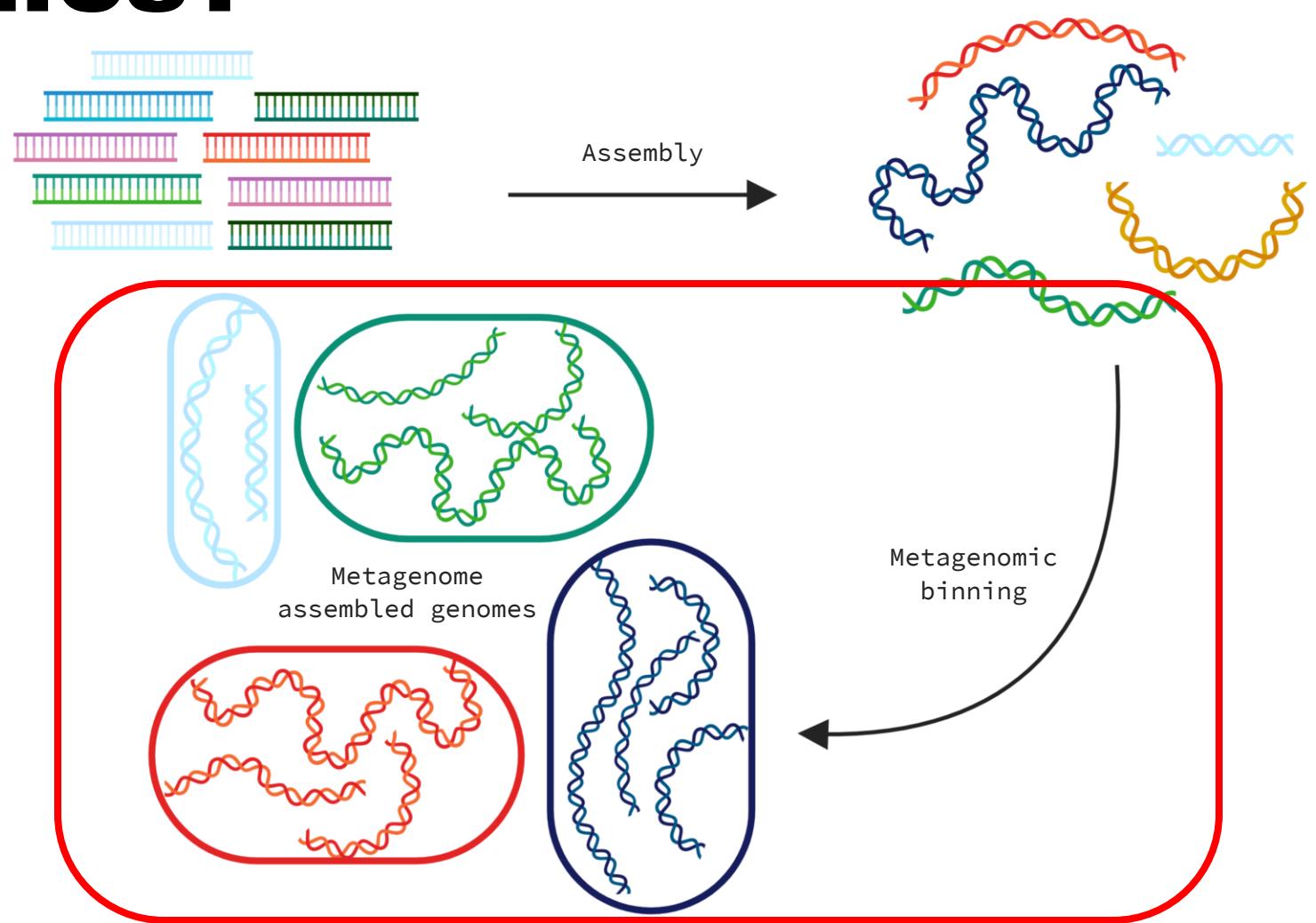


Genome-resolved metagenomics

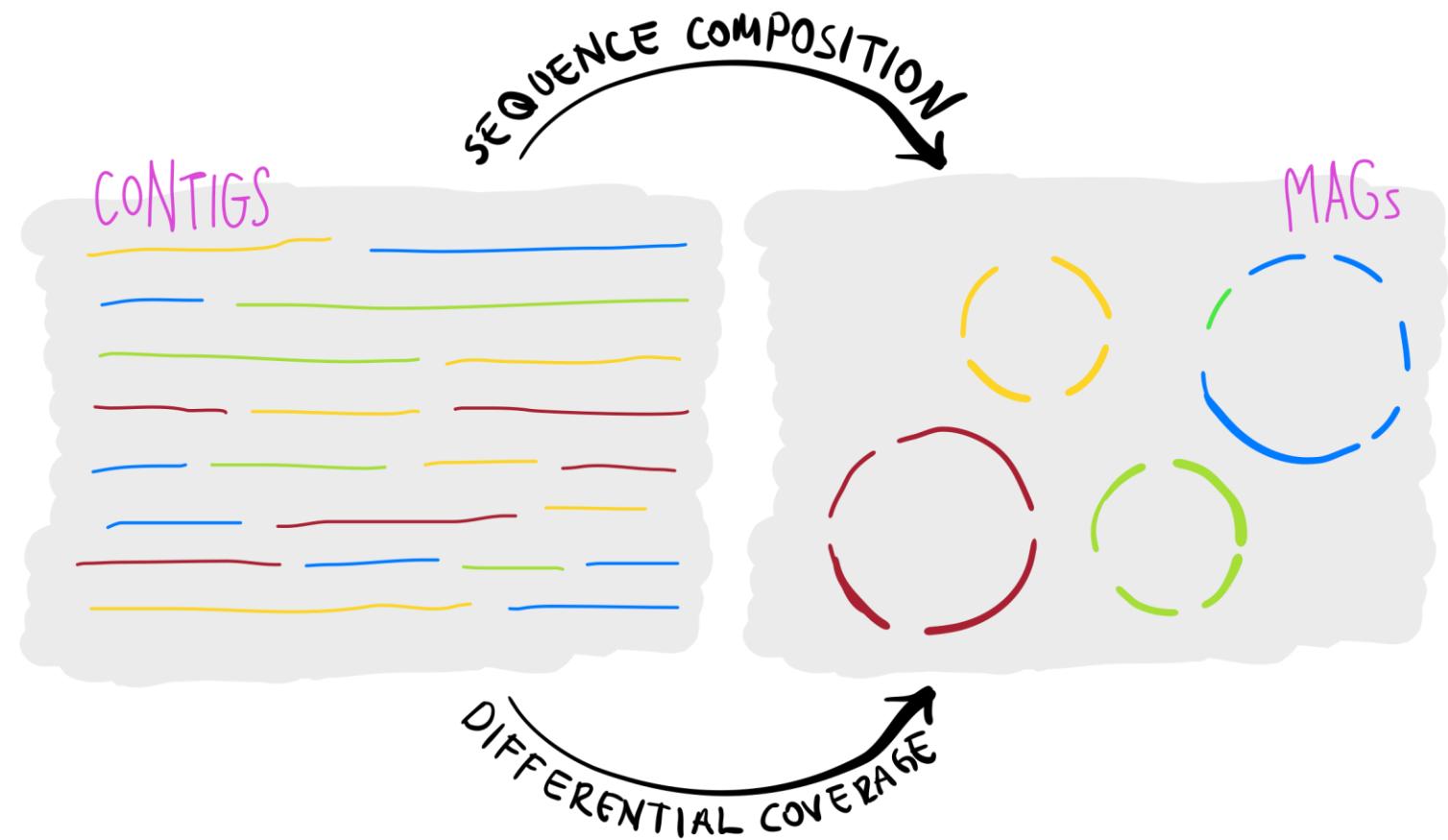
Short introduction

What is genome-resolved metagenomics?

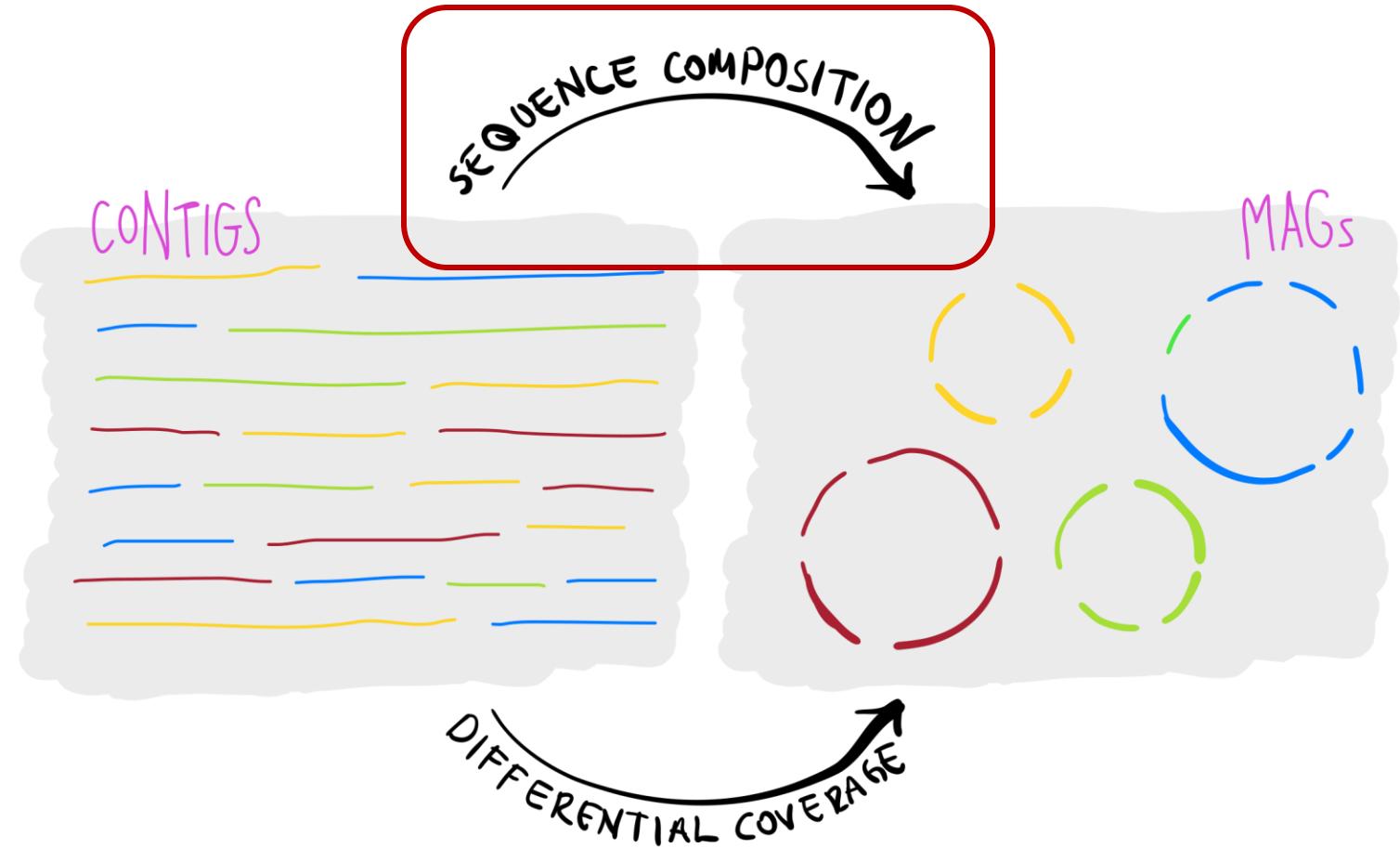
- From contigs to metagenome assembled genomes (MAGs)



Methods to combine contigs



kmer distribution



<http://merenlab.org/momics>

GTTTGCGCATGATTAAGGGAGTTCTTTGTGCTTC

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	0	0	0	0	0

k=2

GT

TTGGCATGATTAAGGGAGTTCTTGTGCTTC

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

TTT	TGGCATGATTAAGGGAGTTCTTTGTGCTTC
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 0 1 0 0 0 1

k=2

GGCATGATTAAGGGAGTTCTTGTGCTTC

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	2

k=2

GTTTGCGCATGATTAAGGGAGTTCTTGTGCTTC

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

GTTTGCGCATGATTAAGGGAGTTCTTTGTGCTTC

	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

AC TTGCGCAGTCGCGCATTACGCGTAGTGGAAATAA

	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										
K										
M										

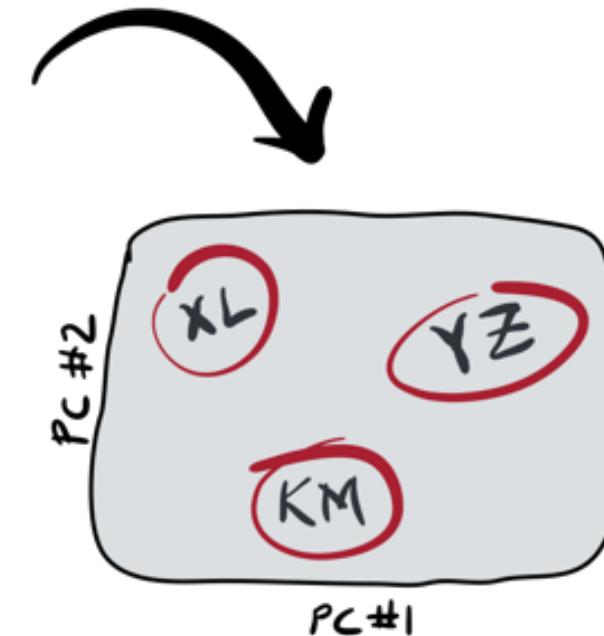
k=2

	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

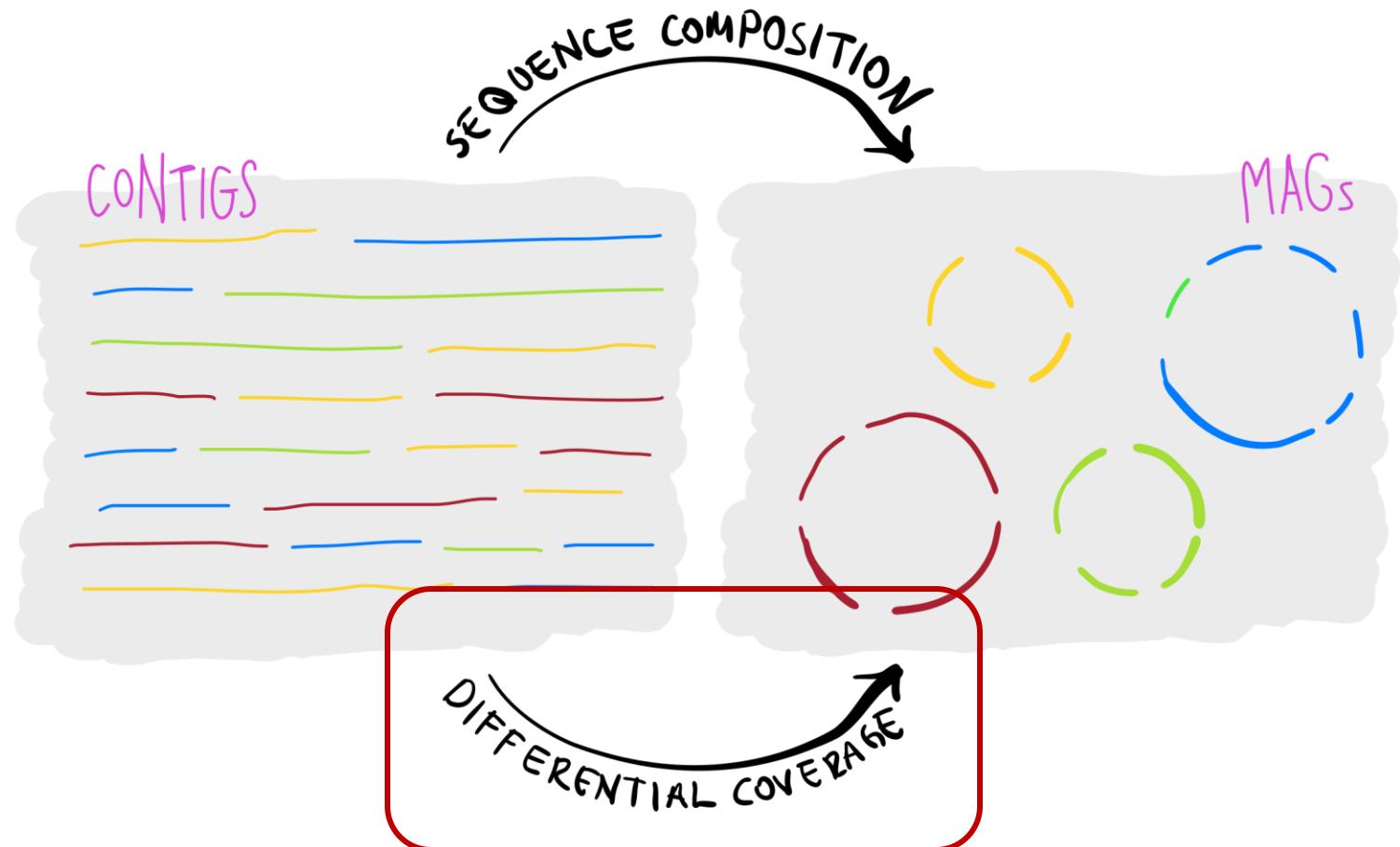
k=2

	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



Differential coverage – read mapping

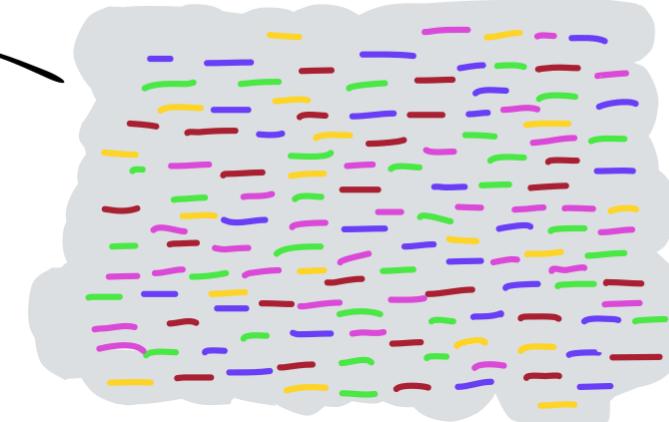


<http://merenlab.org/momics>

CONTIG #1

CONTIG #2

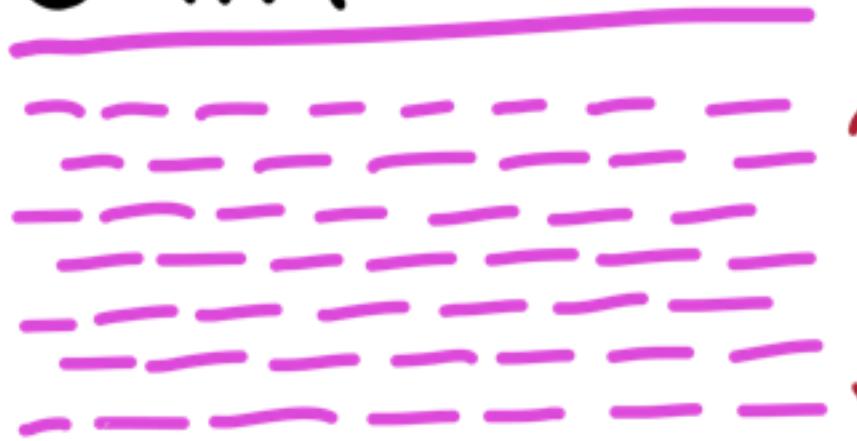
CONTIG #1



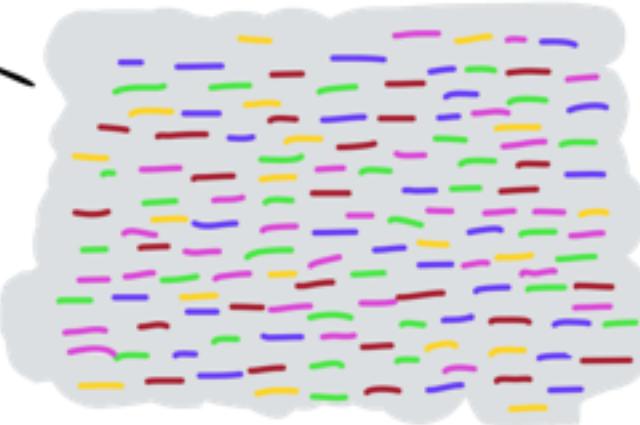
METAGENOMIC READS

CONTIG #2

CONTIG #1



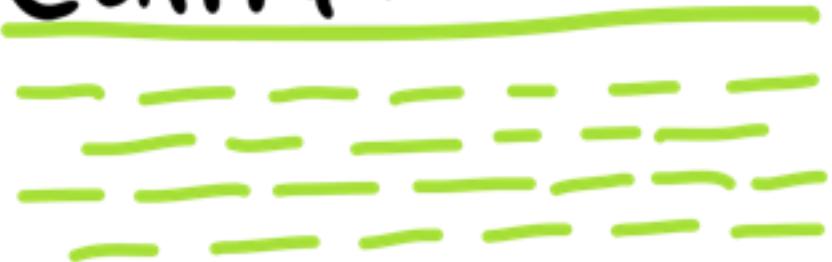
MAPPING



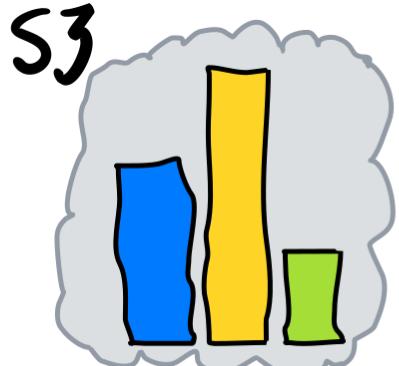
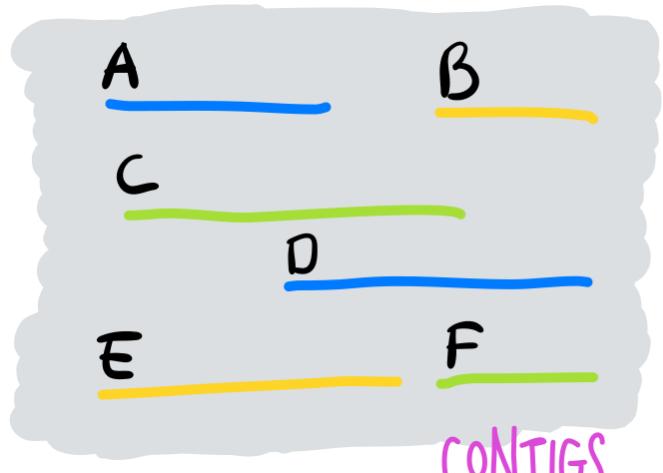
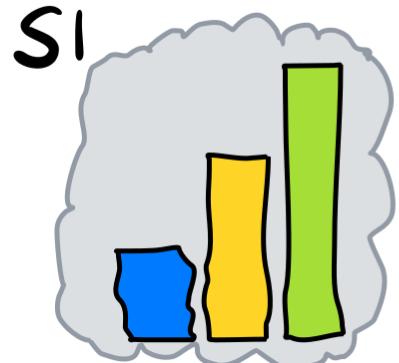
METAGENOMIC READS

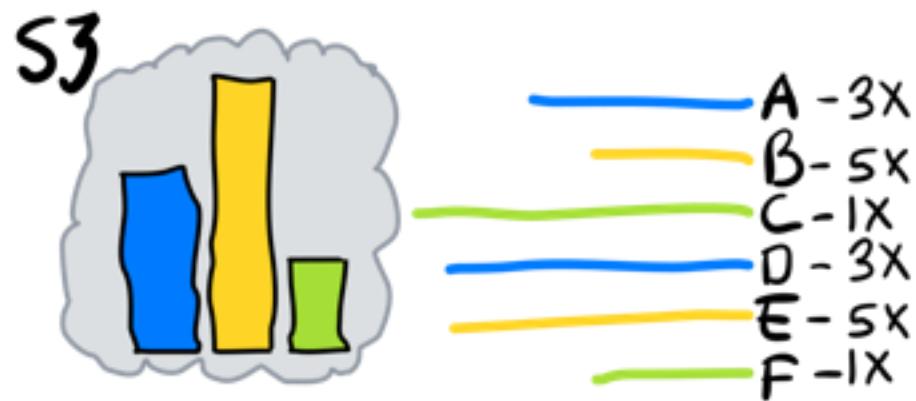
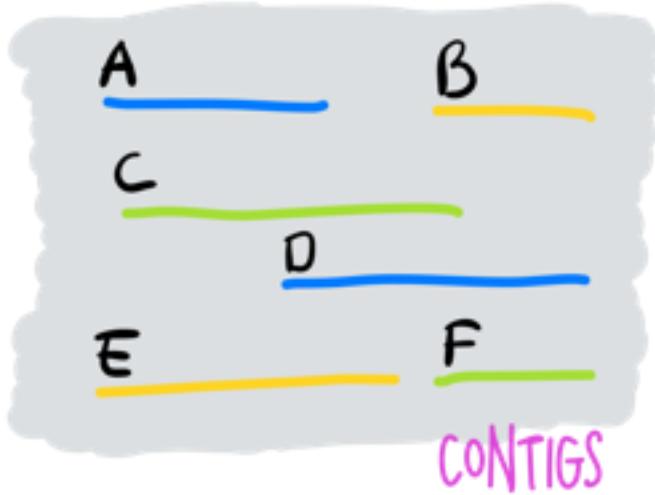
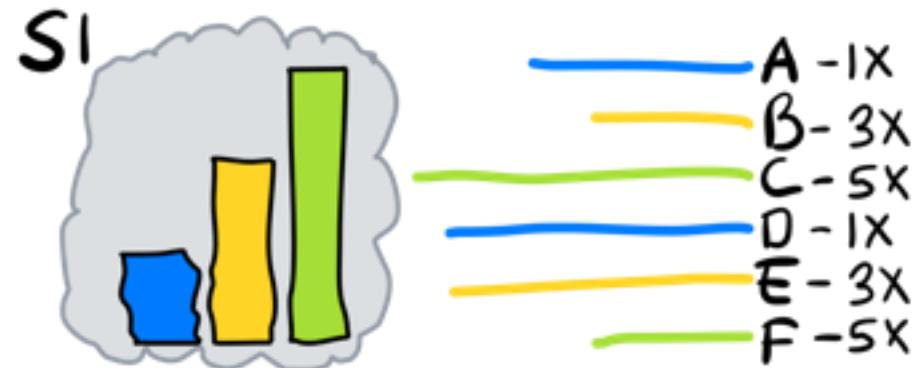
COVERAGE: ~7X

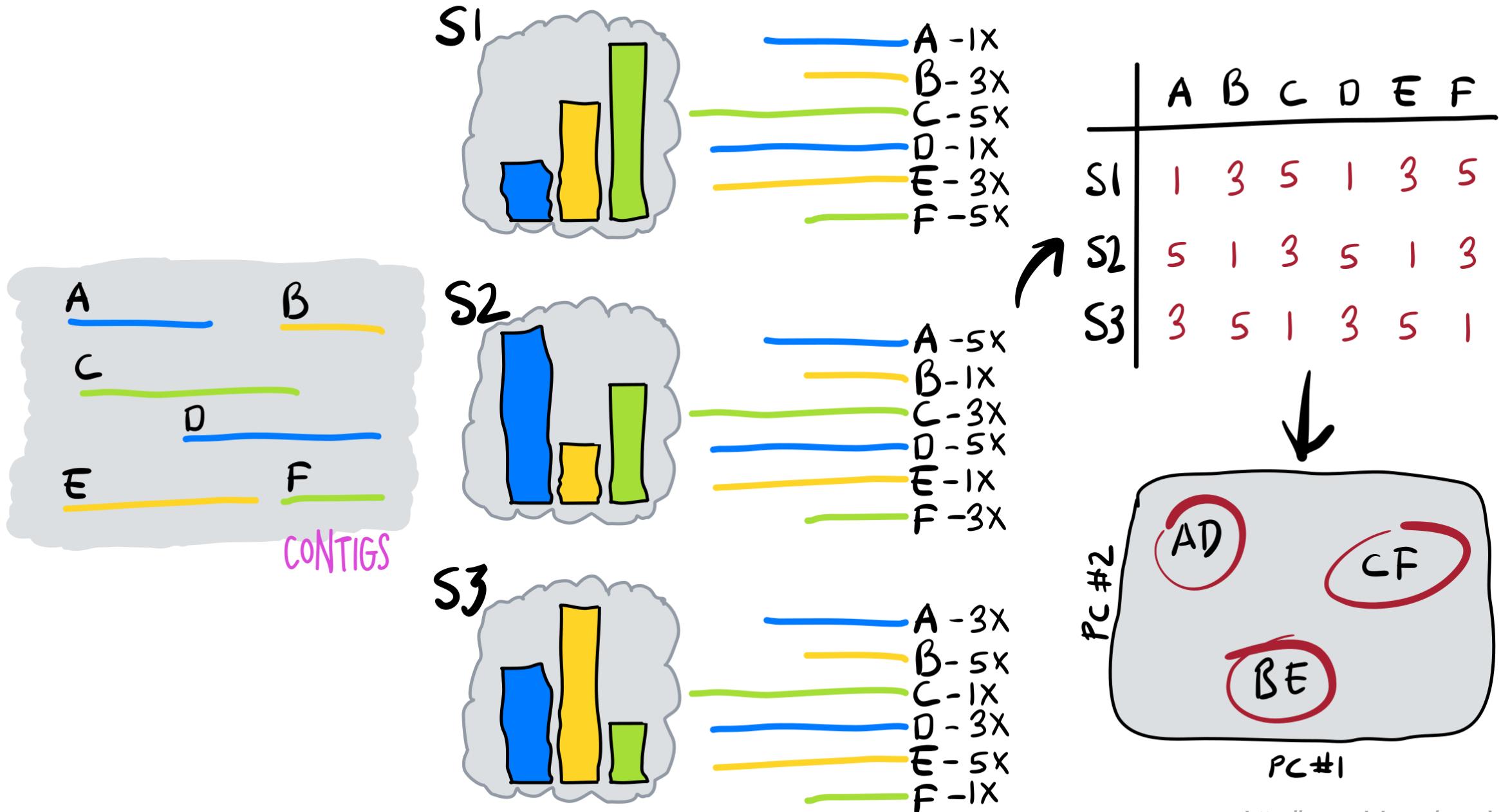
CONTIG #2

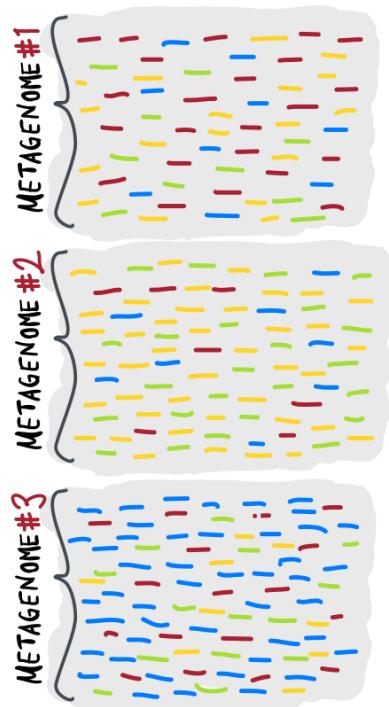


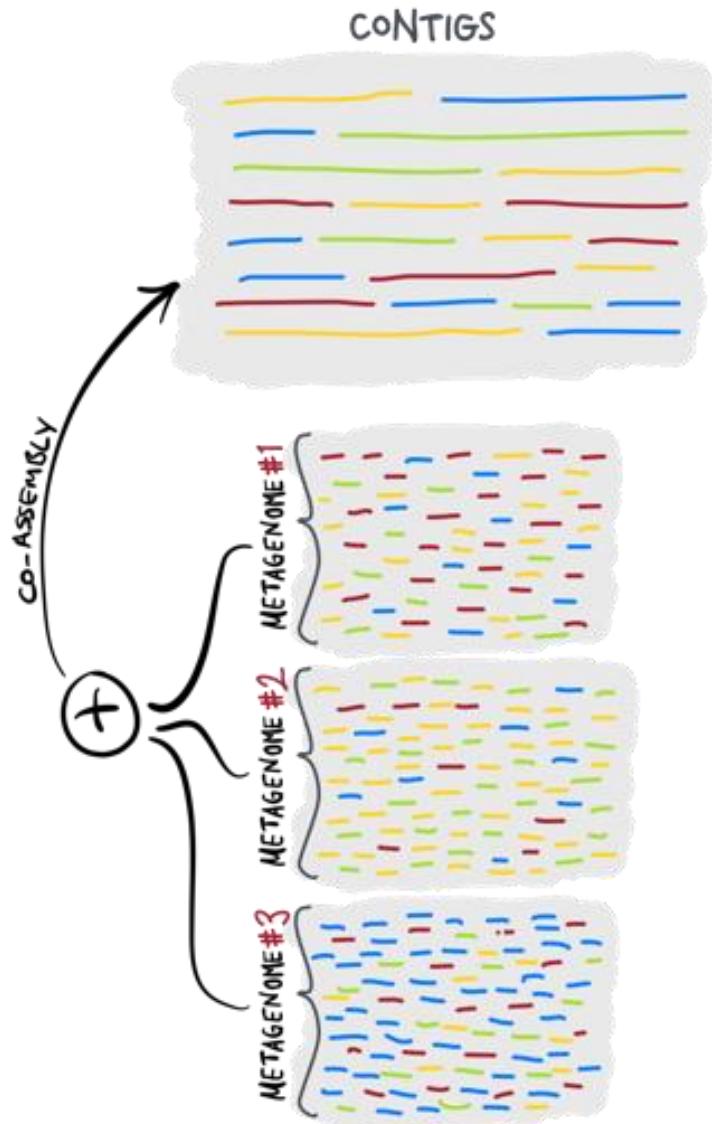
COVERAGE: ~4X

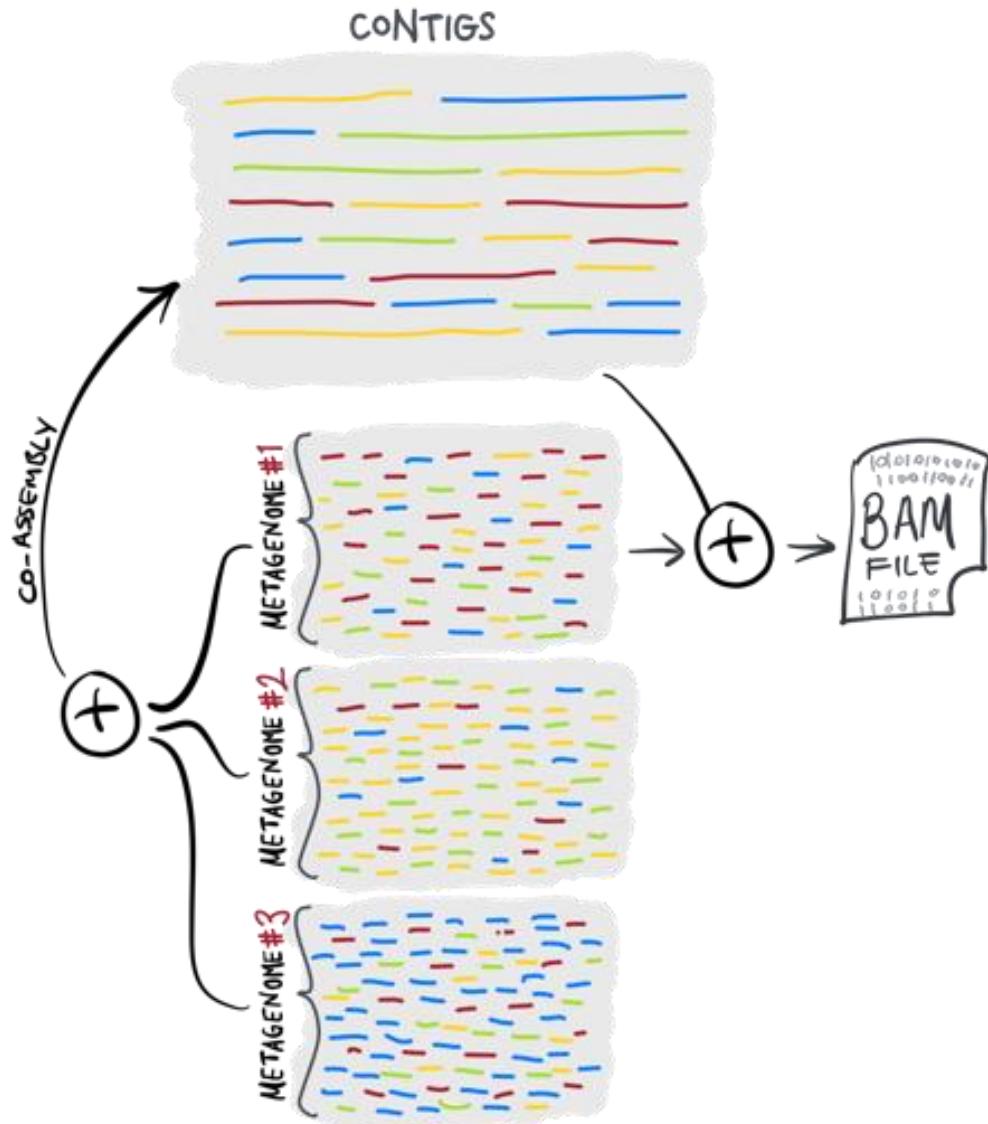




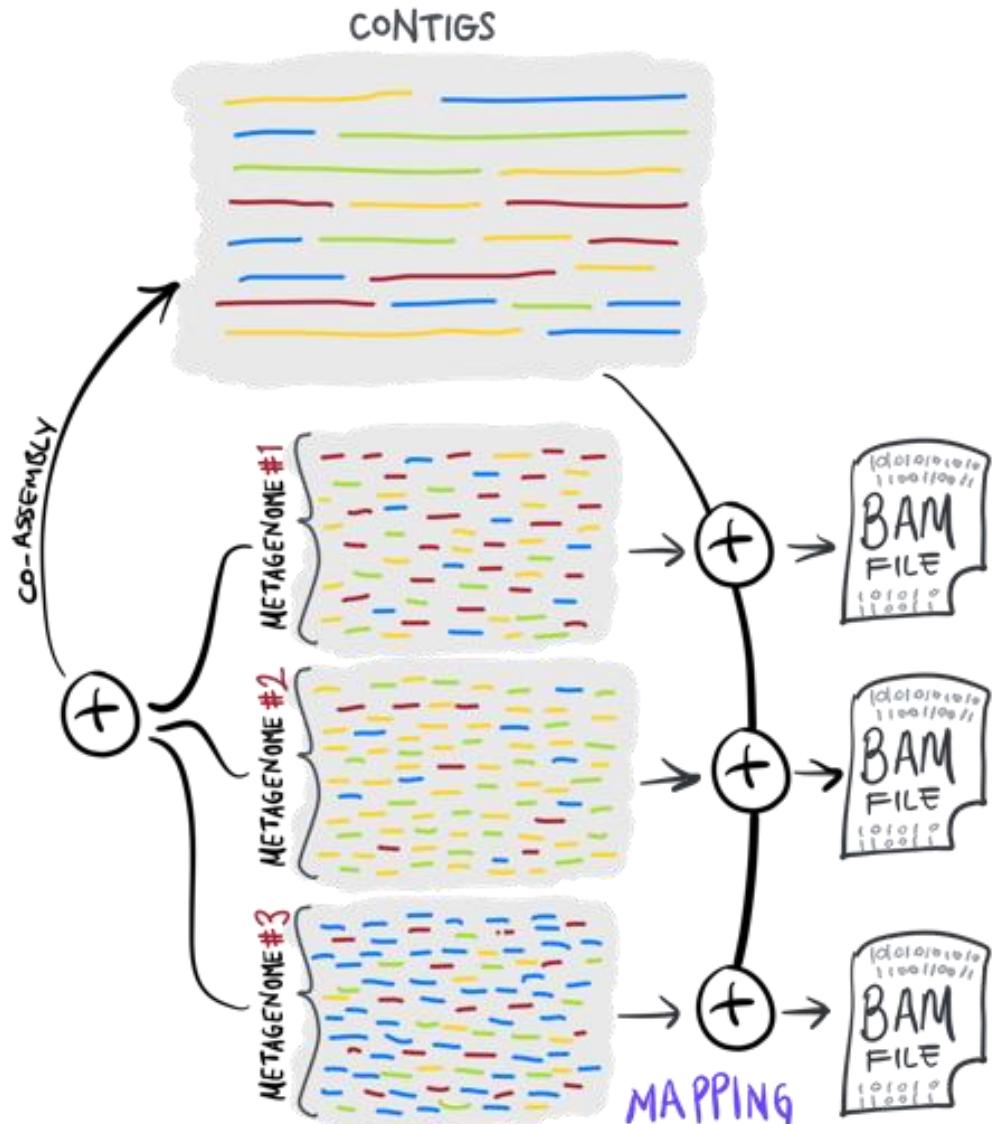


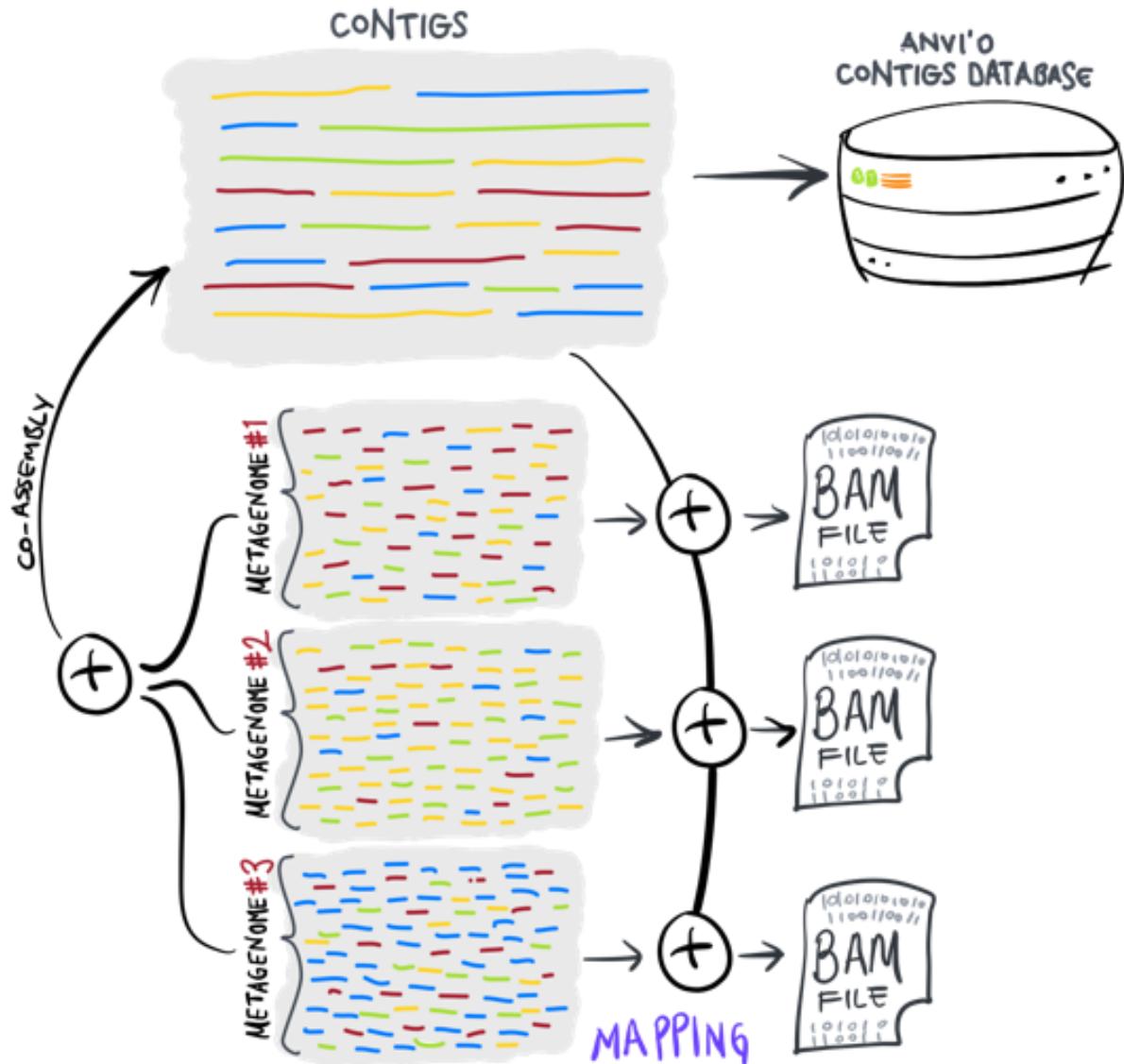


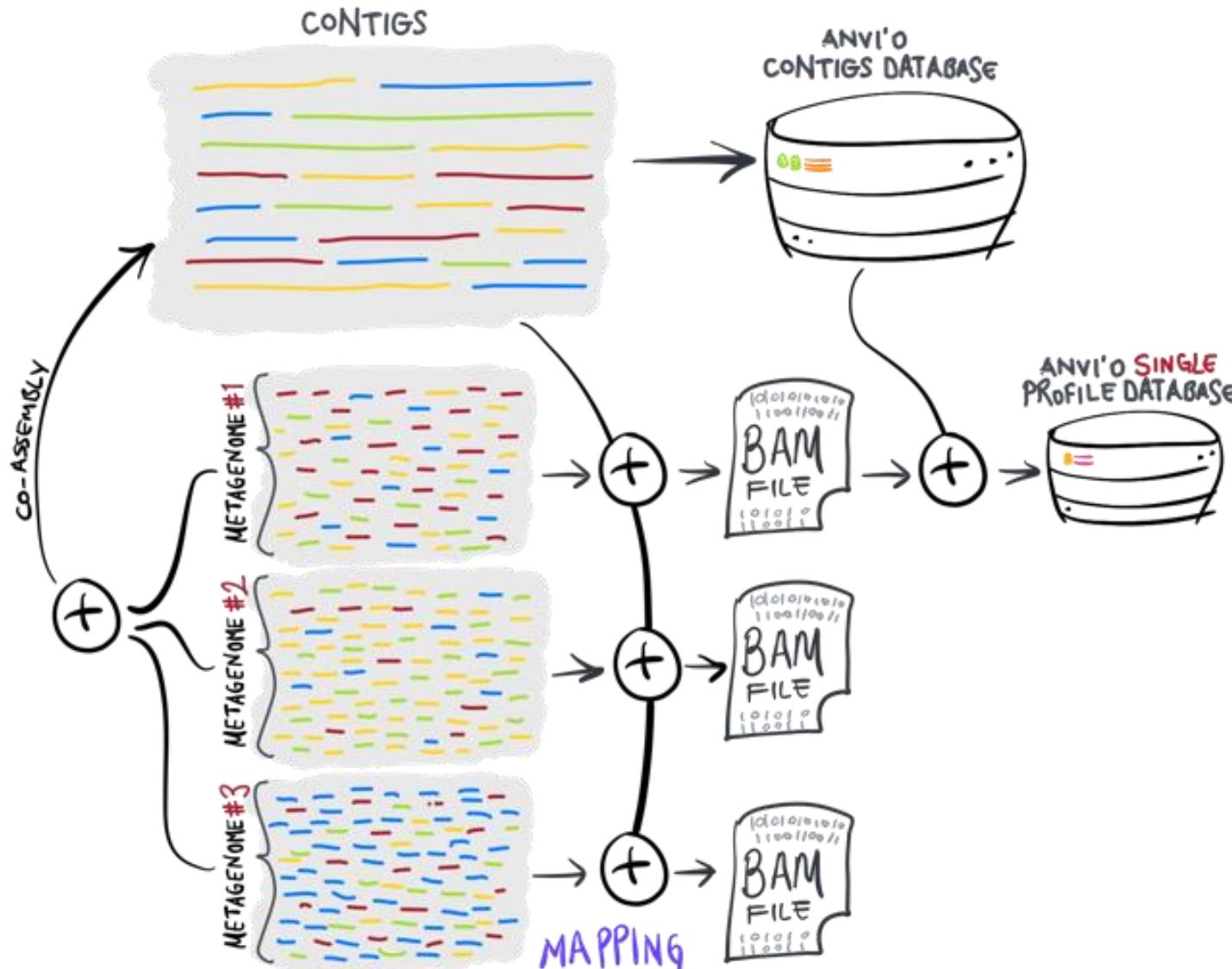


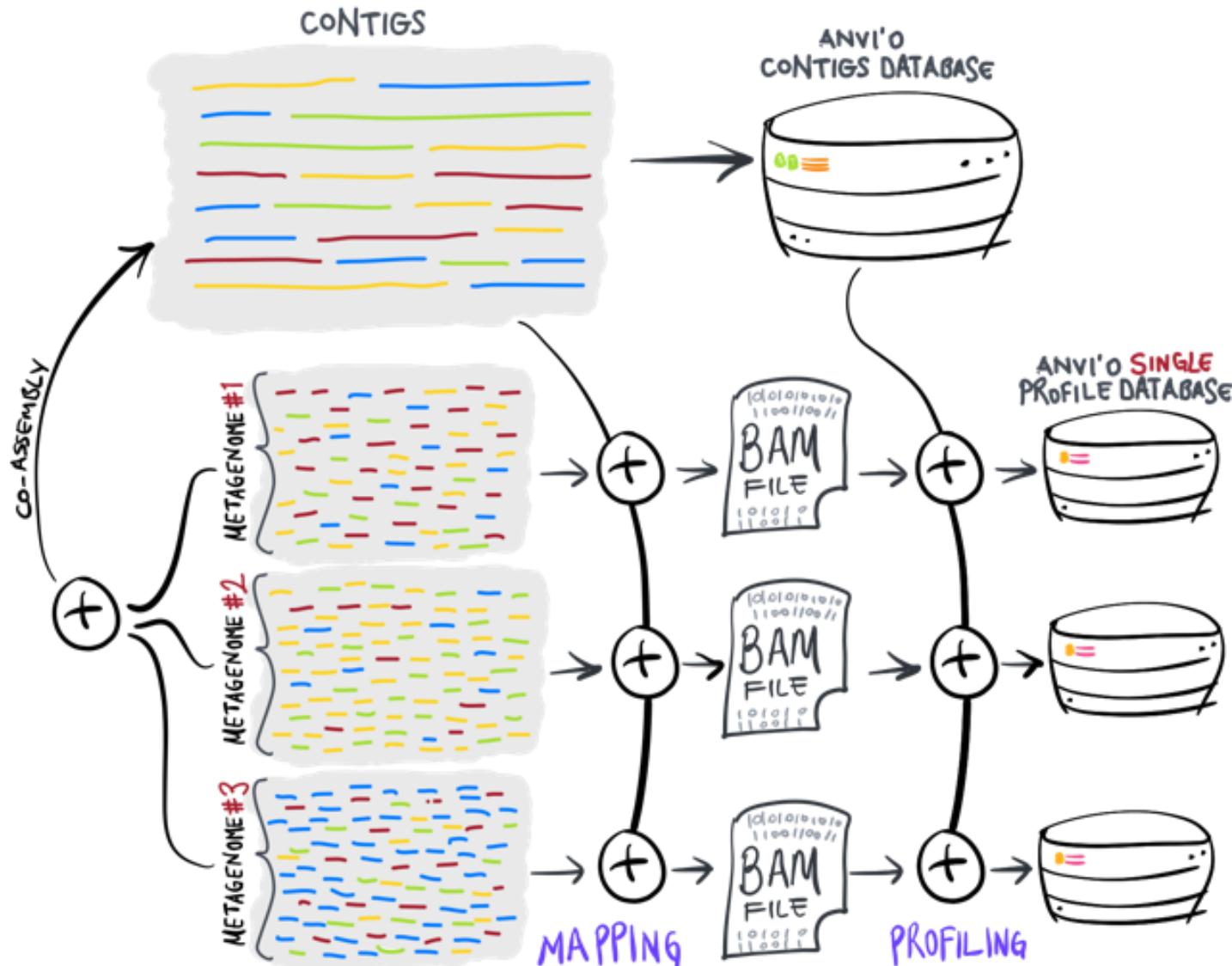


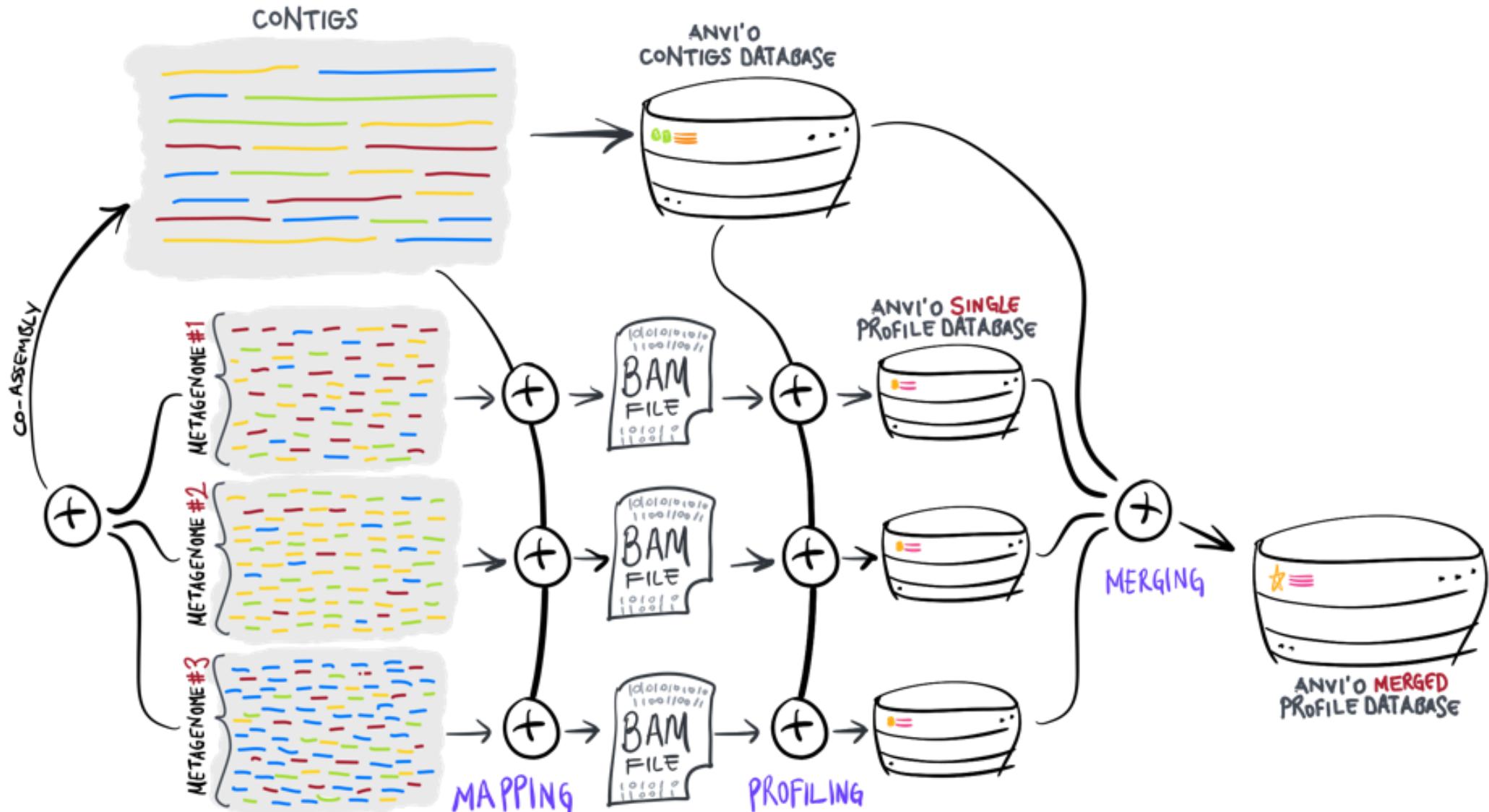
<http://merenlab.org/momics>

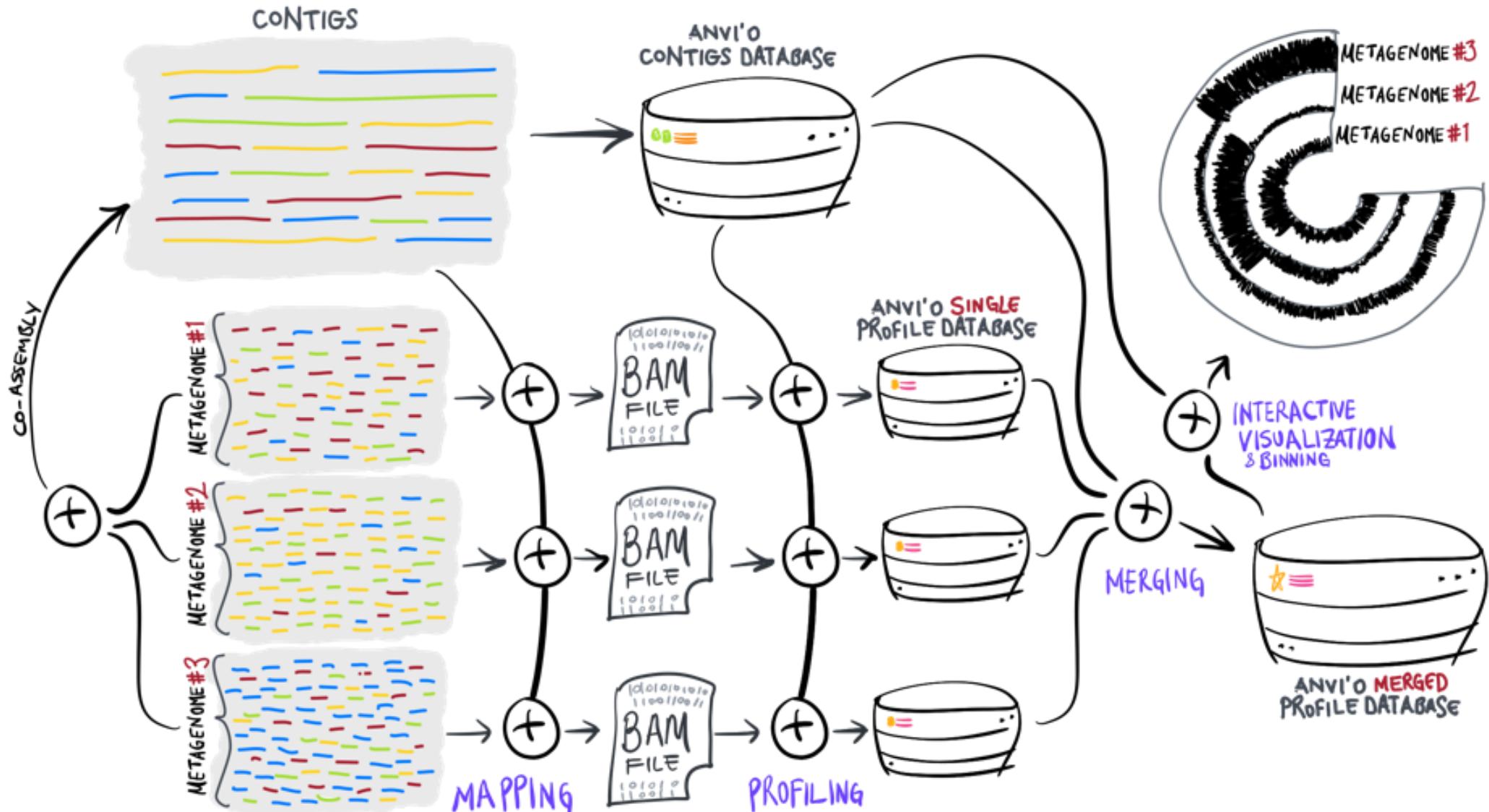




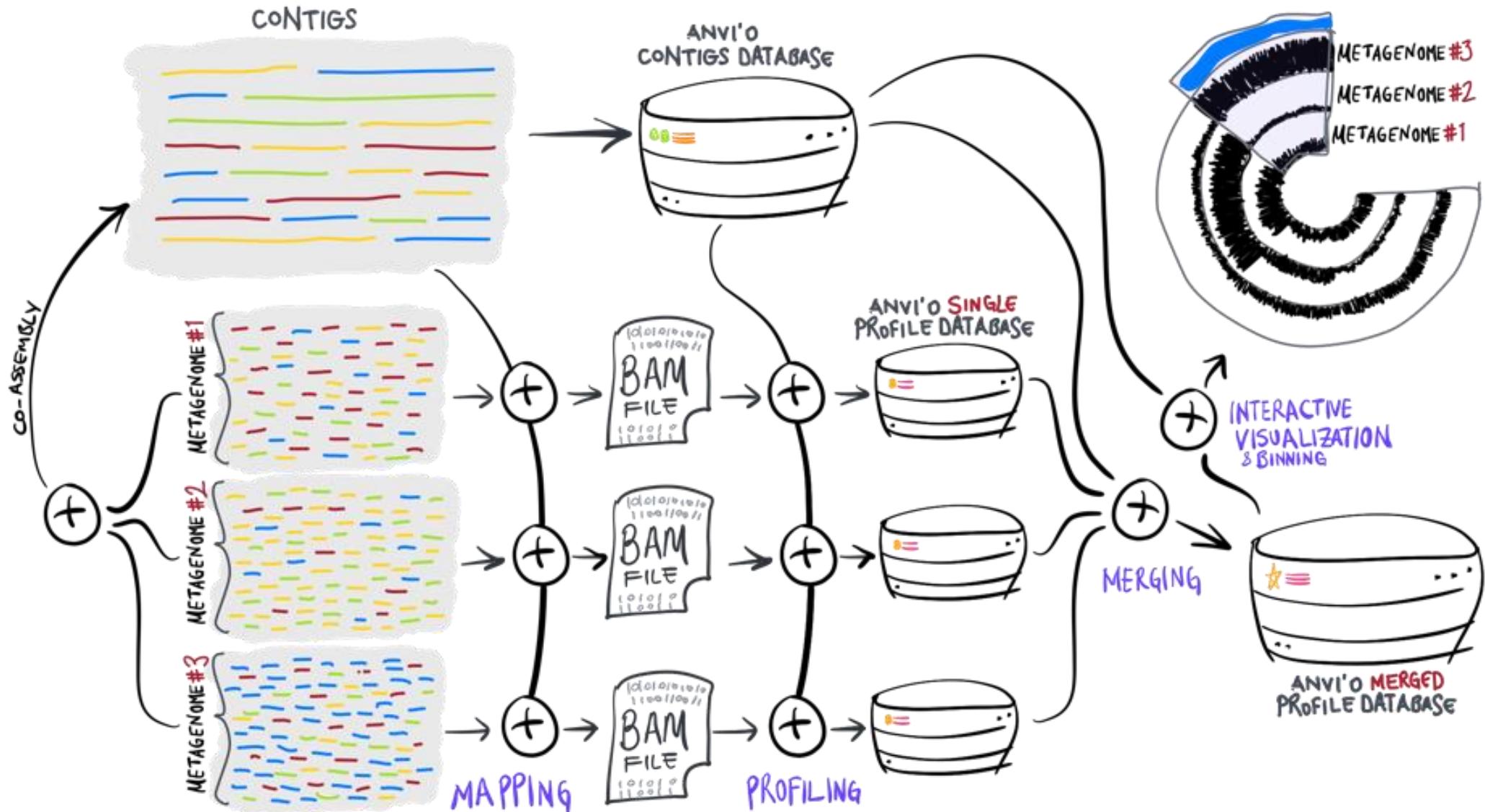


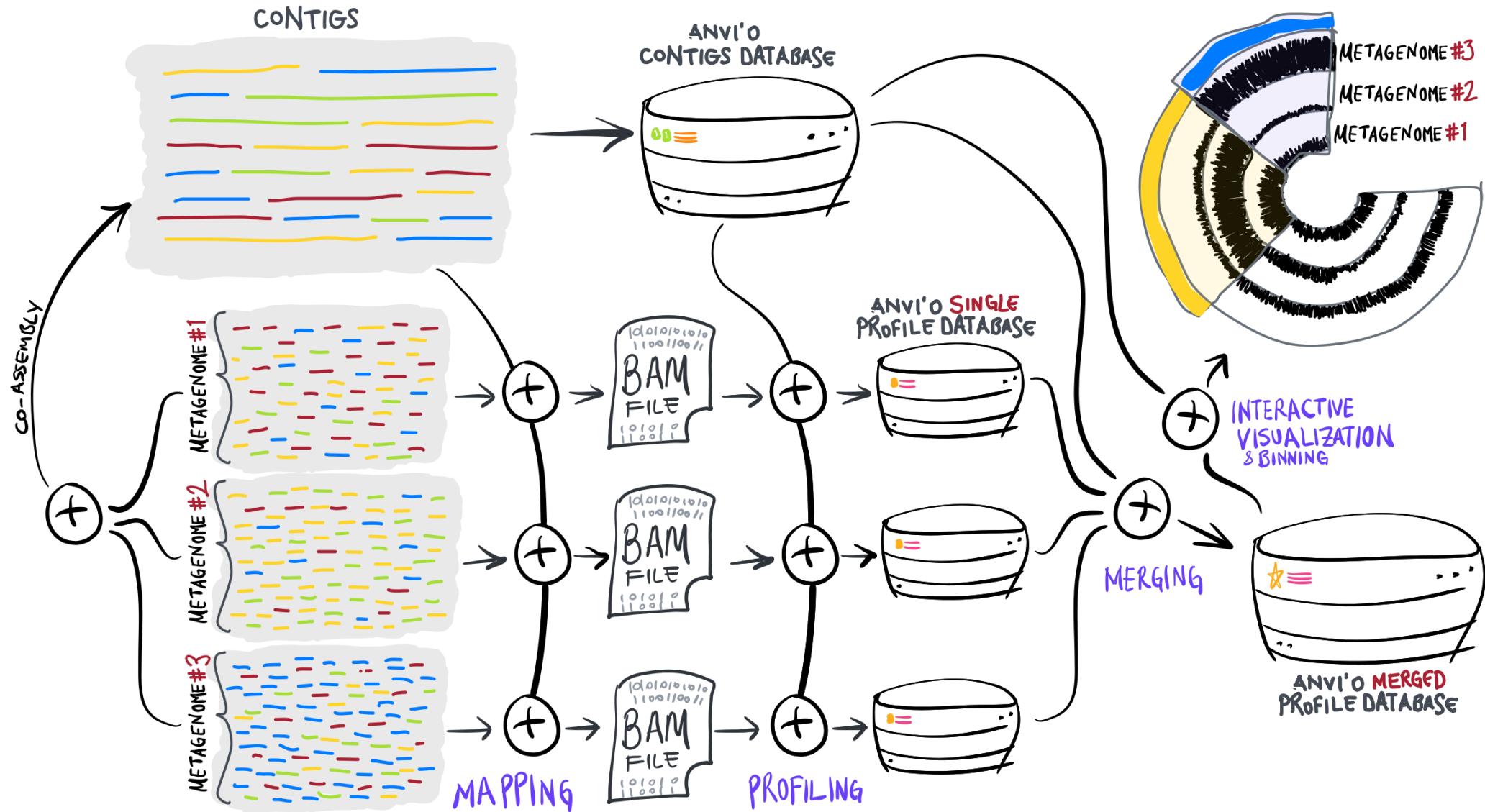




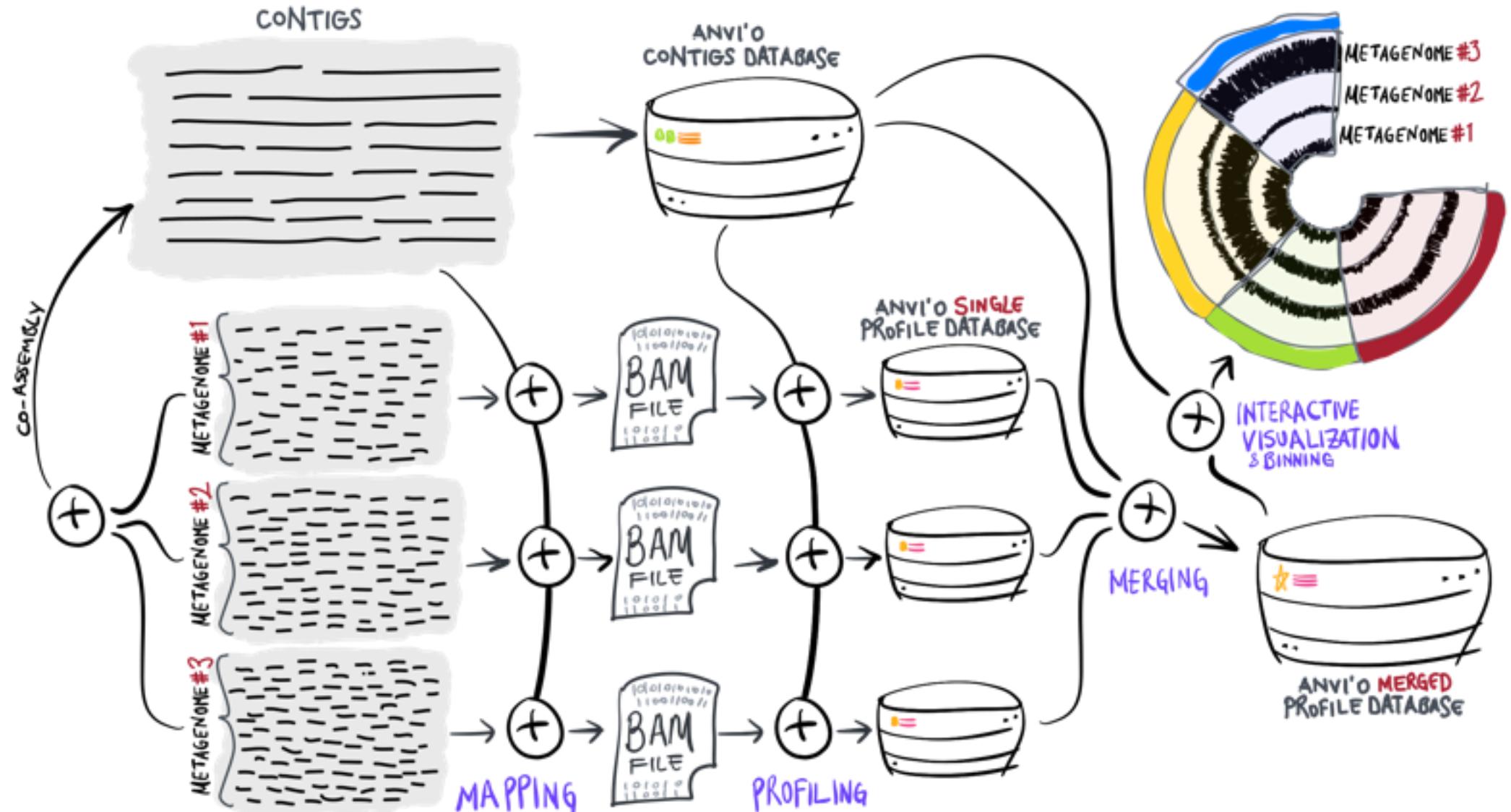


<http://merenlab.org/momics>





<http://merenlab.org/momics>



<http://merenlab.org/momics>

Genome-resolved metagenomics *in action*

- Several automatic binning algorithms available
 - CONCOCT, MetaBat, SemiBin2, MaxBin, BinSanity, DAS Tool, ...
 - Various algorithms, but most rely on kmers and coverage
- Manual binning in anvi'o
 - Tetranucleotide frequency **and/or** differential coverage
 - Also, automatic binning results can be visualised

MAG taxonomy and QC

CheckM2

- Predicts the completeness and contamination of MAGs
- Uses machine learning (not SCGs like e.g. CheckM1)
- Trained on simulated genomes with known completion and contamination – “random-protein-sampling”
- Should work better for lineages with reduced genome sizes compared to SCG-approach

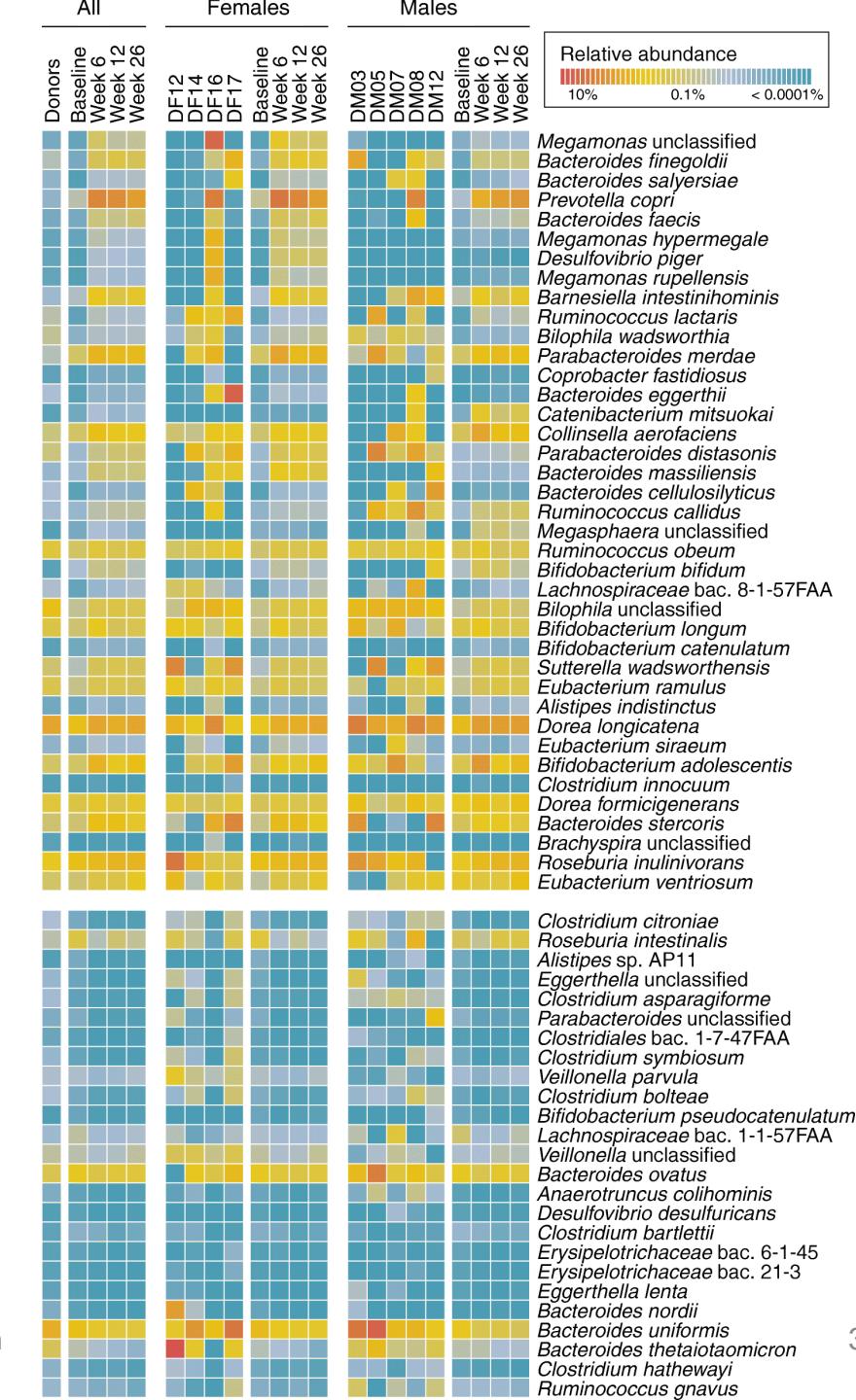
GTDB-Tk

- Genome Taxonomy Database toolkit
- GTDB
 - Standardised microbial taxonomy based on phylogeny
 - Based on SCGs: 120 for bacteria and 122 for archaea
- Taxonomic annotation based on concatenated protein reference trees

Strain engraftment

Selected MAGs

- One either *P. copri* or *Megamonas funiformis*
- Select one additional donor MAG that seems to colonize the recipients
- Original article Figure 2



Next steps in strain engraftment

- Annotation of selected MAGs
- Construct contigs DB from annotated genomes
- Run HMM and add SCG taxonomy annotations to contigs DB
- Fetch all sequence data for 3 recipients
- Map recipient data to each of the selected MAGs
- Construct profile DBs and merge them
- Visualize the results

Let's get to work

<https://github.com/MBDP-bioinformatics-courses/MBDP Metagenomics 2024>