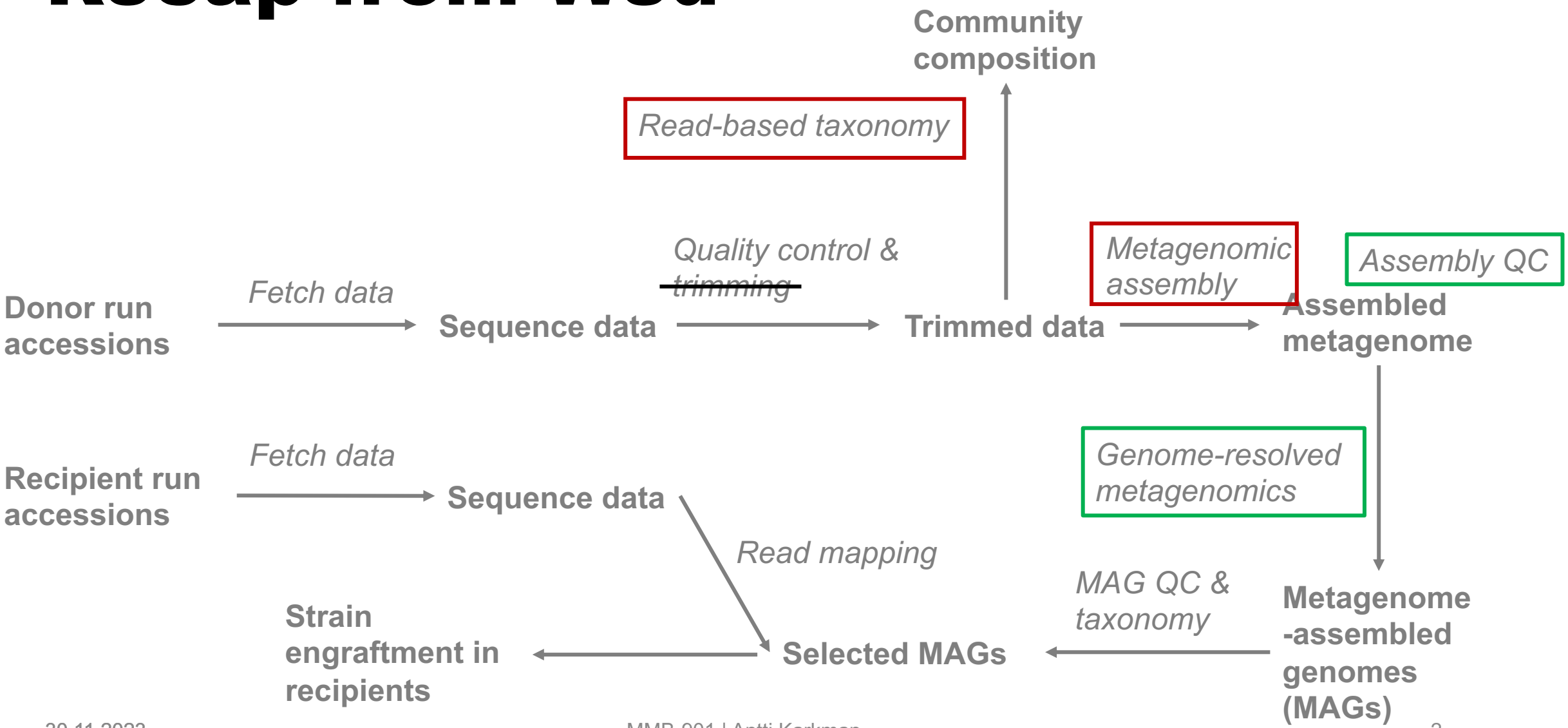
Microbial metagenomics

MMB-901

Recap from Wed



Assembly QC

Is it possible to assess the quality of metagenome assembly?

- Yes/no?
  https://presemo.helsinki.fi/mmb901

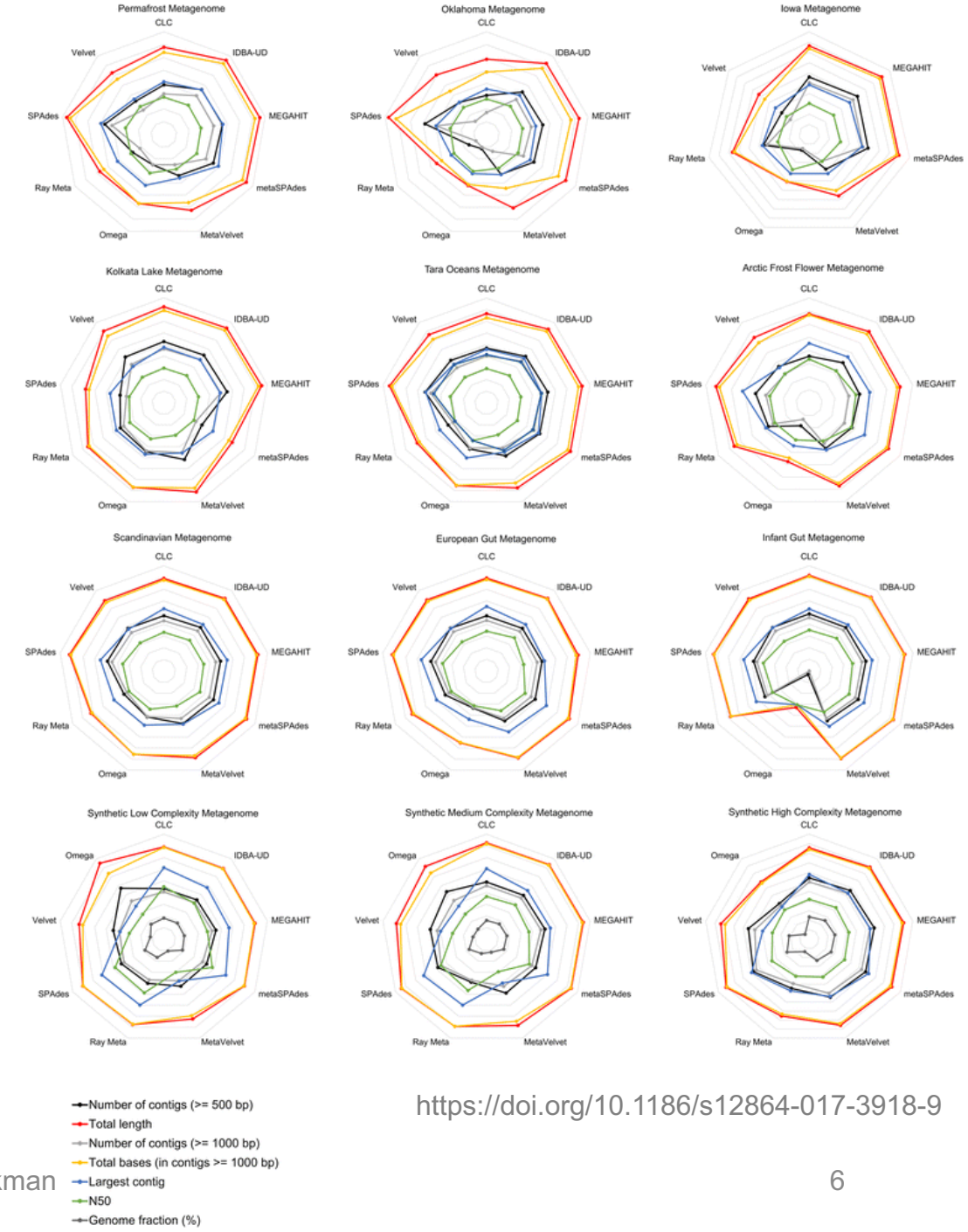
Comparison of assemblers

Table 1 Assembly statistics and computational requirements for assembly of the Tara Oceans metagenome. Time required is given in seconds, minutes and hours for illustrative purposes and memory in GB of RAM required

	Tara Ocean								
	CLC	IDBA-UD	MEGAHIT	metaSPAdes	MetaVelvet	Omega	Ray Meta	SPAdes	Velvet
Number of contigs (≥ 500 bp)	50,716	163,815	216,938	185,419	67,161	15,982	6128	220,178	57,816
Total length	46,069,409	179,686,756	210,621,485	202,770,058	55,972,515	34,861,819	7,277,214	275,920,632	45,425,460
No. of long contigs (≥ 1 kbp)	10,720	50,498	56,243	48,640	12,590	13,305	2179	70,711	8802
No. of ultra-long contigs (≥ 50 kbp)	0	2	1	37	0	9	0	54	0
Largest contig	39,748	101,400	62,649	141,519	30,177	102,255	41,443	197,381	21,980
N50	880	1166	982	1124	805	2691	1329	1415	749
L50	14,113	38,236	58,246	39,033	21,544	2737	1345	39,617	19,631
Mapping rate (%)	38.98	52.24	55.92	64.03	4117	13.64	8.25	64.46	48.19
Time (seconds)	3527	69,782	10,455	125,862	2527	168,213	16,419	80,039	2342
Time (minutes)	58.78	1163.03	174.25	2097.70	42.12	2803.55	273.65	1333.98	39.03
Time (hours)	0.98	19.38	2.90	34.96	0.70	46.73	4.56	22.23	0.65
Memory required (GB)	16.23	42.84	10.58	66.53	109.37	30.7	42	157.75	109.37

https://doi.org/10.1186/s12864-017-3918-9

Comparison of assemblers



https://doi.org/10.1186/s12864-017-3918-9

Which assembler to choose

Availability

The tool should be freely available either as download or webserver.

Usability

The tool should have a proper manual, readme file or help function describing how to use it.

Adoption:

The tool should be widely used or show potential of being widely adopted in the future.

Reference: https://doi.org/10.1038/srep19233

Open questions in metagenomic assembly

- To co-assemble or not?
- No reference – How to define a good assembly?
- Challenging elements for assembly
  - Repeat regions
  - Horizontally transferred genes
  - Extrachromosomal elements
- Others?

Genome-resolved metagenomics

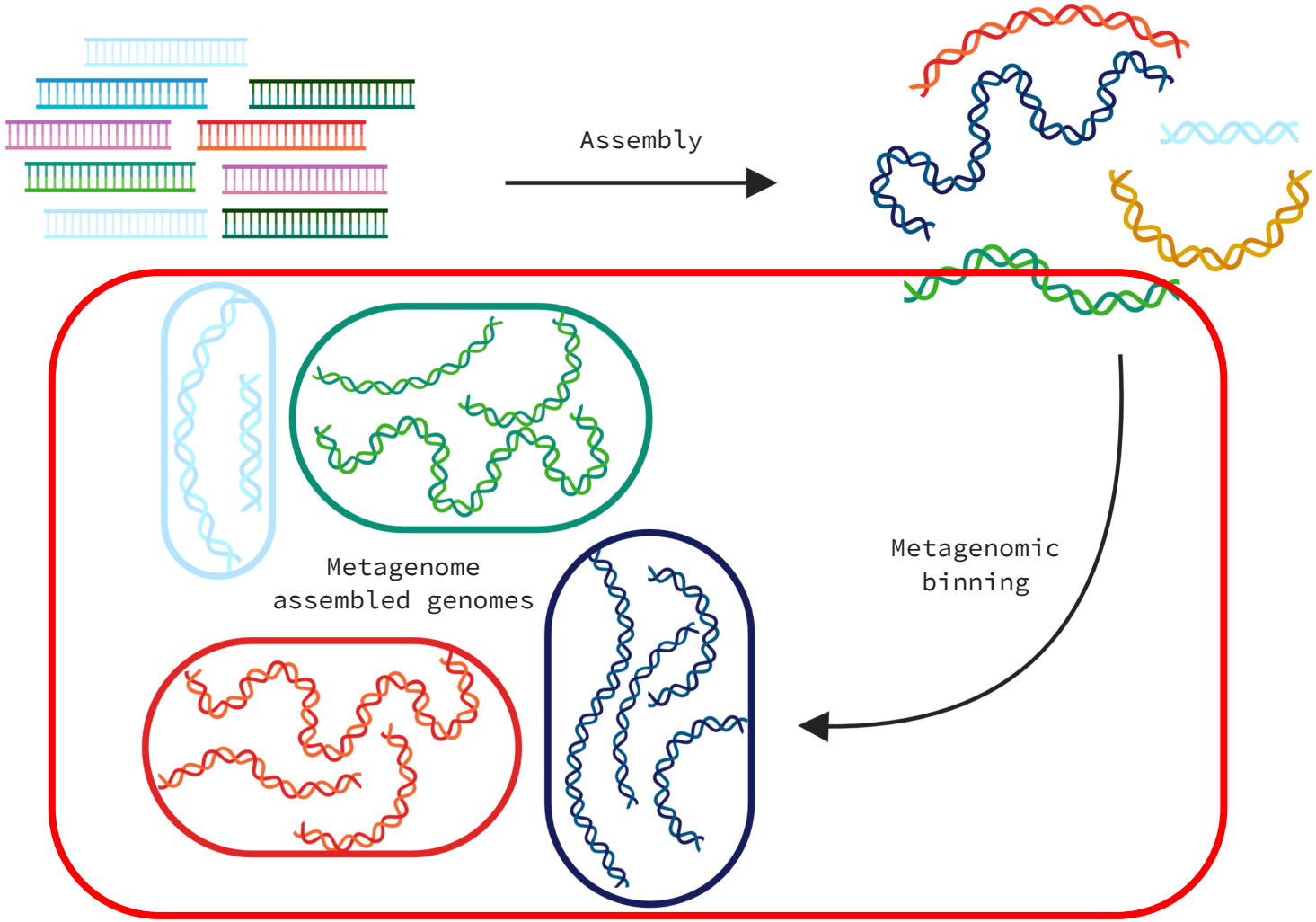
Short introduction

Let’s create the contigs database first

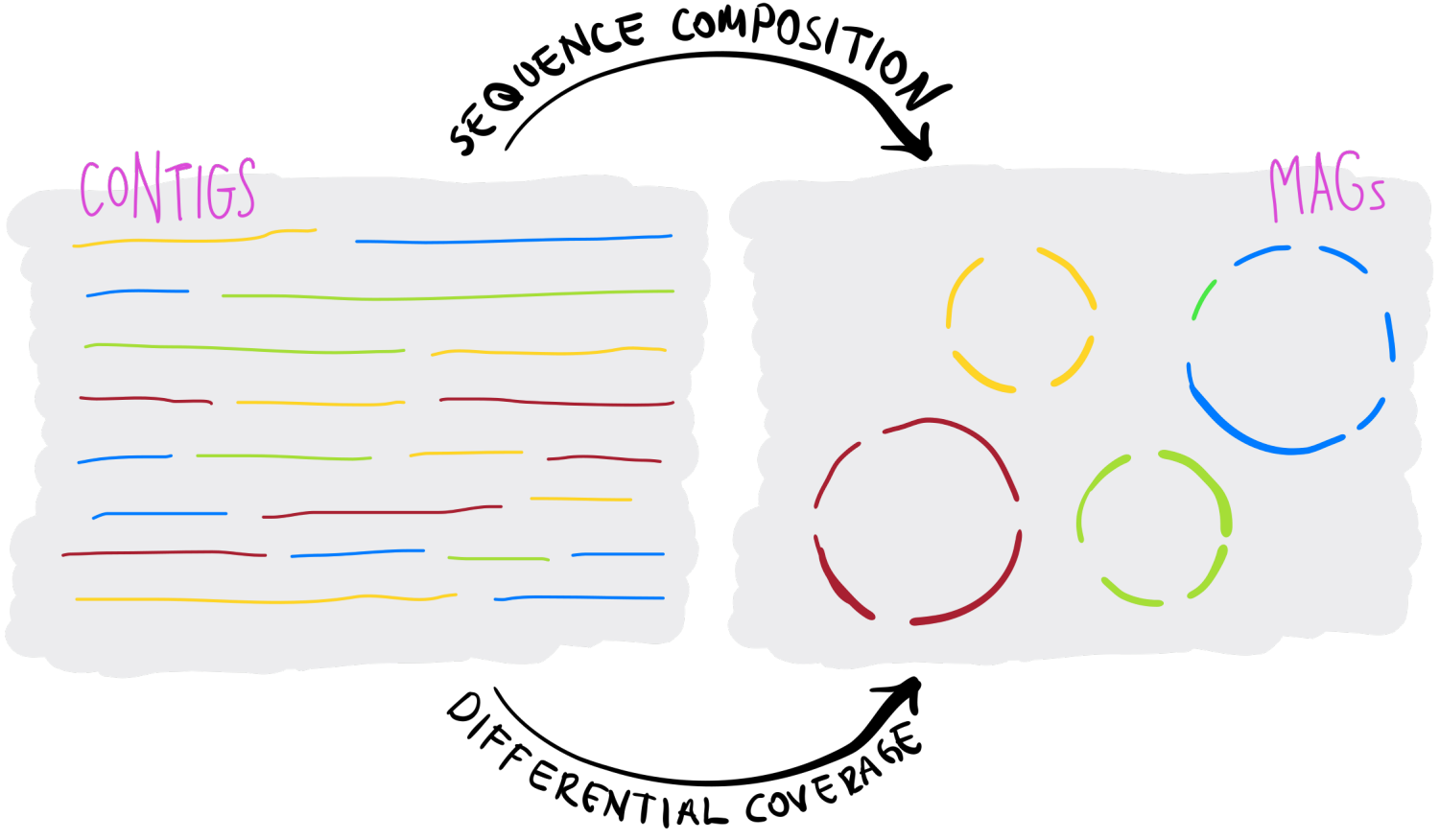
https://github.com/karkman/MMB-901_Metagenomics

What is genome-resolved metagenomics?

- From contigs to metagenome assembled genomes (MAGs)

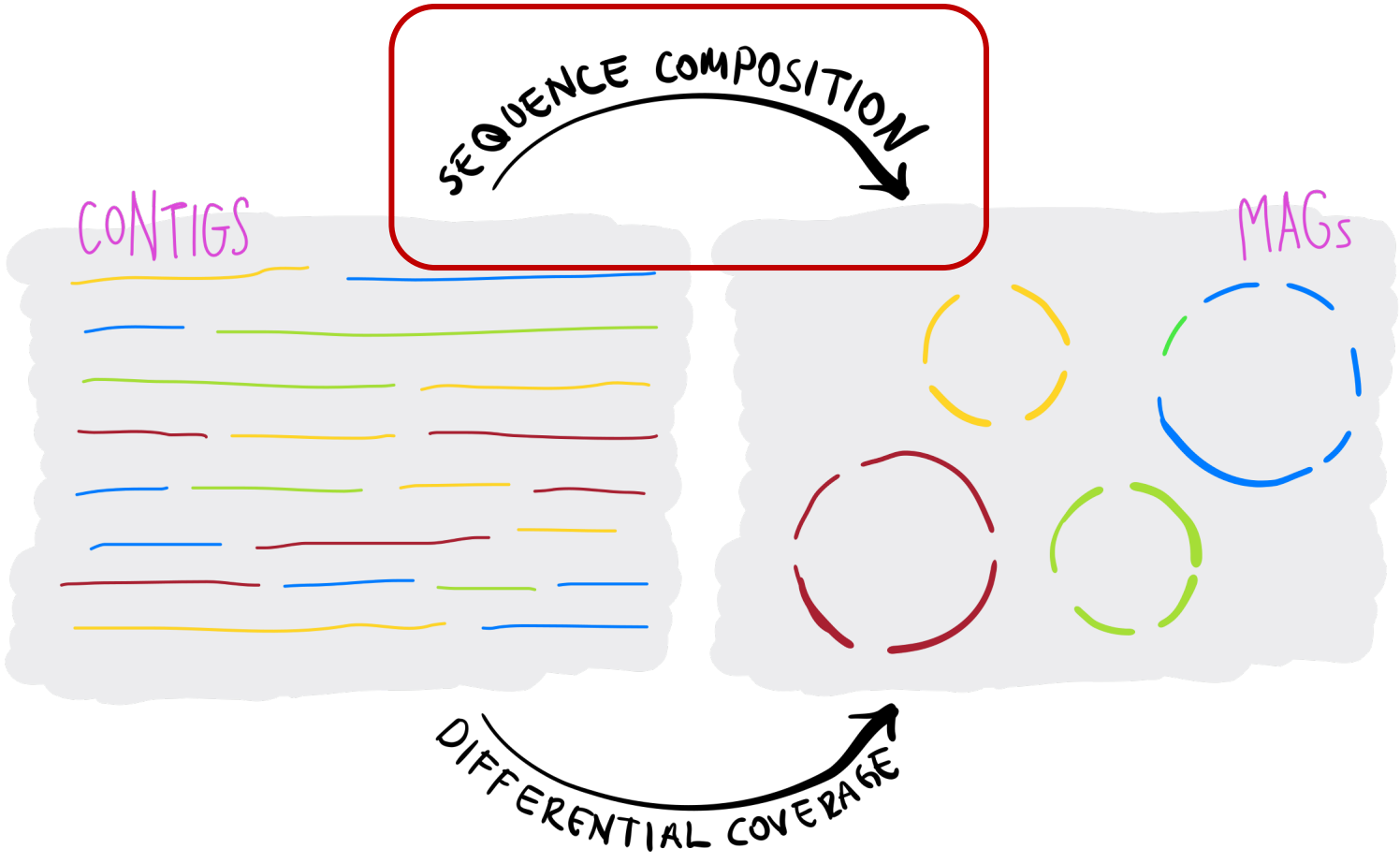


Methods to combine contigs



http://merenlab.org/momics

kmer distribution



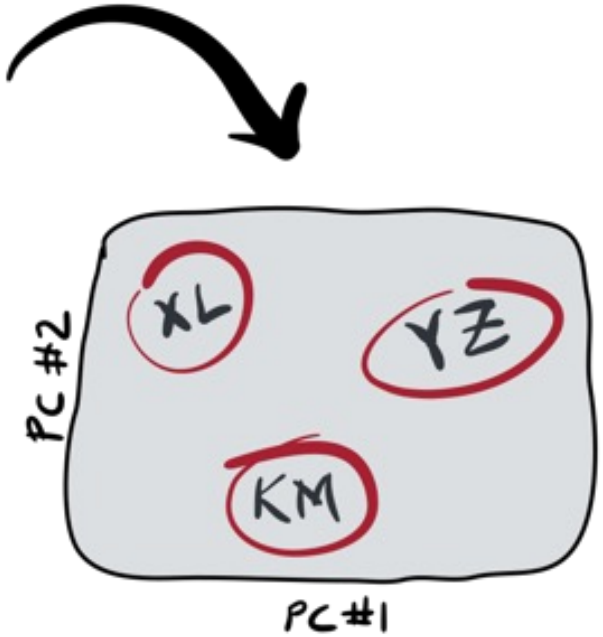
http://merenlab.org/momics

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

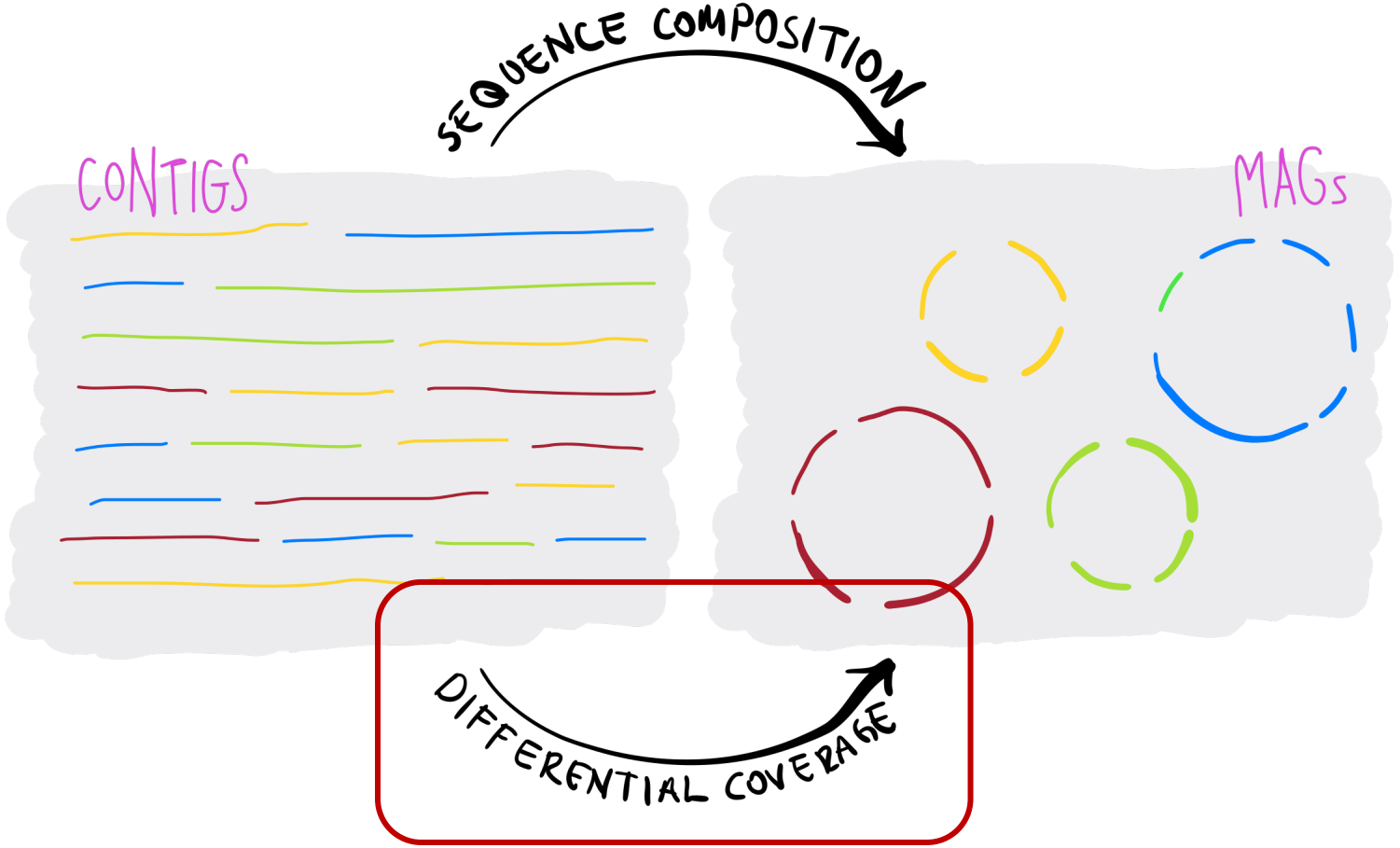
k=2

	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

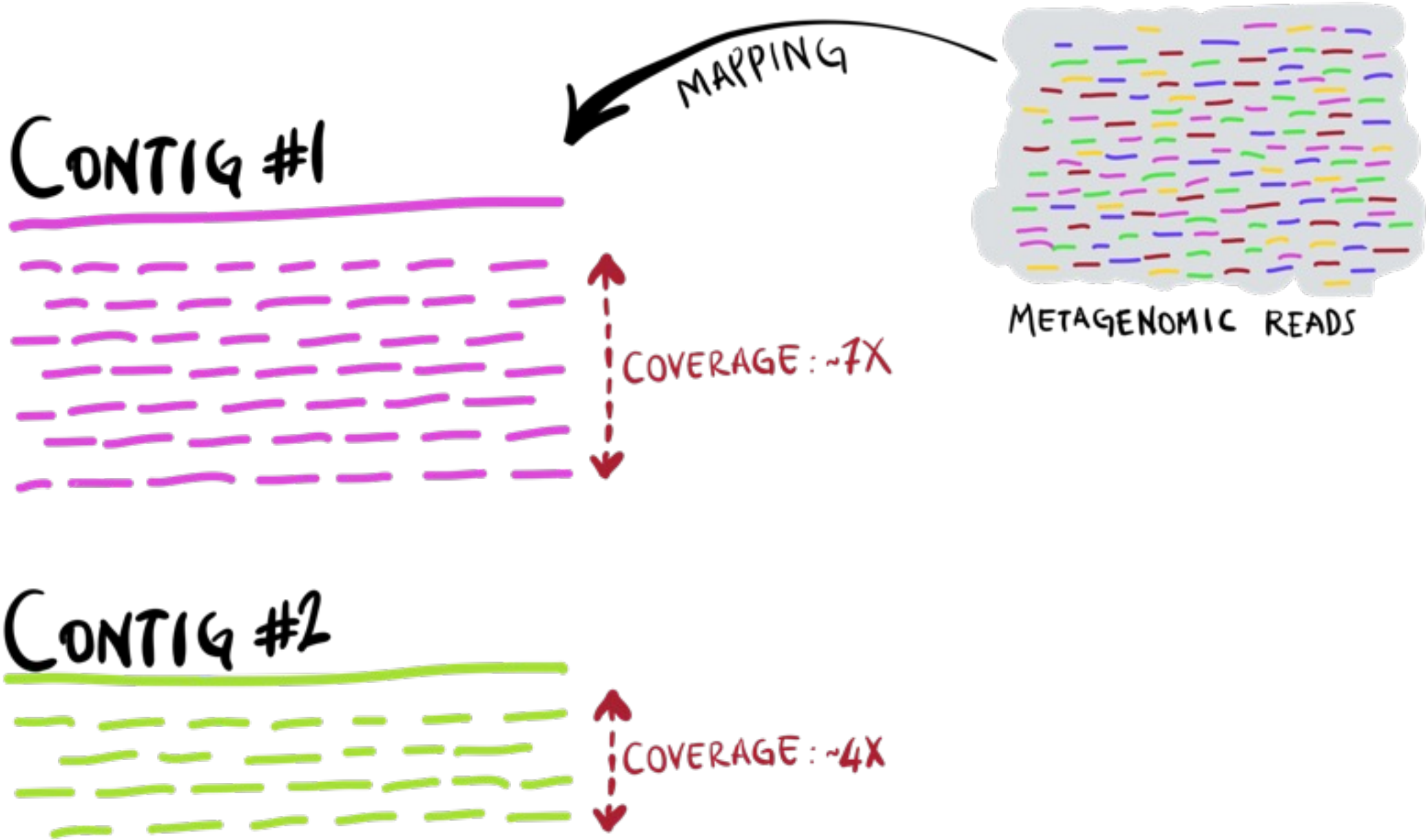


k=2

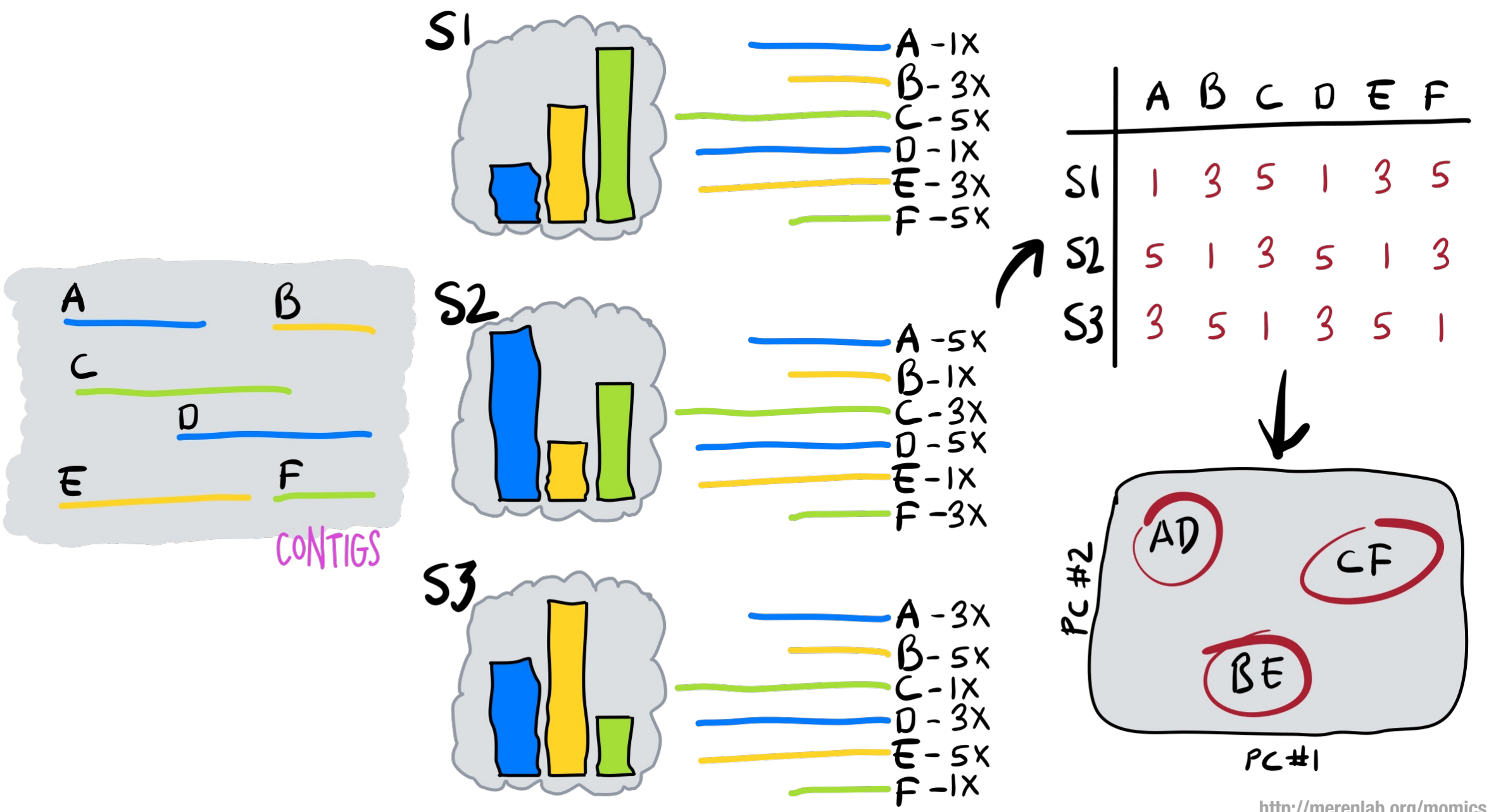
Differential coverage – read mapping



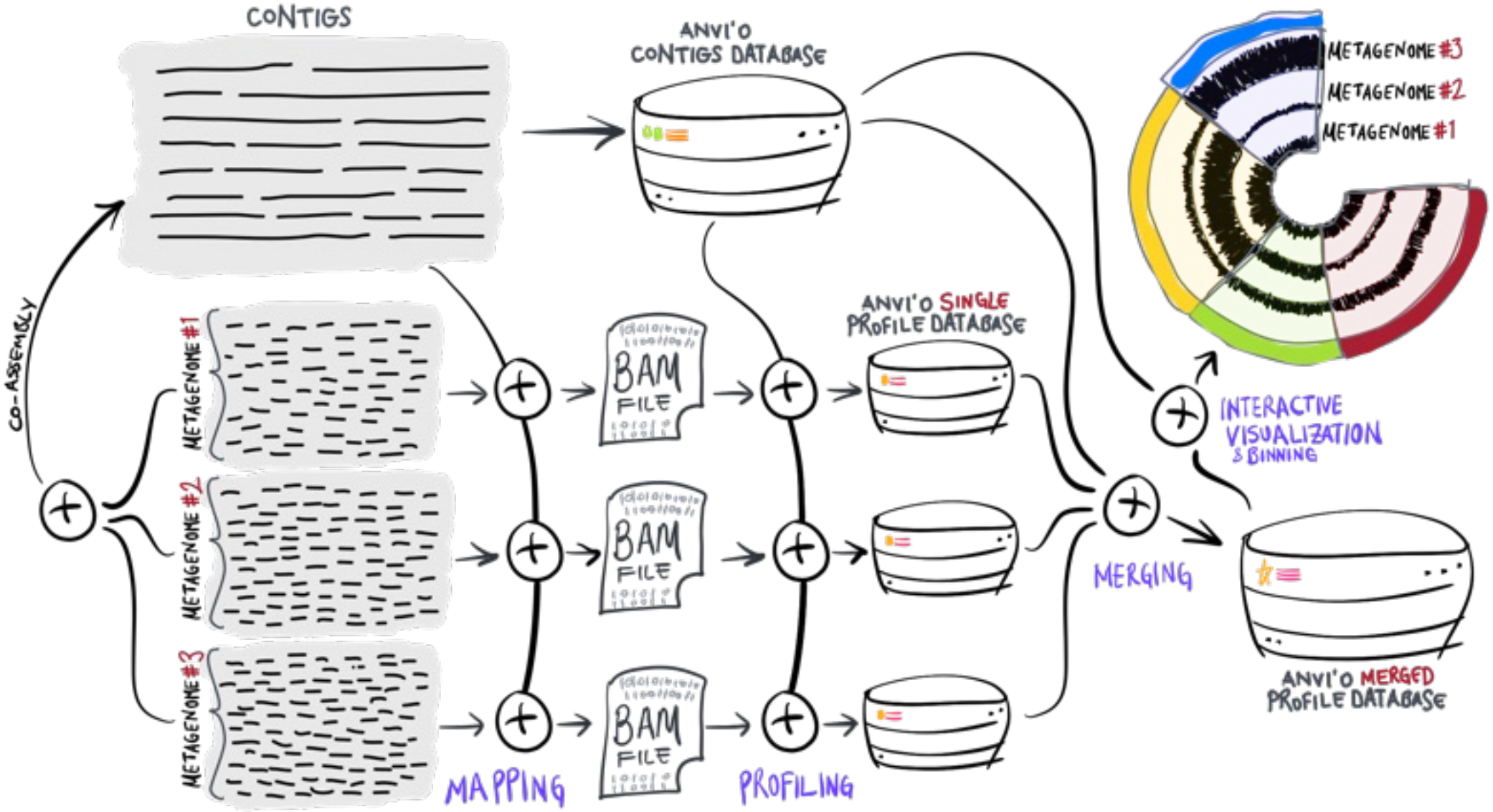
http://merenlab.org/momics



http://merenlab.org/momics



http://merenlab.org/momics



http://merenlab.org/momics

Genome-resolved metagenomics *in action*

- Several automatic binning algorithms availabe
  - 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

Let’s get to work

https://github.com/karkman/MMB-901_Metagenomics