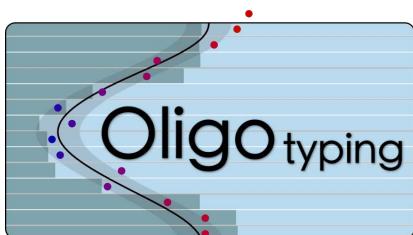
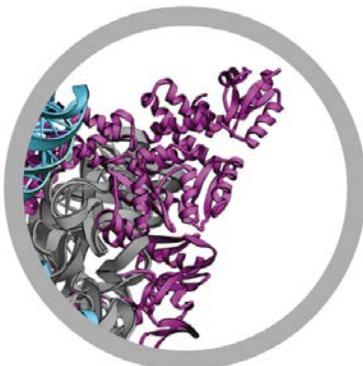


Marker genes, metagenomes, and other things

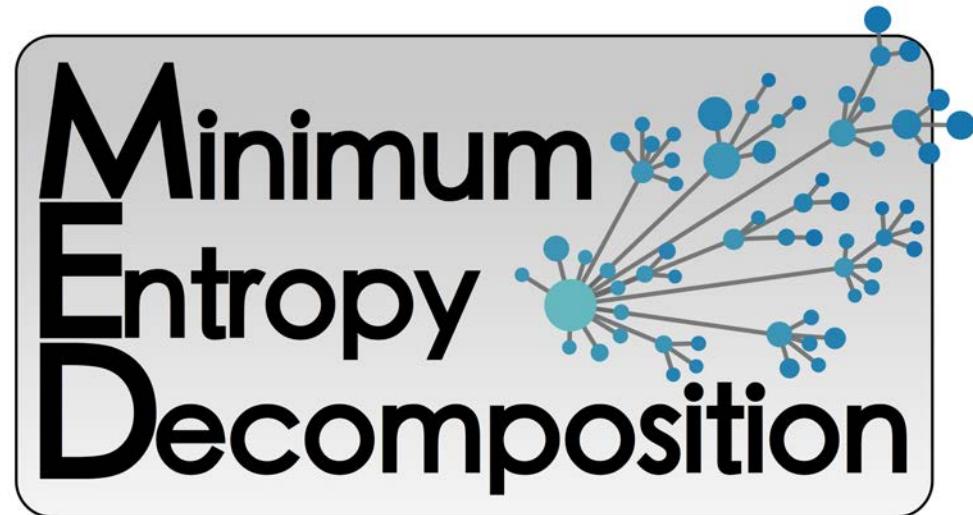


A. Murat Eren (Meren)
<http://merenlab.org>





Eren *et al.* (2013). *Methods in Ecology and Evolution*.

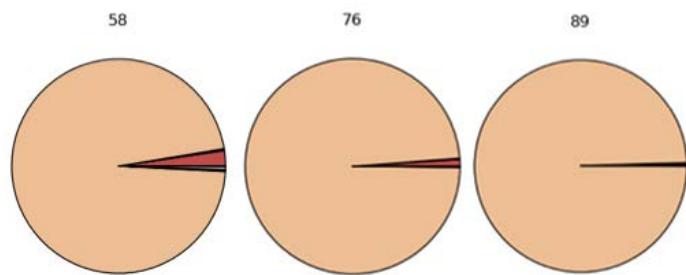
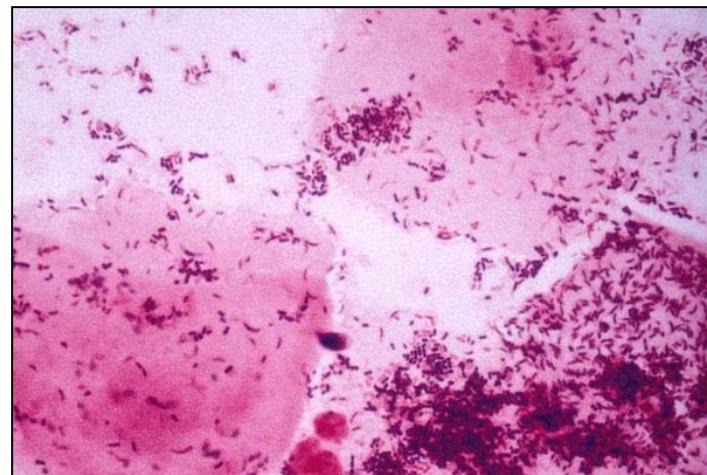
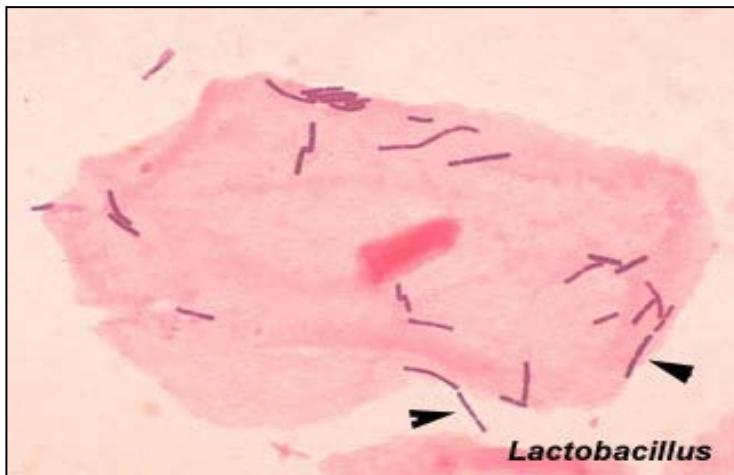


Eren *et al.* (2014). *ISME J.*

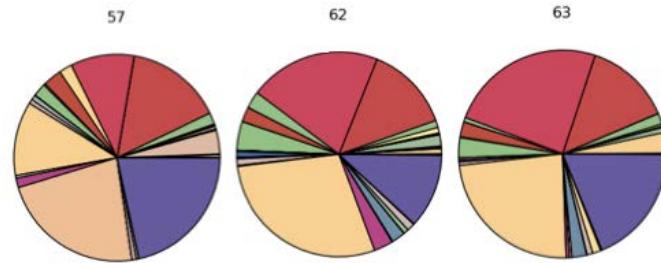
“An arbitrary amount of similarity between 16S reads is enough to group sequences into ecologically relevant units”

--*Bioinformatics*, ~~2011, 2012, 2013, 2014, 2015, 2016 (?)~~, 2017 (!).



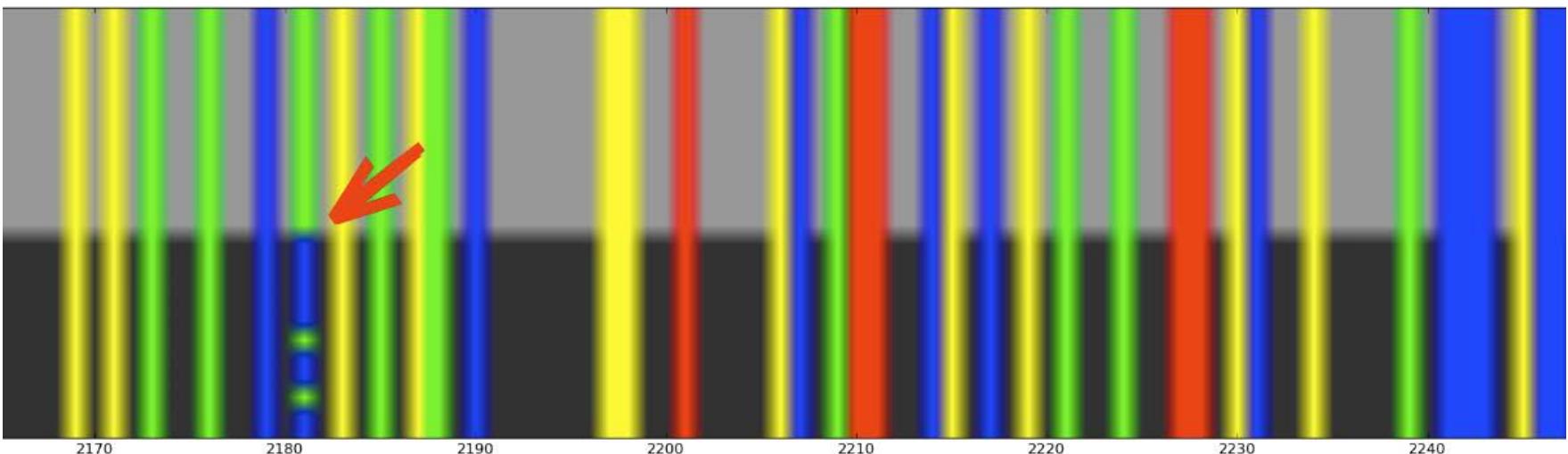


*(genus level bacterial
community composition of
three healthy women)*



*(genus level bacterial
community composition of
three women with BV)*

Reads that are binned into a ‘genus’, ‘species’, or a 3% OTU, usually have very subtle, yet systematical variation.

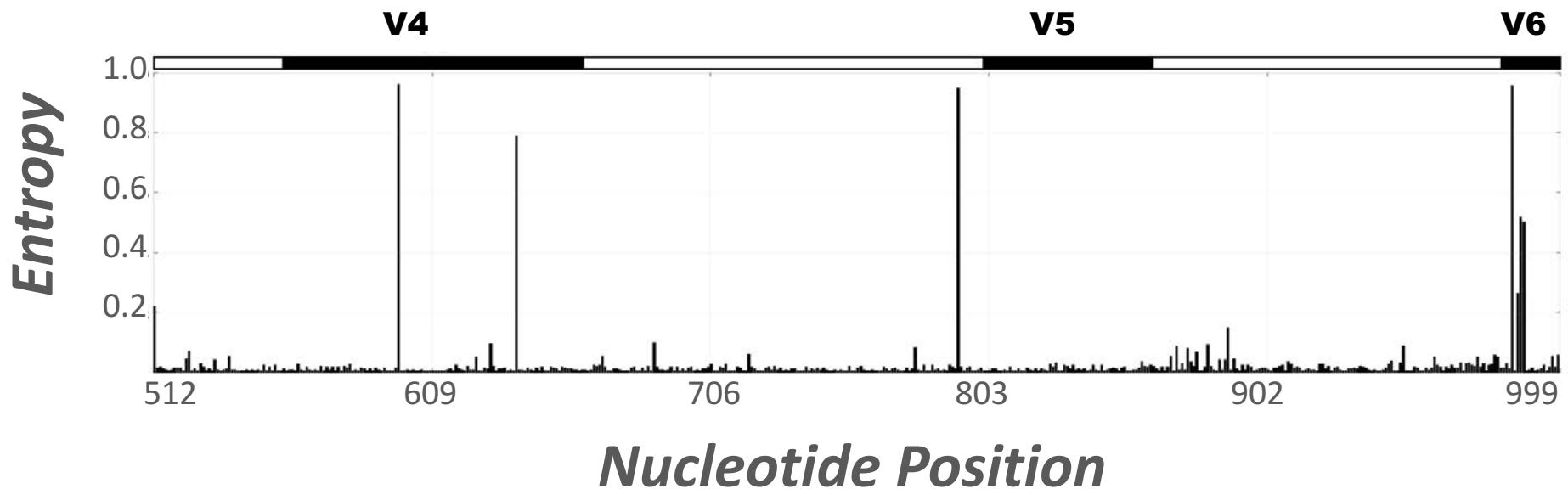




For a random variable X with n outcomes $\{n_i : i = 1, \dots, n\}$, the Shannon entropy, is defined as:

$$H(X) = - \sum_{i=0}^n p(x_i) \log_b(p(x_i))$$

Shannon entropy results for *Gardnerella vaginalis* alignments:



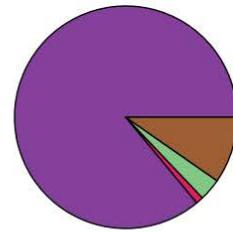
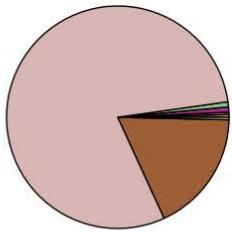
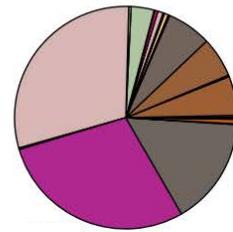
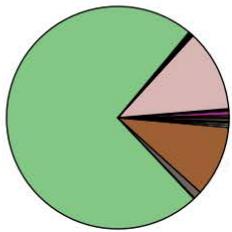
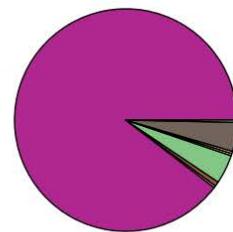
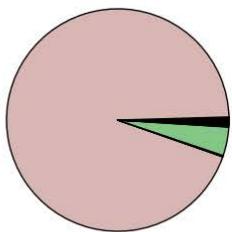
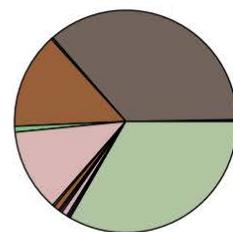
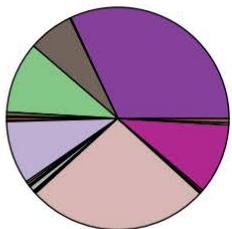
Oligotypes in Samples:

P01	
CCCACAGAT	169
CCCTCGATA	88
TTTTCGAAT	13
TTTACGAGG	9
CTCATGAGG	6
TCTACGAGA	3
TTCATGAGG	1
CCCACGGGA	1

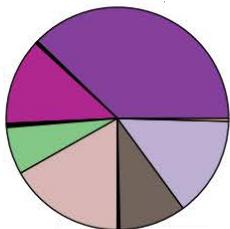
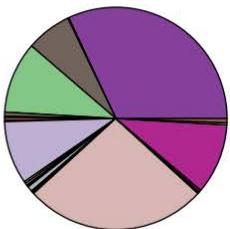
P02	
CCCACAGAT	220
TTTACGATA	26
TTTTCGAGC	12
TCCTCGAGG	6
TTTACAGCC	5
TCCACAGGG	4
CCTACGATT	3
CCCACGGAT	2

P03	
CCCACAGTA	172
CCCTCGAAT	85
TTTACGATT	12
TCTACGAAA	8
TCTTCGAGA	7
CCCACGAGC	3
CTCATGACG	3
CCTACGACC	2

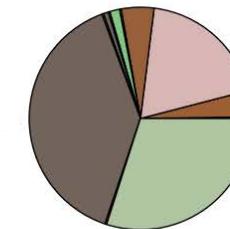
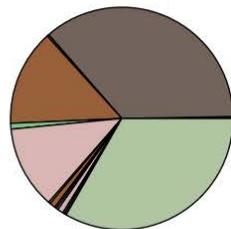
(...)



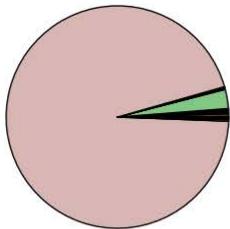
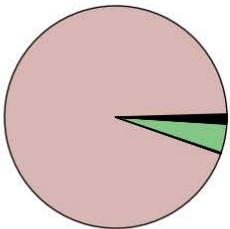
Couple 01, Vaginal and Penile Sample



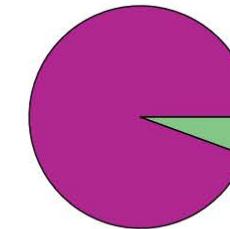
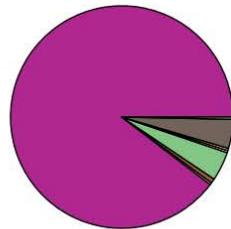
Couple 02, Vaginal and Penile Sample



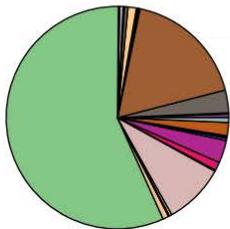
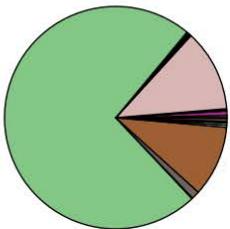
Couple 03, Vaginal and Penile Sample



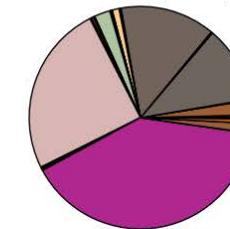
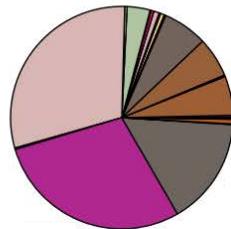
Couple 04, Vaginal and Penile Sample



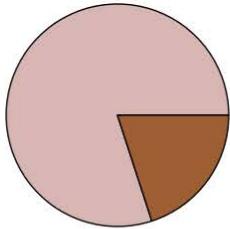
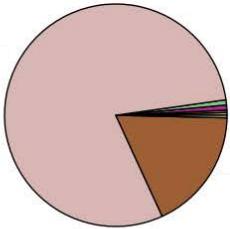
Couple 05, Vaginal and Penile Sample



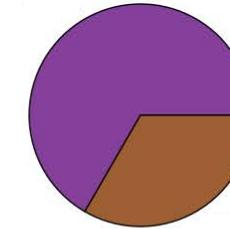
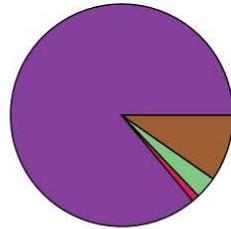
Couple 06, Vaginal and Penile Sample

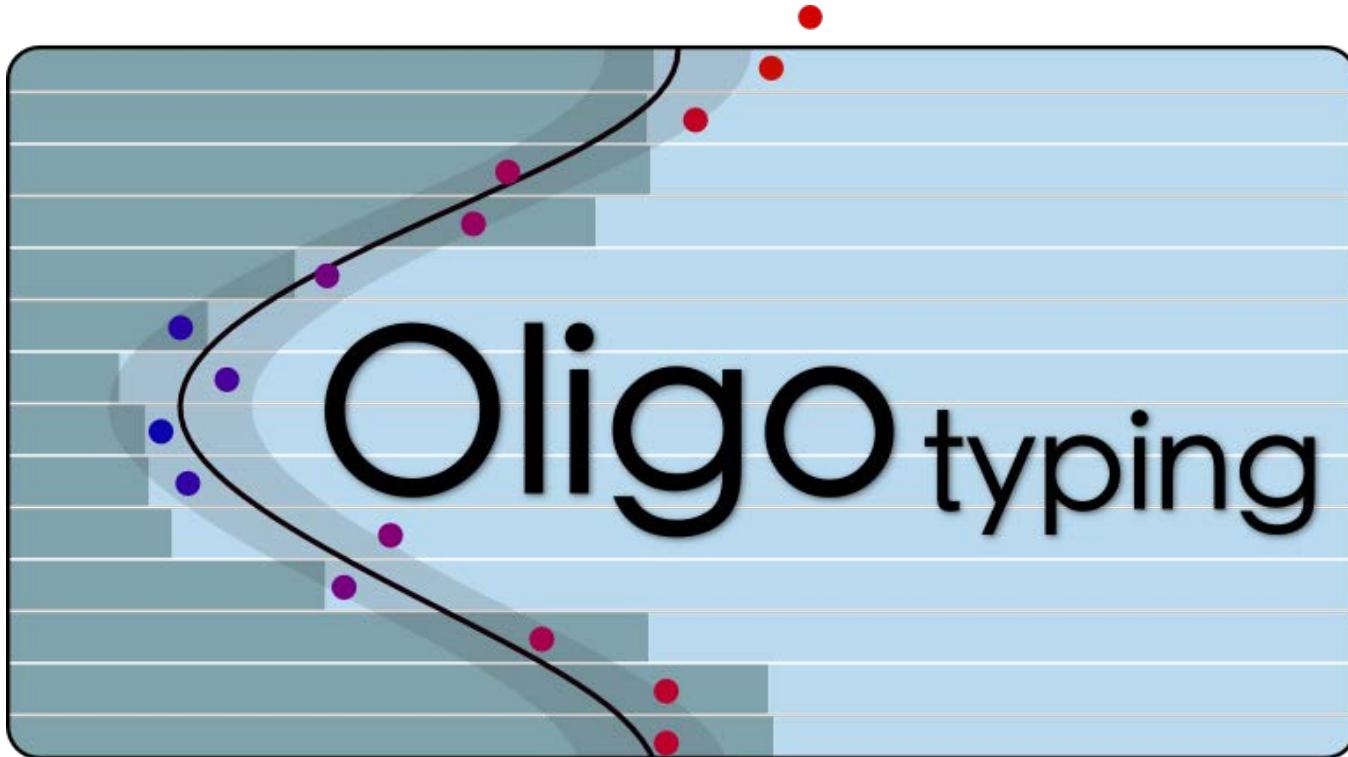


Couple 07, Vaginal and Penile Sample



Couple 08, Vaginal and Penile Sample





Eren et al., (2013). *Methods in Ecology and Evolution*.



1928 – 2012

Carl Woese. Photo courtesy of Don Hamerman (and Robin Tecon, who published it on his blog).

A Phylogenetic Definition of the Major Eubacterial Taxa

C. R. WOESE¹, E. STACKEBRANDT², T. J. MACKE³, and G. E. FOX⁴

¹ Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801, USA

² Institut für Allgemeine Mikrobiologie, Christian-Albrechts-Universität, 2300 Kiel, Federal German Republic

³ Department of Computer Science, University of Illinois, Urbana, Illinois 61801, USA

⁴ Department of Bioophysical Sciences, University of Houston, Houston, TX 77004, USA

Received January 30, 1985

Summary

Through oligonucleotide signature analysis of 16S ribosomal RNAs, it is possible to define ten major groups of eubacteria. These are:

- (1) the Gram positive bacteria,
- (2) the purple photosynthetic bacteria and their relatives,
- (3) the spirochetes and their relatives,
- (4) the sulfur-dependent eubacteria and their relatives,
- (5) the bacteroides, flavobacteria and cytophagas and their relatives,
- (6) the cyanobacteria,
- (7) the green sulfur bacteria,
- (8) the green non-sulfur bacteria and their relatives,
- (9) the radio-resistant micrococci, and
- (10) the planctomyces and their relatives.

Although no consensus exists as regards the taxonomic terminology, these ten groupings are appropriately termed eubacterial Phyla or Divisions. The major subdivisions of those Phyla or Divisions that have been extensively characterized can also be defined by characteristic oligonucleotide signatures.

Certain regions in the rRNA molecule are highly conserved in sequence – some positions therein showing no variation whatever among *all* organisms, other positions varying in composition, but infrequently. For reasons that are not completely understood (Woese et al., 1985b) changes involving these regions of conserved sequence tend to occur mainly during the evolution of major phylogenetic groups. Because of this, sequence in these regions tends to be characteristic of the major taxonomic groups, and so is useful in defining them. Families of sequence variation occurring at these phylogenetically significant locales are shown in Table 1. A set of oligonucleotides that unequivocally defines each of the eubacterial "phyla" can be extracted from these data.

(6) *The Cyanobacteria*

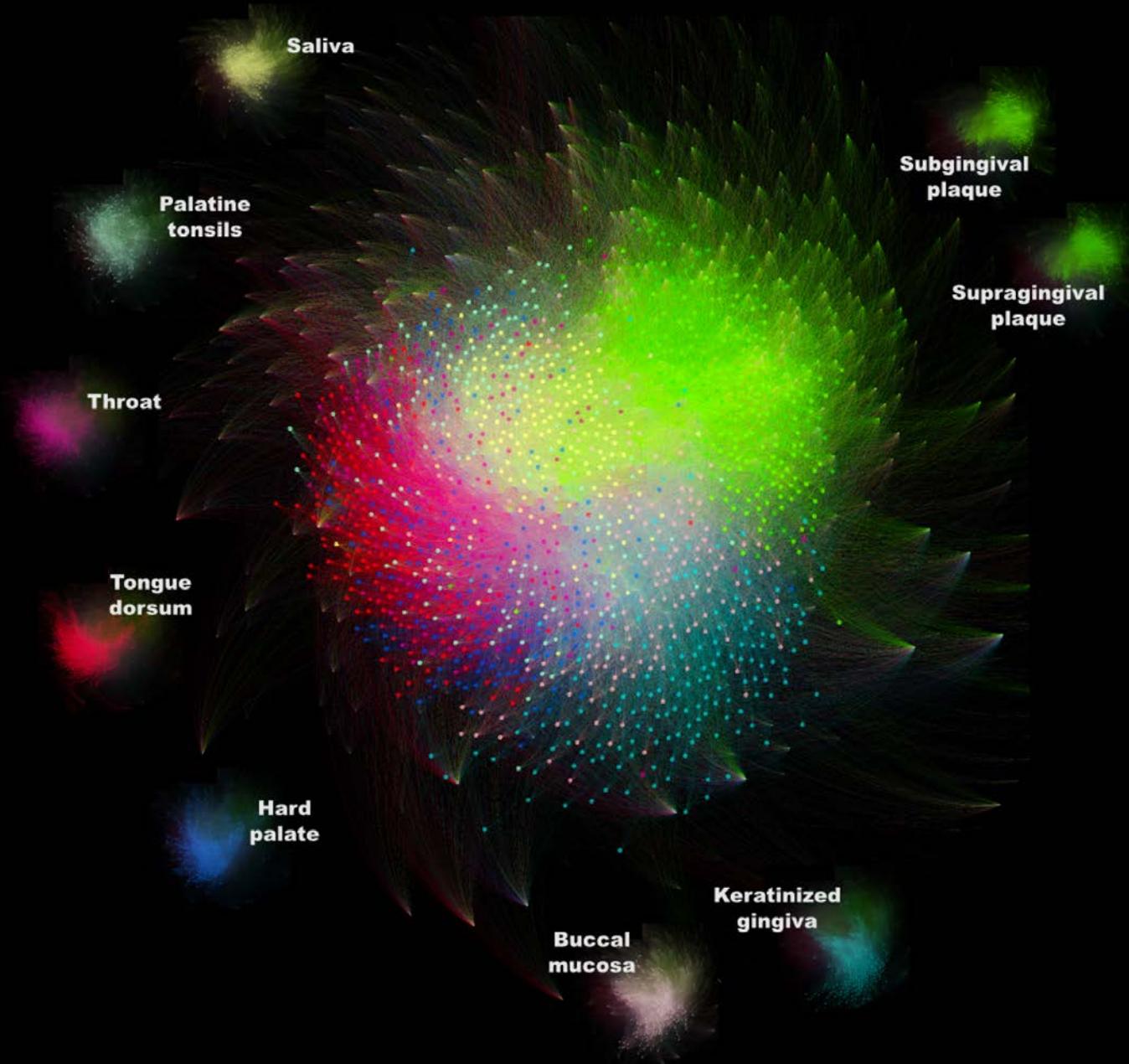
The grouping is well defined by oligonucleotides at positions 365, 795, 1210 and 1240. Too few species of this group have been characterized by cataloging, however, to define its internal structure well (Doolittle et al. 1975).

(8) *The Green Non-sulfur Bacteria and Relatives*

This “phylum” presents one of the more distinctive oligonucleotide signatures; see positions 50, 315, 910 and 1225. Whether this means that *Chloroflexus* and its relatives represent a particularly deep branching in the

(10) *The Planctomyces Group*

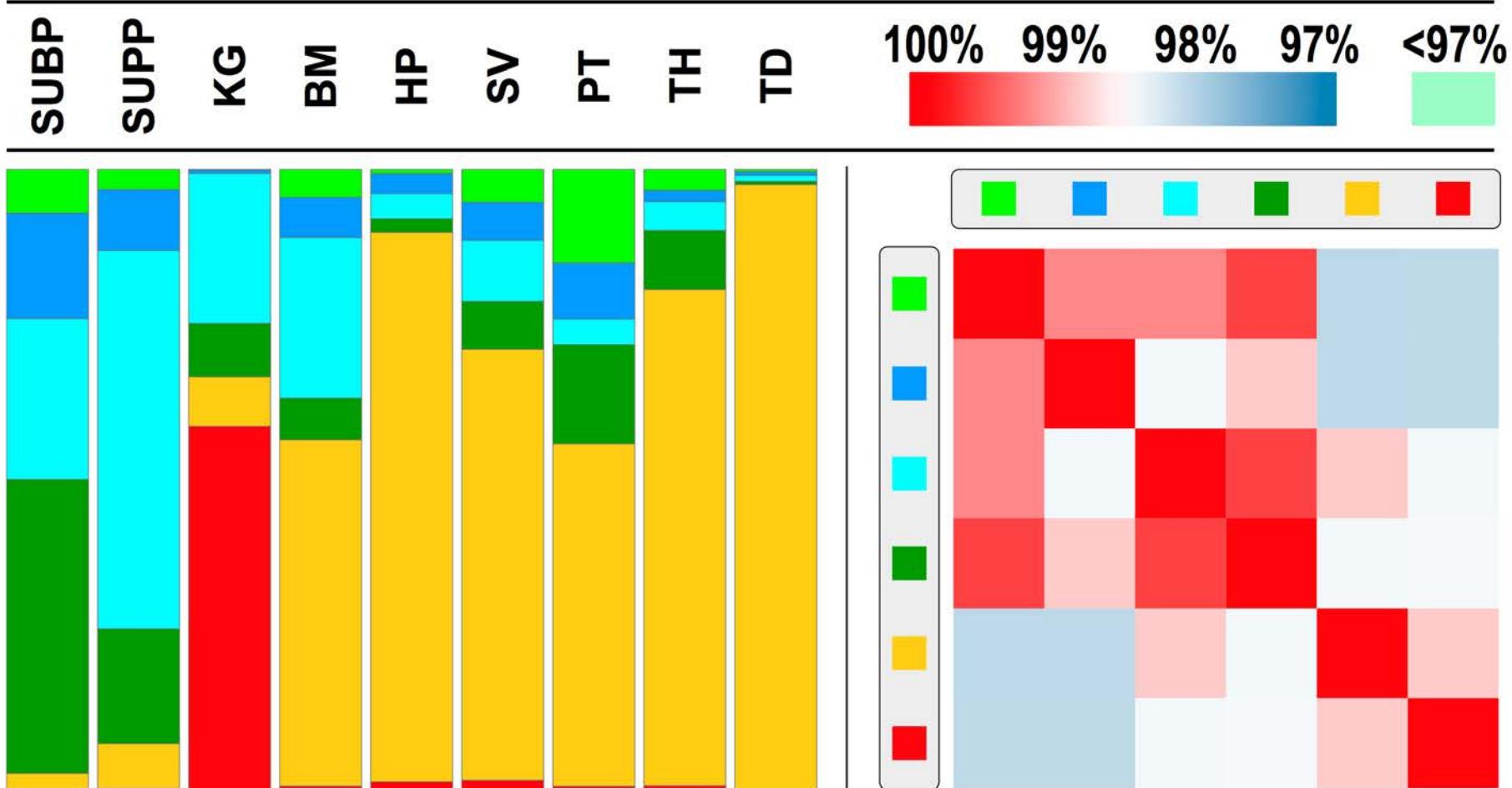
Of all the eubacterial “phyla” this one is the most distinct in terms of its oligonucleotide signature – e.g. positions 340, 570, 935, 960, 985 and 1240 – a uniqueness also reflected in the exceptionally low Sab values. Planctomyces show with other eubacterial rRNAs



Eren *et al.* (2014). PNAS

Oligotypes per Oral Site

Oligotype Identity



Fusobacterium

F. nucl. ss polymorphum

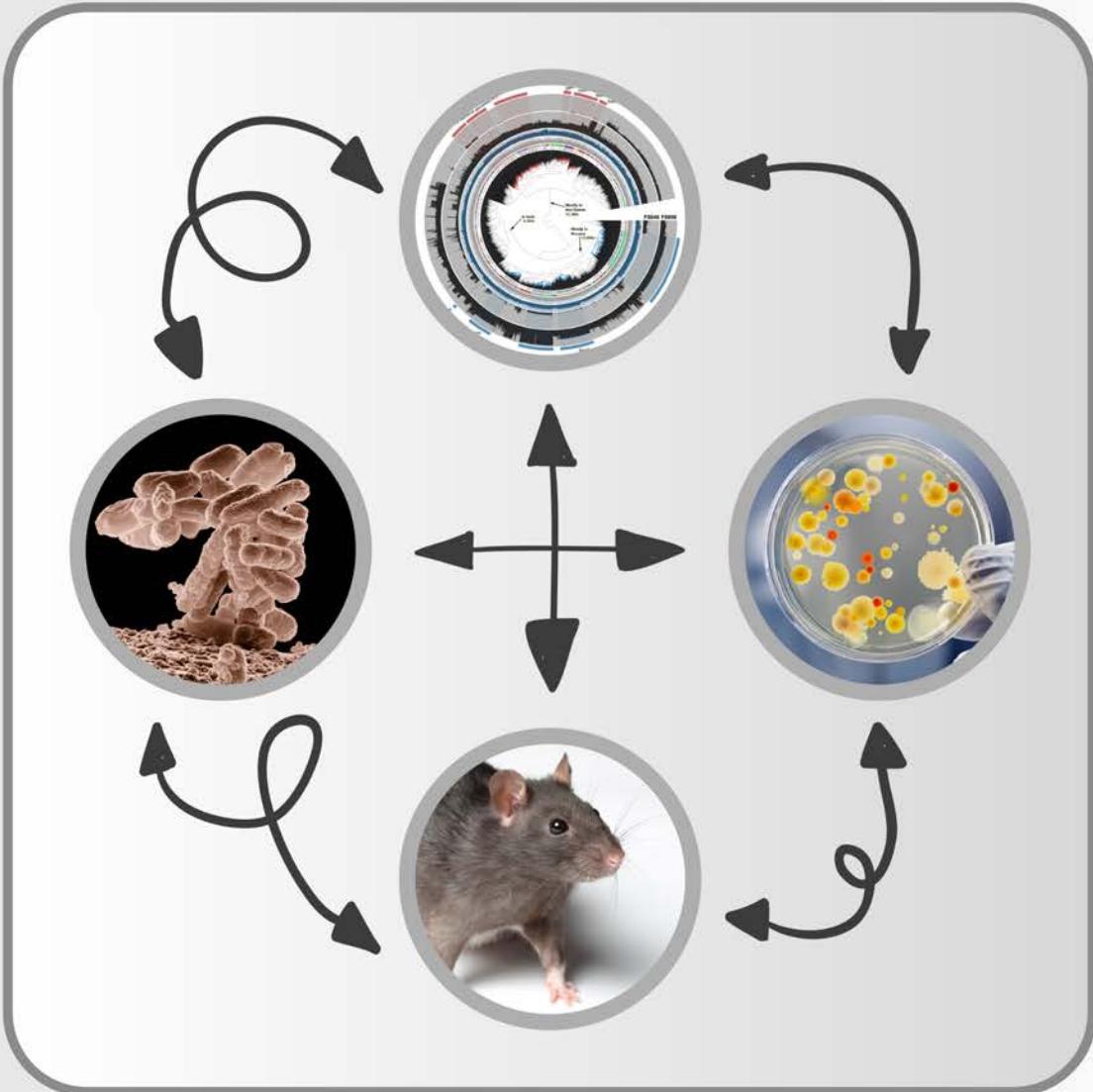
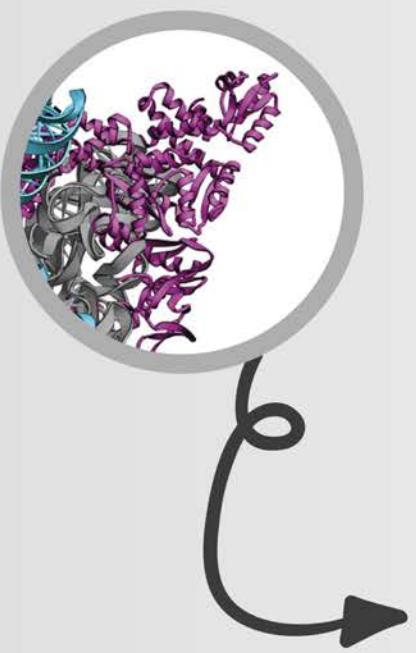
F. nucleatum ss vincentii I

F. nucl. ss vincentii II

F. periodonticum

F. nucleatum ss animalis

F. periodonticum 98.8%



Interactive environments to perform in-depth analyses of metagenomic data including the characterization of subtle nucleotide variation.

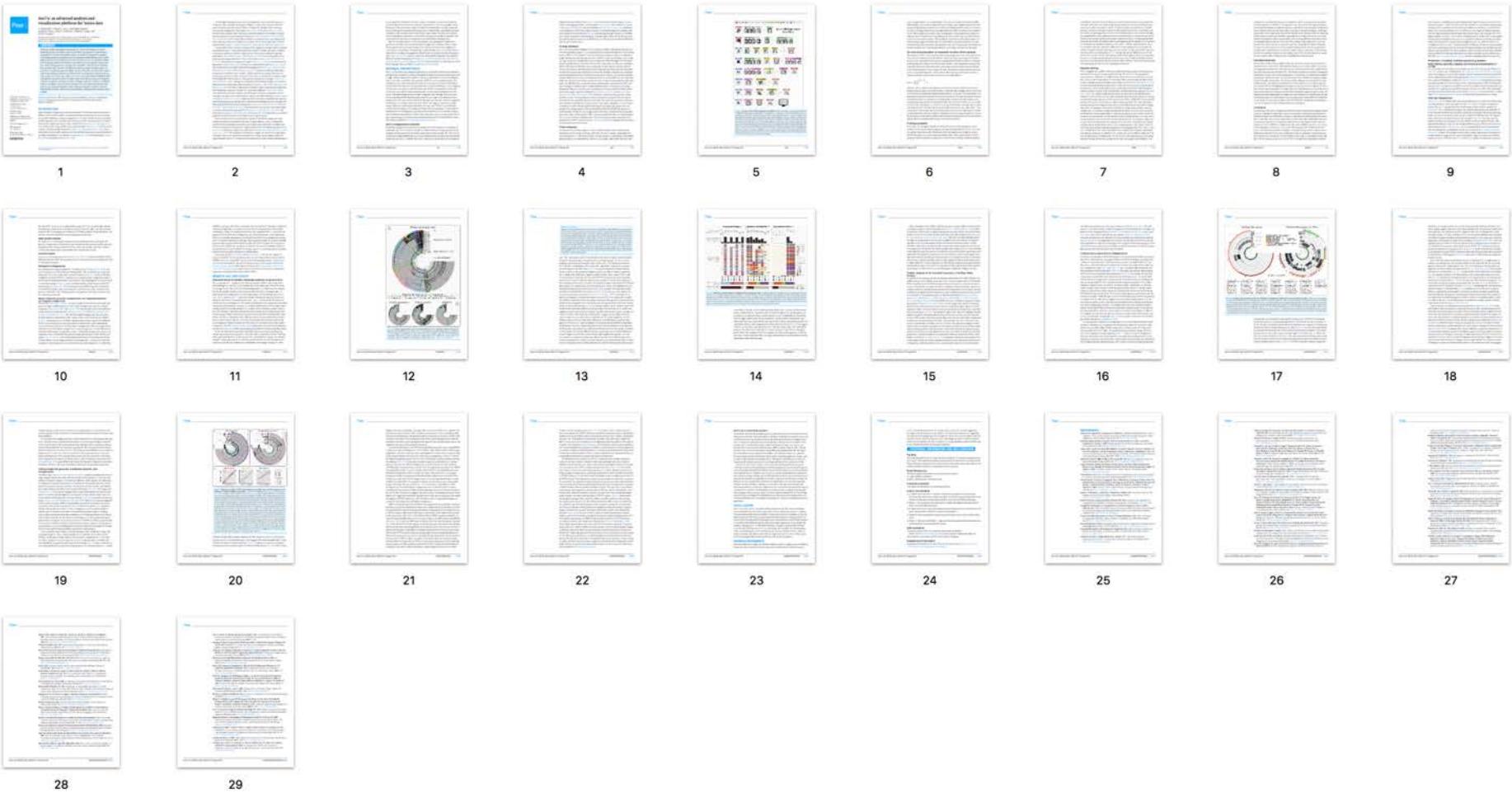
Easy ways to combine datasets of different nature (*i.e.*, putting metagenomic and metatranscriptomic analyses into one intuitive display).

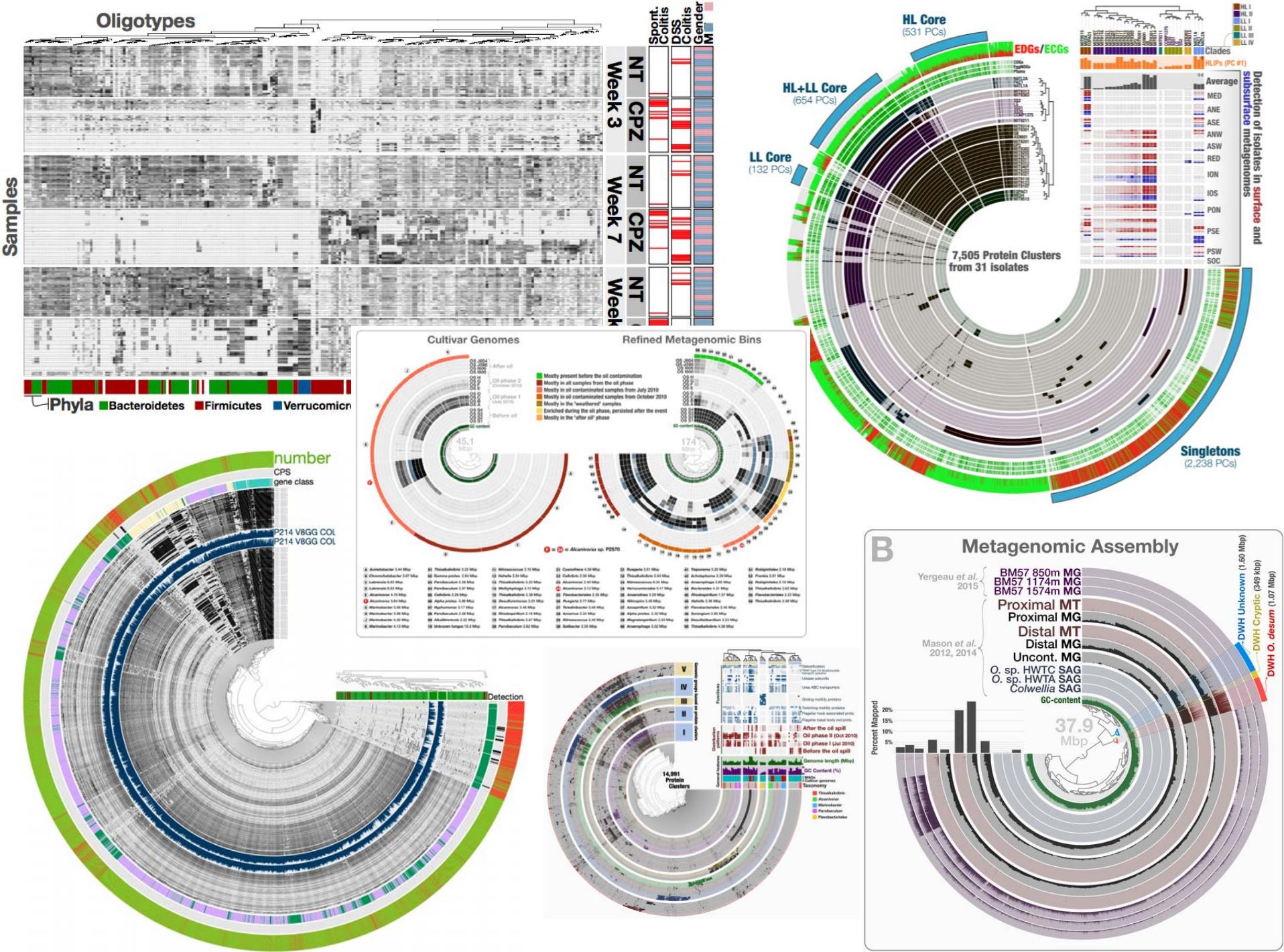
Ability to share / communicate findings.

Access to reasonable data structures to implement complex ideas quickly, without having to deal with boring details of metagenomic / metatranscriptomic workflow.

Anvi'o: an advanced analysis and visualization platform for 'omics data

A. Murat Eren^{1,2}, Özcan C. Esen¹, Christopher Quince³,
Joseph H. Vineis¹, Hilary G. Morrison¹, Mitchell L. Sogin¹ and
Tom O. Delmont¹

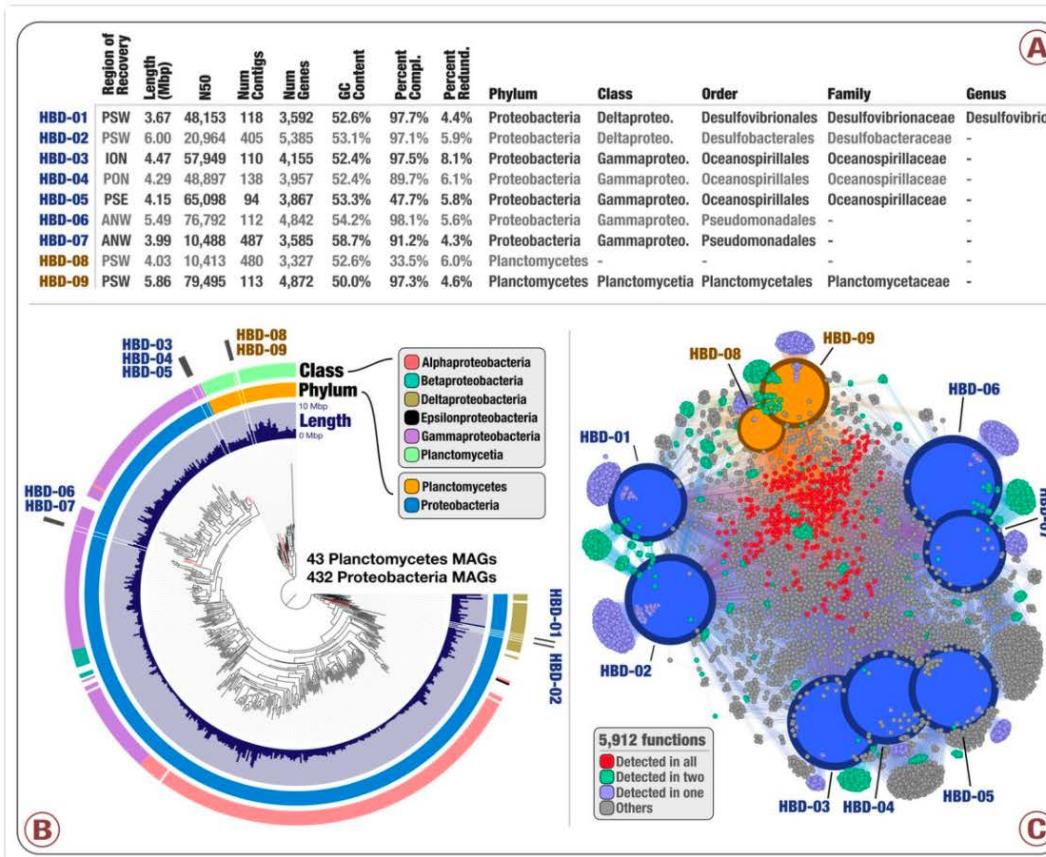






A. Murat Eren
@merenbey

~1,000 microbial genomes from TARA Oceans, including novel nitrogen-fixing populations: [biorxiv.org/content/early/... by @tomodelmont et al.](https://www.biorxiv.org/content/early/2017/04/23/129791)



2:44 PM - 23 Apr 2017

126 Retweets 128 Likes



bioRxiv
beta
THE PREPRINT SERVER FOR BIOLOGY

HOME

Search

New Results

Nitrogen-Fixing Populations Of Planctomycetes And Proteobacteria Are Abundant In The Surface Ocean

Tom O. Delmont, Christopher Quince, Alon Shaiber, Ozcan C. Esen, Sonny T. M. Lee, Sebastian Lucker, A. Murat Eren

doi: <https://doi.org/10.1101/129791>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract Info/History Metrics

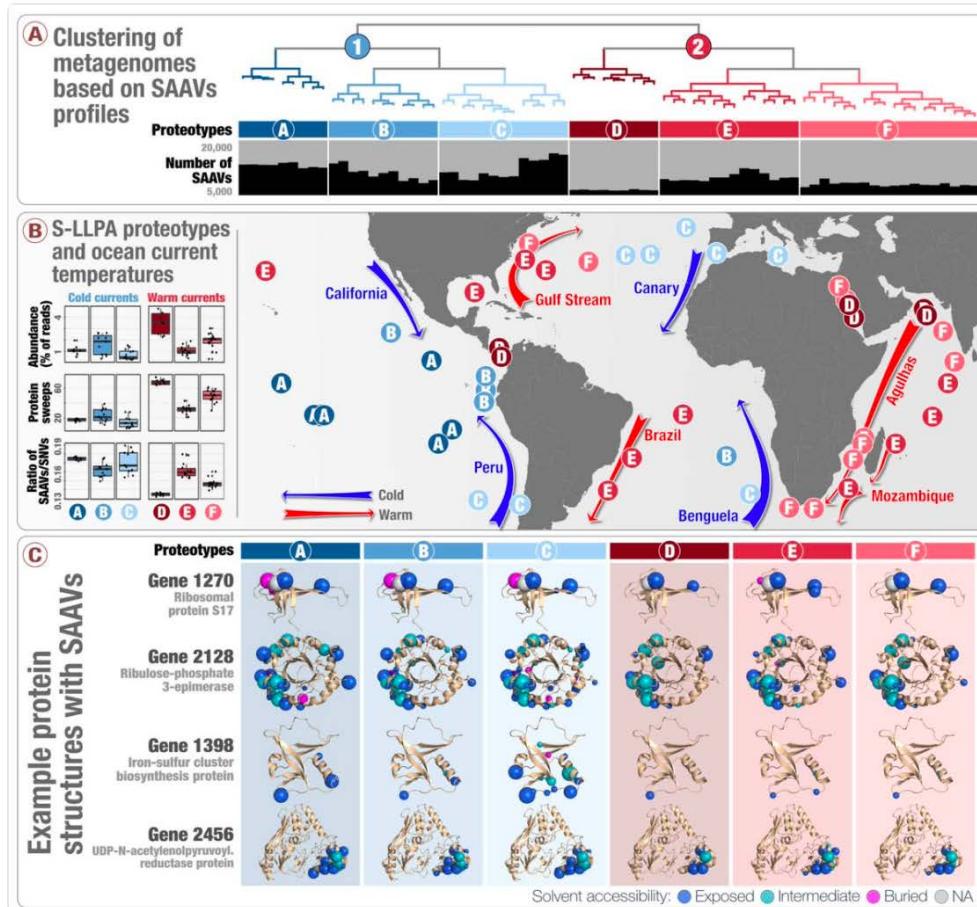
Preview PDF

Abstract

Nitrogen fixation in the surface ocean impacts the global climate by regulating the microbial primary productivity and the sequestration of carbon through the biological pump. Cyanobacterial populations have long been thought to represent the main suppliers of the bio-available nitrogen in this habitat. However, recent molecular surveys of nitrogenase reductase gene revealed the existence of rare non-cyanobacterial populations that can also fix nitrogen. Here, we characterize for the first time the genomic content of some of these heterotrophic bacterial diazotrophs (HBDs) inhabiting the open surface ocean waters. They represent new lineages within Proteobacteria and Planctomycetes, a phylum never linked to nitrogen fixation prior to this study. HBDs were surprisingly abundant in the Pacific Ocean and the Atlantic Ocean northwest, conflicting with decades of PCR surveys. The abundance and widespread occurrence of non-cyanobacterial diazotrophs in the surface ocean emphasizes the need to re-evaluate their role in the nitrogen cycle and primary productivity.

A. Murat Eren
@merenbey

Our new pre-print: "SAR11 amino acid variants are governed by natural selection": [biorxiv.org/content/early/](https://www.biorxiv.org/content/early/) ... with #anvio.
Comments welcome!



bioRxiv
beta
THE PREPRINT SERVER FOR BIOLOGY

HOME | AB

Search

New Results

The global biogeography of amino acid variants within a single SAR11 population is governed by natural selection

Tom O. Delmont, Evan Kiefl, Ozsel Kilinc, Ozcan C. Esen, Ismail Uysal, Michael S. Rappe, Stephen Giovannoni,
A. Murat Eren

doi: <https://doi.org/10.1101/170639>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

Info/History

Metrics

Preview PDF

Abstract

The diversity and geographical distribution of populations within major marine microbial lineages are largely governed by temperature and its co-variables. However, neither the mechanisms by which genomic heterogeneity emerges within a single population nor how it drives the partitioning of ecological niches are well understood. Here we took advantage of billions of metagenomic reads to study one of the most abundant and widespread microbial populations in the surface ocean. We characterized its substantial amount of genomic heterogeneity using single-amino acid variants (SAAVs), and identified systematic purifying selection and adaptive mechanisms governing non-synonymous variation within this population. Our Deep Learning analysis of SAAVs across metagenomes revealed two main ecological niches that reflect large-scale oceanic current temperatures, as well as six prototypes demarcating finer-resolved niches. We identified significantly more protein variants in cold currents and an increased number of protein sweeps in warm currents, exposing a global pattern of alternating genomic diversity for this SAR11 population as it drifts along with surface ocean currents. Overall, the geographic partitioning of SAAVs suggests natural selection, rather than neutral evolution, is the main driver of the evolution of SAR11 in surface oceans.

11:03 AM - 31 Jul 2017

54 Retweets 83 Likes



Microbiome



Sonny T. M. Lee :: Post-doctoral Scientist

 Email

📍 Address: Knapp Center for Biomedical Discovery, 900 E. 57th St., MB 9, RM 9148, Chicago, IL 60637 USA

I am a microbiologist studying the most interesting inhabitants of humans, coral reefs, and the ocean, and using big data to answer vexing problems of our future.

RESEARCH | OPEN ACCESS

Tracking microbial colonization in fecal microbiota transplantation experiments via genome-resolved metagenomics

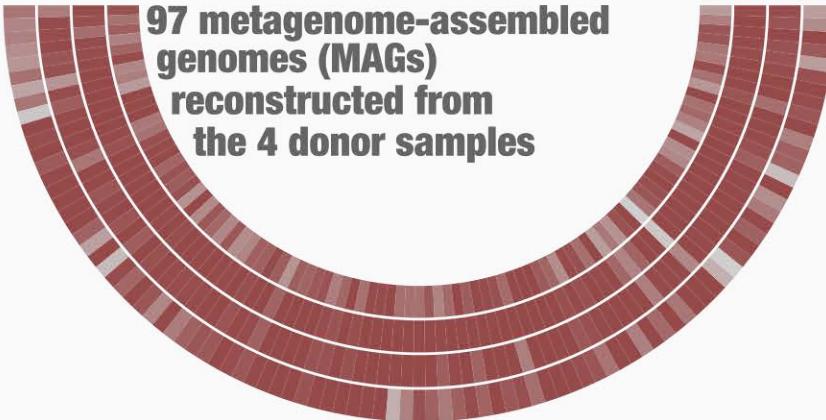
Sonny T. M. Lee[†], Stacy A. Kahn[†], Tom O. Delmont, Alon Shaiber, Özcan C. Esen, Nathaniel A. Hubert, Hilary G. Morrison, Dionysios A. Antonopoulos, David T. Rubin and A. Murat Eren 

[†] Contributed equally

Donor

DS 01
DS 02
DS 03
DS 04

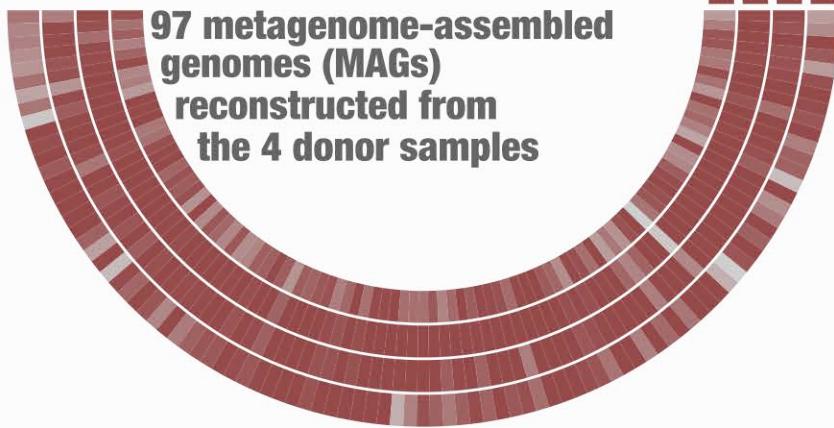
**97 metagenome-assembled
genomes (MAGs)
reconstructed from
the 4 donor samples**



Donor R01

DS 01
DS 02
DS 03
DS 04

97 metagenome-assembled
genomes (MAGs)
reconstructed from
the 4 donor samples

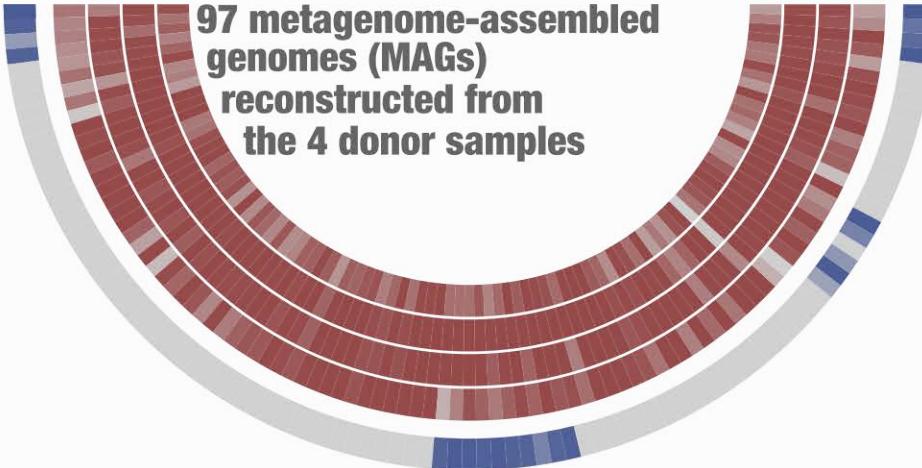


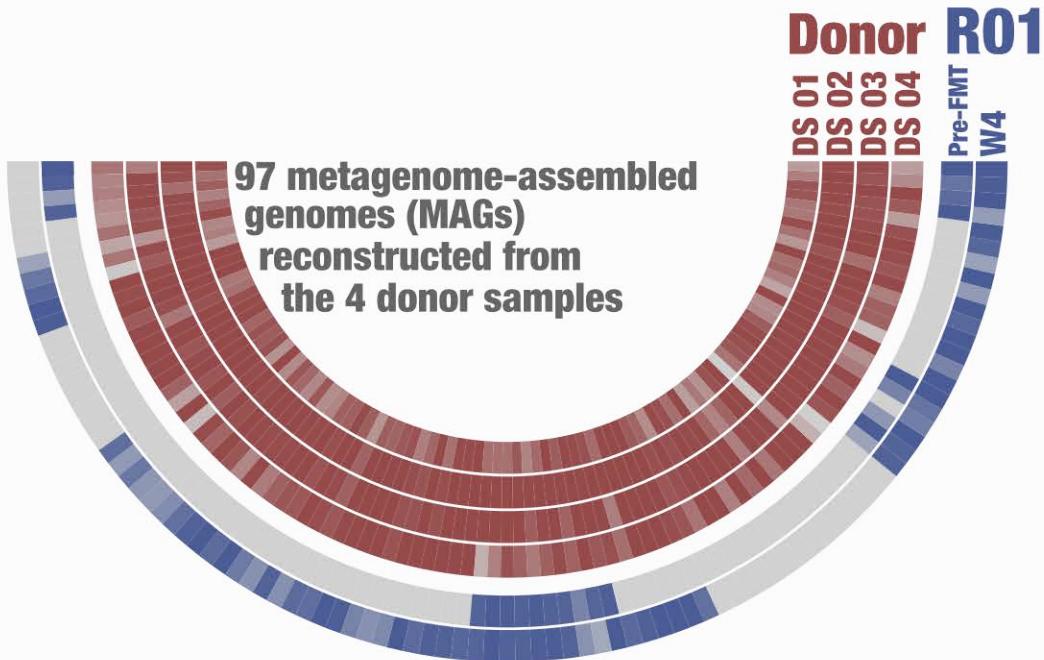
Donor R01

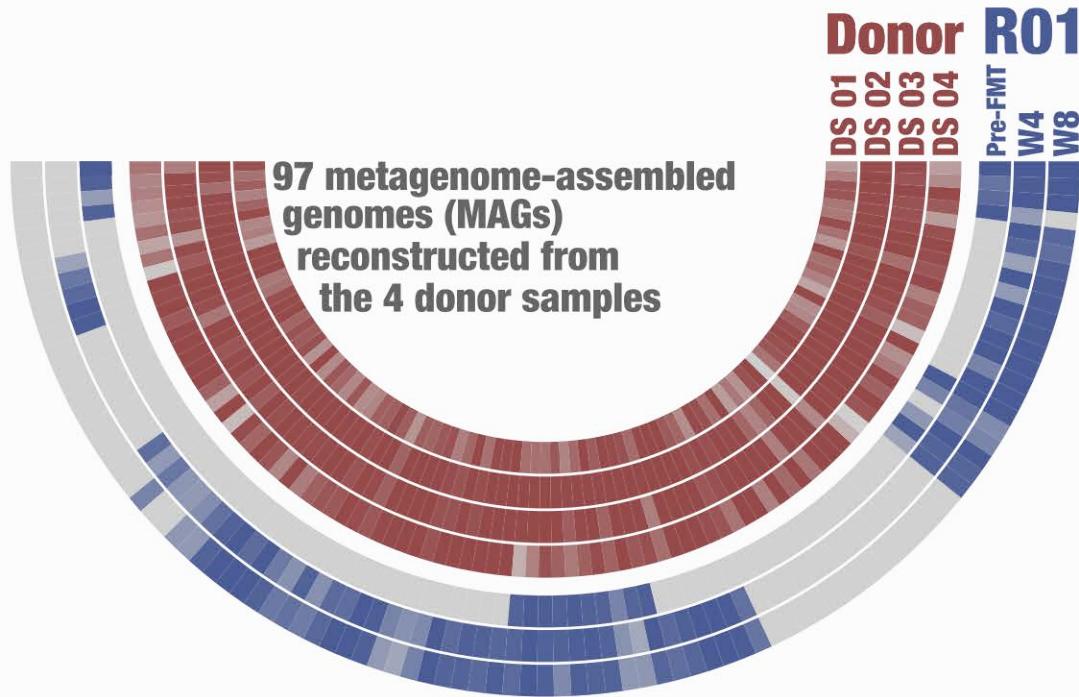
DS 01
DS 02
DS 03
DS 04

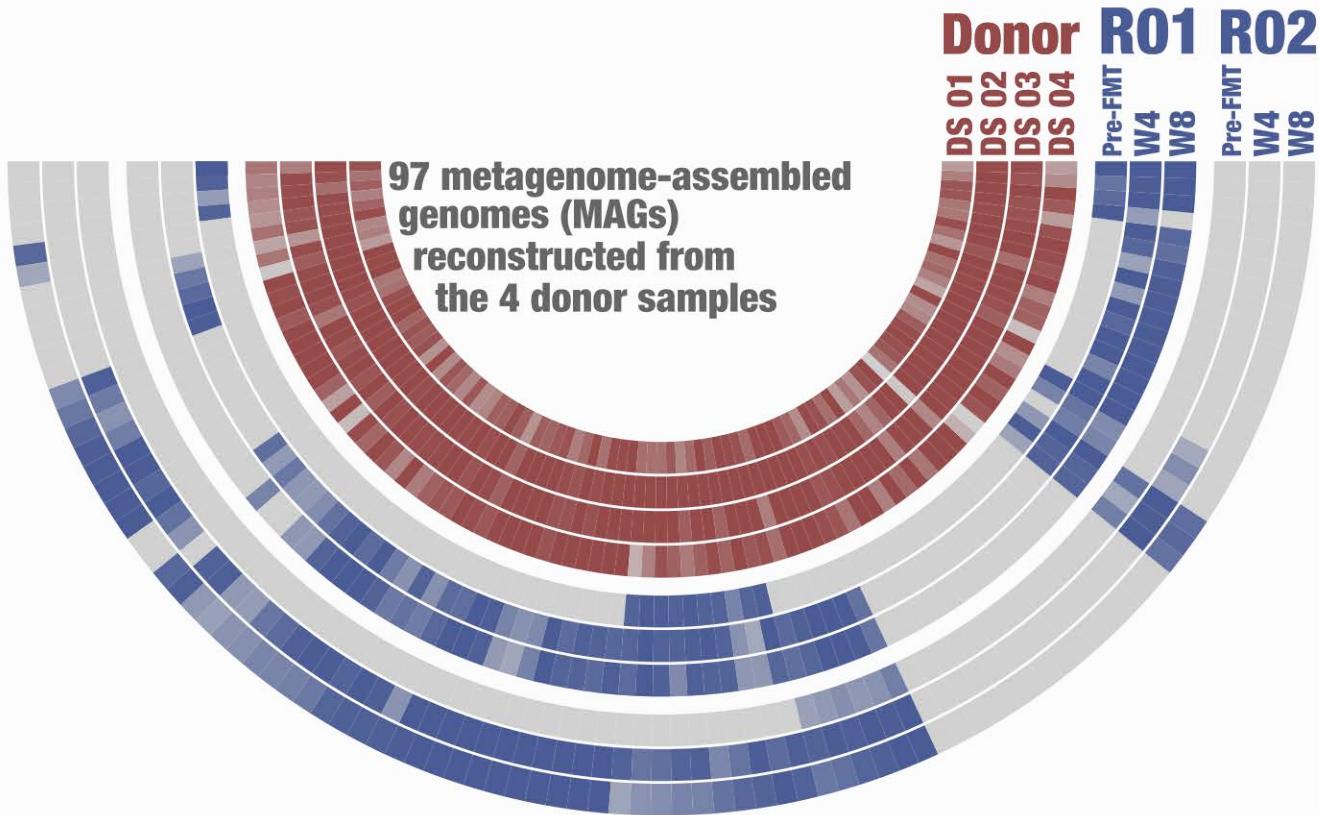
Pre-FMT

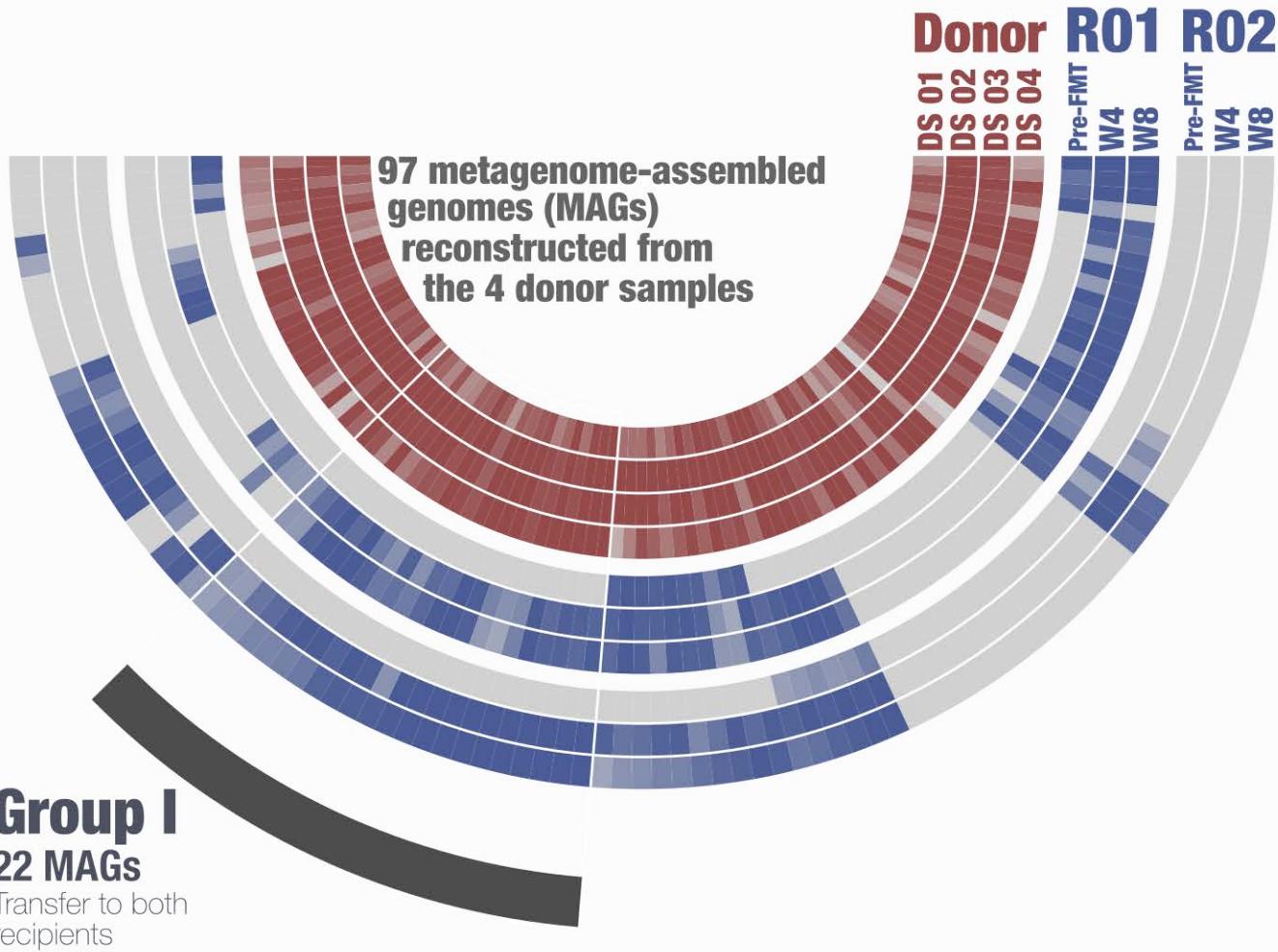
97 metagenome-assembled
genomes (MAGs)
reconstructed from
the 4 donor samples

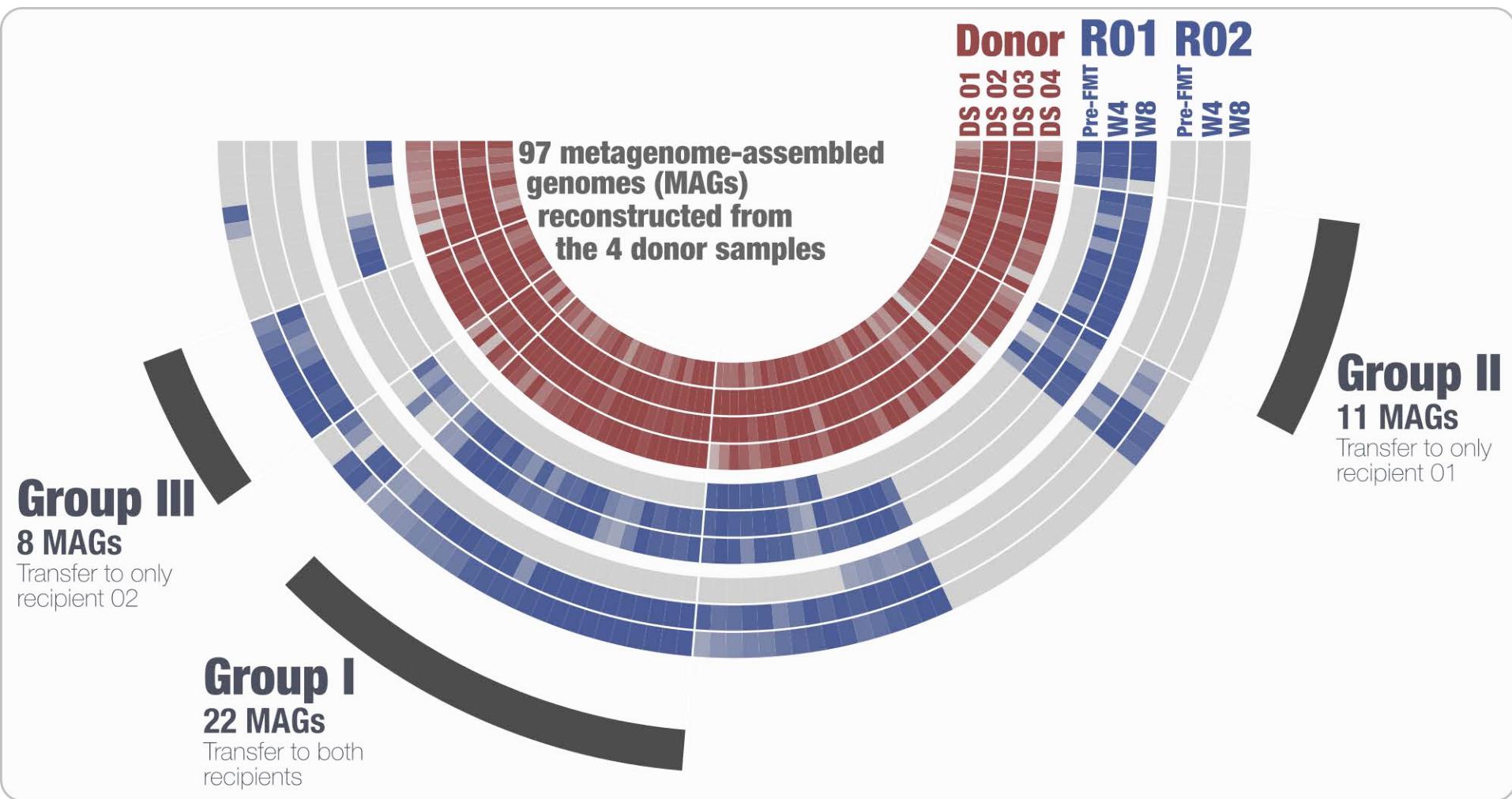


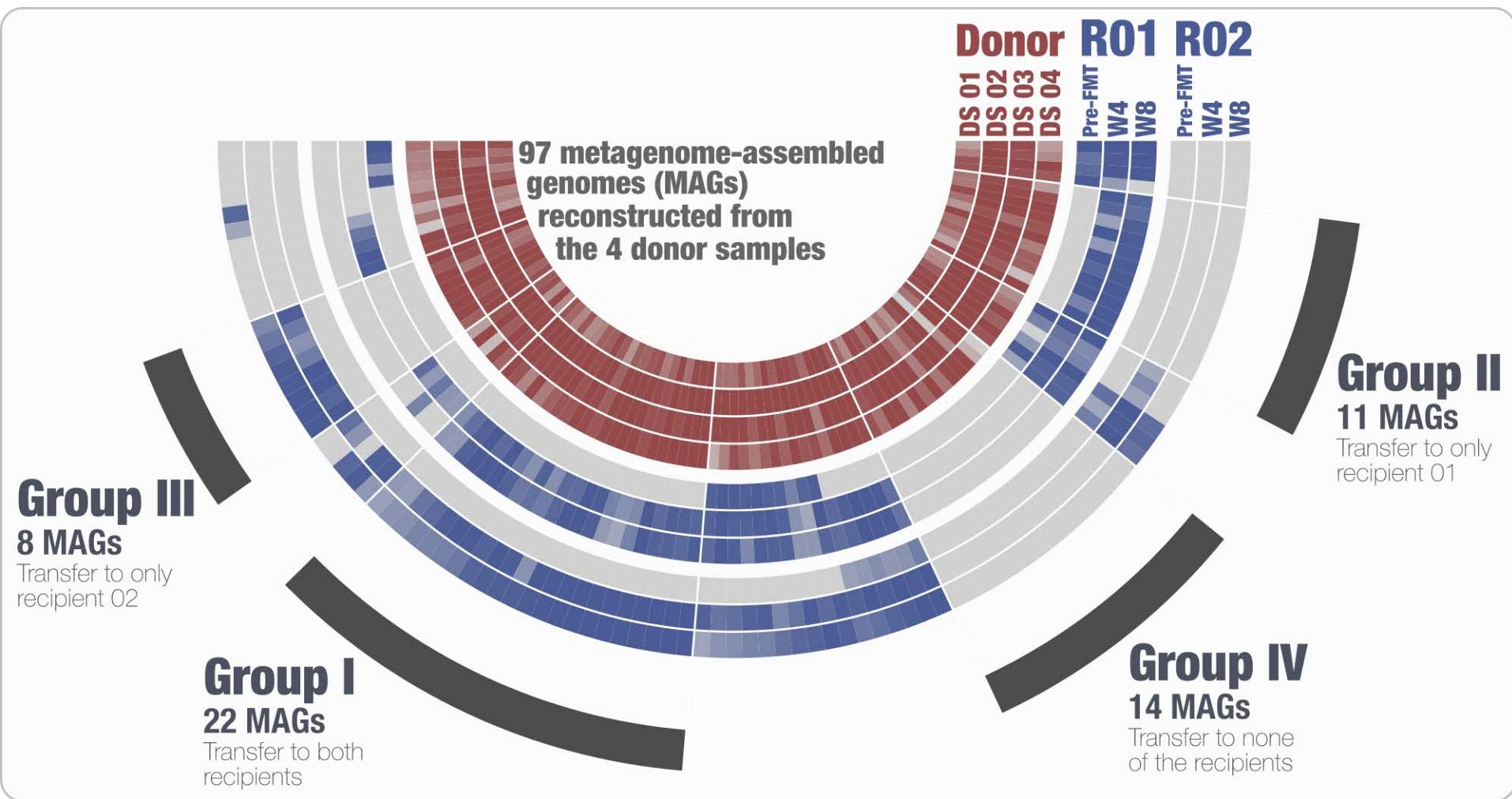


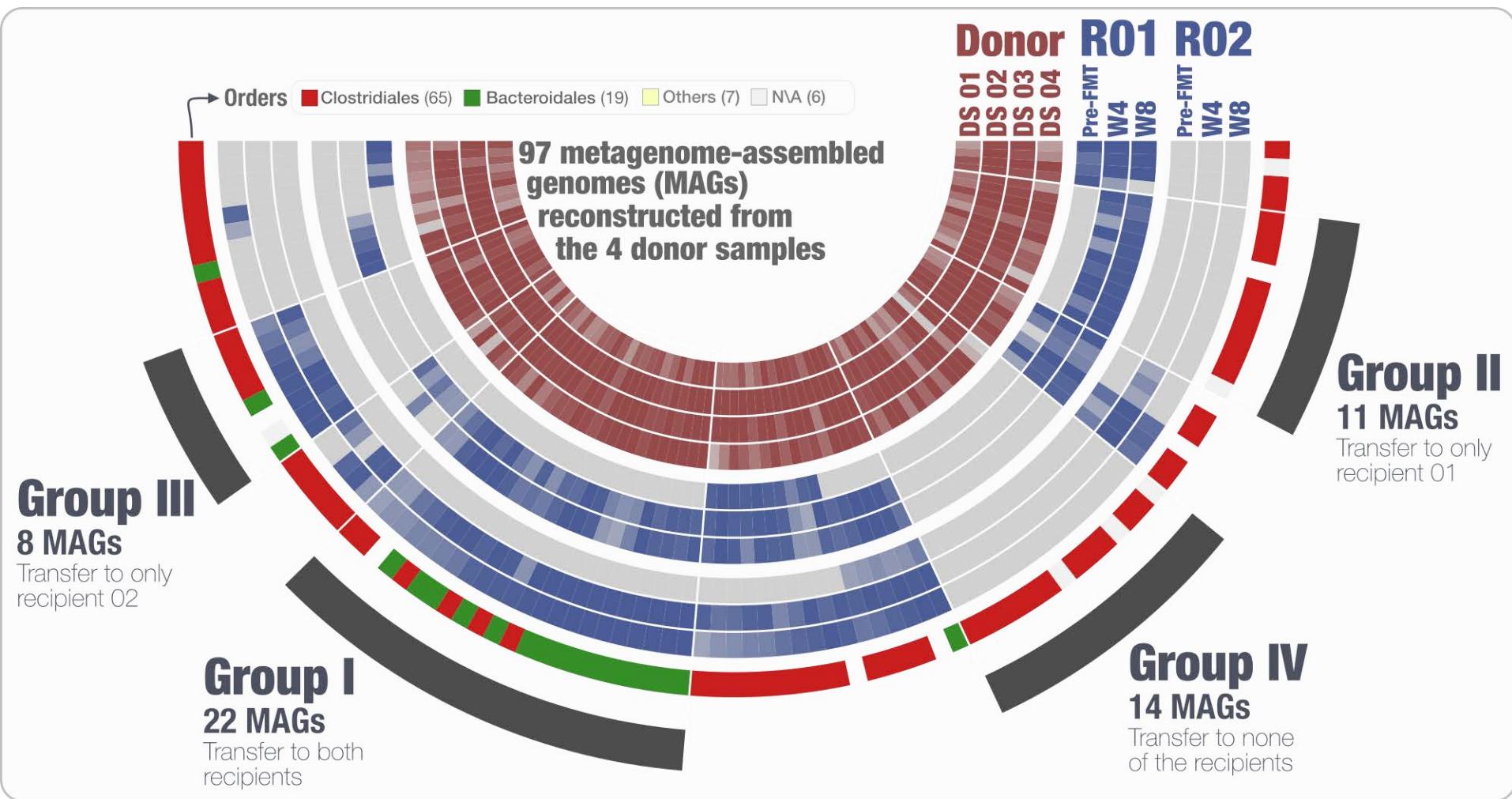


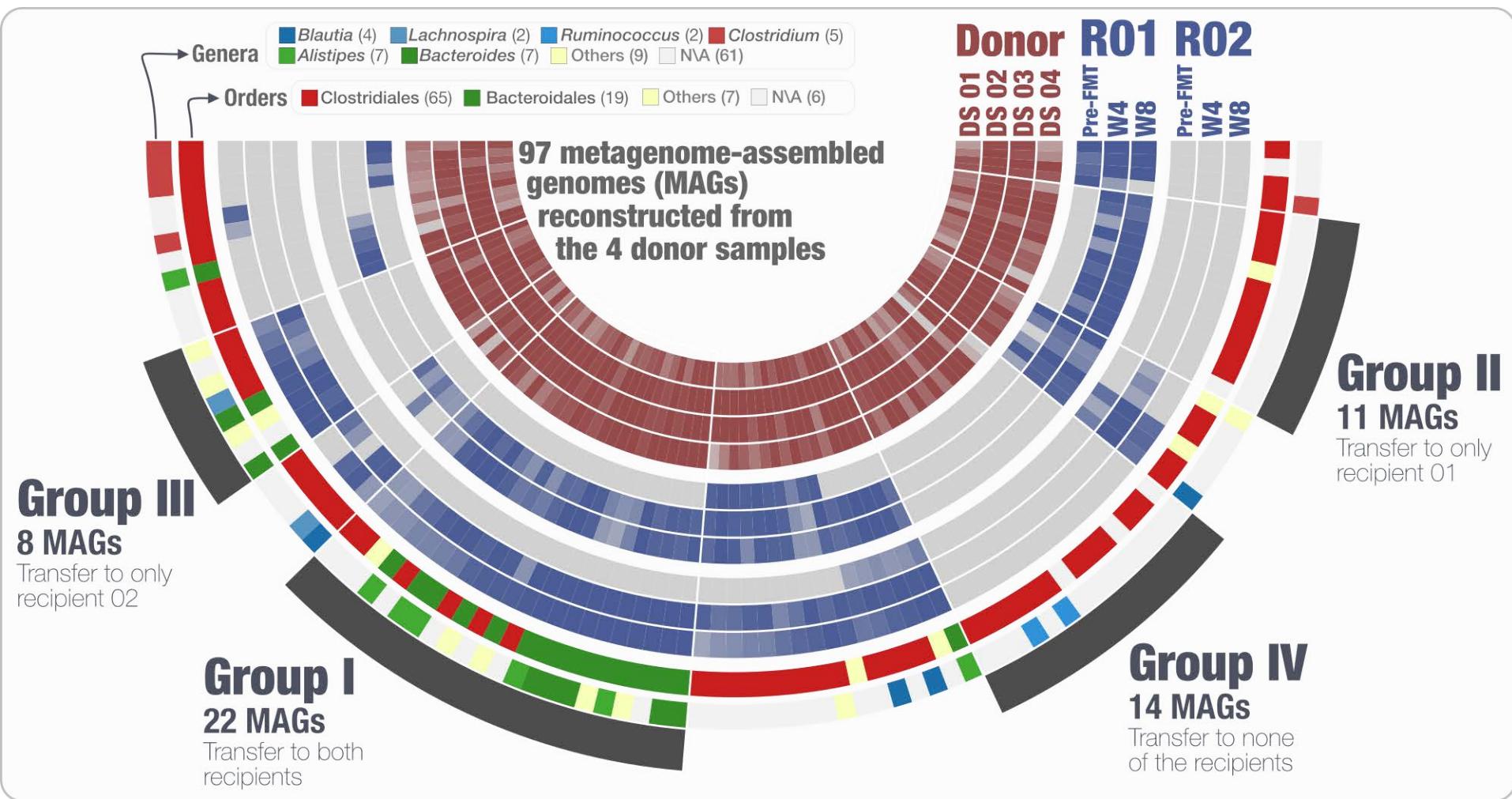






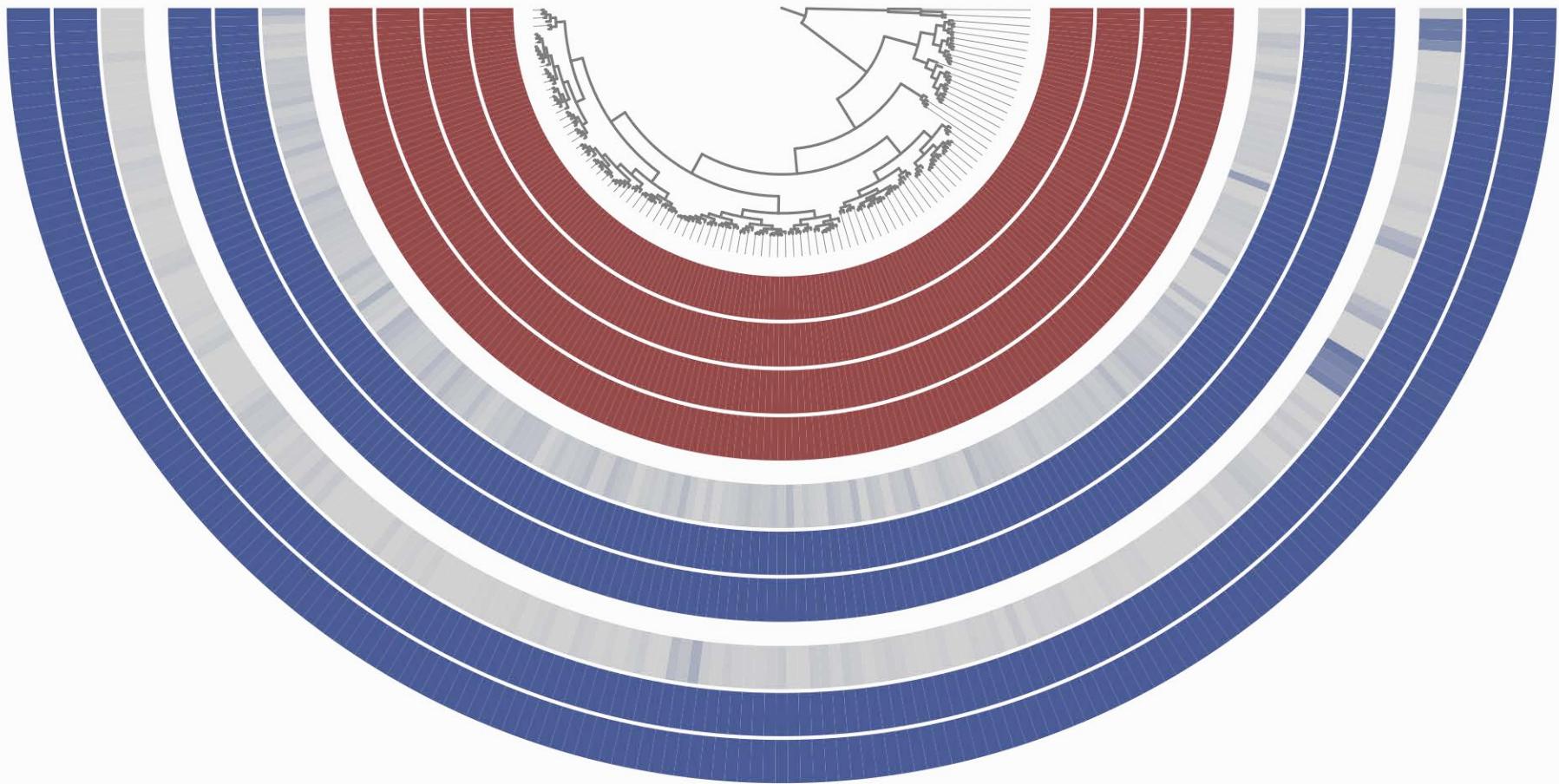






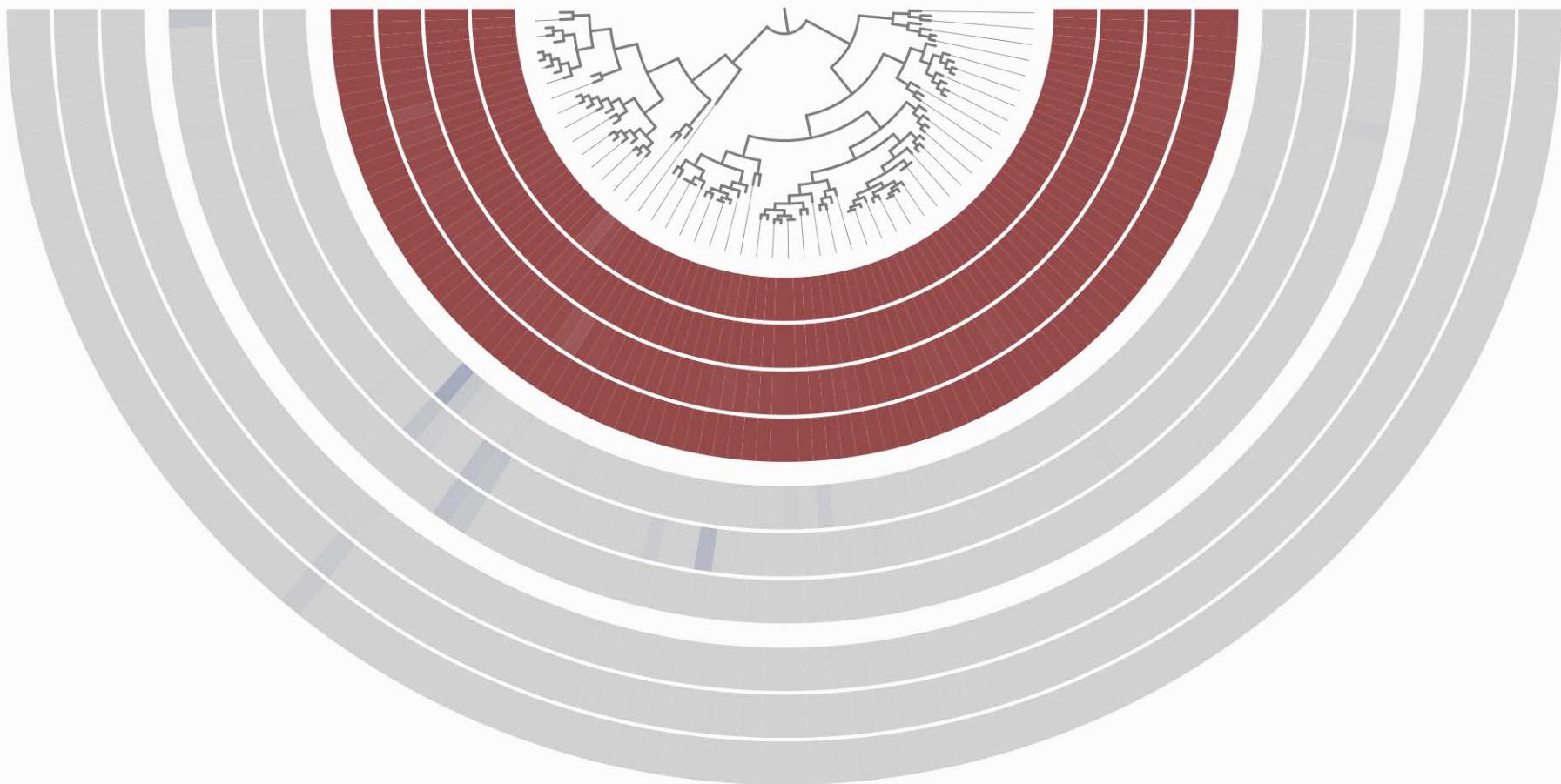
MAG 054 from Group I

2.0 Mbp, 191 contigs
Comp: 91%, Red: 2.1%



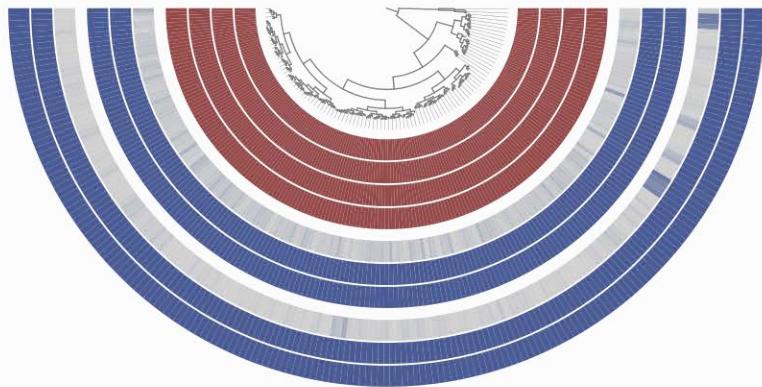
MAG 026 from Group IV

1.3 Mbp, 98 contigs
Comp: 91%, Red: 0.7%



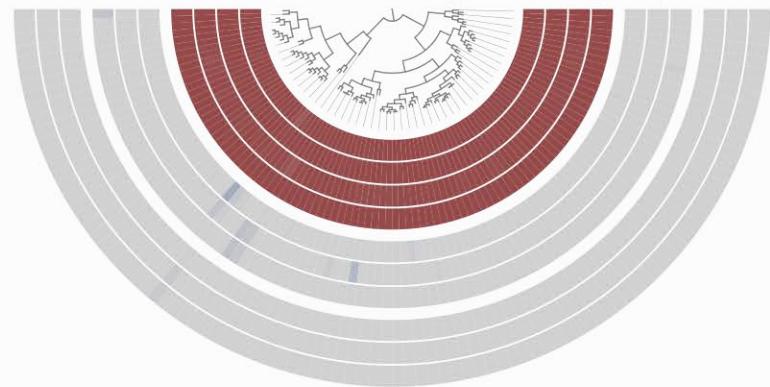
MAG 054 from Group I

2.0 Mbp, 191 contigs
Comp: 91%, Red: 2.1%



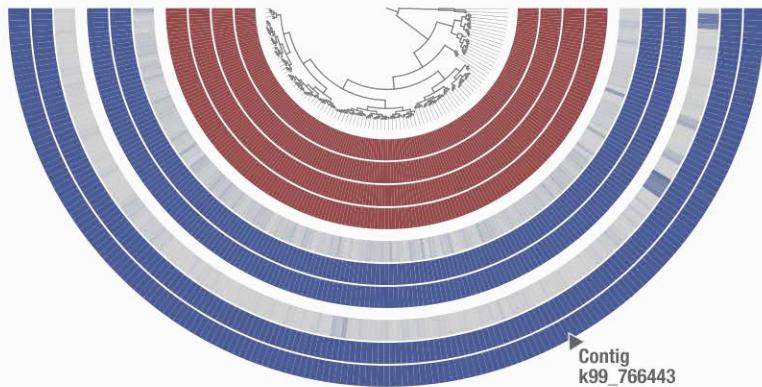
MAG 026 from Group IV

1.3 Mbp, 98 contigs
Comp: 91%, Red: 0.7%



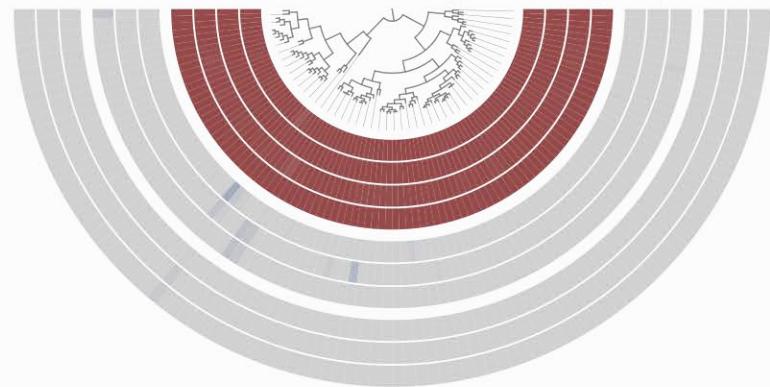
MAG 054 from Group I

2.0 Mbp, 191 contigs
Comp: 91%, Red: 2.1%



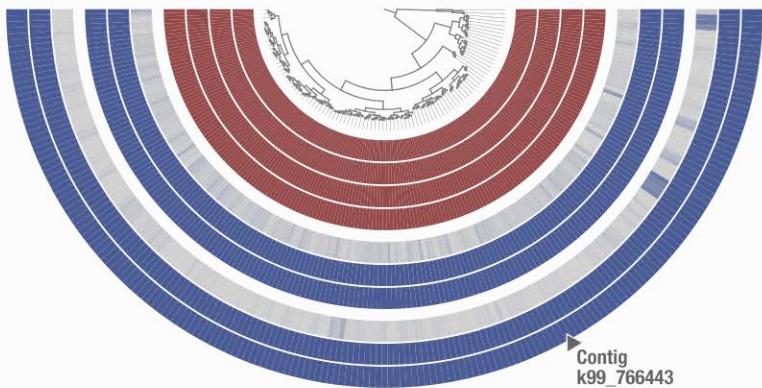
MAG 026 from Group IV

1.3 Mbp, 98 contigs
Comp: 91%, Red: 0.7%



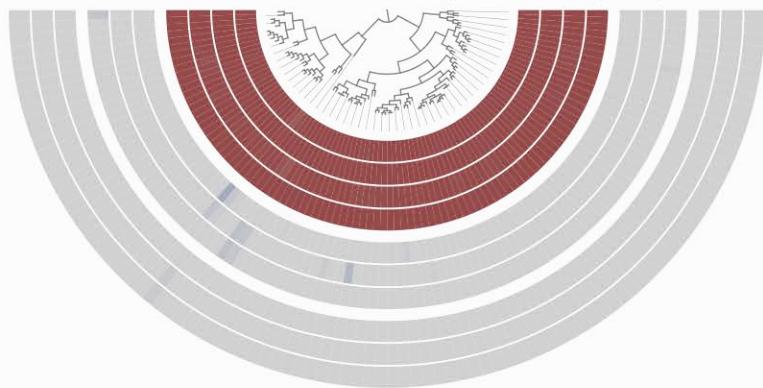
MAG 054 from Group I

2.0 Mbp, 191 contigs
Comp: 91%, Red: 2.1%



MAG 026 from Group IV

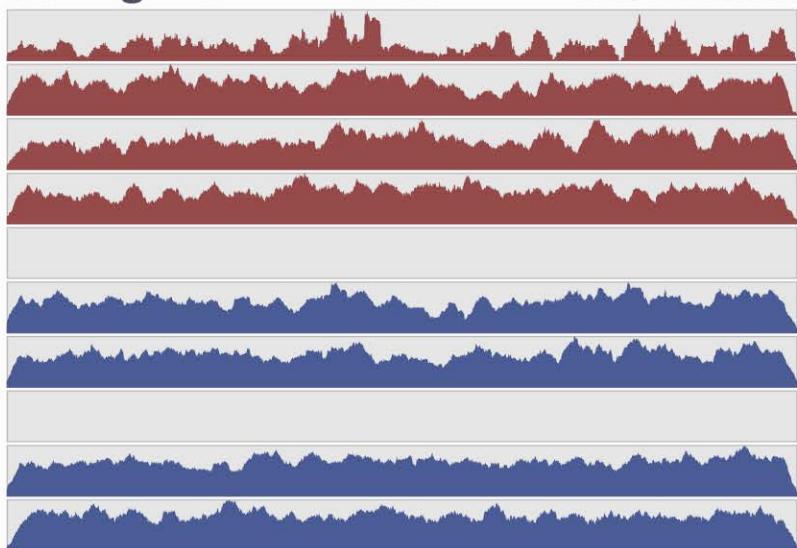
1.3 Mbp, 98 contigs
Comp: 91%, Red: 0.7%



Contig from MAG 054

7,292 nts
Contig ID: k99_766443

Coverage at the short read-level



DS 01

DS 02

DS 03

DS 04

Pre-FMT

W4

W8

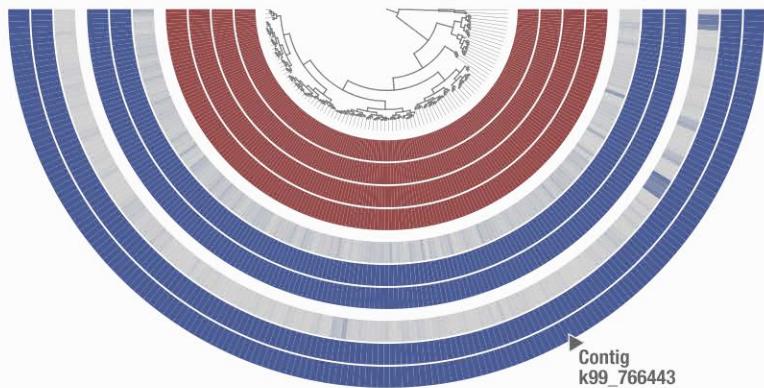
Pre-FMT

W4

W8

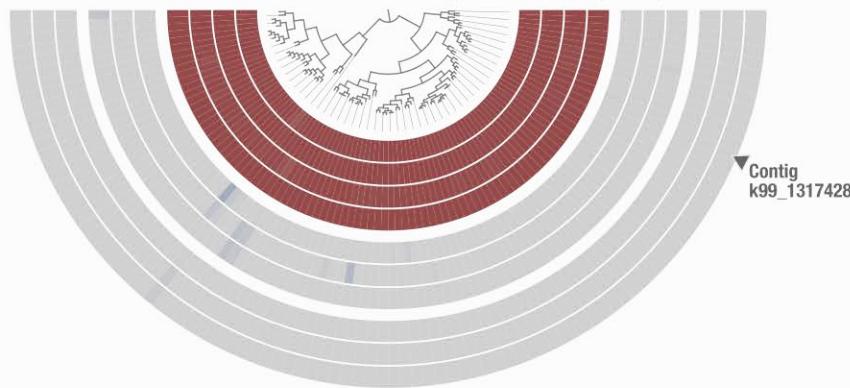
MAG 054 from Group I

2.0 Mbp, 191 contigs
Comp: 91%, Red: 2.1%



MAG 026 from Group IV

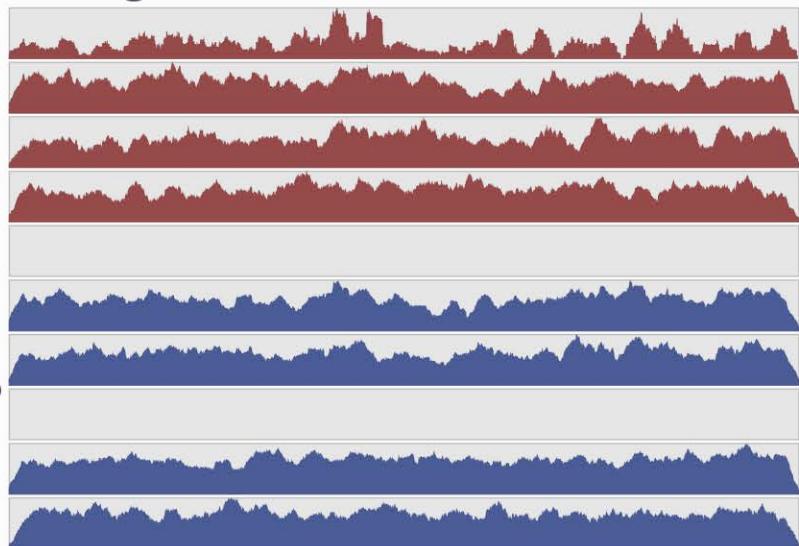
1.3 Mbp, 98 contigs
Comp: 91%, Red: 0.7%



Contig from MAG 054

7,292 nts
Contig ID: k99_766443

Coverage at the short read-level



DS 01

DS 02

DS 03

DS 04

Pre-FMT

W4

W8

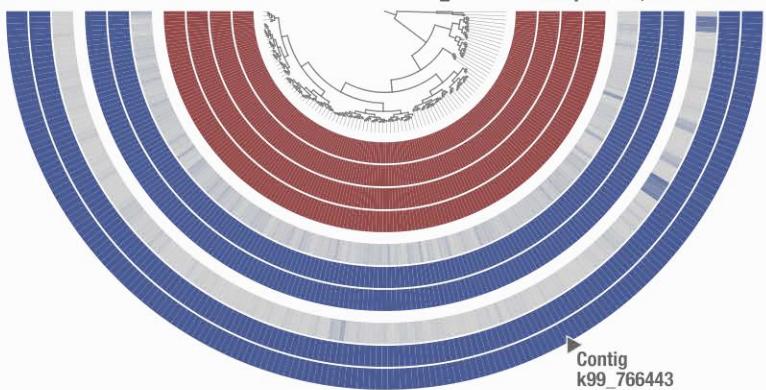
Pre-FMT

W4

W8

MAG 054 from Group I

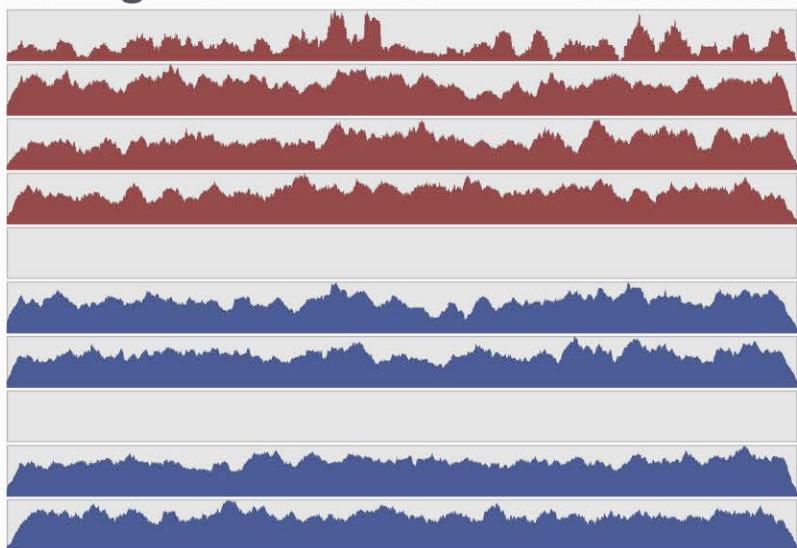
2.0 Mbp, 191 contigs
Comp: 91%, Red: 2.1%



Contig from MAG 054

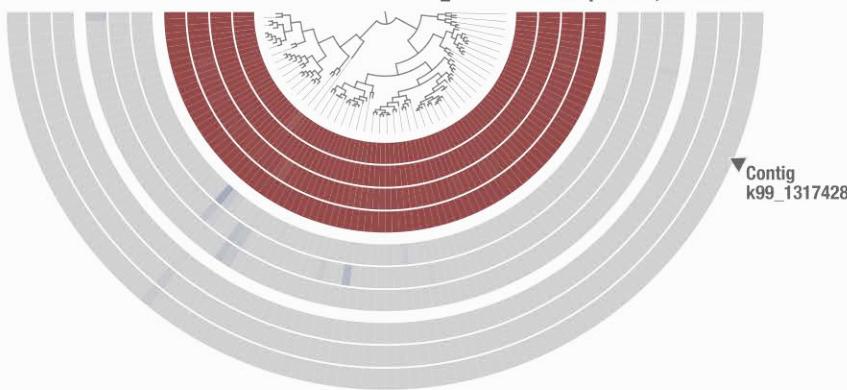
7,292 nts
Contig ID: k99_766443

Coverage at the short read-level



MAG 026 from Group IV

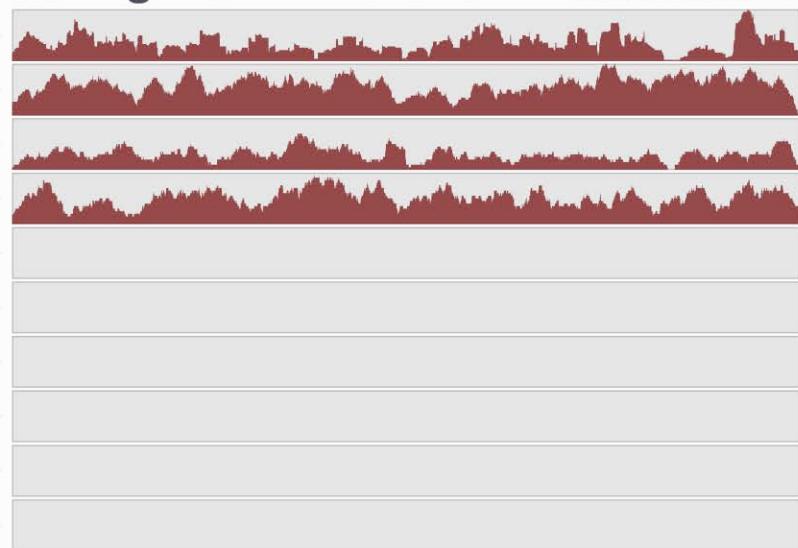
1.3 Mbp, 98 contigs
Comp: 91%, Red: 0.7%

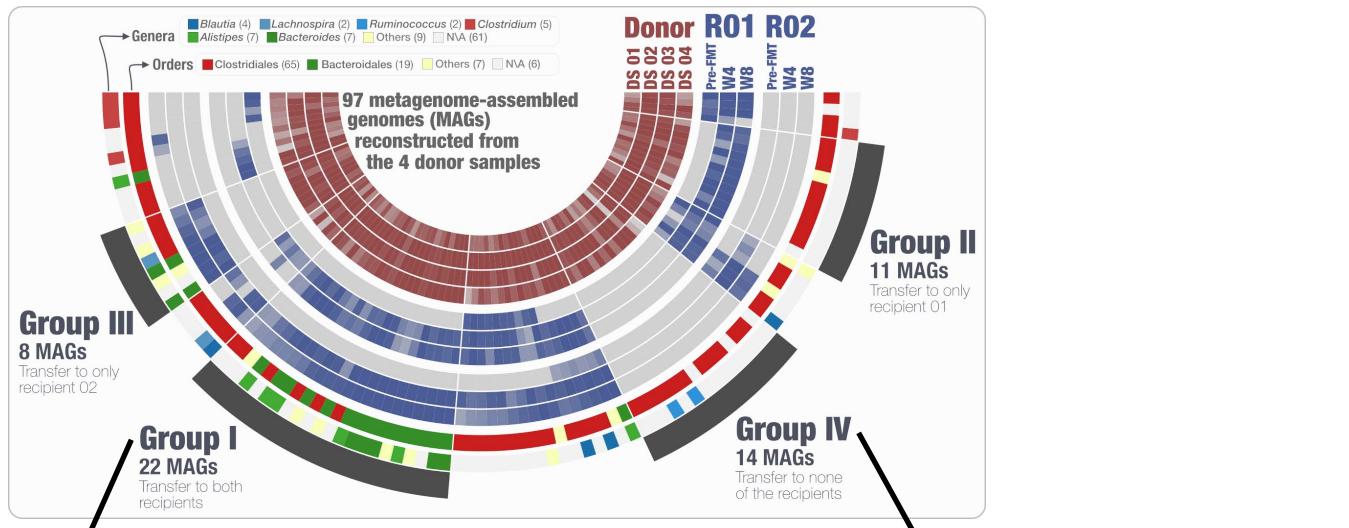


Contig from MAG 026

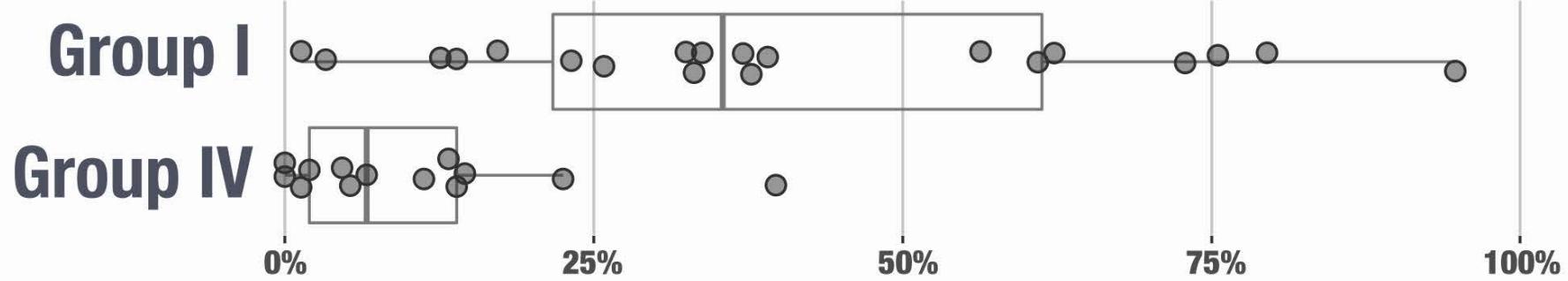
5,942 nts
Contig ID: k99_1317428

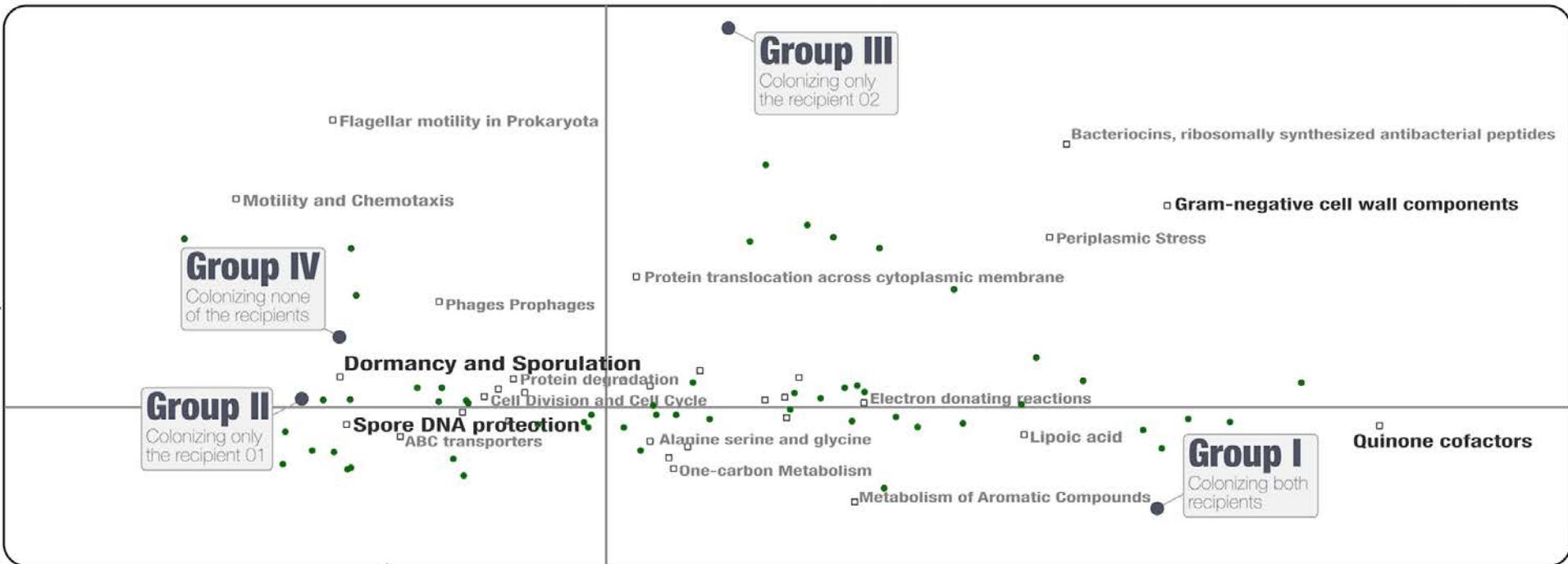
DS 01
DS 02
DS 03
DS 04
Pre-FMT
W4
W8
Pre-FMT
W4
W8





Detection of MAGs across 151 HMP gut metagenomes





**GENOME
RESEARCH**

[HOME](#) | [ABOUT](#) | [ARCHIVE](#) | [SUBMIT](#) | [SUBSCRIBE](#) | [ADVEI](#)

Institution: UNIV OF CHICAGO LIB



[Home](#) | [Current issue](#) | [News & comment](#) | [Research](#) | [Archive](#) ▾ | [Authors & referees](#) ▾ | [About the journal](#)

[home](#) ▶ [archive](#) ▶ [issue](#) ▶ [news](#) ▶ [full text](#)

NATURE BIOTECHNOLOGY | NEWS



An integrated metagenomics p strain profiling reveals novel p bacterial transmission and bio

Stephen Nayfach^{1,2}, Beltran Rodriguez-Mueller², Nandita
Katherine S. Pollard^{1,2,3}

Seres's pioneering microbiome drug fails mid-stage trial

Mark Ratner

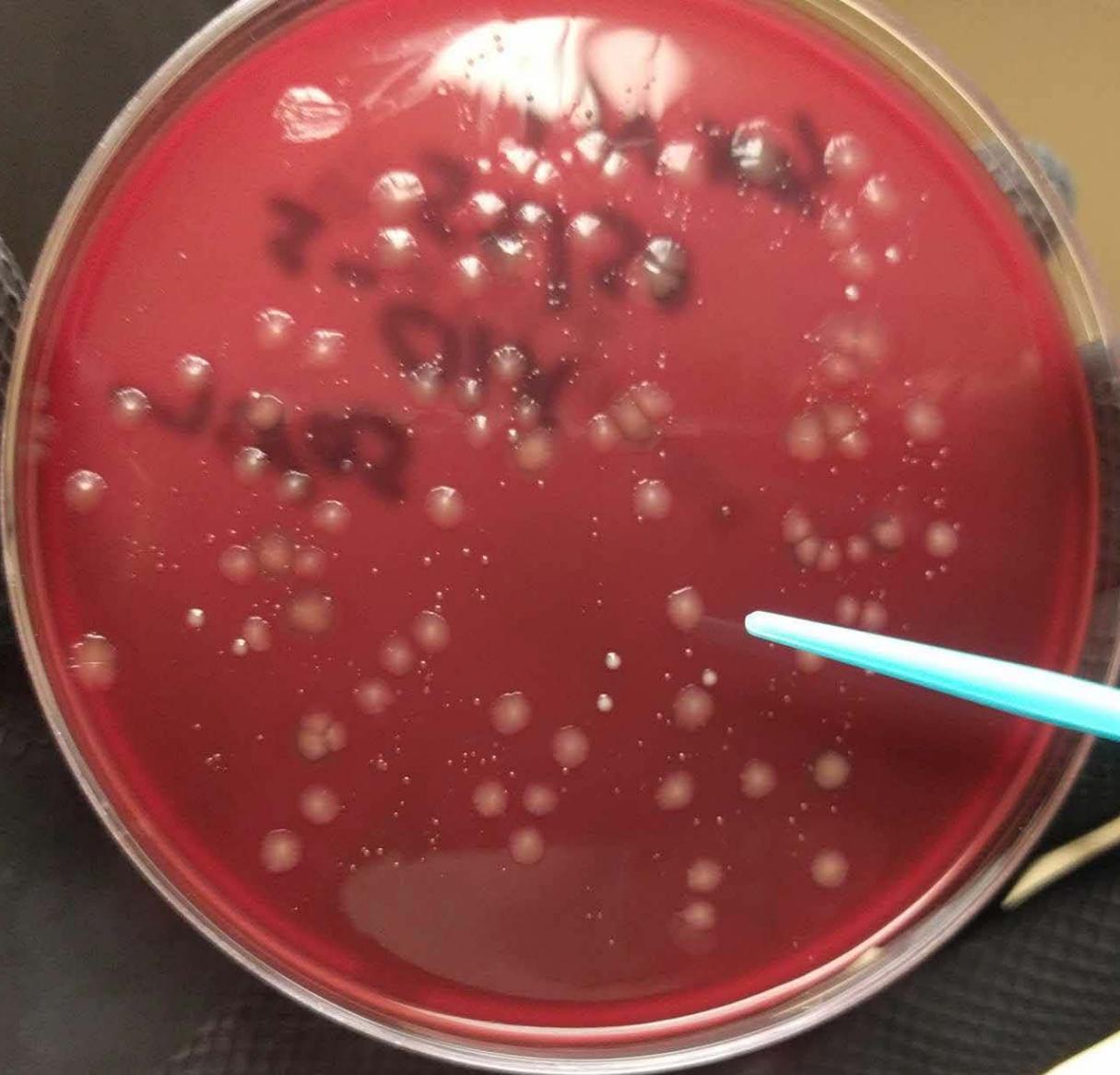
Nature Biotechnology 34, 1004–1005 (2016) | doi:10.1038/nbt1016-1004b

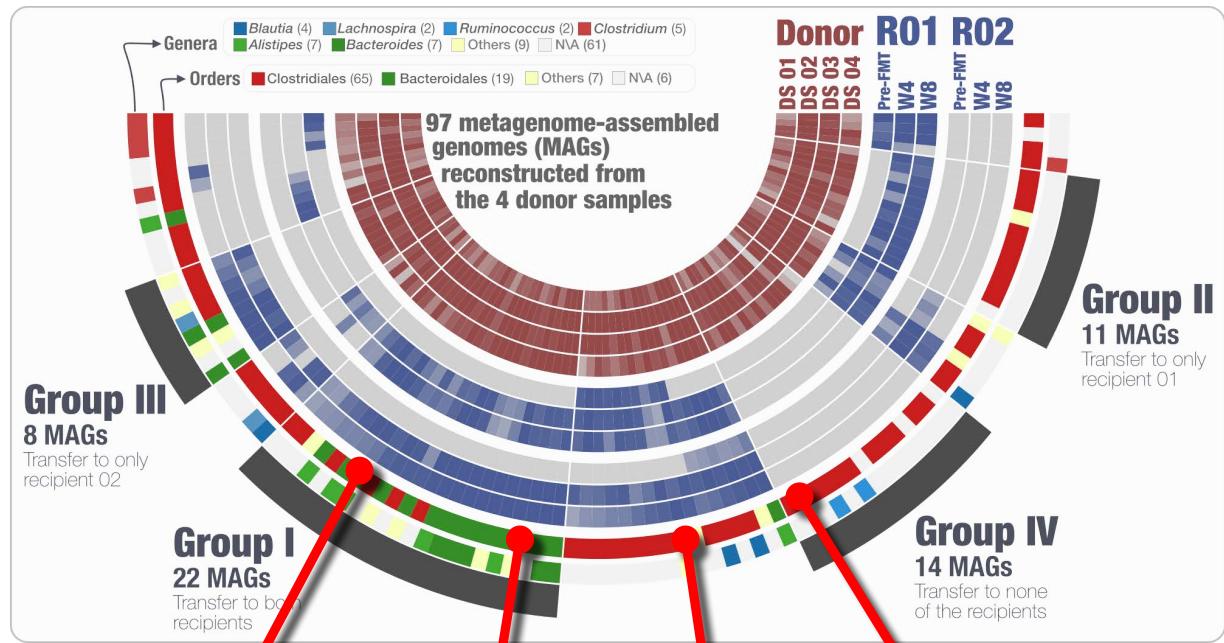
Published online 11 October 2016

SOMMY

XO-6

06/14/17





Thanks!



Tom O. Delmont :: Post-doctoral Scientist

[✉ Email](#) [🐦 Twitter](#)

📍 Address: Knapp Center for Biomedical Discovery, 900 E. 57th St., MB 9, RM 9148, Chicago, IL 60637 USA

I am a microbial ecologist who seeks to understand small and large-scale ecological patterns in the microbial world. My most recent investigations focused on humans, pristine soils, and the Southern Ocean.



Sonny T. M. Lee :: Post-doctoral Scientist

[✉ Email](#)

📍 Address: Knapp Center for Biomedical Discovery, 900 E. 57th St., MB 9, RM 9148, Chicago, IL 60637 USA

I am a microbiologist studying the most interesting inhabitants of humans, coral reefs, and the ocean, and using big data to answer vexing problems of our future.



Alon Shaiber :: Graduate Student

[✉ Email](#) [🐙 Github](#)

📍 Address: Knapp Center for Biomedical Discovery, 900 E. 57th St., MB 9, RM 9148, Chicago, IL 60637 USA

I'm a graduate student at the Bio Physical Sciences Program at UChicago. I'm fascinated by the complexity of microbial ecology and interested in applying and developing computational tools to explore it.



Özcan C. Esen (Software guru, local hacker)

[HomeAs](#) [✉ Email](#) [🐦 Twitter](#) [🐙 Github](#)

📍 Address: Knapp Center for Biomedical Discovery, 900 E. 57th St., MB 9, RM 9148, Chicago, IL 60637 USA

As a computer engineer I strive to utilize my expertise to investigate challenges in microbial ecology, metagenomics, and visualization of complex data.

<http://merenlab.org>

