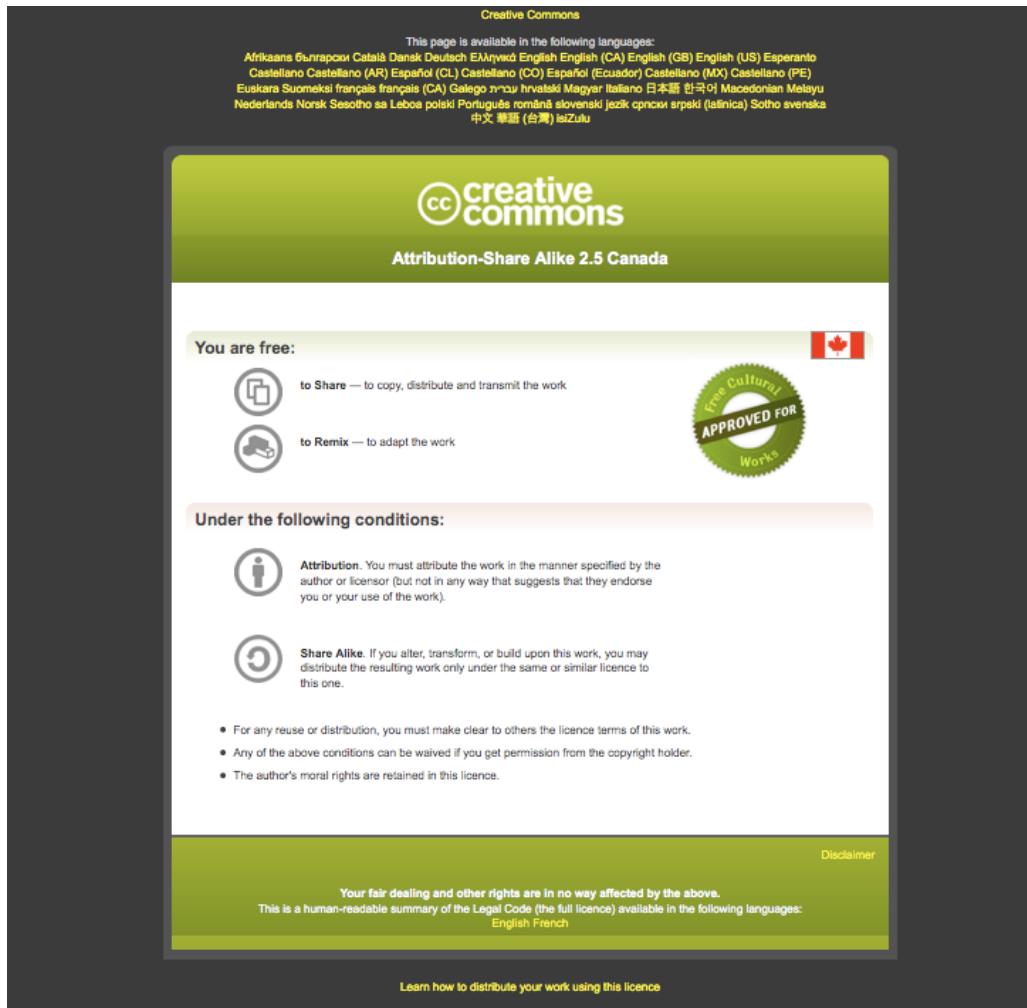




Canadian Bioinformatics Workshops

www.bioinformatics.ca

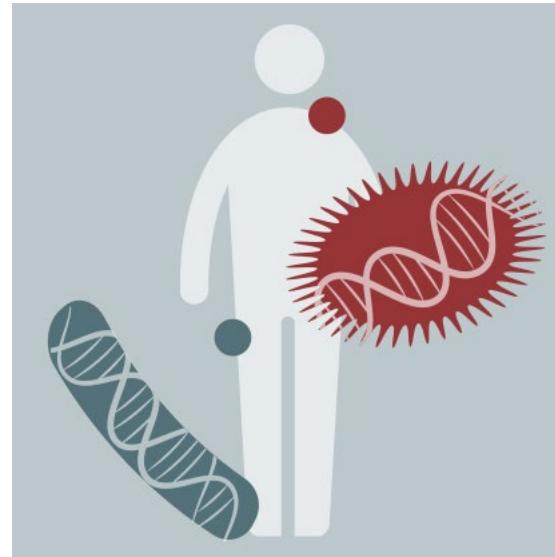
bioinformaticsdotca.github.io





Metagenome Assembly & MAGs

Laura Sycuro, PhD MSc
CBW-IMPAC TT Microbiome Analysis
July 5-7, 2023



**UNIVERSITY OF
CALGARY**



IMC

SYCUR^o LAB



**Adolescent vaginal
microbiome:**

Drivers of *Lactobacillus*
dominance

Risk of STI

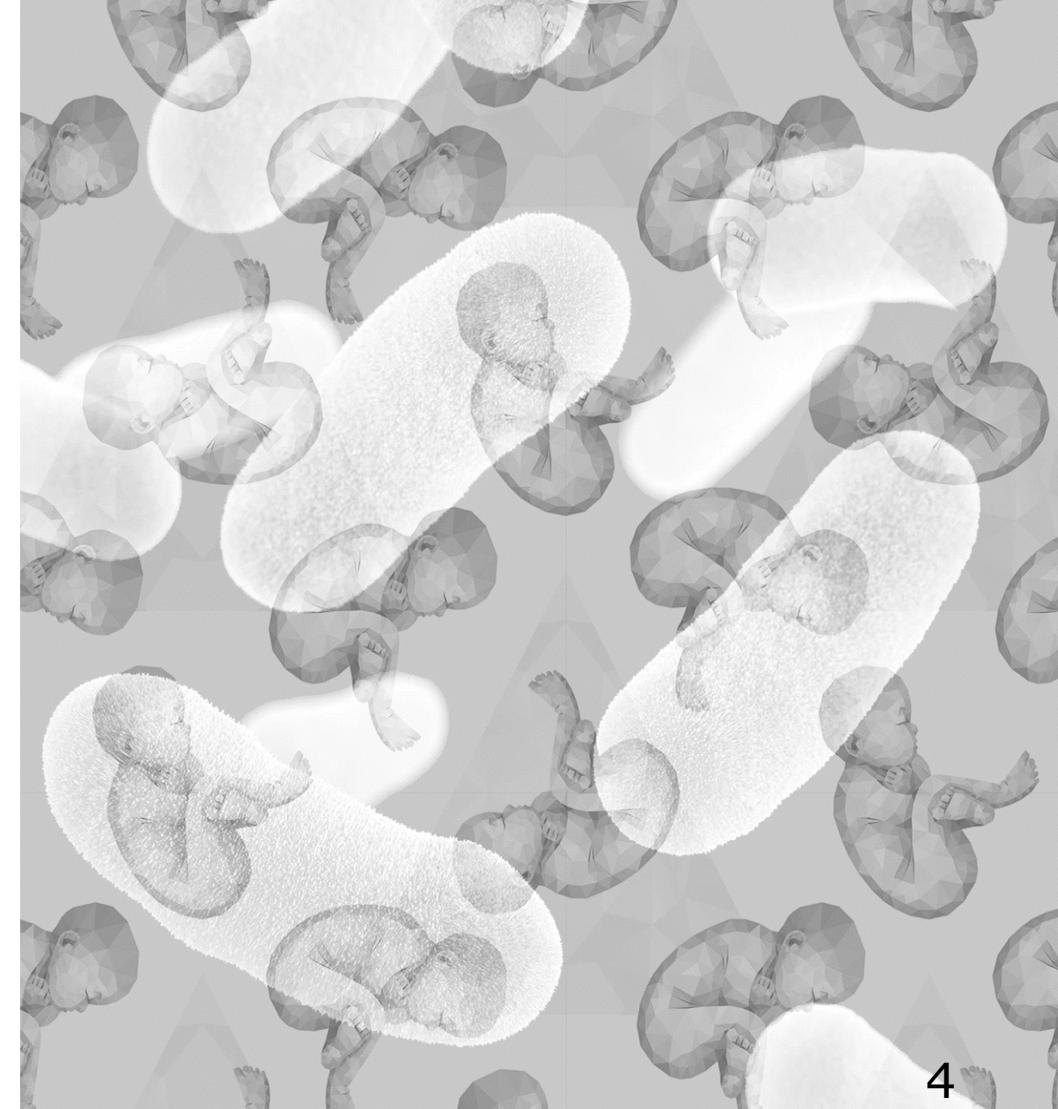
**Vaginal microbes
and preterm birth:**

Novel virulence
mechanisms

**Gut-brain
connection:**

Parkinson's Disease
Neurodevelopment
*Autism, ADHD,
Tourette Syndrome*

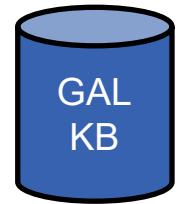
 @sycurol



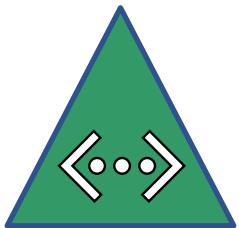
Type of tools/resources presented:



Web-based tool or database



Galaxy/Kbase (cloud-based tools/pipelines)



Command line tool

Learning goals

By the end of this lecture, you will:

Define

- MAG – Metagenome Assembled Genome
- Reads, contigs, scaffolds, coverage, N50, MAG completeness, MAG contamination

Understand

- How metagenomic sequence data is assembled and binned into MAGs
- How to evaluate and maximize the quality of MAGs
- How MAGs can be used to ask biological questions about microorganisms and communities

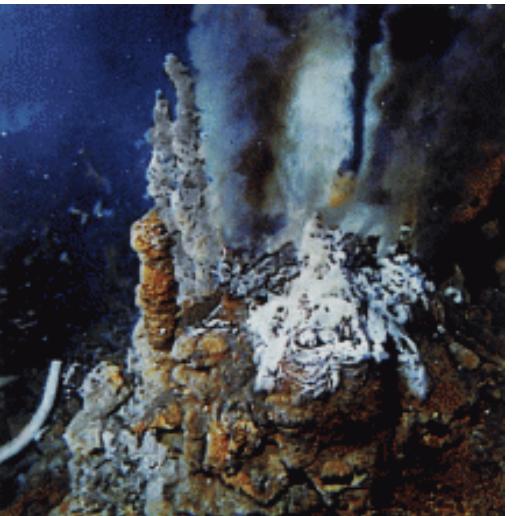
Appreciate

- Why MAGs are useful and when they are obtainable
- The additional information that is accessible from binned metagenomic data

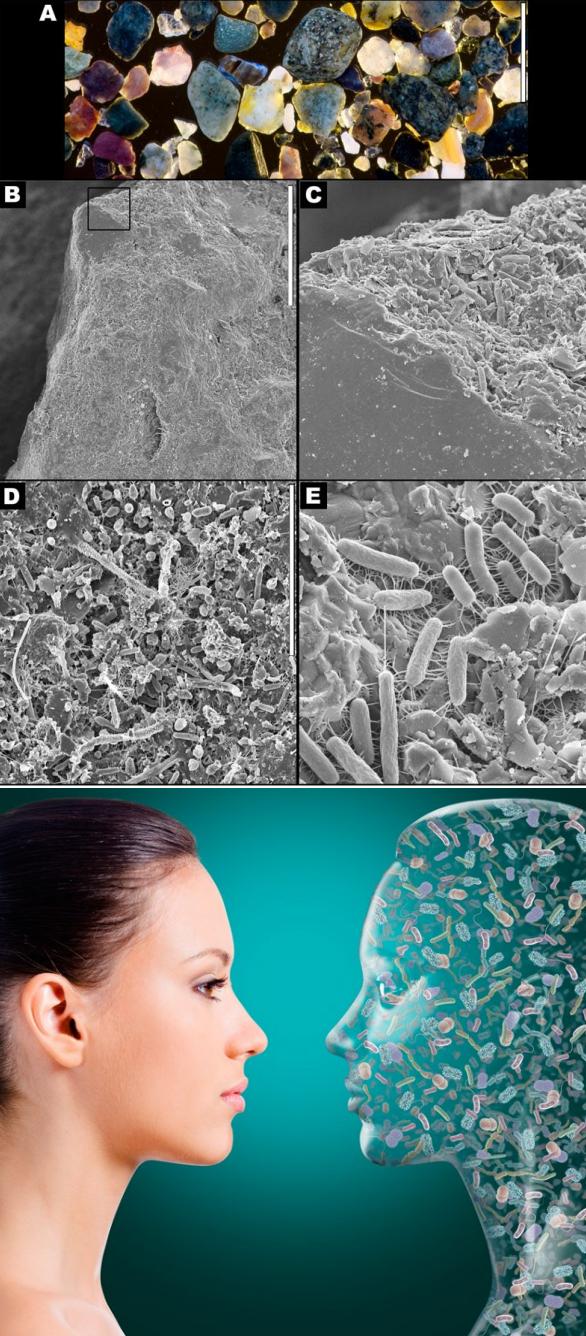
Biomass of microbes
on earth = that of
plants



~1% of microbial
biodiversity has
been cultured



*How do we
discover new
microorganisms
and the
genes/proteins
they encode?*



All tools/approaches in metagenomics have blind spots

Short read metagenomics is 100% reliant on reference databases sequence detection

Until very recently, these databases only contained sequence information from cultured microorganisms

Working with metagenomic assemblies allows you to:

- Detect and study the ‘novel’ content of your metagenome

More finely resolved taxonomic assignment of novel microorganisms

New genes/proteins

- Resolve metagenome assembled genomes (MAGs)

Genome-resolved understanding of novel clades

Allele-resolved understanding of microbial function

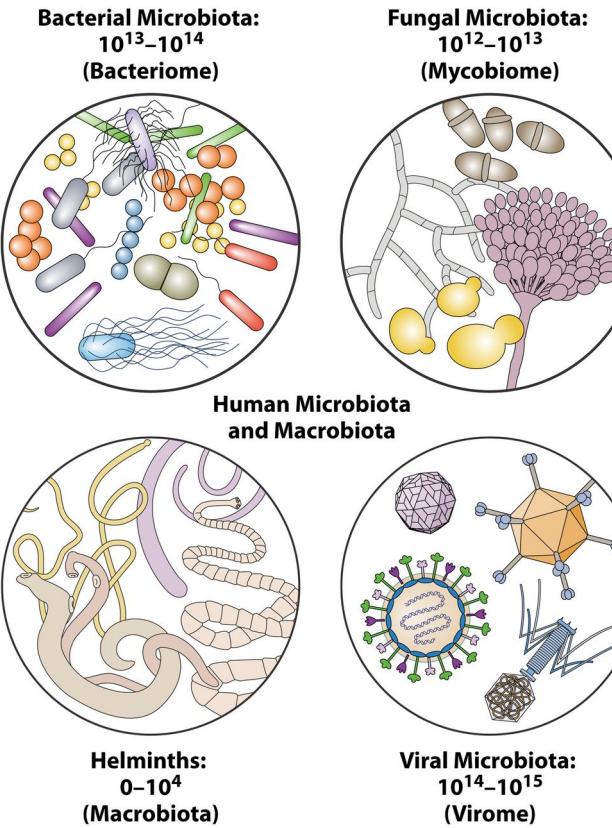
Adaptation & exchange

Microbial metagenomics is a multi-kingdom operation

Does assembly/binning benefit studies of all kingdoms?

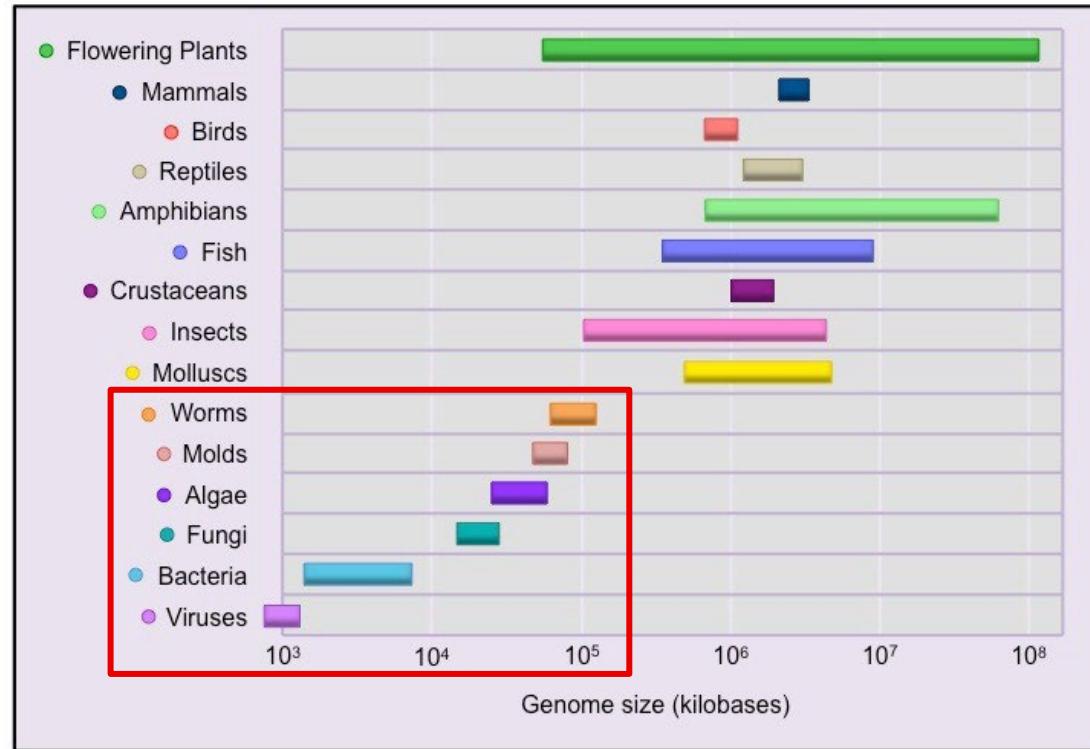
Human Cells: $\sim 10^{13}$

Microbial Cells:



Microbial metagenomics is a multi-kingdom operation

Success of assembly/binning depends on the abundance of the organism(s) in the sample and their genome size



Jangid et al. Microbial Genome Diversity and Microbial Genome Sequencing in Microbial Genomics in Sustainable Agroecosystems 2019 pp. 175-201

Larger genome = more reads (coverage)
Less abundant organism in metagenome = more reads

Where this approach has shone is with bacterial genomes

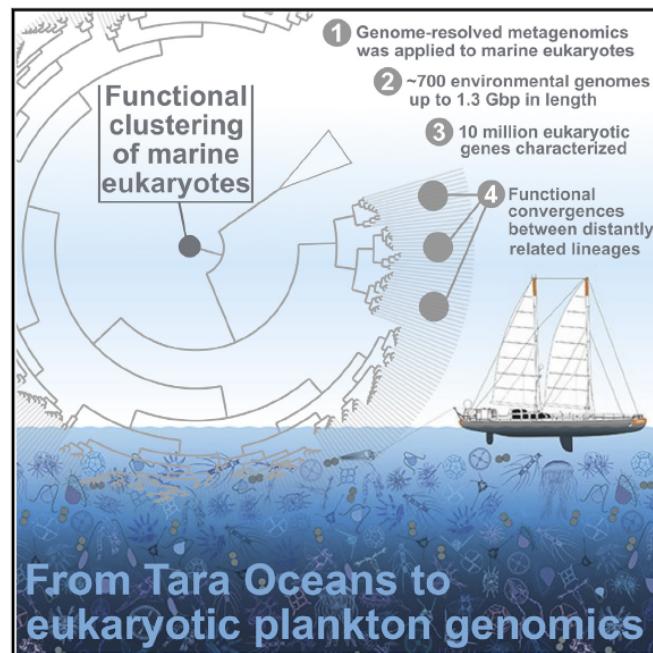


Preview

Metagenome assembled genomes are for eukaryotes too

Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean

Graphical abstract



Authors

Tom O. Delmont, Morgan Gaia,
Damien D. Hinsinger, ..., Eric Pelletier,
Patrick Wincker, Olivier Jaillon

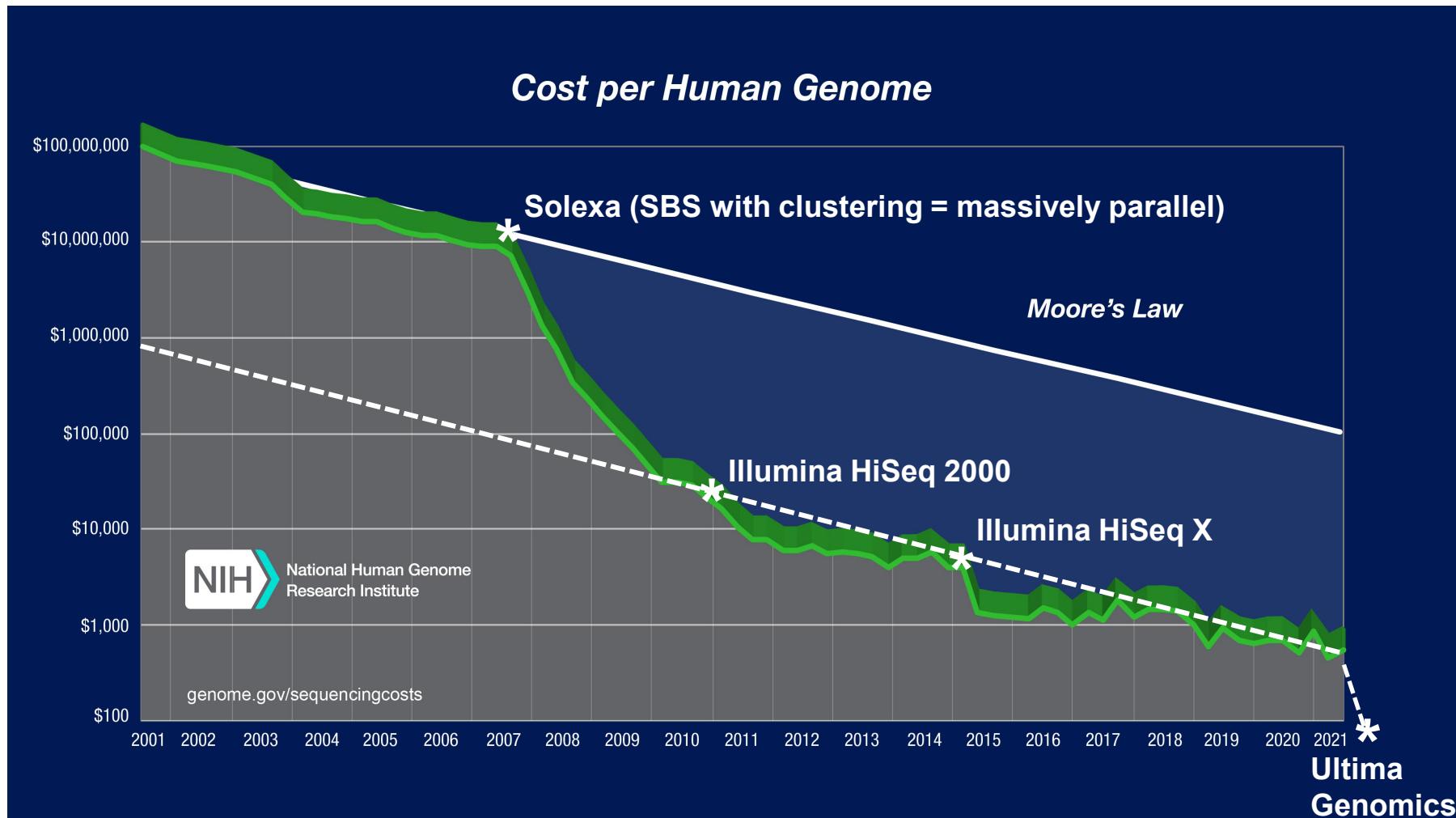
Correspondence

tom.delmont@genoscope.fr

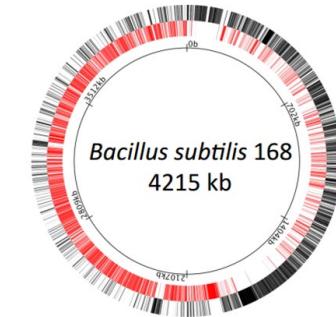
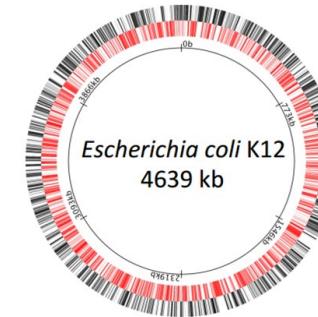
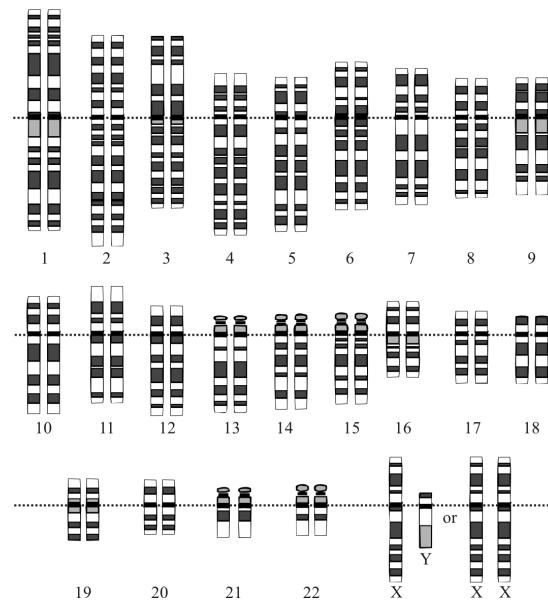
In brief

Delmont et al. use nearly 300 billion metagenomic reads to characterize the genomic content of some of the most abundant and widespread eukaryotic populations in the sunlit ocean. This large genomic resource covers taxa underrepresented in our culture portfolio and exposes a functional convergence of distantly related eukaryotic plankton lineages.

Advances in sequencing technologies have been driven by the quest for the \$1000 >> \$100 >> \$10 human genome



Each genome requires the same library prep process: \$\$



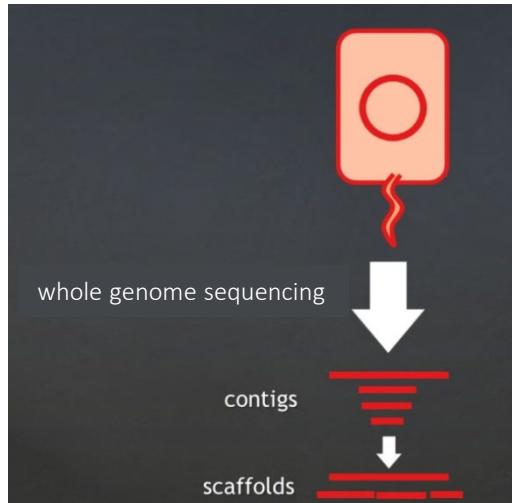
~3,000 Mb ————— ~4 Mb
750X

~\$1000 ————— ~\$100
10X

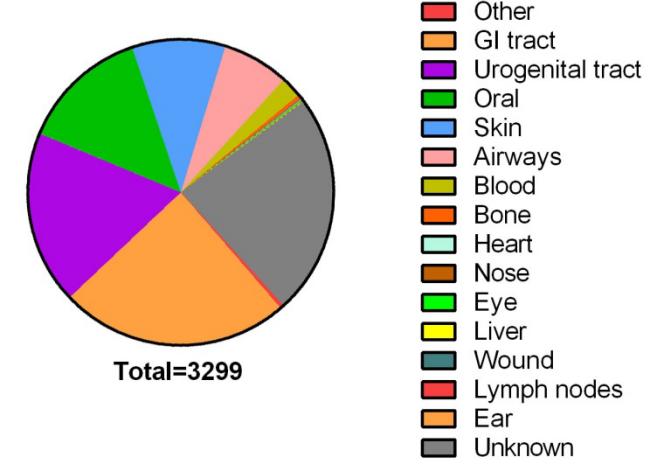
~\$0.33/Mb ————— ~\$25/Mb

What genomes are in our databases?

One Illumina instrument could be sequencing 6,600 genomes/wk...



HMP Reference Genome Project

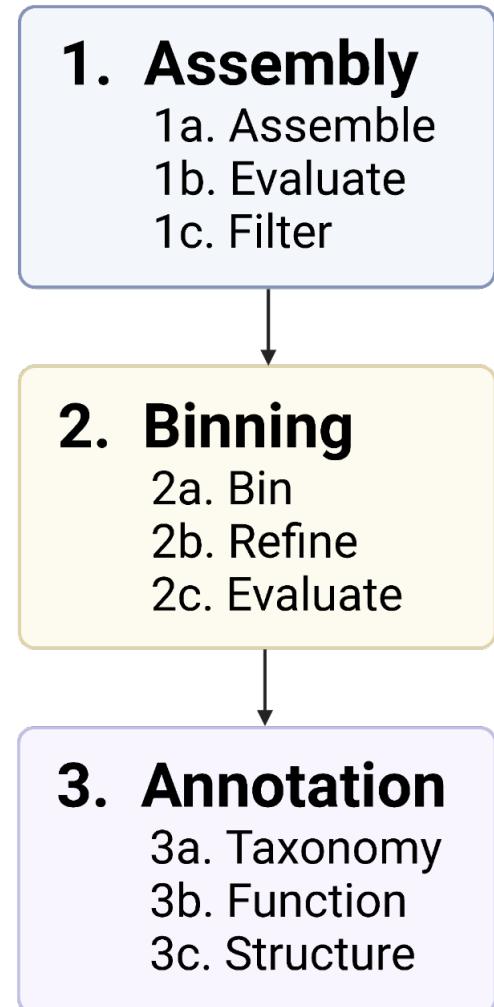


Snapshot of what's in the databases (Oct 2022):

NCBI – 1,186,327	Prokaryotes
899,577	Proteobacteria (75.8%)
175,685	<i>Escherichia coli</i> (14.8%)
114	<i>Phocaeicola</i> (formerly <i>Bacteroides</i>) <i>dorei</i> (0.01%)
1	<i>Sneathia sanguinegens</i> (0.00008%)

Ensembl – 32,332	Prokaryotes
1,232	<i>Escherichia coli</i> (3.8%)
13	<i>Phocaeicola</i> (formerly <i>Bacteroides</i>) <i>dorei</i> (0.04%)
0	<i>Sneathia sanguinegens</i>

Process of generating metagenome assembled genomes:

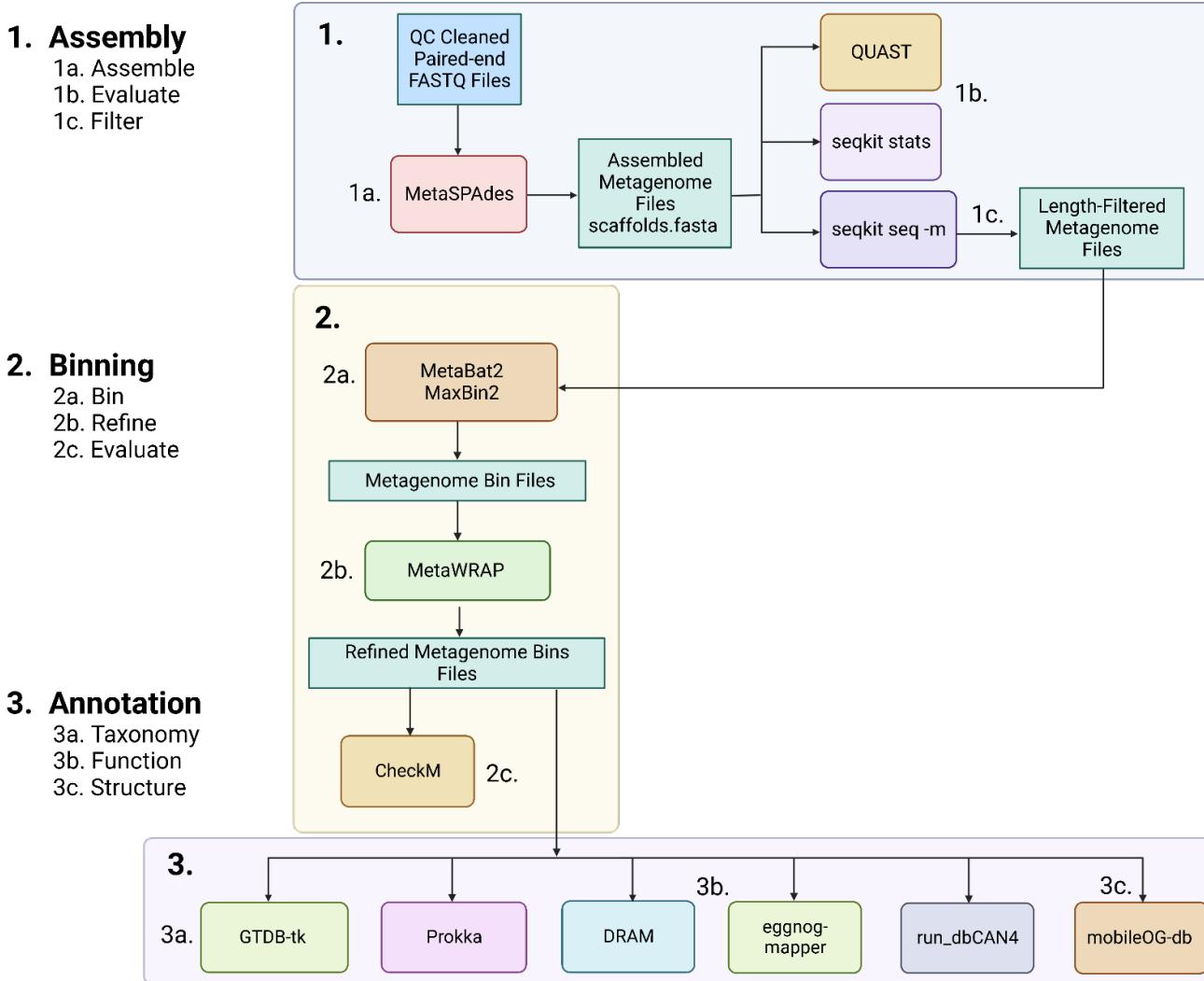


***** Computationally intensive***

***** Requires quality assessment/human intervention within each of the 3 steps***

***** Steps 1-2 relatively 'standardized' but appropriate choices for step 3 depend heavily on system/question***

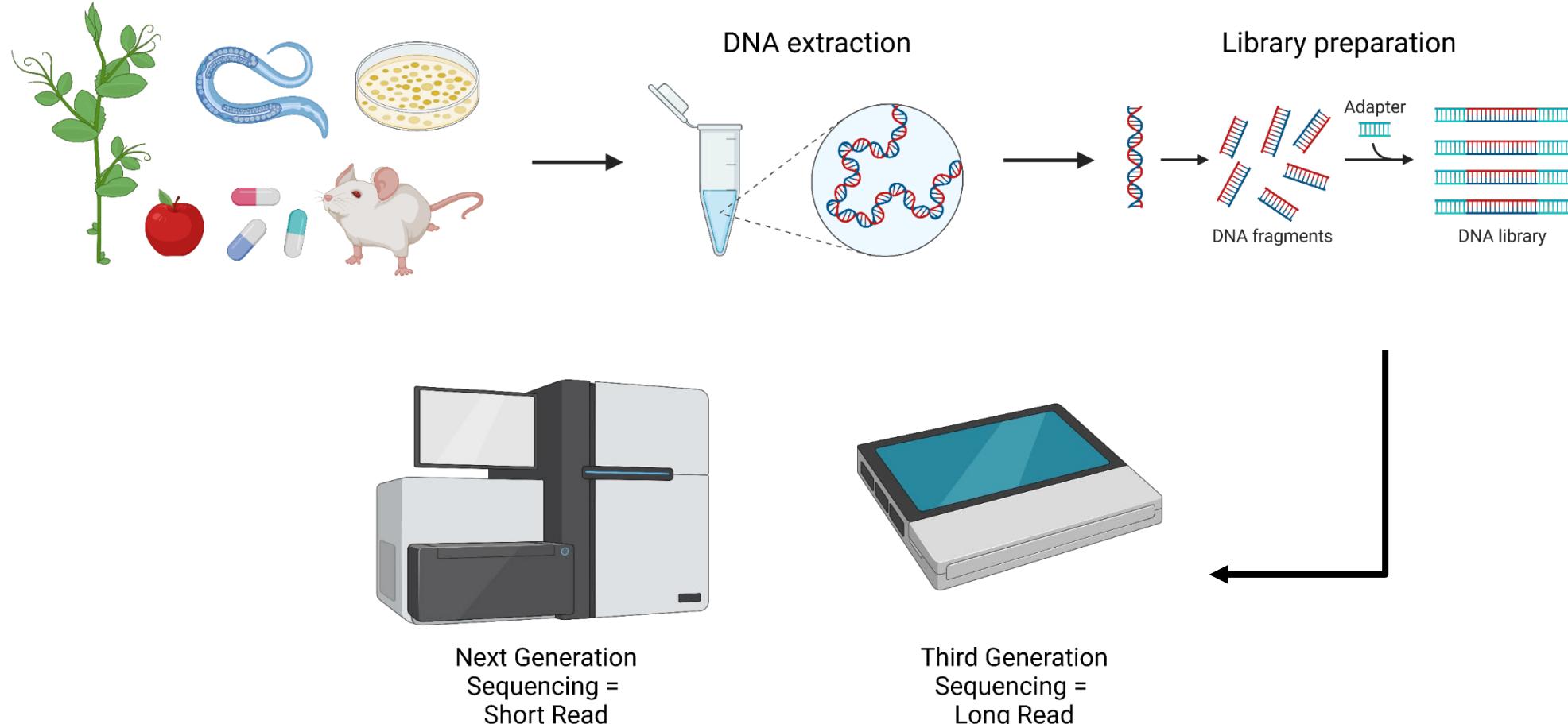
Process of generating metagenome assembled genomes:



Created with biorender.com

What have we done so far?

- 1. Assembly**
 - 1a. Assemble
 - 1b. Evaluate
 - 1c. Filter

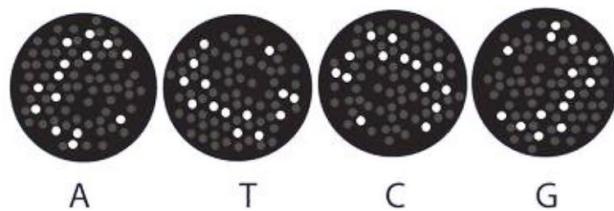
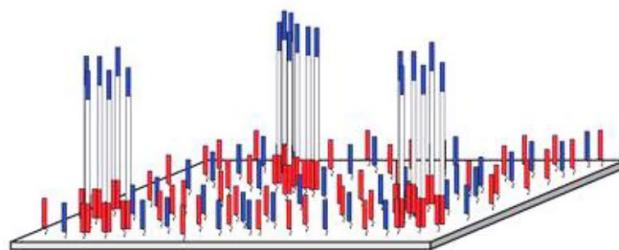


What kind of reads are best for assembly?

'Next Generation Sequencing' = NGS

Second Generation

Massively Parallel Sequencing



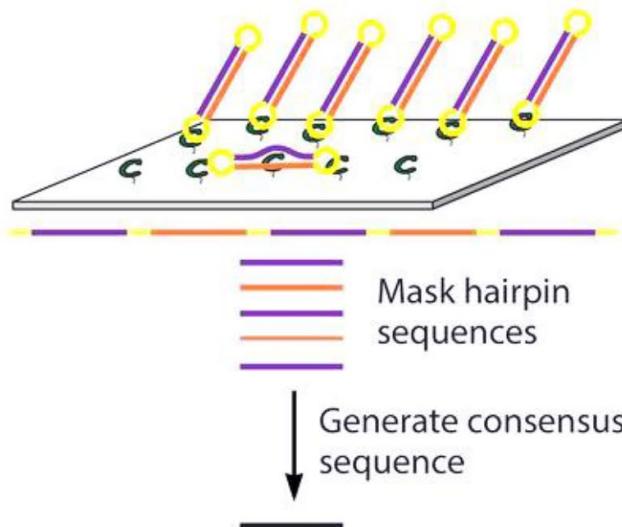
- Sequencing by synthesis
- Amplified templates are generated during sequencing, reducing the requirements for starting material
- High accuracy
- Short read lengths

e.g., MiSeq (Illumina), Ion Torrent
(Thermo Fisher Scientific)

'Long Read Sequencing'

Third Generation

Single-molecule Sequencing



- Single-molecule templates
 - Low accuracy
 - Long read lengths
- e.g., Single-Molecule Real-Time (SMRT) — Sequencing (Pacific Biosciences), MinION (Oxford Nanopore Technologies)

Any type can be used, but match algorithm to technology/read type

Factors that reduce assembly quality:

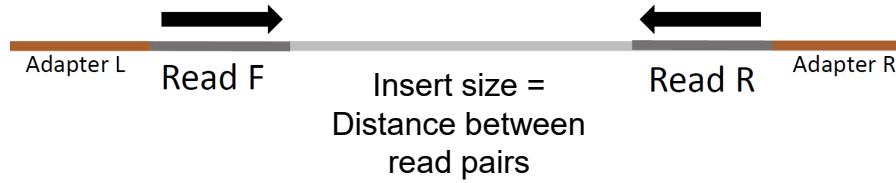
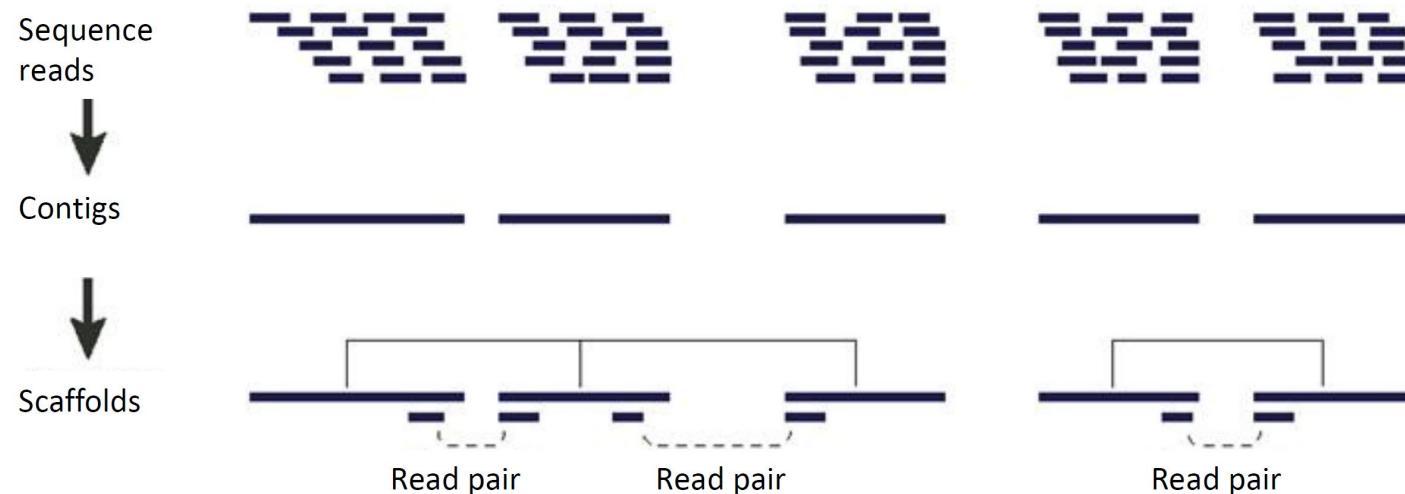
- Poor quality (base calling errors, N's, adapters)
- Small fragment/insert size (relative to read length)
- Short reads (<100 bp)
- Low coverage

Most metagenomic assemblies are done with short reads only, but long reads improve contiguity

Assembly in a nutshell:

De novo assembly

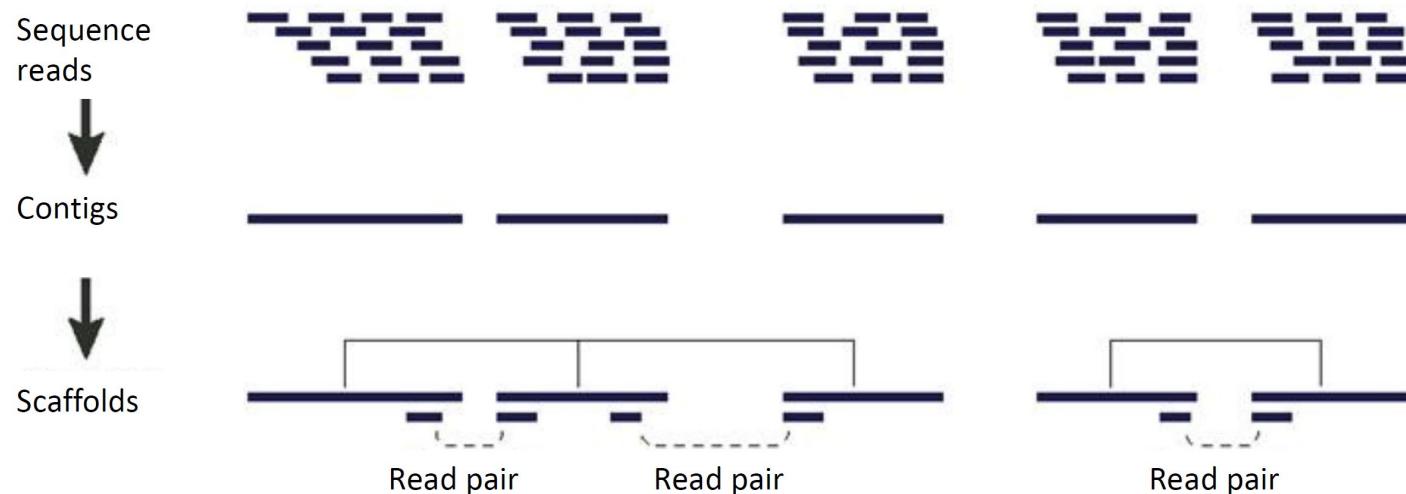
Scaffolding with
paired reads mapped
back to assembly



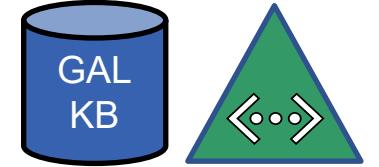
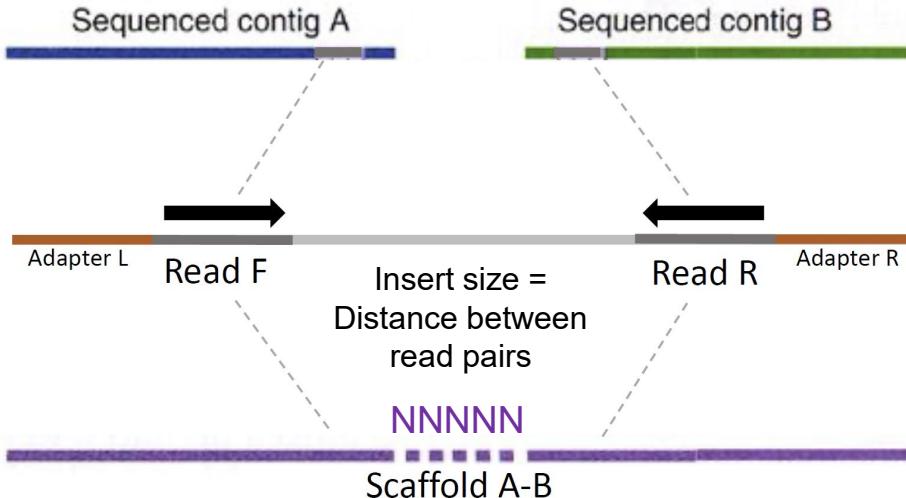
Assembly in a nutshell:

De novo assembly

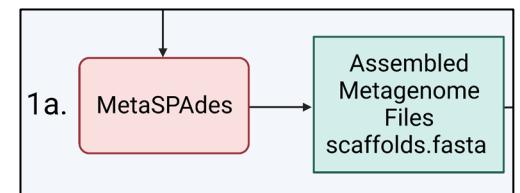
Scaffolding with paired reads mapped back to assembly



Nurk et al., Genome Res. 2017 27(5): 824–834

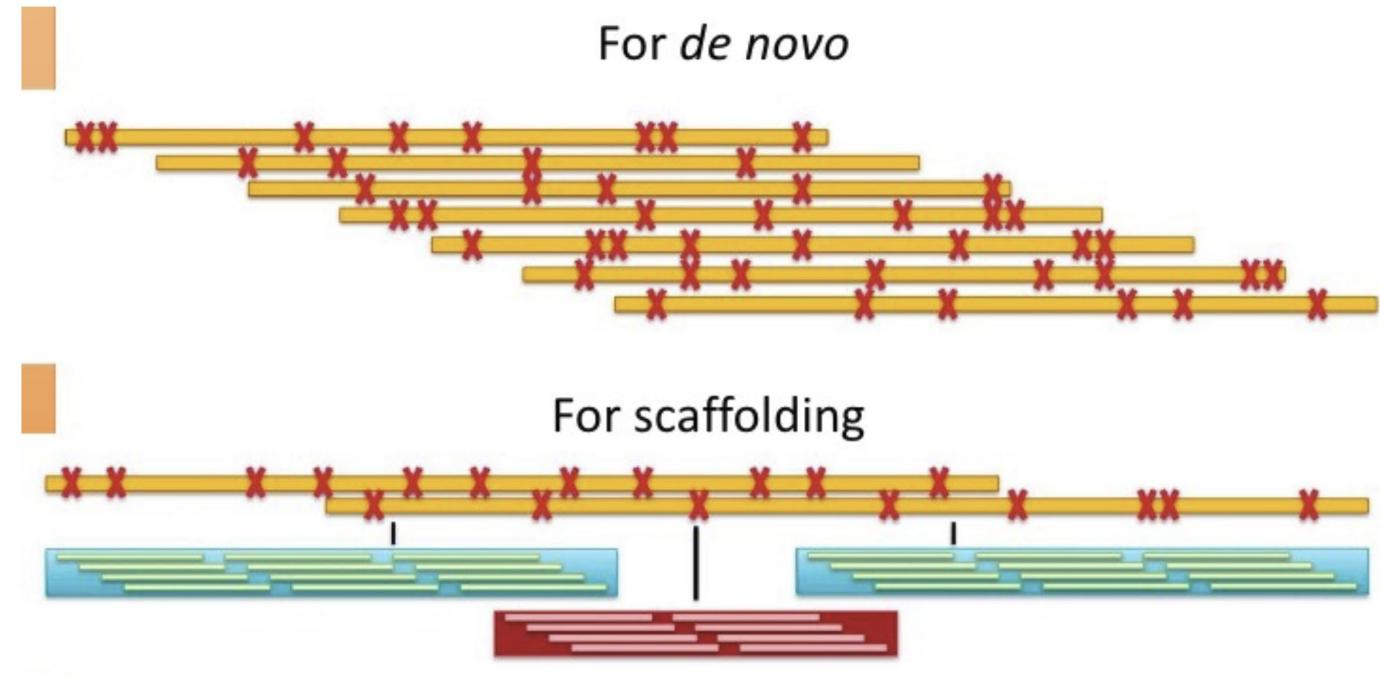
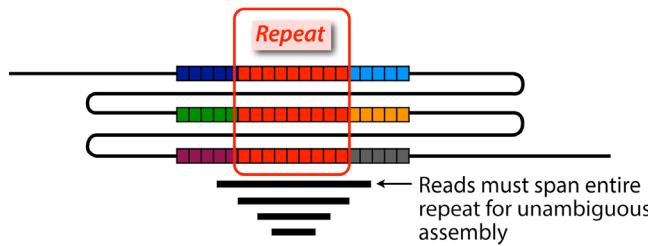


Tutorial preview:



Scaffolding with long read sequencing:

Longer reads can help bridge otherwise ‘intractable’ parts of the genome

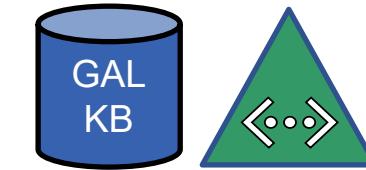


*4 Mb reads have been obtained on Oxford Nanopore MinION,
exceeding the length of a bacterial genome!*

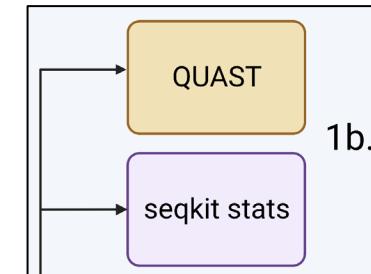
So you've assembled your metagenome, now what?

Check quality of assembly:

- QUAST or MetaQUAST
 - Number of contigs/scaffolds
 - Total assembled length
 - N50 = measure of contiguity
 - at least 50% of nucleotides in assembly belong to contigs/scaffolds N50 length or longer
 - Misassembly
 - requires reference genome(s)



Tutorial preview:



Mikheeno et al., Bioinformatics. 2016 32(7): 1088–1090
Gurevich et al., Bioinformatics. 2013 29(8): 1072-1075

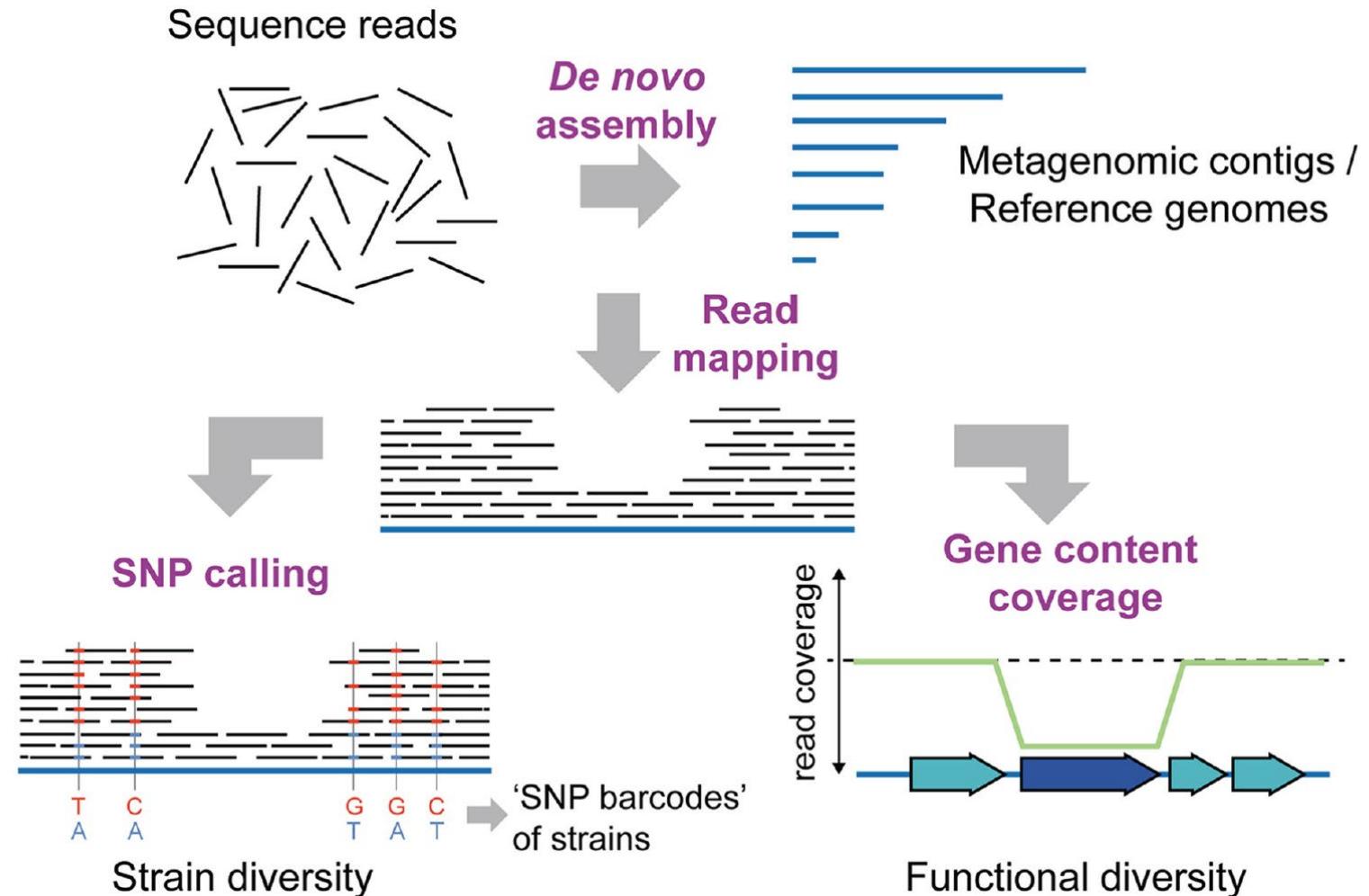
Evaluate choice of assembly algorithm, assembly parameters, implications of length filtering

So you've assembled your metagenome, now what?



1. Assembly

- 1a. Assemble
- 1b. Evaluate
- 1c. Filter



2. Binning

Binning MAGs: *Metagenome Assembled Genomes*

Solving the metagenomics puzzle

Puzzle pieces = *DNA scaffolds*

Strategy for assembly = *binning*

Surface of piece = *patterns in sequence content (tetranucleotide or GC)*

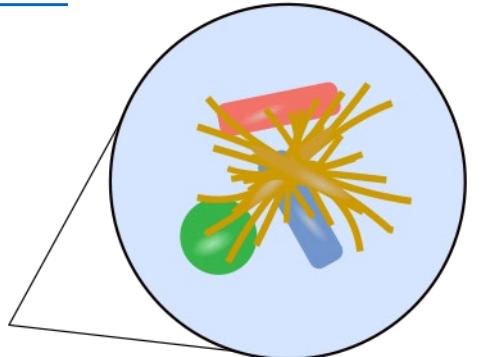
Shape of piece = *abundance of the sequence*

Binning tools:

- Anvi'o
Murat et al. *Nature Microbiology*. 2020 6(1): 3:6
[https://anvio.org/](https:// anvio.org/)
<https://anvio.org/vocabulary/>
- MaxBin2
- MetaBAT2



Mixed
community



Extract
DNA



Shotgun
Metagenomic
Sequencing



Reads

Scaffolds



Assembly

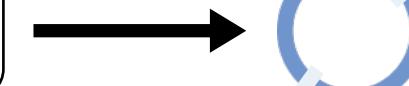
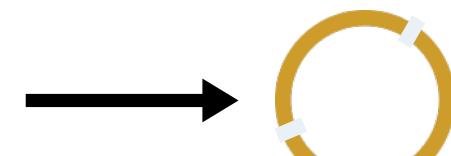
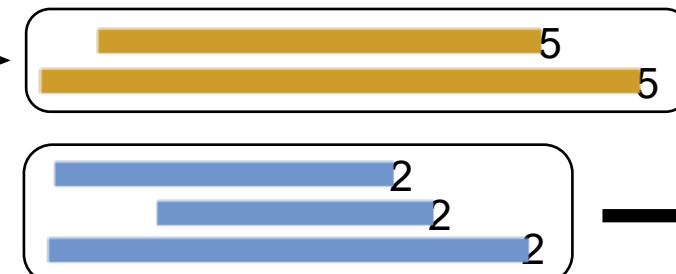
Generate
Scaffold Info

Coverage

	s1	s2	...
AAAA	5	12	...
AAAT	0	6	...
AAAC	14	7	...
AAAG	3	8	...
...

Sequence
Characteristics

Binning
Scaffolds



25

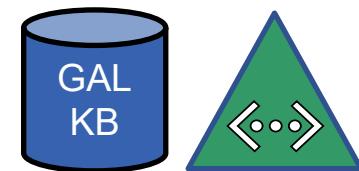
Refining bins:

Binning refinement tools:

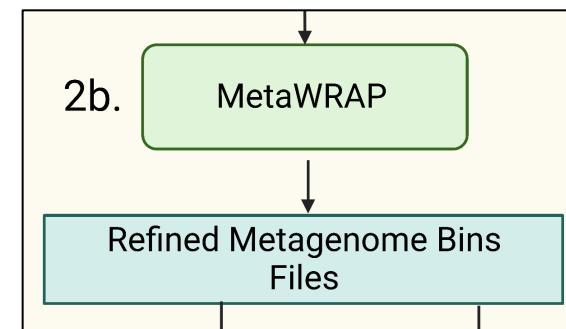
- DAS Tool Sieber et al. *Nature Microbiology*. 2018 3:836–843
- Binning_refiner Song et al. *Bioinformatics*. 2017 33(12):1873–1875

General approaches:

- Iterative aggregating & deduplicating bins
- Bin quality metrics
 - completeness & contamination (based on single copy genes)
- Reassembly



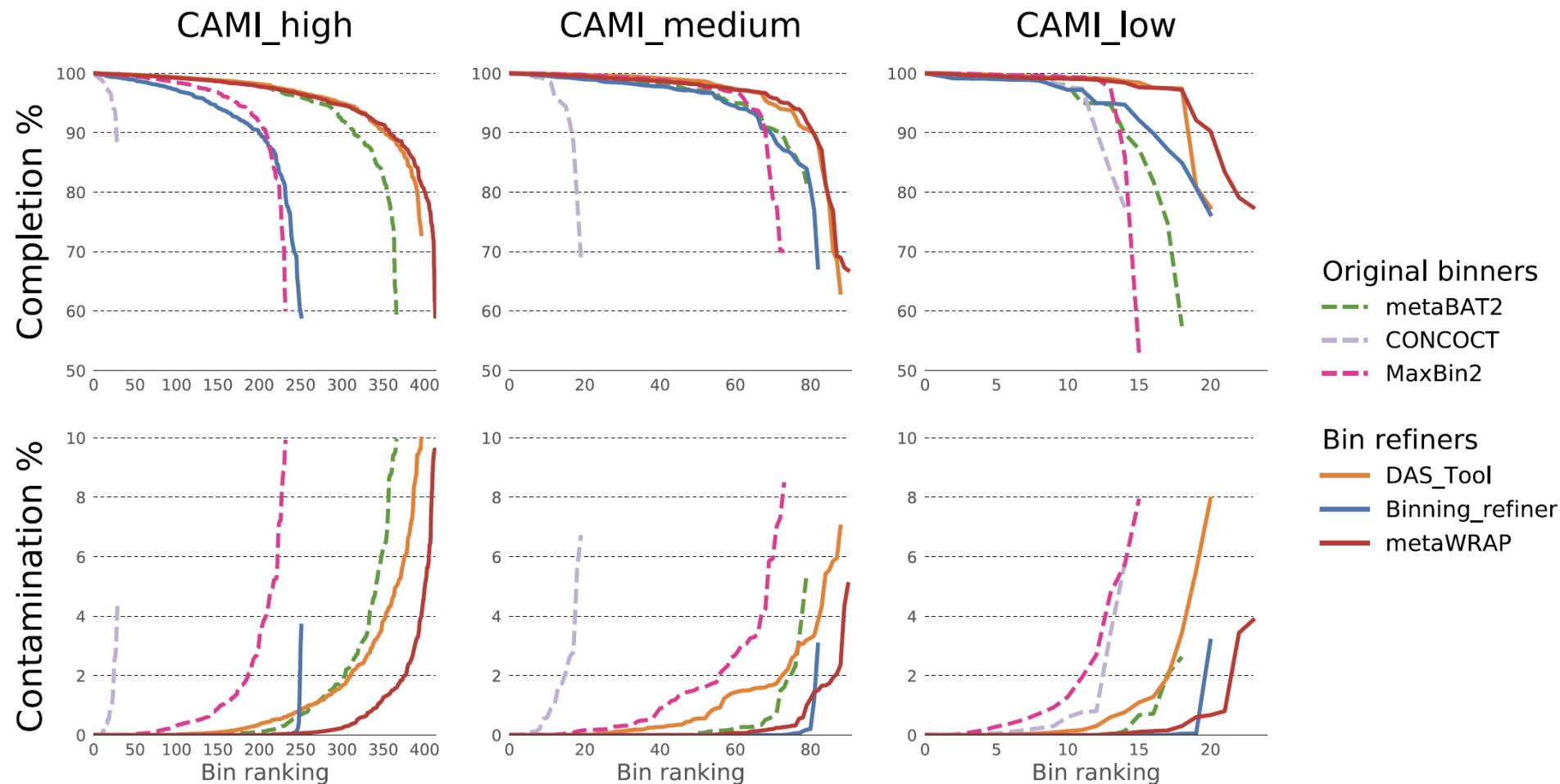
Tutorial preview:



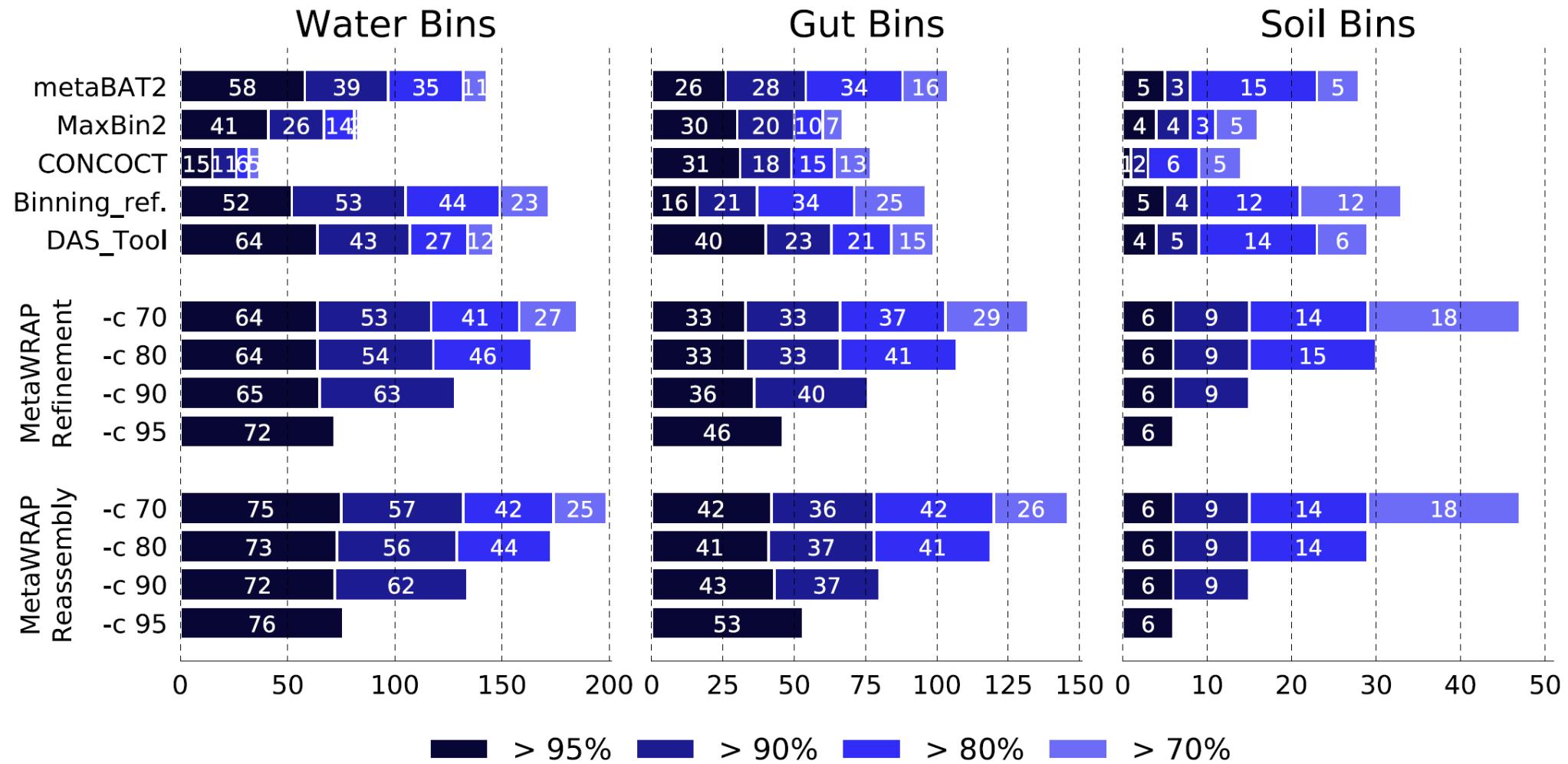
Uritskiy et al. *Microbiome*. 2018 6(158)

Refining bins:

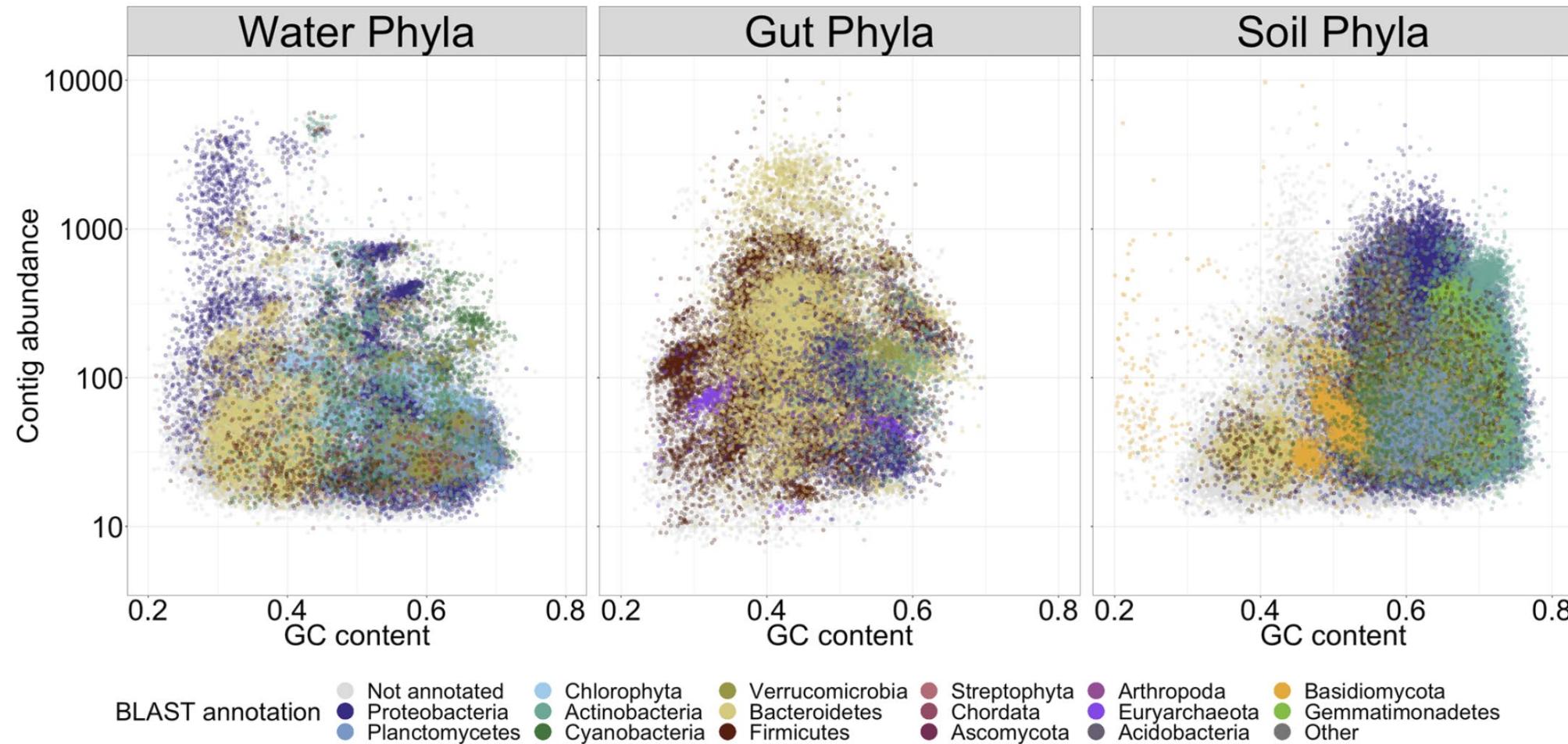
CAMI = Critical Assessment of Metagenome Interpretation



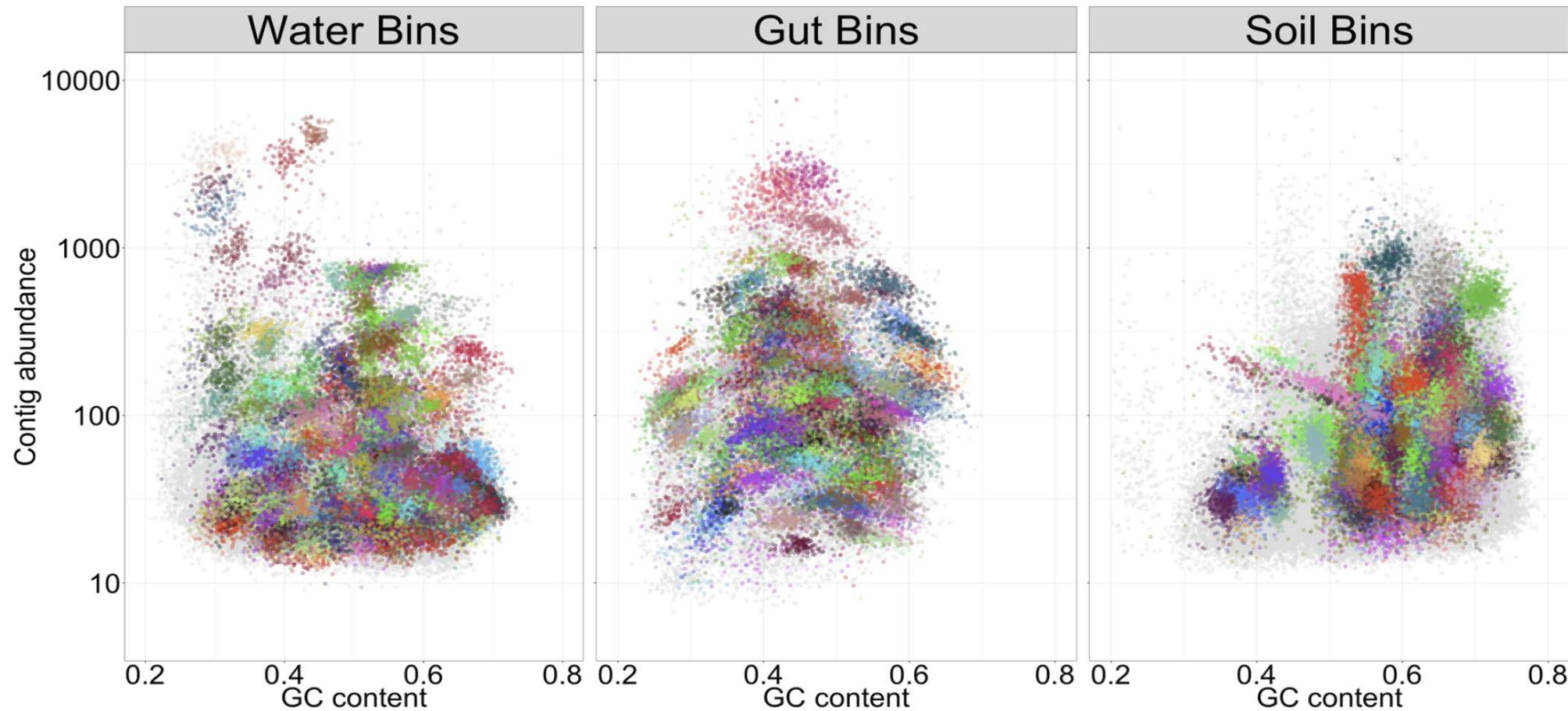
Refining bins:



Refining bins:



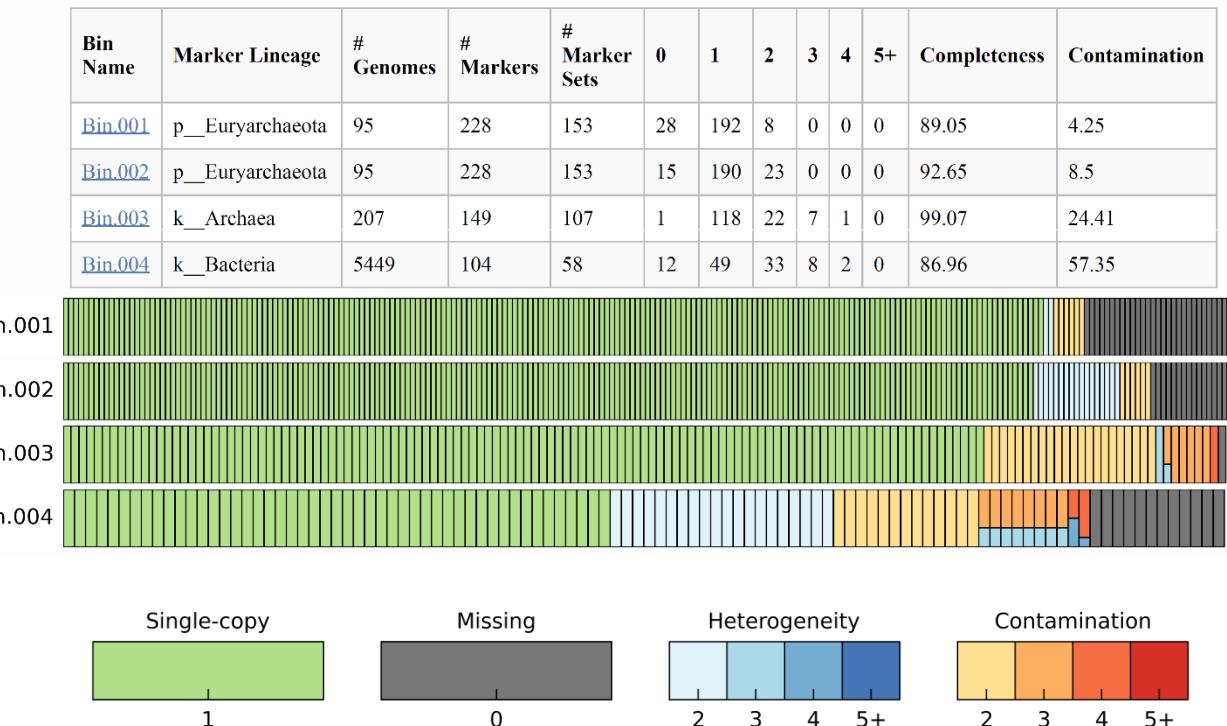
Refining bins:



Evaluating bin quality:

CheckM:

- Define completeness and contamination using lineage-specific single copy genes
 - Place genome on tree universal marker gene tree
 - Define lineage-specific sets of marker genes
 - Account for colocalization (operons)
 - How many are present: % completeness
 - How many are duplicated: % contamination
 - Distinguishes between strain heterogeneity and species contamination
- Other bin quality metrics (QUAST)
 - Total length, number of contigs, N50



Parks et al. Genome Research. 2014 25: 1043-1055

Evaluating bin quality:

CheckM:

- Define completeness and contamination using lineage-specific single copy genes
 - Place genome on tree universal marker gene tree
 - Define lineage-specific sets of marker genes
 - Account for colocalization (operons)
 - How many are present: % completeness
 - How many are duplicated: % contamination
 - Distinguishes between strain heterogeneity and species contamination
- Other bin quality metrics (QUAST)
 - Total length, number of contigs, N50

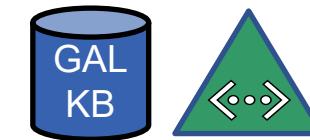
Table 3. Controlled vocabulary of draft genome quality based on estimated genome completeness and contamination

Completeness	Classification	Contamination	Classification
≥90%	Near	≤5%	Low*
≥70% to 90%	Substantial	5% to ≤10%	Medium
≥50% to 70%	Moderate	10% to ≤15%	High
<50%	Partial	>15%	Very high

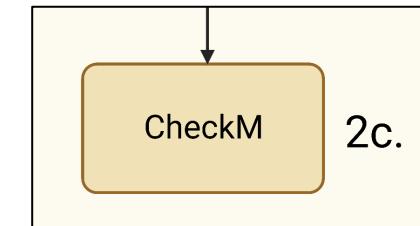
(*) Genomes estimated to have 0% contamination can be designated as having “no detectable contamination”.

QUALITY DEFINITIONS FOR PUBS:

Bowers R et al. Nat Biotechnol. 2017 35:725–731



Tutorial preview:



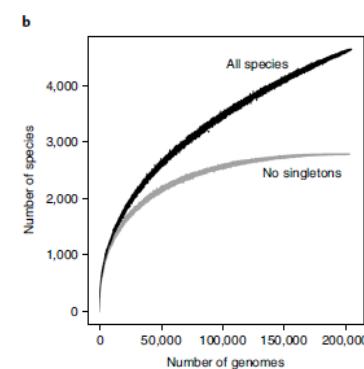
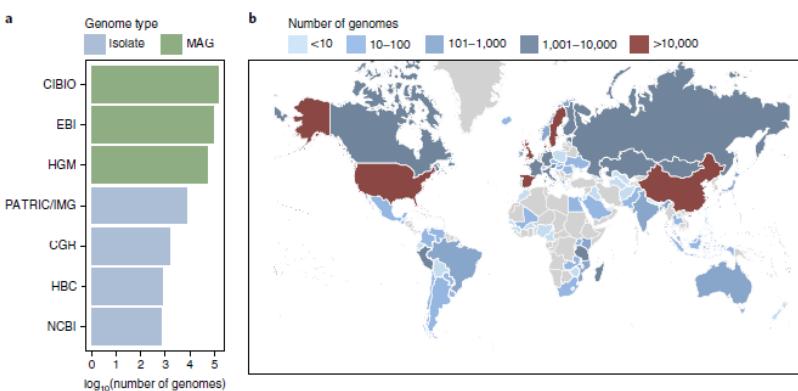
*Happy with parameters and data?
Bins ≥50% (or sometimes 80-90%) complete
with ≤10% (or sometimes 5%) contamination
= MAGs*



2. Binning

- 2a. Bin
- 2b. Refine
- 2c. Evaluate

How it started...



Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life

Donovan H. Parks , Christian Rinke , Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz * and Gene W. Tyson*

NATURE MICROBIOLOGY | VOL 2 | NOVEMBER 2017 | 1533–1542 | www.nature.com/naturemicrobiology

Resource

Cell

Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle

Edoardo Pasolli • Francesco Asnicar ⁸ • Serena Manara ⁸ • ... Christopher Quince • Curtis Huttenhower • Nicola Segata ⁹ • Show all authors • Show footnotes

Open Access • Published: January 17, 2019 • DOI: <https://doi.org/10.1016/j.cell.2019.01.001> •



OPEN

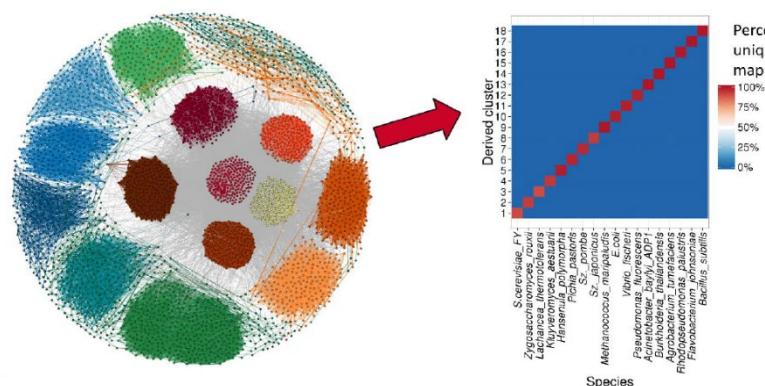
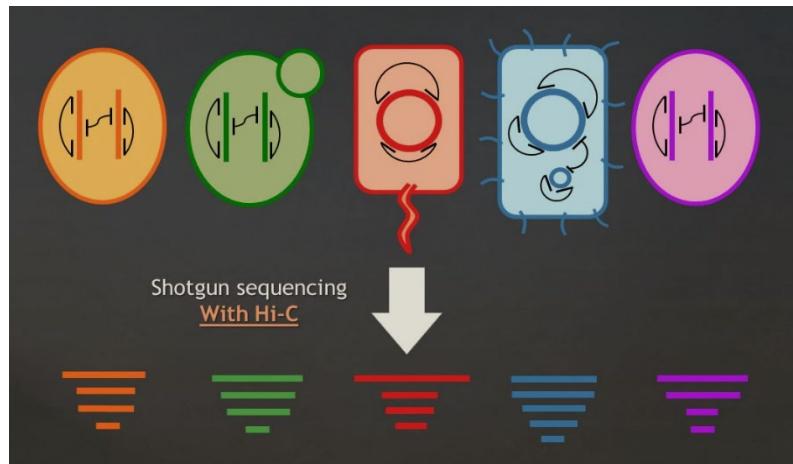
A unified catalog of 204,938 reference genomes from the human gut microbiome

Alexandre Almeida ^{1,2} , Stephen Nayfach ^{3,4}, Miguel Boland ¹, Francesco Strozzi ⁵, Martin Beracocha ¹, Zhou Jason Shi ^{6,7}, Katherine S. Pollard ^{6,7,8,9,10,11}, Ekaterina Sakharova ¹, Donovan H. Parks ¹², Philip Hugenholtz ¹², Nicola Segata ¹³, Nikos C. Kyrpides ^{3,4} and Robert D. Finn ¹



2. Binning

- 2a. Bin
- 2b. Refine
- 2c. Evaluate



New horizons and variations on the theme:

- Serial sampling
 - More reads/taxon, assembly ‘sweet spots’
- Hi-C contact maps
 - Linkage of plasmids with chromosome and phage with host
 - With long reads...
 - Strain/lineage resolution
 - Closed (circular) MAGs

Burton et al. G3. 2014 4(7):1339-46



Bickhart et al. Nat Biotechnol. 2022 40:711–719

- CheckV
 - Viral genomes
- CheckM2
 - ML

Nayfach et al. Nat Biotechnol. 2021 39:578–585

<https://www.biorxiv.org/content/10.1101/2022.07.11.499243v1>

3. Annotation

Annotating MAGs:

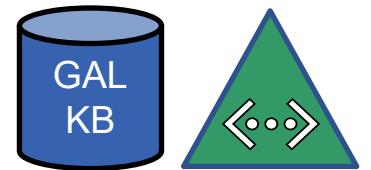
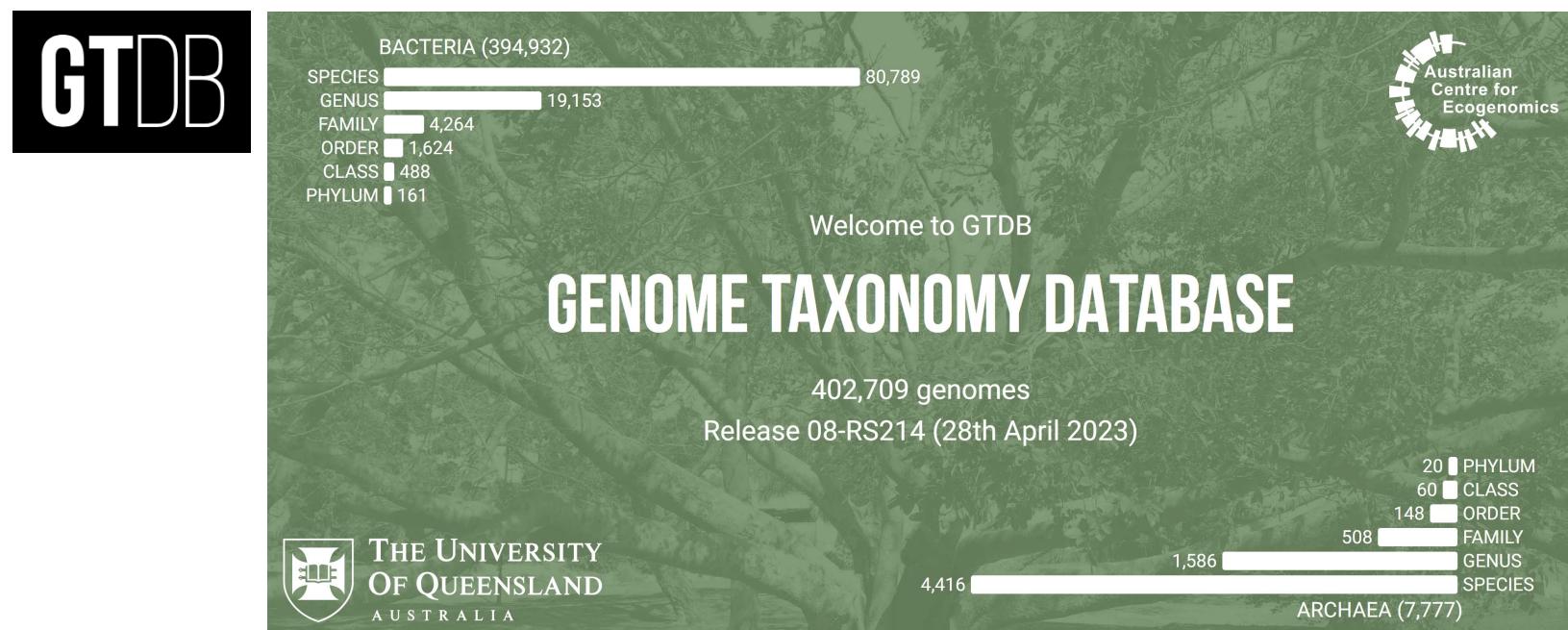
Taxonomy:

- GTDB-tk
 - FastANI v1.3
 - 95-97% Average Nucleotide Identity (ANI)
 - 50% Alignment Fraction

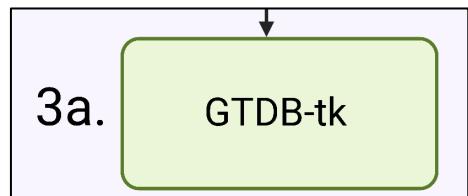
A bit of a thorny problem that depends on which taxonomy you prefer....

... I use the Genome Taxonomy DataBase (GTDB) and their toolkit. If you don't agree with that taxonomic system, there are a wealth of tools available (single gene or multi-gene phylogenies, MEGAN, machine learning approaches) that can be applied to the NCBI taxonomic system.

-Laura Hug, CBW 2022



Tutorial preview:

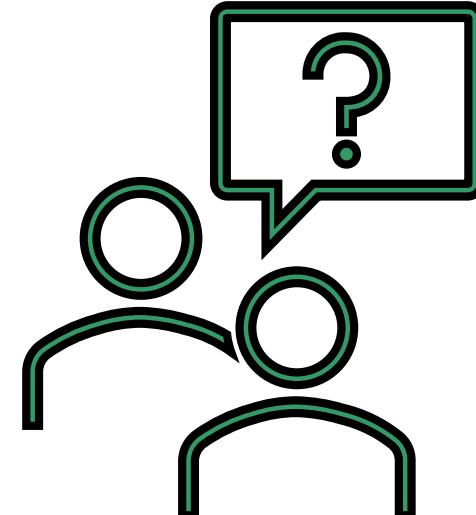


Chaumeil PA et al. Bioinformatics. 2022 btac672

Annotating MAGs:

Calling open reading frames (ORFs) and annotating genes/structural elements:

- What are you looking for?
 - Metabolic pathways, functional groups, certain types of proteins, novel proteins, other genes/structural elements?
 - » Package/pipeline?
 - » Stand-alone/specialized tool?
- Scale of project?
 - Number of MAGs
- How to access tools/databases?
 - Computational resources
 - Database licenses



Annotating MAGs:

Calling open reading frames (ORFs) and annotating gene functions:

- RAST



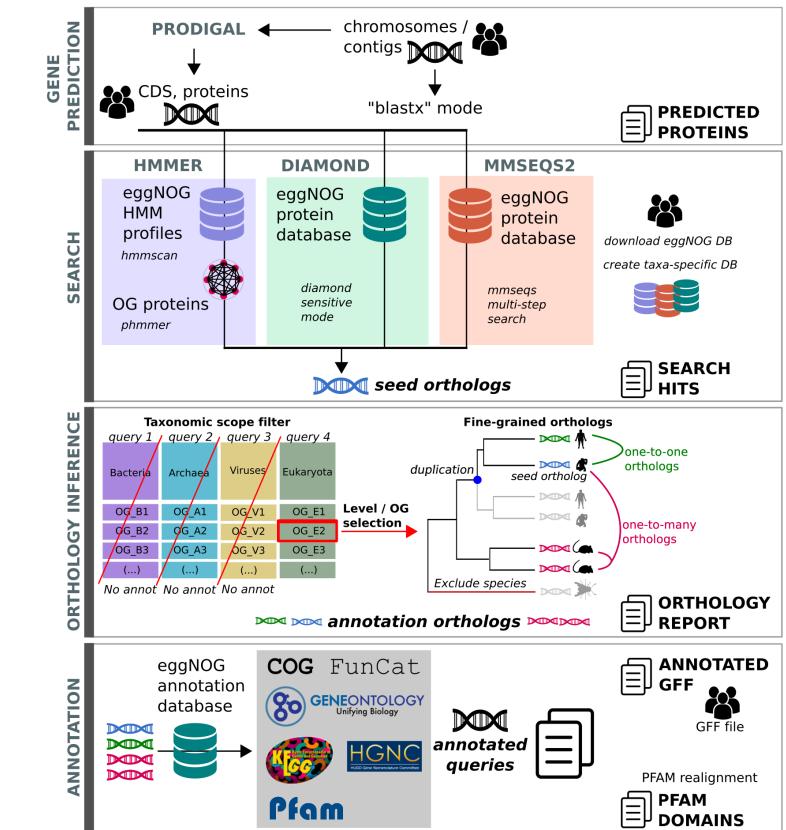
<https://rast.nmpdr.org/>

- SEED database

- eggNOG Mapper



<http://eggnog-mapper.embl.de/>



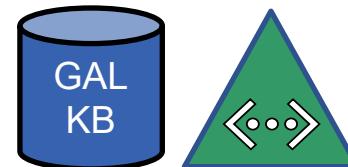
Cantalapiedra et al. Molecular Biology and Evolution.

2021 38(12):5825–5829

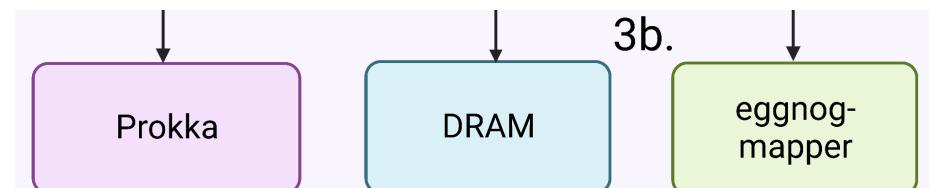
Annotating MAGs:

Calling open reading frames (ORFs) and annotating gene functions:

- RAST    <https://rast.nmpdr.org/>
 - SEED database
- eggNOG Mapper   <http://eggnog-mapper.embl.de/>
- Prokka   <https://github.com/tseemann/prokka>
 - Quick/few databases
- DRAM  <https://github.com/WrightonLabCSU/DRAM>
 - Longer/many databases
 - Metabolic pathways/models
 - Enzymatic activities (CAZymes, proteases)
 - Viruses



Tutorial preview:



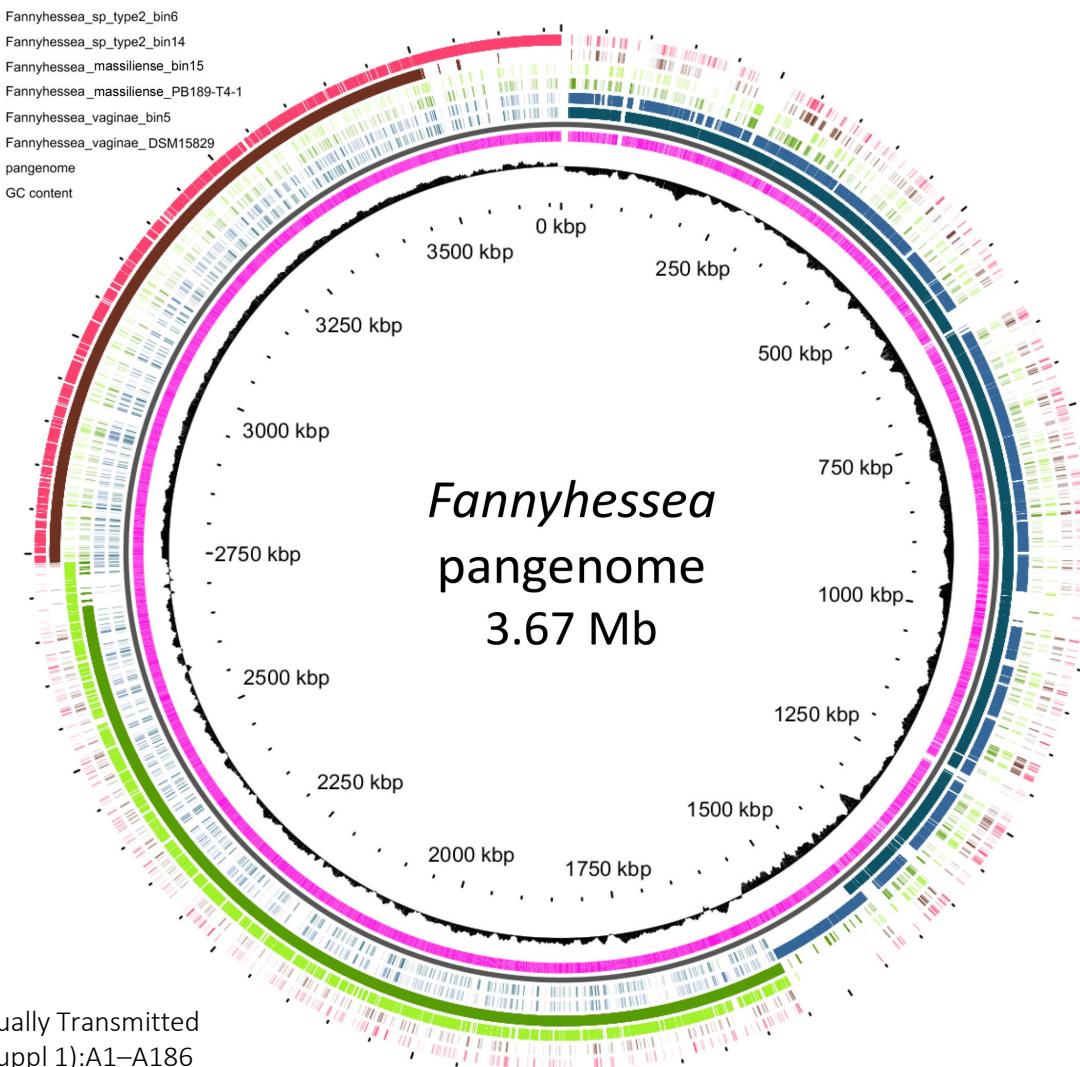
Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades

d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;
o__Coriobacteriales;f__Atopobiaceae;g__Fannyhessea;s__

F. vaginiae, *F. massiliense*, and *F. sp.* type 2 share no more than 73% global average nucleotide identity (gANI) and 45% of their gene content However, full length 16S rRNA gene identity amongst the 3 species is >98.4%

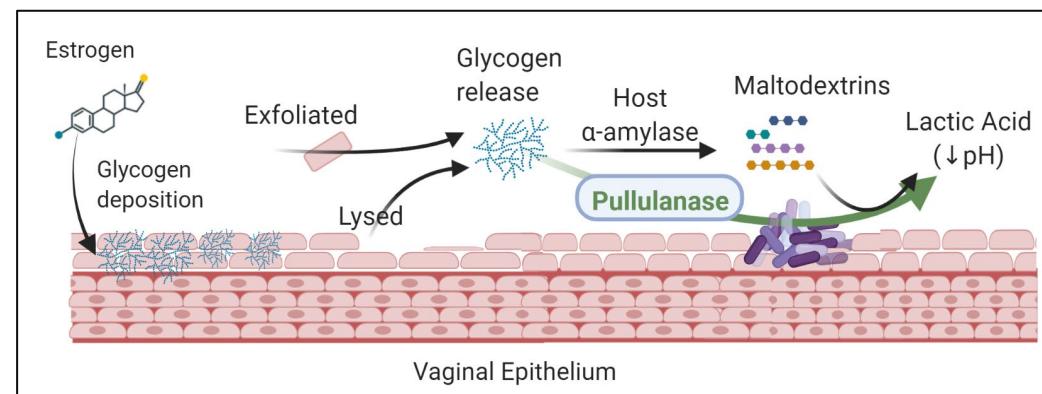
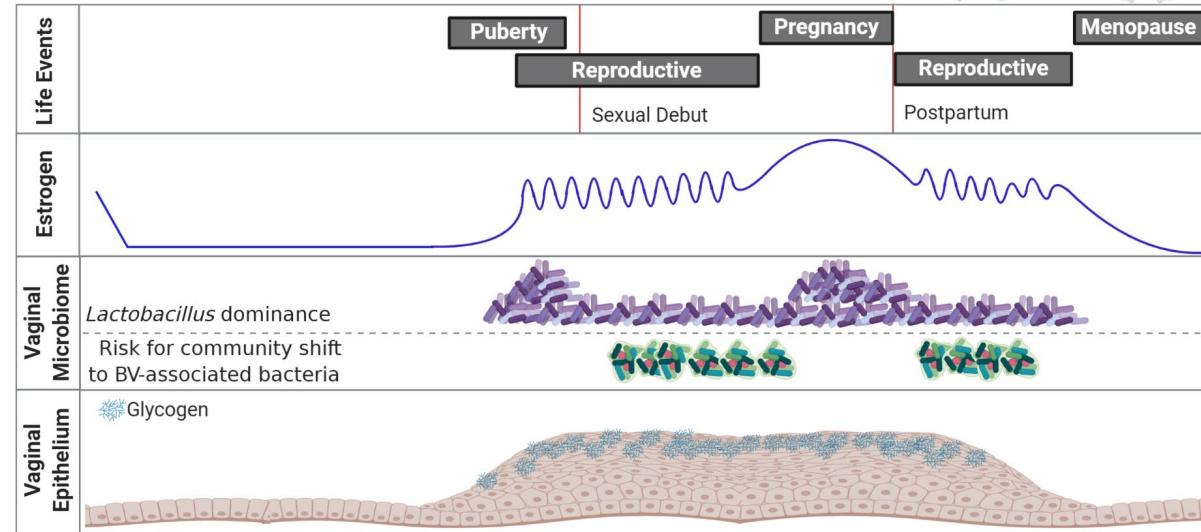
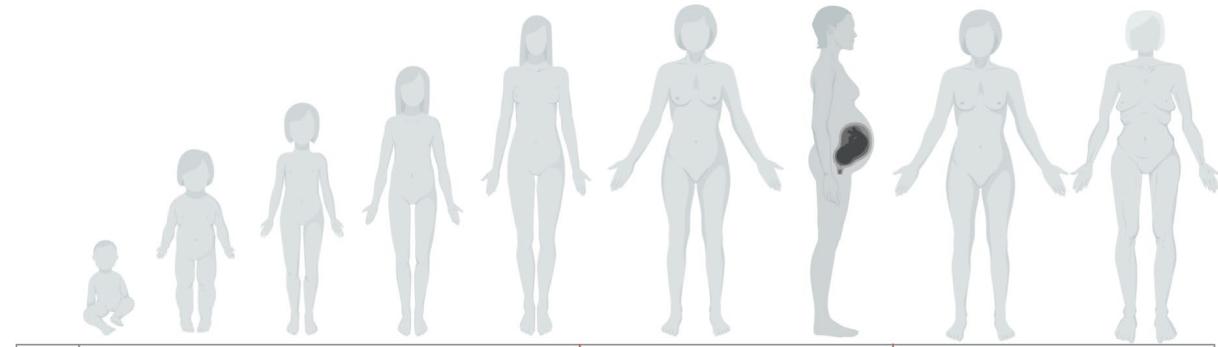


Konschuh S et al. Sexually Transmitted Infections 2021 97(Suppl 1):A1–A186

Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function



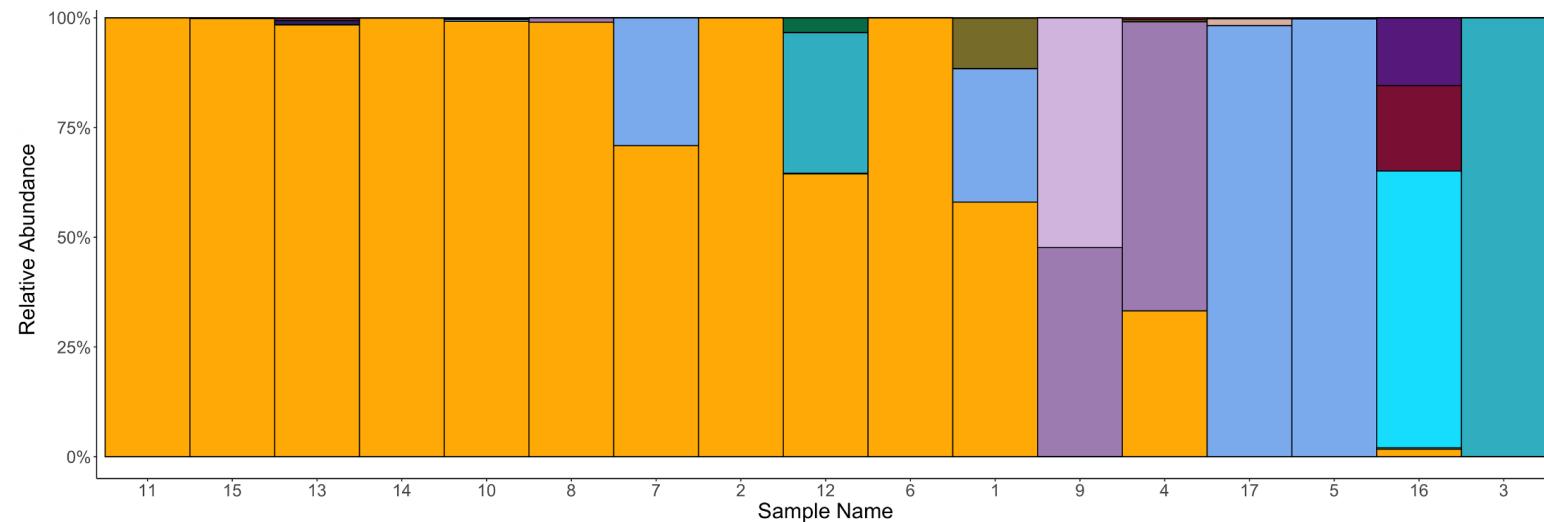
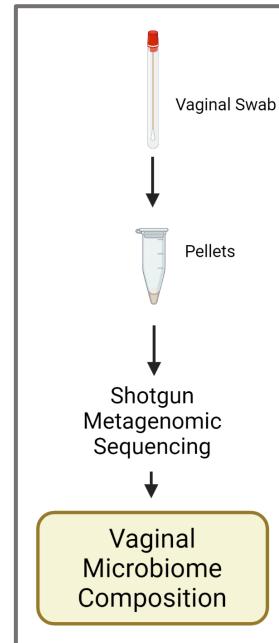
Created with biorender.com

van der Veer et al. *Microbiome*. 2019 7(49)

Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function



Lithgow KV, Cochinamogulos A, Muirhead K et al.
<https://www.biorxiv.org/content/10.1101/2022.03.29.486257v1>

Taxonomy

Lactobacillus crispatus	Gardnerella	Limosilactobacillus vaginalis	Limosilactobacillus reuteri	Ureaplasma parvum
Lactobacillus iners	Lactobacillus johnsonii	Streptococcus agalactiae	Lactobacillus gasseri	Limosilactobacillus oris
Bifidobacterium breve	Lactobacillus paragasseri	Lactobacillus jensenii	Limosilactobacillus coleohominis	

Annotating MAGs:

What you get:

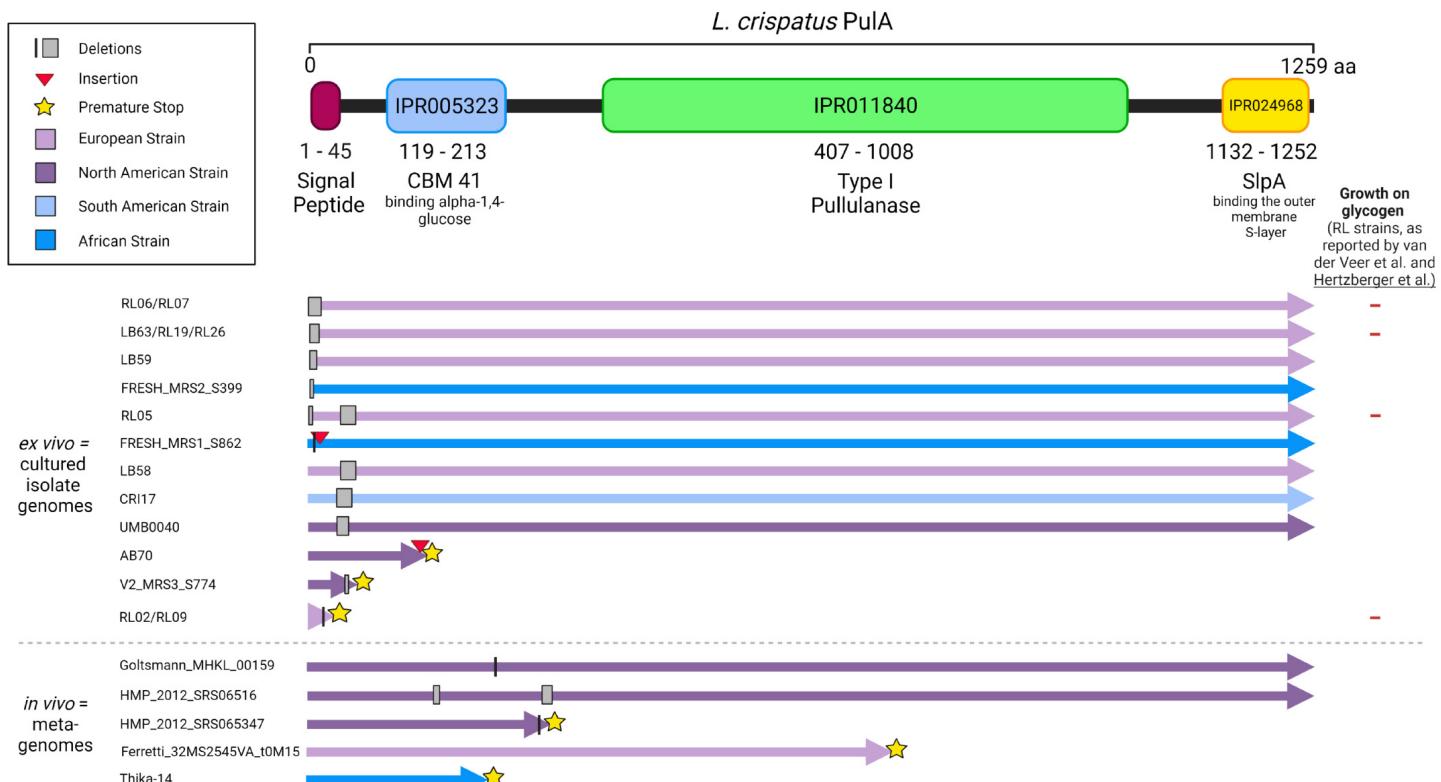
- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function

High rates of functional inactivation (25-30%)

A thousand ways to loose your pullulanse?

Lithgow KV, Cochinamogulos A, Muirhead K et al.
<https://www.biorxiv.org/content/10.1101/2022.03.29.486257v1>

	<i>In Vivo</i> (Metagenomes) N=40 n (%)	<i>Ex Vivo</i> (Isolates) N=123 n (%)	Total <i>In Vivo</i> and <i>Ex Vivo</i>
			N=163 n (%)
Present/Full-Length Gene	30 (75.0)	87 (70.7)	117 (71.8)
Absent/Mutant Gene	10 (25.0)	36 (29.3)	46 (28.2)
Absent	6 (15.5)	19 (15.4)	25 (15.3)
N-terminal Deletion	0 (0)	8 (6.5)	8 (4.9)
Internal Deletion/Frameshift	2 (5.0)	9 (7.3)	11 (6.7)
Nonsense Mutation	2 (5.0)	0 (0)	2 (1.2)

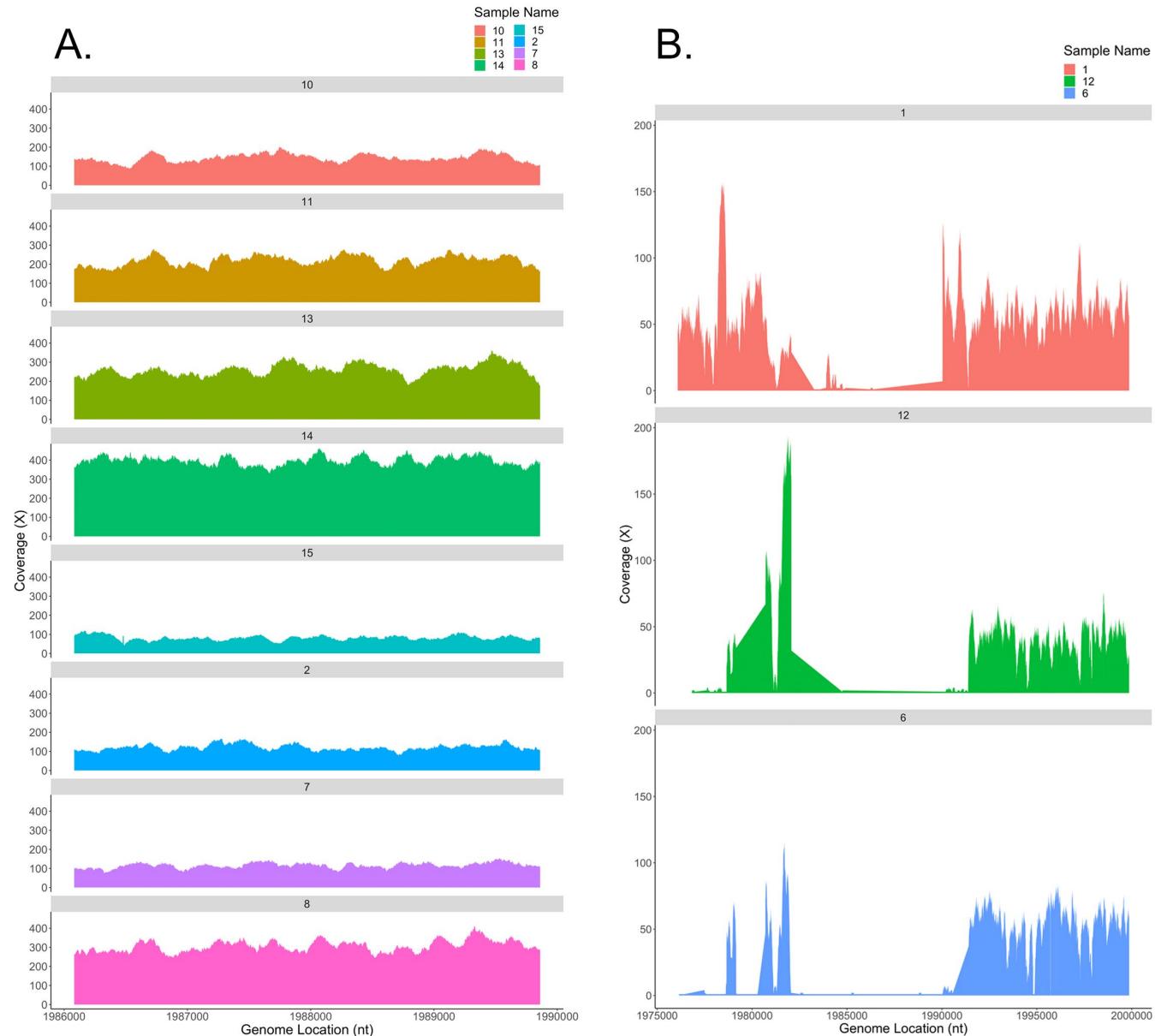


Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function

Gene loss confirmed by mapping reads to closed reference genome



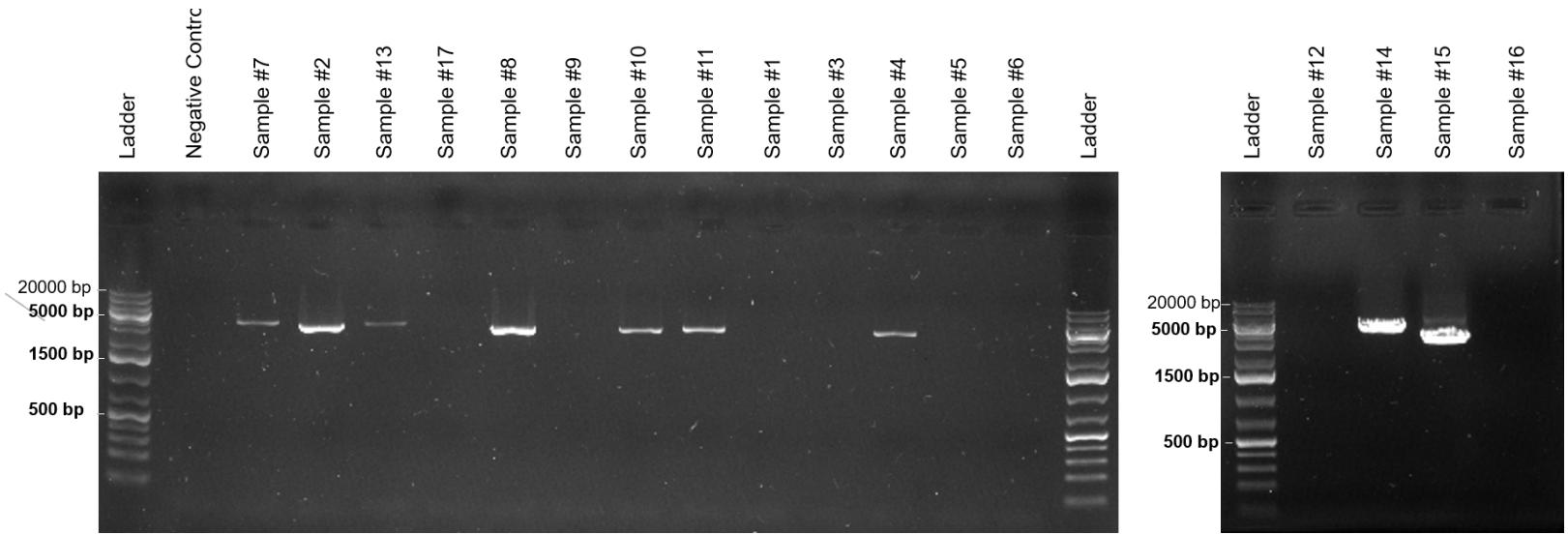
Lithgow KV, Cochinamogulos A, Muirhead K et al.

<https://www.biorxiv.org/content/10.1101/2022.03.29.486257v1>

Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function



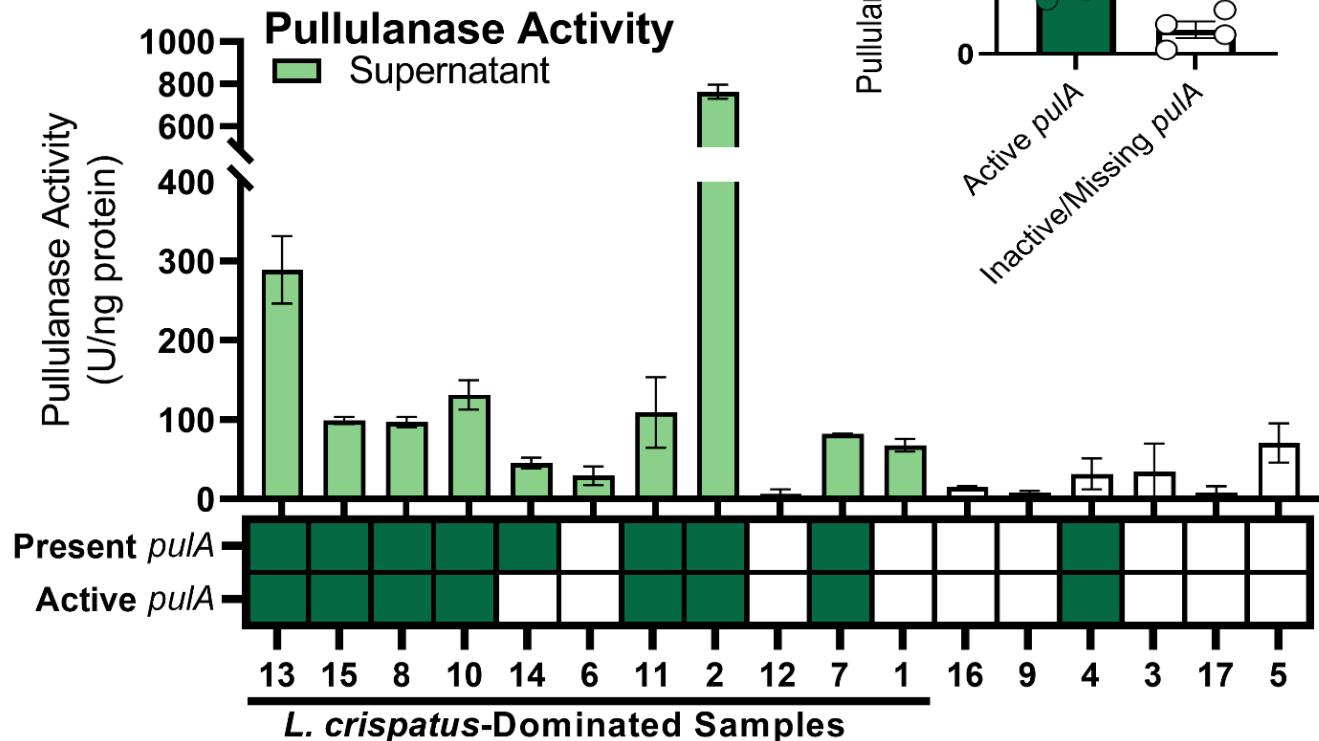
**Gene loss confirmed
by PCR**
*(no amplification from
subdominant strain)*

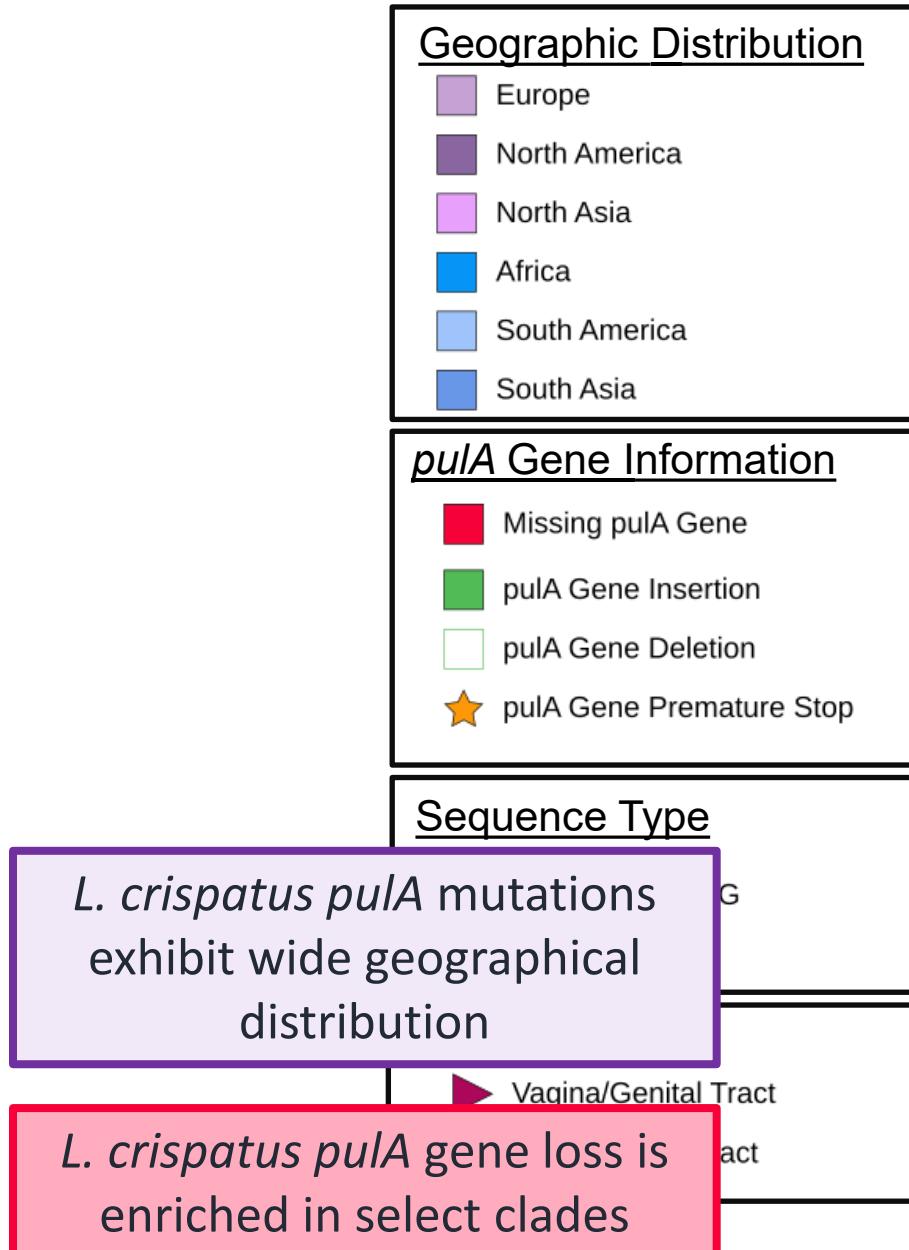
Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function

Pullulanase activity is significantly higher in *L. crispatus*-dominated samples if functional *pulA* gene is present





cpn60 phylogenetic tree

Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function
- Adaptation & exchange

Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands

Finlay Maguire^{1†}, Baofeng Jia^{2†}, Kristen L. Gray², Wing Yin Venus Lau², Robert G. Beiko¹ and Fiona S. L. Brinkman^{2,*}



ARTICLE | VOLUME 185, ISSUE 26, P4921-4936.E15, DECEMBER 22, 2022

Tommi Vatanen ¹⁸ • Karolina S. Jabbar ¹⁸ • Terhi Ruohutula • ... Hera Vlamakis • Mikael Knip • Ramnik J. Xavier   • Show all authors • Show footnotes

Article

Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism

nature communications



Article

<https://doi.org/10.1038/s41467-023-36633-7>

Population-level impacts of antibiotic usage on the human gut microbiome

Received: 16 November 2021

Accepted: 6 February 2023

Kihyun Lee  ^{1,2}, Sébastien Raguideau³, Kimmo Sirén⁴, Francesco Asnicar  ⁵, Fabio Cumbo  ⁵, Falk Hildebrand  ^{3,6}, Nicola Segata  ⁵, Chang-Jun Cha  ^{1,8}  & Christopher Quince  ^{3,6,7,8} 

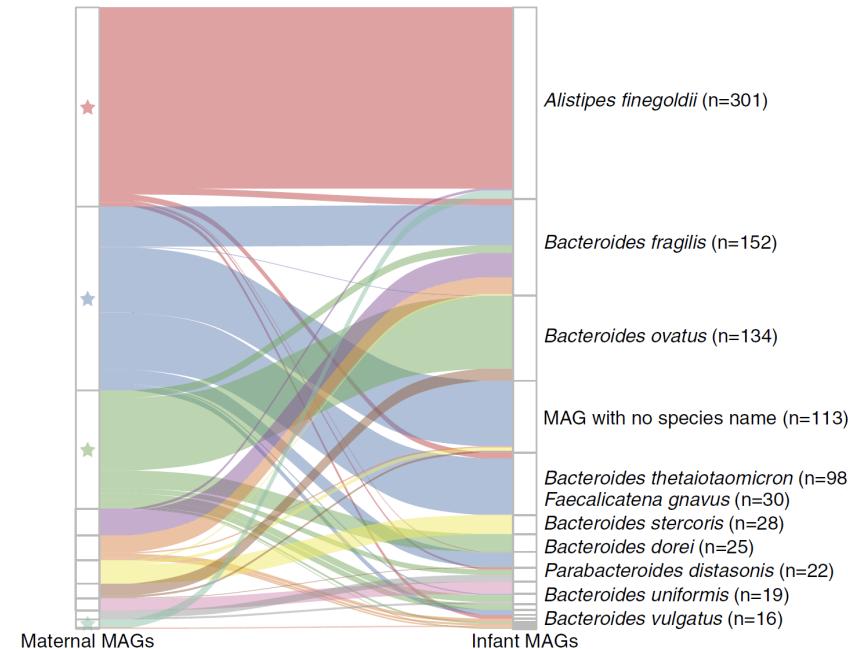
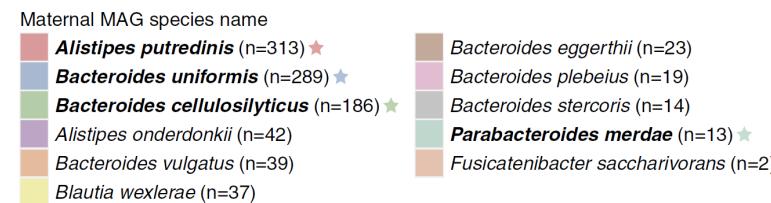
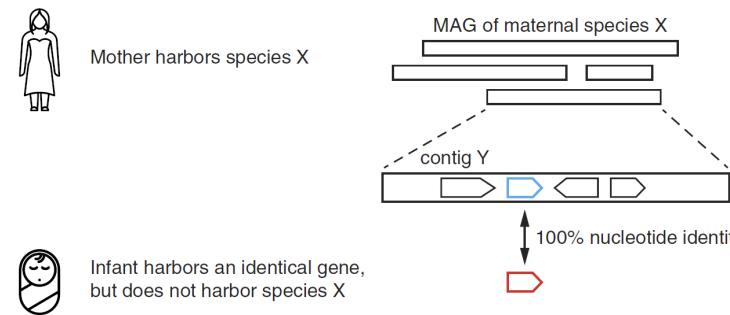
Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function
- Adaptation & exchange

Cell

Article
Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism



Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function
- **Adaptation & exchange**

nature communications

Article

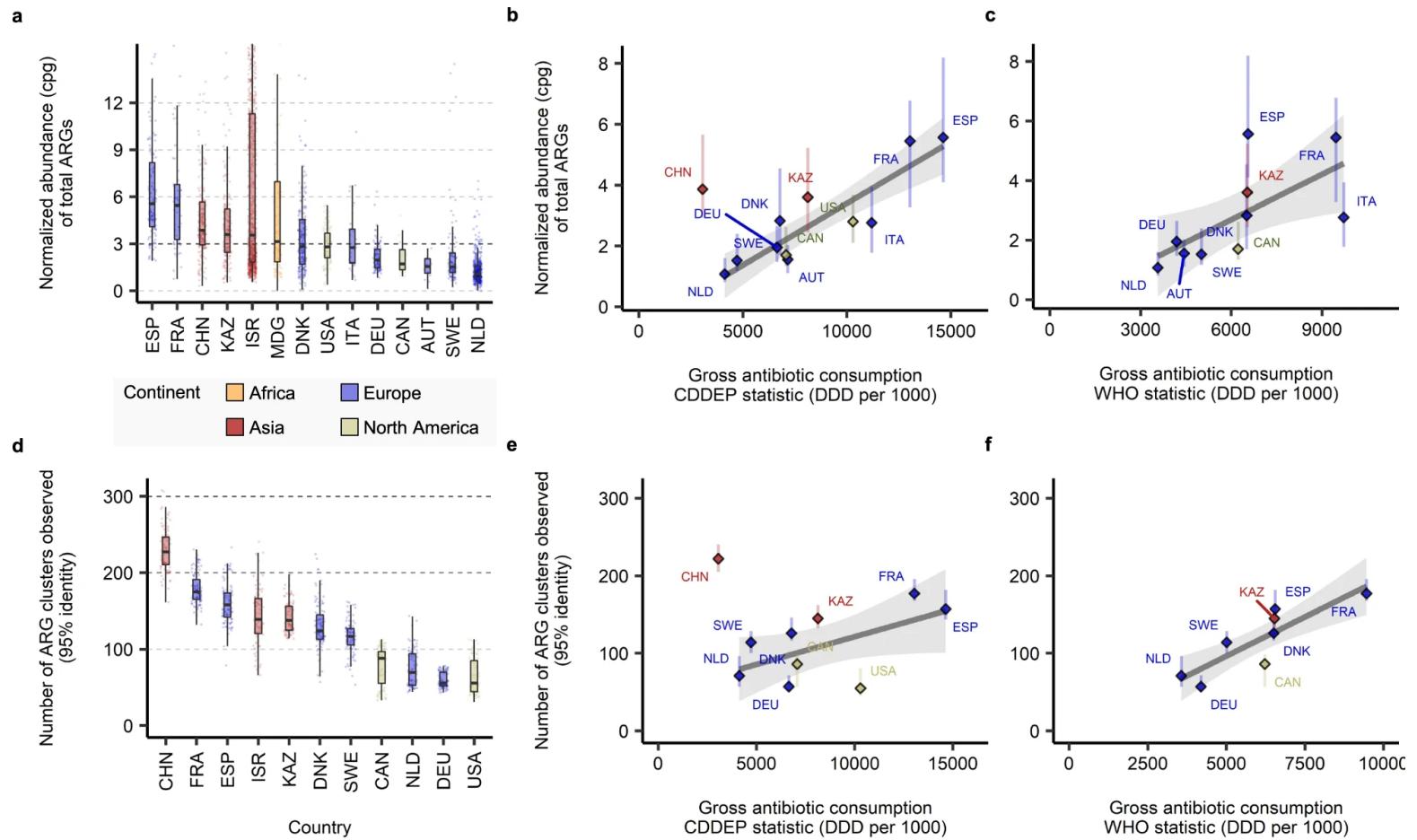
<https://doi.org/10.1038/s41467-023-36633-7>
Population-level impacts of antibiotic usage on the human gut microbiome

Received: 16 November 2021

Kihyun Lee  ^{1,2}, Sébastien Raguideau  ³, Kimmo Sirén  ⁴, Francesco Asnicar  ⁵, Fabio Cubero  ⁶, Falk Hildebrand  ^{6,7}, Nicola Segata  ⁸, Chang-Jun Cha  ^{1,9} & Christopher Quince  ^{3,6,7,8}

Accepted: 6 February 2023

Published online: 02 March 2023

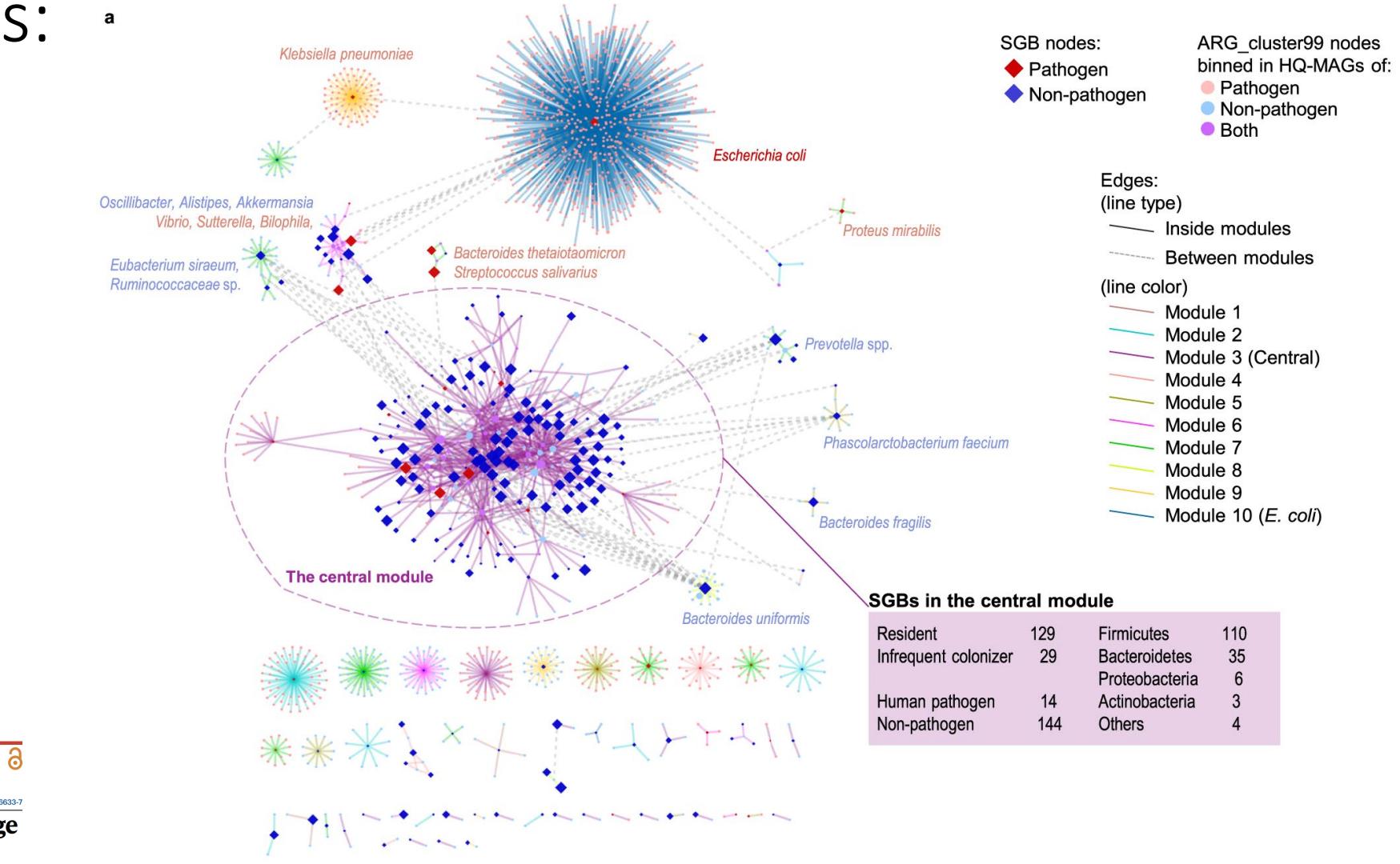


Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function
- Adaptation & exchange

nature communications



Article

<https://doi.org/10.1038/s41467-023-36633-7>

Population-level impacts of antibiotic usage on the human gut microbiome

Received: 16 November 2021

Kihyun Lee ^{1,2}, Sébastien Raguideau ³, Kimmo Sirén ⁴, Francesco Asnicar ⁵,

Accepted: 6 February 2023

Fabio Cumbo ⁶, Falk Hildebrand ^{6,7}, Nicola Segata ⁸, Chang-Jun Cha ^{1,9,10} &

Christopher Quince ^{3,6,7,8}

Published online: 02 March 2023

Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function
- Adaptation & exchange

nature communications

Article

<https://doi.org/10.1038/s41467-023-36633-7>
Population-level impacts of antibiotic usage on the human gut microbiome

Received: 16 November 2021

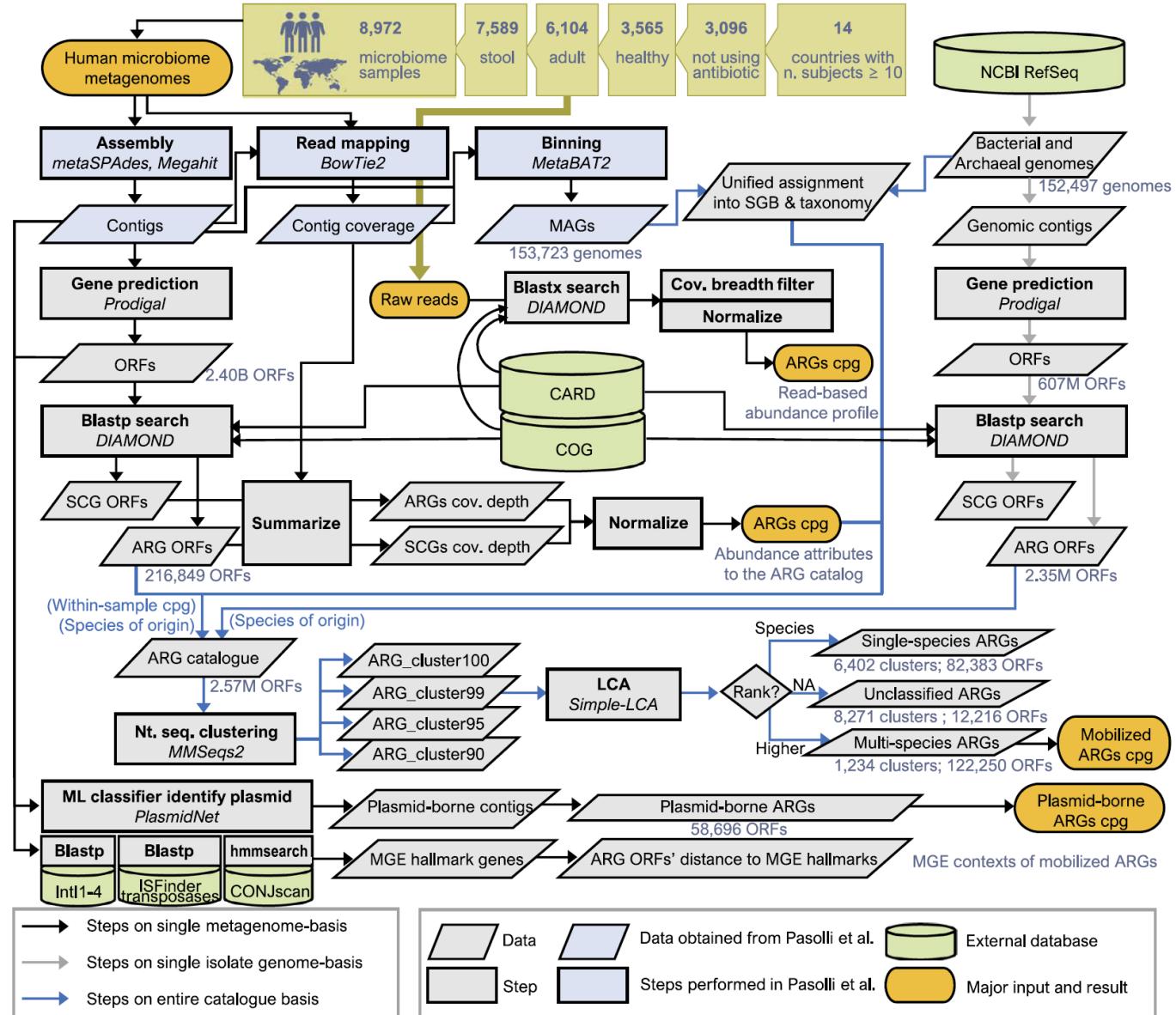
Kihyun Lee ^{1,2}, Sébastien Raguéreau ³, Kimmo Sirén ⁴, Francesco Asnicar ⁵

Accepted: 6 February 2023

Fabio Cubo ⁶, Falk Hildebrand ^{3,6}, Nicola Segata ³, Chang-Jun Cha ^{1,6} &

Christopher Quince ^{3,6,7,8}

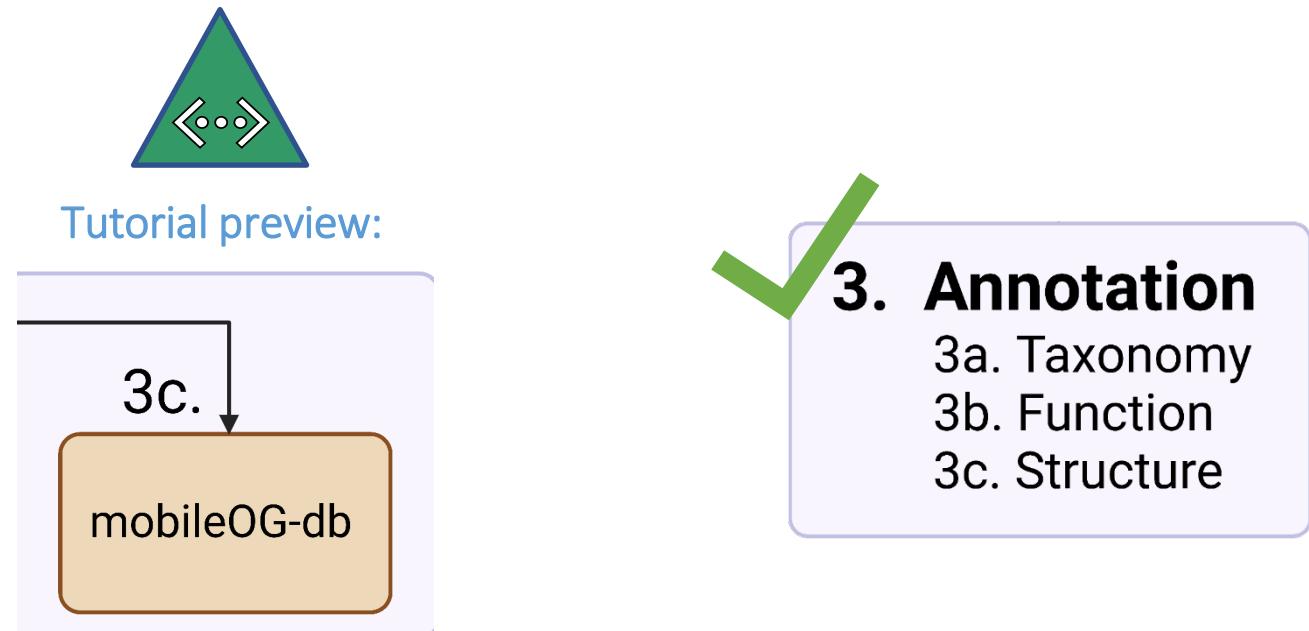
Published online: 02 March 2023



Annotating MAGs:

What you get:

- Genome-resolved understanding of novel clades
- Allele-resolved understanding of function
- Adaptation & exchange



Brown et al. Applied & Environmental Microbiology. 2022 88(18)

My Analysis Carts**: 361 Genomes | 0 Scaffolds | 2 Functions | 7 Genes | 0 Genome Search History | 2 Gene Search History | [?](#)

Home IMG/MER Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Help

NERSC, JGI and IMG will be down during a planned outage, starting from Feb 21, 2020 to Feb 25, 2020

IMG Content

Datasets	JGI	All
Bacteria	19518	92989
Archaea	876	2946
Eukarya	371	751
Plasmids	3	1208
Viruses	1216	9805
Genome Fragments	0	89
Metagenome	12730	23650
Cell Enrichments	2229	2279
Single Particle Sorts	5085	5458
Metatranscriptome	3857	6077
Total Datasets	145315	
My Private Datasets	2	

Last Datasets Added On:

Genome 2020-01-27

Metagenome 2020-02-08

Metagenome Replacement Notice

[Project Map](#)
[Metagenome Projects Map](#)

System Requirements

[Microbial Genomics & Metagenomics Workshop](#)

Tweets by @IMG_DATA

Integrated Microbial Genomes & Microbiomes
Retweeted

JGI Joint Genome Inst.

Our next @jgi Microbial #Genomics & #Metagenomics (MGM) Workshop takes place March 30-April 3 at @BerkeleyLab. Details and registration link at mmg.jgi.doe.gov/ #microbiology #viruses #informatics #bigdata

The **Integrated Microbial Genomes (IMG)** system serves as a community resource for analysis and annotation of genome and metagenome datasets in a comprehensive comparative context. The **IMG data warehouse** integrates genome and metagenome datasets provided by IMG users with a comprehensive set of publicly available genome and metagenome datasets.

IMG/MER provides users with tools ([IMG/MER UI Map](#)) for analyzing their private (password protected access) genome datasets and/or metagenome datasets in the context of all public (free access) genome and metagenome datasets in IMG. ([Nucleic Acids Research, January 2019](#))

If you use IMG web resources or data to assist in research publications or proposals please cite Chen et al., 2019 (PMID: [30289528](#)).

[IMG Statistics](#)

[Data Usage Policy](#)

View Citations



(Only data sets with GOLD metadata were counted.)

Combined assembly data sets were excluded and private data sets were included in the following metagenome and metatranscriptome table statistics.

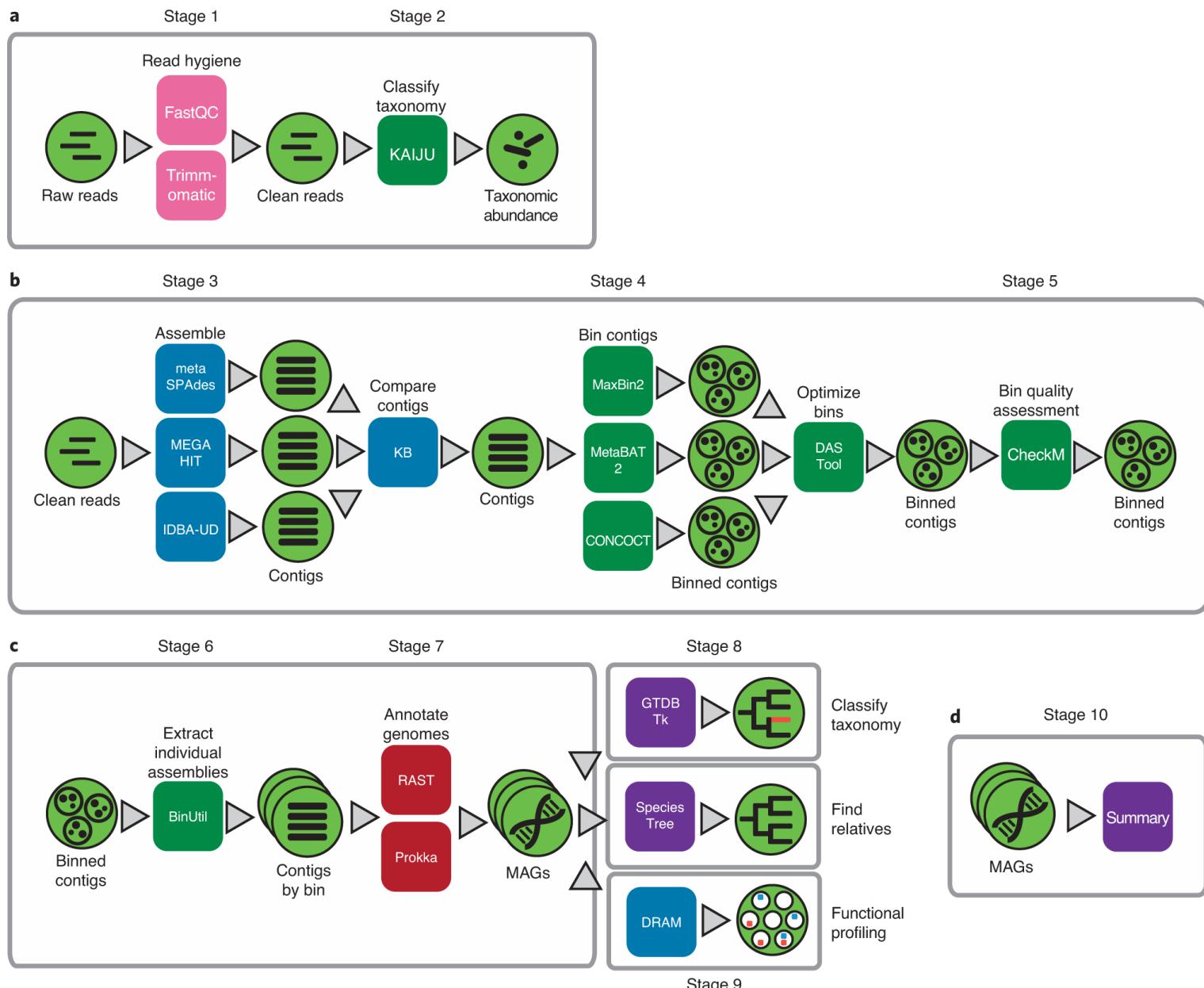
[Metagenome](#) [Metatranscriptome](#)

Engineered	JGI	ALL	Environmental	JGI	ALL	Host-associated	JGI	ALL
Bioreactor	282	406	Air	60	116	Algae	32	124
Bioremediation	177	215	Aquatic	11207	14531	Animal	0	1
Biotransformation	12	25	Terrestrial	5095	5947	Annelida	159	172
Built environment	269	1345	Unclassified	0	3	Arthropoda	123	229
Food production	0	25				Birds	17	48
Industrial production	0	12				Cephalochordata	0	4



Metagenome-assembled genome extraction and analysis from microbiomes using KBase

Dylan Chivian¹✉, Sean P. Jungbluth¹, Paramvir S. Dehal¹, Elisha M. Wood-Charlson¹, Richard S. Canon¹, Benjamin H. Allen², Mikayla M. Clark^{2,4}, Tianhao Gu³, Miriam L. Land¹, Gavin A. Price¹, William J. Riehl¹, Michael W. Sneddon^{1,5}, Roman Sutormin^{1,6}, Qizhi Zhang³, Robert W. Cottingham¹, Chris S. Henry³ and Adam P. Arkin¹✉



Videos illustrating how the sequencing technologies work

Illumina SBS: <https://www.youtube.com/watch?v=womKfikWlxM>

Oxford Nanopore: <https://nanoporetech.com/applications/dna-nanopore-sequencing>

PacBio HiFi Sequencing: https://www.youtube.com/watch?v=_ID8JyAbwEo&t=89s