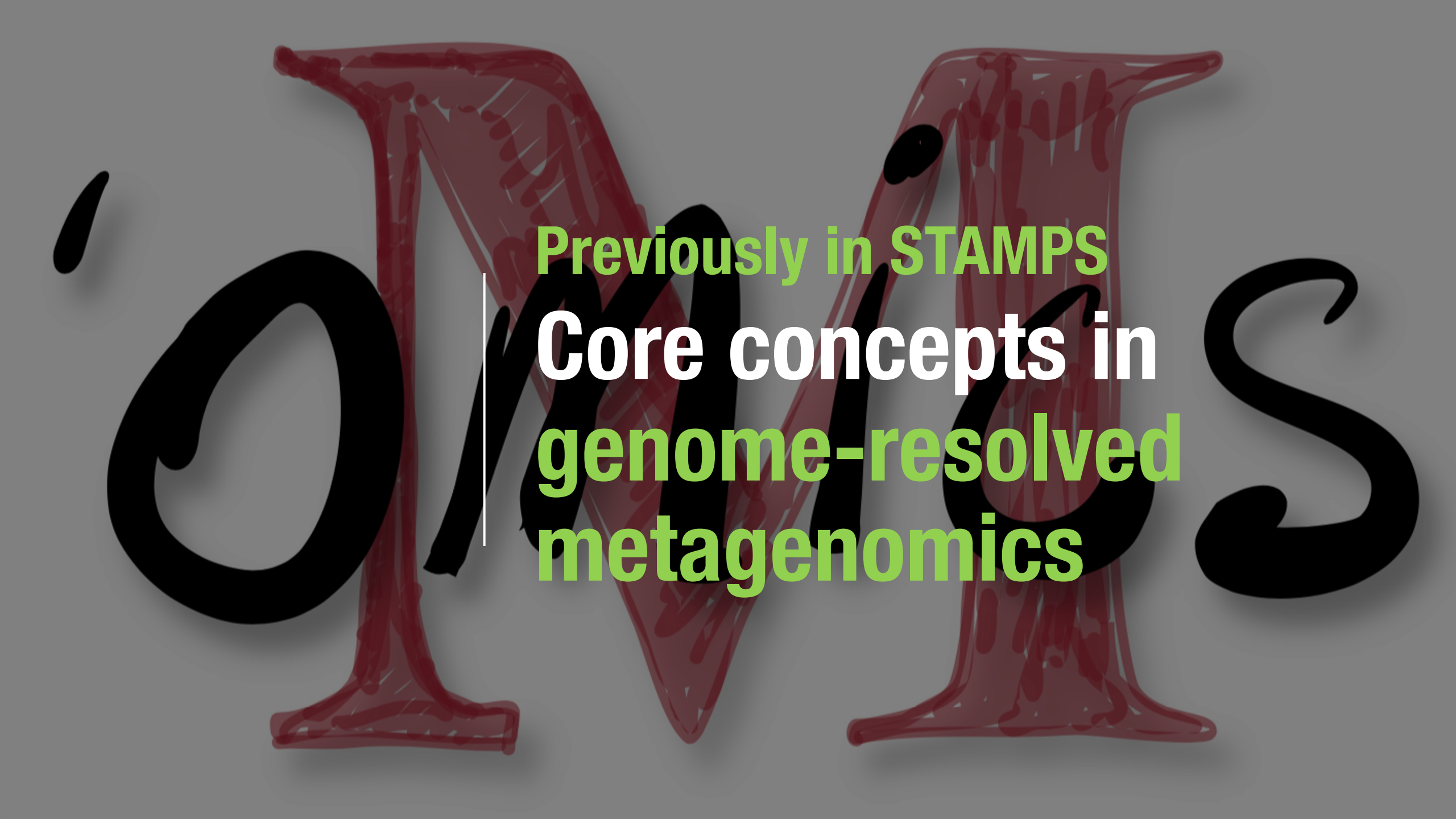
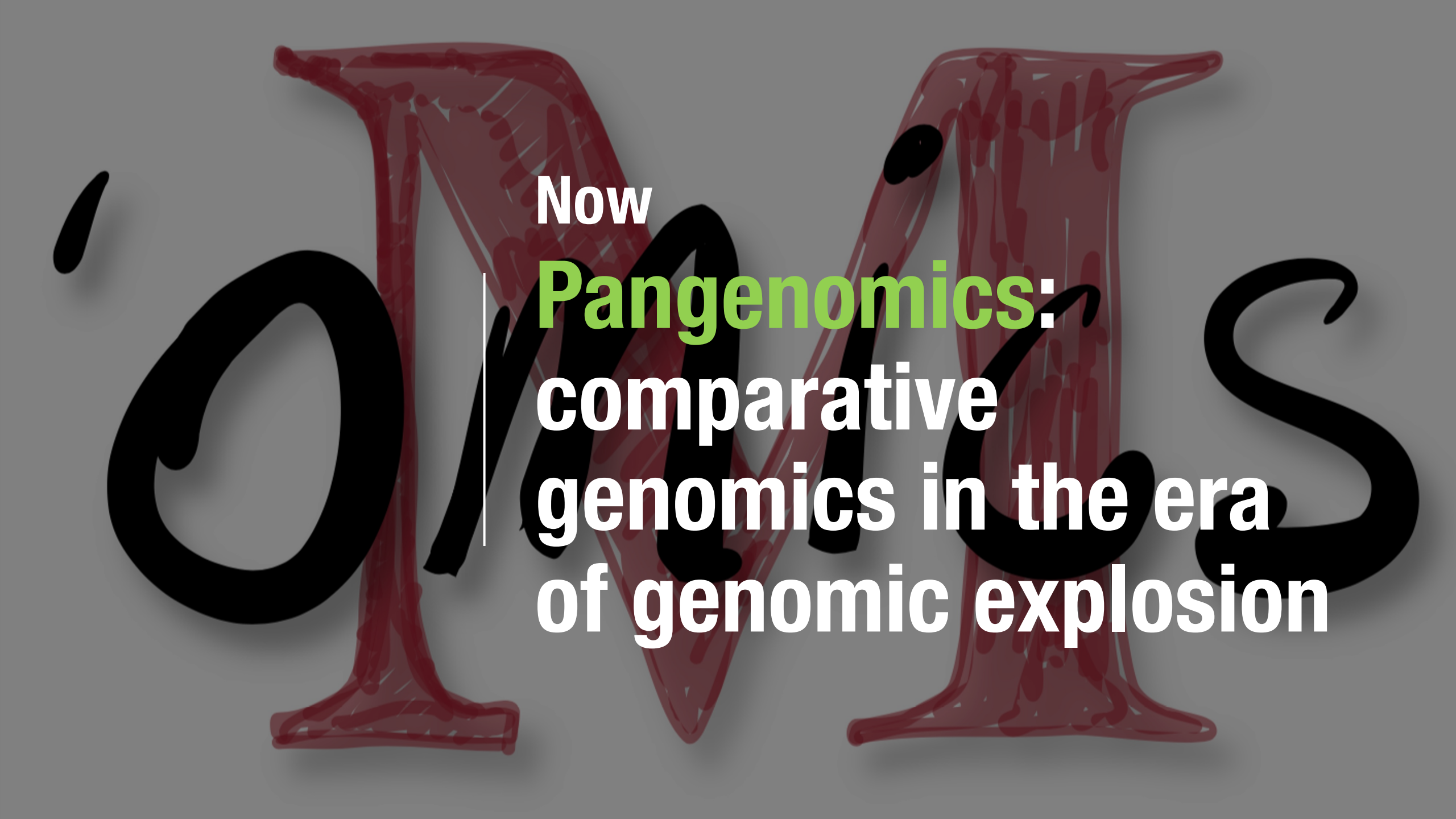
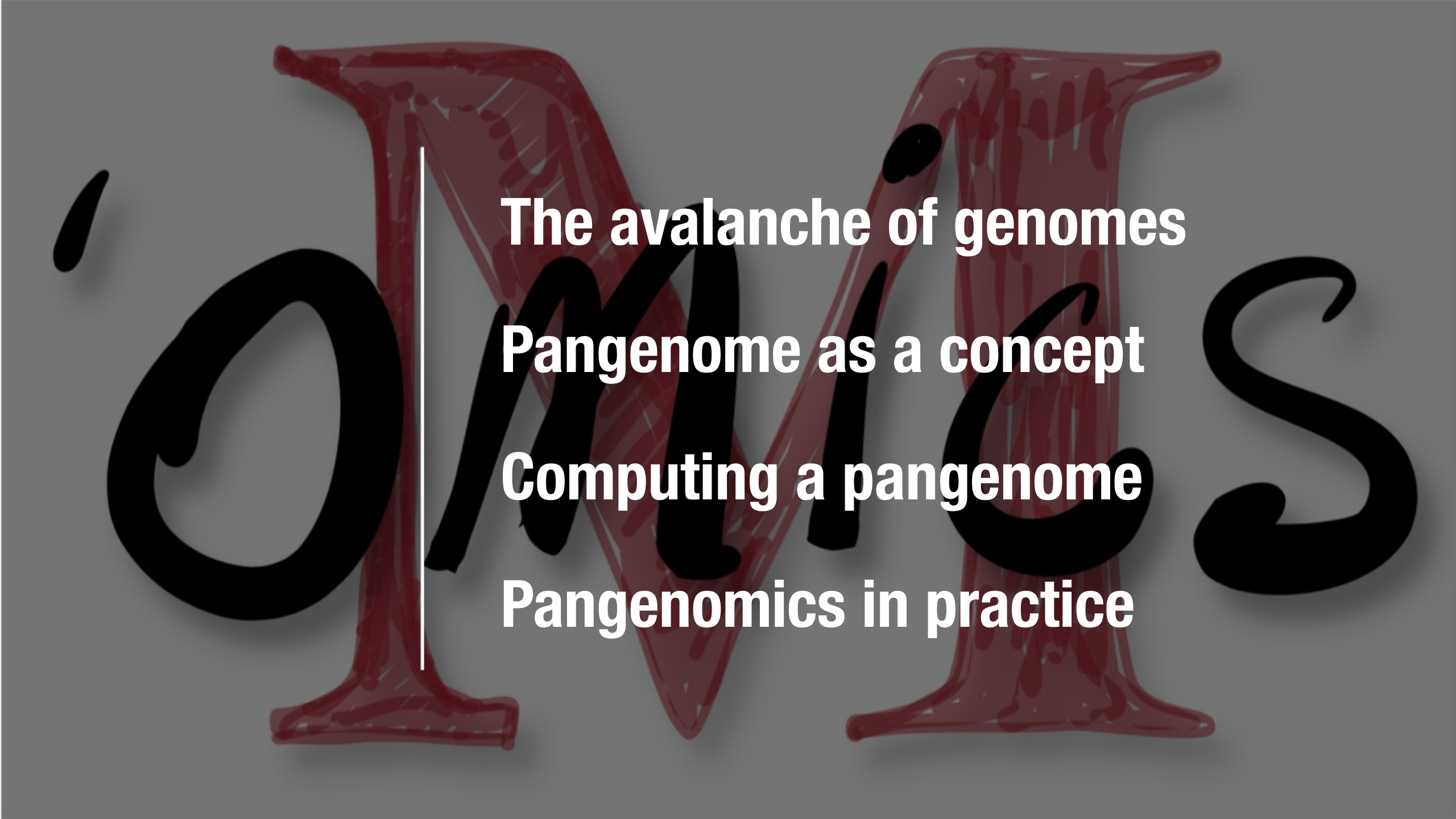
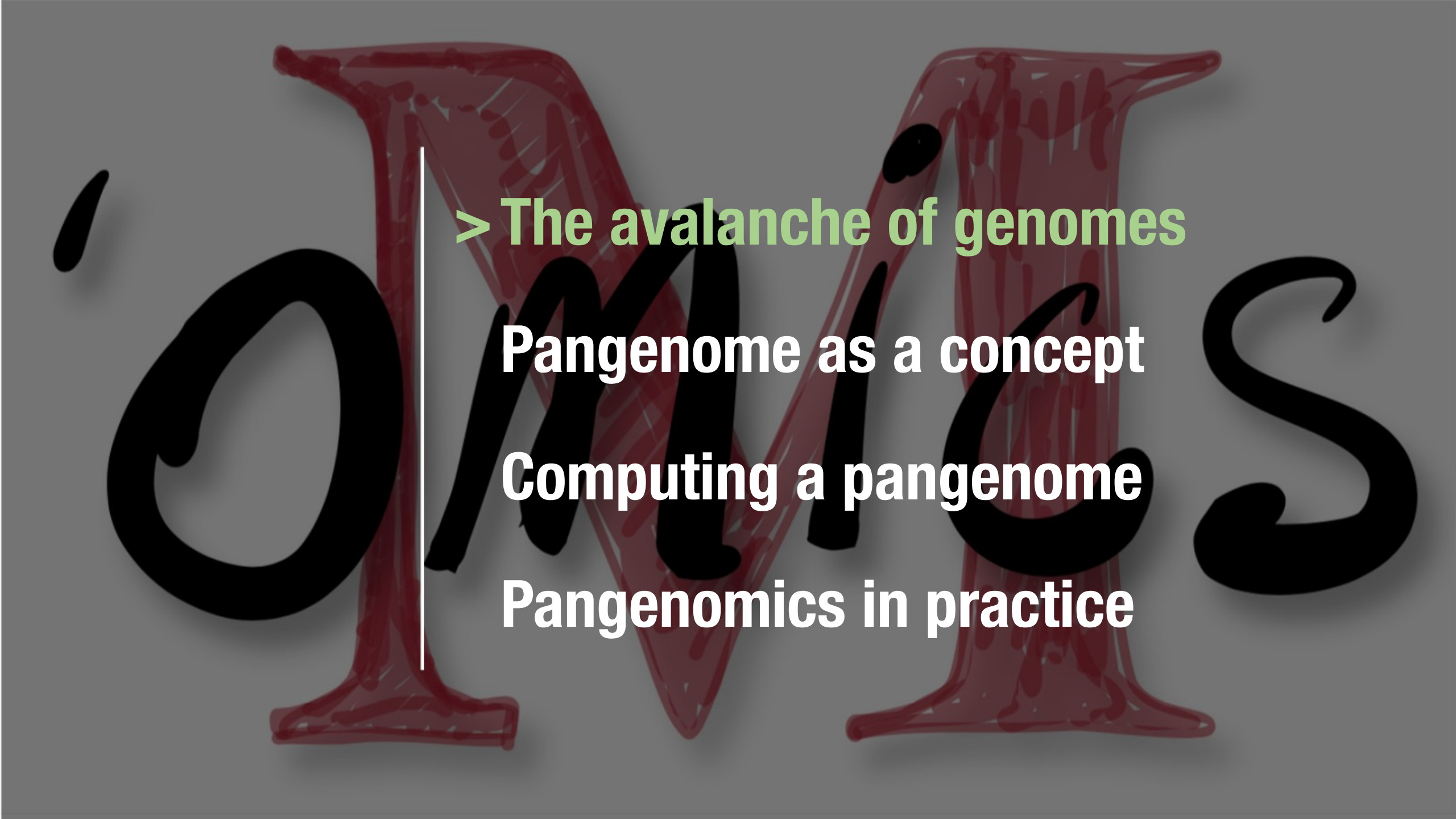
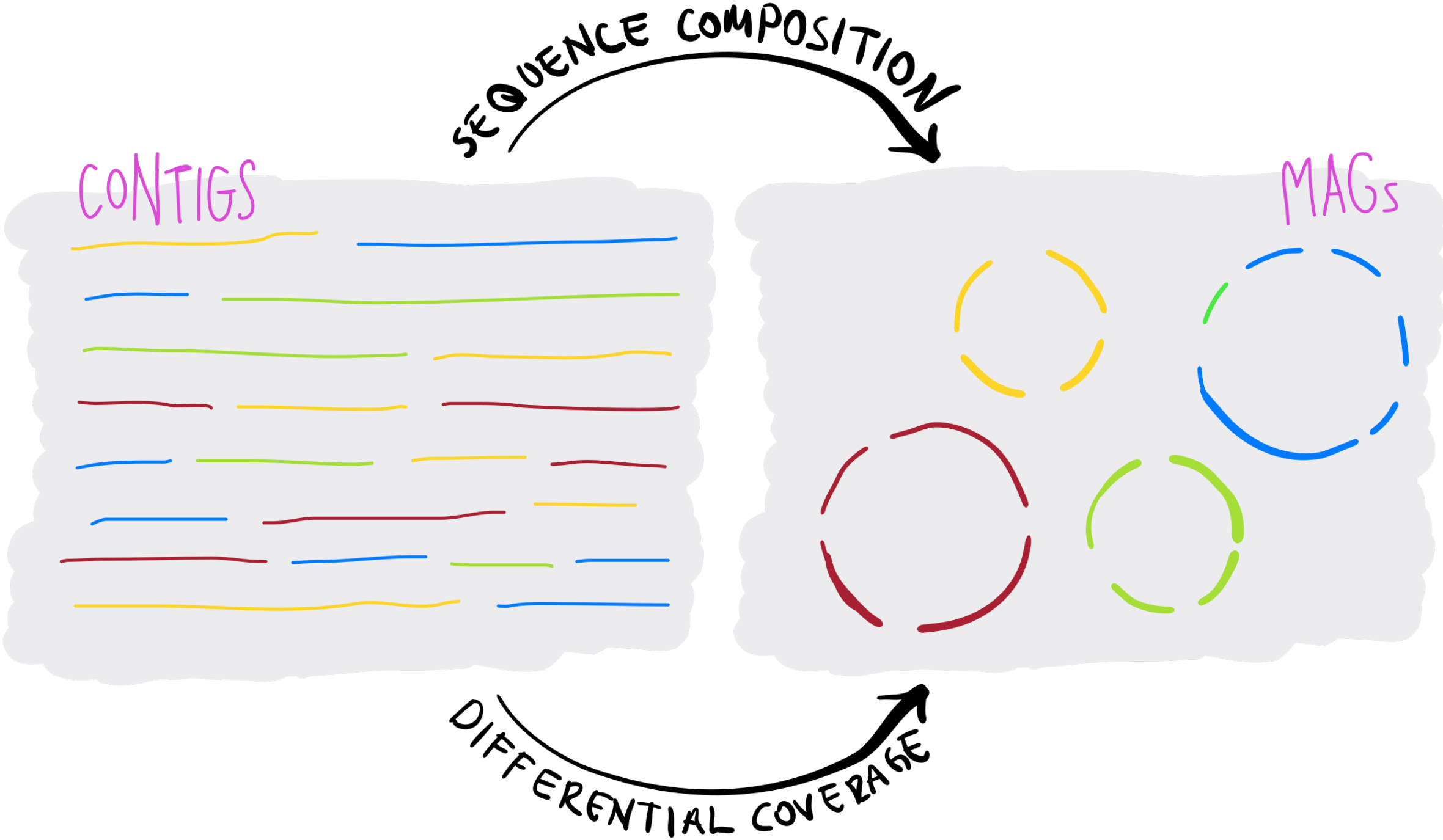

Previously in STAMPS
Core concepts in
genome-resolved
metagenomics


Now
Pangenomics:
comparative
genomics in the era
of genomic explosion


The avalanche of genomes
Pangenome as a concept
Computing a pangenome
Pangenomics in practice


> The avalanche of genomes
Pangenome as a concept
Computing a pangenome
Pangenomics in practice


SEQUENCE COMPOSITION
CONTIGS
MAGs
DIFFERENTIAL COVERAGE

Cell

Log in

RESOURCE | VOLUME 176, ISSUE 3, P649-662.E20, JANUARY 24, 2019

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 •

Highlights

- Large-scale metagenomic assembly uncovered thousands of new human microbiome species
- The new genome resource increases the mappability of gut metagenomes over 87%
- Some of the newly discovered species comprise thousands of reconstructed genomes

nature International journal of science

PDF

Article | OPEN | Published: 11 February 2019

A new genomic blueprint of the human gut microbiota

Alexandre Almeida, Alex L. Mitchell, [...] Robert D. Finn

Nature (2019) | Download Citation

Abstract

The composition of the human gut microbiota is linked to health and disease, but knowledge of individual microbial species is needed to decipher their biological roles. Despite extensive culturing and sequencing efforts, the complete bacterial repertoire of the human gut microbiota remains undefined. Here we identify 1,952 uncultured candidate bacterial species by reconstructing 92,143 metagenome-assembled genomes from 11,850 human gut microbiomes. These uncultured genomes substantially expand the known species repertoire of the

nature International journal of science

PDF

Article | OPEN | Published: 13 March 2019

New insights from uncultivated genomes of the global human gut microbiome

Stephen Nayfach, Zhou Jason Shi, [...] Nikos C. Kyrpides

Nature (2019) | Download Citation

Abstract

The genome sequences of many species of the human gut microbiome remain unknown, largely owing to challenges in cultivating microorganisms under laboratory conditions. Here we address this problem by reconstructing 60,664 draft prokaryotic genomes from 3,810 faecal metagenomes, from geographically and phenotypically diverse humans. These genomes provide reference points for 2,058 newly identified species-level operational taxonomic units (OTUs), which represents a 50% increase over the previously known

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

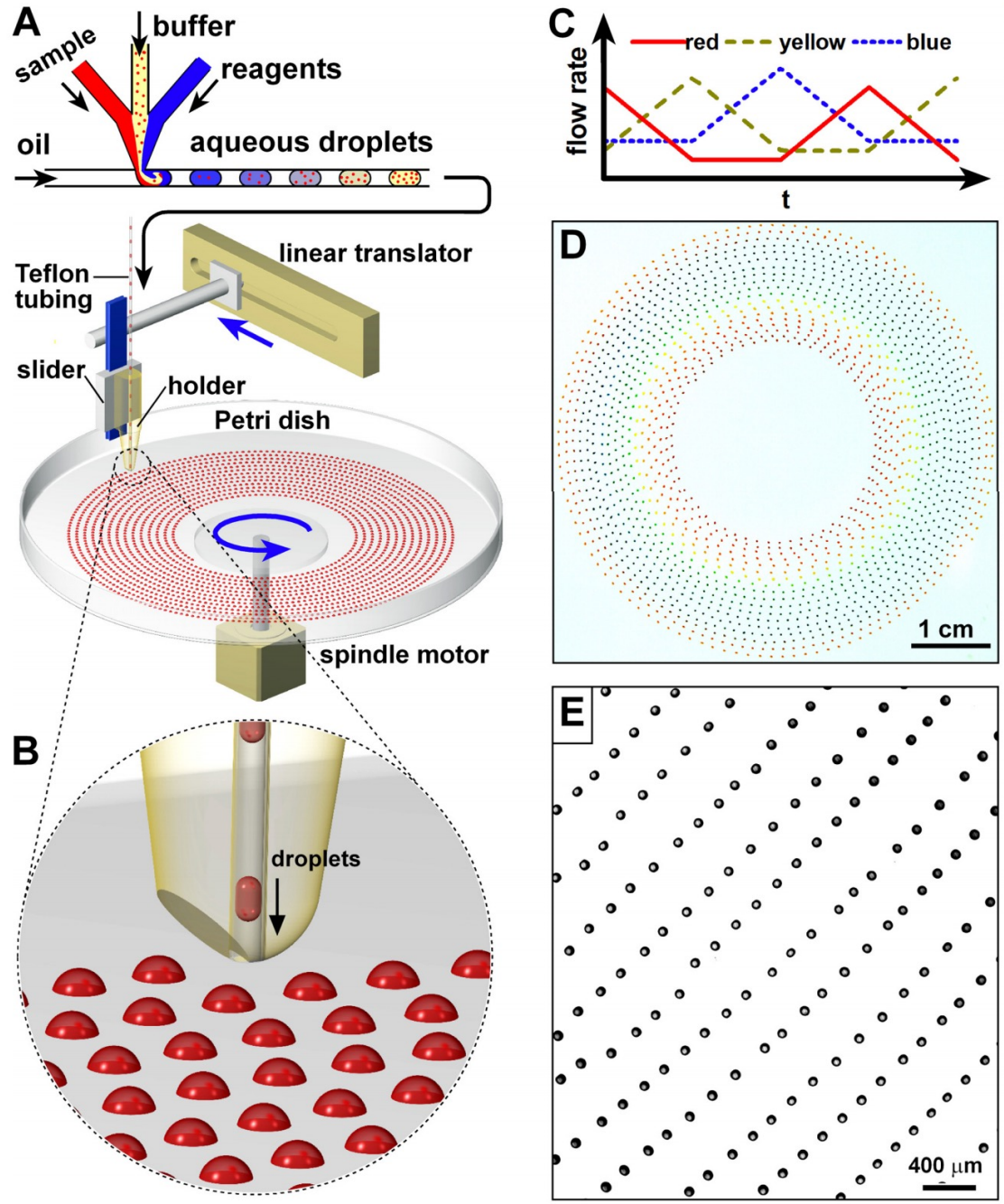
Methods | Spotlight

High-Throughput Single-Cell Cultivation on Microfluidic Streak Plates

Cheng-Ying Jiang, Libing Dong, Jian-Kang Zhao, Xiaofang Hu, Chaohua Shen, Yuxin Qiao, Xinyue Zhang, Yapei Wang, Rustem F. Ismagilov, Shuang-Jiang Liu, Wenbin Du

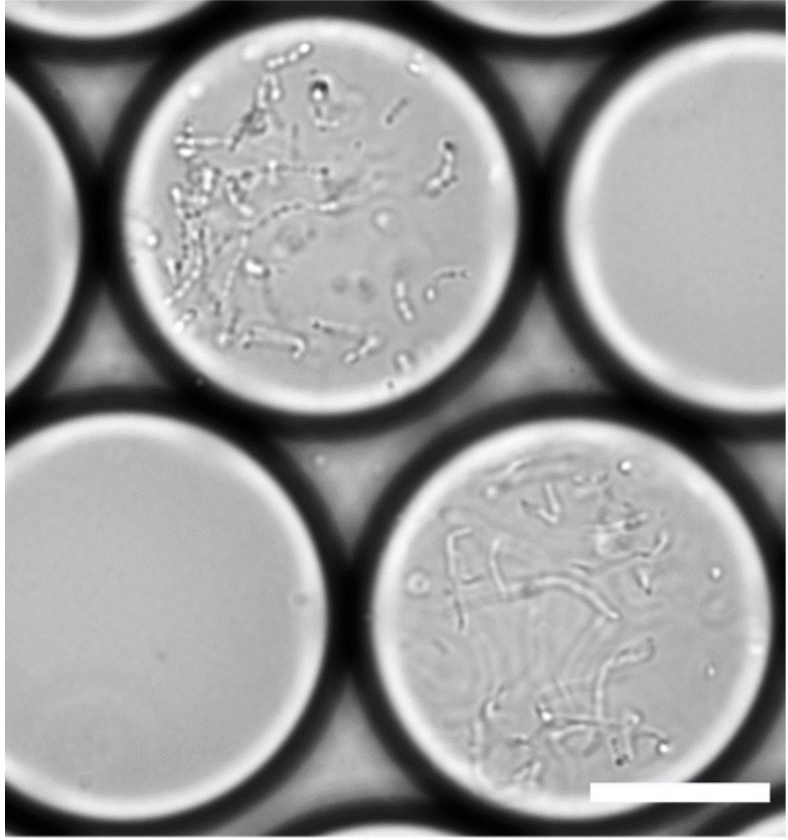
R. E. Parales, Editor

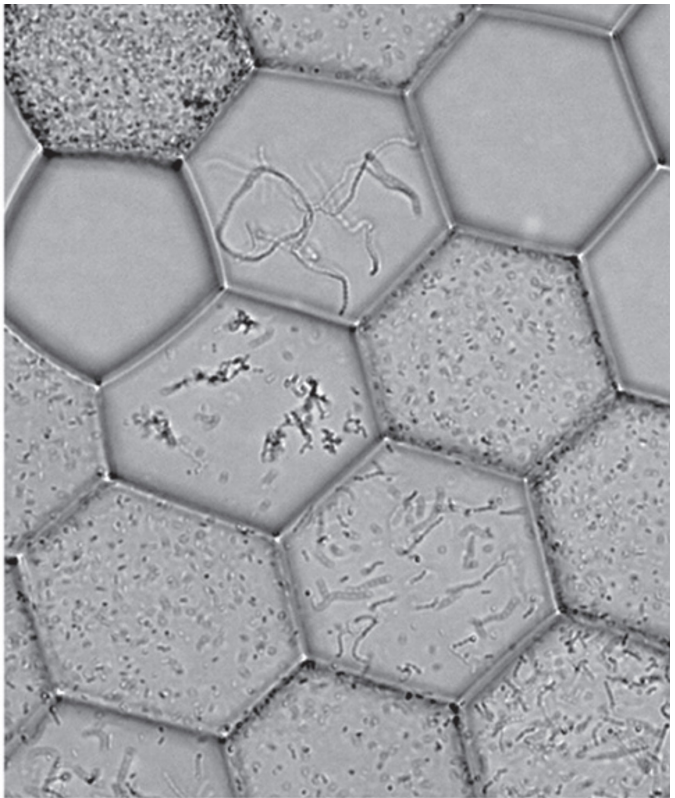
DOI: 10.1128/AEM.03588-15



Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes

William J Watterson[1,2*], Melikhan Tanyeri[1,2,3], Andrea R Watson[4], Candace M Cham[4], Yue Shan[4], Eugene B Chang[4], A Murat Eren[4,5,6*], Savaş Tay[1,2*]





Interindividual Variation in Dietary Carbohydrate Metabolism by Gut Bacteria Revealed with Droplet Microfluidic Culture

Max M. Villa,[a,b] Rachael J. Bloom,[b,c] Justin D. Silverman,[e,h] Heather K. Durand,[a,b] Sharon Jiang,[a,b] Anchi Wu,[f] Eric P. Dallow,[a,b] Shuqiang Huang,[g] Lingchong You,[b,f] Lawrence A. David[a,b,c,d,f]

The trajectory of microbial single-cell sequencing

Tanja Woyke, Devin F R Doud & Frederik Schulz

Over the past decade, it has become nearly routine to sequence genomes of individual microbial cells directly isolated from environmental samples ranging from deep-sea hydrothermal vents to insect guts, providing a powerful complement to shotgun metagenomics in microbial community studies. In this review, we address the technical aspects and challenges of single-cell genome sequencing and discuss some of the scientific endeavors that it has enabled. Specifically, we highlight newly added leaves and branches in the genomic tree of bacterial and archaeal life and illustrate the unique and exciting advantages that single-cell genomics offers over metagenomics, both now and in the near future.

Charting the Complexity of the Marine Microbiome through Single-Cell Genomics

Maria G. Pachiadaki,[1,2] Julia M. Brown,[1] Joseph Brown,[1] Oliver Bezuidt,[1] Paul M. Berube,[3] Steven J. Biller,[3,6] Nicole J. Poulton,[1] Michael D. Burkart,[4] James J. La Clair,[4] Sallie W. Chisholm,[3,5] and Ramunas Stepanauskas[1,7,*]

Summary

Marine bacteria and archaea play key roles in global biogeochemistry. To improve our understanding of this complex microbiome, we employed single-cell genomics and a randomized, hypothesis-agnostic cell selection strategy to recover 12,715 partial genomes from the tropical and subtropical euphotic ocean. A substantial fraction of known prokaryoplankton coding potential was recovered from a single, 0.4 mL ocean sample, which indicates that genomic information disperses effectively across the globe. Yet, we found each genome to be unique, implying limited clonality within prokaryoplankton populations. Light harvesting and secondary metabolite biosynthetic pathways were numerous across lineages, highlighting the value of single-cell genomics to advance the identification of ecological roles and biotechnology potential of uncultured microbial groups. This genome collection enabled functional annotation and genus-level taxonomic assignments for >80% of individual metagenome reads from the tropical and subtropical surface ocean, thus offering a model to improve reference genome databases for complex microbiomes.

Complete, closed bacterial genomes from microbiomes using nanopore sequencing

Eli L. Moss[1,3], Dylan G. Maghini[1,3] and Ami S. Bhatt[1,2] ✉

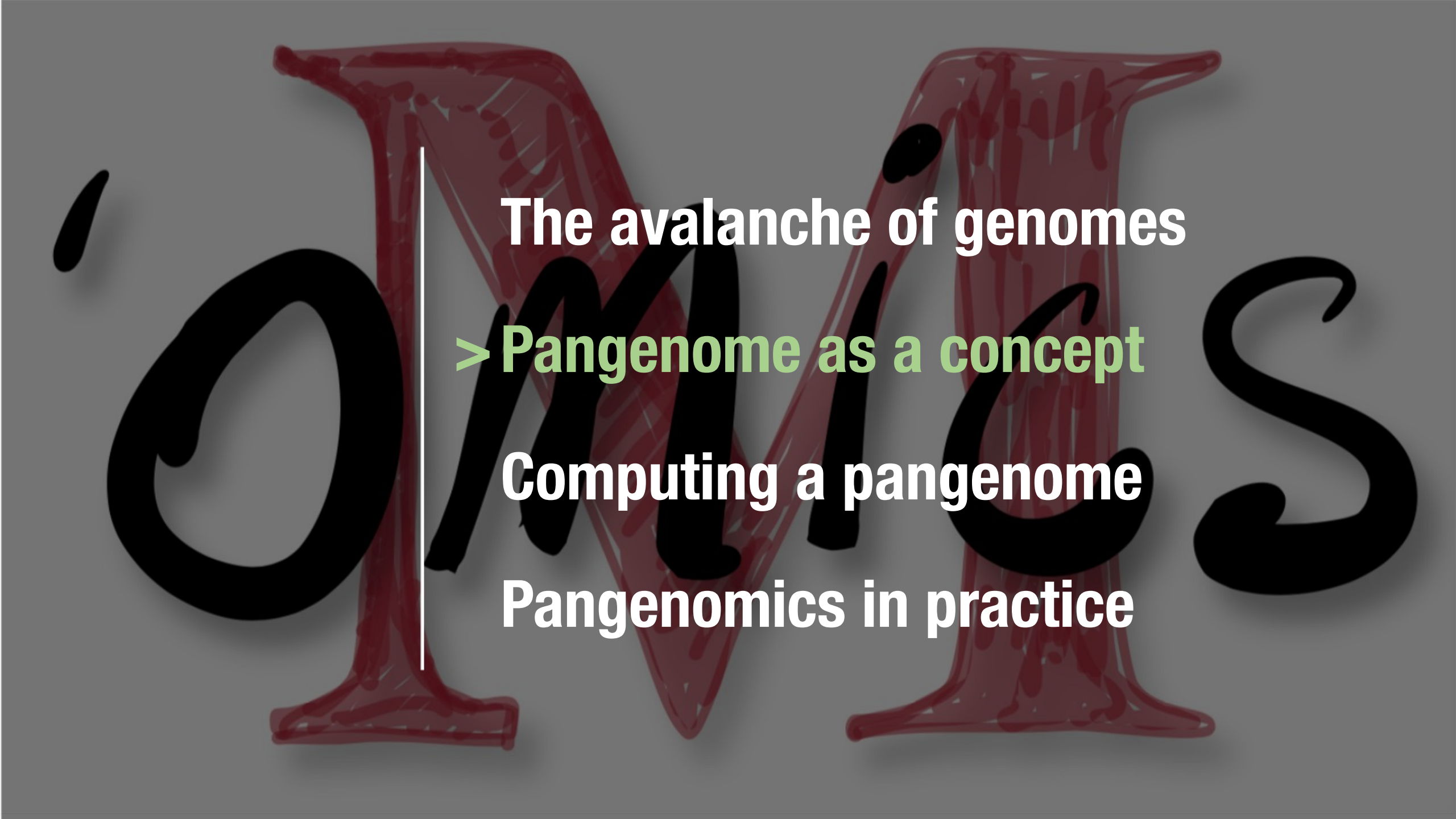
Microbial genomes can be assembled from short-read sequencing data, but the assembly contiguity of these metagenome-assembled genomes is constrained by repeat elements. Correct assignment of genomic positions of repeats is crucial for understanding the effect of genome structure on genome function. We applied nanopore sequencing and our workflow, named Lathe, which incorporates long-read assembly and short-read error correction, to assemble closed bacterial genomes from complex microbiomes. We validated our approach with a synthetic mixture of 12 bacterial species. Seven genomes were completely assembled into single contigs and three genomes were assembled into four or fewer contigs. Next, we used our methods to analyze metagenomics data from 13 human stool samples. We assembled 20 circular genomes, including genomes of *Prevotella copri* and a

elements. Gentle bead beating can reduce shearing, but might fail to extract DNA from organisms that are difficult to lyse. Thus, there is a need for methods to extract long fragments of DNA that can span repetitive elements from both Gram-positive and Gram-negative bacteria to overcome limitations in genome assembly[6].

We present a workflow for nanopore sequencing of stool samples, including protocols for DNA extraction and genome assembly (Supplementary Fig. 1). Our DNA extraction protocol is adapted from extraction methods for cultured bacteria[9], and comprises enzymatic degradation of the cell wall with a cocktail of lytic enzymes, then phenol-chloroform extraction, followed by RNAse A and Proteinase K digestion, gravity column purification and SPRI size selection. This approach produces microgram quantities of pure, HMW DNA suitable for long-read sequencing from as little as 300 mg of stool. Our bioinformatics workflow, Lathe, uses a

Accessing genomes of microbes we have not yet cultivated

Identifying genetic determinants of microbial phenotypes


The avalanche of genomes
> Pangenome as a concept
Computing a pangenome
Pangenomics in practice

Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

Hervé Tettelin et al.

PNAS September 27, 2005 102 (39) 13950-13955; https://doi.org/10.1073/pnas.0506758102

The development of efficient and inexpensive genome sequencing methods has revolutionized the study of human bacterial pathogens and improved vaccine design. Unfortunately, **the sequence of a single genome does not reflect how genetic variability drives pathogenesis within a bacterial species** and also limits genome-wide screens for vaccine candidates or for antimicrobial targets. We have generated the genomic sequence of six strains representing the five major disease-causing serotypes of Streptococcus agalactiae, the main cause of neonatal infection in humans. Analysis of these genomes and those available in databases showed that the S. agalactiae species can be described by a pan-genome consisting of a core genome shared by all isolates, accounting for ≈80% of any single genome, plus a dispensable genome consisting of partially shared and strain-specific genes. Mathematical extrapolation of the data suggests that the gene reservoir available for inclusion in the S. agalactiae pan-genome is vast and that unique genes will continue to be identified even after sequencing hundreds of genomes.

Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

Hervé Tettelin et al.

PNAS September 27, 2005 102 (39) 13950-13955; https://doi.org/10.1073/pnas.0506758102

The development of efficient and inexpensive genome sequencing methods has revolutionized the study of human bacterial pathogens and improved vaccine design. Unfortunately, **the sequence of a single genome does not SAY ANYTHING AT ALL ABOUT THE EXTENT OF GENETIC VARIABILITY, ITS DRIVERS, OR ITS RELEVANCE**. We have generated the genomic sequence of six strains representing the five major disease-causing serotypes of Streptococcus agalactiae, the main cause of neonatal infection in humans. Analysis of these genomes and those available in databases showed that the S. agalactiae species can be described by a pan-genome consisting of a core genome shared by all isolates, accounting for ≈80% of any single genome, plus a dispensable genome consisting of partially shared and strain-specific genes. Mathematical extrapolation of the data suggests that the gene reservoir available for inclusion in the S. agalactiae pan-genome is vast and that unique genes will continue to be identified even after sequencing hundreds of genomes.

(emphases and edits by Meren)

Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

Hervé Tettelin et al.

PNAS September 27, 2005 102 (39) 13950-13955; https://doi.org/10.1073/pnas.0506758102

The development of efficient and inexpensive genome sequencing methods has revolutionized the study of human bacterial pathogens and improved vaccine design. Unfortunately, **the sequence of a single genome does not SAY ANYTHING AT ALL ABOUT THE EXTENT OF GENETIC VARIABILITY, ITS DRIVERS, OR ITS RELEVANCE**. We have generated the genomic sequence of six strains representing the five major disease-causing serotypes of Streptococcus agalactiae, the main cause of neonatal infection in humans. **Analysis of these genomes and those available in databases showed that the S. agalactiae species can be described by a pan-genome consisting of a core genome shared by all isolates, accounting for ≈80% of any single genome, plus a dispensable genome consisting of partially shared and strain-specific genes.** Mathematical extrapolation of the data suggests that the gene reservoir available for inclusion in the S. agalactiae pan-genome is vast and that unique genes will continue to be identified even after sequencing hundreds of genomes.

(emphases and edits by Meren)

Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

Hervé Tettelin et al.

PNAS September 27, 2005 102 (39) 13950-13955; https://doi.org/10.1073/pnas.0506758102

The development of efficient and inexpensive genome sequencing methods has revolutionized the study of human bacterial pathogens and improved vaccine design. Unfortunately, **the sequence of a single genome does not SAY ANYTHING AT ALL ABOUT THE EXTENT OF GENETIC VARIABILITY, ITS DRIVERS, OR ITS RELEVANCE**. We have generated the genomic sequence of six strains representing the five major disease-causing serotypes of Streptococcus agalactiae, the main cause of neonatal infection in humans. **Analysis of these genomes and those available in databases showed that the S. agalactiae species can be described by a pan-genome consisting of a core genome shared by all isolates, accounting for ≈80% of any single genome, plus a dispensable genome consisting of partially shared and strain-specific genes.** Mathematical extrapolation of the data suggests that the gene reservoir available for inclusion in the S. agalactiae pan-genome is vast and that unique genes will continue to be identified even after sequencing hundreds of genomes.

(emphases and edits by Meren)

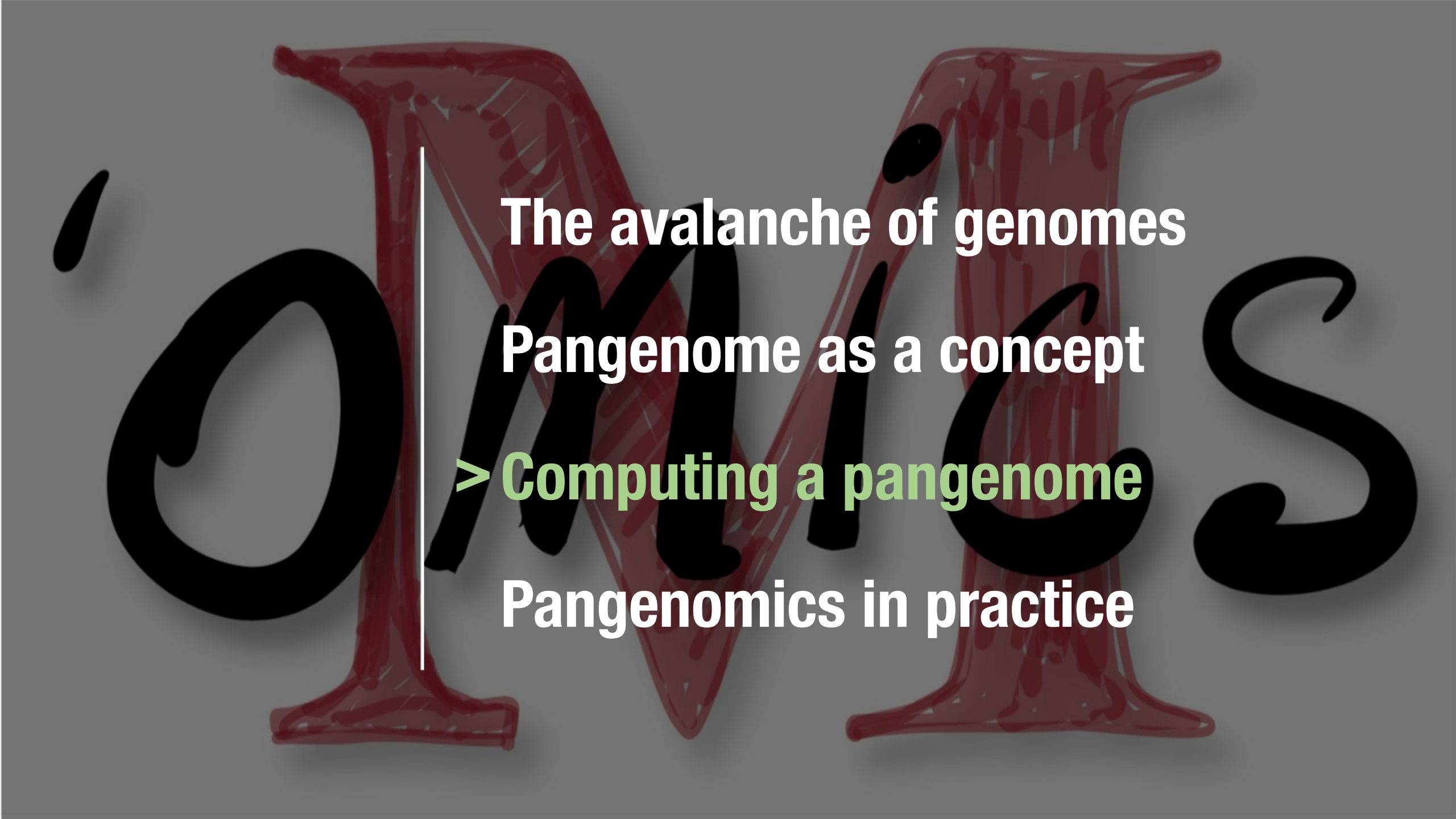
Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

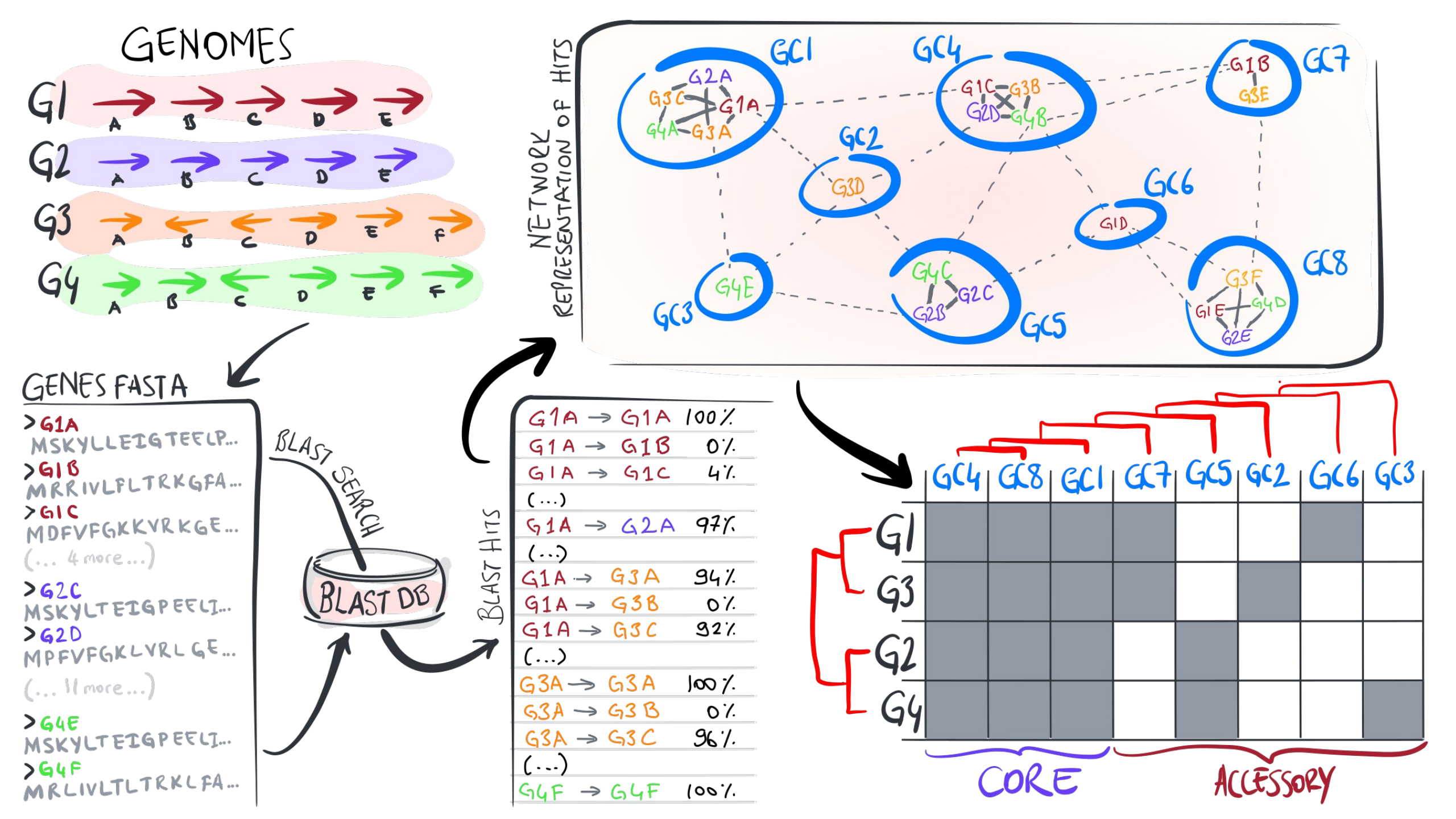
Hervé Tettelin et al.

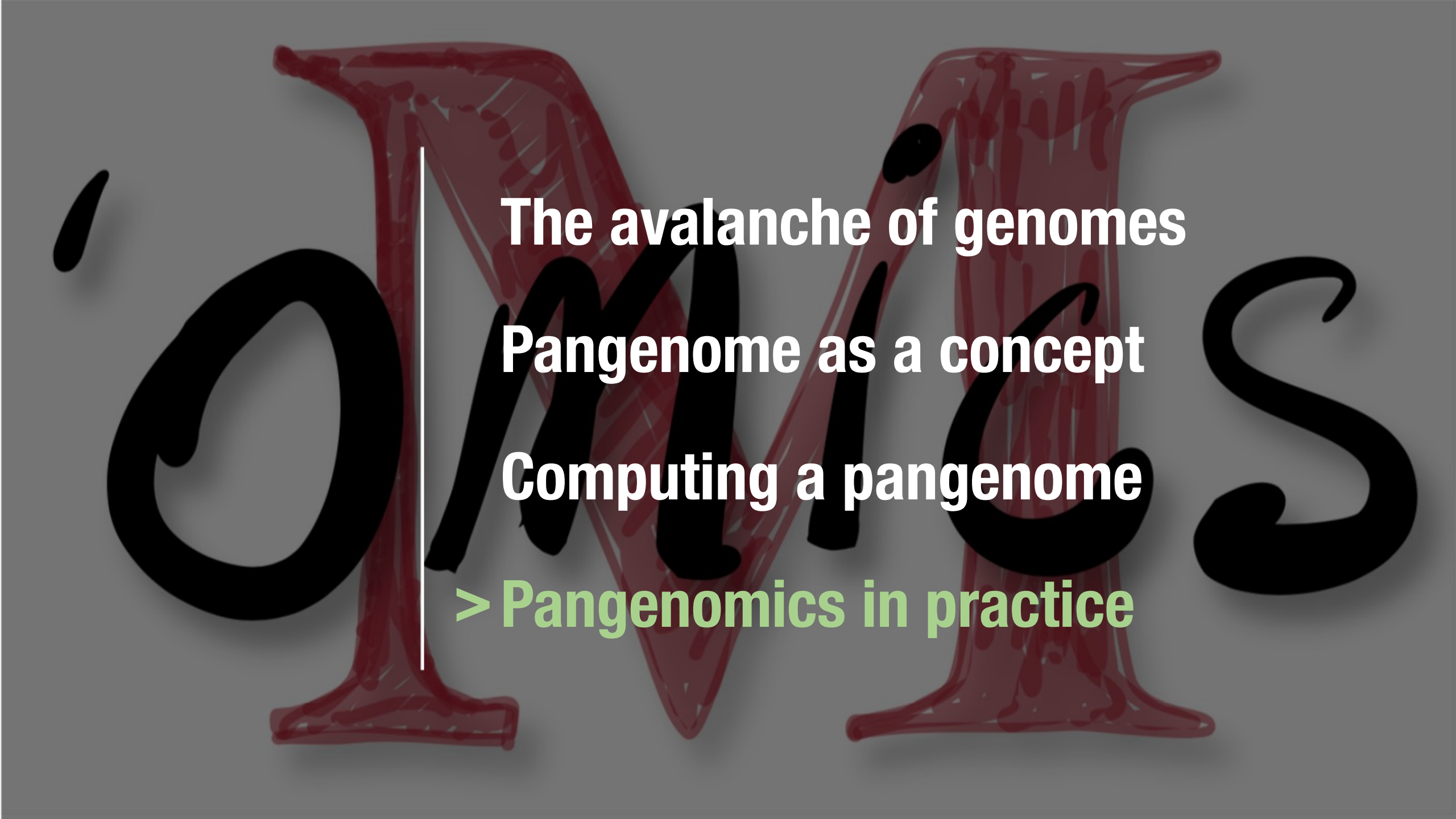
PNAS September 27, 2005 102 (39) 13950-13955; https://doi.org/10.1073/pnas.0506758102

The development of efficient and inexpensive genome sequencing methods has revolutionized the study of human bacterial pathogens and improved vaccine design. Unfortunately, **the sequence of a single genome does not SAY ANYTHING AT ALL ABOUT THE EXTENT OF GENETIC VARIABILITY, ITS DRIVERS, OR ITS RELEVANCE**. We have generated the genomic sequence of six strains representing the five major disease-causing serotypes of Streptococcus agalactiae, the main cause of neonatal infection in humans. **Analysis of these genomes and those available in databases showed that the S. agalactiae species can be described by a pan-genome consisting of a core genome shared by all isolates, accounting for ≈80% of any single genome, plus a dispensable genome consisting of partially shared and strain-specific genes.** Mathematical extrapolation of the data suggests that the gene reservoir available for inclusion in the S. agalactiae pan-genome is vast and that unique genes will continue to be identified even after sequencing hundreds of genomes.

(emphases and edits by Meren)

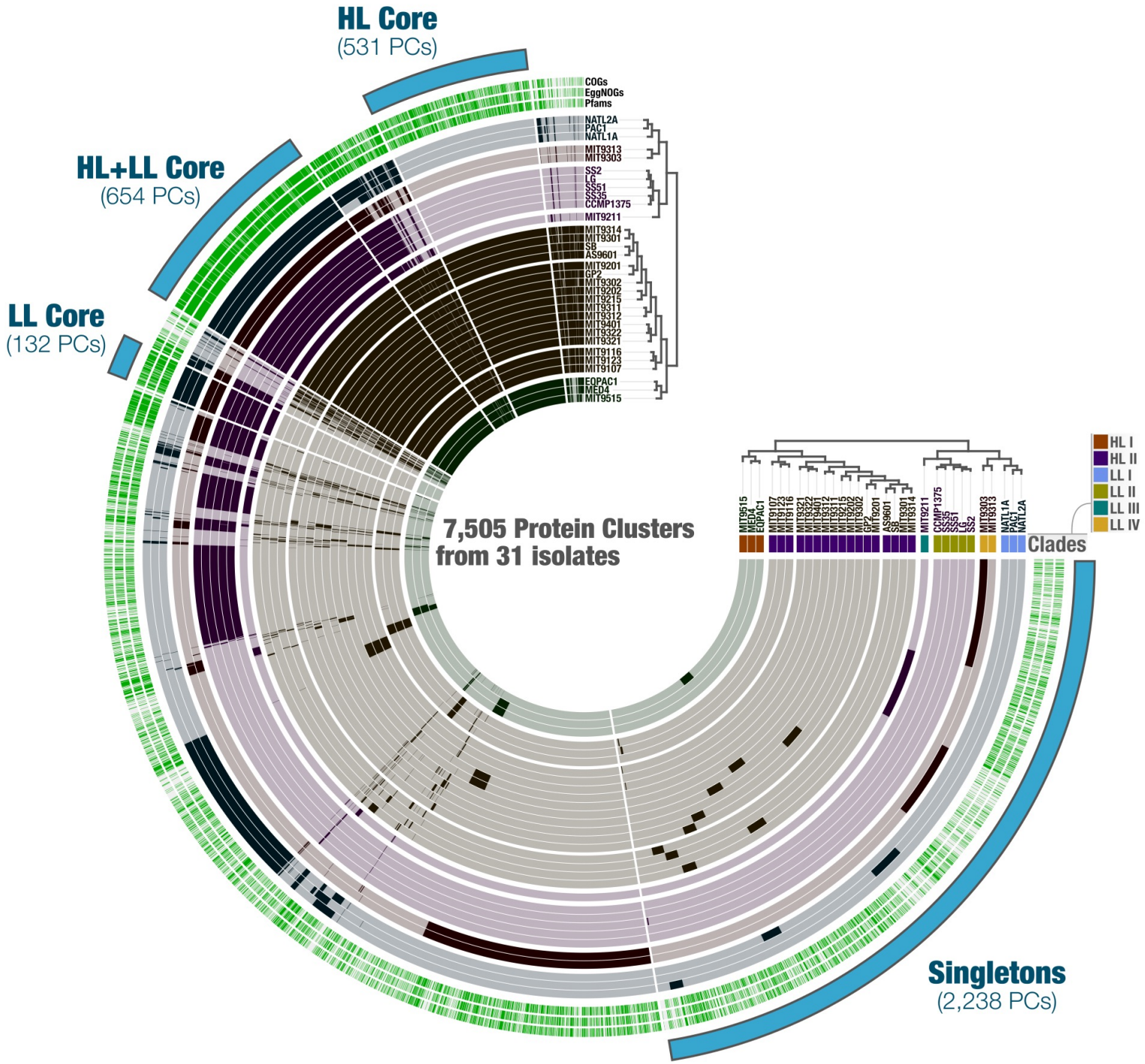

The avalanche of genomes
Pangenome as a concept
> Computing a pangenome
Pangenomics in practice


GENOMES
G1
A
B
C
D
E
G2
A
B
C
D
E
G3
A
B
C
D
E
F
G4
A
B
C
D
E
F
GENES FASTA
>G1A
MSKYLLEIGTEELP...
>G1B
MRRIVLFLTRKGFA...
>G1C
MDFVFGKKVRKGE...
(... 4 more...)
>G2C
MSKYLTEIGPEELI...
>G2D
MPFVFGKLVRLGE...
(... 11 more...)
>G4E
MSKYLTEIGPEELI...
>G4F
MRLIVLTLTRKLFA...
BLAST SEARCH
BLAST DB
BLAST HITS
G1A → G1A 100%
G1A → G1B 0%
G1A → G1C 4%
(...)
G1A → G2A 97%
(...)
G1A → G3A 94%
G1A → G3B 0%
G1A → G3C 92%
(...)
G3A → G3A 100%
G3A → G3B 0%
G3A → G3C 96%
(...)
G4F → G4F 100%
NETWORK REPRESENTATION OF HITS
GC1
G2A
G3C
G1A
G4A
G3A
GC4
G1C
G3B
G2D
G4B
GC7
G1B
G3E
GC2
G3D
GC6
G1D
GC3
G4E
G4C
G2C
G2B
GC5
GC8
G3F
G1E
G4D
G2E
GC4
GC8
GC1
GC7
GC5
GC2
GC6
GC3
G1
G3
G2
G4
CORE
ACCESSORY


The avalanche of genomes
Pangenome as a concept
Computing a pangenome
> Pangenomics in practice

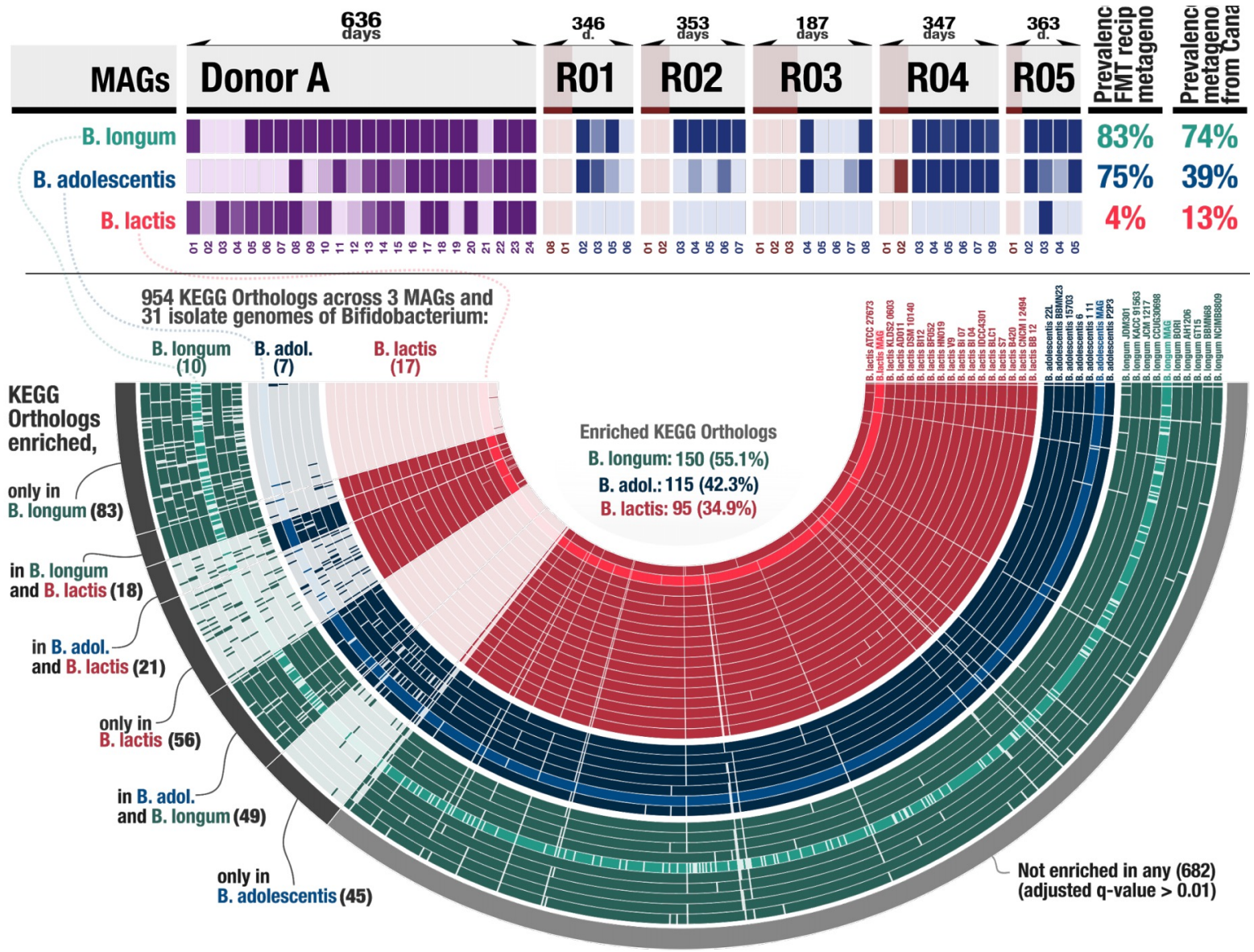
Linking pangenomes and metagenomes: the Prochlorococcus metapangenome

DELMONT TO, EREN AM



Adaptive ecological processes and metabolic independence drive microbial colonization and resilience in the human gut

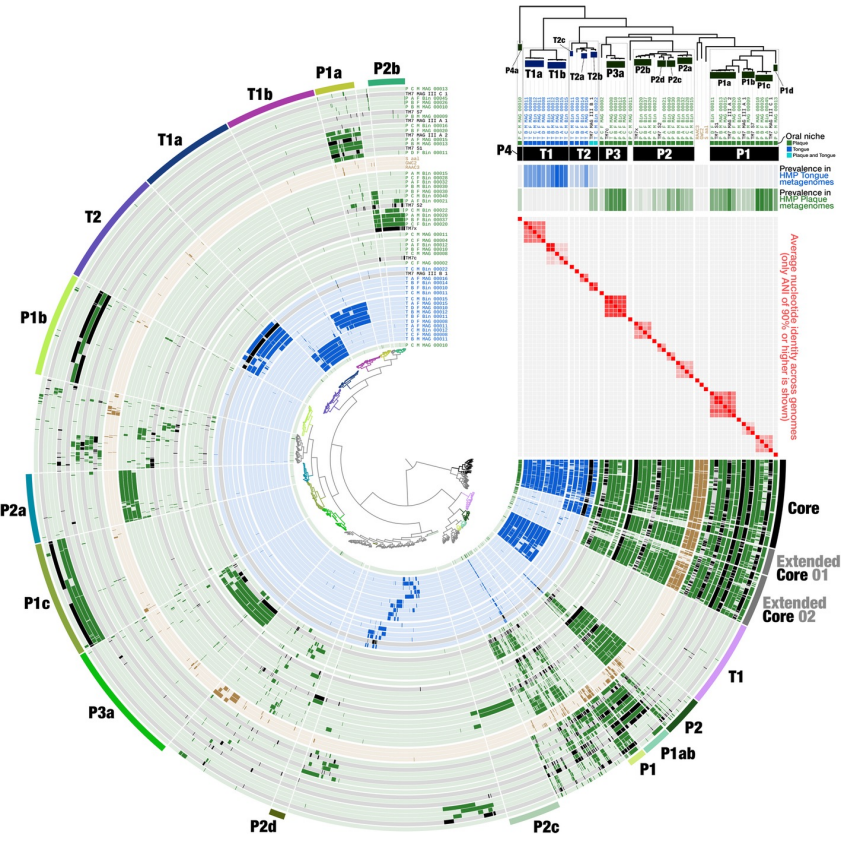
Watson AR, Füssel J, **Veseli I**, DeLongchamp JZ, Silva M, **Trigodet F**, **Lolans K**, **Shaiber A**, **Fogarty EC**, Quince C, Yu MK, Söylev A, Morrison HG, Lee ST, Rubin DT, Jabri B, Louie T, **Eren AM**



Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome

Shaiber A, Willis AD, Delmont TO, Roux S, Chen L, **Schmid AC**, **Yousef M**, **Watson AR**, **Lolans K**, **Esen ÖC**, **Lee STM**, Downey N, Morrison HG, Dewhirst FE, Mark Welch JL‡, **Eren AM**‡

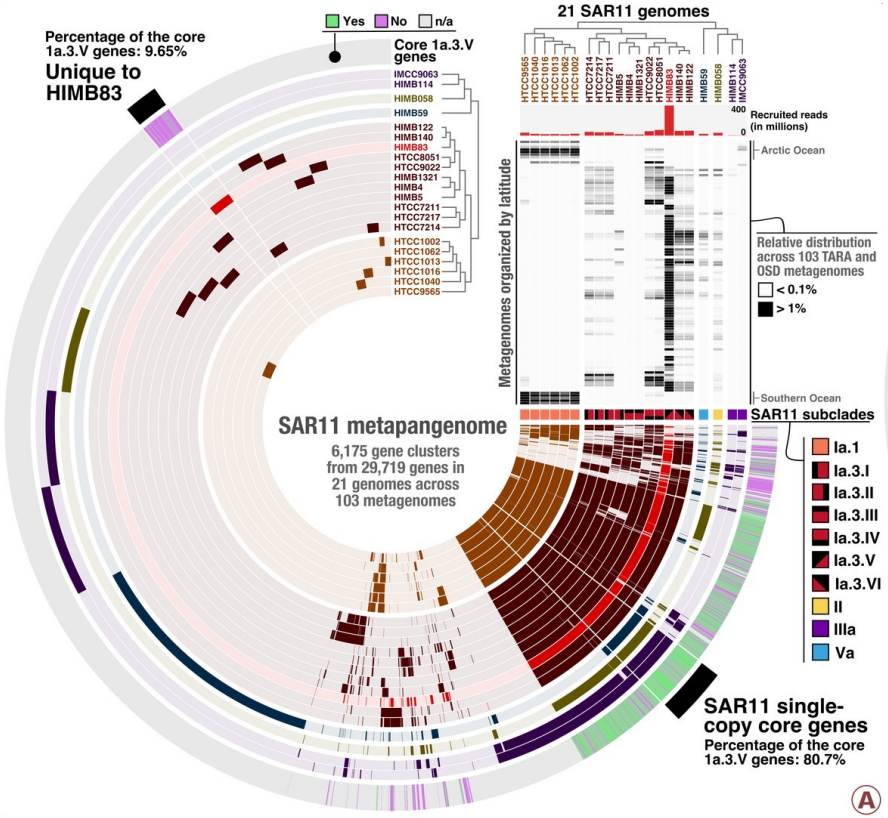
‡*Co-senior authors*



Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade

Delmont TO◒, **Kiefl E**◒, Kilinc O, **Esen ÖC**, Uysal I, Rappé MS, Giovannoni S, **Eren AM**

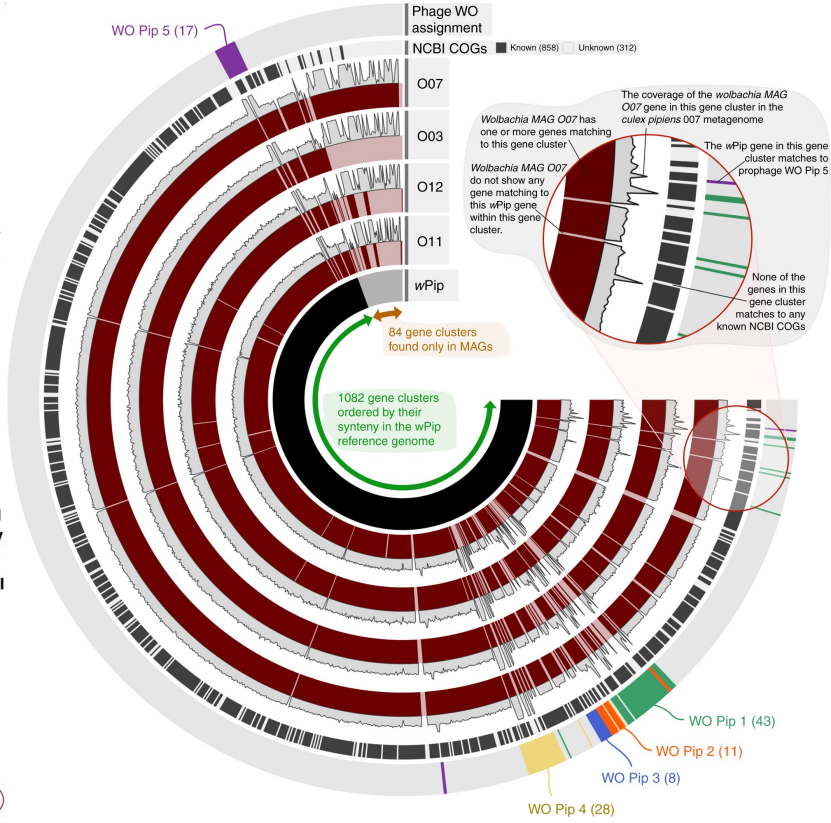
◒*Co-first authors*

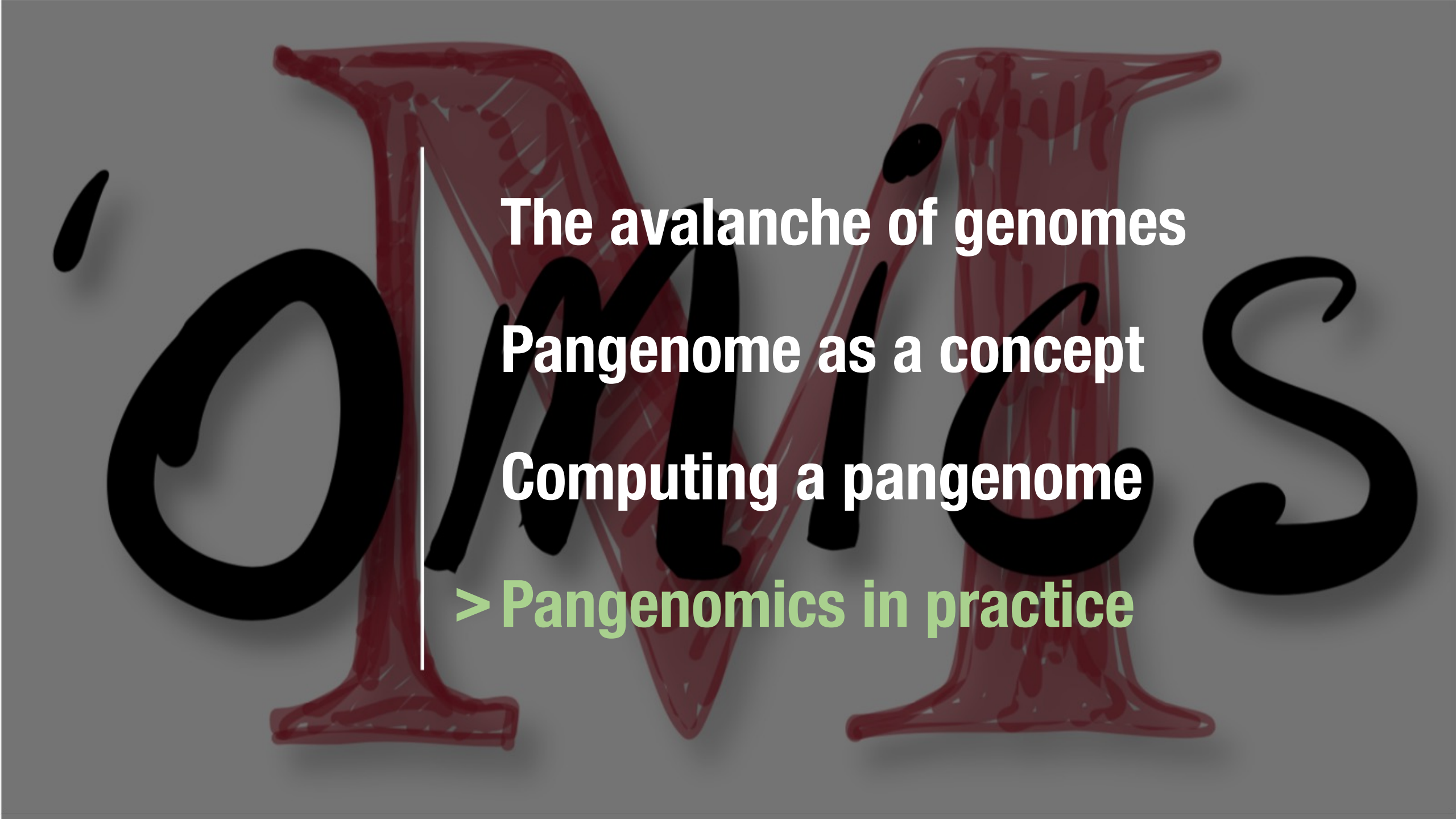


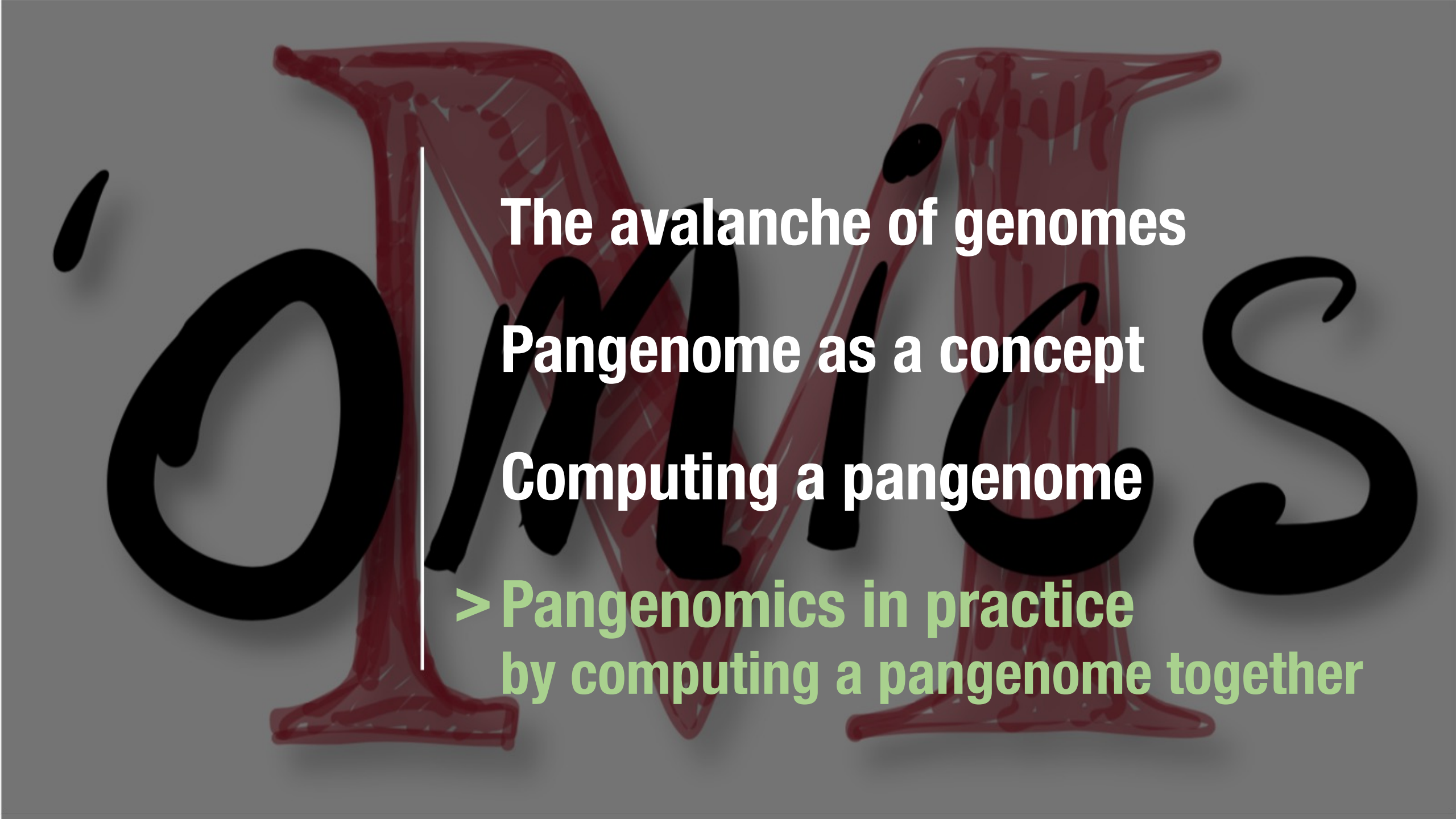
The Wolbachia mobilome in Culex pipiens includes a putative plasmid

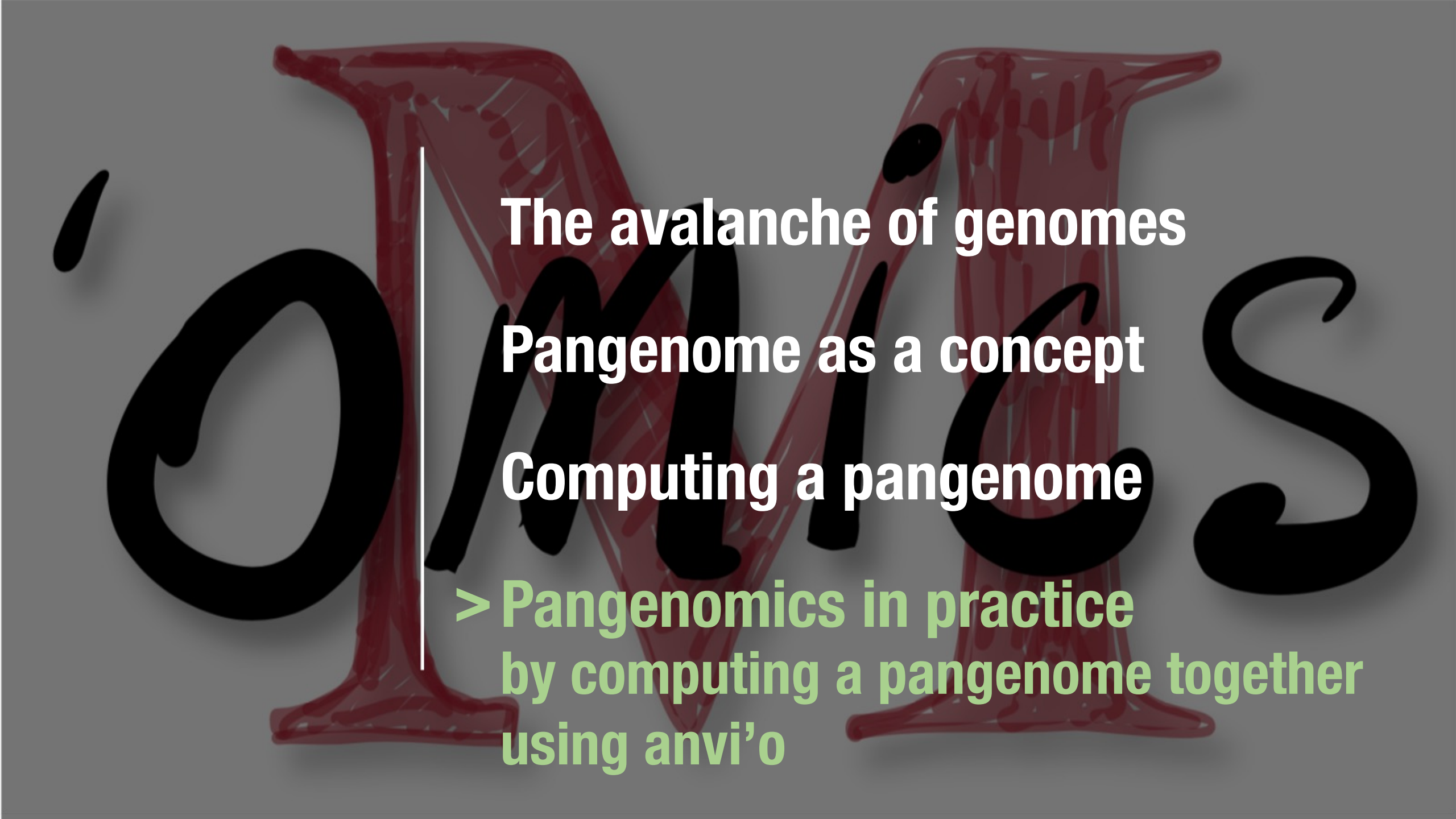
Reveillaud J◒, Bordenstein SR◒, Cruaud C, **Shaiber A**, **Esen ÖC**, Weill M, Makoundou P, **Lolans K**, **Watson AR**, Rakotoarivony I, Bordenstein S, **Eren AM**

◒*Co-first authors*




The avalanche of genomes
Pangenome as a concept
Computing a pangenome
> Pangenomics in practice


'omics
The avalanche of genomes
Pangenome as a concept
Computing a pangenome
> Pangenomics in practice
by computing a pangenome together



The avalanche of genomes

Pangenome as a concept

Computing a pangenome

> Pangenomics in practice
by computing a pangenome together
using anvi'o

'omics

https://merenlab.org/tutorials/vibrio-jasicida-pangenome/