

Supplementary Information

Title: Metagenomic signatures of early life hospitalization and antibiotic treatment in the infant gut microbiota and resistome persist long after discharge

Authors: Andrew J. Gasparrini¹, Bin Wang^{1,2}, Xiaoqing Sun^{1,2}, Elizabeth A. Kennedy¹, Ariel Hernandez-Leyva¹, I. Malick Ndao³, Phillip I. Tarr^{3,4}, Barbara B. Warner³, Gautam Dantas^{1,2,4,5,*}

Contents:

Supplementary Table 1: Clinical and experimental metadata for infants and samples included in this study.

Supplementary Table 2: Assemblies used as outgroups for maximum likelihood phylogenies.

Supplementary Table 3: Antibiotic concentrations used for functional selections.

Supplementary Table 4: Statistics for infant gut isolates sequenced in this study.

- a. Pairwise core-gene single nucleotide polymorphism (SNP) distances for *E. coli* isolates.
- b. Pairwise core-gene SNP distances for *E. cloacae* isolates.
- c. Pairwise core-gene SNP distances for *Klebsiella* isolates.
- d. Resfinder resistance gene annotations in sequenced isolates.

Supplementary Figure 1: Timeline of samples and antibiotic treatments for infants in this study.

Supplementary Figure 2: Rarefaction curve of species identified by Metaphlan by sequencing depth.

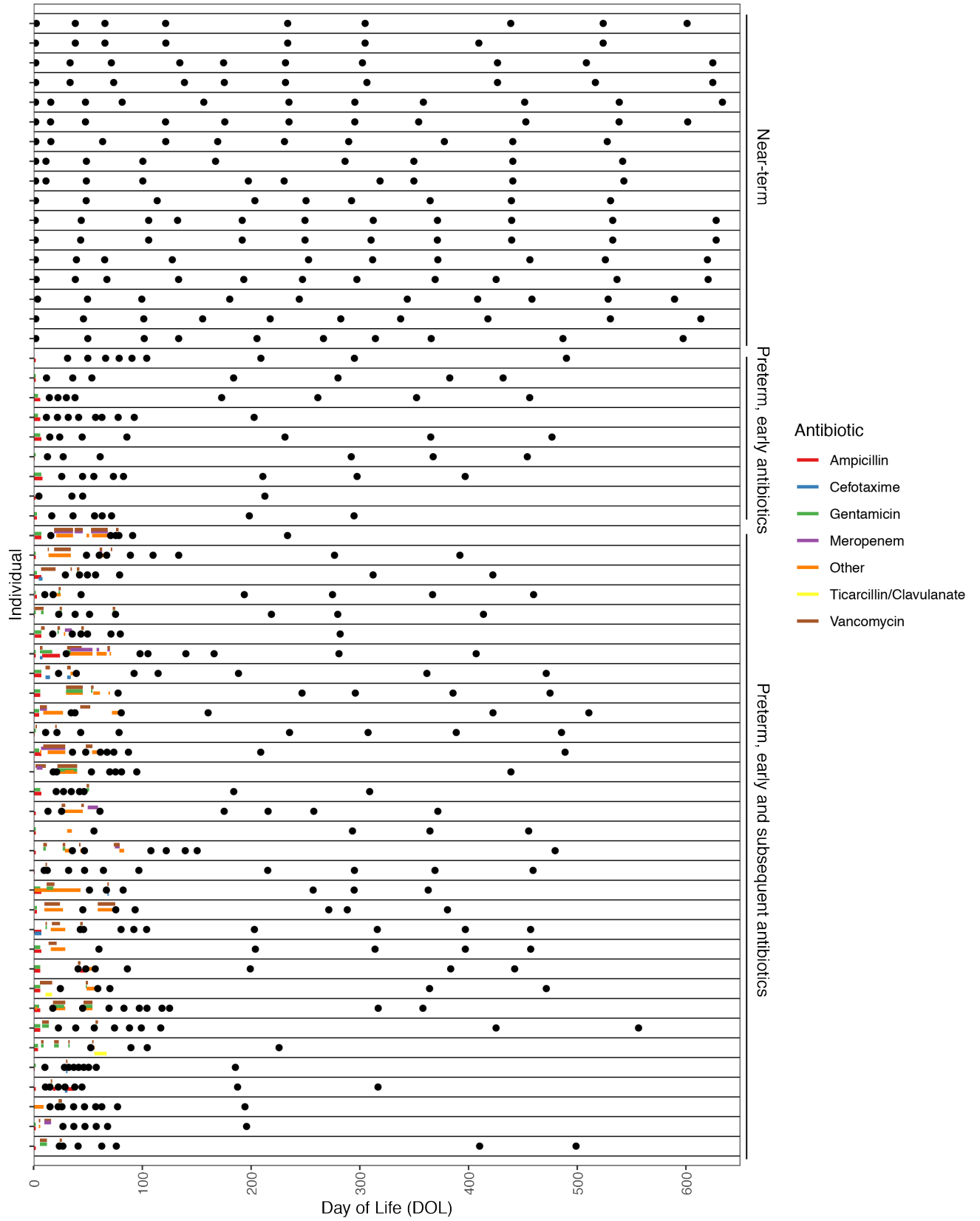
Supplementary Figure 3: Temporal changes in microbiota composition by group.

Supplementary Figure 4: Functional selections of metagenomic libraries constructed from infant stool.

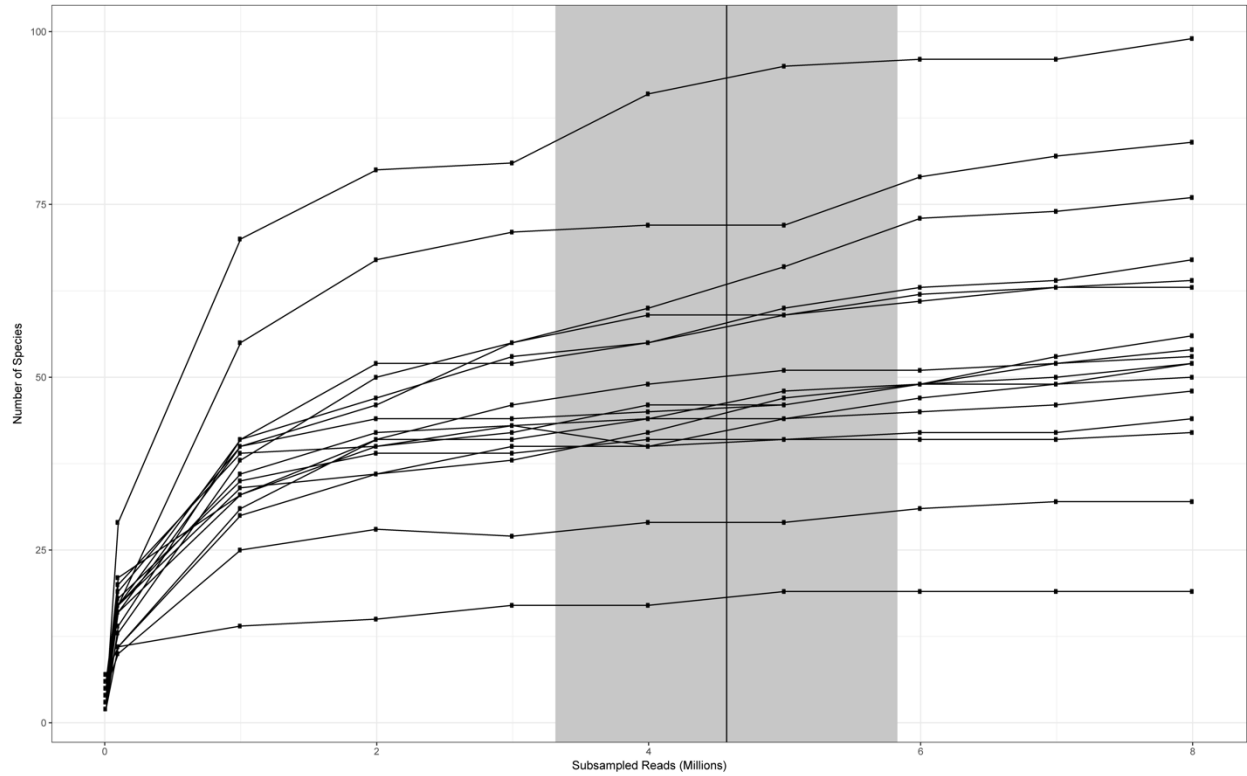
Supplementary Figure 5: Synteny of functionally-selected ARGs with mobile genetic elements.

Supplementary Figure 6: Microbiota with enriched resistome are low in diversity and dominated by Enterobacteriaceae.

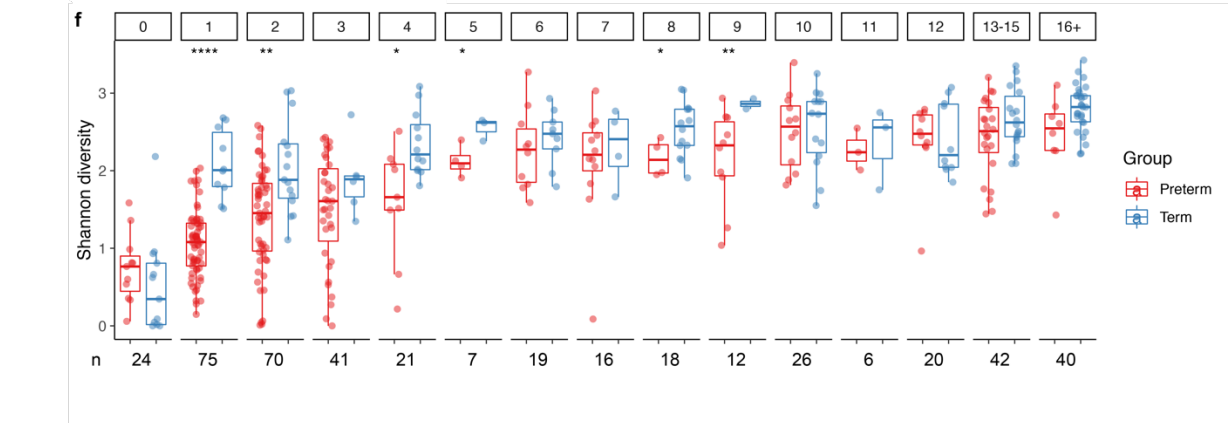
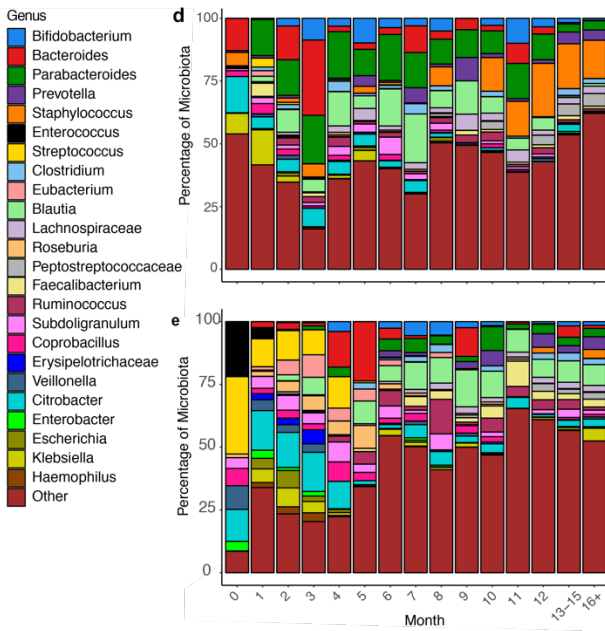
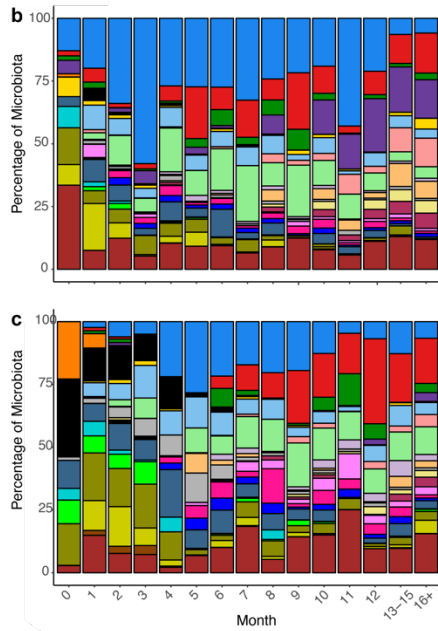
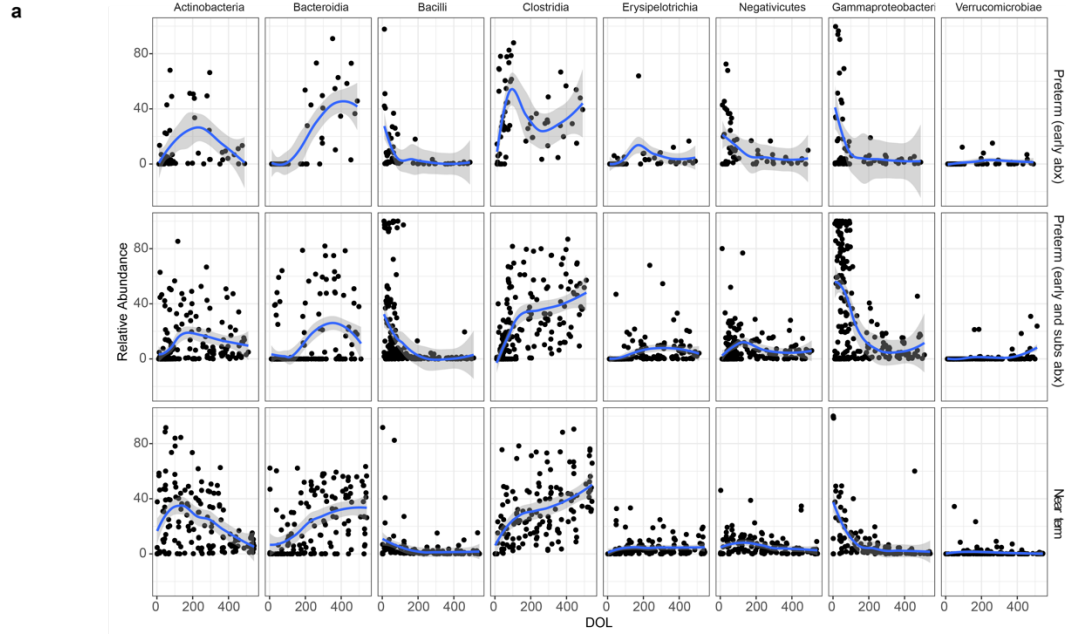
Supplementary Figure 7: Genomic and phenotypic heterogeneity in Enterococcus isolated from infant stool.



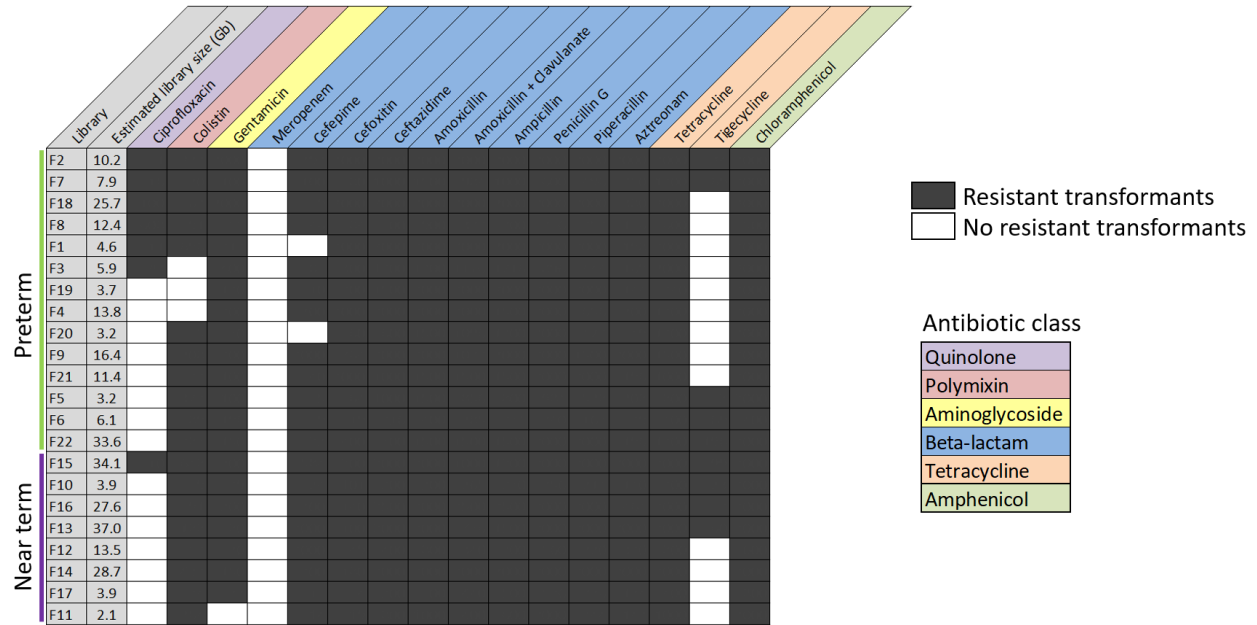
Supplementary Figure 1 | Timeline of samples and antibiotic treatments for infants in this study. Sampling timeline for near-term and preterm infants included in this study. Periods of antibiotic treatment are indicated by colored bars.



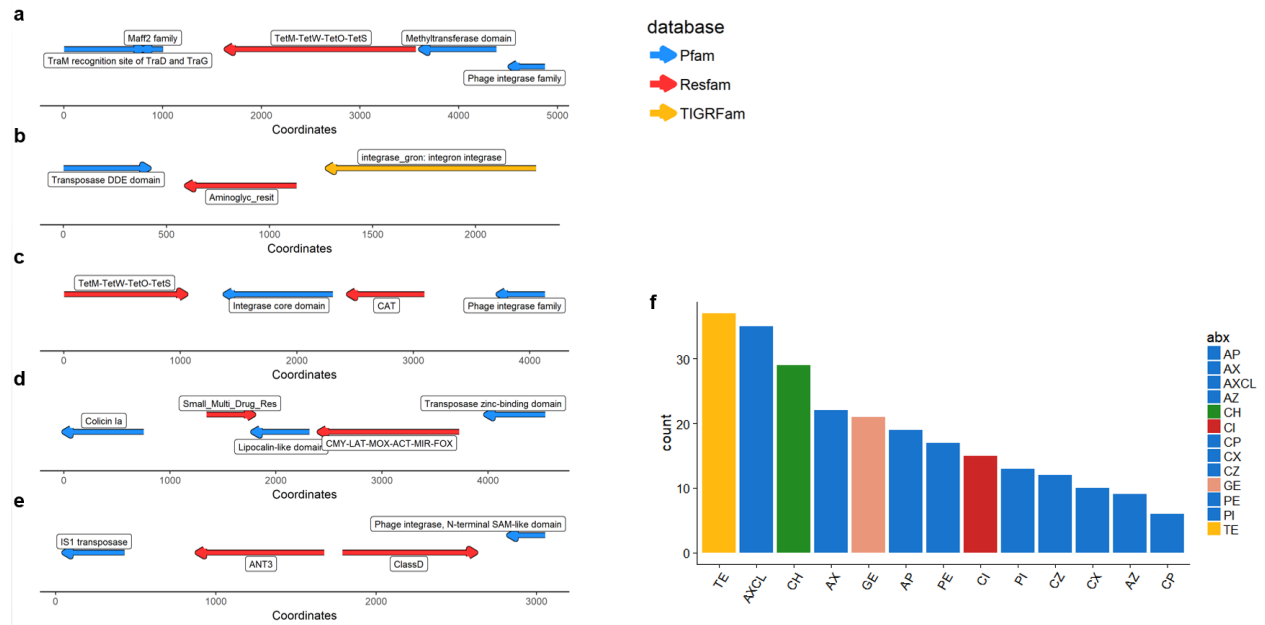
Supplementary Figure 2 | Rarefaction curve of species identified by Metaphlan by sequencing depth. Rarefaction curve for the number of unique species identified by subsampled sequencing depth. Black bar indicates median sequencing depth for all samples, gray shading indicates one standard deviation. n=16 biologically independent samples with sufficient sequencing depth for rarefaction.



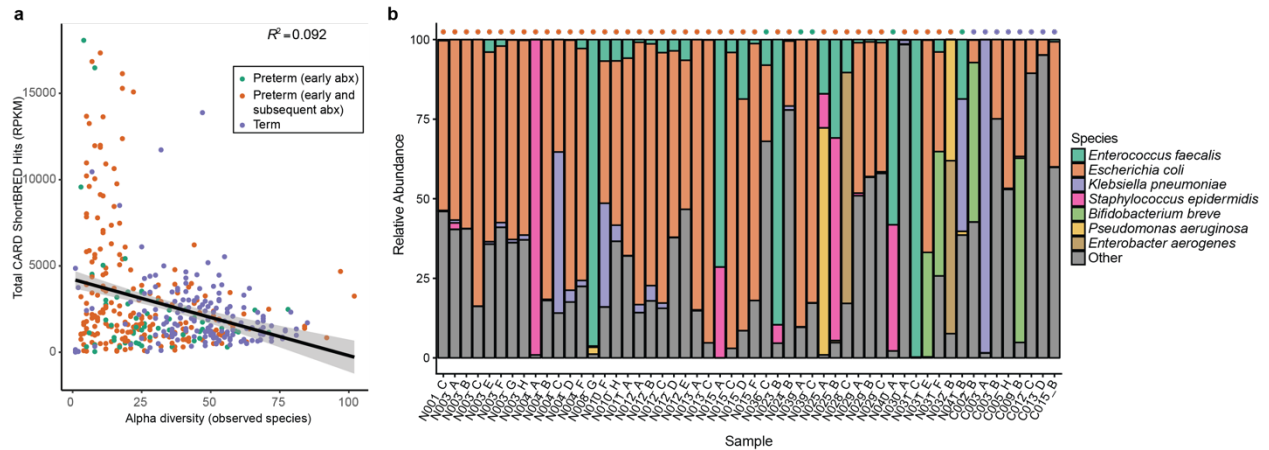
Supplementary Figure 3 | Temporal changes in microbiota composition by group. (a) The changes in relative abundance of the eight most abundant classes of bacteria in the gut microbiota of all infants over time. Curves are loess regression fit to the data, shading indicates 95% confidence interval, n=437 samples. (b) Microbiota composition at the genus level as inferred by MetaPhlAn2 in term infants. (c) Microbiota composition at the genus level as inferred by MetaPhlAn2 in preterm infants. (d) Microbiota composition at the species level as inferred by MetaPhlAn2 in term infants. (e) Microbiota composition at the species level as inferred by MetaPhlAn2 in preterm infants. (f) Shannon diversity comparisons between groups over the first months of life. (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$; two-sided t-test with Holm correction for multiple comparisons). Facet labels correspond to chronological month of life bins and number of samples are listed below each facet. Box plots represent the first quartile, median, and third quartile of the data with whiskers extending to the last data point within $1.5 \times$ the interquartile range.



Supplementary Figure 4 | Functional selections of 22 metagenomic libraries constructed from infant stool. Heatmap of resistance phenotypes observed in functional metagenomic selections. Library size and source (term or preterm infant) is indicated to the left of heatmap.



Supplementary Figure 5 | Synteny of functionally-selected ARGs with mobile genetic elements. Putative mobile genetic elements were identified syntenic to ARGs recovered from tetracycline (a), gentamicin (b), chloramphenicol (c), cefoxitin (d), and piperacillin (e) selections. The x-axis portrays the metagenomic coordinates. f, MGEs were most commonly identified on tetracycline selections. Color corresponds to antibiotic class (tetracycline, yellow; β -lactam, blue; amphenicol, green; aminoglycoside, pink; fluoroquinolone, red). Antibiotic abbreviations are listed in Supplementary Table 3.



Supplementary Figure 6 | Microbiota with enriched resistome are low in diversity and

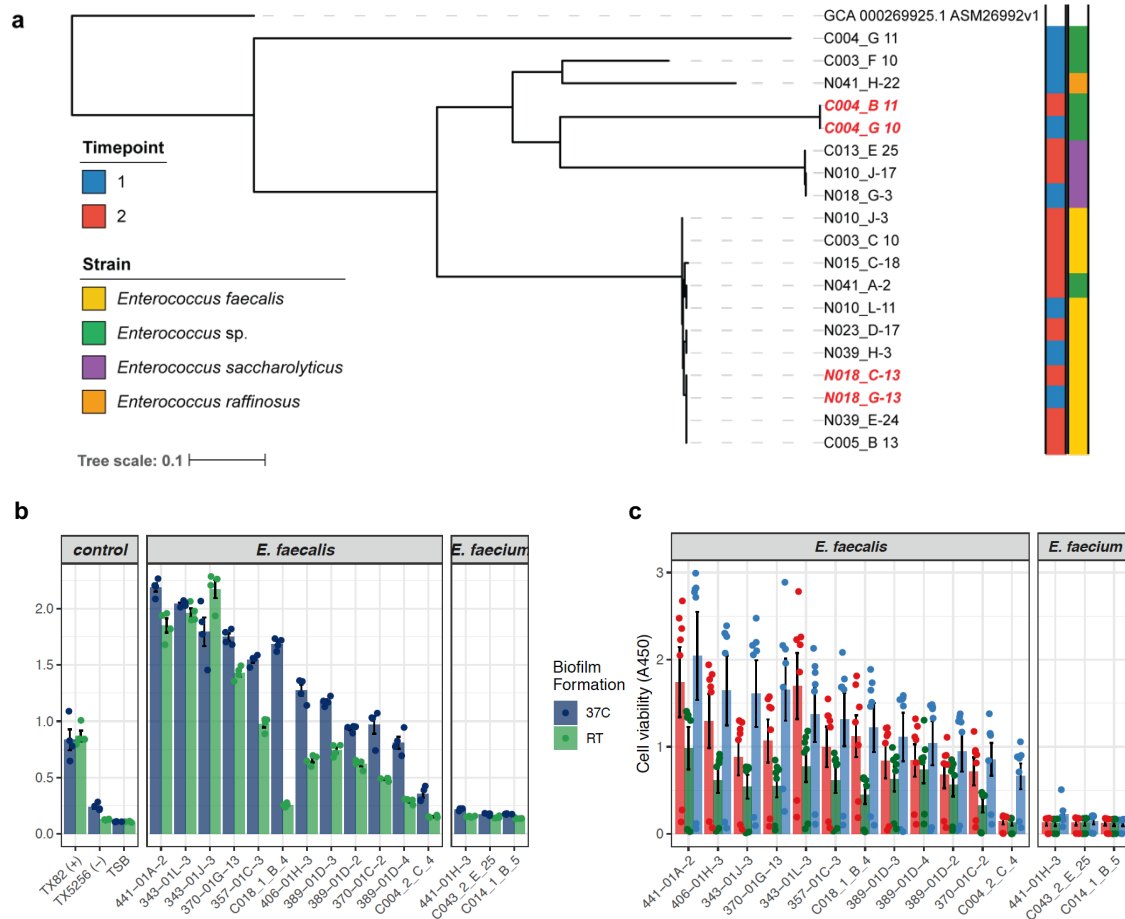
dominated by Enterobacteriaceae. a, There is a negative correlation between cumulative

relative resistome and alpha diversity (n=437, $R^2=0.092$). Black line is linear regression line of number of resistance genes observed against number of species observed in the infant microbiota and gray shading is 95% confidence interval.

b, Composition of the microbiota in samples with a resistome relative abundance > 5000 RPKM. Samples are low in diversity and dominated by a few organisms. In 41 out of 54 samples (76%), a single species comprises >50% of the

microbiota. These organisms are most commonly *E. coli*, *Enterococcus faecalis*, *Klebsiella pneumoniae*, *Staphylococcus epidermidis*, *Enterobacter aerogenes*, and *Bifidobacterium breve*.

Colored dots above barplot indicate the group to which each sample belongs (see legend a bottom of panel a).



Supplementary Figure 7 | Genomic and phenotypic heterogeneity in *Enterococcus* isolated

from infant stool. a, Maximum likelihood core genome phylogeny of *Enterococcus* strains isolated from infant gut microbiota. Persistent strains are highlighted in red. Annotations indicate time point and species. **b**, Biofilm formation phenotypes of *E. faecalis* and *E. faecium* strains at 37°C and at room temperature. TX82 is an *E. faecium* positive biofilm forming control, TX5256 is an *E. faecalis* negative biofilm forming control, TSB is media only control. **c**, Viability of *Enterococcus* strains in biofilms following vancomycin treatment. Bars indicate mean values and error bars indicate s.e.

Supplementary Table 1 | Clinical and experimental metadata for infants and samples included in this study. Detailed clinical and experimental metadata for the infants and samples included in this study. All variables in this table were included in initial fitting of the generalized linear mixed effects model.

Supplementary Table 2 | Assemblies used as outgroups for maximum likelihood phylogenies

Species	Count	Outgroup	Outgroup assembly accession
<i>Escherichia coli</i>	30	<i>Escherichia fergusonii</i> ATCC 35469	ASM2622v1
<i>Enterococcus</i> spp.	19	<i>Lactococcus garvieae</i> ATCC 49156	ASM26992v1
<i>Enterobacter cloacae</i>	11	<i>Enterobacter asburiae</i> ATCC 35953	ASM152171v1
<i>Klebsiella</i> spp.	21	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	ASM19599v1

Each maximum likelihood phylogeny was constructed using the number of assemblies from this study as well as the outgroup listed here.

Supplementary Table 3 | Antibiotic concentrations used for functional selections

Antibiotic	Abbreviation	Screening Concentration - $\mu\text{g/ml}$
Amoxicillin	AX	16
Amoxicillin + Clavulanate	AXCL	16-8
Ampicillin	AP	64
Aztreonam	AZ	8
Cefepime	CP	8
Cefoxitin	CX	64
Ceftazidime	CZ	16
Chloramphenicol	CH	8
Ciprofloxacin	CI	0.5
Colistin	COL	8
Gentamicin	GE	16
Meropenem	ME	16
Penicillin G	PE	128
Piperacillin	PI	16
Tetracycline	TE	8
Tigecycline	TG	2

All functional metagenomic libraries were selected on antibiotics at concentrations listed here.

Supplementary Table 4 | Statistics for infant gut isolates sequenced in this study. Details of comparative genomic and resistance gene analyses of isolates cultured from infant stool and sequenced in this study.

- a. Pairwise core-gene single nucleotide polymorphism (SNP) distances for *E. coli* isolates.
- b. Pairwise core-gene SNP distances for *E. cloacae* isolates.
- c. Pairwise core-gene SNP distances for *Klebsiella* isolates.
- d. Resfinder resistance gene annotations in sequenced isolates.