Hands on introduction to ancient microbiome analysis

Maxime Borry - August 18th 2021

MPI-SHH
SUMMER SCHOOL
2021

Doorway
to Human History

Teaching material

tinyurl.com/ancientmicrobiome

Who am I ?

Maxime Borry - Doctoral researcher at MPI-SHH - MPI-EVA

Ancient DNA microbiome bioinformatics 💩 🍷 🧬 🐱‍💻

🐦 @notmaxib

🐙 @maxibor

🌐 maximeborry.com

How do we analyze microbiomes ?
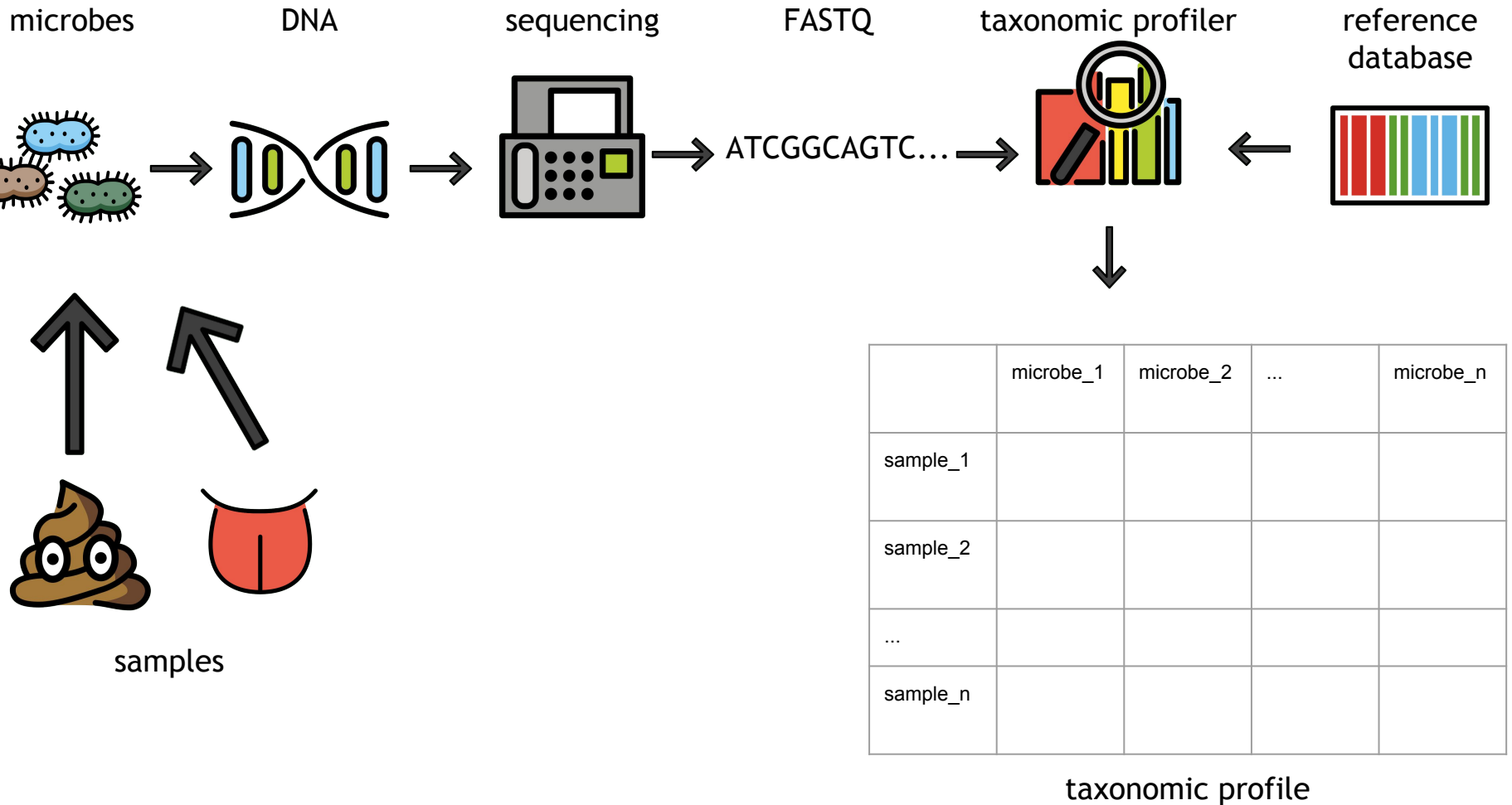
microbes　　　　DNA　　　　sequencing　　　　FASTQ　　　　taxonomic profiler　　　　reference database

ATCGGCAGTC...

samples

	microbe_1	microbe_2	...	microbe_n
sample_1				
sample_2				
...				
sample_n				

taxonomic profile

How do we analyze **ancient** microbiomes ?

ancient DNA · sequencing · FASTQ · taxonomic profiler · reference database

ATCGGCAGTC…

ancient samples

	microbe_1	microbe_2	…	microbe_n
sample_1				
sample_2				
…				
sample_n				

taxonomic profile

How do we analyze **ancient** microbiomes ?

ancient
DNA

sequencing

FASTQ

taxonomic profiler

reference
database

ancient
samples

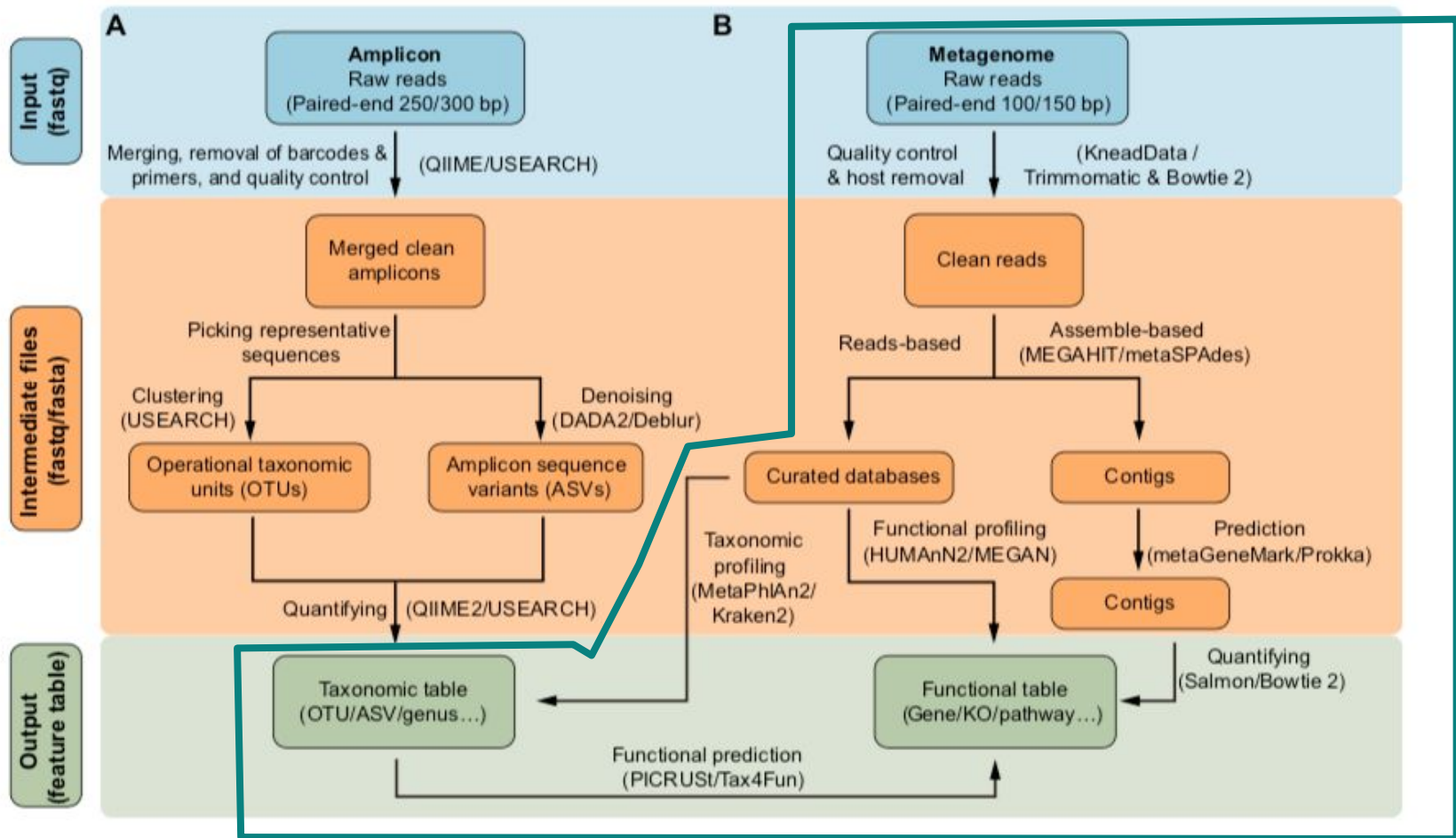...

microbe_n

c profile
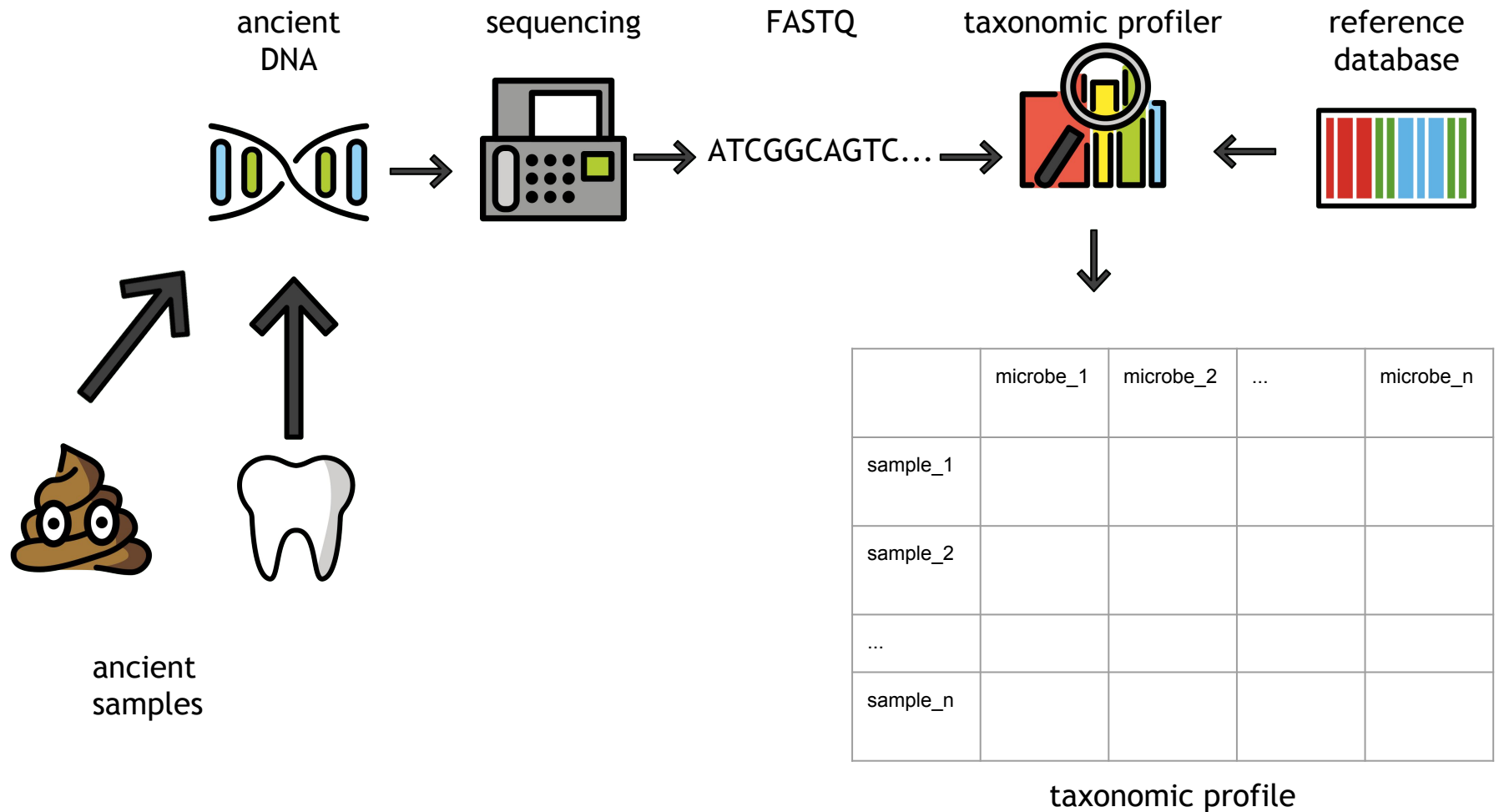
SUMMER
SCHOOL
2021
Doorway to
Human
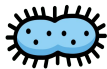History

More in details

Liu Y-X, Qin Y, Chen T, Lu M, Qian X, Guo X, Bai Y. 2021. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein & Cell* 12:315–330. DOI: 10.1007/s13238-020-00724-8.
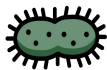
Ambiguity in taxonomic assignation
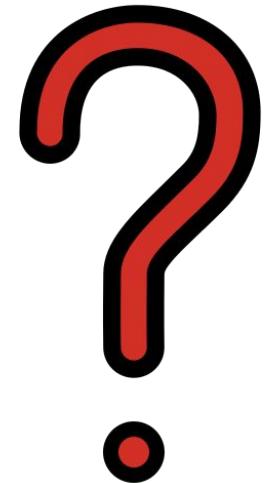
sequence 1 ATGGTC**GGGCAGGACG**TTGCGAGT
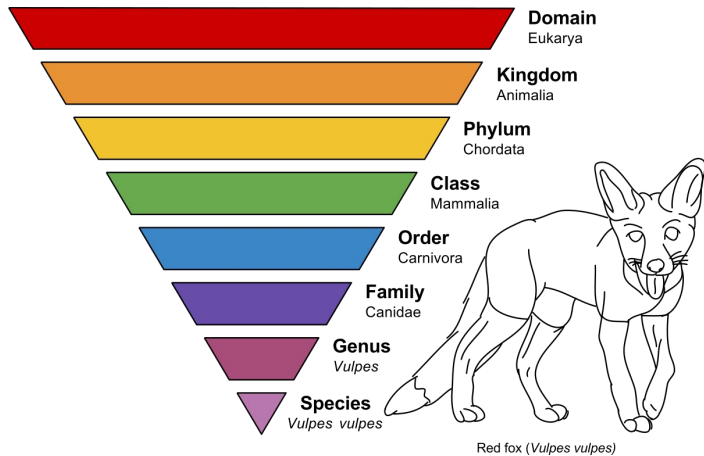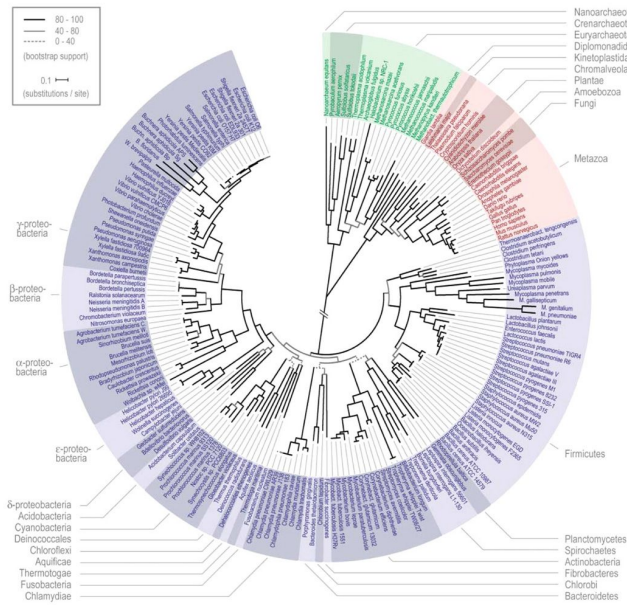sequence 2 CGAGAA**GGGCAGGACG**CCACGTAC

ATGGTC**GGGCAGGACG**TTGCGAGT

CGAGAA**GGGCAGGACG**CCACGTAC

Taxonomy and LCA to the rescue

Red fox (*Vulpes vulpes*)
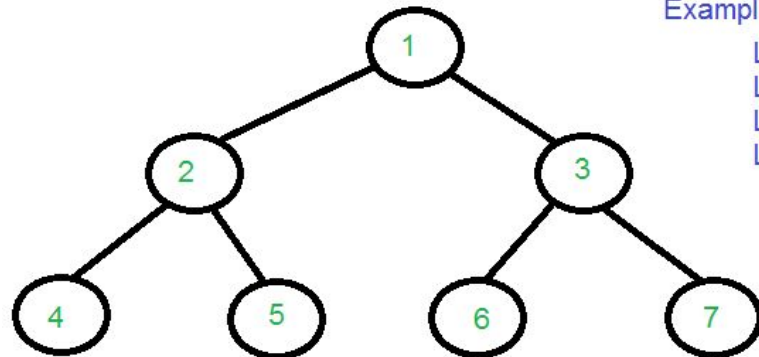
- Species level assignation is not always possible.

- Possibility of hits in different species

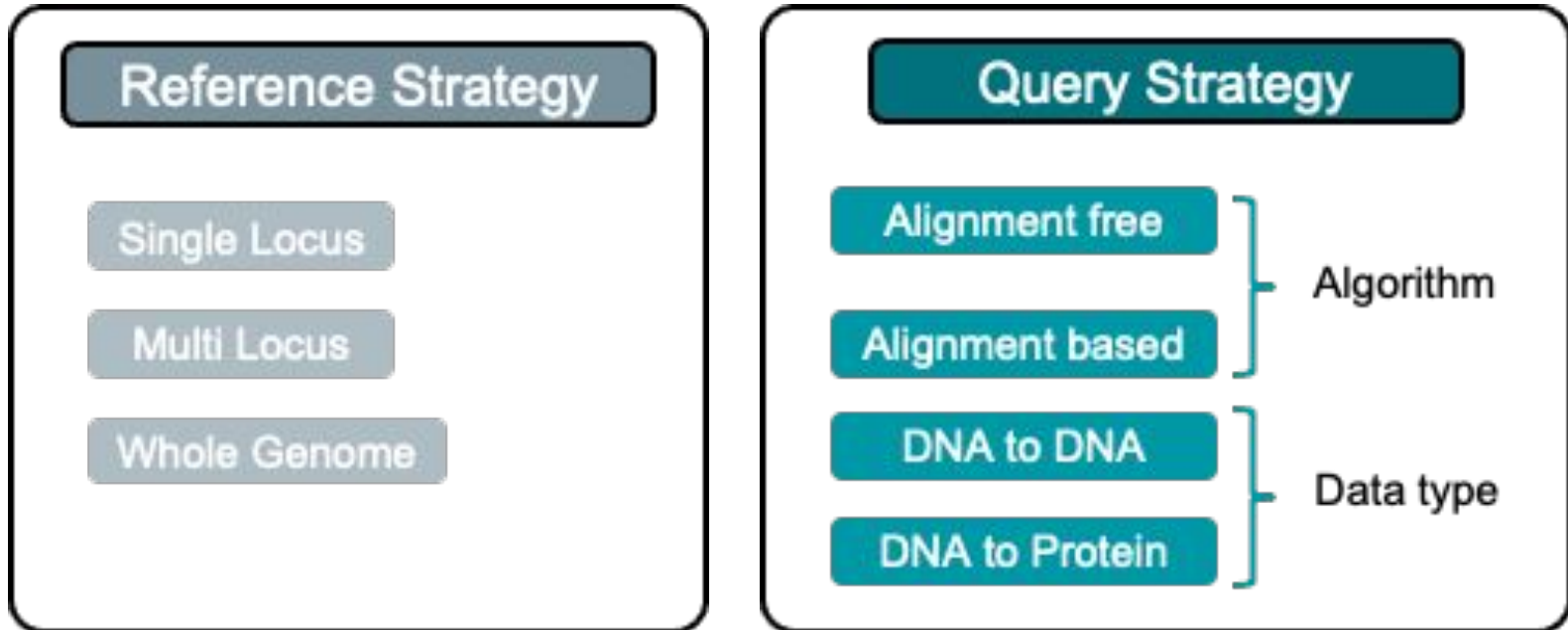- Ambiguities solved by LCA (Lowest Common Ancestor) algorithm.

Examples

LCA(4, 5) = 2
LCA(4, 6) = 1
LCA(3, 4) = 1
LCA(2, 4) = 2

Different taxonomic profilers

maximeborry.com/courses : "Taxonomic classifiers and sequence alignment algorithms"

(most common) taxonomic profilers used in aDNA

Kraken2

- No alignment
- Fast
- Lower specificity (more false positives)

MetaPhlAn

- Custom curated marker database
- Reasonably fast
- Good balance between specificity and sensitivity

MALT

- Alignments
- Slower and resource hungry (if using a big whole genome database)
- Best balance between specificity and sensitivity
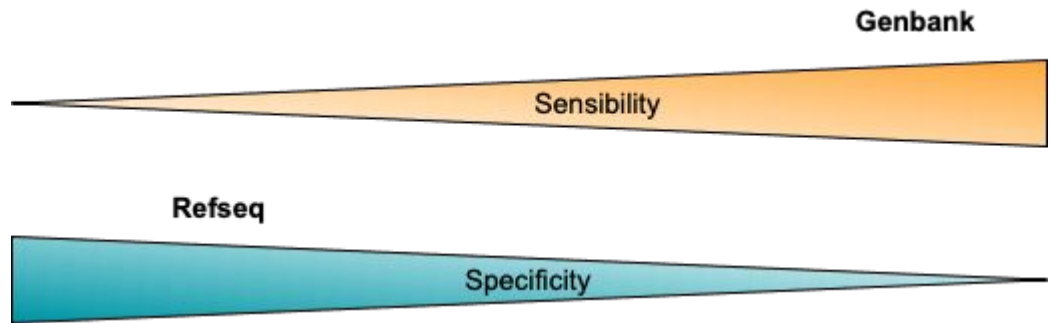
Taxonomic profilers benchmark (and more): CAMI challenge

Reference databases

NCBI databases

- ## NCBI nr/nt
  - the largest database
- ## NCBI RefSeq
  - A curated subset

Custom databases

- Metaphlan: clade specific markers

How is aDNA microbiome different from "modern" microbiome analysis ?

- **We have to show our sample is what we claim it to be:**
  - **time period**
    - Is it actually ancient ?
      - Isotopic dating
  - **source**
    - Is it from the correct host ?
      - microbiome profile, host DNA
    - Is it from the correct ecological niche ?
      - microbiome profile
  - **contamination**
    - How much of the sample is endogenous ?
      - taxonomic composition of bacteria carrying deamination damage
    - Is there a lot of modern contamination (excavation, lab, …) ?
      - taxonomic composition of non deamination damaged bacteria

How does aDNA damage affect taxonomic profiling ?

Damage isn't really an issue
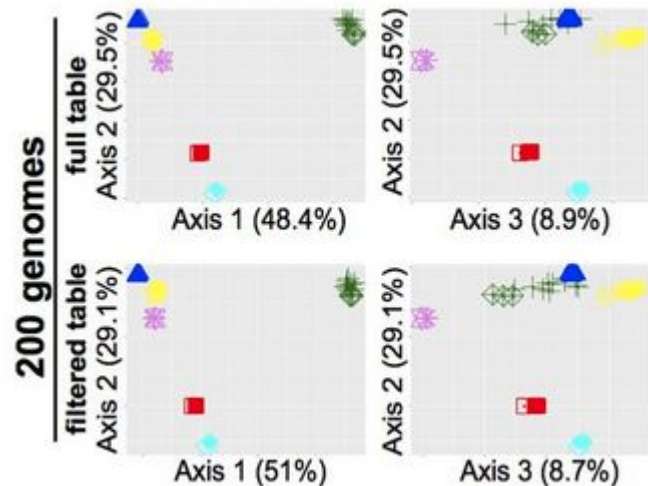
PCoA of simulated community composition with Weighted-Unifrac Distance

Ancient: ✳ True + QIIME/UCLUST ◈ MALT ▲ MetaPhlAn2 ● MIDAS ■ CLARK-S
Modern: ✳ True ◈ QIIME/UCLUST ◇ MALT △ MetaPhlAn2 ◉ MIDAS ▣ CLARK-S

Velsko, Irina M., et al. "Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research." *Msystems* 3.4 (2018): e00080-18.

But very short sequences are more problematic

The problem with short sequences

Ambiguity in taxonomic assignation

sequence 1 ATGGTC**GGGCAGGACG**TTGCGAGT
sequence 2 CGAGAA**GGGCAGGACG**CCACGTAC

ATGGTC**GGGCAGGACG**TTGCGAGT

CGAGAA**GGGCAGGACG**CCACGTAC
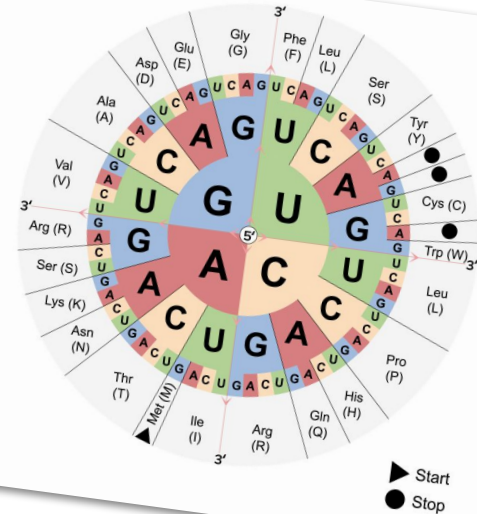
3 DNA bases = 1 amino acid

- [16s rRNA amplification/sequencing](#) is not very good
- Protein alignment is not good enough for very short sequences

Velsko, Irina M., et al. "Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research." *Msystems* 3.4 (2018): e00080-18.
Orlando, Ludovic, et al. "Ancient DNA analysis." *Nature Reviews Methods Primers* 1.1 (2021): 1-26.

aDNA vs modern microbiome

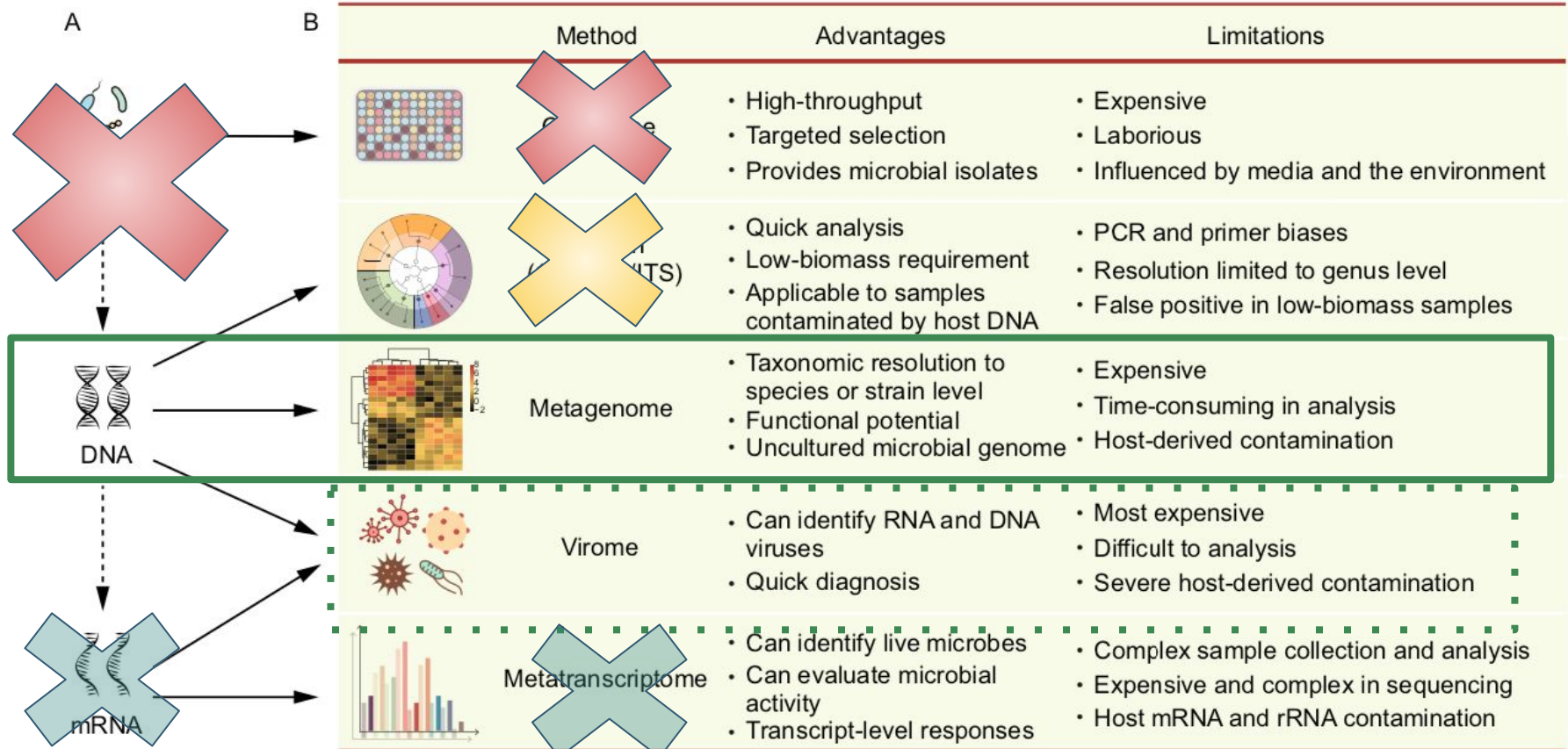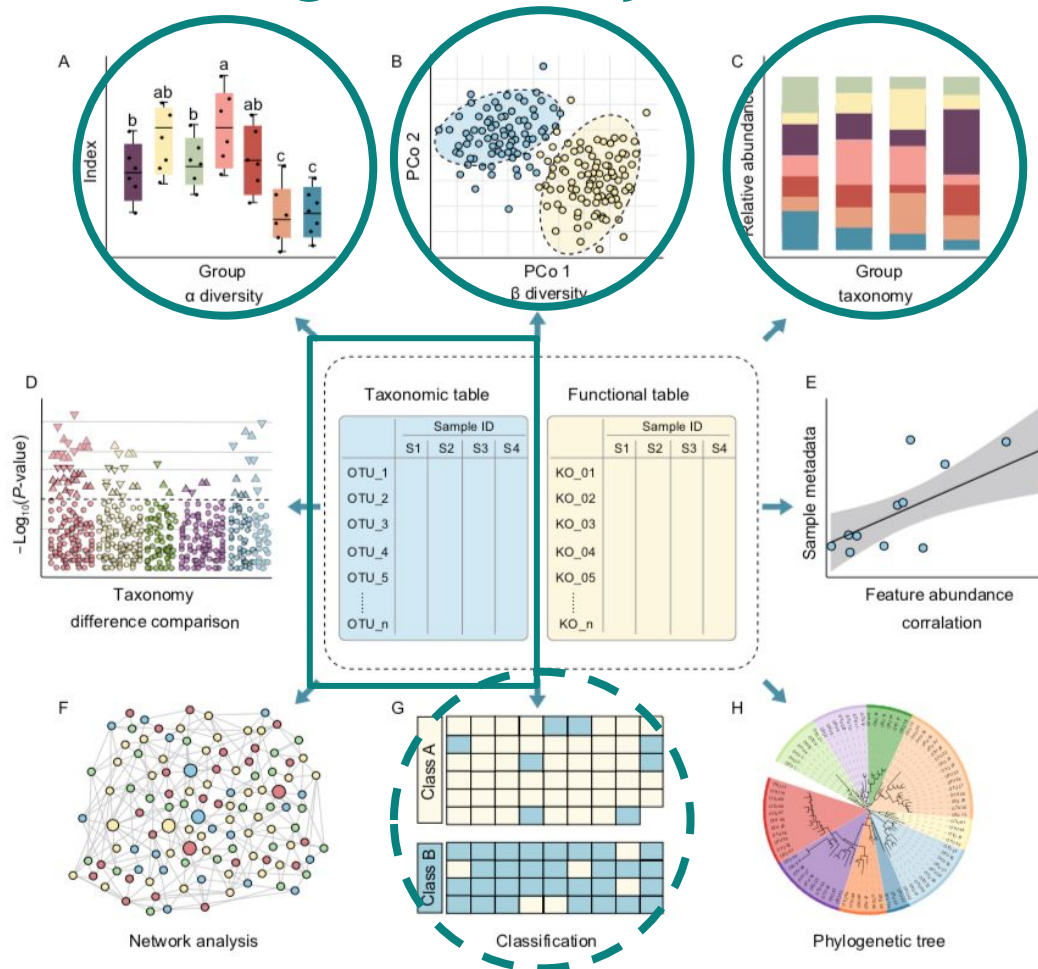| Method	Advantages	Limitations
Culture (crossed out)	• High-throughput • Targeted selection • Provides microbial isolates	• Expensive • Laborious • Influenced by media and the environment
Amplicon (ITS) (crossed out)	• Quick analysis • Low-biomass requirement • Applicable to samples contaminated by host DNA	• PCR and primer biases • Resolution limited to genus level • False positive in low-biomass samples
Metagenome	• Taxonomic resolution to species or strain level • Functional potential • Uncultured microbial genome	• Expensive • Time-consuming in analysis • Host-derived contamination
Virome	• Can identify RNA and DNA viruses • Quick diagnosis	• Most expensive • Difficult to analysis • Severe host-derived contamination
Metatranscriptome (crossed out)	• Can identify live microbes • Can evaluate microbial activity • Transcript-level responses	• Complex sample collection and analysis • Expensive and complex in sequencing • Host mRNA and rRNA contamination

Liu Y-X, Qin Y, Chen T, Lu M, Qian X, Guo X, Bai Y. 2021. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein & Cell* 12:315–330. DOI: 10.1007/s13238-020-00724-8.

SUMMER SCHOOL 2021 Doorway to Human History

What will we get today ?

Liu Y-X, Qin Y, Chen T, Lu M, Qian X, Guo X, Bai Y. 2021. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein & Cell* 12:315–330. DOI: 10.1007/s13238-020-00724-8.

Questions and tutorial !

tinyurl.com/microbiometutorial