Metagenome-assembled genomes (MAGs) reconstruction

Exploring microbiomes with cultivation-independent genome-resolved metagenomics

Samuel Miravet-Verde
smiravet@ethz.ch

Guillem Salazar
guillems@ethz.ch

Adapted from Lucas Paoli
paolil@ethz.ch

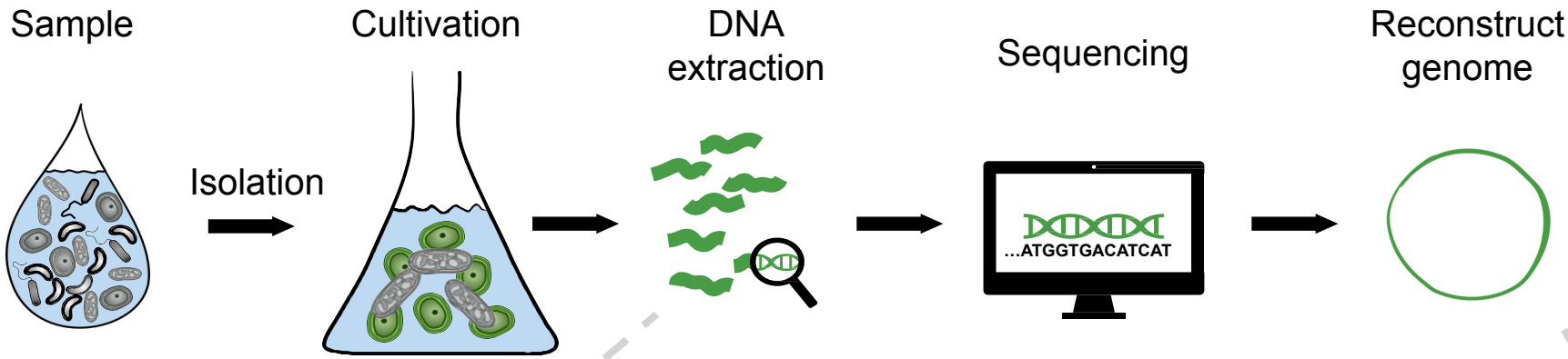
Block Course Fall 2022
551-1119-00L
Microbial Community Genomics

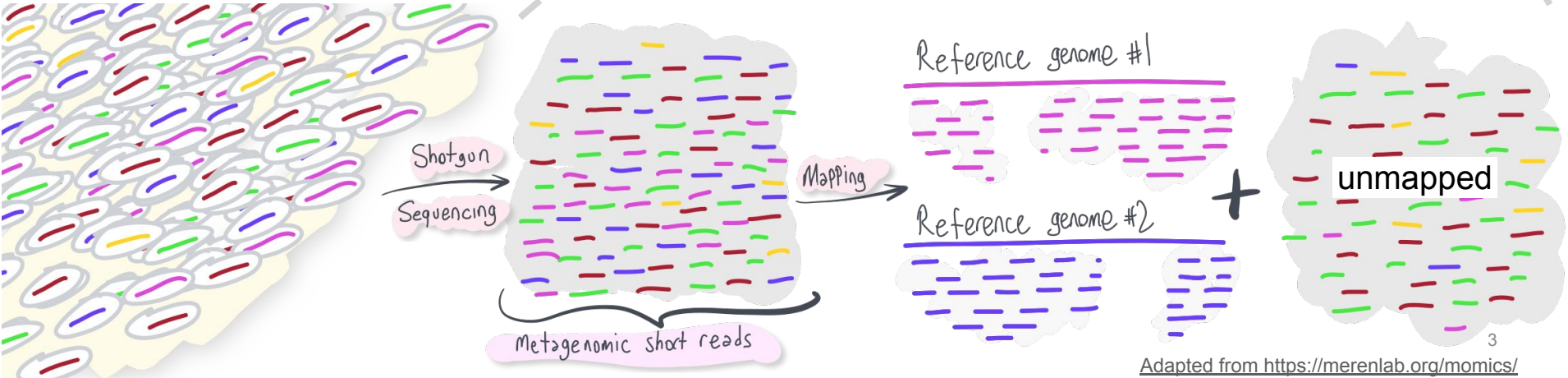
ETH zürich
D BIOL

10-Nov-22

- Ubiquitous across earth's ecosystems
- Support global food webs
- Underpin biogeochemical cycles
- Determine Host's health and disease
- …
- Untapped metabolic diversity

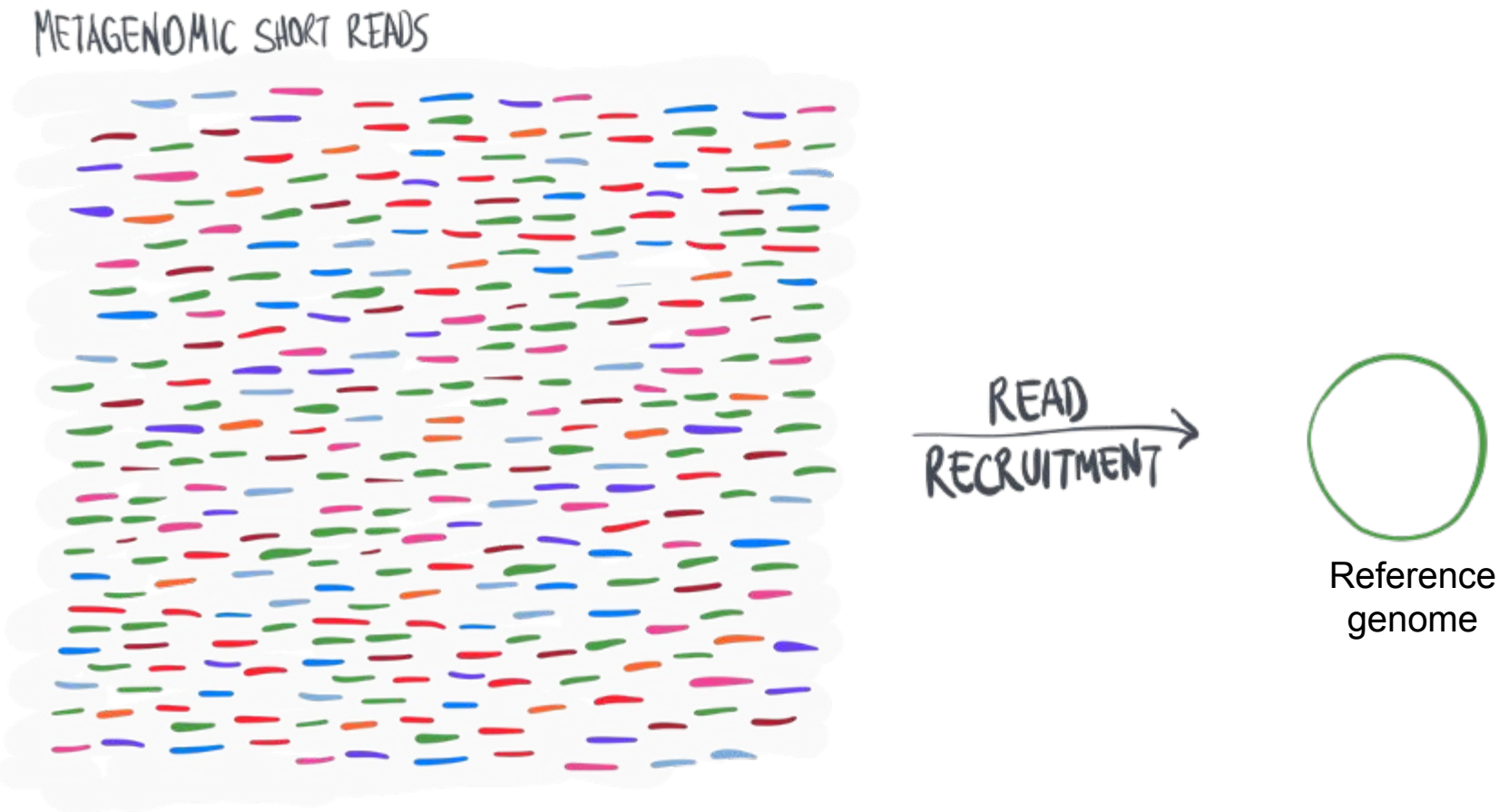
Traditional microbiology | | Culture-based microbiology



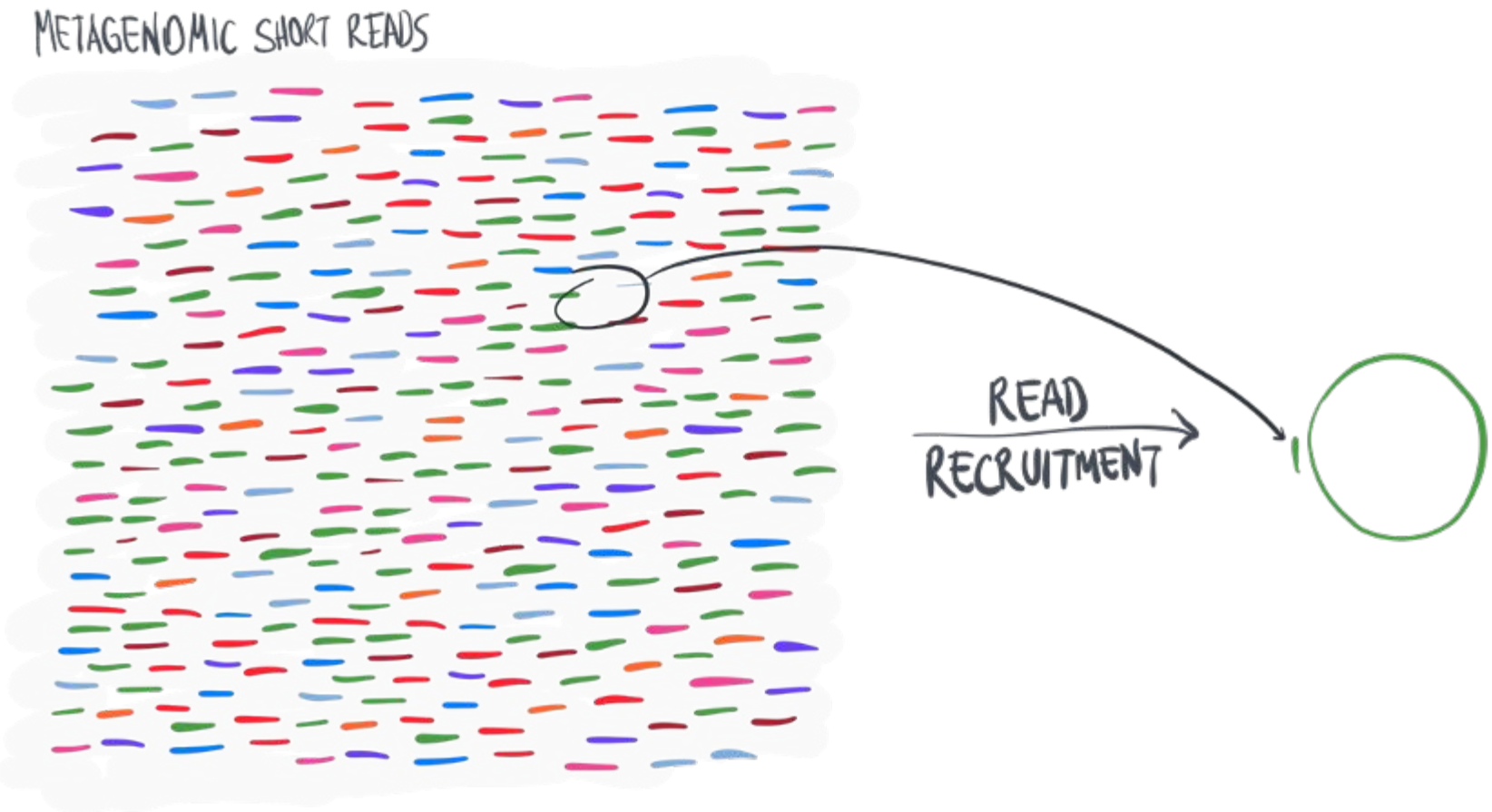


Adapted from https://merenlab.org/momics/

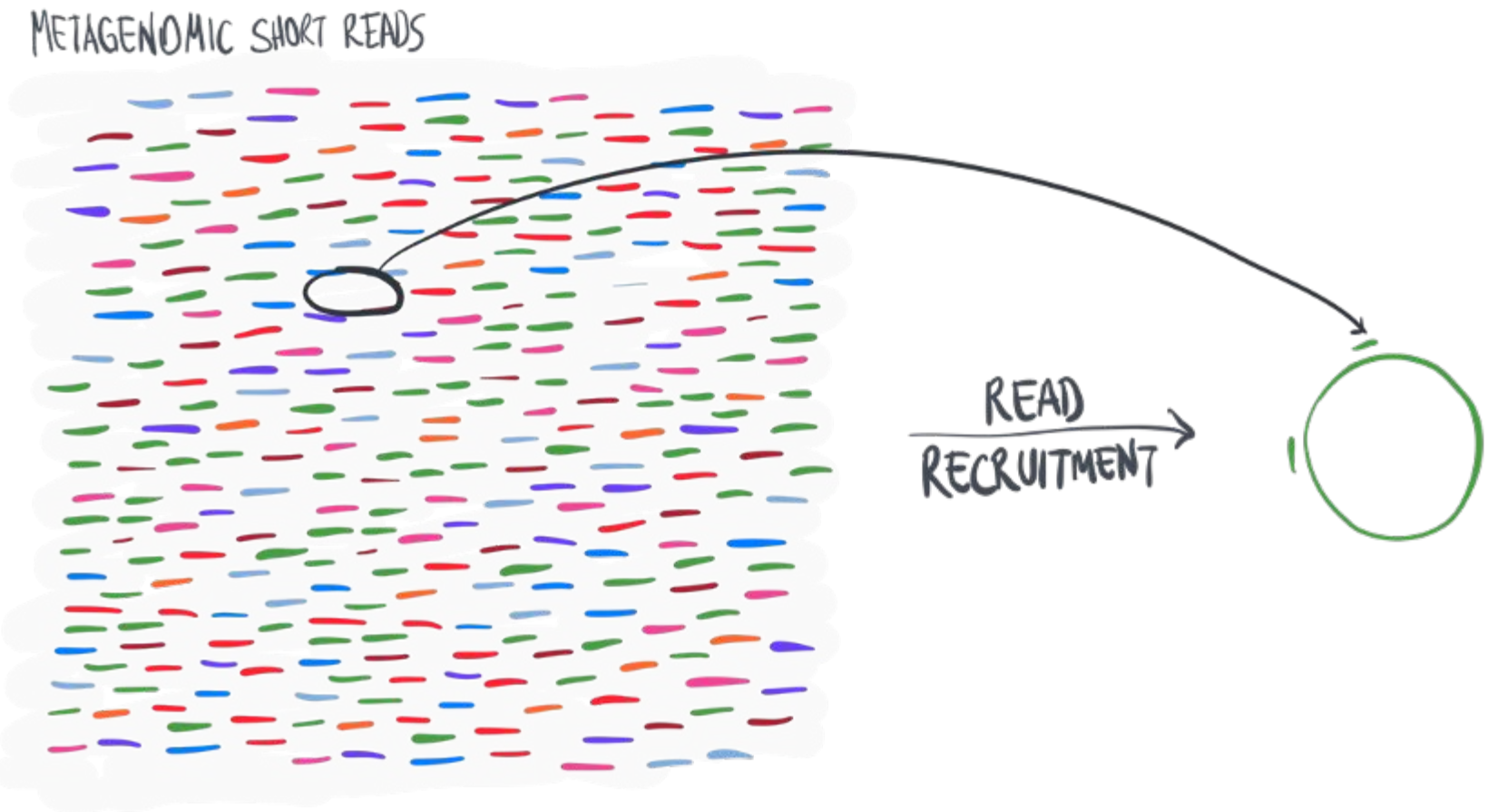
Metagenomics | | Mapping to a reference



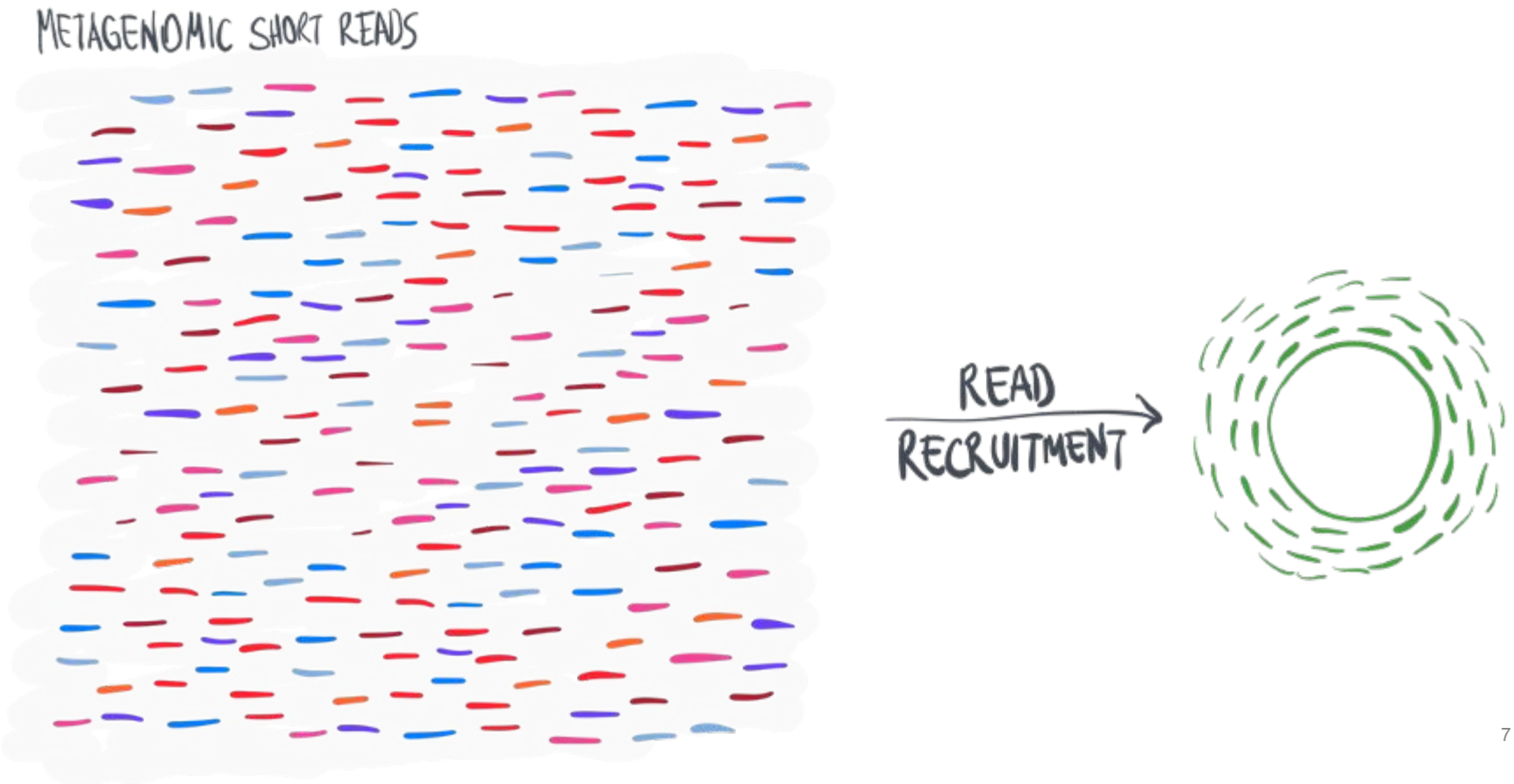
Metagenomics|| Mapping to a reference



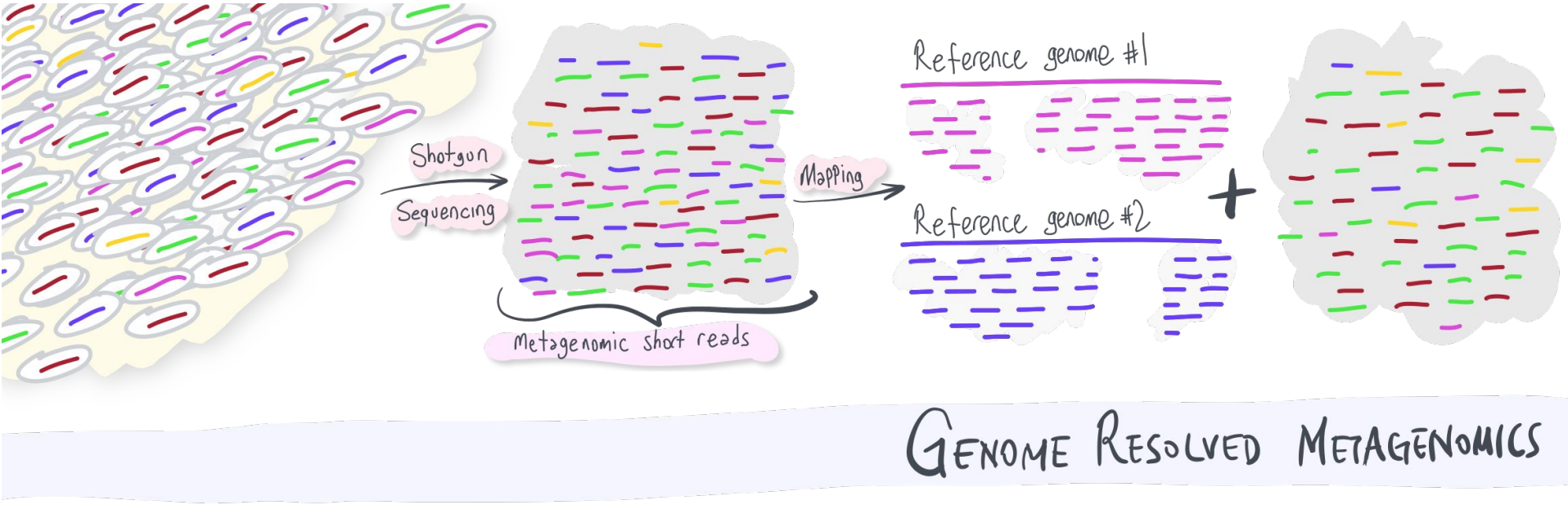
Metagenomics|| Mapping to a reference

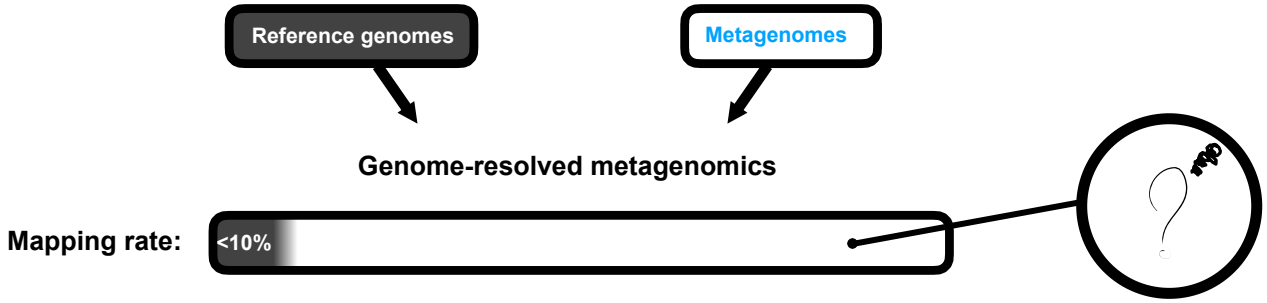


Metagenomics|| Mapping to a reference



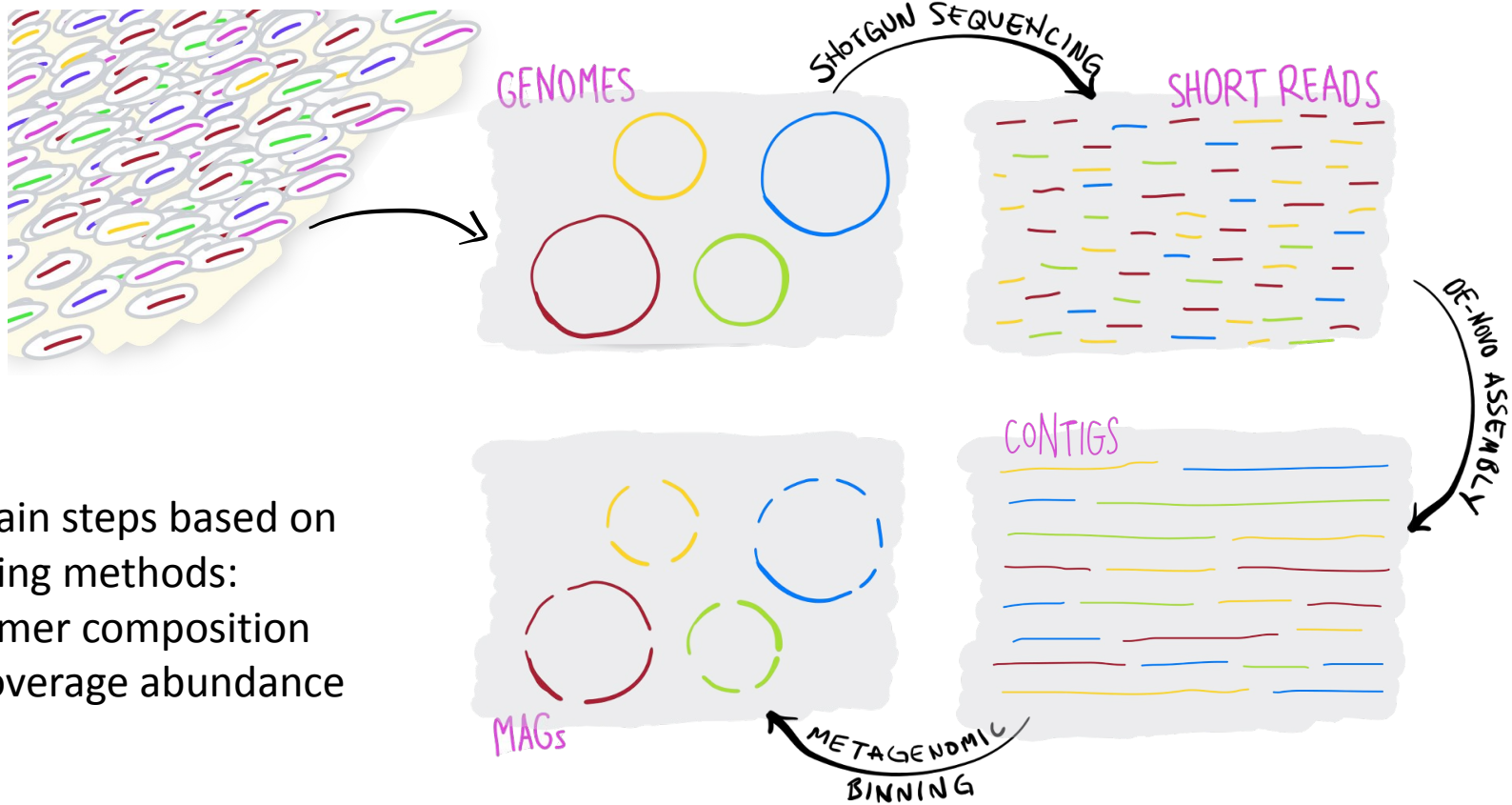
Metagenomics | | Mapping rates





Current state| | Culture-independent microbiology

Current approaches DO NOT require to isolate organisms or reference genomes:



Two main steps based on clustering methods:

- k-mer composition
- coverage abundance

Sequence composition || Computing k-mer frequencies

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

k=2

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT

k=2

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

k=2

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1

k=2

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
1	0	2	2	1	0	0	2	2	2	2	3	1	2	4	10

k=2

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
1	0	2	2	1	0	0	2	2	2	2	3	1	2	4	10

GAAGCACAAAAGAAACTCCTTAATCATGCCAAAAC

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
10	3	2	2	4	2	0	2	2	2	6	0	1	2	1	1

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC
GAAGCAGAAAAGAAACTCCTTAATCATGCCAAAAC

AA	AC	AG	GA	CA	CC	CG	GC	AT	TA
11	3	4	4	5	2	0	2	2	1

→ PALINDROMES :)

k=2

GTTTTGGCATGATTAAGGAGTTTCTTTTGTGCTTC

	AA	AC	AG	GA	CA	CC	CG	GC	AT	TA
X	11	3	4	4	5	2	0	2	2	1
Y										
Z										
L										
K										
M										

k=2

ACTTCCGCAGTCGGGCATTACGCGTTGTGGAATGA

	AA	AC	AG	GA	CA	CC	CG	GC	AT	TA
X	11	3	4	4	5	2	0	2	2	1
Y	4	5	2	4	5	4	4	3	2	1
Z										
L										
K										
M										

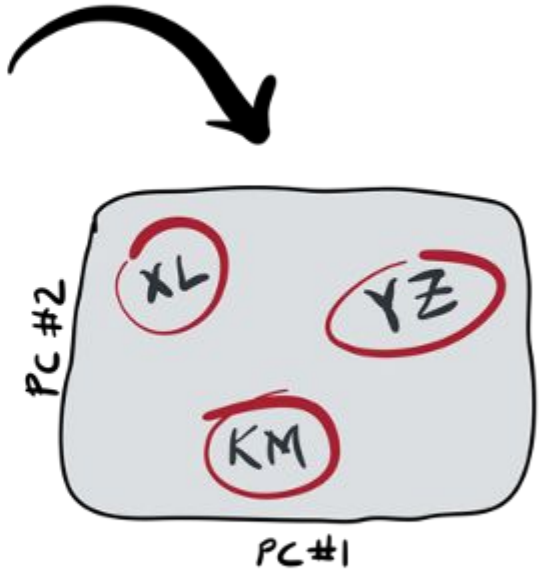
k=2

GGGCCTGCGCCGGTCCAGTCACCCGGCTGCGACCT

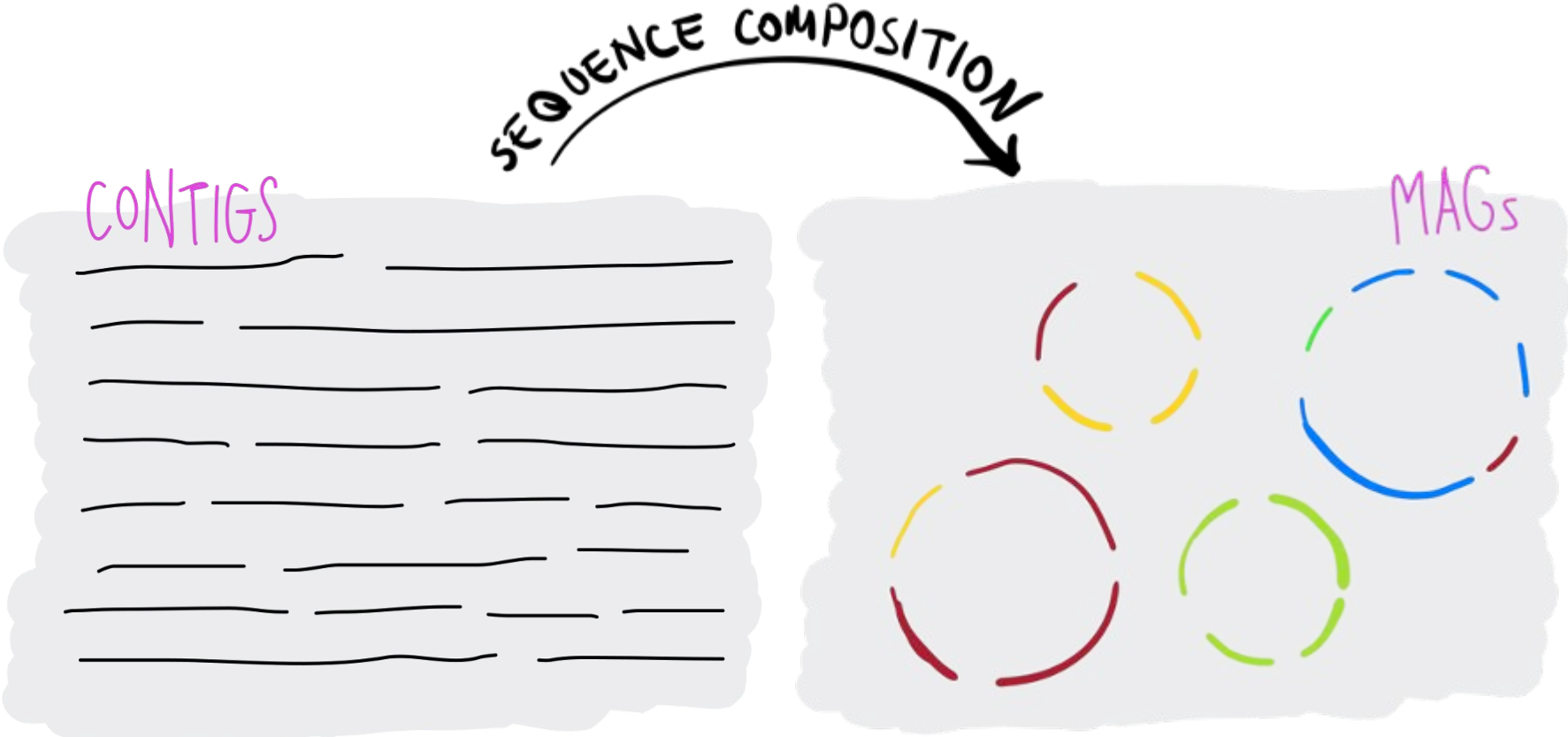
	AA	AC	AG	GA	CA	CC	CG	GC	AT	TA
X	11	3	4	4	5	2	0	2	2	1
Y	4	5	2	4	5	4	4	3	2	1
Z	4	5	3	2	4	1	5	5	2	3
L	11	6	3	2	2	3	2	1	1	4
K	1	1	2	2	1	8	9	10	0	0
M	0	4	4	3	4	10	4	5	0	0

k=2

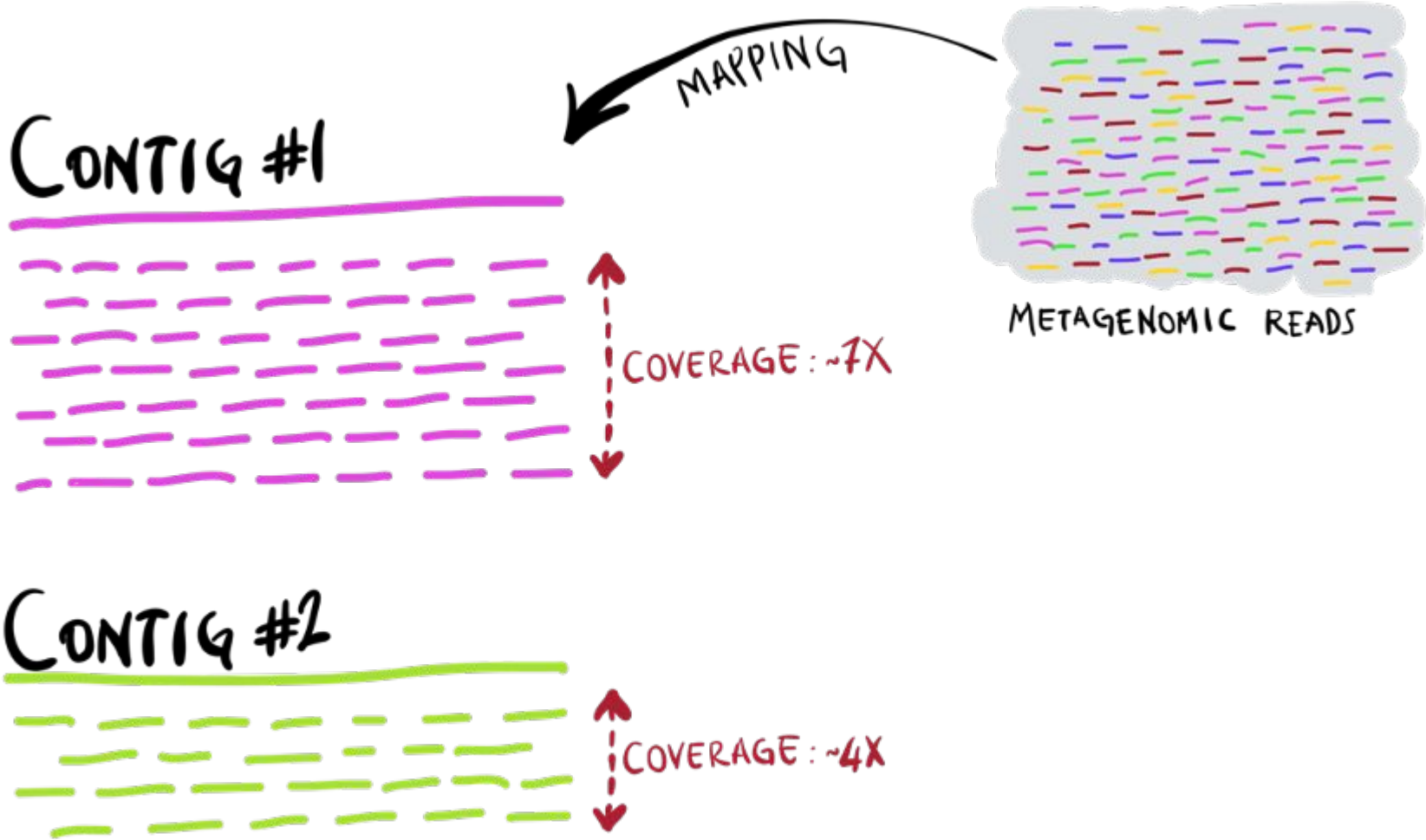
	AA	AC	AG	GA	CA	CC	CG	GC	AT	TA
X	11	3	4	4	5	2	0	2	2	1
Y	4	5	2	4	5	4	4	3	2	1
Z	4	5	3	2	4	1	5	5	2	3
L	11	6	3	2	2	3	2	1	1	4
K	1	1	2	2	1	8	9	10	0	0
M	0	4	4	3	4	10	4	5	0	0

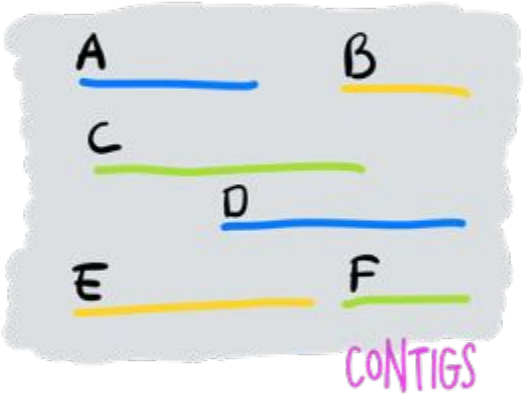


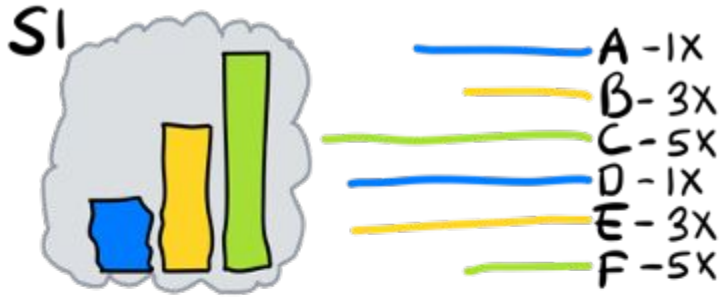
k=2


SEQUENCE COMPOSITION
CONTIGS
MAGs

Abundance correlation|| Counting the content of a genome

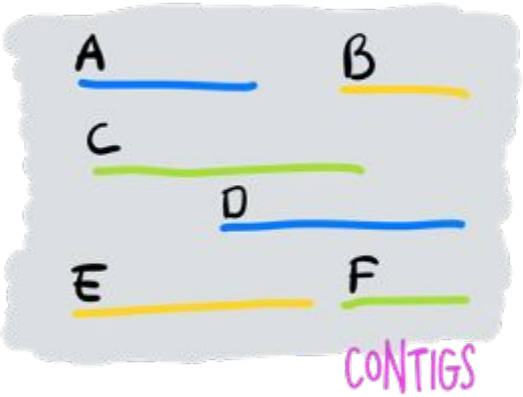

MAPPING
METAGENOMIC READS
CONTIG #1
COVERAGE : ~7X
CONTIG #2
COVERAGE : ~4X

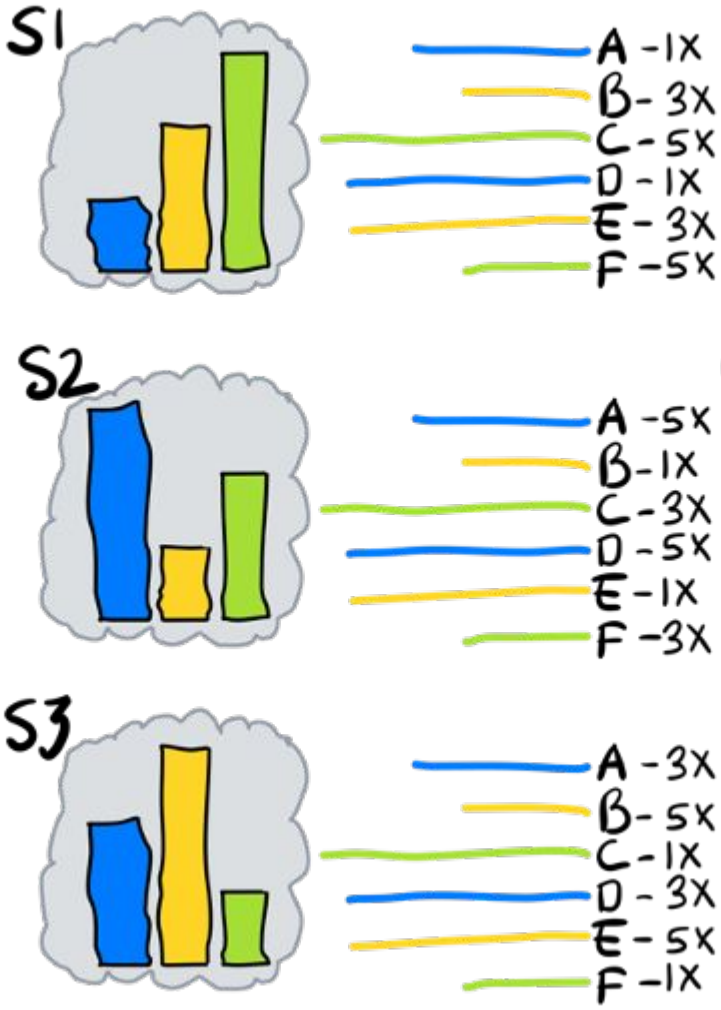

A
B
C
D
E
F
CONTIGS


S1
A -1X
B- 3X
C- 5X
D -1X
E- 3X
F -5X

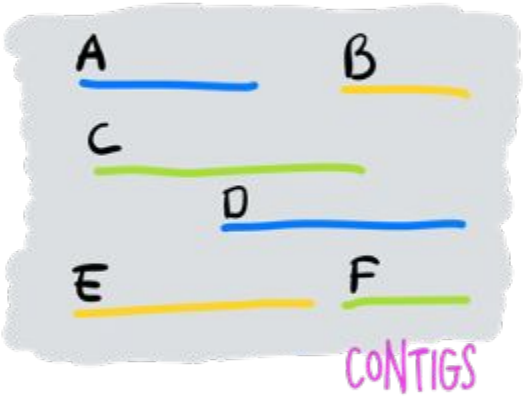

S2
A -5X
B-1X
C-3X
D-5X
E-1X
F -3X

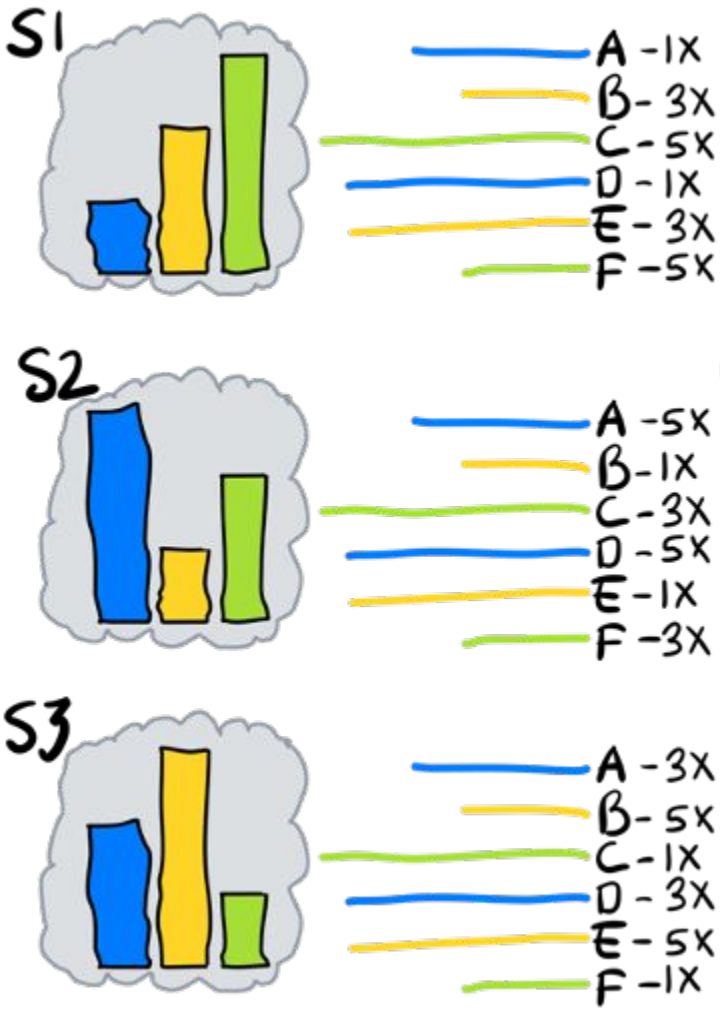

S3
A -3X
B- 5X
C-1X
D- 3X
E- 5X
F -1X



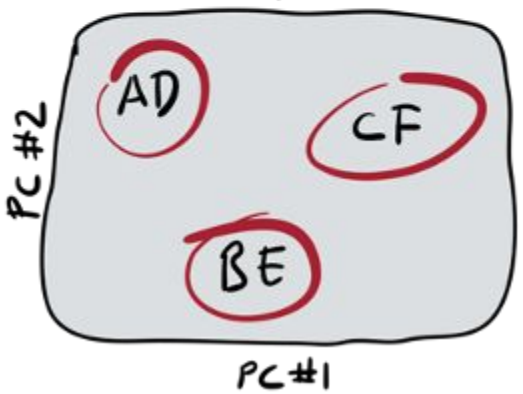


	A	B	C	D	E	F
S1	1	3	5	1	3	5
S2	5	1	3	5	1	3
S3	3	5	1	3	5	1

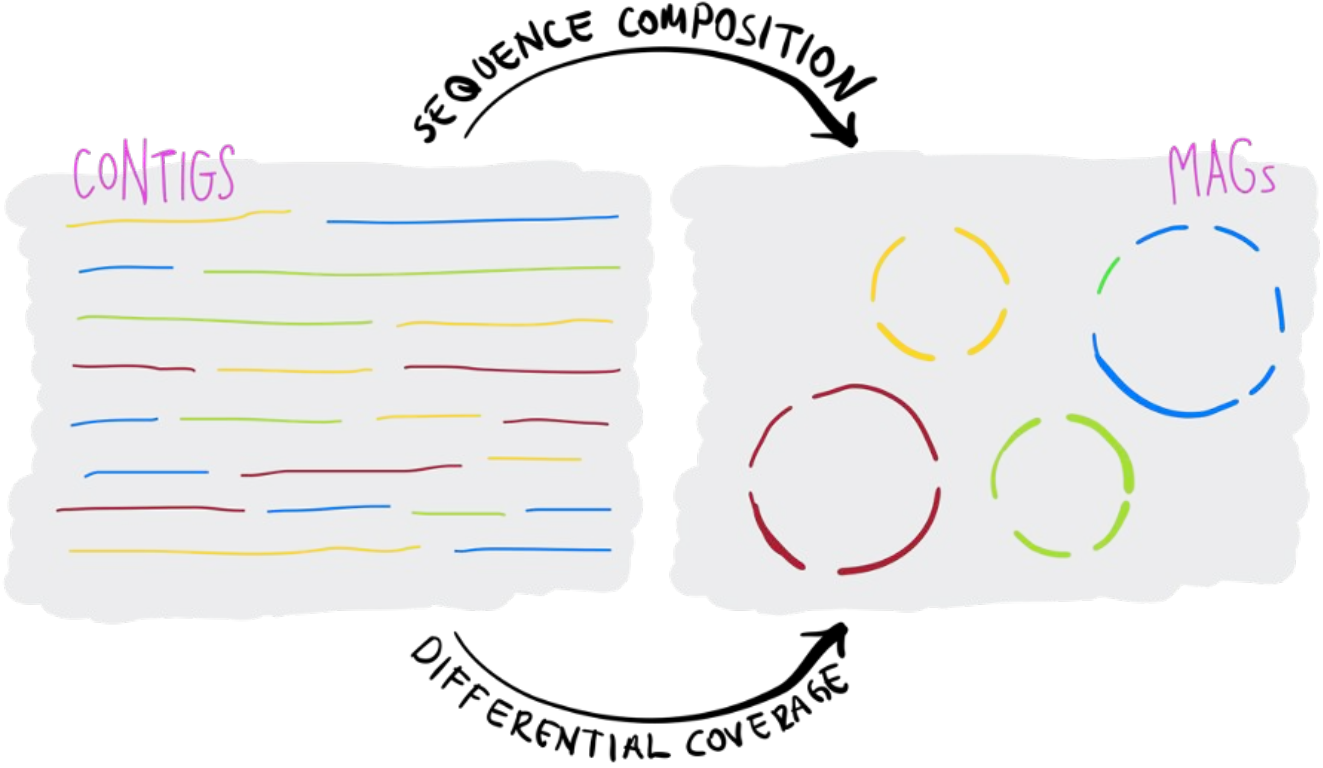




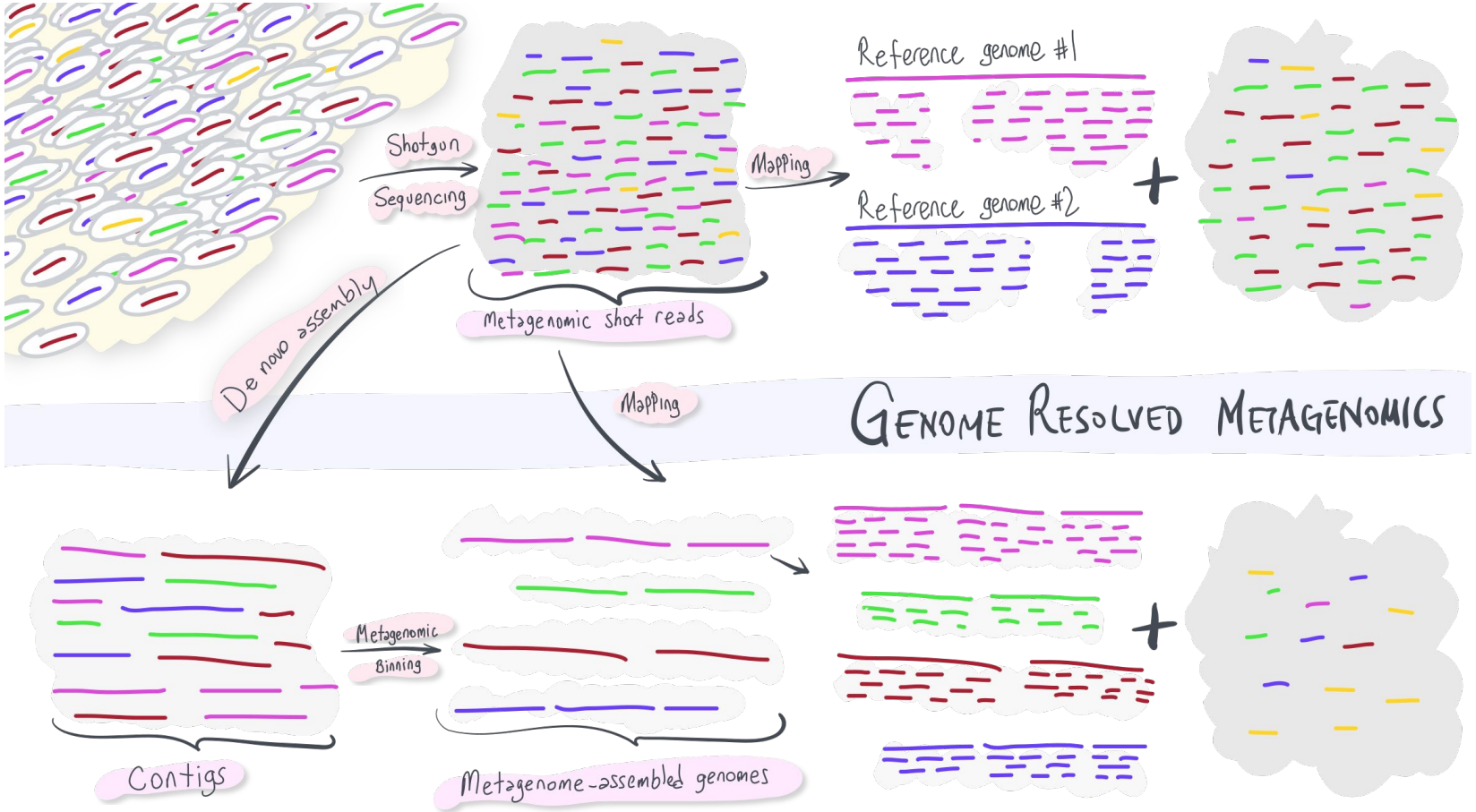
	A	B	C	D	E	F
S1	1	3	5	1	3	5
S2	5	1	3	5	1	3
S3	3	5	1	3	5	1



MAG reconstruction|| Sequence composition & diff. coverage



Metagenomics | | Summary

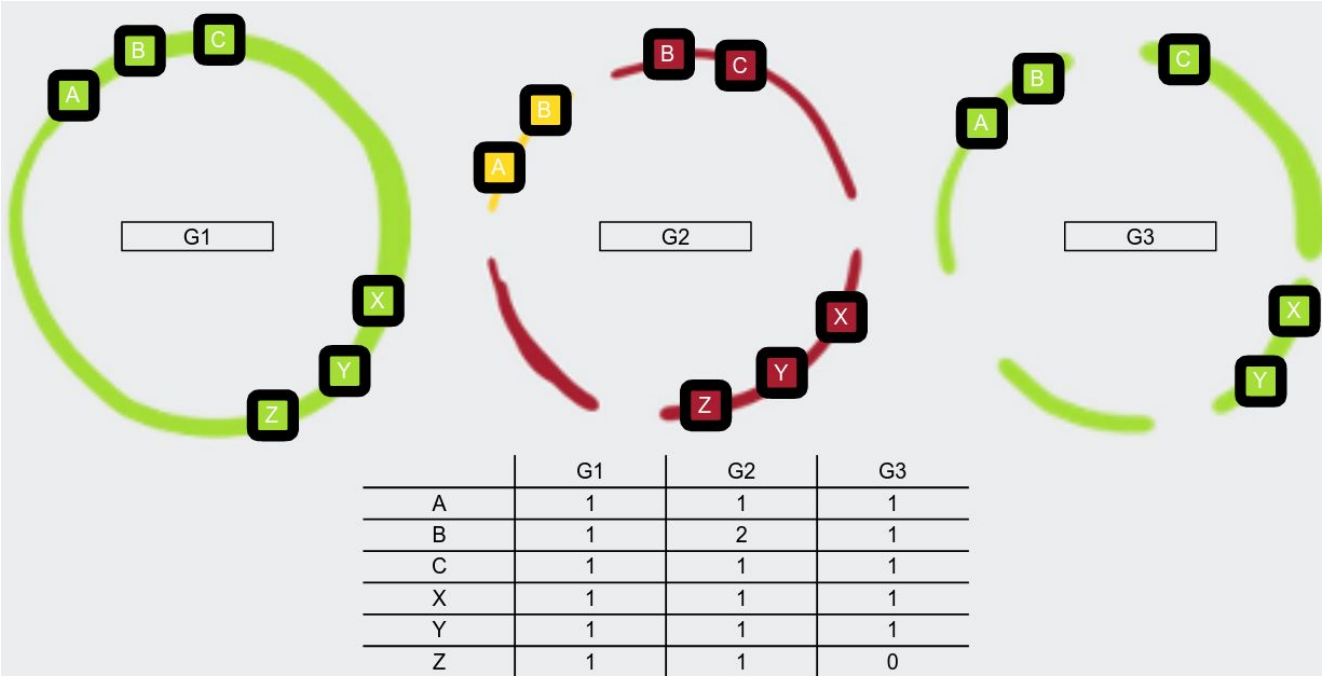


Evaluation of the reconstruction|| How complete are our results?

Evaluation of the reconstruction|| How complete & clean are our results?

Universal single-copy marker genes:

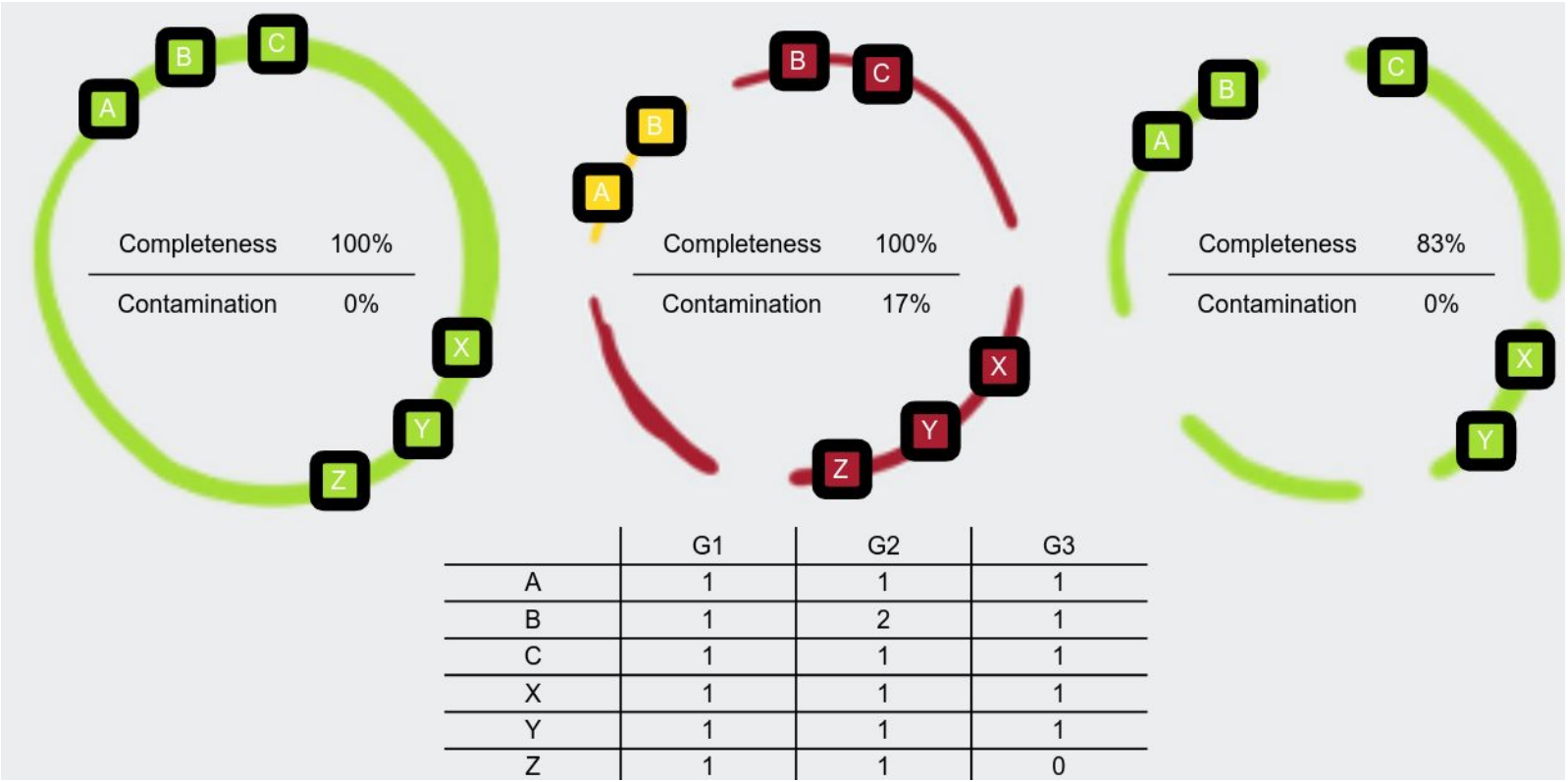
- genes present in every genome
- Between 40 and 120 genes for Bacteria/Archaea depending on cutoffs



	G1	G2	G3
A	1	1	1
B	1	2	1
C	1	1	1
X	1	1	1
Y	1	1	1
Z	1	1	0

Evaluation of the reconstruction|| How complete & clean are our results?

Completeness: % of single-copy marker genes found in the genome
Contamination: % of single-copy marker that are found >1



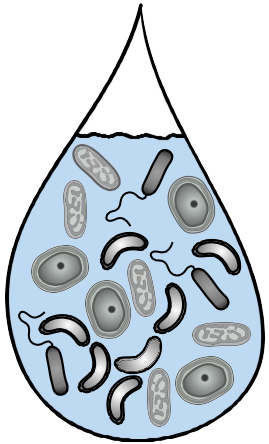
	G1	G2	G3
A	1	1	1
B	1	2	1
C	1	1	1
X	1	1	1
Y	1	1	1
Z	1	1	0

Applications|| the ocean microbiome

Applications|| the ocean microbiome

Oceans cover >70% of the planet

> **500,000 microbial cell per mL**
> **50% of the oxygen production**



nature

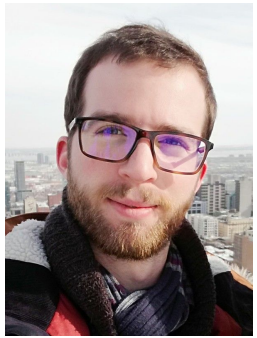
Explore content | About the journal | Publish with us

nature > articles > article

Article | Open Access | Published: 22 June 2022

Biosynthetic potential of the global ocean microbiome

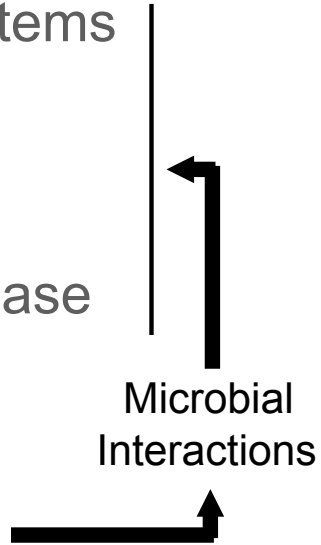
Lucas Paoli, Hans-Joachim Ruscheweyh, Clarissa C. Forneris, Florian Hubrich, Satria Kautsar, Agneya Bhushan, Alessandro Lotti, Quentin Clayssen, Guillem Salazar, Alessio Milanese, Charlotte I. Carlström, Chrysa Papadopoulou, Daniel Gehrig, Mikhail Karasikov, Harun Mustafa, Martin Larralde, Laura M. Carroll, Pablo Sánchez, Ahmed A. Zayed, Dylan R. Cronin, Silvia G. Acinas, Peer Bork, Chris Bowler, Tom O. Delmont, ... Shinichi Sunagawa + Show authors



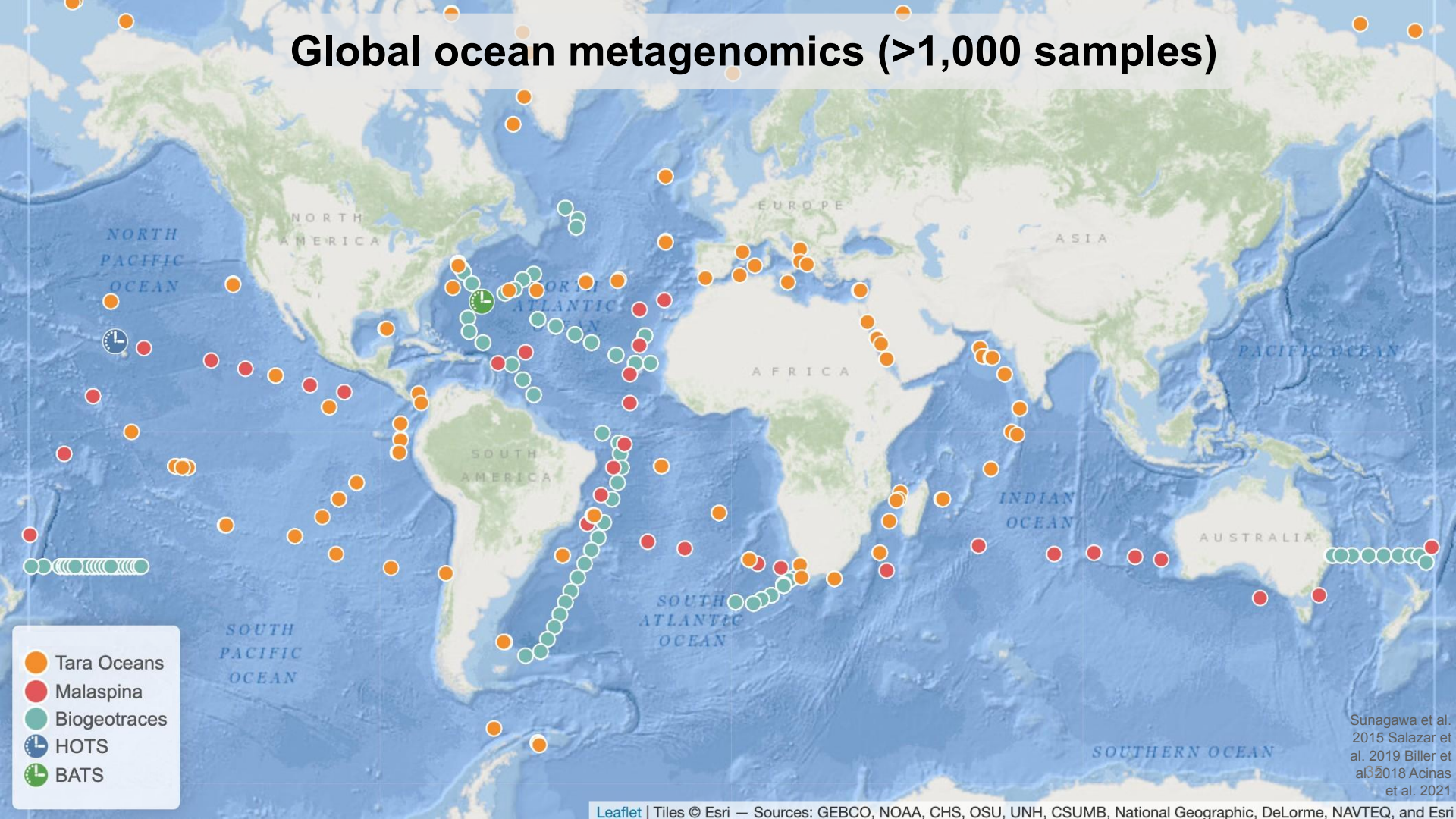
- Ubiquitous across earth's ecosystems
- Support global food webs
- Underpin biogeochemical cycles
- Determine Host's health and disease
- …
- **Untapped metabolic diversity**
  - **New enzymes**
  - **New natural products**

▼

Applications

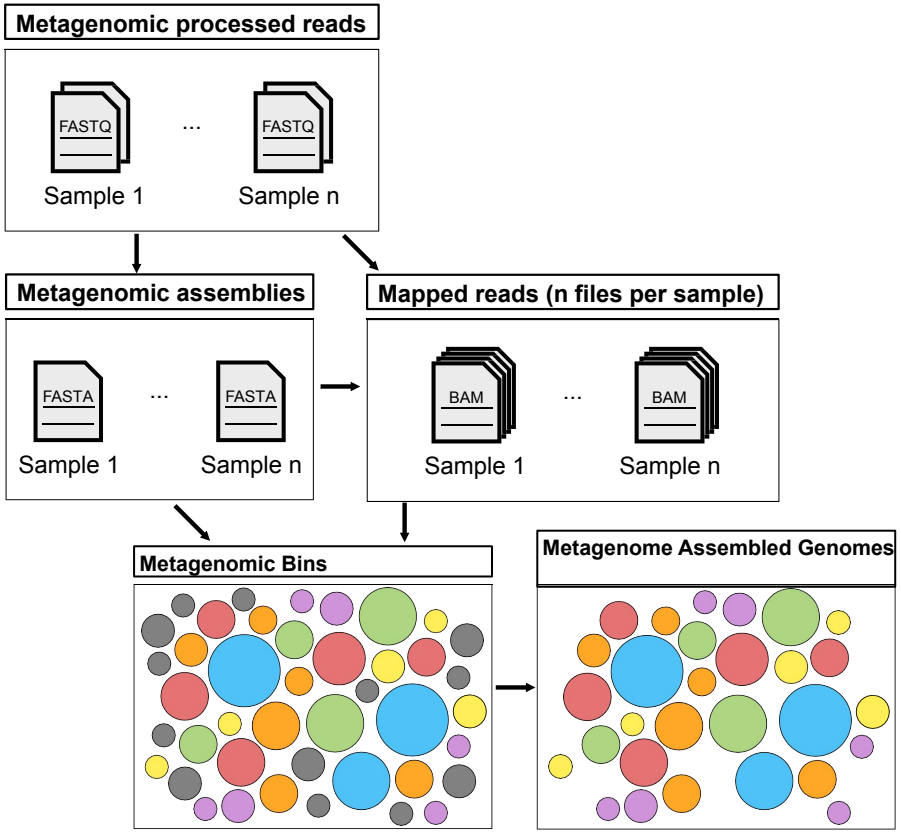


Global ocean metagenomics (>1,000 samples)

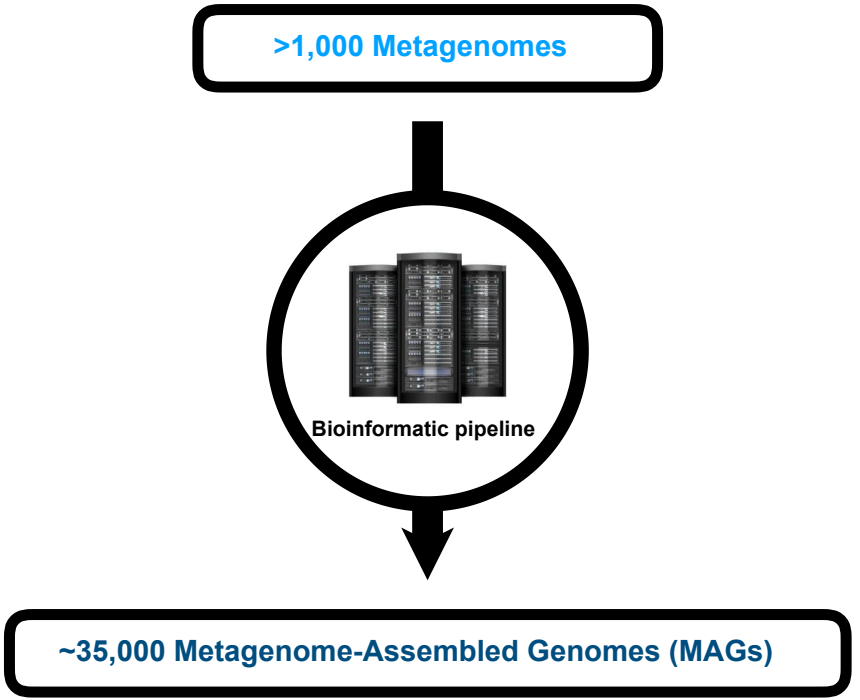


Ocean microbiome|| unveiling the hidden fraction of ocean microbes

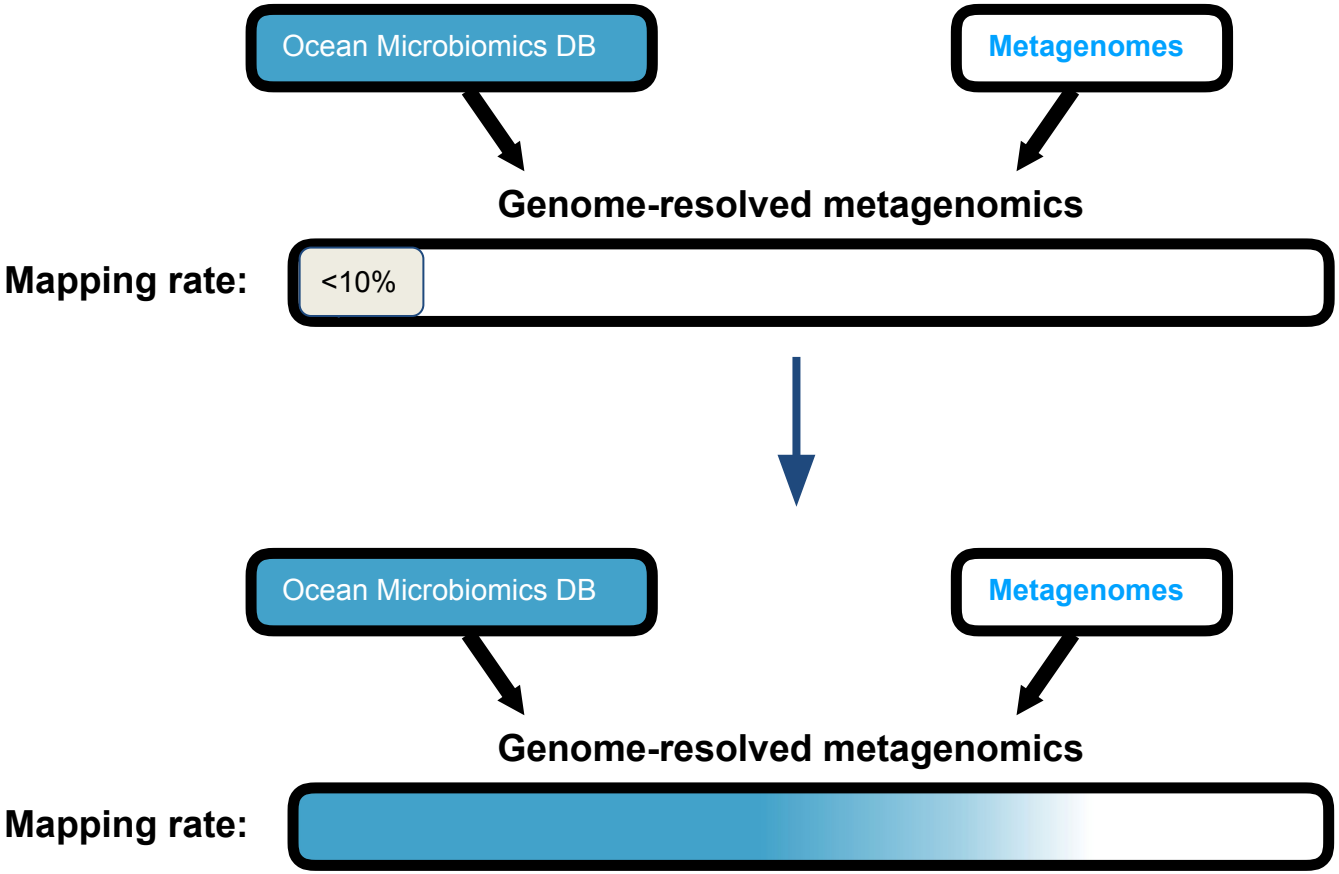
Main analysis idea:



Main outcome:



Ocean microbiome|| improving the representation of ocean microbial genomes



Ocean microbiome | | What can we find in these genomes?

~35k genomes

↓

~40k (mostly new) Biosynthetic Gene Clusters (BGCs)

↓

~7k Gene Cluster Families (GCFs)

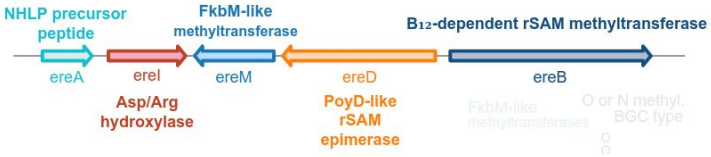
With large potential for new compounds

- RiPPs (Ribosomal Natural Products)
- Non−Ribosomal Peptide Synthases
- Type I Polyketide Synthases
- Type II & III Polyketide Synthases
- Terpenes
- Other

Are there BGC-rich microbial lineages to be discovered in the ocean?

- Eremiobacterota, uncultivated phylum with unsuspected BGC richness

Predict new enzymology



Conclusions | | what can we learn from sequencing uncultured microbes

- **Genome-resolved microbiomics** as a mean to **explore environmental microbiomes** and **discover novel enzymology** and **natural products**
- This approach provides **evolutionary and ecological context** to the **biosynthetic potential**
- Bioinformatics-guided **experimental characterization** is **necessary** and can still lead to **unpredicted discoveries**

Questions

Image: François