Environmental metagenomics

Read-based taxonomic profiling
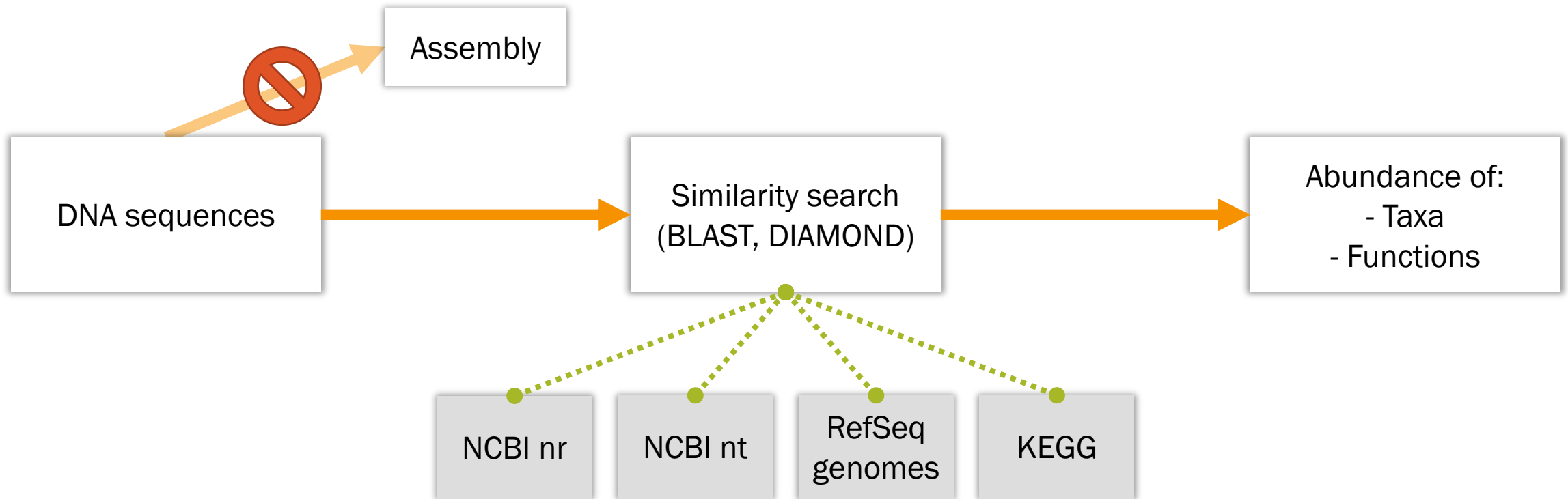
Igor S. Pessi & Antti Karkman
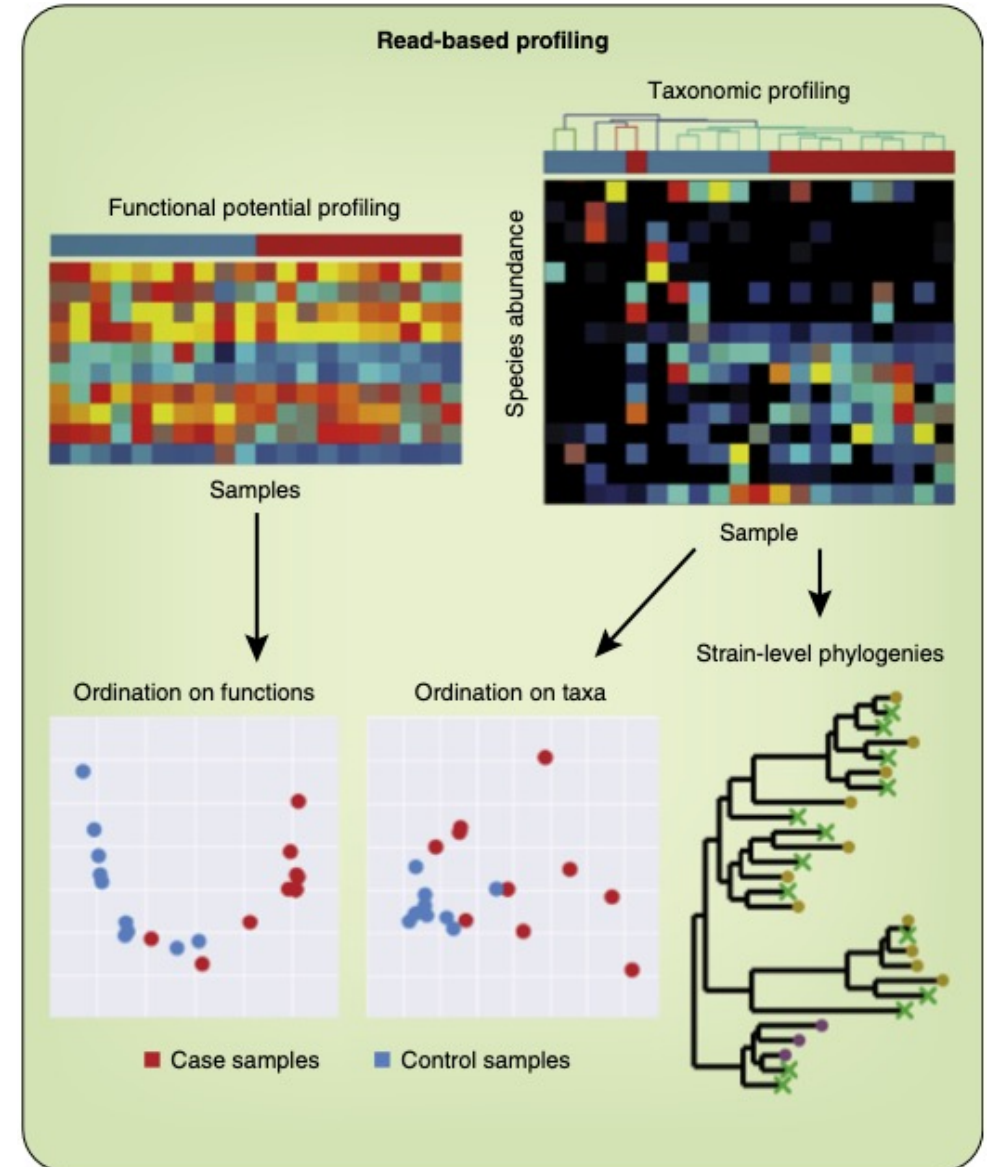
What is read-based profiling?

Read-based profiling is

Fast

Quantitative

Somewhat outdated

- Assembly-based are preferred
- Can give interesting preliminary insights
- Usually done as a "quick-and-dirty" estimate prior to assembly

Quince et al. 2017

Physalia
Courses

Approaches to <u>taxonomic</u> profiling

Read mapping and compositional binning

- Analysis of all reads
- Reference database of sequenced genomes
- <u>Mapping</u>: slow, requires lots of CPU and RAM
- <u>Compositional binning</u>: faster but less accurate

Bengtsson-Palme, 2018

Physalia
Courses
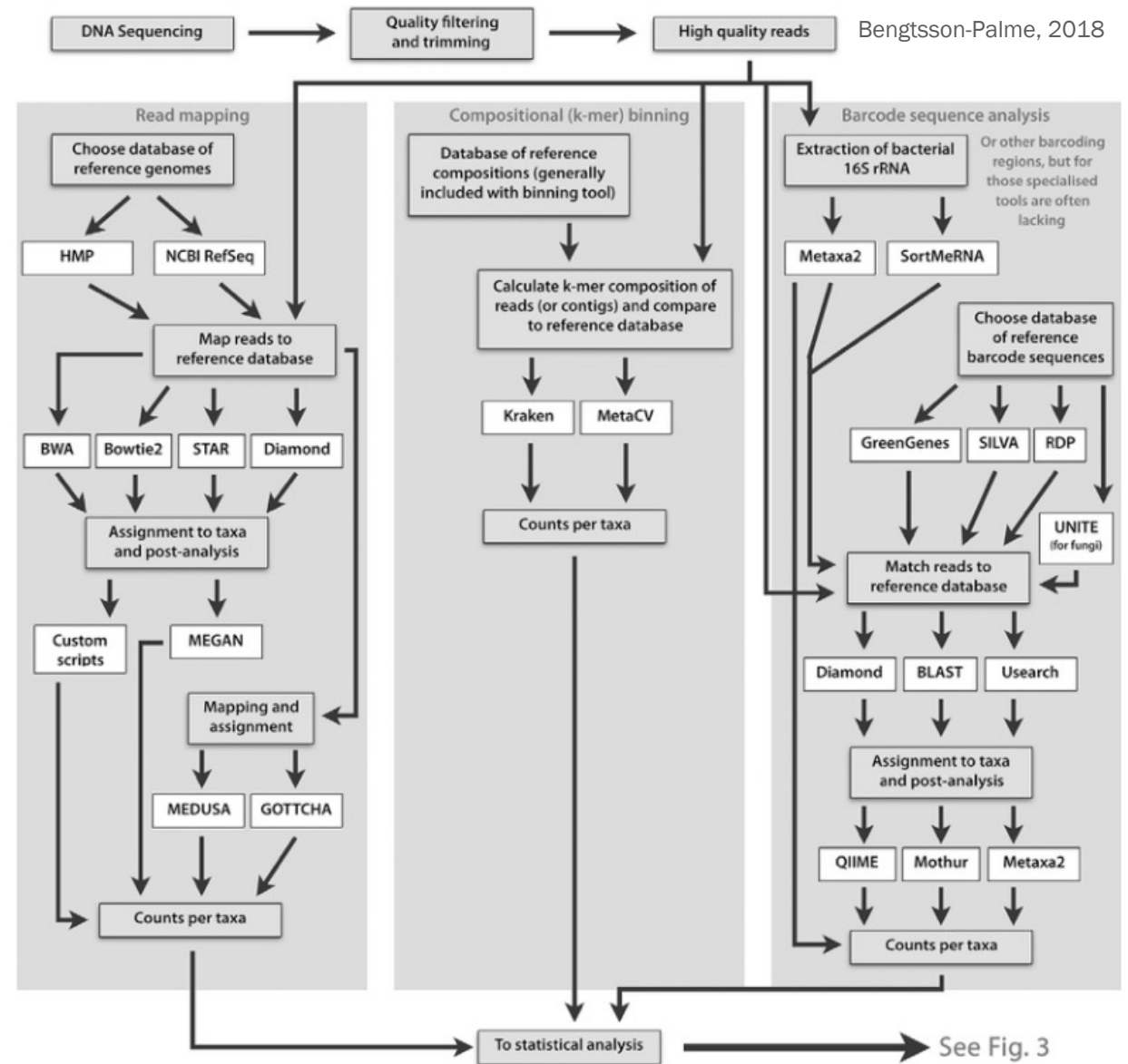
Approaches to <u>taxonomic</u> profiling

Barcode sequence analysis

- Analysis of specific barcode genes (e.g. 16S rRNA)
- Curated database of barcode sequences (e.g. SILVA)
- Much faster than the other approaches, but provides lower resolution

Bengtsson-Palme, 2018

Physalia
Courses
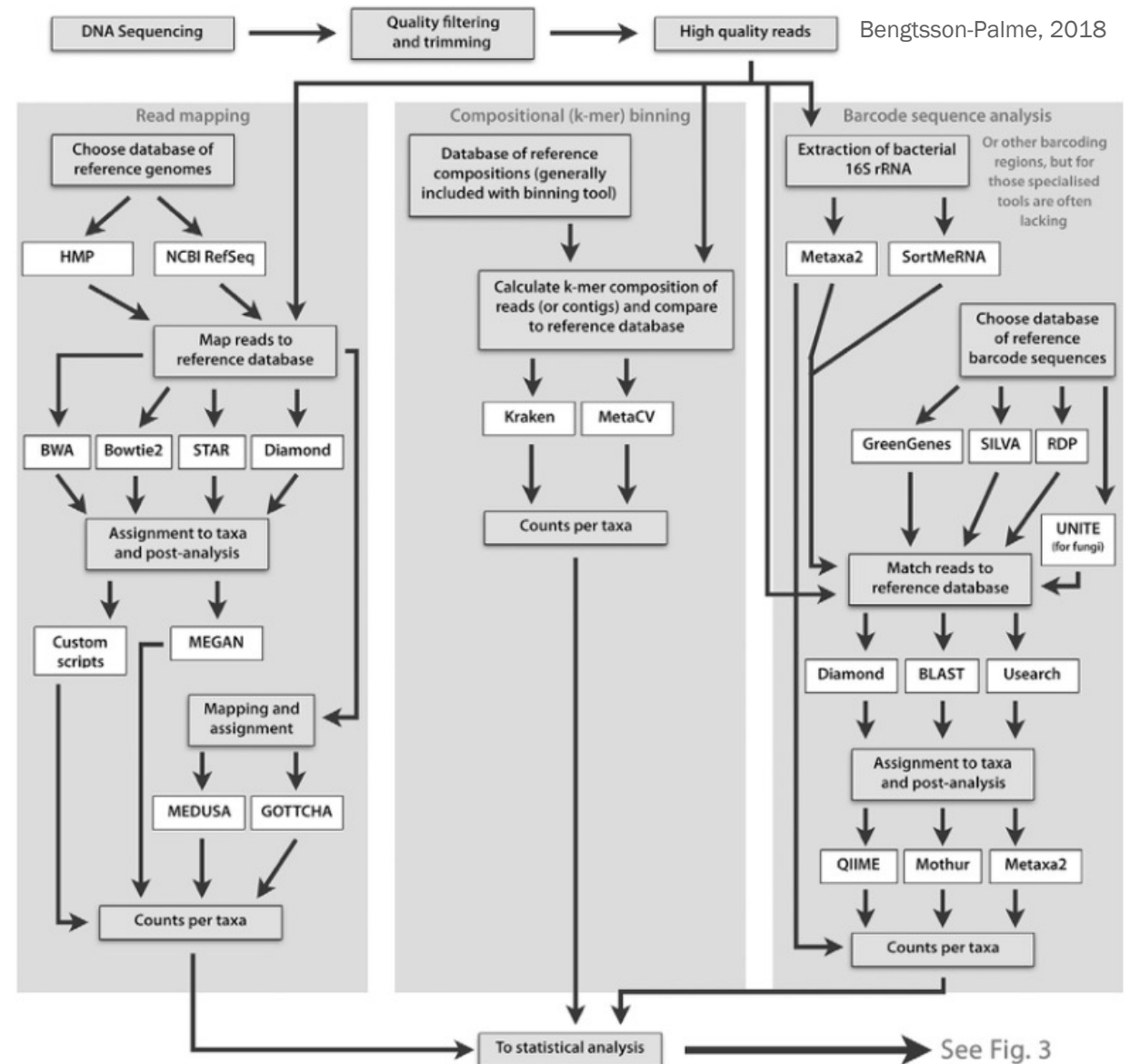
Approaches to taxonomic profiling: how to choose?

Analysis of all reads suffer from limited databases of reference genomes

- More suitable for environments that are better described (e.g. human gut)

Analysis of barcode genes suffer from lower resolution

- More suitable for environments with a high fraction of unknown microorganisms (e.g. soil)

Physalia
Courses
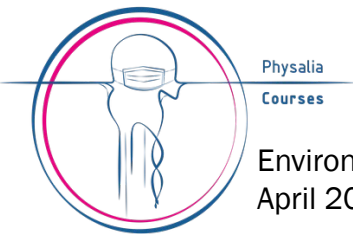
Making sense of read-based analyses

Comparative analyses

Statistics

- Univariate (e.g. ANOVA of specific genes and taxa)
- Multivariate (e.g. PERMANOVA, ordination/clustering, Mantel test)

Normalization!

- Library size
- Bacterial content (e.g *rpoB* gene)

Physalia
Courses