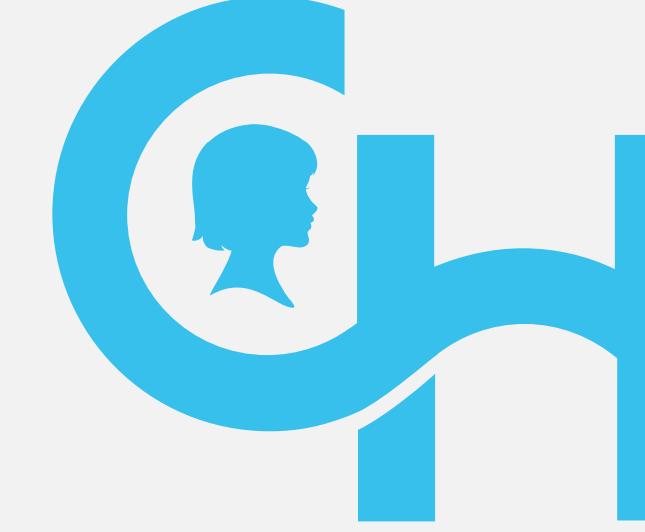




# Metagenomic MAGIC: Dissecting the Structure and Function of the Developing Gut Microbiome



Daniel Chang<sup>1</sup>, Suzie Hoops<sup>1</sup>, Dan Knights<sup>1,2,3</sup>  
<sup>1</sup>Department of Computer Science and Engineering; <sup>2</sup>Bioinformatics and Computational Biology Program; <sup>3</sup>Biotechnology Institute, University of Minnesota

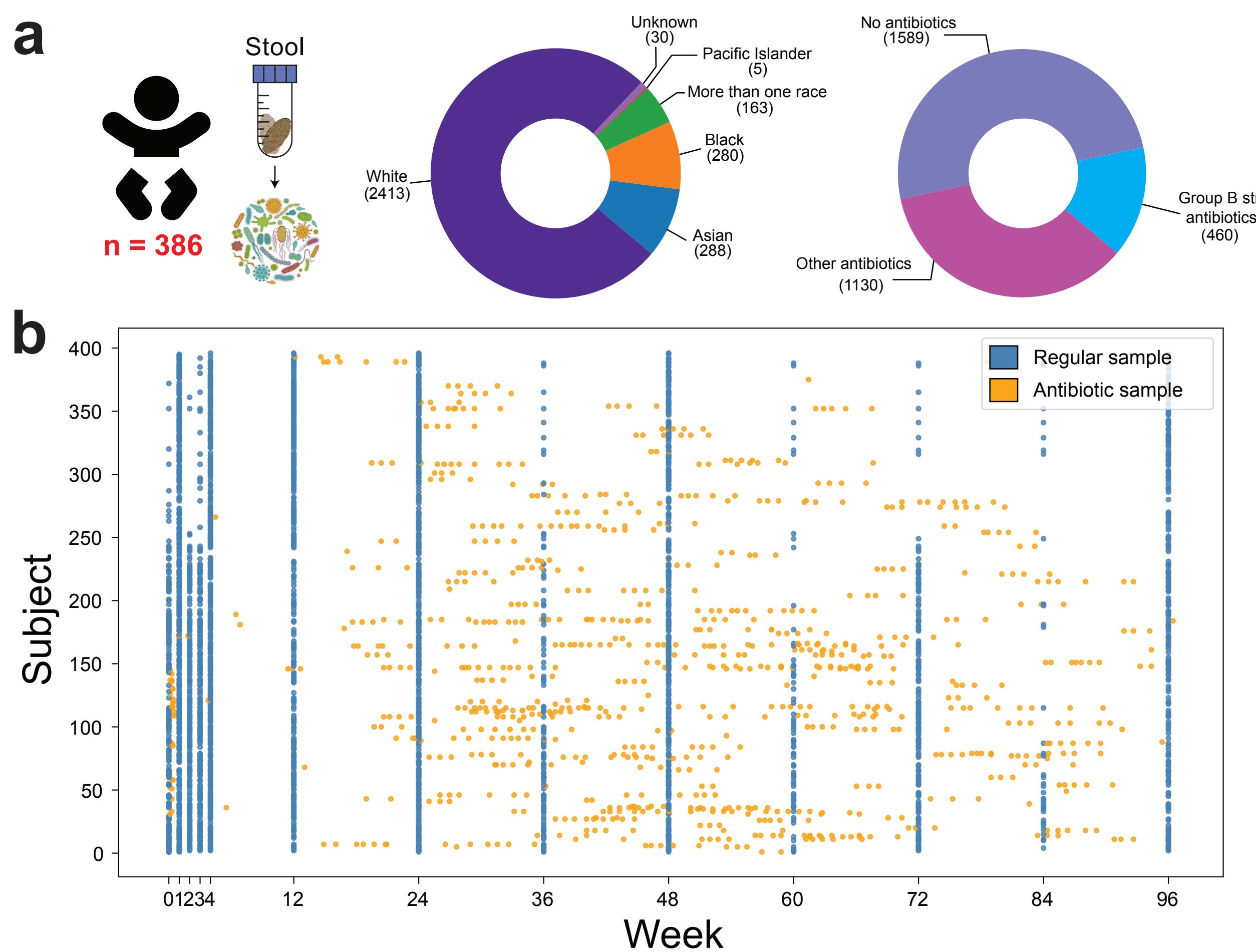
## Background

Perinatal and infant antibiotic exposures are common and have been linked to changes in the gut microbiome, which plays a central role in health and disease. Infants become colonized with trillions of bacteria in the first few hours of life, during which, their nascent immune system develops tolerance to commensal microbes. In this study, we assemble and follow the Microbiome, Antibiotics, and Growth Infant Cohort (MAGIC), a large, diverse birth cohort to determine the relationships between early life antibiotic exposure, microbiome development, and growth.

## Study aims

- To measure the impact of common perinatal and early childhood antibiotic exposures on the structure and function of the developing gut microbiome
- To identify compositional and functional changes of the gut microbiome during early life

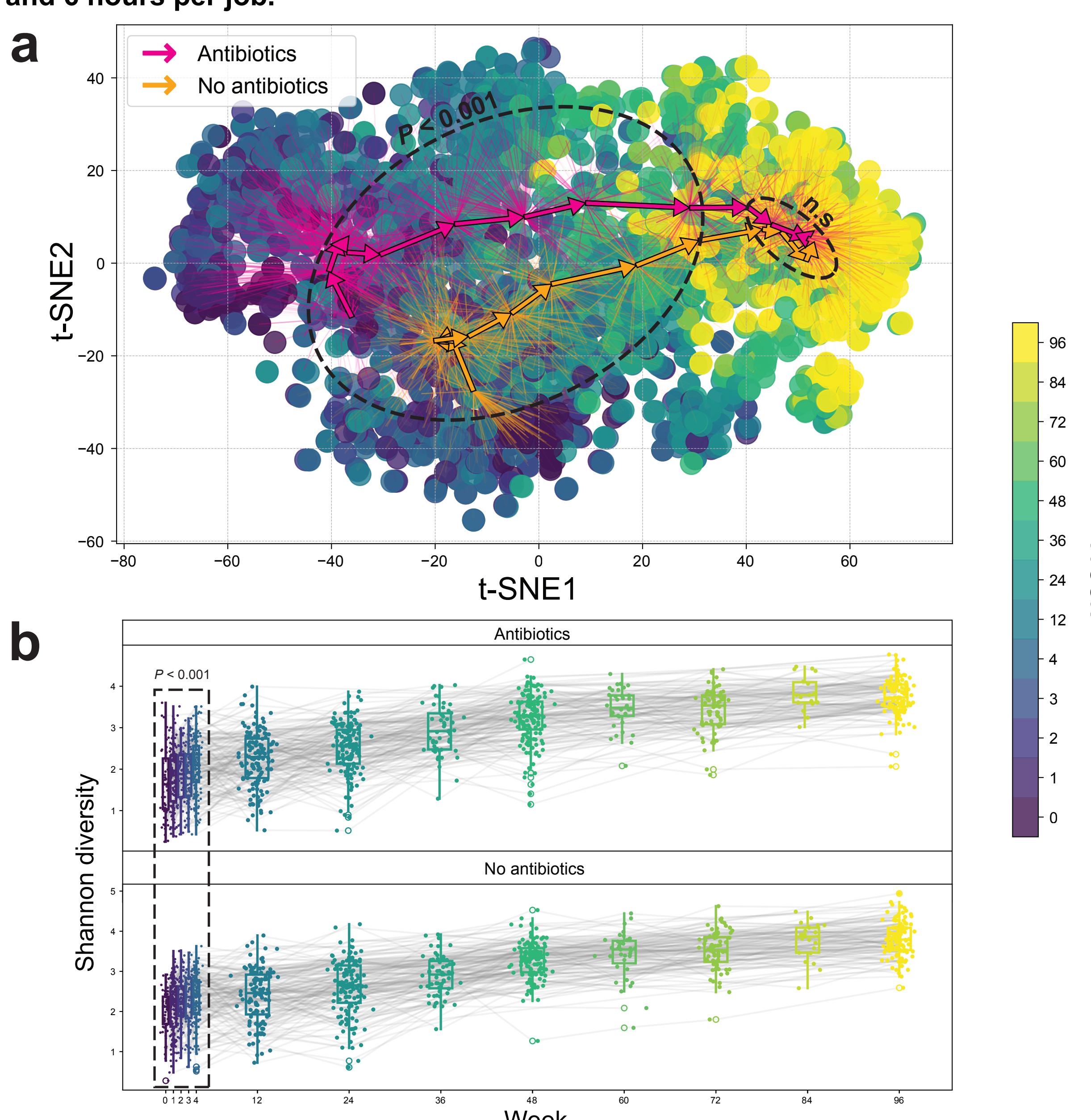
## Study design



A large, diverse birth cohort of 386 children was assembled, spanning various races and intrapartum antibiotic exposure types. **b** Stool shotgun metagenomes were collected during the first four weeks of life, and then once every three months until subjects reached two years of age (blue). Additionally, when subjects would undergo an antibiotic treatment, stool samples would be collected at several time points (0, 2, 5, 10, 30 days) after the beginning of each treatment (orange). After quality control, 3,179 stool shotgun metagenomes from 386 children remained for downstream analysis.

## Taxonomic composition trajectories of intrapartum antibiotic exposure groups take more than a year to converge

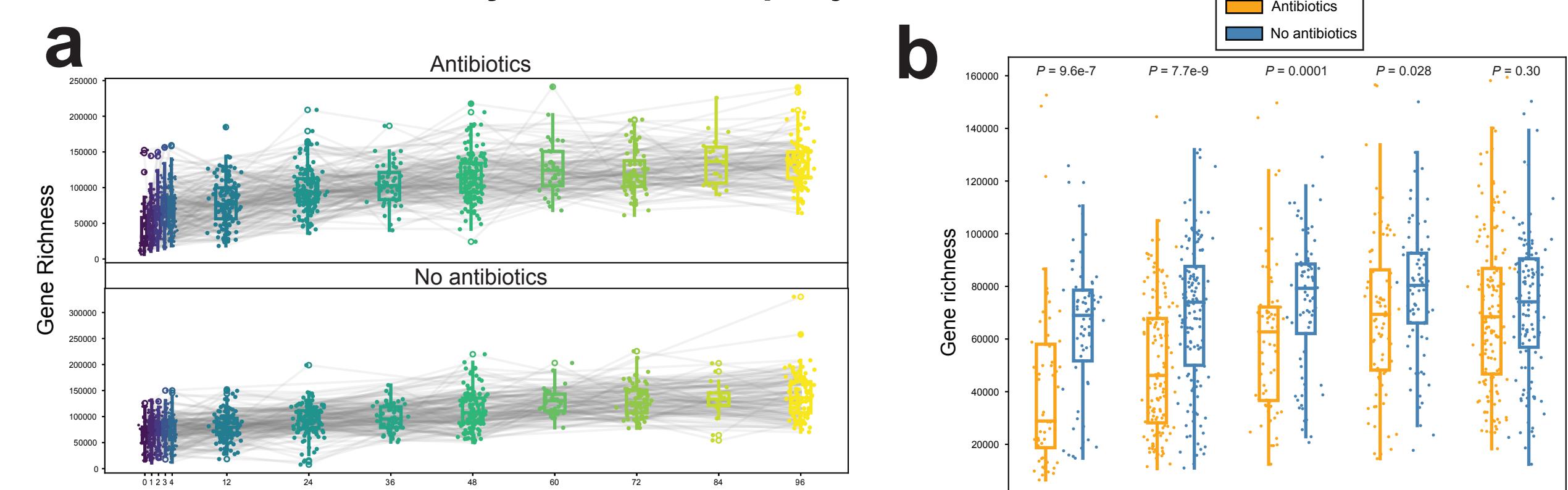
Taxonomic profiling was performed using the BURST sequence aligner and the Genome Taxonomy Database (GTDB): 3,600 jobs on Agate cluster; 24 cores, 48 GB of memory, and 6 hours per job.



t-SNE embedding of clr-transformed relative abundance profiles. Arrows connect centroids of consecutive time points for each intrapartum antibiotic exposure group. Faint lines connect each sample embedding to its respective exposure group centroid. **Species-level compositions between exposure groups are significantly different from each other at each time point from weeks 0 to 48** (PERMANOVA;  $P < 0.001$ ; Aitchinson distance) **but converge at week 60 and beyond** ( $P > 0.05$ ). **b** Shannon diversities between intrapartum antibiotic exposure groups are significantly different at each time point during the first month (Wilcoxon rank-sum;  $P < 0.001$ ) but converge afterwards ( $P > 0.05$ ).

## Functional diversity between antibiotic exposure groups converge quickly

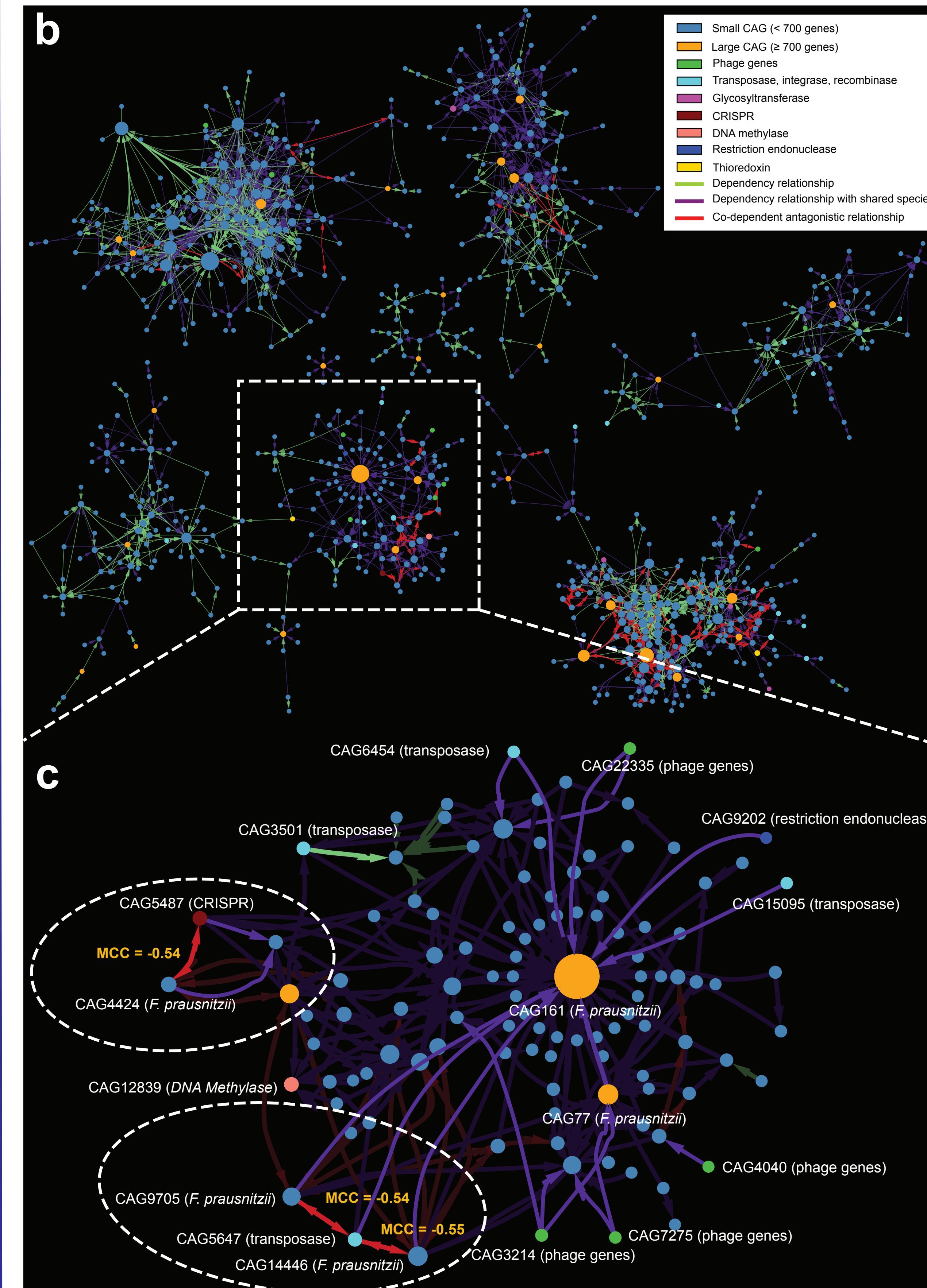
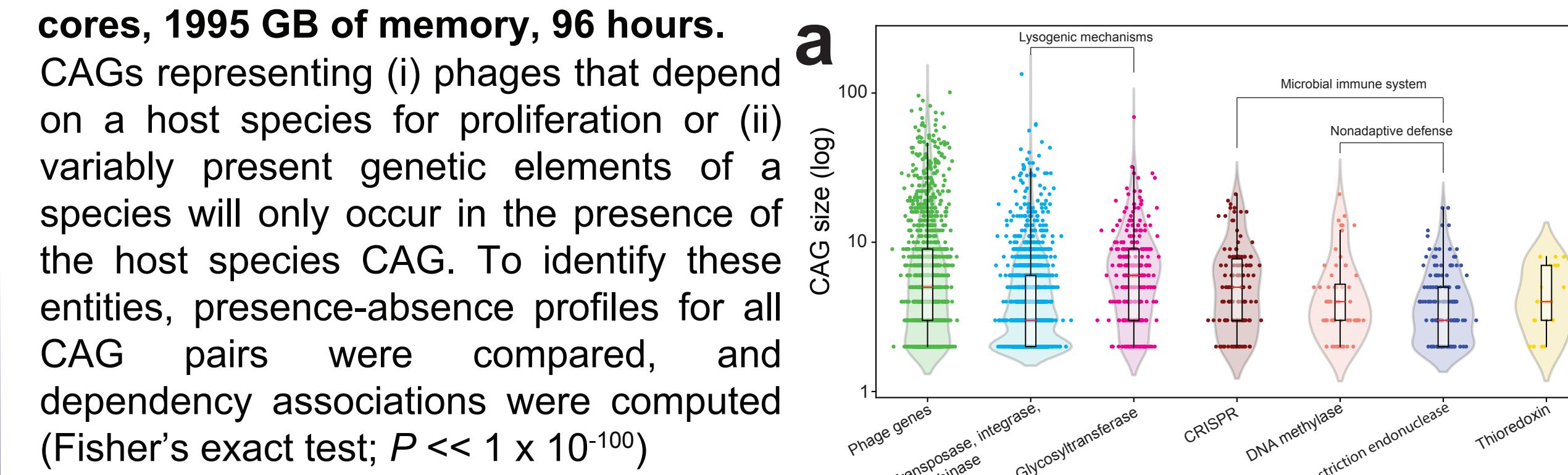
Functional profiling was performed using humann3: 3,600 jobs on Agate cluster; 128 cores, 499 GB of memory, and 8 hours per job.



## Gene co-abundance clustering reveals functional modules, putative viral entities, and clonal differences

Reference genomes cover only a small fraction of the true microbial diversity of complex environments. *De novo* segregation of metagenomic data can potential identify novel biological entities, enabling the insight of finer grained microbiome associations with clinical outcomes and other downstream analyses. As such, co-abundance gene groups (CAGs) were computed using the MGS-canopy clustering algorithm: 1 job on Agate cluster; 128 cores, 1995 GB of memory, 96 hours.

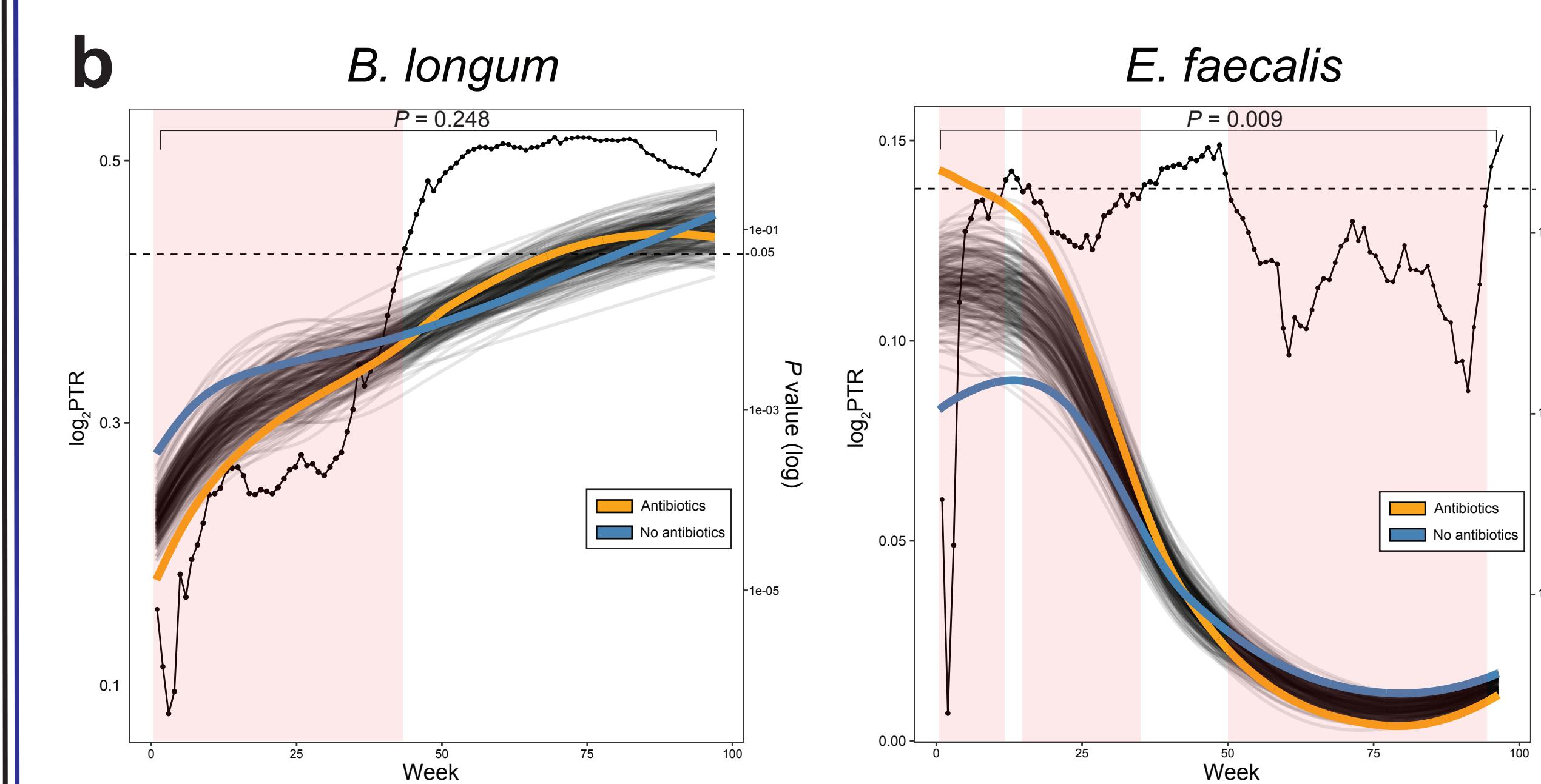
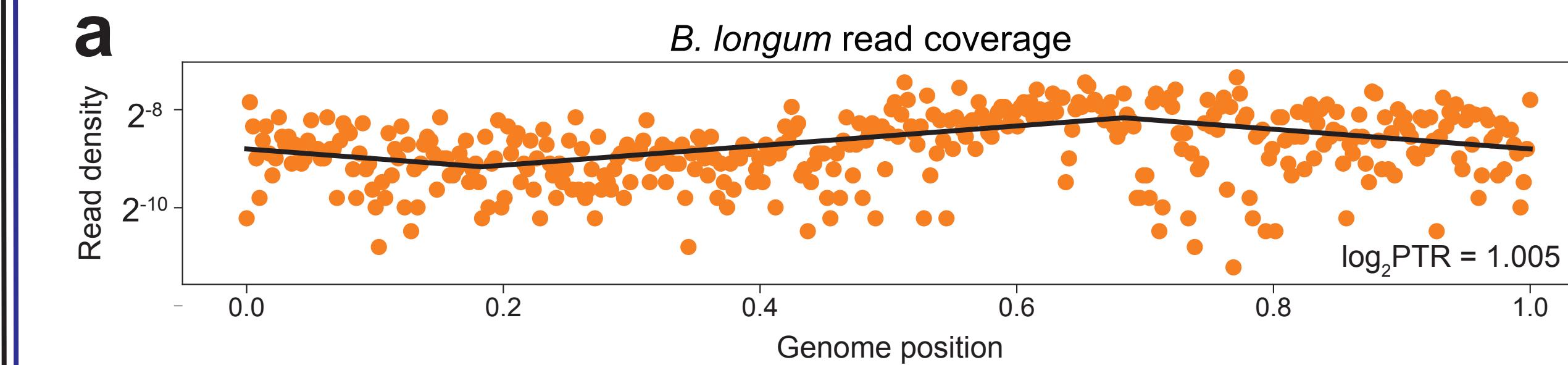
CAGs representing (i) phages that depend on a host species for proliferation or (ii) variably present genetic elements of a species will only occur in the presence of the host species CAG. To identify these entities, presence-absence profiles for all CAG pairs were compared, and dependency associations were computed (Fisher's exact test;  $P < 1 \times 10^{-100}$ )



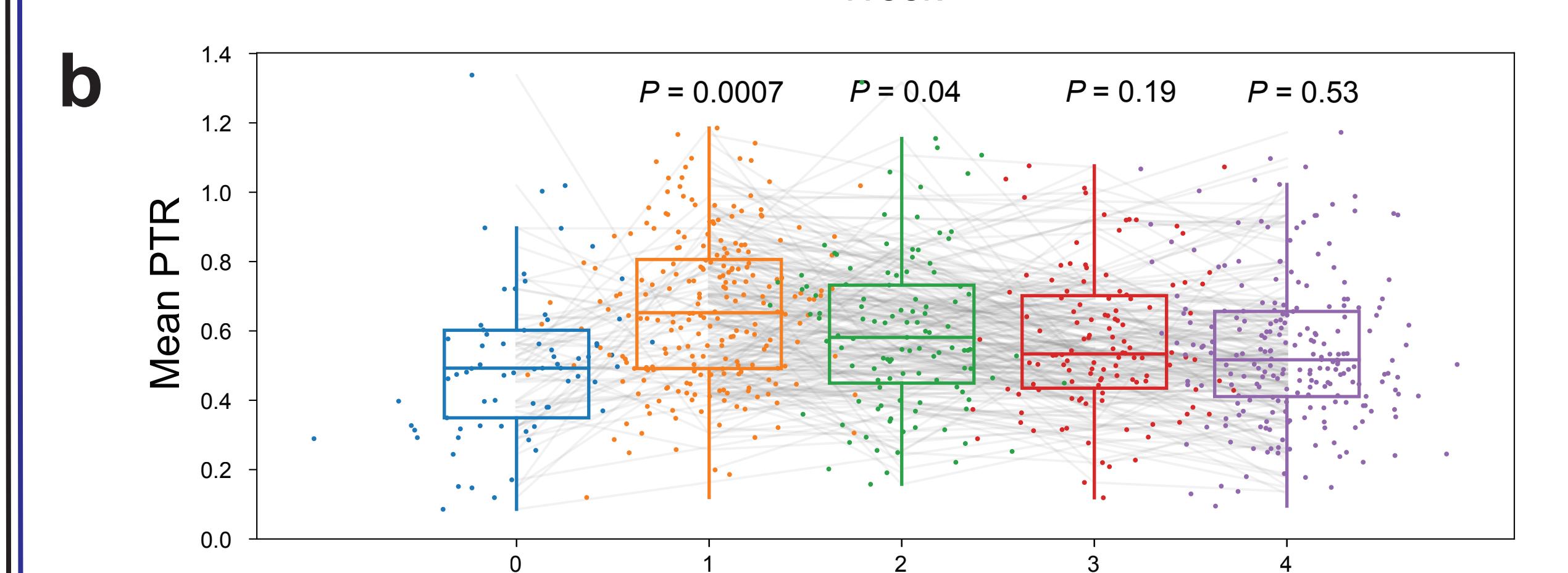
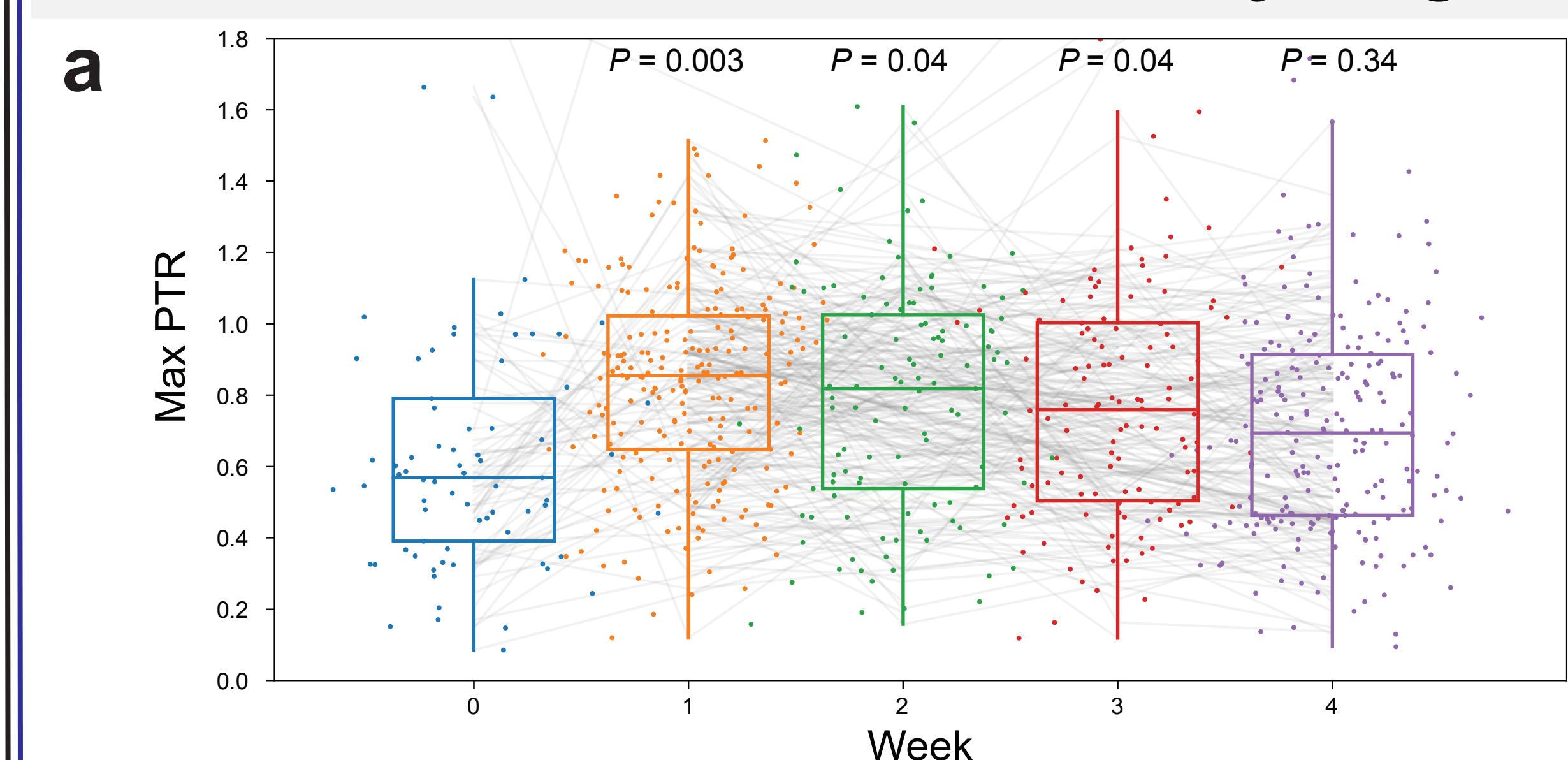
**a** CAG sizes (y-axis) for CAGs with enriched gene sets (x-axis). **b** The 6 largest connected components of the CAG dependency association network. Directed edges indicate CAG dependencies **with** and **without** identical species annotations, and **antagonistic relationships** between CAGs that are both dependent on the same CAG (Matthew's correlation coefficient  $< -0.5$ ). Node colors indicate CAG sizes and gene set enrichment (legend). **c** Subnetwork centered on *F. prausnitzii*. Many CAGs with *F. prausnitzii* genes are only present when the putative host CAG is, potentially representing variably present functional pathways of the species, thus providing insight into intraspecies diversity. **Anti-correlation** (circled) between CRISPR/transposase CAGs with other CAGs identifies putative phages (microbial immune system negatively interacts with phage functions).

## Microbiota growth rate dynamics reveal long term effects of intrapartum antibiotics

Because bacteria replicate their genomes bidirectionally from a single origin, rapidly growing bacterial populations will have a relatively high DNA copy number near the replication origin of the genome. We inferred microbial growth rates by computing peak-to-trough ratios (PTR) of genome read coverage using CoPTR: 3,600 jobs on Agate cluster; 128 cores, 499 GB of memory, and 8 hours per job.



## Microbiota growth rates suggest rapid gut colonization after birth followed by stagnation



**a** Max species PTR values (y-axis) during the first four weeks of life (x-axis) for samples in which at least one species had sufficient read coverage to estimate PTR. Wilcoxon signed-rank test P values for pairwise differences in PTR from week 0 to other timepoints are displayed. Max PTR at weeks 1 to 3 are significantly increased from week 0 ( $P < 0.05$ ) but return to baseline at week 4 ( $P = 0.34$ ). This potentially indicates a period of rapid colonization of the infant gut after birth, followed by a steady decline in growth once microbial communities are established and stabilized. **b** Identical as a but for mean PTR

## Conclusions

We present our progress in data generation and analysis of the gut microbiomes of the **MAGIC** longitudinal birth cohort. We plan to acquire more data from participants and further associate these engineered microbiome features with demographic information and clinical outcomes, potentially elucidating the role of the microbiome in human development during one of the most vital times of life regarding health.

## Acknowledgements

This work is in collaboration with the Children's Hospital of Philadelphia and was supported by the NIH (clinical trial NCT03001167), UMN Research Computing, and Minnesota Supercomputing Institute resources.