Identifying bacteriophage determinants of microbial community dynamics and function

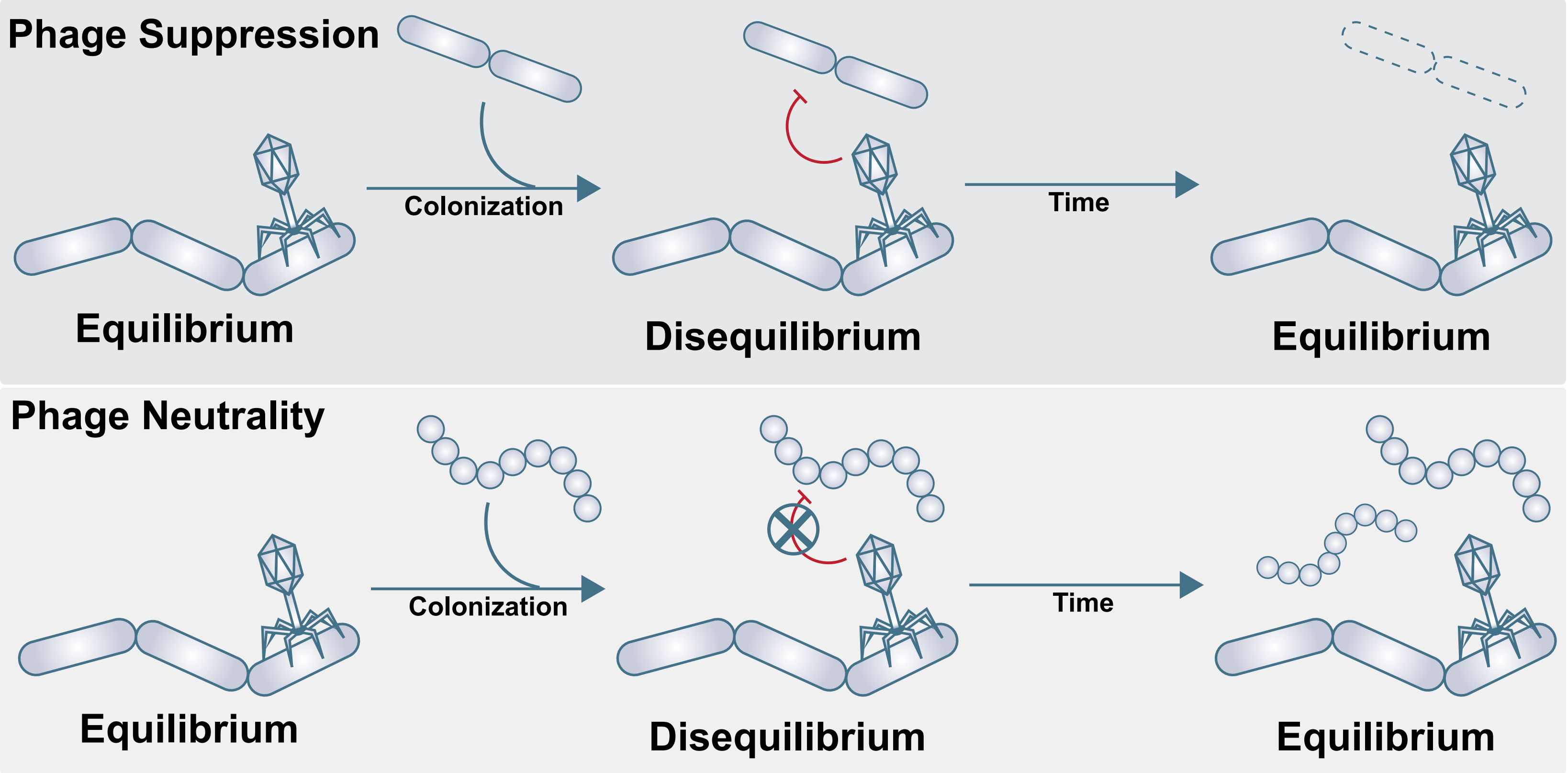
Jack T Sumner[1,2], Jordan Jensen[1,2], Jiaxian Shen[1,3], Long Nguyen[1,3], Eric A Franzosa[1,2],
Kelsey N Thompson[1,2], Curtis Huttenhower[1,2]
[1]Harvard T.H. Chan School of Public Health, [2]Broad Institute of MIT and Harvard, [3]Massachusetts General Hospital

BROAD INSTITUTE

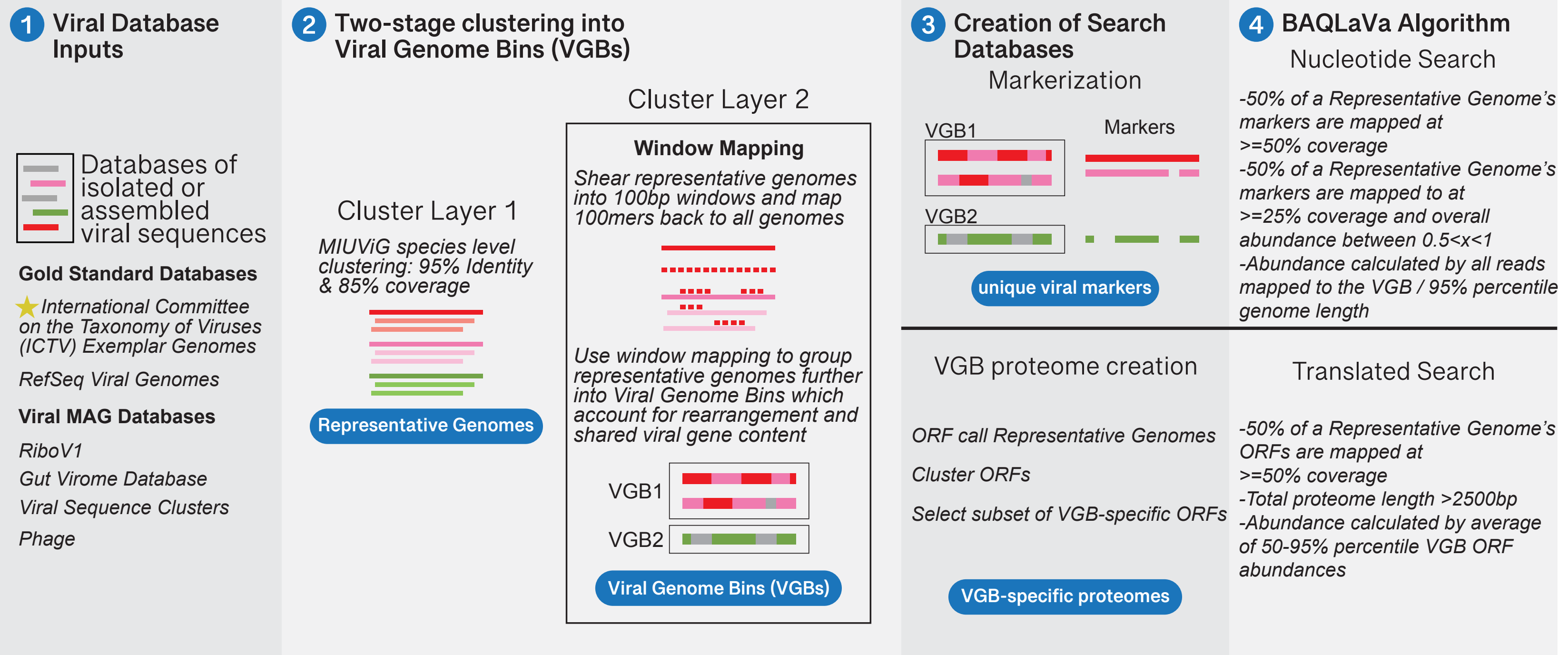
THE HARVARD CHAN MICROBIOME IN PUBLIC HEALTH CENTER

Bacteriophages, the viruses that infect bacteria, are the most abundant biological entities on earth. Despite this, little is known about their ecology and community-level dynamics relative to their bacterial counterparts. Recent bioinformatics advances facilitate previously-infeasible viral community profiling from large-scale shotgun metagenomic and metatranscriptomic sequencing data, providing a new avenue to assess the impact of viral dynamics and function on the microbiome. To that end, we leveraged existing meta'omics datasets (i.e. HMP2) and developments in viral community profiling to identify bacteriophage determinants of microbial ecology and function. By integrating functional metagenomics with phage-host prediction, we provide evidence that promiscuous phage, i.e. those with a broad host range, are abundant in the enteric human microbiome and associate with microbiota functional shifts in a manner consistent with phage-mediated selection and transcriptomic response to phage infection. Moreover, we posit that phage promiscuity acts as an emergent property to constrain and stabilize microbiome community structure, specifically acting to prevent successful colonization of immigrating competitors to a phage's primary bacterial host at the community level. Overall, this work highlights the active role bacteriophages play in shaping microbial communities, and provides a systems-level framework for determining viral drivers of microbial functions. Over the longer term, such efforts contribute to our basic understanding of the complex ecological dynamics driving microbial community stability and composition as well as optimal bacteriophage selection for precision control of the microbiome as a novel therapeutic intervention.

What role do bacteriophage play in shaping the microbiome?



A comprehensive viral database to expand biological discovery



① **Viral Database Inputs**

Databases of isolated or assembled viral sequences

Gold Standard Databases

★International Committee on the Taxonomy of Viruses (ICTV) Exemplar Genomes

RefSeq Viral Genomes

Viral MAG Databases

RiboV1

Gut Virome Database

Viral Sequence Clusters

Phage

② **Two-stage clustering into Viral Genome Bins (VGBs)**

Cluster Layer 1

MIUViG species level clustering: 95% Identity & 85% coverage

Representative Genomes

Cluster Layer 2

Window Mapping

Shear representative genomes into 100bp windows and map 100mers back to all genomes

Use window mapping to group representative genomes further into Viral Genome Bins which account for rearrangement and shared viral gene content

VGB1

VGB2

Viral Genome Bins (VGBs)

③ **Creation of Search Databases**

Markerization

VGB1

VGB2

Markers

unique viral markers

VGB proteome creation

ORF call Representative Genomes

Cluster ORFs

Select subset of VGB-specific ORFs

VGB-specific proteomes

④ **BAQLaVa Algorithm**

Nucleotide Search

-50% of a Representative Genome's markers are mapped at >=50% coverage

-50% of a Representative Genome's markers are mapped to at >=25% coverage and overall abundance between 0.5<x<1

-Abundance calculated by all reads mapped to the VGB / 95% percentile genome length

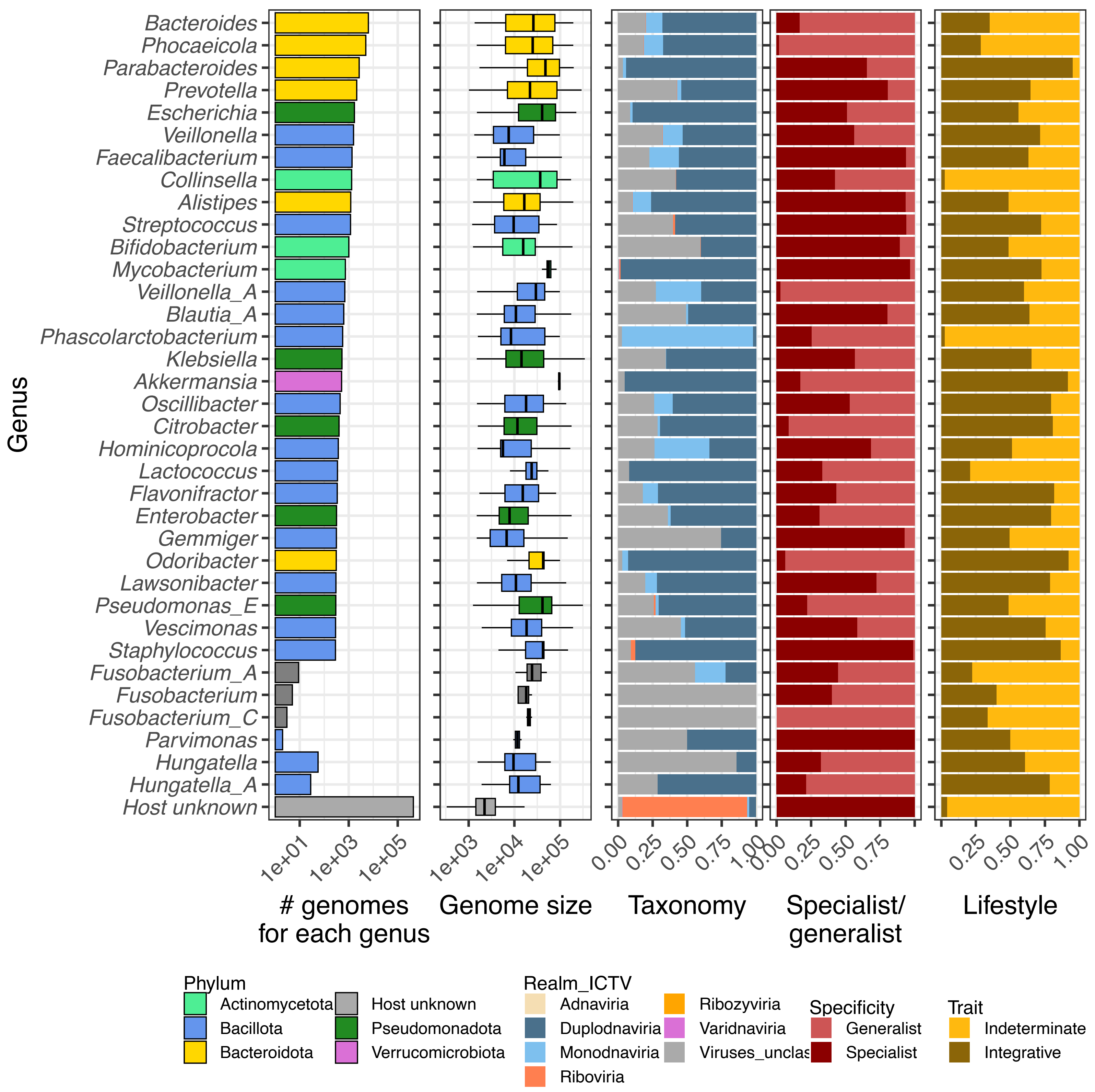
Translated Search

-50% of a Representative Genome's ORFs are mapped at >=50% coverage

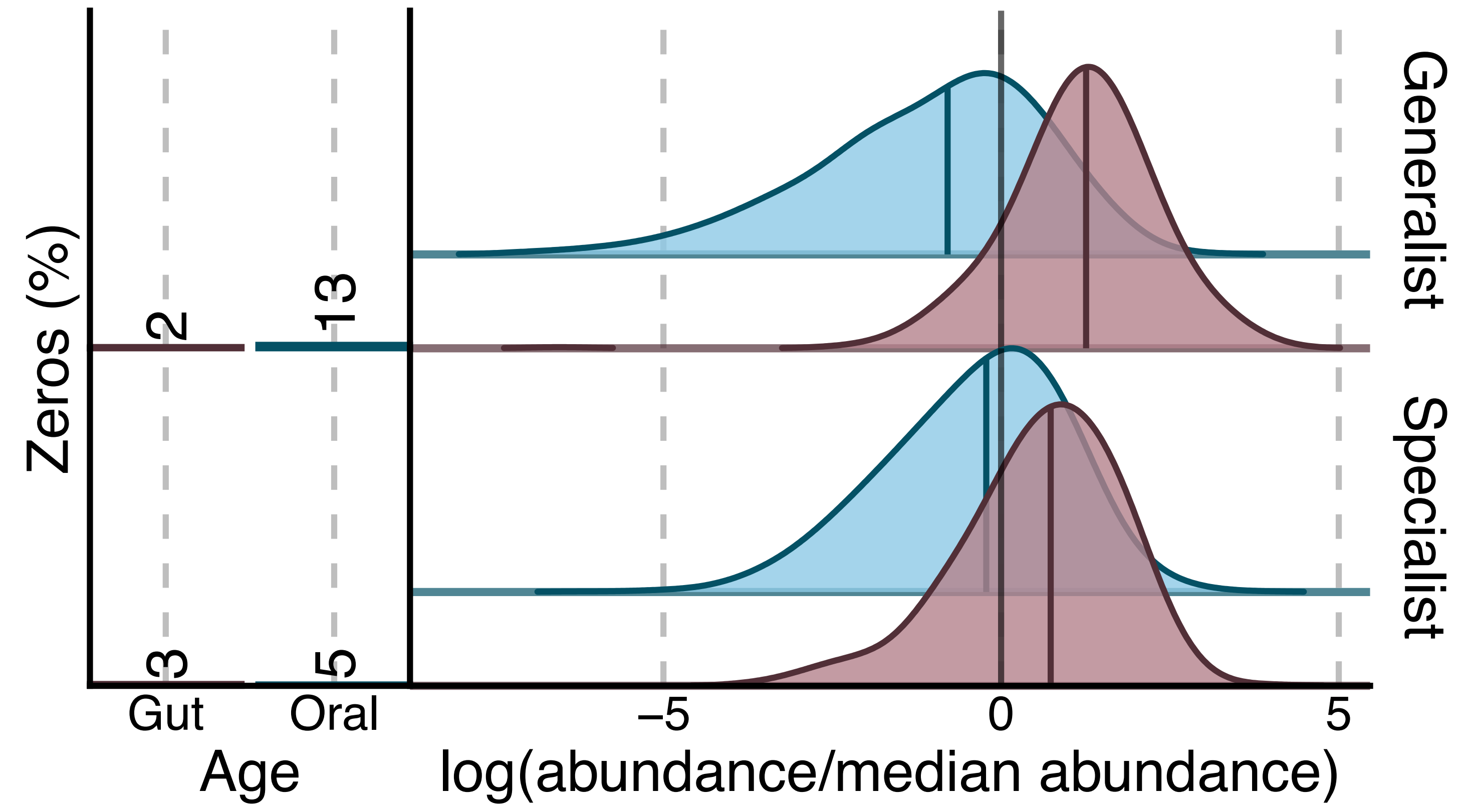
-Total proteome length >2500bp

-Abundance calculated by average of 50-95% percentile VGB ORF abundances

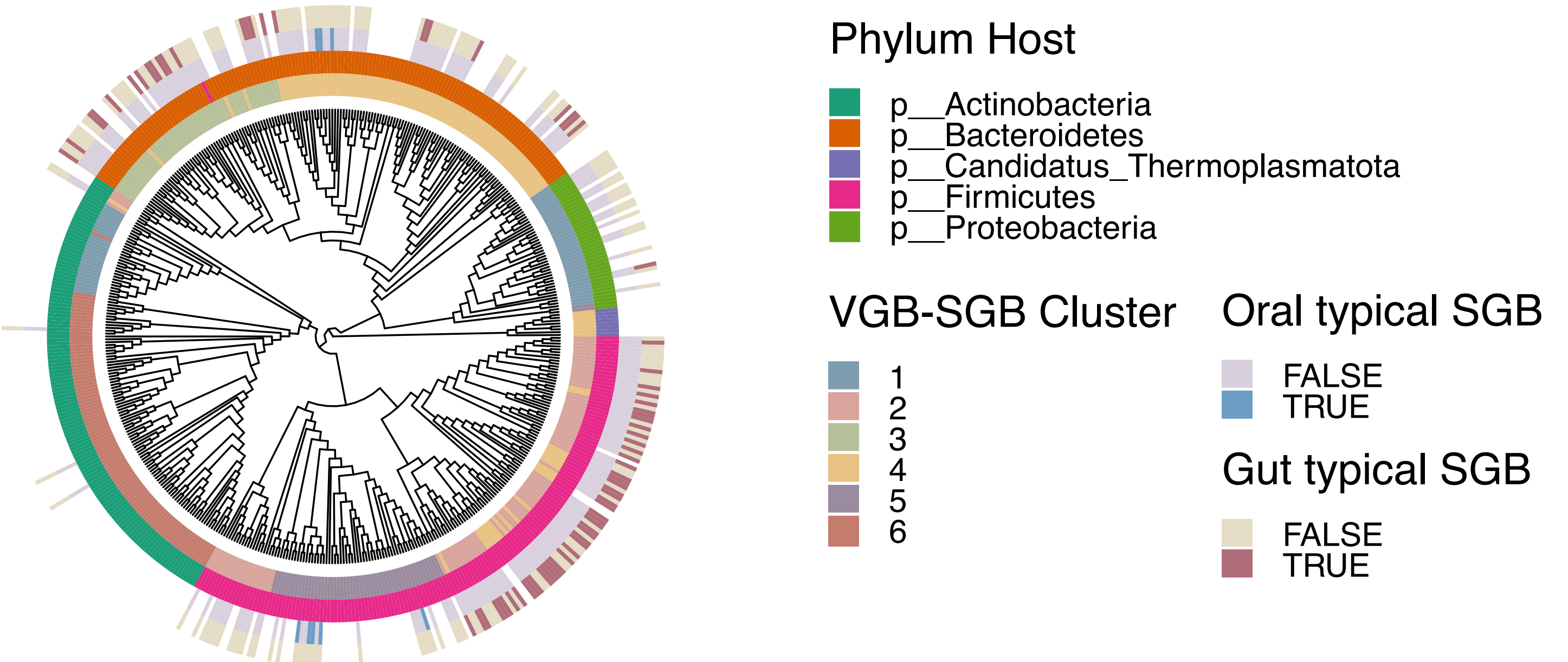
Bacteriophage traits vary across human-associated microbiota



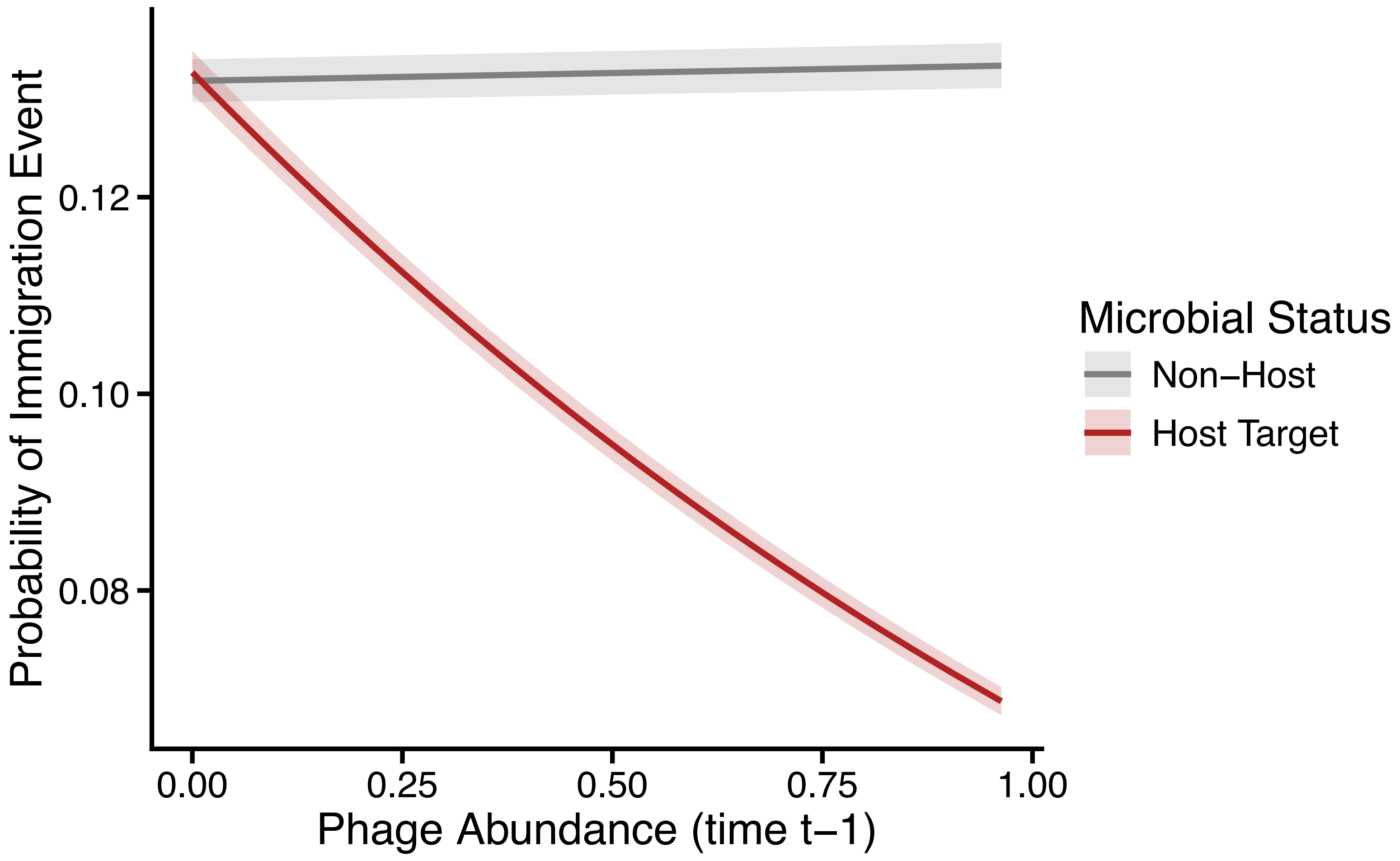
Promiscuous bacteriophage are abundant across the human body



Promiscuous phage clusters may infect related & unrelated hosts



Promiscuous phage suppress colonization of secondary hosts



Conclusions & future directions

- Promiscuous phage are abundant across the human microbiome, and generalist phage abundance differ between human body sites
- Temporal analysis reveals signals consistent with phage-mediated suppression of colonization by secondary hosts, indicating promiscuous phage act to constrain microbiota dynamics
- Further integration of orthogonal data-types, such as metaHiC and correlation analysis will improve robust prediction of phage-host pairs

Acknowledgments

We would like to acknowledge the use of the HMP2, HMP1-II cohorts within this analysis. Further, the computations in this paper were run in part on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

http://huttenhower.sph.harvard.edu