ZMT LEIBNIZ CENTRE for Tropical Marine Research

Leibniz Association

Amplicon sequencing studies

Lecture 12.03.2019

Outline

- Considerations before you sequence
- Reference databases
- Marker gene and primer selection
- Primer design and *in silico* PCR
    - Silva TestPrime
    - NCBI Primer blast
    - ecoPrimers and ecoPCR
- Sequencing errors (incl. chimeras, rare sequences)
- Mock communities
- OTU clustering vs. denoising approaches
- Taxonomic classification
- Resources for amplicon sequence analysis

Before you sequence…

What is your target gene?
- E.g. 16S

How good is the reference database of your target gene?

How much information is contained in the amplified fragment?
- Target variable regions

How specific and how universal is your primer set?
- How well does it cover the diversity of your target group without amplifiying non-target taxa?

How long is the amplified fragment?
- 2x300bp PE Illumina sequencing (without primers): < 500bp

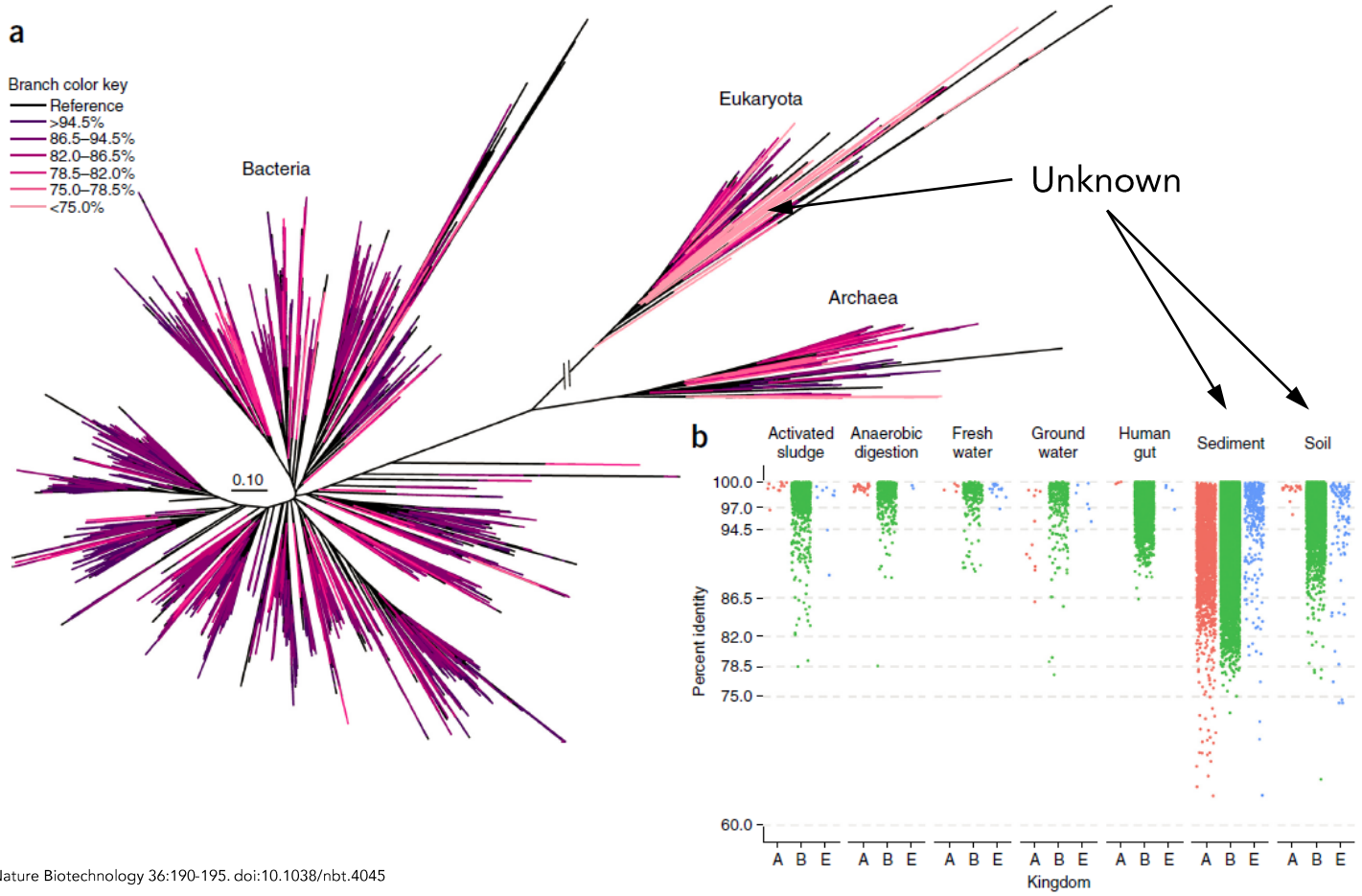
What is the required sequencing depth?
- Depending on the diversity of your sample, e.g. water column vs. sediment
- Depending on your research question, e.g. dominant taxa vs. rare biosphere

Should I sequence all my samples?
- Reduce the number of conditions, never the number of replicates to save sequencing costs

Do I need technical replicates?
- Depending on your budget and study design

Reference databases

We can only study what we know



Karst et al. (2018) Nature Biotechnology 36:190-195. doi:10.1038/nbt.4045

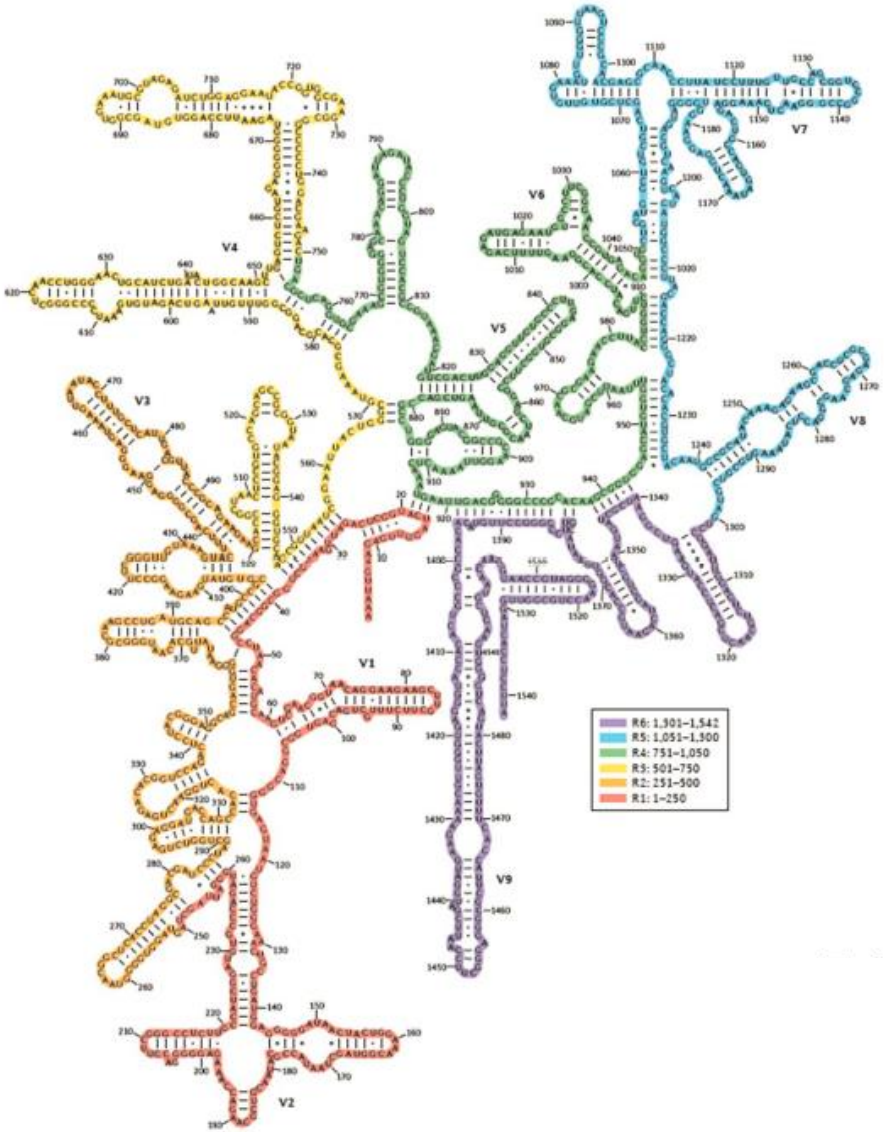
Reference databases

We can only study what we know

- Customized databases:

	Target	Link
SILVA	All domains of life, small and large subunits of ribosomal RNA gene	https://www.arb-silva.de/
RDP	Ribosomal database project, archaea (16S), bacteria (16S), fungi (28S)	https://rdp.cme.msu.edu/
UNITE	Fungi (eukaryotes), internal transcribed spacer 1	https://unite.ut.ee/
PR2	Protist Ribosomal Reference database (18S)	https://github.com/pr2database/pr2database
ITSone	Eukaryotes, internal transcribed spacer 1	http://itsonedb.cloud.ba.infn.it/
ITS2	Eukaryotes, internal transcribed spacer 2	http://its2.bioapps.biozentrum.uni-wuerzburg.de/
Fungene repository	Various functional genes	http://fungene.cme.msu.edu/

Marker gene and primer selection

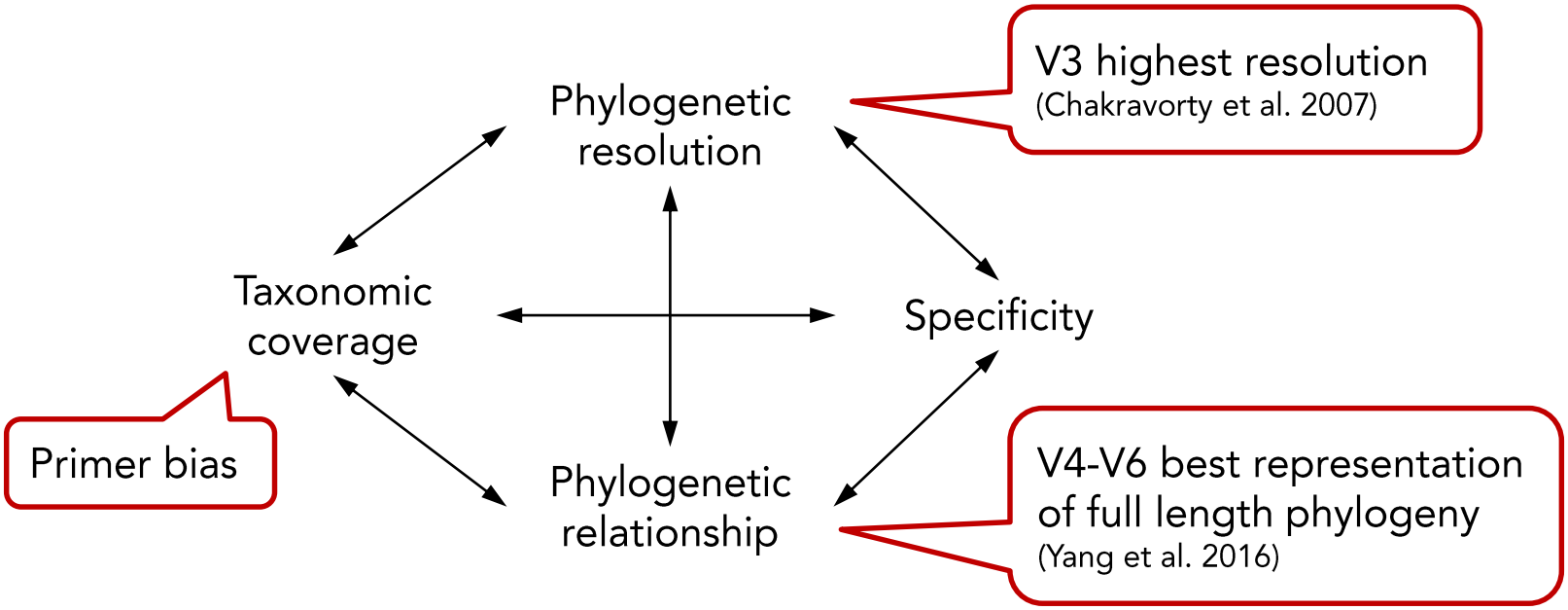


- Small-subunit ribosomal DNA
  - Universal
  - Conserved and hypervariable regions
  - Mutation rate close to species divergence

Yarza et al. (2014) Nature Reviews Microbiology 12:635-645. doi.org/10.1038/nrmicro3330

Marker gene and primer selection

Theoretical concerns:



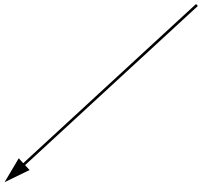
Methodological constrains:

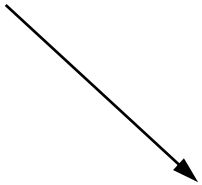
- Fragment length
- PCR conditions

Marker gene and primer selection

Option 1: Use previously published and evaluated primers

Option 2: Design your own primers





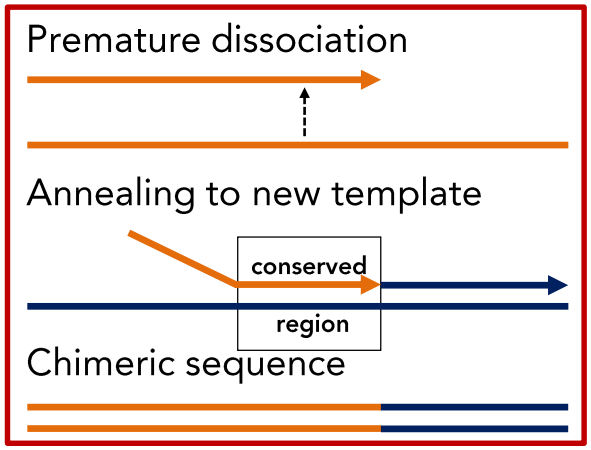
NCBI Primer blast
(Achim Meyer)

ecoPrimers/ecoPCR
(Véronique Helfer)

Sequencing errors

Errors during PCR-library preparation

Chimeras



- In downstream processing estimated based on sequence abundance and alignment quality

Incorporation of wrong base

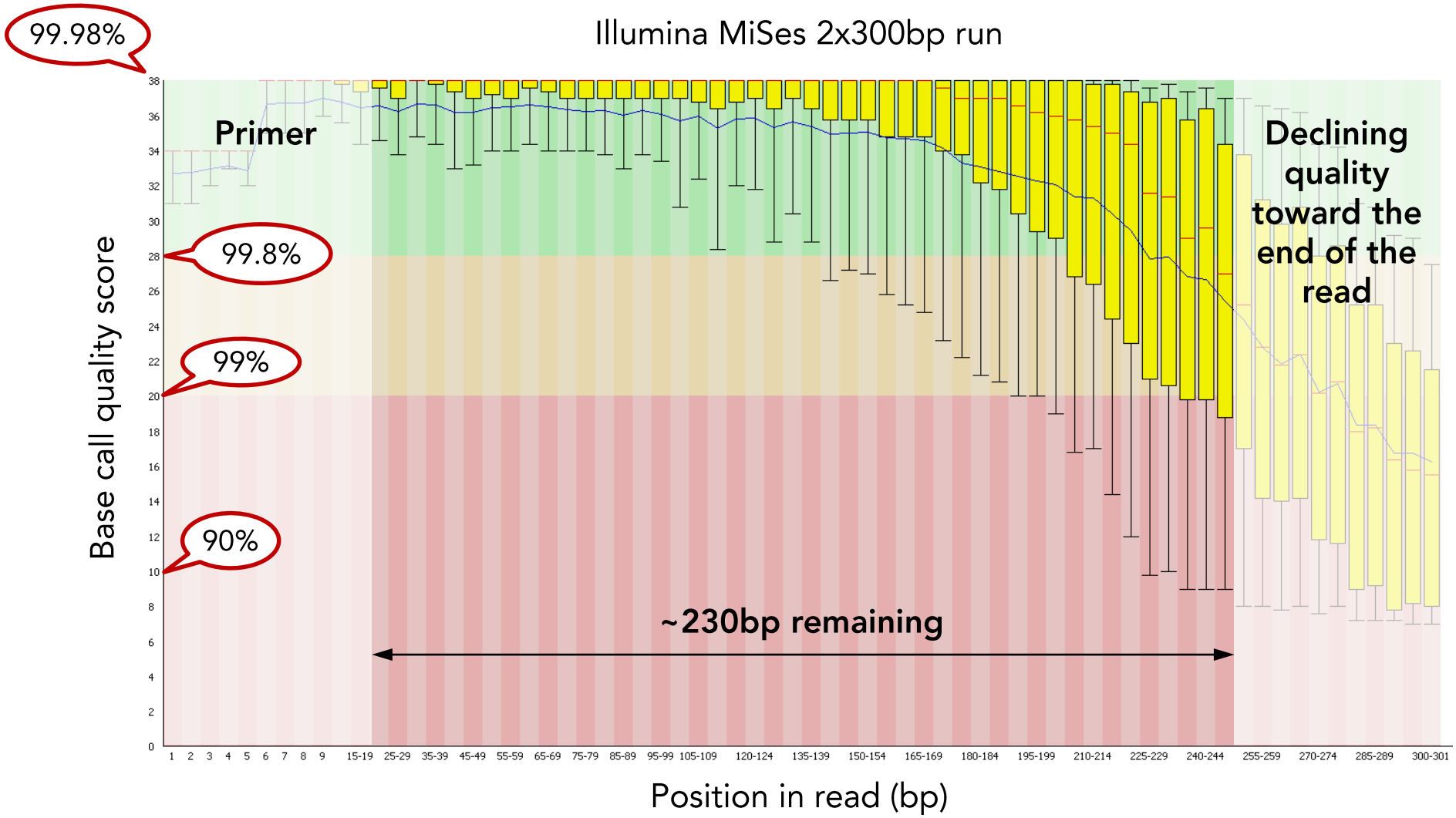
Errors during sequencing

Wrong base call

- Single nucleotide polymorphisms (SNPs)

- Removal based on base quality scores

Sequencing errors

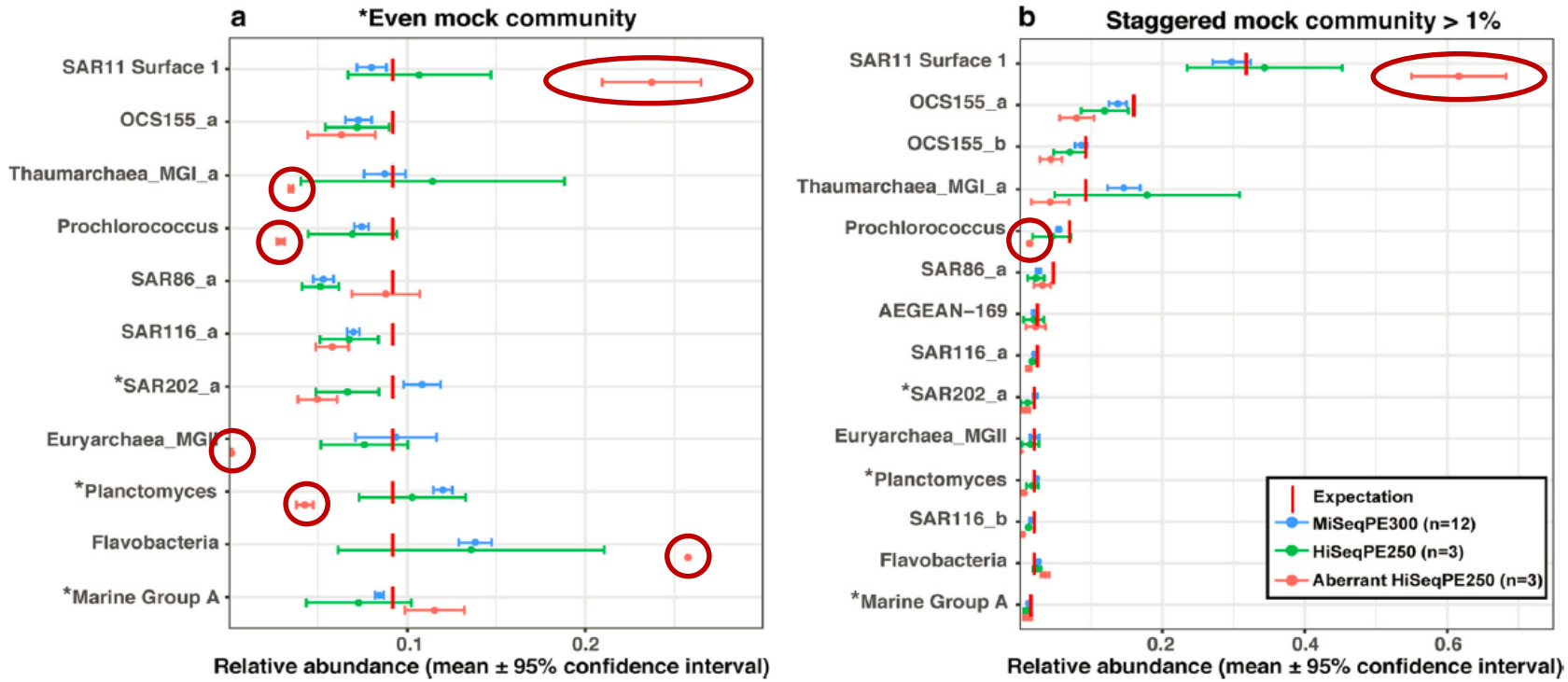


Mock communities

- Artificial sample of known composition and diversity
- Ideally consisting of several taxa that are also expected in the 'real' samples
- Sequenced alongside 'real' samples

→ Assess sequencing error and reliability of bioinformatic sequence analysis during method development

→ Include as routine 'standard' in every sequencing run?



Yeh et al. (2018) mSystems 3:e00023-18. https://doi.org/10.1128/mSystems.00023-18.

Operational taxonomic units (OTUs)…

…are defined as sequences of sufficient similarity that are distinct from other sequences

…are dependent on the amplified region and analysis method

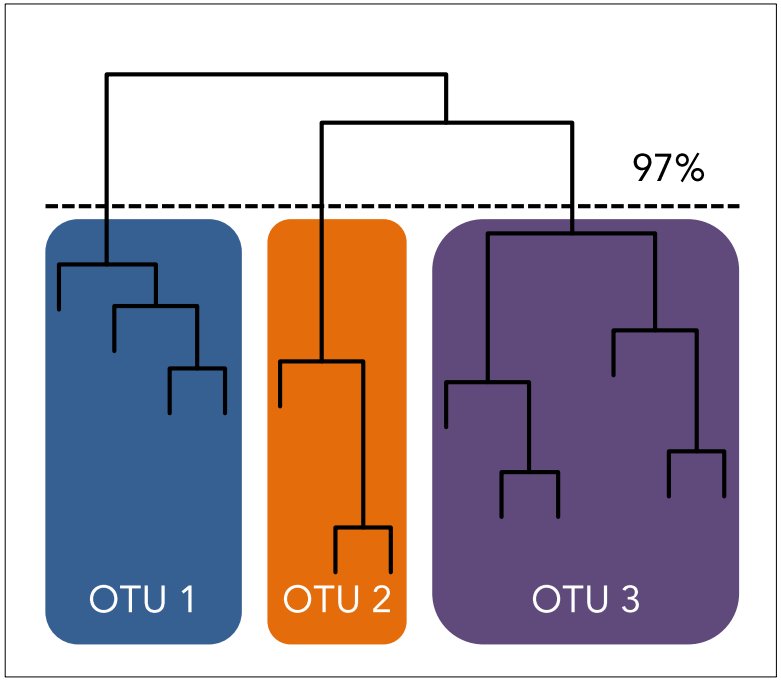
…are NOT comparable across studies (exception: generation via denoising)

…do NOT represent species

…do NOT represent genome divergence

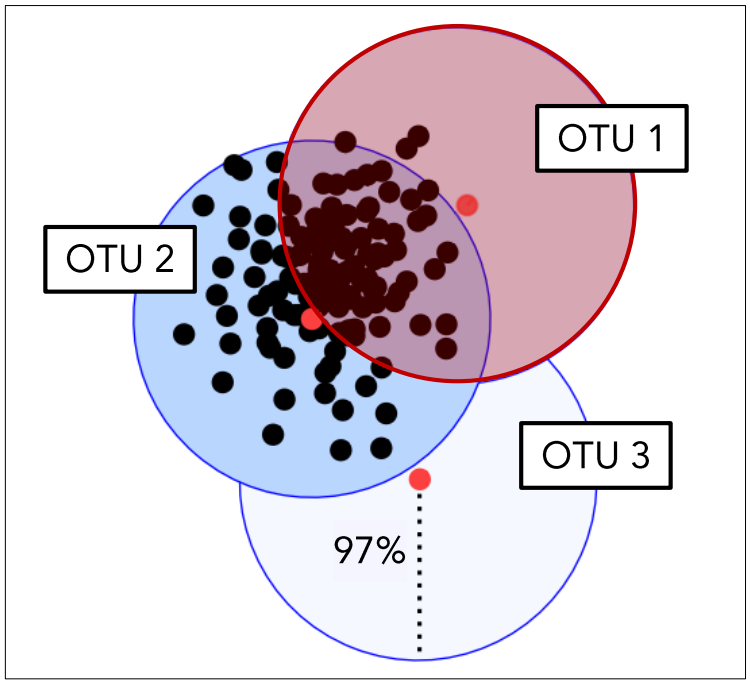
OTU clustering

Hierarchical clustering



- Better defined OTUs than heuristic clustering
- Very slow
- *mothur*

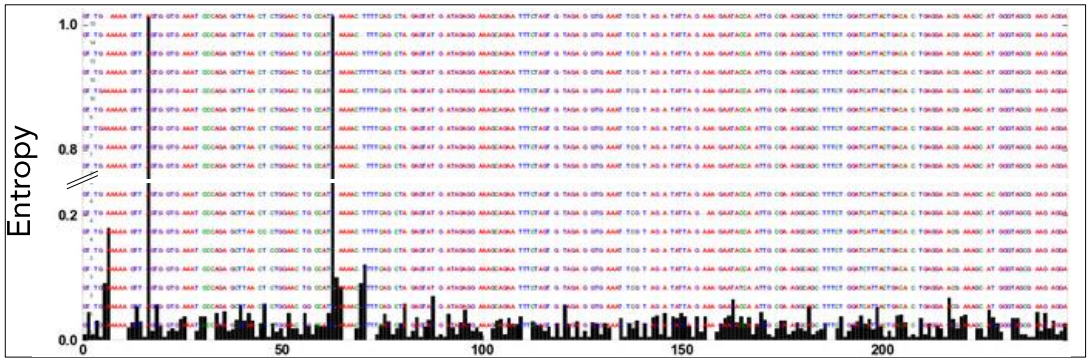
Heuristic clustering



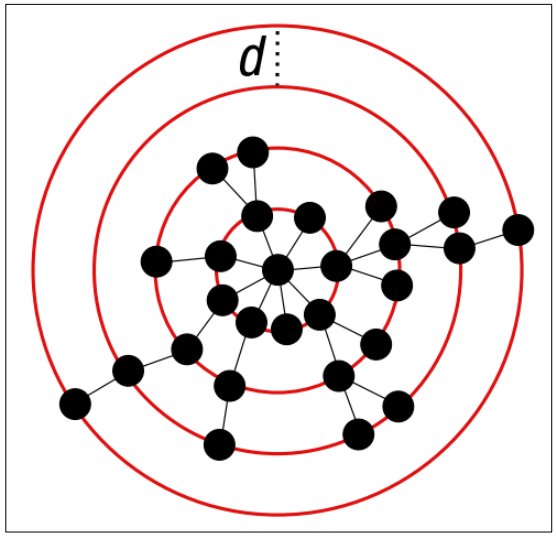
- Fast compared to hierarchical clustering
- Low reproducibility
- *vsearch, qiime*

OTU clustering

- Minimum entropy decomposition (MED)
  - Fast
  - Omits stochastic variation
  - Sub-species resolution (SNPs)
  - No rare biosphere



- Swarming (***swarm***, *OBITools*, *unoise*)
  - Fast
  - Variable OTU cut-off
  - High reproducibility
  - Dimension of swarms depending on sequencing space
  - Spurious OTUs (reduced in unoise algorithm)



- Denoising (***dada2***, *deblur*)
  - Probability that any unique sequence was created by sequencing errors
  - High taxonomic resolution
  - Less rare (spurious?) OTUs than swarm
  - Requires very high quality sequences as input

Taxonomic classification

	Principle	Pro	Con	Implementation
Blast against reference database	Take best hit	Long taxonomic paths	Spurious assignments	*silvangs*
Last common ancestor consensus	Truncate taxonomic path if ambiguous assignment	More conservative	Taxonomic paths often truncated	*sina*
Bootstrap confidence	Calculate confidence of assignment → truncate taxonomic path if confidence is below threshold	More conservative	Taxonomic paths often truncated	RDP Naive Bayesian Classifier (*mothur*, *dada2*)
Phylogenetic placement	Add sequence to phylogenetic tree	Phylogenetic information	Calculation of tree	*pplacer*

What to do with unclassified/uncultured sequences?
→ hypothetical species based on phylgenetically meaning full units

Resources for amplicon sequence analysis

Programs:

- mothur (https://www.mothur.org/)
- OBITools (https://git.metabarcoding.org/obitools/obitools/wikis/home)
- vsearch (https://github.com/torognes/vsearch)
- qiime2 (https://qiime2.org/)
- dada2 (https://benjjneb.github.io/dada2/)

Web resources:

- silvangs (https://www.arb-silva.de/ngs/)
- RDP (https://pyro.cme.msu.edu/)
- qiita (https://qiita.ucsd.edu/static/doc/html/index.html)

Analysis offered by sequencing company:

- **CAUTION: carefully check their approach!**

Get help:

- Google
- seqanswers (http://seqanswers.com/)
- biostars (https://www.biostars.org/)