Metagenomics 101

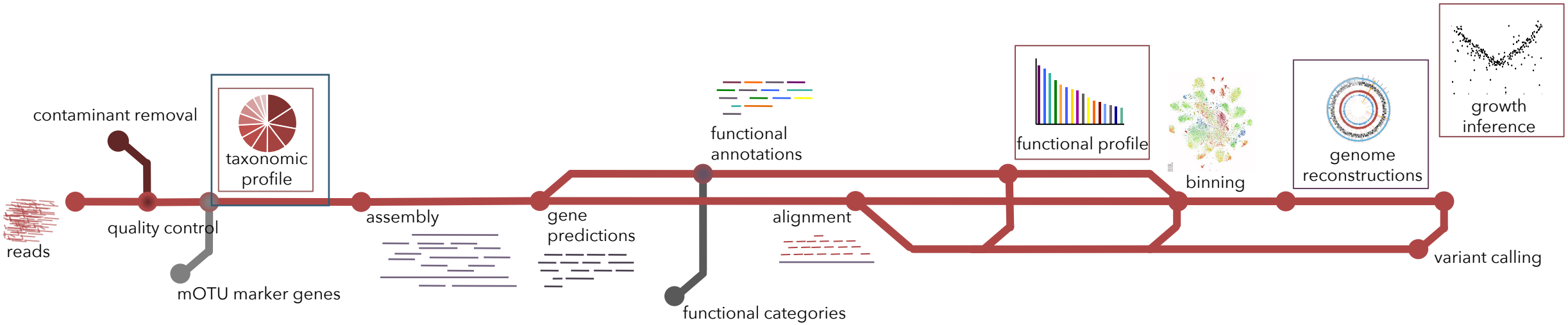
Session 4: Taxonomy I

Anna Heintz-Buschart

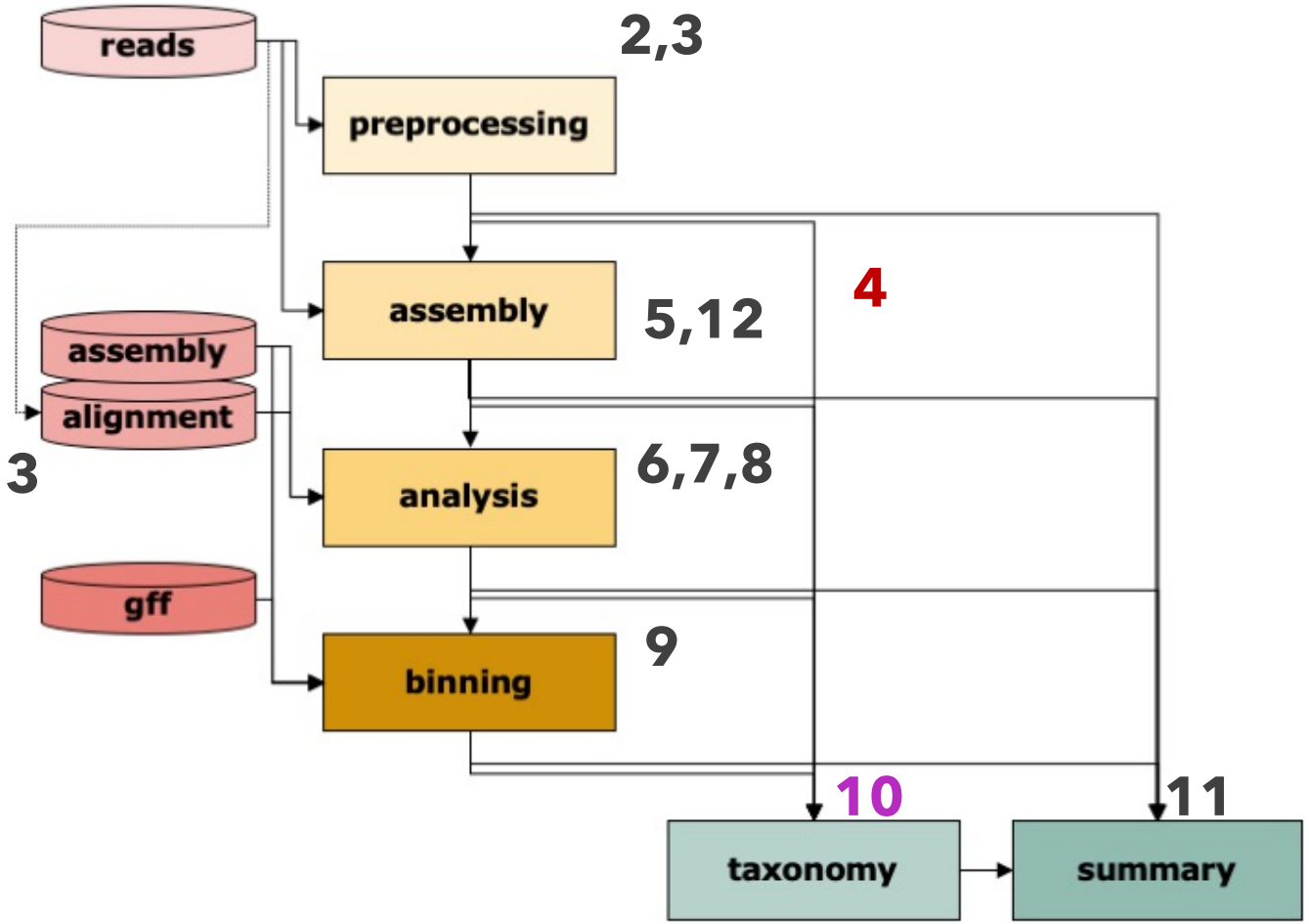
March 2022

University of Amsterdam
Life Sciences

Metagenomics (+ other omics) pipeline



Metagenomics (+ other omics) pipeline



Overview

- considerations: aims and requirements
- taxonomy

- methods:
  - reminder read mapping
  - non alignment based profilers

- benchmarking

Considerations

What do we want from taxonomy?

- accurate list of the taxa that are present
- accurate quantification of the present taxa
- robust semi-quantitative representation of present taxa

- accurate annotation of as many reads as possible
- accurate annotation of as many contigs as possible

What kind of taxa are we interested in?

- specific target taxa
- known organisms
- unknown organisms

- bacteria
- archaea
- eukaryotes – unicellular or multicellular
- viruses

What resolution do we need?

- strains, sub-species
- species-like
- genera
- dynamic (can be a mix of species – phyla per sample)

What kind of microbiome do we have?

- (mostly) known organisms
- likely many unknowns
- unusual taxonomies
- many non-prokaryotic sequences

What kind of data do we have?

- reads
  - metagenomics
  - metatranscriptomics

- assembled contigs
- genome reconstructions

How much computational resources can we use?

- quick or slow?
  - scaling how with the number of reads?
- large memory or not?
- much space for our databases or less?

What kind of taxonomy do we want?

- anything, as long it's got the names

- a community standard
- a custom taxonomy

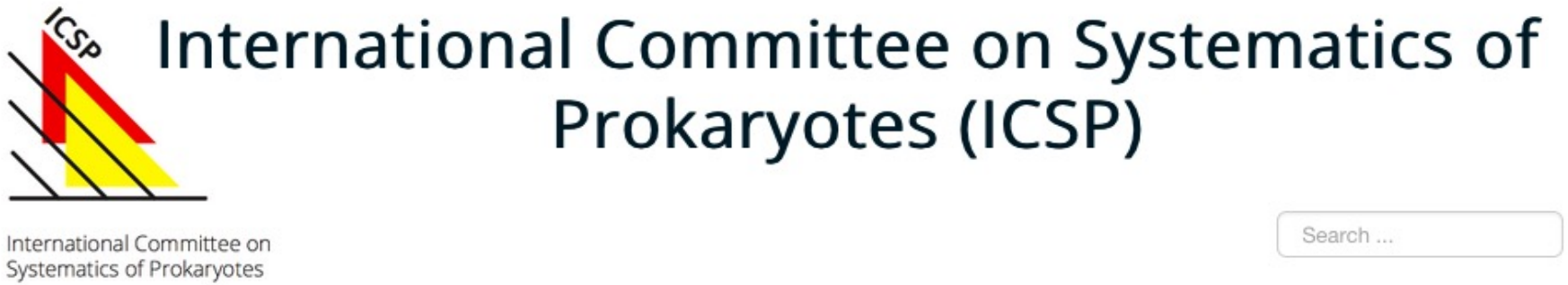
- we need to link it to other databases

What's a taxonomy, anyway?

"taxonomy, in a broad sense the science of classification, but more strictly the classification of living and extinct organisms—i.e., biological classification. The term is derived from the Greek *taxis* ("arrangement") and *nomos* ("law"). ~~Taxonomy is, therefore, the methodology and principles of systematic botany and zoology and sets up arrangements of the kinds of plants and animals in hierarchies of superior and subordinate groups. Among biologists the Linnaean system of binomial nomenclature, created by Swedish naturalist Carolus Linnaeus in the 1750s, is internationally accepted."~~

Cain, A.J.. "taxonomy". Encyclopedia Britannica

Who makes the taxonomy?



Who makes the taxonomy?

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

2019, volume 69, issue 1A, pages S1–S111

MICROBIOLOGY SOCIETY

International Code of Nomenclature of Prokaryotes

Prokaryotic Code (2008 Revision)

Charles T. Parker[1], Brian J. Tindall[2] and George M. Garrity[3] (Editors)

General Consideration 1

The progress of bacteriology can be furthered by a precise system of nomenclature accepted by the majority of bacteriologists of all nations.

General Consideration 4

Rules of nomenclature do not govern the delimitation of taxa nor determine their relations. The Rules are primarily for assessing the correctness of the names applied to defined taxa; they also prescribe the procedures for creating and proposing new names.

Why are there several taxonomies?

Published online 14 September 2021 Nucleic Acids Research, 2022, Vol. 50, Database issue D785–D794
https://doi.org/10.1093/nar/gkab776

GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy

Donovan H. Parks*, Maria Chuvochina, Christian Rinke, Aaron J. Mussig, Pierre-Alain Chaumeil and Philip Hugenholtz*

D590–D596 Nucleic Acids Research, 2013, Vol. 41, Database issue Published online 28 November 2012
doi:10.1093/nar/gks1219

The SILVA ribosomal RNA gene database project: improved data processing and web-based tools

Christian Quast[1], Elmar Pruesse[1,2], Pelin Yilmaz[1], Jan Gerken[1,2], Timmy Schweer[1], Pablo Yarza[3], Jörg Peplies[3] and Frank Oliver Glöckner[1,2,*]

DATABASE
The Journal of Biological Databases and Curation
Database, 2020, 1–21
doi: 10.1093/database/baaa062
Review

Review

NCBI Taxonomy: a comprehensive update on curation, resources and tools

Conrad L. Schoch*, Stacy Ciufo, Mikhail Domrachev, Carol L. Hotton, Sivakumar Kannan, Rogneda Khovanskaya, Detlef Leipe, Richard Mcveigh, Kathleen O'Neill, Barbara Robbertse, Shobha Sharma, Vladimir Soussov, John P. Sullivan, Lu Sun, Seán Turner and Ilene Karsch-Mizrachi

LPSN .dsmz.de

Search taxonomy

Browse by rank

LPSN - List of Prokaryotic names with Standing in Nomenclature

Why are there several taxonomies?

- need to include un-cultured taxa / taxa without specimen cultures

- drive to use more phylogenetic information
- phylogenies based on 16S rRNA or whole genomes

- split taxa based on phylogenetic information or based on literature weight?

Taxonomy, phylogeny, and un-cultured organisms

nature biotechnology

RESOURCE

https://doi.org/10.1038/s41587-020-0501-8

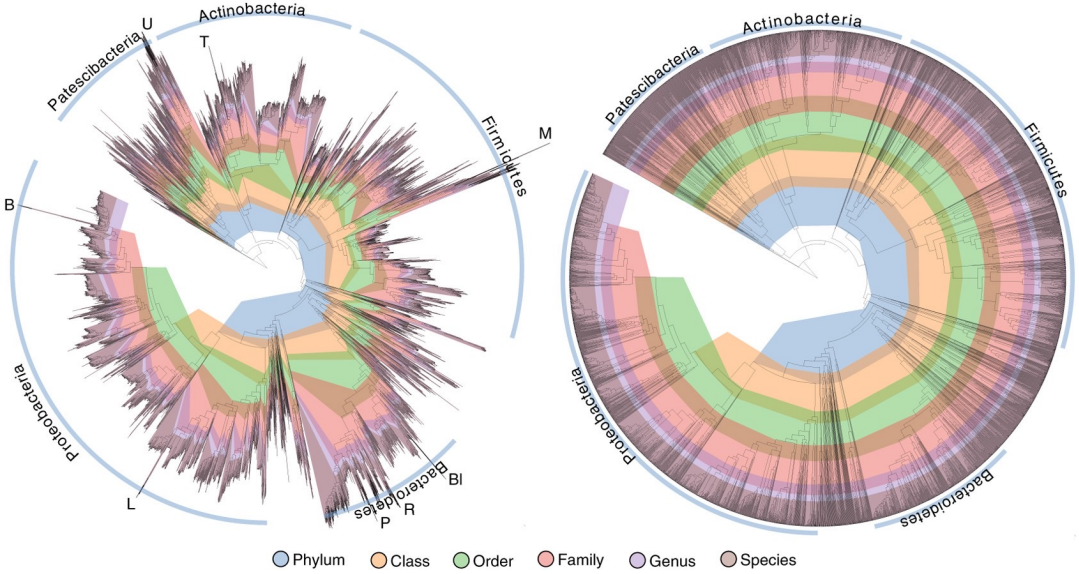
Check for updates

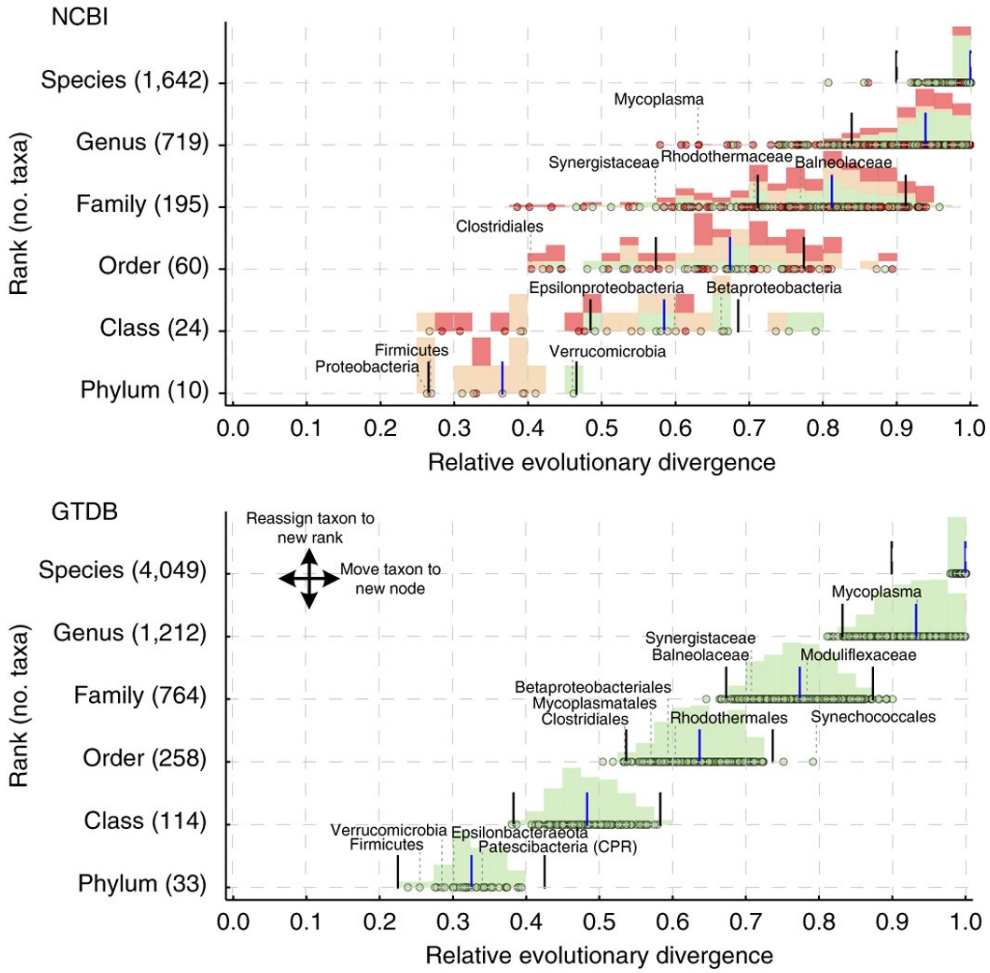
A complete domain-to-species taxonomy for Bacteria and Archaea

Donovan H. Parks ✉, Maria Chuvochina, Pierre-Alain Chaumeil, Christian Rinke, Aaron J. Mussig and Philip Hugenholtz

GTDB

Taxonomy, phylogeny, and un-cultured organisms





GTDB

Parks et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36, 996–1004 (2018)

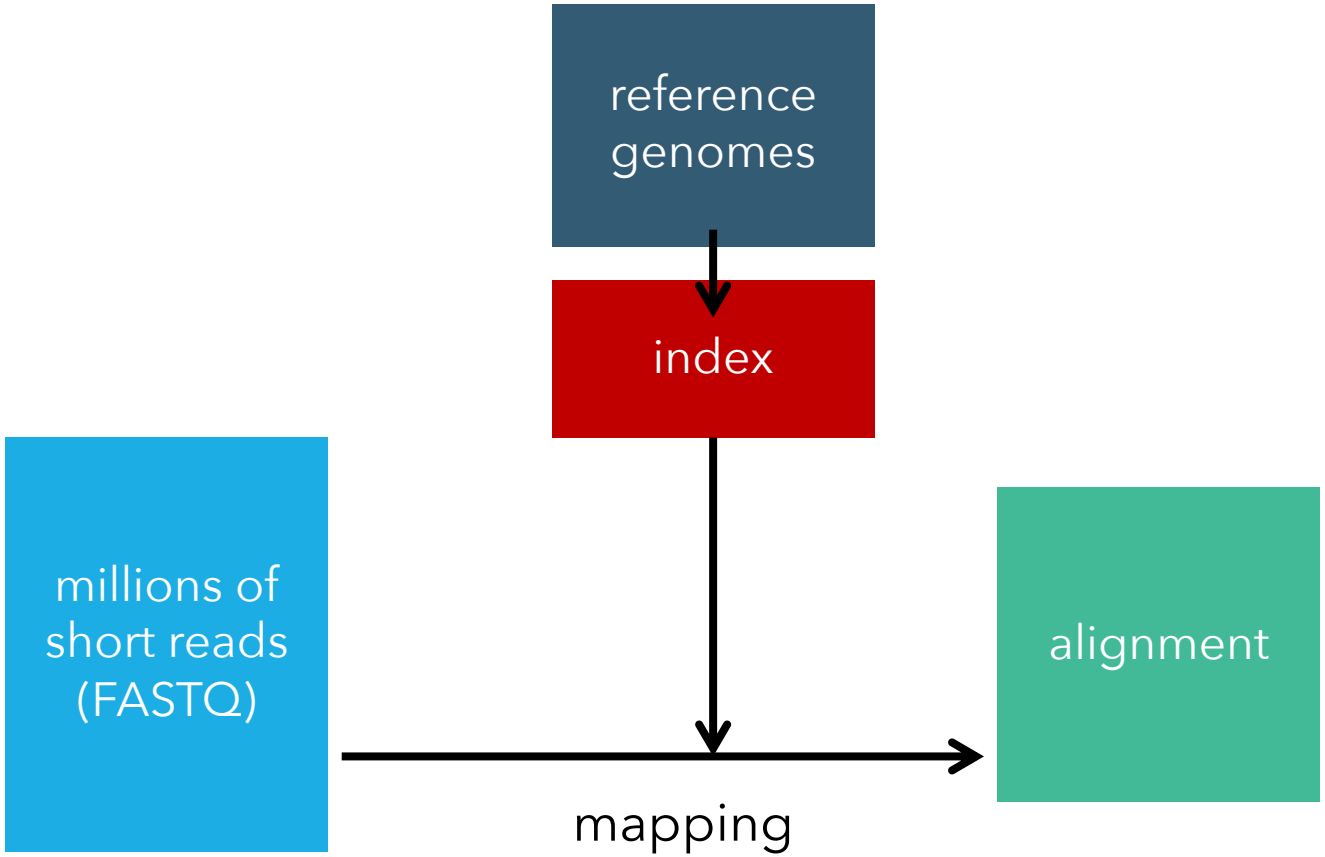
Taxonomy, phylogeny, and un-cultured organisms

Methods

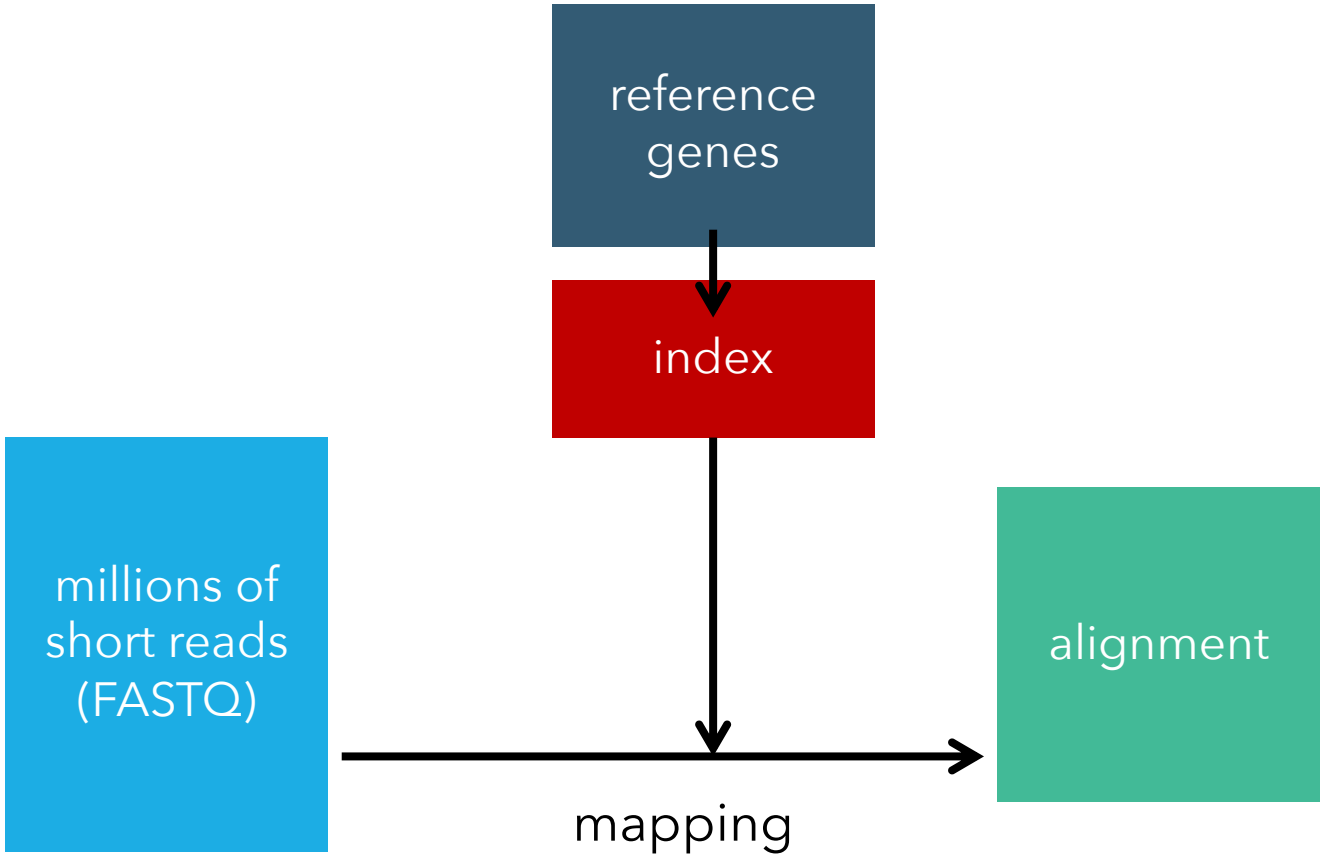
General profiling approaches

- align reads to taxonomically recognized genomes
- align reads to taxonomically annotated genes
- align reads to phylogenetic/specific marker regions
- match reads to taxonomically recognized genomes

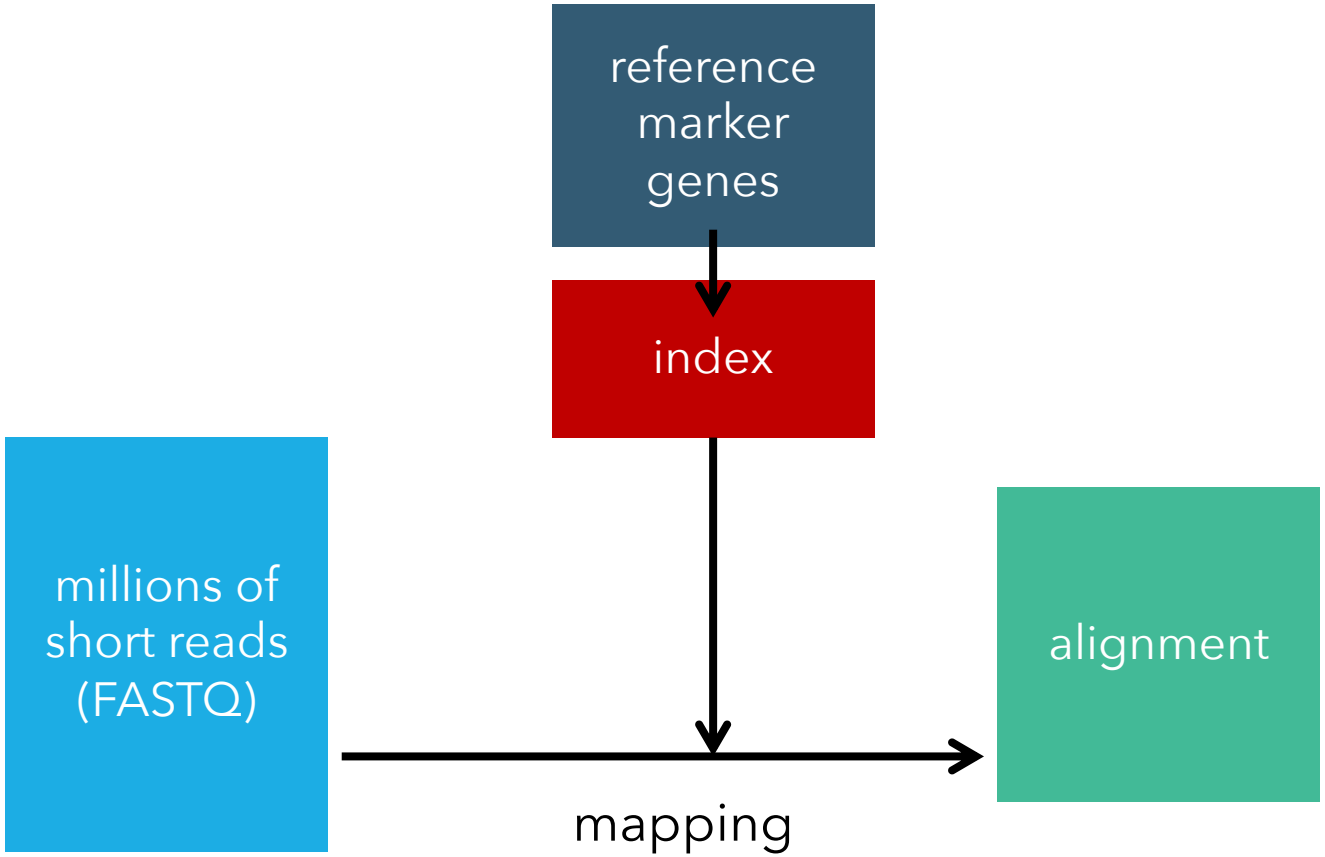
Taxonomic profilers – 1



Taxonomic profilers – 2

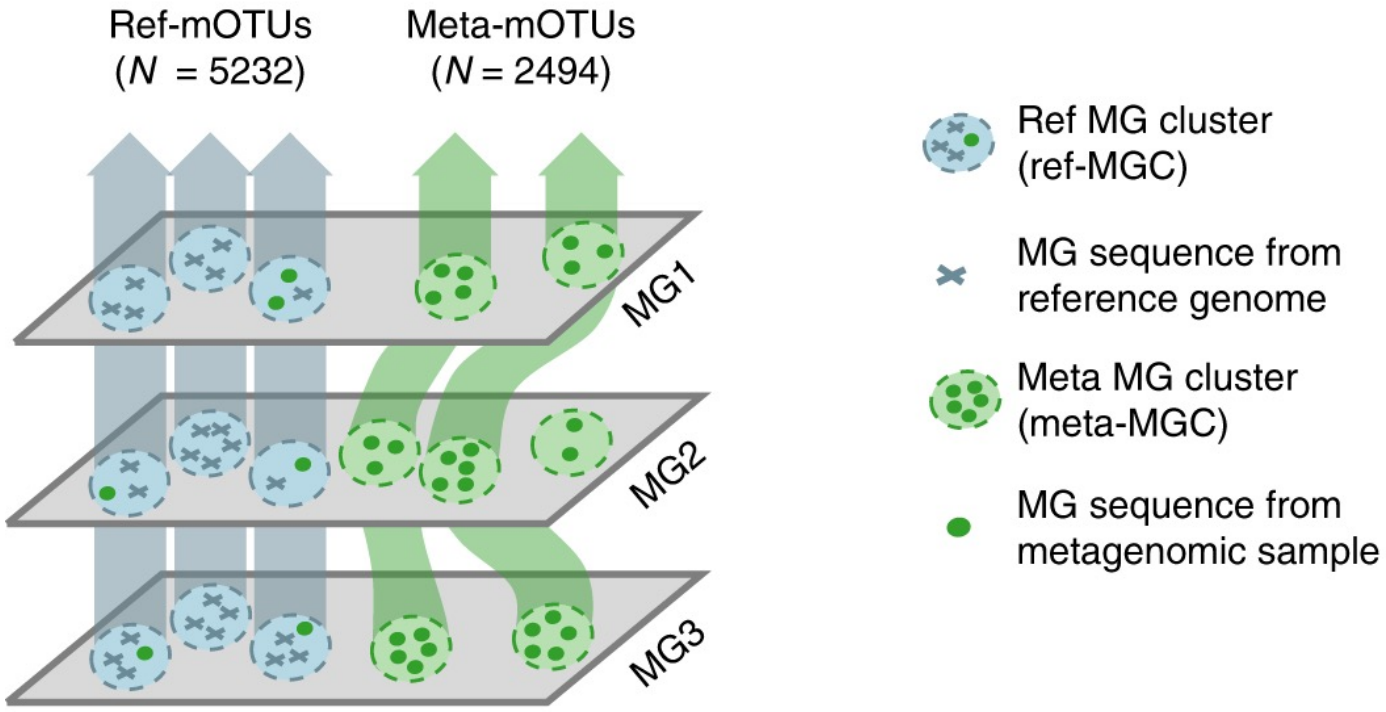


Taxonomic profilers – 3



Taxonomic profilers – 3: marker gene/region approach

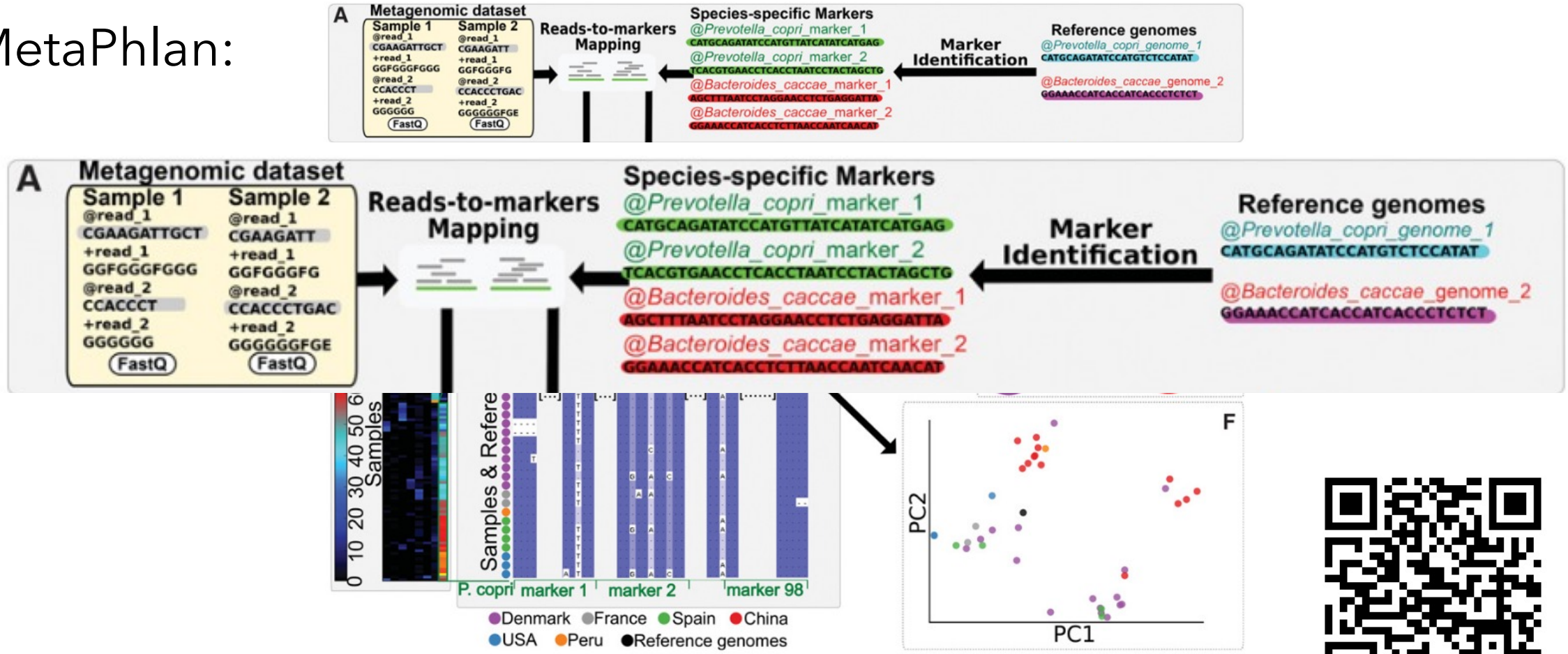
mOTUs:



Milanese, A., Mende, D.R., Paoli, L. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun 10, 1014 (2019).

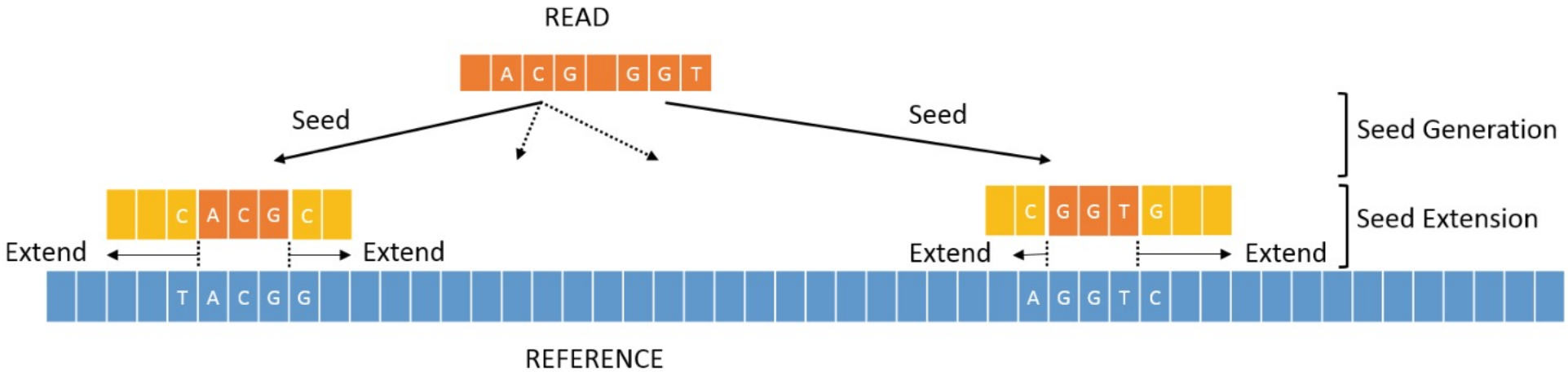
Taxonomic profilers – 3: marker gene/region approach

MetaPhlan:



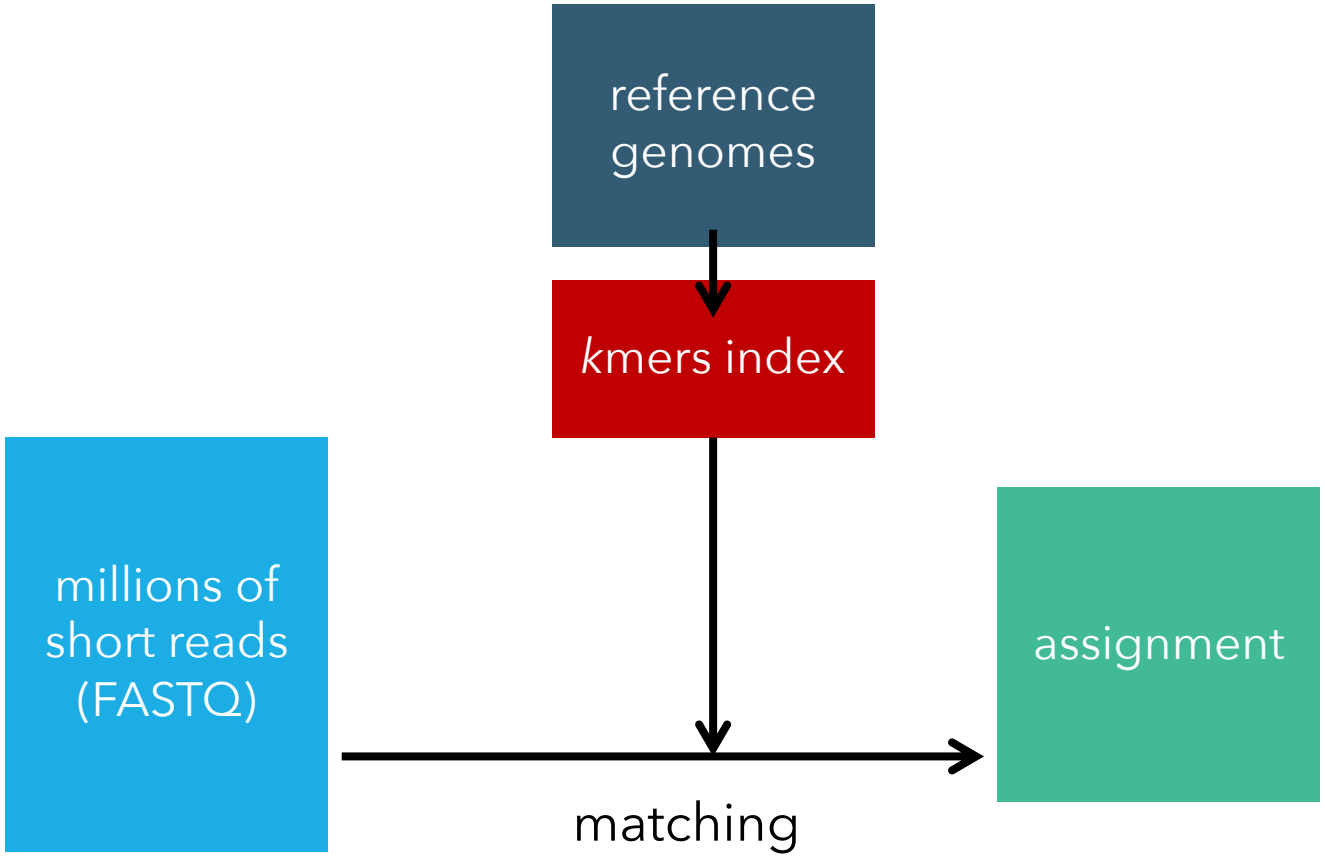
Truong DT, .., Huttenhower C & Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Research 27:626-638 (2017)

Mapping reads = aligning reads



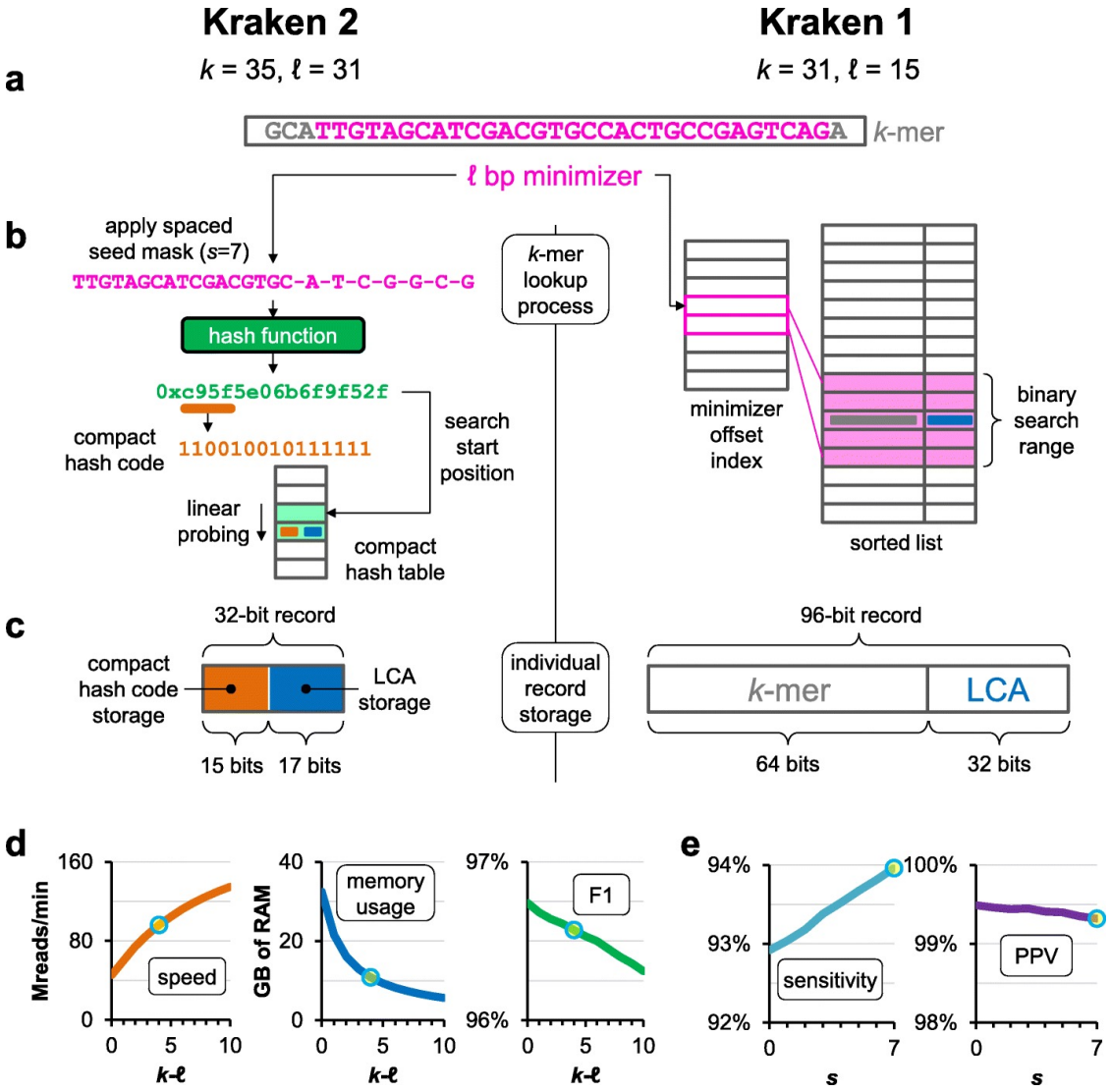
Houtgast et al. 2016. ARCS 2016

Taxonomic profilers – 4



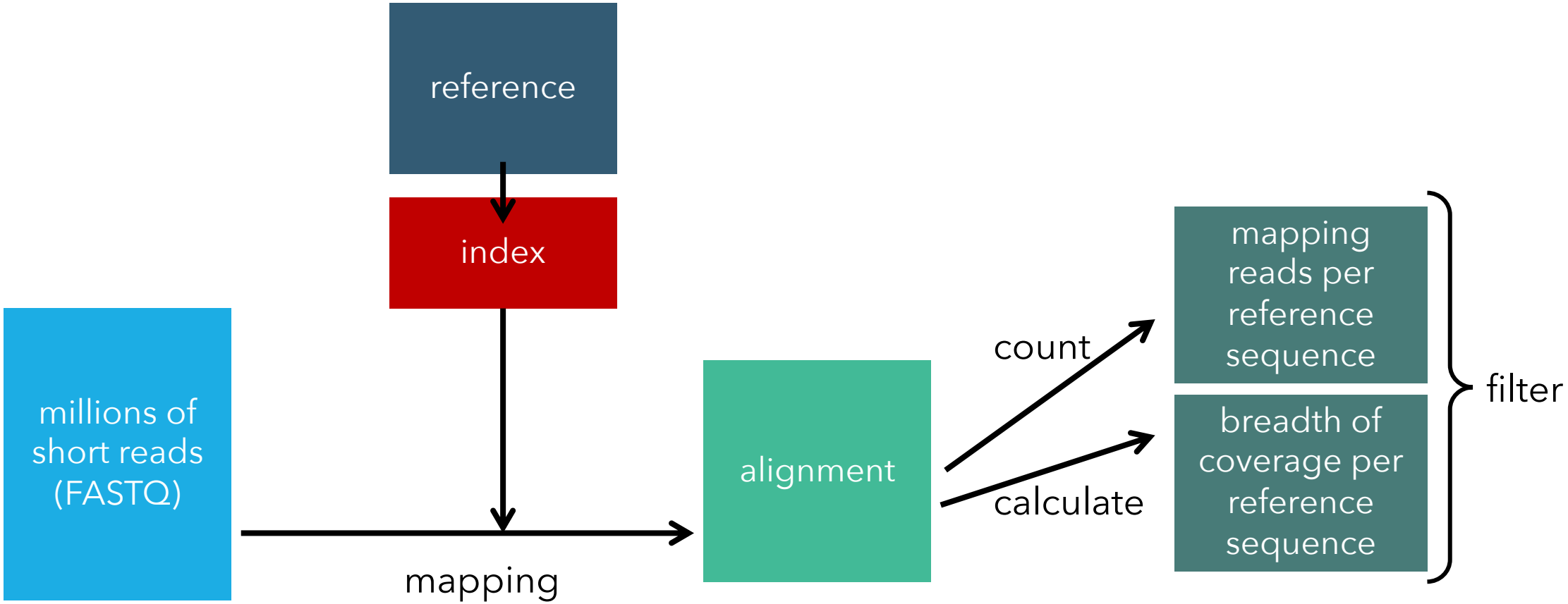
Taxonomic profilers – 4

kraken2:

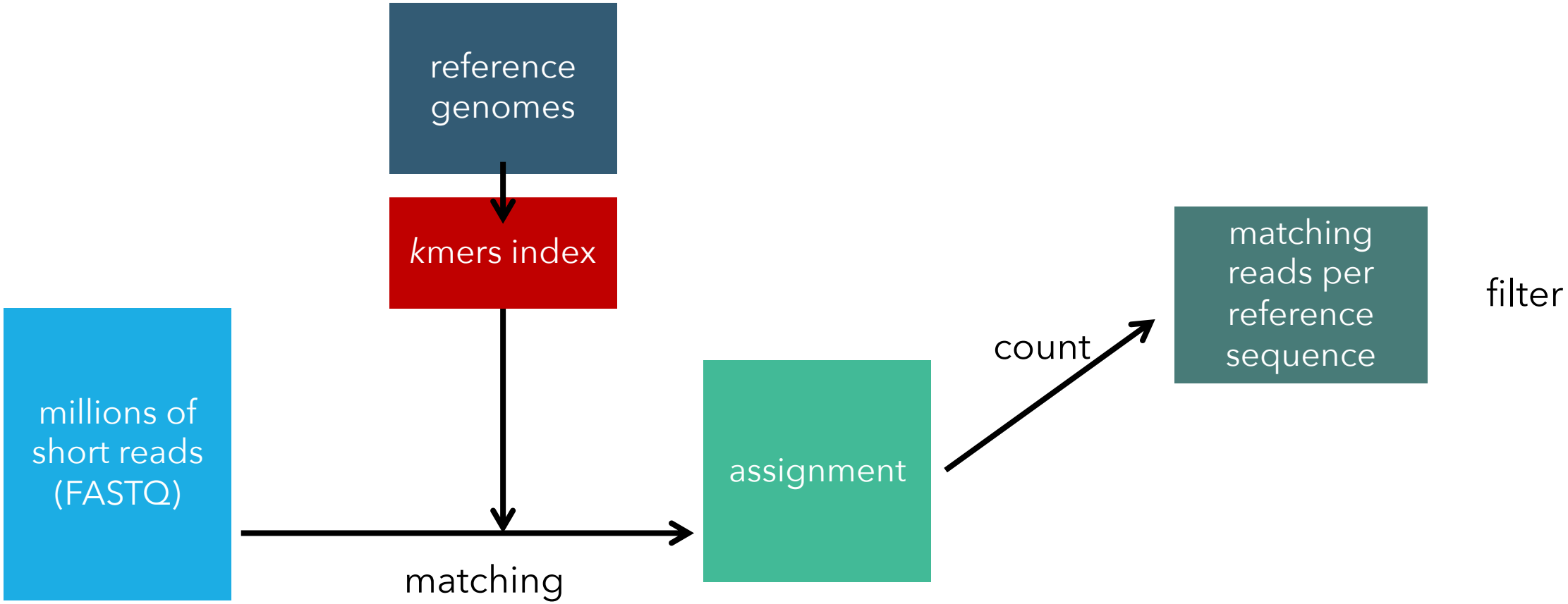


Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257 (2019)

Mapping-based taxonomic profilers: quatification & specificity filtering



*k*mer-based taxonomic profilers : quatification & specificity filtering



How to choose a profiler? Benchmarks

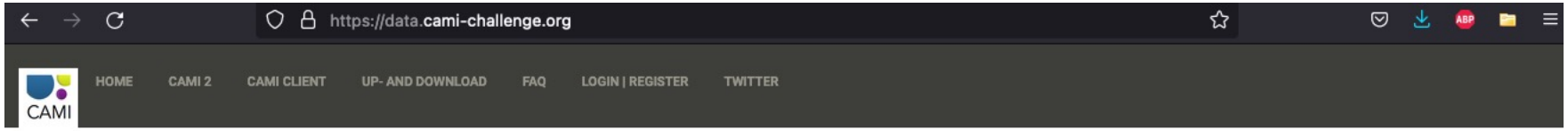
ANALYSIS

OPEN

Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software

Alexander Sczyrba[1,2,48], Peter Hofmann[3–5,48], Peter Belmann[1,2,4,5,48], David Koslicki[6], Stefan Janssen[4,7,8], Johannes Dröge[3–5], Ivan Gregor[3–5], Stephan Majda[3,47], Jessika Fiedler[3,4], Eik Dahms[3–5], Andreas Bremges[1,2,4,5,9], Adrian Fritz[4,5], Ruben Garrido-Oter[3–5,10,11], Tue Sparholt Jørgensen[12–14], Nicole Shapiro[15], Philip D Blood[16], Alexey Gurevich[17], Yang Bai[10,47], Dmitrij Turaev[18], Matthew Z DeMaere[19], Rayan Chikhi[20,21], Niranjan Nagarajan[22], Christopher Quince[23], Fernando Meyer[4,5], Monika Balvočiūtė[24], Lars Hestbjerg Hansen[12], Søren J Sørensen[13], Burton K H Chia[22], Bertrand Denis[22], Jeff L Froula[15], Zhong Wang[15], Robert Egan[15], Dongwan Don Kang[15], Jeffrey J Cook[25], Charles Deltel[26,27], Michael Beckstette[28], Claire Lemaitre[26,27], Pierre Peterlongo[26,27], Guillaume Rizk[27,29], Dominique Lavenier[21,27], Yu-Wei Wu[30,31], Steven W Singer[30,32], Chirag Jain[33], Marc Strous[34], Heiner Klingenberg[35], Peter Meinicke[35], Michael D Barton[15], Thomas Lingner[36], Hsin-Hung Lin[37], Yu-Chieh Liao[37], Genivaldo Gueiros Z Silva[38], Daniel A Cuevas[38], Robert A Edwards[38], Surya Saha[39], Vitor C Piro[40,41], Bernhard Y Renard[40], Mihai Pop[42,43], Hans-Peter Klenk[44], Markus Göker[45], Nikos C Kyrpides[15], Tanja Woyke[15], Julia A Vorholt[46], Paul Schulze-Lefert[10,11], Edward M Rubin[15], Aaron E Darling[19], Thomas Rattei[18] & Alice C McHardy[3–5,11]

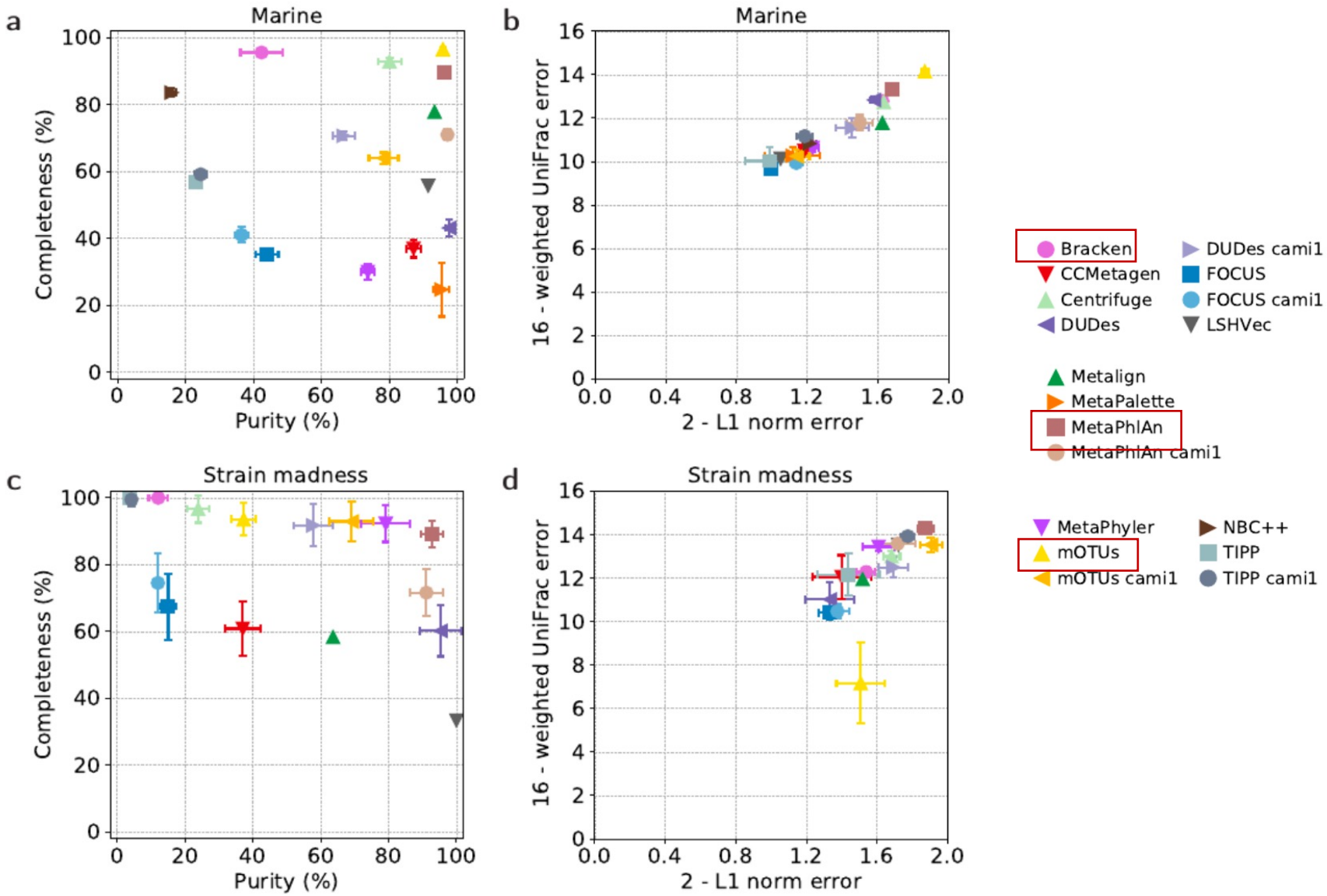
How to choose a profiler? Benchmarks



CAMI

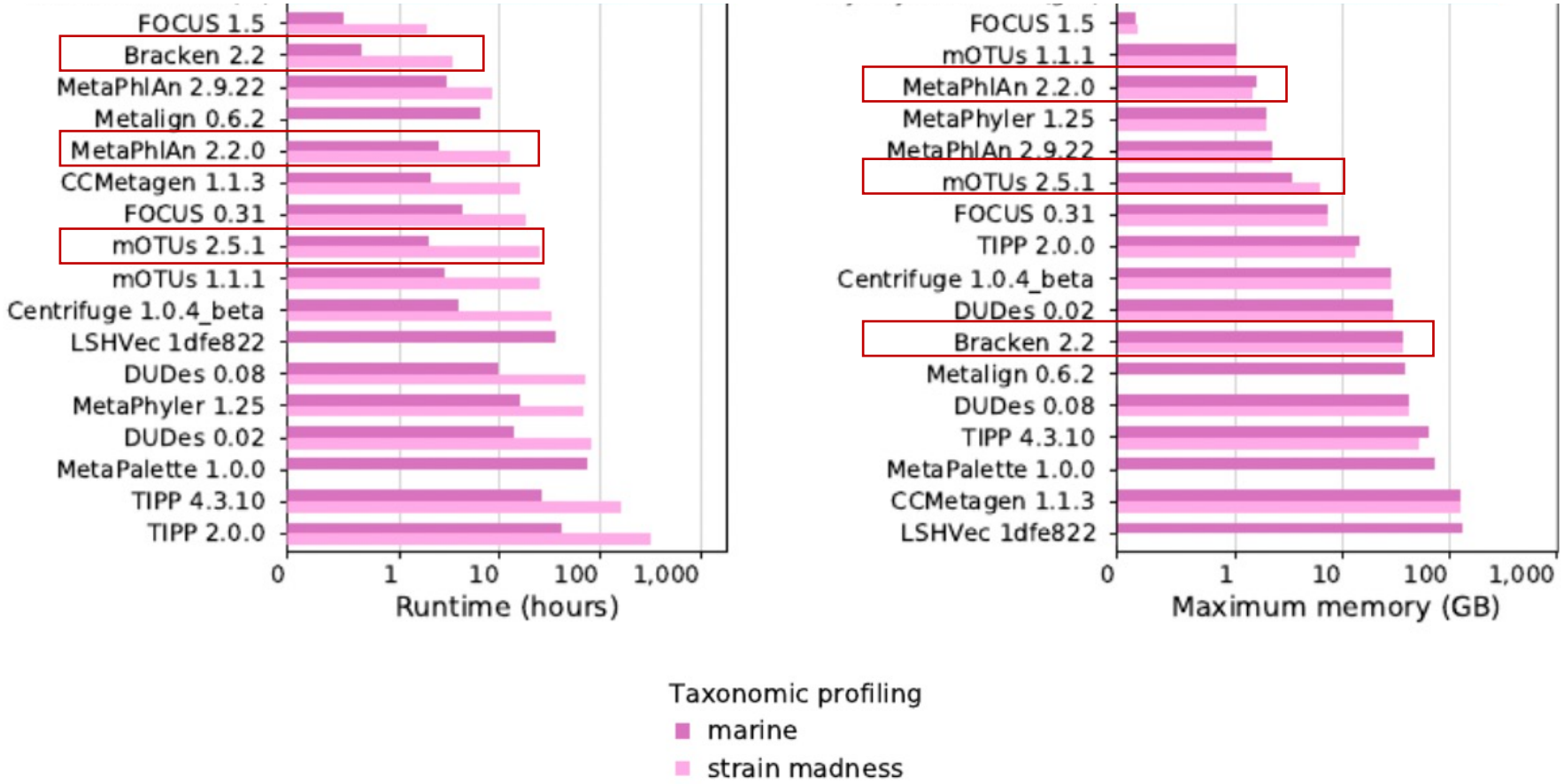
Critical Assessment of Metagenome Interpretation

How to choose a profiler? Benchmarks



F. Meyer, ..., A. C. McHardy: Critical Assessment of Metagenome Interpretation - the second round of challenges bioRxiv 2021.07.12.451567

How to choose a profiler? Benchmarks



F. Meyer, ..., A. C. McHardy: Critical Assessment of Metagenome Interpretation - the second round of challenges bioRxiv 2021.07.12.451567

Thanks for your attention!

a.u.s.heintzbuschart@uva.nl

SP C2.205

github.com/a-h-b

twitter.com/_a_h_b_

University of Amsterdam
Life Sciences