# ORIGINAL ARTICLE

# Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses

François Enault[1,2], Arnaud Briet[3], Léa Bouteille[3], Simon Roux[4], Matthew B Sullivan[4,5] and Marie-Agnès Petit[3]

[1]Clermont Université, Université Blaise Pascal, Laboratoire 'Microorganismes: Génome et Environnement', Clermont-Ferrand, France; [2]CNRS UMR 6023, LMGE, Aubière, France; [3]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France; [4]Department of Microbiology, The Ohio State University, Columbus, OH, USA and [5]Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA

Antibiotic resistance genes (ARGs) are pervasive in gut microbiota, but it remains unclear how often ARGs are transferred, particularly to pathogens. Traditionally, ARG spread is attributed to horizontal transfer mediated either by DNA transformation, bacterial conjugation or generalized transduction. However, recent viral metagenome (virome) analyses suggest that ARGs are frequently carried by phages, which is inconsistent with the traditional view that phage genomes rarely encode ARGs. Here we used exploratory and conservative bioinformatic strategies found in the literature to detect ARGs in phage genomes, and experimentally assessed a subset of ARG predicted using exploratory thresholds. ARG abundances in 1181 phage genomes were vastly overestimated using exploratory thresholds (421 predicted vs 2 known), due to low similarities and matches to protein unrelated to antibiotic resistance. Consistent with this, four ARGs predicted using exploratory thresholds were experimentally evaluated and failed to confer antibiotic resistance in *Escherichia coli*. Reanalysis of available human- or mouse-associated viromes for ARGs and their genomic context suggested that *bona fide* ARG attributed to phages in viromes were previously overestimated. These findings provide guidance for documentation of ARG in viromes, and reassert that ARGs are rarely encoded in phages.
*The ISME Journal* (2017) **11,** 237–247; doi:10.1038/ismej.2016.90; published online 21 June 2016

## Introduction

Antibiotic resistance is widely recognized as problematic for the treatment of infectious diseases, due to the emergence of antibiotic-resistant pathogens (World Health Organization, 2014). However, antibiotic resistance genes (ARGs) are also common outside of pathogens. Among 6179 sequenced microbial genomes, 84% have at least one ARG (Gibson et al., 2015) including many diverse, mostly non-pathogenic bacteria in soils (Riesenfeld et al., 2004; Allen et al., 2009; Torres-Cortes et al., 2011; Forsberg et al., 2012) and activated sludge (Mori et al., 2008; Parsley et al., 2010), as well as in humans and animals (Kazimierczak et al., 2009; Sommer et al., 2009; Penders et al., 2013; Wichmann et al., 2014). Further, ARGs can be transferred between soil bacteria and human pathogens via horizontal gene transfer (Forsberg et al., 2012), such that non-pathogens represent a reservoir of ARG to pathogens

(Canica et al., 2015; Martinez et al., 2015; van Schaik, 2015). It becomes critical then to estimate how often ARGs are present on mobile genetic elements (Martinez et al., 2015), and to understand the conditions under which these elements promote horizontal gene transfer in nature, and lead to the emergence of antibiotic resistance among pathogens.

Numerous mobile elements are implicated in the spread of ARGs (Broaders et al., 2013; Huddleston, 2014). These include plasmids and integrative conjugative elements (ICEs), via conjugation (for reviews see Davies and Davies, 2010; Wozniak and Waldor, 2010), as well as generalized transduction carried out by bacterial viruses, that is, phages (Davies and Davies, 2010; Muniesa et al., 2013; Balcazar, 2014; Volkova et al., 2014). Quantitatively, laboratory experiments with phage P1 suggest that ARG transfer is 1000-fold less common via phage transduction than via conjugative elements (Volkova et al., 2014). This is due to the fact that, contrary to conjugation events that will systematically transfer ARG together with the ICE or plasmid genome, generalized transduction relies on erroneous encapsidation of non-phage DNA. Measurements from phage P1 suggest that this is a rare event as only about 4 out of 10$^4$ phage capsids will encode a

chromosomally encoded ARG gene (Volkova et al., 2014). In addition, ARGs are only rarely directly encoded in phage genomes—2 of 1181 publicly available phage genomes contain an ARG (Pruitt et al., 2007; NCBI RefSeq as of April 2014). These include the twin phages Gamma and Cherry of *Bacillus anthracis*, which encode a fosfomycin resistance gene (Schuch and Fischetti, 2006). Beyond these, some *Staphylococcus aureus* satellite phages, which require another phage to propagate, carry ARGs (Novick et al., 2010). Thus, while phage-encoded ARGs would presumably provide advantage to bacteria hosting the phage in a lysogenic state, there appears to be little selection for phages to carry such genes.

In contrast, recent virome studies have reported high levels of ARG in phages from pulmonary human samples of cystic fibrosis patients (Fancello et al., 2011; Rolain et al., 2011), and from feces samples of antibiotic-treated mice (Modi et al., 2013). These studies call into question how prevalent ARGs are among phages and suggest that phages have far greater roles in spreading ARGs. The higher ARG frequencies in viromes could derive either from genome-sequenced phage isolates misrepresenting naturally occurring ARG frequencies, or virome studies overestimating ARG frequencies.

The first step to identify ARGs in virome sequences involves a homology search against a database dedicated to ARGs. The Antibiotic Resistance Database (ARDB, 7828 proteins at last update in 2009; Liu and Pop, 2009) is commonly used, sometimes together with proteins annotated with the GO term 'antibiotic catabolic process' (Modi et al., 2013). The use of ARDB necessitates caution, as recently observed (van Schaik, 2015), and downstream manual inspection to refine *in silico* functional assignments is required. More recently, expert curated databases have become available including the Comprehensive Antibiotic Resistance Database (CARD; McArthur et al., 2013) and Arg-annot (Gupta et al., 2014), which are restricted to experimentally confirmed proteins conferring antibiotic resistance, and Resfams (Gibson et al., 2015), which updates CARD with recently discovered β-lactamases. The first two contain 2822 and 1689 proteins, respectively, whereas the third contains 2097 proteins grouped into 166 families, using an approach similar to Pfam (Bateman et al., 2000). Of these Resfam families, 119 are 'core' and can be used for ARG discovery without ambiguity, whereas 47 are more challenging to interpret as they will also match non-ARG proteins, such as ABC transporters and transcriptional regulators.

Once a reference database is selected, a relevant sequence similarity threshold to identify ARGs among a collection of new sequences must be chosen. The literature offers both conservative and exploratory options. Conservative criteria require an unknown ORF (open reading frame finder) to match the database with either >40% coverage over the target ARG and >80% nucleotide identity (Zankari et al., 2012), or >85% coverage and >80% amino-acid

identity (Gibson et al., 2015). These stringent criteria will largely only identify known ARGs (Zankari et al., 2012; Gibson et al., 2015). However, in the past decade many new ARGs have been discovered by functional screening, which would not have been found using stringent comparisons to databases (Sommer et al., 2009; Parsley et al., 2010; Moore et al., 2013; Wichmann et al., 2014). To discover distantly related ARGs, one may wish to lower down similarity cutoffs, a common practice for virome studies, where high levels of sequence divergence are routinely observed, due to the high mutation rates of phages and lack of explored 'sequence space' resulting in limited reference genomes. In this case, E-value thresholds below $10^{-5}$ or $10^{-3}$ have been used (Willner et al., 2009; Modi et al., 2013).

Here we compare approaches to detect ARGs in phage genomes, experimentally evaluate four predicted ARGs and assess the impact of conservative and exploratory thresholds for inferring ARGs from 25 published viromes, including those with high reported levels of ARGs (Willner et al., 2009; Modi et al., 2013). We build a case that *bona fide* ARG frequencies are vastly overestimated in virome studies, and suggest that the main path for ARG dissemination by phages is generalized transduction as commonly asserted.

## Materials and methods

*Databases of ARG*
The four databases compared in this analysis are an updated version of ARDB (http://ardb.cbcb.umd.edu/, named hereafter ARDB+, 13 453 different proteins, see details below), Arg-annot (http://www.mediterranee-infection.com/article.php?laref=282&titer=arg-annot, download May 2015), CARD (http://arpcard.mcmaster.ca/download, subset download excluding genes that confer resistance via specific mutations, download May 2015) and Resfams (http://www.dantaslab.org/resfams, v.1.2, updated 27 January 2015). The search against Resfams is not performed with BLAST but hmmscan (Finn et al., 2011), using the –cut_ga parameter that sets the threshold for similarity according to the threshold chosen to aggregate the members of each family. ARDB+ contains the 7828 initial proteins from ARDB, complemented with 5625 Uniprot proteins (September 2014) filtered for the GO:0017001 term «antibiotic catabolic process», and sharing <98% identity with ARDB proteins. Remarkably, 80% of these additional proteins are β-lactamases. We observed *a posteriori* that a partition protein (whose function is to stabilize plasmids and temperate phages replicating autonomously) has been mistakenly incorporated into ARDB (XP_002333050), with the annotation 'tetracycline resistance gene from *Populus trichocarpa*'. Concerning the CARP database, a thymidylate synthase from *Enterococcus faecalis* (AF028811.1) was similarly removed from this analysis. All results with ARDB+ and CARD are given after removal of

these two erroneous matches. ARDB+ being significantly more populated than CARD, Arg-annot and Resfams, we compared the content of these databases. Approximately 88% of ARDB+ proteins were similar (with BLAST and bit-score >70) to those of CARD and Arg-annot (Supplementary Information S1), and 62% only are found in Resfams (with hmmscan and the stringent built-in threshold).

*Reference set of proteins from complete phage genomes, and generation of mock viromes*
Phage proteins from the 1181 genomes (121 505 proteins) were downloaded from the NCBI viral genome database (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=virus&taxid=10239&host=bacteria, 8 April 2014). To generate the mock200 and mock580 reads from this set of genomes, successive DNA fragments of identical sizes were generated sequentially from the genome files.

*Viral metagenomic data*
All viromes composed of at least 50 000 reads available at the beginning of the analysis (2014), and originating from human- or mice-associated bacterial ecosystems, were downloaded and studied (25 viromes in total). Of these, 23 were available as unassembled reads: 10 viromes from human lung (Willner et al., 2009), 9 from human feces (5 from Kim et al. (2011) and 4 from Minot et al. (2011)) and 4 from mice feces (Modi et al., 2013). In addition, two already assembled data sets, both from human feces (Reyes et al., 2010; Minot et al., 2012), were also considered. All unassembled viromes were sequenced using different generations of the pyrosequencing technology (454; Roche, Branford, CT, USA) and the average read length of individual data sets is comprised between 205 and 873.

To normalize read length among viromes and have comparable results for all viromes, comparisons with the three databases were also performed on viromes where each read was randomly truncated to 200 bp.

For the mice viromes (Modi et al., 2013), low-quality ends of the reads were trimmed (quality score <20) and reads shorter than 100 bp were removed.

*Analysis of raw reads for 16S content and bacteria-only COGs*
All reads from the 23 unassembled viromes were truncated to 200 bp and compared (BLASTn, bit-score >200) to the SILVA database (Quast et al., 2013) for the identification of 16S rDNA. To refine bacterial DNA detection, the protein families (clusters of orthologous groups of proteins (COGs)) the most frequently observed in human digestive tract microbiomes were taken from the Qin et al. (2010) gut metagenomic analysis. The 1112 prevalent COGs in all individuals of their study were compared with the 1 72 774 proteins of all sequenced viruses present in the RefSeq database (Pruitt et al., 2007) (http://

www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239). Using a threshold of 50 on the bit-score, the 688 COGs having no similarities with viral proteins were further used and termed 'bacteria-only COGs'.

To estimate the amounts of bacteria-only COG genes in bacterial genomes, 1102 completely sequenced genomes from the KEGG database (version of 2011) were chopped in 580-bp-long reads and compared with the bacteria-only COGs (BLASTx bit-score >50). In all, 21.1% of bacterial reads had a hit against these COGs. To estimate the amount of ARGs in bacterial genomes recovered with the bit-score 70 threshold, the same 580-bp-long bacterial mock reads were compared with BLASTx against ARDB+: 2.2% of these reads had a hit against ARDB+.

*Assembly and analysis of contigs*
All viromes were assembled *de novo* with Newbler 2.6 (454; Life Sciences, Branford, CT, USA), using the threshold of 98% identity on 35 bp. For each contig longer than 500 bp of the 25 assembled viromes (23 assembled in-house, 2 already assembled), an ORF prediction was processed using MetaGeneAnnotator (Noguchi et al., 2008) with default parameters. Predicted ORFs were compared with ARDB+ proteins (bit-score >70) using BLASTp, and to Resfams using hmmer.

*Cloning and testing of four putative phage-encoded ARG*
Strains, plasmid construction and testing are described in Supplementary Information S2.

## Results

*Assessing informatic stringencies to identify candidate ARGs in phage genomes*
To evaluate relevant databases and thresholds for identifying known ARGs, we examined the number of ARGs we could detect among the 1 21 506 proteins (termed 'proteome') from 1181 reference phage genomes. As described above, only two ARGs were expected with this data set—the *fos* genes of phages Gamma and Cherry (Schuch and Fischetti, 2006). Using a conservative threshold (>40% coverage and >80% amino-acid identity), BLASTp screens of the reference phage proteome identified the two positive control *fos* genes, as well as a possible β-lactamase of phage G against ARDB+, and no hits against CARD and Arg-annot (Figure 1a and Supplementary Information S3). Using the same proteome and the built-in conservative threshold of Resfams identified one hit in the *yokD* gene, an aminoglycoside acetyl transferase of *Bacillus subtilis* prophage SPβ. Neither the phage G 'β-lactamase' gene nor the SPβ *yokD* gene is proven to function as ARGs, but represent strong candidates for future experimentation. The lack of detection of the *fos* genes outside the ARDB+
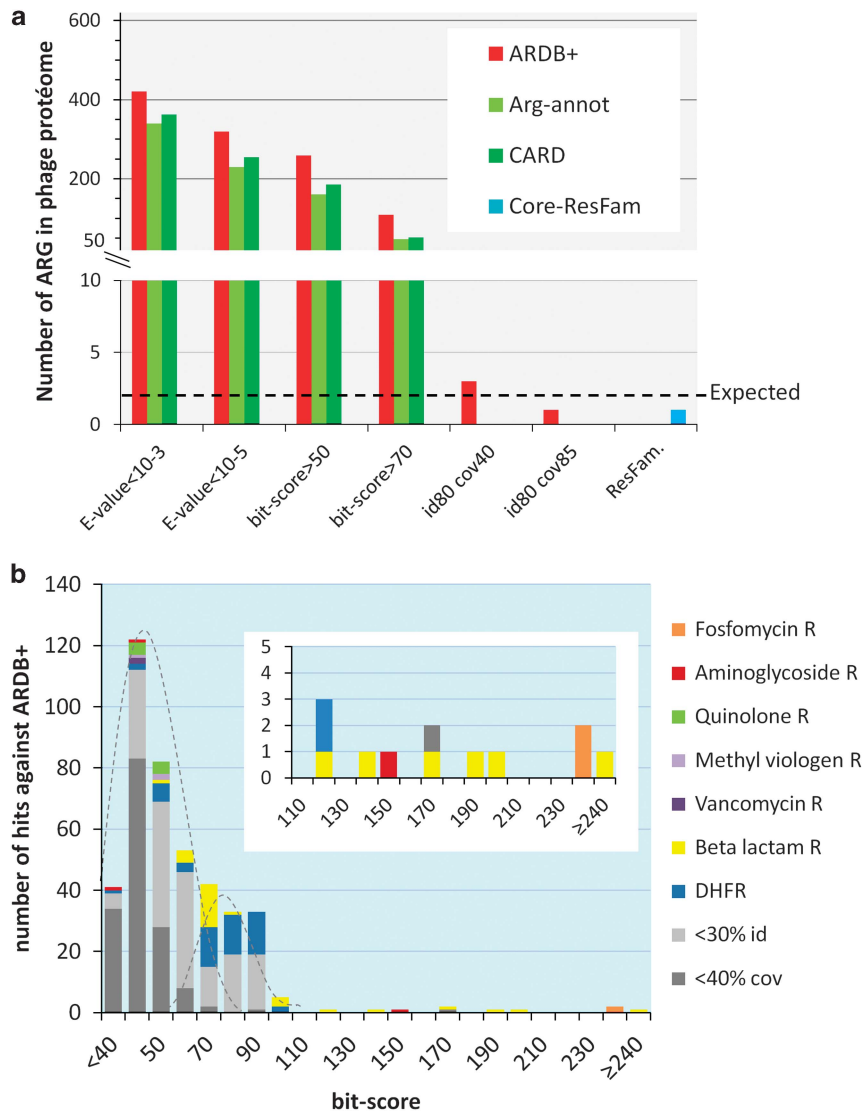
**Figure 1** Analysis of the reference phage proteome against various antibiotic resistance genes database. (**a**) Comparison of the recovery of ARG hits with the four ARG databases, using the conservative thresholds of Resfams, $>40\%$ coverage and $>80\%$ amino-acid identity, or $>85\%$ coverage and $>80\%$ amino-acid identity, as well as the exploratory thresholds E-values of $10^{-3}$ and $10^{-5}$. However, as E-values on large databases cannot be simply transposed for the much smaller ARG databases, two additional bit-score thresholds (a statistics independent of the database size) were introduced, 50 and 70. The dotted line indicates the expected number of hits (2), according to experimental data. (**b**) The 421 hits against ARDB+, obtained with the most exploratory cutoff (E-value $<10^{-3}$) are displayed, as a function of their bit-score value, with a color code for resistance category. In gray, the hits with $<30\%$ identity or $<40\%$ coverage, which are most likely false positives. Zoom inset: hits with bit-scores $>110$.

database was owing to the fact that the *fos* genes present in the other databases were too divergent (57–63% identity). In the following analyses, the most recent Resfams tool was used to detect ARGs among protein data sets, but as Resfam is not adapted to raw DNA reads, BLASTx comparisons against ARDB+ with cutoffs of $>40\%$ coverage and $>80\%$ amino-acid identity was used instead.

In contrast to these conservative threshold results, the exploratory thresholds resulted in hundreds of hits for all databases considered (Figure 1a and Supplementary Information S3). To better understand whether these should reasonably be considered candidate ARGs, the 421 hits recovered with the exploratory

cutoff (E-value $<10^{-3}$) against ARDB+ were further examined. The bit-score distribution of these hits was bimodal with a bit-score cutoff of ∼70 at the junction (Figure 1b), suggesting a first population ($<70$ bit-scores) of random hits, distinct from a second population ($>70$ bit-scores) of significant hits. We therefore manually inspected the 109 hits with bit-scores $>70$. Ninety-six of these could be rejected as nonsignificant because the target ARG was poorly covered ($<40\%$), too highly divergent ($<30\%$ identity) or else had homology to enzymes likely to serve non-ARG functions in phages. On the latter, homology with dihydrofolate reductase was rejected without further consideration, as this enzyme
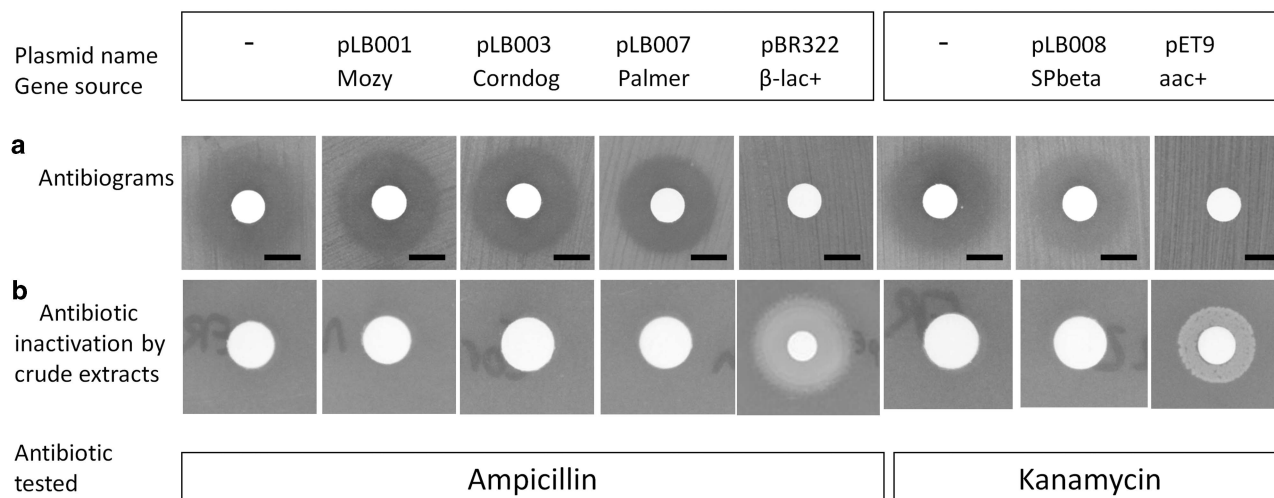
**Figure 2** Experimental testing of four predicted ARGs. (**a**) Antibiograms. The three β-lactamases were tested *in vivo* by spreading bacterial lawns of *E. coli* expressing each protein (100 μM isopropyl β-D-1-thiogalactopyranoside (IPTG)) into top-agar, and spotting 10 μg of ampicillin on 6-mm-diameter disks. For YokD, 30 μg kanamycin was used. Plates were incubated at 30 °C. Confluent growth was observed for the pBR332- (β-lac+: β-lactamase positive) or pET9- (aac+: aminoglycoside acetyl transferase positive) positive controls, but inhibition zones were present for all phage-encoded putative enzymes. Bar, 6 mm. (**b**) Enzymatic tests. Ampicillin (100 μg ml$^{-1}$) or kanamycin (50 μg ml$^{-1}$), together with an indicator ER2566 *E. coli* strain sensitive to both antibiotics, were spread into top-agar. Soluble fractions of total extracts of *E. coli* ER2566 expressing each of the four proteins, or expressing the ARG of plasmid controls, were then spotted on 6-mm-diameter disks. Plates were incubated for 24 h at 30 °C. Bacterial growth around the disk indicates that the protein extract contains antibiotic resistance activity, which degraded sufficient amounts of antibiotic in the top-agar to permit growth of the antibiotic-sensitive *E. coli* strain.

potentially conferring trimethoprim resistance is more likely to be involved in nucleotide metabolism in phages (Asare *et al*., 2015b). β-Lactamases were closely inspected, as several reports have suggested their presence on phages (Quiros *et al*., 2014; Asare *et al*., 2015a), sometimes with atypical length. Indeed, one such protein annotated as a β-lactamase is, in fact, a tail fiber protein (Cresawn *et al*., 2015). Half of the putative β-lactamases were rejected because of atypical lengths and/or similarity to tail proteins (see Supplementary Information S4 for the complete analysis of β-lactamase hits, and Supplementary Information S5 for examples of rejected hits). In total, 13 hits were retained after manual inspection, including the two experimentally proven *fos* genes, and 11 additional candidates that are likely worth experimental follow-up, including the aminoglycoside acetyl transferase from the *B. subtilis* SPβ prophage, and 10 β-lactamases.

Taken together, these results suggest that findings using the conservative threshold, even against the permissive ARDB+ database, will recover only *bona fide* ARGs, whereas the exploratory thresholds may lead to the discovery of novel ARGs, but do so at the expense of hundreds of false positives.

*Experimental testing of four putative phage-encoded ARG*
To test whether the stringency of this manual screening, which removed 90% of the hits, was appropriate, four ARG candidates were chosen for experimental evaluation. Specifically, *yokD* of phage SPβ, encoding a putative aminoglycoside transferase, and three β-lactamases related to those listed in Supplementary

Information S4 (*gp34* of phage Palmer (99% identical amino acids to phage Pony Gp33), *gp62* of mycophage Corndog and *gp20* of Mozy (97% identical amino acids to *gp20* of phage Che8)) were examined. These last two putative β-lactamases were rejected upon manual inspection, and suspected to be rather phage tail proteins (Cresawn *et al*., 2015). All four proteins were expressed in *Escherichia coli* (see Supplementary Information S2 for the origin of genes, and cloning details) and found to be soluble (Supplementary Information S6). However, none of the four proteins conferred antibiotic resistance, in antibiogram assays (Figure 2a). As none of the putative β-lactamases contained predictable signal peptides, and to exclude the possibility that heterologous expression prevented resistance *in vivo*, the activity of crude extracts was tested *in vitro*. For this, 40 μl of a soluble total protein extract was spotted on a lawn of top-agar containing an indicative *E. coli* strain, which growth was prevented by the supplementation of ampicillin (or kanamycin). Inactivation of the antibiotic around the spot containing the modifying enzyme permitted local growth of *E. coli*, as shown with the positive controls (strain expressing the pBR322 encoded β-lactamase, or the pET9 encoded aminoglycoside acetyl transferase). Again, none of the crude extracts with phage-encoded candidate ARG permitted antibiotic degradation (Figure 2b). In addition to ampicillin, the putative β-lactamases were tested on mecillinam (another penicillin, against which a prophage gene similar to Palmer gp34 has reported activity; Ogilvie *et al*., 2013), ceftazidime (cephalosporin) and aztreonam (monobactam) with the same negative result. We conclude that among these four phage genes, none was a *bona fide* ARG. The absence of

resistance with the YokD protein was surprising, given its very good similarity scorings (bit-score = 156, E-value = $10^{-46}$). However, the closely related putative aminoglycoside transferase BA2930 of *B. anthracis* is also unable to degrade kanamycin in antibiogram assays (Klimecka *et al.*, 2011). It is a proven acetyl transferase, but seems to have a substrate distinct from aminoglycoside antibiotics. We conclude that exploratory thresholds, even after manual curation, lead to overestimations of ARG counts.

### Detecting ARGs in virome reads

Given the above informatics benchmarking, and its conflict with experimental tests, we next sought to quantify the number of ARGs in viromes. However, as opposed to the full-length proteins examined above, the length of virome reads is also critical to evaluate, as they impact ARG discovery. To this end, we mimicked currently available virome read lengths by *in silico* fragmenting the 1181 reference phage genomes to create 200 or 580 bp reads. These mock community viromes were termed mock200 and mock580, respectively, and then compared with the same databases (BLASTx) using the conservative and exploratory thresholds (see Supplementary Information S3). Results paralleled those from the full-length proteome analyses for conservative thresholds. For the exploratory threshold based on BLAST bit-score > 70, hits were fewer among short reads (0.09% of all reads for mock200 and 0.60% for mock580, against ARDB+) than among full-length proteome analyses (0.81%), so that specificity increased slightly (Supplementary Information S7). Therefore, conclusions drawn from full genomes hold mostly for short reads, and suggest the use of the > 70 bit-score threshold for exploratory searches, to minimize false positives.

### More ARGs in viromes than in reference phage genomes?

Given the above analyses for selecting appropriate thresholds to identify ARGs in reference genome data sets, we next applied the conservative and exploratory thresholds to 25 publicly available human- or animal-associated viromes. These include 10 from the human lung with and without cystic fibrosis (Willner *et al.*, 2009), 9 from human feces of healthy subjects (Kim *et al.*, 2011; Minot *et al.*, 2011) and 4 from mouse feces with and without antibiotic treatment (Modi *et al.*, 2013), as well as 2 additional viromes from human feces (Reyes *et al.*, 2010; Minot *et al.*, 2012) where only assembled contigs are available.

Comparison of the virome reads against ARDB+ with a conservative threshold gave no ARG hit for all human feces viromes, a single hit for the ciprofloxacin-treated sample of mice viromes and 25 hits for the lung samples (Supplementary Information S8). Application of an exploratory cutoff (BLASTx, bit-score > 70)
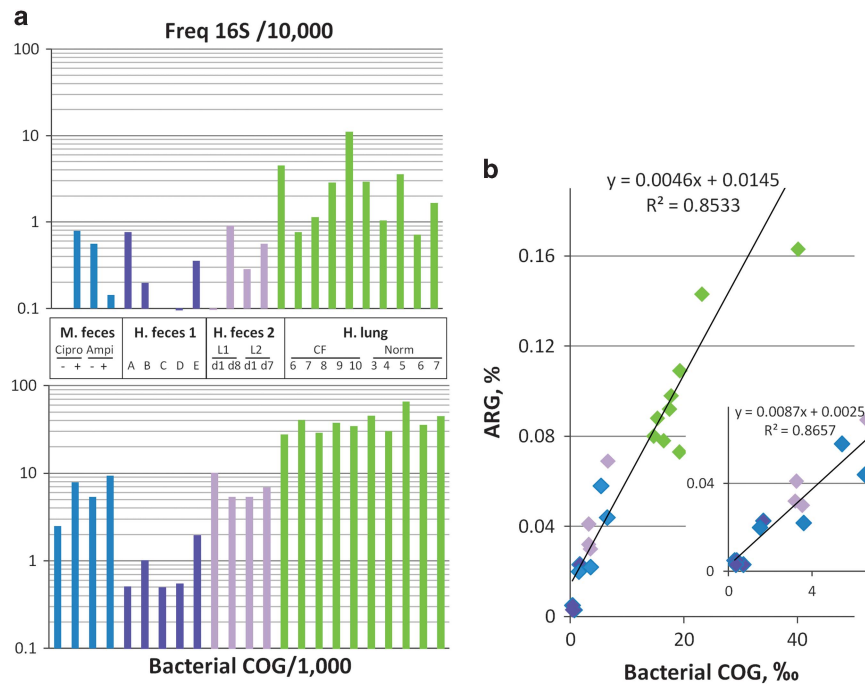


**Figure 3** (**a**) Levels of bacterial DNA content in the 23 unassembled viromes. 'M. feces', mouse feces samples from Modi *et al.* (2013); Amp, ampicillin; Cipro, ciprofloxacin; mice treated ± with without antibiotic. 'H. feces 1', human feces samples from Kim *et al.* (2011). 'H feces 2', human feces viromes from Minot *et al.* (2011); L1 and L2 are two lean subjects, at two different time points. 'H. lung', human lung samples from Willner *et al.* (2009); CF, cystic fibrosis patients; Norm, healthy subjects. Upper panel: Proportion of reads matching against 16S DNA. Lower panel: Proportion of reads matching against bacteria-only COGs. (**b**) Correlation between bacterial (x axis) and ARG DNA (y axis) amounts in human-associated viromes. Data are taken from Supplementary Information S8, ARG matches are with the exploratory threshold, i.e. bit-score > 70. The same color code as in panel a is applied. Inset: Same plot after removal of the lung sample data points.

yielded 0–0.19% of virome reads as possible ARGs, depending upon the virome, with the lung samples again outliers (Supplementary Information S8). Therefore, at this stage, ARGs did not appear more abundant in viromes than in viruses, except possibly in lung samples.

*Variable yet often non-negligible presence of bacterial DNA in viromes*
Bacterial DNA is common in viromes (Roux *et al.*, 2013) and can derive either from phage-encoded bacterial genes (this is what we are looking for in the present study), generalized transduction (a well-known process whereby a bacterial DNA segment of the size of the viral genome is taken mistakenly into the capsid instead of the viral DNA) or from insufficient removal of bacterial DNA (either free or entrapped into vesicles) prior phage particle DNA purification. Discriminating between these possibilities is critical towards understanding the mobility of the bacterial DNA detected in the viromes. Currently, when >0.2‰ of reads match to 16S ribosomal DNA, a non-viral marker of bacterial DNA, the viromes are considered to have high bacterial DNA content (Roux *et al.*, 2013), suggestive of external contamination. Examination of the reads from the 23 unassembled viromes for 16S ribosomal DNA content showed that 17 of the viromes had low (<0.2‰ of 16S DNA reads) levels of bacterial DNA content, whereas 6 lung samples had high levels (Figure 3a, upper panel).

To increase sensitivity relative to the single 16S gene approach, we established 'bacteria-only' protein families by screening the 1112 most frequently observed COGs in gut microbiomes (Qin *et al.*, 2010) against viral proteins derived from NCBI RefSeq genomes. This revealed 688 COGs with no similarities to viral proteins, which we termed 'bacteria-only COGs'. The virome reads were then compared with bacteria-only COGs, which largely corroborated the findings from the 16S analyses (Figure 3a, lower panel), but revealed that all 10 lung samples had similar high levels of bacterial DNA (1.5–4% of matches to bacteria-only COGs).

*The more bacterial DNA, the more ARG detected*
Previously, the mouse feces viromes (Modi *et al.*, 2013) were used to examine the impact of 8-week antibiotic treatments (ampicillin or ciprofloxacin) on the frequency and spread of ARG. The major findings were that treated mice contained ~ 3-fold increase in reads identified as ARG compared with untreated animals, which was interpreted to suggest that ARG were spread through phage genomes. Notably, however, our reanalyses of these data show that there is also a two- to threefold increase in the bacteria-only COGs across these same treatments (Figure 3a, lower panel, blue bars). This suggests that all types of bacterial genes are more frequently detected in the treated mouse viromes, with no particular selection for ARG. Mechanistically, this may be due to the

**Table 1** ARG detected on viral and bacterial contigs >2 kb, among the 25 viromes

| | | Contig classification | | |
|---|---|---|---|---|
| | | *Viral* | *Bacterial* | *Undetermined* |
| Mice feces (Modi *et al.*, 2013) | Total number of genes ARG genes[a] | 4423 (on 465 contigs) 0/0 | 4 (on 1 contig) 0/0 | 375 (on 72 contigs) 0/0 |
| Human feces (Minot *et al.*, 2011) | Total number of genes ARG genes | 3848 (on 475 contigs) 0/2 | 0 0/0 | 584 (on 119 contigs) 0/1 |
| Human feces (Kim *et al.*, 2011) | Total number of genes ARG genes | 2540 (on 270 contigs) 0/0 | 0 0/0 | 586 (on 113 contigs) 0/0 |
| Human feces (Minot *et al.*, 2012) | Total number of genes ARG genes | 20 925 (on 2168 contigs) 0/3[b] | 414 (on 97 contigs) 0/20 | 3240 (on 603 contigs) 3/18 |
| Human feces (Reyes *et al.*, 2010) | Total number of genes ARG genes | 2385 (on 88 contigs) 0/0 | 0 0/0 | 199 (on 5 contigs) 0/0 |
| Human lung (Willner *et al.*, 2009) | Total number of genes ARG genes | 1831 (on 269 contigs) 0/0[c] | 12 871 (on 2164 contigs) 24/450 | 1958 (on 592 contigs) 1/38 |
| Total | Total number of genes ARG genes ARG/total cons/expl | 35 952 (on 3735 contigs) 0/5 $<2.8 \times 10^{-5}/1.39 \times 10^{-4}$ | 13 289 (on 2262 contigs) 24/470 $1.8 \times 10^{-3}/3.5 \times 10^{-2}$ | 6942 (on 1504 contigs) 4/57 $5.7 \times 10^{-4}/8.2 \times 10^{-3}$ |

Abbreviation: ARG, antibiotic resistance gene.
[a]Counts with conservative (core-Resfams)/explorative (ARDB+, bit-score >70 and manual curation) thresholds.
[b]Eleven hits in total, six removed: three vancomycin two component regulator (other van genes missing), three ABC transporters (coverage <50%). Among the five remaining contigs, two were redundant (see Supplementary Information 10).
[c]Four hits in total, four removed: coverages <50% to the target ARG.

antibiotic-treatment-inducing prophages, with some subset performing generalized transduction.

Further, the bacteria-only COGs help uncover background bacterial DNA contamination that confounds interpretations of ARG frequencies across all 23 viromes examined here. Specifically, the percentage of reads matching ARG using the exploratory cutoff (bit-score >70, against ARDB+) is correlated to that for bacteria-only COGs for all 23 viromes (Figure 3b). The fact that the quantities of ARG and bacterial DNA are correlated suggests that the ARG signal in viromes derives from bacterial rather than phage genomes.

*ARG are rare in viral contigs*
At this point, it appears that the majority of the ARG in viromes are confounded with regular bacterial genes, due to the exploratory threshold used in virome publications (Fancello *et al.*, 2011; Modi *et al.*, 2013). However, there may still be some cases of ARG in viruses, as suggested in the mice study, and hypothesized to be due to long-term high-dose antibiotic treatment (Modi *et al.*, 2013).

To separate the bacterial from the viral signal, and to gain further insight into whether phages from viromes encoded ARG in their genomes, we assembled the viromes and examined the resulting contigs for the genomic context of ARG. While this will only evaluate the 'dominant' viral and transducing DNA of the samples (20% of all reads for 454 viromes map to >2-kb-long contigs), it offers yet another opportunity to assess the origin of ARG in viromes. In a first step, all contigs of a size above 2 kb were assigned to one of the three following categories: (i) *viral* when a majority of genes matched to viral proteins or proteins of unknown function (with a minimum of 1 viral gene), (ii) *bacterial* when a majority of genes matched bacterial genes and had no viral-typical genes or (iii) *unknown* where too few genes (<3), a majority of genes of unknown function, or as many bacterial and viral genes (±20%) resided on the contig.

Next, contigs of each category were searched for the presence of ARG, using BLAST bit-score >70 and Resfams cutoffs as exploratory and conservative thresholds, respectively. Again, such analyses revealed them to be very uncommon (summarized in Table 1, and Supplementary Information S9). Among 465 contigs from mouse samples classified as viral, not a single one contained an ARG, even with ARG identified using the exploratory cutoff. In the lung viromes, as expected from the read analysis, most (68%) of the contigs were bacterial, and ARG represented 0.2% (conservative) to 3.5% (exploratory) of the total genes on these bacterial contigs. None of the lung contigs of viral origin contained ARG. Globally across all viral contigs from all 25 viromes, not a single ARG was detected with the conservative cutoff, and only 5 using the exploratory cutoff (Table 1, map of the contigs

Supplementary Information S10). All five were putative β-lactamases, and belonged to the same Pfam family as the Palmer-encoded metal-hydrolase tested above, which did not degrade β-lactams. Still, in some *Bacteroides* prophages, a gene of this same family has reported antibiotic resistance activity (Ogilvie *et al.*, 2013), so that experimental testing is needed to validate the prediction.

Thus, among 35 952 proteins predicted across all available human- and mouse-associated contigs of viral origin, a total of 0 (conservative) and 5 (exploratory) putative ARG were found which suggests an *in silico* frequency of $<2.8 \times 10^{-5}$ (conservative) and $1.4 \times 10^{-4}$ (exploratory) for phage-encoded ARG. In sequenced phages, 1 and 13 putative ARGs were detected *in silico*, with the same cutoffs, for 1 21 506 total proteins, amounting to similar ratio of $0.8 \times 10^{-5}$ and $1.1 \times 10^{-4}$. We conclude that frequencies of phage-encoded ARG in viromes are no higher than those in sequenced phage genomes.

After completion of this analysis, a study focusing on saliva and feces virome samples of human subjects that had been treated or not with antibiotics was published (Abeles *et al.*, 2015). ARG frequency in virome samples was in the 1% range, using a conservative threshold (E-value $<10^{-30}$ at the DNA level, against CARD) and not significantly different between treated and untreated samples (Abeles *et al.*, 2015). Such values being 100-fold higher than in the present study, we included these new viromes in a simplified analysis as follows: (i) bacterial contamination level among raw reads was measured with bacteria-only COGs. Feces samples had generally low levels of matches to such COGs ($0.68 \pm 0.76\%$), but saliva sample levels of bacterial DNA were high (similar to lung virome samples, $4.25 \pm 0.87\%$) despite a very low 16S content (see Supplementary Information S8). (ii) Contigs were assembled, ARGs were searched for with the Resfams database (conservative prediction), and origin of ARG-positive contigs was determined (see Supplementary Information S9). Among all feces viromes, 10 contigs had a putative ARG, two of which were of viral origin. Dividing by the total number of genes found on all feces contigs gave an overall frequency of $7 \times 10^{-5}$. We conclude that the ARG frequency in this data set is in a range similar to other viromes. The prior overestimation might be related to the predominance of the CARD category 'drug transporters' (Abeles *et al.*, 2015). Our conservative analysis used Core-Resfams, in which some transporters have been excluded, owing to their large promiscuity (Gibson *et al.*, 2015).

## Discussion

As sequence analysis of uncultivated viral communities becomes more widespread, we present these analyses as a case study to help establish the boundaries of viromic inference. Earlier work has suggested that ARGs were enriched in the genomes

of antibiotic-treated phage communities, both in the lungs of cystic fibrosis patients (Fancello *et al.*, 2011) and in the feces of mice (Modi *et al.*, 2013). However, reanalysis of these data suggests that the lung sample conclusions were misled by excessive bacterial DNA content, and the mice virome analyses suffered from inflated false positives due to relaxed thresholds for *in silico* detection of ARG (E-value $<10^{-3}$ on ARDB+). To guide future work, we suggest that (i) bacterial DNA contamination be quantified in viromes using analyses such as those presented here and/or automated software now available (VirSorter; Roux *et al.*, 2015), (ii) automated analyses use a conservative threshold to quantify *bona fide* ARG and (iii) discovery-based work proceed with added caution. Specifically, the latter exploratory cutoffs should use a bit-score $>70$ threshold complemented with manual inspection for removing the kind of false positives identified in this study. Additionally, assembly in contigs should be used where possible to confirm that novel ARGs are really present on viral contigs and thereby avoid being misled by general transduction or contaminating bacterial DNA. Finally, only experimental testing will ascertain the function of a predicted ORF as an ARG.

The present analysis of 25 human- and animal-associated viromes does not suggest that a paradigm shift is needed with respect to ARG content of phages: phage genomes rarely carry ARGs, even under intense selection for antibiotic resistance. It will be interesting in the future to investigate whether this remains true in other environments impacted by antibiotics such as soils. We sought to test experimentally whether some of the newly predicted ARGs identified with the exploratory cutoff on complete phage genomes were *bona fide* ARGs, and found that none of them were. This suggests that one should stick to conservative cutoffs for asserting ARG presence, or complement predictions made with exploratory cutoffs with experimental data. That no particular increase in ARG frequency was observed in the antibiotic-treated mouse viromes suggests that ARG did not become part of viral genomes in these samples, at least in the dominant phage population for which contigs could be assembled. Independently, on human salivary and feces virome samples, a recent study reached the same conclusion of an absence of ARG enrichment upon treatment (Abeles *et al.*, 2015). In a context of general warning and concern about the consequences of rampant exposure to antibiotics, this observation is good news as it constrains the spread of ARG by phages predominantly to mechanisms already well known, such as generalized transduction. Our observation of a two- to threefold increase of bacterial DNA content in viromes from mice treated with antibiotics suggests that antibiotics might induce prophages, as already suggested for pig microbiota (Allen *et al.*, 2011), with a concomitant increase of generalized transduction, and low-frequency ARG transfer.

Because metabolic genes are selected for in aquatic phages (Hurwitz *et al.*, 2013, 2014; Roux *et al.*, 2013, 2014; Anantharaman *et al.*, 2014), we posit that ARG acquisition must be largely counterselected in phages. Consistent with this, one study reports an attempt to clone ARG out of an activated sludge viral metagenomic library, with no success (Parsley *et al.*, 2010). Moreover, ARGs are 10-fold less abundant in phages than prophages (Kleinheinz *et al.*, 2014). Mechanistically, it is likely that an ARG on a mobile element invading a prophage will inactivate it, so that the prophage will no longer replenish the pool of 'living' phages. Consistent with this tenant, there are several prophages carrying ARGs for which lytic activity has not been obtained (Billard-Pomares *et al.*, 2014; Wipf *et al.*, 2014). In particular, three well-studied ARGs carrying prophages in *Streptococcus pyogenes* have not been shown yet to be capable of phage lytic activity (Banks *et al.*, 2003; Brenciani *et al.*, 2010; Iannelli *et al.*, 2014). The three elements carry over the antibiotic resistance phenotype across Streptococcal species when cells are put into contact, using a conjugation protocol (Santagati *et al.*, 2003; Giovanetti *et al.,* 2014). Whether these ARG-carrying prophages that may not be lytically active represent an important category of mobile elements in terms of ARG spreading remains an open question.

In summary, these reanalyses present a roadmap for drawing robust conclusions about ARGs, and more generally all types of bacterial genes in virome data sets. As a result, the emerging picture for the spread of ARG suggests that despite the excessive use of antibiotics in humans and animals, ARGs may be among the 'bacterial host genes' that have not (yet?) been selected for in phage genomes, at least in the human and mice-associated environments studied here.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

## References

Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT. (2015). Effects of long term antibiotic therapy on human oral and fecal viromes. *PLoS One* **10**: e0134941.

Allen HK, Looft T, Bayles DO, Humphrey S, Levine UY, Alt D, Stanton TB. (2011). Antibiotics in feed induce prophages in swine fecal microbiomes. *MBio* **2**: e00260-11.

Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. (2009). Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J* **3**: 243–251.

Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science* **344**: 757–760.

Asare PT, Bandara N, Jeong TY, Ryu S, Klumpp J, Kim KP. (2015a). Complete genome sequence analysis and identification of putative metallo-beta-lactamase and SpoIIIE homologs in *Bacillus cereus* group phage BCP8-2, a new member of the proposed Bastille-like group. *Arch Virol* **160**: 2647–2650.

Asare PT, Jeong TY, Ryu S, Klumpp J, Loessner MJ, Merrill BD *et al.* (2015b). Putative type 1 thymidylate synthase and dihydrofolate reductase as signature genes of a novel Bastille-like group of phages in the subfamily Spounavirinae. *BMC Genomics* **16**: 582.

Balcazar JL. (2014). Bacteriophages as vehicles for antibiotic resistance genes in the environment. *PLoS Pathog* **10**: e1004219.

Banks DJ, Lei B, Musser JM. (2003). Prophage induction and expression of prophage-encoded virulence factors in group A Streptococcus serotype M3 strain MGAS315. *Infect Immun* **71**: 7079–7086.

Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. (2000). The Pfam protein families database. *Nucleic Acids Res* **28**: 263–266.

Billard-Pomares T, Fouteau S, Jacquet ME, Roche D, Barbe V, Castellanos M *et al.* (2014). Characterization of a P1-like bacteriophage carrying an SHV-2 extended-spectrum beta-lactamase from an *Escherichia coli* strain. *Antimicrob Agents Chemother* **58**: 6550–6557.

Brenciani A, Bacciaglia A, Vignaroli C, Pugnaloni A, Varaldo PE, Giovanetti E. (2010). Phim46.1, the main *Streptococcus pyogenes* element carrying mef(A) and tet(O) genes. *Antimicrob Agents Chemother* **54**: 221–229.

Broaders E, Gahan CG, Marchesi JR (2013). Mobile genetic elements of the human gastrointestinal tract: potential for spread of antibiotic resistance genes. *Gut Microbes* **4**: 271–280.

Canica M, Manageiro V, Jones-Dias D, Clemente L, Gomes-Neves E, Poeta P *et al.* (2015). Current perspectives on the dynamics of antibiotic resistance in different reservoirs. *Res Microbiol* **166**: 594–600.

Cresawn SG, Pope WH, Jacobs-Sera D, Bowman CA, Russell DA, Dedrick RM *et al.* (2015). Comparative genomics of cluster o mycobacteriophages. *PLoS One* **10**: e0118725.

Davies J, Davies D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**: 417–433.

Fancello L, Desnues C, Raoult D, Rolain JM. (2011). Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota. *J Antimicrob Chemother* **66**: 2448–2454.

Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MO, Dantas G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**: 1107–1111.

Gibson MK, Forsberg KJ, Dantas G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J* **9**: 207–216.

Giovanetti E, Brenciani A, Morroni G, Tiberi E, Pasquaroli S, Mingoia M *et al.* (2014). Transduction of the *Streptococcus pyogenes* bacteriophage Phim46.1, carrying resistance genes mef(A) and tet(O), to other *Streptococcus* species. *Front Microbiol* **5**: 746.

Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L *et al.* (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* **58**: 212–220.

Huddleston JR. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infect Drug Resist* **7**: 167–176.

Hurwitz BL, Brum JR, Sullivan MB. (2014). Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J* **9**: 472–484.

Hurwitz BL, Hallam SJ, Sullivan MB. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* **14**: R123.

Iannelli F, Santagati M, Santoro F, Oggioni MR, Stefani S, Pozzi G. (2014). Nucleotide sequence of conjugative prophage Phi1207.3 (formerly Tn1207.3) carrying the mef(A)/msr(D) genes for efflux resistance to macrolides in *Streptococcus pyogenes*. *Front Microbiol* **5**: 687.

Kazimierczak KA, Scott KP, Kelly D, Aminov RI. (2009). Tetracycline resistome of the organic pig gut. *Appl Environ Microbiol* **75**: 1717–1722.

Kim MS, Park EJ, Roh SW, Bae JW. (2011). Diversity and abundance of single-stranded DNA viruses in human feces. *Appl Environ Microbiol* **77**: 8062–8070.

Kleinheinz KA, Joensen KG, Larsen MV. (2014). Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* **4**: e27943.

Klimecka MM, Chruszcz M, Font J, Skarina T, Shumilin I, Onopryienko O *et al.* (2011). Structural analysis of a putative aminoglycoside *N*-acetyltransferase from *Bacillus anthracis*. *J Mol Biol* **410**: 411–423.

Liu B, Pop M. (2009). ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res* **37**: D443–D447.

Martinez JL, Coque TM, Baquero F. (2015). What is a resistance gene? Ranking risk in resistomes. *Nat Rev Microbiol* **13**: 116–123.

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ *et al.* (2013). The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* **57**: 3348–3357.

Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD. (2012). Hypervariable loci in the human gut virome. *Proc Natl Acad Sci USA* **109**: 3962–3966.

Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD *et al.* (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* **21**: 1616–1625.

Modi SR, Lee HH, Spina CS, Collins JJ. (2013). Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**: 219–222.

Moore AM, Patel S, Forsberg KJ, Wang B, Bentley G, Razia Y *et al.* (2013). Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLoS One* **8**: e78822.

Mori T, Mizuta S, Suenaga H, Miyazaki K. (2008). Metagenomic screening for bleomycin resistance genes. *Appl Environ Microbiol* **74**: 6803–6805.

Muniesa M, Colomer-Lluch M, Jofre J. (2013). Could bacteriophages transfer antibiotic resistance genes from environmental bacteria to human-body associated bacterial populations? *Mobile Genet Elements* **3**: e25847.

Noguchi H, Taniguchi T, Itoh T. (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* **15**: 387–396.

Novick RP, Christie GE, Penades JR. (2010). The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol* **8**: 541–551.

Ogilvie LA, Bowler LD, Caplin J, Dedi C, Diston D, Cheek E et al. (2013). Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences. *Nat Commun* **4**: 2420.

Parsley LC, Consuegra EJ, Kakirde KS, Land AM, Harper Jr WF, Liles MR. (2010). Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Appl Environ Microbiol* **76**: 3753–3757.

Penders J, Stobberingh EE, Savelkoul PH, Wolffs PF. (2013). The human microbiome as a reservoir of antimicrobial resistance. *Front Microbiol* **4**: 87.

Pruitt KD, Tatusova T, Maglott DR. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**: D61–D65.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Quiros P, Colomer-Lluch M, Martinez-Castillo A, Miro E, Argente M, Jofre J et al. (2014). Antibiotic resistance genes in the bacteriophage DNA fraction of human fecal samples. *Antimicrob Agents Chemother* **58**: 606–609.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F et al. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**: 334–338.

Riesenfeld CS, Goodman RM, Handelsman J. (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* **6**: 981–989.

Rolain JM, Fancello L, Desnues C, Raoult D. (2011). Bacteriophages as vehicles of the resistome in cystic fibrosis. *J Antimicrob Chemother* **66**: 2444–2447.

Roux S, Enault F, Hurwitz BL, Sullivan MB. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**: e985.

Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R et al. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* **3**: e03125.

Roux S, Krupovic M, Debroas D, Forterre P, Enault F. (2013). Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol* **3**: 130160.

Santagati M, Iannelli F, Cascone C, Campanile F, Oggioni MR, Stefani S et al. (2003). The novel conjugative transposon tn1207.3 carries the macrolide efflux gene mef(A) in *Streptococcus pyogenes*. *Microb Drug Resist* **9**: 243–247.

Schuch R, Fischetti VA. (2006). Detailed genomic analysis of the Wbeta and gamma phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. *J Bacteriol* **188**: 3037–3051.

Sommer MO, Dantas G, Church GM. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**: 1128–1131.

Torres-Cortes G, Millan V, Ramirez-Saad HC, Nisa-Martinez R, Toro N, Martinez-Abarca F. (2011). Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. *Environ Microbiol* **13**: 1101–1114.

van Schaik W. (2015). The human gut resistome. *Philos Trans R Soc Lond Ser B* **370**: 20140087.

Volkova VV, Lu Z, Besser T, Grohn YT. (2014). Modeling the infection dynamics of bacteriophages in enteric *Escherichia coli*: estimating the contribution of transduction to antimicrobial gene spread. *Appl Environ Microbiol* **80**: 4350–4362.

Wichmann F, Udikovic-Kolic N, Andrew S, Handelsman J. (2014). Diverse antibiotic resistance genes in dairy cow manure. *MBio* **5**: e01017.

Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J et al. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* **4**: e7370.

Wipf JR, Schwendener S, Perreten V. (2014). The novel macrolide-Lincosamide-Streptogramin B resistance gene erm(44) is associated with a prophage in *Staphylococcus xylosus*. *Antimicrob Agents Chemother* **58**: 6133–6138.

World Health Organization (2014). *Antimicrobial Resistance: Global Report on Surveillance*. WHO Press: Geneva, Switzerland.

Wozniak RA, Waldor MK. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* **8**: 552–563.

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O et al. (2012). Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* **67**: 2640–2644.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)