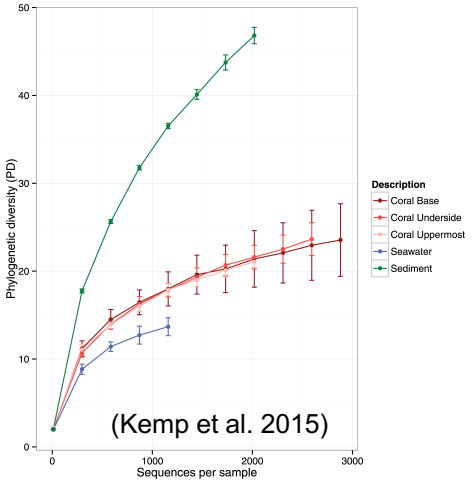
- Alpha, Beta and Gamma diversity (R. H. Whittaker, 1972) are used to describe biodiversity spatially.

- 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



Qiime2 tutorial

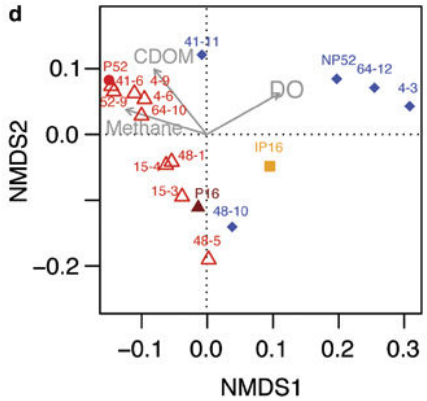


(Kemp et al. 2015)

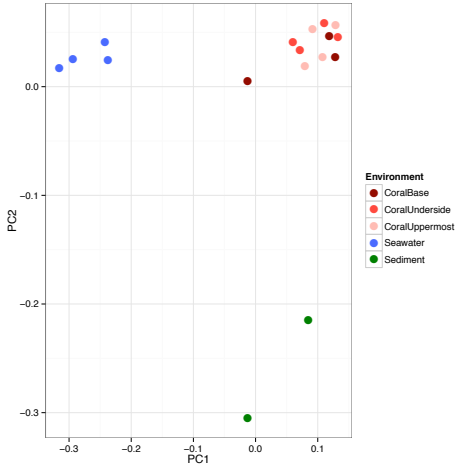
Looking for patterns in communities

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

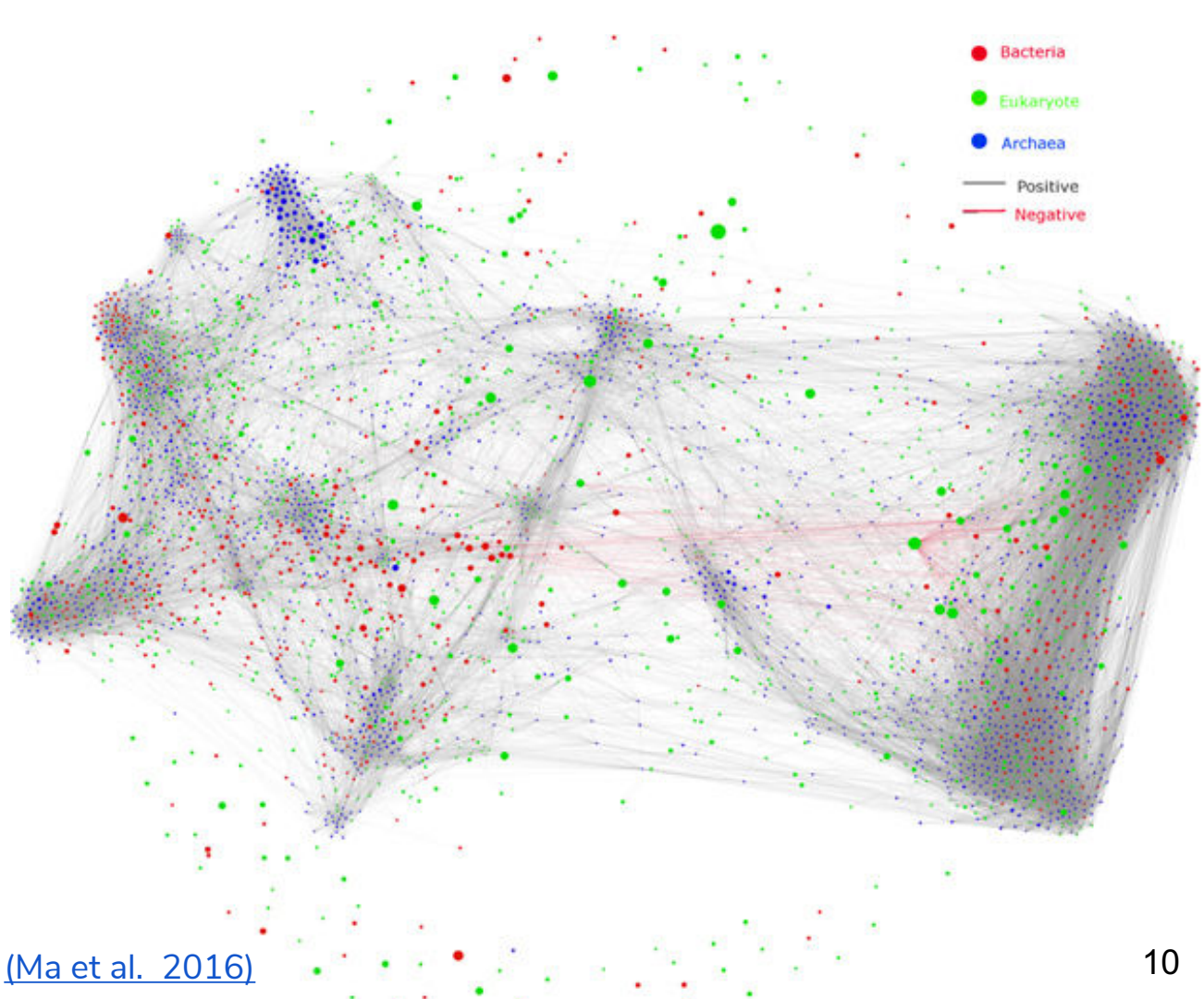


NMDS from [Rivers et al. (2013)]



PCA from [Kemp et al. (2015)]

Looking for co-occurrence



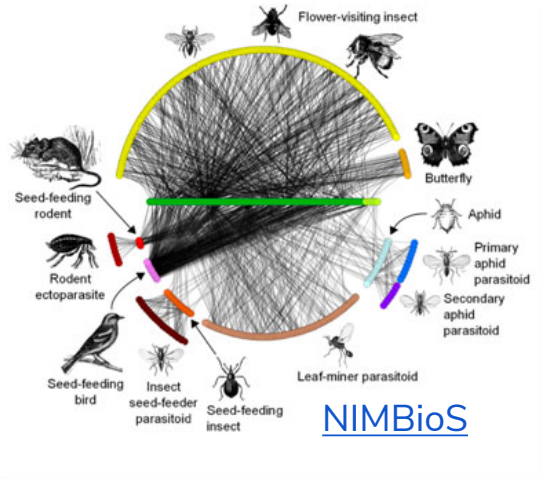
(Ma et al. 2016)

What does co-occurrence mean?

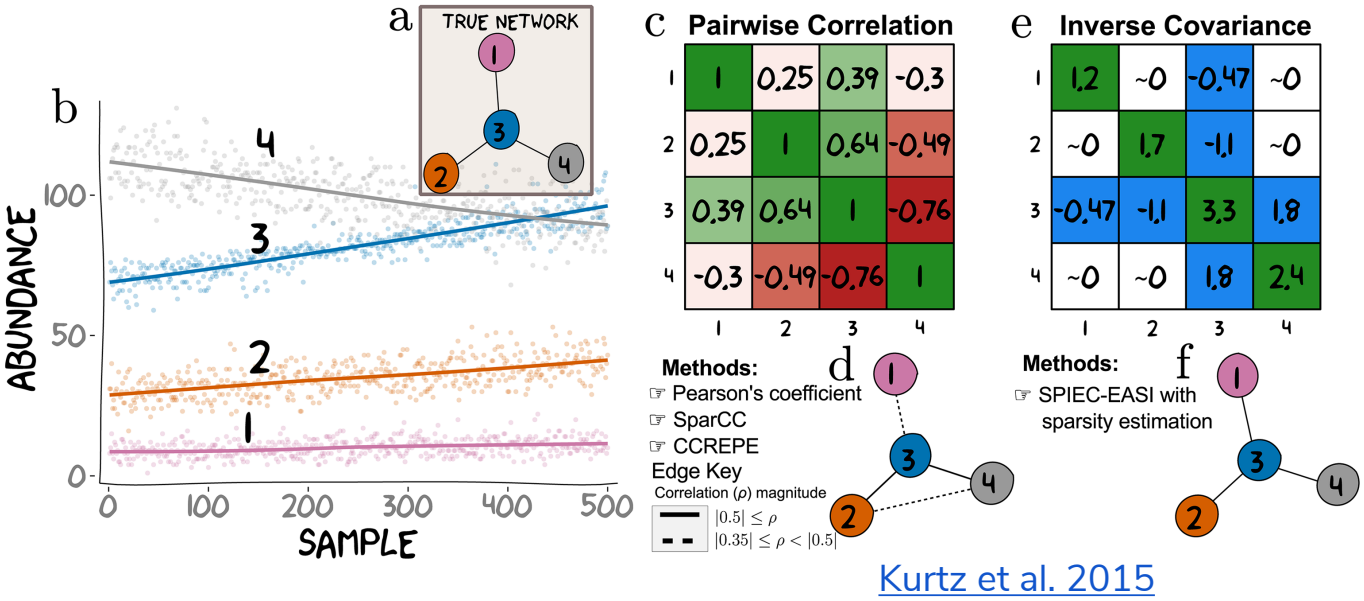
Ideally it maps to an ecological networks

But microbial communities are different...

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



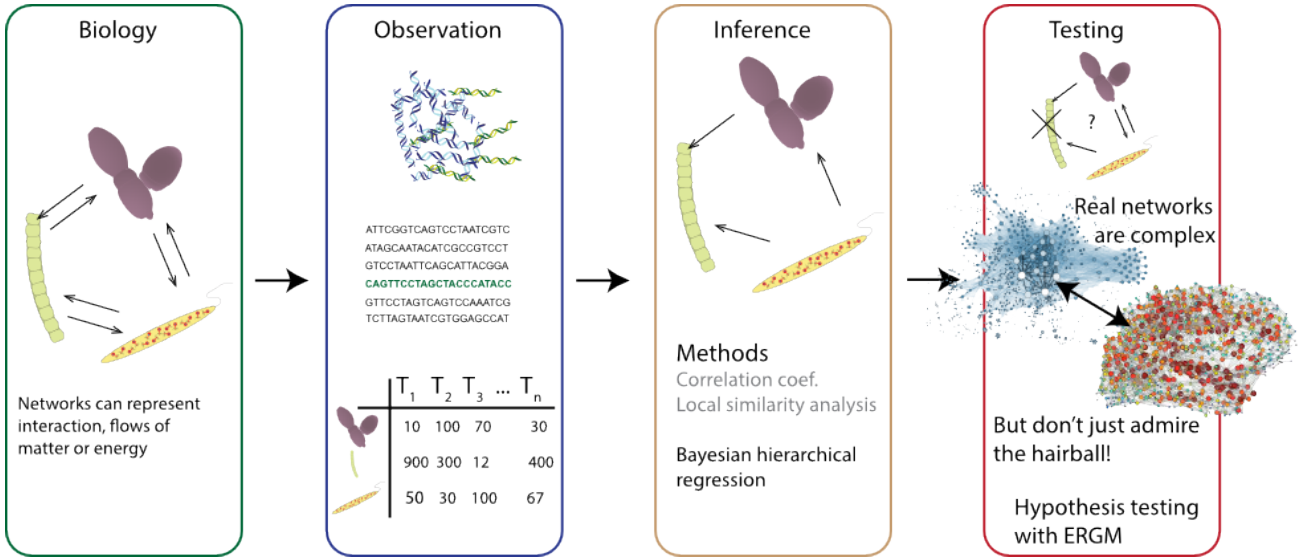
The Covariance problem



Kurtz et al. 2015

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

Ecological networks with predictive capability



Types of Microbiome studies

Observational	• Who's there • Diversity • [C] ~ environmental parameters • Co-varying OTUs and network structure • OTUs with significant relationships to gradient
Time series	• Dynamic networks, seasonal succession • Few studies have the resolution to use ARIMA, etc. • Repeated measures
Spatial	• Spatial scaling latitudinal diversity has been studied • Many GIS tools and methods, Kriging not widely used
Experimental treatments	• 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!


OPEN ACCESS Freely available online

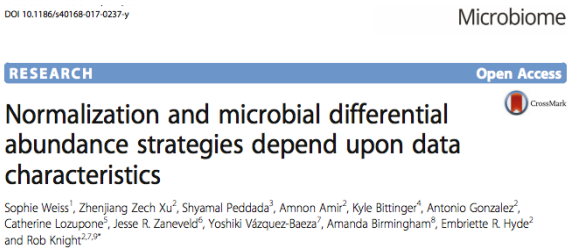
PLOS COMPUTATIONAL BIOLOGY

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

vs.


DOI 10.1186/s40168-017-0237-y

Microbiome

RESEARCH

Open Access

Normalization and microbial differential abundance strategies depend upon data characteristics

Sophie Weiss[1], Zhenjiang Zech Xu[2], Shyamal Peddada[3], Amnon Amir[2], Kyle Bittinger[4], Antonio Gonzalez[2], Catherine Lozupone[5], Jesse R. Zaneveld[6], Yoshiki Vázquez-Baeza[7], Amanda Birmingham[8], Embriette R. Hyde[2] 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.

4ºC Short-term
-20ºC 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 4ºC 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

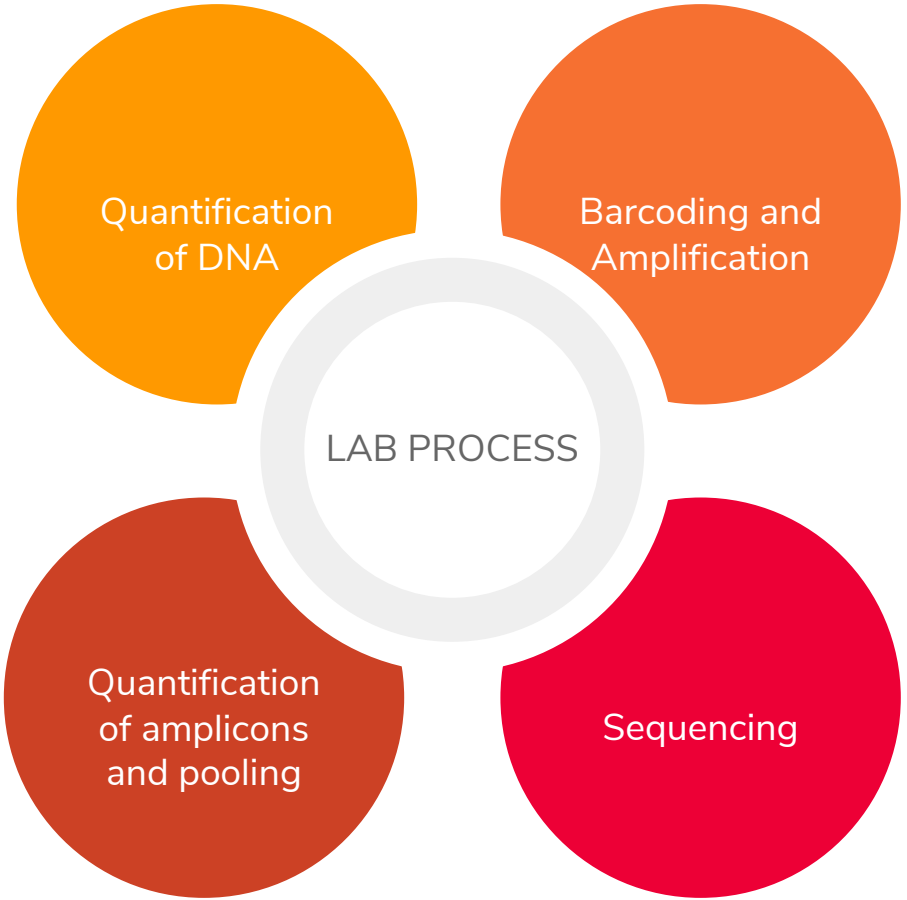
Amplicon sequencing is more forgiving than metagenome sequencing.

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?

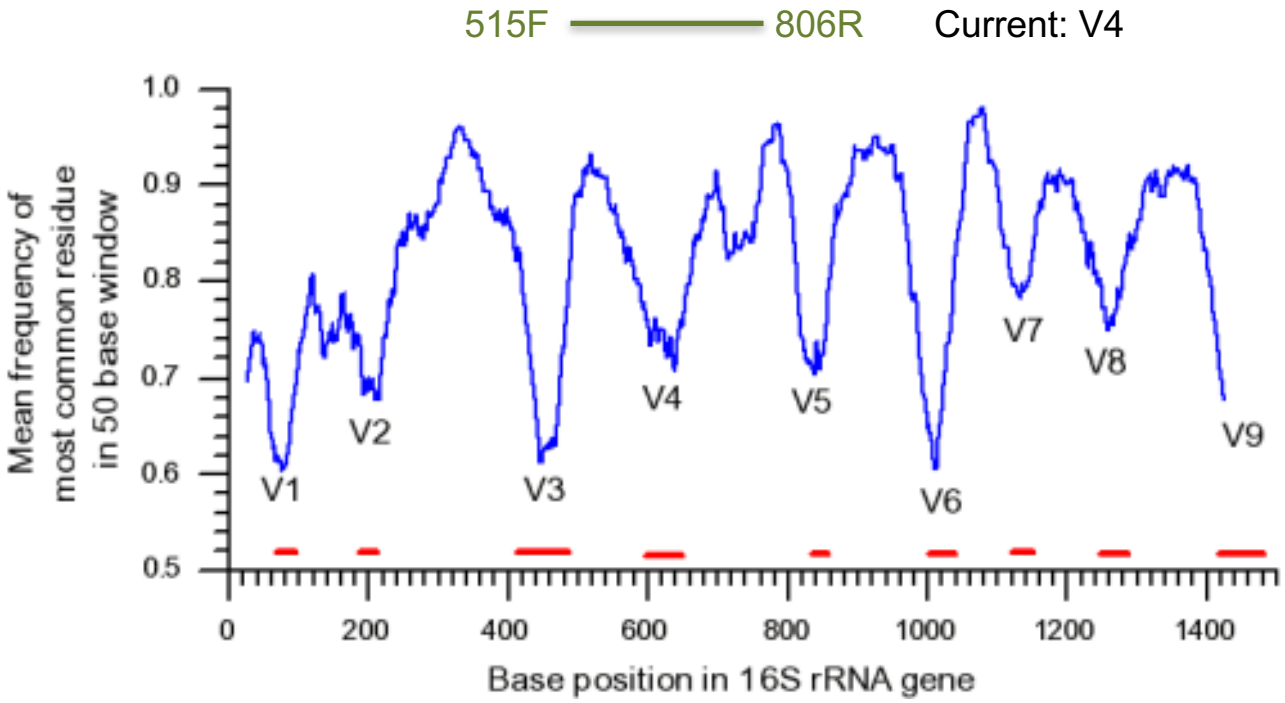
Quality:
Length is less important than amplification.
PCR test with universal part of sequencing primers.
PCR inhibitors like humic acids can be most disruptive.

Internal standards:
For quantitative work control DNA is sometimes added during extraction (Moran et al. 2013).

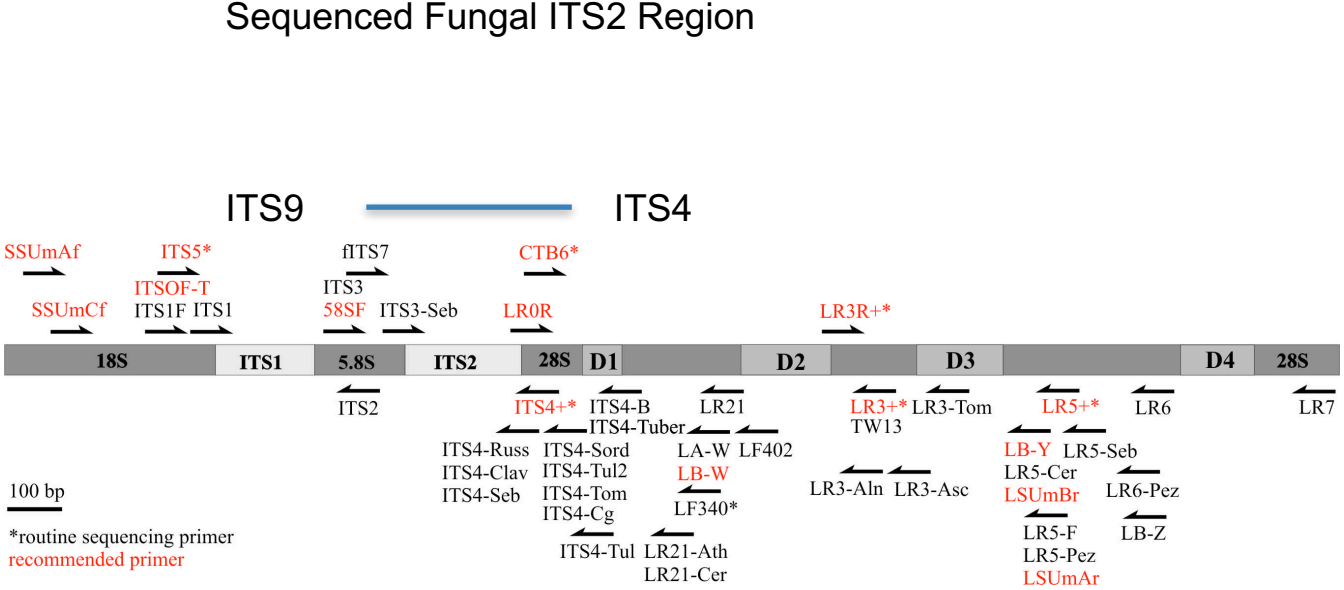
Amplicon Processing



Sequencing primers



Sequencing primers



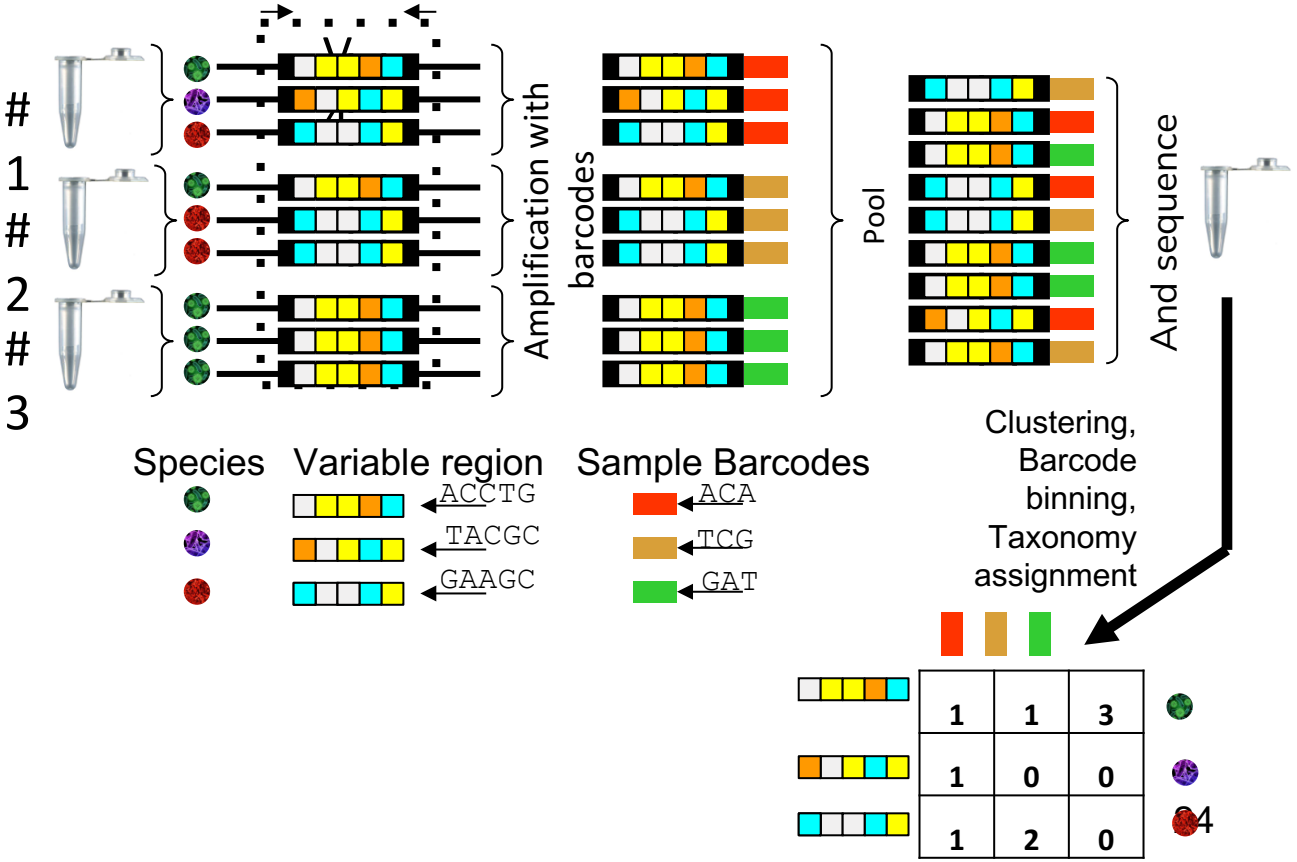
Sequencing primers

515FB GTGYCAGCMGCCGCGGTAA Caporaso et al. 2012 Updated EMP
806RB GGACTACNVGGGTWTCTAAT Caporaso et al. 2012 Updated EMP

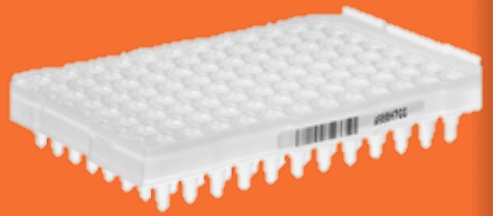
565F CCAGCASCYGCGGTAATTCC Stoeck et al. 2010
948R ACTTTCGTTCTTGATYRA Stoeck et al. 2010

ITS9F GAACGCAGCRAAIIGYGA Menkis et al. 2012
ITS4R TCCTCCGCTTATTGATATGC White et al. 1990

Barcoding


#
1
#
2
#
3
Amplification with barcodes
Pool
And sequence
Species
Variable region
Sample Barcodes
ACCTG
TACGC
GAAGC
ACA
TCG
GAT
Clustering, Barcode binning, Taxonomy assignment
1 1 3
1 0 0
1 2 0
4

Sequencing



2 96 well plates

PCR with 16 forward primer
24 distinct reverse indexes

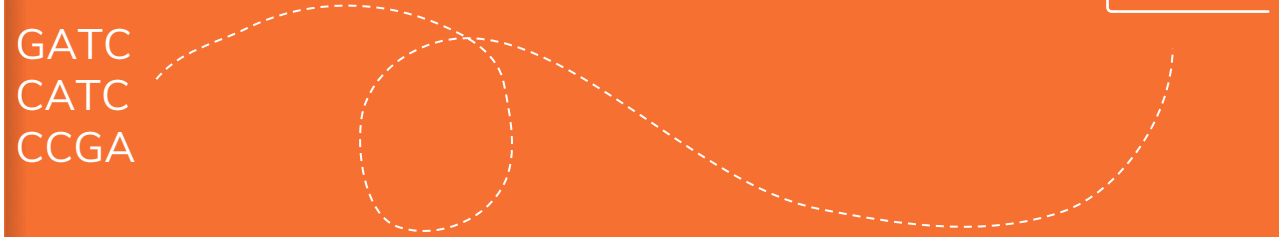
illumina
MiSeq

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

4.

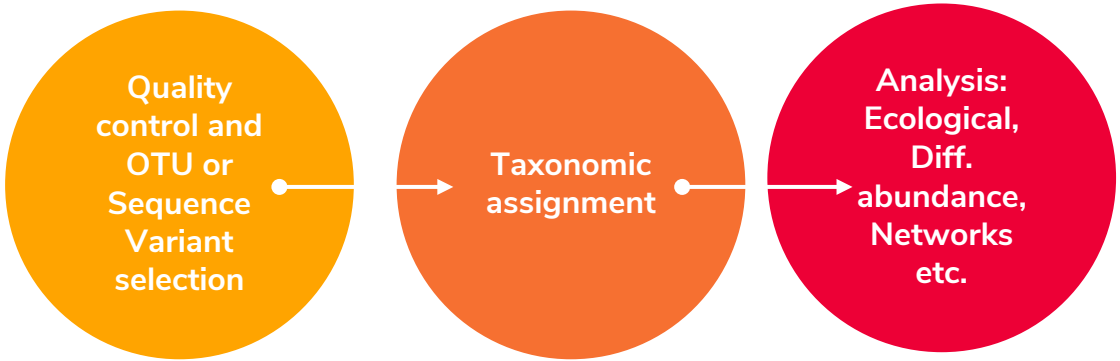
Analysis of amplicon data

What to do with all those fastq.gz files.

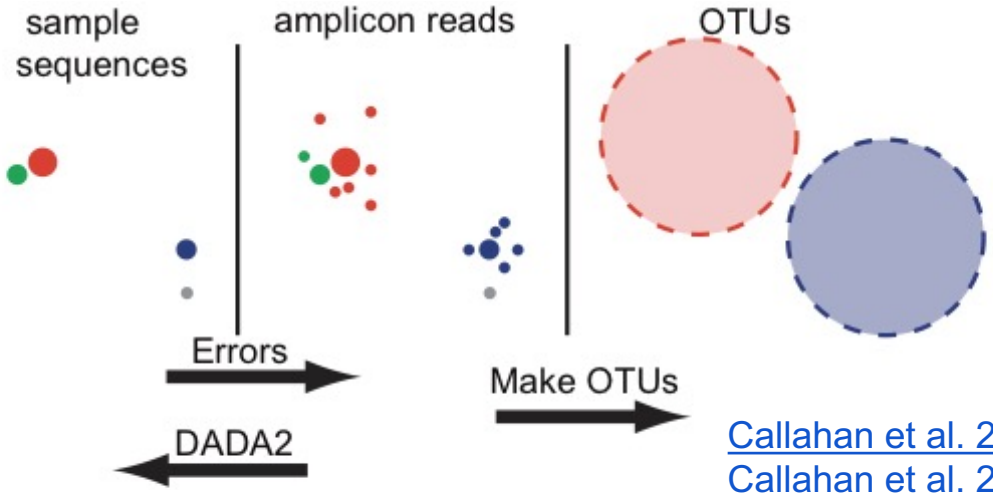


Amplicon sequencing

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



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



Callahan et al. 2016
Callahan et al. 2017

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