Antibiotic resistance potential of the healthy preterm infant gut microbiome

Graham Rose [Corresp., 1], **Alexander G Shaw** [2], **Kathleen Sim** [2], **David J Wooldridge** [1], **Ming-Shi Li** [2], **Saheer Gharbia** [1], **Raju Misra** [1], **John Simon Kroll** [3]

[1] Genomics Research Unit, Public Health England, London, United Kingdom

[2] Department of Medicine, Imperial College London, London, United Kingdom

[3] Section of Paediatrics, Department of Medicine, Imperial College London, London, United Kingdom

Corresponding Author: Graham Rose
Email address: graham.rose@phe.gov.uk

Background: Few studies have investigated the gut microbiome of infants, fewer still preterm infants. In this study we sought to quantify and interrogate the resistome within a cohort of premature infants using shotgun metagenomic sequencing. We describe the gut microbiomes from preterm but healthy infants, characterising the taxonomic diversity identified and frequency of antibiotic resistance genes detected.

Results: Dominant clinically important species identified within the microbiomes included *C. perfringens*, *K. pneumoniae* and members of the *Staphylococci* and *Enterobacter* genera. Screening at the gene level we identified an average 13 genes per preterm infant, ranging across 8 different antibiotic classes, including aminoglycosides and fluoroquinolones. Some antibiotic resistance genes were associated with clinically relevant bacteria, including the identification of *mecA* and high levels of *Staphylococci* within some infants. We were able to demonstrate that in a third of the infants the *S. aureus* identified was unrelated using MLST or metagenome assembly, but low abundance prevented such analysis within the remaining samples.

Conclusions: We found that the healthy preterm infant gut microbiomes in this study harboured a significant diversity of antibiotic resistance genes. This broad picture of resistances and the wider taxonomic diversity identified raises further caution to the use of antibiotics without consideration of the resident microbial communities.

Antibiotic resistance potential of the healthy preterm infant gut microbiome

Graham Rose[1#], Alexander G. Shaw[2], Kathleen Sim[2], David J. Wooldridge[1], Ming-Shi Li[2], Saheer Gharbia[1], Raju V. Misra[1], J. Simon Kroll[2]

[1]Genomic Research Unit, Public Health England, Microbiology Services, 61 Colindale Avenue, London, UK

[2]Department of Medicine, Section of Paediatrics, Imperial College London, London, UK

[#] Corresponding author. Mailing address:

Genomic Research Unit, Public Health England, Microbiology Services, 61 Colindale Avenue, London NW9 5HT, UK.

E-mail: graham.rose@phe.gov.uk

Abstract

Background: Few studies have investigated the gut microbiome of infants, fewer still preterm infants. In this study we sought to quantify and interrogate the resistome within a cohort of premature infants using shotgun metagenomic sequencing. We describe the gut microbiomes from preterm but healthy infants, characterising the taxonomic diversity identified and frequency of antibiotic resistance genes detected.

Results: Dominant clinically important species identified within the microbiomes included *C. perfringens*, *K. pneumoniae* and members of the *Staphylococci* and *Enterobacter* genera. Screening at the gene level we identified an average 13 genes per preterm infant, ranging across 8 different antibiotic classes, including aminoglycosides and fluoroquinolones. Some antibiotic resistance genes were associated with clinically relevant bacteria, including the identification of *mecA* and high levels of *Staphylococci* within some infants. We were able to demonstrate that in a third of the infants the *S. aureus* identified was unrelated using MLST or metagenome assembly, but low abundance prevented such analysis within the remaining samples.

Conclusions: We found that the healthy preterm infant gut microbiomes in this study harboured a significant diversity of antibiotic resistance genes. This broad picture of resistances and the wider taxonomic diversity identified raises further caution to the use of antibiotics without consideration of the resident microbial communities.

Introduction

Over recent years the composition of the gastrointestinal (GI) microbiota has been increasingly implicated in health and disease, with bacterial populations harbouring both

beneficial commensals and pathogens. A diverse bacterial population results in greater genetic content, but some of this additional genetic material is less welcome. Previous studies have implicated the GI microbiota as a reservoir of antimicrobial resistance (AMR) genes (Penders et al., 2013), held by, or capable of being transferred to, potential pathogens. Whilst often benign, during bacterial infection transfer of AMR genes can occur, which - coupled with selection pressures arising through antimicrobial therapy - can make treatment difficult, increasing the time taken to cure the infection. Furthermore, antimicrobial therapies are generally (ideally) tailored towards acute infections targeting a single pathogen, with little consideration of the wider microbial communities which reside in the microbiome, leading to a situation in which the use of antibiotics may cause unintentional harm to the host.

As our understanding of the microbiome has developed, the collection of AMR genes within a bacterial population has recently been defined as the resistome (Penders et al., 2013). Antibiotics have a role in shifting the profile of the resistome within the population (Jernberg et al., 2007), with low antibiotic-use communities harbouring lower AMR gene frequencies (Walson et al., 2001; Bartoloni et al., 2009). Heavy treatment of bacterial populations with antibiotics can lead to the long term overrepresentation of AMR genes. Such dynamics are evident in the microbiome of preterm neonates, who receive multiple antibiotic courses, and are cared for in an Intensive Care Unit environment potentially contaminated with multi-resistant bacteria. Antibiotic treatments for both term and preterm neonates have demonstrated lasting effects on the microbiota (Tanaka et al., 2009; Arboleya et al., 2015), with the trajectory of population development diverging from untreated controls, leading to a potential scenario of prolonged – even life-long - high frequency AMR reservoirs through the selection of bacteria within the population that are most resistant. A wide range of AMR genes have been found in neonatal populations (de Vries et al., 2011; Zhang et al., 2011), some shown to be present from birth

(Alicea-Serrano et al., 2013; Gosalbes et al., 2015), whilst twin pairs have been shown to have GI communities with similar distributions of both organisms and resistance genes (Moore et al., 2015). These observations suggest vertical transmission as a source, with discrepancies between mothers and babies being due to the substantial shifts in the microbiota adapting to the very different environment of a newly born infant's GI tract (Gosalbes et al., 2015).

The GI tract of a premature neonate is a particularly unusual scenario for observation of AMR genes, due to greatly reduced bacterial immigration as a result of the isolated, sterile environment of incubators and very controlled enteral feeds; donor breast milk may be pasteurised and, whilst unpasteurised maternal milk (which harbours specific bacteria (Beasley & Saris, 2004; Jimenez et al., 2008; Martin et al., 2009)) is given where possible, there is a likelihood of little or no breastfeeding due to extreme prematurity.

In these circumstances, the GI community and the resistance genes present are likely in the main to be derived from the mother, and acquired during birth. Whilst limited bacterial numbers and diversity will initially be transferred, mechanisms are available for the dissemination of AMR through the expanding bacterial population (as reviewed by van Hoek *et al* (2011)) with transfer having been documented within the gut environment (Shoemaker et al., 2001; Karami et al., 2007; Trobos et al., 2009). Heavy use of antibiotics in the course of care of premature infants would not only then skew the bacterial population and drive resistance selection, but has been shown to increase the activity of some transposable elements due to stressing of bacterial populations (Beaber, Hochhut & Waldor, 2004).

In this study, we present a detailed investigation of the resistome from the GI microbiota of eleven premature infants, with detailed information on antibiotic receipt and maternal antibiotic use. The microbiota of premature infants has been subjected to such investigations before, but through targeted techniques such as PCR or qPCR (Gueimonde, Salminen & Isolauri,

2006; Alicea-Serrano et al., 2013; von Wintersdorff et al., 2016) or through functional metagenomics (de Vries et al., 2011; Moore et al., 2015), which has the disadvantage of not being able to quantify the antibiotic resistance potential of a community (Forslund et al., 2014). We have used shotgun metagenomic sequencing to describe the resistome in its entirety, moving from species level taxonomic profiling, to characterisation of the resistance landscape, including typing of metagenomes identified as potentially harbouring *mecA*, conferring resistance to methicillin and other β-lactam antibiotics.

Materials & Methods

Study population

The study was approved by West London Research Ethics Committee (REC) Two, United Kingdom, under the REC approval reference number 10/H0711/39. Parents gave written informed consent for their infant to participate in the study.

Faecal samples analysed were collected from premature infants, defined as less than 32 completed weeks of gestation. Premature infants were recruited to the study at the Imperial College Healthcare National Health Service Trust neonatal intensive care unit (NICU), at Queen Charlotte's and Chelsea Hospital, between January 2010 and December 2011.

Sample collection

Almost every faecal sample produced by each participant between recruitment and discharge was collected by nursing staff from diapers using a sterile spatula. Samples were placed in a sterile DNase-, RNase-free Eppendorf tube, stored at -20 °C within two hours of collection

and stored at -80 °C within five days. A single faecal sample from each of twelve infants who had no diagnosis of necrotising enterocolitis or blood-stream infection during their admission was selected for metagenomic sequencing. DNA from one faecal sample did not complete library preparation (see below); clinical characteristics of the remaining eleven infants and faecal sample metadata are presented in Table S1.

DNA extraction and shotgun library preparation

DNA extractions were performed as described previously (Rose et al., 2015), but with the following modifications: DNA extracts were prepared from approximately 200 mg of faeces, which were re-suspended in 10x volume:weight filtered 1x phosphate-buffered saline (PBS), with addition of 1:1 (volume:volume) 2% 2-mercaptoethanol diluted in 1x filtered PBS. The MolYsis selective lysis kit (Molzym) was used for the selective lysis of eukaryotic cells, incorporating the modifications previously described (Rose et al., 2015). Bacterial lysis was performed by addition of 50 µl lysozyme (Sigma), 6 µl mutanolysin (Sigma) and 3 µl lysostaphin (Sigma) to 100 µl of re-suspended bacterial pellet, and incubated at 37 ºC for 1 h. This was followed by addition of 2 µl proteinase K and 150 µl 2x Tissue and Cell lysis buffer (Epicentre) and incubated at 65 ºC for 30 min. Lysates were added to 2 ml tubes containing 0.25 ml of 0.5 mm beads and beaten on a Fast Prep 24 system at 6 m/s for 20 s and repeated once after 5 min. Finally, DNA was purified using the MasterPure complete kit (Epicentre) according to the manufacturer's instructions, eluted in 50 ul 0.1 x TE buffer (Sigma) and stored at -80 ºC.

Extracted DNA was fragmented using the NEBNext dsDNA fragmentase kit (NEB) according to the manufacturer's instructions. Shotgun DNA libraries were subsequently prepared using the KAPA HyperPrep kit (KAPA Biosystems) according to the manufacturer's instructions.

Ligated libraries were amplified by PCR with the number of cycles being dependant on starting material biomass, varying between 2 and 8 (mean 3 cycles). A negative extraction control was included consisting of 1 ml filtered 1x PBS and processed alongside the samples. After library amplification, the negative extraction control and one preterm infant faecal sample required >8 PCR cycles owing to very low starting pre-PCR biomass (DNA concentration <0.1 ng/ul), therefore these samples were excluded from downstream analysis, leaving faecal samples from eleven premature infants.

Shotgun metagenomic sequencing

Library insert size and quantity was assessed for each sample by Bioanalyser and qPCR as described previously (Rose et al., 2015). Library insert size ranged from 244 bp to 288 bp with a mean of 261 bp. Libraries were sequenced on either an Illumina NextSeq 500 system or part of replicate runs on an Illumina MiSeq system. Prior to loading, libraries were normalised, pooled and diluted to either 1.6 pM or 18 pM for sequencing on the NextSeq or on the MiSeq, respectively. Paired end (PE) sequencing was performed on the NextSeq using a v2 300 cycle high output reagent kit (Illumina) and on the MiSeq using the v3 600 cycle reagent kit (Illumina).

Prior to sequencing the complete dataset, three libraries from three infant faecal samples were sequenced as part of separate multiplexed MiSeq runs, generating a mean 8.3 million PE reads and 5.0 Gbp sequence yield per sample. Later sequencing on a NextSeq instrument included the complete 11 sample set, and inclusion of a technical replicate (sample Q89). A mean 10.0 million PE reads were generated per sample, yielding 3.0 Gbp.

142 **Sequencing data availability**

143 All sequencing data generated for this study is available from the EBI European
144 Nucleotide Archive, under study accession PRJEB15257 (see
145 http://www.ebi.ac.uk/ena/data/view/PRJEB15257).

146 **Processing of metagenomic sequences**

147 Sequence quality was calculated using FastQC (v0.11.3) (Andrews, 2010). Read filtering
148 was performed using Trimmomatic (v0.32) (Bolger, Lohse & Usadel, 2014). This consisted of
149 adapter sequence removal based on an in house database of Illumina adapters, primers and index
150 sequences using non-default parameters (*ILLUMINACLIP 2:30:10*), and subsequent read
151 trimming consisting of an initial head crop of the first 15 bp, then iterative removal of leading
152 and trailing bp with phred qualities < 20, and internal bases where mean base phred qualities < 20
153 in 4 bp sliding windows (parameters were: *LEADING:20 TRAILING:20*
154 *SLIDINGWINDOW:4:20*). Finally sequences with less than 40 bp remaining were discarded
155 (*MINLEN:40*).

156 A mapping based approach was used to remove expected human host sequences from the
157 faecal samples, as well as any remaining vector contamination. Using FastQ Screen (v0.4.4)
158 (Babraham Institute) and the short read aligner Bowtie2 (v2.2.6) (Langmead & Salzberg, 2012),
159 reads were mapped against the human genome (GRCh38) and the UniVec (version 8) vector
160 database (non-default parameters were: *--aligner bowtie2*, *--nohits*). All unmapped PE reads were
161 output as new fastq files and continued within downstream analysis.

162 The MiSeq replicate datasets entered an identical workflow, except for an additional step
163 designed to utilise the longer read lengths by joining the read pairs using FLASH (v1.2.11)

(Magoč & Salzberg, 2011) *(--max-overlap 200*), thus generating longer single sequences (mean length 207 bp).

Species identification and relative abundances

Primary metagenomic profiling was performed using DIAMOND (v.0.7.9.8) (Buchfink, Xie & Huson, 2014) and MEGAN (v5.10.6) (Huson et al., 2007). All forward reads, or joined reads in the case of the MiSeq dataset, were aligned against a protein reference database under default parameters. The protein database was built using all 73,055,898 sequences from the NCBI non-redundant (nr) database (downloaded 21-10-15).

Processing and taxonomic analysis of the sequence reads with matches to the nr database was performed within MEGAN and under non-default parameters. Sequences were assigned to the NCBI Taxonomy (1,266,115 individual taxa) using the Lowest Common Ancestor (LCA) and the following thresholds: minimum bit-score: 80; max expectation value: 1.0×10^{-6}; top percentage of hits considered: 10%; minimum taxon support based on all assigned reads: 0.01%. Relative abundances and extraction of species specific binned reads were calculated within MEGAN.

A secondary taxonomic profiling method, MetaPhlAn (v.2.2.0) (Truong et al., 2015) was used on all samples with the following parameters: *--mpa_pkl metaphlan2/db_v20/mpa_v20_m200.pkl --bowtie2db metaphlan2/db_v20/mpa_v20_m200 --input_type fastq*. Relative abundance tables were combined using the packaged MetaPhlAn script - merge_metaphlan_tables.py.

Identifying antimicrobial resistance genes

Presence/absence testing of AMR genes within the samples was performed on unassembled reads. Reads were mapped using bowtie2 (v.2.2.6) (Langmead & Salzberg, 2012) as part of SRST2 (v0.1.7) (Inouye et al., 2014) under default parameters to a clustered ARG-Annot database of acquired resistance genes (Gupta et al., 2014) and all hits recorded (*--gene_db srst2/data/ARGannot.r1.fasta*). Default parameters set AMR gene reporting at 90% minimum coverage cutoff. Computational *S. aureus* sequence typing (ST) was also performed using SRST2 (Inouye et al., 2014) under default scoring parameters, and using the *S. aureus* MLST schema downloaded on 18-04-16 from pubmlst.org. Alleles were flagged uncertain when below threshold depths (--min_edge_depth 2, --min_depth 5).

Metagenome assembly and *S. aureus* phylogeny

Assemblies were performed using spades (v3.7.1) under default parameters except identification of the data as metagenomic (--meta). Assembled contigs were used as blastn queries against the NBCI nt database, and taxonomic labels attached using MEGAN, with all contigs identified as *S. aureus* (NCBI taxon id: 1280) including summarised contigs extracted per sample. As a reference, all *S. aureus* complete genomes were downloaded from PATRIC (Release May 2016), totalling 118 genomes.

An anchor based phylogenetic method, andi (v.0.10) (Haubold, Klötzl & Pfaffelhuber, 2015), was used to estimate the evolutionary distances between the study and global *S. aureus* genomes set, using Phylip (v.3.696) (Felsenstein, 2009) to infer the neighbour-joining phylogeny. Following assembly, very short binned *S. aureus* contigs (<1 kb) and partial assemblies, in this case those with less than half of the median *S. aureus* genome size (<1.5 Mb), were excluded

from andi and phylogenetic tree construction as based on recommended guidelines (Haubold, Klötzl & Pfaffelhuber, 2015).

S. aureus typing

Experimental confirmation of *mecA* was attempted for all eleven samples. Faecal samples were cultivated on the *Staphylococcal* selective growth media manitol salt agar. Sweeps of the presumptive *Staphylococcus* colonies from each sample were propagated and extracted by the following protocol: half a 10 µl loop of overnight growth at 35 ºC was inoculated into 2 ml tubes containing 0.5 mm silica/ zirconia beads filled to the 0.25 ml mark and 350 µl of Master Pure Tissue Cell Lysis Buffer (EpiCentre). Bead beating was performed using a Fast Prep (MPBio) at 6 m/s for 20 seconds. This was repeated three times with a 5 minute pause between each pulse. Lysates were centrifuged at 8,000 x*g* for 10 minutes and 300 µl of supernatant transferred to a new tube. A known *mecA* positive strain (NCTC strain 12232) and a *mecA* negative clinical isolate were used as control strains.

A multiplexed PCR method was used to type the SCCmec element within the samples according to the protocol described previously (Milheiriço, Oliveira & De Lencastre, 2007), but with the following exceptions. Each 50 µl PCR reaction consisted of 1x HotStart Ready Mix (KAPA Biosystems), 25 ng genomic DNA, and primers at the described concentration. The cycling conditions were as follows: 95 ºC for 3 minutes followed by 30 cycles of 98 ºC for 20 seconds, 53 ºC for 30 seconds and 72 ºC for 30 seconds followed by a final extension of 72 ºC for 4 minutes. Amplicons were purified using the AgenCourt AMpure XP PCR purification kit (Beckman Coulter) following manufacturer's instructions. Amplicon sizes were determined measured by BioAnalyser (Agilent) on a high sensitivity DNA chip, with classification of a

positive result based on on fragment sizes ± 5 bp of those expected, and peak concentration >= 500 pg/ul. In addition to the above controls, extraction and PCR negative controls were included, which substituted input genomic DNA for purified water.

Statistics

Species richness and the evenness of their abundance were quantified using the Shannon–Weaver index ecological measure, calculated within MEGAN. Visualisation of samples was performed by hierarchical clustering using the UPGMA method and principal coordinates analysis (PCoA), all based on a matrix of Bray-Curtis distances calculated within MEGAN. Correlations and t-tests were performed within R (v 3.2.5) (R Developement Core Team, 2015).

Results

The healthy preterm metagenome

Using shotgun metagenomic sequencing we have captured an early snapshot of the antimicrobial resistance landscape within the gut microbiota of eleven premature infants who did not have proven sepsis or necrotizing enterocolitis. Infants were born either vaginally (N=6) or by caesarean section (N=5), with gestational ages ranging 24-31 weeks (mean 26.9 weeks). Ages of the infants at which the samples were taken ranged from 5 - 43 days (mean 25.7 days) (Table S1). A mixture of benchtop to medium throughput Illumina platforms were used to generate a dataset of 145.6 million paired end (PE) reads (51.4 Gbp sequence data) (Table 1), enabling us to characterise taxonomic and antimicrobial resistance profiles.

The eleven sequenced samples and four replicates were analysed using a blastx type analysis with filtering by the Lowest Common Ancestor (Huson et al., 2007; Buchfink, Xie & Huson, 2014), which enabled assignment of taxonomic labels for 71.5% of the reads within the complete dataset to at least the level of Kingdom (Table S2). As an alternative method, we also profiled the dataset using a marker based approach (Truong et al., 2015), which was highly congruent to species level relative abundances, as well as higher taxonomic ranks, to the blast based method used (Pearson R = 0.9 - 1.0) (Table S3). Replicate sequencing of samples also demonstrated reproducibility of the method, either by cluster analysis (Figure 1) or pairwise correlations (Figure S1).

Moving to taxonomic composition, each sample was marked by a few highly abundant species, such as sample Q216 with 85.1% *Clostridium perfringens*, Q189 with 73.1% *Klebsiella pneumoniae*, and Q83 with 85.9% *Enterococcus faecalis* (Figure 1A). In terms of prevalence, the previous three species, as well as *Enterobacter cloacae* and *Staphylococcus epidermidis,* were found at over 50% relative abundance in one or more samples. Furthermore, S. *epidermidis* and *S. aureus* were ubiquitous, ranging from 0.06% to 57.1% abundance in all samples (Figure 1B). Principal coordinate analysis (PCoA) demonstrated three loose sample groups based on a high abundance of *S. epidermidis*, *K. pneumoniae*, and either *B. breve*, *S. aureus* or *C. perfringens* (Figure 1C). In total we identified a non-redundant set of 172 species across all samples (see Table S4 for complete dataset).

Prevalence of antimicrobial resistance

Before focusing on individual AMR genes, we measured the α-diversity (Shannon–Weaver index) within each sample and, although a small sample set, compared this to the known

antibiotic exposure of the preterm infants (Figure S2). Eight of the eleven infants had received a course of prophylactic antibiotic treatment consisting of co-amoxiclav (Table S1), whilst a second course was administered to four infants, consisting of combinations of co-amoxiclav, tazocin or vancomycin. In total, exposure ranged from 2 - 8 days of antibiotics before samples were taken, excluding infants Q87 and Q89 which received no antibiotics. Antibiotics were also administered maternally to three infants (Q26, Q117 & Q189), but this did not include the two above infants with no antibiotic treatment. Sample diversity ranged from 0.9 - 2.9 (SD ± 0.5), but when compared to cumulative antibiotic exposure expressed in days, no significant difference was found between the taxonomic diversity and amount of antibiotic exposure for untreated and treated infants (unpaired *t* test, $P = 0.17$) (Figure S2). It is important to stress however that the small and heterogeneous nature of the sample set will have reduced the power to detect differences between antibiotic exposure in this study, and so prevented any meaningful stratification by other clinical variables such as mode of delivery or day of life.

A mapping based approach against a comprehensive collection of acquired antibiotic resistance genes was next used to quantify AMR within the eleven metagenomes (Inouye et al., 2014). In total 143 AMR genes were identified, consisting of a non-redundant set of 39 different AMR genes (Figure 2 and Table S5). Per infant, the average number of genes identified was 13 (ranging 5 – 22 genes), and AMR genes were found across eight different antibiotic classes, including aminoglycosides and fluoroquinolones (Table S6). Mean sequence coverage across the sequence database was 99.0%, and sequence divergence ranged from no difference to 12.3% (Table S6). In total over 1,600 alleles were searched for, and notable AMR genes not detected included those involved in carbapenem and vancomycin resistance, the latter of which was administered to three preterm infants prior to sample collection (Table S1). The class most frequently detected were β-lactamases, comprising ten different genes (Table 2), of which the

293 *blaZ* gene was present in every infant. Interestingly, within this set of β-lactamasae genes, *mecA*
294 was found in four infants (Q87, Q117, Q175, and Q189), and at a mean depth of coverage
295 ranging from 3.9 to 52.2 bp (Figure 2). *mecA* confers resistance to methicillin as well as other β-
296 lactam antibiotics, and is carried on the SCCmec mobile element found across several
297 *Staphylococci* species. Identification of four infants with potential methicillin resistant *S. aureus*
298 (MRSA) or *S. epidermidis* (MRSE) carriage, along with high abundances and prevalence of both
299 *S. aureus* and *S. epidermidis* species across the dataset, could indicate a significant reservoir for
300 AMR transfer between the species, as well as highlight the seeding of the infant gut microbiome
301 from an early stage.

302 **Focus on *S. aureus* species detected**

303 We next wanted to understand the relationship of the *S. aureus* species within the *mecA*
304 positive as well as negative samples, as the premature infants overlapped in time and so could
305 harbour closely related strains. This was undertaken to firstly confirm *in silico* prediction of
306 *mecA* using an established molecular based typing method, but also to push the metagenomic
307 analysis further on what was known to be a challenging dataset owing to the range of identified
308 *S. aureus* as described above, with relative abundance ranging from 0.06% - 39.8% (Table S4).
309 We first tested the computational prediction of *mecA* experimentally using a multiplexed PCR
310 typing method (Milheiriço, Oliveira & De Lencastre, 2007), which provides detection of the
311 *mecA* gene, in addition to typing of the mobile element carrying the gene (SCCmec), although
312 this component of the assay was beyond the scope of this study. Using this method we detected
313 *mecA* presence correctly within the control strains, a methicillin resistant (MRSA) and
314 susceptible (MSSA) strain (see Methods), and three out of the four predicted *mecA* positive
315 samples generated a positive *mecA* result (Table S7). The exception, sample Q87, generated the

316 expected *mecA* amplicon size but the concentration of this fell below the threshold for detection
317 (<500pg/ul) and so was excluded.

318 In an attempt to understand strain relatedness directly from the metagenomic data we
319 undertook *in silico* MLST analysis using an *S.aureus* schema as well as metagenome assembly.
320 The MLST was able to classify four of the eleven samples, all with different ST types – ST8,
321 ST1027, ST22, ST25, although the last two had some degree of uncertainty in their assignment
322 (Table S8). This suggests that for at least these four samples, the *S. aureus* strains are unrelated
323 and unlikely a result of transmission. We were interested to know if *de novo* assembly of the
324 metagenome could be utilised to resolve these and any of the remaining unclassifiable samples
325 further. Following assembly and identification of *S. aureus* contigs (see Methods), we found that
326 it was not possible to capture more than a fifth of the expected genome size for the above
327 unclassified samples, with an abundance of > 3% necessary to achieve over 90% estimated
328 capture, which was achieved in four cases (Table S9). Phylogenetic reconstruction of these four
329 genomes alongside a collection of published *S. aureus* genomes (Table S10), provided
330 confirmation of the diversity of *S. aureus* identified (Figure S3), enabling placement across a
331 global collection of strains.

332 **Discussion**

333 It is recognised that one of the most important public health threats worldwide is
334 antimicrobial resistance. Here we report on the gut composition and AMR diversity for eleven
335 healthy but premature infants. Recent studies have shown that the initial seeding of the infant gut
336 microbiome is influenced by the microorganisms in the immediate environment, and whilst
337 colonisation by bacteria with AMR genes has been demonstrated (Brooks et al., 2014),

338 comparatively far fewer studies have investigated the gut microbiome of infants, fewer still
339 preterm healthy infants. Interest has also increased on how the trajectory of the early gut
340 microbiome is influenced to form the 'stable' adult microbiome. The preterm infant gut
341 microbiome is very different compared to full-term infants (Groer et al., 2014), displaying a
342 much lower diversity, particularly in anaerobes, with an increase in coagulase-negative
343 *Staphylococci* and *Enterobacteriaceae* (Adlerberth & Wold, 2009); adult microbiomes are
344 characterised by several hundred, mostly anaerobic bacterial species (Adlerberth & Wold, 2009).
345 We found a similarly low level of species diversity across all metagenomes, with each sample
346 dominated by a few highly abundant species, including *C. perfringens*, *K. pneumoniae* and
347 members of the *Staphylococci* and *Enterobacter* genera. Presence of such species are in common
348 with previous studies on the premature gut microbiome (Groer et al., 2014; Gibson et al., 2016).

349 Interestingly, each metagenome profile displayed a different dominant few species,
350 clustering into three loose groupings. This could reflect the dynamic nature of the early
351 establishing gut, with the preterm infant microbiome acquiring an increased diversity of bacteria
352 and subjected to a great amount of change until it matures into what is recognised as a more
353 'stable' microbiome. Although we found no correlation between diversity and antibiotic
354 exposure, with infants treated with either no antibiotics (including during pregnancy), to up to 8
355 days of antibiotic administration, effects such as relatively small sample size, as well as day of
356 life of sample and normal gut development are biases to this finding, which is contrary to other
357 studies within infants (Greenwood et al., 2014; Merker et al., 2015). It could be that at this very
358 early stage, the microbiota is influenced to a greater extent by seeding during birth from the
359 mother and environment than antibiotic treatment, or that not enough time has passed to detect
360 differences from the antibiotics administered; larger sample numbers would be required,

alongside longitudinal studies and parallel maternal sampling to better understand the development of diversity.

A threat to this development is the acquisition of antibiotic resistant bacteria, which can potentially seed the infant microbiome. Coupled with the high rate of horizontal gene transfer within the commensal community (Stecher et al., 2012), the preterm infant gut microbiome has the potential to be a reservoir for AMR. With dominance of the preterm gut by species known to carry clinically relevant antibiotic resistance, we next quantified the burden of antibiotic resistance genes within the infant's faecal flora, which identified an average 13 genes per infant. Previous targeted or functional studies based on infants have found some of the AMR genes also identified here, including those for Tetracycline (*tet*) (Gueimonde, Salminen & Isolauri, 2006; Alicea-Serrano et al., 2013) and β-lactam (bla) (Fouhy et al., 2014). In a wider context, it is known that AMR genes are a common feature of bacterial populations, found in communities inhabiting the soil, rivers and even deep-sea sediment (Knapp et al., 2010; Qin et al., 2011; Kittinger et al., 2016). Therefore, whilst their presence in the human gut microbiome should be of little surprise (Bailey et al., 2010), identification of genes such as *mecA* demonstrates the prevalence of some clinically significant resistant bacteria from birth.

One of the advantages of the method used in this study is the utility of the results generated, enabling multiple avenues of questions to be addressed. However, short read sequencing remains a challenge when applied to the linkage of resistance elements, such *mecA*, to specific genome sequences (strains), which is made difficult by the nature of metagenomic samples containing multiple alleles from different closely related species, as well as potentially multiple strains of the same species. Secondly, the methods used here were inherently restricted to identification of known AMR genes found within the ARGannot database used in this study, which contains those genes involved in acquired resistance only, therefore chromosomal

mutations, such as those conferring resistance to rifampicin as well as novel resistance genes would have been missed, leading to potential underrepresentation of resistance in this study.

Conclusions

The healthy preterm infants sampled within this study harboured multiple AMR genes, representing a potential reservoir for later disease onset. In particular, detection of clinically important AMR genes, such as *mecA*, highlights the need to further understand the impact that this reservoir could have on later treatment regimes.. From a methodology point, this approach was able to provide a comprehensive snapshot of the complete taxonomic diversity and resistome in one assay. Although tracking of the movement of such AMR genetic elements would be enhanced by improved handling of the dynamic ranges of abundances; different methods at the level of sample preparation, such as sample normalisation, may offer potential answers to such hurdles. Overall this study leads to questions such as how this resistance potential contributes to later clinical intervention or disease onset, and if antibiotic treatment without knowledge of prior AMR burden could lead to unintentional harm. More broadly, this and other studies show the great promise that shotgun metagenomics holds for clinical microbiology.

Additional Information

Abbreviations

AMR, antimicrobial resistance; GI, gastrointestinal; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; *BLAST,* basic local alignment search tool; SCCmec, staphylococcal chromosome cassette *mec*; LCA, lowest common ancestor; MLST, multilocus

405 sequence typing; MALDI-TOF, Matrix-assisted laser desorption/ionization – time of flight;
406 PCoA, principal coordinates analysis.

407 **References**

408 Adlerberth I., Wold AE. 2009. Establishment of the gut microbiota in Western infants. *Acta*
409 *Paediatrica, International Journal of Paediatrics* 98:229–238. DOI: 10.1111/j.1651-
410 2227.2008.01060.x.

411 Alicea-Serrano AM., Contreras M., Magris M., Hidalgo G., Dominguez-Bello MG. 2013.
412 Tetracycline resistance genes acquired at birth. *Archives of Microbiology* 195:447–451. DOI:
413 10.1007/s00203-012-0864-4.

414 Andrews S. 2010.FastQC: A quality control tool for high throughput sequence data.

415 Arboleya S., Sanchez B., Milani C., Duranti S., Solis G., Fernandez N., De Los Reyes-Gavilan
416 CG., Ventura M., Margolles A., Gueimonde M. 2015. Intestinal microbiota development in
417 preterm neonates and effect of perinatal antibiotics. *Journal of Pediatrics* 166:538–544.
418 DOI: 10.1016/j.jpeds.2014.09.041.

419 Bailey JK., Pinyon JL., Anantham S., Hall RM. 2010. Commensal Escherichia coli of healthy
420 humans: A reservoir for antibiotic-resistance determinants. *Journal of Medical Microbiology*
421 59:1331–1339. DOI: 10.1099/jmm.0.022475-0.

422 Bartoloni A., Pallecchi L., Rodríguez H., Fernandez C., Mantella A., Bartalesi F., Strohmeyer M.,
423 Kristiansson C., Gotuzzo E., Paradisi F., Rossolini GM. 2009. Antibiotic resistance in a very
424 remote Amazonas community. *International Journal of Antimicrobial Agents* 33:125–129.
425 DOI: 10.1016/j.ijantimicag.2008.07.029.

426 Beaber JW., Hochhut B., Waldor MK. 2004. SOS response promotes horizontal dissemination of
427 antibiotic resistance genes. *Nature* 427:72–74. DOI: 10.1038/nature02241.

428 Beasley SS., Saris PEJ. 2004. Nisin-Producing Lactococcus lactis Strains Isolated from Human
429 Milk. *Applied and environmental microbiology* 70:5051–5053. DOI:
430 10.1128/AEM.70.8.5051.

431 Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence
432 data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

433 Brooks B., Firek BA., Miller CS., Sharon I., Thomas BC., Baker R., Morowitz MJ., Banfield JF.
434 2014. Microbes in the neonatal intensive care unit resemble those found in the gut of
435 premature infants. *Microbiome* 2:1. DOI: 10.1186/2049-2618-2-1.

436 Buchfink B., Xie C., Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND.
437 *Nature methods* 12:59–60. DOI: 10.1038/nmeth.3176.

438 Felsenstein J. 2009. PHYLIP - Phylogeny Inference Package, Version 3.69. (Seattle, WA:
439 University of Washington). *The American Naturalist* 171:713–725. DOI: 10.1111/j.1096-
440 0031.1989.tb00562.x.

441 Forslund K., Sunagawa S., Coelho LP., Bork P. 2014. Metagenomic insights into the human gut
442 resistome and the forces that shape it. *BioEssays* 36:316–329. DOI:
443 10.1002/bies.201300143.

444 Fouhy F., Ogilvie LA., Jones B V., Ross RP., Ryan AC., Dempsey EM., Fitzgerald GF., Stanton
445 C., Cotter PD. 2014. Identification of aminoglycoside and β-lactam resistance genes from
446 within an infant gut functional metagenomic library. *PLoS ONE* 9. DOI:
447 10.1371/journal.pone.0108016.

448 Gibson MK., Wang B., Ahmadi S., Burnham C-AD., Tarr PI., Warner BB., Dantas G. 2016.
449 Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome.
450 *Nature Microbiology*:16024. DOI: 10.1038/nmicrobiol.2016.24.

451 Gosalbes MJ., Vallès Y., Jiménez-Hernández N., Balle C., Riva P., Miravet-Verde S., de Vries
452 LE., Llop S., Agersø Y., Sørensen SJ., Ballester F., Francino MP. 2015. High frequencies of
453 antibiotic resistance genes in infants' meconium and early fecal samples. *Journal of*
454 *Developmental Origins of Health and Disease*:1–10. DOI: 10.1017/S2040174415001506.

455 Greenwood C., Morrow AL., Lagomarcino AJ., Altaye M., Taft DH., Yu Z., Newburg DS., Ward
456 D V., Schibler KR. 2014. Early empiric antibiotic use in preterm infants is associated with
457 lower bacterial diversity and higher relative abundance of enterobacter. *Journal of*
458 *Pediatrics* 165:23–29. DOI: 10.1016/j.jpeds.2014.01.010.

459 Groer MW., Luciano AA., Dishaw LJ., Ashmeade TL., Miller E., Gilbert JA. 2014. Development
460 of the preterm infant gut microbiome: a research priority. *Microbiome* 2:38. DOI:
461 10.1186/2049-2618-2-38.

462 Gueimonde M., Salminen S., Isolauri E. 2006. Presence of specific antibiotic (tet) resistance
463 genes in infant faecal microbiota. *FEMS Immunology and Medical Microbiology* 48:21–25.
464 DOI: 10.1111/j.1574-695X.2006.00112.x.

465 Gupta SK., Padmanabhan BR., Diene SM., Lopez-Rojas R., Kempf M., Landraud L., Rolain J-M.
466 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in
467 bacterial genomes. *Antimicrobial agents and chemotherapy* 58:212–20. DOI:
468 10.1128/AAC.01310-13.

469 Haubold B., Klötzl F., Pfaffelhuber P. 2015. Andi: Fast and accurate estimation of evolutionary

distances between closely related genomes. *Bioinformatics* 31:1169–1175. DOI: 10.1093/bioinformatics/btu815.

van Hoek AHAM., Mevius D., Guerra B., Mullany P., Roberts AP., Aarts HJM. 2011. Acquired antibiotic resistance genes: An overview. *Frontiers in Microbiology* 2:203. DOI: 10.3389/fmicb.2011.00203.

Huson DH., Auch AF., Qi J., Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17:377–386. DOI: 10.1101/gr.5969107.

Inouye M., Dashnow H., Raven L-A., Schultz MB., Pope BJ., Tomita T., Zobel J., Holt KE. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome medicine* 6:90. DOI: 10.1186/s13073-014-0090-6.

Jernberg C., Lofmark S., Edlund C., Jansson JK. 2007. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *Isme.J.* 1:56–66. DOI: 10.1038/ismej.2007.3.

Jimenez E., Delgado S., Fernandez L., Garcia N., Albujar M., Gomez A., Rodriguez JM. 2008. Assessment of the bacterial diversity of human colostrum and screening of staphylococcal and enterococcal populations for potential virulence factors. *Research in Microbiology* 159:595–601. DOI: 10.1016/j.resmic.2008.09.001.

Karami N., Martner A., Enne VI., Swerkersson S., Adlerberth I., Wold AE. 2007. Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics. *Journal of Antimicrobial Chemotherapy* 60:1142–1145. DOI: 10.1093/jac/dkm327.

Kittinger C., Lipp M., Baumert R., Folli B., Koraimann G., Toplitsch D., Liebmann A., Grisold

492 AJ., Farnleitner AH., Kirschner A., Zarfel G. 2016. Antibiotic Resistance Patterns of
493 Pseudomonas spp. Isolated from the River Danube. *Frontiers in Microbiology* 7:1–8. DOI:
494 10.3389/fmicb.2016.00586.

495 Knapp CW., Dolfing J., Ehlert PAI., Graham DW. 2010. Evidence of increasing antibiotic
496 resistance gene abundances in archived soils since 1940. *Environmental Science and*
497 *Technology* 44:580–587. DOI: 10.1021/es901221x.

498 Langmead B., Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*
499 9:357–359. DOI: 10.1038/nmeth.1923.

500 Magoč T., Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome
501 assemblies. *Bioinformatics (Oxford, England)* 27:2957–63. DOI:
502 10.1093/bioinformatics/btr507.

503 Martin R., Jimenez E., Heilig H., Fernandez L., Marin ML., Zoetendal EG., Rodriguez JM. 2009.
504 Isolation of bifidobacteria from breast milk and assessment of the bifidobacterial population
505 by PCR-denaturing gradient gel electrophoresis and quantitative real-time PCR. *Applied and*
506 *Environmental Microbiology* 75:965–969. DOI: 10.1128/AEM.02063-08.

507 Merker M., Blin C., Mona S., Duforet-Frebourg N., Lecher S., Willery E., Blum MGB., Rüsch-
508 Gerdes S., Mokrousov I., Aleksic E., Allix-Béguec C., Antierens A., Augustynowicz-Kopeć
509 E., Ballif M., Barletta F., Beck HP., Barry CE., Bonnet M., Borroni E., Campos-Herrero I.,
510 Cirillo D., Cox H., Crowe S., Crudu V., Diel R., Drobniewski F., Fauville-Dufaux M.,
511 Gagneux S., Ghebremichael S., Hanekom M., Hoffner S., Jiao W., Kalon S., Kohl T a.,
512 Kontsevaya I., Lillebæk T., Maeda S., Nikolayevskyy V., Rasmussen M., Rastogi N.,
513 Samper S., Sanchez-Padilla E., Savic B., Shamputa IC., Shen A., Sng L-H., Stakenas P., Toit
514 K., Varaine F., Vukovic D., Wahl C., Warren R., Supply P., Niemann S., Wirth T. 2015.

Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nature Genetics* 47:242–249. DOI: 10.1038/ng.3195.

Milheiriço C., Oliveira DC., De Lencastre H. 2007. Update to the multiplex PCR strategy for assignment of mec element types in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* 51:3374–3377. DOI: 10.1128/AAC.00275-07.

Moore AM., Ahmadi S., Patel S., Gibson MK., Wang B., Ndao MI., Deych E., Shannon W., Tarr PI., Warner BB., Dantas G. 2015. Gut resistome development in healthy twin pairs in the first year of life. *Microbiome* 3:27. DOI: 10.1186/s40168-015-0090-9.

Penders J., Stobberingh EE., Savelkoul PHM., Wolffs PFG. 2013. The human microbiome as a reservoir of antimicrobial resistance. *Frontiers in Microbiology* 4. DOI: 10.3389/fmicb.2013.00087.

Qin Q-L., Li Y., Zhang Y-J., Zhou Z-M., Zhang W-X., Chen X-L., Zhang X-Y., Zhou B-C., Wang L., Zhang Y-Z. 2011. Comparative genomics reveals a deep-sea sediment-adapted life style of Pseudoalteromonas sp. SM9913. *The ISME journal* 5:274–284. DOI: 10.1038/ismej.2010.103.

R Developement Core Team. 2015. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* 1:409. DOI: 10.1007/978-3-540-74686-7.

Rose G., Wooldridge DJ., Anscombe C., Mee ET., Misra R V., Gharbia S. 2015. Challenges of the Unknown : Clinical Application of Microbial Metagenomics. *International Journal of Genomics* 2015:1–10. DOI: 10.1155/2015/292950.

Shoemaker NB., Vlamakis H., Hayes K., Salyers a a. 2001. Evidence for Extensive Resistance Gene Transfer among Bacteroides spp . and among Bacteroides and Other Genera in the

537 Human Colon Evidence for Extensive Resistance Gene Transfer among Bacteroides spp .
538 and among Bacteroides and Other Genera in the Human C. *Appl Environ Microbiol* 67:561–
539 8. DOI: 10.1128/AEM.67.2.561.

540 Stecher B., Denzler R., Maier L., Bernet F., Sanders MJ., Pickard DJ., Barthel M., Westendorf
541 AM., Krogfelt KA., Walker AW., Ackermann M., Dobrindt U., Thomson NR., Hardt W-D.
542 2012. Gut inflammation can boost horizontal gene transfer between pathogenic and
543 commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences of the*
544 *United States of America* 109:1269–74. DOI: 10.1073/pnas.1113246109.

545 Tanaka S., Kobayashi T., Songjinda P., Tateyama A., Tsubouchi M., Kiyohara C., Shirakawa T.,
546 Sonomoto K., Nakayama J. 2009. Influence of antibiotic exposure in the early postnatal
547 period on the development of intestinal microbiota. *FEMS Immunology and Medical*
548 *Microbiology* 56:80–87. DOI: 10.1111/j.1574-695X.2009.00553.x.

549 Trobos M., Lester CH., Olsen JE., Frimodt-M??ller N., Hammerum AM. 2009. Natural transfer
550 of sulphonamide and ampicillin resistance between Escherichia coli residing in the human
551 intestine. *Journal of Antimicrobial Chemotherapy* 63:80–86. DOI: 10.1093/jac/dkn437.

552 Truong DT., Franzosa EA., Tickle TL., Scholz M., Weingart G., Pasolli E., Tett A., Huttenhower
553 C., Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature*
554 *Methods* 12:902–903. DOI: 10.1038/nmeth.3589.

555 de Vries LE., Valles Y., Agerso Y., Vaishampayan PA., Garcia-Montaner A., Kuehl J V.,
556 Christensen H., Barlow M., Francino MP. 2011. The gut as reservoir of antibiotic resistance:
557 Microbial diversity of tetracycline resistance in mother and infant. *PLoS ONE* 6. DOI:
558 10.1371/journal.pone.0021644.

559 Walson JL., Marshall B., Pokhrel BM., Kafle KK., Levy SB. 2001. Carriage of antibiotic-
560 resistant fecal bacteria in Nepal reflects proximity to Kathmandu. *The Journal of infectious*
561 *diseases* 184:1163–1169. DOI: 10.1086/323647.

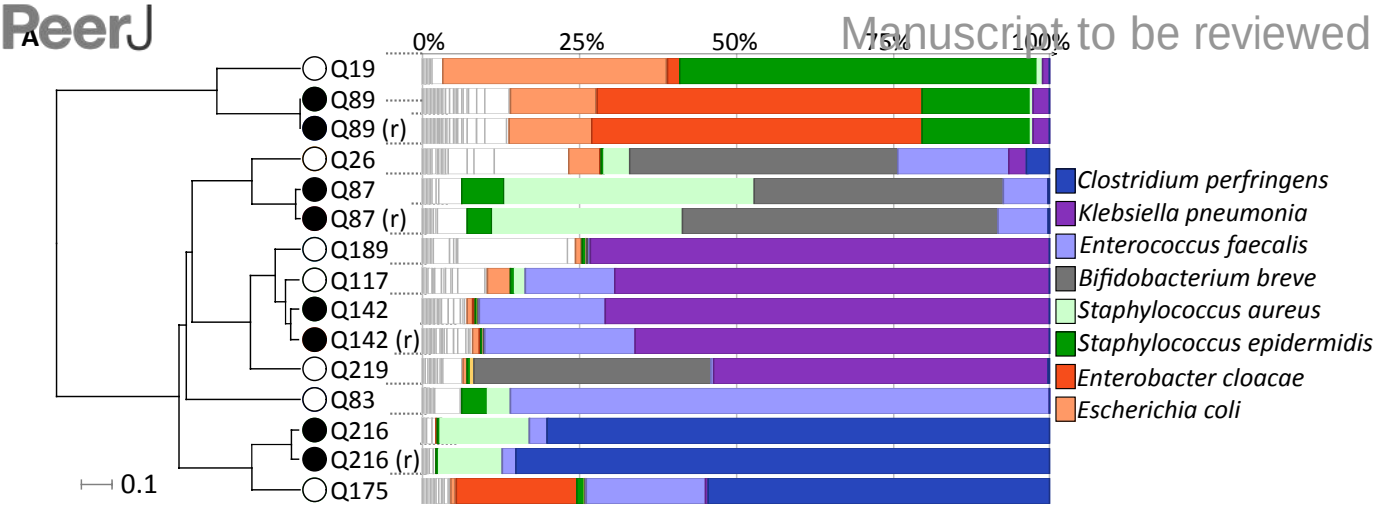
562 von Wintersdorff CJ., Wolffs PF., Savelkoul PH., Nijsen RR., Lau S., Gerhold K., Hamelmann E.,
563 Penders J. 2016. The gut resistome is highly dynamic during the first months of life. *Future*
564 *microbiology* 11:501–510. DOI: 10.2217/fmb.15.154.

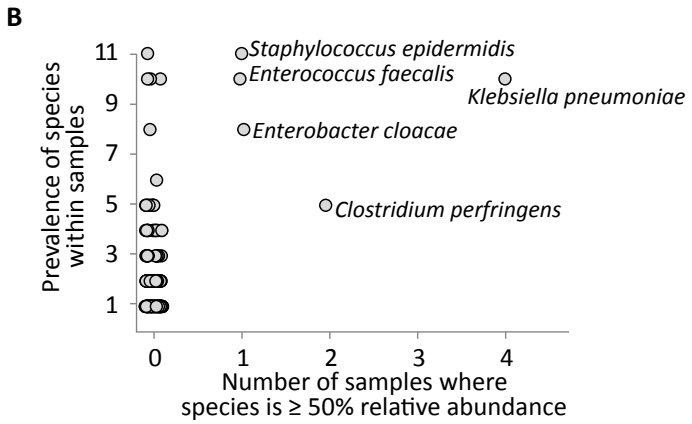
565 Zhang L., Kinkelaar D., Huang Y., Li Y., Li X., Wang HH. 2011. Acquired antibiotic resistance:
566 Are we born with it? *Applied and Environmental Microbiology* 77:7134–7141. DOI:
567 10.1128/AEM.05087-11.

Figure 1(on next page)

Healthy premature infant gut microbiome

A) Metagenomic profiles for the eleven preterm samples and four replicates at the species level. Samples clustered by UPGMA using Bray-Curtis distances shown on left, with replicates highlighted by filled nodes. Relative abundances by rank order shown on right, with the top 8 most abundant species coloured and labelled, leaving remaining species in white. **B**) Dominant species, based on >= 50% abundance, shown on x-axis, with overall prevalence of the species across samples shown on y-axis. Sample number reflects eleven neonates as replicates are averaged. The five labelled species are present in five or more samples with at least one in >50% abundance. **C**) Principle coordinates analysis (PCoA) using Bray-Curtis distances at the species level for all fifteen samples. Separation of the three broad sample groups shown by biplot of the top five species.


A
0%
25%
50%
75%
100%
Q19
Q89
Q89 (r)
Q26
Q87
Q87 (r)
Q189
Q117
Q142
Q142 (r)
Q219
Q83
Q216
Q216 (r)
Q175
0.1
Clostridium perfringens
Klebsiella pneumonia
Enterococcus faecalis
Bifidobacterium breve
Staphylococcus aureus
Staphylococcus epidermidis
Enterobacter cloacae
Escherichia coli


B
Prevalence of species within samples
Number of samples where species is ≥ 50% relative abundance
Staphylococcus epidermidis
Enterococcus faecalis
Klebsiella pneumoniae
Enterobacter cloacae
Clostridium perfringens

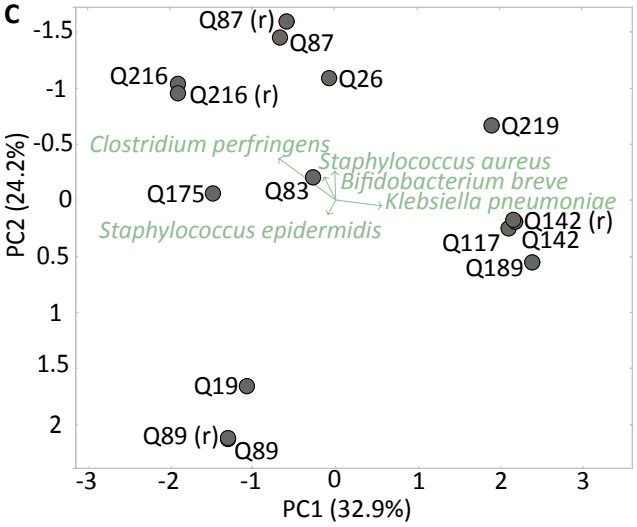

C
PC2 (24.2%)
PC1 (32.9%)
Q87 (r)
Q87
Q216
Q216 (r)
Q26
Q219
Clostridium perfringens
Staphylococcus aureus
Bifidobacterium breve
Klebsiella pneumoniae
Staphylococcus epidermidis
Q175
Q83
Q142 (r)
Q117
Q142
Q189
Q19
Q89 (r)
Q89

Figure 2(on next page)

Antibiotic resistance genes detected.

Heatmap showing distribution of the 39 AMR genes detected within the eleven metagenomic samples. Genes grouped by antibiotic class: AGly (aminoglycosides), Bla (beta-lactamases), Fos (Fosfomycin), Flq (fluoroquinolones), MLS (macrolide-lincosamide- streptogramin), Phe (phenicols), Tet (tetracyclines), Tmt (trimethoprim). Colours show read depth in bp: undetected (grey), 3-199 bp (blue), 200 - 399 bp (yellow), 400 - 1437 (red). Rows clustered by UPGMA method using Euclidean distances.

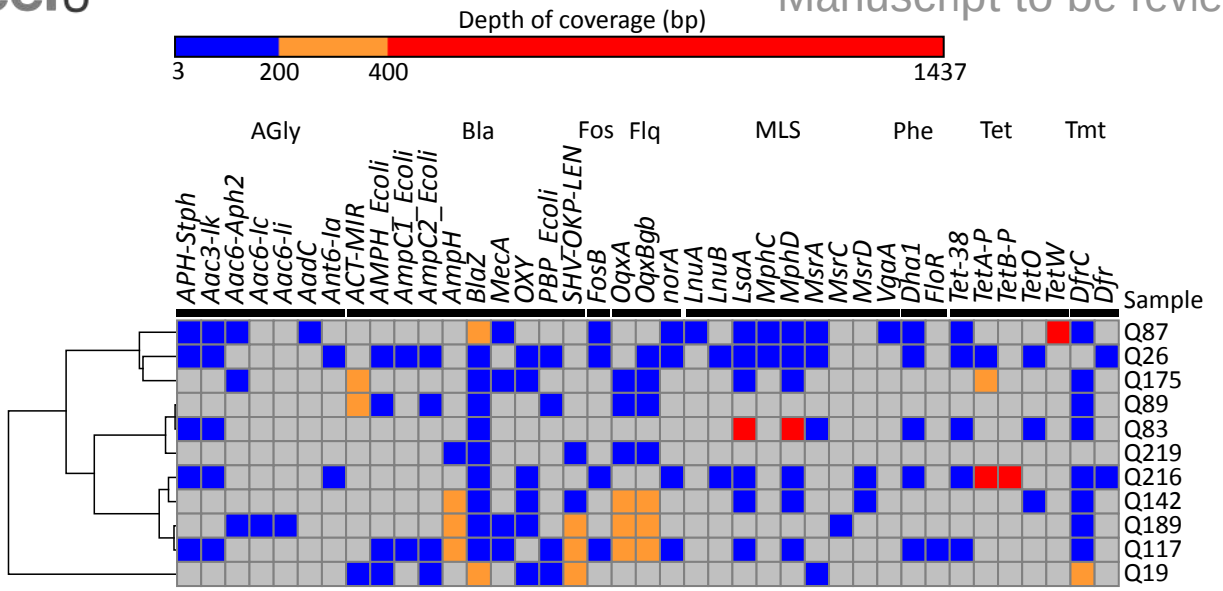

Depth of coverage (bp)
3
200
400
1437
AGly
Bla
Fos
Flq
MLS
Phe
Tet
Tmt
APH-Stph
Aac3-Ik
Aac6-Aph2
Aac6-Ic
Aac6-Ii
AadC
Ant6-Ia
ACT-MIR
AMPH_Ecoli
AmpC1_Ecoli
AmpC2_Ecoli
AmpH
BlaZ
MecA
OXY
PBP_Ecoli
SHV-OKP-LEN
FosB
OqxA
OqxBgb
norA
LnuA
LnuB
LsaA
MphC
MphD
MsrA
MsrC
MsrD
VgaA
Dha1
FloR
Tet-38
TetA-P
TetB-P
TetO
TetW
DfrC
Dfr
Sample
Q87
Q26
Q175
Q89
Q83
Q219
Q216
Q142
Q189
Q117
Q19

Table 1(on next page)

Metagenomic study dataset.

Sequencing results for the eleven preterm samples sequenced and four replicates.

1

Sample	Instrument	Read length	Raw PE reads	Surviving PE reads*	Surviving PE reads %*	Mean read length (bp)	Yield (Gbps)
Q19	NextSeq	151	10,765,181	9,748,620	90.6	118.5	2.3
Q26	NextSeq	151	10,515,261	9,675,994	92.0	125.3	2.4
Q83	NextSeq	151	10,272,541	9,434,574	91.8	128.1	2.4
Q87	NextSeq	151	9,771,928	9,031,166	92.4	126.3	2.3
Q89	NextSeq	151	10,573,383	9,686,447	91.6	130.5	2.5
Q89 (r)	NextSeq	151	10,746,440	9,824,221	91.4	129.1	2.5
Q117	NextSeq	151	9,718,743	8,943,195	92.0	124.4	2.2
Q142	NextSeq	151	9,847,000	9,059,108	92.0	129.0	2.3
Q175	NextSeq	151	11,442,761	10,404,737	90.9	121.1	2.5
Q189	NextSeq	151	9,385,745	8,667,469	92.3	125.2	2.2
Q216	NextSeq	151	11,842,425	10,963,555	92.6	125.8	2.8
Q219	NextSeq	151	5,761,388	5,295,708	91.9	118.4	1.3
Q87 (r)	MiSeq	301	8,061,151	5,710,557	70.8	193.0	1.1
Q142 (r)	MiSeq	301	4,101,014	3,234,121	78.9	204.1	0.7
Q216 (r)	MiSeq	301	12,778,363	10,848,709	84.9	224.0	2.4
Mean	-	-	9,705,555	8,701,879	-	-	2.1
Total	-	-	145,583,324	130,528,181	89.7	-	-

2

3 * Three MiSeq replicate samples paired reads were merged during QC, therefore read number represent
4 single reads

5 (r) signifies replicate samples

Table 2(on next page)

Antibiotic classes identified.

Major antibiotic resistant classes of genes identified within the eleven samples by srst2. Columns show antibiotic type and number of genes found within class.

1

Antibiotic type	Number of identified genes within class
Beta-lactamases (Bla)	10
Macrolide-lincosamide- streptogramin (MLS)	9
Aminoglycosides (AGly)	7
Tetracyclines (Tet)	5
Fluoroquinolones (Flq)	3
Phenicols (Phe)	2
Trimethoprim (Tmt)	2
Fosfomycin (Fcyn)	1

2