FACTORS IMPACTING ANTIBIOTIC RESISTANCE IN THE MICROBIOME

CAPTURING ANTIBIOTIC RESISTANCE GENES IN THE HUMAN GUT MICROBIOME AFTER PROBIOTIC AND ANTIBIOTIC EXPOSURES

By ALLISON KATE GUITOR, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University, DOCTOR OF PHILOSOPHY (2023) Hamilton, Ontario (Health Sciences)

TITLE: Capturing Antibiotic Resistance Genes in the Human Gut Microbiome After Probiotic and Antibiotic Exposures

AUTHOR: Allison K. Guitor, B. Sc. (The University of Waterloo)

SUPERVISOR: Gerard D. Wright

NUMBER OF PAGES: xvii, 274

Foreword

LAY ABSTRACT

Infection rates associated with antibiotic-resistant bacteria are increasing. The traits that allow bacteria to survive antibiotics are encoded by their genetic information or genes. There is a wide diversity of antibiotic resistance mechanisms available to bacteria and comprehensive methods to detect them are needed. My work presents the development and validation of an approach to capture and sequence all known antibiotic resistance genes. With this tool, I investigated the early life exposures that shape antibiotic resistance in the diverse community of bacteria encompassing the human gut microbiome. First, I revealed the protective effects of providing beneficial bacteria, or probiotics, to preterm infants. Next, I uncovered a minimal impact of antibiotic treatment in selecting for antibiotic resistance genes in children with diarrhoea. Improved tools to understand antibiotic resistance in early-life and the factors that modulate its prevalence will help reduce the persistence of antibiotic-resistant bacteria in later life.

ABSTRACT

Antibiotic-resistant bacteria and their often-untreatable infections are increasing globally. The determinants that confer reduced susceptibility to antibiotics are encoded in the genetic information, or genes, of bacteria. Sequencing-based approaches to identify antibiotic resistance genes (ARGs) in diverse environments, such as the gut microbiome, are limited in throughput and inaccessible due to high costs. The microbiome and its associated ARGs are shaped by many factors throughout early life, which have yet to be fully characterized. In this thesis, I developed an approach to capture ARGs in the gut microbiome and study the effect of various early-life exposures.

First, I describe a set of over 37,000 probes to target over 2,000 ARGs. This targeted sequencing approach was designed and validated against samples of DNA isolated from multi-drug resistant bacteria. This probeset proved superior to metagenomic shotgun sequencing in capturing the rare portion of ARGs from a human gut microbiome sample.

Second, I investigated the potential benefits of providing probiotics to infants born preterm in reducing antibiotic resistance in their gut microbiomes. Preterm infants that received probiotics had a reduced burden of ARGs associated with potentially pathogenic bacteria at 5 months of age compared to non-probiotic-supplemented preterm infants.

Finally, I explored the potential consequences of macrolide antibiotic exposure on ARGs in the gut microbiome of children with diarrhoea in Botswana, Africa. Compared to the standard treatment for diarrhoea, a three-day dose of azithromycin did not result in an increased selection for ARGs after 60 days.

This thesis tackles the challenge of comprehensively detecting antibiotic resistance in the gut microbiome. A better understanding of the impact of early-life exposures, including probiotic and antibiotic treatments, in the prevention or maintenance of ARGs in the gut microbiome of children will help reduce unnecessary selective pressures and the persistence of antibiotic resistance into later life.

ACKNOWLEDGEMENTS

Graduate school is quite a journey and there are far too many people to thank.

To Gerry – your passion for science is infectious! Your dedication to pushing the boundaries in research is inspiring and I hope to be half as prosperous as you. I will never forget your many motivational phrases, including "K.I.S.S". I may not have always followed that advice, but I will take it into consideration in all that I do. Thank you for your support, trust in me, and your always insightful career and life advice over the past 6 years.

To my committee members Hendrik and Andrew – Thank you for welcoming me into your own research groups and for being open to collaborations. I appreciate the added support and motivation that you provided outside of our regular committee meetings. The AMR baitset was your brainchild and I am honoured to have been entrusted with this project.

To the Wright Lab! I do not think I will find another environment with as many passionate scientists as this one. You have become like a second family. To the amazing postdocs, especially Min for putting up with being my bench buddy - thank you for providing unattainable examples of what a postdoc can do and for sharing your expertise. To Kalinka, Linda, and Mike: I am grateful for all that you do to ensure the lab runs smoothly. Beth, Emily, Caitlyn, Matt, and Haley – what an inspiring group of PhD students to be a part of! I am excited to see all we achieve in the coming years! Nick - thank you for my first lessons in bioinformatics. Without your mentoring in the early days, I would not have made it this far. To all other members past and present, especially Vishwas and Drew: It's been a blast and I'm glad we can keep in touch!

To my wonderful and supportive Mom and Dad – thank you for believing in me from day one and for giving me the opportunity to follow my dreams! You will always be my biggest cheerleader, and I could never thank you enough for all that you do for me.

To Steven – you are my rock and best friend. You came into my life during the most stressful of times and I am so thankful for your support, kindness, and patience. I am looking forward to all our future adventures together.

To the amazing staff of the McMaster Genomics Sequencing Facility – there is by far no other group that I have emailed more than you! Thank you for humouring my incessant questions and for accepting my over 50 batches of samples throughout the years.

Finally, to so many others that are too numerous to mention, including the graduate students in the department, the Poinar Lab for including me in their group, my friends in Ottawa, and my #38 roomies. You have all contributed to my journey in some way. Thank You!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTS OF ALL ABBREVIATIONS AND SYMBOLS

23S rRNA	23S ribosomal ribonucleic acid
AAC	Aminoglycoside *N*-acetyltransferase
ABC	ATP-binding cassette
ADP	Adenosine diphosphate
AGF	Antimicrobial resistance gene family
AME	Aminoglycoside modifying enzymes
AMR	Antimicrobial resistance
ANT	Aminoglycoside *O*-nucleotidyltransfereases
APH	Aminoglycoside *O*-phosphotransferases
ARG	Antibiotic resistance gene
ARO	Antibiotic resistance ontology
AST	Antibiotic susceptibility testing
ASV	Amplicon sequence variant
ATP	Adenosine triphosphate
β	Beta
Baby and Pre-Mi	Baby and Preterm Microbiota of the Intestine Cohort Study
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CARD	Comprehensive Antibiotic Resistance Database
CLSI	Clinical and Laboratory Standards Institute
CRISPR	Clustered Regularly Interspace Short Palindromic Repeats
DIAMOND	Double Index Alignment of Next-Generation Sequencing Data
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EPEC	Enteropathogenic *Escherichia coli*
ESBLs	Extended-spectrum β-lactamases
ETEC	Enterotoxigenic *Escherichia coli*
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FT	Full term
HGT	Horizontal gene transfer
HiREB	Hamilton Integrated Research Ethics Board
HMM	Hidden Markov models

HSP	high-scoring segment pairs
ICE	Integrative and conjugative element
IIDR	Institute for Infectious Disease Research
IR	Inverted repeats
IS	Insertion sequences
kb	Kilobase
LMAP	Loop-mediate isothermal amplification
LMAT	Livermore Metagenomics Analysis Toolkit
LMICs	Low- and middle-income countries
LPS	Lipopolysaccharide
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MATE	Multidrug and toxic-compound exclusion
MDA	Mass drug administration
MDR	Multidrug resistant
MFS	Major facilitator superfamily
MGE	Mobile genetic element
MGP	Methyl-alpha-D-glucopyranoside
MIC	Minimum inhibitory concentration
MLS_B	Macrolide, lincosamide, streptogramin B
MTB	*Mycobacterium tuberculosis*
MYA	Million years ago
NCBI	National Center for Biotechnology Information
NEC	Necrotizing enterocolitis
NGS	Next-generation sequencing
NICU	Neonatal intensive care unit
NS	Not supplemented
ORF	Open reading frame
PBP	Penicillin-binding protein
PCR	Polymerase chain reaction
PMA	Postmenstrual age
PS	Probiotic supplemented
qPCR	Quantitative PCR
QRDR	Quinolone resistance determining region
RGI	Resistance gene identifier
RGI bwt	Metagenomic read-mapping feature of RGI

RGI kmer_query	Pathogen-of-origin feature of RGI
RGI main	Genome annotation feature of RGI
RIF	Rifampicin
RNA	Ribonucleic acid
RND	Resistance-nodulation-cell division
RPKM	Reads per kilobase of reference gene per million reads
rRNA	Ribosomal RNA
RTT	Rapid Test-and-Treat
SC	Standard care
SCC*mecI*	Staphylococcal cassette chromosomal element
SDG	Sustainability development goal
SMR	Small multidrug resistance
SNP	Single nucleotide polymorphisms
SOS	Save our soul
Tris-HCl	Tris hydrochloride
tRNA	Transfer RNA
USA	United States of America
VRE	Vancomycin-resistant *Enterococcus faecium*
WHO	World Health Organization
XDR	Extant drug resistance

DECLARATION OF ACADEMIC ACHIEVEMENT

I have performed all the research in this body of work except where indicated in the preface of each chapter.

CHAPTER ONE: Introduction

ANTIBIOTICS: MIRACLE DRUGS OVERCOME BY RESISTANCE

Antibiotics are miraculous drugs that have revolutionized human and animal medicine since the early 1900s. Before this breakthrough in clinical treatment, infections were one of the leading causes of human mortality and severely hindered advancements in medicine (Davies & Davies, 2010; Jones et al., 2012). Since this discovery and the widespread adoption of antibiotics alongside modern hygiene practices, many infectious diseases have been rapidly treated, and extensive transmission prevented. Soon after the introduction of antibiotics, a new challenge quickly arose; antibiotics began to lose their efficacy, and microbes triumphed once again (Abraham & Chain, 1940; Finland, 1979; Fleming, 1929). The discovery of resistance to early antibiotics launched global efforts to understand the mechanisms behind this reduced susceptibility and to find new and improved antibiotic candidates.

Despite the past 100 years of research and innovation, microorganisms continue to fight back, and we are threatened by a return to an era without antibiotics in medicine. In 2019, there were an estimated 4.95 million deaths associated with antibiotic-resistant infections, where 1.27 million could be directly attributed to resistance (Murray et al., 2022). While antibiotic resistance is a global issue, the burden is highest in low- and middle-income countries where access to necessities such as clean drinking water, varying availability and quality of antibiotics, lack of regulations on use, and limited surveillance of resistance further exacerbate the problem (Murray et al., 2022). It is estimated that by 2050, more than 10 million human deaths will be attributable to antimicrobial-resistant

infections (O'Neill, 2016). Various global entities, including the World Health Organization (WHO), have proposed measures to alleviate this burden (Division, 2016). The consensus from these recommendations is that a better understanding of antibiotic resistance, from the mechanisms to their distribution and spread at a global level is imperative to prevent a post-antibiotic era.

ANTIBIOTICS: MECHANISMS AND ORIGINS

Anti-infective agents encompass compounds (natural, synthetic, and semi-synthetic) that inhibit cell growth or result in the death of microorganisms, including viruses, bacteria, fungi, and parasites. Some compounds may also have anti-tumour, insecticide, or herbicide activity, among other roles (Davies & Davies, 2010; Demain & Sanchez, 2009). The remainder of this thesis will predominately focus on the subset of agents that target bacteria, the antibacterials or antibiotics. These molecules can be bactericidal (result in bacterial death) or bacteriostatic (limit bacterial replication without causing death) (Leekha et al., 2011). Most antibiotic classes are derived from molecules isolated as natural products from other organisms, including bacteria, fungi, and plants (Brown & Wright, 2016). Throughout this thesis, the terms antibiotic and antibacterial will be used interchangeably. In addition, antibiotic resistance, antimicrobial resistance (AMR), and antibacterial resistance all refer to the feature(s) that provide bacteria with reduced susceptibilities to antibiotics.

To understand the various mechanisms of antibiotic resistance, one must first understand the targets of antibiotics. Many antibiotics target core cellular processes within

bacteria, including cell wall biosynthesis, DNA replication, transcription, and protein translation (Table 1). Each of these systems has a complement of proteins that may be engaged and are often essential for growth (Figure 1A).

The β-lactam family of antibiotics include the penicillins, cephalosporins, monobactams, and carbapenems. These predominantly inhibit penicillin-binding proteins (PBPs) and disrupt peptidoglycan biosynthesis, a key building block of bacterial cell walls (Lewis & Bush, 2015). Glycopeptides, such as vancomycin, also inhibit peptidoglycan biosynthesis but target the *D*-alanyl-*D*-alanyl residue in the cell wall precursor, lipid II (Lewis & Bush, 2015). Quinolones, including ciprofloxacin, target type II DNA topoisomerases, including DNA gyrase encoded by *gyrA*. These antibiotics result in bacterial death by disrupting DNA supercoiling and ultimately interfering with DNA replication (Lewis & Bush, 2015). Finally, rifamycins target the β subunit of the DNA-dependent RNA polymerase (*rpoB*) and prevent DNA transcription.

Protein production is targeted by many classes of antibiotics. Aminoglycosides bind to the 16S rRNA component of the small (30S) ribosomal subunit. This disrupts codon-anticodon pairing and results in improper protein translation and cell death (Carter et al., 2000). Macrolides, such as erythromycin, target the 23S rRNA of the large (50S) ribosomal subunit and block polypeptide elongation (Lewis & Bush, 2015; Tu et al., 2005). Clindamycin, a member of the lincosamide group of antibiotics, also targets the large ribosomal subunit in the same region as macrolides (Lewis & Bush, 2015; Tu et al., 2005). The oxazolidinone family, exemplified by linezolid, prevent the formation of a functional

ribosome by interacting with the 23S rRNA. Streptogramin A and B are used synergistically to target the 50S subunit of the ribosome (Lewis & Bush, 2015). Protein synthesis is also targeted by tetracyclines, which act by binding to the 30S ribosomal subunit and preventing aminoacyl-tRNA entry to the ribosome (Lewis & Bush, 2015).

Other cellular processes targeted in bacteria include the folate pathway essential to generate precursors for DNA synthesis. Sulfonamides were one of the first antibiotics introduced to clinical care and are purely synthetic (Skold, 2000). They target the dihydropteroate synthase (*folP*), while trimethoprim, a pyrimidine-based antibiotic, targets dihydrofolate reductase (*dfr*). Combining these antibiotics as co-trimoxazole has been a popular approach to overcome resistance. Finally, cationic peptides, such as polymyxins, interact with phospholipids to disrupt the bacterial cell membrane. Many classes of antibiotics are not listed in Table 1 for various reasons, including their lack of approval for human use or a limited prevalence and understanding of antibiotic resistance in human clinical pathogens.

Table 1: Major Classes of Antibiotics and Associated Resistance Mechanisms.
Major antibiotic classes used in human medicine and their associated mechanisms of resistance. Information presented in this table were amalgamated from various references (Alcock et al., 2023; Brown & Wright, 2016; Guitor & Wright, 2018; Morar & Wright, 2010; Walsh & Wencewicz, 2016; Wright, 2007).

Target	**Antibiotic Class**	**Example(s)**	**Mode(s) of resistance with example genes**
Cell wall biosynthesis / peptidoglycan biosynthesis	β-lactams (bactericidal)	Pencillins Cephalosporins Penems Monobactams Carbapenems	Hydrolysis (β-lactamases) Efflux or altered porin expression Altered target (mutant PBPs, *mecA*)

	Glycopeptides (bactericidal)	Vancomycin Teicoplanin	Reprogramming peptidoglycan biosynthesis (*vanRSHAX*)
	Epoxide (bactericidal)	Fosfomycin	Inactivation (*fos*)
Cell envelope integrity	Cationic peptides (bactericidal)	Polymyxins (eg. Colistin)	Altered target/Increase in positive charge on the cell surface (*mcr*) Efflux
	Lipopeptides (bactericidal)	Daptomycin	Altered target (*mprF*) Modification of membrane charge
DNA replication and transcription	Quinolones[S] (bactericidal)	Ciprofloxacin	Target mutation (*gyrA*, *parC*), Target protection (*qnr*) Efflux (*acrAB-tolC*, *qepA*, *oxqAB*, *norAB*) Acetylation (*aac(6')-Ib-cr*)
Nucleotide synthesis or folate biosynthesis / C1 metabolism	Sulfonamides[S] (bacteriostatic)	Sulfanilamide Sulfamethoxazole	Target overexpression, mutation or modified dihydropteroate synthase (*folP*) Efflux Acquisition of resistant target (*sul1*)
	Pyrimidines[S] (bacteriostatic)	Trimethoprim	Efflux, altered target (*dfr*)
RNA synthesis / Transcription	Ansamycins / Rifamycins (bactericidal)	Rifampin	Mutations in RNA polymerase (*rpoB*) Glycosylation (*rgt*) ADP-ribosylation (*arr*) Phosphorylation (*rph*) Decomposition or oxidation (*rox*) Efflux
Protein synthesis / Translation	Aminoglycosides (bactericidal)	Gentamicin Streptomycin Spectinomycin Kanamycin A Neomycin	Phosphorylation (*aph*) Acetylation (*aac*) Nucleotidylation/adenylation (*ant*) Efflux

			Altered target: mutation or methylation (*armA, rmtB*) of 16S rRNA
	Tetracyclines (bacteriostatic)	Minocycline Tigecycline Tetracycline Doxycycline	Monooxygenation (*tetX*) Efflux (MexAB-OprM; *tet*(A)) Altered target: 16S rRNA mutation Ribosomal protection proteins (*tetM*)
	Macrolides (bacteriostatic)	Erythromycin Azithromycin Clarithromycin	Esterification (*ere*) Gycosylation (*mgt*), Phosphorylation (*mph*) Efflux (*mef, msr*) Altered target: mutation or methylation (*erm*) of 23S rRNA
	Lincosamides (bacteriostatic)	Clindamycin	Nucleotidylation (*lnu*) Efflux Altered target
	Streptogramins (bactericidal)	Streptogramin A (S_A): dalfopristin Streptogramin B (S_B): quinupristin Pristinamycin IIa	S_A: Acetylation (*vat*) S_B: C-O lyase (*vgb*) S_B: 23S rRNA methylation Efflux
	Oxazolidinones[S] (bacteriostatic)	Linezolid	Efflux (*lmrS*) Altered target: mutation or methylation (*cfr*) of 23S rRNA
	Phenicols (bacteriostatic)	Chloramphenicol	Altered target: 23S rRNA methyltransferase Efflux (*cmlA*) Acetylation (*cat*)
[S]Synthetic antibiotic class			

Antibiotics: natural products of evolution

Most antibiotics in clinical practice today are natural products generated as secondary metabolites by bacteria, predominantly those of the order Actinomycete isolated from soil (Perry et al., 2016). Synthetic compounds also have a history of success as

antibacterials. However, the "Golden Age" of antibiotic discovery in the mid-1900s was dominated by molecules isolated from bacteria (Brown & Wright, 2016). As Brown and Wright elegantly describe, nature has been perfecting efficient molecules throughout millions of years of evolution to create "a cascade of pleiotropic effects ultimately resulting in cell death" (Brown & Wright, 2016).

While antimicrobial peptides and defensins produced by humans and other mammals have roles in host innate immunity, the role of microbially-produced antibiotics in nature remains contested (Ganz, 2003). The primary role of these secondary metabolites may likely be to reduce competition by killing surrounding bacteria when resources are scarce. They are also proposed to serve a role in signalling with neighbouring organisms (Yim et al., 2007). At sub-inhibitory concentrations, antibiotics trigger differential expression of specific genes in targeted bacteria (Goh et al., 2002; Yim et al., 2006). The idea of these molecules as microbial tools of warfare is an attractive one. However, the concentrations of antibiotics used under laboratory conditions to suppress growth may not reflect true environmental amounts (Davies & Davies, 2010; Gottlieb, 1976). Given their low abundance in the environment, would there be sufficient selective pressure for other bacteria to maintain such an expansive diversity of antibiotic resistance genes? (Davies & Davies, 2010). This also brings into question the origin and natural roles of antibiotic resistance mechanisms that provide such high-level protection (Aminov, 2009; Morar & Wright, 2010). Genomics and phylogenetics have begun to shed light on the natural history of antibiotics and antibiotic resistance.

The ancient history of antibiotics and antibiotic resistance

With advances in next-generation sequencing and increasing efforts to decode the genetic information, or genomes, of antibiotic producers, we are beginning to unravel the evolutionary history of antibiotics. Antibiotics are typically encoded by multiple genes closely organized within a region in the genome referred to as a biosynthetic gene cluster (BGC). Genes implicated in β-lactam antibiotic BCGs were first believed to have transferred between fungi and bacteria (directionality unknown) about 370 million years ago (MYA) (Aharonowitz & Cohen, 1992; Weigel et al., 1988). Glycopeptide antibiotic BCGs, including those that produce vancomycin, are predicted to have evolved 150 to 400 MYA (Waglechner et al., 2019). Estimates for the origins of other antibiotics include 610 MYA for streptomycin, 800 MYA for erythromycin, and over a billion years ago for daptomycin (Baltz, 2010; Waglechner et al., 2021).

Since antibiotics have been evolving in microorganisms for millions of years, it is not surprising that the origin of antibiotic resistance can be dated to similar time scales. Phylogenetic analyses of various β-lactam resistance genes suggest certain families are over 2 billion years old (Barlow & Hall, 2002; Hall & Barlow, 2004). Vancomycin resistance genes are also estimated to have evolved around the same time as vancomycin BGCs (Waglechner et al., 2019). Studies of the evolutionary history of tetracycline and erythromycin resistance genes suggest that antibiotic resistance pre-dates human use of these antibiotics (Aminov & Mackie, 2007). A ground-breaking study of 30,000-year-old permafrost confirms that antibiotic resistance is indeed ancient, further increasing interest

in studying ancient DNA from various sources (D'Costa et al., 2011; Santiago-Rodriguez et al., 2015; Warinner et al., 2014). Finally, bacteria resistant to over a dozen antibiotics exist in a cave environment that has been isolated from the surface for over 4 million years (Bhullar et al., 2012; Pawlowski et al., 2016). Therefore, although antibiotic use by humans in the past 80 years has exerted immense selective pressure on bacteria to acquire and maintain antibiotic resistance, the mechanisms and genes involved have been evolving for millions, if not billions, of years.

MECHANISMS OF ANTIBIOTIC RESISTANCE

Antibiotic resistance is predominantly genetically encoded, although certain physiological differences and growth adaptations, such as biofilm production and persistence, can be associated with reduced susceptibility. Gram-negative bacteria generally show wider ranges of AMR compared to Gram-positive bacteria (Walsh & Wencewicz, 2016). The former group of bacteria has a dual membrane cell wall that provides an additional layer of protection from and reduced permeability to antibiotics (Cox & Wright, 2013; Walsh & Wencewicz, 2016). Some bacteria have thick “waxy” coats or cell envelopes, typical of *Mycobacterium tuberculosis*, which can limit the entry of antibiotics into the cell (Brennan, 2003; Walsh & Wencewicz, 2016). Biofilms are collections of bacteria that grow in a community and typically secrete protective extracellular polysaccharides that limit entry or repel antibiotics (Walsh & Wencewicz, 2016). Persistence is a phenomenon whereby bacteria slow their metabolism and become quiescent, thus avoiding antibiotics that act on actively dividing cells (Walsh &

Wencewicz, 2016). Finally, some bacteria can catabolize or subsist on antibiotics as a carbon source (Dantas et al., 2008). General differences in the physiology of bacteria and varying growth adaptations will not be discussed further as mechanisms of antibiotic resistance in this thesis but are further complicating factors in understanding how and when bacteria develop resistance.

Antibiotic resistance is often classified as innate or acquired. Innate, or intrinsic resistance, are mechanisms that naturally exist in bacteria or are typically found in antibiotic producers for self-protection, often encoded within BGCs (Cox & Wright, 2013; Cundliffe, 1989; Hopwood, 2007; Kharel et al., 2004). There is expansive evidence to suggest many resistance mechanisms have transferred from antibiotic producers or other environmental bacteria into modern pathogens (Benveniste & Davies, 1973; Finley et al., 2013; Humeniuk et al., 2002; Jiang et al., 2017; Wright, 2007). Acquired resistance is likened to the development of resistance in previously susceptible organisms, likely due to the selective pressures of antibiotic exposure. This occurs either through the selective growth of a sub-population with an advantageous mutation or the movement of genes, through horizontal transfer, from resistant to susceptible bacteria.

Antibiotic resistance caused by mutations in coding sequences

At the genetic level, antibiotic resistance can be simplified into mutational resistance and resistance that arises due to the acquisition of specific genes (Table 1). At the core of bacterial replication, genomic and accessory (e.g., plasmid) DNA is copied and passed onto daughter cells vertically. This process is predominantly high-fidelity, however,

approximately one error will be fixed in every 10^6 base pairs of DNA (Walsh & Wencewicz, 2016). Subinhibitory concentrations of antibiotics and oxidative stress can increase mutation rates in bacteria (Gillespie et al., 2005; O'Sullivan et al., 2005). Mutations can be synonymous (ie. no change in the resulting coding sequence and protein) or non-synonymous (ie. disrupt the coding sequence by changing an amino acid). Mutations often incur a fitness cost for the bacterium, which may be outcompeted by wildtype members of the population. If, however, the mutation allows the protein to retain activity with little or no fitness cost and prevents an antibiotic from interacting with the target, this mutant will survive antibiotic selective pressure (Andersson & Hughes, 2010).

Specific mutations in the coding sequences for the targets of many antibiotics have been well-documented (Table 1). Mutations often arise in core locations in the gene, likely where antibiotics would interact with the corresponding protein (Figure 1B). For example, mutations resulting in quinolone resistance are frequent in a region of *gyrA* referred to as the quinolone resistance determining region (QRDR) (Hooper & Jacoby, 2015). There are various hotspots for rifampicin-resistant mutations in the DNA-directed RNA polymerase subunit-B (*rpoB*) (O'Sullivan et al., 2005). Sulfonamide and trimethoprim resistance is often associated with mutations in the *folP* and *dfr* genes, respectively, while recombination of *folP* can also arise (Huovinen, 2001; Skold, 2000). Regarding β-lactam resistance, point mutations in PBPs reduce antibiotic affinity to this target (Zapun et al., 2008). Finally, mutations in the 16S rRNA, 23S rRNA, or other ribosomal proteins can confer resistance to aminoglycosides, tetracyclines, or macrolides (Maus et al., 2005).

Although less common, mutational resistance can arise through changes in the regulation of the target or gene amplification. Reduced target protein expression is often the result of mutations in the promoter of the target, as is seen with quinolone resistance (Ince & Hooper, 2003). In terms of gene amplification, chromosomal gene duplications are less frequent than the capture of resistant forms of a target gene on plasmids (Sandegren & Andersson, 2009). However, the duplication of genes involved in folate biosynthesis has been shown to confer resistance to trimethoprim and sulfonamides in *Streptococcus agalactiae* (Brochet et al., 2008).

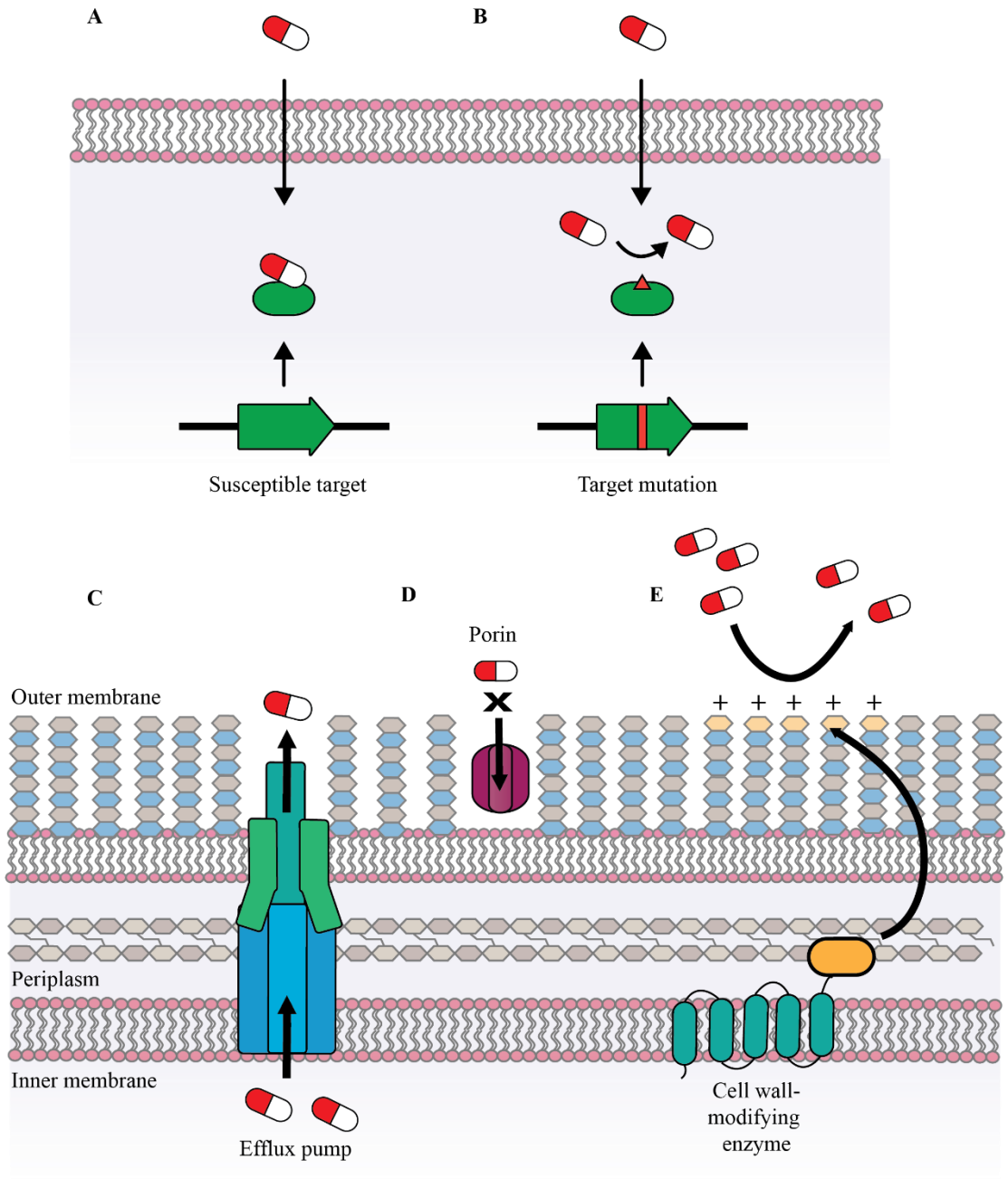


Figure 1: Mechanisms of antibiotic resistance primarily associated with mutations
A) Antibiotics typically interact with a target (either protein, RNA, DNA, or a cell wall component) to result in bacterial growth inhibition or death. B) When a mutation occurs in the gene encoding an antibiotic's target, this can result in reduced binding affinity and

survival of bacteria. C) Membrane-spanning efflux pumps actively remove antibiotics from bacterial cells. D) Reduced expression of porins or modified porins prevent the entry of antibiotics into the bacterial cell. E) Acquisition and altered expression of cell-wall modifying enzymes prevent interaction and entry of antibiotics into the cell. Figure adapted from (Guitor & Wright, 2018).

Changes to the import or export of antibiotics through decreased porin expression, increased efflux pump expression, or modified outer membrane chemistry often arise through mutation. Porins are protein channels that line the outer membrane of Gram-negative bacteria and allow for the often highly selective diffusion or entry of molecules into the cell (Figure 1D) (Cox & Wright, 2013). The expression of porins, including the Omp family in *Escherichia coli* and Opr family in *Pseudomonas aeruginosa*, often decrease in the presence of antibiotics through complex regulatory networks (Pagès et al., 2008). More selective porin proteins may replace promiscuous ones, or mutations can arise in the porins themselves (Pagès et al., 2008). Reduced porin expression has been implicated predominantly with quinolone and β-lactam resistance but is likely to result in a multi-drug resistant (MDR) phenotype (Hooper & Jacoby, 2015; Pagès et al., 2008).

Efflux pumps are membrane-spanning complexes that actively remove substrates, including antibiotics, from the bacterial cell (Figure 1C). These resistance mechanisms can be specific to one class of antibiotics, such as the tetracyclines, or they can confer resistance to multiple antibiotics (ie. MDR efflux pumps) (Cox & Wright, 2013; Piddock, 2006). In Gram-negative bacteria, tripartite efflux systems, such AcrAB-TolC, span the inner membrane, periplasm, and outer membrane to ensure the efficient removal of antibiotics (Piddock, 2006). Gram-positive bacteria also encode efflux systems, for example, NorA in

Staphylococcus aureus confers resistance to quinolones (Hooper & Jacoby, 2015; Piddock, 2006). There are five major families of efflux pumps that are categorized based on the number of components involved, the substrate that is recognized, and the energy source for the system. These are the ATP-binding cassette (ABC), the multidrug and toxic-compound exclusion (MATE), the major facilitator superfamily (MFS), the resistance nodulation division (RND), and the small multidrug resistance (SMR) family (Piddock, 2006; Walsh & Wencewicz, 2016).

Some efflux pumps are expressed only in response to the presence of an antibiotic, while others respond to various environmental signals. There is an intricate network of activators and repressors that are involved in the regulation of efflux pump expression, and mutations in any of these factors can result in increased resistance (Grkovic et al., 2002). In *P. aeruginosa*, mutations in the repressor, *mexR*, result in the upregulation of the MexAB-OprM system and a subsequent increase in resistance to various antibiotics (Ziha-Zarifi et al., 1999). Similar effects are seen with mutations associated with repressors resulting in activation of the NorA efflux system in *S. aureus*, the multidrug efflux system in *Neisseria gonorrhoeae*, and the AcrAB-TolC system in *E. coli* (Grkovic et al., 2002). In contrast, a mutation in an activator, as is the case with *marA* in *E. coli*, leads to increased transcription of the AcrAB-TolC system (Grkovic et al., 2002). While these cases mentioned above are due to mutations in chromosomally encoded efflux systems, there are various efflux genes involved in tetracycline (ex. *tet*(A)), quinolone (ex. *qepA* and *oxqAB*),

and macrolide (ex. *mef*(A)) resistance that are transferable through mobile genetic elements (MGEs) (Chopra & Roberts, 2001; Del Grosso et al., 2002; Hooper & Jacoby, 2015).

Resistance to antibiotics that target components of the cell membrane of bacteria, including daptomycin and polymyxin, arises from mutations that result in changes to the lipid chemistry in the membrane. Daptomycin resistance typically results from variations in the cell membrane that electrostatically repel the antibiotic from the cell. These modifications occur from mutations or deletions in various genes involved in phospholipid synthesis and cell wall biosynthesis (eg. *mprF* in *S. aureus*) (Figure 1E) (Peleg et al., 2012). Polymyxin resistance can arise through the addition of cationic groups (e.g. phosphoethanolamine) to the lipid A component of lipopolysaccharide (LPS) (Velkov et al., 2013). These physiochemical changes are often due to mutations in two-component regulatory systems, including PhoB-PhoQ and PmrA-PmrB, which regulate the operon responsible for the modifications of lipid A (Miller et al., 2011; Moskowitz et al., 2012). Many of these mutations have been identified in isolates of MDR *P. aeruginosa* that frequently cause chronic lung infections in patients with cystic fibrosis (Miller et al., 2011; Moskowitz et al., 2012; Velkov et al., 2013).

Protein-based or acquired antibiotic resistance

Mutational antibiotic resistance is predominantly vertically transmitted; however, resistant versions of target genes can be captured on horizontally transferable elements and replace the susceptible target in the host. This is seen with the mutant PBP2a (*mecA*), which has a lower affinity for methicillin than endogenous PBPs. This gene is clustered with the

regulatory elements *mecI* and *mecR1* in the staphylococcal cassette chromosomal element (SCC*mecI*) (Walsh & Wencewicz, 2016; Zapun et al., 2008). Resistant versions of the targets of trimethoprim (ie. *dfr* genes) and sulfonamides (ie. *sul* genes) mediate high levels of resistance and have been extensively mobilized across many bacterial genera (Huovinen, 2001; Skold, 2000). The *sul* genes were the first mechanisms to be associated with transferable antibiotic resistance and were identified in *Enterobacteriaceae* in Japan and the United Kingdom in the late 1950s (Akiba et al., 1960; Datta, 1962; Davies, 1995).

Protein-based forms of acquired antibiotic resistance encompass the remaining mechanisms and result from the expression of specific enzymes. These enzymes can act on the antibiotic scaffold or the target to introduce modifications or provide protection and limit the interaction between antibiotic and target. These genes are transmitted vertically and horizontally between bacteria of the same species and potentially across genera or families of bacteria. As mentioned previously, some of these genes likely originated in antibiotic-producing organisms or evolved from some similar sequence not originally programmed as a resistance enzyme. Many antibiotics are similar in structure to other compounds that are substrates for certain enzymes. Genomic enzymology has revealed potential evolutionary links between antibiotic resistance enzymes and proteins involved in other cellular functions (Morar & Wright, 2010). Enzymes that are believed to share a common ancestor include: the β-lactamases and penicillin-binding proteins; the aminoglycoside or macrolide phosphotransferases and protein kinases; and the

adenylyltransferases and family X polymerases (Aravind & Koonin, 1999; Morar & Wright, 2010; Walsh & Wencewicz, 2016; Wright, 2007).

For many antibiotic resistance mechanisms, there has been a point in their evolutionary history where they were captured by a MGE and horizontally transferred from their original hosts. ARGs can be captured as gene cassettes on integrons that encode a site-specific recombinase and can insert into the chromosome or on extrachromosomal elements such as plasmids (Domingues et al., 2012). Integrons can be further mobilized by association with insertion sequences (IS) that typically encode a transposase and are flanked by inverted repeats (IRs). When two IS elements flank an integron, this results in a composite transposon. Transposons enable their transfer as integrative and conjugative elements (ICEs) or can be embedded in other MGEs (Domingues et al., 2012). The horizontal transfer of ARGs is mediated through transformation (ie. the uptake of DNA from the environment by naturally competent bacteria), conjugation (ie. contact-dependent transfer of DNA from one bacterium to another), and transduction (ie. bacteriophage-mediated transfer of DNA) (Lerminiaux & Cameron, 2019).

Sub-inhibitory concentrations of antibiotics, activation of the SOS response, and even antidepressants have been shown to promote the horizontal gene transfer (HGT) of antibiotic resistance (Bahl et al., 2004; Beaber et al., 2003; Ding et al., 2022; Knapp et al., 2008). Human antibiotic use, however, is believed to be the greatest selective pressure for the mobilization of AMR. Studies of plasmids in strain collections from the pre-antibiotic era report low to no prevalence of mobilized ARGs (Datta & Hughes, 1982; Hughes &

Datta, 1983; Jones & Stanley, 1992). Conversely, studies of ancient permafrost and human samples have identified potential ARGs on transposons and plasmids, suggesting mobilization of certain ARGs may have occurred in the 'pre-antibiotic' era (Petrova et al., 2009, 2011; Petrova et al., 2014; Warinner et al., 2014).

Antibiotic-inactivating enzymes generally recognize the antibiotic as a substrate and result in the cleavage of a bond, thereby inactivating or destroying the chemical structure and leading to further degradation (Figure 2A). β-lactamases are one of the most rigorously curated groups of antibiotic inactivators, with more than 50,000 sequences reported in NCBI (Frohlich et al., 2021). Only a portion of these sequences has been functionally verified or implicated with clinical pathogens (Berglund et al., 2019). In Gram-negative bacteria, the β-lactamases are typically secreted into the periplasm through an encoded signal peptide, where they form an intermediate with and hydrolyze the β-lactam ring (Walsh & Wencewicz, 2016). β-lactamases are typically classified into four groups based on their hydrolysis mechanism and structure: classes A, C, and D employ an active site serine, while class B metallo-β-lactamases rely on zinc in their active site (Bush, 2018). The evolution of multiple lineages of serine β-lactamases from the PBP superfamily is a prime example of convergent evolution exploring multiple molecular solutions toward β-lactam inactivation (Frohlich et al., 2021).

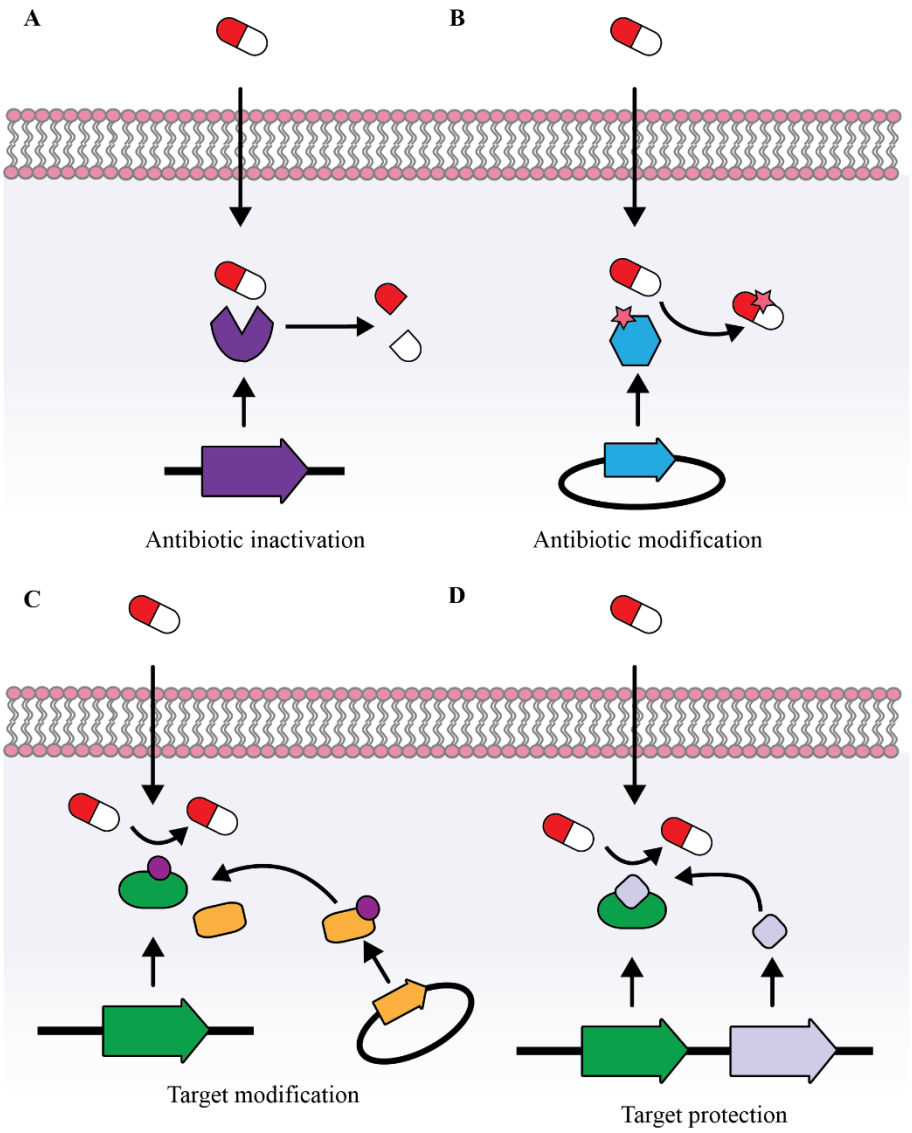


Figure 2: Acquired mechanisms of antibiotic resistance
Many mechanisms of antibiotic resistance are acquired through horizontal gene transfer. A) Antibiotic-inactivating enzymes typically result in the cleavage of a chemical bond within the antibiotic scaffold, ultimately leading to its inactivation or destruction. B) Certain enzymes can modify an antibiotic structure by adding various chemical groups. In this panel the antibiotic resistance gene is depicted on a plasmid. C) Various acquired antibiotic resistance mechanisms result in the modification of the target. D) Finally, other

proteins can specifically interact with an antibiotic's target and result in its protection. Figure adapted from (Guitor & Wright, 2018).

Some β-lactamases are specific to one scaffold, while others have broader activity and are referred to as extended-spectrum β-lactamases (ESBLs). Within each molecular class, there are further subgroups and families. Class A enzymes target various antibiotic substrates and include enzymes from the PC1, TEM, SHV, CTX-M, and KPC families. The PC1 penicillinase is encoded by *blaZ* and has disseminated amongst *Staphylococci sp*. (Bush, 2018). The TEM and SHV β-lactamases target penicillins and cephalosporins, but variants with only a few amino acid differences have emerged with ESBL phenotypes. These families span multiple subgroups and were first identified in *E. coli* (ie. $bla_{\text{TEM-1}}$) and *Klebsiella pneumoniae* (ie. $bla_{\text{SHV-1}}$) (Bush, 2018; Bush & Jacoby, 2010). Another worrisome class A β-lactamase family is the CTX-M family (eg. $bla_{\text{CTX-M-15}}$), which is spread amongst enteric bacteria and confers resistance to extended-spectrum cephalosporins (Bush & Jacoby, 2010). Class C enzymes are cephalosporinases and include the AmpC family that is chromosomally encoded in many members of Enterobacteriaceae, as well as the CMY and ACT families (Bush & Jacoby, 2010). AmpC expression is inducible through β-lactam exposure, but mutations have been associated with de-repression and constitutive resistance (Bush, 2018). Class D includes enzymes that hydrolyze cloxacillin or oxacillin and is exemplified by the OXA family of β-lactamases often associated with MDR Enterobacteriaceae such as *Acinetobacter baumannii* (Bush, 2018; Bush & Jacoby, 2010). Mobile forms of the class B metallo-β-lactamases (IMP and

VIM families) with activity against carbapenems were detected in the 1990s (Bush, 2018). More recently, the NDM family was identified in *K. pneumoniae* and is perhaps the most widespread metallo-β-lactamase (Bush, 2018).

Other antibiotic-inactivating enzymes target rifamycin, tetracycline, streptogramin B, and fosfomycin antibiotics. Decomposition of rifamycin antibiotics is achieved through monooxygenation by enzymes of the Rox or Iri family identified predominantly in species of *Rhodococcus*, *Streptomyces*, and *Nocardia* (Andersen et al., 1997; Hoshino et al., 2010; Koteva et al., 2018; Surette et al., 2021). Other mechanisms that inactivate rifamycin by modifying the antibiotic scaffold include glycosylation (ie. *rgt*), phosphorylation (ie. *rph*), and ADP-ribosylation (ie. *arr*) (Quan et al., 1997; Spanogiannopoulos et al., 2012; Spanogiannopoulos et al., 2014). The addition of any of these groups to the structure of rifampin prevents the interaction with RpoB (Spanogiannopoulos et al., 2012). The Arr family has been detected on MGEs and has thus spread into Gram-negative pathogens, *rph* genes are widespread amongst various bacterial classes, including potential pathogens, while *rgt* genes are more common in environmental bacteria as well pathogenic *Nocardia spp.* (Spanogiannopoulos et al., 2012; Surette et al., 2021).

Regarding tetracycline, the destruction of this antibiotic via oxygenation encoded by the *tetX* gene and related family members is commonly mobilized on transposable elements within *Bacteroides* and other bacteria (Moore et al., 2005; Yang et al., 2004). Streptogramin B is destroyed by a C-O lyase, the virginiamycin factor B hydrolase (vgb genes) (Mukhtar et al., 2001). Finally, epoxide ring opening via glutathione-S-transferases (ie. FosA), bacillithiol-S-transferases (ie. FosB), or hydrolyases (ie. FosX) results in the inactivation of fosfomycin

(Fillgrove et al., 2003; Silver, 2017). The FosA family is linked to MGEs and has been identified in MDR strains of *E. coli*, *K. pneumoniae*, and *P. aeruginosa* (Silver, 2017). Finally, macrolides are hydrolyzed via esterification by Ere enzymes (Roberts et al., 1999).

While many of the modifications mentioned above destroy the antibiotic scaffold, other additions impede the interaction with the target while maintaining the core structure of the antibiotic (Figure 2B). Modifications affect aminoglycosides, chloramphenicol, macrolides, streptogramins, clindamycin, and rifampicins (as noted above). Perhaps the largest group of modifying enzymes target aminoglycosides with acetyl, phosphoryl, adenylate amino, and hydroxyl group additions possible at many sites on the antibiotic (Krause et al., 2016). These group transfer reactions result in reduced affinity towards the ribosomal target. These modifications are encoded by the aminoglycoside *N*-acetyltransferases (AACs), aminoglycoside *O*-nucleotidyltransferases (ANTs), and aminoglycoside *O*-phosphotransferases (APHs) families (Krause et al., 2016). Aminoglycoside modifying enzymes (AMEs) are notorious for their dissemination through MGEs and association with other resistance elements, particularly in Enterobacteriaceae, *A. baumannii*, and *P. aeruginosa* (Krause et al., 2016). One aminoglycoside acetyltransferase (AAC(6')-Ib-cr) has evolved activity against quinolone antibiotics as well (Hooper & Jacoby, 2015).

Acetylation, or the addition of an acetyl group, also targets chloramphenicol via the *cat* gene family and Streptogramin A by *vat* (virginiamycin factor A acetylation) genes (Roberts et al., 1999; Shaw, 1983). Adenylation, or the addition of a nucleotidyl group, has been associated with lincosamides via the lincosamide nucleotidyltransferases (*lnu* genes) (Roberts et al.,

1999). Finally, macrolides can be modified via phosphorylation through the Mph family and glycosylation by Mgt enzymes (Cundliffe, 1992; Roberts et al., 1999).

Antibiotics interact intimately with their target molecules (ie. DNA, RNA, or proteins). Similar to the effect of mutations to rRNA, methylation of 16S or 23S rRNAs can prevent antibiotics from interacting with the ribosome and preventing translation (Figure 2C). Methyltransferases confer resistance to aminoglycosides via *arm* and *rmt* genes, or MLS_B type (macrolides, lincosamides, streptogramin B) antibiotics via *erm* genes (Galimand et al., 2003; Roberts et al., 1999). These genes are often plasmid-borne and have spread among many Gram-negative pathogens (Galimand et al., 2003; Roberts et al., 1999). More recently, a 23S rRNA methyltransferase (*cfr*) that confers resistance to chloramphenicol was identified in *S. aureus* and on plasmids (Arias et al., 2008).

Rather than direct modification of the target, some proteins bind to the target and prevent interaction with the antibiotic (Figure 2D). Ribosomal protection proteins (ex. *tet*(O), *tet*(M)) are homologous to elongation factors and bind to the ribosome to provide tetracycline resistance (Chopra & Roberts, 2001). Finally, Qnr proteins, commonly found on plasmids, provide quinolone resistance by interacting with and protecting DNA gyrase (Hooper & Jacoby, 2015).

The final category of protein-based antibiotic resistance includes modifications to the structure of the cell wall and precursors of peptidoglycan that results in polymyxin and vancomycin resistance, respectively. Like the changes that occur due to mutations in various regulatory genes, mobilizable colistin resistance genes (*mcr*) result in the addition of

phosphoethanolamine to lipid A, leading to a reduction in the negative charge on the cell membrane (Liu et al., 2016). The primary mechanism of vancomycin resistance involves the replacement of the D-alanyl-D-alanyl target within peptidoglycan with a D-alanyl-D-lactate, which reduces the binding affinity of the antibiotic by 1,000-fold (Bugg et al., 1991). The enzymes involved in this process have been identified in both vancomycin producers and pathogens, including vancomycin-resistant *Enterococcus faecium* (VRE). In VRE, vancomycin resistance has been captured on a transposable element and includes a regulatory system (*vanRS*) that controls the genes involved in the modification: an oxido-reductase (*vanH*), an ATP-dependent ligase (*vanA*), and a dipeptidase (*vanX*) (Marshall et al., 1998). Other operons characterized by the ligases VanB, VanD, and VanM result in the same D-alanyl-D-lactate modification. Alternatively, a D-alanyl-D-serine terminus is encoded by ligases VanC, VanE, VanG, VanL, and VanN (Patel & Richter, 2015).

Factors influencing the maintenance of antibiotic resistance

Certain genes may have a higher propensity to be captured and mobilized over others (Nielsen et al., 2022; Waglechner & Wright, 2017). Similar to mutations, the success of dissemination of these genes can be related to the fitness costs incurred in the host, but also whether the MGE contains the correct machinery for replication and expression, as well as compatibility between the codon usage of the gene and the host (Sorek et al., 2007). Fitness costs are often due to the additional energy and resources required to express an ARG. Without antibiotic pressure, this may reduce the ability to compete with sensitive bacteria. One could argue that ARGs with a low fitness cost and a broad spectrum of

activity are more favourable to maintain and more likely to be mobilized then disseminated in a population.

Regulation can alleviate some amount of fitness cost. Certain ARGs are constitutively expressed, while others are under strict regulatory control. Efflux-mediated tetracycline resistance due to TetA is activated by release of the repressor TetR upon binding to the antibiotic (Grkovic et al., 2002). The two-component regulatory system VanRS senses vancomycin in the environment and activates the expression of VanHAX (Walsh & Wencewicz, 2016). Certain, predominantly chromosomally encoded β-lactamases (AmpC and blaZ) and *mecA* are activated following the recognition of β-lactam antibiotics by their corresponding regulators (Walsh & Wencewicz, 2016; Zapun et al., 2008). Finally, Erm ribosomal methyltransferases have a unique mechanism of translational attenuation that is relieved by macrolides (Roberts et al., 1999). The evolution of various pathways to ensure ARGs are only expressed in the presence of an antibiotic may select for maintenance.

Co-selection of ARGs occurs when genes conferring resistance to different antibiotics are located within the same genetic material, whether that be within the chromosome or an MGE. Thus, one ARG is selected for given the exposure of an antibiotic that targets the other ARG (Andersson & Hughes, 2010). The capture and maintenance of multiple ARGs within single MGEs has led to organisms acquiring multi-drug and extant-drug resistance (XDR). Many plasmids encode other beneficial genetic information,

including virulence factors or metal resistance, and exposure to selective pressures for these other markers will help maintain ARGs in the population.

Perhaps over time, without further antibiotic selection, ARGs may be lost. In theory, if an antibiotic resistance gene confers a fitness cost, it would be beneficial for the bacteria to lose that gene when antibiotics are removed from their environment (Andersson & Hughes, 2010). In reality, resistance reversal is infrequent due to additional features such as compensatory mutations and co-selection (Handel et al., 2006). It may take months to years for resistance to be reduced in the community after experiencing a selective pressure (Sjölund et al., 2005; Sjölund et al., 2003). There are some cases where a resistant mutation in an essential gene confers a fitness advantage in the absence of antibiotics and thus is overall beneficial (Andersson & Hughes, 2010; Luo et al., 2005; Wright, 2007). Given our reliance on antibiotics and the lack of regulation for their use in various countries, it will be difficult to limit bacterial exposure to antibiotics and reverse antibiotic resistance.

WHY SHOULD WE STUDY THE ANTIBIOTIC RESISTOME?

Given the complexity of the various mechanisms implicated in antibiotic resistance, the genetic component of this phenomenon has been coined the resistome (D'Costa et al., 2006). The resistome refers to the collection of all genes involved in reducing the susceptibility of bacteria to antibiotics. This may focus on a sole bacterium, a population of strains from the same species, or a more complex metagenome (ie. all genetic material from an environment or microbiome). A microbiome is often defined as the collection of microorganisms within a defined environment and considers the abiotic factors in addition

to microbial interactions within this niche (Berg et al., 2020). The metagenomic fraction of the microbiome describes the compendium of all genetic information from bacteria, fungi, parasites, viruses, and eDNA in that environment (Adu-Oppong et al., 2017; Turnbaugh et al., 2007). The resistome encompasses genes from pathogens, commensals, and environmental bacteria. All bona-fide antibiotic resistance genes, cryptic genes with low to no expression or detectable resistance, and genes that may evolve into ARGs, referred to as proto-resistance or precursor genes, encompass the resistome (Wright, 2007). Understanding the resistome in any environment may enable the prediction of new and emerging threats of AMR.

In the last 20 years, the resistome of many environments has been extensively characterized; from the soil, various human microbiomes, animals and agriculture, wastewater, pristine or untouched environments, and even historical samples that pre-date the Anthropocene (Allen et al., 2010; Allen et al., 2009; Bengtsson-Palme et al., 2016; Bhullar et al., 2012; D'Costa et al., 2007; D'Costa et al., 2011; Danko et al., 2021; Pawlowski et al., 2016; Riesenfeld et al., 2004; Sommer et al., 2009). A major shift in this research has been due to the development of high-throughput methods, including next-generation sequencing. One study of 303 million prokaryotic genes found that, in general, most genes are unique to a given habitat, but ARGs are among the most likely to span multiple habitats (Coelho et al., 2022). This speaks to the dissemination of AMR across various environments and the frequency of ARG capture within MGEs.

Antibiotic resistance is a One Health problem, meaning it requires an understanding of AMR in all environments, from the clinic to agriculture and the soil (Cunningham et al., 2017). The large diversity of environmental resistance has highlighted this important reservoir of potential sources for clinically associated ARGs (Forsberg et al., 2014; Forsberg et al., 2012; Gibson et al., 2015; Perry et al., 2016; Surette & Wright, 2017; Waglechner & Wright, 2017; Wright, 2019). Anthropogenic human activities, including mass prophylaxis in farming, animal husbandry and agriculture, have inevitably selected for AMR maintenance and dissemination in these environments (Knapp et al., 2010; Udikovic-Kolic et al., 2014; Zhu et al., 2013). To prevent a future without antibiotics in medicine, all levels of government, academia, industry, health care, and the public must work together to prioritize understanding antibiotic resistance, practicing antibiotic stewardship, and improving incentives for antibiotic drug discovery (Perez & Villegas, 2015). We must understand which resistance mechanisms exist in bacterial pathogens circulating in human niches and those that await in the environment for an opportunity to spill over into human pathogens (Manaia, 2017; Martinez & Baquero, 2014). The following sections will describe current approaches to studying the resistome in infections and more complex environments, such as the gut microbiome.

ANTIBIOTIC RESISTANCE IN THE CLINIC

When a patient presents in the clinic with a serious infection, it can take beyond 24 hours to identify the pathogen and the antibiotics it is susceptible to (Bissonnette & Bergeron, 2010; Leekha et al., 2011). Before receiving this information, the patient is often

given broad-spectrum antibiotics, which can be unnecessary (ie. if the infection-causing agent is non-bacterial), have off-target effects on the natural microbial flora or microbiota, or can select for resistant sub-populations of bacteria. Rapid and accurate diagnostic tools and susceptibility tests can ensure narrow-spectrum treatment, which may prevent some of these consequences.

GENERAL APPROACHES TO DETECTING ANTIBIOTIC RESISTANCE

Antibiotic susceptibility testing (AST) measures the ability of antibiotics to suppress the growth of an organism (Leekha et al., 2011). When an organism can grow in concentrations of an antibiotic that would typically result in cell death or growth inhibition, they are considered resistant (Figure 3A). In the clinical microbiology laboratory, AST is typically performed on pure bacterial cultures using standardized methods (eg. broth microdilution method) from the Clinical and Laboratory Standards Institute (CLSI) in the USA or the European Committee on Antimicrobial Susceptibility Testing (EUCAST) (CLSI, 2018; EUCAST, 2003). These methods determine the minimum inhibitory concentration (MIC), which is the minimum amount of antibiotic required to inhibit further growth of that bacteria (Leekha et al., 2011; Wiegand et al., 2008). Automated systems such as the VITEK 2 from bioMérieux or the Phoenix from BD Diagnostics improve throughput and efficiency in AST in large clinical laboratories (Didelot et al., 2012; Karlowsky & Richter, 2015; Wiegand et al., 2008). Finally, Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry methods are

alternative rapid approaches to identifying the pathogen causing infection and certain AMR mechanisms (Hrabak et al., 2011; Wieser et al., 2012).

Ultimately, culture-based methods to test susceptibility rarely detect the mechanism implicated in the resistance phenotype, but a few exist. Determination of a MIC in the presence or absence of an inhibitor can detect the production of ESBLs (CLSI, 2018; Wiegand et al., 2008). Additional β-lactamase tests rely on changes in pH or colorimetric substrates, including a nitrocefin-based test and the Carba NP test (Escamilla, 1976; Nordmann et al., 2012; O'Callaghan et al., 1972). Vancomycin resistance in enterococci due to *vanC* can be detected using a methyl-alpha-D-glucopyranoside (MGP) reagent, while a latex agglutination test can identify *mecA* in MRSA (Brown & Walpole, 2001; Ramotar et al., 2000). Finally, some automated AST systems detect several antibiotic resistance mechanisms, including ESBLs (Karlowsky & Richter, 2015).

MOLECULAR METHODS TO DETECT ANTIBIOTIC RESISTANCE

All these aforementioned techniques rely on the phenotypic detection of antibiotic resistance and typically work best for bacteria that can be cultured under standard laboratory conditions. Culture-based methods are less amenable to large epidemiological studies and surveillance of transmission of AMR. For more complicated environments such as the soil or human gut, alternate methods are required to profile a broader diversity of antibiotic resistance. Molecular methods such as the amplification of ARGs, probe-based capture, and next-generation sequencing assay the genetic material, either DNA or RNA, to study the resistome of various environments. This thesis will predominantly focus on

DNA-based approaches. One major limitation of studying DNA alone is the inability to identify if a gene is being expressed and contributing to a resistance phenotype in the host bacterium or environmental sample. This can only be accomplished with transcriptomics or RNA sequencing.

PCR and Probe-based approaches

Rapid diagnostic approaches often rely on polymerase chain reaction (PCR) to exponentially amplify and detect a select few ARGs from a pure bacterial culture or patient sample (Figure 3B). This method relies on designing primers to amplify the sequence(s) of interest. Several automated systems can identify a limited set of ARGs and pathogen markers from a positive culture tube, including the BioFire FilmArray system (Berinson et al., 2021; Ledeboer & Hodinka, 2011). PCR is coupled with a probe-based approach in the GeneXpert MTB/RIF assay to detect rifampicin resistance caused by mutations in *Mycobacterium tuberculosis* (Boehme et al., 2010). A challenge with diagnostics, is the ability to provide point-of-care results and technology that is accessible in low- and middle-income settings. Loop-mediated isothermal amplification (LAMP)-based tests can be incorporated onto microfluidic chips, however, as with PCR, these approaches are limited in the number of ARGs they can target and sensitive to inhibition (Abbott & Fang, 2015; Bissonnette & Bergeron, 2010; WHO, 2019).

An alternative approach that allows for a higher number of targets with lower sensitivity than PCR, is the solid phase microarray. These platforms contain oligonucleotides attached to a slide that are complementary to the targets of interest.

Through hybridization followed by a secondary detection probe, a signal is emitted corresponding to the presence of the target (Abbott & Fang, 2015; Beal et al., 2013; Tojo et al., 2014). The Nanosphere Verigene platform couples DNA extraction, hybridization to a microarray, and detection in one system. Various panels aimed at different infections or sample types target a variety of pathogens and ARGs (Beal et al., 2013; Samuel et al., 2013; Tojo et al., 2014). Another method, molecular padlock probes, uses biotinylated oligonucleotides to capture targets of interest, followed by rolling-circle amplification and fluorescent labelling of the amplified target (Mezger et al., 2015). In general, PCR, microarrays, and padlock probes are limited in the number of ARGs they target and often cannot identify the genetic context. The latter result can inform on the pathogen that hosts an ARG and is critical when deciding on antibiotic therapy or tracking the spread of AMR in the environment.

Sequencing-based approaches to characterize the resistome

DNA sequencing technologies can identify all bacteria and genetic factors associated with antibiotic resistance in any environment imaginable. This revolutionary method has pushed the boundaries beyond studying only culturable organisms and has enabled epidemiological studies of AMR worldwide (Didelot et al., 2012; WHO, 2020a). Whole genome sequencing provides information on the genomic content of cultured isolates, including their MGEs and ARGs (Didelot et al., 2012; Hasman et al., 2014; Joensen et al., 2017). For environments such as the soil, direct patient samples, or the human gut microbiome, metagenomic sequencing provides a more holistic view of the total

DNA from both culturable and unculturable organisms, although there exist some biases with this approach (Adu-Oppong et al., 2017; Leo et al., 2020; Monier et al., 2011). Amplicon-based sequencing of microbiome samples typically relies on amplification and then sequencing of the hypervariable region of the 16S ribosomal RNA gene that is present in all bacteria (Leo et al., 2020). The AmpliSeq™ technology by Illumina offers a panel of primers to amplify over 400 ARGs for subsequent sequencing (Guernier-Cambert et al., 2021; Urbaniak et al., 2018). While these are limited targeted methods, whole metagenomic shotgun sequencing can theoretically capture information from all bacteria and other organisms in the original sample from which DNA can be extracted (Leo et al., 2020).

Next-generation sequencing technology typically refers to Illumina sequencing, which provides short, 100 to 300 base pair (bp) in length but highly accurate reads (Petersen et al., 2019). Limitations to short-read data include producing highly fragmented genome assemblies and difficulties in resolving highly repetitive regions (Baker, 2022). Recently, long-read sequencing technology (ie. third-generation) from Oxford Nanopore and PacBio has opened the potential to achieve reads greater than 10 kilobases (kb) in length (Petersen et al., 2019). Oxford Nanopore sequencing typically has a higher error rate than Illumina and PacBio sequencing; however, the portability of the MinION sequencer and rapid time to results are attractive features for clinical diagnostic potential (Petersen et al., 2019). This technology can provide rapid and accurate resistome profiling directly from patient and environmental samples (Baker, 2022; Schmidt et al., 2017; van der Helm et al., 2017).

While sequencing-based approaches can improve the speed of diagnosing both the pathogen and AMR compared to traditional culture-based approaches, there are still some limitations. In many settings, the infrastructure is not available, the cost is still too high, and the expertise to run the equipment and then analyze the data remain challenges to implementing this technology (Papaiakovou et al., 2022). The source of the sample, whether stool, soil, sputum, or blood, can greatly impact the ability to detect a bacterium of interest and its associated resistance genes due to host DNA and inhibitors (Hayden et al., 2022). To successfully capture the entire resistome of a complex metagenomic sample, ultra-deep sequencing may be needed, which results in excessive costs. One approach to prevent sequencing off-target DNA or host-derived DNA is hybridization-based enrichment before sequencing.

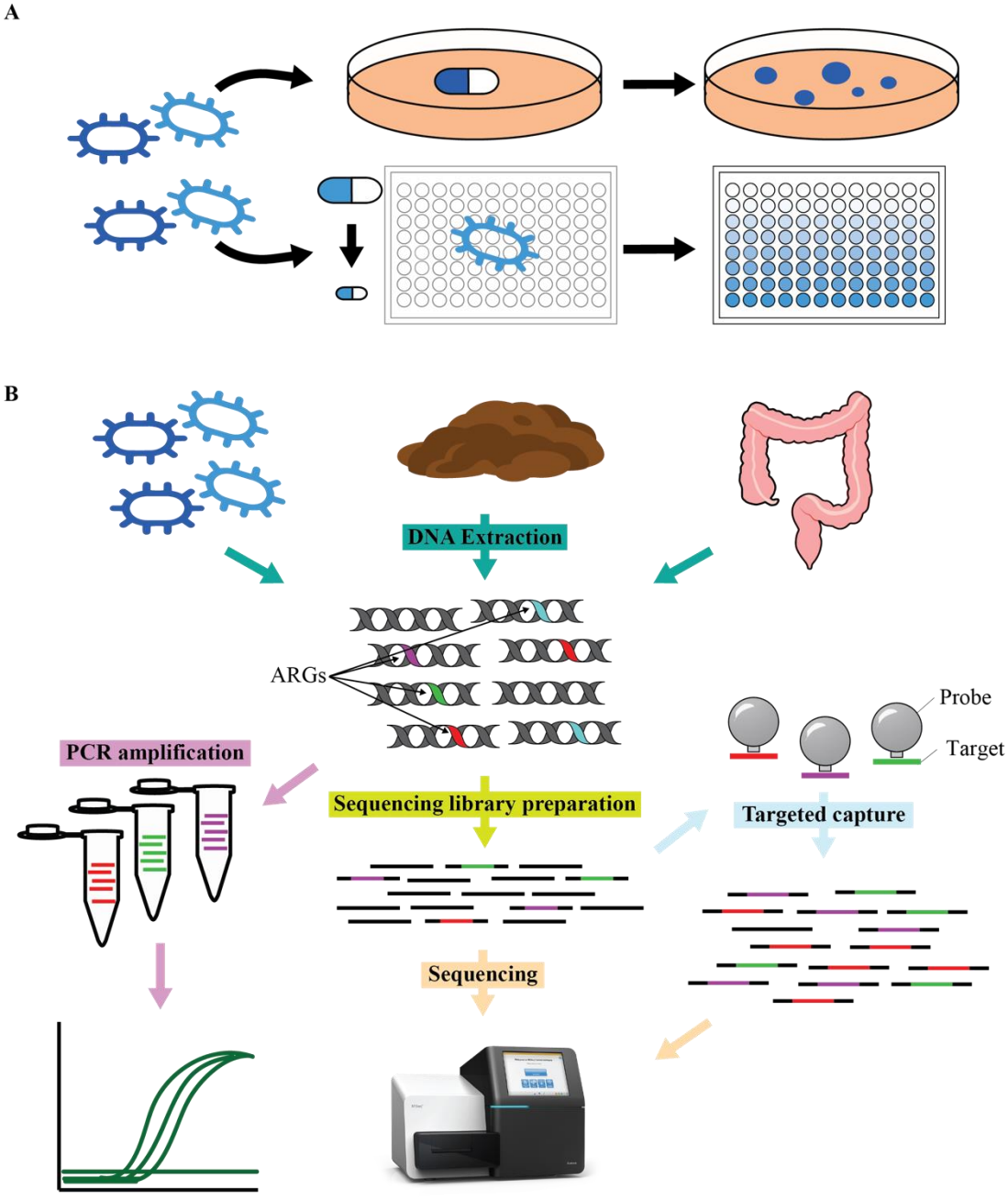


Figure 3: Methods to detect and characterize antibiotic resistance.

A) Phenotypic or culture-based methods either on solid media containing an antibiotic to test for susceptibility or in liquid broth to measure the minimum inhibitory concentration of an antibiotic. B) DNA-based methods start with the extraction of DNA from a sample (eg. cultured organisms, soil, or stool representing the gut microbiota). This DNA likely contains a variety of ARGs that can be amplified using specific primers through PCR (pink arrows) and further quantified with a fluorescent dye. Alternatively, the DNA can be prepared as a sequencing library (green arrow) that can be sequenced directly (orange arrow) or subject to targeted capture to enrich ARGs before sequencing (blue arrows).

Targeted capture for sequencing the resistome

In human stool samples, ARGs typically represent <1% of the data retrieved through metagenomic shotgun sequencing (Guitor et al., 2019). A targeted strategy such as hybridization-based sequencing typically results in on-target percentages greater than 50% and over 1,000-fold enrichment of target genes (Allicock et al., 2018; Guitor et al., 2019). This method relies on the design of RNA or DNA biotinylated probes complementary to the target sequences (Mertes et al., 2011). This technology has been employed to enrich and sequence DNA from ancient samples and capture whole genomes of specific bacteria from clinical samples, among many other applications (Brown et al., 2015; Enk et al., 2014; Gasc et al., 2016; Melnikov et al., 2011). The probes are typically designed by the researcher based on known sequences of interest and then synthesized by a manufacturer that provides all the necessary reagents (Garcia-Garcia et al., 2016). In other cases, the user can generate probes from PCR amplicons from their targets of interest (Hayden et al., 2022; Peñalba et al., 2014).

Our group and others have harnessed this technology to target over 1000s of ARGs from various sample types (Allicock et al., 2018; Beaudry, Thomas, et al., 2021; Ferreira

et al., 2021; Guitor et al., 2019; Lanza et al., 2018; Noyes et al., 2017). Our targeted capture approach avoids detecting AMR associated with mutations in the target genes (Guitor et al., 2019). In contrast, others have specifically used enrichment to monitor allelic diversity and strain diversity, opening the possibility of detecting mutational resistance (Hayden et al., 2022). It is also possible to capture larger DNA fragments and couple this with long-read sequencing to provide a broader context of the ARG in a metagenomic sample (Dapprich et al., 2016; Eckert et al., 2016; Slizovskiy et al., 2022). The design and implementation of a targeted capture probeset for Illumina sequencing of AMR is presented in Chapter 2.

Any sequence-based or hybridization-based sequencing method cannot predict completely new resistance mechanisms. To achieve this objective, functional metagenomics is an approach where DNA from any sample or environment is cloned into susceptible hosts that are then screened for the acquisition of resistance, and corresponding genetic determinants are sequenced (Boolchandani et al., 2019). This method has successfully identified new ARGs in soil, the human gut microbiome, and beyond (Allen et al., 2009; Forsberg et al., 2014; Forsberg et al., 2012; Gibson et al., 2015; Moore et al., 2013; Pehrsson et al., 2013; Sommer et al., 2009). Large functional screens, however, are quite laborious and rely on the compatibility and expression of the captured ARGs within the heterologous host (Boolchandani et al., 2019).

Analyzing sequencing data to identify antibiotic resistance

All the sequencing-based methods mentioned above rely on bioinformatic analysis of large datasets to identify ARGs. The two most common approaches to analyzing sequence data for AMR are assembly-based and read-based methods (Boolchandani et al., 2019). Assembly-based methods rely on the *de novo* reconstruction of larger fragments (contigs) from the sequencing reads (Boolchandani et al., 2019). Various tools have been developed to predict AMR from assemblies (Boolchandani et al., 2019). The following approaches are used to identify ARGs: comparing nucleotide or amino acid similarities using alignment tools such as BLAST or DIAMOND (eg. the Resistance Gene Identifier – RGI and ResFinder), using hidden Markov models (HMMs) (eg. Resfams), or applying machine-learning (eg. DeepARG) (Alcock et al., 2022; Arango-Argoty et al., 2018; Bortolaia et al., 2020; Gibson et al., 2015). Read-based methods rely on short-read aligners to map reads to a reference database (Boolchandani et al., 2019). Several tools, including SRST2, KmerResistance, and ShortBRED, use reads, kmers, and short marker sequences, respectively, to compare sequence data to a database of ARGs (Clausen et al., 2016; Inouye et al., 2014; Kaminski et al., 2015). While these methods rely on databases of curated ARGs and, therefore, can only predict known ARGs, fARGene claims to be able to predict unknown ARGs from metagenomic data (Berglund et al., 2019).

Databases for antibiotic resistance

Because antibiotic resistance mechanisms are genetically encoded and diverse, it is imperative to catalogue all the mutations and genes involved for easy and accurate referencing. There are over a dozen various specialized and general databases for AMR.

The Lahey list of β-lactamases was the first to catalogue these genes and has since been incorporated into NCBI's Bacterial Antimicrobial Resistance Reference Gene Database, which contains over 6,386 ARGs (NCBI BioProject: PRJNA313047). Other general ARG databases include Resfams, ARDB, and ARG-ANNOT (Gibson et al., 2015; Gupta et al., 2014; Liu & Pop, 2009). These, however, are no longer active or have not been updated in recent years. Databases that reflect the increasing understanding of AMR and are updated frequently include ResFinder, the NCBI database, and the Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2022; Bortolaia et al., 2020). Although databases such as CARD are highly curated and include only rigorously validated ARGs, they are selectively enriched for AMR identified in clinical pathogens. There is merit in including other ARGs that are identified from environmental sources as well as those captured through functional metagenomics in databases; this has been attempted through ResFinderFG (Dunivin et al., 2019; Wallace et al., 2017). These types of resources may help identify ARGs with the highest risk of transfer from the environment to the clinic.

ANTIBIOTIC RESISTANCE AND THE GUT MICROBIOME

One important niche where antibiotic resistance has been extensively studied is the human gut microbiome. Unsurprisingly, given the diversity of bacteria in the gut microbiome, this environment is also rich in antibiotic resistance potential (Sommer et al., 2009; van Schaik, 2015). One goal of understanding the human gut resistome is to track how it changes over the course of human life and development and what implications it may have for human health. From an early age, humans are introduced to the

microorganisms that will colonize the gut and remain their microbial partners throughout their lives (Martino et al., 2022). Many factors shape the development of the microbiome and its associated resistome. A better understanding of these factors and how they may select for the persistence of certain ARGs is important to alleviate the burden of AMR in the gut microbiome.

Estimates of the number of bacterial cells in the human body range from $10^{12} – 10^{14}$, with most of these bacteria concentrated in the large intestine or colon (Sender et al., 2016; Walsh & Wencewicz, 2016). Bacterial cells are as numerous, if not more so, than human cells (Sender et al., 2016). Analysis of over 50,000 gut metagenomes identified over 3,500 potential bacterial species in this niche, while an individual is likely to harbour fewer species, somewhere in the range of hundreds (Leviatan et al., 2022). The human gut microbiota serves invaluable functions in metabolism, extracting certain nutrients from otherwise indigestible foods, developing the immune system, and providing protection from pathogens (Turnbaugh et al., 2007; Valdes et al., 2018). Differences between healthy and diseased individuals have implicated a role of the microbiome in diabetes, obesity, allergies, irritable bowel disease, cardiovascular disease, and neurological disorders (Flint et al., 2012; Valdes et al., 2018).

The gut microbiota is not established prenatally; the first colonization occurs during birth and soon afterwards (Kennedy et al., 2021). The first few years of life are crucial in the gut microbiome's development and colonization. By the third year of life, the diversity of the microbiota has increased to the point of stabilization and resembles that of an adult

(Arrieta et al., 2014; Yatsunenko et al., 2012). Apart from their genetics, the mode of delivery, whether via caesarian section or vaginally, is the first major impact on an infant's gut microbial community (Dominguez-Bello et al., 2010). Gestational age is also an important factor, as infants born early or preterm have significant differences in their microbiota compared to full-term infants, and the gut's developmental trajectory is delayed (Arboleya et al., 2012). Pre-natal exposures of the mother, such as antibiotics and diet, also have an effect, likely through contact from skin and breastfeeding (Arrieta et al., 2014; Sonnenburg et al., 2016). From birth onwards, how infants are fed (ie. formula vs. breast milk), their subsequent solid food diets, supplementation with probiotics and prebiotics, environmental exposures, and of course, antibiotic treatments all factor into the microbiome's development (Flint et al., 2012; Sonnenburg et al., 2016; Valdes et al., 2018).

Antibiotics perturb the gut microbiome and select for resistance

Reports as early as the 1960s describe antibiotic disruption of the diversity of the human microflora and selection for AMR (Finland, 1979). Antibiotics deplete sensitive commensal organisms, alter the function of the gut microbiota, and may allow pathobionts to thrive (Zimmermann & Curtis, 2019). The time to complete restoration of the gut microbiota ranges from months to years (MacPherson et al., 2018; Zimmermann & Curtis, 2019). Longer durations of treatment or repeated antibiotic exposures may eradicate certain members of the microbiota (Blaser, 2018). For children, the effect of antibiotic disruptions may persist longer, impede the colonization of beneficial commensals, and enrich for AMR and MDR bacteria (Gasparrini et al., 2016; Gibson et al., 2016; Lebeaux et al., 2022). The

consequences of this on future health are unknown (Lebeaux et al., 2022; Leo et al., 2022; McDonnell et al., 2021). Preterm infants, compared to infants born full term, typically receive more antibiotics to prevent infections due to their immaturity (Arrieta et al., 2014; Gasparrini et al., 2016; Gibson et al., 2016). Similar to adults, recovery of the microbiota and depletion of AMR after disruption by antibiotics can take months to several years, although depending on the duration of exposure, the microbial community may be permanently altered (Arrieta et al., 2014; Fouhy et al., 2012).

Given that antibiotics have striking effects on the bacterial diversity and resistome of the gut, excessive treatment may have other negative consequences. The gut microbiome is a prime location for HGT of many genetic functions, including antibiotic resistance (Kent et al., 2020; Smillie et al., 2011). This environment is also a reservoir of potential pathogens that, if given the appropriate conditions, can cause infections (Donskey, 2004). Therefore, antibiotics may select for the transfer of ARGs between commensal and pathogenic organisms in the gut microbiome. The latter would then have the potential to cause difficult-to-treat infections.

Probiotics to prevent persistence of multi-drug resistant organisms

Interventions to promote the development of the gut microbiome, reduce rates of AMR, and prevent colonization by MDR bacteria have been proposed, including supplementation with probiotics. Probiotics are formulations of live microorganisms, or microbial-derived products, associated with benefits for the gut microbiome (McFarland, 2009). They are often indicated for recovery from oral antibiotic treatment that can impact

the gut microbiota and generally have positive outcomes (McFarland, 2009). The evidence supporting the impact of restoring the gut microbiome and preventing antibiotic resistance in adults and children is conflicting, and there are still many unknowns (Montassier et al., 2021; Su et al., 2020; Suez et al., 2019). For preterm infants, probiotics have been shown to reduce rates of necrotizing enterocolitis and promote the development of the microbiome (Samara et al., 2022). Few studies assess the possibility of reducing AMR in the gut microbiome of preterm infants through probiotic supplementation. Our study on this effect and those of others are discussed in Chapter 3.

Does mass drug administration do more harm than good?

Antimicrobial resistance is a global One Health issue; however, certain countries are disproportionally affected. Sub-Saharan Africa has the highest rates of deaths attributable to AMR (Murray et al., 2022). Many systemic issues have led to this, including unregulated access to antibiotics, antibiotics of reduced quality, poor sanitation practices, and lack of clean drinking water. Unfortunately, this region and other low- and middle-income countries (LMICs) have some of the highest childhood mortality rates (UNICEF, 2021). Antibiotics, however, are still life-saving medicines, and their potential for mass drug administration or prophylaxis in children in developing countries is currently recommended to control certain infectious diseases (O'Brien et al., 2019). To reduce mortality in children under five in regions with high death rates, the WHO has recommended mass administration of the macrolide antibiotic azithromycin (WHO, 2020b). A major concern with this recommendation is the impact this prophylactic

exposure to an antibiotic will have on selecting for AMR in these children. Studies looking at the effect of azithromycin exposure on the gut resistome, including our study of children in Botswana, are discussed in Chapter 4.

PURPOSE AND GOALS OF THIS THESIS

Antimicrobial resistance is a complicated phenomenon that requires improved methods to comprehensively detect and understand its various intricacies. Furthermore, many factors impact the development of the gut microbiome and corresponding resistome, including antibiotics and probiotics. The extent of the effect of these exposures in early life in certain populations and geographic regions is underexplored. In Chapter 2, I assess a novel set of probes designed to capture antimicrobial resistance genes from metagenomic samples. Using hybridization capture and next-generation sequencing, I show how this probeset can selectively enrich ARGs from mock metagenomic samples of MDR clinical strains of bacteria and a human stool sample. In Chapter 3, I interrogated the impact of probiotic supplementation on the gut resistome of preterm infants. In this study, I demonstrated that early life exposure to probiotics reduced the diversity of AMR in this vulnerable population at 5 months of age. In Chapter 4, I examined the effect of azithromycin exposure on the gut resistome of children treated for diarrhea in Botswana, Africa. Comparisons of the differences in prevalence and abundance of ARGs to children that received the standard of care revealed slight variations in the macrolide resistome, but no major selections for AMR. The commonality amongst these chapters is the probeset designed to target AMR and its utility in enriching the rare components of the resistome in

the gut microbiome. Chapters 3 and 4 rely on having validated this approach in Chapter 2. In the concluding chapter, I discuss future applications for this targeted capture technology, its potential role in reducing the burden of AMR, and other strategies to prevent unnecessary antibiotic use and its associated consequences.

CHAPTER TWO: Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes

CHAPTER TWO PREFACE

The work presented in this chapter has been published as:

Guitor AK, Raphenya AR, Klunk J, Kuch M, Alcock B, Surette MG, McArthur AG, Poinar HN, Wright GD. Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. *Antimicrob Agents Chemother*. 2019 Dec 20;64(1):e01324-19. doi: 10.1128/AAC.01324-19. PMID: 31611361; PMCID: PMC7187591.

A.K.G, G.D.W, H.N.P, A.G.M and M.G.S conceived the study. B.A and A.G.M designed the probeset. Experiments were designed by A.K.G with input from H.N.P, J.K, M.K and G.D.W. Linda Ejim cultured clinical isolates, Susan McCusker from the Center for Microbial Chemical Biology at McMaster University isolated genomic DNA from clinical isolates, and Laura Rossi isolated DNA from human stool. A.K.G prepared DNA libraries and performed targeted capture. The McMaster Genomics Facility performed next-generation sequencing. A.K.G and A.R.R wrote code and designed the analysis pipeline. A.K.G generated the figures. A.K.G and G.D.W wrote the manuscript with input from all authors. Additional contributions include Kara Tsang and Biren Dave for genome assembly and annotation of the clinical isolates, Nicholas Waglechner for helpful discussions and guidance pertaining to analyses, Christy Groves for valuable input on the figures, and all the members of the McMaster Ancient DNA Center.

ABSTRACT

Identification of the nucleotide sequences encoding antibiotic resistance elements and determination of their association with antibiotic resistance are critical to improve surveillance and monitor trends in antibiotic resistance. Current methods to study antibiotic resistance in various environments rely on extensive deep sequencing or laborious culturing of fastidious organisms, both of which are heavily time-consuming operations. An accurate and sensitive method to identify both rare and common resistance elements in complex metagenomic samples is needed. Referencing the sequences in the Comprehensive Antibiotic Resistance Database, we designed a set of 37,826 probes to specifically target over 2,000 nucleotide sequences associated with antibiotic resistance in clinically relevant bacteria. Testing of this probe set on DNA libraries generated from multidrug-resistant bacteria to selectively capture resistance genes reproducibly produced higher numbers of reads on target at a greater length of coverage than shotgun sequencing. We also identified additional resistance gene sequences from human gut microbiome samples that sequencing alone was not able to detect. Our method to capture the resistome enables a sensitive means of gene detection in diverse environments where genes encoding antibiotic resistance represent less than 0.1% of the metagenome.

INTRODUCTION

Antimicrobial resistance (AMR) is one of the most pressing challenges of the 21st century that poses a threat to modern medicine and food security (1). The challenge of AMR is amplified by the movement of genes between bacteria, coupled with the movement

of people and goods across the planet (2–4). One of the gaps in addressing the antibiotic resistance crisis is a lack of suitable tools to catalog the complete resistome (the entire AMR gene contingent) in various environments and associated microbiomes. Detecting the resistome of an individual bacterium, a microbiome, and other environmental settings (sediment, hospitals, etc.) will aid in tracking the spread of resistance and monitoring the emergence of new resistance alleles associated with the use of antibiotics or other bioactive compounds (5–10). This information can guide antibiotic use, in addition to informing stewardship programs and public health decisions.

Profiling the resistomes of bacteria that are culturable is reasonably straightforward using whole-genome sequencing followed by analysis using algorithms, such as the Resistance Gene Identifier (RGI) in the Comprehensive Antibiotic Resistance Database (CARD) (11). In metagenomes, where resistance determinants are relatively rare, deep sequencing, requiring millions of sequencing reads, followed by careful filtering is needed. This resource-intensive strategy can be alleviated by the targeted detection of selected genes, e.g., via PCR, microarrays, or CRISPR/Cas9-based methods (12–16). However, such highly targeted approaches suffer from the fact that they are rarely comprehensive and generally cannot account for the continual emergence of gene variants and/or completely novel mechanisms of resistance (17–19).

A more appropriate approach for the identification of resistomes in metagenomes is the use of a probe-and-capture strategy (20). Using this strategy, we and others have captured, sequenced, and reconstructed human mitochondrial sequences as well as the

genomes of infectious agents and extinct species from various environments, including highly degraded archeological and historical samples (21–26). In a probe-and-capture experiment, target RNA baits are designed to be complementary to the target DNA sequences of interest. Synthesized probes are biotin labeled and are incubated with the DNA from metagenomic or genomic libraries, where they hybridize to related sequences (Fig. 1a and b). Targets are captured using streptavidin-coated magnetic bead separation, and then the reaction mixtures are pooled and the sequences are determined on a next-generation sequencing (NGS) platform (Fig. 1c to e). This strategy offers significant advantages for the sampling of resistomes in a variety of environments where resistance genes are generally rare and genetically diverse. Indeed, recently, the use of this approach for resistance gene capture has been explored by other groups (27–29). However, these accounts target many other genes that are not rigorously associated with resistance, increasing the sequencing cost and the opportunity for false-positive gene identification.

Here, we chronicle our targeted method for the analysis of antibiotic resistomes. We based our probe set design on stringently curated AMR gene (ARG) sequences from CARD (v1.0.1, 2015), tiled across ARG sequences, combined with rigorous analysis to suppress off-target hybridization. This design enables a more cost-effective and sensitive method to sample the known resistance gene landscape (11). We tested the efficacy of this probe set and our strategy using both a panel of pathogenic bacteria with known resistance genotypes and uncharacterized human metagenomic stool samples. Our method

demonstrates the superior design and methodology of the approach, which is readily applicable to both clinical and nonclinical settings.

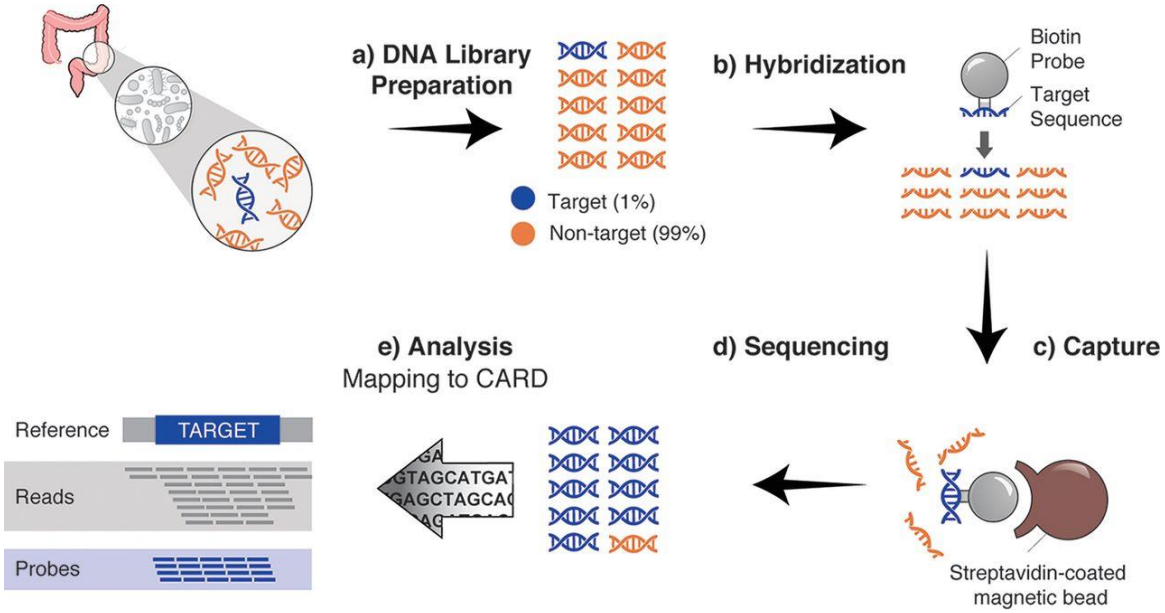


Figure 1: Platform for capture and identification of diverse antibiotic resistance genes.
The targeted capture sequencing work flow begins with DNA isolation from a sample of interest (stool from a healthy donor, in this example). (a) DNA is fragmented through sonication and prepared as a sequencing library. (b, c) Target sequences representing less than 1% of the total DNA are captured through hybridization with biotinylated probes and streptavidin-coated magnetic beads. (d, e) The captured and amplified library fragments are sequenced, and reads are analyzed for AMR gene sequence content by mapping to the sequences in CARD.

RESULTS

Design and characterization of resistance gene probes

A set of 80-mer nucleotide probes was custom designed and synthesized through the use of the myBaits platform (Arbor Biosciences, Ann Arbor, MI). The probes (n = 37,826) span the protein homolog model of curated ARGs from CARD and represent

nucleotide sequences (n = 2,021) that are well characterized in the literature. Probes targeting genes for resistance conferred through single point mutations (single nucleotide polymorphisms [SNPs], e.g., sequences contained in the protein variant model in CARD) in chromosomal metabolic genes (including DNA gyrase [*gyrA*] mutations associated with fluoroquinolone resistance and RNA polymerase subunit [*rpoB*] mutations associated with rifampin resistance) were purposefully not included in our design. Of the genes targeted by our probes, 78.03% mirrored the breakdown in CARD, dominated by genes encoding antibiotic inactivation mechanisms and by genes encoding the beta-lactamases (Fig. 2). The majority of the probes (n = 24,767) target a single gene, and the remainder target multiple genes ranging up to a maximum of 211 genes (average, 5.96 genes) due to sequence conservation within gene families (see Fig. S1A in the supplemental material). For example, a single probe initially designed to target 80 nucleotides of the beta-lactamase gene $bla_{\text{SHV-52}}$ also targets an additional 208 genes, including other members of the SHV, LEN, and OKP-A/-B beta-lactamases, due to homology between these nucleotide sequences within AMR gene families. The combination of overlap in the utility of some 80-mer probes and partial hybridization can allow probes to target sequences that are divergent from their reference sequences and thus identify new alleles at the SNP level up to 15% divergence.

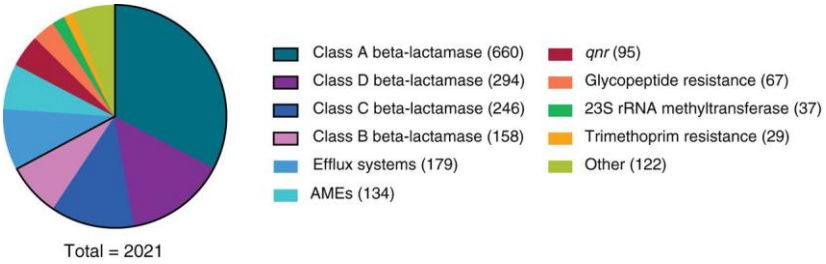


Figure 2: Design of a probe set to target over 2,000 antibiotic resistance genes. Breakdown of resistance gene classes from CARD that are targeted by probes. A legend for the top 10 classes is shown. AME, aminoglycoside-modifying enzymes; *qnr*, quinolone resistance genes. The remaining 122 genes belong to various classes. The beta-lactamase genes make up the majority of genes targeted by probes and are highlighted with a black border.

At the individual determinant level, the number of probes per gene (average, 105 probes per gene; range, 1 to 309 probes per gene) and the length coverage of a gene (average, 96.20%; range, 3.17% to 100%) vary (Fig. S1B and C). The majority of the targeted genes (2,004/2,021; 99.16%) are covered by at least 10 or more probes (Fig. S1B). Members of the beta-lactamase families (*bla*$_{\text{CTX-M}}$, *bla*$_{\text{TEM}}$, *bla*$_{\text{OXA}}$, *bla*$_{\text{GES}}$, *bla*$_{\text{SHV}}$) are among the genes with the highest probe coverage. The majority of genes (1,970/2,021) have greater than 80% length coverage by probes, 26 genes have less than 50% length coverage by probes, and only 1 (*mexW*, Antibiotic Resistance Ontology [ARO] ID: 30003031) has less than 5% length coverage by probes (Fig. S1C). Only 28 sequences from CARD have no probe coverage, due to filtering of candidate probes during the design. Overall, this probe set targets ~1.77 megabases of antibiotic resistance-associated nucleotide sequences and greater than 83% of the nucleotide sequences curated in CARD. Additional metrics of the probe set are given in Fig. S1D to H.

ARG enrichment from bacterial genomes with a range of antibiotic resistance determinants

To characterize the sensitivity and the selectivity of this probe set, we conducted a series of control experiments using a panel of sequenced multidrug-resistant Gram-positive and Gram-negative bacteria. The proportion of the genomes targeted by our probe set ranged from 0.21 to 0.97%, consisting of 13 to 65 ARGs representing 102 unique genes among the isolates tested (Table S1). Genomic DNA from four different species was tested individually via enrichment on two different library preparations (the NEBNext Ultra II library preparation versus the modified Meyer and Kircher library preparation) of various insert sizes (average library fragment size range, 396 to 1,257), referred to here as trial 1 and trial 2 (Table S2). Our enrichment approach is insensitive and tractable to different insert sizes, as there was a strong correlation between the read count on targeted regions for bacterial genomes enriched individually between the two trials (Pearson correlation, 0.811 to 0.975) (Table S3; Fig. S2).

This probe set is selective for regions associated with antibiotic resistance in these isolates, given that over 90% of the reads mapped to the respective draft bacterial genomes and the majority (greater than 85% in all cases) of the reads mapped to the small proportion (<1%) of the genome associated with resistance (Fig. 3A; Table S3). We successfully captured 100% of the targeted genes in both library preparation methods with at least 10 reads and with 100% length coverage for the four species of bacteria tested (Table S3). This represents a sensitivity ranging from 0.21% to 0.97% of the total DNA in these samples,

with successful enrichment of regions as small as 97 bp (*mexW* in *Pseudomonas aeruginosa* C0060 with a probe coverage of 2 had greater than 10 reads in both trials) and 80 bp (*crp* in *Klebsiella pneumoniae* C0050 had greater than 100 reads in both trials). Other genes that had low probe length coverage included *mdtA* (22.4% coverage by 11 probes) in *Escherichia coli* C0002, which still retained over 100 reads in both trials, and a 140-bp region of *aad(6)* (16.8% coverage by 4 probes) in *Staphylococcus aureus* C0018 that was recovered with over 1,000 reads in both trials.

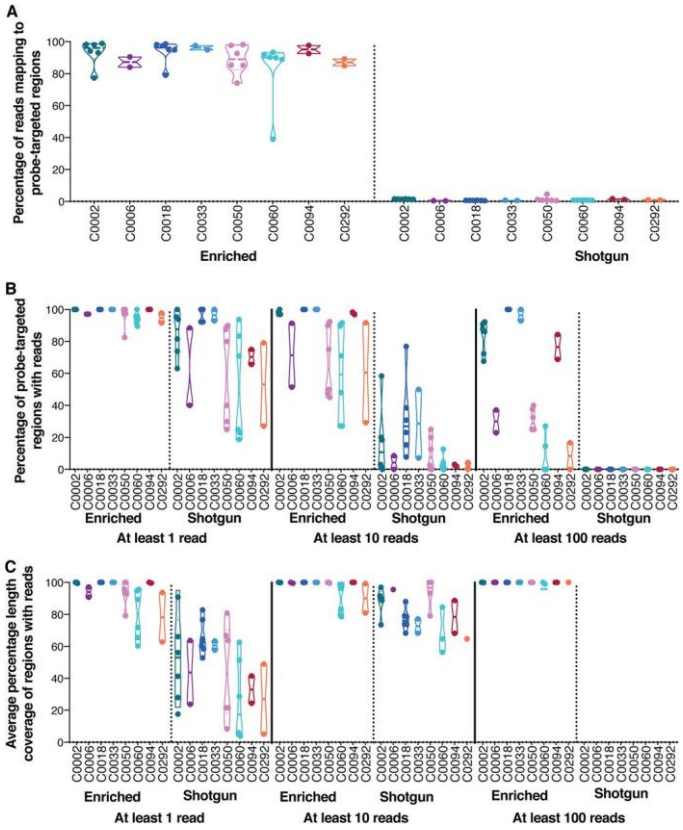


Figure 3: Comparison of enriched and shotgun sequencing results for on-target mapping, recovery, and length coverage.
Each point on the graph represents the results of a replicate experiment for either a genome that was enriched individually or a genome pooled with other genomes across

both trials. The horizontal line for each isolate represents the mean. (A) Percentage of reads on target for each bacterium tested in various sample types for both enriched and shotgun samples. (B) Percent recovery of regions predicted to be targeted by probes for each bacterial genome tested in both enriched and shotgun samples (1 versus 10 versus 100 reads per probe-targeted region). (C) Average percent length of coverage of probe-targeted regions with reads from isolates tested individually and in pools in both enriched and shotgun samples (1 versus 10 versus 100 reads). If samples did not have any probe-targeted regions with a given read coverage, the results were excluded from panel C. This represents eight samples in the panel labeled "At least 10 reads" (all from the shotgun data [strain C0002, n = 1; strain C0050, n = 2; strain C0060, n = 3; strain C0006, n = 1; strain C0292, n = 1]), all samples for the shotgun data in the panel labeled "At least 100 reads," and five samples for the enriched data (strain C0060, n = 4; strain C0292, n = 1).

Successful enrichment of ARGs in mock metagenomes

Genomic DNA from multiple bacteria was pooled at various ratios of 4 or 8 isolates, with the sequences of some bacteria representing less than 10% of the total mock metagenome (Table S4). In 28/32 enrichments, 80% or more of the sequencing reads mapped to probe-targeted regions within the individual bacterial genome regardless of the pooling ratios (Fig. 3A; Table S5). The one exception was trial 1, pool 2 (enriched), where on-target mapping was not as effective (~70%); nevertheless, even the results of this trial remained over 50-fold better than those of the trial with the shotgun-sequenced samples (shotgun samples) (Table S5). In all shotgun samples, the percentage of reads on target never exceeded 5%, and in 31/32 cases, it was less than 2% of the total sequencing data (Fig. 3A; Table S5).

At the isolate level, the percentage of the mock metagenome that was represented by probe-targeted regions in an individual isolate ranged from 0.0015 to 0.63% of the total DNA (Table S4; Fig. S3). In 21/32 enriched cases, over 90% of the probe-targeted regions

were captured by 10 reads or more (Fig. 3B; Table S5). In contrast, none of the shotgun-sequenced samples recovered more than 80% of the probe-targeted regions with at least 10 reads. The cases in which enrichment underperformed were associated with two species in particular: *K. pneumoniae* and *P. aeruginosa* (Fig. 3B; Table S5). We defined the sensitivity of detection of AMR for a given isolate to be the percentage of total DNA represented by probe-targeted regions of a given genome at which greater than 90% of the probe-targeted regions were recovered with at least 10 reads. These values ranged from 0.033% for *S. aureus* C0018 to 0.11% for *P. aeruginosa* C0060 (Fig. S3). With these bacterial species tested, our probe set could successfully capture the resistome of these isolates, which represents less than 0.1% of the total DNA and even less at the individual gene level.

Target gene recovery from mock metagenomes by enrichment exceeds that by shotgun sequencing

We recovered significantly more targeted genes with at least 1, 10, or 100 reads mapping (mapping quality ≥ 41 MAPQ, length ≥ 40 bp) by enrichment than by shotgun sequencing (Fig. 3B; Table S5). Furthermore, the average percent coverage of the probe-targeted regions with at least 1, 10, or 100 reads in all isolates enriched individually or in pools was always higher than that for the shotgun samples and ranged from being 1.05- to 18.3-fold higher (Fig. 3C; Table S5). For all genomes in all pooled libraries across both trials, the average normalized read count and the depth of the reads on probe-targeted ARGs from enriched libraries were over 50 times (57.09 to 25,683.42) higher than those from the

shotgun sequencing control (Table S5). In 31/32 cases, the fold increase in read counts exceeded 2 orders of magnitude and was over 4 orders of magnitude for some probe-targeted regions (Table S5). The one case that did not conform (from trial 1, pool 2; see above) reflects a minor and nonreproducible variability in the quality of the capture for unknown reasons. Nonetheless, there was a clear distinction between the shotgun and enriched samples, with the enriched data showing a more consistent agreement between normalized read counts per probe-targeted region than the shotgun data (Fig. 4). A similar trend was observed when the raw read counts for each sample were used (Fig. S4).

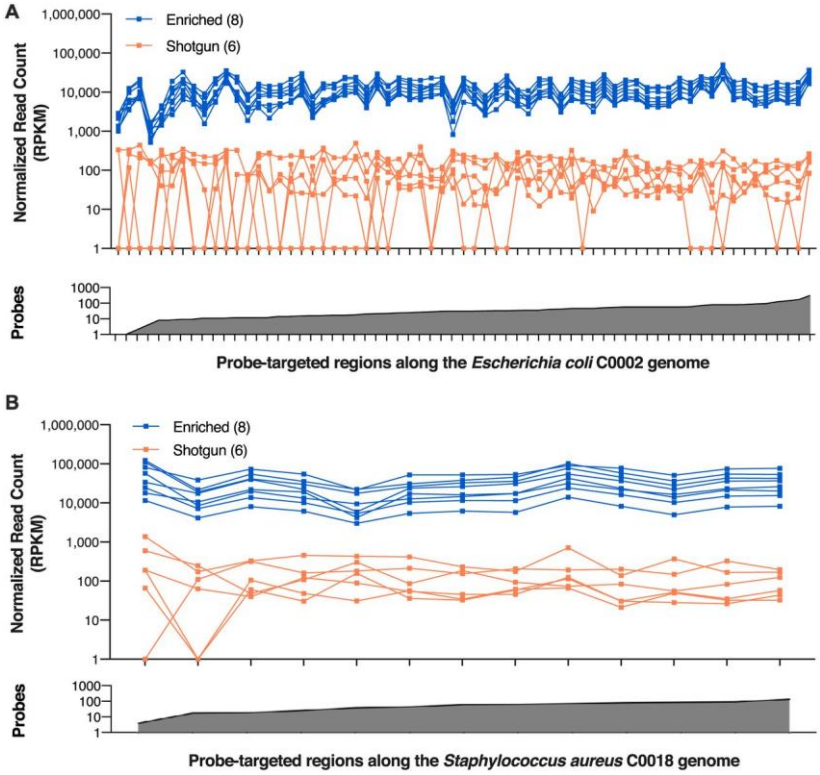


Figure 4: Enrichment results in higher read counts for antibiotic resistance genes than shotgun sequencing.

Normalized read counts at each probe-targeted region within the *Escherichia coli* C0002 genome (A) and *Staphylococcus aureus* C0018 genome (B) in enriched and shotgun samples, including individual and mock metagenomes of multiple isolates, are shown. Among the enriched and shotgun pairs, reads were subsampled to equal depths and mapped to the individual isolate's genome. Read counts were normalized by the number of reads mapping per target length (in total number of reads per kilobase per million [RPKM]). The predicted number of probes for each region along the genome is shown at the bottom of each panel. The y axes are in the logarithmic scale.

ARG analysis of a human gut metagenome

In order to test the efficacy and the reproducibility of our enrichment in more complex samples, we performed enrichments on replicates from metagenomic libraries with DNA isolated from a healthy individual's stool sample. We compared the results of the experiments with those of traditional shotgun sequencing, whereby selected libraries were sequenced to a depth of over 3.5 million paired reads (Table S2). We included a series of positive-control enrichments with genomic DNA from *E. coli* C0002, which was previously used for enrichment with the mock metagenomes. In all cases, we identified the same genes, with a consistent number of reads mapping among these replicate enrichments (when subsampled to equal depths among sets), proving reproducibility regardless of the probe and library ratio (Table S6; Fig. S5). Within each set, we found an excellent correlation with the previous results seen with *E. coli* C0002 (Pearson correlations, >0.923 for all pairs in set 1, >0.924 for all pairs in set 2, and >0.901 for all pairs in set 3) (Fig. S5).

Across the enriched gut microbiome samples with the full number of reads and no filters, on average, 50.69% of reads mapped to sequences from CARD and 68 genes with at least 10 reads were identified, whereas 0.03% of reads mapped, on average, and 32 genes

with at least 10 reads were identified in the shotgun libraries (Fig. S6A and B; Table S7). We found significantly more genes with at least 1, 10, and 100 reads from each enriched sample than from the shotgun samples, and the average percent coverage of a gene by the number of reads in the enriched samples was 1.5-fold higher (Fig. S6B and C). When subsampled to the same depth as their enriched pairs (between 22,324 and 149,320 reads), we identified, on average, 1 (range, 0 to 2) antibiotic resistance determinant with at least 10 reads after filtering in the shotgun samples, making comparisons at this level unrealistic (Table S8). Conversely, when subsampled to the depth of the sample with the lowest enriched read coverage (22,324 reads), we identified, on average, 28 ARGs with at least 10 reads in the enriched libraries postfiltering (Table S8).

High fold enrichment of ARGs from human stool

We combined the read counts for genes with at least 10 reads that passed the chosen filters within each set to compare the probe and library ratios in subsampled and full-read samples through both enrichment and shotgun sequencing. With the full number of reads, 24/70 (34.28%) of genes detected overlapped across all enriched libraries (n = 27), while we identified 16 genes of a total 32 (50.00%) that overlapped across all the shotgun libraries (n = 6) (Tables S7 and S9). When subsampled to the lowest enriched read coverage (22,324 reads), there were no genes that overlapped across all 6 shotgun libraries, while 13/47 (27.66%) of the genes overlapped across all 27 enriched libraries (Table S10). Comparing the subsampled enriched libraries (22,324 reads), the majority (31/34) of the genes missing in at least one sample were those with, on average, less than 20 reads across the 27 libraries

(Table S10; Fig. S7). The order of genes with higher read counts (i.e., a higher abundance and a higher gene copy number) was consistent among the enriched and shotgun samples, and there was a more significant discrepancy between the two sets of samples for reads associated with lower-abundance genes (Fig. S7 and S8). Thus, enrichment, in the same way as shotgun sequencing, does not in some way bias the prevalence of the rank order of AMR genes in these samples. Finally, both methods resulted in an excellent correlation among technical replicates individually (Pearson correlations, 0.871 for shotgun samples and 0.972 for enriched samples; Fig. S7 and S8).

We found that the performance of enrichment exceeded that of shotgun sequencing by identifying more unique antibiotic resistance genes at much lower sequencing depths. The enriched samples provided a more diverse representation of ARGs at less than 100,000 paired reads, compared to over 5 million reads in the shotgun samples (Fig. S8 and S9). With the full number of reads in both methods (between 66- and 389-fold more in the shotgun samples than in the enriched samples), the average fold enrichment was >600-fold, and there were still 18 to 50 fewer genes in the shotgun samples than in the enriched samples (Fig. 5A; Table 1). In most cases, there were only a few genes found via shotgun sequencing that were missing in the paired enriched sample (between 9 and 15; 22 unique genes). Only between 1 and 5 of these genes (7 total unique genes) in each sample were predicted to be targeted by probes (Table 1). Of these, only one, *novA*, was missing from all enriched samples but was present in all shotgun samples with >10 reads, a mapping quality of ≥11, and percent length coverage by reads of ≥10%. The other 6 genes (*macB*,

vanRG, *vanSG*, *smeE*, *cfxA6*, *cepA*) were found in only a few shotgun samples with less than 30 reads and less than 20% read length coverage, on average (Table 1; Table S13). When the two sample types were combined for hierarchical clustering analysis, the enriched libraries clustered separately from the shotgun libraries with a stronger correlation (0.9957 compared to 0.8712 for the shotgun libraries; Fig. S8).

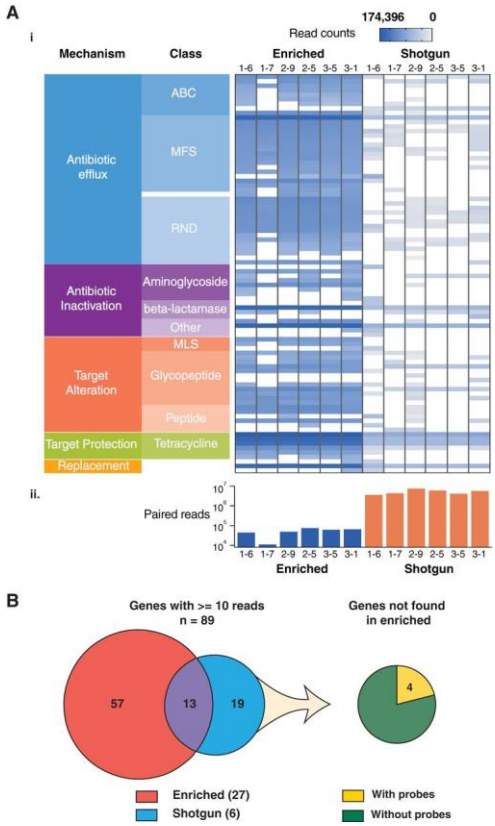


Figure 5: Comparison of resistance elements between enriched and shotgun libraries. For the enriched and shotgun samples, the full number of reads for each sample was mapped to the sequences in CARD using the *rgi bwt* tool, and the results were filtered for genes with probes mapping with reads with an average mapping quality of ≥11 and a percent length coverage of a gene by reads greater than or equal to 10%. (A) (i) Read counts were normalized per kilobase of reference gene per million reads sequenced (RPKM) and log transformed to produce the heatmap. The rows are grouped based on resistance mechanisms, as annotated in CARD (not all mechanisms and classes are labeled). ABC, ATP-binding cassette antibiotic efflux pump; MFS, major facilitator superfamily antibiotic

efflux pump; RND, resistance-nodulation-cell division antibiotic efflux pump; MLS, macrolides, lincosamides, and streptogramins. (ii) The number of reads used for mapping in each sample. (B) (Left) Overlap of genes found with at least 10 reads, a percent coverage greater than or equal to 10%, and an average mapping quality of reads greater than or equal to 11 in the 27 enriched and 6 shotgun samples. Between all samples, enriched or shotgun sequenced, there were 89 genes with reads passing these filters; 13 overlapped, 57 were unique to the enriched samples, and 19 were unique to the shotgun samples. (Right) Of the 19 genes identified only through shotgun sequencing, only 4 of these genes are predicted to be targeted by probes.

Table 1: Comparing genes with reads for shotgun and enriched stool library pairs[a]

Set	Amt (ng)		Fold difference in no. of reads (enriched vs shotgun)	No. of genes:				Fold enrichment (minimum–maximum)
	Probes	Library		Found in shotgun samples	Found in enriched samples	Overlapping	With probes missing in enriched samples	
1	200	100	389.70	18	49	9	1	1,054.92 (0–10,905.8)
	100	200	82.24	20	25	7	5	1,171.32 (0–6,459.8)
2	400	200	154.93	27	55	12	4	879.87 (0–9,612.1)
	100	100	80.73	23	61	11	1	868.16 (0–8,193.3)
3	100	100	66.67	19	57	9	2	732.16 (0–6,962.7)
	25	50	88.26	22	58	9	2	690.19 (0–7,319.6)

[a]We mapped the full number of reads from shotgun and enriched pairs to the sequences in CARD using the rgi bwt tool. The results for the samples were filtered for genes with at least 10 reads, those to which probes mapped (only for the enriched samples), an average read mapping quality of ≥11, and an average read length coverage of ≥10%. Filtered genes and their normalized read counts (RPKM) from each enriched sample/shotgun sample pair were combined to compare and determine the fold enrichment.

We then compared the overlap between all 27 enriched samples and the 6 shotgun-sequenced samples and included genes found through shotgun sequencing without any probes mapping. We found a total of 89 genes with at least 10 reads between all libraries, of which 13 overlapped between methods, 57 were unique to the enriched libraries, and 19 were unique to the shotgun libraries (Fig. 5B; Table S13). Of the 19 genes not found in any enriched library, only 4 were predicted to be targeted by probes, while the remaining were not in CARD when the probes were initially designed (n = 8 gene sequences) or had probes that were removed during design and filtering (n = 7 gene sequences). Of the four genes with predicted probes, *cfxA6* was present in all enriched samples but was filtered out by mapping quality, *vanSG* was present in only 2/6 shotgun samples at less than 20% gene length coverage by reads, and *cepA* was found in enriched samples but at less than 10 reads; finally, we identified *novA* in all shotgun samples but in only a few enriched samples at less than 10 reads and less than 10% read length coverage. Despite the few (n = 4) genes that were missing from the enriched samples, even a 200-fold greater sequencing depth of our shotgun libraries could not provide results that match those shown by our enrichment data (Fig. S9).

Negative-control results

To track and measure the contamination in the laboratory environment at McMaster University within the Michael G. DeGroote Institute for Infectious Disease Research, commercial kits, environment, and reagents, we included negative controls consisting of a blank DNA extraction and negative reagent controls in enrichments that we processed in a

manner identical to that used for our samples in phases 1 and 2. For phase 1 in both trial 1 and trial 2, we found a negligible amount of library DNA in the blank after enrichment, and very few of the sequenced reads were associated with the indexes used for the blank library (between 2.46% and 8.96% of sequenced reads; Tables S2 and S11). Only the blank samples from phase 1, trial 1, and phase 2, set 2, resulted in genes with at least 10 reads mapping (10 and 19, respectively; Table S12).

DISCUSSION

Increased interest in targeted capture approaches has resulted in the design of probe sets for the detection of viruses, bacteria, and, more recently, antibiotic resistance elements (26–29). Although our study is not the first to employ targeted capture for antibiotic resistance genes, we have focused on a rigorous probe design that includes choosing an appropriate reference database, robust probe set validation, and experimental considerations for enrichment, including reduced input library and probe concentrations (25, 30–33). Our probe design and the application of in-solution targeted capture ultimately result in a cost-effective alternative to shotgun sequencing for identifying antibiotic resistance genes in complex environmental and clinical metagenomes.

Reference database for probe design and analysis

CARD was chosen as the reference database for our probe design (v1.0.1) and analysis (v3.0.0) due to its rigorous curation of antibiotic resistance determinants. We excluded some genes (e.g., *gyrA*, EF-Tu genes, etc.) that are likely to be found as homologs across many families of bacteria and that would likely have overwhelmed the probe set and

sequencing effort with abundant, nonmutant antibiotic-susceptible alleles. Instead, we chose CARD's protein homolog model (v1.0.1, n = 2,010 genes) to focus our approach on genes that are likely to be acquired (i.e., genes associated with mobile genetic elements) and those that are unique to individual families of bacteria. Therefore, although we were unable to detect resistance conferred by SNPs in chromosomal metabolic genes, our probe set was capable of capturing the vast majority of resistance elements and those that were at a higher risk of being mobile. In future probe designs, the protein variant model of CARD (v1.0.1, n = 77 genes; v3.0.0, n = 141 genes) can be targeted using probes specific to the regions of a gene associated with a given set of SNPs, but they will need to be carefully tested in silico to ensure that they do not enrich unintended targets. Given that in certain populations (e.g., metagenomes) these variant sequences may be less abundant than their susceptible counterparts, careful and rigorous analysis will need to be implemented to identify the relevant variants (i.e., RGI developments).

To address our probe set's compatibility with a frequently updated database, we chose a more recent version of CARD (v3.0.0, n = 2,238 genes) for comparative analysis with our bait set designed in 2015. Since the design of our probes against v1.0.1 of CARD, the database has been updated and includes 264 additional genes. Despite these changes, our probe set targeted the majority (2,021/2,238) of known antibiotic resistance gene sequences from CARD (v3.0.0). In reality, the probes should target sequences with up to 15% nucleotide sequence divergence from a reference sequence, suggesting a wider applicability and target capacity toward newly characterized members of AMR gene

families, which often differ from other members by only a few nucleotides. Of the 264 genes added to CARD (v3.0.0), our existing probes capture 75 of these genes that are sufficiently similar to other targeted members of the same AMR gene family (e.g., genes for aminoglycoside acetyltransferases, chloramphenicol acetyltransferases, and beta-lactamases [*bla*$_{ACT}$, *bla*$_{CARB}$, *bla*$_{CMY}$, *bla*$_{LEN}$, *bla*$_{NDM}$, *bla*$_{OXA}$, *bla*$_{PDC}$. *bla*$_{SHV}$, *bla*$_{TEM}$, *bla*$_{VEB}$]). Of the remaining genes, the sequences of 60 have been newly identified since 2015, and the other genes, although mentioned in the literature prior to 2015, were added due to increased efforts of curation of CARD.

Other approaches targeting ARGs have included probes for species identifiers, plasmid markers, and biocide or metal resistance (27–29). These probe sets range in target capacity from 5,557 genes (3.34 Mb) (28) to over 78,600 genes (88.13 Mb) (27) and comprise up to 4 million probes (29). Other strategies involve designing one probe per gene, tiling probes across a gene without overlap (1× coverage), or interprobe distances of up to 121 nucleotides (28, 29). Our approach is more conservative in probe design (1.77 Mb for 2,021 genes), but the dense tiling allows for more probes per gene (99.16% of genes had greater than 10 probes) and an increased depth of probe coverage (average, 9.47 times). We believe that the design approach increases the specificity, sensitivity, and likelihood of capturing rare DNA molecules common in complex metagenomes (34). We also performed extensive filtering of candidate probes against the human genome and other eukaryote, archaeal, and weakly matching bacterial sequences to provide a probe set that is bacterial ARG specific and avoids off-target hybridization. Focusing on one highly curated database

of antibiotic resistance determinants (CARD) increases the likelihood of capturing bona fide sequences that are associated with known resistance and reduces the overall cost of the probe set and sequencing effort. When updates to CARD are released or if additional markers are of interest, probes can easily be designed and added to the existing probe set.

Experimental considerations in targeted capture methods

For our trials, we tested amounts of inputs (25 ng to 400 ng) significantly smaller than the amount recommended by the manufacturer (up to 2 μg of DNA for metagenomic samples), setting our method apart from other methods for the targeted capture of AMR genes (27, 28). Others have looked at reducing the amount of input DNA from the manufacturer's recommended amount of 3,000 ng to 500 ng and saw no significant differences in results (35). Despite a 16-fold reduction in the DNA input amount (25 ng versus the recommended 2,000 ng), we saw no visible differences in the order of genes captured in the stool sample, and the normalized read counts were comparable among the different library and probe amounts, suggesting that our approach is robust to tremendous fluctuations yet still identifies all antibiotic resistance genes in samples with a low DNA yield (e.g., clinical and environmental samples). Furthermore, a lower input concentration of probes also reduces the cost per reaction.

Reproducibility, sensitivity, and performance with clinical isolates

The sensitivity of our probe set was tested using individual bacterial genomes and mock metagenomes, wherein the percentage of total DNA represented by probe-targeted antibiotic resistance genes ranged from 0.0015% to 0.97%. A successful enrichment in our

trials was considered when greater than or equal to 90% of the probe-targeted regions with 10 or more reads were captured. When tested individually, enrichment was able to successfully capture all probe-targeted ARGs (100% with more than 10 reads) in the four bacterial species tested, with >85% of sequenced reads mapping to the targeted regions (<1%) of the genome. With the mock metagenomes, the probe-targeted regions of each isolate represented a smaller proportion, and there were 11 cases in which enrichment was not successful under the above-described criterion. In 7 instances, the given isolate represented less than 10% of the total pool, and many of the probe-targeted regions that were missing were short (<200 bp), and less than 5 probes for these regions were designed (see Table S4 in the supplemental material). One particular predicted resistance gene that was not captured in 2 cases, *fosA2* (ARO:3002804) in *K. pneumoniae* C0050, retained good probe coverage, despite a low percent identity (71.32%) to the CARD reference sequence. The poor performance in enrichment may suggest the limit of sequence similarity (>30%) that can be captured by probes designed against a single reference sequence. In addition, the high GC content of certain genes in the *K. pneumoniae* isolates and of many regions of the *P. aeruginosa* isolates (average GC content, 67%) likely reduced the capture efficiency in the more complex pooled samples, resulting in less than 10 reads for targeted genes. The conditions of hybridization may need to be further optimized for targets with higher GC contents. Regardless of this limitation, the enriched data provide significantly more read coverage for antibiotic resistance genes at a lower depth of sequencing than shotgun sequencing of these mock metagenomes.

Standardization and controls in metagenomics

Standardization (including reproducibility) in enrichment studies remains sorely lacking. In this study, we attempted to reduce bias and assess enrichment by using the same DNA extract, library preparations, and enrichment in triplicate. Even among replicate libraries and shotgun sequencing runs, the differences in the number of genes identified at various sequencing depths highlight the inherent variability in metagenomics (Fig. S8). The positive control (*E. coli* C0002), processed alongside the other samples, ensured that our methodology and probes were performing optimally at the time of hybridization. We also introduced negative controls to measure the extent of exogenous DNA contamination, which is ubiquitous in all laboratory settings and reagents (36, 37). Between 86.07 and 100% of the sequenced reads from our negative controls had corresponding index sequences from experimental samples, suggesting that DNA exchange among samples during enrichment or cross-contamination is the primary concern with our method (Tables S2 and S11). Notably, the genes identified in the results for the blanks not arising from cross-contamination and also found in the enriched and shotgun results are commonly associated with bacteria identified in negative controls in microbiome studies (mainly *Escherichia coli*) and encode efflux systems or other intrinsic resistance determinants (*mdtEFHOP*, *emrKY*, *cpxA*, *acrDEFS*, *pmrF*, *eptA*, *tolC*). The two genes that were unique to the results for the blanks [*drfA17* had 11 reads with 85.86% coverage; *aph(3")-Ib* had 16 reads with 57.46% coverage] are associated with mobile genetic elements in *Enterobacteriaceae*, and the latter has previously been associated with laboratory reagent

contamination (38, 39). Despite the use of standard methods to control for contamination (i.e., filter pipette tips, PCR cabinets, and sterile DNA- and RNA-free consumables), we still found limited contamination, likely stemming from reagents and/or the surrounding laboratory environment, further highlighting the importance of negative controls in all targeted capture experiments and meticulous reporting and publishing of a laboratory-based resistome (Table S6) (36, 37, 40).

Enrichment in the gut microbiome

Our enrichment of resistance genes in the human gut microbiome samples resulted in a higher average percentage on target (50.69%) compared to that obtained by other published capture-based methods, 30.26% (range, 20.27% to 41.83%) (27) and a median of 15.8% (range, 0.28% to 68.2%) (28), highlighting the increased specificity of our probe design. Overall, our probe set and method identified a greater diversity of antibiotic resistance genes in the human gut microbiome, despite having been sequenced at a 66- to 389-fold lower depth than the shotgun-sequenced correlate. With a reduced depth of sequencing, it is evident that enrichment offers more valuable information in terms of both the number of genes with reads as well as the depth and breadth of coverage of those genes (Fig. 5).

Although shotgun sequencing can provide additional information on other functions and genes of interest, our targeted capture provides a more robust and a more reproducible profile of antibiotic resistance genes from metagenomes at a fraction of the sequencing cost. Only a few genes were absent in the enriched libraries but present in the shotgun

libraries. In the case of *novA*, which has a 70.51% GC content, there was a gap in the tiling of probes across the gene, and the hybridization conditions were perhaps not sufficient to capture this gene by our method. Additional probes or denser tiling along high-GC-content (>65%) sequences may facilitate successful capture. Another gene that we could not identify, the variant of the *vanS* (GC content, 36.7%) sensor from vancomycin resistance gene clusters, was covered by less than 20 reads in the shotgun samples, suggesting a very low abundance in the metagenome. Finally, the beta-lactamase genes *cepA* and *cfxA6* had been excluded from the enriched results after filtering due to low mapping quality or less than 10 reads. The low mapping quality suggests that reads are mapping to other beta-lactamase genes in the reference database.

All current methods to detect antibiotic resistance genes have limitations. Culturing, although time-consuming, remains the standard for diagnosing infections through the identification of both the pathogen and its susceptibility to a panel of antibiotics. Other biochemical techniques have been developed but are often organism specific and require additional assays for confirming ARGs (41). When studying the microbiome and the resistome of various environments, a culture-based approach is not feasible and, thus, high-throughput methods are needed (19). Sequencing-based approaches (i.e., PCR-based assays, microarray-based assays, and in-solution targeted capture) and quantitative PCR (qPCR) methods offer selective and sensitive means to identify a larger contingent of antibiotic resistance genes than other methods but can be (or are designed to be) heavily biased or selective. While PCR is highly sensitive, many panels for AMR genes target only

a range of between 200 and 400 genes (42). As we have shown here, probe-based hybridization methods enable the detection of over 2,000 ARGs in a single assay.

Compared to the other probe sets designed for AMR (27–29), ours offers a highly curated specific set of probes with a high coverage of ARGs and works exceptionally well on samples with low inputs. We have also included crucial controls to validate our findings. Whereas shotgun sequencing requires millions of reads to detect a few antibiotic resistance genes, we have shown that targeted capture can detect the same genes and more with ~50-fold less sequencing effort. A reduced amount of sequencing allows more samples to be processed per individual sequencing run, reducing sequencing costs overall and increasing throughput. One limitation to targeted approaches is that the probe design relies on known reference sequences, while shotgun sequencing can reveal additional information not captured by the probes, but at an added cost (depth). All sequencing-based methods are limited in the inability to characterize completely novel antibiotic resistance determinants, whereas a functional metagenomics approach is ideal in this regard (19).

In conclusion, we have rigorously measured the performance of our probe design and methods to satisfy many of the parameters in targeted capture routinely discussed (43). The sensitivity and specificity of our probe set are evident from the consistently high percentage of reads on target and the high recovery of probe-targeted sequences representing <0.1% of the total DNA. Our approach results in the uniform recovery of ARGs across bacterial genomes and is reproducible between library preparations. We believe that our targeted capture serves a critical role in the surveillance and detection of

ARGs across complex environmental settings, hospitals, and clinics. Profiling of these resistomes will provide invaluable information that can be used to target antibiotic and resistance inhibitor discovery but that at the same time can be used to keep abreast of the rapidly shifting rise of local and global antibiotic resistance.

MATERIALS AND METHODS

Nucleotide probe design and filtering to prevent off-target hybridization

Our reference for probe design was the protein homolog model of antibiotic resistance determinants (n = 2,129) from CARD (v1.0.1, released 14 December 2015) (11). Using PanArray (v1.0) software, we designed probes with a length of 80 nucleotides across all genes with a sliding window of 20 nucleotides and acceptance of 1 mismatch across probes (32). To prevent off-target hybridization between the probes and nonbacterial sequences, the candidate set of probe sequences (n = 38,980) was compared against the human reference genome and GenBank's nonredundant nucleotide database through BLAST (blastn) analysis (44, 45). Probes with high sequence similarity (>80%) and probes with high-scoring segment pairs (HSPs) of greater than 50 nucleotides of a possible 80 were discarded (human genome sequences, n = 158; eukaryotic sequences, n = 1,617; viral sequences, n = 774; archaeal sequences, n = 30). Probes with HSPs of less than 50/80 nucleotides to bacterial sequences were additionally discarded, resulting in a set of 32,066 probes. The candidate list was further filtered to omit probes that had bacterial HSPs that had <95% identity, resulting in a candidate list of 21,911 probes.

Optimizing probe density and redundancy

Probe sequences, along with 1 to 100 nucleotides upstream and downstream of the probe location on the target gene, were sent to Arbor Biosciences (Ann Arbor, MI) for probe design. These sequences are contained within the open reading frame of the target gene and allow probe sequences to be modified, if needed (i.e., polynucleotides at the termini), ensuring that the desired probe coverage of the target genes is attained. Additional 80 nucleotide probes were created across the candidate probe and flanking sequences at a tiling density of four times, resulting in 226,440 probes. Sequences with 99% identity over 87.5% of their length were collapsed using the USEARCH program (settings, usearch -cluster_fast -query_cov 0.875 -target_cov 0.875 -id 0.99 -centroids), resulting in a set of 37,826 final probes (46). Filtering against the human genome was performed by a method similar to that described above; no probes were found to be similar. Arbor Biosciences (Ann Arbor, MI) synthesized this final set of 37,826 80-nucleotide biotinylated single-stranded RNA probes by use of the custom myBaits kit (catalog number 300248; Arbor Biosciences, Ann Arbor, MI).

Probe assessment and predicted target genes

To predict the genes that can be targeted by the probes, a Bowtie2 program (the settings used included bowtie2 --end-to-end -N 1 '-L 32' -a) (47) alignment was performed to compare the set of 37,826 probe sequences to the 2,238 nucleotide reference sequences of the protein homolog models in CARD (v3.0.0, released 11 October 2018). The alignment file was manipulated through the use of samtools and bedtools to determine the number of instances that a probe mapped to a nucleotide sequence in CARD, the fraction of each gene

sequence covered by probes (length coverage by probes), and the depth of coverage by probes of each gene (bedtools genomecov, bedtools coverage -mean) (48, 49). The GC content of the probe sequences and the nucleotide sequences in CARD was calculated using a Python3 script from https://gist.github.com/wdecoster/8204dba7e504725e5bb249ca77bb2788. The melting temperature (Tm) was determined using the OligoArray function melt.pl (settings, -n RNA, -t 65 -C 1.89e−9) (50). We used Prism (v8) software for macOS (GraphPad Software) to generate the plots shown in Fig. S1 in the supplemental material.

Bacterial isolates, samples, and DNA extraction

Clinical bacterial isolates were obtained from the IIDR clinical isolate collection, which consists of isolates from the core clinical laboratory at the Hamilton Health Sciences Centre (Table S1). Genomic DNA was isolated from a cell pellet using the PureLink genomic DNA mini kit (catalog number K182002; Invitrogen, Carlsbad, CA). If DNA was not isolated on the same day, we stored the cell pellets at −80°C. While genomic DNA from all other isolates was extracted only once, DNA from a cell pellet of *Pseudomonas aeruginosa* C0060 was additionally extracted using a varied genomic lysis/binding buffer (30 mM EDTA, 30 mM Tris-HCl, 800 mM guanidine thiocyanate, 5% Triton X-100, 5% Tween 20, pH 8.0). We obtained a human stool sample from a healthy volunteer for the purpose of culturing the microbiome with consent from the Hamilton Integrated Research Ethics Board (HiREB approval number 5513-T). DNA was extracted on the same day following a modified protocol described elsewhere (51). Briefly, samples were bead beaten

and centrifuged, and the supernatant was further processed using a MagMax Express 96-well deep well magnetic particle processor from Applied Biosystems (Foster City, CA) with a multisample kit (catalog number 4413022; Life Technologies). DNA was stored at −20°C until it was used for library preparation.

Isolate genome sequencing

Library preparation for genome sequencing of the clinical bacterial genomes was completed by the McMaster Genomics Facility in the Farncombe Institute at McMaster University (Hamilton, ON, Canada) using the Nextera XT DNA library preparation kit (catalog number FC-131-1024; Illumina, San Diego, CA). Libraries were sequenced using an Illumina HiSeq 1500 or Illumina MiSeq v3 platform and v2 (2 × 250-bp) chemistry. Paired sequencing reads were processed through a Trimmomatic (v0.39) trimmer to remove adapters, checked for quality using the FASTQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and *de novo* assembled using SPAdes (v3.9.0) software (52, 53). The Livermore metagenomics analysis toolkit (LMAT; v1.2.6) was used to identify the bacterial species and screen for contamination or a mixed culture, while the Resistance Gene Identifier (RGI; v4.2.2) from CARD was used on the contigs obtained with SPAdes software to identify perfect (100% match) and strict (<100% match but within CARD similarity cutoffs) hits to CARD's curated antibiotic resistance genes (54).

Trials for enrichment

We performed two phases of experiments. The first was with genomic DNA from cultured multidrug-resistant bacteria (phase 1), and the second was with metagenomic DNA from a human stool sample (phase 2). The two trials in phase 1 differed in their library preparation methods, as described below (the major difference being the library fragment size obtained by sonication). In both trials, we tested genomic DNA from isolates individually (*Escherichia coli* C0002, *Pseudomonas aeruginosa* C0060, *Klebsiella pneumoniae* C0050, and *Staphylococcus aureus* C0018) (Tables S1 and S3). In addition, various nanogram amounts (based on the absorbance; Thermo Fisher Nanodrop spectrophotometer [Waltham, MA]) of each genome were combined prior to library preparation to create mock metagenomes, referred to as pool 1 (with the genomes of strains C0002, C0018, C0050, and C0060), pool 2 (with the genomes of strains C0002, C0018, C0050, and C0060), and pool 3 (with the genomes of strains C0002, C0018, C0050, C0060, *Klebsiella pneumoniae* C0006, *Staphylococcus aureus* C0033, *Escherichia coli* C0094, and *Pseudomonas aeruginosa* C0292). The amounts of the genome of each isolate in each pool varied between trials (Table S4). Phase 2 consisted of 3 replicates, referred to as set 1, set 2, and set 3, wherein a DNA extract from one individual human stool sample was split evenly into each set. From these aliquots, we generated 9 individually indexed sequencing libraries and performed capture with various library and probe ratios (Table S3). In all trials and sets, a blank DNA extract was carried throughout library preparation and enrichment, while an additional negative reagent control was introduced during enrichment.

Library preparation for enrichment sequencing

Library preparations were performed in a PCR clean hood, using bleached equipment, and the equipment was UV irradiated before use to prevent nonendogenous DNA contamination. Trial 1 library preparations were performed through the McMaster Genomics Facility using the NEBNext Ultra II DNA library preparation kits for Illumina (catalog number E7645L; New England BioLabs, Ipswich, MA). Based on absorbance and fluorometer values (QuantiFluor; Promega, Madison, WI), we sonicated approximately 1 μg of individual bacterial genomic DNA or pools of genomic DNA to 600 bp and prepared dual-index libraries with a size selection for 500- to 600-bp inserts. Post-library quality and quantity verification was performed using a high-sensitivity DNA kit for the Agilent 2100 bioanalyzer (catalog number 5067-4626; Agilent Technologies, Santa Clara, CA) and quantitative PCR using a Kapa SYBR Fast qPCR master mix for Bio-Rad machines (catalog number SFBRKB; Sigma-Aldrich, St. Louis, MO), primers for the distal ends of Illumina adapters, and the following cycling conditions: (i) 95°C for 3 min, (ii) 95°C for 10 s, (iii) 60°C for 30 s, (iv) a repeat of steps (ii) and (iii) for 30 cycles total, (v) 60°C for 5 min, and (vi) hold at 8°C. We used Illumina's PhiX control library (catalog number FC-110-3001; Illumina, San Diego, CA) as a standard for quantification.

In trial 2, the same genomic DNA, except for that of *P. aeruginosa* C0060, which was reisolated, was used for library construction through a modified protocol (see the supplemental material) (55). Briefly, we performed blunt-end repair, adapter ligation, library size selection, and indexing PCR on ~200 ng of sonicated DNA (250 to 300 bp). The McMaster Genomics Facility performed library quality control as described above.

Library preparation from a human stool sample

We divided one DNA extract from a donor stool sample into three 50-µl aliquots of approximately 3,150 ng each (based on QuantiFluor fluorometer results). DNA was sonicated to 600 bp and split into 9 individual library reaction mixtures (350 ng in 5.55 µl). We prepared dual-index libraries (NEBNext Ultra II DNA library preparation kits for Illumina [catalog number E7645L; New England BioLabs, Ipswich, MA]) with a size selection for 700- to 800-bp library fragments and 6 (set 1), 7 (set 2), or 8 (set 3) cycles of amplification. The McMaster Genomics Facility performed library quality control (with an Agilent 2100 bioanalyzer and by quantitative PCR, as described above). We generated positive-control libraries using *Escherichia coli* C0002 genomic DNA (40 ng of sonicated DNA) and a negative control with a blank DNA extract.

Targeted capture of bacterial isolates

We performed enrichments in a PCR clean hood, with a water bath, thermal cyclers, and heat blocks being located nearby. The probe set was provided by Arbor Biosciences (Ann Arbor, MI) and diluted with deionized water. For enrichment of bacterial genomes in trial 1, we used 100 ng of probes and 100 ng of each library, following the instructions in the myBaits manual (v3; Arbor Biosciences, Ann Arbor, MI), at a hybridization temperature of 65°C for 16 h (see the methods in the supplemental material for more details). After hybridization and capture with Dynabeads MyOne streptavidin C1 beads (catalog number 65001; Thermo Fisher, Waltham, MA), the resulting enriched library was amplified through 30 cycles of PCR (cycling conditions are described in the supplemental

material) using Kapa HiFi HotStart polymerase with library-nonspecific primers (Kapa library amplification primer mix [10×]; catalog number KK2620; Roche Canada). A 2-μl aliquot of this library was amplified in an additional PCR for 3 cycles (under the same conditions described above) and then purified. We performed the capture in trial 2 in the same manner described above for trial 1 but applied 17 cycles of amplification post-capture (see the PCR conditions in the supplemental material). The McMaster Genomics Facility performed library quality control as described above. The libraries were pooled in equimolar amounts and sequenced to an average of 94,117 clusters by use of an Illumina MiSeq sequencer (v2; 2 × 250-bp reads). Pre-enrichment libraries for the mock metagenomes were sequenced in a separate Illumina MiSeq (v2; 2 × 250-bp reads) run from the enriched libraries to an average of 93,195 clusters each. From both trial 1 and trial 2, negative controls consisting of blank extractions carried through library preparation and enrichment were sequenced on separate individual Illumina MiSeq (2 × 250-bp) runs. After demultiplexing of the blank, all possible index combinations were retrieved to identify potential cross-contamination of libraries as well as exogenous bacterial contamination.

Targeted capture of the stool sample

Based on qPCR values and the average fragment sizes of each library generated from the human stool DNA extract, we combined various nanogram amounts of library (50, 100, 200 ng) and probes (25, 50, 100, 200, 400 ng) for enrichment (Table S3). Along with the negative-control (blank) library, we introduced additional negative controls during enrichment, using distilled H2O to replace the volume normally required for library input.

We performed enrichment following the instructions in the myBaits manual (v4; Arbor Biosciences, Ann Arbor, MI) at a hybridization temperature of 65°C for 24 h. After hybridization and capture with Dynabeads (MyOne streptavidin C1 beads; catalog number 65001; Thermo Fisher, Waltham, MA), the resulting enriched library was amplified through 14 cycles of PCR using Kapa HiFi HotStart ReadyMix polymerase with library-nonspecific primers and the same conditions described above (see the enrichment methods in the supplemental material). The resulting products were purified using Kapa Pure beads (catalog number KK8000; Roche Canada) at a 1× volume ratio and eluted in 10 mM Tris, pH 8.0. Purified libraries were quantified through qPCR using 10× SYBR Select master mix (catalog number 4472942; Applied Biosystems, Foster City, CA) for Bio-Rad Cfx machines, Illumina specific primers (a 10× primer mix from Kapa; catalog number KK4809; Roche Canada), and Illumina's PhiX control library (catalog number FC-110-3001; Illumina, San Diego, CA) as a standard. Cycling conditions were as follows: (i) 50°C for 2 min, (ii) 95°C for 2 min, (iii) 95°C for 15 s, (iv) 60°C for 30 s, and (v) a repeat of steps (iii) and (iv) for 40 cycles total. We pooled the enriched libraries in equimolar amounts based on qPCR values, and the McMaster Metagenomic Sequencing facility performed library quality control as described above. Finally, we sequenced the enriched libraries (average, 97,286 clusters) and the pre-enrichment libraries (average, 5,325,185 clusters) with an Illumina MiSeq sequencer (v2; 2 × 250 bp). The negative controls consisting of blank extractions carried through library preparation and enrichment were sequenced on

separate individual Illumina MiSeq (2 × 250-bp) runs. After demultiplexing, all possible index combinations were retrieved.

Analysis of bacterial isolate sequencing data

In order to identify probe-targeted regions and coordinates that overlap predicted resistance genes based on RGI results for the individual bacterial genomes, we aligned our probe set to the draft reference genome sequence using the Bowtie2 (v2.3.4.1) program (47). We used the Skewer (v0.2.2) program (skewer -m pe -q 25 -Q 25) to trim sequencing reads (enriched or shotgun) and the bbmap (v37.93) program tool dedupe2.sh to remove duplicates and mapped the reads to the bacterial genomes using the Bowtie2 (v2.3.4.1) program (settings, –very-sensitive-local, unique sites only) (https://github.com/BioInfoTools/BBMap) (47, 56). Aligned reads were filtered based on mapping quality (≥41 MAPQ) and length (≥40 bp) using various tools: samtools (v1.4), bamtools (v2.4.1), and bedtools (v2.27.1) (48, 49, 57). We determined the number of reads mapping to the reference genome overall and the number of reads mapping within a predicted probe-targeted region using genomic coordinates and bedtools (intersectBed) (50). The percent length coverage and the average depth of coverage of each probe-targeted region with at least one read were determined using bedtools coverage (settings, -counts, -mean and default function) (49). We normalized the read counts by the number of reads mapping per kilobase of targeted region per the total number of reads mapping to a particular genome. The number of genes with at least 1, 10, or 100 reads was counted, and their percent length coverage by reads was determined.

Analysis of stool sample sequencing data

We processed the enriched and shotgun reads for the human stool sample as described above for the bacterial isolates. Subsampling of reads was performed using the seqtk (v1.2-r94) program (settings, seqtk sample -s100; https://github.com/lh3/seqtk). We used the bwt feature in RGI (the beta version of v5.0.0; http://github.com/arpcard/rgi) to map trimmed reads, using the Bowtie2 (v2.3.4.1) program, to the sequences in CARD (v3.0.0), generating alignments and results without any filters (47). We parsed the gene mapping and allele mapping files to determine the number of genes in CARD with reads mapping (at least 1, 10, and 100 reads) under various filters. After plotting the mapping quality for each read in every sample across the 3 sets, we chose an average mapping quality (mapq) filter of 11. We assessed a percent length coverage filter of a gene by reads of 10, 50, and 80% and chose the most permissive (10%) for comparison between the shotgun and enriched samples. These low thresholds were necessary for analyzing the shotgun data to obtain any reasonable results at all. Finally, we used a filter to check for the probes mapping to the reference sequences in most comparisons, except to identify genes in the shotgun samples that would not be captured by our probe set. We repeated the same analysis process for the negative-control (blank) libraries. In phase 2, set 1, there were very few reads associated with the blank library after enrichment, so we used the raw sequencing reads for analysis. For the blank in set 2, we omitted deduplication, and we could not identify any reads associated with the blank indexes after sequencing for set 3. Read counts were normalized using the “all mapped reads” column in the gene mapping file and the

reference length (in kilobases) along with the total number of reads per kilobase per million (RPKM) available for mapping. Hierarchical clustering was performed using the Gene Cluster (v3.0) and Java Tree View (v1.1.6r4) (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) programs, log transformation, and clustering arrays with an uncentered correlation (Pearson) and average linkage. For rarefaction analysis, we first aligned trimmed reads against the sequences in CARD (v3.0.0) using the Bowtie2 program, followed by filtering for a mapping quality of ≥11 (47). This file, along with an annotation file for CARD, was analyzed with the AmrPlusPlus rarefaction analyzer (http://megares.meglab.org/amrplusplus) (58), with subsampling every 1% of total reads and a gene read length coverage of at least 10%. The average number of genes identified after rarefaction was plotted and fit to a logarithmic curve to allow for simplified extrapolation. We generated heat maps and figures in Prism (v8) software for macOS (GraphPad Software).

Data availability

Raw sequencing reads (FASTQ) for the IIDR clinical isolate collection bacterial isolate genome assembly were deposited in NCBI under BioProject accession number PRJNA532924. All metagenomic sequencing results, enriched or shotgun, were deposited in NCBI under BioProject accession number PRJNA540073. The probe set sequences and annotations are available at https://card.mcmaster.ca/download.

ACKNOWLEDGMENTS

This research was funded by the Canadian Institutes of Health Research (grant MT-14981 to G.D.W. and grant PJT-156214 to A.G.M.) and Genome Canada (to A.G.M.), by a Canada Research Chair (to G.D.W., H.N.P., and M.G.S.), and by a Cisco Research Chair in Bioinformatics supported by Cisco Systems Canada (to A.G.M.), an Ontario Research Fund grant, and the Boris Family Foundation. A.K.G. was supported by a Canada graduate scholarship (master's) from the Natural Sciences and Engineering Council of Canada, an Ontario graduate scholarship, and an Ontario graduate fellowship. Computer resources were supplied by the McMaster Service Lab and Repository computing cluster, funded in part by grants to A.G.M. from the Canadian Foundation for Innovation (grant 34531).

We acknowledge Linda Ejim for clinical isolate culturing, Susan McCusker from the Center for Microbial Chemical Biology at McMaster University for genomic DNA isolation from clinical isolates, Laura Rossi for DNA isolation from human stool, and the McMaster Genomics Facility for next-generation sequencing. Additional thanks go to Kara Tsang and Biren Dave for genome assembly and annotation of the clinical isolates, Nicholas Waglechner for helpful discussions and guidance pertaining to analyses, Christy Groves for valuable input on the figures, and all the members of the McMaster Ancient DNA Center.

We declare no conflicts of interest.

REFERENCES

1. Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, Vlieghe E, Hara GL, Gould IM, Goossens H, Greko C, So AD, Bigdeli M, Tomson G, Woodhouse W, Ombaka E, Peralta AQ, Qamar FN, Mir F, Kariuki S, Bhutta ZA, Coates A, Bergstrom R, Wright GD, Brown ED, Cars O. 2013. Antibiotic resistance—the need for global solutions. Lancet Infect Dis 13:1057–1098. 10.1016/S1473-3099(13)70318-9. PubMed.
2. Levy SB, Marshall B. 2004. Antibacterial resistance worldwide: causes, challenges and responses. Nat Med 10:S122–S129. 10.1038/nm1145. PubMed.
3. Schwartz KL, Morris SK. 2018. Travel and the spread of drug-resistant bacteria. Curr Infect Dis Rep 20:29. 10.1007/s11908-018-0634-9. PubMed.
4. Gaze WH, Krone SM, Larsson DGJ, Li X-Z, Robinson JA, Simonet P, Smalla K, Timinouni M, Topp E, Wellington EM, Wright GD, Zhu Y-G. 2013. Influence of humans on evolution and mobilization of environmental antibiotic resistome. Emerg Infect Dis 19:120871. 10.3201/eid1907.120871.
5. Surette MD, Wright GD. 2017. Lessons from the environmental antibiotic resistome. Annu Rev Microbiol 71:309–329. 10.1146/annurev-micro-090816-093420. PubMed.
6. Allen HK, Donato J, Wang HH, Cloud-Hansen KA, Davies J, Handelsman J. 2010. Call of the wild: antibiotic resistance genes in natural environments. Nat Rev Microbiol 8:251–259. 10.1038/nrmicro2312. PubMed.
7. Fitzpatrick D, Walsh F. 2016. Antibiotic resistance genes across a wide variety of metagenomes. FEMS Microbiol Ecol 92:fiv168. 10.1093/femsec/fiv168. PubMed.
8. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science 337:1107–1111. 10.1126/science.1220761. PubMed.
9. Luo Y, Yang F, Mathieu J, Mao D, Wang Q, Alvarez P. 2014. Proliferation of multidrug-resistant New Delhi metallo-β-lactamase genes in municipal wastewater treatment plants in northern China. Environ Sci Technol Lett 1:26–30. 10.1021/ez400152e.
10. Pal C, Bengtsson-Palme J, Kristiansson E, Larsson D. 2016. The structure and diversity of human, animal and environmental resistomes. Microbiome 4:54. 10.1186/s40168-016-0199-5. PubMed.
11. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573. 10.1093/nar/gkw1004. PubMed.
12. Walsh F, Duffy B. 2013. The culturable soil antibiotic resistome: a community of multi-drug resistant bacteria. PLoS One 8:e65567. 10.1371/journal.pone.0065567.
13. Mezger A, Gullberg E, Göransson J, Zorzet A, Herthnek D, Tano E, Nilsson M, Andersson DI. 2015. A general method for rapid determination of antibiotic susceptibility

and species in bacterial infections. J Clin Microbiol 53:425–432. 10.1128/JCM.02434-14. PubMed.
14. Zumla A, Al-Tawfiq JA, Enne VI, Kidd M, Drosten C, Breuer J, Muller MA, Hui D, Maeurer M, Bates M, Mwaba P, Al-Hakeem R, Gray G, Gautret P, Al-Rabeeah AA, Memish ZA, Gant V. 2014. Rapid point of care diagnostic tests for viral and bacterial respiratory tract infections—needs, advances, and future prospects. Lancet Infect Dis 14:1123–1135. 10.1016/S1473-3099(14)70827-8. PubMed.
15. Pulido MR, García-Quintanilla M, Martín-Peña R, Cisneros JM, McConnell MJ. 2013. Progress on the development of rapid methods for antimicrobial susceptibility testing. J Antimicrob Chemother 68:2710–2717. 10.1093/jac/dkt253. PubMed.
16. Quan J, Langelier C, Kuchta A, Batson J, Teyssier N, Lyden A, Caldera S, McGeever A, Dimitrov B, King R, Wilheim J, Murphy M, Ares LP, Travisano KA, Sit R, Amato R, Mumbengegwi DR, Smith JL, Bennett A, Gosling R, Mourani PM, Calfee CS, Neff NF, Chow ED, Kim PS, Greenhouse B, DeRisi JL, Crawford ED. 2019. FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. Nucleic Acids Res 47:e83. 10.1093/nar/gkz418.
17. Boolchandani M, Patel S, Dantas G. 2017. Functional metagenomics to study antibiotic resistance. Methods Mol Biol 1520:307–329. 10.1007/978-1-4939-6634-9_19. PubMed.
18. Boolchandani M, D'Souza AW, Dantas G. 2019. Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet 20:356–370. 10.1038/s41576-019-0108-4. PubMed.
19. Crofts TS, Gasparrini AJ, Dantas G. 2017. Next-generation approaches to understand and combat the antibiotic resistome. Nat Rev Microbiol 15:422–434. 10.1038/nrmicro.2017.28. PubMed.
20. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27:182–189. 10.1038/nbt.1523. PubMed.
21. Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN, Kuch M, Lumibao C, Poinar D, Pearson T, Fourment M, Golding B, Riehm JM, Earn DJD, DeWitte S, Rouillard JM, Grupe G, Wiechmann I, Bliska JB, Keim PS, Scholz HC, Holmes EC, Poinar H. 2014. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. Lancet Infect Dis 14:319–326. 10.1016/S1473-3099(13)70323-2. PubMed.
22. Patterson Ross Z, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, Poinar D, Douglas MW, Eden J-S, Holmes EC, Poinar HN. 2018. The paradox of HBV evolution as revealed from a 16th century mummy. PLoS Pathog 14:e1006750. 10.1371/journal.ppat.1006750.
23. Duggan AT, Perdomo MF, Piombino-Mascali D, Marciniak S, Poinar D, Emery MV, Buchmann JP, Duchêne S, Jankauskas R, Humphreys M, Golding GB, Southon J,

Devault A, Rouillard JM, Sahl JW, Dutour O, Hedman K, Sajantila A, Smith GL, Holmes EC, Poinar HN. 2016. 17th century variola virus reveals the recent history of smallpox. Curr Biol 26:3407–3412. 10.1016/j.cub.2016.10.061. PubMed.

24. Devault AM, Mortimer TD, Kitchen A, Kiesewetter H, Enk JM, Golding GB, Southon J, Kuch M, Duggan AT, Aylward W, Gardner SN, Allen JE, King AM, Wright G, Kuroda M, Kato K, Briggs DEG, Fornaciari G, Holmes EC, Poinar HN, Pepperell CS. 2017. A molecular portrait of maternal sepsis from Byzantine Troy. Elife 6:e20983. 10.7554/eLife.20983. PubMed.

25. Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard J-M, Poinar HN. 2014. Ancient whole genome enrichment using baits built from modern DNA. Mol Biol Evol 31:1292–1294. 10.1093/molbev/msu074. PubMed.

26. Depledge DP, Palser AL, Watson SJ, Lai I-C, Gray ER, Grant P, Kanda RK, Leproust E, Kellam P, Breuer J. 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. PLoS One 6:e27805. 10.1371/journal.pone.0027805. PubMed.

27. Lanza VF, Baquero F, Martínez JL, Ramos-Ruíz R, González-Zorn B, Andremont A, Sánchez-Valenzuela A, Ehrlich SD, Kennedy S, Ruppé E, van Schaik W, Willems RJ, de la Cruz F, Coque TM. 2018. In-depth resistome analysis by targeted metagenomics. Microbiome 6:11. 10.1186/s40168-017-0387-y. PubMed.

28. Noyes NR, Weinroth ME, Parker JK, Dean CJ, Lakin SM, Raymond RA, Rovira P, Doster E, Abdo Z, Martin JN, Jones KL, Ruiz J, Boucher CA, Belk KE, Morley PS. 2017. Enrichment allows identification of diverse, rare elements in metagenomic resistome-virulome sequencing. Microbiome 5:142. 10.1186/s40168-017-0361-8. PubMed.

29. Allicock OM, Guo C, Uhlemann A-C, Whittier S, Chauhan LV, Garcia J, Price A, Morse SS, Mishra N, Briese T, Lipkin WI. 2018. BacCapSeq: a platform for diagnosis and characterization of bacterial infections. mBio 9:e02007-18. 10.1128/mBio.02007-18. PubMed.

30. Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS. 2014. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc 9:989–1009. 10.1038/nprot.2014.058. PubMed.

31. Metsky HC, Siddle KJ, Gladden-Young A, Qu J, Yang DK, Brehio P, Goldfarb A, Piantadosi A, Wohl S, Carter A, Lin AE, Barnes KG, Tully DC, Corleis B, Hennigan S, Barbosa-Lima G, Vieira YR, Paul LM, Tan AL, Garcia KF, Parham LA, Odia I, Eromon P, Folarin OA, Goba A, Simon-Lorière E, Hensley L, Balmaseda A, Harris E, Kwon DS, Allen TM, Runstadler JA, Smole S, Bozza FA, Souza TML, Isern S, Michael SF, Lorenzana I, Gehrke L, Bosch I, Ebel G, Grant DS, Happi CT, Park DJ, Gnirke A, Sabeti PC, Matranga CB. 2019. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. Nat Biotechnol 37:160–168. 10.1038/s41587-018-0006-x. PubMed.

32. Phillippy AM, Deng X, Zhang W, Salzberg SL. 2009. Efficient oligonucleotide probe selection for pan-genomic tiling arrays. BMC Bioinformatics 10:293. 10.1186/1471-2105-10-293. PubMed.
33. Chafin TK, Douglas MR, Douglas ME. 2018. MrBait: universal identification and design of targeted-enrichment capture probes. Bioinformatics 34:4293–4296. 10.1093/bioinformatics/bty548. PubMed.
34. Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. 2011. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol 29:908–914. 10.1038/nbt.1975. PubMed.
35. Shearer AE, Hildebrand MS, Ravi H, Joshi S, Guiffre AC, Novak B, Happe S, LeProust EM, Smith R. 2012. Pre-capture multiplexing improves efficiency and cost-effectiveness of targeted genomic enrichment. BMC Genomics 13:618. 10.1186/1471-2164-13-618. PubMed.
36. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends Microbiol 27:105–117. 10.1016/j.tim.2018.11.003. PubMed.
37. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87. 10.1186/s12915-014-0087-z. PubMed.
38. Sandalli C, Buruk CK, Sancaktar M, Ozgumus OB. 2010. Prevalence of integrons and a new *dfrA17* variant in Gram-negative bacilli which cause community-acquired infections. Microbiol Immunol 54:164–169. 10.1111/j.1348-0421.2010.00197.x. PubMed.
39. Wally N, Schneider M, Thannesberger J, Kastner MT, Bakonyi T, Indik S, Rattei T, Bedarf J, Hildebrand F, Law J, Jovel J, Steininger C. 2019. Plasmid DNA contaminant in molecular reagents. Sci Rep 9:1652. 10.1038/s41598-019-38733-1. PubMed.
40. de Goffau MC, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnock-Jones DS, Peacock SJ, Smith GCS, Parkhill J. 2018. Recognizing the reagent microbiome. Nat Microbiol 3:851–853. 10.1038/s41564-018-0202-y. PubMed.
41. Chiu CY, Miller SA. 2019. Clinical metagenomics. Nat Rev Genet 20:341–355. 10.1038/s41576-019-0113-7. PubMed.
42. Waseem H, Jameel S, Ali J, Saleem Ur Rehman H, Tauseef I, Farooq U, Jamal A, Ali MI. 2019. Contributions and challenges of high throughput qPCR for determining antimicrobial resistance in the environment: a critical review. Molecules 24:E163. 10.3390/molecules24010163. PubMed.
43. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. Nat Methods 7:111–118. 10.1038/nmeth.1419. PubMed.
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local Alignment Search Tool. J Mol Biol 215:403–410. 10.1016/S0022-2836(05)80360-2. PubMed.

45. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2017. GenBank. Nucleic Acids Res 45:D37–D42. 10.1093/nar/gkw1070. PubMed.
46. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. 10.1093/bioinformatics/btq461. PubMed.
47. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. 10.1038/nmeth.1923. PubMed.
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. 10.1093/bioinformatics/btp352. PubMed.
49. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. 10.1093/bioinformatics/btq033. PubMed.
50. Rouillard JM, Zuker M, Gulari E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res 31:3057–3062. 10.1093/nar/gkg426. PubMed.
51. Whelan FJ, Verschoor CP, Stearns JC, Rossi L, Luinstra K, Loeb M, Smieja M, Johnstone J, Surette MG, Bowdish D. 2014. The loss of topography in the microbial communities of the upper respiratory tract in the elderly. Ann Am Thorac Soc 11:513–521. 10.1513/AnnalsATS.201310-351OC. PubMed.
52. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. 10.1093/bioinformatics/btu170. PubMed.
53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. 10.1089/cmb.2012.0021. PubMed.
54. Ames SK, Hysom DA, Gardner SN, Lloyd GS, Gokhale MB, Allen JE. 2013. Scalable metagenomic taxonomy classification using a reference genome database. Bioinformatics 29:2253–2260. 10.1093/bioinformatics/btt389. PubMed.
55. Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010:pdb.prot5448. 10.1101/pdb.prot5448. PubMed.
56. Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182. 10.1186/1471-2105-15-182. PubMed.
57. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. Bamtools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691–1692. 10.1093/bioinformatics/btr174. PubMed.

58. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C. 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. Nucleic Acids Res 45:D574–D580. 10.1093/nar/gkw1009. PubMed.

SUPPLEMENTARY INFORMATION

Supplementary methods:

Modified Library Preparation for Trial 2

The same genomic DNA (except for *Pseudomonas aeruginosa* C00060) used in Trial 1 was used for library construction through a modified library preparation protocol (1). Briefly, ~200 nanograms of sonicated genomic DNA (250 - 300 bp) was used for library preparation at a volume of 25 µL. Deviations from the protocol include the use of KAPA Pure Beads (KK8000, Roche Canada, Laval, QC) instead of AMPure beads, oligonucleotides ordered from Integrated DNA Technologies (IDT, Coralville, IA, USA) (the ready-to-use adapter mix was diluted to 10uM of each adapter prior to use), and a total reaction volume of 50 µL for blunt end repair. Briefly, blunt end repair with T4 polynucleotide kinase (10U/µL) and T4 DNA polymerase (5U/µL) was carried out for 15 min at 25ºC and 15 min at 12ºC in a modified buffer (50mM NaCl, 10mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9 at 25C) with 1 mM DTT, 100 µM dNTPs (2.5 mM each), and 1 mM ATP. Adapter ligation was performed using a final adapter concentration of 2.5µM, with T4 DNA ligase (0.125U/µL) in T4 DNA ligase buffer (1X) and PEG-4000 (5%) for 15 hours at 16ºC, then held at 4ºC. The reaction was purified using KAPA Pure beads at 1.8X and eluted in 20 µL of EBT (10 mM Tris-Cl, pH 8.3, 0.05% Tween-20). This

was followed with adapter fill-in with Bst polymerase, large fragment (8U/μL) in ThermoPol reaction buffer (10X) with dNTPs (final concentration of 250 μM each) (30 mins at 37ºC; 80ºC for 20 mins; hold at 4ºC). Purification with library size-selection was performed using KAPA Purebeads (0.6X-0.8X ratio of beads to reaction volume) and products were eluted in 30 μL of 10mM Tris-HCl, pH 8.3. Indexing PCR was performed using 12.5 μL of template, KAPA HiFi Hotstart DNA polymerase ready mix (2X), (750nM of each indexing primer – 8 bp primers ordered from IDT diluted in nuclease free water at a stock concentration of 100μM) in a 40 μL reaction with the following conditions: 1) 98ºC for 45 sec; 2) 98ºC for 15 sec; 3) 60ºC for 45 sec; 4) 72ºC for 30 sec; 5) Repeat 2-4 for 12 cycles total; 6) 72ºC for 1 min. Libraries were purified with 1.5X KAPA Pure beads then eluted in 33 μL 10 mM Tris pH 8.0. Re-amplification of libraries for 2 cycles was performed using 1 μL of each library sample and KAPA HiFi Hotstart DNA polymerase ready mix (2X) and Library Amplification Primer Mix (10X) supplied by KAPA in the Library Amplification Kit with the followed conditions: 1) 98ºC for 45 sec; 2) 98ºC for 15 sec; 3) 60ºC for 45 sec; 4) 72ºC for 30 sec; 5) Repeat 2 - 4 for 2 cycles total; 6) 72ºC for 1 min. Libraries were purified using 1.5X Kapa Purebeads and eluted in 33 μL of 10 mM Tris-HCl pH 8.3. A negative control was also included consisting of a blank DNA extraction carried throughout the DNA extraction of bacterial DNA. Bioanalyzer and qPCR analyses were performed through the McMaster Farncombe Metagenomics Facility on all libraries including the blanks.

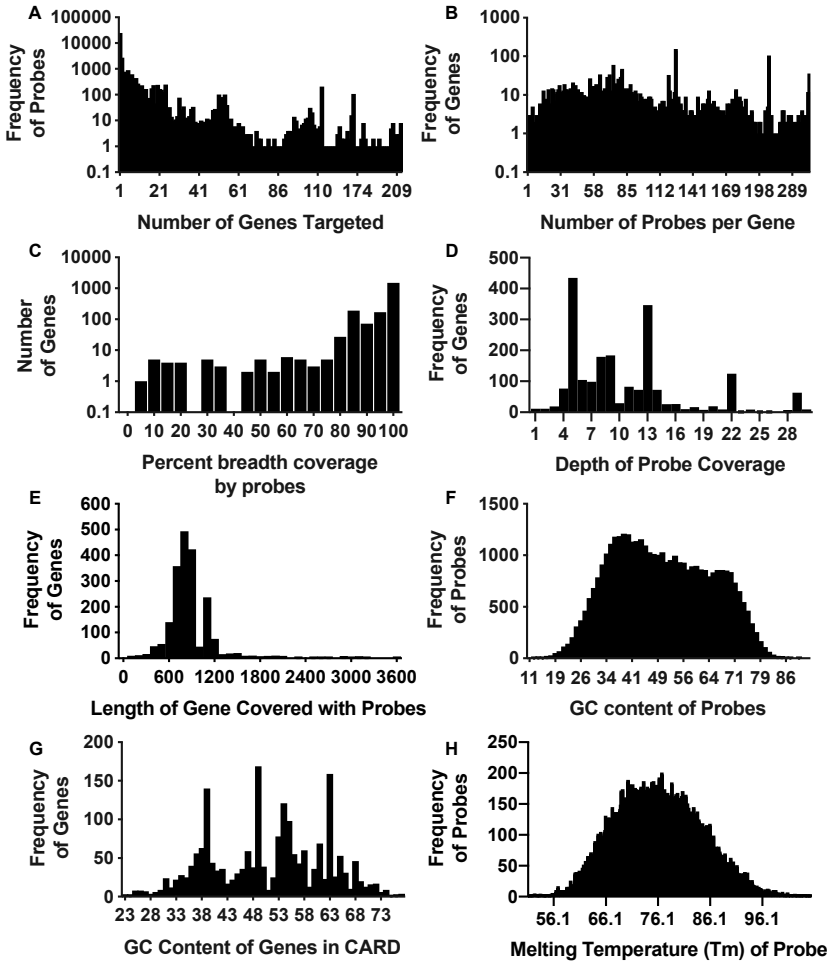
Enrichment methods

For Phase 1 Trial 1, the hybridization mix and blockers mix were prepared as described in the myBaits® manual version 3 (Arbor Biosciences, Ann Arbor, MI). 7 μL of DNA library (~100 ng) was added to the blockers mix and libraries were denatured at 95ºC for 5 minutes. The hybridization mix and libraries with blockers were brought to the hybridization temperature of 65ºC for 5 minutes. Hybridization mixes including the appropriate concentration of probeset were mixed with libraries and incubated at 65ºC for 16 hours. After 16 hours, the wash buffer was prepared and heated to 65ºC. Dynabeads MyOne Streptavidin C1 beads were prepared with binding buffer (10 mM Tris-HCl pH7.5, 1 mM EDTA, 2M NaCl in nuclease-free water) by three washes with binding buffer, and finally resuspended in binding buffer then heated to 65ºC (Thermo Fisher, Waltham, MA). Prepared beads were added to capture reactions, incubated at 65ºC for 30 minutes with frequent mixing, then pelleted on a magnetic rack. Beads were washed with preheated wash buffer (Supplied in the myBaits® kit), incubated for 10 minutes at 65ºC and the wash was repeated two times. The beads were finally resuspended in 30 μL 10mM Tris-Cl, 0.05% Tween-20, pH 8.3. Post-capture amplification was performed on 15 μL of the bead-bound library with for 30 cycles of amplification using KAPA HiFi Hotstart DNA polymerase ready mix (2X), Library Amplification Primer Mix (10X) with the followed conditions: 1) 98ºC for 45 sec; 2) 98ºC for 15 sec; 3) 60ºC for 45 sec; 4) 72ºC for 30 sec; 5) Repeat 2-4 for 30 cycles total; 6) 72ºC for 1 min. Reactions were purified with KAPA Purebeads (0.8X cleanup) and eluted in 30 μL of 10 mM Tris-HCl, pH 8.3. These reaction products were

purified with KAPA pure beads (1X) and eluted in 30 μL 10 mM Tris-HCl, pH 8.3. This is an over-amplified product and likely contains heterodimers of library fragments. An aliquot (2 μL) was used for an additional 3 cycles of amplification with the same primers and conditions as above. A 1X KAPA Purebeads purification was performed and products were eluted in 30 μL TE buffer. The concentration was verified by absorbance, followed by the Bioanalyzer and qPCR through the McMaster Metagenomics Facility. For Trial 2, conditions were similar to the enrichment for Trial 1 libraries except for post-enrichment which was performed using 7.5 μL of enriched library, 12.5 μL of KAPA Ready Mix 2X, 1 μL of 10X primer mix (1) 98ºC 45 sec; 2) 98ºC 30sec; 3) 60ºC for 45sec; 4) 72ºC for 45sec; 5) Repeat step 2 - 4 for 17 cycles; 6) 72ºC for 5 min; 7) 4ºC hold. PCR products were purified using KAPA Pure beads (1.5X) and eluted in 10 mM Tris-HCl pH 8.0. The concentration was determined via absorbance and additional library quality control was performed through the McMaster Metagenomics Facility (Bioanalyzer and qPCR). Products were sequenced via Illumina MiSeq 2 x 250 bp.

References

1. Meyer M, Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb Protoc 2010:pdb.prot5448-pdb.prot5448.



Supplementary Figure 1: Probe and CARD nucleotide sequence statistics
A) Number of genes targeted by probes through mapping with Bowtie2. B) Number of probes targeting genes through mapping with Bowtie2. C) Percent length coverage of genes with probes. D) Mean depth of probe coverage across individual genes in CARD. E) Length of sequence targeted by probes in genes in CARD. F) GC content of probes. G) GC content of genes in CARD. H) Melt temperature of final list of probes. Figures generated using Prism 8 for macOS.

Supplementary Table 1: Bacterial strains used in control experiments
Clinical bacterial isolates obtained through the Wright Clinical Collection. Bacterial genomes were sequenced, and draft genome assemblies were analyzed through the Resistance Gene Identifier in CARD to predict the number of resistance genes. The total probeset was mapped against the draft assembled genome and the number of genes with

probe coverage, percentage of genome covered by probes and overlap between predicted RGI genes and probe coverage were determined.

Bacterial strain	Genome size (Mb)	GC Content (%)	Predicted genes by RGI	Region predicted by RGI (%)	# Probe-targeted sites	Length of probe-targeted site (average and range)	Region with probe coverage (%)	RGI genes with probes	Region predicted by RGI and targeted by probes (%)
Escherichia coli C0002	5.29	50.62	67	1.64	65	797.75 (80 - 3595)	0.97	43	0.81
Klebsiella pneumoniae C0006	5.45	57.23	30	0.55	35	331.54 (80 - 877)	0.21	17	0.17
Staphylococcus aureus C0018	2.92	32.66	16	0.55	13	1127.54 (140 – 2013)	0.50	12	0.41
Staphylococcus aureus C0033	2.92	32.77	16	0.64	14	1143.07 (155 – 2130)	0.52	13	0.44
Klebsiella pneumoniae C0050	5.60	57.05	34	0.63	40	346.18 (80 – 900)	0.25	18	0.19
Pseudomonas aeruginosa C0060	6.80	66.19	53	1.18	48	933.35 (97 – 3415)	0.66	33	0.54
Escherichia coli C0094	5.22	50.74	67	1.65	64	779.86 (80 – 3003)	0.95	41	0.79
Pseudomonas aeruginosa C0292	6.81	66.21	54	1.17	48	938.71 (97 – 3415)	0.66	33	0.57

Supplementary Table 2: Library and sequencing information

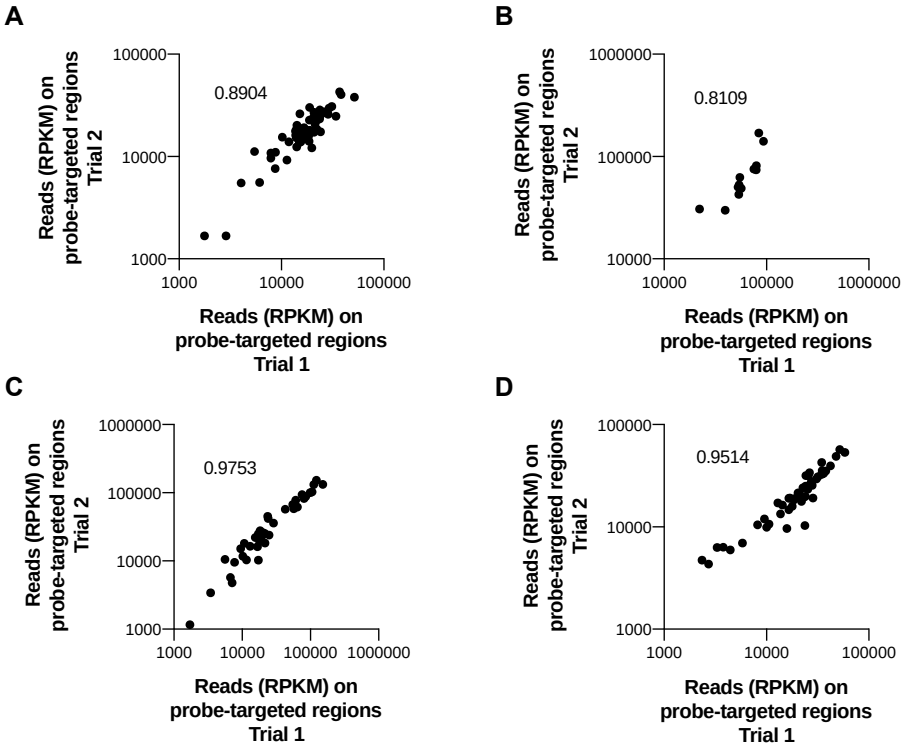
The amount in nanograms of each library and the corresponding amount of probes used for enrichment. The average size of library fragments prior to enrichment was determined through the Agilent Bioanalyzer 2100. The number of clusters (paired-end reads) that were generated for each library when sequenced by Illumina's MiSeq V2 2x250. Blanks for each trial were included and sequenced on a separate run; many of the blank libraries did not generate peaks on the Bioanalyzer nor any signal by quantitative PCR therefore their values are N/A. In Phase 2, three positive controls for enrichment were included with genomic DNA from *Escherichia coli* C0002 and varying library and probe amounts.

Phase	Trial/Set	Library	Amount of Probes (ng)	Amount of Library (ng)	Average Library Size (bp)	Clusters sequenced enriched	Clusters sequenced shotgun
Phase 1	Trial 1	C0002	100	100	988	66926	
		C0018	100	100	994	75860	
		C0050	100	100	1222	73941	
		C0060	100	100	1225	81810	
		Pool 1	100	100	1257	61568	218008
		Pool 2	100	100	1158	61658	159059
		Pool 3	100	100	1216	58308	109194
		Negative Control - Blank	100	N/A	632	170565	
	Trial 2	C0002	100	100	435	99748	
		C0018	100	100	438	143804	
		C0050	100	100	416	153673	
		C0060	100	100	403	124971	
		Pool 1	100	100	429	86023	29241
		Pool 2	100	100	413	124170	33488
		Pool 3	100	100	427	127682	32560
		Negative Control - Blank	100	N/A	345	44026	
Phase 2	Set 1	1 - 1	25	50	952	89768	
		1 - 2	50	50	968	77117	
		1 - 3	100	50	919	65746	
		1 - 4	50	100	1044	55783	
		1 – 5	100	100	972	64761	
		1 - 6	200	100	940	71099	3652948
		1 - 7	100	200	915	15211	4405779
		1 - 8	200	200	1020	59409	
		1 - 9	400	200	998	25911	
		Negative Control - Blank	50	N/A	276	2590	
	Positive Controls	C0002 – 1 - 1	100	50	986	80647	
		C0002 – 1 - 2	50	50	939	116965	
		C0002 – 1 - 3	25	50	976	112881	
	Set 2	2 - 1	25	50	955	158710	
		2 - 2	50	50	887	100590	
		2 - 3	100	50	891	102689	
		2 - 4	50	100	902	120764	
		2 - 5	100	100	956	141994	6151998
		2 - 6	200	100	941	159192	
		2 - 7	100	200	790	96211	
		2 - 8	200	200	944	129333	
		2 - 9	400	200	871	76195	7660355
		Negative Control - Blank	50	N/A	N/A	3804	
	Positive Controls	C0002 – 2 -1	100	33	993	139909	
		C0002 – 2 - 2	50	50	935	235429	
		C0002 – 2 - 3	25	50	876	129070	
	Set 3	3 - 1	25	50	854	82778	5866495
		3 - 2	50	50	888	158968	
		3 - 3	100	50	910	65675	
		3 - 4	50	100	889	103671	
		3 - 5	100	100	882	78251	4213540
		3 - 6	200	100	943	68331	
		3 - 7	100	200	820	96722	
		3 - 8	200	200	934	79036	
		3 - 9	400	200	917	82375	
		Negative Control - Blank	50	N/A	N/A	5962	
	Positive Controls	C0002 – 3 -1	100	38	846	54117	
		C0002 – 3 - 2	50	32	881	96258	
		C0002 – 3 - 3	25	38	779	110746	

Supplementary Table 3: Individual strain enrichment results

Strains were enriched individually in two trials with different library sizes. For each strain the regions predicted to be targeted by probes were determined through mapping the probeset to each individual genome). Enrichment results across two trials were determined by mapping trimmed and filtered reads to genome, calculating the percentage on-target and normalizing reads and depth per kb per million reads for probe-targeted regions with at least 10 reads.

Strain	Average % mapping to genome	Average % mapping to probe-targeted sites	% of targeted regions with at least 10 reads	Average % coverage of targeted regions	Average reads per kb per million reads on probe-targeted region	Average depth per kb per million reads on probe-targeted region
Escherichia coli C0002	96.67 (±2.72)	95.07 (±1.54)	100	100	18975.73 (±414.91)	6192.13 (±297.27)
Staphylococcus aureus C0018	97.99 (±1.98)	94.89 (±2.31)	100	100	67615.06 (±4360.20)	19968.28 (±2670.37)
Klebsiella pneumoniae C0050	95.60 (±3.96)	85.74 (±4.68)	100	100	40531.43 (±2516.77)	17315.24 (±1630.66)
Pseudomonas aeruginosa C0060	91.45 (±5.49)	90.73 (±0.95)	100	100	22725.67 (±32.97)	6497.48 (±61.46)



Supplementary Figure 2: Consistency in library prep methods and trials
Comparing read counts normalized in subsampled individual enrichment trials through different library prep methods. Reads from enrichment of individual genomes of *Escherichia coli* C0002 (A), *Staphylococcus aureus* C0018 (B), *Klebsiella pneumoniae*

C0050 (C) and *Pseudomonas aeruginosa* C0060 (D) in Trial 2 were subsampled to same depth as reads in Trial 1. The reads were mapped to the respective bacterial genome, filtered for mapping quality and then the number of reads on each RGI and probe-targeted region were counted and normalized per kb per million reads. Pearson correlation coefficients are shown.

Supplementary Table 4: Pooling of genomic DNA to create "mock metagenomes"
We pooled various nanogram amounts of genomic DNA from bacteria and estimated the percentage of each strain in the respective pools based on total genome size of each strain. With reads generated through shotgun sequencing and after enrichment, we calculated the percentage of reads mapping to a particular genome by mapping to a combined reference of the genomes used in a given pool and counting the reads that mapped to each respective genome (= reads mapping to genome A / reads mapping to all genomes). The percentage represented by targeted regions of each genome in each pool is determined by taking the percentage of each isolate's genome that is targeted by probes and the percentage of reads that map to that particular genome from the shotgun sequencing data.

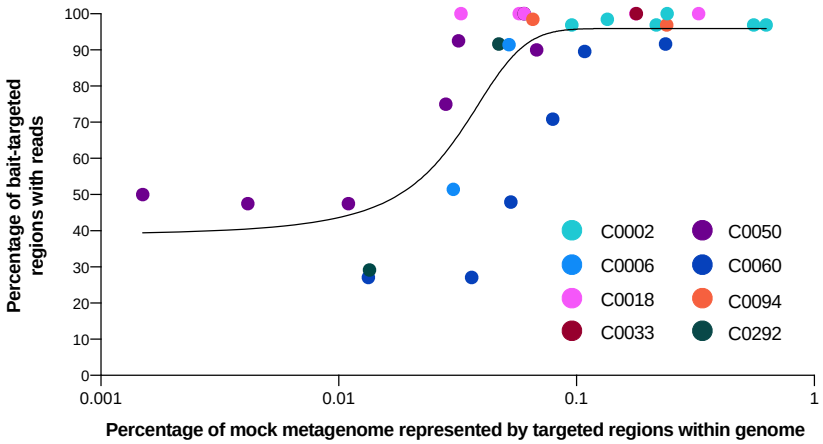
Pool	Strain	Amount of genomic DNA pooled (ng)	Estimated % of pool	% of reads mapping from shotgun	Targeted % of pool	% of reads mapping from enriched
Trial 1 Pool 1	C0002	312	21.98	24.82	0.24	52.55
	C0018	312	40.00	12.06	0.060	32.12
	C0050	312	20.74	27.18	0.068	8.86
	C0060	312	17.28	35.93	0.24	6.47
Trial 2 Pool 1	C0002	112	18.77	22.30	0.22	33.95
	C0018	174	53.01	65.29	0.33	62.88
	C0050	106	16.79	4.39	0.011	1.54
	C0060	88	11.43	8.02	0.053	1.63
Trial 1 Pool 2	C0002	1250	66.30	64.73	0.63	71.26
	C0018	180	17.22	11.96	0.060	19.69
	C0050	180	9.07	11.28	0.028	4.75
	C0060	180	7.41	12.03	0.079	4.30
Trial 2 Pool 2	C0002	264	48.04	57.31	0.56	65.39
	C0018	102	33.92	35.54	0.18	33.24
	C0050	62	10.75	1.66	0.0042	0.44
	C0060	51	7.29	5.49	0.036	0.94
Trial 1 Pool 3	C0002	125	11.01	13.91	0.13	38.50
	C0006	125	10.70	24.75	0.052	2.34
	C0018	125	19.88	6.54	0.033	11.62
	C0033	125	19.88	11.59	0.060	22.81
	C0050	125	10.40	12.75	0.032	2.73
	C0060	125	8.56	16.40	0.11	2.16
	C0094	125	11.01	6.90	0.066	18.78
	C0292	125	8.56	7.15	0.047	1.07
Trial 2 Pool 3	C0002	46	8.65	9.84	0.095	14.80
	C0006	83	8.16	14.44	0.030	1.53
	C0018	43	28.17	11.49	0.057	12.49
	C0033	36	28.15	34.36	0.18	34.58
	C0050	45	7.68	0.60	0.0015	0.13
	C0060	83	5.20	2.02	0.013	0.42
	C0094	46	8.78	25.21	0.24	35.67
	C0292	36	5.21	2.04	0.013	0.39

Supplementary Table 5: Enrichment results to probe-targeted regions in pooled samples

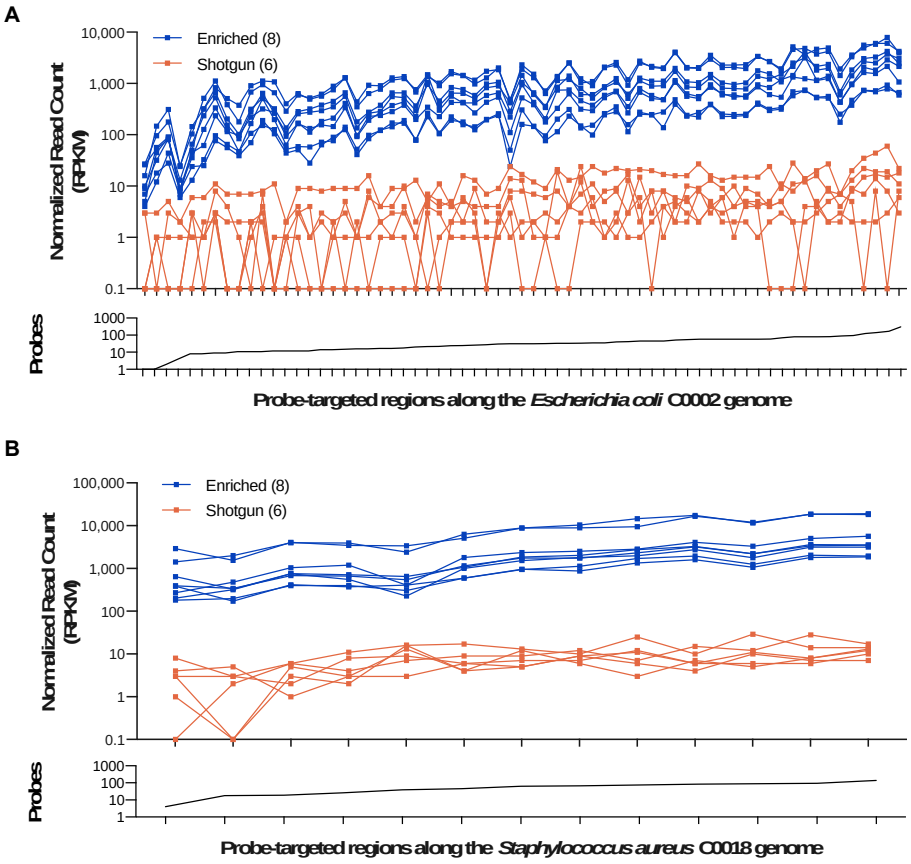
Genomic DNA from individual strains was pooled in various ratios to produce "mock metagenomes" for enrichment. For each strain, the regions predicted be targeted by probes (determined through mapping the probeset to each individual genome) are considered the targeted region for analysis. Trimmed and filtered reads from paired enriched and shotgun pools were subsampled to same read depth. The resulting reads were mapped to the individual strain's genomes, counted on-target and normalized per kb per million reads mapping. Percentage on-target, percentage of probe-targeted regions with at least 10 reads

as well as their percent coverage, average reads, and average depth were determined for each strain at the probe-targeted region level. The fold enrichment is based on all genes regardless of read counts.

Sample	Strain	% of reads in Pool	% Mapping to probe-targeted regions	% of probe-targeted regions with reads	% coverage of probe-targeted regions	Average reads per kb per million reads	Average depth per kb per million reads	Fold-enrichment in reads (average and range)
Trial 1 Pool 1 Enriched	C0002	52.75	93.06	100	100	19097.95	6091.42	810.18 (2.66 – 16590.95)
	C0018	20.05	94.84	100	100	67393.09	19715.42	135.84 (31.11 – 291.78)
	C0050	18.73	85.44	90	100	41944.82	16304.97	1341.88 (3.77 – 23020.26)
	C0060	3.40	90.26	91.67	98.73	24920.46	6697.48	994.87 (0 – 21945.61)
Trial 1 Pool 1 Shotgun	C0002	21.61	1.56	18.46	90.13	671.09	153.52	
	C0018	10.32	0.70	15.38	88.03	820.59	161.15	
	C0050	23.56	0.82	25.00	100	762.34	190.87	
	C0060	28.70	0.81	12.50	84.54	301.55	44.92	
Trial 2 Pool 1 Enriched	C0002	35.84	98.90	96.92	100	20081.94	6630.47	4972.95 (2.84 – 35942.31)
	C0018	56.55	98.56	100	100	74814.49	24542.74	144.41 (41.36 – 332.17)
	C0050	7.72	97.63	47.50	99.75	74609.44	24141.06	18991.42 (0 – 170582.07)
	C0060	1.31	93.37	47.92	83.22	30865.24	7310.50	17166.87 (0 – 70414.91)
Trial 2 Pool 1 Shotgun	C0002	23.52	1.49	1.54	91.65	471.34	30.86	
	C0018	57.30	0.71	76.92	79.03	570.56	98.30	
	C0050	5.19	0.88	0	0	0	0	
	C0060	6.65	0.65	0	0	0	0	
Trial 1 Pool 2 Enriched	C0002	68.39	77.35	96.92	100	15928.54	4982.54	57.09 (2.57 – 192.18)
	C0018	12.69	79.11	100	100	56570.38	16316.11	2614.81 (15.93 – 32565.71)
	C0050	12.61	74.13	75.00	99.93	41711.08	15702.37	2727.71 (0 – 39495.86)
	C0060	2.34	38.95	70.83	96.27	11523.24	2820.94	2382.94 (0 – 19387.19)
Trial 1 Pool 2 Shotgun	C0002	58.69	1.34	58.46	96.92	321.15	81.43	
	C0018	10.64	0.74	30.77	78.51	896.24	141.82	
	C0050	11.48	1.33	20	100	1745.41	464.15	
	C0060	9.72	0.75	2.08	56.38	266.69	18.15	
Trial 2 Pool 2 Enriched	C0002	65.64	98.29	96.92	100	19970.52	6708.67	1190.08 (7.74 – 29085.20)
	C0018	28.13	98.15	100	100	75034.93	24899.52	210.58 (32.41 – 596.02)
	C0050	10.26	98.23	47.50	100	77537.34	26906.17	8270.19 (0 – 50937.25)
	C0060	0.73	88.86	27.08	78.56	37440.00	8936.77	18933.20 (0 – 106732.35)
Trial 2 Pool 2 Shotgun	C0002	56.47	1.38	20.00	73.49	404.35	72.86	
	C0018	29.19	0.57	23.08	73.76	698.55	125.73	
	C0050	3.01	4.51	2.50	79.03	10409.44	2093.37	
	C0060	4.27	0.73	0	0	0	0	
Trial 1 Pool 3 Enriched	C0002	38.74	94.12	98.46	100	19755.27	6312.06	2493.04 (3.05 – 22767.27)
	C0006	13.66	84.08	91.43	100	51010.68	22066.06	3295.94 (0 – 61249.67)
	C0018	29.65	95.22	100	100	63154.77	15991.26	2909.12 (54.61 – 35638.08)
	C0033	33.17	94.82	100	100	56232.72	13178.66	156.78 (28.17 – 314.91)
	C0050	14.84	85.22	92.5	100	43478.45	18486.32	2475.78 (4.87 – 47799.65)
	C0060	2.45	91.97	89.58	98.78	26022.10	7430.52	3742.84 (3.65 – 62302.44)
	C0094	35.52	92.59	98.44	100	19949.59	6561.88	3526.16 (2.48 – 23220.26)
	C0292	2.78	84.96	91.67	99.29	28432.58	10574.24	4014.72 (0 – 54962.31)
Trial 1 Pool 3 Shotgun	C0002	9.83	1.63	3.08	88.69	1449.60	308.97	
	C0006	25.19	0.36	8.57	95.63	3450.36	1206.18	
	C0018	11.94	0.51	7.69	68.26	413.96	50.49	
	C0033	12.81	0.59	7.14	68.25	424.67	47.08	
	C0050	24.09	0.48	12.5	93.25	853.91	300.04	
	C0060	17.84	0.90	4.17	64.69	222.28	16.28	
	C0094	8.25	1.67	3.125	88.69	1726.91	368.08	
	C0292	16.78	0.94	4.17	64.69	1141.24	84.87	
Trial 2 Pool 3 Enriched	C0002	32.65	98.09	96.92	99.97	20307.57	6847.06	7369.15 (4.14 – 66339.3)
	C0006	7.75	90.49	51.43	99.50	86220.71	36708.00	25683.46 (0 – 271673.69)
	C0018	45.46	97.45	100	100	65485.29	17173.26	5819.09 (29.42 – 74023.04)
	C0033	52.11	97.53	100	100	58846.80	13719.18	698.58 (72.34 – 8084.37)
	C0050	8.22	92.65	50.00	99.55	74207.10	29767.85	21813 (0 – 256173.72)
	C0060	0.86	90.00	27.08	79.68	39544.66	8226.37	16172.91 (0 – 70505.29)
	C0094	34.91	97.65	96.87	100	20612.44	7021.48	7479.75 (2.67 – 61794.38)
	C0292	0.89	89.30	29.17	80.95	44281.92	13985.84	18128.93 (0 – 120321.02)
Trial 2 Pool 3 Shotgun	C0002	16.88	1.38	0	0	0	0	
	C0006	15.36	0.47	0	0	0	0	
	C0018	41.07	0.70	38.46	73.84	525.28	55.49	
	C0033	44.54	0.79	50.00	77.22	703.43	113.13	
	C0050	12.76	0.64	0	0	0	0	
	C0060	4.54	0.77	0	0	0	0	
	C0094	21.50	1.23	1.56	68.13	404.24	25.04	
	C0292	4.59	0.86	0	0	0	0	



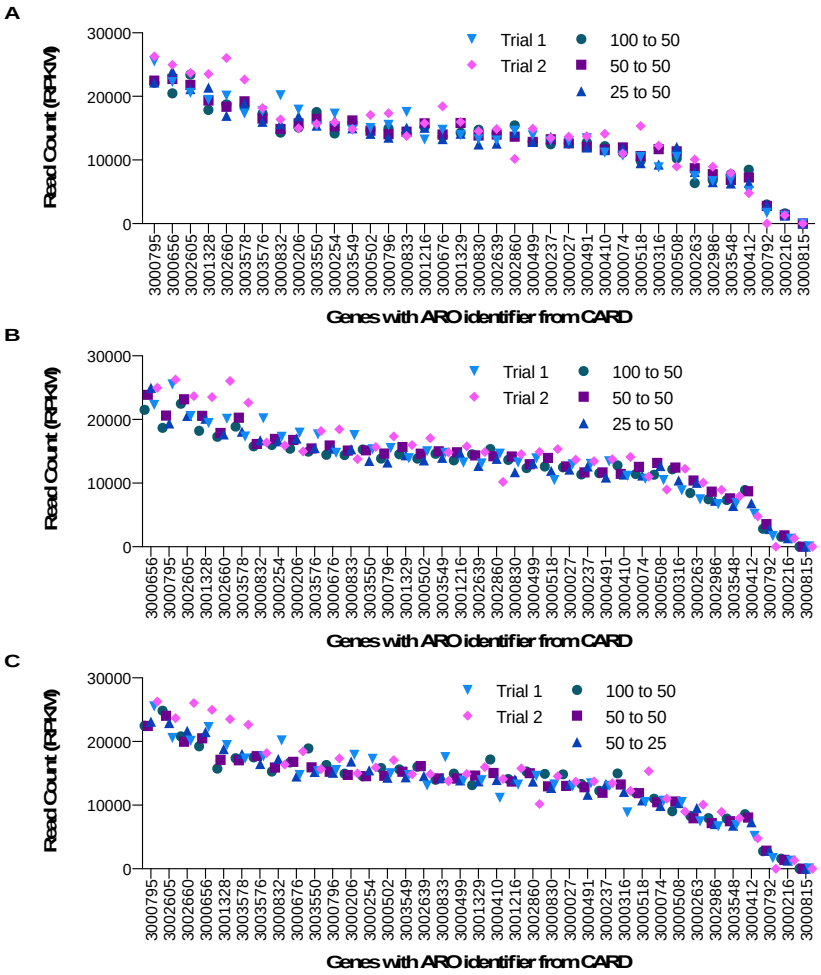
Supplementary Figure 3: Sensitivity in recovery of bait-targeted regions with reads
Using the estimated percentage of a pool represented by each isolate's genome from the percentage of reads mapping in the shotgun dataset, the estimate percentage of the mock metagenome represented by that isolate's bait-targeted complement was determined. This represents the limit of detection or sensitivity for each isolate/genome tested. The Y-axis is the percentage of probe-targeted regions of a given isolate's genome that were successfully captured/recovered with at least 10 reads after filtering.



Supplementary Figure 4: Enrichment results in higher read counts on antibiotic resistance genes compared to shotgun sequencing
Raw read counts at each probe-targeted region within the *Escherichia coli* C0002 genome (A) and *Staphylococcus aureus* C0018 genome (B) in enriched and shotgun samples including individual and "mock metagenomes" of multiple strains. Among enriched and shotgun pairs, reads were subsampled to equal depths and mapped to the individual strain's genome. The predicted number of probes for each region along the genome are shown in the panels below. The Y axes are in the logarithmic scale.

Supplementary Table 6: Control enrichment with *Escherichia coli* C0002.
Enrichment results from the positive control of *E. coli* C0002 control used in Phase 2. Trimmed and deduplicated reads were mapped to CARD using RGIBWT, filtered by genes with probe coverage, an average read mapping quality >=11, and percent length coverage of a gene with reads >=80%.

	Probes (ng)	Library (ng)	% reads mapping to CARD	Total number of genes	Genes with map quality >=11	Genes with probes	Genes with length coverage with reads >=80%	Genes with probes and map quality >=11	Genes passing all filters
C0002 – Set 1	25	50	63.52	164	51	53	86	39	36
	50	50	64.81	164	54	53	84	39	36
	100	50	63.75	154	53	53	80	40	36
C0002 – Set2	25	50	61.10	179	62	54	82	42	36
	50	50	65.77	195	60	59	84	44	36
	100	33	60.31	170	59	57	87	42	36
C0002 – Set 3	25	38	65.46	182	58	57	86	39	36
	50	32	65.77	172	58	53	88	40	36
	100	38	67.98	147	54	56	83	42	36

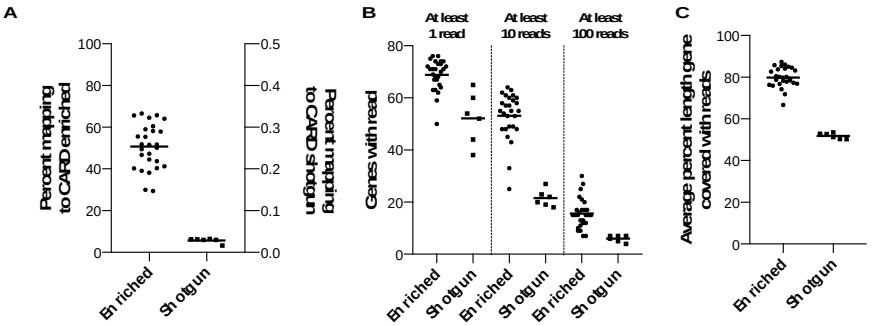


Supplementary Figure 5: Genes identified in positive control enrichments

Normalized read counts from C0002 control enrichments from rgi bwt results from 3 samples in each Set (A) Set 1, B) Set 2, C) Set 3) to the two trials of individual enrichment. Genes with reads were filtered based on read mapping quality greater than or equal to 11, percent length coverage of a gene with reads greater than or equal to 80% and genes with probes mapping. Reads were normalized per kb of reference gene per million reads mapping. Genes are ordered by sum of read counts from highest to lowers (left to right) with the ARO identifier shown along the X axis.

Supplementary Table 7: Phase 2 enrichment results with the full number of reads
For the enriched samples, trimmed and deduplicated reads were mapped to CARD using RGIBWT, filtered by genes with at least 10 reads, those with probes, an average read mapping quality >=11, and length coverage of a gene with reads >=10%. For the shotgun samples, trimmed and deduplicated reads were mapped to CARD using RGIBWT, filtered by genes with an average read mapping quality >=11 and read length coverage of a gene >=10%. EN = enriched, UN = shotgun.

	Probes (ng)	Library (ng)	Reads mapping to CARD (%)	Total number of genes	Genes with read map quality >=11	Genes with probes	Genes with read length coverage >=10%	Genes passing all filters
Sample Set 1								
EN	25	50	55.36	60	50	51	58	48
	50	50	65.73	62	54	52	60	49
	100	50	55.59	60	50	50	60	48
	50	100	65.63	56	47	46	55	43
	100	100	51.85	61	51	51	60	48
	200	100	58.21	64	56	53	61	49
	100	200	51.52	34	26	27	34	25
	200	200	66.57	60	50	48	59	45
	400	200	49.44	45	37	36	43	33
UN	200	100	0.030	26	19	N/A	24	18
	100	200	0.030	32	22	N/A	29	20
Sample Set 2								
EN	25	50	64.07	78	67	64	76	61
	50	50	64.60	72	64	61	71	58
	100	50	57.96	75	64	61	74	57
	50	100	46.75	78	66	66	76	62
	100	100	58.99	79	69	64	77	61
	200	100	44.52	85	72	69	80	63
	100	200	60.43	76	66	62	73	59
	200	200	47.27	82	71	67	81	64
	400	200	41.22	70	59	58	69	55
UN	400	200	0.016	41	28	N/A	37	27
	100	100	0.032	34	24	N/A	32	23
Sample Set 3								
EN	25	50	50.16	72	63	61	70	58
	50	50	38.19	79	66	64	76	60
	100	50	51.73	69	59	59	68	55
	50	100	29.46	78	66	63	76	60
	100	100	40.28	74	65	60	72	57
	200	100	39.06	67	57	57	67	53
	100	200	29.97	69	57	58	68	54
	200	200	40.32	72	60	58	71	55
	400	200	43.74	69	58	56	67	53
UN	100	100	0.031	29	19	N/A	26	19
	25	50	0.031	34	23	N/A	30	22



Supplementary Figure 6: Comparing enriched and shotgun ARG recovery

For the enriched and shotgun samples, the full number of reads for each sample were mapped to CARD using rgi bwt. A) The percentage of reads mapping to CARD. B) Genes were counted with at least 1, 10 and 100 reads and filtered for mapping quality (>=11), percent coverage by reads (>=10) and probes mapping (only for the enriched samples). C) The average percent coverage of all genes with at least 10 reads in each sample after the same filters used in B.

Supplementary Table 8: Phase 2 enrichment results with subsampled reads.
For the enriched samples, reads were subsampled to 22,324 reads and mapped to CARD using RGIBWT. Results were filtered by genes with at least 10 reads, those with probes, an average read mapping quality >=11, and length coverage of a gene with reads >=10%. For the shotgun samples, reads were subsampled to their paired enriched sample and mapped to CARD using RGIBWT. Results were filtered by genes with an average read mapping quality >=11 and read length coverage of a gene >=10%. EN = enriched, UN = shotgun.

	Probes (ng)	Library (ng)	Reads mapping to CARD (%)	Total number of genes	Genes with read map quality >=11	Genes with probes	Genes with read length coverage >=10%	Genes passing all filters
Sample Set 1								
EN	25	50	55.24	34	26	27	34	25
	50	50	65.84	39	31	31	37	28
	100	50	56.11	46	37	37	45	34
	50	100	66.01	39	32	32	39	30
	100	100	51.94	40	32	32	37	28
	200	100	57.93	38	30	30	37	28
	100	200	51.52	34	26	27	34	25
	200	200	66.99	42	34	33	39	30
	400	200	49.39	33	26	26	33	24
UN	200	100	0.038	2	2	N/A	2	2
	100	200	0.054	0	0	N/A	0	0
Sample Set 2								
EN	25	50	64.25	41	33	34	40	32
	50	50	64.11	43	36	35	40	31
	100	50	58.80	43	36	35	43	33
	50	100	46.95	40	32	33	38	29
	100	100	59.13	42	35	34	41	31
	200	100	44.64	45	35	34	41	31
	100	200	60.55	50	42	42	49	39
	200	200	47.29	45	38	37	45	35
	400	200	41.56	43	34	35	41	32
UN	400	200	0.029	1	1	N/A	1	1
	100	100	0.035	2	2	N/A	2	2
Sample Set 3								
EN	25	50	50.64	37	29	30	36	27
	50	50	37.85	27	19	20	27	18
	100	50	51.41	36	27	28	33	24
	50	100	29.56	29	21	22	28	20
	100	100	40.77	34	26	26	33	24
	200	100	38.86	37	30	30	37	28
	100	200	30.08	31	23	24	30	21
	200	200	40.62	34	26	26	32	23
	400	200	44.35	37	30	29	35	26
UN	100	100	0.023	0	0	N/A	0	0
	25	50	0.023	1	1	N/A	1	1

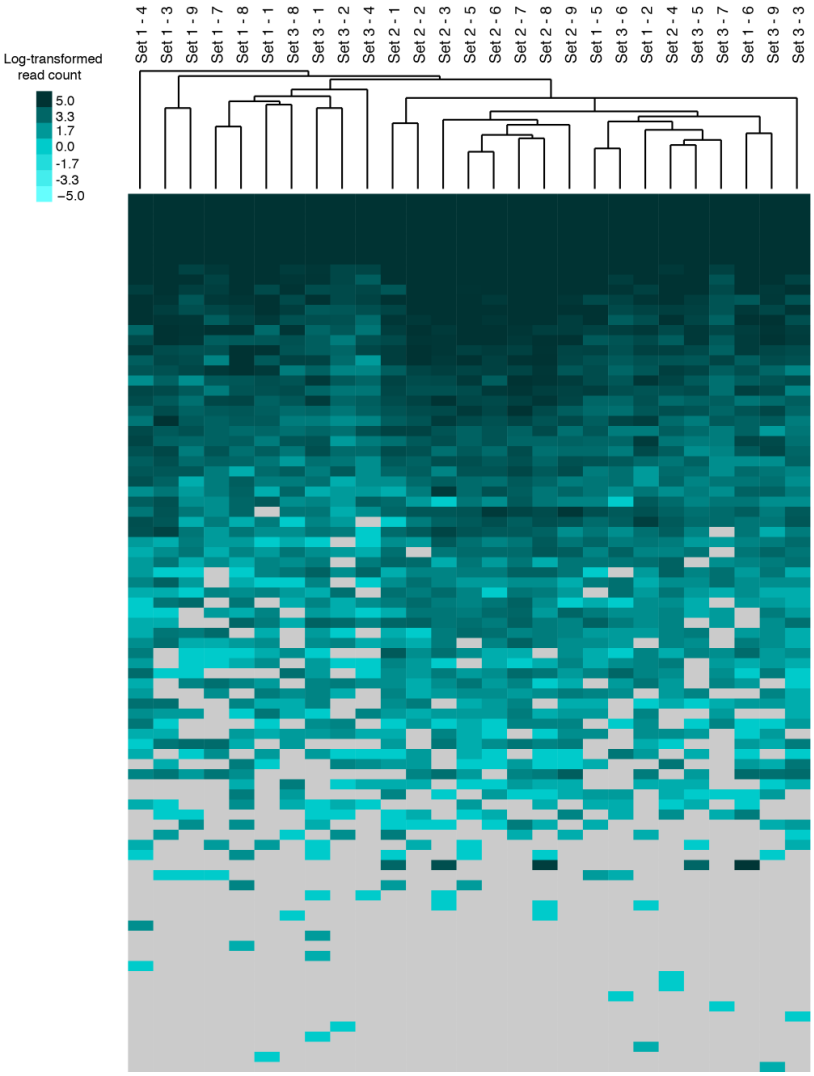
Supplementary Table 9: Phase 2 overlapping genes with the full number of reads
We calculated the overlap of genes with at least 10 reads passing the percent length coverage by reads (>=10%), average read mapping quality (>=11) and probe mapping (except for shotgun libraries) filters.

Samples	Total genes	Genes found in all	Genes found in 2/3 or more	Genes found in 1/3 or more	Overlap in All Samples (%)
Set 1 Enriched	62	24	38	53	38.71
Set 2 Enriched	68	50	57	64	73.53
Set 3 Enriched	70	41	53	60	58.57
All Enriched	70	24	52	60	34.28
All Shotgun	32	16	18	28	50.00

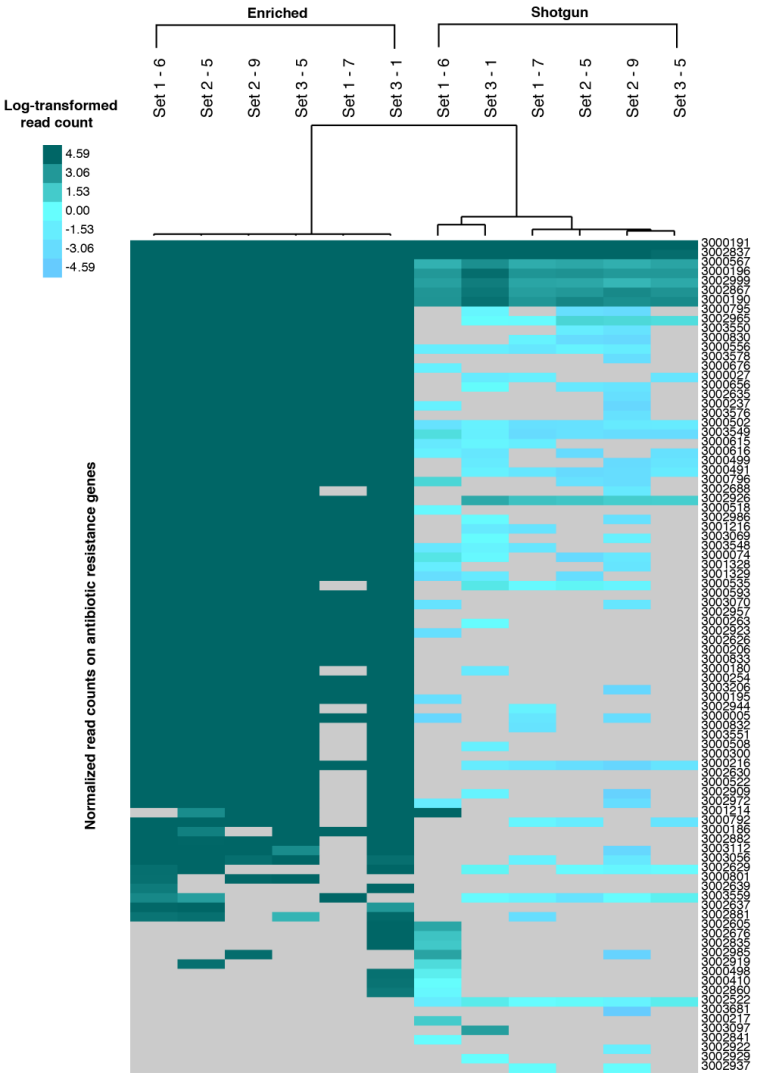
Supplementary Table 10: Phase 2 overlapping genes with subsampled reads

Libraries were subsampled to the same number of reads within sets and overall (22,324 reads). Shotgun libraries were subsampled to the same number of reads as the lowest enriched library overall. Resulting genes with at least 10 reads were filtered for percent coverage by reads (>=10%), average mapping quality (>=11) and probe mapping (except for the shotgun samples).

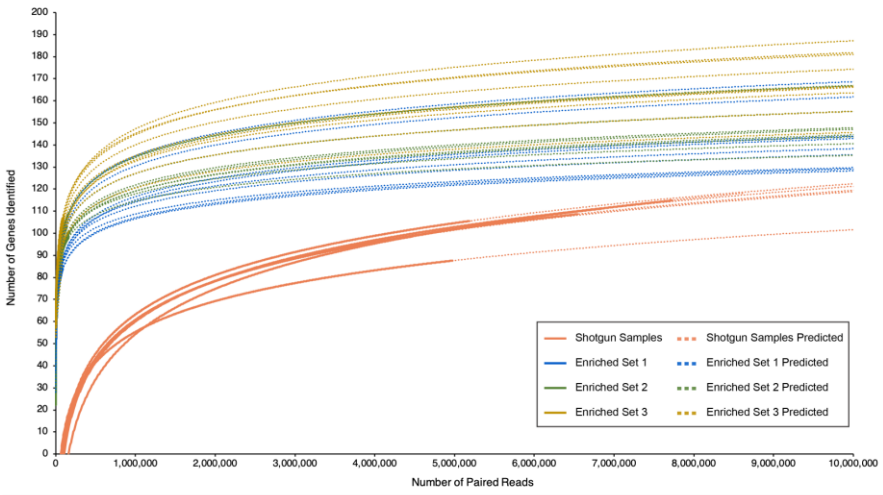
Samples	Total genes	Genes found in all	Genes found in 2/3 or more	Genes found in 1/3 or more	Overlap in All Samples (%)
Set 1 Enriched	38	16	26	32	42.10
Set 2 Enriched	45	22	30	36	48.89
Set 3 Enriched	37	13	20	26	35.14
All Enriched	47	13	24	31	27.66
All Shotgun	2	0	1	2	0



Supplementary Figure 7: Hierarchical clustering of enriched libraries
Enriched reads from 27 libraries were subsampled to 22,324 reads, mapped to CARD through rgi bwt. The reads were mapped to CARD through rgi bwt and filtered for genes with probes mapping, with greater than or equal to 10% length coverage by reads and an average read mapping quality >=11. Read counts were log-transformed and combined into a heatmap ordered by average read counts across the 27 enriched samples.



Supplementary Figure 8: Hierarchical clustering of enriched and shotgun libraries
The full number of reads from the 6 enriched and shotgun pairs were mapped to CARD through rgi bwt. The results were filtered for genes with greater than or equal to 10% read length coverage and an average read mapping quality greater >= 11. Read counts were normalized by kb of gene and reads available for mapping, log-transformed, and combined into a heatmap. Genes are ordered by sum of read counts. ARO numbers from CARD are shown on the right-hand side of the heatmap.



Supplementary Figure 9: Rarefaction curves for identification of antibiotic resistance genes
The AmrPlusPlus Rarefaction Analyzer was used with subsampling every 1% of the total reads and a gene read length of at least 10% to identify antibiotic resistance genes. The solid lines show individual sequencing experiments and the dotted lines are the logarithmic extrapolations beyond the experimental sequencing depth.

Supplementary Table 11: Sequencing reads identified in the Blank samples
Enriched negative control blank libraries were sequenced on separate MiSeq 2 x 250 runs. After de-multiplexing, we pulled the reads that were associated with various index combinations used alongside the Blank Negative control throughout library preparation within the same trials and sets.

Sample	Samples processed alongside the blank library	Number of paired reads sequenced on run with Blank	Percentage of Blank
Blank Trial 1	C0002	1575	0.92
	C0018	0	0.00
	C0050	435	0.26
	C0060	379	0.22
	Pool1	3064	1.80
	Pool2	110959	65.05
	Pool3	36390	21.33
	Additional barcodes	2487	1.46
	Blank	15276	8.96
Blank Trial 2	C0002	6611	15.02
	C0018	11763	26.72
	C0050	5194	11.80
	C0060	4491	10.20
	Pool1	1178	2.68
	Pool2	4800	10.90
	Pool3	5862	13.31
	Additional barcodes	3044	6.91
	Blank	1083	2.46
Blank Set 1	1 - 1	456	17.61
	1 - 2	94	3.63
	1 - 3	174	6.72
	1 - 4	101	3.90
	1 – 5	316	12.20
	1 - 6	82	3.17
	1 - 7	683	26.37
	1 - 8	173	6.68
	1 - 9	35	1.35
	Negative Control - Blank	28	1.08
	C0002 – 1 - 1	120	4.63
	C0002 – 1 - 2	37	1.43
	C0002 – 1 - 3	291	11.24
Blank Set 2	2 - 1	367	9.65
	2 - 2	22	0.58
	2 - 3	44	1.16
	2 - 4	119	3.13
	2 - 5	40	1.05
	2 - 6	0	0.00
	2 - 7	39	1.03
	2 - 8	271	7.12
	2 - 9	137	3.60
	Negative Control - Blank	530	13.93
	C0002 – 2 -1	207	5.44
	C0002 – 2 - 2	34	0.89
	C0002 – 2 - 3	1994	52.42
Blank Set 3	3 - 1	224	3.76
	3 - 2	286	4.80
	3 - 3	71	1.19
	3 - 4	1653	27.73
	3 - 5	282	4.73
	3 - 6	23	0.39
	3 - 7	42	0.70
	3 - 8	128	2.15
	3 - 9	1198	20.09
	Negative Control - Blank	0	0.00
	C0002 – 3 -1	161	2.70
	C0002 – 3 - 2	817	13.70
	C0002 – 3 - 3	1077	18.06

Supplementary Table 12: Negative control enrichment with Blank samples
Enriched reds were divided among index combinations used during the respective Phase, Trial or Set (Supplementary Table 7). The reads belonging to each Negative Control – Blank library were trimmed and duplicates were removed then mapped to CARD through rgibwt. The number of genes with 1, at least 10 and at least 100 reads as well as genes with probes mapping, with average read mapping quality >=11 and gene length coverage with reads >=10% are shown. In Phase 2 Set 1, raw sequencing reads were used for analysis, in Set 2, deduplication was omitted, and for Set 3, there were no reads associated with the Blank indexes after sequencing.

Sample	Paired reads	Paired reads after trimming and de-duplication	Percent of reads mapping to CARD	Total genes with reads	Genes with 10 or more reads	Genes with 100 or more reads	Genes with at least 10 reads, >10% read coverage, MQ >=11 and probes
Blank Phase 1 Trial 1	15276	2716	80.34	153	82	9	10: *cpxA, mefA, arlS, mdtO, mdtE, mdtN, acrD, armA, AAC(3)-IV, APH(7'')-Ia,*
Blank Phase 1 Trial 2	1083	341	97.21	106	9	1	0
Phase 2 Set 1	28	N/A	0	0	0	0	0
Phase 2 Set 2*	530	412	76.46	94	26	0	**19**: *APH(3'')-Ib, acrD, acrE, acrF, acrS, cpxA, dfrA17, emrK, emrY, eptA, evgS, mdtE, mdtF, mdtH, mdtO, mdtP, pmrF, tetQ, tolC*
Phase 2 Set 3	0	0	0	0	0	0	0

*Genes in Set 2 Blank found in enriched and shotgun libraries: *tetQ, acrF*. Genes found in blank and enriched: *acrD, acrE*, *acrS*, *cpxA*, *emrK, emrY, eptA, evgS, mdtE, mdtF, mdtH, mdtO, mdtP*, *pmrF*, *tolC*. Genes found in blank only: *APH(3'')-Ib, dfrA17,*

Supplementary Table 13 (omitted from Thesis due to size)

CHAPTER THREE: Capturing the antibiotic resistome of preterm infants reveals new benefits of probiotic supplementation

CHAPTER THREE PREFACE

The work presented in this chapter has been published as:

Guitor AK, Yousuf EI, Raphenya AR, Hutton EK, Morrison KM, McArthur AG, Wright GD, Stearns JC. Capturing the antibiotic resistome of preterm infants reveals new benefits of probiotic supplementation. *Microbiome*. 2022 Aug 26;10, 136. doi: 10.1186/s40168-022-01327-7. PMID: 36008821; PMCID: PMC9414150.

The studies in which samples were collected were led by EKH and KMM; EIY oversaw participant recruitment and data and sample collection. EIY and JCS generated and analyzed the microbiome data. AKG planned the experiments, performed DNA library preparation, and targeted enrichment and preparation of samples for sequencing. ARR and AKG wrote the code. AGM provided advice on analyses and computational support. AKG performed the analyses and generated the tables and figures for the resistome capture data. AKG wrote the manuscript with primary input from EIY, JCS, and GDW. The authors read and approved the final manuscript.

ABSTRACT

Background

Probiotic use in preterm infants can mitigate the impact of antibiotic exposure and reduce rates of certain illnesses; however, the benefit on the gut resistome, the collection of antibiotic resistance genes, requires further investigation. We hypothesized that probiotic supplementation of early preterm infants (born < 32-week gestation) while in hospital reduces the prevalence of antibiotic resistance genes associated with pathogenic bacteria in the gut. We used a targeted capture approach to compare the resistome from stool samples collected at the term corrected age of 40 weeks for two groups of preterm infants (those that routinely received a multi-strain probiotic during hospitalization and those that did not) with samples from full-term infants at 10 days of age to identify if preterm birth or probiotic supplementation impacted the resistome. We also compared the two groups of preterm infants up to 5 months of age to identify persistent antibiotic resistance genes.

Results

At the term corrected age, or 10 days of age for the full-term infants, we found over 80 antibiotic resistance genes in the preterm infants that did not receive probiotics that were not identified in either the full-term or probiotic-supplemented preterm infants. More genes associated with antibiotic inactivation mechanisms were identified in preterm infants unexposed to probiotics at this collection time-point compared to the other infants. We further linked these genes to mobile genetic elements and *Enterobacteriaceae*, which were also abundant in their gut microbiomes. Various genes associated with aminoglycoside and

beta-lactam resistance, commonly found in pathogenic bacteria, were retained for up to 5 months in the preterm infants that did not receive probiotics.

Conclusions

This pilot survey of preterm infants shows that probiotics administered after preterm birth during hospitalization reduced the diversity and prevented persistence of antibiotic resistance genes in the gut microbiome. The benefits of probiotic use on the microbiome and the resistome should be further explored in larger groups of infants. Due to its high sensitivity and lower sequencing cost, our targeted capture approach can facilitate these surveys to further address the implications of resistance genes persisting into infancy without the need for large-scale metagenomic sequencing.

INTRODUCTION

Preterm infants (born < 37-week gestation) have an immature gut microbiome that is shaped by various factors, including the immaturity of the gastrointestinal tract at birth, maternal and postnatal antibiotic exposure, delivery mode, and feeding method [1,2,3,4,5,6,7,8]. Preterm infants often have reduced gut microbial diversity compared to full-term infants. Their microbiota can be dominated by a few potentially pathogenic bacteria, including *Enterococcus faecalis*, *Staphylococcus epidermidis*, *Klebsiella pneumoniae, Escherichia coli,* and *Pseudomonas aeruginosa* [3, 6, 9, 10]. The initial colonizers of the preterm infant gut have been linked to the environment, including the neonatal intensive care unit (NICU) [11, 12]. The potential exposure to multidrug-resistant

(MDR) strains, their persistence in the infant gut, and possible transfer of antibiotic resistance genes (ARGs) in this hospital niche are of concern [13,14].

Due to their overall immune immaturity, preterm infants are at high risk of infection and can be exposed to antibiotics for prolonged periods, usually as empiric treatment for suspected sepsis [15]. Among the most frequently used medications in the NICU are the antibiotics ampicillin, gentamicin, amikacin, and vancomycin [15,16]. The consequences of antibiotic exposure on preterm infants' gut microbiota and resistome, or ARG content, have been explored [8,17]. Exposure to various broad-spectrum antibiotics not only reduces the overall diversity of bacteria in the infant gut but also can select for pathogenic *Enterobacteriaceae* and reduce beneficial organisms, including *Bifidobacteriaceae* [18,19,20,21]. This change in microbial diversity is also related to a selection for MDR strains and promotes persistence of ARG-carrying bacteria over time [13,18,21,22,23].

The intestinal immaturity of preterm infants, the reduced microbial diversity in the gut, and frequent exposures to antibiotics increase the risk of necrotizing enterocolitis (NEC) for these infants [24,25,26]. NEC is a disease that affects around 7% of very low birth-weight infants, is associated with longer hospital stays, and has a high mortality rate (15–30%) [24,27]. Probiotics have been shown to reduce the incidence of NEC, sepsis, and mortality in preterm infants, in addition to reducing the impact of extensive antibiotic exposure on the gut microbiota [2, 24, 28,29,30,31,32,33,34,35,36,37].

A higher prevalence of *Bifidobacterium* species in the infant gut is associated with reduced colonization by taxa commonly associated with antimicrobial resistance (AMR)

such as *Enterobacteriaceae* [1, 38,39,40]. By association, we hypothesized that a probiotic supplement containing *Bifidobacterium* species will reduce the diversity of ARGs detected in the preterm infant gut. A few studies have focused on probiotics and the resistome of preterm infants and reported that *Bifidobacterium*-containing probiotics reduced the abundance of MDR bacteria and their associated ARGs [41,42,43]. These studies used either polymerase chain reaction (PCR)-based approaches [42] or shotgun metagenome sequencing [41, 43] to characterize the resistome. Because ARGs usually represent less than 1% of a metagenome, the latter approach of shotgun sequencing requires a high number of sequencing reads, which becomes expensive for longitudinal cohort studies [44]. PCR-based resistome analyses typically only target a few antibiotic resistance genes of interest [45]. Our previous work showed how a targeted capture probe set of over 37,000 nucleotide baits designed against 2000 antibiotic resistance genes is superior to shotgun sequencing for surveying the resistome [46]. We employed this probe set to profile ARGs of preterm infants supplemented with probiotics and highlight the reduction of these genes commonly associated with pathogenic bacteria.

A subset of preterm infants from the Baby and Preterm Microbiota of the Intestine Cohort Study (Baby & Pre-Mi) at McMaster University received a commercial probiotic supplement (FloraBABY, Renew Life Canada) that contained four species of *Bifidobacterium* and one *Lactobacillus* species [47]. We have previously shown that at term age, the gut microbiota of these preterm infants supplemented with this probiotic was more similar to that of 10-day-old healthy infants born full term (> 37-week gestation) [4, 5, 47].

In this study, we assessed the resistome of the gut microbiota in a subset of samples from preterm and full-term infants from the Baby & Pre-Mi and Baby & Mi studies [4, 5, 47], including stool samples collected during hospitalization and at follow-up visits up to 5-month corrected age from 8 preterm infants who were supplemented with the probiotic in hospital (PS), 13 preterm infants that were not supplemented with the probiotic during hospitalization (NS) [47], and stool samples from nine 10-day-old full-term (FT) infants that did not receive probiotics or antibiotics [4, 5]. DNA extracts were prepared for Illumina sequencing and enriched for ARGs using targeted capture [46]. After enrichment and sequencing, the Resistance Gene Identifier's (RGI) metagenomic feature was used to map reads to the Comprehensive Antibiotic Resistance Database (CARD) [48]. ARGs were compared across patient cohorts and study time-points to detect differences in the resistome after probiotic supplementation. We highlight differences in the diversity of ARGs rather than individual ARG abundances, given that even rare ARGs in the microbiome can provide selection for antibiotic-resistant bacteria and the future implications of ARG persistence on preterm infant health are unknown.

MATERIALS AND METHODS

Study participants and sampling

A detailed description of the study participants and design can be found in Yousuf *et al.* [47] and Stearns *et al.* [5]. Four stool samples were chosen for preterm infants in our study, based on time-point sample availability and distribution to provide a longitudinal survey spanning their time spent in hospital up to 5 months of age. The last sample collected

in hospital before the infant was discharged or they reached their expected due date (i.e., term age) is referred to as the in-hospital time-point. The other samples included are the first study visit (visit 1) that took place as close to term age as possible and subsequent samples collected at around 6-week (visit 2), 12-week (visit 3), and 5-month (visit 4) corrected age (with corrected age referring to age of the infant from the expected due date). Postmenstrual age (PMA) in weeks at the time of sample collection was calculated as the sum of gestational age at birth (based on the expected due date and infant birth date) and postnatal age.

Part way through the Baby & Pre-Mi study (November 2017), the McMaster Children's Hospital NICU changed their policy such that the probiotic FloraBABY (Renew Life Canada, Brampton, ON, Canada) was routinely given to infants born at less than 34-week gestation or weighing less than 2 kg. This probiotic contains 0.5 g (2 billion CFU bacteria) per single-dose sachet, including the following: *Bifidobacterium breve* (HA-129), *Lactobacillus rhamnosus* (HA-111), *Bifidobacterium bifidum* (HA-132), *Bifidobacterium longum* subsp. *infantis* (HA-116), and *Bifidobacterium longum* subsp. *longum* (HA-135). For our resistome analysis, infants born early preterm (< 32-week gestation) that were admitted to the NICU and had samples available in hospital and at around term age were studied. This includes 8 probiotic-supplemented (PS) preterm infants born at an average gestational age of 28.14 weeks and 13 not supplemented (NS) preterm infants born at an average of 27.49 weeks. The PS infants were exposed to the probiotic FloraBABY for an average of 8.27 weeks (Table 1). One PS infant (PS4) continued supplementation with the

probiotic BioGaia, which contains *Limosilactobacillus reuteri* strain DSM 17938, throughout visit 1 and visit 2 after stopping FloraBABY administration post-discharge from hospital. Two NS infants (NS3 and NS4) received the probiotic BioGaia between visit 3 and visit 4 for an unknown duration. As a comparator to the term age (visit 1) sample collected from preterm infants, a 10-day stool samples from 9 full-term (FT) infants from the Baby & Mi study were included (Table S1, Additional file 1) [5]. These full-term infants had not received probiotics or antibiotics prior to stool collection. From our chosen set of preterm infants (n = 21), a subset of samples from PS (n = 6) and NS (n = 6) infants were matched for antibiotic exposure and sample availability (Table S2, Figs. S2, S3, Additional file 1). We present the results from this subset alongside the results from the entire cohort to determine if the results were replicated.

Table 1: Characteristics of infant cohorts and samples used in this study

	NS preterm (*n* = 13)	**PS preterm (*n* = 8)**	***p*-value**
Gestational age at birth, weeks	27.49 ± 2.03	28.14 ± 1.54	0.47
Probiotic exposure, weeks	0.00	8.27 ± 3.19	< 0.0001
Antibiotic exposure during sample collection (types and number of infants exposed)	Amo(1), Amp(13), Az(2), Cefa(3), Cefo(5), Cefu(1), Cl(5), G(13), Mer(2), Met(1), T(1), V(5)	Amp(6), Cefa(1), Cefo(2), Cl(3), G(6), Met(2), V(1)	N/A
Antibiotic exposure, weeks	1.98 ± 1.83	1.11 ± 1.20	0.20
In-hospital sample, weeks in PMA (*N*)	37.20 ± 3.80 (12)	37.86 ± 1.69 (3)	0.66

	NS preterm (*n* = 13)	**PS preterm (*n* = 8)**	***p*-value**
Visit 1 sample, weeks in PMA (*N*)	41.96 ± 2.25 (10)	42.63 ± 1.69 (8)	0.52
Visit 2 sample, weeks in PMA (*N*)	46.83 ± 1.75 (9)	46.43 ± 0.50 (5)	0.90
Visit 3 sample, weeks in PMA (*N*)	52.29 ± 2.45 (7)	54.21 ± 1.85 (6)	0.18
Visit 4 sample, weeks in PMA (*N*)	62.46 ± 2.56 (10)	59.68 ± 0.62 (4)	0.07

PMA is the postmenstrual age in weeks, and SD is the standard deviation. The data are presented as mean ± SD. *P*-values < 0.05 using Student's t-test or Mann-Whitney were considered to be statistically significant.

16S rRNA gene profile analysis

The amplicon sequence variant (ASV) table from [47, 49] representing bacterial 16S rRNA gene v3 region amplicons was analyzed by plotting the relative abundance of ASVs for each full-term infant at 10 days and the subset of preterm infants included in our study at all available collection time-points using R. ASVs were grouped together as one category (< 1% abundance) if they represented less than 1% of the total relative abundance across all infants.

DNA library preparation, enrichment, and sequencing

DNA extracted from stool for the previous study [5, 47, 50] was used here. Library preparation and enrichment for ARGs were performed as described in Guitor *et al.* [46]. When available, up to 500 ng of dsDNA was used for library preparation with the NEBNext Ultra II dsDNA library kits (Additional files 2, 3). After library preparation, a High

Sensitivity DNA ScreenTape Analysis (Agilent Technologies) was performed to estimate the DNA available for enrichment. For most samples, at least 100 ng of DNA was available for enrichment (Additional file 2). All samples were enriched for 24 h at 65 °C using a probe set of 37,826 probes designed to target over 2000 ARGs. After 3 rounds of washing of the streptavidin beads, 12.5 μL of captured DNA was amplified by PCR for 14 cycles, purified using KAPA Pure Beads, and then eluted in 30 μL of 10 mM Tris HCl, pH 8.3. Enriched libraries were quantified by quantitative PCR (qPCR) and then pooled in equimolar ratios. Sequencing was performed by the Farncombe Metagenomics sequencing facility at the McMaster University on an Illumina HiSeq or MiSeq with 2 × 250 bp sequencing chemistry with a targeted depth of 250,000 clusters per library.

Analysis of targeted capture sequencing data

Paired sequencing reads were trimmed using *skewer* version 0.2.2 [51, 52] and deduplicated using *dedupe*.sh from BBMap version 38.57 [53]. Reads were then subsampled to 50,000 pairs, or 100,000 paired reads total, using the sample command from *seqtk* version 1.3 [54]. Using the beta read mapping to CARD (RGI *bwt*) feature of RGI version 5.1.1 [55], reads were mapped to a combined reference of 179,050 nucleotide sequences from CARD (nucleotide sequences, protein homolog model, version 3.1.0) and the Resistomes & Variants database version 3.0.7 using *bowtie2* version 2.3.5.1 [56, 57]. Both databases are available for download [58]. We additionally used RGI's beta feature for the k-mer prediction of pathogen of origin for AMR genes or reads (RGI *kmer_query*) with the default 61-mer database, to predict bacterial species that may harbor an ARG.

Lastly, using in-house scripts, we generated *de novo* assemblies of the enriched metagenomes using the *metaSPAdes* option in SPAdes v. 3.13.1 [59, 60]. Resistance genes were predicted from these assemblies using the *main* feature of RGI version 5.1.1 and CARD (version 3.1.0–2702 nucleotide sequences). Detailed code is available at https://github.com/AllisonGuitor/AMR-metatools. An important distinction to note is that RGI's *bwt* read-mapping algorithm in version 5.1.1 is unable to detect resistance conferred by point mutation in chromosomally encoded ARGs, whereas RGI *main* can predict resistance via this mechanism.

The RGI *bwt* results for each infant were filtered for genes with at least 100 mapped reads. If a *de novo* assembly was successfully generated, all perfect and strict predicted genes from the RGI analysis were included in downstream analyses. These results were then combined across all infants. Significant differences between infant groups were determined by unpaired t-test in GraphPad Prism version 9.0.1, with a p-value cutoff of below 0.05. To assess whether the types of resistance genes differ between infant groups, we further categorized the genes identified through RGI *bwt* based on their AMR resistance mechanisms and gene families in CARD. Unique genes identified in each infant cohort in the RGI *bwt* analysis and the *de novo* assembly + RGI *main* analysis were compared. Venn diagrams were generated using BioVenn [61]. To compare ARGs that we differentially detected in one group of infants compared to the others, we excluded AMR gene families that were present in all cohorts at the visit 1 time-point (term age for preterm and 10 days of age for full-term infants). For the longitudinal analysis, we removed AMR gene families

that were present in 4/5 time-points in both preterm infant cohorts, as these likely represent the general resistome of preterm infants.

Mobile genetic element detection and bacterial host identification

Given the nature of our probe set design and method, the genomic context surrounding targeted resistance genes can also be captured. This analysis relies on assembling contigs from the targeted capture data and successfully predicting open reading frames (ORFs). Therefore, in some cases, a particular gene might not be detected. In many cases, large enough contigs were obtained to predict an ARG through RGI *main* and to annotate neighboring genes using *Prokka* version 1.14.5 [62]. Only contigs greater than 1.2 kb with perfect or strict hits from CARD's protein homolog model were considered. Contigs greater than 1.2 kb containing ARGs were analyzed using *mob_recon* from MOB-suite v3.0.1 [63, 64] to predict potential plasmid sequences and mobile genetic elements (MGEs). In addition to these potentially mobile ARGs, we selected contigs with AMR gene families that were unique to one infant cohort when comparing preterm infants at term to FT infants at 10 days of age or to one infant cohort when comparing preterm infants up to 5 months of age. This included the following: *aac(3)*, *ant(2'')*, *ant(4')*, *aph(3')*, *arr-3*, *bla*$_{\text{ACT}}$, *bla*$_{\text{CTX-M}}$, *bla*$_{\text{CblA}}$, *bla*$_{\text{DHA}}$, *bla*$_{\text{LEN}}$, *bla*$_{\text{MIR}}$, *bla*$_{\text{MOX}}$, *bla*$_{\text{OXA}}$ *kdpDE*, streptothricin acetyltransferase, tetracycline inactivation enzymes, vancomycin resistance genes, and streptogramin vat acetyltransferase. We retrieved the most similar nucleotide hit of these contigs using Nucleotide BLAST (blastN) [65] with the nonredundant nucleotide collection in NCBI [66]. We also recorded the pathogen-of-origin prediction by RGI *kmer_query*, and

results from read mapping to CARD's Resistomes & Variants database which provides potential hosts for each ARG based on reported sequencing data in NCBI [67]. From the blastN, RGI *kmer_query*, CARD's Resistomes & Variants database, and the *mob_recon* analysis, a consensus prediction of a bacterial host for each ARG was inferred if at least 2 of the results agreed. Annotated contigs containing unique ARGs of interest were compared using *clinker* version 0.0.21 [68].

Negative controls sequencing

We included 12 negative controls throughout our study to account for potential contamination from reagents and laboratory environment. The libraries were analyzed via a High Sensitivity DNA ScreenTape Analysis (Agilent Technologies) and quantified by qPCR as described previously [46]. Only 3 of the negative controls displayed signatures of DNA sequencing libraries, and one had sufficient concentration for sequencing (Additional file 5). This sample was sequenced on an Illumina MiSeq run (2 × 250 bp chemistry) separately from other libraries generated in this study. The unique index combination corresponding to the negative control library did not generate any read data.

RESULTS

Exposure of preterm infants to various antibiotics and probiotics in early life

All preterm infants included in this study were treated with ampicillin and gentamicin, except two PS infants that did not receive antibiotics during sample collection. Many received up to five different antibiotics within the first 10 weeks of life (Fig. S1, Additional file 1). On average, NS infants received more prolonged doses of antibiotics

(1.98 weeks compared to 1.11 weeks for PS infants; however, the difference was not profound), with one NS infant receiving a total of 5.86 weeks of antibiotic dosing between birth and 5-month corrected age (Table 1). No PS infant received antibiotics after beginning probiotic administration during the study period. A timeline of sample collection and probiotic administration is shown in Fig. 1. Three of the preterm infants were still receiving probiotics during the visit 1 or term age time-point, and PS4 continued to receive BioGaia at the visit 1 and visit 2 time-points. In previous work, the impact of probiotics on the gut microbiota of this cohort of infants is described in detail [47]. Our study found that bacteria belonging to the family *Bifidobacteriaceae* are present at higher relative abundances at earlier collection time-points in PS infants than in NS infants, and that *Enterobacteriaceae* and *Clostridiaceae* dominate the gut microbiota of NS infants (Fig. S4, Additional file 1).

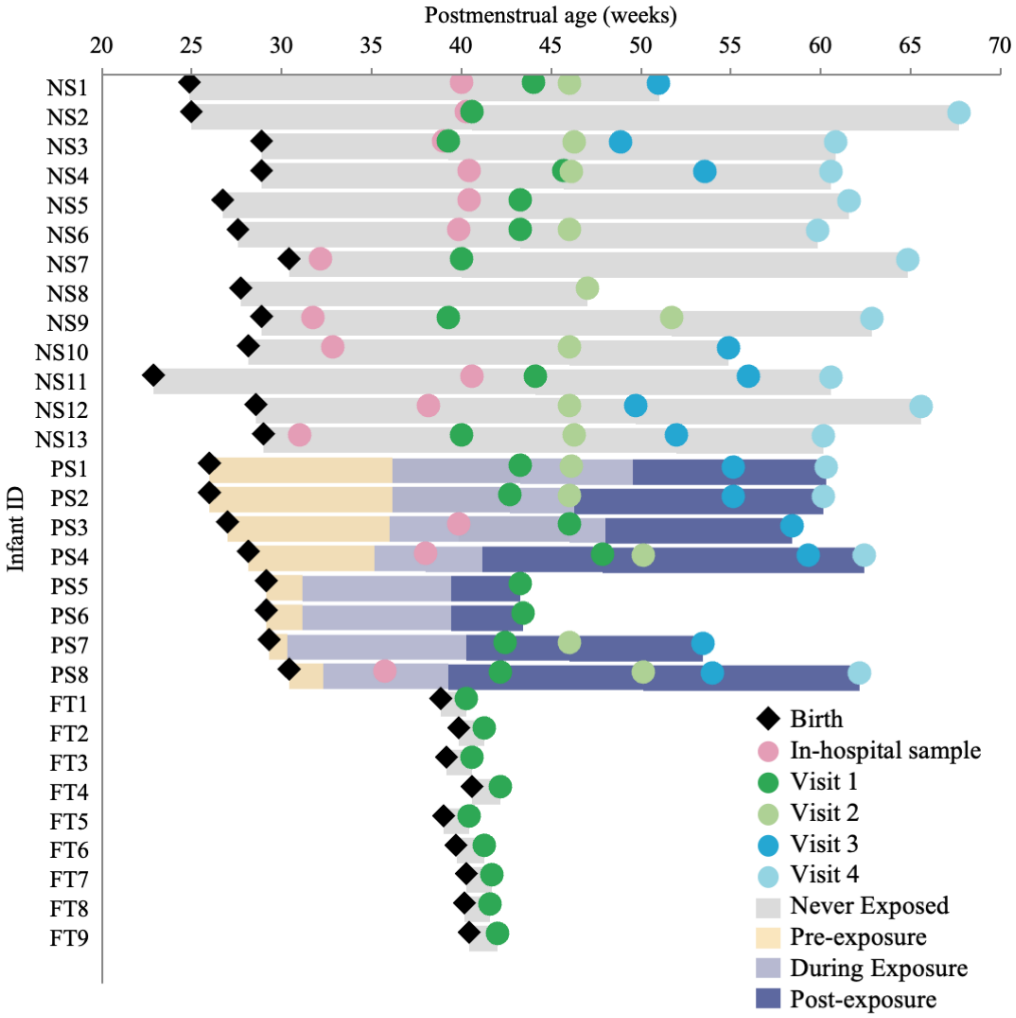


Figure 1: Sample collection and probiotic exposure of preterm infants.
Timelines from birth to final sample collection for all infants are included in this study. The duration of exposure to probiotics (lavender bar) and timing of sample collection in relation to postmenstrual age in weeks are shown for non-probiotic-supplemented (NS), probiotic-supplemented (PS) preterm infants, and full-term (FT) infants.

Similar number of antibiotic resistance genes in preterm and full-term infants at an early age

We first wanted to compare the resistome of preterm infants to full-term infants at an early age. We found that neither preterm birth nor probiotic exposure in preterm infants resulted in significant differences in the number of ARGs recovered from these infant gut microbiomes at the term age as compared to full-term infants at 10 days of age (Fig. S5 B–C, Additional file 1). Given the variability in the number of samples between the NS, PS, and FT infants, and differences in antibiotic treatments of the preterm infants, we reanalyzed our results in a subset of preterm infant samples matched for time-point and antibiotic administration along with the FT infants. Again, we found no significant differences in the number of ARGs between the infant groups (Fig. S6 B–C, Additional file 1). To ensure our results were not biased, we compared the percentages of sequencing reads that map to ARGs in CARD between infant groups. We found a significantly higher number of reads mapping in the FT compared with the NS infants at visit 1 (10 days/term age) ($P = 0.0014$; $P = 0.0318$, Figs. S5A and S6A, Additional file 1). The higher percentage of mapped reads to CARD did not correspond to increased numbers of ARGs in the full-term infants. Therefore, percentage of mapped reads to CARD cannot be used as a measure of ARG load after enrichment as it is in shotgun sequencing but is instead used as a measure to determine whether enrichment was successful [41].

Preterm infants not supplemented with probiotics have a greater diversity of antibiotic resistance genes

We next compared the types of ARGs found in each infant group at an early age (visit 1). In both the read-mapping RGI *bwt* and *de novo* assembly with RGI *main* analysis

approaches, over 200 ARGs were identified (226 for RGI *bwt* and 243 for RGI *main*) in all infants (Figs. 2A, S7A, Additional file 1). Many of these ARGs (81 for RGI *bwt* and 94 for RGI *main*) were only ever identified in the NS infants (Figs. 2A, S7A, Additional file 1). We then looked at the number of ARGs at the individual level that was unique to their given infant group and found significantly fewer unique ARGs in each PS infant compared to the NS and FT infants ($P = 0.0047$ for NS vs PS, $P = 0.0262$ for PS vs FT; Figs. 2B, S7B, Additional file 1). Many of the NS infants had more than 10 ARGs that were not identified in the two other infant groups (Figs. 2B, S7B, Additional file 1). We found that probiotic exposure in preterm infants resulted in a reduced number of unique ARGs in the infant gut microbiome as compared to other preterm infants at the term age and full-term infants at 10 days of age.

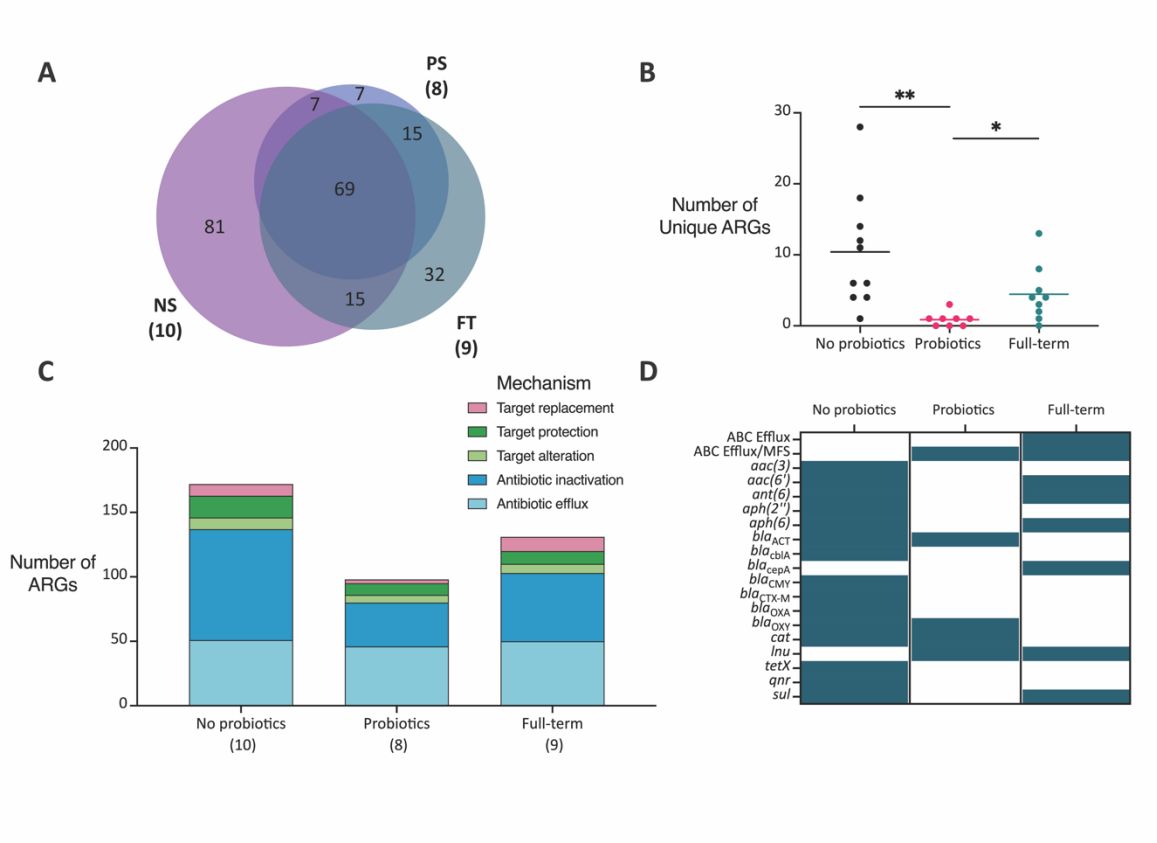


Figure 2: Differences in the resistome identified through RGI *bwt* in infants at visit 1. Reads were mapped to CARD using *bowtie2*, and antibiotic resistance genes with at least 100 reads were reported. The data presented is from the full set of preterm and full-term infants and at visit 1. A) Unique and overlapping ARGs identified in each infant group. The number of infant samples included in each is shown next to the sample type. B) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared (P = 0.0047 for NS vs PS, P = 0.0262 for PS vs FT). C) A breakdown of the mechanisms of antibiotic resistance identified in each infant group. The number of infant samples included in each is shown next to the sample type. D) The presence or absence of selected AMR gene families in each infant group. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples (NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants).

When classified under their respective AMR mechanism, we identified more antibiotic inactivation genes in the NS infant group compared to the other infants (Figs. 2C,

S7C, Additional file 1). Finally, we refined our results to the AMR gene family level and determined the presence of AMR gene families in the infant groups (Figs. S8A, S9A, Additional file 1). Various AMR gene families were found uniquely in the NS infants including aminoglycoside-modifying enzymes (*aac(3)*, *aph(2")*) and beta-lactamases ($bla_{\text{CTX-M}}$, bla_{CMY}, bla_{OXA}) (Figs. 2D, S7D, Additional file 1). Again, when looking at the subset of preterm infant samples matched for time-point and antibiotic exposure along with the full-term infants, we found similar results to those above (Figs. S8B, S9B, S10, S11, Additional file 1). We found that probiotic exposure in preterm infants resulted in a reduced diversity of ARGs in the infant gut microbiome as compared to other preterm infants at the term age and full-term infants at 10 days of age.

Probiotics reduce the diversity of the preterm gut resistome up to 5 months of age

Next, we sought to compare the impact of probiotics on the resistome of preterm infants up to 5 months of age. Apart from the differences at visit 1 noted above, we found no significant differences in the percentage of reads mapping on target or the number of ARGs detected at any time-point between the NS and PS infants (Figs. S5DEF, S6DEF, Additional file 1). We did, however, identify more ARGs in the NS infants as a group compared to the PS infants at all study time-points up to 5 months of age (Figs. 3, S12, Additional file 1). These results were recapitulated in the matched subset of infants except at the 12-week corrected age (visit 3) where ~50% of the unique genes (21/44) were identified in one PS infant (results not included) highlighting the potential for large individual variation in the resistome among infants (Figs. S13, S14, Additional file 1).

Again, we looked at the number of ARGs at the individual level that were unique to their given infant group and, in addition to the results previously described for visit 1, found significantly fewer unique ARGs in each PS infant compared to the NS infants at visit 2 ($P = 0.0180$) and visit 4 ($P = 0.0144$, $P = 0.0105$) (Figs. 4A, S15A, Additional file 1). When we looked at the matched subset of preterm infants, however, we only found significant differences in the unique number of ARGs at the visit 1 time-point (Figs. S15B, S16A, Additional file 1).

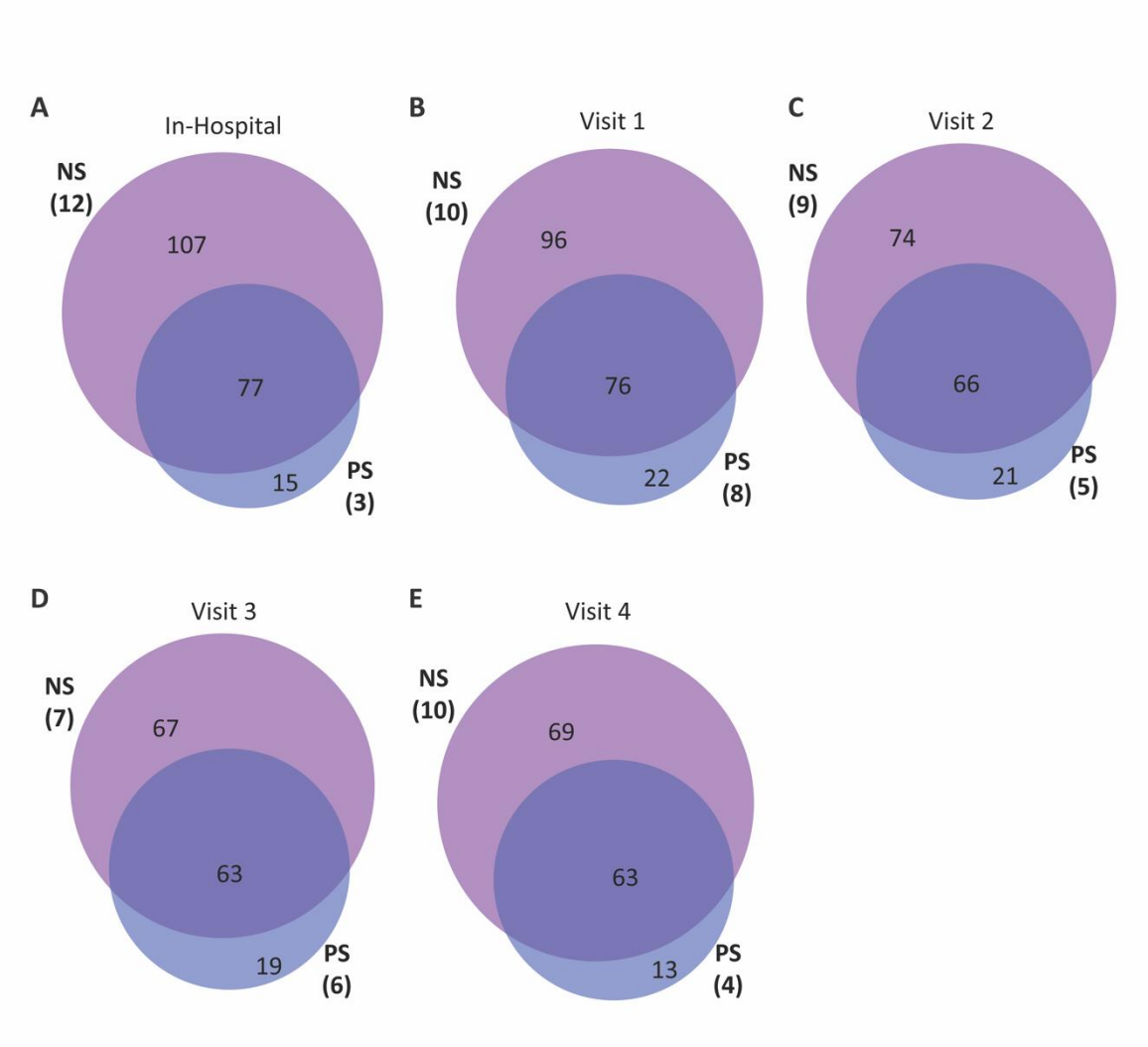


Figure 3: Number of unique genes in preterm infants at various time-points.
These gene counts are from mapping reads to CARD using *bowtie2* and counting the number of genes with at least 100 reads. Data are from NS and PS infants at the in-hospital collection (A), visit 1 (B), visit 2 (C), visit 3 (D), and visit 4 (E) time-points. The number of infants included in each time-point is indicated (NS = non-probiotic-supplemented preterm, PS = probiotic-supplemented preterm).

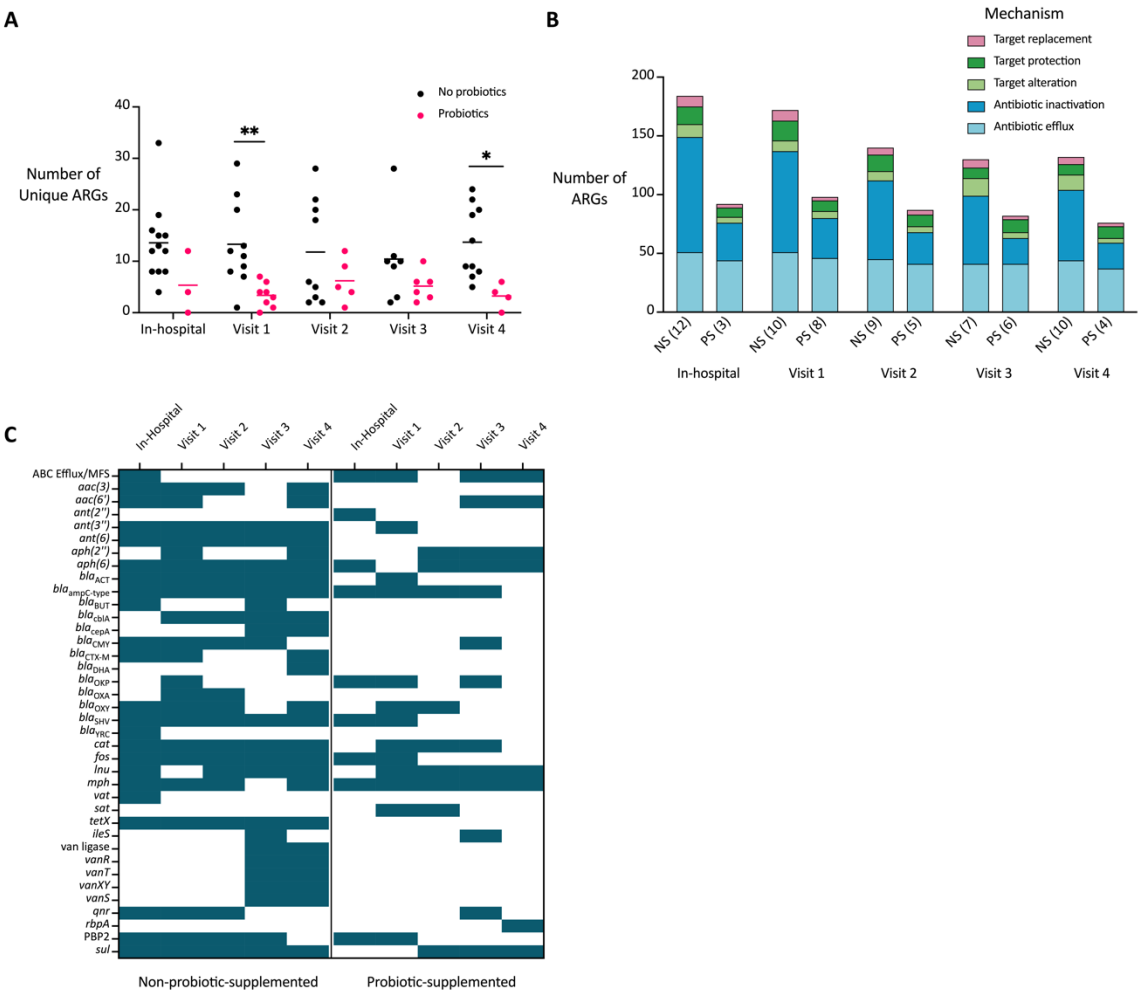


Figure 4: Unique ARGs, mechanisms, and families in preterm infants up to 5 months of age.

Reads were mapped to CARD using bowtie2, and ARGs with at least 100 reads were reported. The data presented is for all preterm infants at all visits. A) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared (P = 0.0052 for visit 1, P = 0.0144 for visit 4). B) The number of ARGs identified in each infant group classified by resistance gene mechanism. The number of infant samples included in each is shown next to the sample type (NS = non-probiotic-supplemented preterm, PS = probiotic-supplemented preterm). C) A selected subset of detected AMR gene families in preterm infants. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infants at that time-point.

Antibiotic inactivating genes are more prevalent in preterm infants not supplemented with probiotics

Overall, we identified a reduced diversity of ARGs in both preterm infant groups at visit 4 (5-month corrected age) compared to in early life (Figs. 3, 4B, S15C, Additional file 1). In the NS group of infants, we found more ARGs classified as antibiotic inactivating genes at all time-points compared to their PS counterparts (Figs. 4B, S15C, Additional file 1). This result was corroborated in our matched subset of preterm infants (Figs. S15D, S16B, Additional file 1). Therefore, probiotic supplementation reduced the diversity of antibiotic inactivation genes detected in the preterm infant gut resistome up to 5 months of age.

Preterm infants not supplemented with probiotics retain certain AMR gene families

We next looked at the diversity of AMR gene families in the preterm infants up to 5 months of age (Figs. S17, S18, Additional file 1). We found certain AMR genes that were unique to the NS infants across many time-points and never identified in the PS infants. These genes belong to the following AMR gene families: *aac(3)*, *bla*$_{\text{CblA}}$, *bla*$_{\text{CTX-M}}$, *bla*$_{\text{OXA}}$, streptogramin vat acetyltransferase, tetracycline inactivation enzymes (*tetX* and *tet(X4)*), and various vancomycin resistance genes (Figs. 4C, S15C, Additional file 1). Also, there were certain AMR gene families that, although they were identified at early time-points in both preterm groups, appeared to persist longer in the NS infants than in the PS infants. These families included the aminoglycoside resistance families *ant(3")* and *ant(6)*, the SHV-type beta-lactamases, and the fosfomycin thiol transferases (Figs. 4C, S15E). We

found similar trends in the matched subset of infants and compared the similarities across both analysis approaches (Figs. S15F, S16C, S17B, S18B, Additional file 1). These results suggests that these preterm infants are exposed to similar ARGs at an early age, but that probiotic supplementation prevents prolonged retention of these ARGs in the gut resistome.

Genetic context of antibiotic resistance genes was retained in the infant gut

A unique feature of targeted capture is the ability to increase not only the depth of sequencing coverage of ARGs but also the surrounding genes. Using the k-mer-based pathogen-of-origin prediction feature of RGI (RGI *kmer_query*) and MOB-suite's *mob_recon* algorithm, we identified the genetic context of specific ARGs, predicted potential bacterial hosts of ARGs, and identified MGEs carrying these ARGs in the infant gut microbiome (Figs. 5, 6, S19–S24, Additional files 1 and 4). In many cases for the NS infants, the genetic context of ARGs was conserved within individuals over the various study time-points indicating the persistence of the same host organism or mobile genetic element housing the ARGs. This was also found between individual infants. These cases included *aac(3)-IId* in NS5 and NS12 over multiple time-points, *bla*$_{\text{CTX-M-14}}$ in NS5 from in-hospital up to 5 months of age, and various *bla*$_{\text{SHV}}$ genes that were detected from in-hospital up to 3 and 5 months of age only in NS infants (Figs. 5AB, 6). With both *aac(3)-IId* and *bla*$_{\text{CTX-M-14}}$, the genes were near IS4, IS6, or IS1 family transposases, highlighting the potential mobility of these genes. The contig containing *aac(3)-IId* in infant NS5 at visit 1 is similar to a plasmid identified in *K. pneumoniae* and other *Gammaproteobacteria* based on the MOB-suite analysis (Additional file 4). This contig also contained the ARGs *dfrA17*,

aadA5, and *sul1*. In other NS and PS infants, these additional genes were found in different genomic contexts and potential plasmids identified in *Gammaproteobacteria*, *Enterobacteriaceae*, and *Enterobacterales*, highlighting the mobility of these ARGs (Fig. S21, Additional files 1 and 4). The SHV beta-lactamases likely originated from *Klebsiella* spp. based on the RGI *kmer_query* results but may be associated with plasmids with a broad host range in *Gammaproteobacteria* and *Enterobacterales* (Additional file 4).

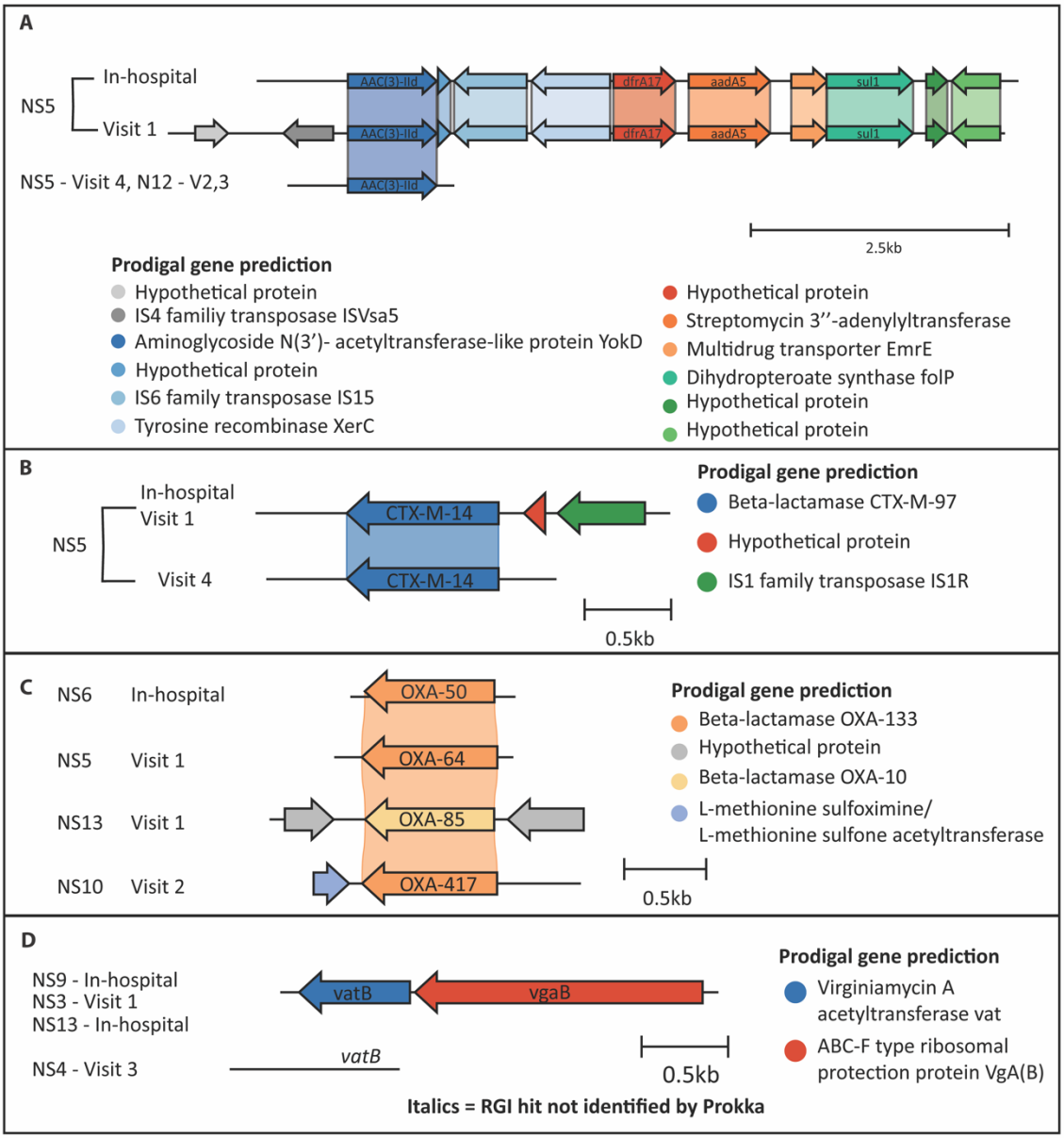


Figure 5: Genetic context of AMR gene families unique to NS infants.
From the de novo assembly, open reading frames were annotated using Prokka, and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows, and the RGI main predictions are labeled on each ORF. The genes are shown grouped into their respective AMR gene families: (A) AAC(3) gene family, (B) CTX-M beta-lactamase family, (C) OXA beta-lactamase family, (D) streptogramin vat acetyltransferase family (NS = not supplemented preterm).

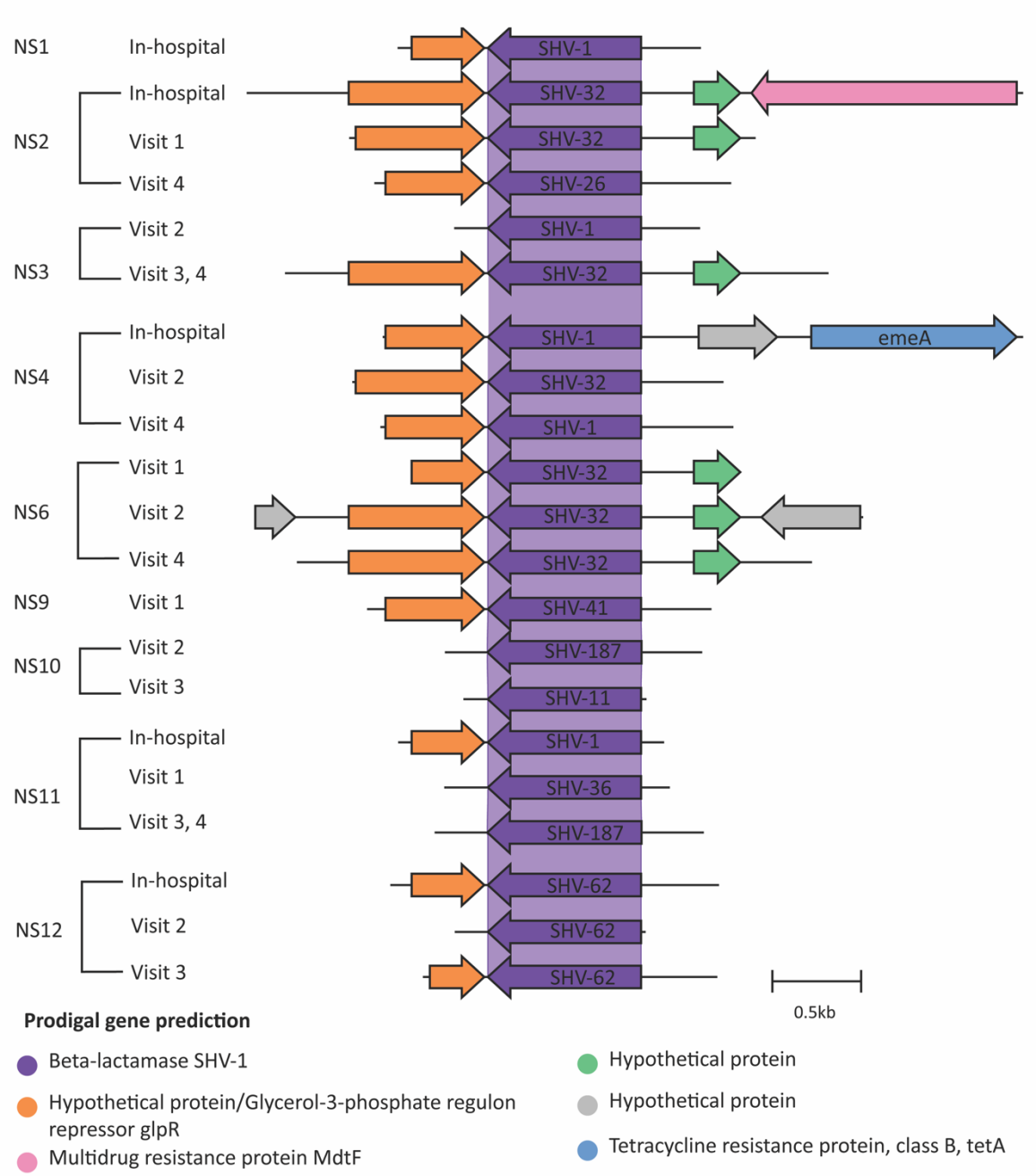


Figure 6: Genetic context of the SHV beta-lactamases.
From the *de novo* assembly, open reading frames were annotated using Prokka, and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows, and the RGI main predictions are labeled on each ORF (NS = not supplemented preterm).

Other clusters of AMR genes were more prevalent in NS infants compared to the PS and FT infants. Vancomycin resistance genes possibly originating from *Enterococcus* spp. were prominent at later time-points in various NS infants and were not identified on plasmids (Fig. S19, Additional files 1 and 4). Tetracycline inactivation enzymes (i.e., *tetX*) flanked by the rRNA methyltransferase *ermD* were detected in NS infants at various time-points, likely originated from *Bacteroides fragilis*, and were not associated with any known plasmids (Fig. S20B, Additional files 1 and 4). CblA beta-lactamases were most similar to sequences from uncultured bacteria or *Bacteroides uniformis* in NCBI, and contigs containing this ARG did not show similarity to plasmids in MOB suite (Fig. S20A, Additional files 1 and 4). Finally, the ANT(6) AMR family, consisting of *ant(6)-Ia, aad(6),* and *aadS*, and the combination of the streptogramin vat acetyltransferase, *vatB* and *vgaB,* were likely found in Gram-positive organisms such as *Staphylococcus aureus* and *Enterococcus* spp. and have been associated with plasmids in these genera (Figs. S22, 5D, Additional files 1 and 4).

For the few genes that were more prominent in the PS infants, the OKP beta-lactamases likely originated from *Klebsiella* spp., based on the RGI *kmer_query* results. Based on the MOB-suite analysis, these genes are likely associated with plasmids with a broad host range in *Gammaproteobacteria*, *Enterobacteriaceae*, and *Enterobacterales* (Fig. S23, Additional files 1 and 4). There are two genes identified from the APH(2″) family: *aph(2")-IIa* and *aph(2")-Iva* (Fig. S24, Additional file 1). From the RGI *kmer_query* analysis, the latter was likely found in *Enterococcus* spp., while the former has

a broader host range, including *Clostridiodes* spp. (Additional file 4). Neither were similar to plasmids via the MOB-suite analysis.

Including these highlighted ARGs, over 200 instances of ARGs were predicted to be on a plasmid (Additional file 4). Many of the genes originated from *S. aureus*, including *mecA*, *mecI*, *mecR1*, *ermC*, *msrA*, *mphC*, bla_{PC1}, *lnuA*, *dfrC*, *qacA*, and *qacB*. These genes were not unique to any infant group nor sample collection time-point. The *mecA*, *mecI*, *mecR1*, and *ermC* genes were more commonly found in NS infants. ARGs potentially found on plasmids from *Enterobacteriaceae* and *Gammaproteobacteria* included *sul2*, bla_{TEM-1}, *aac(6')-Ib-cr*, *aph(3'')-Ib*, *aph(6)-Id*, *dfrA1*, *E. coli* bla_{ampC}, and *tetO*. The *sul2*, *aac(6')-Ib-cr*, and *aph(6)-Id* genes were more common in NS infants compared to the other infants we investigated.

DISCUSSION

Given the consequences of antibiotic exposure on the gut microbiome and resistome, and the evidence suggesting that probiotics reduce the risk of NEC in preterm infants, we sought to study the effect of probiotics on the preterm infant gut resistome. While the overall number of ARGs between the groups of infants did not differ significantly, we found differences in the unique types of ARGs and resistance mechanisms. The NS infants harbored more unique ARGs associated with antibiotic inactivation mechanisms of resistance than the PS and FT infants. We also identified ARGs that persisted longer throughout the study period in NS infants compared to PS infants. Finally, by harnessing the unique aspects of targeted capture and our analysis approaches,

we could predict potential hosts and plasmid sequences associated with ARGs. Therefore, our survey of the resistome in preterm infants suggests that probiotics given within the first 12 weeks of life reduce the diversity of ARGs in the preterm infant gut and prevent the persistence of ARGs up to 5 months of age.

Antibiotic resistance genes were found in all infant gut microbiome samples, regardless of whether they were preterm or received probiotics. This is not surprising given that a diverse gut resistome containing beta-lactam, tetracycline, aminoglycoside, and chloramphenicol resistance genes has been found in both preterm and full-term infants [21, 69, 70]. We identified many of the same beta-lactamases, aminoglycoside-modifying enzymes, and tetracycline protection proteins as those detected previously in preterm and full-term infants, suggesting similar exposures to ARG-carrying bacteria throughout early life, perhaps from the NICU or other shared environments [8, 11, 12, 22, 41, 43].

We found more unique ARGs in NS preterm infants than in PS preterm infants at visit 1 (term age) and full-term infants at 10 days of age. This contradicts a previous study, where the resistome of preterm infants encoded fewer unique ARGs; however, the relative abundance of these ARGs was higher than in full-term infants [21], highlighting the importance of going beyond enumerating ARGs alone. Few studies have compared probiotic-supplemented preterm infants to full-term infants. One reported a much lower diversity of ARGs in all infants than our study did (99 vs over 200) [43]. This disparity is likely due to the limited detection of antibiotic resistance genes with shallow shotgun metagenomic sequencing. They noted that non-probiotic-supplemented preterm infants had

a higher abundance of certain beta-lactamases and efflux pumps compared to preterm infants supplemented with a similar probiotic to our study and full-term infants at 7 days [43]. Another group surveyed the resistome of preterm infants supplemented with *Bifidobacterium longum* subsp. *infantis* EVC001 using much deeper shotgun metagenomics (average 33.6 million reads per sample) and detected 315 unique ARGs [41]. They found that the burden of AMR was lower in PS preterm infants, and 67 unique ARGs, including a chloramphenicol acetyltransferase and macrolide resistance genes, were more abundant in NS infants [41]. One final study, however, reported no significant difference in the resistome of PS and NS preterm infants [42]. This study used the very limited approach of PCR to detect a small set of resistance genes [42].

Our design is distinct from others because we applied a powerful hybridization-based sequencing approach to specifically target over 2000 ARGs with 50–300× less sequencing effort than shotgun metagenomics, which reduces the cost of expanding to larger sample sets. As others have noted, we also found that NS preterm infants had a distinct resistome compared to PS preterm infants and full-term infants. This was evident in the number of unique ARGs we identified and the differences in resistance mechanisms and AMR gene families that were present. While we detected the same genes in our infant samples as other studies [41, 43], we did not highlight these same genes as differences between the infant groups. Alternatively, we found the association of unique antibiotic inactivation genes including the beta-lactamases *bla*$_{\text{CblA}}$, *bla*$_{\text{CTX-M}}$, and *bla*$_{\text{OXA}}$ and aminoglycoside-modifying enzyme *aac(3)* with only the NS infants and a reduced number

of unique ARGs in the PS and FT infants. Preterm infants that received probiotics were more comparable to full-term infants than NS infants in terms of the distribution of resistance mechanisms that were detected in the gut microbiome, in particular the numbers of antibiotic efflux and antibiotic inactivation genes. These results suggest that probiotic supplementation with *Bifidobacterium* and *Lactobacillus* species soon after birth reduces the diversity of antibiotic resistance genes in the preterm infant gut, resulting in a resistome that is more similar to full-term infants at 10 days of age than that of other preterm infants.

While previous longitudinal studies have investigated the impact of probiotics on the preterm infant gut resistome up to 4 months of age, we had a later time-point at 5-month corrected age. Esaiassen and co-workers reported that the resistome of PS infants was not significantly different from that of more mature infants at 4 months of age, suggesting that probiotics remediated the effects of premature birth and antibiotic exposure [43]. Nguyen and colleagues did not follow-up with infants after discharge from hospital but found that a longer stay in the NICU resulted in greater accumulation of ARGs, and probiotic supplementation reduced this effect [41]. While we did not have full-term infants to compare to at older time-points, we found that PS infants consistently had fewer unique ARGs up to 5 months of age as compared to NS infants. We, and others, also observed that the diversity of ARGs was higher at earlier points in life and decreased over time in both groups of preterm infants [8, 10]. Others have not noted, as we have, that certain aminoglycoside-modifying enzymes (*aac(3)*) and beta-lactamases ($bla_{\text{CTX-M-14}}$, bla_{SHV}) persisted longer in NS infants than PS infants. The reduced diversity and persistence of

ARGs associated with probiotic supplementation should be further monitored beyond 5 months of age. An approach similar to ours could facilitate these studies in a cost-effective and sensitive way.

To go beyond what has been accomplished with previous shotgun sequencing studies, we used the unique benefit of targeted capture in increasing the coverage of genetic regions surrounding ARGs to predict potential hosts and MGEs associated with those ARGs. Potentially mobile ARGs were found in all infants, as has been reported previously for both full-term and preterm infants [21, 22, 43, 69]. Indeed, the antibiotic inactivation genes that persisted in the NS infants in our study up to 5-month corrected age (*aac(3)-IId*, $bla_{\text{CTX-M-14}}$, bla_{SHV}) were associated with MGEs in various MDR *Enterobacteriaceae* and enterococci (Additional file 4) [71,72,73]. Bacteria belonging to the *Enterobacteriaceae* family have proportionately high levels of ARGs and can facilitate the transfer of ARGs through MGEs to other pathogens [73,74,75,76,77]. The association of these genes with MGEs and resistant pathobionts highlights the potential risk of dissemination of ARGs in the preterm infant gut. Various ARGs associated with MGEs and *Enterobacteriaceae* were more common in NS infants than PS infants and therefore suggest probiotics reduce the diversity of potentially mobile antibiotic resistance.

Similar to another study, we detected vancomycin ARGs in infants that did not receive vancomycin at later collection time-points as well as the *mecA* gene (Fig. S19, Additional files 1 and 4) [43]. We captured the entire vancomycin resistance gene cluster (consisting of 5 genes) in NS infants at visit 3 (12-week corrected age) and 4 (5-month

corrected age) that likely originated from *Enterococcus gallinarum* or other enterococci (Fig. S19, Additional files 1 and 4). Methicillin-resistant *Staphylococcus aureus* (MRSA) harbor the staphylococcal cassette chromosome *mec* that consists of the genes *mecA*, *mecI*, and *mecR1* [78]. The combination of two or more of these genes on the same contig was identified in 3 NS infants and 1 full-term infant in our study (Additional file 4). *Enterococcaceae* and *Staphylococcaceae* are often reported in preterm infants; however, only a few infants in our study had a relative abundance of > 1% of these families (Fig. S4, Additional file 1) [3, 6, 9, 10, 21, 79, 80]. Despite the low abundance of these families of bacteria, we still detected ARGs that likely originated from *Staphylococcus* spp. and *Enterococcus* spp. at various time-points. These two examples highlight the sensitivity of our approach and the potential to monitor rates of vancomycin-resistant enterococci or MRSA colonization and infection. We did not try isolating these organisms to test their antibiotic susceptibility.

Another interesting result is the prevalence and persistence of the SHV beta-lactamases in NS infants (Fig. 6). SHV beta-lactamases confer intrinsic resistance to penicillins and first-generation cephalosporins and are core chromosomal genes in a group of *Klebsiella pneumoniae* [73, 81, 82]. These genes have since been mobilized on plasmids in other members of the *Enterobacteriaceae*. The presence of SHV beta-lactamases at multiple study time-points in NS infants could suggest competitive inhibition of certain *K. pneumoniae* strains by the probiotic bacteria. We detected other genes that likely originated from *Klebsiella* spp. including bla_{OKP} in PS infants and therefore cannot rule out the

absence of *K. pneumoniae* in this infant group. Indeed, in the microbiota of the subset of infants, various ASVs associated with strains of *K. pneumoniae* were detected, although they were more abundant in NS infants. This may be a result of exposures to different *K. pneumoniae* strains in the NICU by these two groups of preterm infants.

Strengths of our study were its longitudinal nature, where samples were collected both in-hospital and after discharge (up to 5-month corrected age) and the timing of our study that captured a change in protocol in the NICU to provide probiotic supplementation as standard procedure. The sensitivity of our sequence capture method allowed us to detect ARGs at low prevalence with a small amount of sequencing data, and the ability to enrich the genetic context of ARGs allowed us to predict potential hosts and mobilization of ARGs. The limitations of our study included the small number of samples at each time-point and variability in antibiotic exposures of the infants. Compared to other approaches, targeted capture does not reflect the microbiome's functional genes or species diversity and cannot detect previously uncharacterized antibiotic resistance genes [46, 83, 84]. Finally, both analysis approaches used in our study have limitations when detecting antibiotic resistance in metagenomes [85] and rely on a reference database that requires curation and frequent updates. In general, more ARGs belonging to the AMR mechanism group of antibiotic efflux and antibiotic inactivation were reported in all infants. This reflects the biased distribution of resistance genes curated in the CARD, which is itself based on ARGs reported in the scientific literature [46].

CONCLUSIONS

In this study of preterm infants, we have highlighted the potential of a *Bifidobacterium* spp. and *Lactobacillus* spp. containing probiotic to reduce the diversity of AMR in the preterm infant gut. When compared with probiotic-supplemented infants, infants that did not receive probiotics had a higher number of unique ARGs that were predominantly associated with the mechanism of antibiotic inactivation and a greater diversity of antibiotic resistance genes that persisted up to 5 months of age in the gut microbiome. Furthermore, we highlight how the unique combination of targeted capture and analysis approaches can resolve individual ARGs, their surrounding genetic context, and predict potential bacterial hosts. This allowed us to associate many of the persistent antibiotic resistance genes in non-probiotic-supplemented infants with *Enterobacteriaceae* and MGEs. Our results suggest that probiotics can be used as a supplement during hospitalization to reduce the diversity of AMR in preterm infants that are exposed to a variety of multidrug resistant pathogens at an early age. Our study highlights the feasibility and advantages of using targeted capture in longitudinal cohort studies of the resistome and why it should be considered to further improve preterm infant care.

Availability of data and materials

Sequence data that support the findings of this study have been deposited in NCBI's Sequence Read Archive with the BioProject accession code: PRJNA805248. Two participants (NS2 and NS5) did not consent to the release of their data. Code used to analyze the data is available at https://github.com/AllisonGuitor/AMR-metatools or https://zenodo.org/badge/latestdoi/444553774.

ABBREVIATIONS

NICU: Neonatal intensive care unit

MDR: Multidrug resistant

ARG(s): Antibiotic resistance gene(s)

NEC: Necrotizing enterocolitis

PCR: Polymerase chain reaction

Baby and Pre-Mi: Baby and Preterm Microbiota of the Intestine Cohort Study

PS: Probiotic supplemented

NS: Not supplemented

FT: Full term

RGI: Resistance Gene Identifier

CARD: Comprehensive Antibiotic Resistance Database

qPCR: Quantitative PCR

ASV: Amplicon sequence variant

AMR: Antimicrobial resistance

RGI *bwt* : Metagenomic read-mapping feature of RGI

RGI *kmer_query* : Pathogen-of-origin feature of RGI

RGI *main* : Genome annotation feature of RGI

ORF(s): Open reading frame(s)

MGE(s): Mobile genetic element(s)

PMA: Postmenstrual age

ESBLs: Extended-spectrum beta-lactamases

REFERENCES

1. Pärnänen K, Karkman A, Hultman J, Lyra C, Bengtsson-Palme J, Larsson DGJ, et al. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. Nat Commun. 2018;9:1–11.
2. van den Akker CHP, van Goudoever JB, Shamir R, Domellöf M, Embleton ND, Hojsak I, et al. Probiotics and preterm infants: a position paper by the European Society for Paediatric Gastroenterology Hepatology and Nutrition Committee on Nutrition and the European Society for Paediatric Gastroenterology Hepatology and Nutrition Working Group for Pr. J Pediatr Gastroenterol Nutr. 2020;70:664–80.
3. Matamoros S, Gras-Leguen C, Le Vacon F, Potel G, de La Cochetiere M-F. Development of intestinal microbiota in infants and its impact on health. Trends Microbiol. 2013;21:167–73.
4. Simioni J, Hutton EK, Gunn E, Holloway AC, Stearns JC, McDonald H, et al. A comparison of intestinal microbiota in a population of low-risk infants exposed and not exposed to intrapartum antibiotics: the Baby & Microbiota of the Intestine Cohort Study protocol. BMC Pediatr. 2016;16:1–7.
5. Stearns JC, Simioni J, Gunn E, McDonald H, Holloway AC, Thabane L, et al. Intrapartum antibiotics for GBS prophylaxis alter colonization patterns in the early infant gut microbiome of low risk infants. Sci Rep. 2017;7:1–9.
6. Aguilar-Lopez M, Dinsmoor AM, Ho TTB, Donovan SM. A systematic review of the factors influencing microbial colonization of the preterm infant gut. Gut Microbes. 2021;13:1–33.
7. Healy DB, Ryan CA, Ross RP, Stanton C, Dempsey EM. Clinical implications of preterm infant gut microbiome development. Nat Microbiol. 2022;7:22–33.
8. Lebeaux RM, Coker MO, Dade EF, Palys TJ, Morrison HG, Ross BD, et al. The infant gut resistome is associated with *E. coli* and early-life exposures. BMC Microbiol. 2021;21:201.
9. Arboleya S, Sánchez B, Milani C, Duranti S, Solís G, Fernández N, et al. Intestinal microbiota development in preterm neonates and effect of perinatal antibiotics. J Pediatr. 2015;166:538–44.
10. Moles L, Gómez M, Jiménez E, Fernández L, Bustos G, Chaves F, et al. Preterm infant gut colonization in the neonatal ICU and complete restoration 2 years later. Clin Microbiol Infect. 2015;21:936.e1–936.e10.
11. Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, et al. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat Commun. 2017;8:1814.

12. Raveh-Sadka T, Firek B, Sharon I, Baker R, Brown CT, Thomas BC, et al. Evidence for persistent and shared bacterial strains against a background of largely unique gut colonization in hospitalized premature infants. ISME J. 2016;10:2817–30.
13. Hagbø M, Ravi A, Angell IL, Sunde M, Ludvigsen J, Diep DB, et al. Experimental support for multidrug resistance transfer potential in the preterm infant gut microbiota. Pediatr Res. 2020;88:57–65.
14. Ravi A, Estensmo ELF, Abée-Lund TML, Foley SL, Allgaier B, Martin CR, et al. Association of the gut microbiota mobilome with hospital location and birth weight in preterm infants. Pediatr Res. 2017;82:829–38.
15. Prusakov P, Goff DA, Wozniak PS, Cassim A, Scipion CEA, Urzúa S, et al. A global point prevalence survey of antimicrobial use in neonatal intensive care units: the no-more-antibiotics and resistance (NO-MAS-R) study. EClinicalMedicine. 2021;32:100727.
16. Hsieh EM, Hornik CP, Clark RH, Laughon MM, Benjamin DK, Smith PB. Medication use in the neonatal intensive care unit. Am J Perinatol. 2014;31:811–21.
17. Gibson MK, Crofts TS, Dantas G. Antibiotics and the developing infant gut microbiota and resistome. Curr Opin Microbiol. 2015;27:51–6.
18. Fjalstad JW, Esaiassen E, Juvet LK, van den Anker JN, Klingenberg C. Antibiotic therapy in neonates and impact on gut microbiota and antibiotic resistance development: a systematic review. J Antimicrob Chemother. 2018;73:569–80.
19. Greenwood C, Morrow AL, Lagomarcino AJ, Altaye M, Taft DH, Yu Z, et al. Early empiric antibiotic use in preterm infants is associated with lower bacterial diversity and higher relative abundance of enterobacter. J Pediatr. 2014;165:23–9.
20. Zou ZH, Liu D, Li HD, Zhu DP, He Y, Hou T, et al. Prenatal and postnatal antibiotic exposure influences the gut microbiota of preterm infants in neonatal intensive care units. Ann Clin Microbiol Antimicrob. 2018;17:1–11.
21. Gasparrini AJ, Wang B, Sun X, Kennedy EA, Hernandez-Leyva A, Ndao IM, et al. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. Nat Microbiol. 2019;4:2285–97.
22. Gibson MK, Wang B, Ahmadi S, Burnham CAD, Tarr PI, Warner BB, et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. Nat Microbiol. 2016;1:16024.
23. Gasparrini AJ, Crofts TS, Gibson MK, Tarr PI, Warner BB, Dantas G. Antibiotic perturbation of the preterm infant gut microbiome and resistome. Gut Microbes. 2016;7:443–9.
24. Neu J, Walker WA. Necrotizing enterocolitis. N Engl J Med. 2011;364:255–64.
25. Esaiassen E, Fjalstad JW, Juvet LK, van den Anker JN, Klingenberg C. Antibiotic exposure in neonates and early adverse outcomes: a systematic review and meta-analysis. J Antimicrob Chemother. 2017;72:1858–70.

26. Michael Cotten C, Taylor S, Stoll B, Goldberg RN, Hansen NI, Sanchez PJ, et al. Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. Pediatrics. 2009;123:58–66.
27. Pammi M, Cope J, Tarr PI, Warner BB, Morrow AL, Mai V, et al. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. Microbiome. 2017;5:1–15.
28. Athalye-Jape G, Patole S. Probiotics for preterm infants – time to end all controversies. J Microbial Biotechnol. 2019;12:249–53.
29. Underwood MA. Arguments for routine administration of probiotics for NEC prevention. Curr Opin Pediatr. 2019;31:188–94.
30. Chang H-Y, Chen J-H, Chang J-H, Lin H-C, Lin C-Y, Peng C-C. Multiple strains probiotics appear to be the most effective probiotics in the prevention of necrotizing enterocolitis and mortality: An updated meta-analysis. PLoS One. 2017;12:e0171579.
31. Deshmukh M, Patole S. Current status of probiotics for preterm infants. Indian J Pediatr. 2021;88(July):703–8.
32. Underwood MA, Umberger E, Patel RM. Safety and efficacy of probiotic administration to preterm infants: ten common questions. Pediatr Res. 2020;88:48–55.
33. Aceti A, Beghetti I, Maggio L, Martini S, Faldella G, Corvaglia L. Filling the gaps: Current research directions for a rational use of probiotics in preterm infants. Nutrients. 2018;10:1–10.
34. Nolan LS, Rimer JM, Good M. The role of human milk oligosaccharides and probiotics on the neonatal microbiome and risk of necrotizing enterocolitis: a narrative review. Nutrients. 2020;12(10):3052.
35. Patole SK, Rao SC, Keil AD, Nathan EA, Doherty DA, Simmer KN. Benefits of *Bifidobacterium breve* M-16V supplementation in preterm neonates - a retrospective cohort study. PLoS One. 2016;11:e0150775.
36. Janvier A, Malo J, Barrington KJ. Cohort study of probiotics in a North American neonatal intensive care unit. J Pediatr. 2014;164:980–5.
37. Olsen R, Greisen G, Schrøder M, Brok J. Prophylactic probiotics for preterm infants: a systematic review and meta-analysis of observational studies. Neonatology. 2016;109:105–12.
38. Taft DH, Liu J, Maldonado-Gomez MX, Akre S, Huda MN, Ahmad SM, et al. Bifidobacterial dominance of the gut in early life and acquisition of antimicrobial resistance. mSphere. 2018;3:1–24.
39. Wuethrich I, Pelzer BW, Khodamoradi Y, Vehreschild MJGT. The role of the human gut microbiota in colonization and infection with multidrug-resistant bacteria. Gut Microbes. 2021;13:1–13.

40. Alcon-Giner C, Dalby MJ, Caim S, Ketskemety J, Shaw A, Sim K, et al. Microbiota supplementation with Bifidobacterium and lactobacillus modifies the preterm infant gut microbiota and metabolome: an observational study. Cell Rep Med. 2020;1:100077.
41. Nguyen M, Holdbrooks H, Mishra P, Abrantes MA, Eskew S, Garma M, et al. Impact of probiotic *B. infantis* EVC001 feeding in premature infants on the gut microbiome, nosocomially acquired antibiotic resistance, and enteric inflammation. Front Pediatr. 2021;9(February):1–19.
42. van Best N, Trepels-Kottek S, Savelkoul P, Orlikowsky T, Hornef MW, Penders J. Influence of probiotic supplementation on the developing microbiota in human preterm neonates. Gut Microbes. 2020;12:1–16.
43. Esaiassen E, Hjerde E, Cavanagh JP, Pedersen T, Andresen JH, Rettedal SI, et al. Effects of probiotic supplementation on the gut microbiota and antibiotic resistome development in preterm infants. Front Pediatr. 2018;6:347.
44. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep. 2018;8:1–11.
45. Waseem H, Jameel S, Ali J, Saleem Ur Rehman H, Tauseef I, Farooq U, et al. Contributions and challenges of high throughput qPCR for determining antimicrobial resistance in the environment: a critical review. Molecules. 2019;24:163.
46. Guitor AK, Raphenya AR, Klunk J, Kuch M, Alcock B, Surette MG, et al. Capturing the resistome: a targeted capture method to reveal antibiotic resistance determinants in metagenomes. Antimicrob Agents Chemother. 2019;64:e01324–19.
47. Yousuf EI, Carvalho M, Dizzell SE, Kim S, Gunn E, Twiss J, et al. Persistence of suspected probiotic organisms in preterm infant gut microbiota weeks after probiotic supplementation in the NICU. Front Microbiol. 2020;11:2305.
48. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48:D517–25.
49. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. Appl Environ Microbiol. 2011;77:3846–52.
50. Stearns JC, Davidson CJ, Mckeon S, Whelan FJ, Fontes ME, Schryvers AB, et al. Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. ISME J. 2015;9:1246–59.
51. Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014;15:1–12.
52. Jiang H. skewer. 2021. https://github.com/relipmoc/skewer. Accessed 25 Jun 2021.

53. Bushnell B. BBMap – Bushnell B. https://github.com/BioInfoTools/BBMap. Accessed 11 Jan 2021.
54. Li H. seqtk. https://github.com/lh3/seqtk. Accessed 6 Jul 2021.
55. Raphenya AR. RGI. https://github.com/arpcard/rgi. Accessed 14 Oct 2020.
56. Langmead B. Bowtie2. https://github.com/BenLangmead/bowtie2. Accessed 6 Jul 2021.
57. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357–9.
58. McArthur AG. CARD Downloads. https://card.mcmaster.ca/download. Accessed 4 Aug 2021.
59. SPAdes Genome Assembler. https://github.com/ablab/spades. Accessed 25 Jun 2021.
60. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinformatics. 2020;70:e102.
61. Hulsen T, de Vlieg J, Alkema W. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics. 2008;9:488.
62. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
63. PHAC-NML. mob-suite. https://github.com/phac-nml/mob-suite). Accessed 3 Jul 2021.
64. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom. 2018;4:e000206.
65. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
66. Bethesda (MD): National Library of Medicine (US) NC for BI. Nucleotide. https://www.ncbi.nlm.nih.gov/nucleotide/. Accessed 4 June 2021.
67. McArthur AG. Resistomes, variants, and prevalence. https://card.mcmaster.ca/resistomes. Accessed 27 Jun 2021.
68. Gilchrist C. Clinker. https://github.com/gamcil/clinker. Accessed 17 Dec 2020.
69. Moore AM, Patel S, Forsberg KJ, Wang B, Bentley G, Razia Y, et al. Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. PLoS One. 2013;8:e78822.
70. Moore AM, Ahmadi S, Patel S, Gibson MK, Wang B, Ndao IM, et al. Erratum: gut resistome development in healthy twin pairs in the first year of life. Microbiome. 2015;3:1–10.
71. Cocchi S, Grasselli E, Gutacker M, Benagli C, Convert M, Piffaretti JC. Distribution and characterization of integrons in *Escherichia coli* strains of animal and human origin. FEMS Immunol Med Microbiol. 2007;50:126–32.

72. Navon-Venezia S, Kondratyeva K, Carattoli A. *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. FEMS Microbiol Rev. 2017;41:252–75.
73. Wyres KL, Holt KE. *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Curr Opin Microbiol. 2018;45:131–9.
74. Shin NR, Whon TW, Bae JW. Proteobacteria: microbial signature of dysbiosis in gut microbiota. Trends Biotechnol. 2015;33:496–503.
75. Lee K, Kim DW, Lee DH, Kim YS, Bu JH, Cha JH, et al. Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome. 2020;8:1–14.
76. Goldstone RJ, Smith DGE. A population genomics approach to exploiting the accessory resistome of *Escherichia coli*. Microb Genom. 2017;3(4):e000108.
77. Bengtsson-Palme J, Kristiansson E, Larsson DGJ. Environmental factors influencing the development and spread of antibiotic resistance. FEMS Microbiol Rev. 2018;42:68–80.
78. Ito T, Hiramatsu K, Oliveira DC, De Lencastre H, Zhang K, Westh H, et al. Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrob Agents Chemother. 2009;53:4961–7.
79. Groer MW, Luciano AA, Dishaw LJ, Ashmeade TL, Miller E, Gilbert JA. Development of the preterm infant gut microbiome: a research priority. Microbiome. 2014;2:1–8.
80. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe. 2015;17:690–703.
81. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A. 2015;112:E3574–81.
82. Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. Nat Rev Microbiol. 2020;18:344–59.
83. Crofts TS, Gasparrini AJ, Dantas G. Next-generation approaches to understand and combat the antibiotic resistome. Nat Rev Microbiol. 2017;15:422–34.
84. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6:23.
85. Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet. 2019;20:356–70.

ACKNOWLEDGEMENTS

We acknowledge the McMaster Genomics Facility for next-generation sequencing.

FUNDING

This research was funded by Canadian Institutes of Health Research grants (FRN-148463 to J. C. S & PJT-156214 to A. G. M), by a Canada Research Chair to G. D. W, a David Braley Chair in Computational Biology to A. G. M, and by the Boris Family Fund for Health Research Excellence. A. K. G was supported by a Canadian Institutes of Health Research Doctoral Research Award (GSD-164145). Computer resources were supplied by the McMaster Service Lab and Repository computing cluster, funded in part by grants to A. G. M. from the Canadian Foundation for Innovation (grant 34531). J. C. S is supported by the Farncombe Chair in Microbial Ecology and Bioinformatics. E. I. Y is supported by an Ontario Graduate Scholarship.

AUTHOR INFORMATION

Authors and Affiliations

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Canada

Allison K. Guitor, Amogelang R. Raphenya, Andrew G. McArthur, Gerard D. Wright & Jennifer C. Stearns

Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada

Allison K. Guitor, Amogelang R. Raphenya, Andrew G. McArthur & Gerard D. Wright

David Braley Centre for Antibiotic Discovery, McMaster University, Hamilton, Canada

Allison K. Guitor, Amogelang R. Raphenya, Andrew G. McArthur & Gerard D. Wright

Department of Pediatrics, McMaster University, Hamilton, Canada

Efrah I. Yousuf & Katherine M. Morrison

Department of Obstetrics & Gynecology, McMaster University, Hamilton, Canada

Eileen K. Hutton

The Baby & Mi and the Baby & Pre-Mi Cohort Studies, Hamilton, Canada

Eileen K. Hutton, Katherine M. Morrison & Jennifer C. Stearns

Department of Medicine, McMaster University, Hamilton, Canada

Jennifer C. Stearns

Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, Canada

Jennifer C. Stearns

CONTRIBUTIONS

The studies in which samples were collected were led by EKH and KMM; EIY oversaw participant recruitment and data and sample collection. EIY and JCS generated and analyzed the microbiome data. AKG planned the experiments, performed DNA library preparation, and targeted enrichment and preparation of samples for sequencing. ARR and AKG wrote the code. AGM provided advice on analyses and computational support. AKG performed the analyses and generated the tables and figures for the resistome capture data.

AKG wrote the manuscript with primary input from EIY, JCS, and GDW. The authors read and approved the final manuscript.

Ethics declarations

Ethics approval and consent to participate

Ethics approval was obtained from the Hamilton Integrated Research Ethics Board (no. 1909), and parents provided consent.

SUPPLEMENTARY INFORMATION

Supplementary Methods:

Library preparation methods - The numbers correspond to values used in File S3.

1. No size-selection, only a clean-up step during library preparation;
   10 cycles of PCR amplification;
   Final library eluted in 33 ul 0.1XTE
2. No size selection, only a clean-up step during library preparation;
   8 cycles of PCR amplification;
   Final library eluted in 33 ul 0.1XTE
3. Size selection for 500-700 bp inserts
   7 cycles of PCR amplification;
   Final library eluted in 33 ul 0.1XTE.
4. No size selection, only a clean-up step during library preparation;
   7 cycles of PCR amplification;
   Final library eluted in 33 ul 0.1XTE
5. Size selection for 500-700 bp inserts
   4 cycles of PCR amplification;
   Final library eluted in 33 ul 0.1XTE.

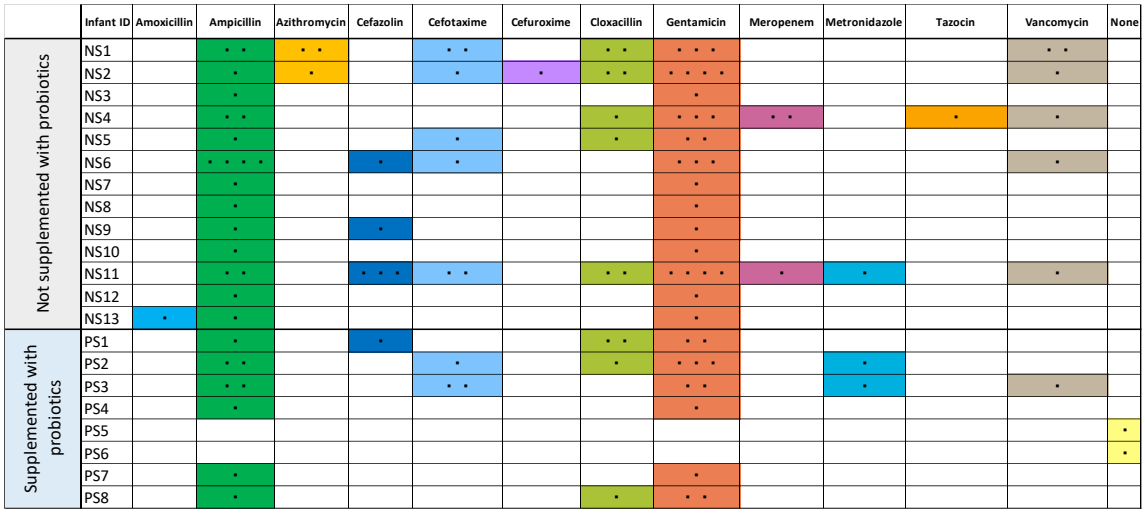


Figure S1: Antibiotic exposure of preterm infants.
The antibiotics preterm infants were exposed to throughout the duration of the study. A colored box indicates that infant received the antibiotic. The dots within each box represent the number of times they were given that antibiotic. The length of treatment is not reflected in this figure.

Table S1: Information regarding the full-term infants.

	Full-term (n = 9)
Gestational age at birth, weeks	39.78 ± 0.61
Probiotic exposure, weeks	None
Antibiotic exposure during sample collection (types and number of infants exposed)	None
Antibiotic exposure, weeks	None
In-hospital sample, weeks in PMA (N)	N/A
Visit 1 sample, weeks in PMA (N)	41.25 ± 0.64 (9)
Visit 2 sample, weeks in PMA (N)	N/A
Visit 3 sample, weeks in PMA (N)	N/A
Visit 4 sample, weeks in PMA (N)	N/A
PMA is the postmenstrual age in weeks and SD is the standard deviation. The data are presented as mean ± SD.	

Table S2: Information regarding the subset of preterm infants.

	NS preterm (n = 6)	**PS preterm (n = 6)**	***p*-value**
Gestational age at birth, weeks	28.26 ± 0.84	27.81 ± 1.65	0.60
Probiotic exposure, weeks	0	8.21 ± 3.68	**<0.0001**
Antibiotic exposure during sample collection (types and number of infants exposed)	Amo(1) Amp(6) Cefa(2) Cefo(2) Cl(1) G(6) V(1)	Amp(6) Cefa(1) Cefo(2) Cl(3) G(6) Met(2) V(1)	N/A
Antibiotic exposure, weeks	1.02 ± 0.80	1.48 ± 1.17	0.49
In-hospital sample, weeks in PMA (N)	36.69 ± 3.84 (6)	37.86 ± 1.69 (3)	0.96
Visit 1 sample, weeks in PMA (N)	41.03 ± 1.86 (5)	42.38 ± 1.88 (6)	0.31
Visit 2 sample, weeks in PMA (N)	47.26 ± 2.23 (5)	46.43 ± 0.50 (5)	>0.99
Visit 3 sample, weeks in PMA (N)	50.19 ± 1.33 (3)	54.21 ± 1.85 (6)	**0.04**
Visit 4 sample, weeks in PMA (N)	61.81 ± 1.95 (6)	59.68 ± 0.62 (4)	0.08
PMA is the postmenstrual age in weeks and SD is the standard deviation. The data are presented as mean ± SD.			

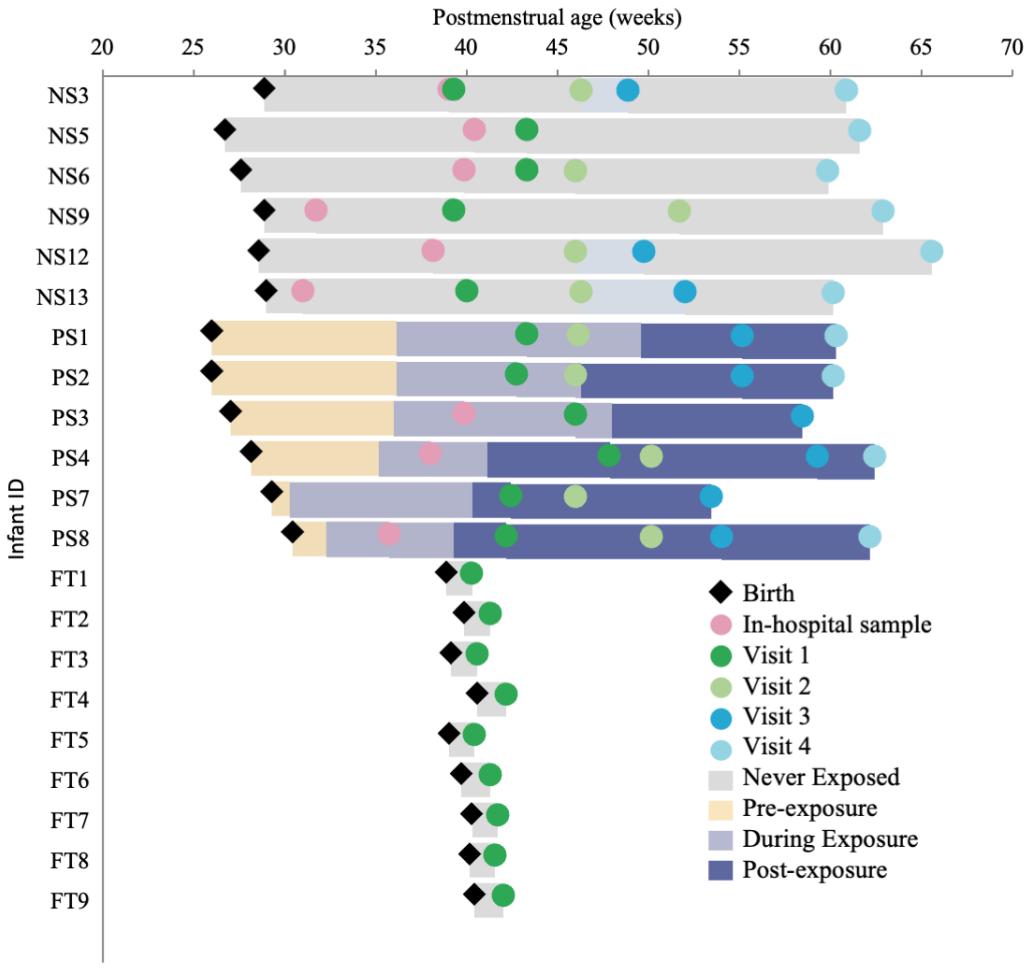


Figure S2: Sample collection and probiotic exposure of the subset of infants.
Timelines from birth to final sampling point for the subset of infants. The duration of exposure to probiotics (lavender bar) as well as all study points included in this study (up to 4 timepoints) for not supplemented (NS) and probiotic-supplemented (PS) preterm infants as well as full-term (FT) infants are shown in relation to postmenstrual age in weeks.

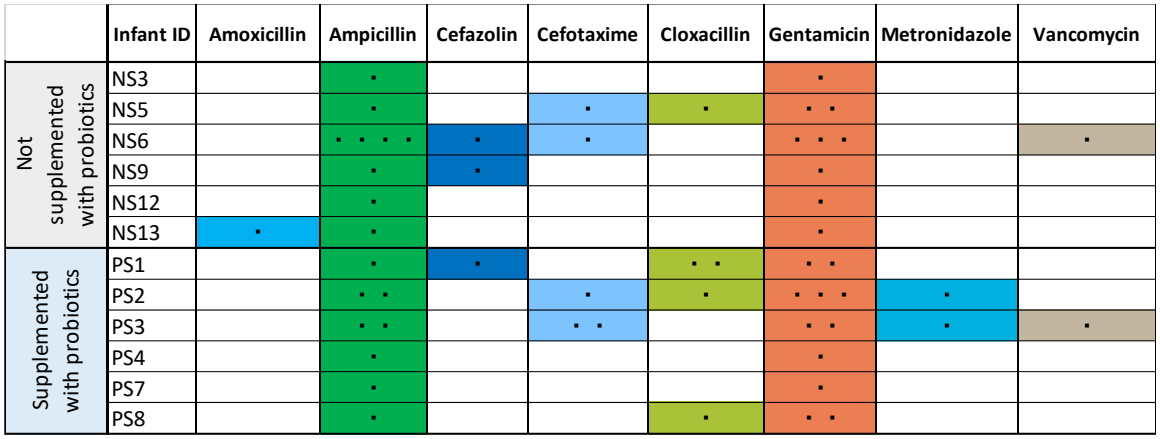


Figure S3: Antibiotic exposure of the subset of preterm infants.
The antibiotics preterm infants were exposed to throughout the duration of the study. A colored box indicates that infant received the antibiotic. The dots within each box represent the number of times they were given that antibiotic. The length of treatment is not reflected in this figure.

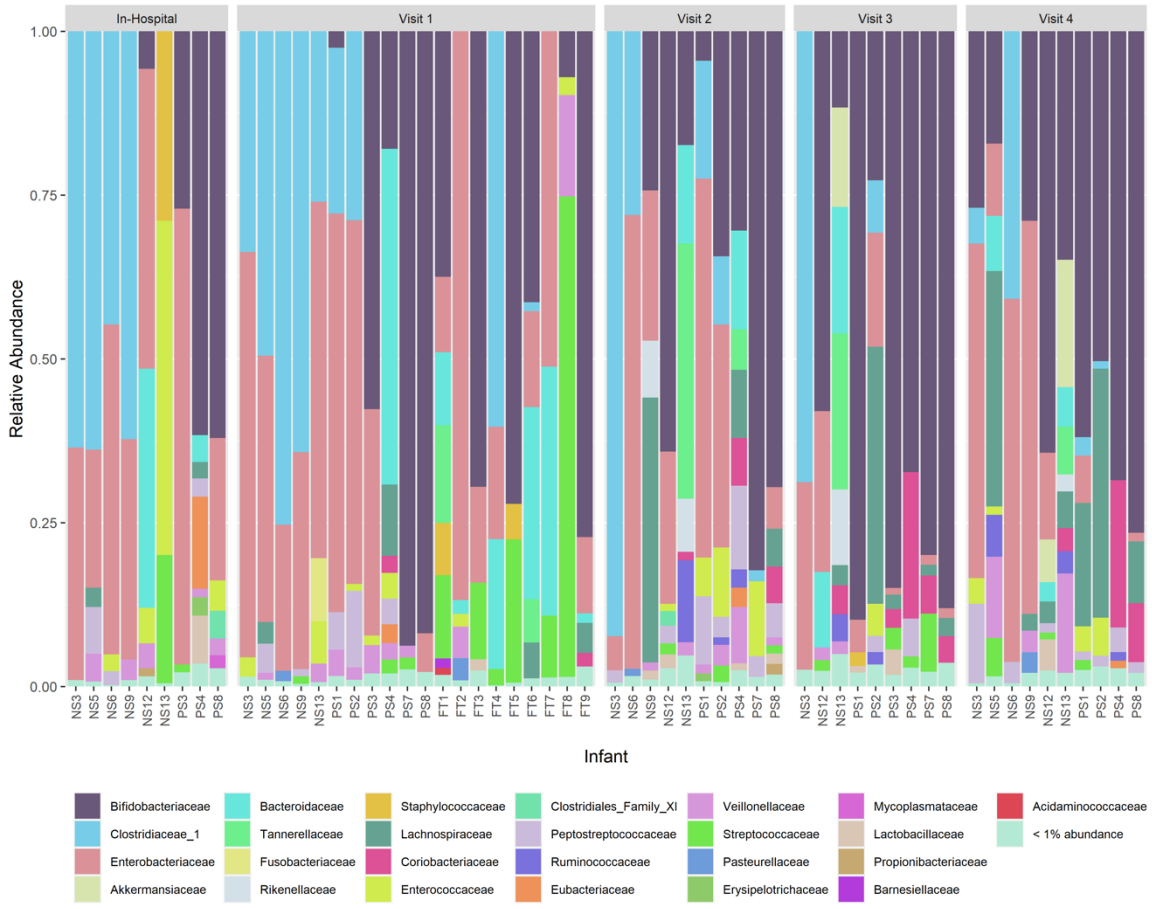


Figure S4: Relative abundance of bacterial families in the infant gut microbiome. Stacked bar charts representing the relative abundance of bacterial families in the subset of preterm infants up to 5 months corrected age and for full-term infants at 10 days postpartum. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants. The category <1% abundance represents all ASVs that were present at less than 1% of the relative abundance across all infants.

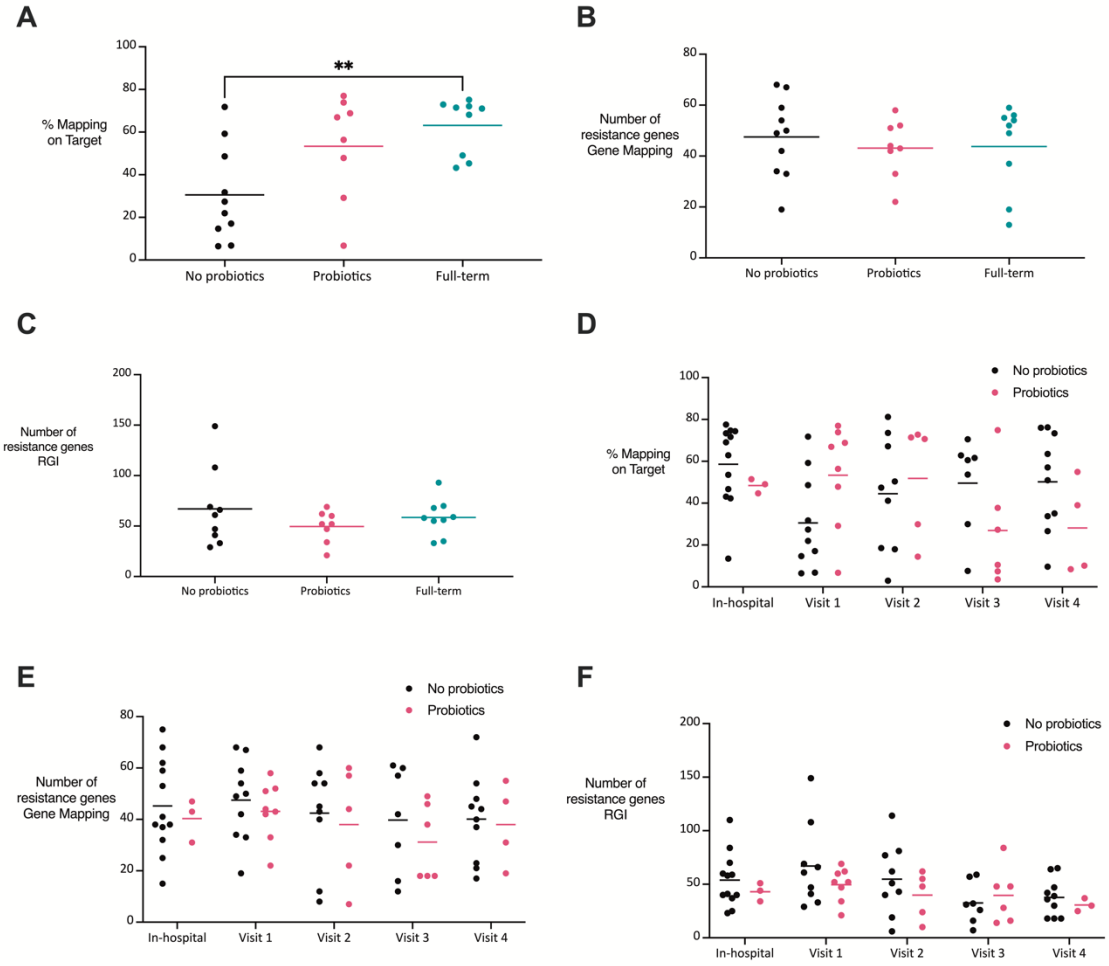


Figure S5: Targeted capture of resistance genes in the full set of infants.
A) A subsampled number of reads were mapped to CARD using RGI *bwt*. The percentage of reads mapping to the target of ARGs is reported for each infant at Visit 1. Significant differences are indicated by a line and asterisk(s) above the groups compared. P = 0.0014 for no probiotic vs full-term. B) Using the results of mapping reads to CARD, the number of ARGs with at least 100 mapped reads is reported for Visit 1. C) Subsampled reads from each infant sample were subjected to *de novo* assembly and prediction of ARGs using RGI *main*. The number of genes predicted at each Visit 1 is reported. D) The percentage of reads mapping to the target of ARGs is reported for preterm infants across all visits. E) Using the results of mapping reads to CARD, the number of ARGs with at least 100 mapped reads is reported for preterm infants at all visits. F) Subsampled reads from each infant sample were subjected to *de novo* assembly and prediction of ARGs using RGI *main*. The number of genes predicted in each preterm infant at all visit timepoints is reported.

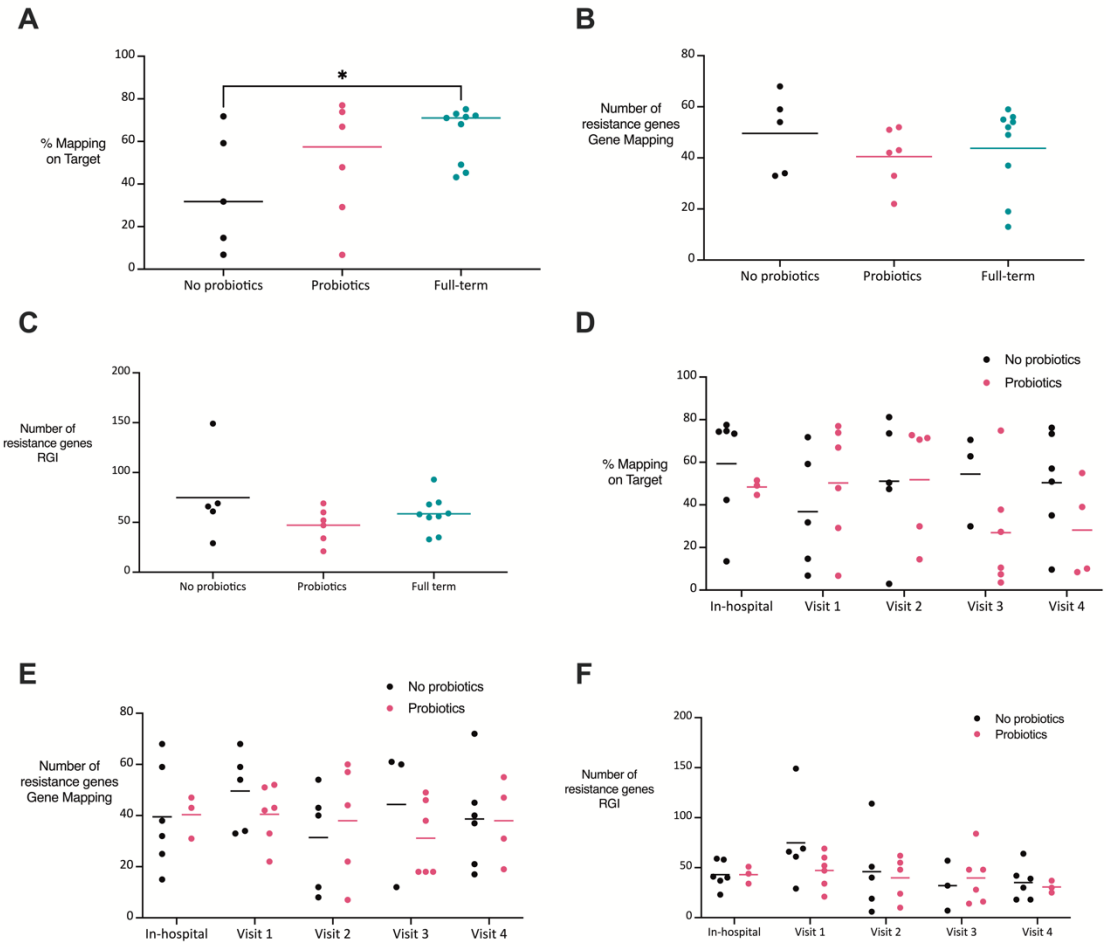


Figure S6: Targeted capture of resistance genes in the subset of infants.
A) For the subset of preterm infant samples and full-term infants at all visits, a subsampled number of reads were mapped to CARD using RGI *bwt*. The percentage of reads mapping to the target of ARGs is reported for each infant at Visit 1. Significant differences are indicated by a line and asterisk(s) above the groups compared. P = 0.0318 for Visit 1 no probiotic vs full-term. B) Using the results of mapping reads to CARD, the number of ARGs with at least 100 reads is reported for all infants at Visit 1. C) Subsampled reads from each infant sample were also subjected to *de novo* assembly and prediction of ARGs with RGI *main*. The number of genes predicted at Visit 1 is reported. D) The percentage of reads mapping to the target of ARGs is reported for preterm infants at all visits. E) Using the results of mapping reads to CARD, the number of ARGs with at least 100 reads is reported for all preterm infants at all visits. F) Subsampled reads from each infant sample were also subjected to *de novo* assembly and prediction of ARGs with RGI *main*. The number of genes predicted at all visits for the preterm infants is reported.

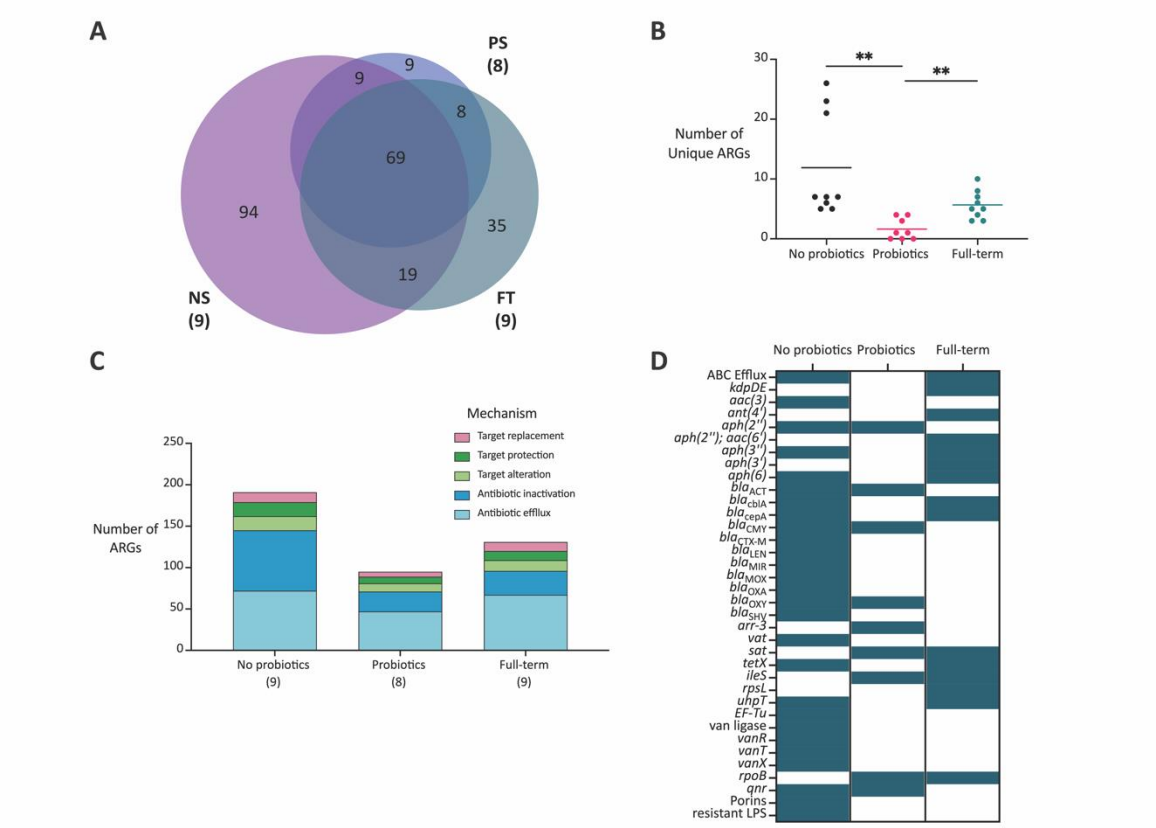


Figure S7: Differences in the resistome identified through RGI *main* at Visit 1.
These ARG counts are from the results of *de novo* assembly and ARG prediction using RGI *main* and CARD. Perfect and strict hits are reported. The data presented is from all preterm and full-term infants at Visit 1. A) Unique and overlapping ARGs identified in each infant group. The number of infant samples included in each is shown next to the sample type. B) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared. P = 0.0052 for NS vs PS, P = 0.0012 for PS vs FT. C) A breakdown of the mechanisms of antibiotic resistance identified in each infant group. The number of infant samples included in each is shown next to the sample type. D) The presence or absence of selected AMR gene families in each infant group. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

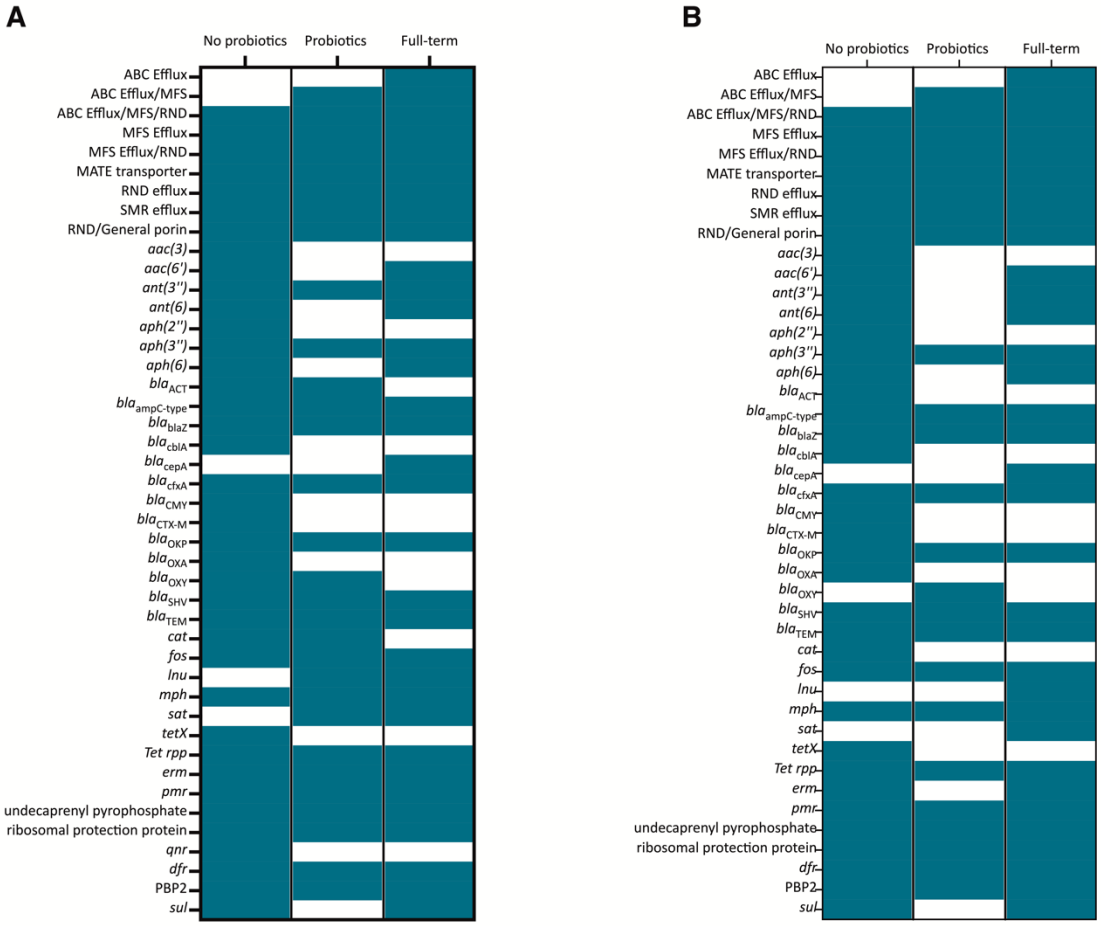


Figure S8: Distribution of genes detected at the AMR gene family level through RGI *bwt*.

A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples at that timepoint through read mapping to CARD. A) For all preterm and full-term infants. B) For the subset of preterm infants and all full-term infants.

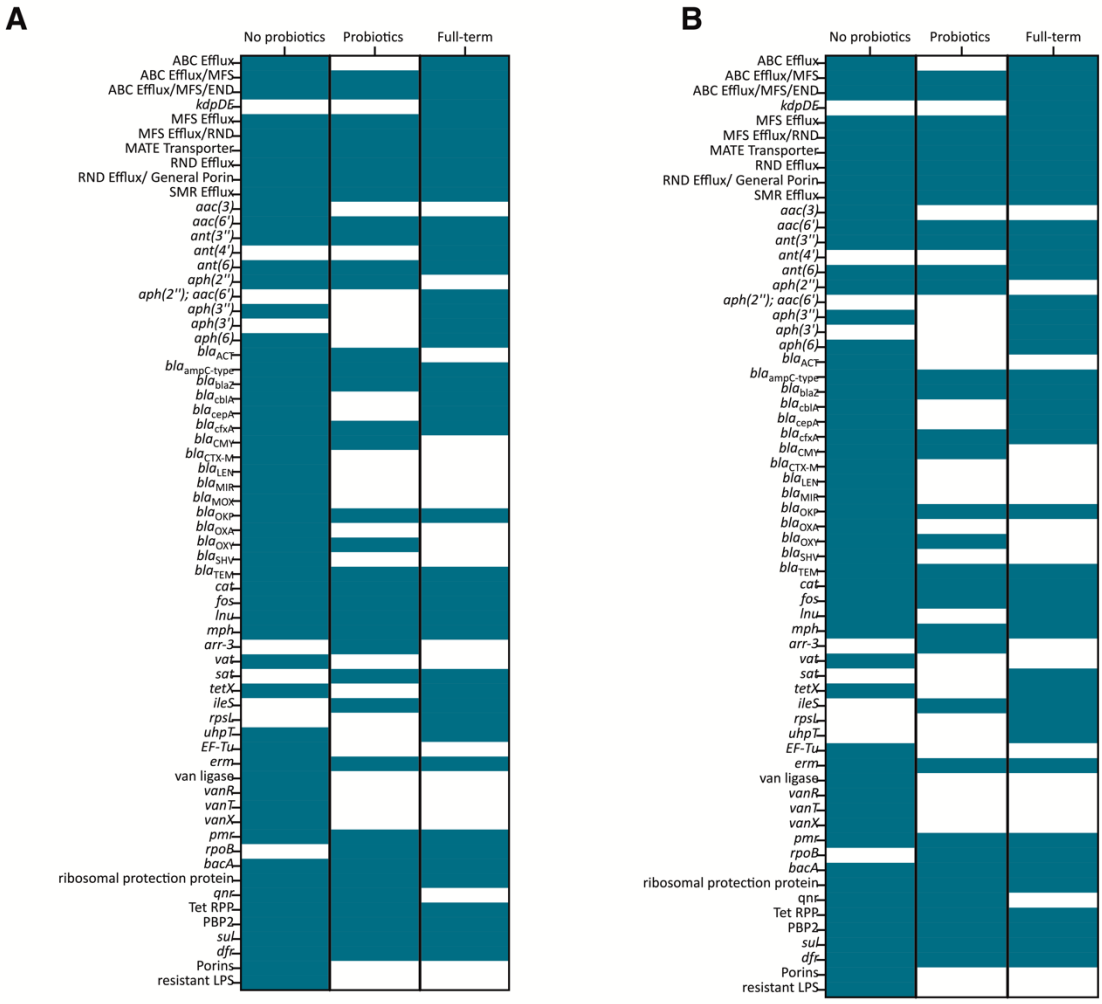


Figure S9: Distribution of genes detected at the AMR gene family level through RGI *main*.

A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples at that timepoint through *de novo* assembly and prediction of ARGs through RGI *main*. A) For all preterm and full-term infants. B) For the subset of preterm infants and all full-term infants.

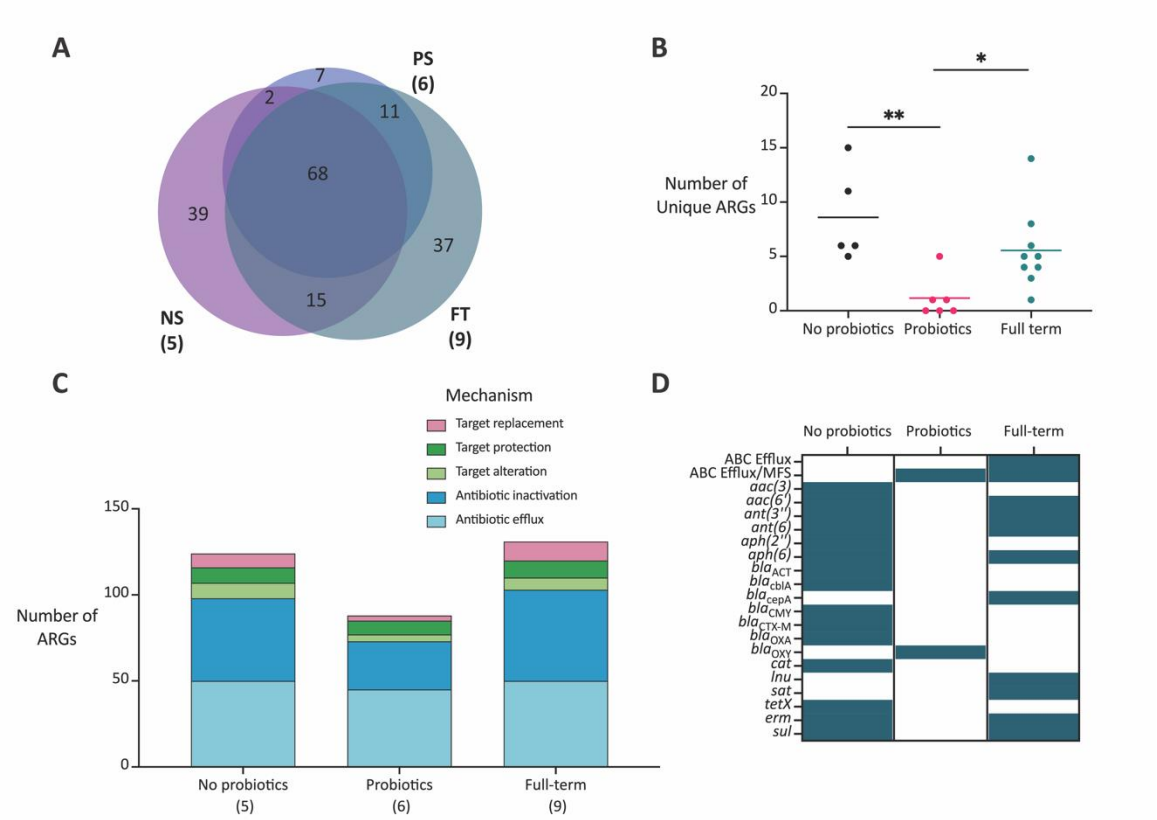


Figure S10: Differences in the resistome identified through RGI *bwt* in the subset at Visit 1.

Reads were mapped to CARD using bowtie2 and antibiotic resistance genes with at least 100 reads were reported. The data presented is from the subset of preterm infants and all full-term infants at Visit 1. A) Unique and overlapping ARGs identified in each infant group. The number of infant samples included in each is shown next to the sample type. B) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared. P = 0.0040 for NS vs PS, P = 0.0203 for PS vs FT. C) A breakdown of the mechanisms of antibiotic resistance identified in each infant group. The number of infant samples included in each is shown next to the sample type. D) The presence or absence of selected AMR gene families in each infant group. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

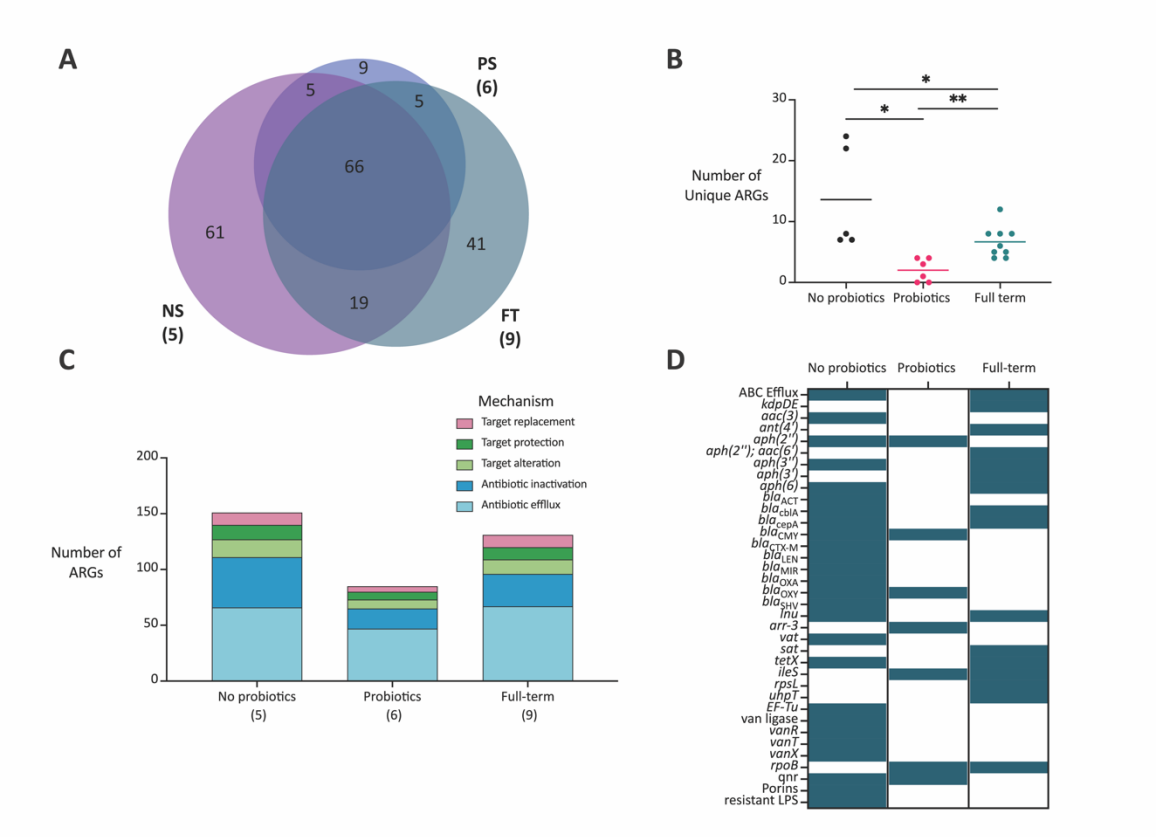


Figure S11: Differences in the resistome identified through RGI *main* in the subset at Visit 1.

These ARG counts are from the results of *de novo* assembly and ARG prediction using RGI *main* and CARD. Perfect and strict hits are reported. The data presented is from the subset of preterm infants and all full-term infants at Visit 1. A) Unique and overlapping ARGs identified in each infant group. The number of infant samples included in each is shown next to the sample type. B) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared. P = 0.0404 for NS vs FT, P = 0.0102 for NS vs PS, P = 0.0024 for PS vs FT. C) A breakdown of the mechanisms of antibiotic resistance identified in each infant group. The number of infant samples included in each is shown next to the sample type. D) The presence or absence of selected AMR gene families in each infant group. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

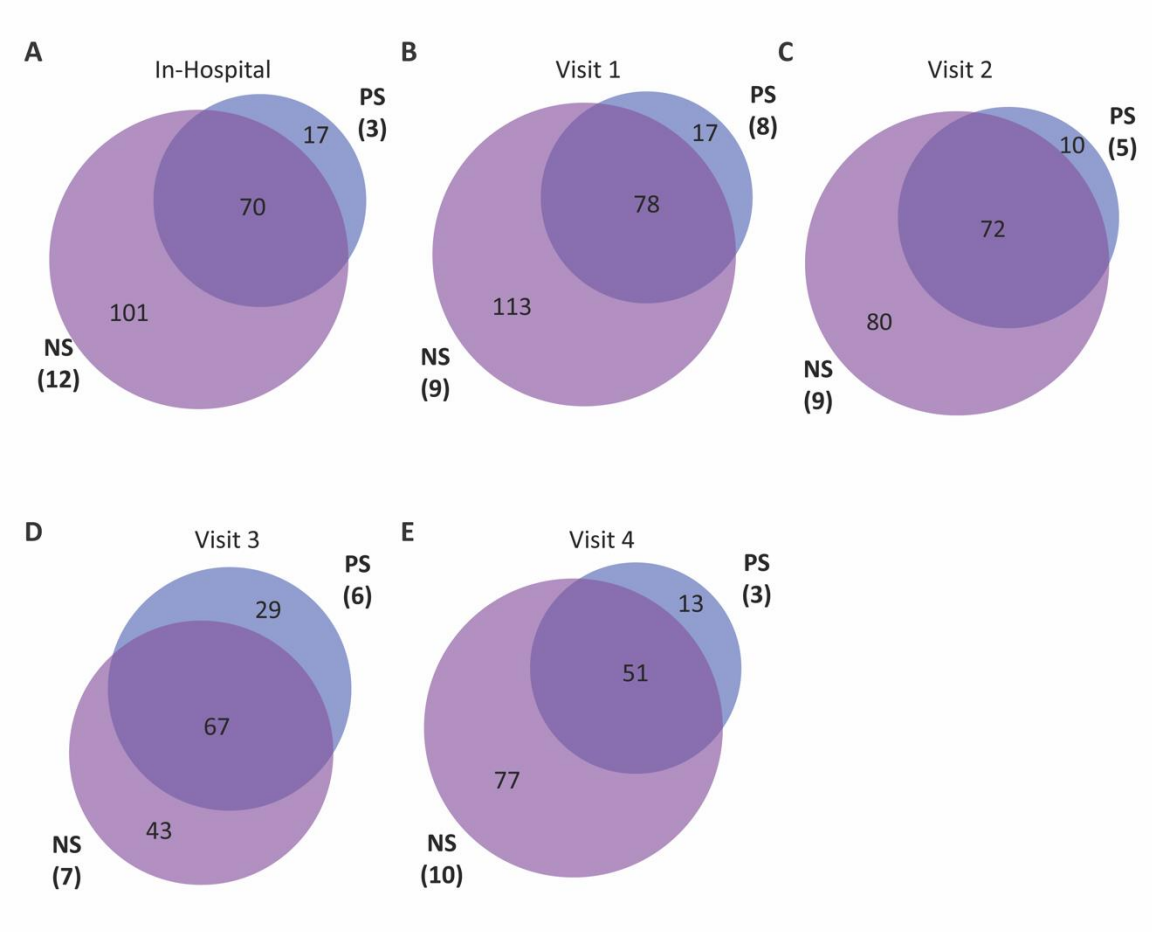


Figure S12: Unique genes in each infant group at various timepoints for the preterm infants.

These gene counts are from the results of *de novo* assembly and ARG prediction using RGI *main* and CARD. Perfect and strict hits are reported. Data from all preterm infants the in-hospital collection (A), Visit 1 (B), Visit 2 (C), Visit 3 (D), and Visit 4 (E) timepoints. The number of infants included in each time point is indicated. NS = not supplemented preterm, PS = probiotic-supplemented preterm.

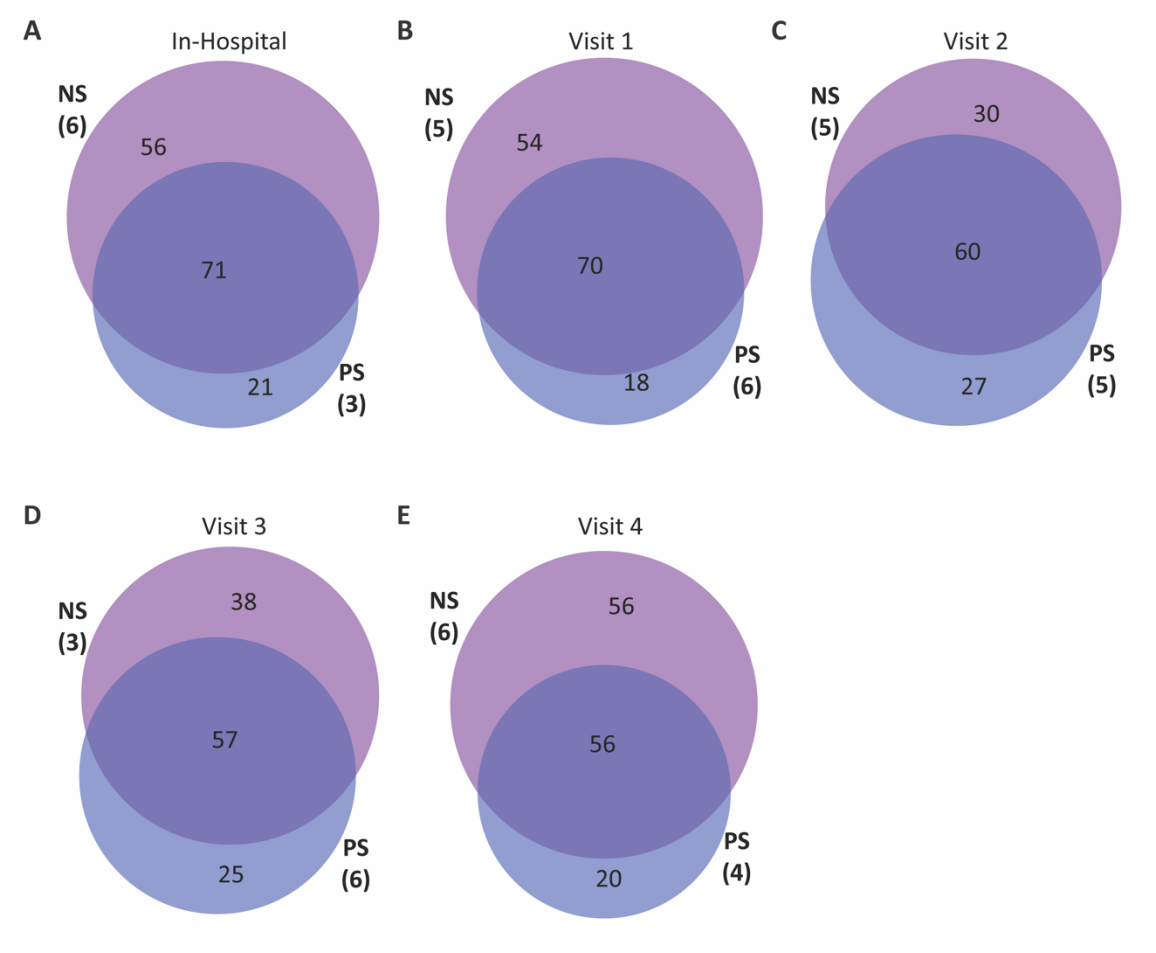


Figure S13: Unique genes in each infant group for the subset of preterm infants at various timepoints.
These gene counts are from the results of mapping reads to CARD using RGI *bwt* and counting the number of genes with at least 100 reads. Data from the subset of preterm infants at the in-hospital collection (A), Visit 1 (B), Visit 2 (C), Visit 3 (D), and Visit 4 (E) timepoints. The number of infants included in each time point is indicated. NS = not supplemented preterm, PS = probiotic-supplemented preterm.

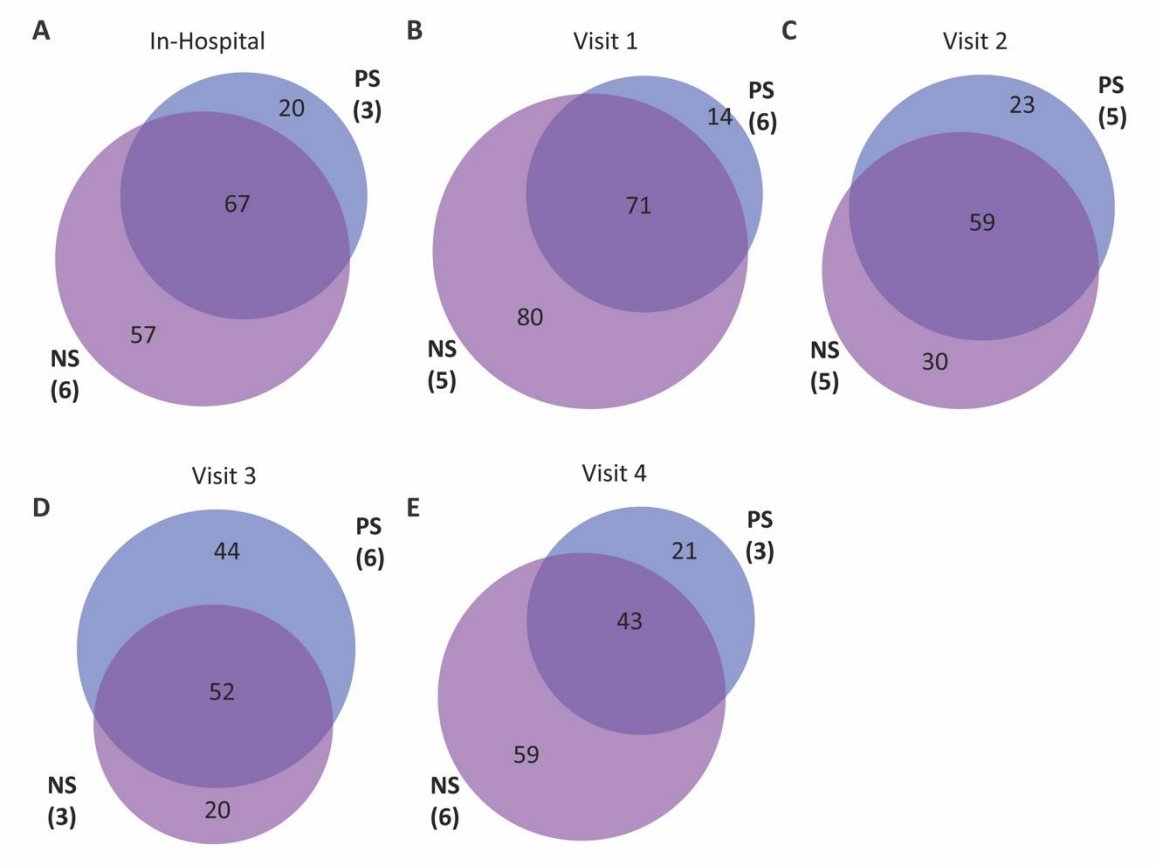


Figure S14: Unique genes in each infant group at various timepoints for the subset of preterm infants.

These gene counts are from the results of *de novo* assembly and ARG prediction using RGI *main* and CARD. Perfect and strict hits are reported. A) Data from the subset of preterm infants and all full-term infants at all time points. B) Data from the subset of preterm infants and all full-term infants at Visit 1. C-H) Data from the subset of preterm infants at all time points, in-hospital collection, Visit 1, Visit 2, Visit 3, and Visit 4. The number of infants included in each time point is indicated. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

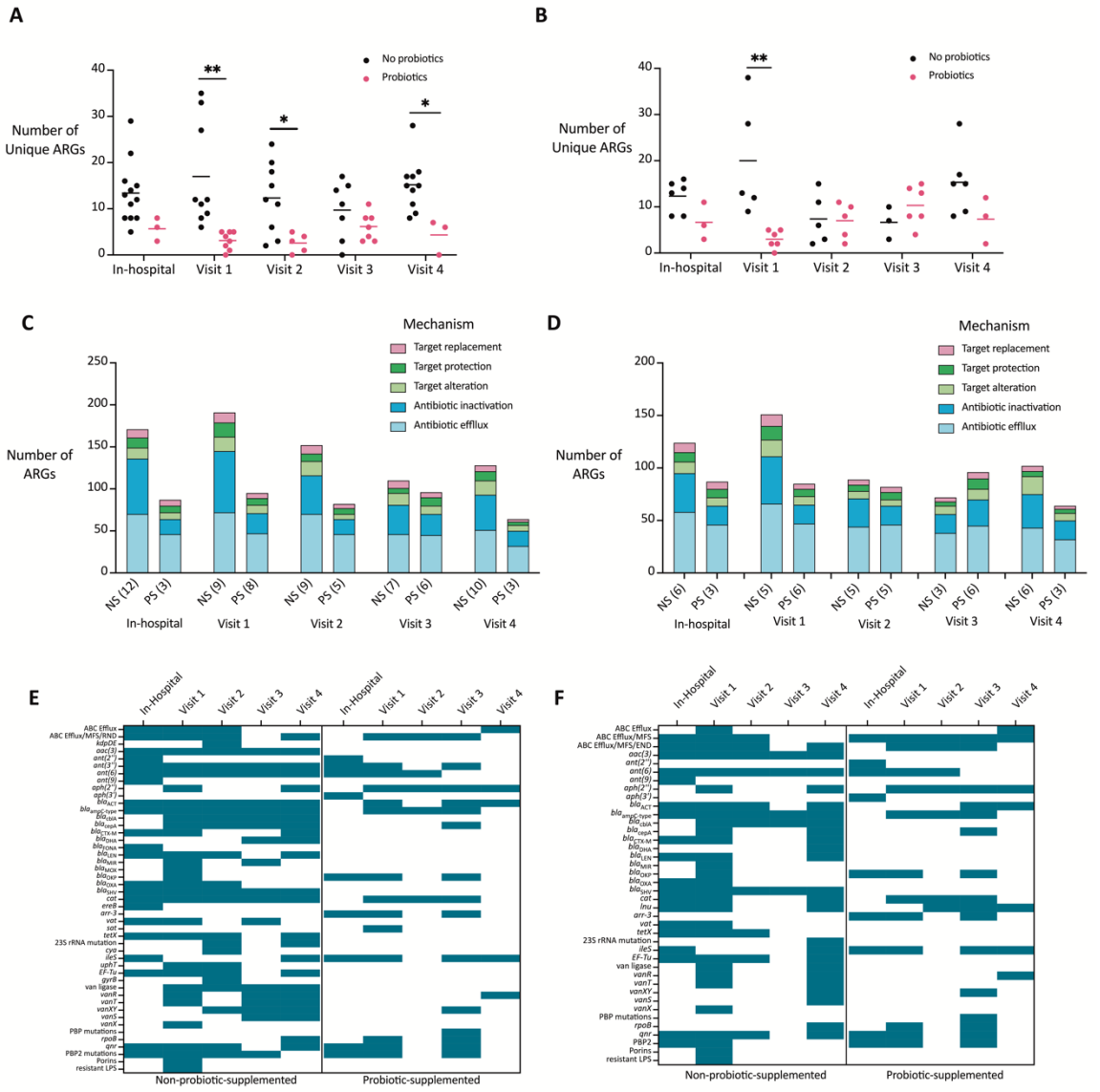


Figure S15: Unique ARGs, mechanisms, and families identified in preterm infants through RGI *main*.

These ARG counts are from the results of *de novo* assembly and ARG prediction using RGI *main* and CARD. Perfect and strict hits are reported. Figures A, C, and E represent the full set of preterm infants. Figures B, D, and F represent the subset of preterm infants. AB) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared. P = 0.0040 for Visit 1, P = 0.0180 for Visit 2, P = 0.0105 for Visit 4 for panel A. P = 0.0089 for Visit 1 in panel B. CD) The

number of ARGs identified in each infant group classified by resistance gene mechanism. The number of infant samples included in each is shown next to the sample type. NS = non-probiotic-supplemented preterm, PS = probiotic-supplemented preterm. EF) A selected subset of detected AMR gene families in preterm infants. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infants at that time-point.

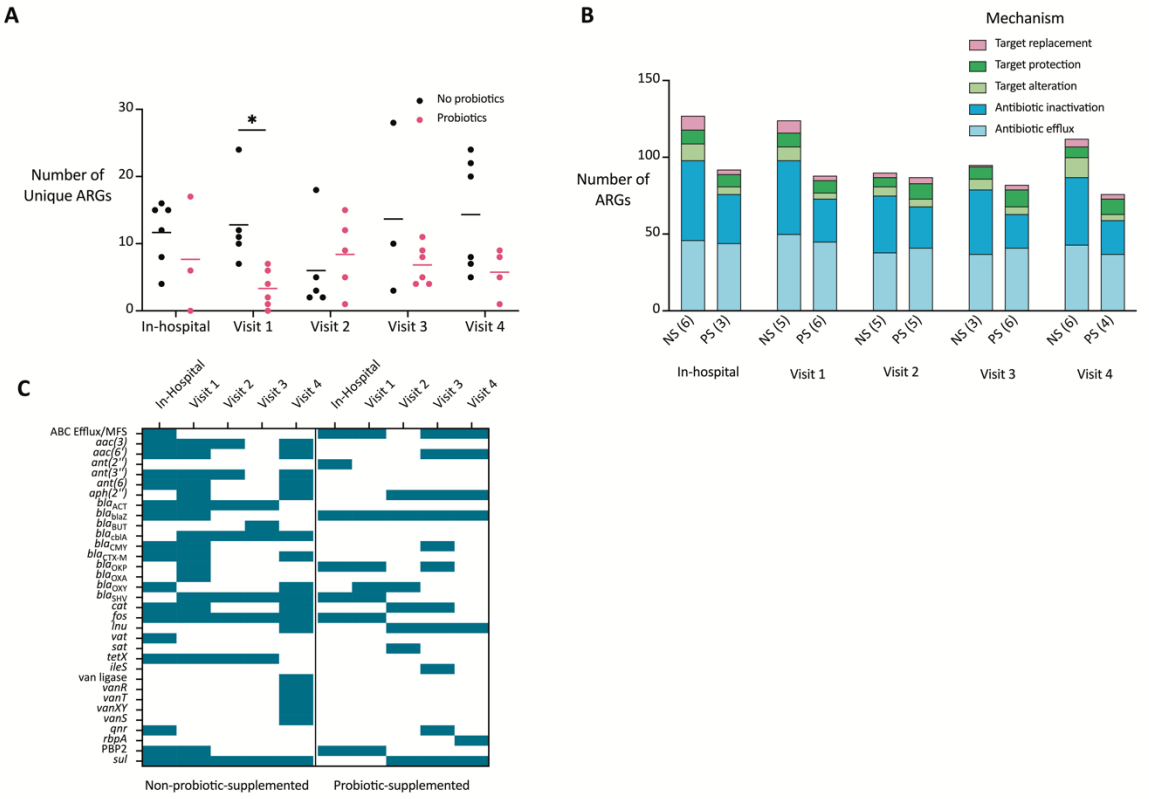


Figure S16: Unique ARGs, mechanisms, and families identified in the subset of preterm infants through RGI *bwt*.
Reads were mapped to CARD using RGI *bwt* and antibiotic resistance genes with at least 100 reads were reported. The data presented is from the subset of preterm infants at all visits. A) The number of unique ARGs identified in each infant. Significant differences are denoted by a line and asterisk(s) above the groups that were compared. P = 0.0102 for Visit 1. B) The number of ARGs identified in each infant group classified by resistance gene mechanism. The number of infant samples included in each is shown next to the sample type. NS = non-probiotic-supplemented preterm, PS = probiotic-supplemented preterm. C) A selected subset of detected AMR gene families in preterm infants. A teal box indicates that at least one gene from that AMR gene family was identified in any of the infants at that time-point.

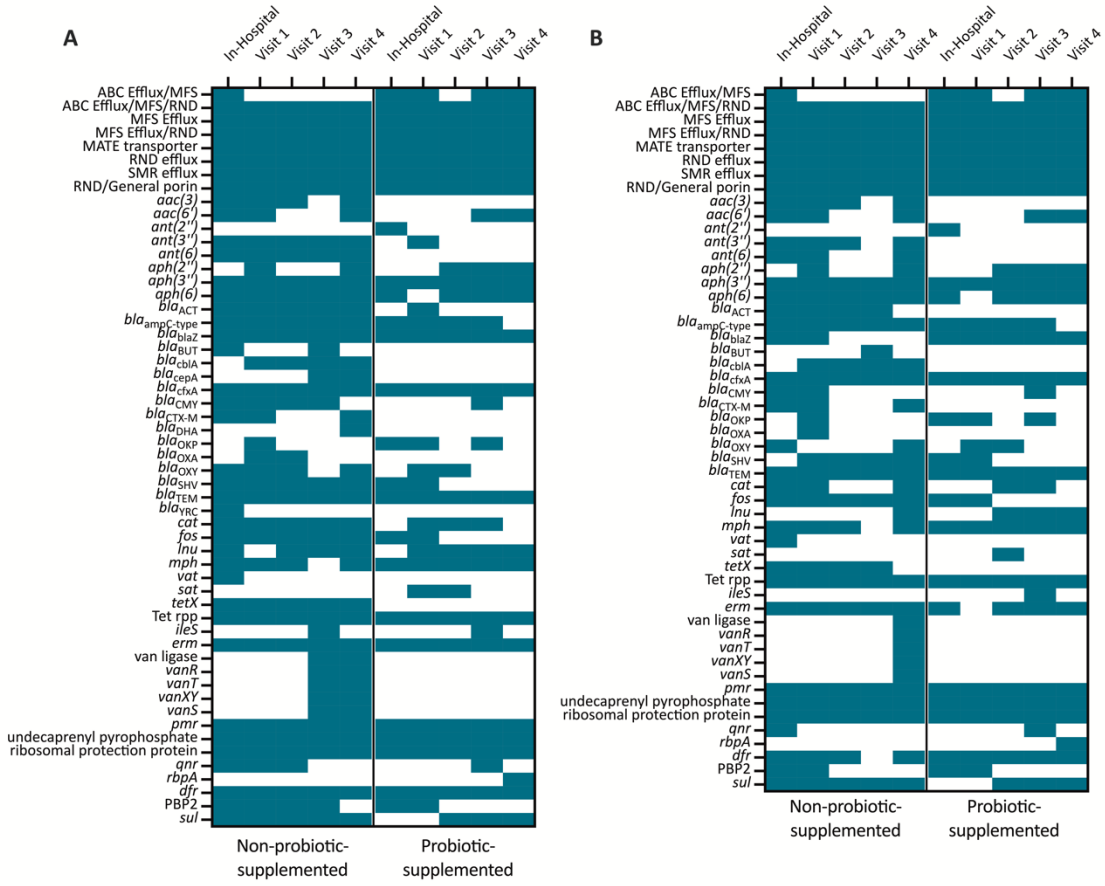


Figure S17: AMR gene families identified through RGI *bwt*.
A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples at that timepoint through read mapping to CARD. Each column is an individual study timepoint and the figure is divided between infants that received probiotics and those that did not. A) Results for the full set of preterm infants. B) Results for the subset of preterm infants.

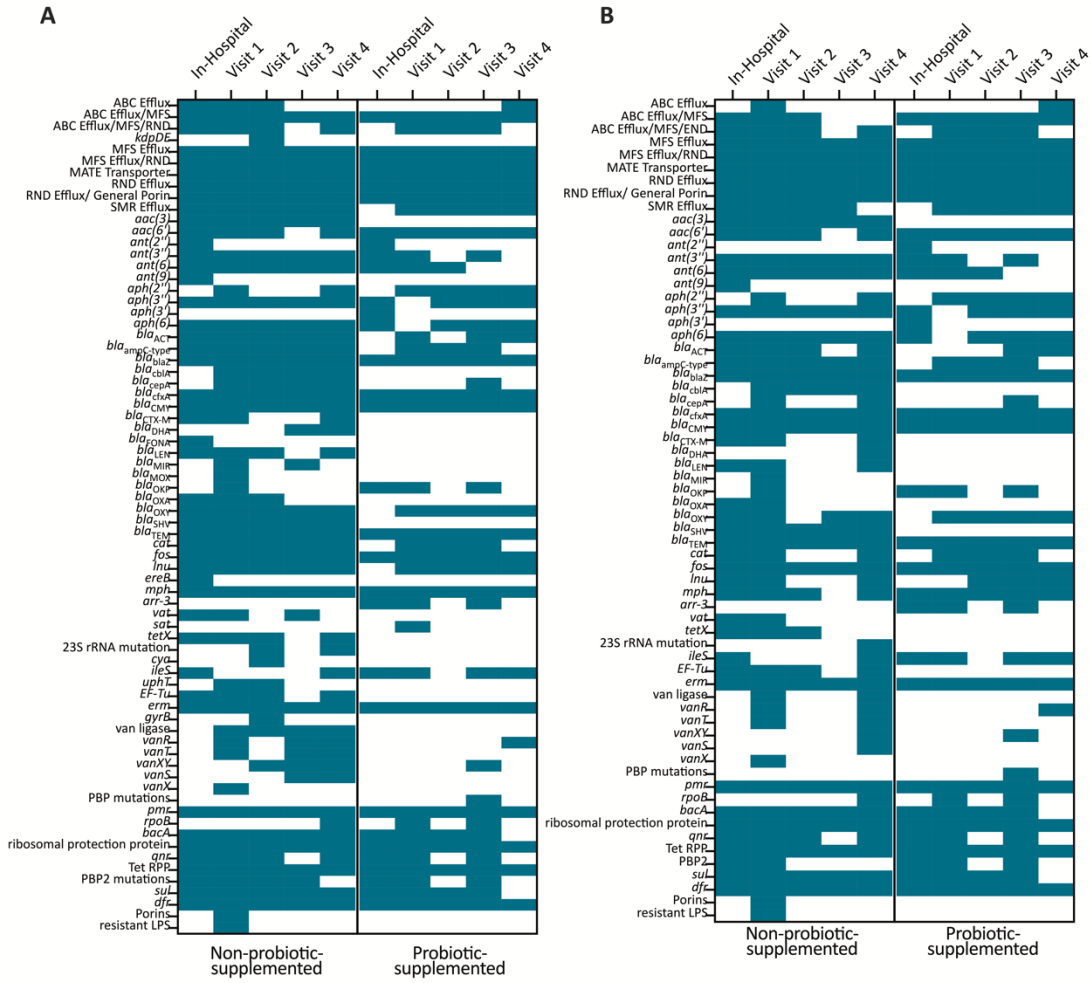


Figure S18: AMR gene families identified through RGI *main*.
A teal box indicates that at least one gene from that AMR gene family was identified in any of the infant samples at that timepoint through *de novo* assembly and prediction with RGI *main*. Each column is an individual study timepoint and the figure is divided between infants that received probiotics and those that did not. A) Results for the full set of preterm infants. B) Results for the subset of preterm infants.

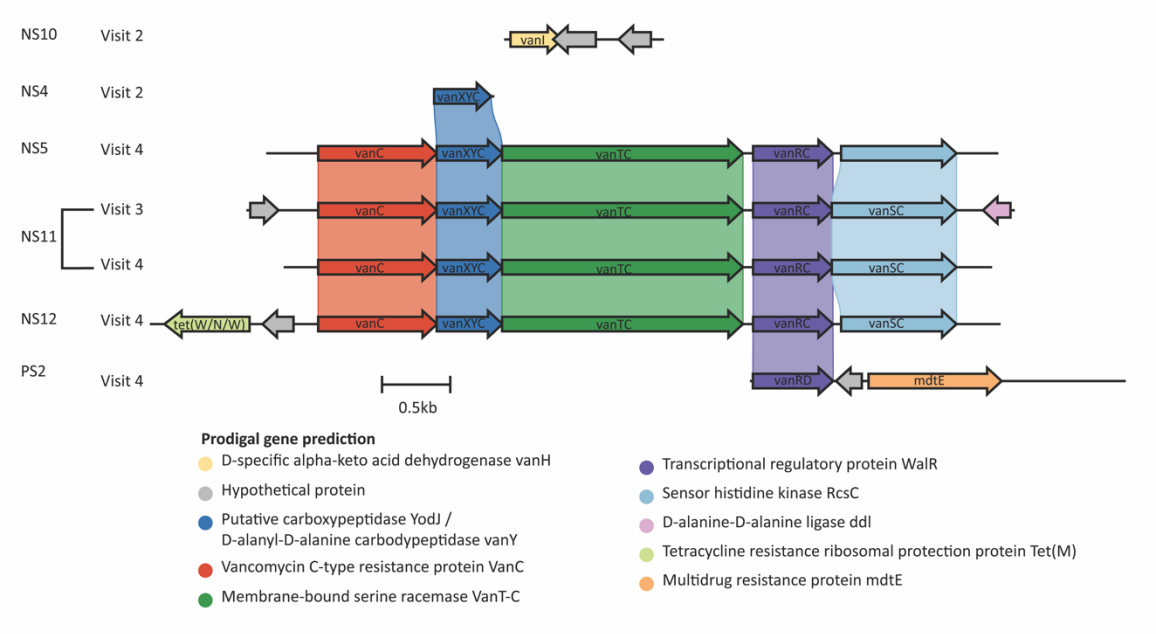


Figure S19: Genetic context of vancomycin resistance gene families detected in all infants.

From the *de novo* assembly, open reading frames were annotated using Prokka and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows and the RGI *main* predictions are labelled on each ORF. NS = not supplemented preterm and PS = probiotic-supplemented preterm.

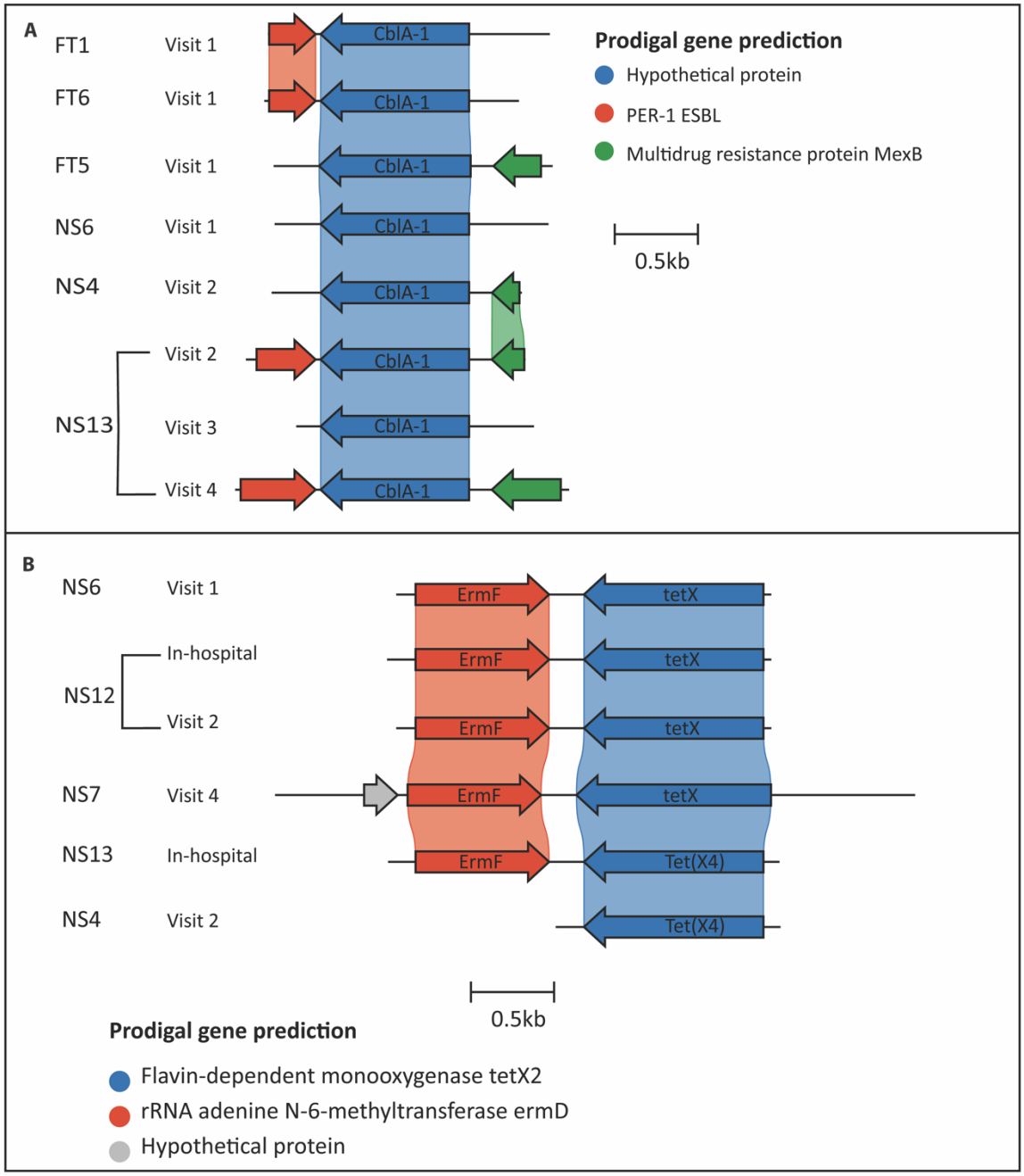


Figure S20: Genetic context of AMR families more prominent in NS infants.
From the *de novo* assembly, open reading frames were annotated using Prokka and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows and the RGI *main* predictions are labelled on each ORF. A) CblA beta-lactamase

family. B) tetracycline inactivation enzymes. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

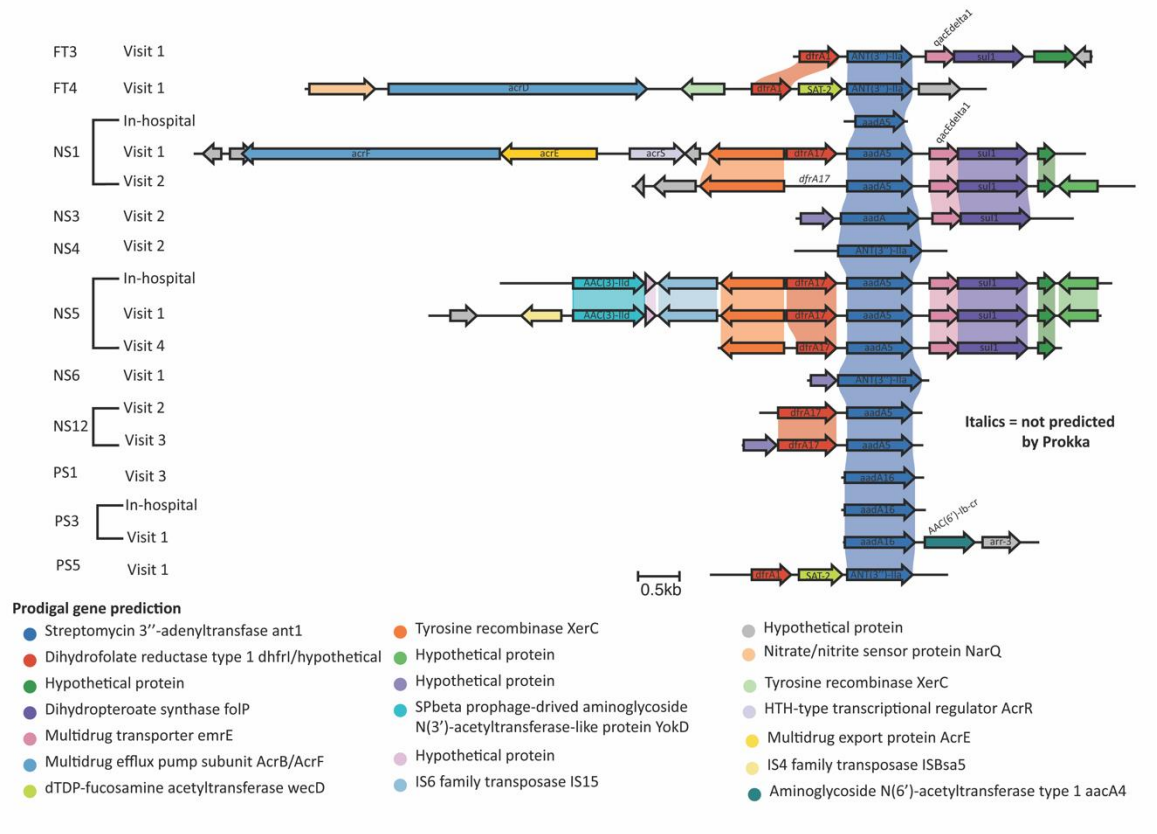


Figure S21: Genetic context of ANT(3'') resistance gene families detected in all infants.

From the *de novo* assembly, open reading frames were annotated using Prokka and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows and the RGI *main* predictions are labelled on each ORF. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

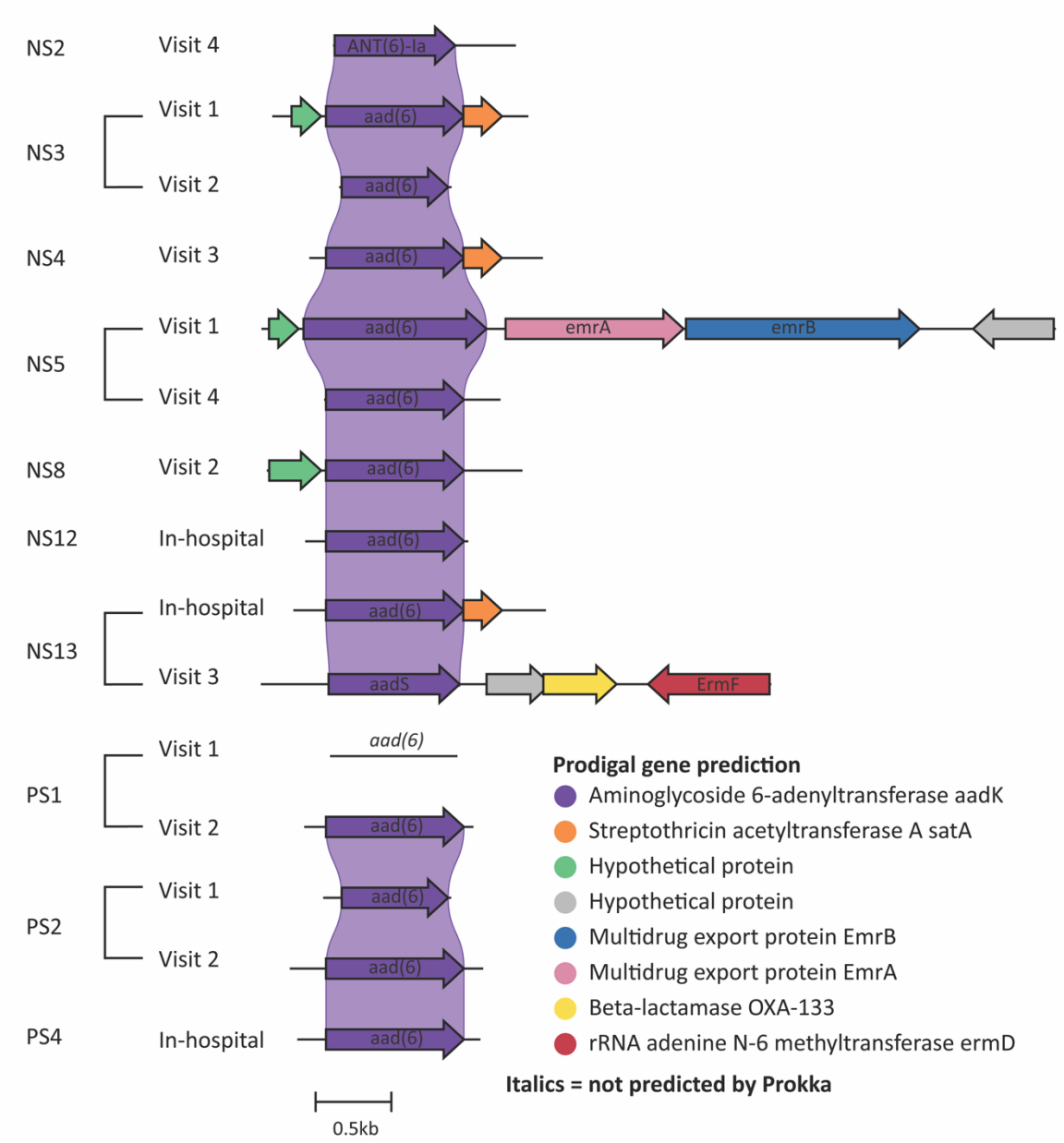


Figure S22: Genetic context of the ANT(6) gene family in all infants.
From the *de novo* assembly, open reading frames were annotated using Prokka and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows and the RGI *main* predictions are labelled on each ORF. NS = not supplemented preterm and PS = probiotic-supplemented preterm.

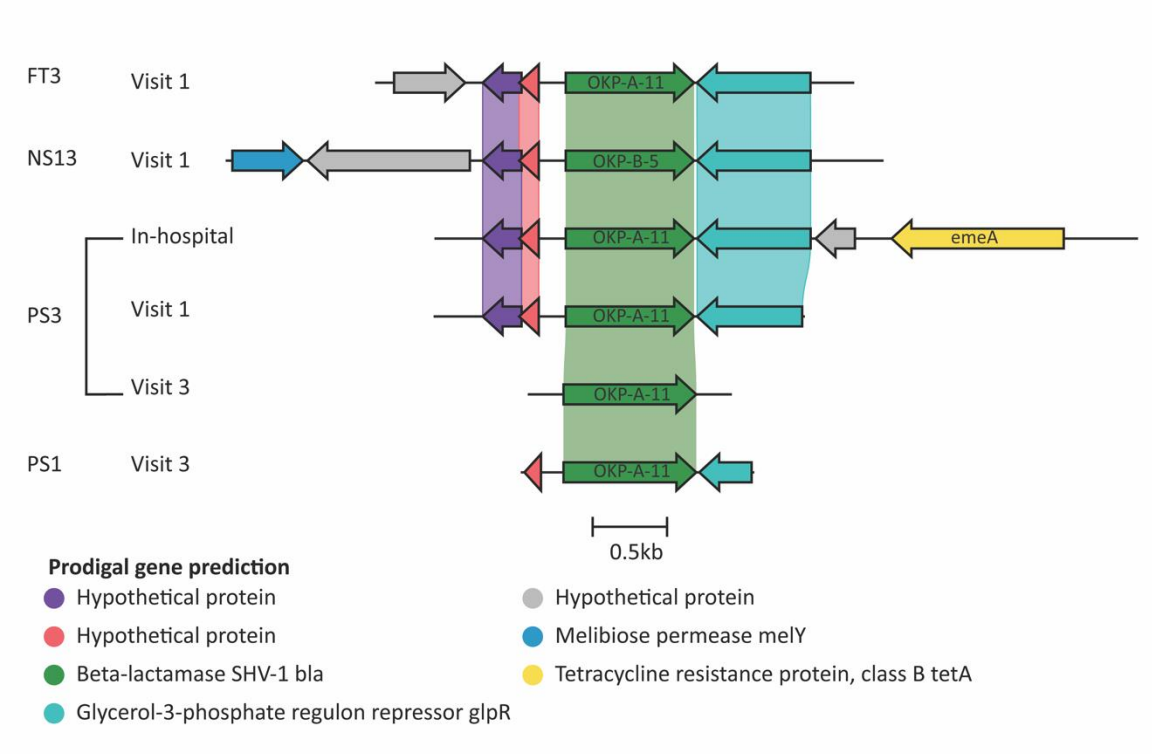


Figure S23: Genetic context of OKP beta-lactamases detected in all infants.
From the *de novo* assembly, open reading frames were annotated using Prokka and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows and the RGI *main* predictions are labelled on each ORF. NS = not supplemented preterm, PS = probiotic-supplemented preterm, and FT = full-term infants.

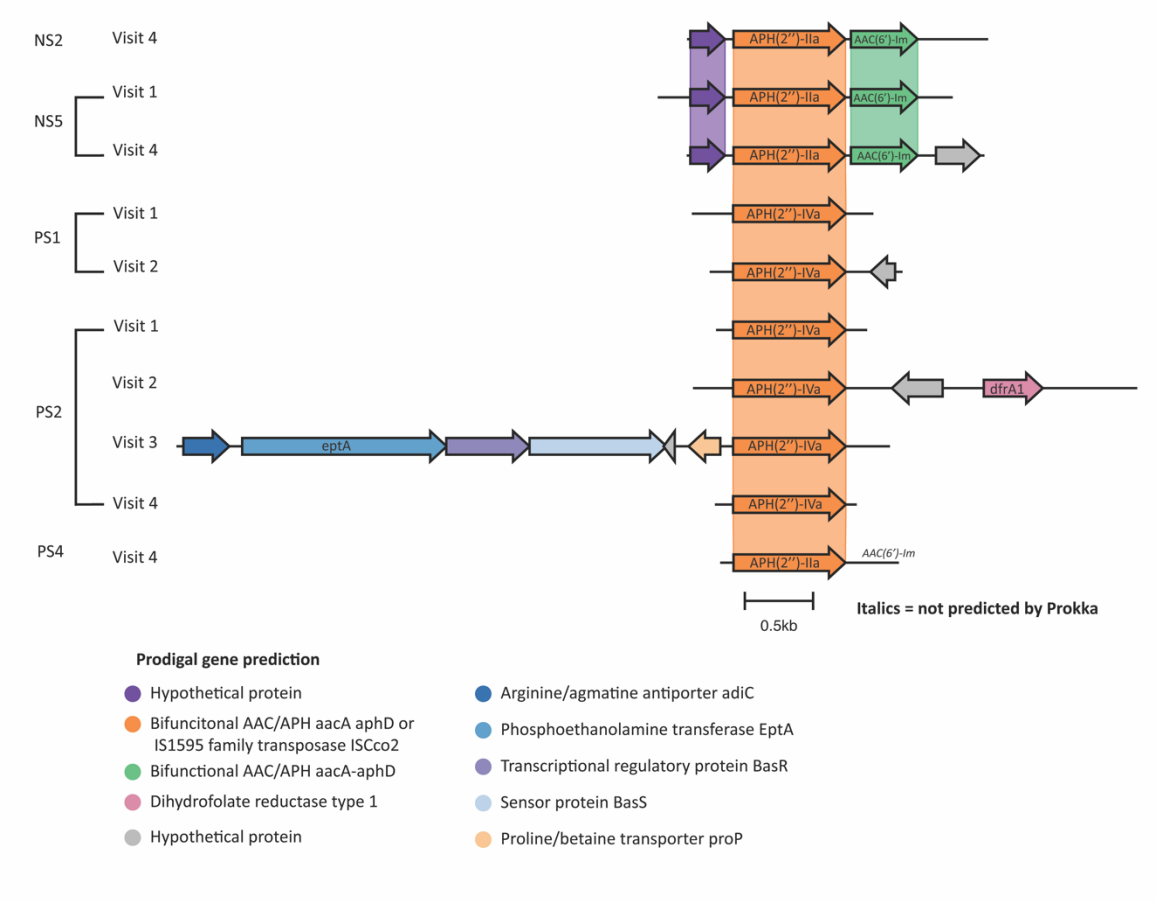


Figure S24: Genetic context of the APH(2'') gene family detected in all infants. From the *de novo* assembly, open reading frames were annotated using Prokka and resistance genes were predicted using RGI *main*. The Prokka annotations are the colored arrows and the RGI *main* predictions are labelled on each ORF. NS = not supplemented preterm and PS = probiotic-supplemented preterm.

Additional File S3 – Results of enrichment for ARGs – Details on library preparation, enrichment, sequencing, and analysis results for each sample (Not included in thesis)

Additional File S4 - Predicted bacterial host of ARGs – For a subset of ARGs, the bacterial host was predicted through various analysis approaches (Not included in thesis)

Additional File S5 - Negative control results – Results of negative controls included throughout the workflow and sequencing results for one sample (Not included in thesis)

CHAPTER FOUR: A three-day macrolide treatment of children with diarrhoea in Botswana has minimal effect on antibiotic resistance genes in the gut microbiome

CHAPTER FOUR PREFACE

The work presented in this chapter is in preparation as of December 14, 2022:

Guitor AK, Katyukhina A, Goldfarb DM, Wright GD, McArthur AG, Pernica JM. A three-day macrolide treatment of children with diarrhoea in Botswana has minimal effect on antibiotic resistance genes in the gut microbiome. *In preparation*.

The randomized control trial from which the samples were selected was led by DMG and JMP. AGM, JMP, DMG, GDW, and AKG conceived the study and designed the experiments. AKG and AK performed the DNA extraction, library preparation, and targeted enrichments. AKG analyzed the data, generated tables and figures, and wrote the manuscript with primary input from AGM, GDW, and JMP. The authors read and approved the final manuscript.

ABSTRACT

Mass distribution of the antibiotic azithromycin has been recommended to reduce under-five mortality rates in certain countries in Sub-Saharan Africa. Diarrhoea remains one of the leading causes of death for children in this region. While rehydration and zinc therapies are the recommended treatment, rapid diagnosis followed by targeted antibiotic therapy may prevent adverse outcomes of childhood diarrhoea. However, mass administration and imprudent prescription of antibiotics can select for antibiotic-resistant bacteria in the gut microbiota of children. The long-term implications of this selection are unknown and worrisome.

Our previous randomized control trial of children with diarrhoea in Botswana, Africa, evaluated the use of rapid diagnostics in guiding appropriate treatment. In these studies, children with rectal swabs positive for specific pathogens were treated with azithromycin daily for three days. A comparator arm treated children with the standard of care. Stool samples were collected at baseline and 60 days later. In this current study, DNA from 136 stool samples was enriched and sequenced to detect changes in the resistome, otherwise known as the collection of antibiotic resistance genes.

At baseline, the gut microbiota of these children contained a diverse complement of azithromycin resistance genes that increased in prevalence in both treatment groups by 60 days. Certain 23S rRNA methyltransferases were associated with other resistance genes and mobile genetic elements, highlighting the potential for the transfer of macrolide resistance in the gut microbiome. There were other minor changes in non-azithromycin

resistance genes; however, the trends were not specific to the antibiotic-treated children. In conclusion, a three-day azithromycin treatment for diarrhoea in one-year-old children in Botswana did not increase the prevalence of azithromycin-specific antibiotic resistance genes at 60 days. The gut microbiota of these children appeared primed for macrolide resistance, and repeated exposures may further select resistant bacteria. This suggests that the selection of antibiotic resistance in the short term should not threaten the overall benefits of mass administration of azithromycin in children in Sub-Saharan Africa.

INTRODUCTION

In 2020, the under-five mortality rate in Sub-Saharan Africa was 74 deaths per 1000 live births, totaling 2.7 million deaths, or 54% of the global value [1]. The World Health Organization's (WHO) proposed Sustainability Development Goal (SDG) aims to reduce the child under-five mortality rate to at least 25 deaths per 1000 live births worldwide by 2030 [1]. Infectious diseases are associated with almost half of the global under-five deaths, with malaria, lower respiratory infections, and diarrhoea among the leading causes [2]. Diarrhoea, or acute gastroenteritis, is the third greatest cause of disease burden in children under 10 years of age and is associated with other developmental challenges, including stunted linear growth [3-6]. Viruses (i.e., rotavirus), parasites (i.e., *Cryptosporidium* spp.), and bacteria (i.e., *Shigella* spp.) are the most common aetiological agents of this disease [4]. Despite decreases in childhood mortality and diarrhoeal-related deaths in the past 30 years due to various interventions, including vaccination, alternative approaches to reducing this burden are needed to attain the SDG within the next decade [1, 7, 8].

Azithromycin, a commonly used macrolide antibiotic, targets many bacterial pathogens associated with diarrhoea and respiratory infections [9]. This antibiotic also targets the bacterial-like ribosome of the apicoplast, a vestigial plastid-like organelle, rendering it effective against malarial parasites [10]. Campaigns to control *Chlamydia trachomatis* through mass drug administration (MDA) of azithromycin relieved the burden of trachoma and consequently reduced all-cause mortality [11-13]. Given these benefits, the MDA of azithromycin has been recommended in certain countries with high under-five mortality rates to improve overall child survival [14]. Trials in Niger, Ethiopia, Malawi, and Tanzania typically relied on biannual azithromycin distribution and noted an overall reduction in mortality of 14.4% [15-17].

There are obvious concerns with the MDA of an antibiotic, including unwanted impacts on the gut microbiota, selection for antibiotic-resistant organisms, and potential pressure for the mobilization and spread of antimicrobial resistance (AMR) [18-20]. In 2019, over 250,000 deaths in Sub-Saharan Africa were attributable to AMR, with about half occurring in children under the age of five [21]. Macrolide antibiotics target the 23S rRNA of the 50S large ribosomal subunit of bacteria and interfere with protein translation [9]. Resistance typically arises through mutations to the target, methylation of the 23S rRNA through Erm methyltransferases, efflux, or modification of the antibiotic [22]. Bacteria encode variable copies of the 23S rRNA locus *rrn*; therefore, the extent of resistance and the relevance of mutations depends on how many copies are present and modified [22]. While many studies have assessed the selection for AMR in cultured

pathogens, including *Streptococcus pneumoniae* and *Escherichia coli*, few have investigated the impacts on the gut microbiota and total resistome (i.e., collection of all antibiotic resistance genes (ARGs)) [18, 23-27]. It is imperative to weigh the benefits of MDA of azithromycin against potential impacts on selection for AMR in the long term.

Our previous trial of children with non-bloody diarrhoeal disease in southern Botswana, Africa, compared a rapid test-and-treat strategy against the standard of care, which consisted of oral rehydration and zinc therapy [28, 29]. While the latter has been an effective approach to reducing the duration of diarrhoea and fatalities, rapid diagnostics and targeted treatment may further improve outcomes [5, 23, 29-34]. As probiotics have been suggested to alleviate the burden of diarrhoea, this trial also included a *Lactobacillus reuteri* intervention [35-37]. We collected rectal swabs and bulk stool from children around 1 year of age. Children in the Rapid Test-and-Treat arm positive for specific infectious agents through a rapid diagnostic were treated with azithromycin (10 mg/kg/day) by mouth once daily for three days. The Standard Care group was not tested and did not receive azithromycin. Follow-up stool samples were collected from both groups 60 days later. We hypothesized that azithromycin exposure would select for an increase in the prevalence and abundance of macrolide ARGs in the gut microbiome of these children compared to those that received the standard care. Metagenomic DNA from these stool samples was assayed with a targeted-capture method to selectively sequence ARGs. The azithromycin-specific and total gut resistome of these children were compared at baseline and 60 days to identify potential consequences of antibiotic exposure.

METHODS

Study Design and Sample collection

A full description of the multicentre, randomised, controlled trial was described previously [28, 29]. Children with acute diarrhoea were randomised to either a Rapid Test-and-Treat (RTT) or Standard Care (SC) arm. Each group was further allocated to receive a probiotic supplementation of *L. reuteri* DSM 17938 ($1x10^8$ colony forming units by mouth) once daily within 24 hours of enrolment or a placebo [37]. Caregivers were asked to administer the probiotic or placebo for 60 days until follow-up. In the RTT arm, rectal swabs were assayed with the BioFire FilmArray GI Panel [38] or a PCR assay targeting a variety of viruses, parasites, and bacteria [39, 40]. Children with positive swabs for *Shigella*, enterotoxigenic *Escherichia coli* (ETEC), enteropathogenic *E. coli* (EPEC), *Campylobacter*, and/or *Vibrio cholerae* were treated orally with azithromycin (10 mg/kg once daily for three days), and those positive for *Cryptosporidium* were treated with nitazoxanide. Children in the SC arm received routine practice of fluid rehydration and zinc therapy. Bulk stool samples were collected at baseline (before treatment) and 60 days after enrolment and kept at -80°C until further processing.

Characteristics of participants chosen for this study

A subset of 68 children was chosen for this study. Thirty-four participants from the RTT arm were positive for either *Campylobacter*, *Shigella*, EPEC, ETEC, and/or *Cryptosporidium* and received azithromycin daily for 3 days. Thirty-four children from the

SC arm were not tested and did not receive azithromycin. Sixteen children in each arm of the study received the probiotic supplementation.

DNA extraction and library preparation

DNA was extracted from 0.1 – 0.2 g of stool as previously described [41, 42]. This method involved mechanical lysis with 2.8 mm ceramic beads, enzymatic lysis with lysozyme, proteinase K, and RNase A, followed by phenol-chloroform extraction and purification using the Zymo DNA Clean and Concentrator 25-kit. The DNA was quantified using a NanoDrop 2000c Spectrophotometer (Thermo Scientific, Mississauga, ON Canada) and Qubit 1X dsDNA high sensitivity assay, and the quality of the DNA extract was assessed via agarose gel electrophoresis. When available, up to 500 ng of dsDNA was used for input into library preparation with the NEBNext Ultra II dsDNA library kit (Additional File 1). Most libraries received 5 rounds of indexing PCR (Additional File 1).

Library enrichment and sequencing

Enrichment for ARGs was performed as previously described [43, 44]. To normalize the DNA input for in-solution targeted capture, a High Sensitivity DNA ScreenTape Analysis (Agilent Technologies) was performed to estimate the concentration of each library (Additional File 1). All samples were enriched for 24 hours at 65°C using a previously described probe set to target over 2,000 ARGs [43]. Enriched libraries were quantified by quantitative PCR (qPCR), pooled, then sequenced by the Farncombe Metagenomics sequencing facility at McMaster University on an Illumina MiSeq with 2 x 300 bp sequencing chemistry to a targeted depth of 250,000 clusters per library.

Analysis of captured antibiotic resistance genes

Demultiplexed reads were trimmed using skewer v 0.2.2 and string deduplicated using bbtools (dedupe.sh) [45, 46]. Reads were subsampled to 150,000 clusters or 300,000 reads using seqtk (v1.3-r117-dirty) [47]. Reads were mapped to the Comprehensive Antibiotic Resistance Database (CARD) v 3.2.1 (4891 sequences) using the Resistance Gene Identifier (RGI)*bwt (v 5.2.0) with kma version 1.3.4 [48-50]. ARGs reported were filtered for those with at least 85% length coverage of the gene with a least 50 reads mapped or 100% length coverage of the gene with at least 10 reads mapped. Reads were *de novo* assembled using SPAdes v 3.13.0 (metaspades default option), and ARGs predicted using RGI*main v 5.2.0 with CARD v 3.2.1 (using BLAST and with the --exclude_nudge and --low-quality flags included) [51]. The potential bacterial hosts of ARGs were predicted using RGI's beta feature for the k-mer prediction of pathogen of origin (RGI*kmer_query) with the default 61-mer database.

We compared results between the two treatment groups at the ARG and the AMR gene family (AGF) levels, which is a higher classification of ARGs. For example, *ermF* (ARO: 3000498) and *ermG* (ARO: 3000522) are both members of the "Erm 23S ribosomal RNA methyltransferase" AGF in CARD. The prevalence of ARGs associated with macrolide resistance, as well as 163 non-macrolide ARGs that were present in at least 10% of children in either treatment group at either timepoint, was determined from the RGI*main results. The $\log_2$ fold change in prevalence of an ARG was calculated using: $\log_2$([prevalence at 60 days]/[prevalence at baseline]). The number of reads mapping to a

given ARG was used as an approximation for the abundance of that ARG, given that all samples were subsampled to the same depth (number of reads) and there were no significant differences in the percentage of reads mapping to CARD between groups and time points (Figure 1A). The abundance differences for ARGs associated with macrolide resistance and the top 100 ARGs based on the total number of reads mapping across all samples were determined from RGI*bwt results. The RGI*main and RGI*bwt results almost always reported the same AGFs; however, different variants may be reported at the ARG level, given the nature of both analysis approaches. For example, RGI*main reported TEM-1 and TEM-116, while RGI*bwt predominantly detected TEM-206 and TEM-104. The difference amongst these alleles is 1-5 SNPs over 681 bp, making RGI*bwt susceptible to the allele network problem during read alignment [52].

For the 23S rRNA methyltransferases that confer resistance to macrolides (*ermFGQTX*, AROs: 3000498, 3000522, 3000593, 3000595, 3000596) in which there were appreciable increases in prevalence or differences in abundance between groups and at 60 days, the prevalence of predicted bacterial hosts with the greatest number of k-mer hits from the RGI*kmer_query was compared across cohorts and timepoints. The surrounding genetic context of these methyltransferases, along with the macrolide phosphotransferases (*mphAC,* AROS:3000316, 3000319), was predicted using Prokka version 1.14.5, and the potential origin inferred from the hit with the greatest query coverage and highest percent identity from Nucleotide BLAST (blastN) results against the nonredundant nucleotide collection in NCBI (access Nov 29, 2022) [53-55]. Representative contigs showing the

diversity of genes surrounding these ARGs were visualized using clinker version 0.0.25 [56].

Negative controls

Ten negative controls consisting of a buffer-only extraction blank were included to account for potential contamination from reagents or the laboratory environment. These were processed in the same manner as the stool DNA extracts, except that 10 rounds of indexing PCR amplification were performed during library preparation due to their low-input (Additional File 1). After enrichment, the libraries were analyzed with a High Sensitivity DNA ScreenTape Analysis (Agilent Technologies) and quantified by qPCR. Five libraries with sufficient concentration were sequenced on an Illumina MiSeq with 2 x 300 bp chemistry separately from the other stool libraries. Sequencing data was analyzed in the same fashion as the stool samples. From the RGI*bwt read mapping, 46 ARGs were identified, while 122 ARGs were identified through *de novo* assembly and RGI*main (Additional Files 5, 6). The most likely source of these ARGs in the negative controls is cross-contamination during DNA extraction and library preparation. Given the variability in the resistome profiles of the negative controls, a contaminated reagent was not likely. To link the contamination seen in the negative controls with the stool samples, we inspected an alignment of contigs containing *emrB* (ARO: 3000074), a gene involved in an antibiotic efflux system present in most stool samples and all negative controls (results not included). Only one negative control (#3) shared 100% similarity for *emrB* with three other samples that were processed in the same batch and located in nearby wells on the 96-well plate

during library preparation. Given the negligible DNA concentration of the negative extractions as well as the additional rounds of amplification during library preparation, even a small amount of cross-contamination would appear enriched in these controls. The higher levels of endogenous DNA in the stool samples and fewer rounds of indexing amplification are not likely to capture the low level of cross-contamination. Finally, the ARGs of interest that we highlight in this study were of low frequency or not detected in the negative controls; therefore, we do not believe that a small amount of cross-contamination between samples impacted our results.

RESULTS

Patient characteristics

Our study focused on 68 children from the original randomized control trial assessing rapid diagnostics for non-bloody diarrhea in Botswana, Africa [28]. These children were, on average, 1 year of age when they presented to the clinic and were randomized to either the Rapid Test-and-Treat (N = 34) or Standard Care (N = 34) arm of the study (Table 1). This trial also included randomization to a daily probiotic supplementation or placebo. Approximately 50% of children in each initial arm of the study received the probiotic for 60 days. For the remainder of our study, we did not group children based on probiotic supplementation because there was no distinct clustering of the resistome based on this factor (Figures S3, S4). Almost 95% of tested children were positive for either EPEC or ETEC, and a small percentage (8/34) were positive for *Campylobacter* spp. Three children tested positive for *Cryptosporidium*, in addition to

another pathogen, and received nitazoxanide and azithromycin. Many children in both arms of the study received antibiotics before enrolment, including cefotaxime, amoxicillin, co-trimoxazole, and ampicillin (Additional File 1). One child in the SC group received erythromycin before enrolment. Previous antibiotic exposure was not considered a factor in our analysis.

Table 1: Participant characteristics of children included in the resistome study.

Factor	**Standard Care (n = 34)**	**Rapid Test-and-Treat (n = 34)**
Male, N (%)	17 (50.0%)	20 (58.8%)
Age (yrs) at baseline (average +/- SD)	0.988 +/- 0.736	0.995 +/- 0.505
Age (yrs) at 60 day sample (average +/- SD)	1.168 +/- 0.736	1.167 +/- 0.501
Difference between 60 day and baseline (yrs) (average +/- SD)	0.179 +/- 0.020	0.172 +/- 0.015
Probiotic supplementation, N (%)	16/34 (47.0%)	16 (47.0%)
Prior antibiotic exposure N (%)	14/34 (41.2%)	7 (20.6%)
Campylobacter positive N (%)	Not tested	8 (23.5%)
Shigella positive N (%)	Not tested	4 (11.8 %)
Cryptosporidium positive N (%)	Not tested	3 (8.8%)
Enterotoxigenic *E. coli* positive N (%)	Not tested	11 (32.4%)
Enteropathogenic *E. coli* positive N (%)	Not tested	25 (73.5%)
Other pathogen positive N (%)	Not tested	22 (64.7%)
Azithromycin exposure, N (%)	0 (0%)	34 (100%)

Capturing the resistome of children treated with azithromycin

Using a method to selectively target and sequence over 2,000 ARGs from metagenomic DNA extracted from stool, we detected, on average, between 78 to 82 ARGs per child through read mapping to CARD, and slightly fewer through the more stringent method of *de novo* assembly and RGI (67 to 75 ARGs on average per child) (Figure 1BD). Overall, we found no appreciable differences in the number of ARGs or AGFs between the two cohorts of children. Through the *de novo* assembly and RGI*main analysis, the number of ARGs per child increased on average by 5.6 genes for the SC cohort and 8.8 for the RTT cohort (Figure 1D). We found the increase in ARGs to be significant only in the azithromycin-treated group (Figure 1D). In general, the number of AGFs also increased from baseline to 60 days in both cohorts. We reported an average increase of 4.2 (RGI*bwt) or 4.3 (RGI*main) AGFs for the RTT children and 2.3 (RGI*bwt) or 2.0 (RGI*main) AGFs for the SC children. Again, we found this increase to be significant only in azithromycin-treated children (Figure 1CE). There were many ARGs that were unique to one group; however, these were typically rare ARGs in the entire group and only found in a few individuals (Figures S1AB, S2AB; Additional Files 2, 3). Despite the individual differences in AGFs identified in azithromycin-treated children by 60 days, there were very few families that were unique to one group, and these were often only identified in one or two children (Figures S1CD, S2CD; Additional Files 2, 3). We also assessed the beta-diversity of the resistome and did not identify any clustering between individuals associated with azithromycin exposure, probiotic supplementation, or age (Figures S3, S4). With the read

mapping results, there was a small resistome cluster at baseline that may be associated with the presence of *E. coli*, but this was not tested further (Figure S3B).

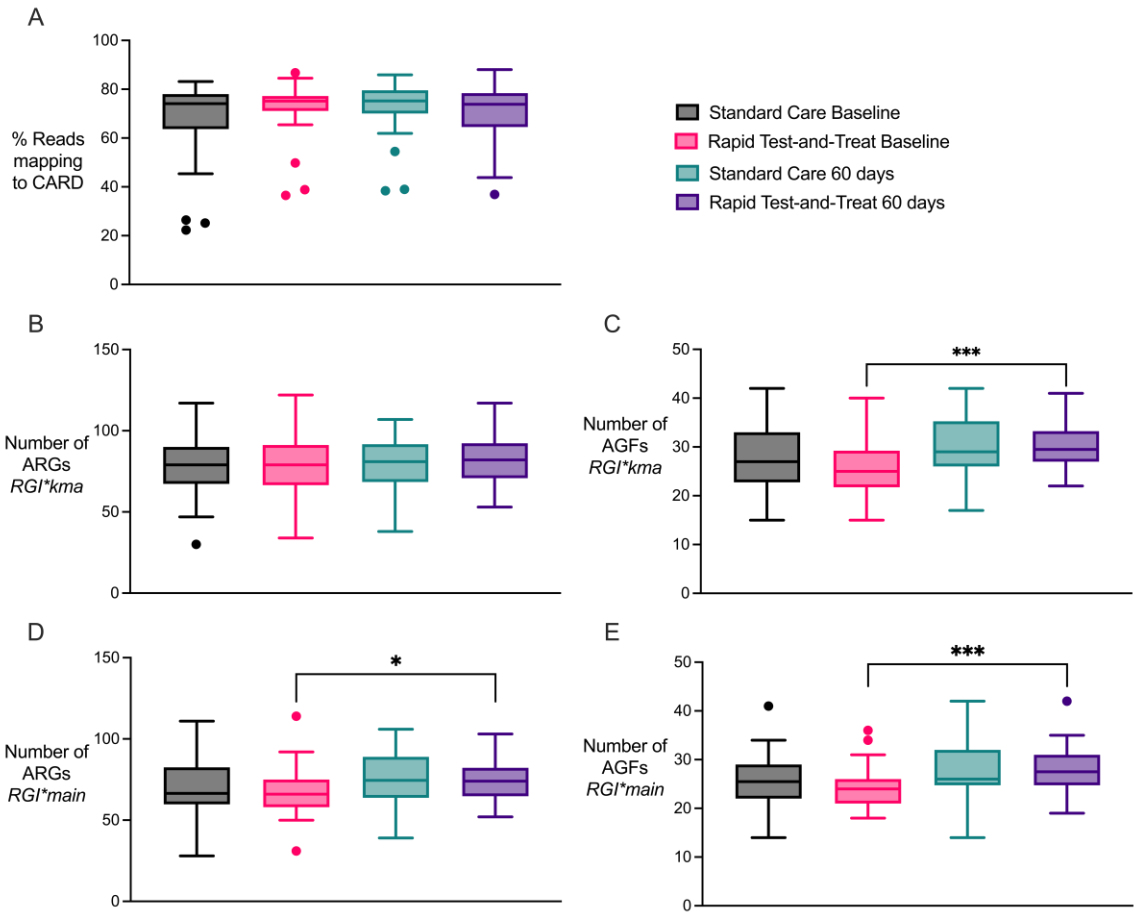


Figure 1: Capturing antibiotic resistance genes in the gut microbiome.
DNA from stool samples collected at Baseline and around 60 days later was enriched for antibiotic resistance genes prior to sequencing. A) Reads were mapped to CARD and the percentage of reads mapping to CARD in each child was determined. The number of ARGs (B) and the number of AGFs (C) per infant with at least 85% length coverage by at least 50 reads or 100% length coverage by at least 10 reads as determined through RGI*bwt. Paired t-test between Rapid Baseline and Rapid 60 days: $p < 0.001$. The number of ARGs (D) and the number of AGFs (E) per infant identified as Perfect or Strict hits as determined through *de novo* assembly and RGI*main. Paired t-test between Rapid Baseline and Rapid 60 days: $p = 0.013$ (D) and $p < 0.001$ (E).

Increase in the prevalence of macrolide resistance genes in both cohorts

At baseline, the macrolide resistome appeared similar across azithromycin-treated infants and those who received standard care (Figures 2A, S5). There were distinct changes in the prevalence and abundance of 23S ribosomal RNA methyltransferases that confer resistance to macrolide, lincosamide, and streptogramin antibiotics. In both cohorts, the prevalence of *ermQ*, *ermF*, *ermT*, *ermX*, and *ermG* increased at 60 days by 15% to 55% (RGI presence/absence results) (Figure 2). Many children maintained these genes from baseline to 60 days, while in some children, they appeared throughout the study period (Figure 2B). We found no significant differences in the prevalence of these genes between cohorts at 60 days (Figure 2A). Before antibiotic treatment, the children in the RTT arm had more reads mapping to *ermF* (ARO: 3000498) than those in the SC arm (Figure S6A). While both cohorts experienced an increase in reads mapping to *ermF* (average of 2419 for RTT and 2137 for SC), this gene remained more abundant at 60 days in the azithromycin-treated children (average of 4422 reads in RTT vs 2454 in SC) (Figure S6). At 60 days, on-average, more reads were mapped to *ermT* (ARO: 3000595) in the SC group (3614 for SC vs 1514 for RTT), which had increased from baseline (average reads of 73 for SC vs 1309 for RTT) (Figure S6). Whereas *ermX* (ARO: 3000596) read counts remained higher than the standard care children and increased in the azithromycin-treated children by 60 days (average of 5726 reads and an increase of 3939), this gene was slightly decreased (average of 79 reads) in SC children at 60 days (Figure S6).

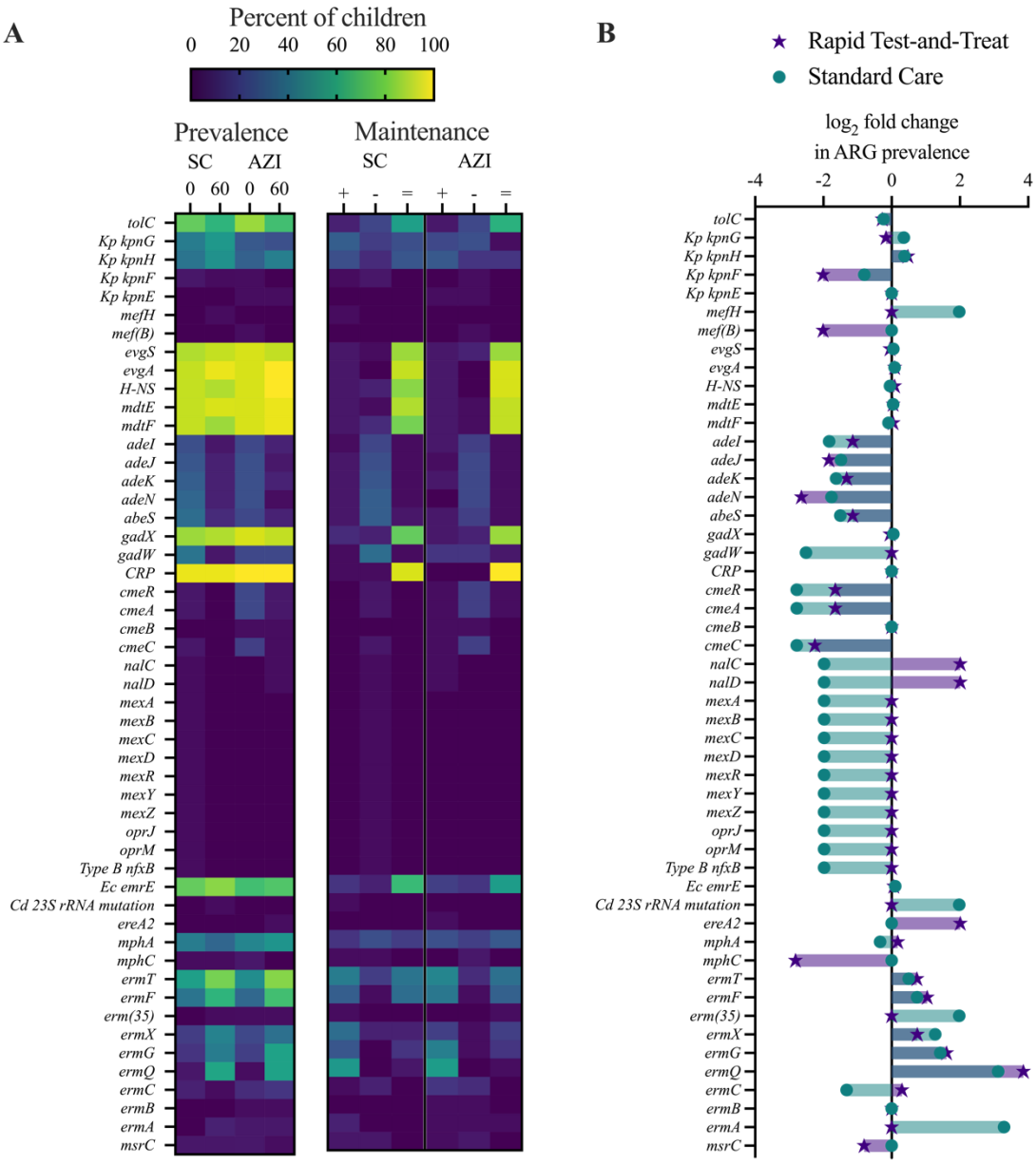


Figure 2: Prevalence of macrolide resistance genes in infants and changes at 60 days. A) Heatmap showing the prevalence and maintenance of macrolide ARGs in each group of treated children. In the left panel, the values correspond to the percentage of children in each group at each time point for which a given ARG was detected through *de novo* assembly and RGI*main analysis. The right panel shows the percentage of children in which a gene appeared (+), disappeared (-) or was maintained (=) by 60 days in each

treatment group. B) The $\log_2$ fold change in macrolide ARG prevalence by 60 days in both groups. SC = Standard Care, AZI = Rapid Test-and-Treat, Kp = *Klebsiella pneumoniae*, Ec = *Escherichia coli*, Cd = *Clostridioides difficile*.

In addition to the ribosomal methyltransferases, there were few changes in other genes implicated in macrolide resistance (Figures S5, S6). At baseline, the macrolide phosphotransferase *mphA* (ARO: 3000316) was more prevalent in the RTT arm (45%) than the SC arm (41%) and remained so at 60 days (52% for RTT vs 32% for SC) (Figure 2A). Before antibiotic treatment, the children in the RTT arm had more reads mapping to the protection protein *msrC* (ARO: 3002819) than those in the SC arm (Figure S5). By 60 days, *msrC* read counts were depleted in both groups but more considerably in the azithromycin-treated individuals (Figure S6B). Also, at this later timepoint, the number of reads mapping to the efflux-associated gene *mel* was slightly increased by an average of 248 reads in the RTT group and 523 reads in the SC group (Figures S5, S6). Genes from the *cme* efflux system found in *Campylobacter sp*. *jejuni*, including *cmeB*, were more prevalent and abundant in the RTT children at baseline but are depleted or disappeared by 60 days post-azithromycin treatment (Figure 2, Figures S5, S6). The children with these signatures were positive for *Campylobacter* at baseline. Finally, *ade* genes commonly associated with *Acinetobacter* species appear to be reduced in prevalence and disappear from the gut microbiome of children in both groups by 60 days (Figure 2).

Potential bacterial hosts of macrolide resistance genes

Given the rising frequency of 23S rRNA methyltransferases in both cohorts, we next predicted the bacterial hosts of these genes based on reported genetic contexts (Figure 3). The predicted hosts for a given ARG are similar between the azithromycin-treated children and those that received standard care. Most genes (*ermF*, *ermG*, and *ermX*) had only 1 – 2 predicted hosts at baseline, and these hosts either remained the same or 1 – 2 new hosts were detected by 60 days. *Bacteroides spp.* were common for both *ermF* and *ermG*, while *ermX* likely originated in *Bifidobacterium spp.* or *Corynebacterium spp*. (Additional File 4). Interestingly, in both groups of children at 60 days, *ermG* was predicted in the additional host *Klebsiella sp.*. A more diverse range of bacteria was detected for *ermT* including *Enterococcus faecium, Escherichia sp., Klebsiella sp., Staphylococcus sp*. and *Streptococcus suis*. The results between the k-mer prediction for pathogen-of-origin and the Nucleotide BLAST results were consistent across genes (Additional File 4). Additional potential hosts identified through the BLAST analysis included *Clostridiodes spp.* for *ermG*, and *Intestinibacter sp.*, *Peptacetobacter sp.*, or *Clostridium perfringens* for *ermQ* (Additional File 4). We further annotated the genetic context of these genes using Prokka and identified *ermTFX* alongside insertion sequences (IS) in some children (Figure 4; Additional File 4).

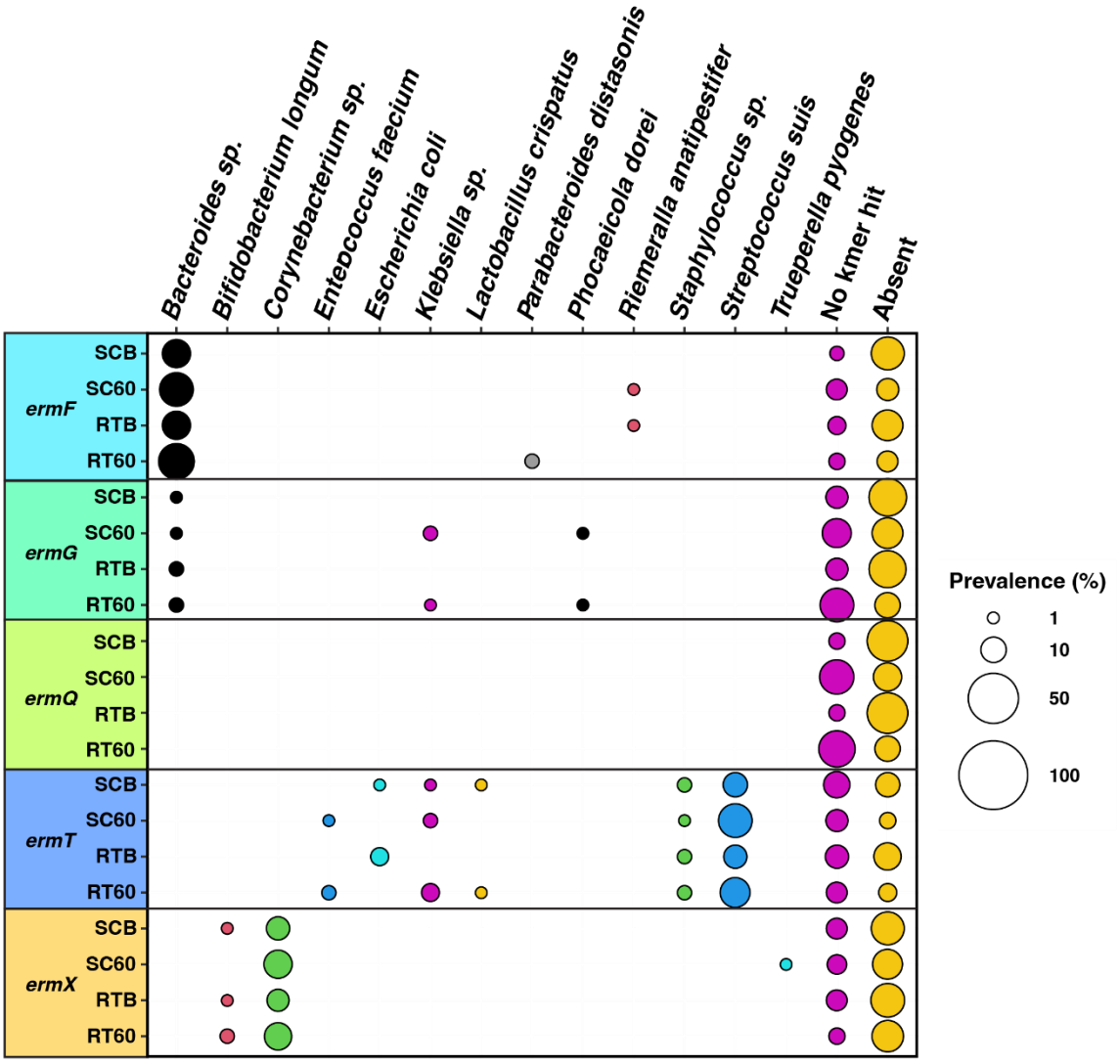


Figure 3: Predicted bacterial hosts of selected macrolide resistance genes in infants. Bacterial hosts were predicted through RGI*kmer_query from the *de novo* assembly and RGI*main results. Each bubble represents the percentage of children in each cohort at a given time point for which a specific host was predicted for a given ARG. "No kmer hit" signifies that the ARG did not have any significant kmer matches with the current database. "Absent" signifies the percentage of children in which the ARG gene not identified through the RGI*main analysis. Only the *erm* ARGs with distinct changes in prevalence or abundance are shown. SCB = Standard Care Baseline, SC60 = Standard Care 60 days, RTB = Rapid Test-and-Treat Baseline, RT60 = Rapid Test-and-Treat 60 days.

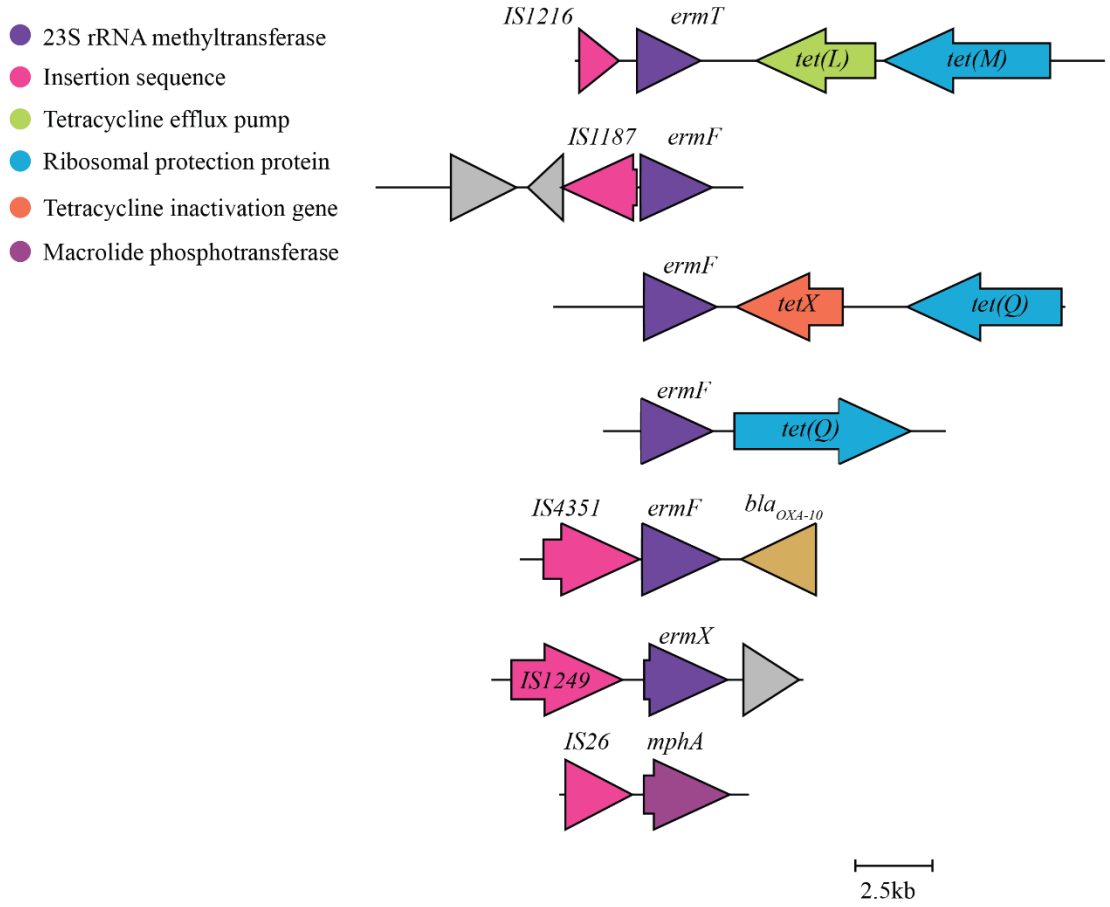


Figure 4: Representative genetic context of selected macrolide resistance genes in infants.
Contigs obtained through *de novo* assembly and predicted to contain various macrolide resistance genes were annotated using Prokka. Any ARGs present within the same contig were identified through RGI*main. While a variety of genetic contexts were observed in both the Standard Care and Rapid-Test-and-Treat cohorts, a representative contig that was found in both groups is shown here as an example. Other contexts are described in Additional File 4.

Changes in non-macrolide resistance genes

Next, we sought to compare differences in ARGs not associated with macrolide resistance (Figure S7; Figure 5). From baseline to 60 days, there is an increase in the prevalence and abundance of tetracycline-resistant ribosomal protection proteins (*tetO*, *tetQ*, *tetS*, *tetB*(P), *tet32*; ARO: 3000190, 3000191, 300192, 3000195, 3000196) in both groups (Figures 5, S7, S8AB). The average number of reads mapping per group at 60 days was higher for *tetQ* in the RTT group, while *tetO, tetS, tetB(P), and tet32* were higher in the SC children (Figure S8A). The number of reads mapping to an efflux pump associated with tetracycline resistance, *tet(40)* (ARO: 3000567), also increased in both groups over time, but was, on-average, higher at 60 days in children that received azithromycin (Figure S8A). The gene, *tetX* (ARO:3000205), encoding a tetracycline inactivating monooxygenase, also increased in prevalence and abundance in both groups by 60 days (from 15% to 38% in the SC group and 12% to 48% in the RTT children) (Figures 5, S7, S8AB). When we assessed the genetic context of the Erm methyltransferases, we found that *ermT* was commonly associated with *tetM* and tetracycline efflux pumps, including *tet(45) and tet(L)*. We also identified *ermF* on assembled DNA fragments containing *tetQ* and *tetX* (Figure 4).

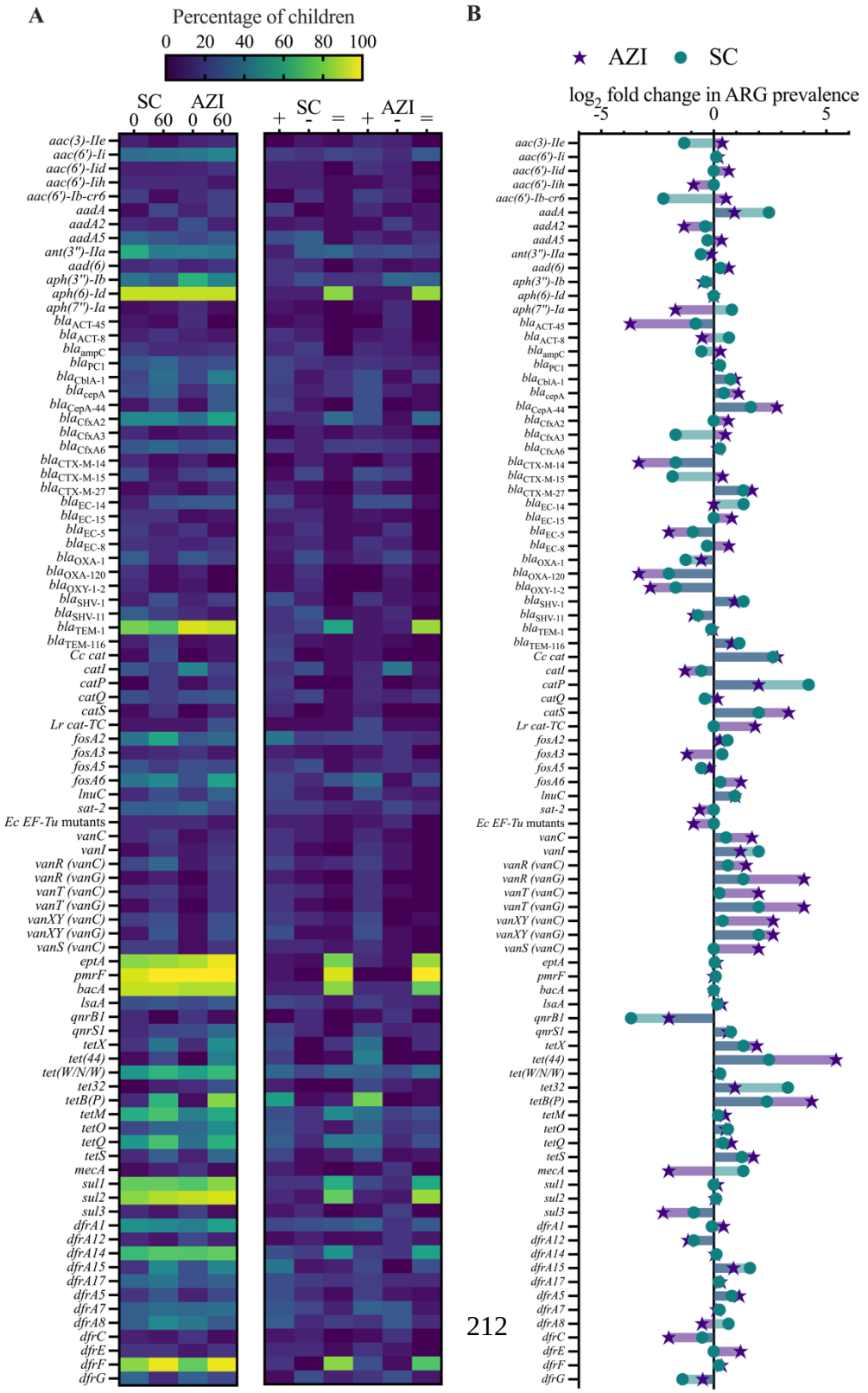

A
Percentage of children
0 20 40 60 80 100
SC AZI SC AZI
0 60 0 60 + - = + - =
B
AZI SC
log2 fold change in ARG prevalence
-5 0 5
aac(3)-IIe
aac(6')-Ii
aac(6')-Iid
aac(6')-Iih
aac(6')-Ib-cr6
aadA
aadA2
aadA5
ant(3'')-IIa
aad(6)
aph(3'')-Ib
aph(6)-Id
aph(7'')-Ia
blaACT-45
blaACT-8
blaampC
blaPC1
blaCblA-1
blacepA
blaCepA-44
blaCfxA2
blaCfxA3
blaCfxA6
blaCTX-M-14
blaCTX-M-15
blaCTX-M-27
blaEC-14
blaEC-15
blaEC-5
blaEC-8
blaOXA-1
blaOXA-120
blaOXY-1-2
blaSHV-1
blaSHV-11
blaTEM-1
blaTEM-116
Cc cat
catI
catP
catQ
catS
Lr cat-TC
fosA2
fosA3
fosA5
fosA6
lnuC
sat-2
Ec EF-Tu mutants
vanC
vanI
vanR (vanC)
vanR (vanG)
vanT (vanC)
vanT (vanG)
vanXY (vanC)
vanXY (vanG)
vanS (vanC)
eptA
pmrF
bacA
lsaA
qnrB1
qnrS1
tetX
tet(44)
tet(W/N/W)
tet32
tetB(P)
tetM
tetO
tetQ
tetS
mecA
sul1
sul2
sul3
dfrA1
dfrA12
dfrA14
dfrA15
dfrA17
dfrA5
dfrA7
dfrA8
dfrC
dfrE
dfrF
dfrG

Figure 5: Prevalence of ARGs present in at least 10% of infants in at least one time point and changes at 60 days.
A) Heatmap showing the prevalence and maintenance of ARGs in each group of treated children. In the left panel, the values correspond to the percentage of children in each group at each time point for which a given ARG was detected through *de novo* assembly and RGI*main analysis. The right panel shows the percentage of children in which a gene appeared (+), disappeared (-) or was maintained (=) by 60 days in each treatment group. B) The $\log_2$ fold change in macrolide ARG prevalence by 60 days in both groups. ARGs classified as Efflux are not shown. Macrolide ARGs were omitted because they are included in Figure 2. SC = Standard Care, AZI = Rapid Test-and-Treat, Ec = *Escherichia coli*, Lr = *Limosilactobacillus reuteri*, Cc = *Campylobacter coli*.

In general, the gut microbiome of children in both cohorts acquired trimethoprim-resistant dihydrofolate reductase (*dfr*) genes by the 60-day follow-up (Figure 5). At baseline, *dfrF* (ARO: 3002867) and *dfrA15* (ARO: 3003013) had more reads on-average mapping in the RTT group compared to the SC group (Figures S7, S8A). The abundance of *dfrF* increased at 60 days in both groups (Figure S8A). Finally, *dfrA14* (ARO: 3002859) decreased in abundance by an average of 841 reads in the RTT and 1108 reads in the SC group by 60 days (Figures S8AB).

After 60 days, more azithromycin-treated children acquired the beta-lactamase bla_{cfxA3}, and in both cohorts there was a slight increase in the number of reads mapping to members of the bla_{cfxA} family in general (Figures 5, S8AB). Although we did not find any significant differences in the average number of reads mapping to members of the following AGFs, they appear to change in prevalence over time (Figures 5, S7): vancomycin resistance genes belonging to the *vanG* and *vanC* clusters, and the *cepA* beta-lactamases appear in the gut microbiome of more children in both groups by 60 days, whereas the

prevalence of genes belonging to the *bla*$_{OXA}$ gene family was reduced in the gut microbiome of children in both groups by 60 days (Figure 5).

Does azithromycin treatment reduce the burden of diarrhea-causing pathogens?

As mentioned above, although ARGs associated with *Campylobacter* were more prevalent in children in the RTT arm, by 60 days, the prevalence of these ARGs, and perhaps by association this pathogen, was reduced to similar levels of the SC group, likely due to the azithromycin treatment. Over 90% of tested children were positive for either EPEC, ETEC, *Shigella spp.*, or a combination of pathogens, including *E. coli* (Table 1). Many genes that were detected at the baseline sampling in both cohorts can be associated with *E. coli* (Figures 5, S7, S8CD). In general, genes intrinsic to *E. coli* and associated with antibiotic efflux including *mdtEFOP*, *acrDEF*, *emrKY*, *cpxA*, *mdfA*, and *evgS* remain stable in children over time and at a similar prevalence in both groups (prevalence data not shown). The number of reads mapping to these genes, however, decreased over time in both groups of children (Figures S7, S8CD). An aminoglycoside phosphotransferase, *aph*(6)-Id that is often associated with plasmids in *E. coli* [57], is prominent across both cohorts at both timepoints (Figure 5). However, reads mapping to this gene decreased slightly in both groups by 60 days (Figures S8CD). Finally, other ARGs that are often associated with *E. coli* and are reduced by 60 days in both groups include: *bla*$_{TEM-104}$, *bla*$_{TEM-206}$, *bla*$_{CTX-M-15}$, *eptA*, *pmrF*, and members of the ANT(3'') family (*aadA5*, *aadA24*, *ant(3'')*-IIa).

DISCUSSION

A concerning threat of the MDA of azithromycin and its use for the targeted treatment of diarrhoea in children in Sub-Saharan Africa is the selection for AMR. Our study is, to our knowledge, the first to comprehensively investigate the effect of azithromycin on the gut resistome of children with diarrhoea in Botswana, Africa. Although the children that received azithromycin experienced a greater increase in the total number of AGFs by 60 days, this was not associated with a difference in macrolide resistance genes when compared to the children that received the standard treatment for diarrhoea. In both groups, certain macrolide- and non-macrolide ARGs persisted or increased in prevalence regardless of antibiotic exposure. We also noted an overall decrease in the abundance of ARGs commonly associated with *E. coli* in both groups of children at the follow-up time point. Finally, certain macrolide resistance determinants were linked to insertion sequences and other ARGs highlighting the potential for dissemination and spread of AMR within the gut microbiome of these children. In this survey of children under five years of age, we revealed the diversity of the gut resistome and showed that a three-day exposure to azithromycin did not provide additional pressure to retain macrolide resistance genes two months later.

We did not find a significant increase in the total number of ARGs from baseline to 60 days for either group of infants. There was a general increase in the number of AGFs in both groups, although this was only significant in the azithromycin-treated children and may reflect small inter-individual differences. Given that these children presented to the clinic with acute gastroenteritis, their gut microbiota was likely dominated by the infection-

causing organism at this time [58, 59]. Over 90% of the tested children were infected by a pathogenic *E. coli*, either EPEC or ETEC, and many likely experienced co-infections. It is possible that by 60 days, the gut microbiota would have recovered from the infection or was in the process of being restored to a more diverse ecosystem where members of the genera *Bacteroides* and *Prevotella* are more abundant [58, 60]. Although we did not measure changes in the microbial diversity of these children from baseline to 60 days, the microbiota likely transformed dramatically throughout recovery [60, 61]. The increase in AGFs we observed may reflect an increase in commensal and other beneficial bacteria of the gut microbiota that encode a variety of intrinsic ARGs distinct from those of diarrhoeal-causing pathogens [62].

A disappearance of *Campylobacter*-associated ARGs and decrease in abundance of *E. coli*-associated ARGs by 60 days in the gut microbiota of these children was observed irrespective of the treatment. This suggests either equal efficacy of azithromycin and the standard therapy or self-limitation and natural recovery of the gut microbiota over time. However, *E. coli*-associated ARGs were not completely depleted and remained prevalent at 60 days in both groups of children. This ubiquitous organism is a common member of the gut microbiota of children at one year of age [59]. Commensal strains of *E. coli* likely share similar ARG profiles as pathogenic ones making their distinction difficult through our ARG-targeted approach. It is also possible that the children continued to carry and shed the diarrhoeal-causing agent 60 days post-infection [63].

With a focus on macrolide resistance genes, we identified a diversity of Erm 23S rRNA methyltransferases in both groups of children. At baseline, these genes were similar in prevalence in both groups of children and increased by 60 days, irrespective of the treatment received. Data on the prevalence of macrolide resistance in children in Botswana is limited, however, a recent study of wastewater found high levels of resistant bacteria in the environment [64]. Although the abundance of certain *erm* genes may have changed more dramatically in one treatment group compared to the other, there is large individual variability within each group that is likely skewing the results. By analyzing the genetic surroundings of these *erm* genes, we inferred they were likely associated with *Bacteroides*, *Bifidobacteria*, *Corynebacteria*, *Peptacetobacter*, *Enterococci*, or *Streptococci*. Bacteria belonging to these genera are often found in the gut microbiota and are associated with the general restoration of the microbiome after a diarrhoeal infection [60, 61].

In addition to the high prevalence of the Erm 23S rRNA methyltransferases upon arrival at the clinic, macrolide phosphotransferases were also prevalent in these children. The gene, *mphA*, slightly decreased in prevalence in the SC group of children and increased in the azithromycin-treated children; it was also more prevalent in this latter group at baseline. Therefore, it is difficult to associate the increase in prevalence with azithromycin treatment alone. What is concerning, however, is the association of this macrolide phosphotransferase with IS elements. Plasmids containing *mphA* have been identified in ETEC isolates and are common in bacteria circulating in wastewater and river environments in Botswana [64, 65]. Finally, many of the *erm* genes identified in this group

of children were commonly associated with insertion sequences and mobile elements. Although these genes are often found in commensal gut bacteria, their ability to transfer within genera to potential pathogenic strains is worrisome. For example, enterotoxigenic *Bacteroides fragilis* is an opportunistic human pathogen that has been associated with diarrhoea [66]. This diverse macrolide resistome and its association with IS elements might put these young children at future risk for the transfer of ARGs within their gut microbiota and the persistence of resistant strains if they are repeatedly exposed to macrolides.

Finally, tetracycline resistance genes were common in both groups of children and increased in prevalence by 60 days. When we analyzed the *erm* genes, we found many were associated with tetracycline efflux pumps and tetracycline resistance ribosomal protection proteins, likely explaining the increased prevalence of both these ARGs by 60 days. Tetracycline resistance is common in the gut microbiome of children and even in adults [67]. This association between these two genes and their linkage to a common bacterial host highlights the potential for co-selection of tetracycline and macrolide resistance, given any future oral use of these antibiotics. A point prevalence survey in Botswana of antibiotic prescriptions revealed, however, that tetracyclines and macrolides are less frequently prescribed than other antibiotics such as cefotaxime and metronidazole [68].

Our study focused on the total gut resistome after azithromycin exposure in children with diarrhoea in Botswana, Africa. A similar study of a three-day dose of azithromycin for diarrhoea in children in 7 countries did not identify differences in antibiotic susceptibilities in bacterial isolates from participants that received the antibiotic compared

to those that did not [23]. In cultured isolates of *Streptococcus pneumoniae* from nasopharyngeal swabs and *E. coli*, resistance increased after azithromycin use for trachoma prevention [18, 69-75]. Others have found that macrolide resistance determinants, including *ermB* and *mefA,* increased in prevalence by up to 60% following azithromycin biannual distribution for trachoma [69]. Trials focused on using the MDA of azithromycin to reduce mortality in children have also assessed the impact of this repeated exposure on resistance [24-26]. Using metagenomics and culture, they noted longer-term impacts on the gut resistome and microbiome, including decreased microbial diversity, increased macrolide gene expression, and macrolide resistance in *S. pneumoniae* after 2 years of biannual azithromycin administration [26, 27]. Finally, after 4 years of biannual exposure, they observed the selection of non-macrolide resistance genes towards aminoglycosides, beta-lactams, trimethoprim, and metronidazole [25]. The disparities between our results and those of these trials on biannual distribution may relate to the resilience of the gut microbiome to short-term courses of antibiotics compared to repeated exposures [76].

Our study of children in Botswana with acute gastroenteritis used a targeted and rapid Test-and-Treat approach to reduce the selection for antibiotic resistance and off-target effects from unwarranted antibiotic exposure (e.g., if a virus was the causative agent). We used a unique targeted enrichment approach before sequencing to capture the majority of known elements of the antibiotic resistome with minimal sequencing depth and cost. Our sensitive method selectively enriches clinically relevant and potentially mobilized macrolide resistance genes. One limitation is the inability to capture resistance due to

mutations in target genes (i.e., 23S rRNA mutations). These mutations may be selected rapidly after azithromycin treatment and would be missed in our analysis. However, we used the increased depth of coverage of each ARG obtained through enrichment and various analysis tools to infer the most likely bacterial host. Another limitation is that we only assessed stool samples at baseline and 60 days and, therefore, may have missed changes in the gut resistome of these children immediately after azithromycin treatment and in the longer term. Our study and others highlight the importance of longitudinal sampling to fully characterize the dynamics and changes that occur in the gut resistome post-antibiotic exposure. Shotgun metagenomics is still a relatively expensive approach to these longitudinal studies. A targeted sequencing method, such as the one used in our study, could provide sensitive and high-quality resolution of the entire resistome in these larger sample sets.

Overall, we were surprised to identify such a diverse set of macrolide resistance genes in children at a young age and that these generally increased in prevalence in both cohorts regardless of the treatment they received. Bystander exposure to antibiotics or other unknown factors may have contributed to this result [77]. Given the extreme benefit of biannual MDA of azithromycin in reducing the under-five mortality rates in sub-Saharan Africa, there appears to be a limited impact of this antibiotic use in the short-term on the gut resistome of children. Many of the previous comprehensive resistome studies of the gut microbiome after azithromycin exposure relied on pooled samples for their analysis, which could be biased by a few individuals [25, 26]. Improved longitudinal studies monitoring

both the development of resistance due to mutations and the transfer of mobile macrolide resistance genes after antibiotic treatment, as well as increased general surveillance of resistance, are needed to fully understand the implications of MDA of azithromycin in children under five. While azithromycin is a relatively cheap and effective intervention to reduce childhood mortality, we believe the unknowns associated with selection for antibiotic resistance, the linkage of these resistance genes with mobile elements, and the early warnings from other trials on resistance, warrant investigating alternative interventions. The incidence of diarrhoea and other infectious diseases may be reduced through vaccination, access to safe drinking water, nutrient supplementation, improved sanitation, and access to quality health care [4, 8, 78-81]. These alternative interventions should be further explored to prevent the MDA of antibiotics as the longer-term impacts of antibiotic treatment on the resistome of children have yet to be fully explored.

CONCLUSIONS

In our survey of children with diarrhoea in Botswana, Africa, we did not identify any specific increases in acquired macrolide resistance determinants 60 days post-azithromycin exposure compared to children that received the standard treatment. Certain macrolide resistance genes increased in prevalence in both treatment groups at 60 days. These genes were predominantly associated with gut commensal organisms and found in proximity to tetracycline resistance genes and mobile genetic elements. The increase in these determinants may have resulted from a general recovery of the gut microbiota following acute gastroenteritis. Co-selection and the potential dissemination of these

macrolide resistance genes are still a concern. Although a short three-day course of azithromycin for diarrhoea did not have an appreciable selective effect on the resistome, repeated exposures to this antibiotic risk further selection in an environment that is primed with resistance. Alternative approaches to reducing under-five mortality in sub-Saharan Africa should be pursued to prevent additional selective pressure for macrolide resistance in these vulnerable populations.

AVAILABILITY OF DATA AND MATERIALS

Sequence data that support the finding of this study will be deposited in NCBI's Sequence Read Archive upon publication. Code used to analyze the data is available at: https://github.com/AllisonGuitor/AMR-metatools.

ABBREVIATIONS

23S rRNA: 23S ribosomal ribonucleic acid
AGF: Antimicrobial resistance gene family
AMR: Antimicrobial resistance
ARG: antibiotic resistance gene
ARO: Antibiotic Resistance Ontology
CARD: Comprehensive Antibiotic Resistance Database
DNA: deoxyribonucleic acid
EPEC: Enteropathogenic *Escherichia coli*
ETEC: Enterotoxigenic *Escherichia coli*
MDA: mass drug administration
PCR: polymerase chain reaction
qPCR: quantitative PCR
RGI: Resistance Gene Identifier

RTT: Rapid Test-and-Treat

SC: Standard Care

SDG: Sustainability Development Goal

WHO: World Health Organization

REFERENCES

1. UNICEF, *Levels & Trends in Child Mortality - Report 2021*. United Nations Children's Fund (UNICEF), 2021.
2. Perin, J., et al., *Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals.* The Lancet Child & Adolescent Health, 2022. **6**(2): p. 106-115.
3. Vos, T., et al., *Global burden of 369 diseases and injuries in 204 countries and territories, 1990 - 2019: a systematic analysis for the Global Burden of Disease Study 2019.* The Lancet, 2020. **396**(10258): p. 1204-1222.
4. Troeger, C., et al., *Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015.* The Lancet Infectious Diseases, 2017. **17**(9): p. 909-948.
5. Nasrin, D., et al., *Pathogens Associated With Linear Growth Faltering in Children With Diarrhea and Impact of Antibiotic Treatment: The Global Enteric Multicenter Study.* J Infect Dis, 2021. **224**(12 Suppl 2): p. S848-S855.
6. Rogawski, E.T., et al., *Use of quantitative molecular diagnostic methods to investigate the effect of enteropathogen infections on linear growth in children in low-resource settings: longitudinal analysis of results from the MAL-ED cohort study.* Lancet Glob Health, 2018. **6**(12): p. e1319-e1328.
7. Mokomane, M., et al., *The global problem of childhood diarrhoeal diseases: emerging strategies in prevention and management.* Ther Adv Infect Dis, 2018. **5**(1): p. 29-43.
8. Malande, O.O., et al., *Barriers to effective uptake and provision of immunization in a rural district in Uganda.* PLoS One, 2019. **14**(2): p. e0212270.
9. Parnham, M.J., et al., *Azithromycin: mechanisms of action and their relevance for clinical applications.* Pharmacol Ther, 2014. **143**(2): p. 225-45.
10. Burns, A.L., et al., *Targeting malaria parasites with novel derivatives of azithromycin.* Front Cell Infect Microbiol, 2022. **12**: p. 1063407.
11. Porco, T.C., et al., *Effect of Mass Distribution of Azithromycin for Trachoma Control on Overall Mortality in Ethiopian Children: A Randomized Trial.* JAMA, 2009. **302**(9): p. 962-968.
12. Keenan, J.D., et al., *Childhood mortality in a cohort treated with mass azithromycin for trachoma.* Clin Infect Dis, 2011. **52**(7): p. 883-8.

13. O'Brien, K.S., et al., *Childhood Mortality After Mass Distribution of Azithromycin: A Secondary Analysis of the PRET Cluster-randomized Trial in Niger.* Pediatr Infect Dis J, 2018. **37**(11): p. 1082-1086.
14. WHO, *WHO guideline on mass drug administration of azithromycin to children under five years of age to promote child survival.* Geneva: World Health Organization, 2020.
15. Keenan, J.D., et al., *Longer-Term Assessment of Azithromycin for Reducing Childhood Mortality in Africa.* N Engl J Med, 2019. **380**(23): p. 2207-2214.
16. Keenan, J.D., et al., *Azithromycin to Reduce Childhood Mortality in Sub-Saharan Africa.* N Engl J Med, 2018. **378**(17): p. 1583-1592.
17. Oldenburg, C.E., et al., *Mass Azithromycin Distribution to Prevent Childhood Mortality: A Pooled Analysis of Cluster-Randomized Trials.* Am J Trop Med Hyg, 2019. **100**(3): p. 691-695.
18. O'Brien, K.S., et al., *Antimicrobial resistance following mass azithromycin distribution for trachoma: a systematic review.* The Lancet Infectious Diseases, 2019. **19**(1): p. e14-e25.
19. Keenan, J.D., A.M. Arzika, and T.M. Lietman, *Azithromycin and Childhood Mortality in Africa.* New England Journal of Medicine, 2018. **379**(14): p. 1382-1384.
20. Hooda, Y., et al., *Mass azithromycin administration: considerations in an increasingly resistant world.* BMJ Glob Health, 2020. **5**(6).
21. Murray, C.J.L., et al., *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.* The Lancet, 2022. **399**(10325): p. 629-655.
22. Gomes, C., et al., *Macrolide resistance mechanisms in Enterobacteriaceae: Focus on azithromycin.* Crit Rev Microbiol, 2017. **43**(1): p. 1-30.
23. Antibiotics for Children With Diarrhea Study, G., et al., *Effect of 3 Days of Oral Azithromycin on Young Children With Acute Diarrhea in Low-Resource Settings: A Randomized Clinical Trial.* JAMA Netw Open, 2021. **4**(12): p. e2136726.
24. Arzika, A.M., et al., *Gut Resistome of Preschool Children After Prolonged Mass Azithromycin Distribution: A Cluster-randomized Trial.* Clin Infect Dis, 2021. **73**(7): p. 1292-1295.
25. Doan, T., et al., *Macrolide and Nonmacrolide Resistance with Mass Azithromycin Distribution.* N Engl J Med, 2020. **383**(20): p. 1941-1950.
26. Doan, T., et al., *Macrolide Resistance in MORDOR I — A Cluster-Randomized Trial in Niger.* New England Journal of Medicine, 2019. **380**(23): p. 2271-2273.
27. Doan, T., et al., *Gut microbiome alteration in MORDOR I: a community-randomized trial of mass azithromycin distribution.* Nat Med, 2019. **25**(9): p. 1370-1376.
28. Pernica, J.M., et al., *Optimising the management of childhood acute diarrhoeal disease using a rapid test-and- treat strategy and/or Lactobacillus reuteri DSM 17938: a multicentre, randomised, controlled, factorial trial in Botswana.* BMJ Glob Health, 2022. **7**(4).

29. Pernica, J.M., et al., *Rapid enteric testing to permit targeted antimicrobial therapy, with and without Lactobacillus reuteri probiotics, for paediatric acute diarrhoeal disease in Botswana: A pilot, randomized, factorial, controlled trial.* PLoS One, 2017. **12**(10): p. e0185177.
30. Dhingra, U., et al., *Lower-Dose Zinc for Childhood Diarrhea - A Randomized, Multicenter Trial.* N Engl J Med, 2020. **383**(13): p. 1231-1241.
31. Nalin, D.R. and R.A. Cash, *50 years of oral rehydration therapy: the solution is still simple.* Lancet, 2018. **392**(10147): p. 536-538.
32. Lazzerini, M. and H. Wanzira, *Oral zinc for treating diarrhoea in children* Cochrane Database Syst Rev, 2016. **12**.
33. Liu, J., et al., *Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study.* Lancet, 2016. **388**(10051): p. 1291-301.
34. Pernica, J.M., et al., *Correlation of Clinical Outcomes With Multiplex Molecular Testing of Stool From Children Admitted to Hospital With Gastroenteritis in Botswana.* J Pediatric Infect Dis Soc, 2016. **5**(3): p. 312-8.
35. Collinson, S., et al., *Probiotics for treating acute infectious diarrhoea.* Cochrane Database Syst Rev, 2020. **12**(12): p. Cd003048.
36. Florez, I.D., et al., *Comparative effectiveness and safety of interventions for acute diarrhea and gastroenteritis in children: A systematic review and network meta-analysis.* PLoS One, 2018. **13**(12): p. e0207701.
37. Francavilla, R., et al., *Randomised clinical trial: Lactobacillus reuteri DSM 17938 vs. placebo in children with acute diarrhoea--a double-blind study.* Aliment Pharmacol Ther, 2012. **36**(4): p. 363-9.
38. Walker, C.R., et al., *Evaluation of Anatomically Designed Flocked Rectal Swabs for Use with the BioFire FilmArray Gastrointestinal Panel for Detection of Enteric Pathogens in Children Admitted to Hospital with Severe Gastroenteritis.* J Clin Microbiol, 2019. **57**(12).
39. Mokomane, M., et al., *A comparison of flocked swabs and traditional swabs, using multiplex real-time PCR for detection of common gastroenteritis pathogens in Botswana.* Diagn Microbiol Infect Dis, 2016. **86**(2): p. 141-3.
40. Goldfarb, D.M., et al., *Evaluation of anatomically designed flocked rectal swabs for molecular detection of enteric pathogens in children admitted to hospital with severe gastroenteritis in Botswana.* J Clin Microbiol, 2014. **52**(11): p. 3922-7.
41. Yousuf, E.I., et al., *Persistence of Suspected Probiotic Organisms in Preterm Infant Gut Microbiota Weeks After Probiotic Supplementation in the NICU.* Front Microbiol, 2020. **11**: p. 574137.
42. Stearns, J.C., et al., *Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age.* ISME J, 2015. **9**(5): p. 1246-59.

43. Guitor, A.K., et al., *Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes.* Antimicrob Agents Chemother, 2019. **64**(1).
44. Guitor, A.K., et al., *Capturing the antibiotic resistome of preterm infants reveals new benefits of probiotic supplementation.* Microbiome, 2022. **10**(1): p. 136.
45. Jiang, H., et al., *Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads.* BMC Bioinformatics, 2014. **15**(1): p. 182.
46. Bushnell, B. *BBMap - Bushnell B.* 2022 [cited 2022; Available from: https://github.com/BioInfoTools/BBMap.
47. Li, H. *seqtk*. 2022 [cited 2022; Available from: https://github.com/lh3/seqtk.
48. Alcock, B.P., et al., *CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database.* Nucleic Acids Res, 2022.
49. Clausen, P., F.M. Aarestrup, and O. Lund, *Rapid and precise alignment of raw reads against redundant databases with KMA.* BMC Bioinformatics, 2018. **19**(1): p. 307.
50. Raphenya, A.R. *RGI*. 2022 [cited 2022; Available from: https://github.com/arpcard/rgi.
51. Prjibelski, A., et al., *Using SPAdes De Novo Assembler.* Current Protocols in Bioinformatics, 2020. **70**(1): p. e102.
52. Lanza, V.F., et al., *In-depth resistome analysis by targeted metagenomics.* Microbiome, 2018. **6**(1): p. 11.
53. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-2069.
54. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
55. NCBI. *Nucleotide*. [cited 2022 November 29]; Available from: https://www.ncbi.nlm.gov/nucleotide.
56. Gilchrist, C. *Clinker*. [cited 2022; Available from: https://github.com/gamcil/clinker.
57. Mbanga, J., et al., *Genomic Insights of Multidrug-Resistant Escherichia coli From Wastewater Sources and Their Association With Clinical Pathogens in South Africa.* Front Vet Sci, 2021. **8**: p. 636715.
58. Pop, M., et al., *Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition.* Genome Biology, 2014. **15**(6): p. R76.
59. Mizutani, T., et al., *Gut microbiota signature of pathogen-dependent dysbiosis in viral gastroenteritis.* Sci Rep, 2021. **11**(1): p. 13945.
60. David, L.A., et al., *Gut microbial succession follows acute secretory diarrhea in humans.* mBio, 2015. **6**(3): p. e00381-15.
61. Chung The, H. and S.H. Le, *Dynamic of the human gut microbiome under infectious diarrhea.* Curr Opin Microbiol, 2022. **66**: p. 79-85.

62. van Schaik, W., *The human gut resistome.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1670): p. 20140087.
63. McMurry, T.L., et al., *Duration of Postdiarrheal Enteric Pathogen Carriage in Young Children in Low-resource Settings.* Clin Infect Dis, 2021. **72**(11): p. e806-e814.
64. Tapela, K. and T. Rahube, *Isolation and antibiotic resistance profiles of bacteria from influent, effluent and downstream: A study in Botswana.* African Journal of Microbiology Research, 2019. **13**(15): p. 279-289.
65. Xiang, Y., et al., *A new plasmid carrying mphA causes prevalence of azithromycin resistance in enterotoxigenic Escherichia coli serogroup O6.* BMC Microbiol, 2020. **20**(1): p. 247.
66. Jasemi, S., et al., *Antibiotic resistance pattern of Bacteroides fragilis isolated from clinical and colorectal specimens.* Ann Clin Microbiol Antimicrob, 2021. **20**(1): p. 27.
67. de Vries, L.E., et al., *The gut as reservoir of antibiotic resistance: microbial diversity of tetracycline resistance in mother and infant.* PLoS One, 2011. **6**(6): p. e21644.
68. Anand Paramadhas, B.D., et al., *Point prevalence study of antimicrobial use among hospitals across Botswana; findings and implications.* Expert Review of Anti-infective Therapy, 2019. **17**(7): p. 535-546.
69. Keenan, J.D., et al., *The Effect of Antibiotic Selection Pressure on the Nasopharyngeal Macrolide Resistome: A Cluster-randomized Trial.* Clin Infect Dis, 2018. **67**(11): p. 1736-1742.
70. Leach, A.J., et al., *A Prospective Study of the Impact of Community-Based Azithromycin Treatment of Trachoma on Carriage and Resistance of Streptococcus pneumoniae.* Clinical Infectious Diseases, 1997. **24**(3): p. 356-362.
71. Fry, A.M., et al., *Adverse and Beneficial Secondary Effects of Mass Treatment with Azithromycin to Eliminate Blindness Due to Trachoma in Nepal.* Clinical Infectious Diseases, 2002. **35**(4): p. 395-402.
72. Ho, D.K., C. Sawicki, and N. Grassly, *Antibiotic Resistance in Streptococcus pneumoniae after Azithromycin Distribution for Trachoma.* J Trop Med, 2015. **2015**: p. 917370.
73. Seidman, J.C., et al., *Increased carriage of macrolide-resistant fecal E. coli following mass distribution of azithromycin for trachoma control.* Int J Epidemiol, 2014. **43**(4): p. 1105-13.
74. Coles, C.L., et al., *Mass distribution of azithromycin for trachoma control is associated with increased risk of azithromycin-resistant Streptococcus pneumoniae carriage in young children 6 months after treatment.* Clin Infect Dis, 2013. **56**(11): p. 1519-26.
75. Bloch, E.M., et al., *Antibiotic Resistance in Young Children in Kilosa District, Tanzania 4 Years after Mass Distribution of Azithromycin for Trachoma Control.* Am J Trop Med Hyg, 2017. **97**(3): p. 815-818.

76. Dethlefsen, L. and D.A. Relman, *Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**(Suppl 1): p. 4554-61.
77. Rogawski McQuade, E.T., et al., *Frequency of bystander exposure to antibiotics for enteropathogenic bacteria among young children in low-resource settings.* Proc Natl Acad Sci U S A, 2022. **119**(36): p. e2208972119.
78. Luby, S.P., et al., *Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial.* Lancet Glob Health, 2018. **6**(3): p. e302-e315.
79. Null, C., et al., *Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial.* Lancet Glob Health, 2018. **6**(3): p. e316-e329.
80. Grembi, J.A., et al., *Effect of water, sanitation, handwashing and nutrition interventions on enteropathogens in children 14 months old: a cluster-randomized controlled trial in rural Bangladesh.* J Infect Dis, 2020.
81. Gakidou, E., et al., *Improving Child Survival Through Environmental and Nutritional InterventionsThe Importance of Targeting Interventions Toward the Poor.* JAMA, 2007. **298**(16): p. 1876-1887.

ACKNOWLEDGMENTS

We acknowledge Michelle Shah for help with DNA extractions and the McMaster Genomics Facility for next-generation sequencing. We thank Dr. David Speicher (Redeemer University, Canada) for assistance in the early stages of planning this work. We thank Dr. Margaret Mokomane and Dr. Leoto Mazhani for their involvement in the conceptualization and supervision of the trial.

FUNDING

This research was funded by a Canadian Institutes of Health Research grant (FRN-148463) to GDW. AKG was supported by a CIHR Doctoral Research Award (GSD-164145). AGM holds McMaster's inaugural David Braley Chair in Computational Biology. Computational support was provided by the McMaster Service Lab and Repository (MSLR) computing

cluster, supplemented by hardware donations and loans from Cisco Systems Canada, Hewlett Packard Enterprise, and Pure Storage.

CONTRIBUTIONS

The randomized control trial from which the samples were selected was led by DMG and JMP. AGM, JMP, DMG, GDW, and AKG conceived the study and designed the experiments. AKG and AK performed the DNA extraction, library preparation, and targeted enrichments. AKG analyzed the data, generated tables and figures, and wrote the manuscript with primary input from AGM, GDW, and JMP. The authors read and approved the final manuscript.

SUPPLEMENTARY INFORMATION*

Additional File 1 – Sample details - Details of patient characteristics, DNA extraction, library preparation, enrichment, and sequencing results for each sample.

Additional File 2 – ARGs identified through RGI*bwt read mapping

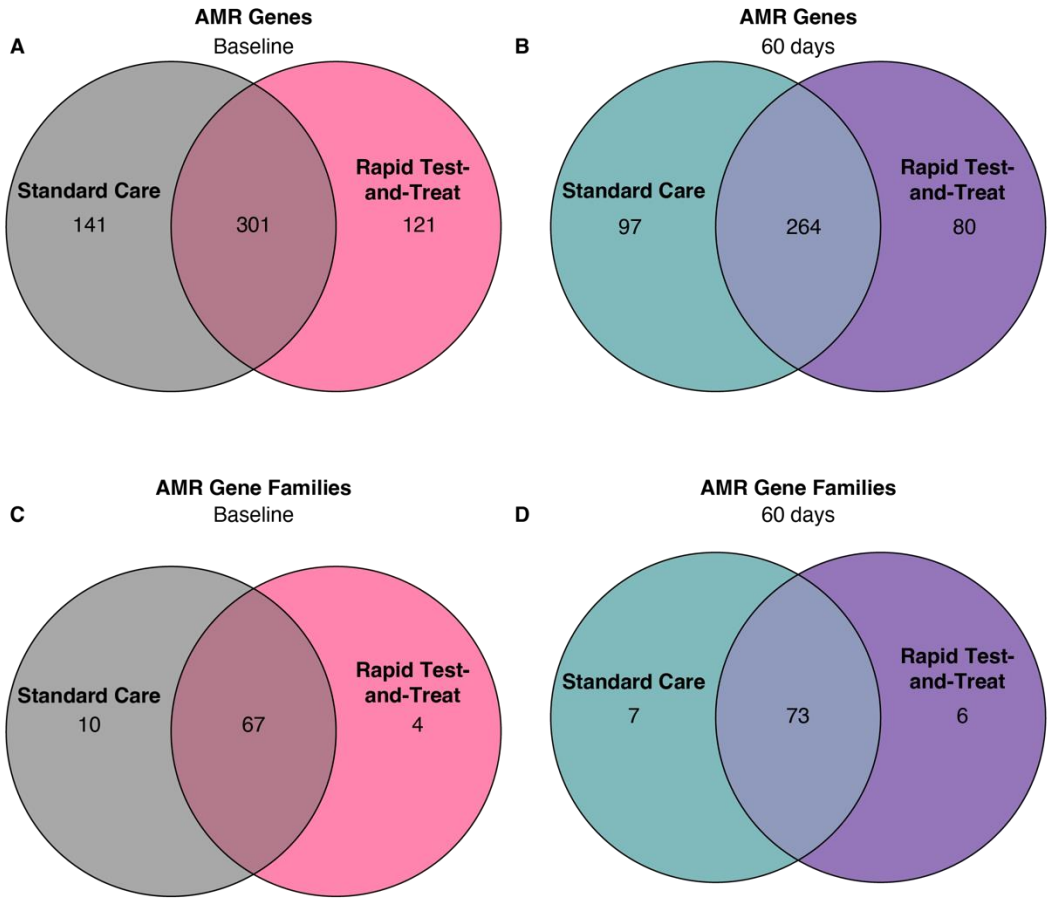
Additional File 3 – ARGs identified through assembly and prediction with RGI

Additional File 4 – Predicted bacterial host of ARGs – For a subset of ARGs, the bacterial host and surrounding ARGs were predicted.

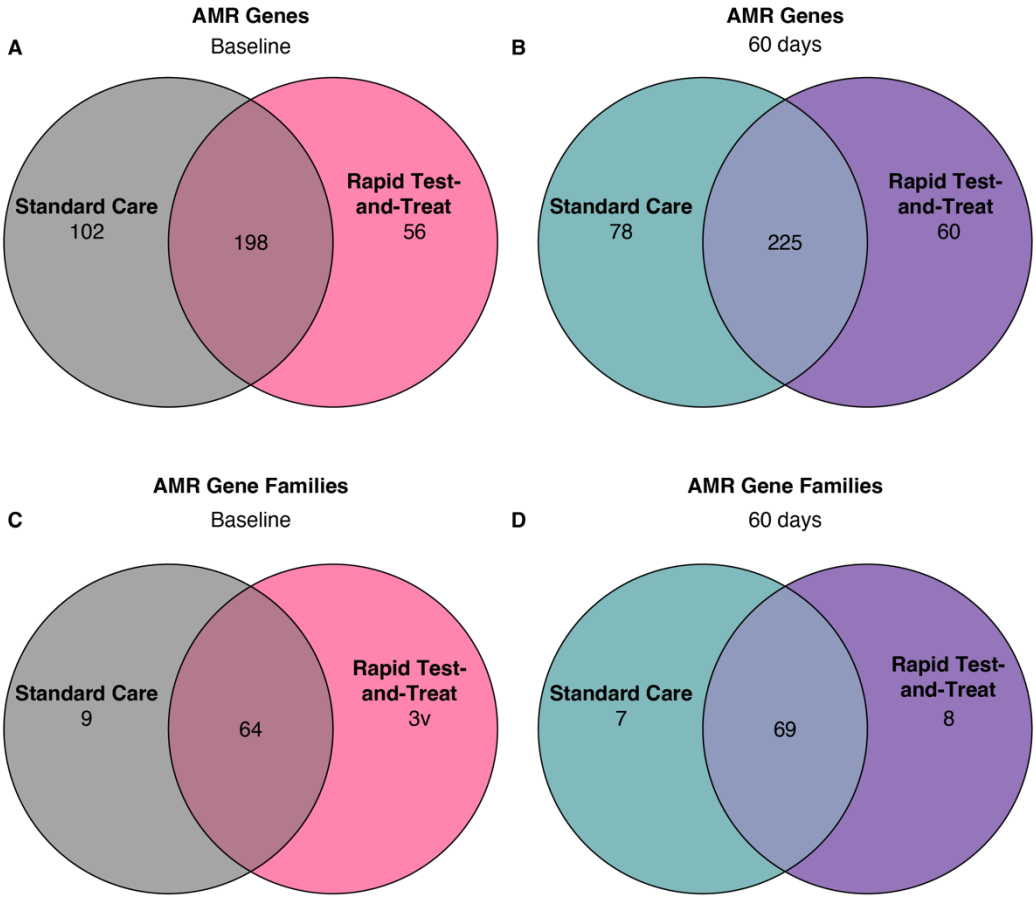
Additional File 5 – Negative control results RGI*bwt – Results of mapping reads from negative controls

Additional File 6 – Negative control results RGI*main – Results of assembly and prediction of ARGs from negative control samples.

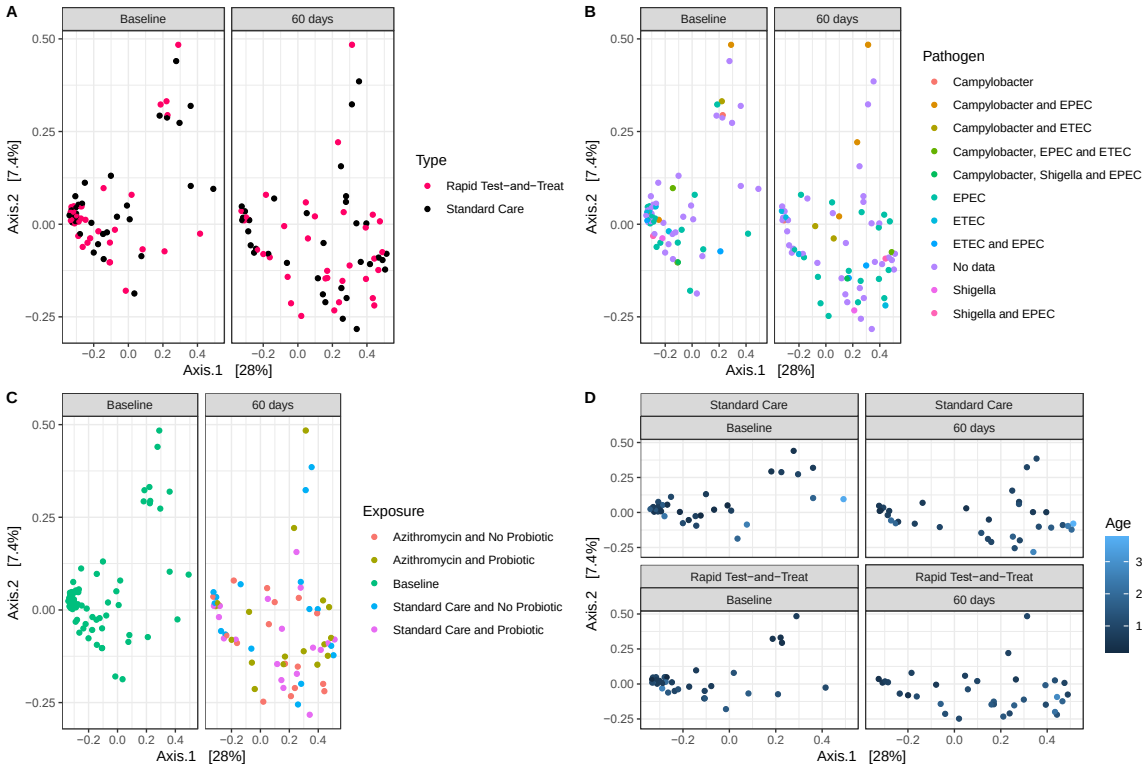
*Additional files are not included in this thesis due to size constraints.



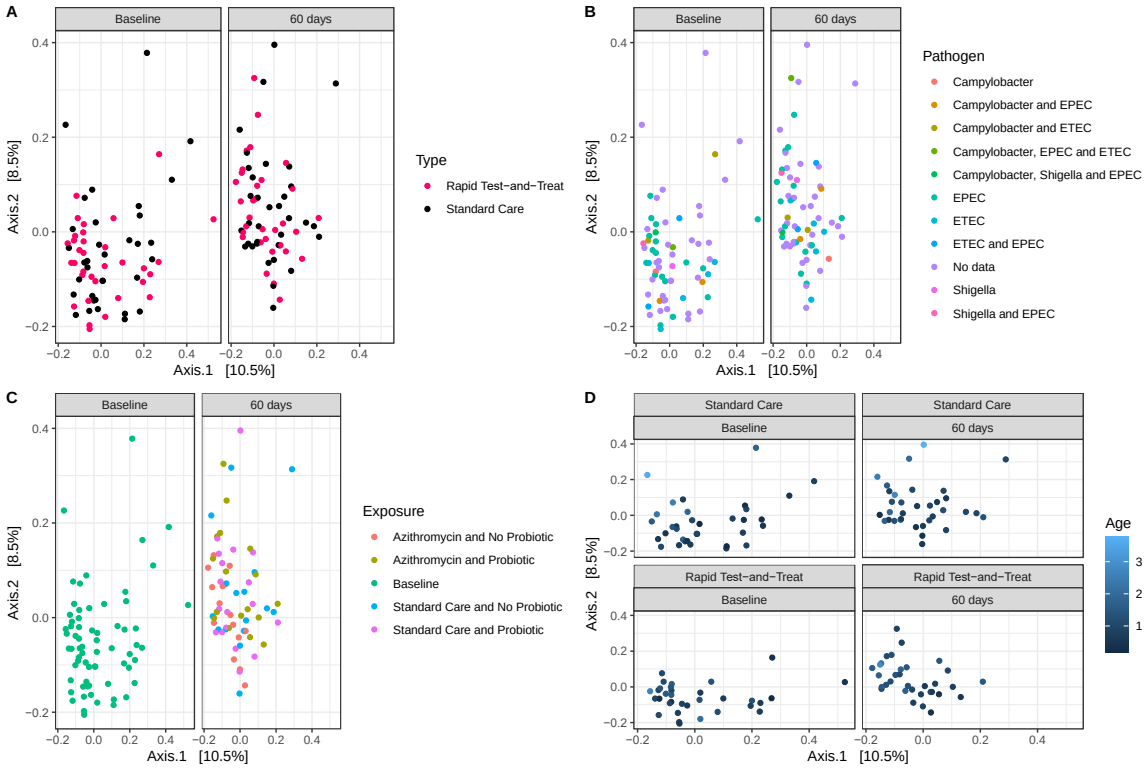
Supplementary Figure 1: Unique ARGs and AMR gene families identified in each cohort at baseline and 60 days post-treatment.
Results shown are from the RGI*bwt results.



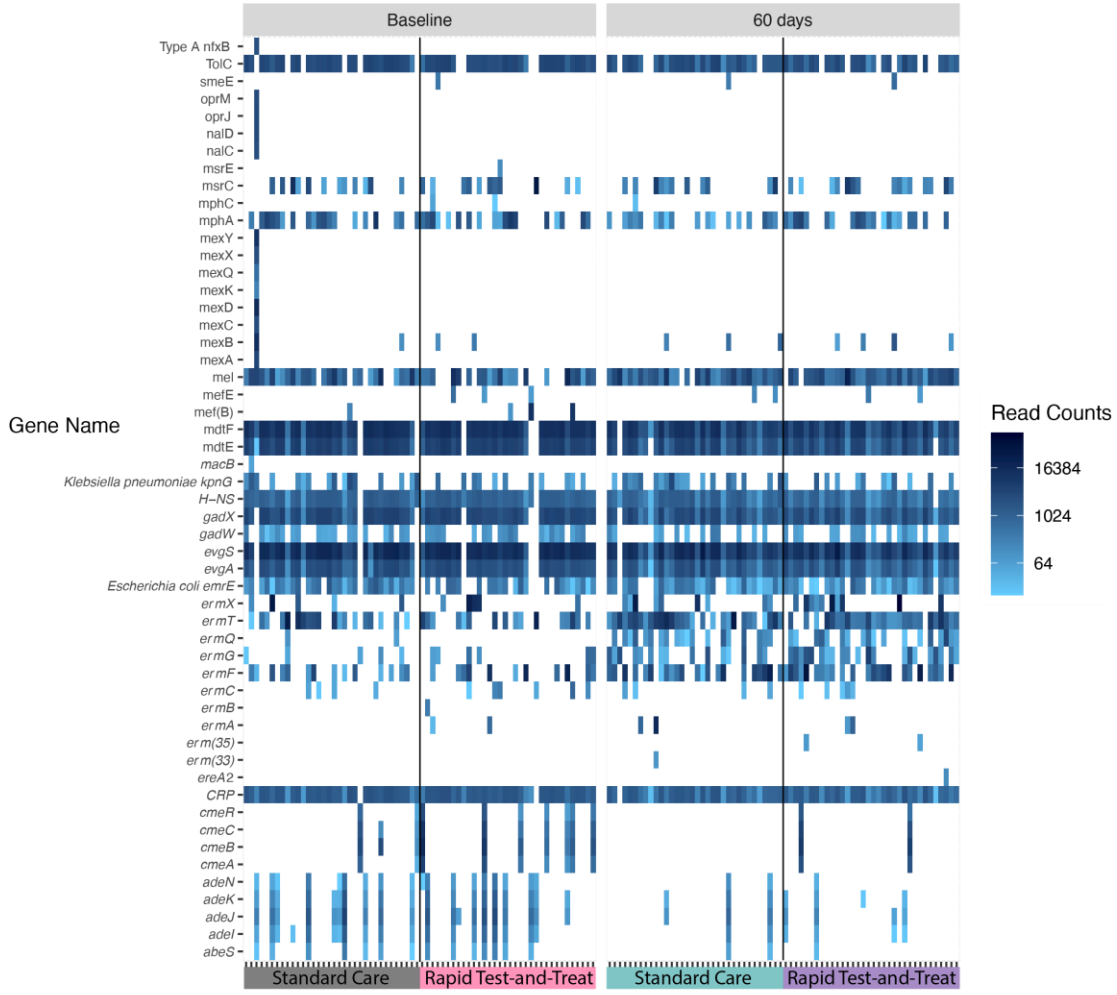
Supplementary Figure 2: Unique ARGs and AMR gene families identified in each cohort at baseline and 60 days post-treatment.
Results shown are from the RGI*main data.



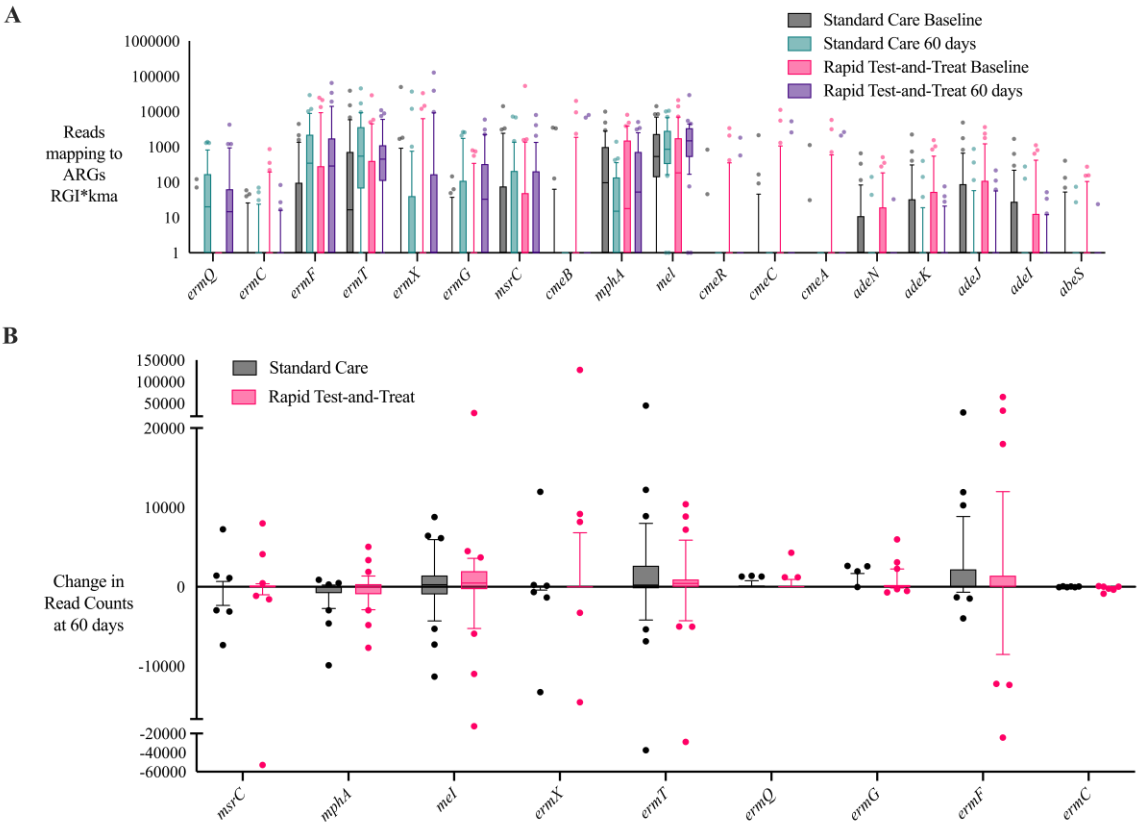
Supplementary Figure 3: PCoA of Jaccard distances of the resistome of children. Reads were mapped to CARD using RGI**bwt*. Samples are separated by baseline and 60-day timepoints and then coloured according to cohort (A), the pathogen detected through rapid molecular test before azithromycin exposure (B), azithromycin and probiotic exposures (C) or the age at sample collection (D).



Supplementary Figure 4: PCoA of Jaccard distances of the resistome of children. Results were obtained from *de novo* assembly and prediction of ARGs through RGI**main*. Samples are separated by baseline and 60-day timepoints and then coloured according to cohort (A), the pathogen detected through the rapid molecular test before azithromycin exposure (B), azithromycin and probiotic exposures (C) or the age at sample collection (D).



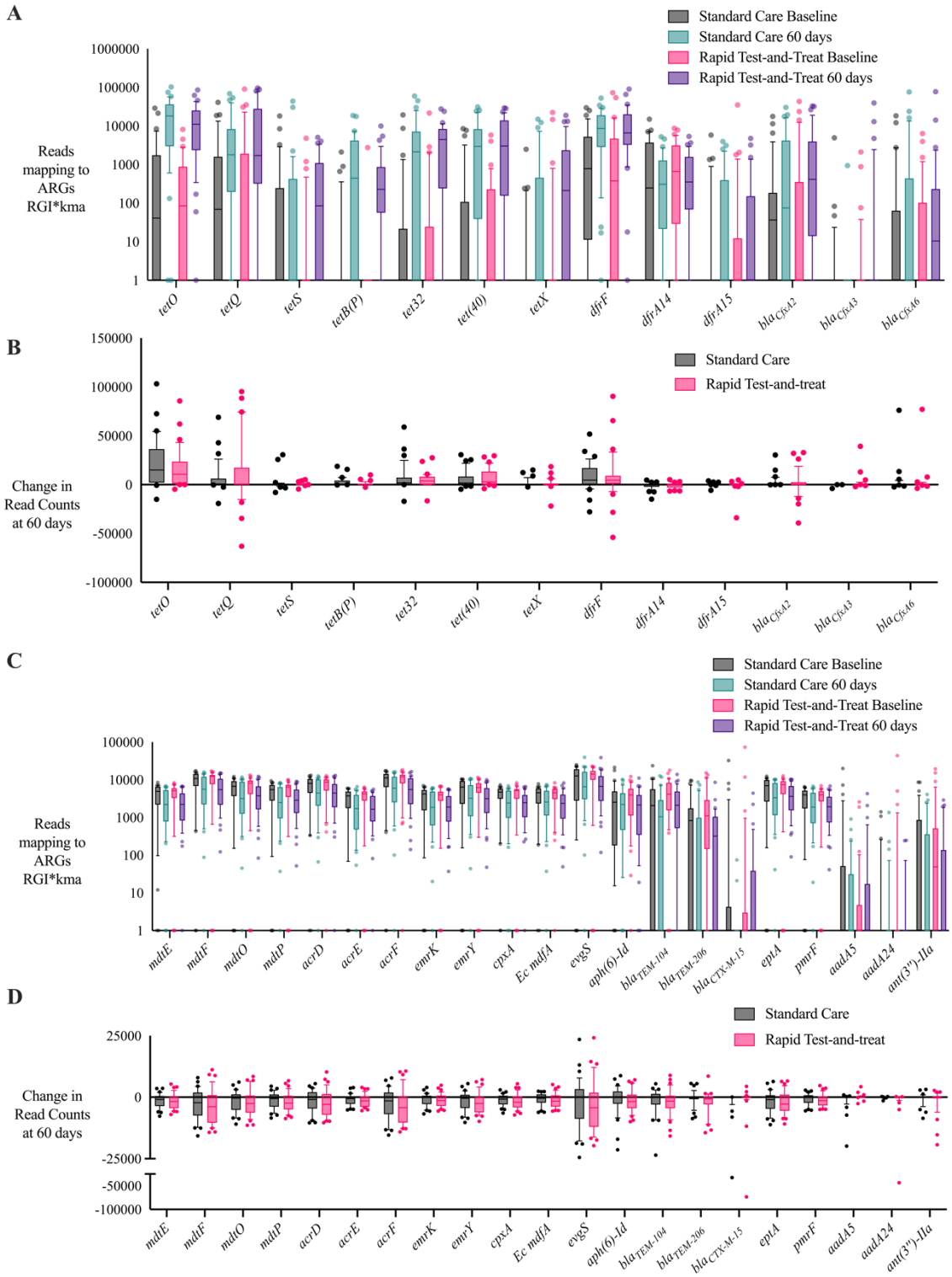
Supplementary Figure 5: Abundance of macrolide resistance genes in infants and changes at 60 days.
Heatmap showing the number of reads mapping to each ARG as determined through RGI*bwt.



Supplementary Figure 6: Differences in read count abundance for a subset of ARGs associated with macrolide resistance.
Results were obtained from read mapping to CARD using RGI*bwt. A) Each point represents the number of reads mapping to the specific ARG in one child separated by cohort and timepoint. This set of genes does not include all ARGs associated with macrolide resistance, only the subset of genes that changed in prevalence or abundance in either group of children over time. B) Each point represents the change in read counts mapping to the specific ARG in one child from baseline to 60 days. This plot only encompasses the most prevalent macrolide resistance genes. Box plots represent data within 10^{th} to 90^{th} percentiles.


Baseline
60 days
Gene Name
Read Counts
16384
1024
64
Standard Care
Rapid Test-and-Treat
Standard Care
Rapid Test-and-Treat

Supplementary Figure 7: Abundance of top 100 antibiotic resistance genes in infants and changes at 60 days.
Heatmap showing the number of reads mapping to each ARG as determined through RGI*bwt. These were the top 250 ARGs based on overall read abundance across all children and timepoints. ARGs associated with macrolide resistance were omitted and can be found in Supplementary Figure 5. Certain names were shortened; See Additional File 2 for full names and ARO accession numbers.


A
Standard Care Baseline
Standard Care 60 days
Rapid Test-and-Treat Baseline
Rapid Test-and-Treat 60 days
Reads mapping to ARGs RGI*kma
1000000
100000
10000
1000
100
10
1
tetO
tetQ
tetS
tetB(P)
tet32
tet(40)
tetX
dfrF
dfrA14
dfrA15
blaCfxA2
blaCfxA3
blaCfxA6
B
Standard Care
Rapid Test-and-treat
Change in Read Counts at 60 days
150000
100000
50000
0
-50000
-100000
C
Standard Care Baseline
Standard Care 60 days
Rapid Test-and-Treat Baseline
Rapid Test-and-Treat 60 days
Reads mapping to ARGs RGI*kma
100000
10000
1000
100
10
1
mdtE
mdtF
mdtO
mdtP
acrD
acrE
acrF
emrK
emrY
cpxA
Ec mdfA
evgS
aph(6)-Id
blaTEM-104
blaTEM-206
blaCTX-M-15
eptA
pmrF
aadA5
aadA24
ant(3'')-IIa
D
Standard Care
Rapid Test-and-treat
Change in Read Counts at 60 days
25000
0
-25000
-50000
-100000

Supplementary Figure 8: Differences in read count abundance for a subset of ARGs. Results were obtained from read mapping to CARD using RGI*bwt. A and C) Each point represents the number of reads mapping to the specific ARG in one child separated by cohort and timepoint. B and D) Each point represents the change in read counts mapping to the specific ARG in one child from baseline to 60 days. A and B represent a subset of genes that changed in prevalence or abundance in at least one group of children over time. C and D represent ARGs often associated with *Escherichia coli*. Ec = *Escherichia coli.* Box plots represent data within 10^{th} to 90^{th} percentiles.

CHAPTER FIVE: Conclusions and future directions

SUMMARY

Antibiotics and antibiotic resistance have been evolving in concert for millennia. In the short time since the human discovery and use of antibiotics, there has been an unprecedented increase in the mobilization and transfer of ARGs. This has resulted in MDR pathogens capable of causing intractable infections. Deaths attributable to antibiotic resistance are increasing, and a return to an era where simple cuts and infections are deadly threatens humanity. While the environment is a rich source of ARGs that can transfer into human-associated pathogens, AMR is also diverse in the human microbiome. The work presented in this thesis describes a comprehensive and selective sequencing-based method to detect most known ARGs (Chapter 2), which allowed for the characterization of the impact of probiotics (Chapter 3) and antibiotics (Chapter 4) on the gut resistome. Probiotics prevented the long-term persistence of ARGs that preterm infants were likely exposed to in the hospital (Chapter 3). While a three-day dose of the macrolide antibiotic azithromycin did not select for AMR in the gut microbiome of children two months after the exposure (Chapter 4). These studies are just beginning to unearth the early-life factors that shape the gut microbiome and resistome. In the following sections, I will discuss prospects for sequencing-based methods for the surveillance of AMR, as well as other interventions to reduce the burden and selection of ARGs in the gut microbiome in early life.

Expanding our ability to detect antibiotic resistance genes

Our set of over 37,000 probes targets over 2,000 known ARGs and can expand as new mechanisms are characterized. While ours was not the first probeset to be designed for

AMR, we have established a precedent for rigorous validation of this method using mock metagenomic samples. We have also highlighted the importance of including negative controls in the form of extraction buffer blanks and water blanks throughout DNA extraction, library preparation, and enrichment to identify potential contamination. Finally, we were the first to apply a probeset for AMR to study the resistome in longitudinal gut microbiome samples.

Since the publication of our probeset, it has been downloaded by over 1800 unique users globally. Many others have designed similar approaches to capture AMR (Beaudry, Thomas, et al., 2021; Ferreira et al., 2021; Hayden et al., 2022). Increasing interest in designing probesets has led to approaches to detect viruses, parasites, and even virulence factors (Beaudry, Wang, et al., 2021; Lasa et al., 2019; Melnikov et al., 2011). As a pioneer of targeted enrichment of ARGs, it is exciting to see the development of this method, including improvements in bait design and its application in various environments. There remains, however, much to consider when applying targeted capture for detecting AMR.

Addressing contamination in sequencing data

A common theme throughout my three projects was to prevent the false reporting of ARGs. Avoiding contamination is an aspect of my work that has since been patented in Canada. Common laboratory reagents and kits used in next-generation sequencing preparation may be inherently contaminated by bacteria that encode ARGs (de Goffau et al., 2018; Wally et al., 2019). After an enrichment step, these rare contaminants become inflated. In many cases, a low level of contamination is unavoidable, and any unexpected

data should be reported. Working in “clean-room” conditions, as is the norm for ancient DNA work, is an extreme measure to avoid contamination (Eisenhofer et al., 2019). This intervention is not feasible in all laboratory settings, and therefore, contamination is predominantly dealt with bioinformatically (de Goffau et al., 2018).

For metagenomic samples, computational approaches can identify similar strains resulting from cross-contamination (Lou et al., 2021). Alternatively, AMR markers identified in negative controls and blanks can be removed from the analysis to avoid inflating the sensitivity (Ferreira et al., 2021). In Chapters 2 and 3, I reported the ARGs identified in my negative controls. In Chapter 4, I expanded this analysis by comparing the allelic diversity of ARGs identified in the negative controls with those in the enriched stool samples. Another approach to removing data originating from contaminants would be identifying exact sequence duplicates from one's samples and controls and removing those reads. For any sequencing-based method to become a diagnostic for infections and antibiotic susceptibility profiles, careful sampling and reporting of ARGs in negative controls must be required.

Limitations of targeted capture for AMR and future directions

Currently, the biggest competition for targeted sequencing is metagenomic shotgun sequencing. This method provides additional information not specific to AMR, however large amounts of sequence data are required to capture the rare portion of the resistome. Hybridization-based enrichment before sequencing reduces the depth of sequencing required by upwards of 1,000-fold. In the field of AMR detection, a rapid and

comprehensive diagnostic is extremely desirable. While our approach requires at least one additional day before sequencing, it is possible that with optimization, this time can be reduced (Dave & Liu, 2010). Combining targeted capture with real-time sequencing, such as Oxford Nanopore Technologies (ONT), could provide a specific and rapid method to detect AMR that would also be accessible in LMICs (Slizovskiy et al., 2022). Other emerging technologies with promise for increasing our understanding of AMR include single-cell genomics (Jin et al., 2022). Finally, cross-linking techniques, such as Hi-C, coupled with targeted capture could provide additional information on the surrounding genetic context of ARGs that is lost through traditional targeted sequencing (Kent et al., 2020). As new tools are developed, combining our probeset with these technologies will be essential to ensure its longevity and relevance.

Barriers to sequence-based approaches for the surveillance of AMR remain, including the required equipment, expertise, and high cost. This is especially true in LMICs, where a better understanding of the scope of AMR could inform policies and improve general health in the population. While probe synthesis and reagents remain expensive, it is possible to generate probes "in-house" or use home-brew recipes for library preparation (Hayden et al., 2022; Meyer & Kircher, 2010). Another cost-related barrier is the size of the probeset. Alternative design approaches to ensure unique regions within ARGs are targeted or designing different modules for certain subsets of ARGs can reduce the number of probes (Dickson et al., 2021). Finally, as sequencing becomes cheaper, there will be

increased reliance on metagenomics for surveying AMR. Hybridization-based sequencing costs must also decrease for this approach to remain relevant and cost-effective.

Another challenge with DNA-based targeted sequencing approaches is the inability to discern DNA derived from active bacteria versus extracellular DNA (eDNA) (Carini et al., 2016; Crofts et al., 2017; Larsson & Flach, 2022). Specific treatments to ensure DNA is extracted from live bacteria or single-cell genomics techniques are beginning to shed light on the active fraction of bacteria and their ARGs (Li et al., 2022). However, eDNA may continue to play a role in transferring AMR, as plasmids or other genetic elements can be acquired through natural transformation (Mell & Redfield, 2014).

In the same regard, DNA-based approaches do not identify whether an ARG is expressed in a particular environment. To achieve this, our probeset and others can also target AMR from RNA extracts (Mercer et al., 2014). One could argue, however, that even if an ARG is not currently expressed, it could be captured downstream of a constitutive promoter or activated given the appropriate conditions. Furthermore, unlike RNA-based sequencing, DNA-based capture can provide additional information on sequences beyond the barriers of targeted ARGs. Analysis of the surrounding genetic context can provide insights into potential bacterial hosts of ARGs or their presence on MGEs. This was another unique aspect of our targeted sequencing approach and analysis described in Chapters 3 and 4.

Finally, one of the challenges with comprehensively cataloguing ARGs from sequence data from any environment is the worrisome nature of presenting alarming estimates of the

diversity of AMR. A typical output from an analysis of sequence data can result in 10s to 100s of ARGs. Not all resistance genes are created equal and pose the same risk level. A study of the resistome should focus on the context of the ARGs, whether they are likely to be expressed or silent, and particularly if they are mobilized into clinical pathogens (Martinez et al., 2015). This was the goal of the results reported in Chapters 3 and 4. Rather than discuss all ARGs, I focused on specific types of AMR mechanisms, including antibiotic inactivation genes (Chapter 3) and macrolide resistance genes (Chapter 4).

Modulating antibiotic resistance in the gut microbiome

The studies presented in this thesis were the first to use a targeted-sequencing approach for the analysis of large (>100) longitudinal sample sets representing the gut microbiome of preterm infants and children under five years of age. While not surprising given other studies of similarly aged populations, the gut resistomes of our participants were extensive and diverse. Increasing awareness of early-life perturbations of the gut microbiome and their implications for future health outcomes highlight the need for comprehensive studies of infants and children (Healy et al., 2022).

In Chapter 3, I revealed how a probiotic containing four strains of *Bifidobacterium* spp. and one *Lactobacillus rhamnosus* strain reduced the persistence of ARGs associated with Enterobacteriaceae and *Klebsiella spp.* in preterm infants at 5 months of age. The mechanisms behind this reduction are currently unknown, although, certain bacteria confer niche exclusion or colonization resistance against other bacteria (Osbelt et al., 2021). While the NICU was not sampled in this study, the children were likely exposed to MDR strains

during their extended stays in this environment (Brooks et al., 2017). In future studies, targeted capture of ARGs from environmental swabs may identify sources of AMR within the NICU. Furthermore, given the extensive prophylactic antibiotic treatment of preterm infants, it is worth investigating whether probiotics can be used in lieu of antibiotics to prevent nosocomial infections. Given the small sample size of our study and individual variability, further research into the ideal formulations and dosing schedules of probiotics for preterm infants is still needed (Beck et al., 2022). Ideally, changes to the microbiome and resistome of preterm infants following probiotic supplementation would be monitored in both the short- and long-term.

The primary goal of Chapter 4 was to assess whether azithromycin exposure would result in increased macrolide resistance. Perhaps surprisingly, the resistome of children with diarrhoea that received a three-day dose of azithromycin did not differ from those that received the standard of care 60 days later. A limitation of our approach is the inability to capture antibiotic resistance that arises due to mutations in target genes due to their initial exclusion from the probeset. Therefore, these potential mutations are not detected in both of our resistome studies. Targeted capture can be used to monitor allelic diversity, or separate PCR amplification of these target genes followed by sequencing could incorporate these targets into our method (Hayden et al., 2022).

In this thesis, both studies of the gut resistome highlight the diverse set of ARGs in infants and children at baseline sampling and their persistence into later life. Although we did not measure strain persistence specifically, the maintenance of these ARGs is likely

associated with persistent antibiotic-resistant bacteria in the gut microbiome. These bacteria pose a risk for potential future infections. It has been noted previously that antibiotic-resistant *P. aeruginosa* strains from the gut can translocate to the lungs, where they could potentially cause infections (Wheatley et al., 2022). Therefore, a reservoir of pathogenic bacteria with associated AMR in the microbiome is a concern that should be revealed before antibiotic treatment. Monitoring baseline levels of AMR in an individual upon entry to the hospital could prevent future selection of MDR bacteria. Perhaps a future of personalized antibiotic treatments to prevent unwanted selection through understanding a patient's past antibiotic exposures is possible (Stracy et al., 2022). Finally, the persistence of ARGs in these young children at the measured time points in our studies raises concerns of how long this colonization lasts and whether it can be reversed (Kang et al., 2022).

Reducing the need for antibiotics

Another common theme between chapters 3 and 4 is the need for antibiotic alternatives. We highlight how probiotics are one potential intervention to not only seed the microbiome with beneficial organisms but also reduce the burden of certain ARGs in preterm infants. Although a three-day dose of azithromycin for the treatment of diarrhoea in children did not result in an increase in macrolide resistance genes, it is possible that repeated exposures may select for AMR in the longer term. Therefore, alternative therapies to treat infections, such as phages, probiotics that target specific pathogens, and combination therapies, should be further explored (Shkoporov et al., 2022). The significance of studying azithromycin exposure in Chapter 4 is the proposed

recommendation of MDA of this antibiotic to reduce childhood mortality. If we are to learn anything from our past it should be that the MDA of antibiotics, as we have seen in agriculture, selects for antibiotic resistance. The high level of baseline macrolide resistance identified in the gut microbiome of these children and the potential for repeated selection should encourage alternative interventions to reduce under-five mortality. These could include improved nutrition, access to clean drinking water, and vaccination.

To prevent further selection for AMR and reduce the number of deaths attributable to resistance, it is important that we continue to change our practices. Improved sanitation, better policies surrounding antibiotic use, widespread research into antibiotics and AMR, improved infection prevention, and antibiotic stewardship programs must be implemented (Murray et al., 2022). Infection prevention has received tremendous interest and uptake throughout the COVID-19 pandemic, where non-pharmaceutical interventions have limited the spread of respiratory viruses. Vaccination effectively prevents the need for pharmaceuticals that will inevitably select for resistance (Murray et al., 2022). However, very few bacterial vaccines have been successfully developed and distributed apart from one targeting *Streptococcus pneumoniae* (Murray et al., 2022). The future is bright, however, with the potential for personalized medicine through rapid diagnostics and an understanding of baseline levels of AMR in the microbiome. Controlling the expansion of antibiotic resistance is ultimately a One Health issue that requires dedicated efforts from individuals at all levels and all sectors of government, health care, as well as academic and pharmaceutical research.

Future directions

While the surveys of the gut resistome presented in this thesis focused on a small subset of individuals at early age, it would be ideal to track AMR in later life. Given the initial promising results, it would be intuitive to follow up on larger cohorts with samples beyond 5 months of age and 60 days. While our studies using a targeted probeset for ARGs have focused on the gut microbiome of humans, this approach can be extended to target AMR in any environment, including soil and wastewater. It is likely that the probeset used in these environments will differ from the current set, given it is largely biased towards ARGs identified in clinically relevant bacteria. Therefore, improved curation of environmentally derived ARGs should inform future probeset design. Ultimately, as technologies improve, I believe the most powerful approach would be to couple hybridization-based capture for ARGs with long-read sequencing to better elucidate the bacterial hosts of the targeted resistance mechanisms.

Concluding remarks

For years humanity was naïve to the impact that the inappropriate use of antibiotics in agriculture and human and animal medicine would have on the emergence of resistance. Now that we face bacteria with a diverse and hardy arsenal of defensive mechanisms, we must find new ways to protect one of the greatest gifts Nature has given us: antibiotics. The comprehensive detection tool for AMR presented in this thesis can expand the surveillance of ARGs in the human microbiome and beyond. Ultimately this targeted sequencing approach for AMR can facilitate large longitudinal studies to provide an improved

understanding of the diversity of ARGs in the gut microbiome from early-life into adulthood. By monitoring which factors and exposures (eg. probiotics or antibiotics) impact the spread and maintenance of AMR, we can prevent unnecessary selective pressures, infer future health implications, and perhaps even relieve the gut microbiome of ARGs of concern entirely.

References

Abbott, A. N., & Fang, F. C. (2015). Molecular Detection of Antibacterial Drug Resistance. In J. H. Jorgensen, K. C. Carroll, G. Funke, M. A. Pfaller, M. L. Landry, S. S. Richter, & D. W. Warnock (Eds.), *Manual of Clinical Microbiology, 11th Edition*. https://doi.org/10.1128/9781555817381.ch77

Abraham, E. P., & Chain, E. (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, *146*(3713), 837. https://doi.org/10.1038/146837a0

Adu-Oppong, B., Gasparrini, A. J., & Dantas, G. (2017). Genomic and functional techniques to mine the microbiome for novel antimicrobials and antimicrobial resistance genes. *Ann N Y Acad Sci*, *1388*(1), 42-58. https://doi.org/10.1111/nyas.13257

Aharonowitz, Y., & Cohen, G. (1992). Pencillin and Cephalosporin Biosynthetic Genes: Structure, Organization, Regulation, and Evolution. *Annu. Rev. Microbiol.*, *46*.

Akiba, T., Koyama, K., Ishiki, Y., Kimura, S., & Fukushima, T. (1960). On the mechanism of the development of multiple-drug-resistant clones of Shigella. *Jpn J Microbiol*, *4*, 219-227. https://doi.org/10.1111/j.1348-0421.1960.tb00170.x

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., . . . McArthur, A. G. (2022). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res*. https://doi.org/10.1093/nar/gkac920

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., . . . McArthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res*, *51*(D1), D690-d699. https://doi.org/10.1093/nar/gkac920

Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J., & Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol*, *8*(4), 251-259. https://doi.org/10.1038/nrmicro2312

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *ISME J*, *3*(2), 243-251. https://doi.org/10.1038/ismej.2008.86

Allicock, O. M., Guo, C., Uhlemann, A., Whittier, S., Chauhan, L. V., Garcia, J., Price, A., Morse, S. S., Mishra, N., Briese, T., & Lipkin, W. I. (2018). BacCapSeq: a Platform for Diagnosis and Characterization of Bacterial Infections. *mBio*, *9*(5). https://doi.org/10.1128/mBio

Aminov, R. I. (2009). The role of antibiotics and antibiotic resistance in nature. *Environ Microbiol*, *11*(12), 2970-2988. https://doi.org/10.1111/j.1462-2920.2009.01972.x

Aminov, R. I., & Mackie, R. I. (2007). Evolution and ecology of antibiotic resistance genes. *FEMS Microbiol Lett*, *271*(2), 147-161. https://doi.org/10.1111/j.1574-6968.2007.00757.x

Andersen, S. J., Quan, S., Gowan, B., & Dabbs, E. R. (1997). Monooxygenase-Like Sequence of a *Rhodococcus equi* Gene Conferring Increased Resistance to Rifampin by Inactivating This Antibiotic. *Antimicrob Agents Chemother*, *41*(1). https://doi.org/10.1128/AAC.41.1.218

Andersson, D. I., & Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol*, *8*(4), 260-271. https://doi.org/10.1038/nrmicro2319

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, *6*(1), 23. https://doi.org/10.1186/s40168-018-0401-z

Aravind, L., & Koonin, E. V. (1999). DNA polymerase β-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Research*, *27*(7). https://doi.org/10.1093/nar/27.7.1609

Arboleya, S., Binetti, A., Salazar, N., Fernandez, N., Solis, G., Hernandez-Barranco, A., Margolles, A., de Los Reyes-Gavilan, C. G., & Gueimonde, M. (2012). Establishment and development of intestinal microbiota in preterm neonates. *FEMS Microbiol Ecol*, *79*(3), 763-772. https://doi.org/10.1111/j.1574-6941.2011.01261.x

Arias, C. A., Vallejo, M., Reyes, J., Panesso, D., Moreno, J., Castaneda, E., Villegas, M. V., Murray, B. E., & Quinn, J. P. (2008). Clinical and microbiological aspects of linezolid resistance mediated by the *cfr* gene encoding a 23S rRNA methyltransferase. *J Clin Microbiol*, *46*(3), 892-896. https://doi.org/10.1128/JCM.01886-07

Arrieta, M. C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., & Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Front Immunol, 5*, 427. https://doi.org/10.3389/fimmu.2014.00427

Bahl, M. I., Sorensen, S. J., Hansen, L. H., & Licht, T. R. (2004). Effect of tetracycline on transfer and establishment of the tetracycline-inducible conjugative transposon Tn916 in the guts of gnotobiotic rats. *Appl Environ Microbiol*, *70*(2), 758-764. https://doi.org/10.1128/AEM.70.2.758-764.2004

Baker, J. L. (2022). Using Nanopore Sequencing to Obtain Complete Bacterial Genomes from Saliva Samples. *mSystems*, *7*(5). https://doi.org/10.1128/msystems.00491-22

Baltz, R. H. (2010). Genomics and the ancient origins of the daptomycin biosynthetic gene cluster. *J Antibiot (Tokyo)*, *63*(8), 506-511. https://doi.org/10.1038/ja.2010.82

Barlow, M., & Hall, B. G. (2002). Phylogenetic analysis shows that the OXA β-lactamase genes have been on plasmids for millions of years. *J Mol Evol*, *55*(3), 314-321. https://doi.org/10.1007/s00239-002-2328-y

Beaber, J. W., Hochhut, B., & Waldor, M. K. (2003). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature*, *427*. https://doi.org/10.1038/nature02241

Beal, S. G., Ciurca, J., Smith, G., John, J., Lee, F., Doern, C. D., & Gander, R. M. (2013). Evaluation of the Nanosphere Verigene Gram-positive Blood Culture Assay with the VersaTREK Blood Culture system and assessment of possible impact on selected patients. *J Clin Microbiol*, *51*(12), 3988-3992. https://doi.org/10.1128/JCM.01889-13

Beaudry, M. S., Thomas, J. C., Baptista, R. P., Sullivan, A. H., Norfolk, W., Devault, A., Enk, J., Kieran, T. J., Rhodes Jr, O. E. R., Perry-Dow, K. A., Rose, L. J., Bayona-Vásquez, N. J., Oladeinde, A., Lipp, E. K., Sanchez, S., & Glenn, T. C. (2021). Escaping the fate of Sisyphus: assessing resistome hybridization baits for antimicrobial resistance gene capture. *Environ Microbiol*, *23*(12). https://doi.org/10.1111/1462-2920.15767

Beaudry, M. S., Wang, J., Kieran, T. J., Thomas, J., Bayona-Vasquez, N. J., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, J. S., Rhodes, O. E., Jr., & Glenn, T. C. (2021). Improved Microbial Community Characterization of 16S rRNA via Metagenome Hybridization Capture Enrichment. *Front Microbiol*, *12*, 644662. https://doi.org/10.3389/fmicb.2021.644662

Beck, L. C., Masi, A. C., Young, G. R., Vatanen, T., Lamb, C. A., Smith, R., Coxhead, J., Butler, A., Marsland, B. J., Embleton, N. D., Berrington, J. E., & Stewart, C. J. (2022). Strain-specific impacts of probiotics are a significant driver of gut microbiome development in very preterm infants. *Nat Microbiol*, *7*(10), 1525-1535. https://doi.org/10.1038/s41564-022-01213-w

Bengtsson-Palme, J., Hammaren, R., Pal, C., Ostman, M., Bjorlenius, B., Flach, C. F., Fick, J., Kristiansson, E., Tysklind, M., & Larsson, D. G. J. (2016). Elucidating

selection processes for antibiotic resistance in sewage treatment plants using metagenomics. *Sci Total Environ*, *572*, 697-712. https://doi.org/10.1016/j.scitotenv.2016.06.228

Benveniste, R., & Davies, J. (1973). Aminoglycoside Antibiotic-Inactivating Enzymes in Actinomycetes Similar to Those Present in Clinical Isolates of Antibiotic-Resistant Bacteria. *Proc Natl Acad Sci U S A*, *70*(8), 5. https://doi.org/10.1073/pnas.70.8.2276

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., . . . Schloter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, *8*(1), 103. https://doi.org/10.1186/s40168-020-00875-0

Berglund, F., Osterlund, T., Boulund, F., Marathe, N. P., Larsson, D. G. J., & Kristiansson, E. (2019). Identification and reconstruction of novel antibiotic resistance genes from metagenomes. *Microbiome*, *7*(1), 52. https://doi.org/10.1186/s40168-019-0670-1

Berinson, B., Both, A., Berneking, L., Christner, M., Lutgehetmann, M., Aepfelbacher, M., & Rohde, H. (2021). Usefulness of BioFire FilmArray BCID2 for Blood Culture Processing in Clinical Practice. *J Clin Microbiol*, *59*(8), e0054321. https://doi.org/10.1128/JCM.00543-21

Bhullar, K., Waglechner, N., Pawlowski, A., Koteva, K., Banks, E. D., Johnston, M. D., Barton, H. A., & Wright, G. D. (2012). Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One*, *7*(4), e34953. https://doi.org/10.1371/journal.pone.0034953

Bissonnette, L., & Bergeron, M. G. (2010). Diagnosing infections - current and anticipated technologies for point-of-care diagnostics and home-based testing. *Clin Microbiol Infect*, *16*(8), 1044-1053. https://doi.org/10.1111/j.1469-0691.2010.03282.x

Blaser, M. J. (2018). Our missing microbes: Short-term antibiotic courses have long-term consequences. *Cleve Clin J Med*, *85*(12), 928-930. https://doi.org/10.3949/ccjm.85gr.18005

Boehme, C. C., Nabeta, P., Hillemann, D., Nicol, M. P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., Milovic, A., Jones, M., O'Brien, S. M., Persing, D. H., Ruesch-Gerdes, S., Gotuzzo, E., Rodrigues, C., Alland, D., & Perkins, M. D. (2010). Rapid Molecular Detection of Tuberculosis and Rifampin Resistance. *N Eng J Med*, *363*(11). https://doi.org/10.1056/NEJMoa0907847

Boolchandani, M., D'Souza, A. W., & Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet*, *20*(6), 356-370. https://doi.org/10.1038/s41576-019-0108-4

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T.,

Neumann, B., Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., . . . Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*, *75*(12), 3491-3500. https://doi.org/10.1093/jac/dkaa345

Brennan, P. J. (2003). Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis*, *83*(1-3), 91-97. https://doi.org/10.1016/s1472-9792(02)00089-6

Brochet, M., Couve, E., Zouine, M., Poyart, C., & Glaser, P. (2008). A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *J Bacteriol*, *190*(2), 672-680. https://doi.org/10.1128/JB.01357-07

Brooks, B., Olm, M. R., Firek, B. A., Baker, R., Thomas, B. C., Morowitz, M. J., & Banfield, J. F. (2017). Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nat Commun*, *8*(1), 1814. https://doi.org/10.1038/s41467-017-02018-w

Brown, A. C., Bryant, J. M., Einer-Jensen, K., Holdstock, J., Houniet, D. T., Chan, J. Z., Depledge, D. P., Nikolayevskyy, V., Broda, A., Stone, M. J., Christiansen, M. T., Williams, R., McAndrew, M. B., Tutill, H., Brown, J., Melzer, M., Rosmarin, C., McHugh, T. D., Shorten, R. J., . . . Breuer, J. (2015). Rapid Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples. *J Clin Microbiol*, *53*(7), 2230-2237. https://doi.org/10.1128/JCM.00486-15

Brown, D. F. J., & Walpole, E. (2001). Evaluation of the Mastalex latex agglutination test for methicillin resistance in *Staphylococcus aureus* grown on different screening media. *Journal of Antimicrobial Chemotherapy*, *47*(2). https://doi.org/10.1093/jac/47.2.187

Brown, E. D., & Wright, G. D. (2016). Antibacterial drug discovery in the resistance era. *Nature*, *529*(7586), 336-343. https://doi.org/10.1038/nature17042

Bugg, T. D. H., Wright, G. D., Dutka-Malen, S., Arthur, M., Courvalin, P., & Walsh, C. T. (1991). Molecular Basis for Vancomycin Resistance in *Enterococcus faecium* BM4147: Biosynthesis of a Depsipeptide Peptidoglycan Precursor by Vancomycin Resistance Proteins VanH and VanA. *Biochemistry*, *30*. https://doi.org/10.1021/bi00107a007,

Bush, K. (2018). Past and Present Perspectives on β-Lactamases. *Antimicrob Agents Chemother*, *62*(10). https://doi.org/10.1128/AAC.01076-18

Bush, K., & Jacoby, G. A. (2010). Updated functional classification of β-lactamases. *Antimicrob Agents Chemother*, *54*(3), 969-976. https://doi.org/10.1128/AAC.01009-09

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S., & Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol*, *2*, 16242. https://doi.org/10.1038/nmicrobiol.2016.242

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, *407*. https://doi.org/10.1038/35030019

Chopra, I., & Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev*, *65*(2), 232-260 https://doi.org/10.1128/MMBR.65.2.232-260.2001

Clausen, P. T., Zankari, E., Aarestrup, F. M., & Lund, O. (2016). Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *J Antimicrob Chemother*, *71*(9), 2484-2488. https://doi.org/10.1093/jac/dkw184

CLSI. (2018). Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically. *CLSI Standard M07, 11 ed.* .

Coelho, L. P., Alves, R., Del Rio, A. R., Myers, P. N., Cantalapiedra, C. P., Giner-Lamia, J., Schmidt, T. S., Mende, D. R., Orakov, A., Letunic, I., Hildebrand, F., Van Rossum, T., Forslund, S. K., Khedkar, S., Maistrenko, O. M., Pan, S., Jia, L., Ferretti, P., Sunagawa, S., . . . Bork, P. (2022). Towards the biogeography of prokaryotic genes. *Nature*, *601*(7892), 252-256. https://doi.org/10.1038/s41586-021-04233-4

Cox, G., & Wright, G. D. (2013). Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *Int J Med Microbiol*, *303*(6-7), 287-292. https://doi.org/10.1016/j.ijmm.2013.02.009

Crofts, T. S., Gasparrini, A. J., & Dantas, G. (2017). Next-generation approaches to understand and combat the antibiotic resistome. *Nat Rev Microbiol*, *15*(7), 422-434. https://doi.org/10.1038/nrmicro.2017.28

Cundliffe, E. (1989). How antibiotic-producing organisms avoid suicide. *Annu. Rev. Microbiol.*, *43*. https://doi.org/annurev.mi.43.100189.001231

Cundliffe, E. (1992). Resistance to macrolides and lincosamides in *Streptomyces lividans* and to aminoglycosides in *Micromonospora purpurea*. *Gene*, *115*. https://doi.org/10.1016/0378-1119(92)90543-x

Cunningham, A. A., Daszak, P., & Wood, J. L. N. (2017). One Health, emerging infectious diseases and wildlife: two decades of progress? *Philos Trans R Soc Lond B Biol Sci*, *372*(1725). https://doi.org/10.1098/rstb.2016.0167

D'Costa, V. M., Griffiths, E., & Wright, G. D. (2007). Expanding the soil antibiotic resistome: exploring environmental diversity. *Curr Opin Microbiol*, *10*(5), 481-489. https://doi.org/10.1016/j.mib.2007.08.009

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., Golding, G. B., Poinar, H. N., & Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature*, *477*(7365), 457-461. https://doi.org/10.1038/nature10388

D'Costa, V. M., McGrann, K. M., Hughes, D. W., & Wright, G. D. (2006). Sampling the Antibiotic Resistome. *Science*, *311*(5759). https://doi.org/10.1126/science.1120800

Danko, D., Bezdan, D., Afshin, E. E., Ahsanuddin, S., Bhattacharya, C., Butler, D. J., Chng, K. R., Donnellan, D., Hecht, J., Jackson, K., Kuchin, K., Karasikov, M., Lyons, A., Mak, L., Meleshko, D., Mustafa, H., Mutai, B., Neches, R. Y., Ng, A., . . . International Meta, S. U. B. C. (2021). A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell*, *184*(13), 3376-3393 e3317. https://doi.org/10.1016/j.cell.2021.05.002

Dantas, G., Sommer, M. O. A., Oluwasegun, R. D., & Church, G. M. (2008). Bacteria Subsisting on Antibiotics. *Science*, *320*. https://doi.org/10.1126/science.1155157

Dapprich, J., Ferriola, D., Mackiewicz, K., Clark, P. M., Rappaport, E., D'Arcy, M., Sasson, A., Gai, X., Schug, J., Kaestner, K. H., & Monos, D. (2016). The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics*, *17*, 486. https://doi.org/10.1186/s12864-016-2836-6

Datta, N. (1962). Transmissible drug resistance in an epidemic strain of *Salmonella typhimurium*. *Journal of Hygiene*, *60*. https://doi.org/10.1017/s0022172400020416

Datta, N., & Hughes, V. M. (1982). Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics. *Nature*, *306*. https://doi.org/10.1038/306616a0

Dave, N., & Liu, J. (2010). Fast Molecular Beacon Hybridization in Organic Solvents with Improved Target Specificity. *The Journal of Physical Chemistry B*, *114*(47), 15694-15699. https://doi.org/10.1021/jp106754k

Davies, J. (1995). Vicious Circles: Looking Back on Resistance Plasmids. *Genetics*, *139*. https://doi.org/10.1093/genetics/139.4.1465

Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev*, *74*(3), 417-433. https://doi.org/10.1128/MMBR.00016-10

de Goffau, M. C., Lager, S., Salter, S. J., Wagner, J., Kronbichler, A., Charnock-Jones, D. S., Peacock, S. J., Smith, G. C. S., & Parkhill, J. (2018). Recognizing the reagent microbiome. *Nat Microbiol*, *3*(8), 851-853. https://doi.org/10.1038/s41564-018-0202-y

Del Grosso, M., Iannelli, F., Messina, C., Santagati, M., Petrosillo, N., Stefani, S., Pozzi, G., & Pantosti, A. (2002). Macrolide efflux genes *mef(A)* and *mef(E)* are carried by different genetic elements in *Streptococcus pneumoniae*. *J Clin Microbiol*, *40*(3), 774-778. https://doi.org/10.1128/JCM.40.3.774-778.2002

Demain, A. L., & Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *J Antibiot (Tokyo)*, *62*(1), 5-16. https://doi.org/10.1038/ja.2008.16

Dickson, Z. W., Hackenberger, D., Kuch, M., Marzok, A., Banerjee, A., Rossi, L., Klowak, J. A., Fox-Robichaud, A., Mossmann, K., Miller, M. S., Surette, M. G.,

Golding, G. B., & Poinar, H. (2021). Probe design for simultaneous, targeted capture of diverse metagenomic targets. *Cell Reports Methods*, *1*(6), 100069. https://doi.org/https://doi.org/10.1016/j.crmeth.2021.100069

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet*, *13*(9), 601-612. https://doi.org/10.1038/nrg3226

Ding, P., Lu, J., Wang, Y., Schembri, M. A., & Guo, J. (2022). Antidepressants promote the spread of antibiotic resistance via horizontally conjugative gene transfer. *Environ Microbiol*. https://doi.org/10.1111/1462-2920.16165

Division, W. A. R. (2016). Global action plan on antimicrobial resistance. *World Health Organization*, 45.

Domingues, S., da Silva, G. J., & Nielsen, K. M. (2012). Integrons: Vehicles and pathways for horizontal dissemination in bacteria. *Mob Genet Elements*, *2*(5), 211-223. https://doi.org/10.4161/mge.22967

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*, *107*(26), 11971-11975. https://doi.org/10.1073/pnas.1002601107

Donskey, C. J. (2004). The Role of the Intestinal Tract as a Reservoir and Source for Transmission of Nosocomial Pathogens. *Clin Infect Dis*, *39*. https://doi.org/10.1086/422002

Dunivin, T. K., Choi, J., Howe, A., & Shade, A. (2019). RefSoil+: a Reference Database for Genes and Traits of Soil Plasmids. *mSystems*, *4*(1). https://doi.org/10.1128/mSystems.00349-18

Eckert, S. E., Chan, J. Z., Houniet, D., The Pathseek, C., Breuer, J., & Speight, G. (2016). Enrichment by hybridisation of long DNA fragments for Nanopore sequencing. *Microb Genom*, *2*(9), e000087. https://doi.org/10.1099/mgen.0.000087

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends in Microbiology*, *27*(2), 105-117. https://doi.org/https://doi.org/10.1016/j.tim.2018.11.003

Enk, J. M., Devault, A. M., Kuch, M., Murgha, Y. E., Rouillard, J. M., & Poinar, H. N. (2014). Ancient whole genome enrichment using baits built from modern DNA. *Mol Biol Evol*, *31*(5), 1292-1294. https://doi.org/10.1093/molbev/msu074

Escamilla, J. (1976). Susceptibility of *Haemophilus influenzae* to ampicillin as determined by use of a modified, one-minute beta-lactamase test. *Antimicrob Agents Chemother*, *9*(1). https://doi.org/10.1128/aac.9.1.196

EUCAST. (2003). Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. *Clinical Microbiology and Infection*, *9*(8), ix-xv. https://doi.org/10.1046/j.1469-0691.2003.00790.x

Ferreira, I., Lepuschitz, S., Beisken, S., Fiume, G., Mrazek, K., Frank, B. J. H., Huber, S., Knoll, M. A., von Haeseler, A., Materna, A., Hofstaetter, J. G., Posch, A. E., & Weinberger, J. (2021). Culture-Free Detection of Antibiotic Resistance Markers from Native Patient Samples by Hybridization Capture Sequencing. *Microorganisms*, *9*(8). https://doi.org/10.3390/microorganisms9081672

Fillgrove, K. L., Pakhomova, S., Newcomer, M. E., & Armstrong, R. N. (2003). Mechanistic Diversity of Fosfomycin Resistance in Pathogenic Microorganisms. *J. Am. Chem. Soc*, *125*. https://doi.org/10.1021/ja039307z

Finland, M. (1979). Emergence of Antibiotic Resistance in Hospitals, 1935-1975. *Reviews of infectious Diseases*, *1*(1), 19. https://doi.org/10.1093/clinids/1.1.4

Finley, R. L., Collignon, P., Larsson, D. G., McEwen, S. A., Li, X. Z., Gaze, W. H., Reid-Smith, R., Timinouni, M., Graham, D. W., & Topp, E. (2013). The scourge of antibiotic resistance: the important role of the environment. *Clin Infect Dis*, *57*(5), 704-710. https://doi.org/10.1093/cid/cit355

Fleming, A. (1929). On the antibacterial action of cultures of a Penicillium, with special reference to their use in the isolation of *B. influenzae Br J Exp Pathol*, *10*(3), 11.

Flint, H. J., Scott, K. P., Louis, P., & Duncan, S. H. (2012). The role of the gut microbiota in nutrition and health. *Nature reviews. Gastroenterology & hepatology*, *9*(10). https://doi.org/10.1038/nrgastro.2012.156

Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., & Dantas, G. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature*, *509*(7502), 612-616. https://doi.org/10.1038/nature13377

Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. A., & Dantas, G. (2012). The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science*, *337*(6098). https://doi.org/10.1126/science.1220761

Fouhy, F., Guinane, C. M., Hussey, S., Wall, R., Ryan, C. A., Dempsey, E. M., Murphy, B., Ross, R. P., Fitzgerald, G. F., Stanton, C., & Cotter, P. D. (2012). High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. *Antimicrob Agents Chemother*, *56*(11), 5811-5820. https://doi.org/10.1128/AAC.00789-12

Frohlich, C., Chen, J. Z., Gholipour, S., Erdogan, A. N., & Tokuriki, N. (2021). Evolution of β-lactamases and enzyme promiscuity. *Protein Eng Des Sel*, *34*. https://doi.org/10.1093/protein/gzab013

Galimand, M., Courvalin, P., & Lambert, T. (2003). Plasmid-mediated high-level resistance to aminoglycosides in *Enterobacteriaceae* due to 16S rRNA methylation. *Antimicrob Agents Chemother*, *47*(8), 2565-2571. https://doi.org/10.1128/AAC.47.8.2565-2571.2003

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol*, *3*(9), 710-720. https://doi.org/10.1038/nri1180

Garcia-Garcia, G., Baux, D., Faugere, V., Moclyn, M., Koenig, M., Claustres, M., & Roux, A. F. (2016). Assessment of the latest NGS enrichment capture methods in clinical context. *Sci Rep*, *6*, 20948. https://doi.org/10.1038/srep20948

Gasc, C., Peyretaillade, E., & Peyret, P. (2016). Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Res*, *44*(10), 4504-4518. https://doi.org/10.1093/nar/gkw309

Gasparrini, A. J., Crofts, T. S., Gibson, M. K., Tarr, P. I., Warner, B. B., & Dantas, G. (2016). Antibiotic perturbation of the preterm infant gut microbiome and resistome. *Gut Microbes*, *7*(5), 443-449. https://doi.org/10.1080/19490976.2016.1218584

Gibson, M. K., Forsberg, K. J., & Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J*, *9*(1), 207-216. https://doi.org/10.1038/ismej.2014.106

Gibson, M. K., Wang, B., Ahmadi, S., Burnham, C. A., Tarr, P. I., Warner, B. B., & Dantas, G. (2016). Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat Microbiol*, *1*, 16024. https://doi.org/10.1038/nmicrobiol.2016.24

Gillespie, S. H., Basu, S., Dickens, A. L., O'Sullivan, D. M., & McHugh, T. D. (2005). Effect of subinhibitory concentrations of ciprofloxacin on *Mycobacterium fortuitum* mutation rates. *J Antimicrob Chemother*, *56*(2), 344-348. https://doi.org/10.1093/jac/dki191

Goh, E. B., Yim, G., Tsui, W., McClure, J., Surette, M. G., & Davies, J. (2002). Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci U S A*, *99*(26), 17025-17030. https://doi.org/10.1073/pnas.252607699

Gottlieb, D. (1976). The production and role of antibiotics in soil. *The Journal of Antibiotics*, *29*. https://doi.org/10.7164/antibiotics.29.987

Grkovic, S., Brown, M. H., & Skurray, R. A. (2002). Regulation of Bacterial Drug Export Systems. *Microbiol Mol Biol Rev*, *66*(4). https://doi.org/10.1128/MMBR.66.4.671-701.2002

Guernier-Cambert, V., Chamings, A., Collier, F., & Alexandersen, S. (2021). Diverse Bacterial Resistance Genes Detected in Fecal Samples From Clinically Healthy Women and Infants in Australia-A Descriptive Pilot Study. *Front Microbiol*, *12*, 596984. https://doi.org/10.3389/fmicb.2021.596984

Guitor, A. K., Raphenya, A. R., Klunk, J., Kuch, M., Alcock, B., Surette, M. G., McArthur, A. G., Poinar, H. N., & Wright, G. D. (2019). Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. *Antimicrob Agents Chemother*, *64*(1). https://doi.org/10.1128/AAC.01324-19

Guitor, A. K., & Wright, G. D. (2018). Antimicrobial Resistance and Respiratory Infections. *Chest*, *154*(5), 1202-1212. https://doi.org/10.1016/j.chest.2018.06.019

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., & Rolain, J. M. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother*, *58*(1), 212-220. https://doi.org/10.1128/AAC.01310-13

Hall, B. G., & Barlow, M. (2004). Evolution of the serine β-lactamases: past, present and future. *Drug Resist Updat*, *7*(2), 111-123. https://doi.org/10.1016/j.drup.2004.02.003

Handel, A., Regoes, R. R., & Antia, R. (2006). The role of compensatory mutations in the emergence of drug resistance. *PLoS Comput Biol*, *2*(10), e137. https://doi.org/10.1371/journal.pcbi.0020137

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C. A., Frimodt-Moller, N., & Aarestrup, F. M. (2014). Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J Clin Microbiol*, *52*(1), 139-146. https://doi.org/10.1128/JCM.02452-13

Hayden, H. S., Joshi, S., Radey, M. C., Vo, A. T., Forsberg, C., Morgan, S. J., Waalkes, A., Holmes, E. A., Klee, S. M., Emond, M. J., Singh, P. K., & Salipantec, S. J. (2022). Genome Capture Sequencing Selectively Enriches Bacterial DNA and Enables Genome-Wide Measurement of Intrastrain Genetic Diversity in Human Infections. *mBio*, *13*(5). https://doi.org/10.1128/mbio.01424-22

Healy, D. B., Ryan, C. A., Ross, R. P., Stanton, C., & Dempsey, E. M. (2022). Clinical implications of preterm infant gut microbiome development. *Nat Microbiol*, *7*(1), 22-33. https://doi.org/10.1038/s41564-021-01025-4

Hooper, D. C., & Jacoby, G. A. (2015). Mechanisms of drug resistance: quinolone resistance. *Ann N Y Acad Sci*, *1354*, 12-31. https://doi.org/10.1111/nyas.12830

Hopwood, D. A. (2007). How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol Microbiol*, *63*(4), 937-940. https://doi.org/10.1111/j.1365-2958.2006.05584.x

Hoshino, Y., Fujii, S., Shinonaga, H., Arai, K., Saito, F., Fukai, T., Satoh, H., Miyazaki, Y., & Ishikawa, J. (2010). Monooxygenation of rifampicin catalyzed by the *rox* gene product of *Nocardia farcinica*: structure elucidation, gene identification and role in drug resistance. *J Antibiot (Tokyo)*, *63*(1), 23-28. https://doi.org/10.1038/ja.2009.116

Hrabak, J., Walkova, R., Studentova, V., Chudackova, E., & Bergerova, T. (2011). Carbapenemase activity detection by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, *49*(9), 3222-3227. https://doi.org/10.1128/JCM.00984-11

Hughes, V. M., & Datta, N. (1983). Conjugative plasmids in bacteria of the 'pre-antibiotic' era. *Nature*, *302*. https://doi.org/Conjugative plasmids in bacteria of the 'pre-antibiotic' era

Humeniuk, C., Arlet, G., Gautier, V., Grimont, P., Labia, R., & Philippon, A. (2002). β-lactamases of *Kluyvera ascorbata*, probable progenitors of some plasmid-encoded

CTX-M types. *Antimicrob Agents Chemother*, *46*(9), 3045-3049. https://doi.org/10.1128/AAC.46.9.3045-3049.2002

Huovinen, P. (2001). Resistance to Trimethoprim-Sulfamethoxazole. *Clin Infect Dis*, *32*, 7. https://doi.org/10.1086/320532

Ince, D., & Hooper, D. C. (2003). Quinolone resistance due to reduced target enzyme expression. *J Bacteriol*, *185*(23), 6883-6892. https://doi.org/10.1128/JB.185.23.6883-6892.2003

Inouye, M., Dashnow, H., & Raven, L. A. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med*, *6*. https://doi.org/10.1186/s13073-014-0090-6

Jiang, X., Ellabaan, M. M. H., Charusanti, P., Munck, C., Blin, K., Tong, Y., Weber, T., Sommer, M. O. A., & Lee, S. Y. (2017). Dissemination of antibiotic resistance genes from antibiotic producers to pathogens. *Nat Commun*, *8*, 15784. https://doi.org/10.1038/ncomms15784

Jin, J., Yamamoto, R., Takeuchi, T., Cui, G., Miyauchi, E., Hojo, N., Ikuta, K., Ohno, H., & Shiroguchi, K. (2022). High-throughput identification and quantification of single bacterial cells in the microbiota. *Nat Commun*, *13*(1), 863. https://doi.org/10.1038/s41467-022-28426-1

Joensen, K. G., Engsbro, A. L. O., Lukjancenko, O., Kaas, R. S., Lund, O., Westh, H., & Aarestrup, F. M. (2017). Evaluating next-generation sequencing for direct clinical diagnostics in diarrhoeal disease. *Eur J Clin Microbiol Infect Dis*, *36*(7), 1325-1338. https://doi.org/10.1007/s10096-017-2947-2

Jones, C., & Stanley, J. (1992). Salmonella plasmids of the pre-antibiotic era. *Journal of General Microbiology*, *138*. https://doi.org/10.1099/00221287-138-1-189

Jones, D. S., Podolsky, S. H., & Greene, J. A. (2012). The Burden of Disease and the Changing Task of Medicine. *The New England Journal of Medicine*, *366*(25), 6. https://doi.org/10.1056/NEJMp1113569

Kaminski, J., Gibson, M. K., Franzosa, E. A., Segata, N., Dantas, G., & Huttenhower, C. (2015). High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *PLoS Comput Biol*, *11*(12), e1004557. https://doi.org/10.1371/journal.pcbi.1004557

Kang, J. T. L., Teo, J. J. Y., Bertrand, D., Ng, A., Ravikrishnan, A., Yong, M., Ng, O. T., Marimuthu, K., Chen, S. L., Chng, K. R., Gan, Y. H., & Nagarajan, N. (2022). Long-term ecological and evolutionary dynamics in the gut microbiomes of carbapenemase-producing Enterobacteriaceae colonized subjects. *Nat Microbiol*, *7*(10), 1516-1524. https://doi.org/10.1038/s41564-022-01221-w

Karlowsky, J. A., & Richter, S. S. (2015). Antimicrobial Susceptibility Testing Systems. In J. H. Jorgensen, K. C. Carroll, G. Funke, M. A. Pfaller, M. L. Landry, S. S. Richter, & D. W. Warnock (Eds.), *Manual of Clinical Microbiology, 11th Edition*.

Kennedy, K. M., Gerlach, M. J., Adam, T., Heimesaat, M. M., Rossi, L., Surette, M. G., Sloboda, D. M., & Braun, T. (2021). Fetal meconium does not have a detectable

microbiota before birth. *Nat Microbiol*, *6*(7), 865-873. https://doi.org/10.1038/s41564-021-00904-0

Kent, A. G., Vill, A. C., Shi, Q., Satlin, M. J., & Brito, I. L. (2020). Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat Commun*, *11*(1), 4379. https://doi.org/10.1038/s41467-020-18164-7

Kharel, M. K., Subba, B., Basnet, D. B., Woo, J. S., Lee, H. C., Liou, K., & Sohng, J. K. (2004). A gene cluster for biosynthesis of kanamycin from *Streptomyces kanamyceticus*: comparison with gentamicin biosynthetic gene cluster. *Arch Biochem Biophys*, *429*(2), 204-214. https://doi.org/10.1016/j.abb.2004.06.009

Knapp, C. W., Dolfing, J., Ehlert, P. A. I., & Graham, D. W. (2010). Evidence of Increasing Antibiotic Resistance Gene Abundances in Archived Soils since 1940. *Environ. Sci. Technol.*, *44*. https://doi.org/10.1021/es901221x

Knapp, C. W., Engemann, C. A., Hanson, M. L., Keen, P. L., Hall, K. J., & Graham, D. W. (2008). Indirect Evidence of Transposon-Mediated Selection of Antibiotic Resistance Genes in Aquatic Systems at Low-Level Oxytetracycline Exposures. *Environ Sci Technol*, *42*(14). https://doi.org/10.1021/es703199g

Koteva, K., Cox, G., Kelso, J. K., Surette, M. D., Zubyk, H. L., Ejim, L., Stogios, P., Savchenko, A., Sorensen, D., & Wright, G. D. (2018). Rox, a Rifamycin Resistance Enzyme with an Unprecedented Mechanism of Action. *Cell Chem Biol*, *25*(4), 403-412 e405. https://doi.org/10.1016/j.chembiol.2018.01.009

Krause, K. M., Serio, A. W., Kane, T. R., & Connolly, L. E. (2016). Aminoglycosides: An Overview. *Cold Spring Harb Perspect Med*, *6*(6). https://doi.org/10.1101/cshperspect.a027029

Lanza, V. F., Baquero, F., Martinez, J. L., Ramos-Ruiz, R., Gonzalez-Zorn, B., Andremont, A., Sanchez-Valenzuela, A., Ehrlich, S. D., Kennedy, S., Ruppe, E., van Schaik, W., Willems, R. J., de la Cruz, F., & Coque, T. M. (2018). In-depth resistome analysis by targeted metagenomics. *Microbiome*, *6*(1), 11. https://doi.org/10.1186/s40168-017-0387-y

Larsson, D. G. J., & Flach, C. F. (2022). Antibiotic resistance in the environment. *Nat Rev Microbiol*, *20*(5), 257-269. https://doi.org/10.1038/s41579-021-00649-x

Lasa, A., di Cesare, A., Tassistro, G., Borello, A., Gualdi, S., Furones, D., Carrasco, N., Cheslett, D., Brechon, A., Paillard, C., Bidault, A., Pernet, F., Canesi, L., Edomi, P., Pallavicini, A., Pruzzo, C., & Vezzulli, L. (2019). Dynamics of the Pacific oyster pathobiota during mortality episodes in Europe assessed by 16S rRNA gene profiling and a new target enrichment next-generation sequencing strategy. *Environ Microbiol*, *21*(12), 4548-4562. https://doi.org/10.1111/1462-2920.14750

Lebeaux, R. M., Karalis, D. B., Lee, J., Whitehouse, H. C., Madan, J. C., Karagas, M. R., & Hoen, A. G. (2022). The association between early life antibiotic exposure and the gut resistome of young children: a systematic review. *Gut Microbes*, *14*(1), 2120743. https://doi.org/10.1080/19490976.2022.2120743

Ledeboer, N. A., & Hodinka, R. L. (2011). Molecular Detection of Resistance Determinants. *Journal of Clinical Microbiology*, *49*(9_Supplement). https://doi.org/10.1128/jcm.00771-11

Leekha, S., Terrell, C. L., & Edson, R. S. (2011). General principles of antimicrobial therapy. *Mayo Clin Proc*, *86*(2), 156-167. https://doi.org/10.4065/mcp.2010.0639

Leo, S., Cherkaoui, A., Renzi, G., & Schrenzel, J. (2020). Mini Review: Clinical Routine Microbiology in the Era of Automation and Digital Health. *Front Cell Infect Microbiol*, *10*, 582028. https://doi.org/10.3389/fcimb.2020.582028

Leo, S., Curtis, N., & Zimmermann, P. (2022). The neonatal intestinal resistome and factors that influence it-a systematic review. *Clin Microbiol Infect*. https://doi.org/10.1016/j.cmi.2022.07.014

Lerminiaux, N. A., & Cameron, A. D. S. (2019). Horizontal transfer of antibiotic resistance genes in clinical environments. *Can J Microbiol*, *65*(1), 34-44. https://doi.org/10.1139/cjm-2018-0275

Leviatan, S., Shoer, S., Rothschild, D., Gorodetski, M., & Segal, E. (2022). An expanded reference map of the human gut microbiome reveals hundreds of previously unknown species. *Nat Commun*, *13*(1), 3863. https://doi.org/10.1038/s41467-022-31502-1

Lewis, J. S., & Bush, K. (2015). Antibacterial Agents. In J. H. Jorgensen & K. C. Carroll (Eds.), *Manual of Clinical Microbiology, 11th Edition*. ASM Press.

Li, H. Z., Yang, K., Liao, H., Lassen, S. B., Su, J. Q., Zhang, X., Cui, L., & Zhu, Y. G. (2022). Active antibiotic resistome in soils unraveled by single-cell isotope probing and targeted metagenomics. *Proc Natl Acad Sci U S A*, *119*(40), e2201473119. https://doi.org/10.1073/pnas.2201473119

Liu, B., & Pop, M. (2009). ARDB - Antibiotic Resistance Genes Database. *Nucleic Acids Res*, *37*(Database issue), D443-447. https://doi.org/10.1093/nar/gkn656

Liu, Y.-Y., Wang, Y., Walsh, T. R., Yi, L.-X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L.-F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J.-H., & Shen, J. (2016). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases*, *16*(2), 161-168. https://doi.org/10.1016/s1473-3099(15)00424-7

Lou, Y. C., Olm, M. R., Diamond, S., Crits-Christoph, A., Firek, B. A., Baker, R., Morowitz, M. J., & Banfield, J. F. (2021). Infant gut strain persistence is associated with maternal origin, phylogeny, and traits including surface adhesion and iron acquisition. *Cell Rep Med*, *2*(9), 100393. https://doi.org/10.1016/j.xcrm.2021.100393

Luo, N., Pereira, S., Sahin, O., Lin, J., Huang, S., Michel, L., & Zhang, Q. (2005). Enhanced in vivo fitness of fluoroquinolone-resistant *Campylobacter jejuni* in the absence of antibiotic selection pressure. *Proc Natl Acad Sci U S A*, *102*(3), 541-546. https://doi.org/10.1073/pnas.0408966102

MacPherson, C. W., Mathieu, O., Tremblay, J., Champagne, J., Nantel, A., Girard, S. A., & Tompkins, T. A. (2018). Gut Bacterial Microbiota and its Resistome Rapidly Recover to Basal State Levels after Short-term Amoxicillin-Clavulanic Acid Treatment in Healthy Adults. *Sci Rep*, *8*(1), 11192. https://doi.org/10.1038/s41598-018-29229-5

Manaia, C. M. (2017). Assessing the Risk of Antibiotic Resistance Transmission from the Environment to Humans: Non-Direct Proportionality between Abundance and Risk. *Trends Microbiol*, *25*(3), 173-181. https://doi.org/10.1016/j.tim.2016.11.014

Marshall, C. G., Lessard, I. A. D., Park, I.-S., & Wright, G. D. (1998). Glycopeptide Antibiotic Resistance Genes in Glycopeptide-Producing Organisms. *Antimicrob Agents Chemother*, *42*(9). https://doi.org/10.1128/AAC.42.9.2215

Martinez, J. L., & Baquero, F. (2014). Emergence and spread of antibiotic resistance: setting a parameter space. *Ups J Med Sci*, *119*(2), 68-77. https://doi.org/10.3109/03009734.2014.901444

Martinez, J. L., Coque, T. M., & Baquero, F. (2015). What is a resistance gene? Ranking risk in resistomes. *Nat Rev Microbiol*, *13*(2), 116-123. https://doi.org/10.1038/nrmicro3399

Martino, C., Dilmore, A. H., Burcham, Z. M., Metcalf, J. L., Jeste, D., & Knight, R. (2022). Microbiota succession throughout life from the cradle to the grave. *Nat Rev Microbiol*, *20*(12), 707-720. https://doi.org/10.1038/s41579-022-00768-z

Maus, C. E., Plikaytis, B. B., & Shinnick, T. M. (2005). Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, *49*(8), 3192-3197. https://doi.org/10.1128/AAC.49.8.3192-3197.2005

McDonnell, L., Gilkes, A., Ashworth, M., Rowland, V., Harries, T. H., Armstrong, D., & White, P. (2021). Association between antibiotics and gut microbiome dysbiosis in children: systematic review and meta-analysis. *Gut Microbes*, *13*(1), 1-18. https://doi.org/10.1080/19490976.2020.1870402

McFarland, L. V. (2009). Evidence-based review of probiotics for antibiotic-associated diarrhea and *Clostridium difficile* infections. *Anaerobe*, *15*(6), 274-280. https://doi.org/10.1016/j.anaerobe.2009.09.002

Mell, J. C., & Redfield, R. J. (2014). Natural competence and the evolution of DNA uptake specificity. *J Bacteriol*, *196*(8), 1471-1483. https://doi.org/10.1128/JB.01293-13

Melnikov, A., Galinsky, K., Rogov, P., Fennell, T., Van Tyne, D., Russ, C., Daniels, R., Barnes, K. G., Bochicchio, J., Ndiaye, D., Sene, P. D., Wirth, D. F., Nusbaum, C., Volkman, S. K., Birren, B. W., Gnirke, A., & Neafsey, D. E. (2011). Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol*, *12*(8), R73. https://doi.org/10.1186/gb-2011-12-8-r73

Mercer, T. R., Clark, M. B., Crawford, J., Brunck, M. E., Gerhardt, D. J., Taft, R. J., Nielsen, L. K., Dinger, M. E., & Mattick, J. S. (2014). Targeted sequencing for

gene discovery and quantification using RNA CaptureSeq. *Nat Protoc*, *9*(5), 989-1009. https://doi.org/10.1038/nprot.2014.058

Mertes, F., Elsharawy, A., Sauer, S., van Helvoort, J. M., van der Zaag, P. J., Franke, A., Nilsson, M., Lehrach, H., & Brookes, A. J. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics*, *10*(6), 374-386. https://doi.org/10.1093/bfgp/elr033

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*, *2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Mezger, A., Gullberg, E., Goransson, J., Zorzet, A., Herthnek, D., Tano, E., Nilsson, M., & Andersson, D. I. (2015). A general method for rapid determination of antibiotic susceptibility and species in bacterial infections. *J Clin Microbiol*, *53*(2), 425-432. https://doi.org/10.1128/JCM.02434-14

Miller, A. K., Brannon, M. K., Stevens, L., Johansen, H. K., Selgrade, S. E., Miller, S. I., Hoiby, N., & Moskowitz, S. M. (2011). PhoQ mutations promote lipid A modification and polymyxin resistance of *Pseudomonas aeruginosa* found in colistin-treated cystic fibrosis patients. *Antimicrob Agents Chemother*, *55*(12), 5761-5769. https://doi.org/10.1128/AAC.05391-11

Monier, J. M., Demaneche, S., Delmont, T. O., Mathieu, A., Vogel, T. M., & Simonet, P. (2011). Metagenomic exploration of antibiotic resistance in soil. *Curr Opin Microbiol*, *14*(3), 229-235. https://doi.org/10.1016/j.mib.2011.04.010

Montassier, E., Valdes-Mas, R., Batard, E., Zmora, N., Dori-Bachash, M., Suez, J., & Elinav, E. (2021). Probiotics impact the antibiotic resistance gene reservoir along the human GI tract in a person-specific and antibiotic-dependent manner. *Nat Microbiol*, *6*(8), 1043-1054. https://doi.org/10.1038/s41564-021-00920-0

Moore, A. M., Patel, S., Forsberg, K. J., Wang, B., Bentley, G., Razia, Y., Qin, X., Tarr, P. I., & Dantas, G. (2013). Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLoS One*, *8*(11), e78822. https://doi.org/10.1371/journal.pone.0078822

Moore, I. F., W., H. D., & D., W. G. (2005). Tigecycline Is Modified by the Flavin-Dependent Monooxygenase TetX. *Biochemistry*, *44*(35), 6. https://doi.org/10.1021/bi0506066

Morar, M., & Wright, G. D. (2010). The genomic enzymology of antibiotic resistance. *Annu Rev Genet*, *44*, 25-51. https://doi.org/10.1146/annurev-genet-102209-163517

Moskowitz, S. M., Brannon, M. K., Dasgupta, N., Pier, M., Sgambati, N., Miller, A. K., Selgrade, S. E., Miller, S. I., Denton, M., Conway, S. P., Johansen, H. K., & Hoiby, N. (2012). PmrB mutations promote polymyxin resistance of *Pseudomonas aeruginosa* isolated from colistin-treated cystic fibrosis patients. *Antimicrob Agents Chemother*, *56*(2), 1019-1030. https://doi.org/10.1128/AAC.05829-11

Mukhtar, T. A., Koteva, K. P., Hughes, D. W., & Wright, G. D. (2001). Vgb from *Staphylococcus aureus* Inactivates Streptogramin B Antibiotics by an Elimination Mechanism Not Hydrolysis. *Biochemistry*, *40*(30), 8877-8886. https://doi.org/10.1021/bi0106787

Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., . . . Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, *399*(10325), 629-655. https://doi.org/10.1016/s0140-6736(21)02724-0

Nielsen, T. K., Browne, P. D., & Hansen, L. H. (2022). Antibiotic resistance genes are differentially mobilized according to resistance mechanism. *Gigascience*, *11*. https://doi.org/10.1093/gigascience/giac072

Nordmann, P., Poirel, L., & Dortet, L. (2012). Rapid detection of carbapenemase-producing *Enterobacteriaceae*. *Emerg Infect Dis*, *18*(9), 1503-1507. https://doi.org/10.3201/eid1809.120355

Noyes, N. R., Weinroth, M. E., Parker, J. K., Dean, C. J., Lakin, S. M., Raymond, R. A., Rovira, P., Doster, E., Abdo, Z., Martin, J. N., Jones, K. L., Ruiz, J., Boucher, C. A., Belk, K. E., & Morley, P. S. (2017). Enrichment allows identification of diverse, rare elements in metagenomic resistome-virulome sequencing. *Microbiome*, *5*(1), 142. https://doi.org/10.1186/s40168-017-0361-8

O'Brien, K. S., Emerson, P., Hooper, P. J., Reingold, A. L., Dennis, E. G., Keenan, J. D., Lietman, T. M., & Oldenburg, C. E. (2019). Antimicrobial resistance following mass azithromycin distribution for trachoma: a systematic review. *The Lancet Infectious Diseases*, *19*(1), e14-e25. https://doi.org/10.1016/s1473-3099(18)30444-4

O'Sullivan, D. M., McHugh, T. D., & Gillespie, S. H. (2005). Analysis of *rpoB* and *pncA* mutations in the published literature: an insight into the role of oxidative stress in *Mycobacterium tuberculosis* evolution? *J Antimicrob Chemother*, *55*(5), 674-679. https://doi.org/10.1093/jac/dki069

O'Callaghan, C. H., Morris, A., Kirby, S. M., & Shingler, A. H. (1972). Novel Method for Detection of β-Lactamases by Using a Chromogenic Cephalosporin Substrate. *Antimicrob Agents Chemother*, *1*(4). https://doi.org/10.1128/AAC.1.4.283

O'Neill, J. (2016). Tackling drug-resistant infections globally: Final report and recommendations. *London: Review on Antimicrobial Resistance*.

Osbelt, L., Wende, M., Almasi, E., Derksen, E., Muthukumarasamy, U., Lesker, T. R., Galvez, E. J. C., Pils, M. C., Schalk, E., Chhatwal, P., Farber, J., Neumann-Schaal, M., Fischer, T., Schluter, D., & Strowig, T. (2021). *Klebsiella oxytoca* causes colonization resistance against multidrug-resistant *K. pneumoniae* in the gut via cooperative carbohydrate competition. *Cell Host Microbe*, *29*(11), 1663-1679 e1667. https://doi.org/10.1016/j.chom.2021.09.003

Pagès, J. M., James, C. E., & Winterhalter, M. (2008). The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. *Nat Rev Microbiol*, *6*(12), 893-903. https://doi.org/10.1038/nrmicro1994

Papaiakovou, M., Littlewood, D. T. J., Doyle, S. R., Gasser, R. B., & Cantacessi, C. (2022). Worms and bugs of the gut: the search for diagnostic signatures using barcoding, and metagenomics-metabolomics. *Parasit Vectors*, *15*(1), 118. https://doi.org/10.1186/s13071-022-05225-7

Patel, J. B., & Richter, S. S. (2015). Mechanisms of Resistance to Antibacterial Agents. In J. H. Jorgensen, K. C. Carroll, G. Funke, M. A. Pfaller, M. L. Landry, S. S. Richter, D. W. Warnock, S. S. Richter, & J. B. Patel (Eds.), *Manual of Clinical Microbiology*.

Pawlowski, A. C., Wang, W., Koteva, K., Barton, H. A., McArthur, A. G., & Wright, G. D. (2016). A diverse intrinsic antibiotic resistome from a cave bacterium. *Nat Commun*, *7*, 13803. https://doi.org/10.1038/ncomms13803

Pehrsson, E. C., Forsberg, K. J., Gibson, M. K., Ahmadi, S., & Dantas, G. (2013). Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front Microbiol*, *4*, 145. https://doi.org/10.3389/fmicb.2013.00145

Peleg, A. Y., Miyakis, S., Ward, D. V., Earl, A. M., Rubio, A., Cameron, D. R., Pillai, S., Moellering, R. C., Jr., & Eliopoulos, G. M. (2012). Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of *Staphylococcus aureus*. *PLoS One*, *7*(1), e28316. https://doi.org/10.1371/journal.pone.0028316

Peñalba, J. V., Smith, L. L., Tonione, M. A., Sass, C., Hykin, S. M., Skipwith, P. L., McGuire, J. A., Bowie, R. C. K., & Moritz, C. (2014). Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Mol Ecol Resource*, *14*. https://doi.org/10.1111/1755-0998.12249

Perez, F., & Villegas, M. V. (2015). The role of surveillance systems in confronting the global crisis of antibiotic-resistant bacteria. *Curr Opin Infect Dis*, *28*(4), 375-383. https://doi.org/10.1097/QCO.0000000000000182

Perry, J., Waglechner, N., & Wright, G. (2016). The Prehistory of Antibiotic Resistance. *Cold Spring Harb Perspect Med*, *6*(6). https://doi.org/10.1101/cshperspect.a025197

Petersen, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., & Tsongalis, G. J. (2019). Third-Generation Sequencing in the Clinical Laboratory- Exploring the Advantages and Challenges of Nanopore Sequencing. *Journal of Clinical Microbiology*, *58*(1). https://doi.org/10.1128/JCM.01315-19

Petrova, M., Gorlenko, Z., & Mindlin, S. (2009). Molecular structure and translocation of a multiple antibiotic resistance region of a *Psychrobacter psychrophilus*

permafrost strain. *FEMS Microbiol Lett*, *296*(2), 190-197. https://doi.org/10.1111/j.1574-6968.2009.01635.x

Petrova, M., Gorlenko, Z., & Mindlin, S. (2011). Tn5045, a novel integron-containing antibiotic and chromate resistance transposon isolated from a permafrost bacterium. *Res Microbiol*, *162*(3), 337-345. https://doi.org/10.1016/j.resmic.2011.01.003

Petrova, M., Kurakov, A., Shcherbatova, N., & Mindlin, S. (2014). Genetic structure and biological properties of the first ancient multiresistance plasmid pKLH80 isolated from a permafrost bacterium. *Microbiology (Reading)*, *160*(Pt 10), 2253-2263. https://doi.org/10.1099/mic.0.079335-0

Piddock, L. J. V. (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nature Reviews Microbiology*, *4*. https://doi.org/10.1038/nrmicro1464

Quan, S., Venter, H., & Dabbs, E. R. (1997). Ribosylative Inactivation of Rifampin by *Mycobacterium smegmatis* Is a Principal Contributor to Its Low Susceptibility to This Antibiotic. *Antimicrob Agents Chemother*, *14*(11). https://doi.org/10.1128/AAC.41.11.2456

Ramotar, K., Woods, W., Larocque, L., & Toye, B. (2000). Comparison of phenotypic methods to identify enterococci intrinsically resistant to Vancomycin (VanC VRE). *Diagnostic Microbiology and Infectious Disease*, *36*. https://doi.org/10.1016/s0732-8893(99)00126-1

Riesenfeld, C. S., Goodman, R. M., & Handelsman, J. (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol*, *6*(9). https://doi.org/10.1111/j.1462-2920.2004.00664.x

Roberts, M. C., Sutcliffe, J., Courvalin, P., Jensen, L. B., Rood, J., & Seppala, H. (1999). Nomenclature for Macrolide and Macrolide-Lincosamide- Streptogramin B Resistance Determinants. *Antimicrob Agents Chemother*, *43*(12). https://doi.org/10.1128/AAC.43.12.2823

Samara, J., Moossavi, S., Alshaikh, B., Ortega, V. A., Pettersen, V. K., Ferdous, T., Hoops, S. L., Soraisham, A., Vayalumkal, J., Dersch-Mills, D., Gerber, J. S., Mukhopadhyay, S., Puopolo, K., Tompkins, T. A., Knights, D., Walter, J., Amin, H., & Arrieta, M. C. (2022). Supplementation with a probiotic mixture accelerates gut microbiome maturation and reduces intestinal inflammation in extremely preterm infants. *Cell Host Microbe*, *30*(5), 696-711 e695. https://doi.org/10.1016/j.chom.2022.04.005

Samuel, L. P., Tibbetts, R. J., Agotesku, A., Fey, M., Hensley, R., & Meier, F. A. (2013). Evaluation of a microarray-based assay for rapid identification of Gram-positive organisms and resistance markers in positive blood cultures. *J Clin Microbiol*, *51*(4), 1188-1192. https://doi.org/10.1128/JCM.02982-12

Sandegren, L., & Andersson, D. I. (2009). Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nature Reviews Microbiology*, *7*. https://doi.org/10.1038/nrmicro2174

Santiago-Rodriguez, T. M., Fornaciari, G., Luciani, S., Dowd, S. E., Toranzos, G. A., Marota, I., & Cano, R. J. (2015). Gut Microbiome of an 11th Century A.D. Pre-Columbian Andean Mummy. *PLoS One*, *10*(9), e0138135. https://doi.org/10.1371/journal.pone.0138135

Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., Khan, A. M., Woodford, N., Saunders, N. J., Wain, J., O'Grady, J., & Livermore, D. M. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother*, *72*(1), 104-114. https://doi.org/10.1093/jac/dkw397

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Shaw, W. V. (1983). Chloramphenicol Acetyltransferase- Enzymology and Molecular Biology. *CRC Critical Reviews in Biochemistry*, *14*(1). https://doi.org/10.3109/10409238309102789

Shkoporov, A. N., Turkington, C. J., & Hill, C. (2022). Mutualistic interplay between bacteriophages and bacteria in the human gut. *Nat Rev Microbiol*, *20*(12), 737-749. https://doi.org/10.1038/s41579-022-00755-4

Silver, L. L. (2017). Fosfomycin: Mechanism and Resistance. *Cold Spring Harb Perspect Med*, *7*(2). https://doi.org/10.1101/cshperspect.a025262

Sjölund, M., Tano, E., Blaser, M. J., Andersson, D. I., & Engstrand, L. (2005). Persistence of Resistant *Staphylococcus epidermidis* after Single Course of Clarithromycin. *Emerging Infectious Diseases*, *11*(9). https://doi.org/10.3201/eid1109.050124

Sjölund, M., Wreiber, K., Andersson, D. I., Blaser, M. J., & Engstrand, L. (2003). Long-term persistence of resistant *Enterococcus* species after antibiotics to eradicate *Helicobacter pylori*. *Annals of Internal Medicine*, *139*(6). https://doi.org/10.7326/0003-4819-139-6-200309160-00011

Skold, O. (2000). Sulfonamide resistance: mechanisms and trends. *Drug Resist Updat*, *3*(3), 155-160. https://doi.org/10.1054/drup.2000.0146

Slizovskiy, I. B., Oliva, M., Settle, J. K., Zyskina, L. V., Prosperi, M., Boucher, C., & Noyes, N. R. (2022). Target-enriched long-read sequencing (TELSeq) contextualizes antimicrobial resistance genes in metagenomes. *Microbiome*, *10*(1), 185. https://doi.org/10.1186/s40168-022-01368-y

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, *480*(7376), 241-244. https://doi.org/10.1038/nature10571

Sommer, M. O. A., Dantas, G., & Church, G. M. (2009). Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science*, *325*(5944). https://doi.org/10.1126/science.1176950

Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S., & Sonnenburg, J. L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature*, *529*(7585), 212-215. https://doi.org/10.1038/nature16504

Sorek, R., Yiwen Zhu, Y., Creevey, C. J., Francino, P., Bork, P., & Rubin, E. M. (2007). Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer. *Science*, *318*. https://doi.org/10.1126/science.1147112

Spanogiannopoulos, P., Thaker, M., Koteva, K., Waglechner, N., & Wright, G. D. (2012). Characterization of a rifampin-inactivating glycosyltransferase from a screen of environmental actinomycetes. *Antimicrob Agents Chemother*, *56*(10), 5061-5069. https://doi.org/10.1128/AAC.01166-12

Spanogiannopoulos, P., Waglechner, N., Koteva, K., & Wright, G. D. (2014). A rifamycin inactivating phosphotransferase family shared by environmental and pathogenic bacteria. *Proc Natl Acad Sci U S A*, *111*(19), 7102-7107. https://doi.org/10.1073/pnas.1402358111

Stracy, M., Snitser, O., Yelin, I., Amer, Y., Parizade, M., Katz, R., Rimler, G., Wolf, T., Herzel, E., Koren, G., Kuint, J., Foxman, B., Chodick, G., Shalev, V., & Kishony, R. (2022). Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. *Science*, *375*(6583), 889-894. https://doi.org/10.1126/science.abg9868

Su, G. L., Ko, C. W., Bercik, P., Falck-Ytter, Y., Sultan, S., Weizman, A. V., & Morgan, R. L. (2020). AGA Clinical Practice Guidelines on the Role of Probiotics in the Management of Gastrointestinal Disorders. *Gastroenterology*, *159*(2), 697-705. https://doi.org/10.1053/j.gastro.2020.05.059

Suez, J., Zmora, N., Segal, E., & Elinav, E. (2019). The pros, cons, and many unknowns of probiotics. *Nat Med*, *25*(5), 716-729. https://doi.org/10.1038/s41591-019-0439-x

Surette, M. D., Spanogiannopoulos, P., & Wright, G. D. (2021). The Enzymes of the Rifamycin Antibiotic Resistome. *Acc Chem Res*, *54*(9), 2065-2075. https://doi.org/10.1021/acs.accounts.1c00048

Surette, M. D., & Wright, G. D. (2017). Lessons from the Environmental Antibiotic Resistome. *Annu. Rev. Microbiol.*, *71*. https://doi.org/10.1146/annurev-micro-

Tojo, M., Fujita, T., Ainoda, Y., Nagamatsu, M., Hayakawa, K., Mezaki, K., Sakurai, A., Masui, Y., Yazaki, H., Takahashi, H., Miyoshi-Akiyama, T., Totsuka, K., Kirikae, T., & Ohmagari, N. (2014). Evaluation of an automated rapid diagnostic assay for detection of Gram-negative bacteria and their drug-resistance genes in positive blood cultures. *PLoS One*, *9*(4), e94064. https://doi.org/10.1371/journal.pone.0094064

Tu, D., Blaha, G., Moore, P. B., & Steitz, T. A. (2005). Structures of MLS_BK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell*, *121*(2), 257-270. https://doi.org/10.1016/j.cell.2005.02.005

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804-810. https://doi.org/10.1038/nature06244

Udikovic-Kolic, N., Wichmann, F., Broderick, N. A., & Handelsman, J. (2014). Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proc Natl Acad Sci U S A*, *111*(42), 15202-15207. https://doi.org/10.1073/pnas.1409836111

UNICEF. (2021). Levels & Trends in Child Mortality - Report 2021. *United Nations Children's Fund (UNICEF)*.

Urbaniak, C., Sielaff, A. C., Frey, K. G., Allen, J. E., Singh, N., Jaing, C., Wheeler, K., & Venkateswaran, K. (2018). Detection of antimicrobial resistance genes associated with the International Space Station environmental surfaces. *Sci Rep*, *8*(1), 814. https://doi.org/10.1038/s41598-017-18506-4

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ*, *361*. https://doi.org/10.1136/bmj.k2179

van der Helm, E., Imamovic, L., Hashim Ellabaan, M. M., van Schaik, W., Koza, A., & Sommer, M. O. A. (2017). Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res*, *45*(8), e61. https://doi.org/10.1093/nar/gkw1328

van Schaik, W. (2015). The human gut resistome. *Philos Trans R Soc Lond B Biol Sci*, *370*(1670), 20140087. https://doi.org/10.1098/rstb.2014.0087

Velkov, T., Roberts, K. D., Nation, R. L., Thompson, P. E., & Li, J. (2013). Pharmacology of polymyxins: new insights into an 'old' class of antibiotics. *Future Microbiol*, *8*(6), 711-724. https://doi.org/10.2217/fmb.13.39

Waglechner, N., Culp, E. J., & Wright, G. D. (2021). Ancient Antibiotics, Ancient Resistance. *EcoSal Plus*, *9*(2). https://doi.org/10.1128/ecosalplus.ESP-0027-2020

Waglechner, N., McArthur, A. G., & Wright, G. D. (2019). Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat Microbiol*, *4*(11), 1862-1871. https://doi.org/10.1038/s41564-019-0531-5

Waglechner, N., & Wright, G. D. (2017). Antibiotic resistance: it's bad, but why isn't it worse? *BMC Biol*, *15*(1), 84. https://doi.org/10.1186/s12915-017-0423-1

Wallace, J. C., Port, J. A., Smith, M. N., & Faustman, E. M. (2017). FARME DB: a functional antibiotic resistance element database. *Database (Oxford)*, *2017*. https://doi.org/10.1093/database/baw165

Wally, N., Schneider, M., Thannesberger, J., Kastner, M. T., Bakonyi, T., Indik, S., Rattei, T., Bedarf, J., Hildebrand, F., Law, J., Jovel, J., & Steininger, C. (2019). Plasmid DNA contaminant in molecular reagents. *Sci Rep*, *9*(1), 1652. https://doi.org/10.1038/s41598-019-38733-1

Walsh, C., & Wencewicz, T. A. (2016). Antibiotics: challenges, mechanisms, opportunities. *ASM Press*.

Warinner, C., Rodrigues, J. F., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S.,

Luder, H. U., Salazar-Garcia, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A., . . . Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nat Genet*, *46*(4), 336-344. https://doi.org/10.1038/ng.2906

Weigel, B. J., Burgett, S. G., Chen, V. J., Skatrud, P. L., Frolik, C. A., Queener, S. W., & Ingolia, T. D. (1988). Cloning and expression in Escherichia coli of isopenicillin N synthetase genes from Streptomyces lipmanii and Aspergillus nidulans. *J Bacteriol*, *170*(9), 3817-3826. https://doi.org/10.1128/jb.170.9.3817-3826.1988

Wheatley, R. M., Caballero, J. D., van der Schalk, T. E., De Winter, F. H. R., Shaw, L. P., Kapel, N., Recanatini, C., Timbermont, L., Kluytmans, J., Esser, M., Lacoma, A., Prat-Aymerich, C., Oliver, A., Kumar-Singh, S., Malhotra-Kumar, S., & Craig MacLean, R. (2022). Gut to lung translocation and antibiotic mediated selection shape the dynamics of *Pseudomonas aeruginosa* in an ICU patient. *Nat Commun*, *13*(1), 6523. https://doi.org/10.1038/s41467-022-34101-2

WHO. (2019). Molecular methods for antimicrobial resistance (AMR) diagnostics to enhance the Global Antimicrobial Resistance Surveillance System. *Geneva: World Health Organization*.

WHO. (2020a). GLASS whole-genome sequencing for surveillance of antimicrobial resistance. *Geneva: World Health Organization*.

WHO. (2020b). WHO guideline on mass drug administration of azithromycin to children under five years of age to promote child survival. *Geneva: World Health Organization*.

Wiegand, I., Hilpert, K., & Hancock, R. E. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc*, *3*(2), 163-175. https://doi.org/10.1038/nprot.2007.521

Wieser, A., Schneider, L., Jung, J., & Schubert, S. (2012). MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). *Appl Microbiol Biotechnol*, *93*(3), 965-974. https://doi.org/10.1007/s00253-011-3783-4

Wright, G. D. (2007). The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol*, *5*(3), 175-186. https://doi.org/10.1038/nrmicro1614

Wright, G. D. (2019). Environmental and clinical antibiotic resistomes, same only different. *Curr Opin Microbiol*, *51*, 57-63. https://doi.org/10.1016/j.mib.2019.06.005

Yang, W., Moore, I. F., Koteva, K. P., Bareich, D. C., Hughes, D. W., & Wright, G. D. (2004). TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. *J Biol Chem*, *279*(50), 52346-52352. https://doi.org/10.1074/jbc.M409573200

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A.,

Lauber, C., Clemente, J. C., Knights, D., . . . Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222-227. https://doi.org/10.1038/nature11053

Yim, G., Wang, H. H., & Davies, J. (2006). The truth about antibiotics. *Int J Med Microbiol*, *296*(2-3), 163-170. https://doi.org/10.1016/j.ijmm.2006.01.039

Yim, G., Wang, H. H., & Davies, J. (2007). Antibiotics as signalling molecules. *Philos Trans R Soc Lond B Biol Sci*, *362*(1483), 1195-1200. https://doi.org/10.1098/rstb.2007.2044

Zapun, A., Contreras-Martel, C., & Vernet, T. (2008). Penicillin-binding proteins and β-lactam resistance. *FEMS Microbiol Rev*, *32*(2), 361-385. https://doi.org/10.1111/j.1574-6976.2007.00095.x

Zhu, Y. G., Johnson, T. A., Su, J. Q., Qiao, M., Guo, G. X., Stedtfeld, R. D., Hashsham, S. A., & Tiedje, J. M. (2013). Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proc Natl Acad Sci U S A*, *110*(9), 3435-3440. https://doi.org/10.1073/pnas.1222743110

Ziha-Zarifi, I., LLANES, C., KÖHLER, T., PECHERE, J.-C., & PLESIAT, P. (1999). In Vivo Emergence of Multidrug-Resistant Mutants of *Pseudomonas aeruginosa* Overexpressing the Active Efflux System MexA-MexB-OprM. *Antimicrob Agents Chemother*, *43*(2). https://doi.org/10.1128/aac.43.2.287

Zimmermann, P., & Curtis, N. (2019). The effect of antibiotics on the composition of the intestinal microbiota - a systematic review. *J Infect*, *79*(6), 471-489. https://doi.org/10.1016/j.jinf.2019.10.008