Johannes Dröge, johannes.droege@chalmers.se
EMBO practical course **Plant Microbiota**
Cologne, March/April 2017

Taxonomic Characterization and Diversity Analyses

CHALMERS
UNIVERSITY OF TECHNOLOGY

Taxonomy Structure

- Hierarchical ordering (topology)

- Introduced for higher-level organisms

  - Carl von Linné

  - classification by morphological traits

- Microbial taxonomy

  - mostly by nucleotide sequence
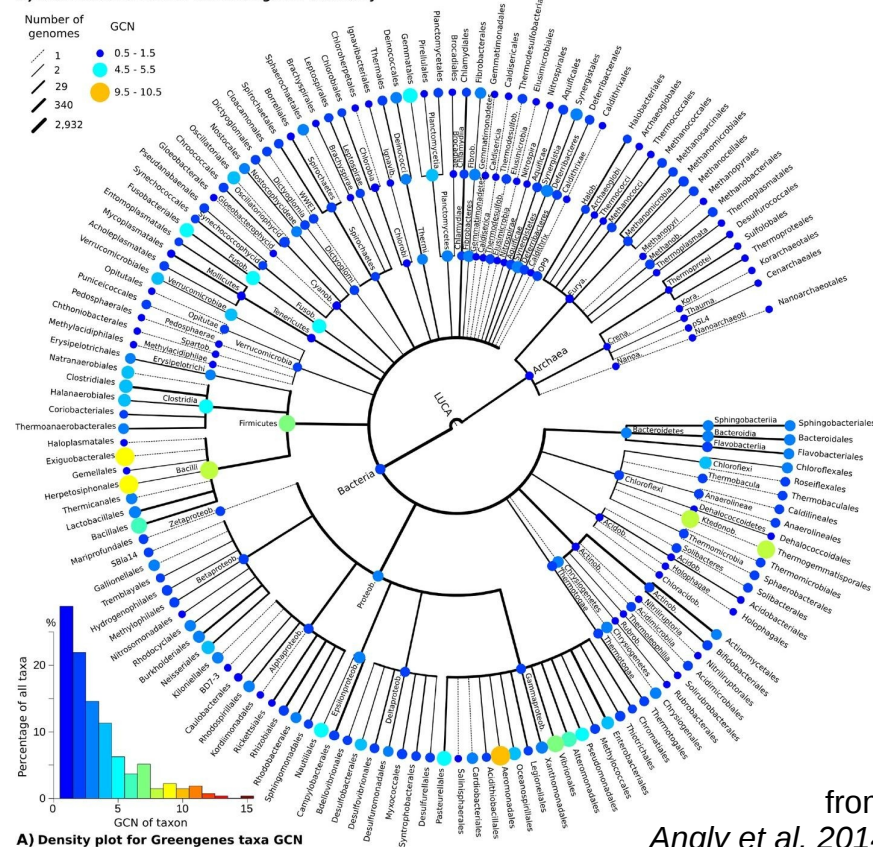
  - approximates **species tree of life**
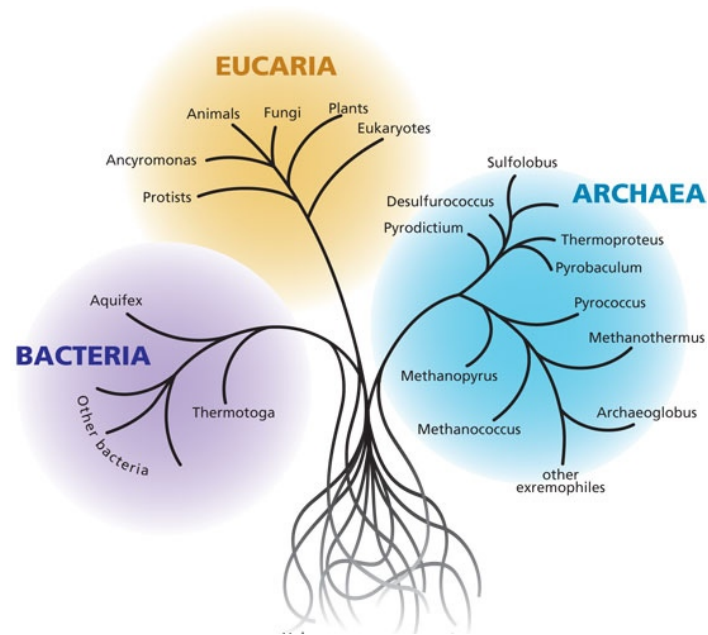
from *Woods Hole Currents, Vol. 8*

Using Taxonomy

- Simplify reasoning

  - group OTUs by evolution

  - transfer knowledge

- Create features for analysis

  - compare communities

from
Angly et al. 2014

Microbial Taxonomy Providers

from www.ezbiocloud.net

- GreenGenes (08/2013)

- Ribosomal Database Project (09/2016)

- SILVA rRNA database project (09/2016)

- EzBioCloud (03/2017)

- NCBI taxonomy (continuous updates)

- **check usage restrictions!**
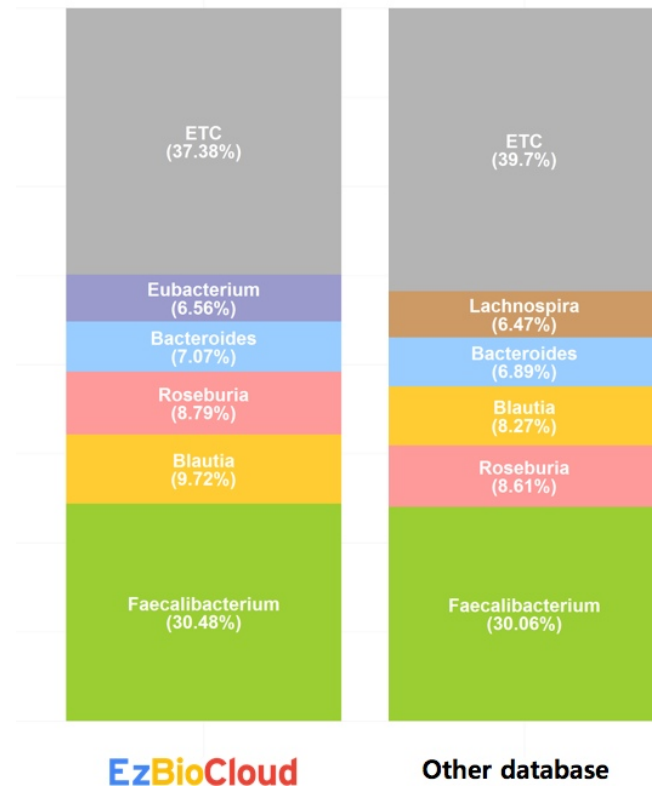
Taxonomic Classifications

- Basic ingredients

  - reference sequences

  - corresponding taxonomy

  - similarity search program

- Do **nearest neighbor classification**

  - find close references
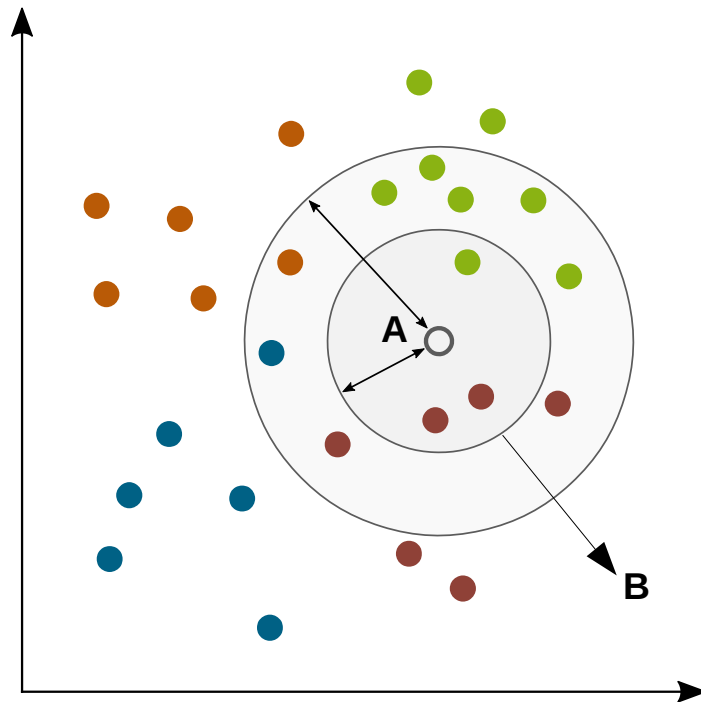
  - carry over taxonomic annotation
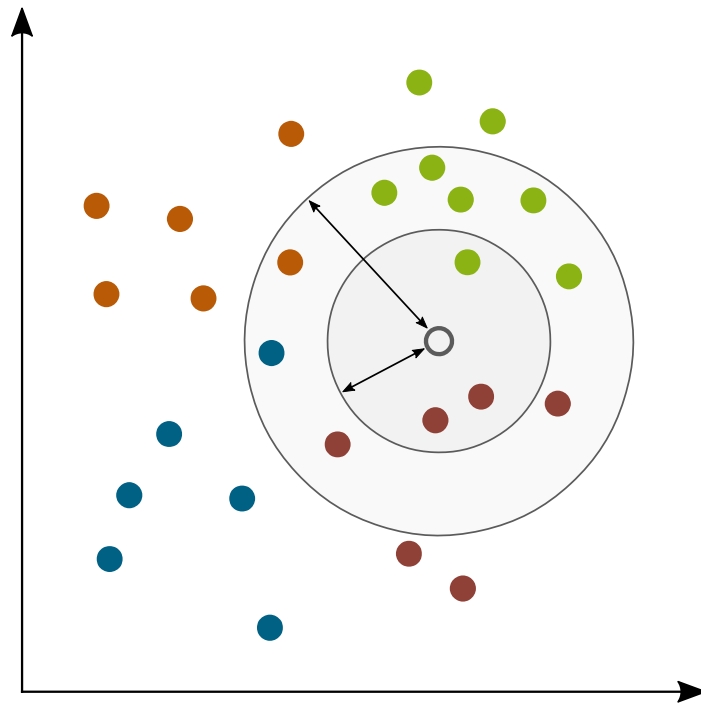
Methods for Classification

A) Similarity search

- local alignment (BLAST)

- fast k-mer comparison

B) Taxonomy assignment

- majority vote

- **lowest common ancestor (LCA)**

- phylogenetic inference

Programs

- all-inclusive:

  - **QIIME:** wrappers

  - **mothur**: re-implementations

  - download reference sequences+ taxonomy and search+classify

- stand-alone: RDP classifier

- online: RDP, SILVA/SINA

RDP Classifier

- Naïve Bayesian classifier

- Original: all predictions
  at genus level

- *mothur* re-implementation:
  dynamic rank classifications

- Naïve Bayesian classifier

- Frequencies of shared 7-mers

- Bootstrapping confidence scores
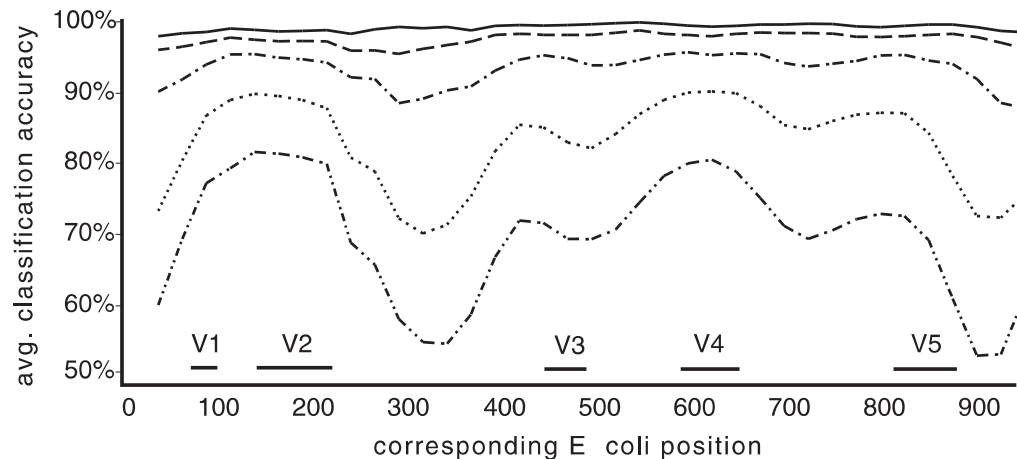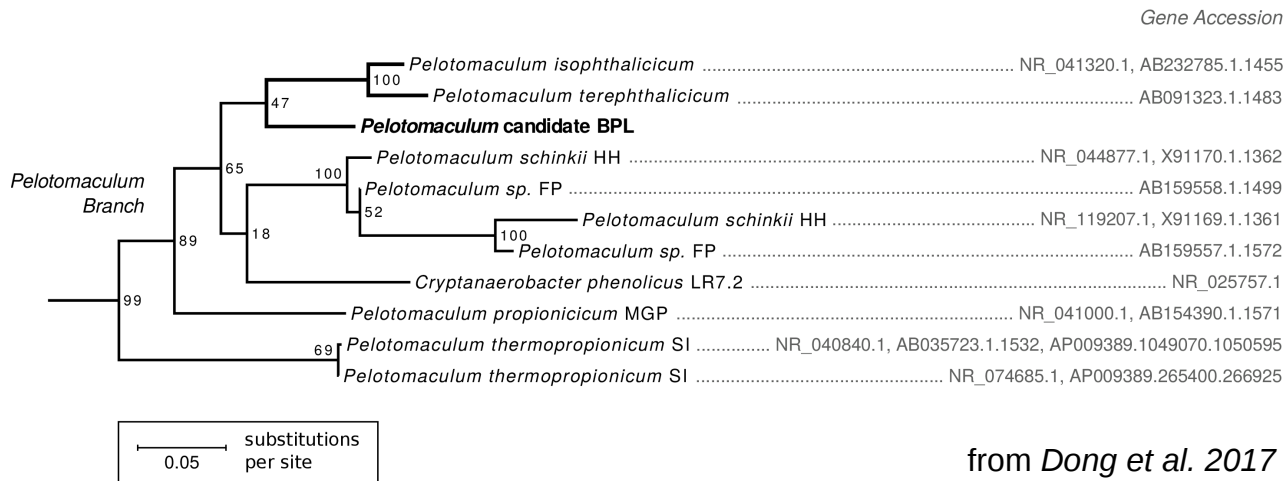

from *Wang et al. 2007*

Phylogenetic Reconstruction

- Learn about evolution of novel, deep-branching OTUs

- Collect homologs

- Create **Multiple Sequence Alignment**

- Construct **phylogenetic tree**

from *Dong et al. 2017*

Shotgun Metagenomes

- Characterize community by extracting genes for classification

    - ribosomal: 5S/16S/18S

    - single-copy marker genes
      e.g. Campbell et al. 2016

- Local alignment (*BLAST*)

- Hidden Markov models (HMMs)

Whole-genome Classification

- genome bins/partial genomes

  - nucleotide composition (k-mers): *NBC, PhyloPythiaS(+)*

  - sequence similarity: C*entrifuge*, *taxator-tk*

- full genomes

  - average nucleotide identity (ANI)

  - *MUMMER*

Abundance and Taxonomy

- combine number of reads and taxa

- compare profiles over communities, conditions etc.

- ordination techniques to visualize similarity

- diversity estimation
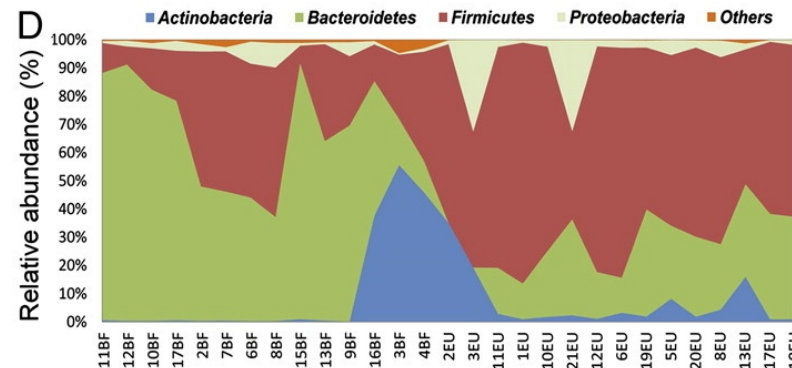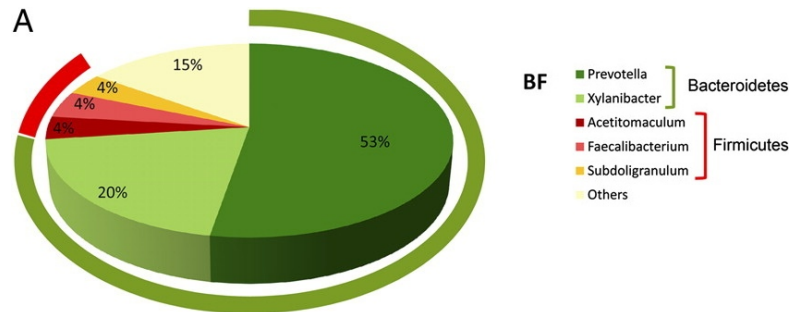
from *Filippo et al. 2010*

Keep in mind

- Absolute profiles are skewed by (16S) gene copy number variations

- Primer bias (amplicions)

- Species (16S) vs. gene evolution

- Specifics of assignment algorithm

  - selected taxonomy

  - level of false positives

- **Don't put too much trust into taxa**

- **Calculate fold change**

Hands-on Session

Taxonomic classification

Load Tutorial Enviroment

- includes QIIME scripts and shell functions

```
source base_folder/activate.sh
```

QIIME assign_taxonomy.py

- See online documentation at http://qiime.org/scripts/assign_taxonomy.html

- use different taxonomy and reference sequences

  - GreenGenes, SILVA, RDP (folder references)

- use different methods

  - uclust, mothur/rdp, sortmerna

Inspect Taxonomic Annotation

- compare output with

  `show_tax_assignments output1.txt output2.txt | less -S`

- run mothur/rdp twice and compare results between runs

- compare with whole genome annotations

  - in references/wgs.txt

  - uses NCBI taxonomy

  - generated with *taxator-tk*

Compare to Online Annotation Services

- Extract first 10 sequences using

```
seqkit head 10 < in.fna > out.fna
```

- SILVA/SINA aligner
  https://www.arb-silva.de/aligner/

- RDP Seqmatch
  https://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp

Annotate Your OTU-picked Data (morning session)

- See online documentation for

  - `pick_otus.py`

  - `assign_taxonomy.py`

  - `make_otu_table.py`

  - `summarize_taxa.py`