



Standards,
Precautions &
Advances in
Ancient
Metagenomics

Intro to microbial ecology for ancient DNA

Maxime Borry
SPAAM summer school 2022

Who am I ?



Maxime Borry - Doctoral researcher at MPI-EVA

Ancient DNA microbiome bioinformatics



@notmaxib

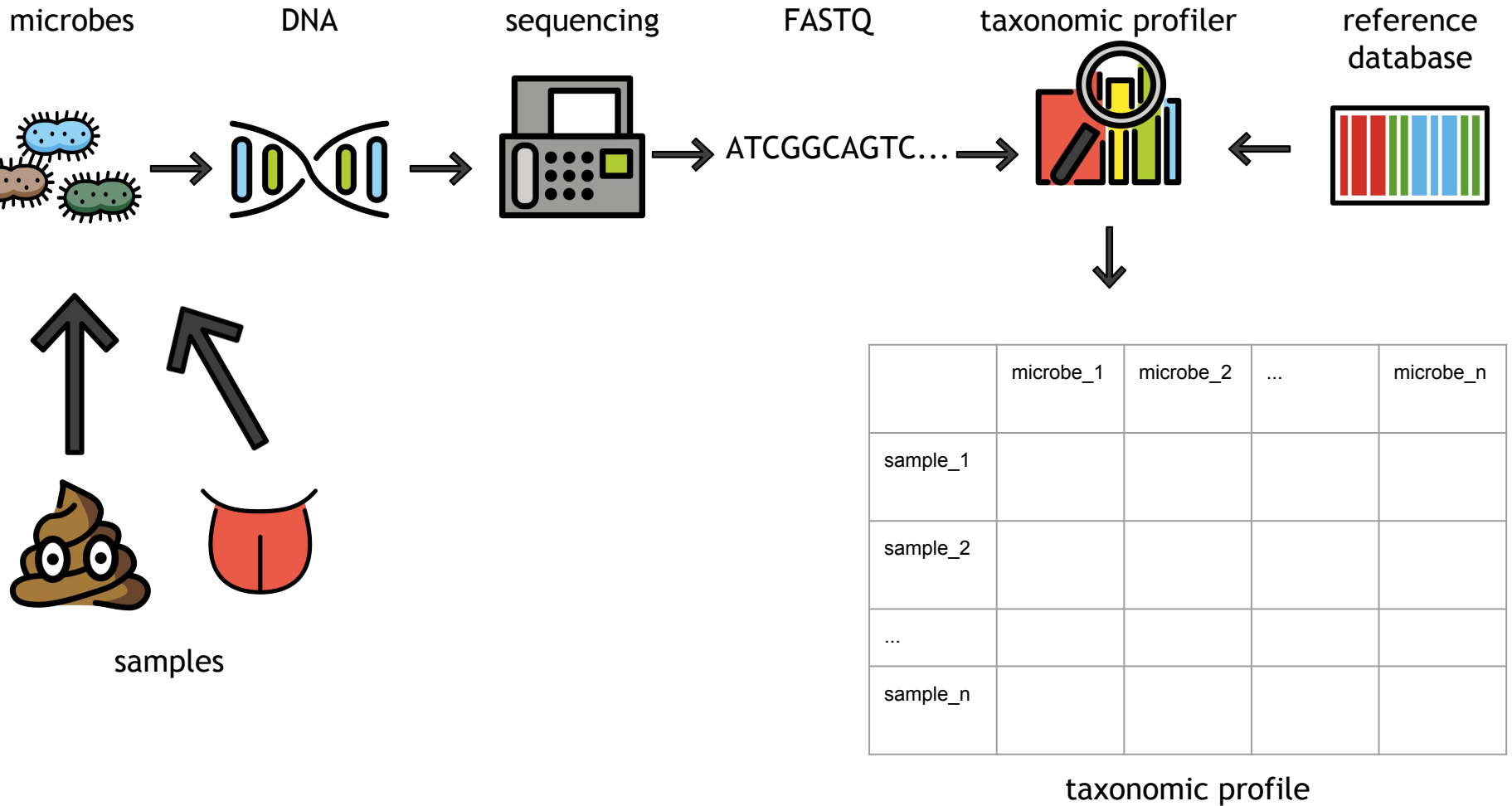


@maxibor

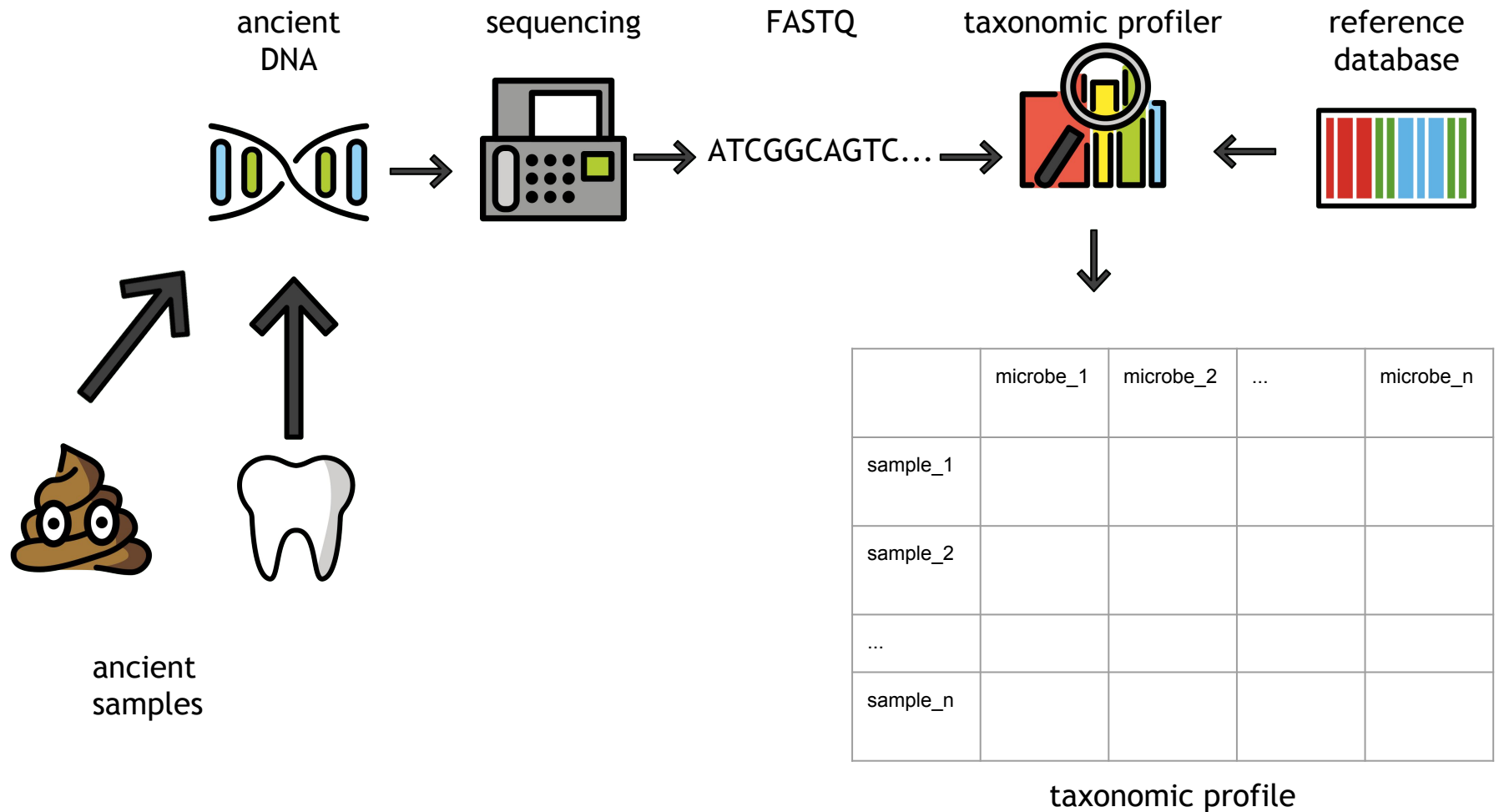


maximeborry.com

How do we analyze microbiomes ?



How do we analyze ancient microbiomes ?

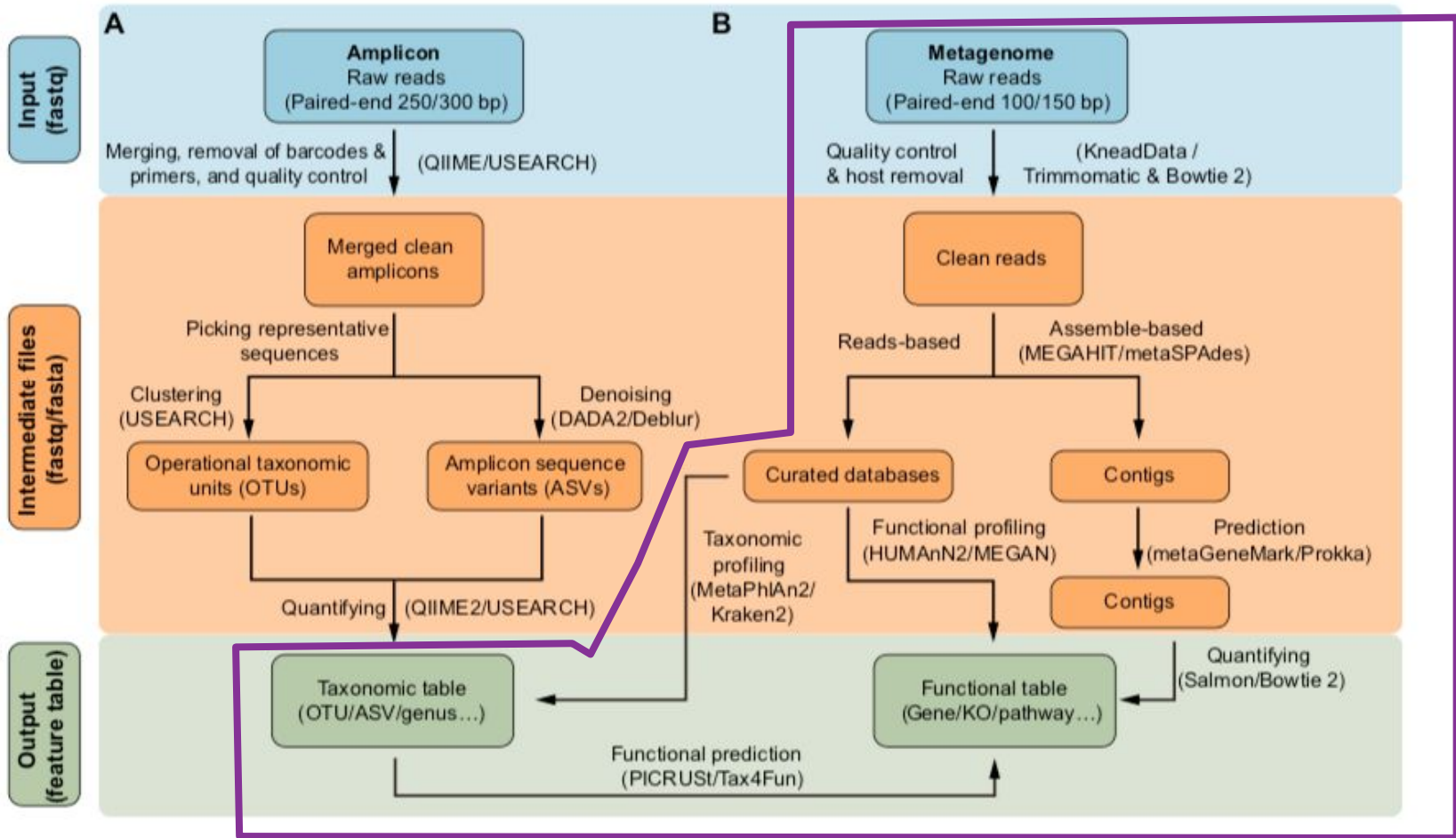


reference
database



c profile

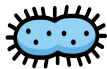
More in details



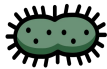
Ambiguity in taxonomic assignment

sequence 1 ATGGTCGGGCAGGACGTTGCGAGT

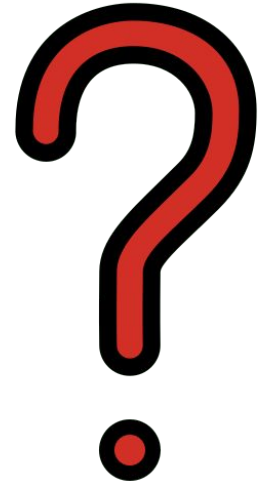
sequence 2 CGAGAAGGGCAGGACGCCACGTAC



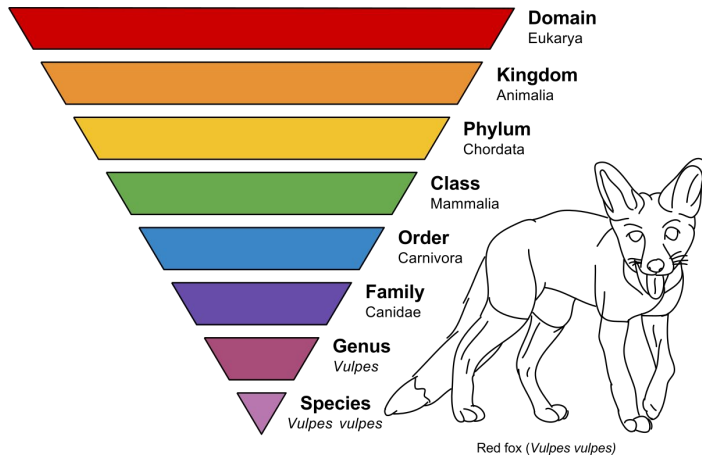
ATGGTCGGGCAGGACGTTGCGAGT



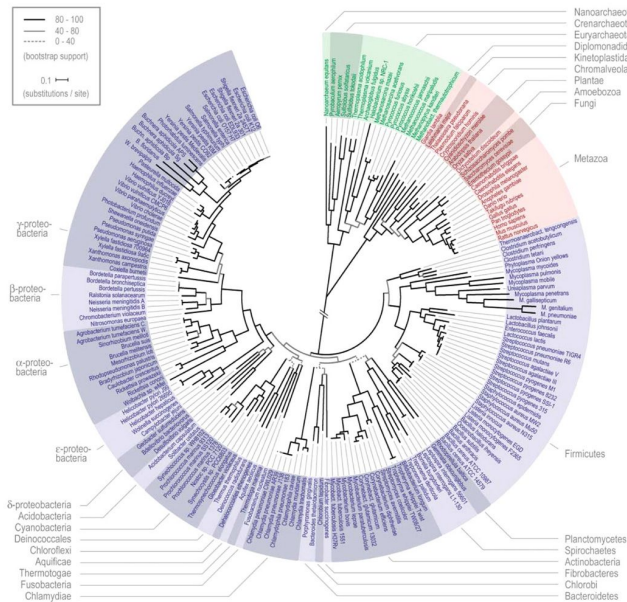
CGAGAAGGGCAGGACGCCACGTAC



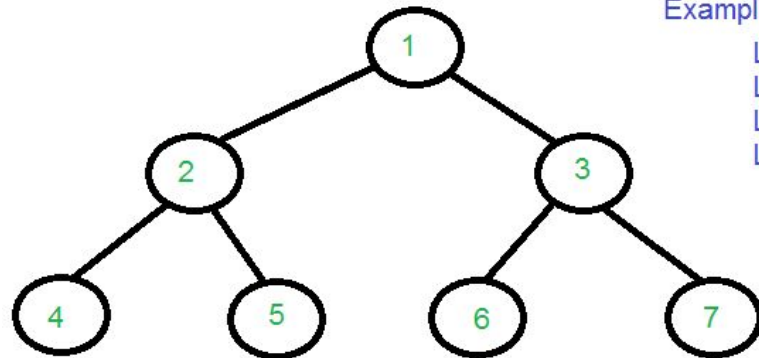
Taxonomy and LCA to the rescue



- Species level assignment is not always possible.
- Possibility of hits in different species
- Ambiguities solved by LCA (Lowest Common Ancestor) algorithm.

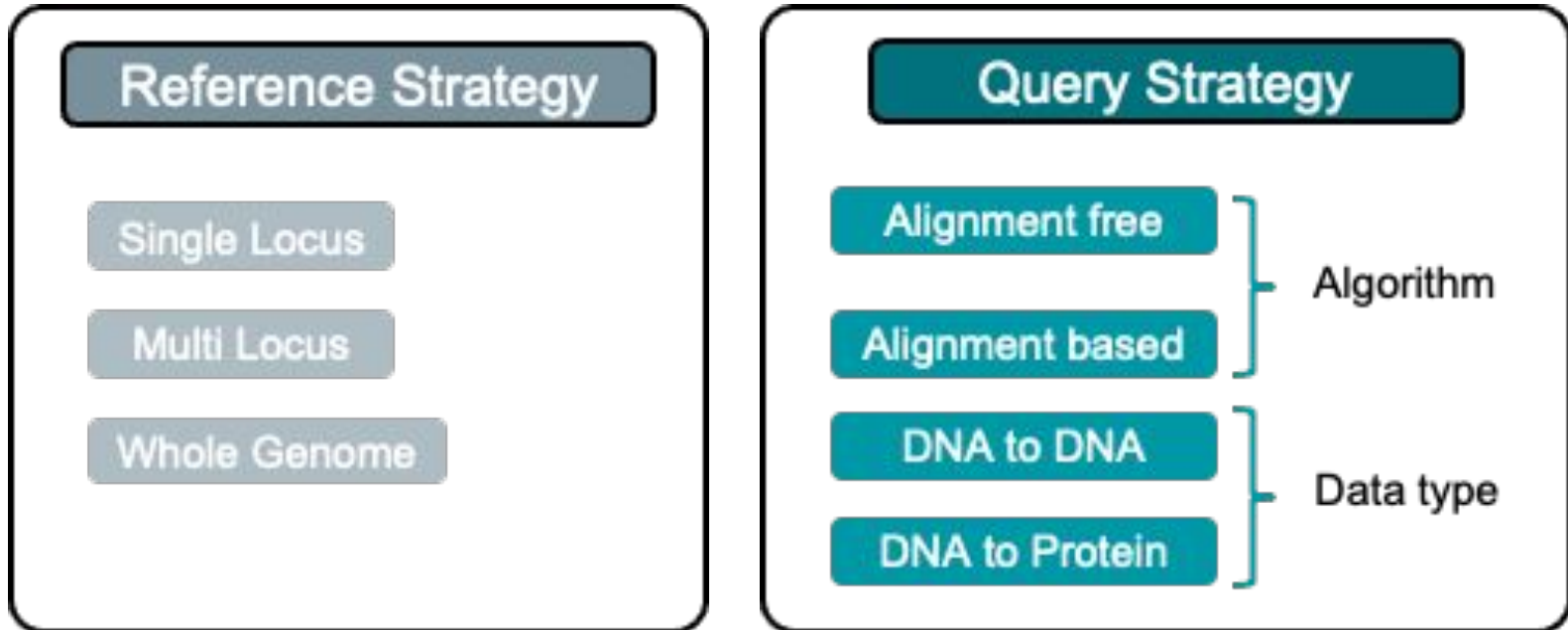


Examples



$$\begin{aligned} \text{LCA}(4, 5) &= 2 \\ \text{LCA}(4, 6) &= 1 \\ \text{LCA}(3, 4) &= 1 \\ \text{LCA}(2, 4) &= 2 \end{aligned}$$

Different taxonomic profilers



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

(most common) taxonomic profilers used in aDNA

Kraken family (Centrifuge, KrakenUniq, Kraken2)

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

MetaPhlAn

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

MALT

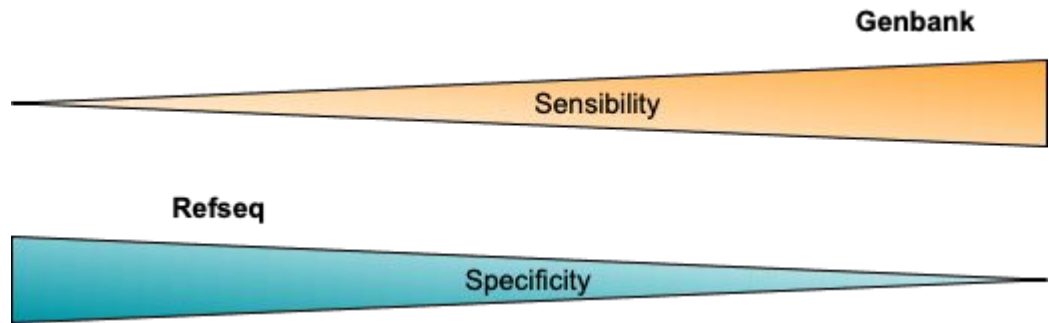
- Alignments
- Slower and resource hungry (if using a big whole genome database)
- Best balance between specificity and sensitivity

Taxonomic profilers benchmark (and more): CAMI challenge

Reference databases

NCBI databases

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



Custom databases

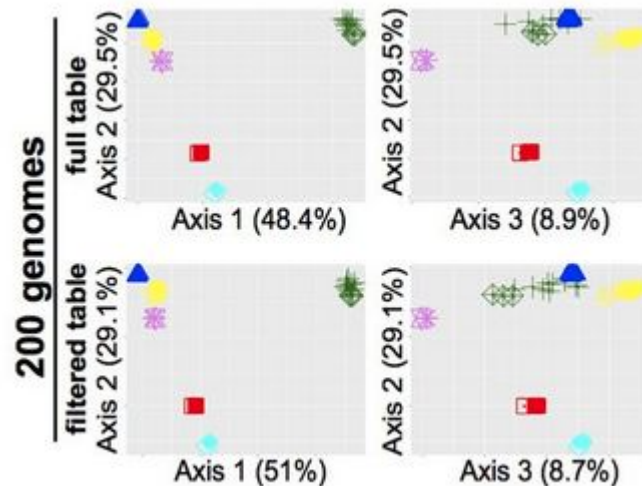
- Metaphlan: clade specific markers

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

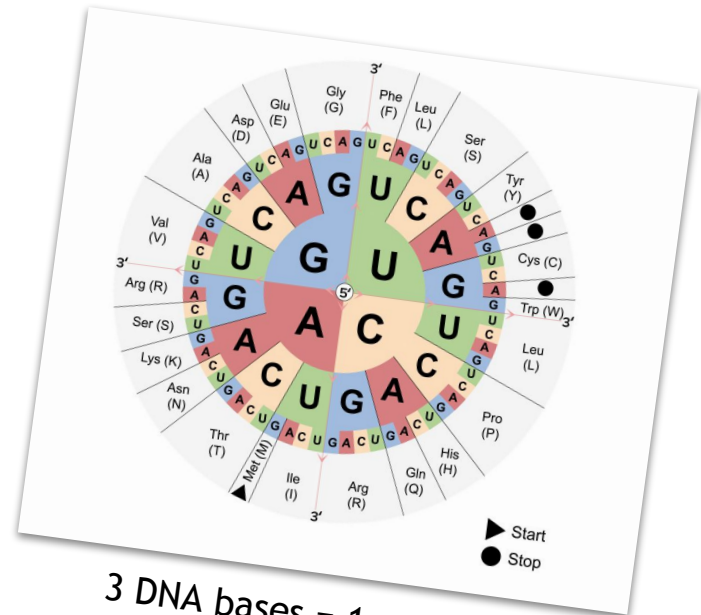
sequence 1 ATGGTCGGGCAGGACGTTGCGAGT
sequence 2 CGAGAAAGGCAGGACGCCACGTAC



ATGGTCGGGCAGGACGTTGCGAGT



CGAGAAGGCAGGACGCCACGTAC

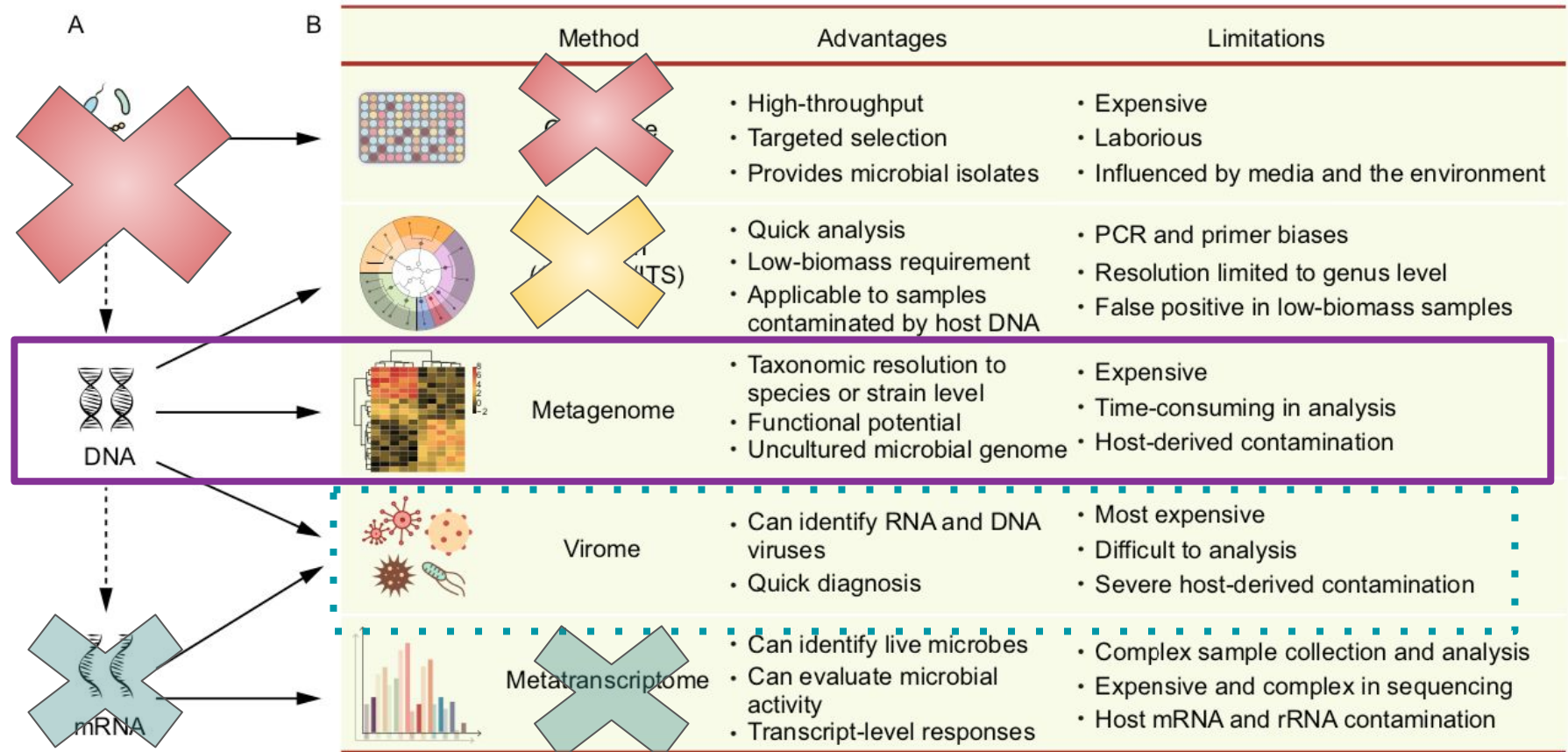


3 DNA bases = 1 amino acid

- 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



What will we get today ?

