Microbiome: general aspects

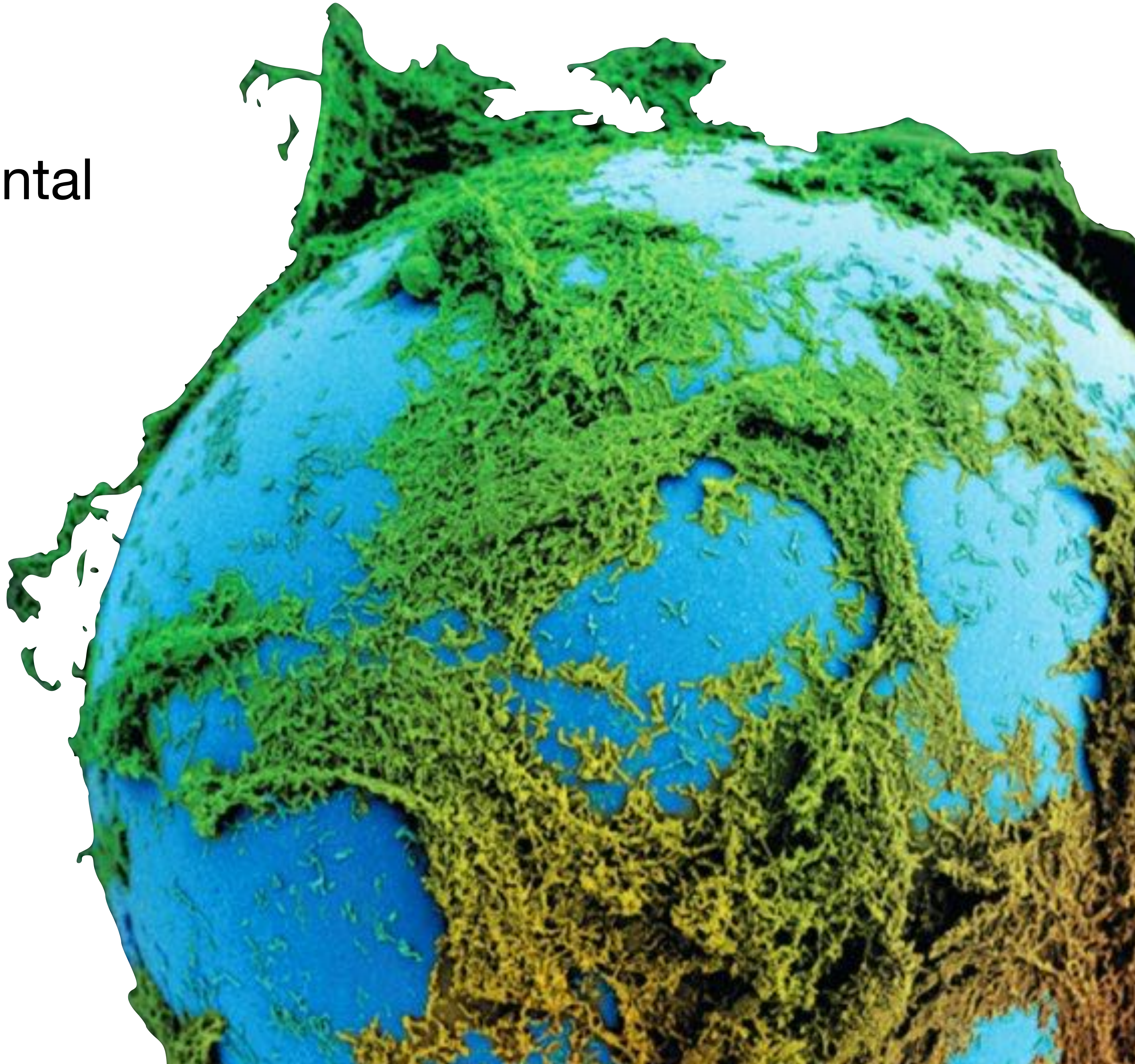
Meeting Toscano di Bioinformatica e Systems Biology

Bioinformatiha 11

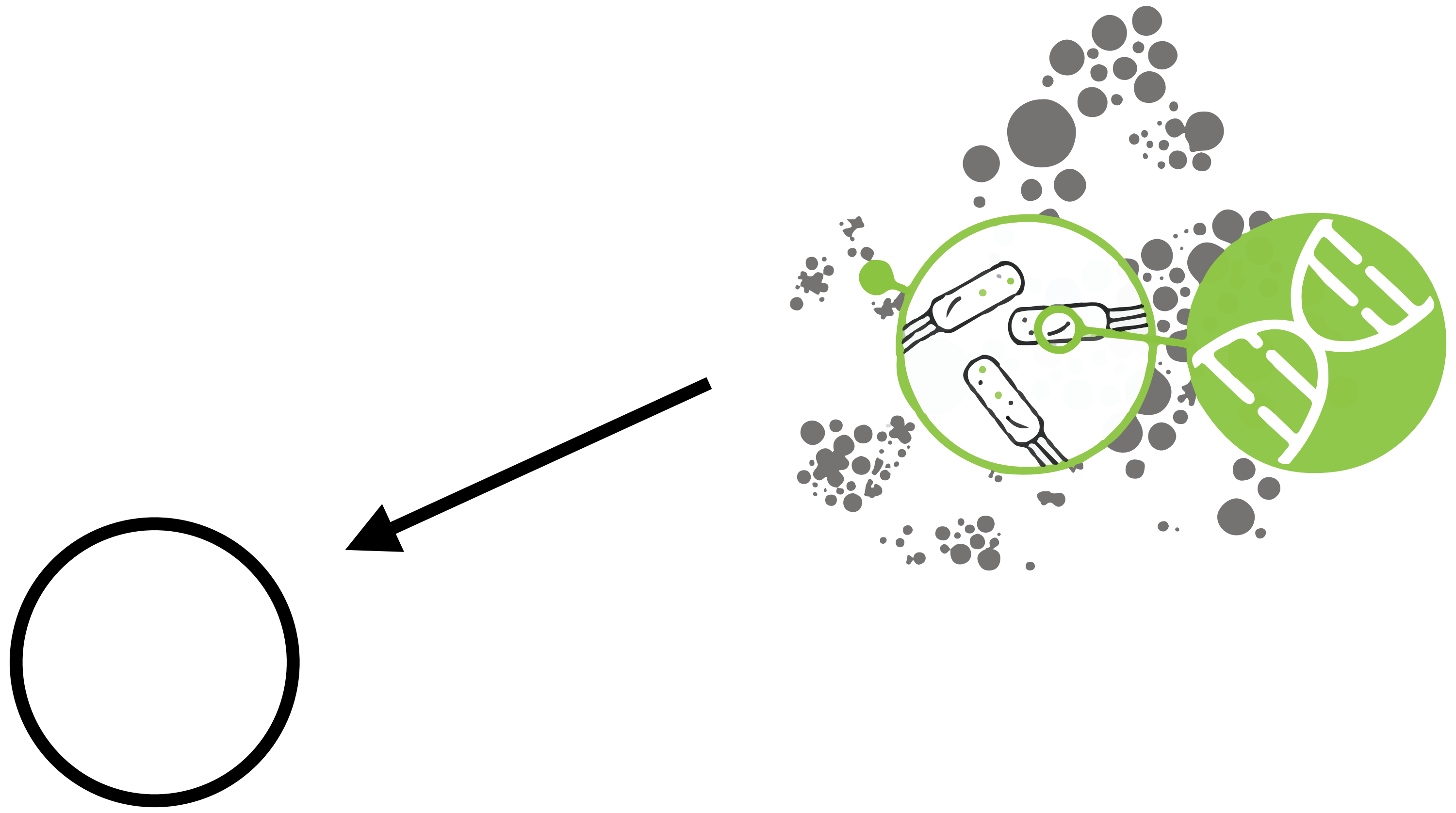
Giovanni Bacci - giovanni.bacci@unifi.it

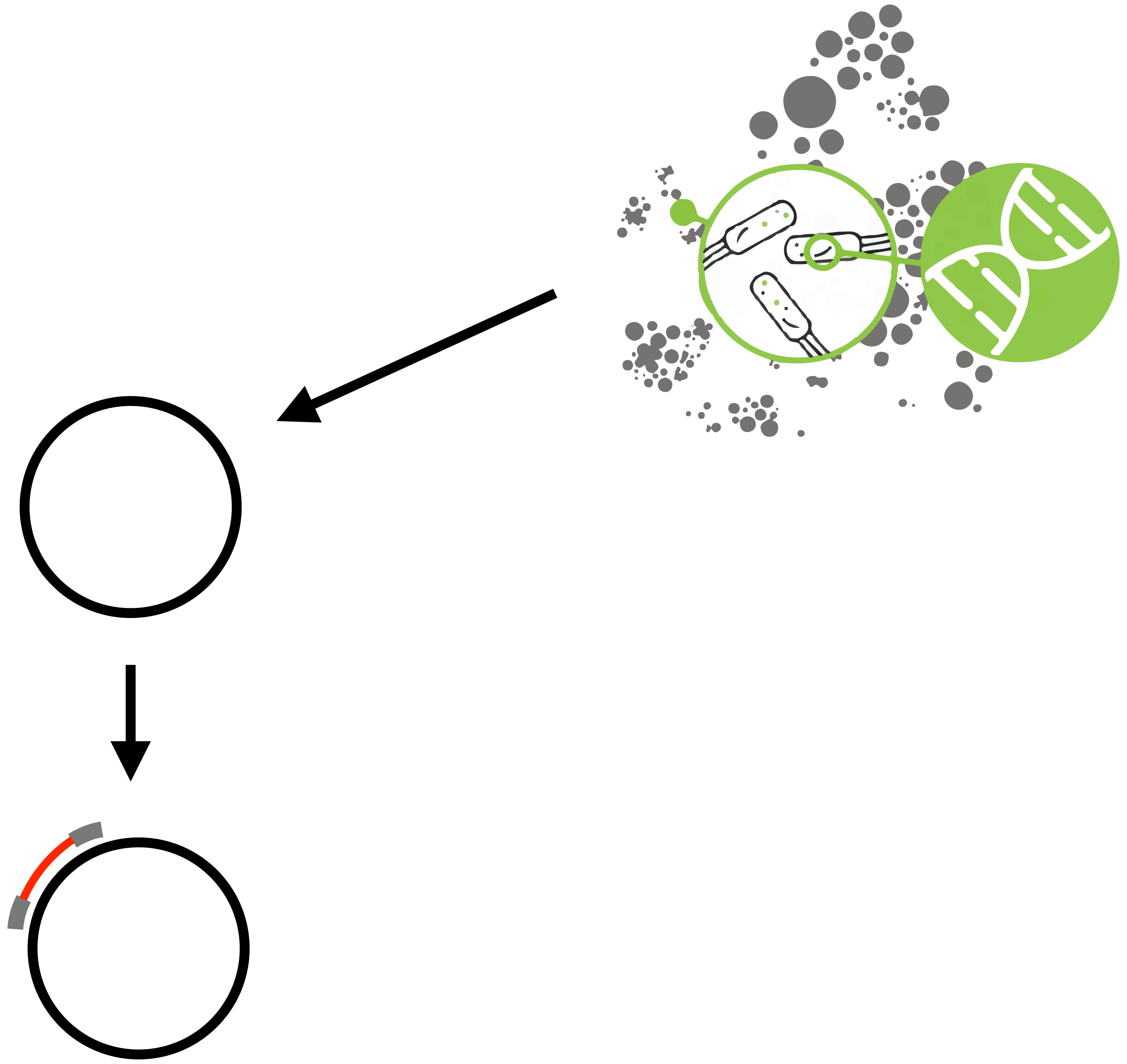
Microbes are everywhere

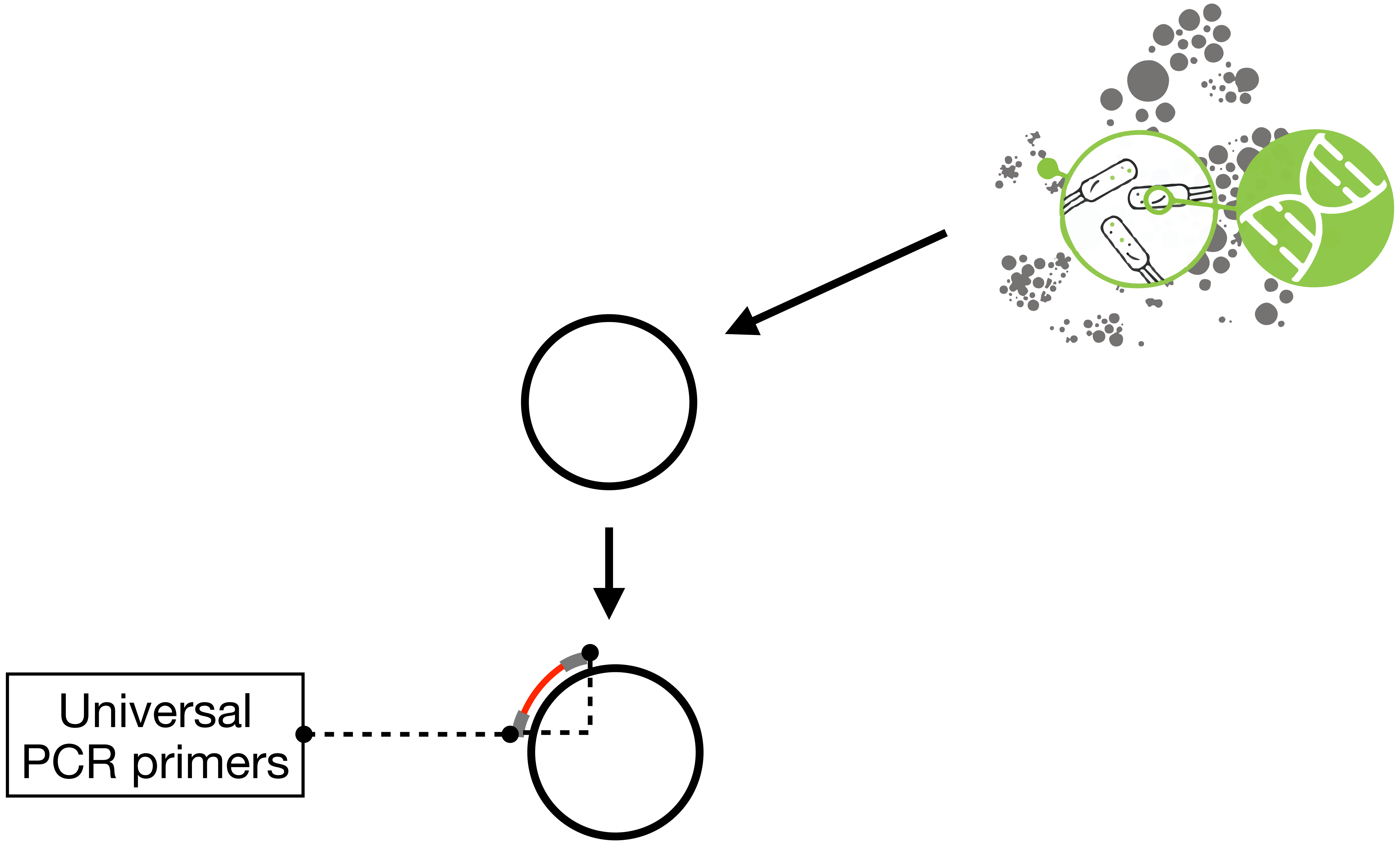
- They thrive in every known environmental niche
- Animals, plants and other eukaryotic organisms are no exception
- The association between microbial and eukaryotic organisms is called holobiont and it represents the unit of selection for evolution



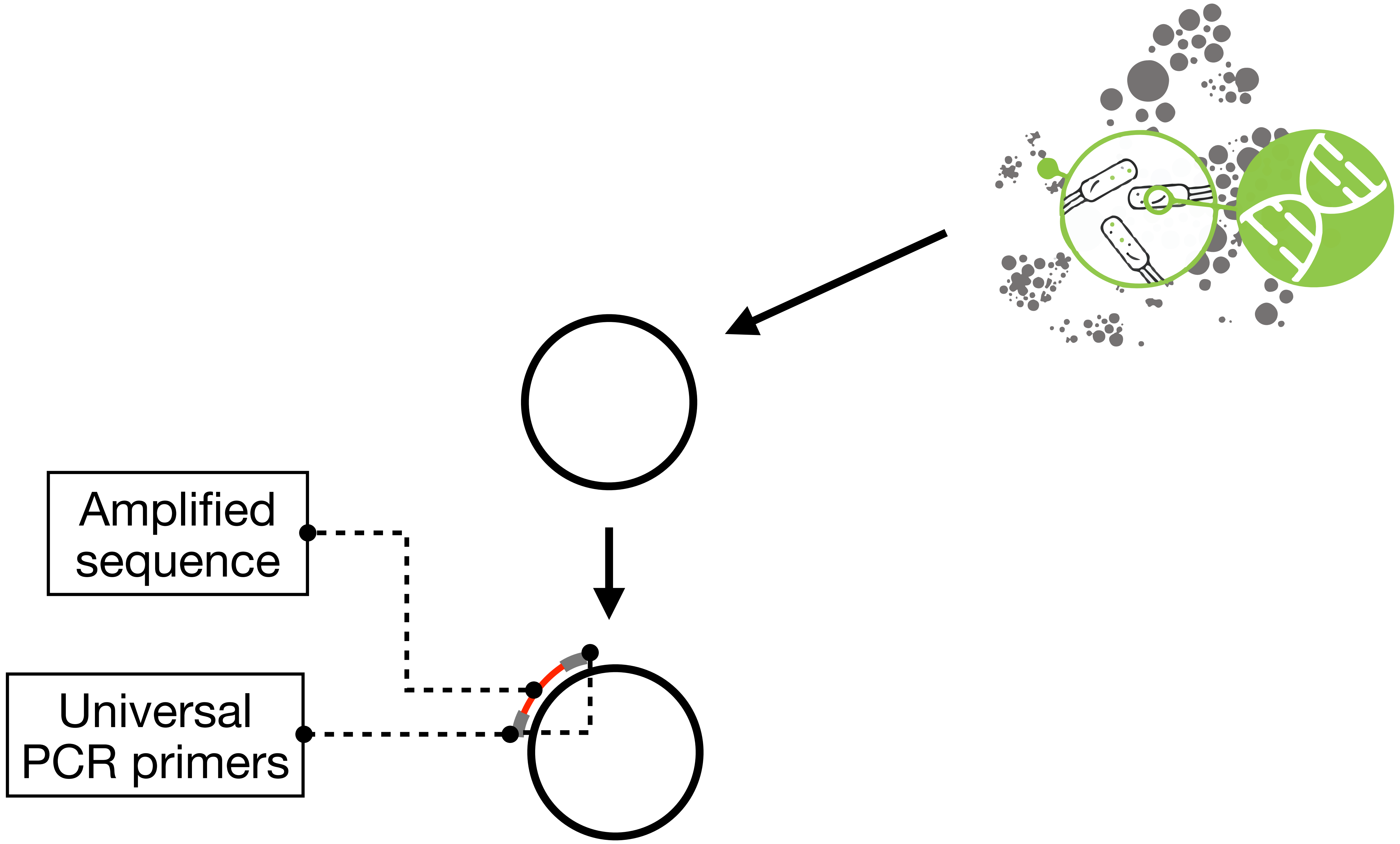
Profiling





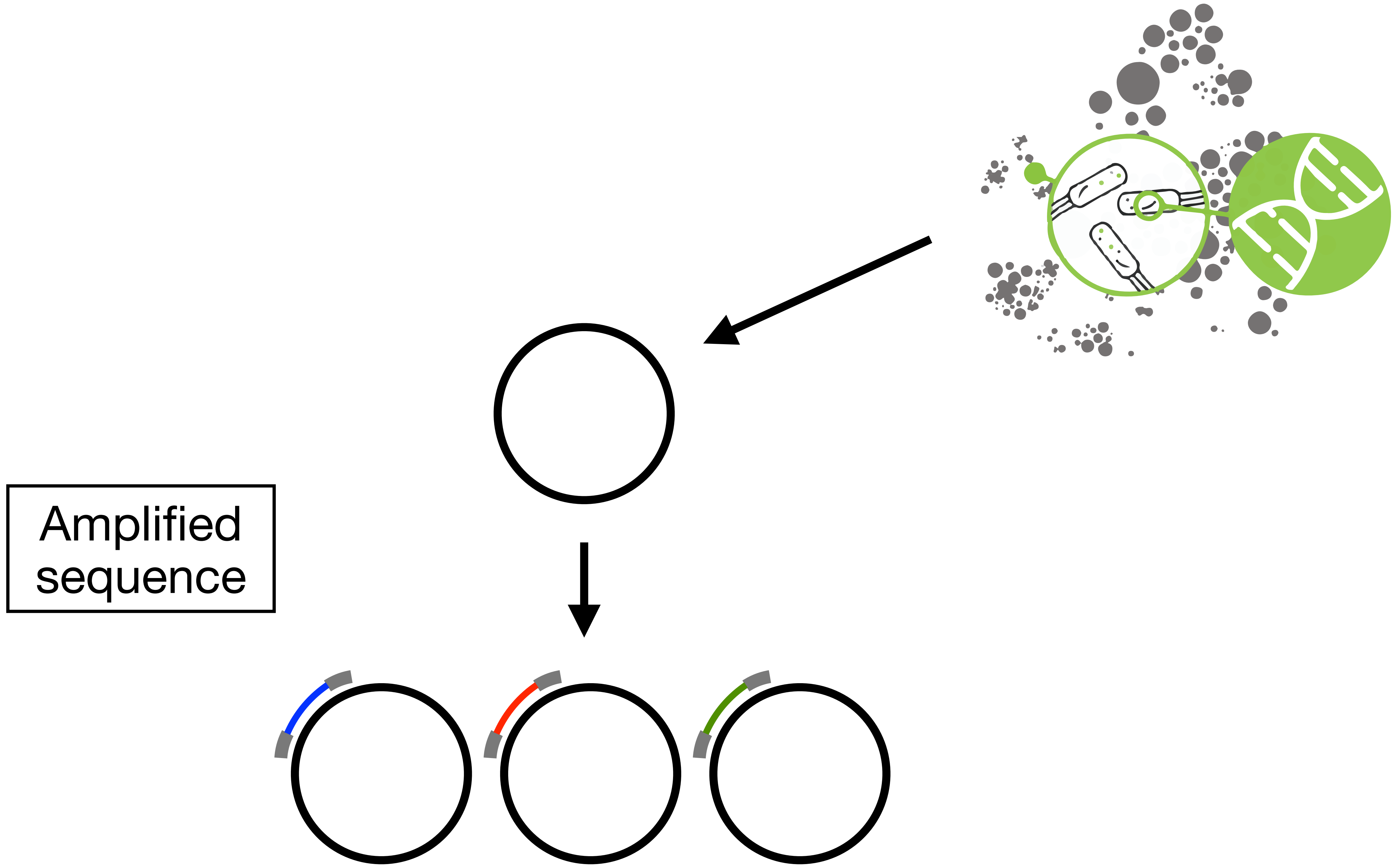


Universal
PCR primers

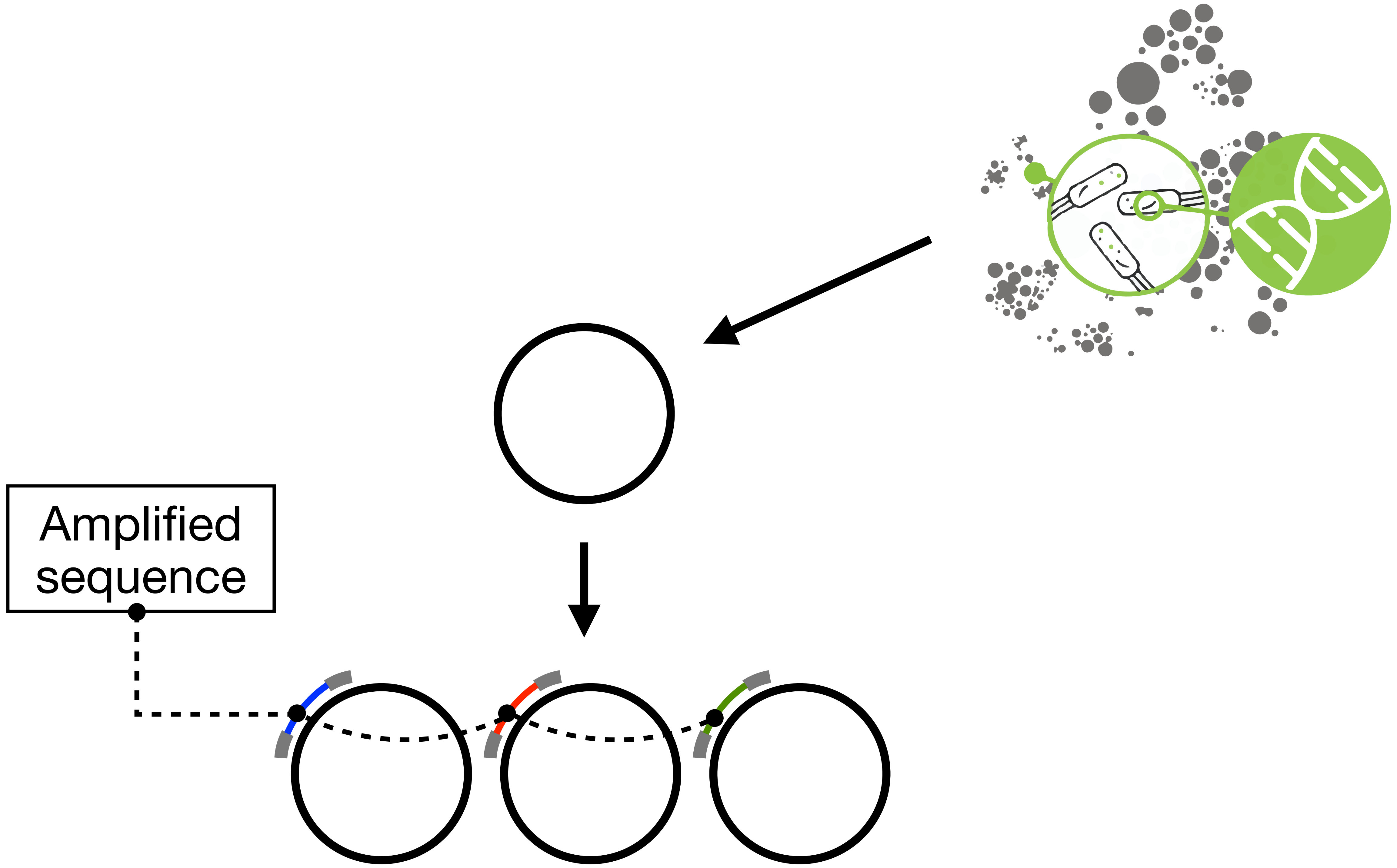


Amplified sequence

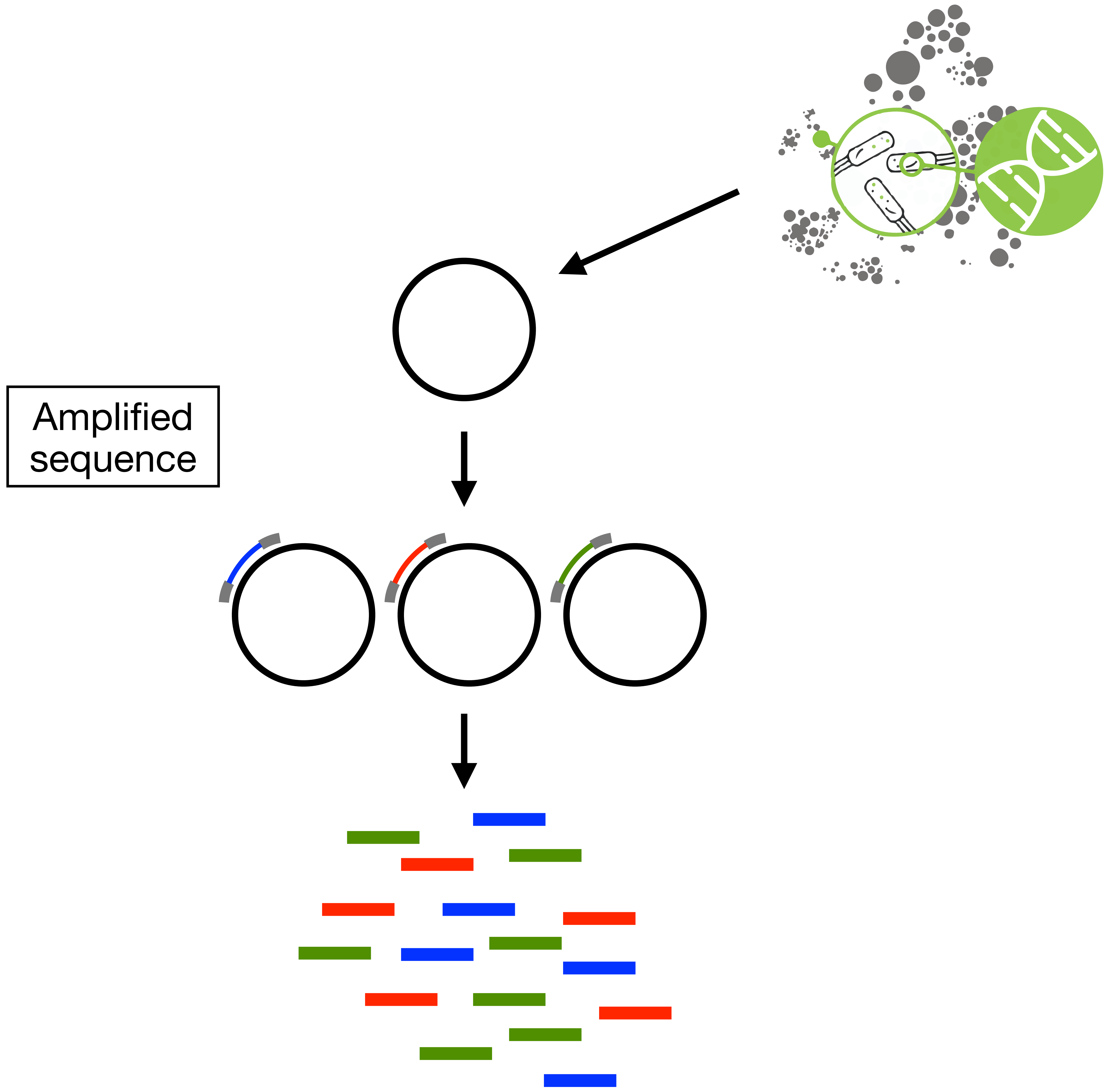
Universal PCR primers



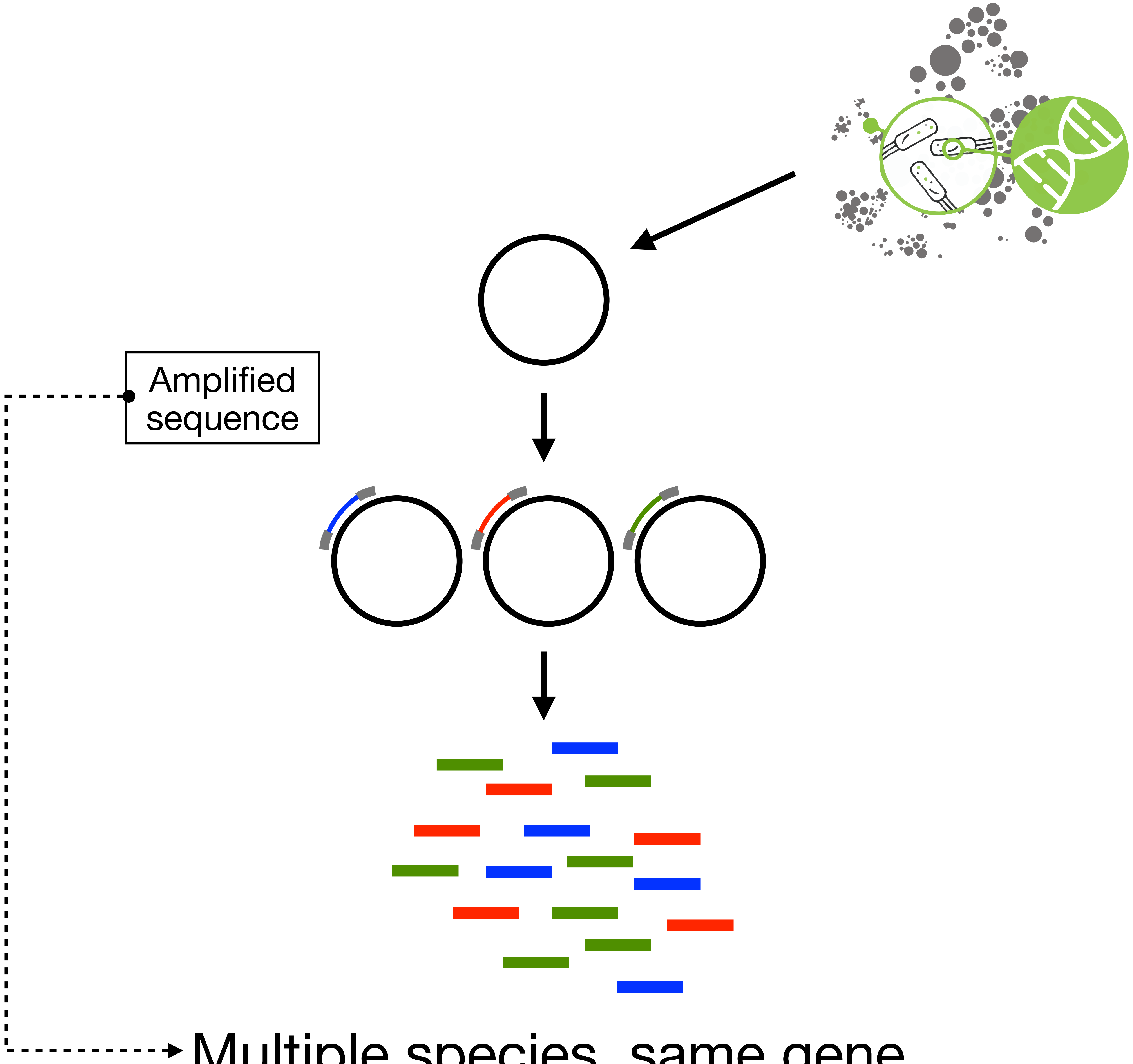
Amplified sequence



Amplified sequence



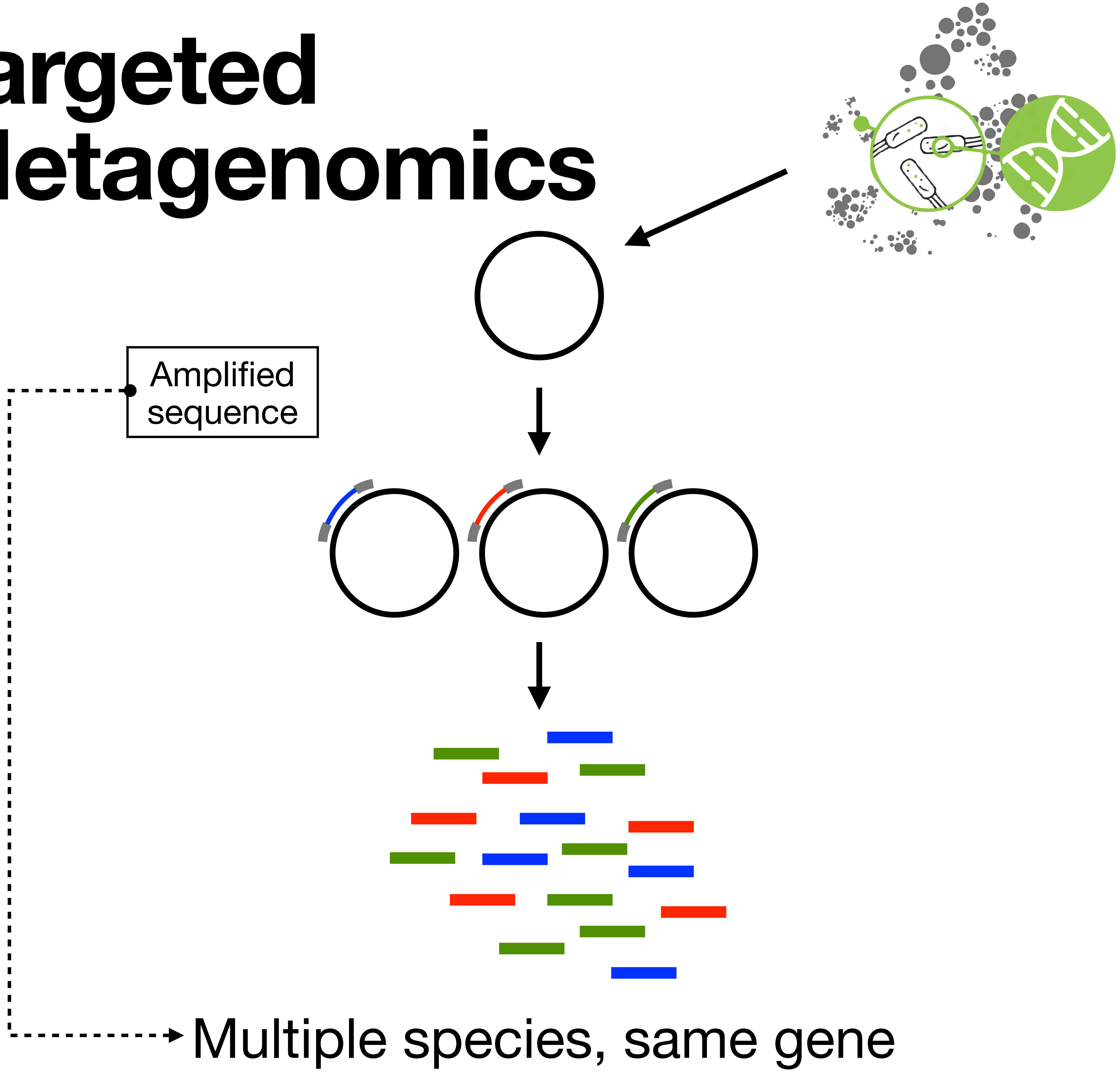
Amplified sequence



Amplified sequence

Multiple species, same gene

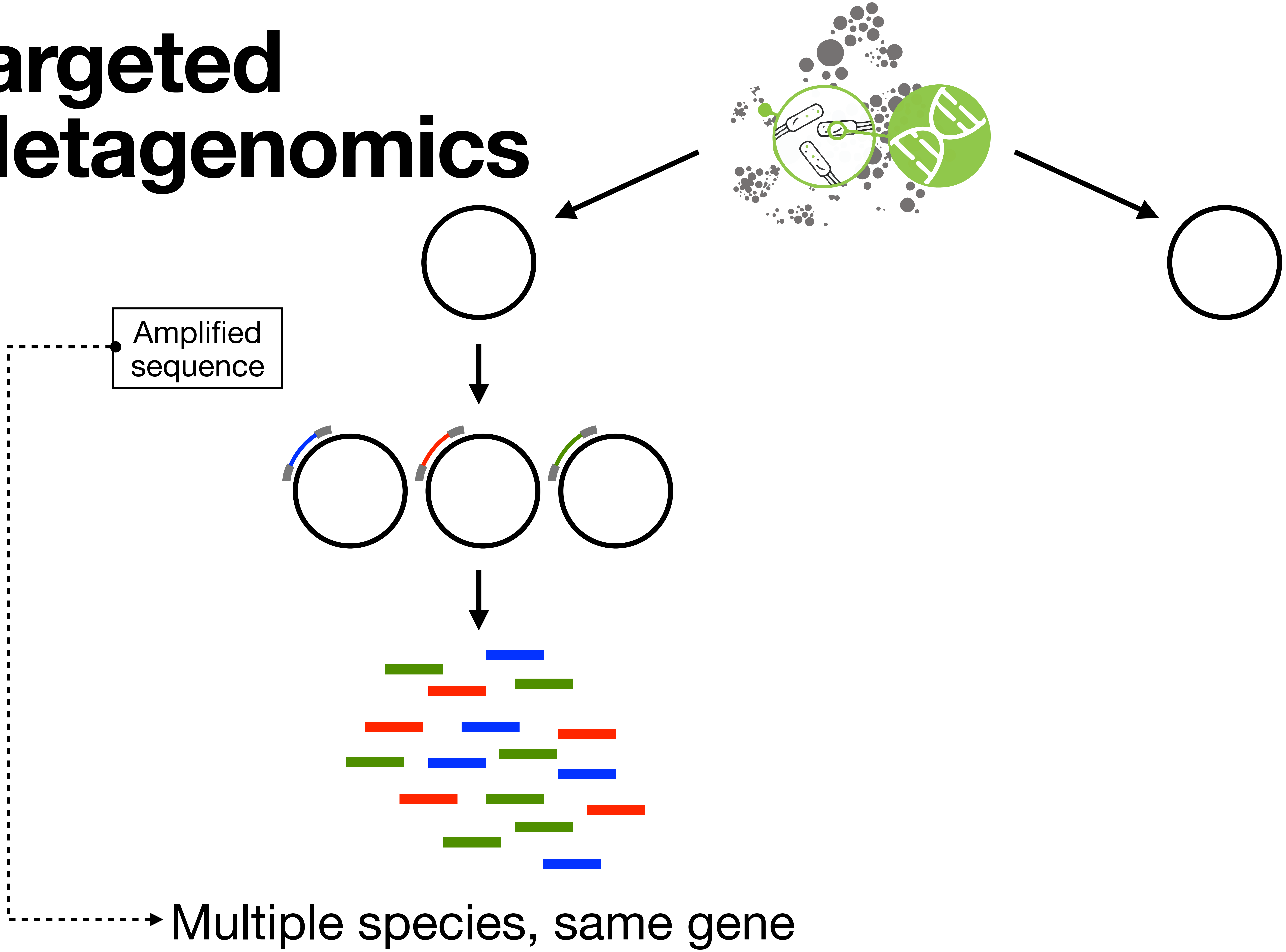
Targeted Metagenomics



Amplified sequence

Multiple species, same gene

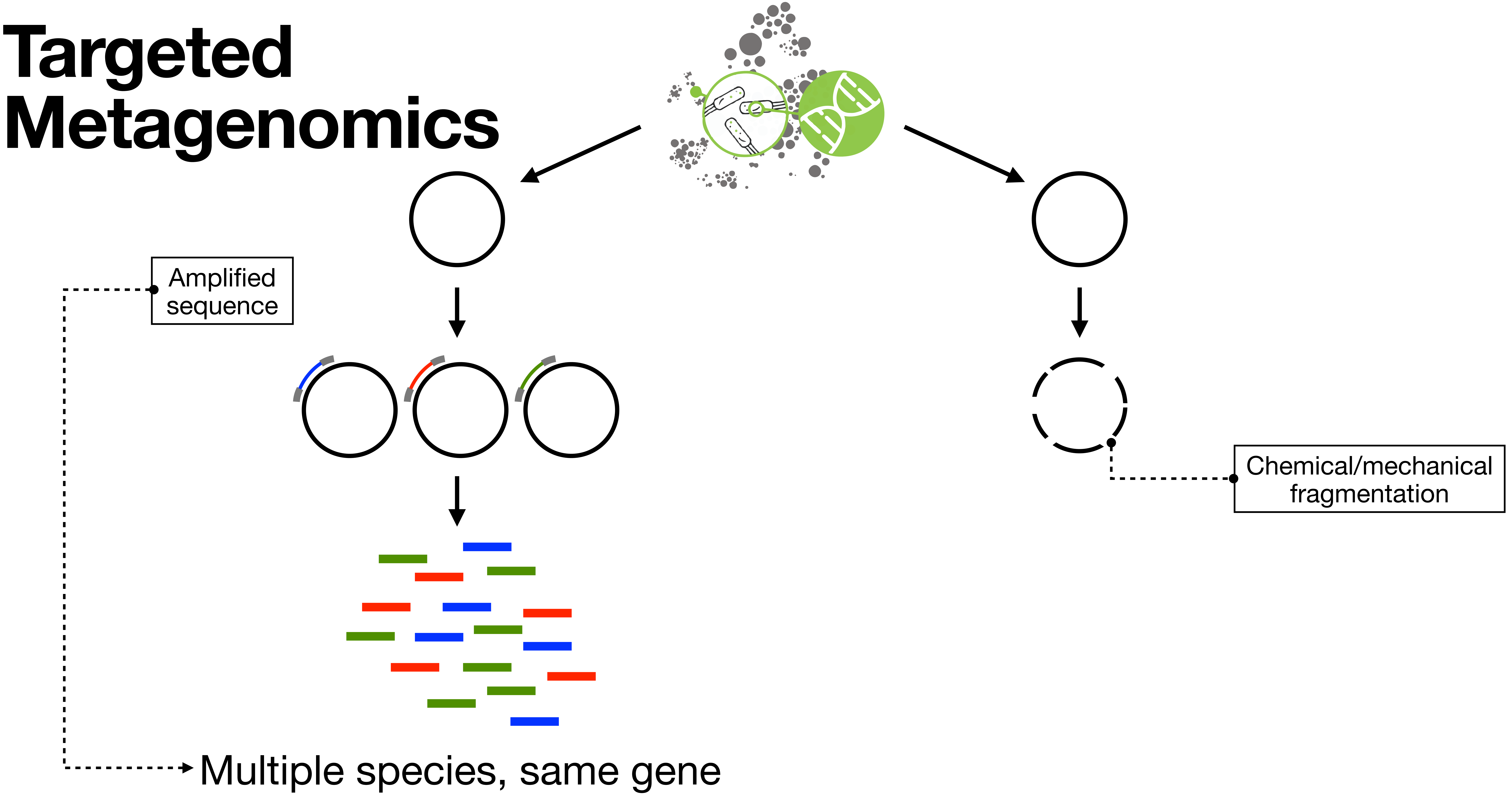
Targeted Metagenomics



Amplified sequence

Multiple species, same gene

Targeted Metagenomics

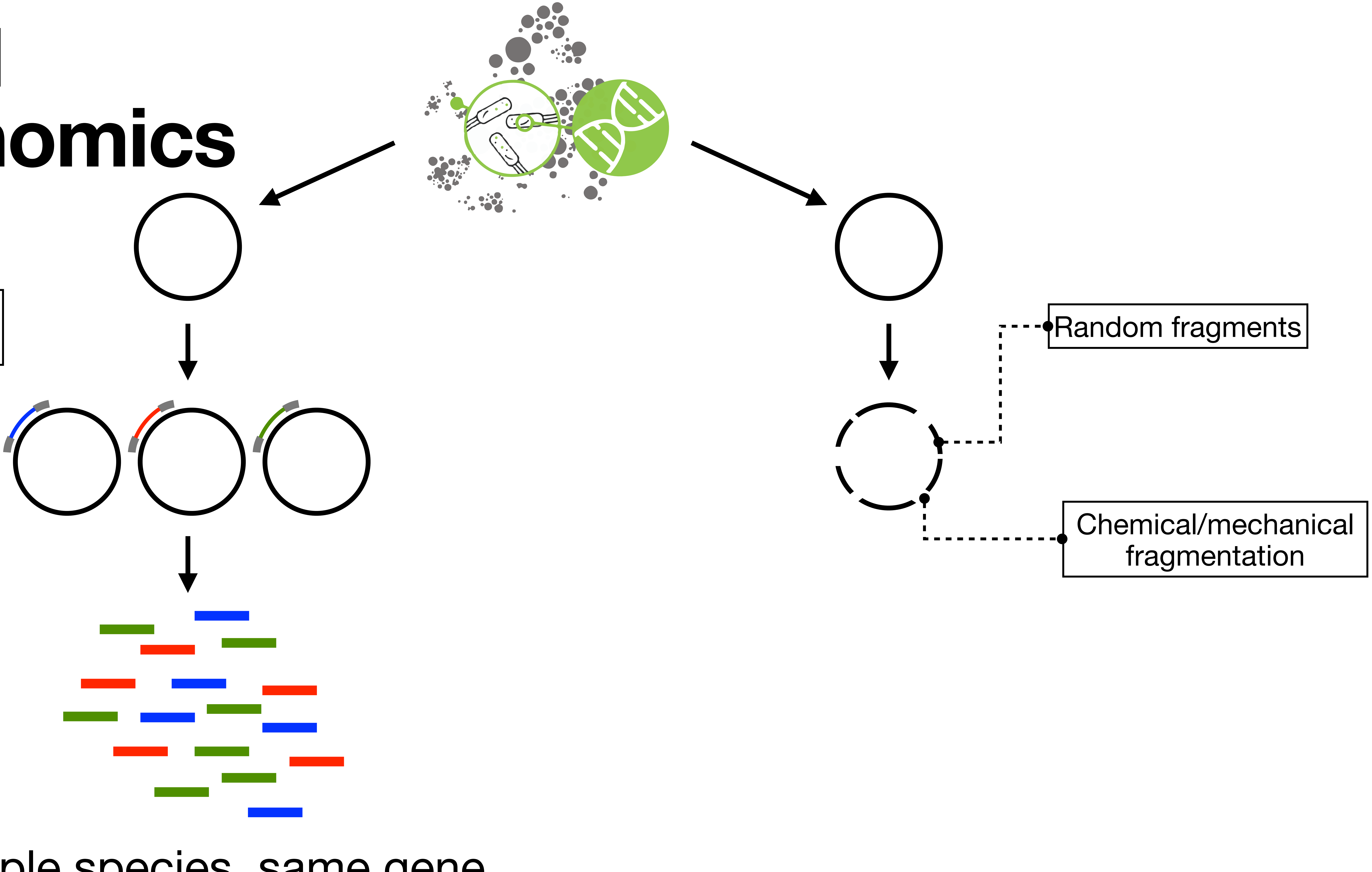


Amplified sequence

Chemical/mechanical fragmentation

Multiple species, same gene

Targeted Metagenomics



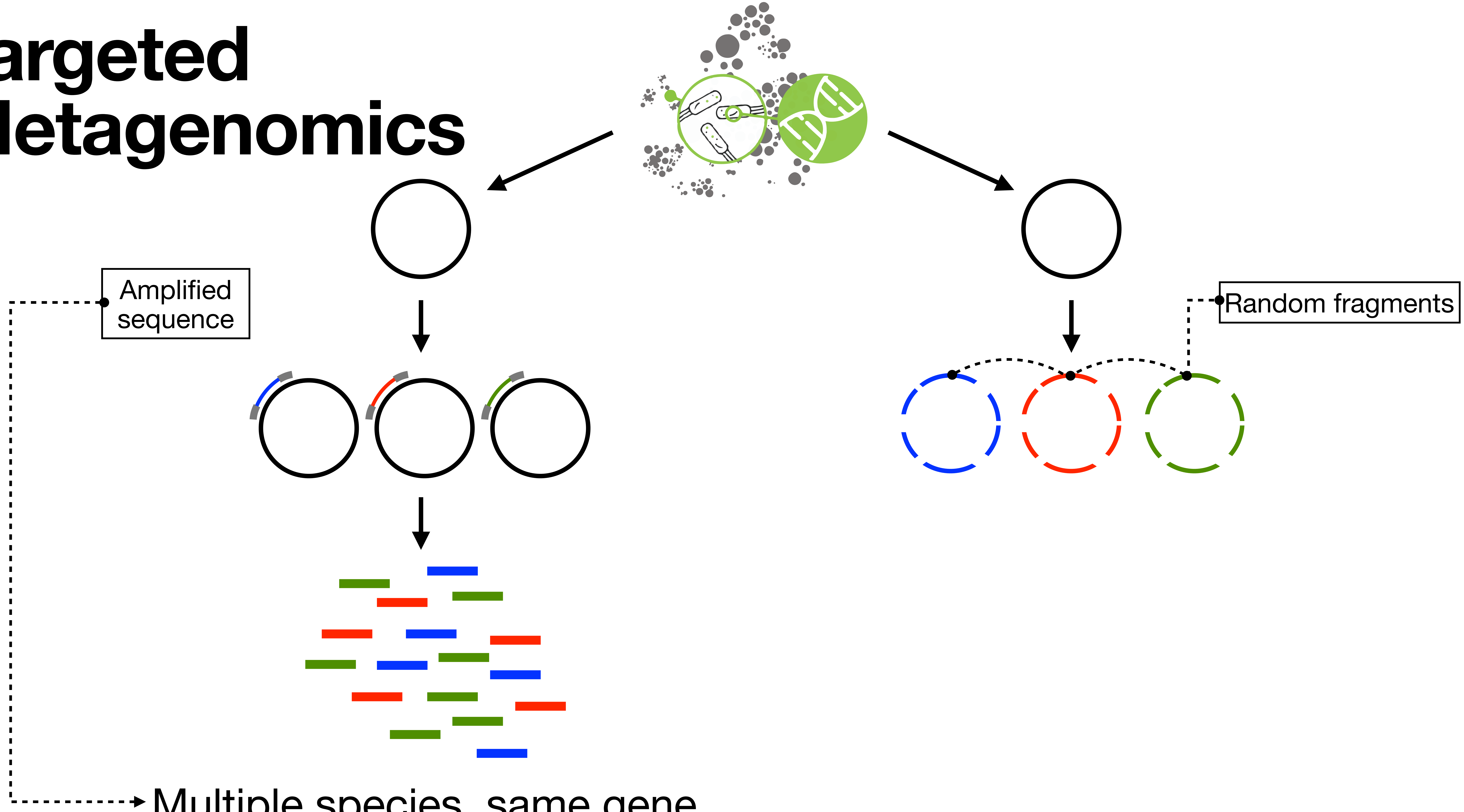
Amplified sequence

Random fragments

Chemical/mechanical fragmentation

Multiple species, same gene

Targeted Metagenomics

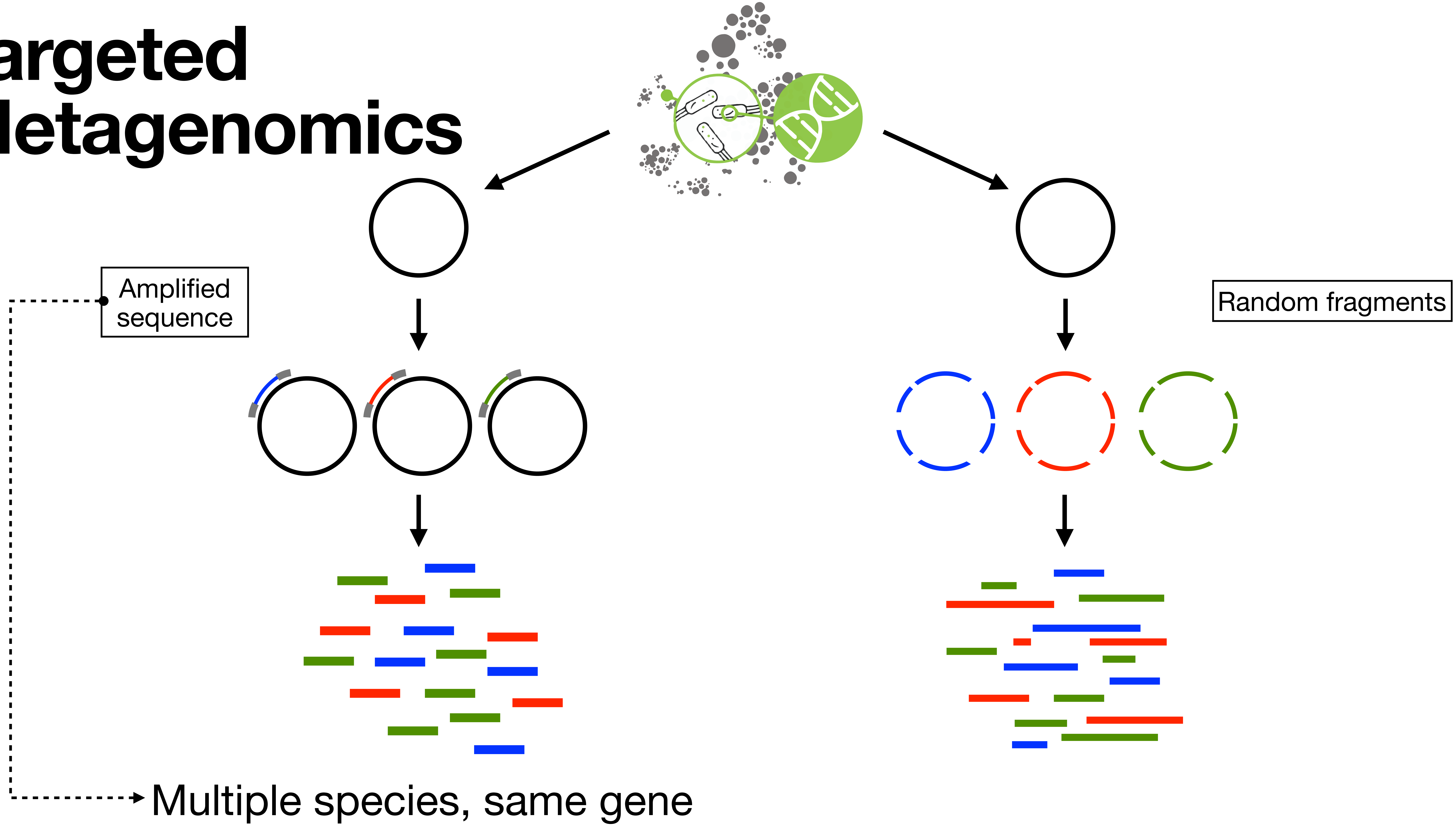


Amplified sequence

Random fragments

Multiple species, same gene

Targeted Metagenomics

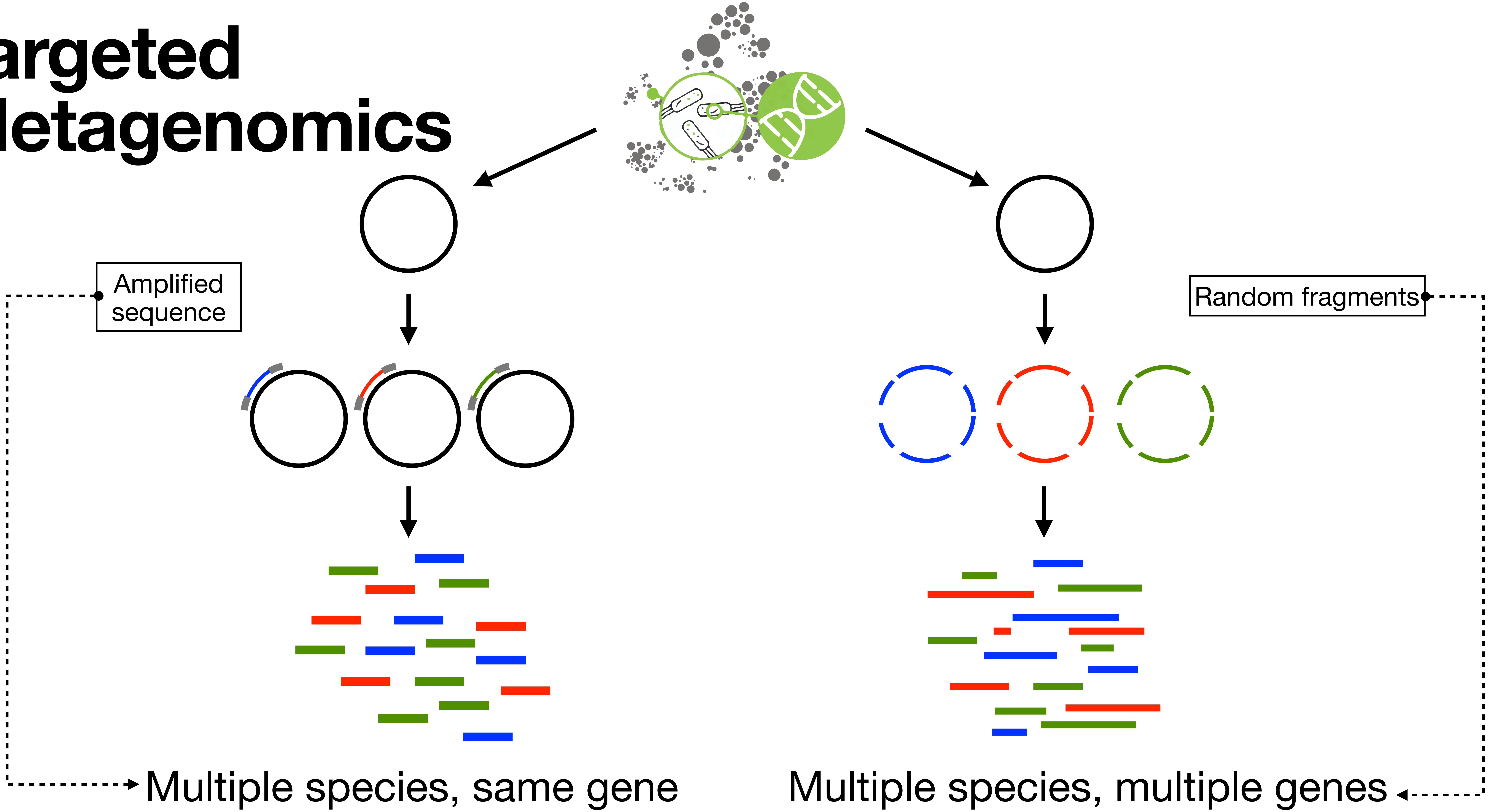


Amplified sequence

Random fragments

Multiple species, same gene

Targeted Metagenomics



Amplified sequence

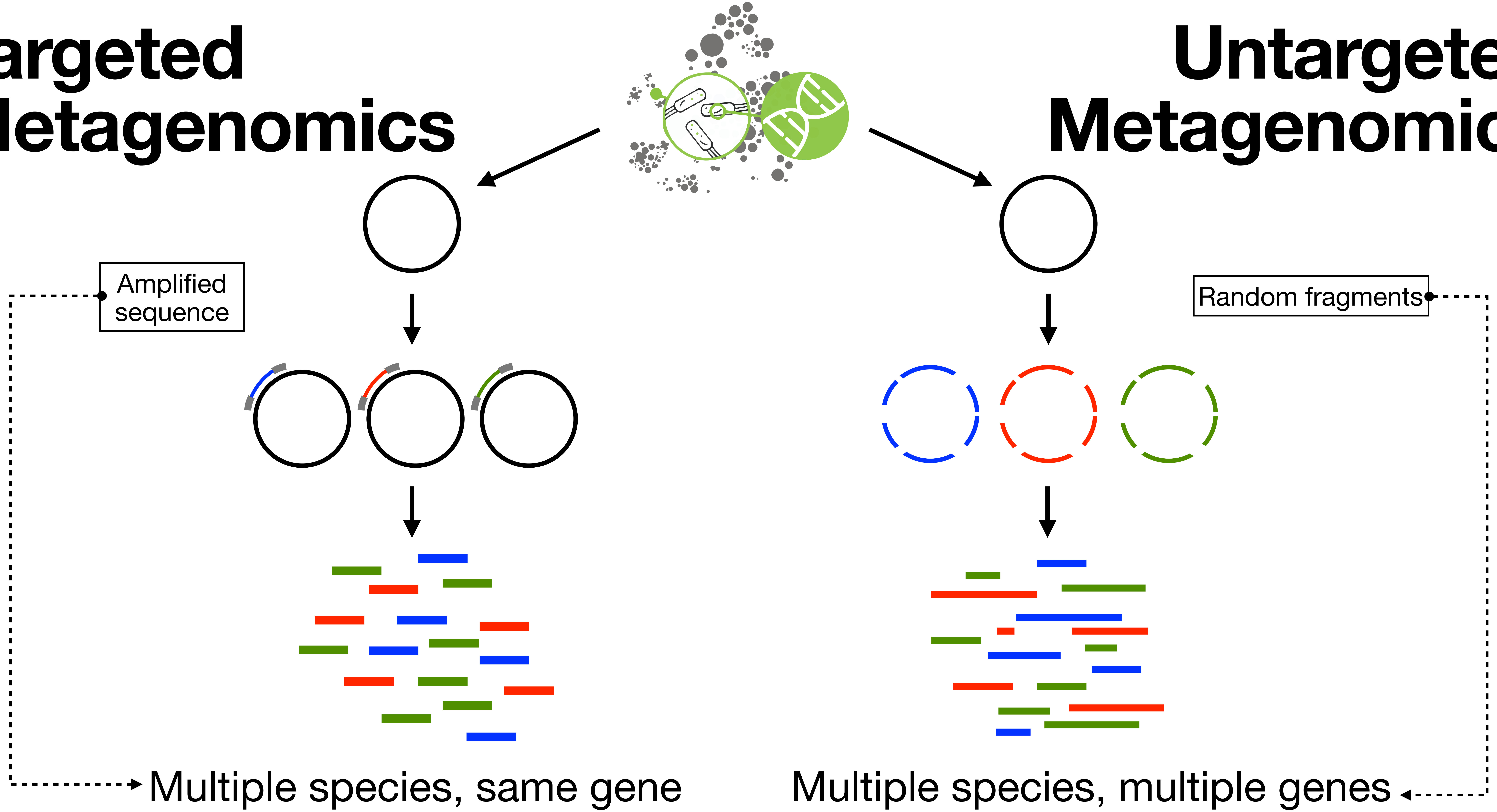
Random fragments

Multiple species, same gene

Multiple species, multiple genes

Targeted Metagenomics

Untargeted Metagenomics



Amplified sequence

Random fragments

Multiple species, same gene

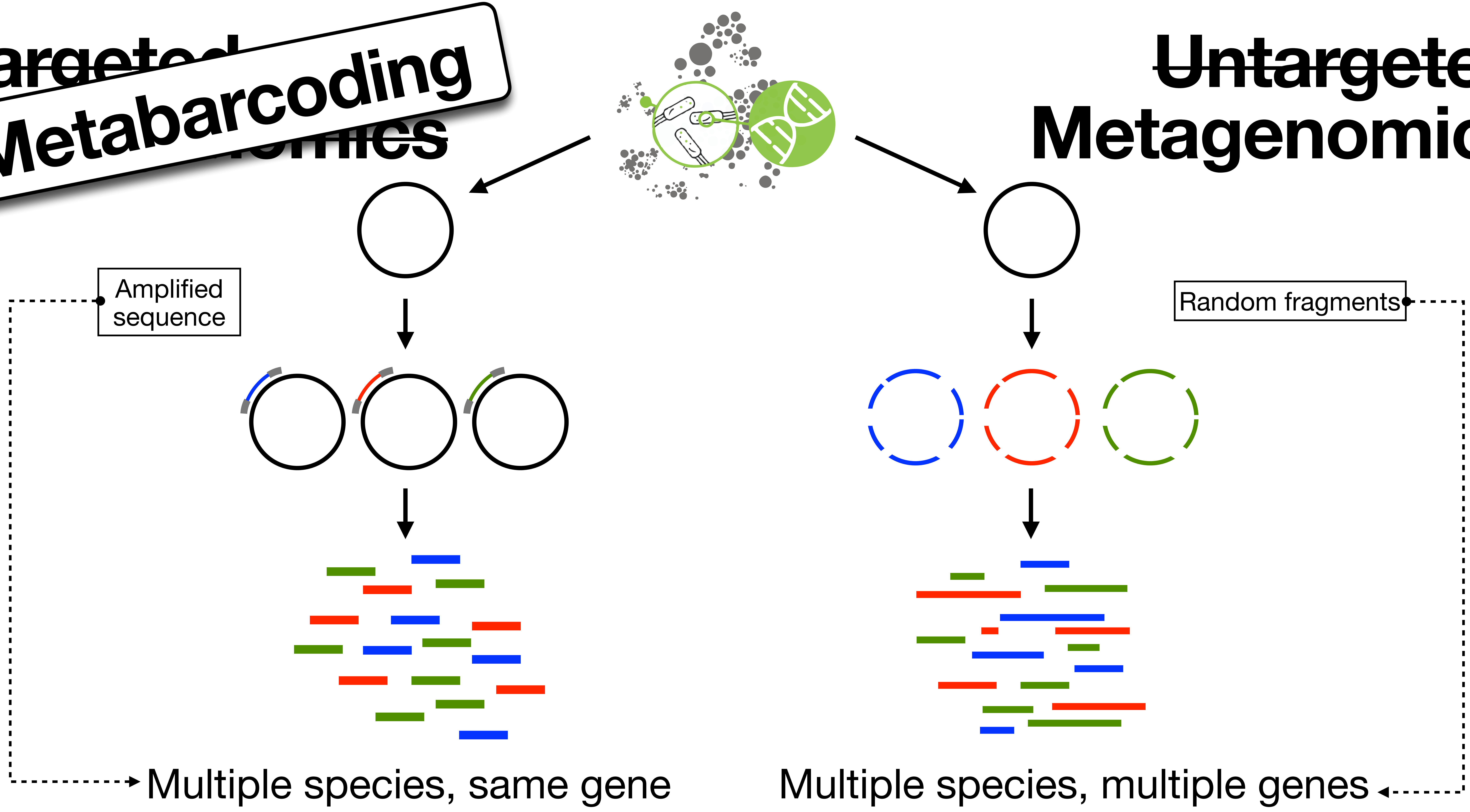
Multiple species, multiple genes

~~Targeted~~

Metabarcoding

~~omics~~

~~Untargeted~~ Metagenomics



Amplified sequence

Random fragments

Multiple species, same gene

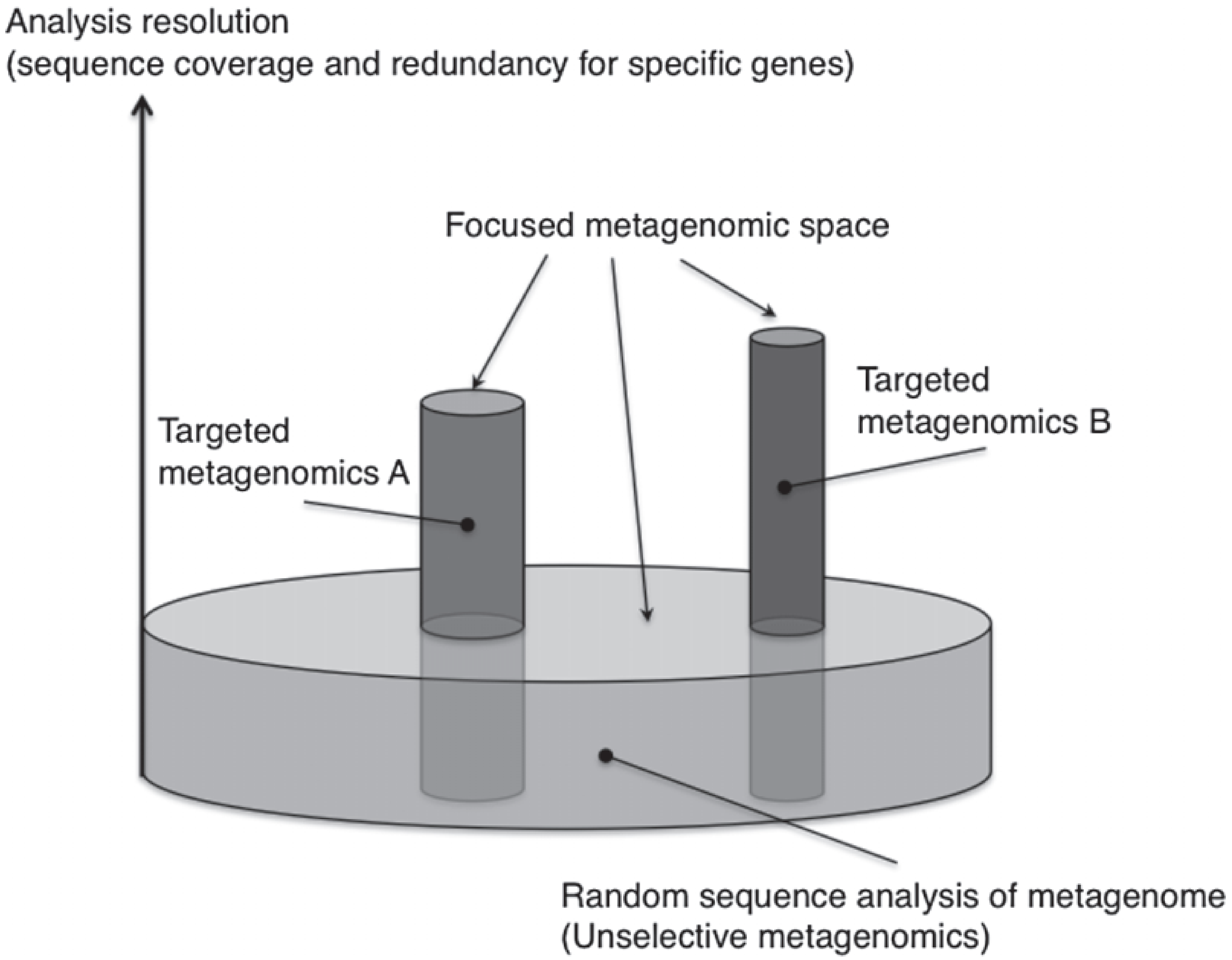
Multiple species, multiple genes

Obtainable information

Analysis space in meta-analyses

- **Targeted metagenomics**
  Focused on a single “analysis space” for increasing analysis resolution

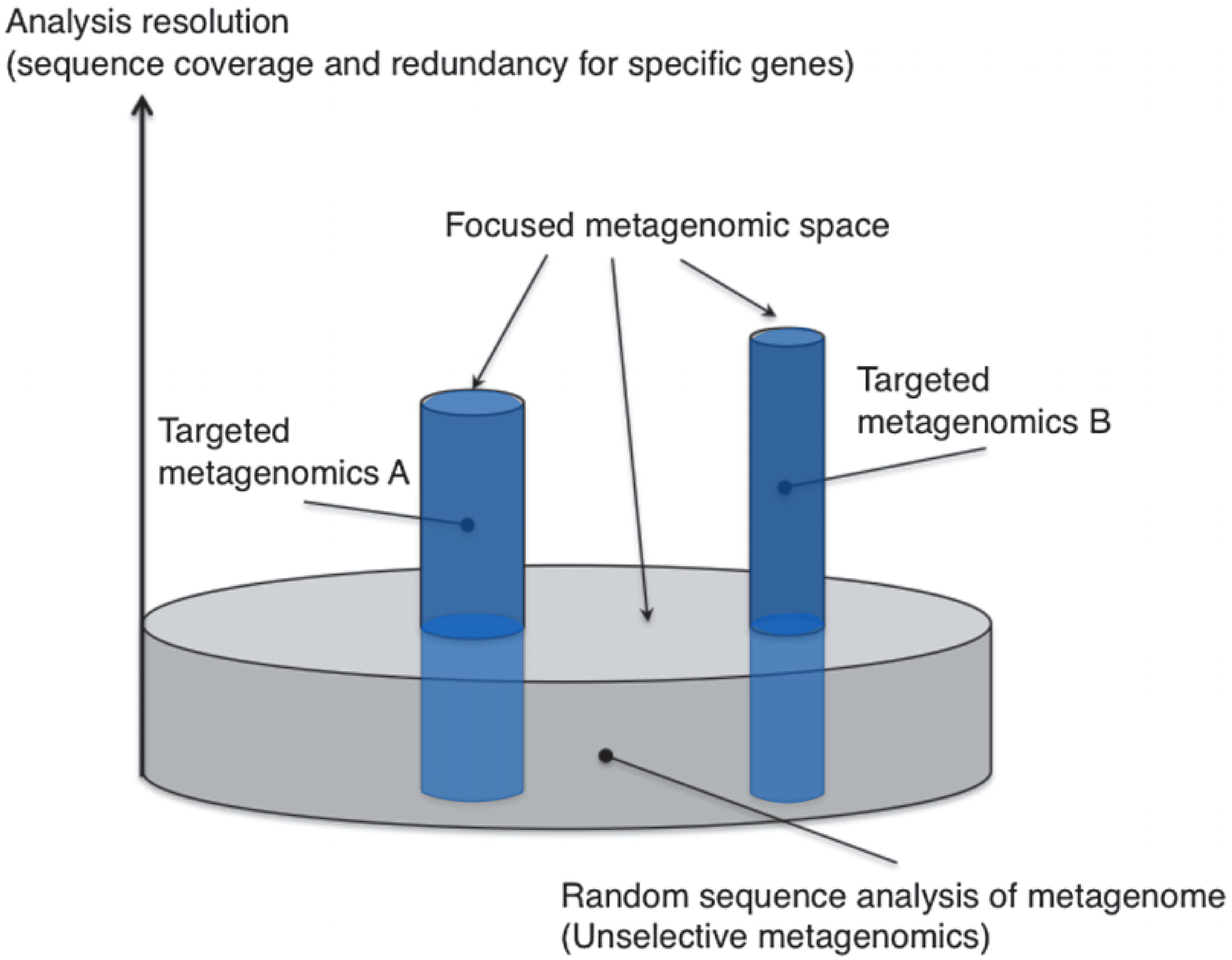
- **Untargeted metagenomics**
  All information available with relatively low resolution



Obtainable information

Analysis space in meta-analyses

- **Targeted metagenomics**
  Focused on a single “analysis space” for increasing analysis resolution
- **Untargeted metagenomics**
  All information available with relatively low resolution

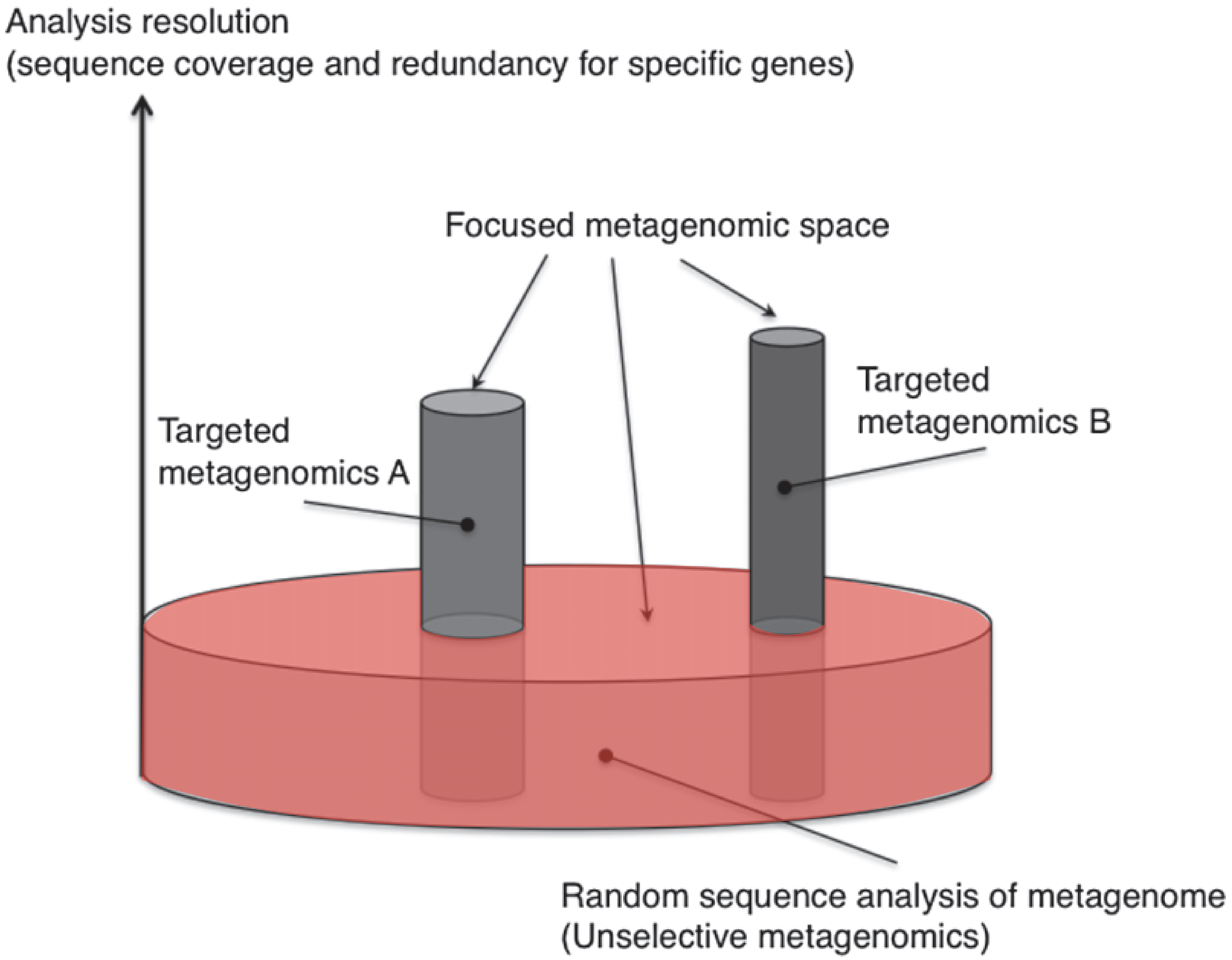


Obtainable information

Analysis space in meta-analyses

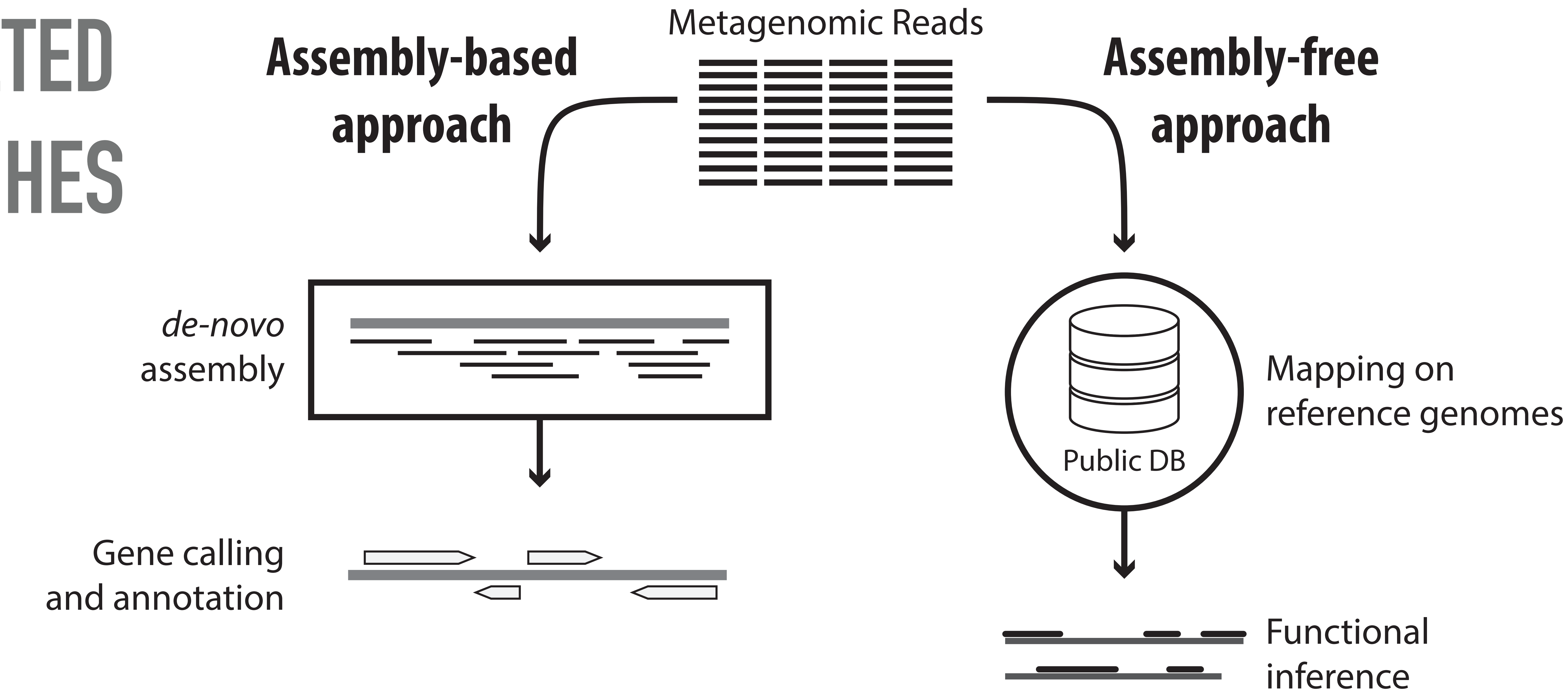
- **Targeted metagenomics**
  Focused on a single "analysis space" for increasing analysis resolution

- **Untargeted metagenomics**
  All information available with relatively low resolution



Untargeted Metagenomics

UNTARGETED
APPROACHES


Metagenomic Reads

Assembly-based approach

de-novo assembly

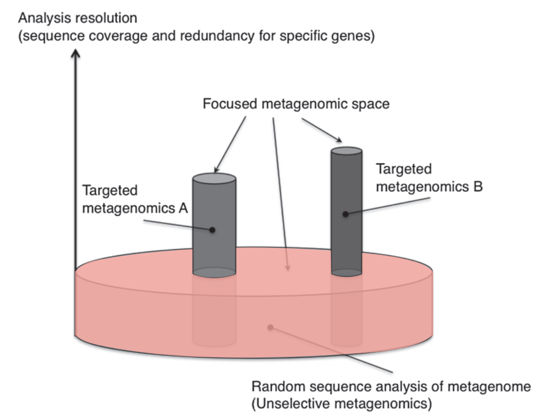
Gene calling and annotation

Assembly-free approach

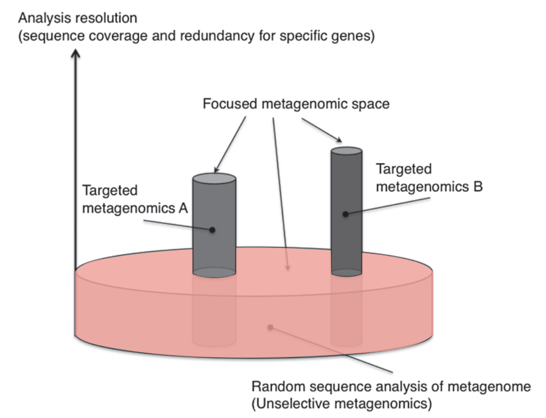
Public DB

Mapping on reference genomes

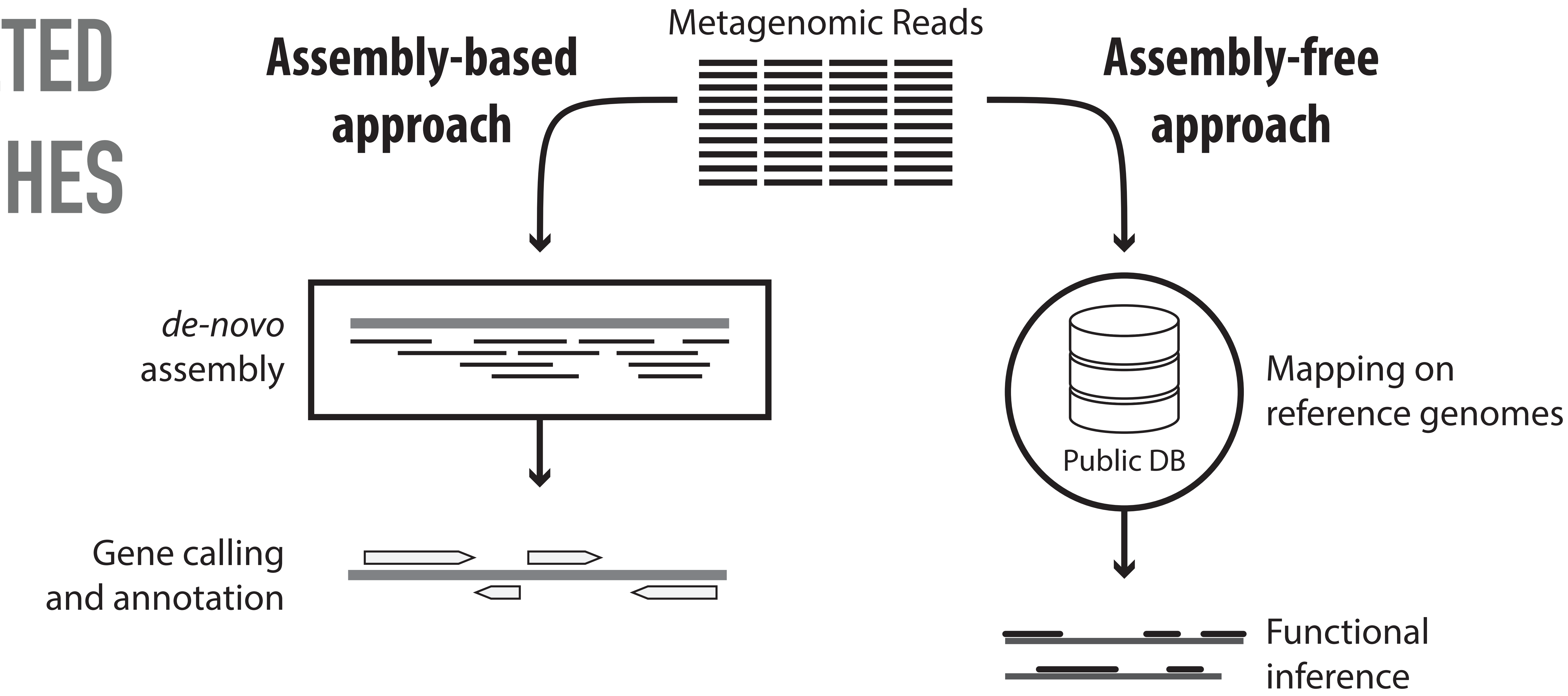
Functional inference



Targeted vs. Untargeted Metagenomics



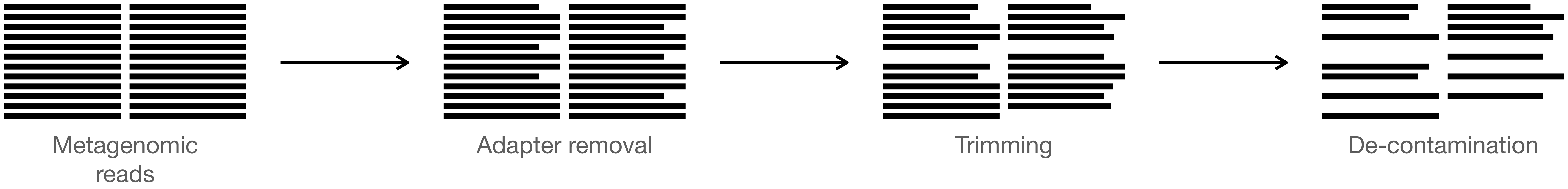
UNTARGETED APPROACHES



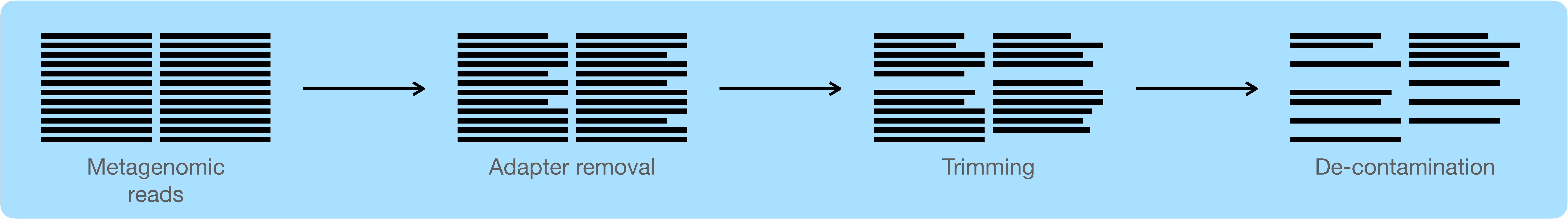
- Novel genes and pathways can be identified
- Useful if no reference genomes is available
- Biased towards higher-abundance community members

- More sensitive
- Less time consuming
- Less specific in functional identification

Assembly-based approaches

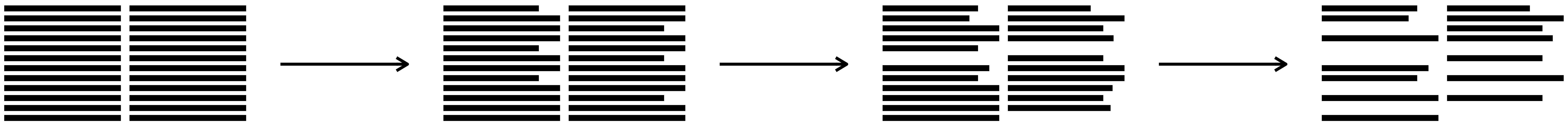


Assembly-based approaches



Quality control

Assembly-based approaches



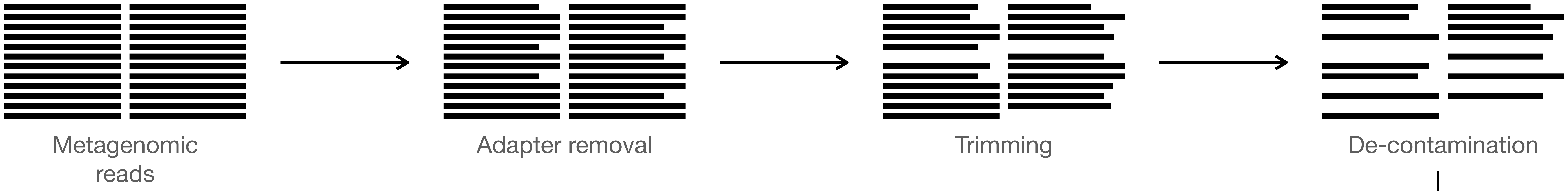
Metagenomic reads

Adapter removal

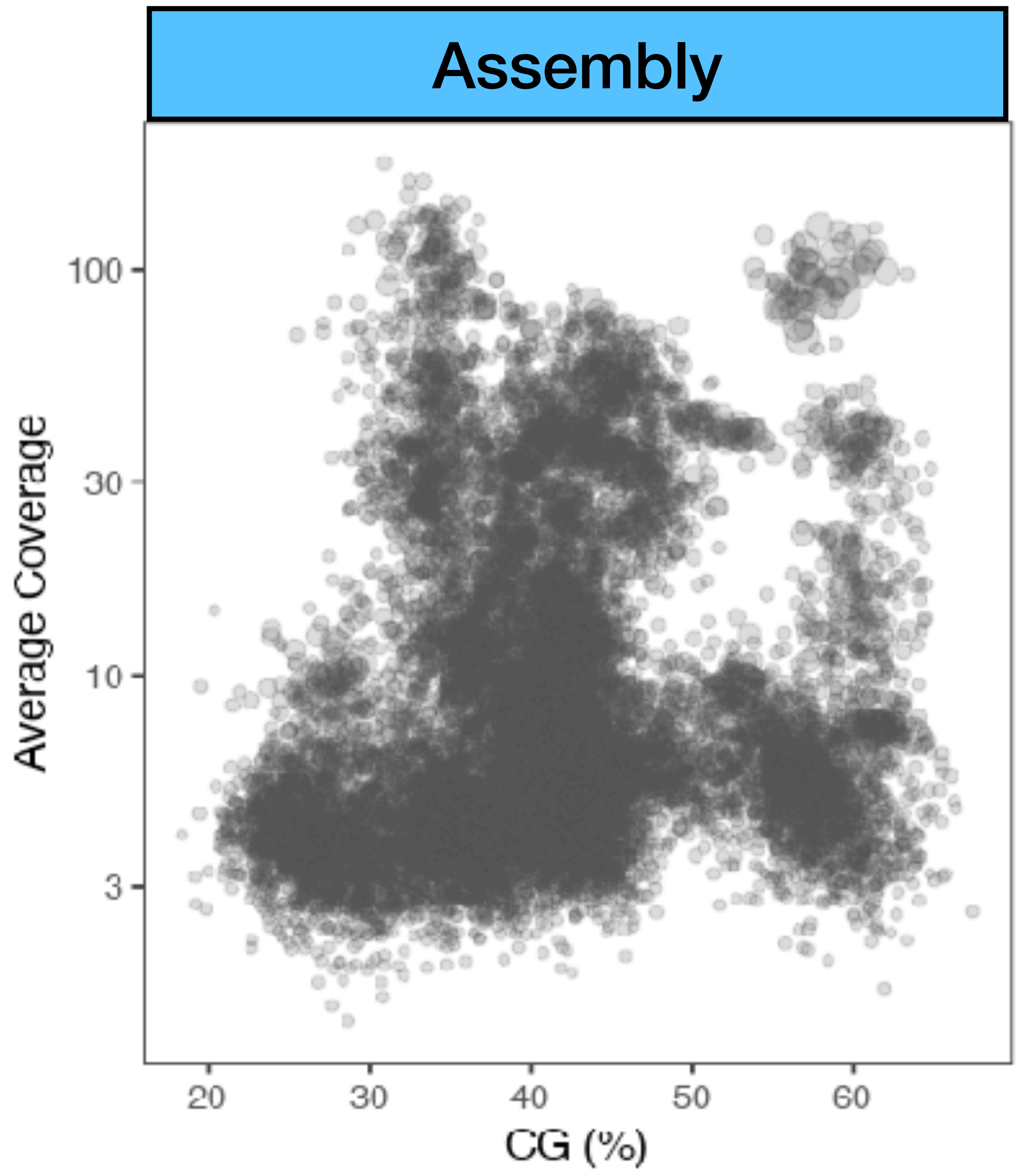
Trimming

De-contamination

Assembly-based approaches



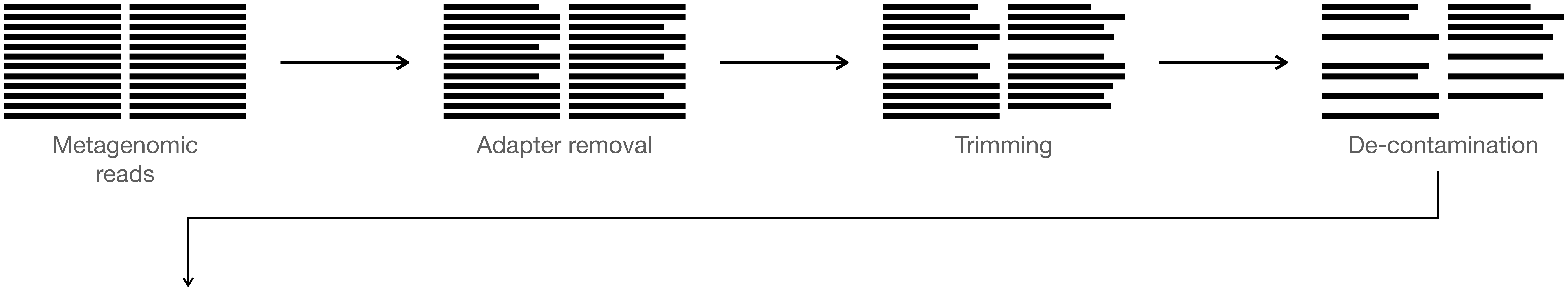
Metagenomic reads → Adapter removal → Trimming → De-contamination



Assembly

- metaSPAdes
- MEGAHIT
- Athena
- Canu

Assembly-based approaches

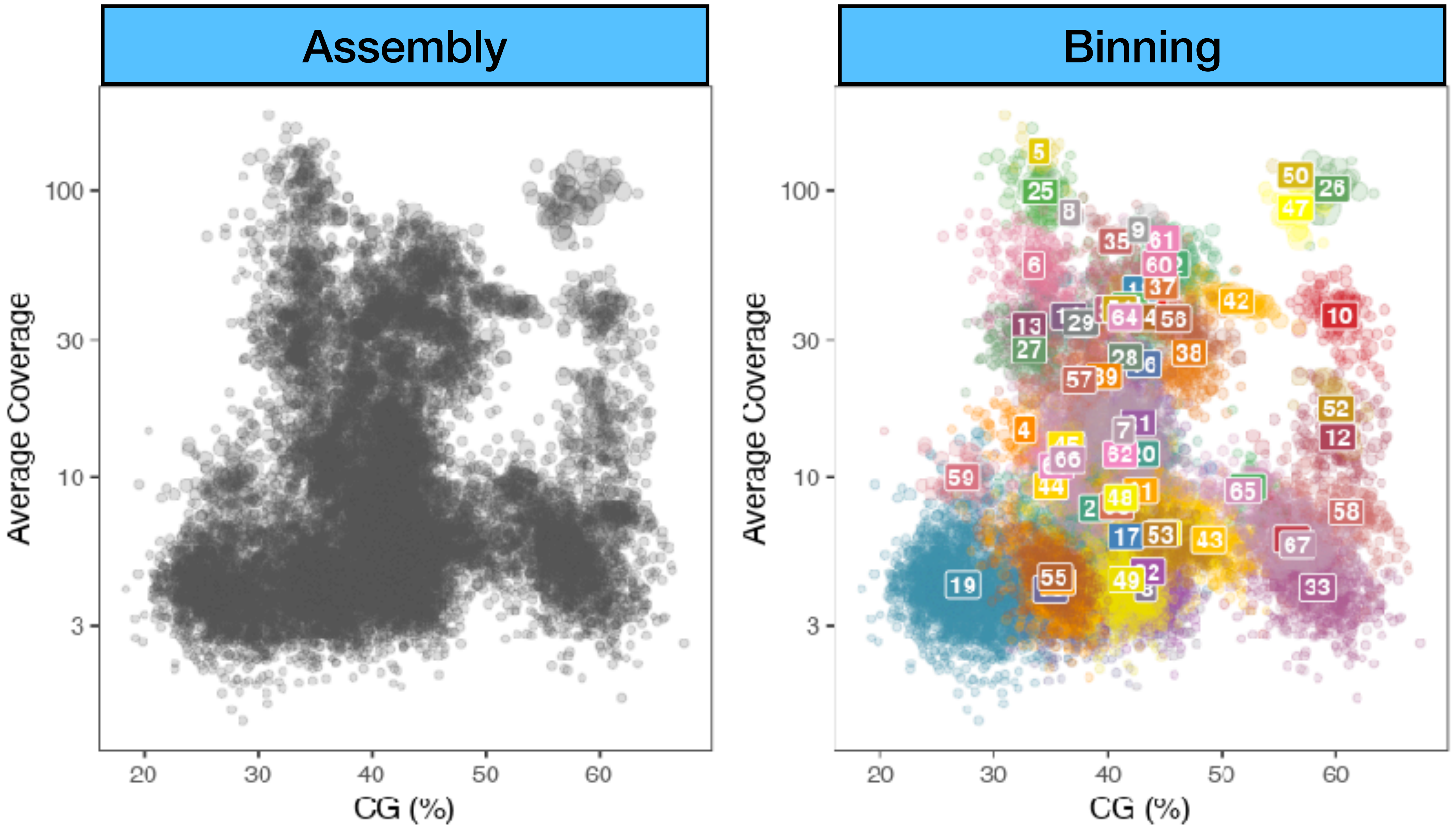


Metagenomic reads

Adapter removal

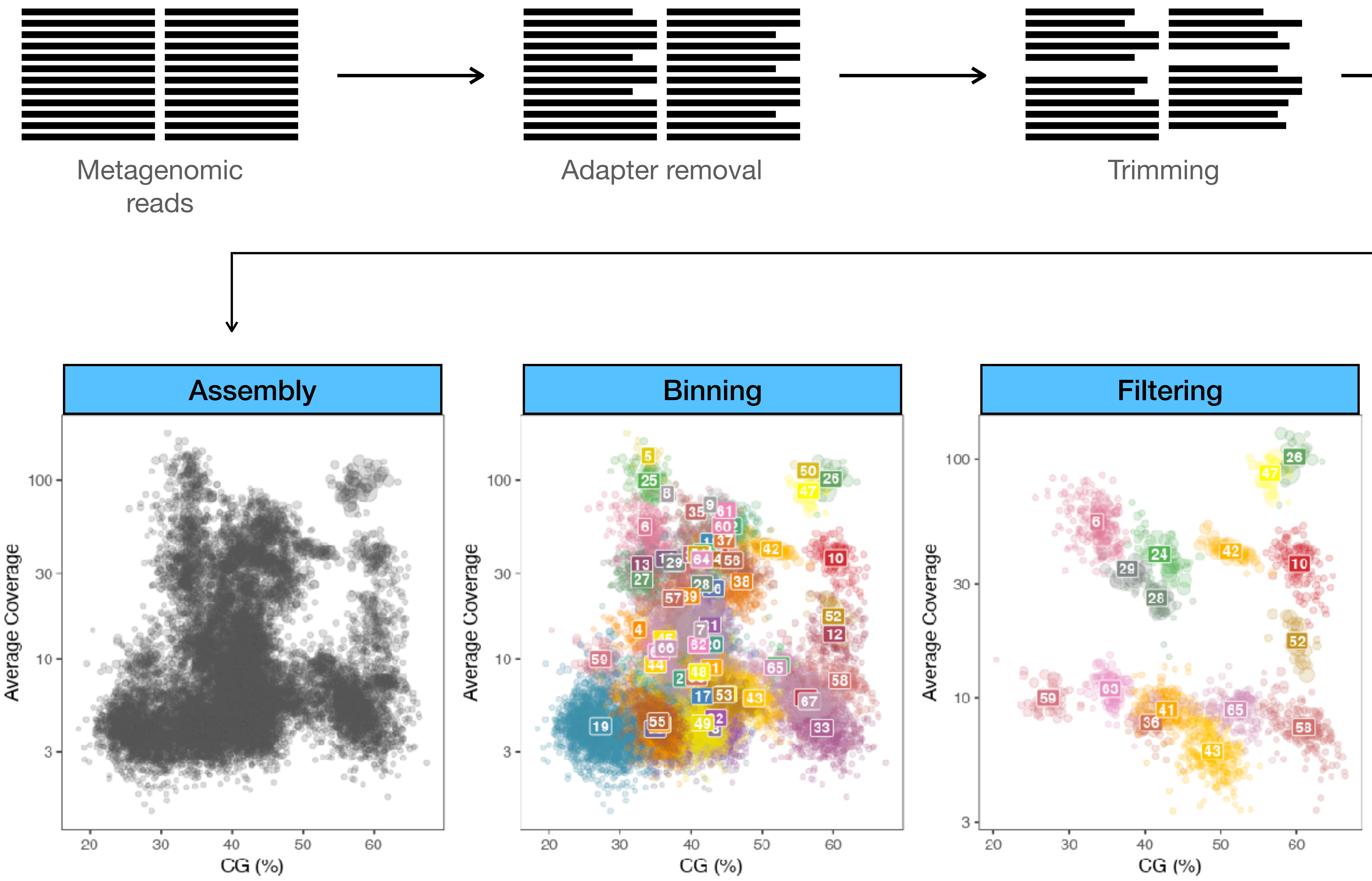
Trimming

De-contamination



- Metabat2
- TETRA
- CD-HIT
- COMEBin

Assembly-based approaches



Metagenomic reads

Adapter removal

Trimming

De-contamination

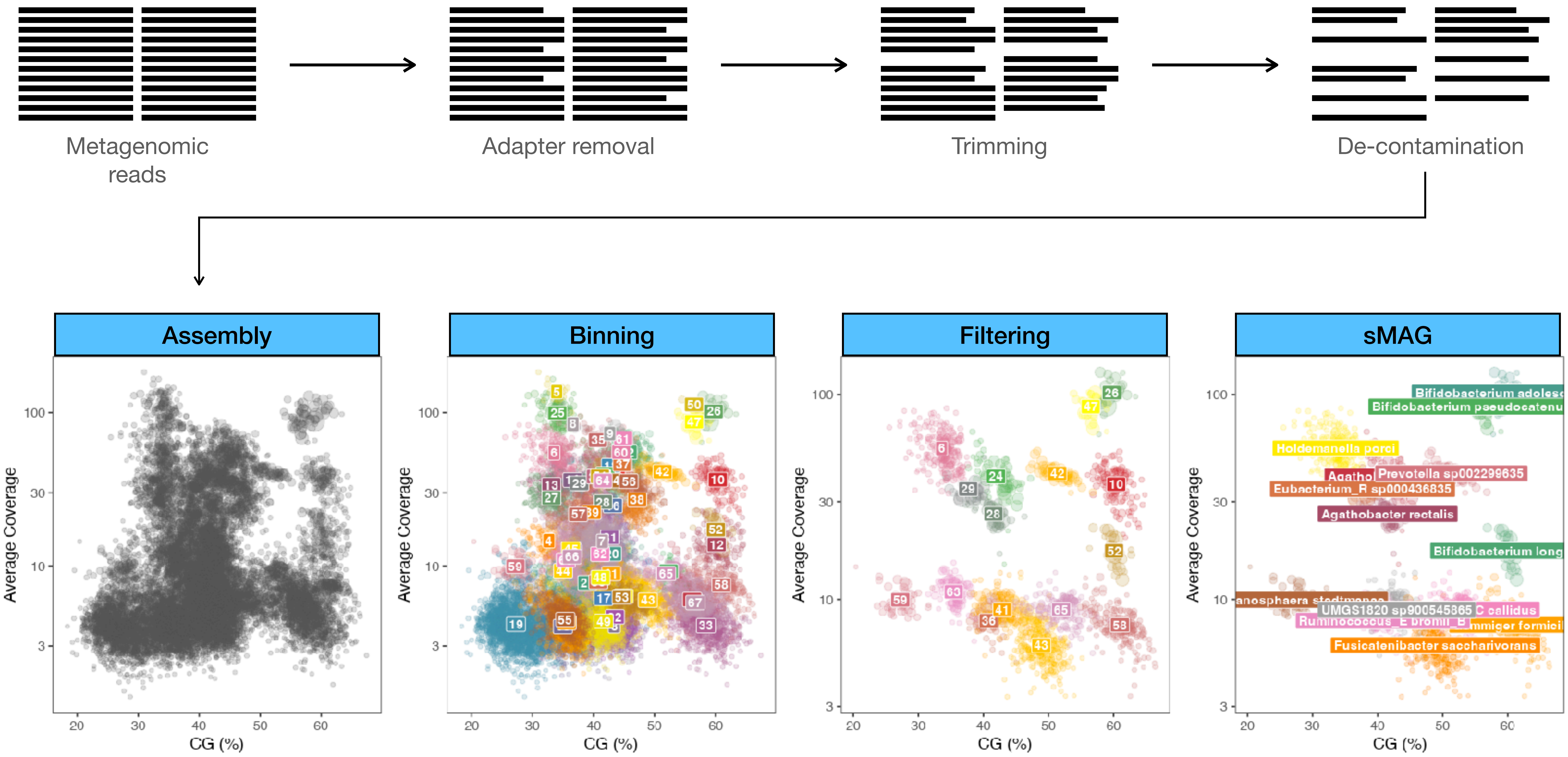
Assembly

Binning

Filtering

- CheckM
- ConFindR
- GUNC

Assembly-based approaches



Metagenomic reads

Adapter removal

Trimming

De-contamination

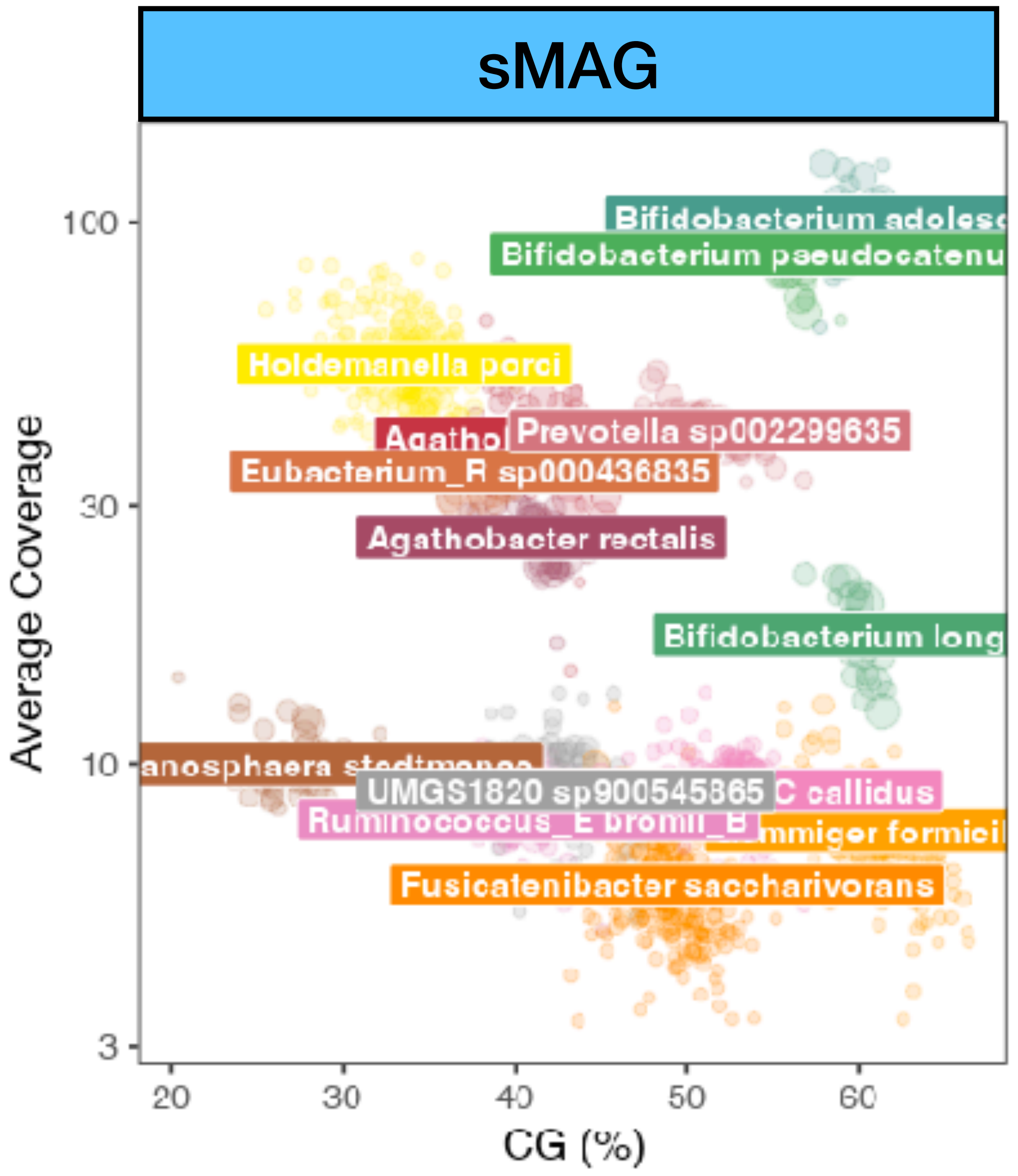
Assembly

Binning

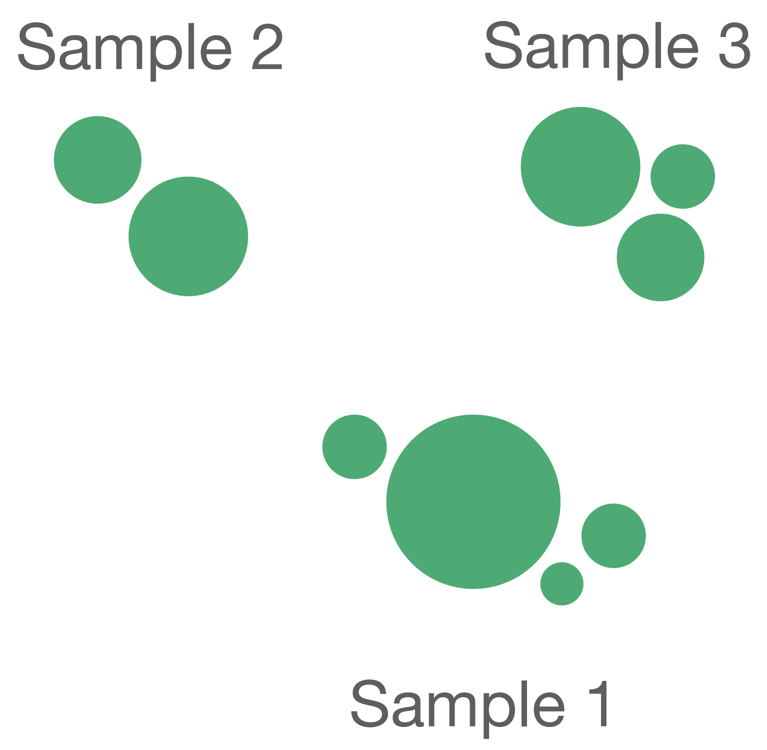
Filtering

sMAG

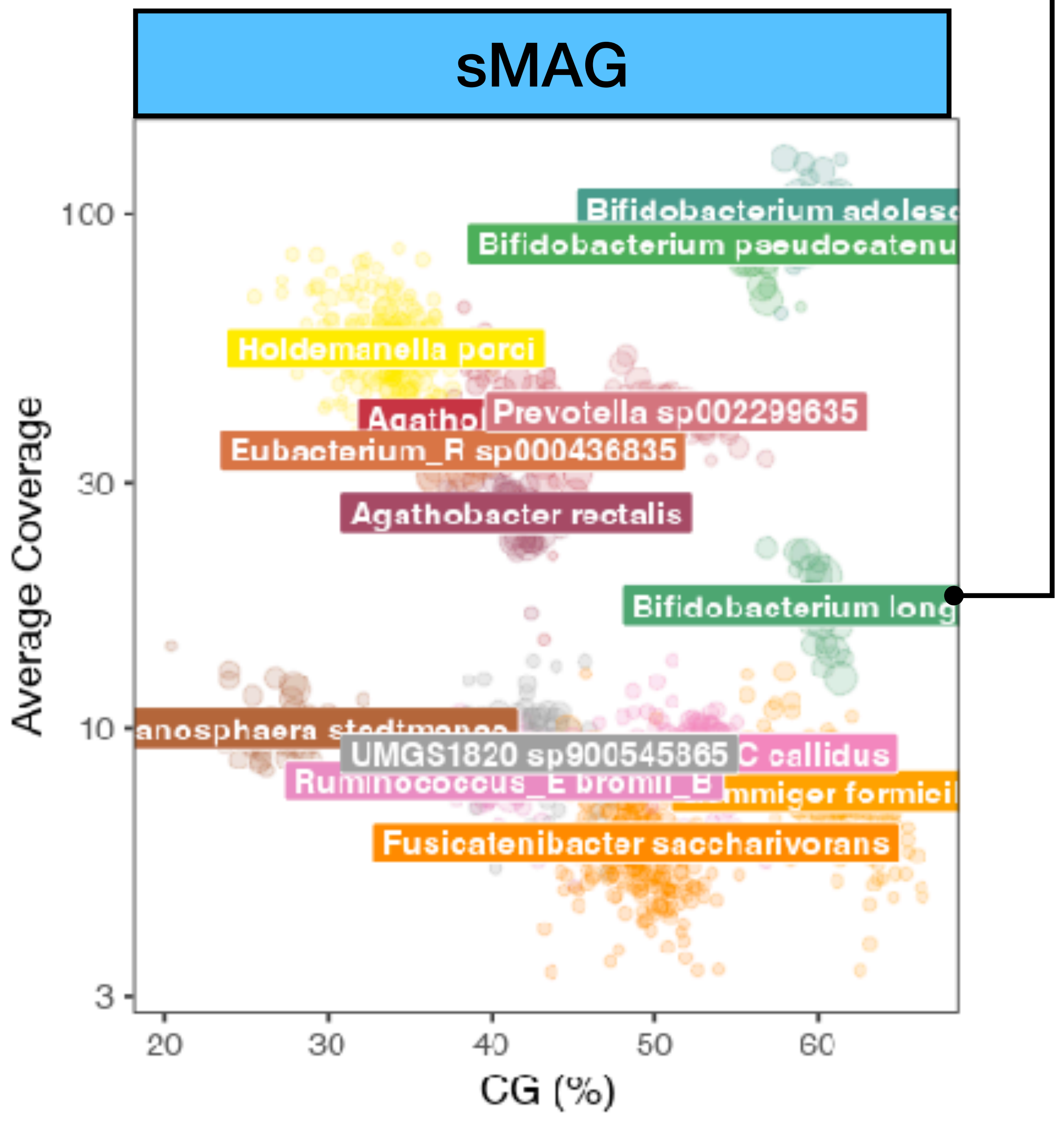
Assembly-based approaches



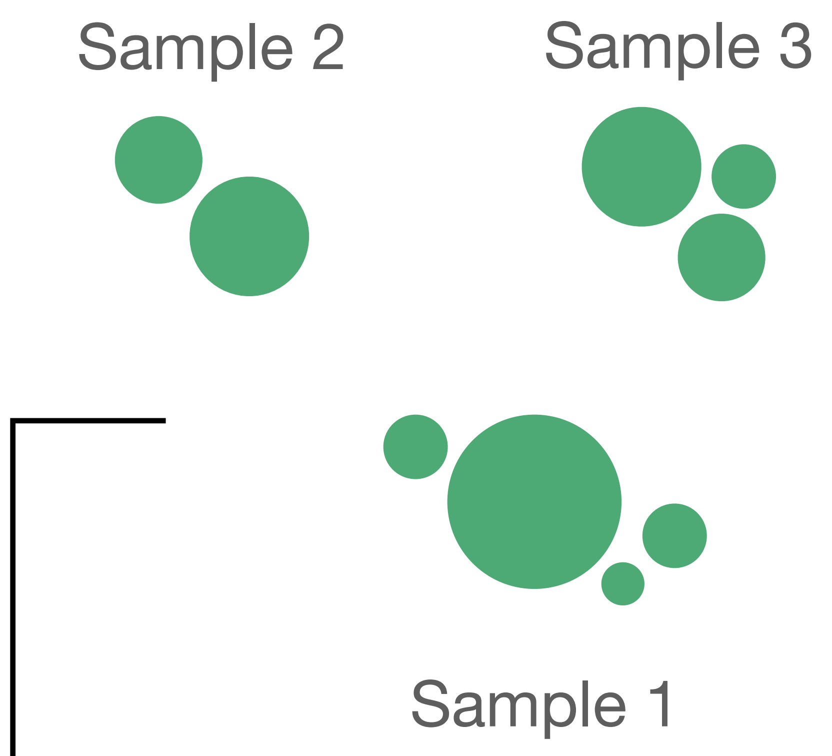
Assembly-based approaches



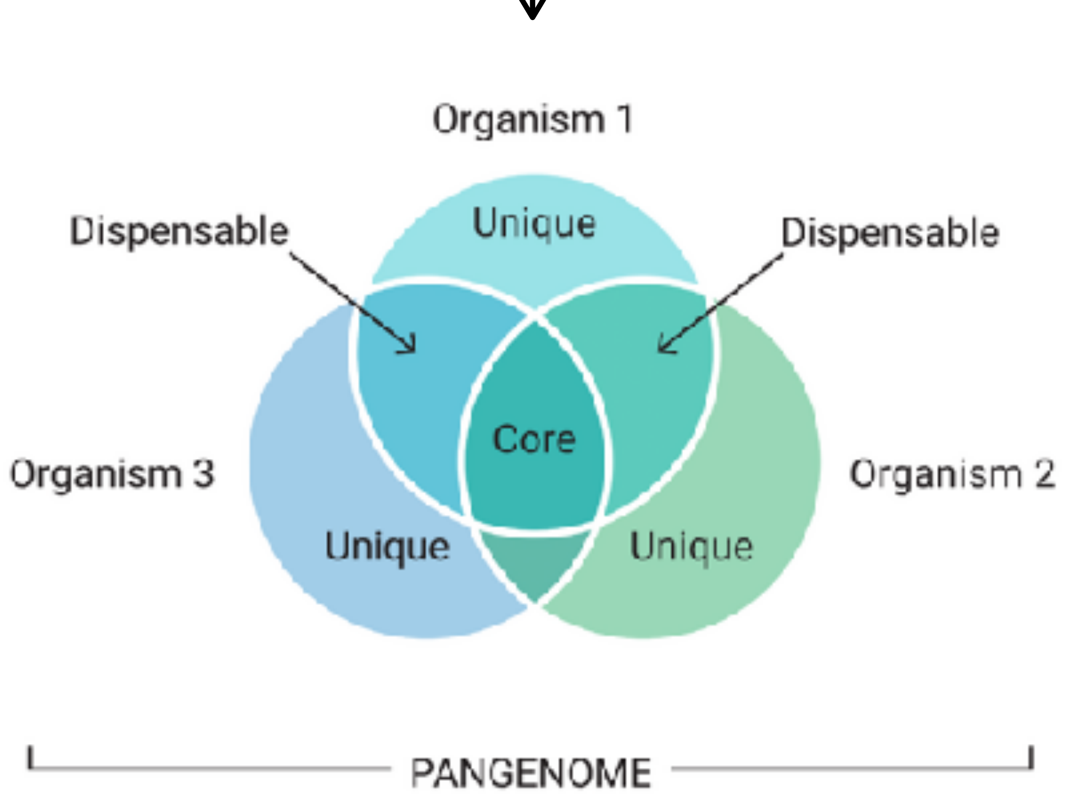
Select the same species in multiple samples

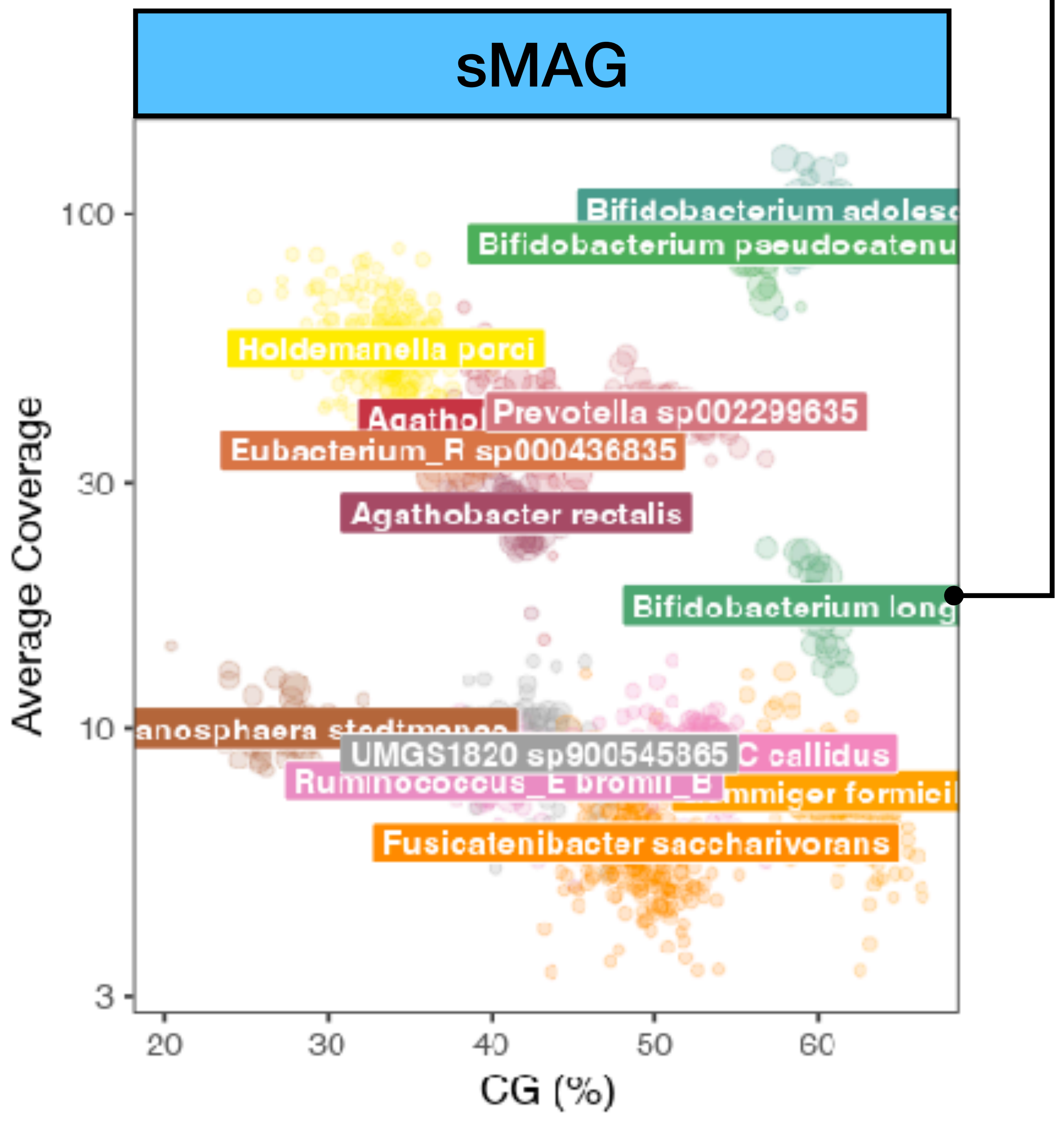


Assembly-based approaches

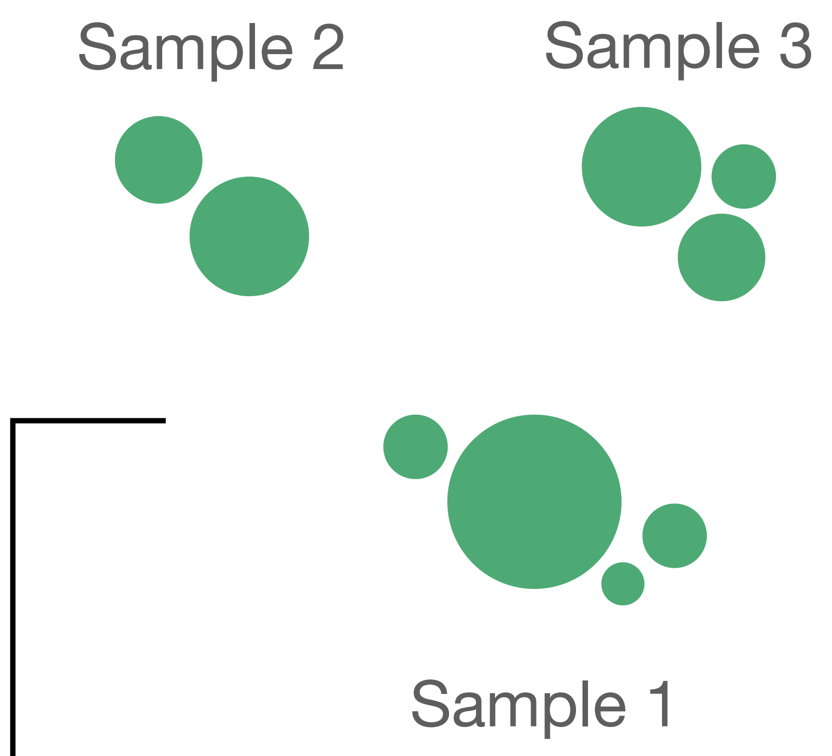


Select the same species in multiple samples

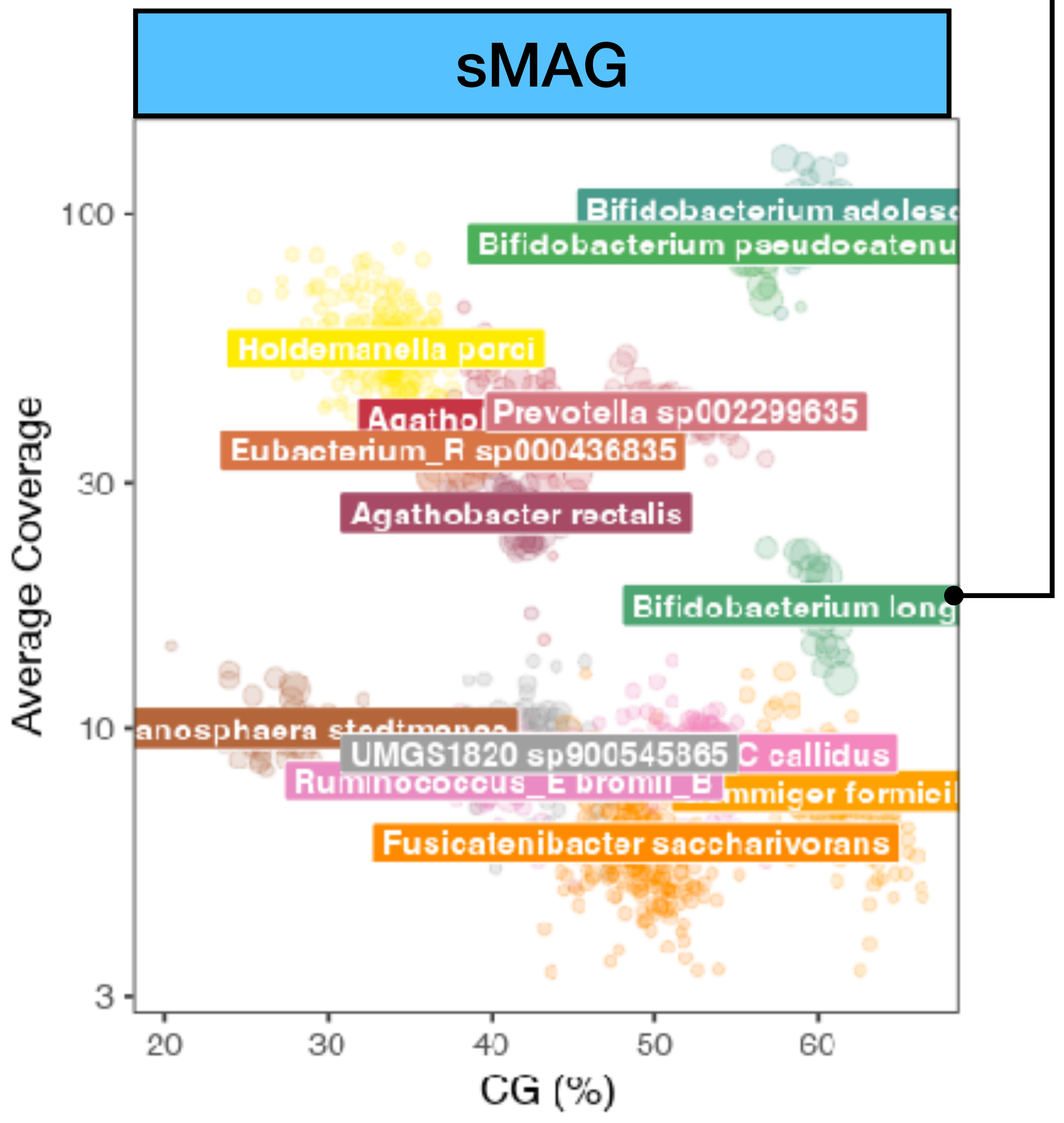


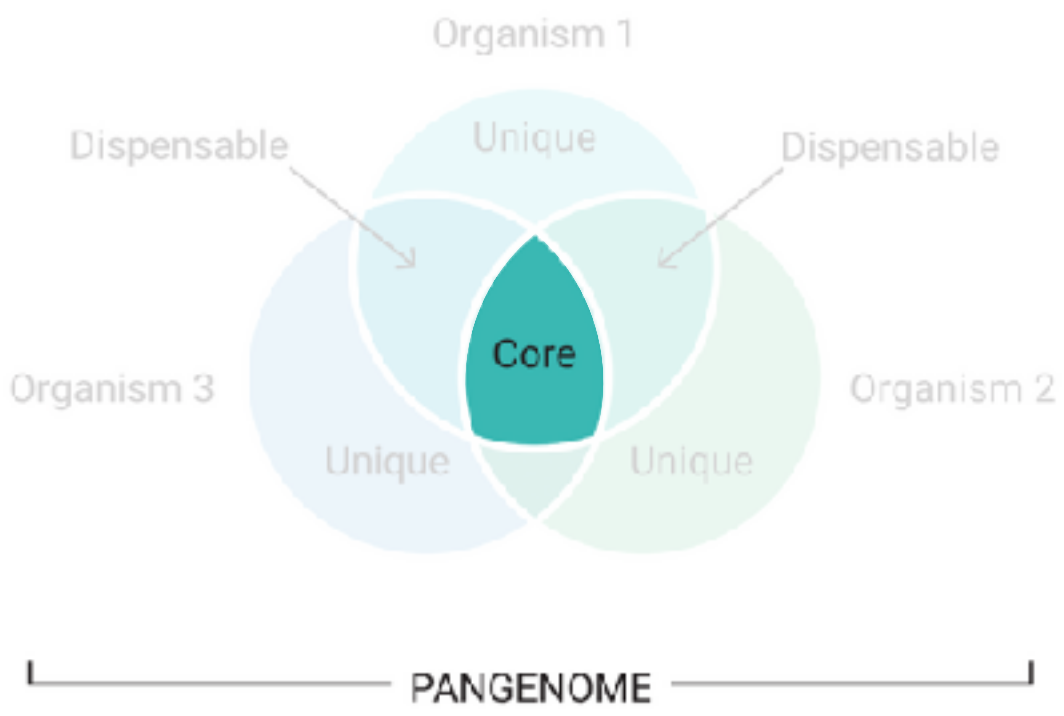


Assembly-based approaches



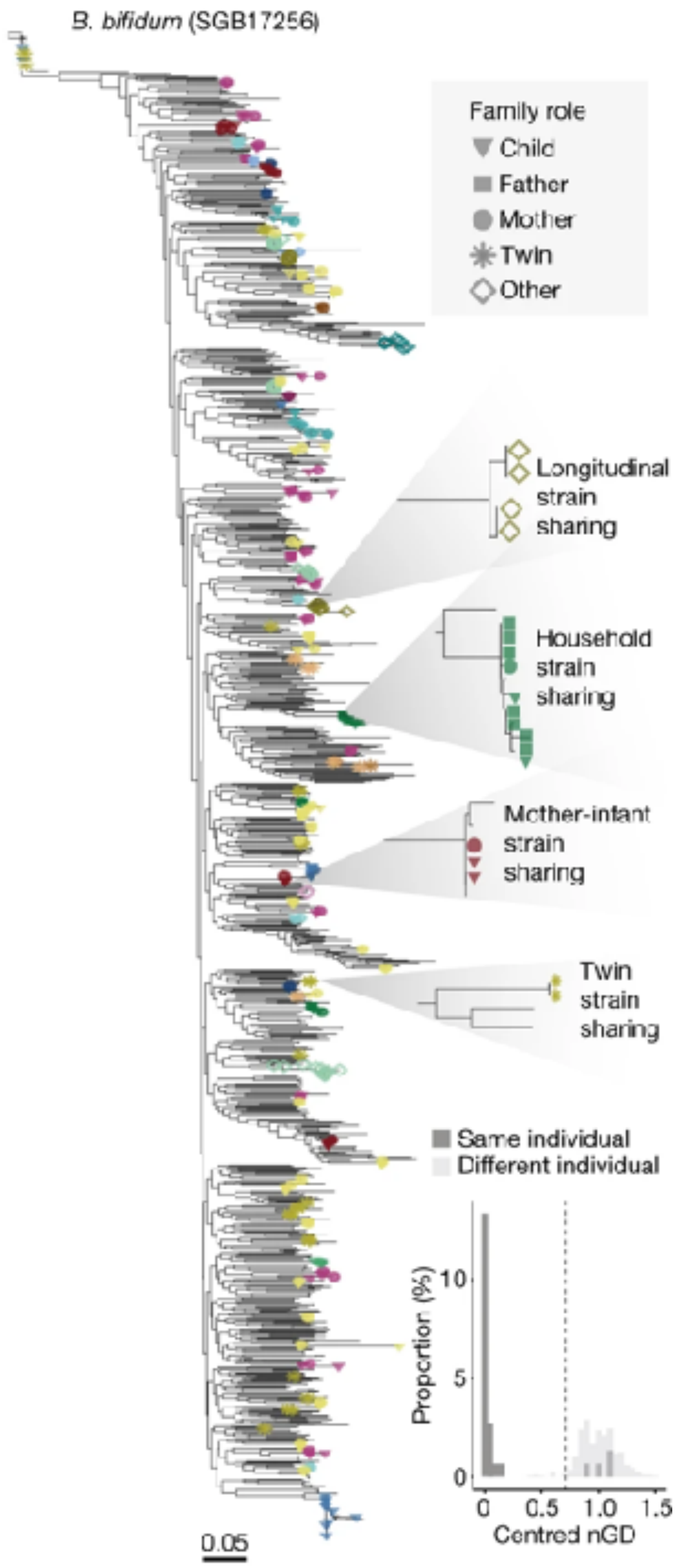
Select the same species in multiple samples





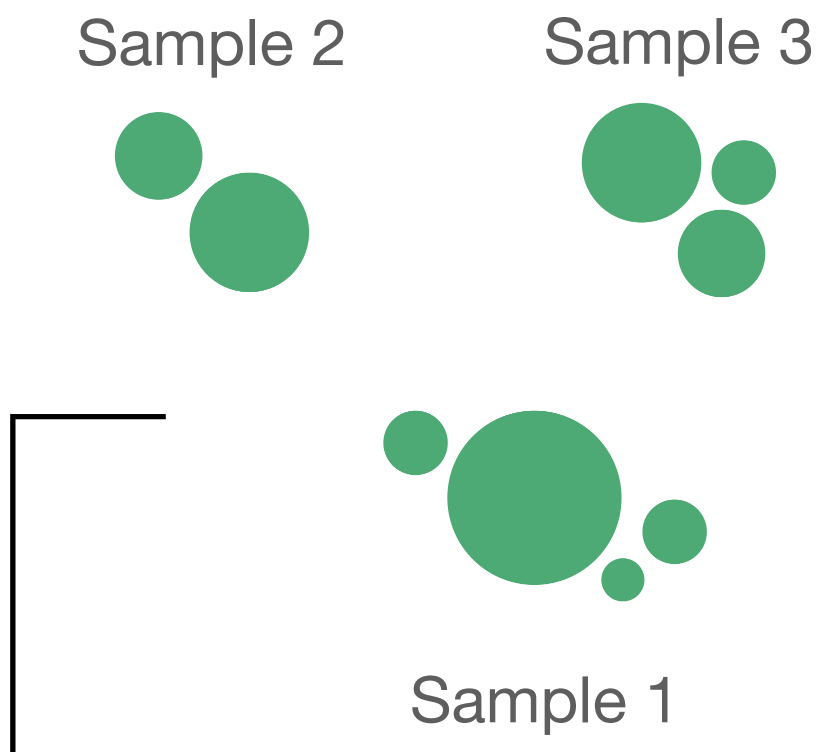
Get genes present in all “strains”

Assembly-based approaches

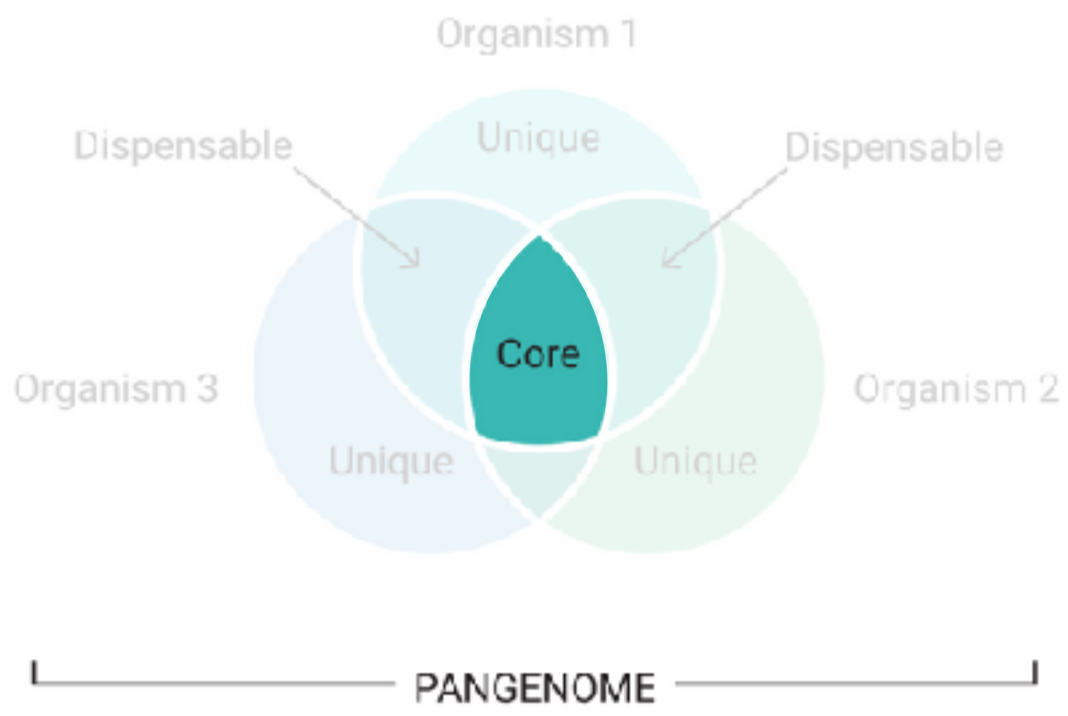


(Valles-Colomer et al 2023)

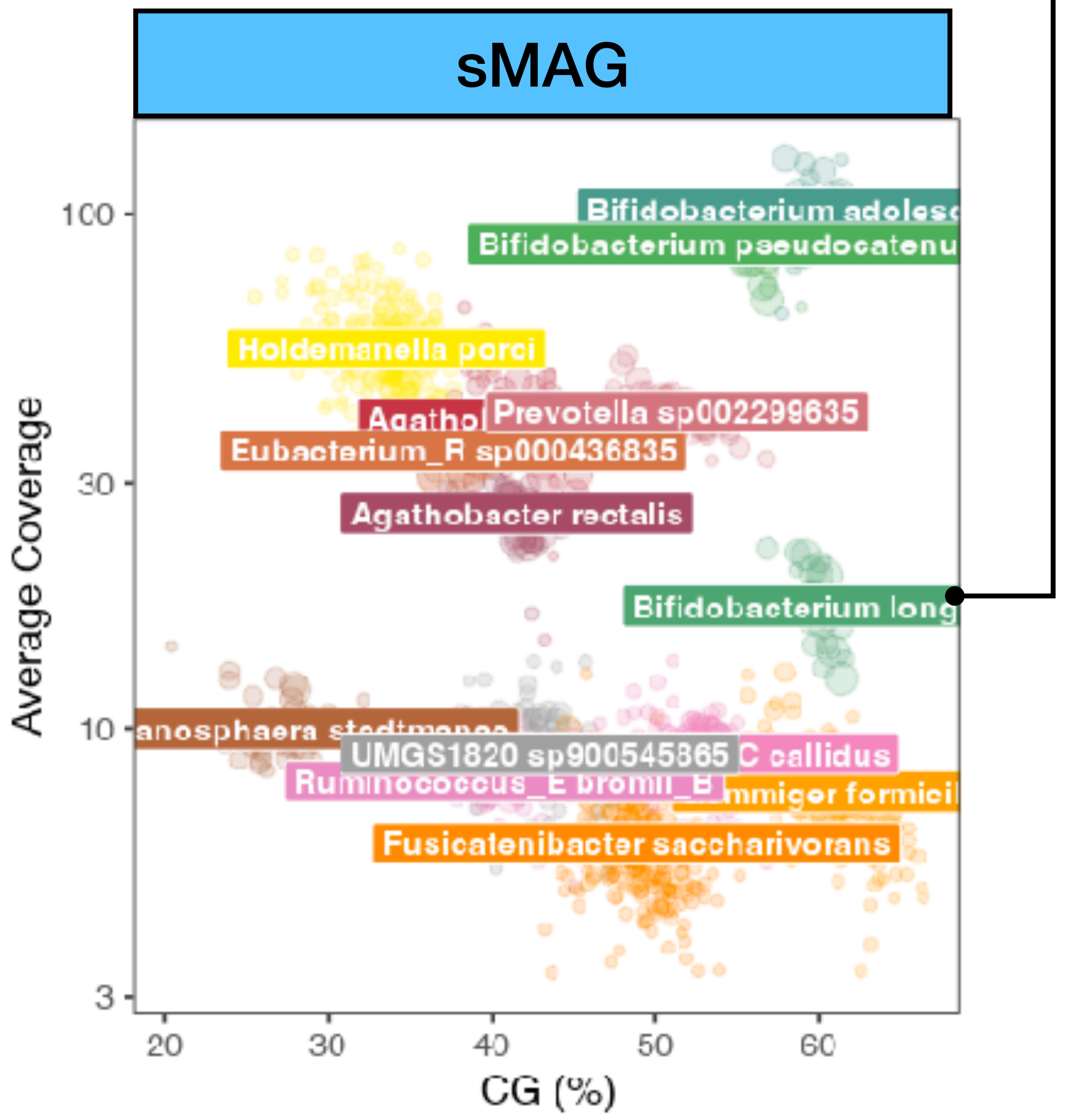
Strain typing



Select the same species in multiple samples

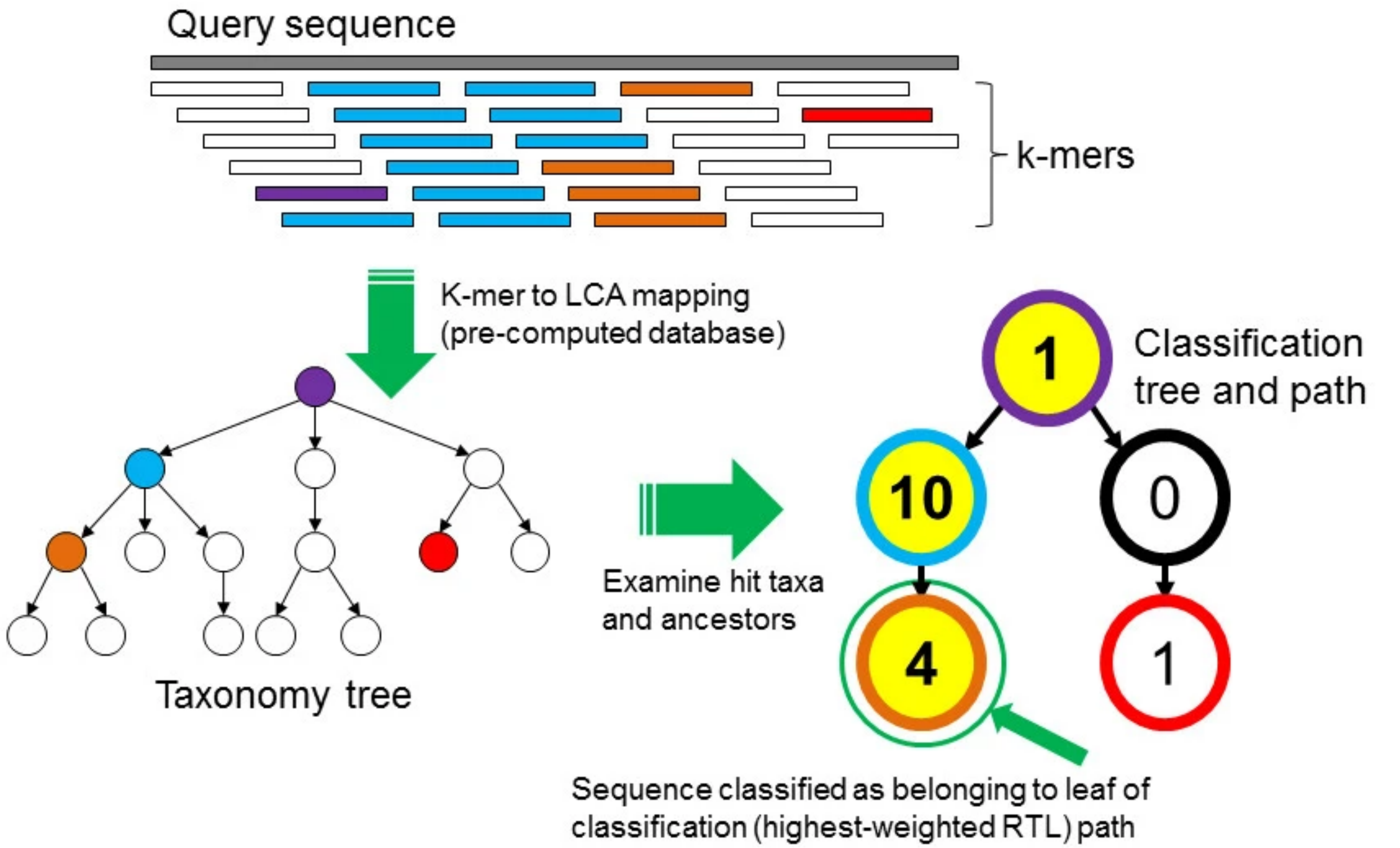


Get genes present in all “strains”



Assembly-free approaches

Kraken



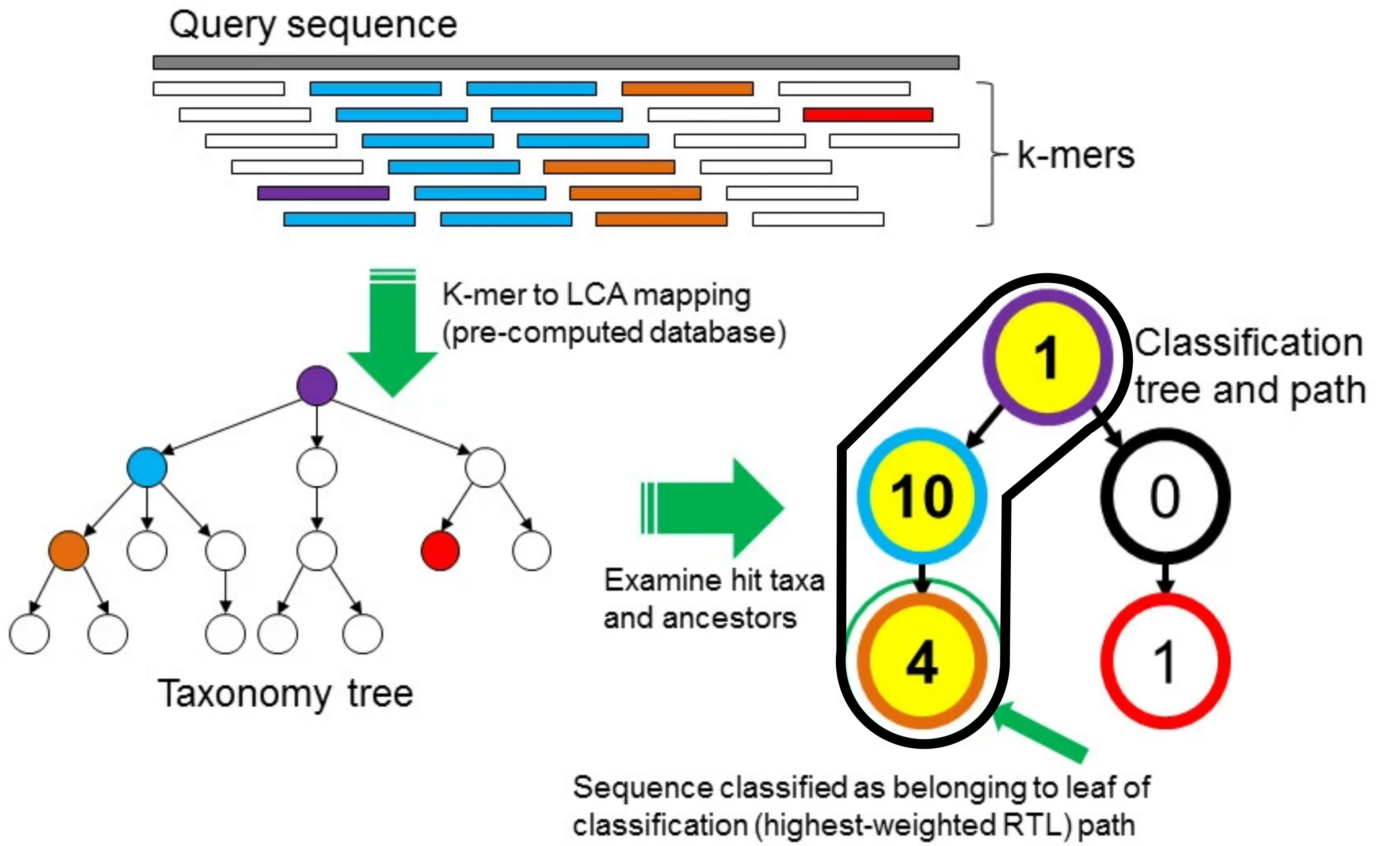
(Wood et al 2014)

Metagenomic reads are split into *k*-mers and mapped onto a pre-computed tree using “Lowest Common Ancestor" algorithm (LCA)

The taxonomic composition of each sample is reconstructed following the highest “Root To Leaf” (RTL) path

Assembly-free approaches

Kraken



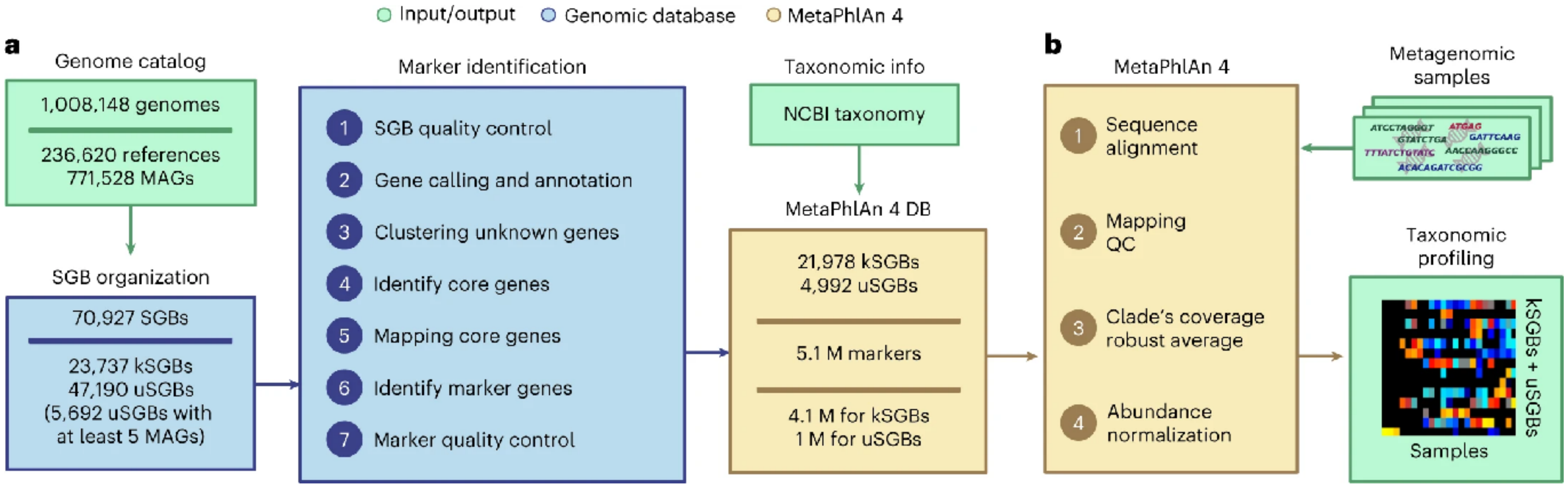
(Wood et al 2014)

Metagenomic reads are split into *k*-mers and mapped onto a pre-computed tree using “Lowest Common Ancestor" algorithm (LCA)

The taxonomic composition of each sample is reconstructed following the <u>highest “Root To Leaf” (RTL) path</u>

Assembly-free approaches

MetaPhlAn



○ Input/output ○ Genomic database ○ MetaPhlAn 4

a

Genome catalog

1,008,148 genomes

236,620 references
771,528 MAGs

SGB organization

70,927 SGBs

23,737 kSGBs
47,190 uSGBs
(5,692 uSGBs with
at least 5 MAGs)

Marker identification

1. SGB quality control
2. Gene calling and annotation
3. Clustering unknown genes
4. Identify core genes
5. Mapping core genes
6. Identify marker genes
7. Marker quality control

Taxonomic info

NCBI taxonomy

MetaPhlAn 4 DB

21,978 kSGBs
4,992 uSGBs

5.1 M markers

4.1 M for kSGBs
1 M for uSGBs

b

MetaPhlAn 4

1. Sequence alignment
2. Mapping QC
3. Clade's coverage robust average
4. Abundance normalization

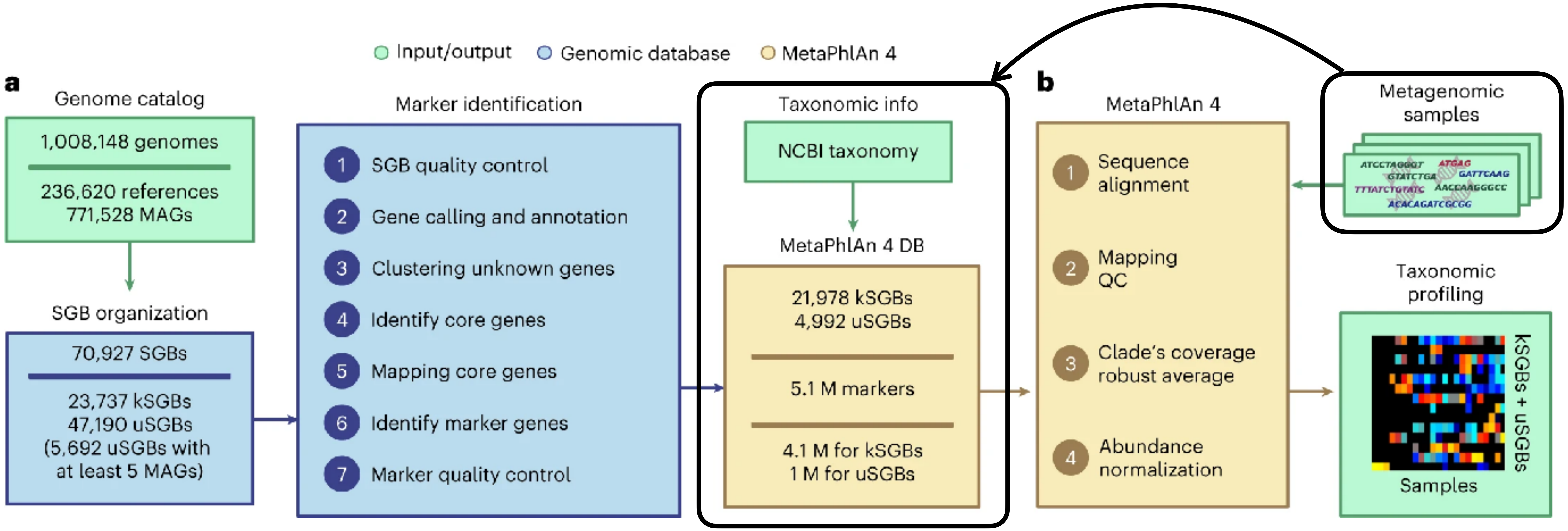
Metagenomic samples

Taxonomic profiling

(Blanco-Míguez et al 2023)

Assembly-free approaches

MetaPhlAn

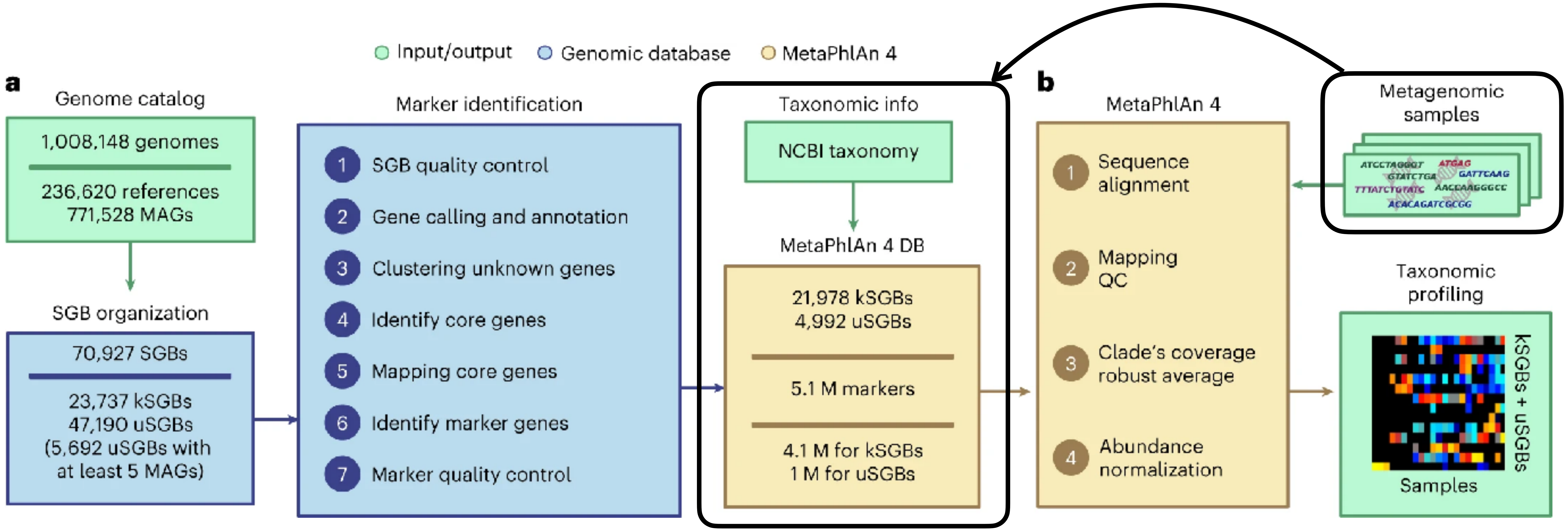


(Blanco-Míguez et al 2023)

Metagenomic reads are mapped to a taxonomically curated marker gene set

Assembly-free approaches

MetaPhlAn



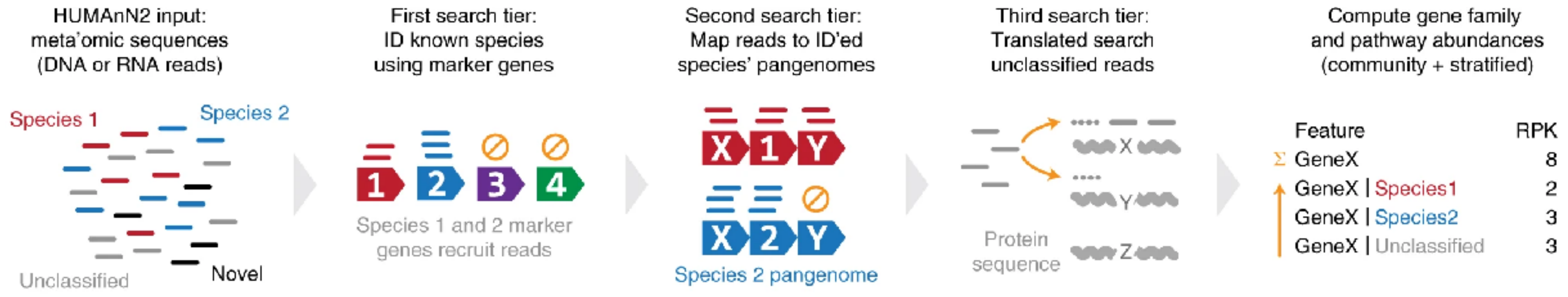
(Blanco-Míguez et al 2023)

Metagenomic reads are mapped to a taxonomically curated marker gene set

→ a set of unique genes for each clade:
one gene = one taxon

Assembly-free approaches

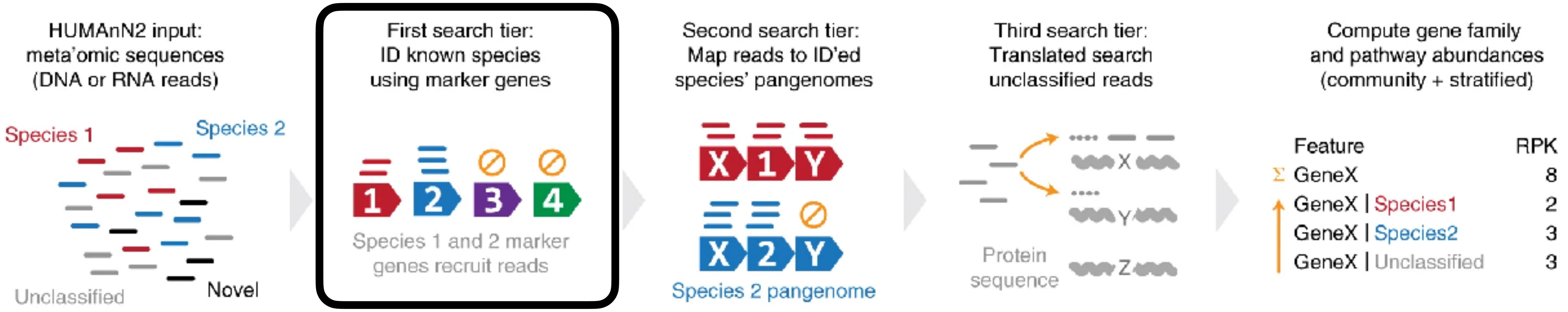
HUMAnN



(Franzosa et al 2018)

Assembly-free approaches

HUMAnN

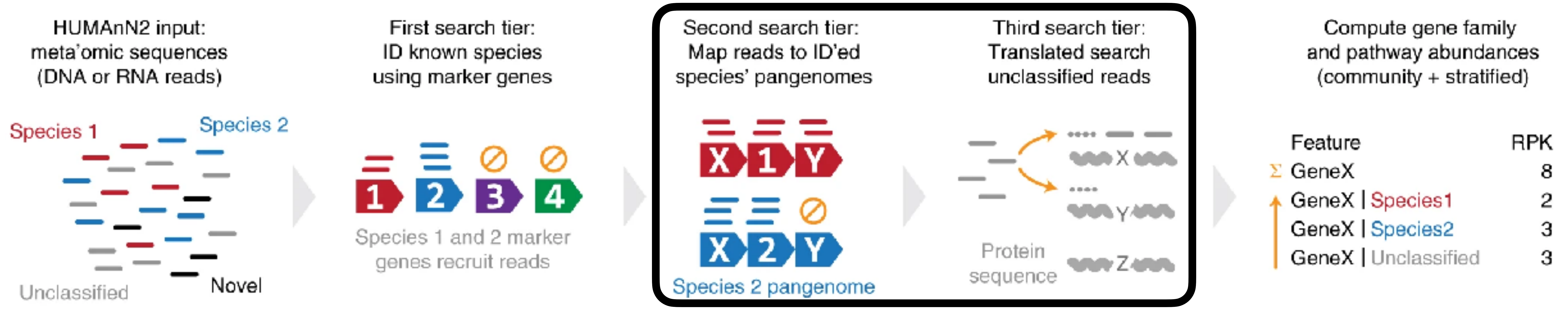


(Franzosa et al 2018)

Taxonomic profiling obtained with MetaPhlAn are used to subset functional genes

Assembly-free approaches

HUMAnN

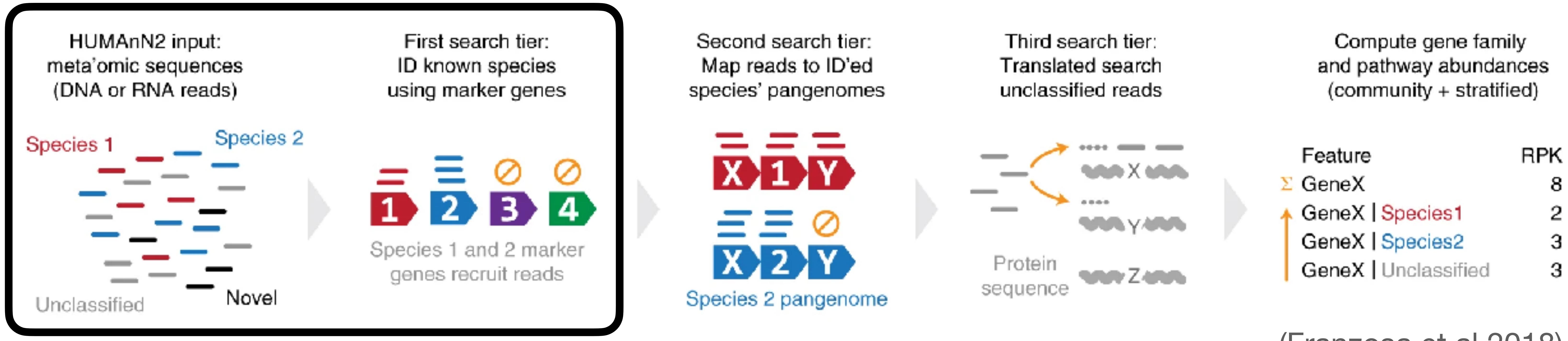


(Franzosa et al 2018)

Taxonomic profiling obtained with MetaPhlAn are used to subset functional genes

A translated search is then performed against the reduced database produced above

What we are going to do today

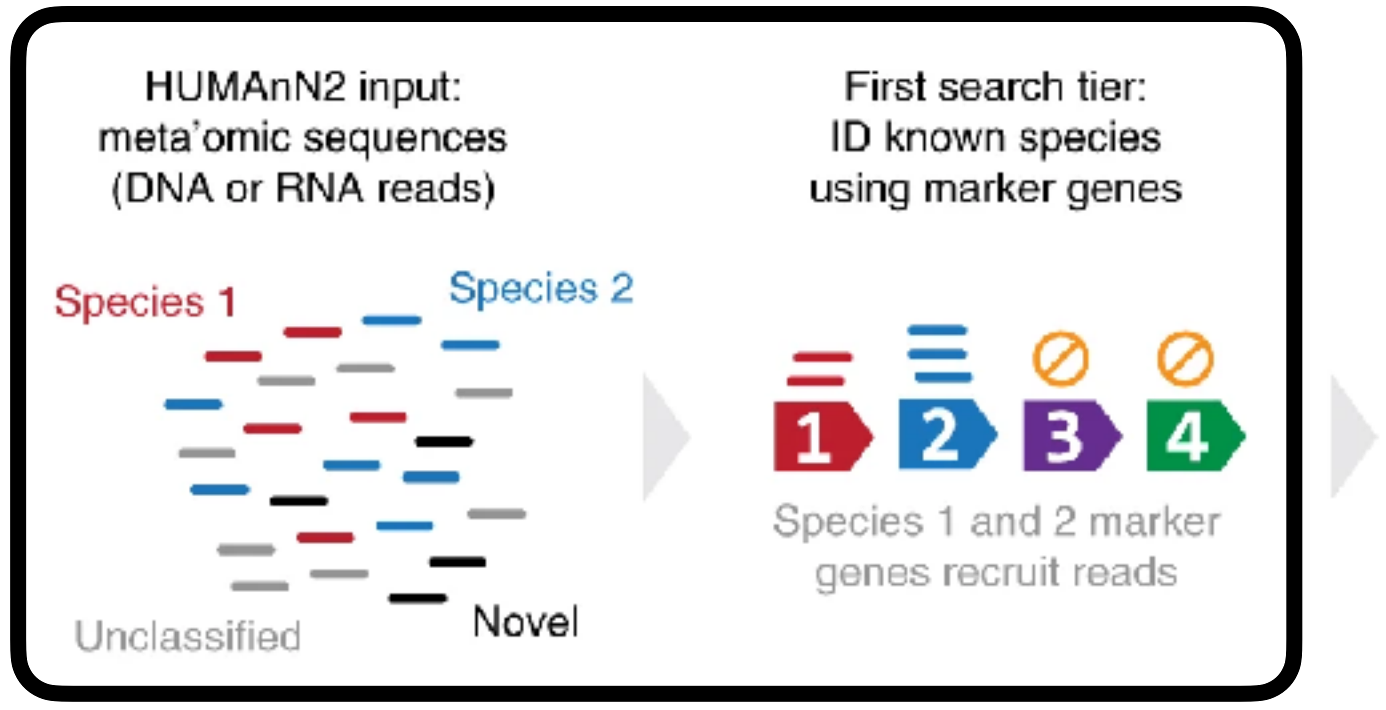


Feature	RPK
GeneX	8
GeneX | Species1	2
GeneX | Species2	3
GeneX | Unclassified	3

(Franzosa et al 2018)

Taxonomic profiling with MetaPhlAn

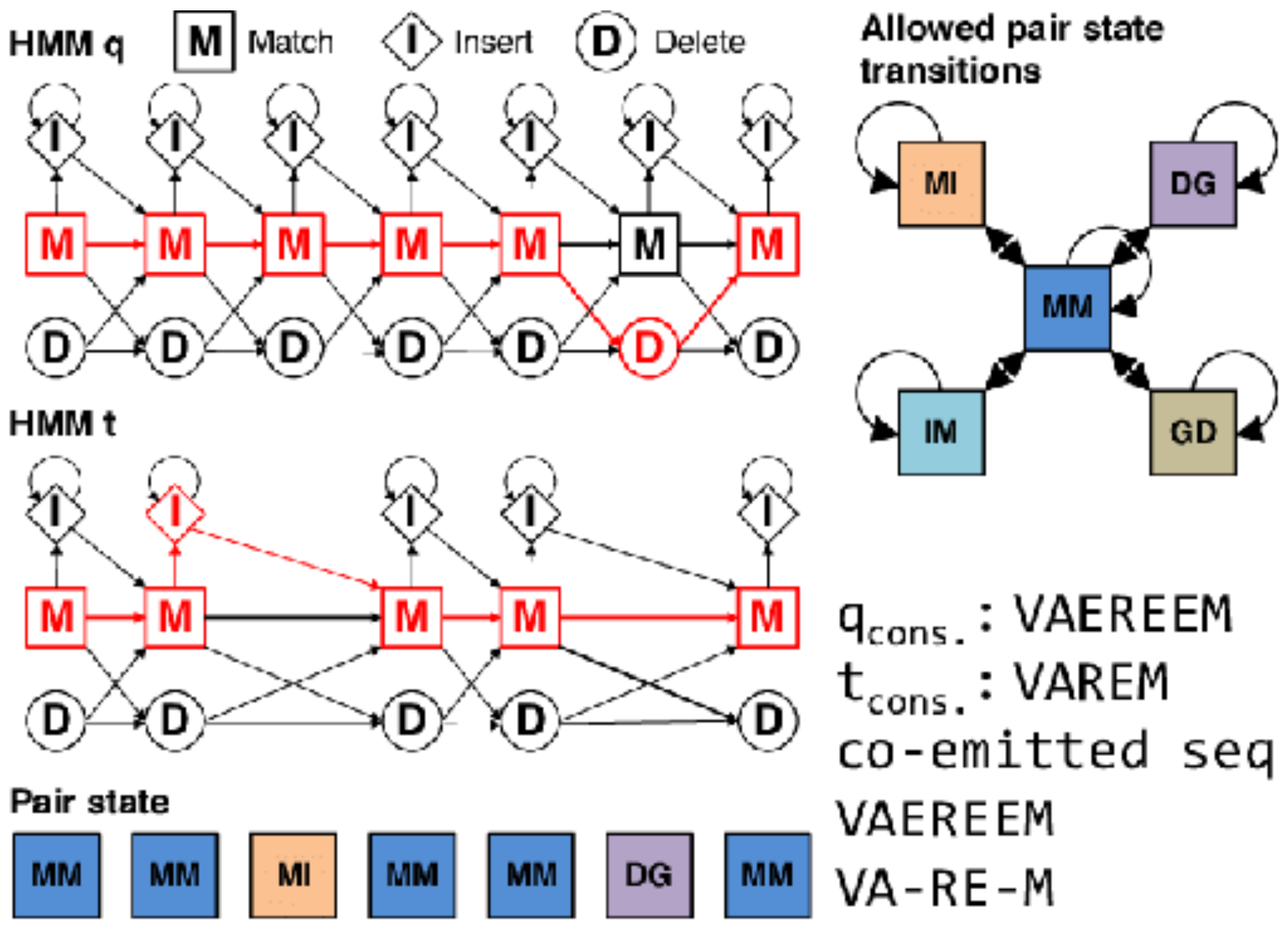
What we are going to do today



HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)

First search tier:
ID known species
using marker genes

Species 1 and 2 marker genes recruit reads

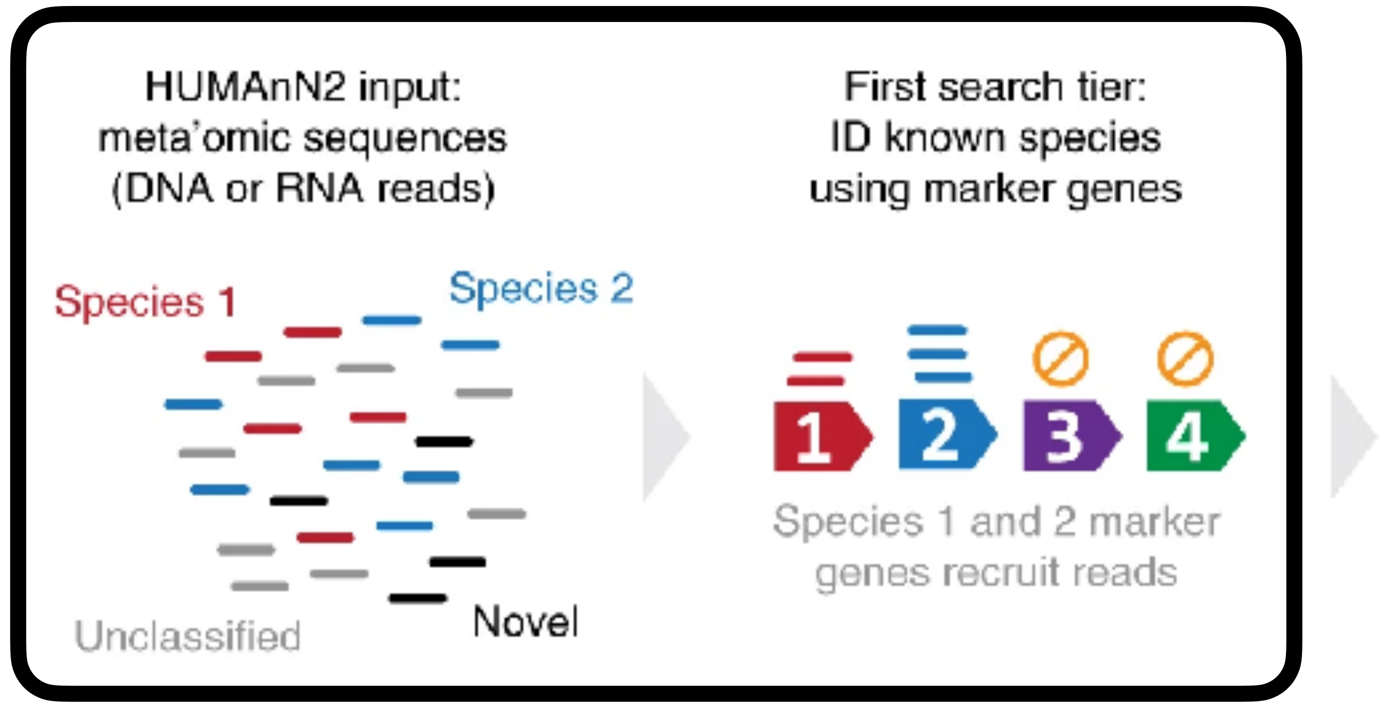


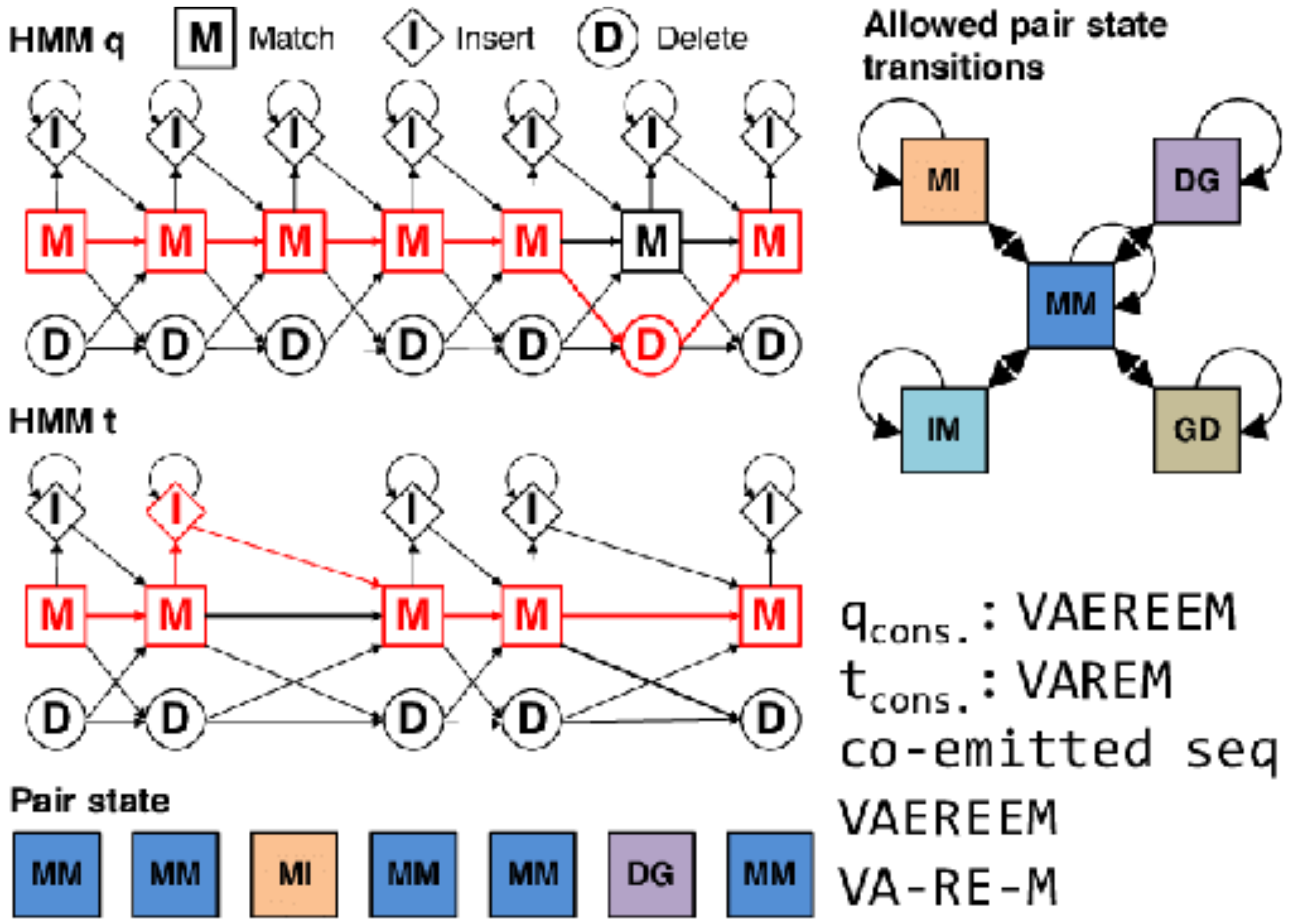
$q_{cons.}$: VAEREEM
$t_{cons.}$: VAREM
co-emitted seq
VAEREEM
VA-RE-M

Taxonomic profiling with MetaPhlAn

Inspecting functions with HMMER

What we are going to do today





Taxonomic profiling with MetaPhlAn

Inspecting functions with HMMER