HAL open science

Analysis of viromes and microbiomes from pig fecal sample reveals that phages and prophages rarely carry antibiotic resistance genes

Maud Billaud, Quentin Lamy-Besnier, Julien Lossouarn, Elisabeth Moncaut, Moira Dion, Sylvain S. Moineau, Fatoumata Traoré, Emmanuelle Le Chatelier, Catherine Denis, Jordi Estellé, et al.

To cite this version:

Maud Billaud, Quentin Lamy-Besnier, Julien Lossouarn, Elisabeth Moncaut, Moira Dion, et al.. Analysis of viromes and microbiomes from pig fecal sample reveals that phages and prophages rarely carry antibiotic resistance genes. The International Society of Microbiologial Ecology Journal, 2021, 1 (55), 10.1038/s43705-021-00054-8 . hal-03299373

HAL Id: hal-03299373

https://agroparistech.hal.science/hal-03299373v1

Submitted on 9 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

ISME

www.nature.com/ismecomms

ARTICLE OPEN

Analysis of viromes and microbiomes from pig fecal samples reveals that phages and prophages rarely carry antibiotic resistance genes

Maud Billaud[1,2], Quentin Lamy-Besnier[1], Julien Lossouarn[1], Elisabeth Moncaut[1], Moira B. Dion[3,4], Sylvain Moineau[3,4,5], Fatoumata Traoré[6], Emmanuelle Le Chatelier[6], Catherine Denis[7], Jordi Estelle[7], Caroline Achard[8], Olivier Zemb[8] and Marie-Agnès Petit[1]✉

© The Author(s) 2021

Understanding the transmission of antibiotic resistance genes (ARGs) is critical for human health. For this, it is necessary to identify which type of mobile genetic elements is able to spread them from animal reservoirs into human pathogens. Previous research suggests that in pig feces, ARGs may be encoded by bacteriophages. However, convincing proof for phage-encoded ARGs in pig viromes is still lacking, because of bacterial DNA contaminating issues. We collected 14 pig fecal samples and performed deep sequencing on both highly purified viral fractions and total microbiota, in order to investigate phage and prophage-encoded ARGs. We show that ARGs are absent from the genomes of active, virion-forming phages (below 0.02% of viral contigs from viromes), but present in three prophages, representing 0.02% of the viral contigs identified in the microbial dataset. However, the corresponding phages were not detected in the viromes, and their genetic maps suggest they might be defective. We conclude that among pig fecal samples, phages and prophages rarely carry ARG. Furthermore, our dataset allows for the first time a comprehensive view of the interplay between prophages and viral particles, and uncovers two large clades, inoviruses and Oengus-like phages.

ISME Communications (2021)1:55 ; https://doi.org/10.1038/s43705-021-00054-8

INTRODUCTION

The spread of antibiotic resistance genes (ARGs) is a major health concern, and powerful sequencing techniques are now making it possible to examine the contribution of mobile genetic elements to their spread. The main mobile vectors currently recognized as responsible for spreading ARGs are conjugative plasmids and integrative conjugative elements (ICEs, also known as conjugative transposons) [1]. In both cases, ARG spread occurs through conjugation, a process that involves the formation of a conjugative pilus, the contact between a donor and a recipient bacterium, and the transfer of the genetic element. ARG are frequently found in pig fecal samples [2], and their associated mobile genetic elements start to be well documented. For instance, analyses were conducted on a set of metagenomics contigs that were cloned on bacterial artificial chromosomes. Eleven of them were conferring resistance to tetracycline to *Escherichia coli*, and nine of the *tet* genes were present on mobile genetic elements (plasmids or ICEs) [3].

Interestingly, bacteriophages, another category of mobile genetic elements, apparently rarely encode ARGs [4]. Phages can usually be divided into two large groups based on their lifestyle. Virulent phages inject their genetic material into bacteria, replicate and lyse their host cells, thereby releasing several new virions at the end of this lytic cycle. Temperate phages are able to alternate between the lytic cycle and a dormant stage where they maintain their genetic material within the bacterial genome as prophages (following an integration step or as plasmids). While dormant, prophages may still express a few of their genes, including some which are beneficial to their host, such as *bor*, which is involved in resistance to serum complement killing [5], and the iron transporter *sitABCD* genes [6]. In theory, a temperate phage and its expression profile in a prophage state would offer an efficient process for the dissemination of ARGs. So, why have they so rarely been detected?

Initially, their absence was noted on the basis of single phage biology methods, such as phage culturing and sequencing, and therefore the investigations lacked breadth. On the other hand, qPCR studies have reported the presence of ARGs in virus-enriched environmental samples from natural waters [7], wastewater plants [8], or human/animal fecal samples [9, 10]. However, virome preparations are often highly contaminated by bacterial DNA [11, 12], making it difficult to discern whether the ARG originates from viral or bacterial DNA using qPCR alone. A distinct process through which phages can mobilize ARG is generalized

[1]Université Paris- Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas, France. [2]Pherecydes Pharma 22 Bd Benoni Goullin, Nantes, France. [3]Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Quebec City, QC G1V 0A6, Canada. [4]Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Quebec City, QC G1V 0A6, Canada. [5]Felix D'Hérelle Reference Center for Bacterial Viruses, Université Laval, Quebec City, QC G1V 0A6, Canada. [6]Université Paris-Saclay, INRAE, MGP, Jouy en Josas, France. [7]Université Paris-Saclay, INRAE, GABI, Jouy-en-Josas, France. [8]GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326 Castanet-Tolosan, France. ✉email: Marie-agnes.petit@inrae.fr

Received: 25 May 2021 Revised: 6 September 2021 Accepted: 13 September 2021
Published online: 25 October 2021

transduction. During this process, instead of packaging their own genome in their capsid, some phage particles will package bacterial DNA (sometimes at high frequencies [13]). This process is far from universal in phages, and the fraction of phages capable of performing such task is unknown.

Shotgun metagenomics studies of enriched viral fractions (viromes) and improved programs for assembling short reads into quality contigs can now help determine whether ARGs are found in phage genomes. ARGs were recently investigated in the virome from wastewater samples, and they were found to be scarce compared to other genetic elements [14, 15]. However, several reports have raised a concern with pig fecal samples, which appeared to be rich in phage encoded ARG [9, 16, 17].

Another approach for investigating the ARG content of phages is to study prophage regions of completely assembled bacterial genomes. Recent reports on pathogenic bacterial species such as *Acinetobacter baumanii* or *Streptococcus suis* found prophage regions with ARGs [18–20]. In these in silico studies, the difficulty lies in the ability to discern whether such prophages are defective or still functional and able to, for instance, complete lytic cycles, form viral particles, and therefore spread ARG in bacterial populations.

Here, with respect to the ARG content in pig fecal viromes, we elected to extract and analyse new samples, because the presence of bacterial DNA contamination was either clearly mentioned in previous reports [17], or it was poorly documented, by the absence of detection of a DNA band after PCR with 16 S universal primers. Moreover, we sequenced in parallel the purified virome and the total microbiota of the same samples to have access to the prophage content and be able to test whether they were active (i.e., present also in the virome fraction). Our study presents the results of these analyses performed using 14 pig fecal samples collected from farms with varying degrees of antibiotic usage and reports the presence and absence of ARG among microbiota and viral particles, respectively.

MATERIAL AND METHODS

Sample preparation sequencing and assembly

Starting from 14 pig or piglet fecal samples, parallel DNA extractions were conducted for the purified viral fraction, named VP, and the total microbiota, named P, using standard procedures. All details are available in the Supplemental Information. Raw sequencing data and SPAdes assembly outputs (>2 kb, contigs for VP samples, scaffolds for P,134 before the clustering step at 95% identity) are accessible through the NCBI, Bioproject PRJNA741980.

Analysis and labeling of the bona fide viral contigs within the VP dataset

The analysis of all VP contigs of a size above 2 kb was performed in two steps. From the 7755 VP contigs, the 2480 with the highest RPKM (above 100) were first analyzed. Several programs and methods were applied to identify non viral contigs (bacteria and plasmids, see Supplementary Fig. 1). Virsorter [21] recognized 23% (568 contigs) of these contigs as viral (all categories were accepted, cat1 and cat2 were vastly dominant). The Inovirus detector [22] was then run on the remaining contigs and 143 inoviruses (putative or confirmed, most were circular). To add contigs corresponding to phages recently added into public databases, Mega-BLAST was run against the nt database of NCBI (April 2020, E-val below 10^{-16}). As bacterial hits could also correspond to prophages, all BLASTn hits, as well as the remaining unclassified contigs were taken for a search centered on proteins, making use of PHROG, a database of phage proteins [23] (see https://phrogs.lmge.uca.fr). For this, an HMMER hmmscan [24] of proteins from all remaining contigs was run against PHROG profiles, and all contigs with at least 5 orfs and 35% of them homologous to a PHROG family (E-value below 10^{-12}) were counted as phages. In total, this BLASTn + PHROG step allowed for the recognition of 966 additional contigs from the following origins: 654 phage, 232 eukaryotic viral (collectively designated below as "viruses" for short, while "phages" is kept for bacterial viruses), 72 bacterial and 8 plasmid fragments. Next, the VIBRANT classifier [25] was used on the remaining uncharacterized contigs, adding 325 phages and 29 viruses. In the end, 1951 contigs were found to be of viral origin, 72 bacterial, 8 plasmidic, and only 449 contigs of the 2480 (18%) remained uncharacterized.

In a parallel study on infant gut viromes, various tools dedicated to the recognition of viral contigs were compared, and VIBRANT was among the best [26]. The 1951 viral contigs detected in the top 2480 contigs with high abundance (as described above, see also Supplementary Fig. 1) were therefore verified using VIBRANT. A high degree of overlap was observed between the pre-VIBRANT sorting and VIBRANT detection for all *Caudovirales* and *Petitvirales* genomes (92.5% and 99% were detected by VIBRANT, respectively). VIBRANT was also successful at detecting eukaryotic viruses, as 79% of them were detected. VIBRANT was less successful at detecting inoviruses (35% detected).

In a second step, we therefore used VIBRANT and the Inovirus detector on the remaining 5275 low-abundance contigs and obtained an additional 3654 viral and 201 inovirus contigs. In total, 5806 VP contigs (75%) were recognized as viral (and 77.3% of the contigs treated by VIBRANT, which filters for contigs with ≥2 kb and 4 ORFs). Of these 5806 viral contigs, 44% were considered complete or high quality by CheckV [27].

Other bio-informatic methods

ARG search, VP viral contig classification, host prediction for the viral contigs, as well as the procedure to partition active phage versus inactive prophage contigs within the microbial P dataset, are all described in the Supplementary Information.

RESULTS

Among most abundant virome contigs, only 18% correspond to unclassified contigs

Viral fractions of the 14 pig fecal samples were purified, separated on gradients, their ss- and ds-DNA extracted, amplified by multiple displacement amplification (MDA), and sequenced, as described in Supplementary Information. Reads were assembled into contigs, those of a size above 2 kb were retained, and clustered at the species level (>95% nt identity). Viral reads were mapped back onto the 7755 contigs obtained, and relative abundance of these contigs were computed to generate the so-called "VP" (for "virome-de-porc") matrix.

The 2480 contigs with total RPKM values above 100 were then selected for in-depth viral sequence identification (see "Methods" and Supplementary Fig. 2A). Overall, 82% of all abundant contigs (2031) could be classified either as viral or, in a few cases, bacterial or plasmidic, therefore reducing unclassified contigs (the so-called "viral dark matter") to only 18% of the sub-set of abundant VP contigs (Supplementary Fig. 2B). Among these classified contigs, 83% were phages, and 13% eukaryotic viruses (Supplementary Fig. 2C), and the remaining contigs were bacterial or plasmid DNA. Most phages belonged to the *Petitvirales* order (52% of phages), which is consistent with previous research [28], and the most common viruses were from the *Smacoviridae* family (31 % of viruses, Supplementary Fig. 2D).

An overview of the abundance matrix of these 2480 most abundant contigs, grouped by clades, is shown Fig. 1 (see full matrix in Supplementary Table 1). It should be noted that ssDNA viruses abundances cannot be compared to dsDNA ones (*Caudovirales*), due to their over-amplification during the MDA step [29, 30]. Still, *Petitvirales* represented 30% of total abundance, followed by non-bacterial viruses (21%) and *Tubulavirales* (7%). *Tubulavirales* contigs (described in detail below) were grouped according to their type of morphogenesis protein, encoded by *gpI*. We noted a gradient of abundance in *Tubulavirales*, with adults containing most of them (samples are organized in the matrix following this gradient). Among *Caudovirales*, some of the contigs were grouped into higher order taxons (see below). In these cases, the contigs were labeled with the name of a similar reference phage genome when available, and were otherwise assigned a vCONTACT2 [31] cluster number (followed by the size of the circular or largest contig). Still, most signals came from

	Adult pigs									Piglets					
	17	9	13	15	3	5	16	11	6	4	8	2	7	12	%
Propionibacterium B5 gpl	47 698	7 236	5 645	5 048	5 760	3 586	4 736	1 455	1 950	2 867	2 583	629			7
Acholeplasma MV-L1 gpl	7 732	255		852				344	756				299		
Ralstonia RSM1 gpl		968	283			826		708		9 574		138			
Other *Tubulavirales*	226	788	1 095	1 519	1 904	1 578		3 280	2 155		2 264				
Gokushovirinae	32 158	7 434	5 914	13 292	5 783	4 844	13 628	1 249	9 902	1 497	4 485		323	4 581	30
Alphatrevirus									1 137						
Other *Petitvirales*	15 140	21 885	15 620	54 182	12 192	91 843	56 344	17 350	38 736	39 194	22 729	4 223	8 399	66 471	
Smacoviridae	17 770			10 204		428	3 524	1 237	5 703	3 709	5 946		2 623	7 734	21
Circoviridae			819	5 736	807	4 464	775	11 510	2 610	2 418	2 623				
Gemycircularvirus				997							2 499		205		
Genomoviridae				2 337							792				
CRESS	525	3 749	4 091	1 000	13 726	36 216	2 129	5 162	12 577	1 413		16 019	397	16 056	
Other non-bacterial viruses	4 592	3 133	5 975	6 288	1 680	7 855	30 853	36 067	85 695	6 148	2 602	2 235	902	459	
cluster734_21kb	103														14
cluster62_125kb	137														
cluster484_62kb			154												
cluster448_24kb								311							
cluster_423_48kb								1 057							
cluster374_104kb										168					
cluster725_20kb											127				
cluster_r1t			317				307			113	204				
cluster535_29kb											274				
cluster23_37kb											321	111			
cluster737_65kb												158			
cluster444_37kb												210			
cluster 100_29kb							113			443		226			
cluster643_24kb												437			
cluster_Oengus		90		62	460			52				686		99	
cluster_T1															
cluster_440_30kb													473		
cluster361_37kb													530		
cluster657_26kb													10 627		
cluster119_75kb															
cluster394_40kb															
cluster_IME207														872	
cluster_CrAssphage01									1 023					1 087	
Other *Caudovirales*	15946	16708	11373	19869	3360	6918	22230	20309	8588	16261	18952	28474	25965	36019	
Bacteroidetes			96									13 487		260	1
Firmicutes		709	253	179				974			1 337	768	268		
Proteobacteria													1 654		
Plasmids		2 263	153		374	772		770	243					101	0
Unclassified contigs	39868	18 142	62 847	13 538	138 247	55948	42 011	29 526	21 235	30 790	12 308	27 463	25 563	5 898	27

RPKM: 30 000, 3000, 300, 30

Fig. 1 Abundance matrix of the 2480 most abundant virome contigs, grouped into clades. Abundances are expressed as RPKM (gray scale on the right). The last column shows the percentage of total RPKM in each contig category: *Tubulavirales*, *Petitvirales*, non-bacterial viruses, *Caudovirales*, Bacteria, Plasmids, and Unclassified.

unclustered contigs. We noted that piglets had significantly higher abundances of *Caudovirales* compared to adults ($P = 0.0036$, Student t-test), and fewer ssDNA eukaryotic viruses (not significant, $P = 0.07$). The remaining lower abundance contigs were also classified,and 72% were found to be viral. We conclude that the majority of contigs from these highly purified viromes were indeed viral in nature, with a substantial proportion being ssDNA viruses.

The few antibiotic resistance genes found in viromes are not encoded by phages

ARG were then searched among the complete set of 8090 VP contigs of a size above 2 kb (before clustering, and regardless of their abundance) in viromes. Contigs that were not predicted to be viral were also included (Table 1, see Supplementary Information). The search against ResFam profiles [32] revealed 8 VP contigs encoding 10 putative ARG, 5 of which were also retrieved with a BLASTp search against MUSTARD [33]. The 5 missing ARG encoded ABC transporters, a category which is simply not included in MUSTARD. Finally, a more stringent search of confirmed ARG with Resfinder [34] reported only 3 of the 10 initially uncovered ones (Table 1). Interestingly, in one of the 8 ARG-positive contigs (VP7_NODE302), Resfinder identified a *tet(40)* gene undetected with the two previous tools.

Detailed inspection of the 8 contigs revealed that none of them originated from phages, despite a positive VIBRANT [25] prediction for one of them (contig VP12_NODE394, last line in Table 1). This phage candidate, however, turned out to be 99.8% identical to plasmid pCd13 of *Clostridioides difficile* (Accession MH229772). In addition, this 5.4 kb contig was circular and contained a 1.5 kb deletion compared to pCd13. We therefore conclude that it is rather a plasmid than a phage. Of note, the second circular contig, VP8_NODE804, was 100% identical to synthetic plasmid pET11a with its *Nde*I restriction site mutated. We think this laboratory vector comes from the control DNA of the Genomiphi kit used to prepare DNA prior sequencing, suggesting that the sequencing depth was sufficient to detect even minor contaminants. Homologs for all remaining, non phage contigs were searched with BLASTn against NCBI nt database (May 2019) and most of them had a taxonomic association that is typical of intestinal bacterial species (Table 1).

We noted that the six ARG-positive contigs of chromosomal origin came from the viromes with less viral enrichment (Supplementary Table 2). Given the lower level of purity, it seems likely that these fragments came from remaining free bacterial DNA rather than viral particles. In contrast, the plasmid contig homologous to pCd13 from *Clostridioides difficile* came from a highly purified virome, making bacterial contamination unlikely. This plasmid may have been carried over by generalized transduction.

Taken altogether, the few ARG-positive contigs we found in highly purified pig viromes were not phage-encoded, even in samples originating from farms with heavy antibiotic use, such as sample 16. Compared to the 5806 contigs positively recognized as viral, this places the frequency of ARG-encoded viral contigs below 0.02%. The maximal ARG gene ratio over total viral orfs was below 1.2 10^{-5}, or below 0.015 per Mb of assembled viral contigs.

Table 1. The eight ARG-positive contigs assembled from viromes.

NODE	ARG name ResFam	ResFam (E-val)	Resfinder	CONTIG properties			
				phage VIBRANT	best BLASTn hit	size, nt (©=circular)	RPKM in viromes
VP8_NODE183	ABC_efflux	2.9e–97	NO	No	*Parabacter. distasonis*	16 427	15
VP2_NODE298	MexH	2.2e–42	NO	No	Alistipes	13 316	7
	RND_efflux	0	NO				
VP7_NODE302	tet(40)	NO	YES	No	*Eubacterium maltosivorans*	7 945	11.9
	ANT9	1.3e–80	NO				
VP7_NODE1	ABC_efflux	7.9e–89	NO	No	Clostridiales	301 151	35.9
VP7_NODE2	ABC_efflux	5.1e–84	NO	No	–	230 004	37.9
VP7_NODE3	tet(W)	0	YES	No	–	226 564	33.4
VP8_NODE804	TEM	1.8e–205	YES	No	*E. coli* cloning vector pET11a	5 676 ©	32.1
	tet_MFS_efflux	1.9e–80	NO				
VP12_NODE394	APH3'	3.2e–113	YES	Yes	*C. difficile* plasmid pCd13	5 473 ©	57.7

Pig viromes are rich in inoviruses and Oengus-like phages

Following the examination of ARGs, we explored the main characteristics of these virome assemblies. Among the rich viral content, two main clades emerged. First, 344 VP contigs (272 of them circular) were found to correspond to inoviruses—filamentous phages performing chronic bacterial infections [35]. Taxonomically, the best characterized family of inoviruses, *Inoviridae*, is now included into a *Tubulavirales* order and the richness of this order has recently been put to light [36]. Most of inovirus contigs (181 of the 344, 52%) possessed the *gpI* type of Propionibacterium phage B5 (Fig. 1), the only known filamentous phage infecting a Gram-positive host [37]. Adult pig viromes were particularly rich in this group, compared to piglets. The 272 complete circular genomes were taken for a deeper analysis (Fig. 2, details about all circular genomes are reported in Suppl. Table 3). Their clustering using vCONTACT2 [31] indicated that they were very diverse, forming 50 clusters of a quality above 0.5, and leaving aside 93 singletons. None of the clusters included any of the 10 reference inoviruses available, not even phage B5. The 118 circular and clustered genomes with a B5 *gpI* are highlighted in yellow on the vCONTACT2 map (Fig. 2a).

We tried to predict the bacterial hosts of these inoviruses with circular genomes. A search of nearly-identical CRISPR spacers led to the prediction of four possible *Firmicute* hosts: two *Roseburia*, a *Ruminococcus* and a *Lachnoclostridium*. A complementary search using WiSH [38] or BLASTn added 67 predictions, most often in the *Clostridium* genus (Fig. 2a, right). Consistent with these predictions, a filamentous phage infecting *Clostridium acetobutylicum* NCIB6444 was isolated (but its genome not sequenced) 30 years ago [39].

Pairwise tBLASTx genome comparisons between B5 and some of the VP genomes with a similar *gpI* revealed similar genetic organizations, although the B5-encoded *gpI* was too distantly related to those of pig genomes to appear in the results (Fig. 2b). Finally, we noticed that electron transmission microscopy images of the VP17 virome, which contained the most inoviruses, revealed some long filaments that are typical of inoviruses (Fig. 2c). We conclude that novel, *Firmicute*-infecting inoviruses are present in pig fecal samples.

The second notable characteristic of these pig viromes was identified by performing a vCONTACT2 clustering analysis [31] of the 2991 *Caudovirales* VP contigs, together with 2214 reference phage genomes (see Supplementary Information). Among the seven most populated clusters (listed in Suppl. Table 4), there was a particularly large one of 98 elements, 94 of which originated from pig virome contigs. Of note, four reference phage genomes were connected to them. Three of these four phages are virulent and infect *Actinobacteria*, including the Rhodococcus phages ReqiPoco6 and ReqiPepy6 [40], and Arthrobacter phage Mudcat (accession NC_031224). The fourth is the temperate phage Oengus, which infects the *Firmicute Faecalibacterium prausnitzii* [41].

The 17 largest genomes of this cluster (45–62 kb) were then compared to already classified phage genomes with ViPTree [42] (Fig. 3a). The ViPTree branch lengths were used to delimit family boundaries (branch lengths extending into the 0.01–0.05 region), and subfamily boundaries (between 0.1 and 0.5). Members of the large vCONTACT2 cluster were connected by deep branches (0.1), suggesting the clade includes more than one viral genus. Furthermore, phage Oengus was closest to all but one of these VP contigs. Surprisingly, among the 38 nearly complete genomes (>40 kb), only three contained an integrase, suggesting a clade of mostly virulent lifestyle. Alignments of Oengus with two complete genomes of this clade showed that only a few genes were shared, confirming further that this cluster represents a taxon beyond the genus level (Fig. 3b). Finally, most frequently predicted hosts belonged to the *Clostridiaceae* family (Fig. 3c). In addition, for two of them, a few matching CRISPR spacers were found in strains of *Anaeromassilibacillus*, a genus that belongs to the same *Oscillospiraceae* family as *F. prausnitzii*, the host of phage Oengus. We recently proposed the family name Sisseviridae for this phage clade (extending up to Lactococcus phage 1706, Fig. 3a), which currently comprises 72 genera and 254 species [27].

Pig microbiota contains a high abundance of viruses in addition to six main bacterial phyla

We next investigated whether ARGs could be detected on prophages that are embedded in bacterial genomes. For this, we extracted, sequenced and assembled the DNA from the global fecal samples, following a distinct protocol for DNA extraction (see Supplementary Information, and Supplementary Fig. 2a for the overall analysis pipeline). After assembly and dereplication, a final binning step was performed to aggregate contigs from the same microbial species, using the VAMB tool [43]. To estimate the main phyla present in the 14 pig fecal samples, a taxonomic affiliation with Kaiju was performed on the 220,000 contigs of a size above 2 kb.

In addition, we searched systematically for viral contigs among this microbial contig collection, using a combination of 3 criteria: (i) positive viral result according to VIBRANT [25], (ii) homologous to a viral VP contig according to BLASTn results, (iii) viral affiliation by Kaiju (see details in Supplementary Information). Since huge phages are often found in pig microbiota, we searched for them in our dataset, and found one distant homolog in the P set, and three additional among the virome VP contigs (Section 2 of Supplementary Information). Of the 220 000 contigs, 16 940

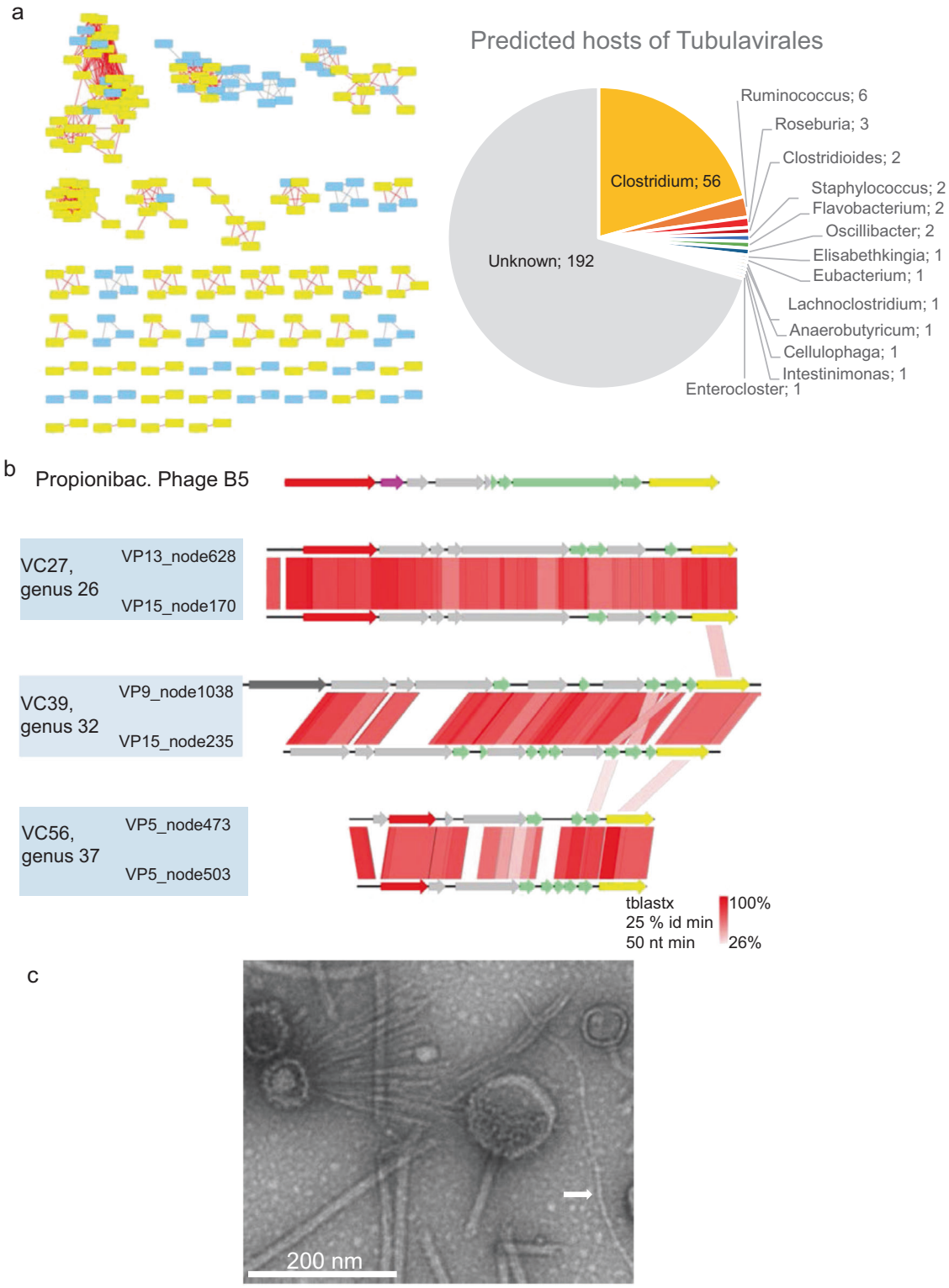

Fig. 2 **Properties of circular pig inoviruses. a**, left VContact2 map of all complete genomes that could be clustered. Those encoding a *gpI* homologous to the one of Propionibacterium phage B5 are highlighted in yellow; **a**, right, host predicted for all circular genomes. **b** Genome alignments (tBLASTx) and Easyfig map of several inoviruses encoding a B5-like *gpI* gene. Yellow: gene encoding morphogenesis protein I, red: *rep* gene for replication initiation, green: structural genes (with transmembrane domains), pink: *ssb* (single-strand DNA binding) gene. Gray: gene of unknown function. Dark gray: putative phosphoadenosine phosphosulfate reductase. **c** Transmission electron microscope image of virome VP17, the white arrows point to a thin filament that may correspond to an inovirus (length> 420 nm, width, 8 nm).

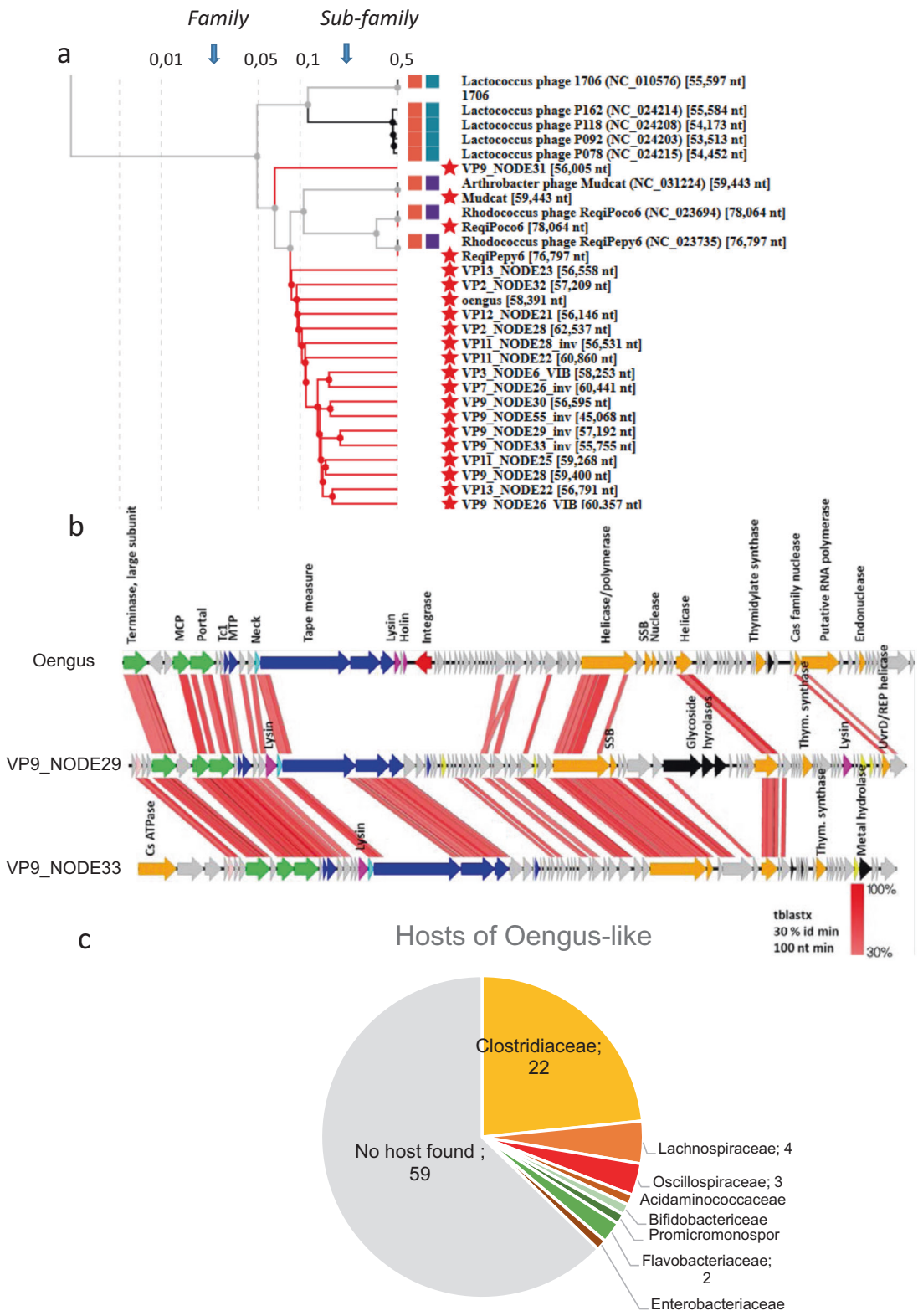

Fig. 3 **Characteristics of the Oengus cluster. a** Tree representation of the 17 largest Oengus-like contigs together with close reference phage neighbors in the global VipTree [42], which builds a distance based on shared genes, starting from nt sequences and tBLASTx comparisons. Red stars indicate all genomes that were added to the reference VipTree. **b** Whole genome alignments (Easyfig map, with tBLASTx option) comparing two relatively close VP nodes (**a**) and Oengus. MCP: major capsid protein, Tc1: Tail completion 1, SSB: single-strand binding protein. **c** Predictions of infecting hosts for the 94 contigs of the Oengus-like cluster.

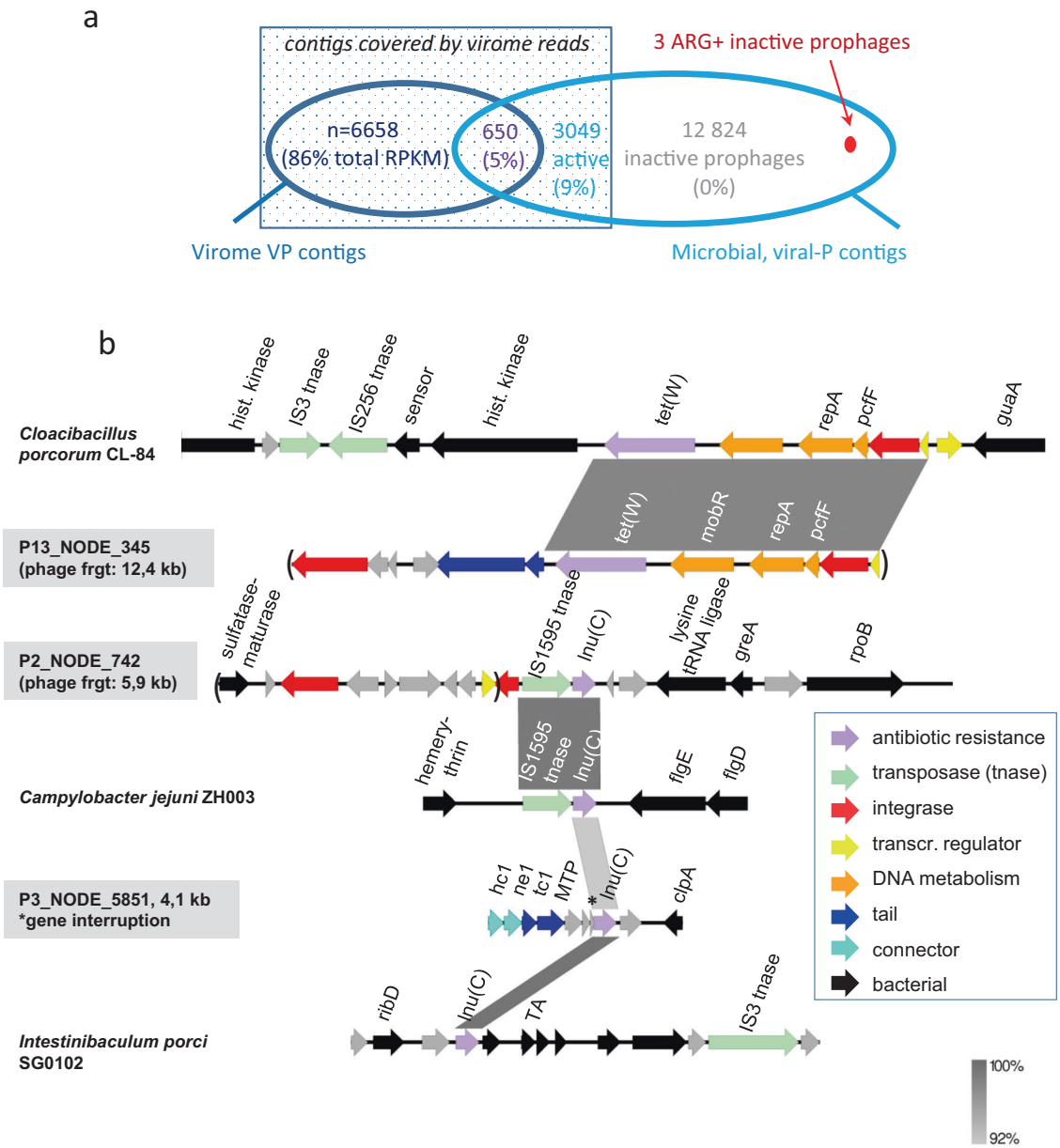

Fig. 4 Overview of the ARG found in pig phages and prophages contigs. a Overlap between viral contigs assembled from virome fractions (VP) and those recognized as viral within microbiota assemblies (viral-P). Among 23 181 dereplicated viral contigs (see Supplementary Information for details), 650 were present in both contig sets, while 6658 were present exclusively in viromes. No ARG was found in these two first subsets. Among the viral-P contigs, we distinguish those "covered" by viromes reads, and therefore considered active phages (square of dark blue shade, abundance in viromes indicated as % of RPKM among the 3 subsets of contigs covered by reads), from those that correspond to dormant prophages The 3 ARG positive viral contigs (indicated in red) are among dormant prophages. **b** Genetic maps of these three prophage regions. Easyfig [53] maps (BLASTn) of shared regions with other mobile elements are shown. Phage hallmark genes (tail and connector) are colored in blue. GNAT stands for GNAT family N-acetyltransferase, TA for toxin-antitoxin system, GGDEF for GGDEF domain-containing protein.

(7.7 %) were deemed of viral origin (labeled viral-P contigs below). We reasoned however, that some of these contigs might correspond to uninduced prophages and some may even be defective prophages. We therefore used the collection of virome reads to determine which of the viral-P contigs were forming viral particles (see Supplementary Information). Among these 16 940 viral-P contigs, 23.7% were covered by a significant amount of virome reads and classified as "active phages" (Fig. 4 and Supplementary Fig. 2).

Figure 5 reports a synthetic representation of the main phyla detected in the pig microbiota, among OTU with abundance above 10 RPKM across the 14 samples (see details in Supplementary Table 5). In this matrix, active phages were binned and counted separately (798 contigs, binned into 547 "active phages"). The remaining viral contigs, for which no proof of activity was obtained, were either kept under their bacterial host taxonomy when available (329 contigs), or placed under "inactive prophages" otherwise.

In most samples (13/14), the three main phyla were *Firmicutes* (39%), *Bacteroidetes* (21%) and surprisingly, active phages (16%). The next microbial core phyla with abundances above 1% were *Proteobacteria*, *Spirochaetes*, *Actinobacteria* and Archaea. Aside from the active phage category, previous reports on pig microbiota described similar phyla distributions [2, 44, 45]. The

	N	Adults									Piglets						
	OTUs	P15	P3	P13	P5	P11	P9	P6	P16	P17	P12	P8	P7	P4	P2	av	
FIRMICUTES	2559	1 968	2 446	1 882	2 043	2 722	1 976	2 236	4 123	6 752	4 338	4 059	3 197	3 953	6 231	39%	
BACTEROIDETES	1304	1 707	2 215	1 117	1 333	1 609	1 888	1 514	2 803	1 514	1 848	1 980	683	1 927	1 601	21%	15 000
ACTIVE PHAGES	547	956	1 219	1 641	658	547	1 955	2 044	5 287	3 209	198	956	12 551	917	695	16%	5 000
UNCLASSIFIED	429	630	399	1 197	960	729	583	987	837	2 282	378	580	4 248	780	491	9%	1 000
INACTIVE PROPHAGES	247	1 094	107	1 125	1 954	196	377	586	731	348	422	71	2 100	108	92	7%	200
PROTEOBACTERIA	221	262	185	433	432	354	190	329	575	389	245	214	813	289	213	4%	
SPIROCHAETES	187	490	256	302	143	239	327	206	39	49	45	128	557	147	146	3%	
ACTINOBACTERIA	83	115	58	83	57	79	88	117	112	139	57	107	143	74	148	1%	
ARCHAEA	59	106	38	33	34	76	29	34	125	186	118	24	151	60	45	1%	
MOLLICUTES	15	7	10	7	8	13	18	177	4	20	6	7	57	7	11	0%	
SUMS		7 335	6 934	7 821	7 621	6 566	7 431	8 230	14 635	14 887	7 655	8 126	24 499	8 262	9 672	100%	

Fig. 5 Main pig microbiota phyla. Abundances are expressed as RPKM (gray scale on the right). The average column does not include sample P7.

Mollicute class was added in the matrix to highlight sample P6, which contained a particularly high level of a single *Mycoplasma* OTU, suggesting pig 6 was sick.

The last sample (piglet P7) had a distinct profile, as virus reads represented 51% of the total. More precisely, a bin of two contigs (12.6 kb total) was responsible for 52% of all viral signal. Upon closer inspection, these contigs had a typical *Salasmaviridae/Sarlesviridae* organization (Supplementary Fig. 3), but its host could not be predicted. Interestingly, the P7 sample also contained a very large 14 Mb OTU (bin of 401 contigs). A BLASTn search against the nt database of the NCBI indicated that its closest (77% identity) and most frequent (18% of contigs from this bin) parent was *Blastocystis hominis* (Eukaryote, Stramenopiles). We also noted that in this sample, richness and abundance in *Bacteroides* and *Prevotella* species was low. This suggests that there is intense predation on bacteria in the microbiota from sample P7.

A small fraction of ARGs are found in prophages in pig microbiota

An ARG search among the 271 163 P contigs (before dereplication) retrieved 1764 ARG with ResFam and 279 with Resfinder (see Supplementary Table 6 for overall counts, and Suppl. Table 7 for the details of each contig). The large increase in ARG with ResFam is due to the inclusion of ABC transporters, which probably include false positives. We therefore concentrated on ResFinder numbers. The ResFinder counts correspond globally to 1.8 (+/−0.5) 10^{-4} ARG gene per total genes, and 0.18 (±0.05) ARG gene per Mb.

To further investigate the samples with the highest abundances of ARG-positive bacterial contigs and the category of antibiotic resistance that was prevalent, we focused on the abundance matrix of the 168 contigs that harbored the 279 ARG genes (Supplementary Table 17). Overall, a four-fold difference was observed between the most and least ARG-abundant microbiota samples (Supplementary Fig. 4a). Piglets also had higher prevalence of ARGs compared to adults (not significant, Wilcoxon-Mann Whitney test, $p = 0.109$). Contigs encoding tetracycline resistance genes were by far the most abundant, followed by aminoglycoside and macrolide resistance genes (Supplementary Fig. 4b–e), consistent with previous reports [2]. Among the *tetR*-encoding contigs, three were particularly abundant, in all samples: the most abundant one encodes a *tet(*W) gene (in P4_NODE_19877) which apparently belongs to an integrative conjugative element (ICE, see Supplementary Information). The two next most abundant contigs (*tet (40)* genes in P12_NODE_17183, 2004 bp, and P8_NODE_13797, 2606 bp) were partially overlapping and matched a plasmid reported from a study on the tetracycline resistome of bio pig farms [3]. Clearly, and as concluded also by Kazimierczak et al. [3], even though tetracycline which was used as a growth factor, is no longer used as such in Europe, the previously selected resistance genes still remain in pig microbiota. Some samples, however, and particularly for the pig from farm 6, were markedly ARG-poor. What parameter of the farm settings could explain this good record is not known at present. We conclude that globally, the microbial ARG content of these pig fecal samples was similar to that previously reported in the literature [2, 46].

We next sought whether some of the ARG-positive contigs were viral. Three were indeed viral (representing 0.02% of the 16 940 viral-P contigs), but belonged to the inactive prophages subset (Fig. 4a). These 3 contigs represent 1.7% of all 168 ARG-positive contigs, while 30 at least were present on other mobile genetic elements (plasmids and ICE, see Supplementary Table 7, 8). A genetic analysis of the 3 ARG + viral contigs revealed an ICE-prophage hybrid in one case, an IS just next to the prophage region in the second case, and a possible phage gene interruption by the ARG in the last case (star above the gene, Fig. 5b). Altogether, we conclude that our search for ARG-positive viral contigs across viromes and microbiomes retrieved none in the viromes, and 3 on inactive prophages.

DISCUSSION

In-depth characterization of 14 pig fecal samples revealed that ARGs are present at 0.18 genes per Mb in the total microbiota of these samples (168 contigs, 279 ARGs). However, none of the ARG-containing contigs appears to belong to active phages, since ARGs were not detected in the viral contigs assembled from viromes reads (detection limit of 0.015 ARGs per Mb of viral contigs). Assuming an average genome size of 3 Mb for bacteria, and 30 kb for phages, these ARG numbers correspond to 0.5 ARG per bacterial genome, and below 3.6 10^{-4} per phage genome. This suggests that ARG genes are at least 1000-fold less likely per viral genome, compared to bacterial genomes. We therefore conclude that even in pigs, and in samples originating from farms with heavy antibiotic usage (such as sample 16), the probability of detecting an ARG on a phage contig is at least 1000-fold lower compared to a bacterial genome.

An explanation for this contrast could be that ARG, whenever they enter prophage genomes (by transposition or recombination), tend to mutate the recipient phage, which loses then rapidly its fitness. From the bacterial point of view, one could also argue that an ARG-containing prophage is so beneficial for the host that phage domestication rates speed up in order to retain the ARG.

Indeed, comparison of the 168 ARG-positive contigs to the pool of prophages present in microbiota revealed that three of them were prophages. Among these prophages, we introduced a distinction between those corresponding to functional and active phages, producing virions and therefore covered by virome reads, and inactive prophages. A large majority of these prophages were inactive (12 824, 77% of the total pool of prophage contigs), and the three ARG-positive prophage fragments belonged to these inactive prophages (Fig. 4a). This, together with the genetic map of these fragments (Fig. 4b), suggests they are being domesticated.

In fact, the subset of 'active phage' contigs in total microbiota samples entailed more than prophages. DNA from virulent phages was also present, such as the dominant phage of sample P7, belonging either to *Salasmaviridae* or *Sarlesviridae* family

(Supplementary Fig. 3) We can propose that the category of "active phages" also includes the "virocell" fraction, namely the fraction of phage particles replicating in bacteria at sampling time. A recent report suggests that in human microbiota, this fraction represents 25% of all viral contigs detected in total microbiota [47]. Interestingly, in our pig samples, the proportion of 'active phage' contigs was similar (23%). This pool of active phages constituted an important fraction of the ecosystem, as it represented in average 16% of genome abundance (RPKM) in each sample (Fig. 5). If phage burst size is around 10 in the intestine (it is estimated at 16 for the Lambda phage in the murine intestine [48]), this proportion of 16% would correspond to a maximum of 1.6% of bacteria hosting an actively replicating phage. In one case (piglet P7 sample), the active phage fraction included 50% of total RPKM. This sample hosted as well a putative Stramenopile, and its bacterial richness was strongly decreased, suggesting dysbiosis.

The comparison between viral contigs found in the virome fraction and the total microbiota is somewhat puzzling, though. Only 1097 of the 7755 virome contigs (14%) have homologs in the total microbiota. Two studies have reported similar comparisons. A limited 10% overlap has been reported by Gregory et al. [49] on 10 human fecal samples sequenced both in bulk and after viral particle filtration [50]. In a study involving 662 kid fecal samples however, the overlap was larger (30% of virome contigs overlapped microbiota viral contigs, and 50% the other way round) [51]. A technical reason for the lack of overlap between microbial and virome viral contigs is certainly phage incomplete assemblies, which may produce different fragments of the same phage in the two subsets. The fact that in the present study, some 3000 pig virome contigs, while not overlapping microbiota contigs, are nevertheless covered by microbiota reads (and reciprocally, see Supplementary Fig. 1) comforts this hypothesis. Besides this first explanation, the protocol used to extract total microbial DNA may also have extracted the DNA present in viral particles, among which ssDNA viruses would not be sequenced (indeed, 70% of the virome contigs not covered by microbiota reads corresponded to ssDNA viruses). Unless a specific step is added to bulk sequencing, virome sequencing will continue to bring more viral information relative to bulk sequencing for this reason. Finally, even if the two preceding technical problems are solved in the future, complete overlap between the two types of viral contigs will never occur, due to domesticated prophages that will never replenish the pool of active phages. A way to focus on the active temperate phages and exclude dormant prophages was recently documented [52]. Even though a complete view on the viral world is not possible upon microbiota sequencing as performed here, it offered the invaluable benefit of a glimpse at phage-bacteria interactions, and revealed the striking prevalence of phages reads (50% of total) in a dysbiotic piglet (P7).

The significant progress of all tools designed to sort out viral genomes allowed us to characterize 82% of the 2480 most abundant contigs of the 14 virome samples. Among them, two prominent phage clades were distinguished, inovirus infecting Gram-positive hosts (Fig. 2), and dsDNA phages similar to Faecalibacterium phage Oengus (Fig. 3 [41]). Even though viral samples had been purified on gradients, some bacterial and plasmid DNA was still present. We could not determine convincingly whether such DNA corresponded to generalized transduction, or to remaining contamination. Among them, three contigs encoded ARG (according to the Resfinder results): two had a *tetR* gene on bacterial DNA, both coming from the dysbiotic VP7 piglet sample which was also one of the most contaminated samples, one had a *kanR* gene on a plasmid (Table 1).

Given the limited number of samples analyzed, few conclusions could be drawn concerning the differences between adult pigs and piglets, nor between animals treated with many/few antibiotics. We noted however that piglets had significantly higher abundances of *Caudovirales* compared to adult pigs, as well as fewer ssDNA eukaryotic viruses. The abundance of ARG genes found in microbiota did not correlate with the (recently applied) antibiotic treatments. The *tetR* genes were the most abundant overall, and most likely reflected the past practice, now forbidden in Europe, of antibiotic use for animal growth.

In conclusion, the consistent observations that ARGs are rarely encoded by phages is encouraging for the fight against antibiotic resistance, as well as for phage therapy. These results suggest that this type of mobile genetic element should not pose a threat, except in phages with high levels of generalized transduction. This is also good news in terms of safety for the use of phages to combat recalcitrant infections, as with the exception of those with high transduction levels, phages should not lead to ARG spread.

DATA AVAILABILITY

Raw sequencing data and SPAdes assembly outputs (>2 kb, contigs for VP samples, scaffolds for P samples, before the clustering step at 95% identity) are accessible through the NCBI, Bioproject PRJNA741980.

REFERENCES

1. Checcucci A, Trevisi P, Luise D, Modesto M, Blasioli S, Braschi I, et al. Exploring the animal waste resistome: the spread of antimicrobial resistance genes through the use of livestock manure. Front Microbiol. 2020;11:1416.
2. Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, et al. A reference gene catalogue of the pig gut microbiome. Nat Microbiol. 2016;16161:16161.
3. Kazimierczak KA, Scott KP, Kelly D, Aminov RI. Tetracycline resistome of the organic pig gut. Appl Environ Microbiol. 2009;75:1717–22.
4. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit MA. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. 2016;1:237–47.
5. Barondess JJ, Beckwith J. bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. J Bacteriol. 1995;177:1247–53.
6. Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, et al. Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. Nat Commun. 2020;11:378.
7. Blanco-Picazo P, Roscales G, Toribio-Avedillo D, Gomez-Gomez C, Avilla C, Ballesté E, et al. Antibiotic resistance genes in phage particles from Antarctic and Mediterranean seawater ecosystems. Microorganisms. 2020;8:1293.
8. Zhang A, Call DR, Besser TE, Liu J, Jones L, Wang H, et al. beta-lactam resistance genes in bacteriophage and bacterial DNA from wastewater, river water, and irrigation water in Washington State. Water Res. 2019;161:335–40.
9. Wang M, Liu P, Zhou Q, Tao W, Sun Y, Zeng Z. Estimating the contribution of bacteriophage to the dissemination of antibiotic resistance genes in pig feces. Environ Pollut. 2018;238:291–8.
10. Colomer-Lluch M, Imamovic L, Jofre J, Muniesa M. Bacteriophages carrying antibiotic resistance genes in fecal waste from cattle, pigs, and poultry. Antimicrob Agents Chemother. 2011;55:4908–11.
11. Zolfo M, Pinto F, Asnicar F, Manghi P, Tett A, Bushman FD, et al. Detecting contamination in viromes using ViromeQC. Nat Biotechnol. 2019;37:1408–12.
12. Roux S, Krupovic M, Debroas D, Forterre P, Enault F. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. Open Biol. 2013;3:130160.
13. Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB, Westh H, et al. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. Nat Commun. 2016;7:13333.
14. Petrovich ML, Zilberman A, Kaplan A, Eliraz GR, Wang Y, Langenfeld K, et al. Microbial and viral communities and their antibiotic resistance genes throughout a hospital wastewater treatment system. Front Microbiol. 2020;11:153.
15. Maestre-Carballa L, Lluesma Gomez M, Angla Navarro A, Garcia-Heredia I, Martinez-Hernandez F, Martinez-Garcia M. Insights into the antibiotic resistance dissemination in a wastewater effluent microbiome: bacteria, viruses and vesicles matter. Environ Microbiol. 2019;21:4582–96.
16. Calero-Caceres W, Ye M, Balcazar JL. Bacteriophages as environmental reservoirs of antibiotic resistance. Trends Microbiol. 2019;27:570–7.
17. Lekunberri I, Subirats J, Borrego CM, Balcazar JL. Exploring the contribution of bacteriophages to antibiotic resistance. Environ Pollut. 2017;220:981–4.
18. Du F, Lv X, Duan D, Wang L, Huang J. Characterization of a Linezolid- and Vancomycin-Resistant Streptococcus suis Isolate That Harbors optrA and vanG Operons. Front Microbiol. 2019;10:2026.
19. Gabashvili E, Osepashvili M, Koulouris S, Ujmajuridze L, Tskhitishvili Z, Kotetishvili M. Phage transduction is involved in the intergeneric spread of antibiotic

resistance-associated blaCTX-M, mel, and tetM loci in natural populations of some human and animal bacterial pathogens. Curr Microbiol. 2020;77:185–93.
20. Lopez-Leal G, Santamaria RI, Cevallos MA, Gonzalez V, Castillo-Ramirez S. Letter to the editor: prophages encode antibiotic resistance genes in Acinetobacter baumannii. Microb Drug Resist. 2020;26:1275–7.
21. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. 2015;3:e985.
22. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019;4:1895–906.
23. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genom Bioinform. 2021;3:lqab067.
24. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46:W200–W204.
25. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020;8:90.
26. Shah SA, Deng L, Thorsen J, Pedersen AG, Dion MB, Castro-Mejia JL, et al. Manual resolution of virome dark matter uncovers hundreds of viral families in the infant gut. 2021.07.02.450849 (2021).
27. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578–85.
28. Sachsenroder J, Twardziok SO, Scheuch M, Johne R. The general composition of the faecal virome of pigs depends on age, but not on feeding with a probiotic bacterium. PLoS One. 2014;9:e88888.
29. Kim MS, Park EJ, Roh SW, Bae JW. Diversity and abundance of single-stranded DNA viruses in human feces. Appl Environ Microbiol. 2011;77:8062–70.
30. Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ. 2016;4:e2777.
31. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37:632–9.
32. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 2015;9:207–16.
33. Ruppé E, Ghozlane A, Tap J, Pons N, Alvarez AS, Maziers N, et al. Prediction of the intestinal resistome by a three-dimensional structure-based method. Nat Microbiol. 2019;4:112–23.
34. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75:3491–3500.
35. Hay ID, Lithgow T. Filamentous phages: masters of a microbial sharing economy. EMBO Rep. 2019;20:e47427.
36. Roux S, Krupovic M, Daly RA, Nayfach S, Schulz F, Sharrar A, et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019;4:1895–906.
37. Chopin MC, Rouault A, Ehrlich SD, Gautier M. Filamentous phage active on the grampositive bacterium Propionibacterium freudenreichii. J Bacteriol. 2002;184:2030–3.
38. Galiez C, Siebert M, Enault F, Vincent J, Soding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics. 2017;33:3113–4.
39. Kim AY, Blaschek HP. Isolation and characterization of a filamentous viruslike particle from Clostridium acetobutylicum NCIB 6444. J Bacteriol. 1991;173:530–5.
40. Summer EJ, Liu M, Gill JJ, Grant M, Chan-Cortes TN, Ferguson L, et al. Genomic and functional analyses of Rhodococcus equi phages ReqiPepy6, ReqiPoco6, ReqiPine5, and ReqiDocB7. Appl Environ Microbiol. 2011;77:669–83.
41. Cornuault JK, Petit MA, Mariadassou M, Benevides L, Moncaut E, Langella P, et al. Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes. Microbiome. 2018;6:65.
42. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33:2379–80.
43. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros J, Grønbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. Nat Biotechnol. 2021;39:555–60.
44. Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G, et al. Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. ISME J. 2016;10:2973–7.
45. Yang H, Xiao Y, Wang J, Xiang Y, Gong Y, Wen X, et al. Core gut microbiota in Jinhua pigs and its correlation with strain, farm and weaning age. J Microbiol. 2018;56:346–55.
46. Joyce A, McCarthy CGP, Murphy S, Walsh F. Antibiotic resistomes of healthy pig faecal metagenomes. Microb Genom. 2019;5:e000272.
47. Marbouty M, Thierry A, Millot GA, Koszul R. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. Elife. 2021;10:e60608.
48. De Paepe M, Tournier L, Moncaut E, Son O, Langella P, Petit MA. Carriage of lambda latent virus is costly for its bacterial host due to frequent reactivation in monoxenic mouse intestine. PLoS Genet. 2016;12:e1005861.
49. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe. 2020;28:724–40 e8.
50. Shkoporov AN, Clooney AG, Sutton T, Ryan FJ, Daly KM, Nolan JA, et al. The human gut virome is highly diverse, stable, and individual specific. Cell Host Microbe. 2019;26:527–41 e5.
51. Johansen J, Plichta D, Nissen JN, Jespersen ML, Shah SA, Deng L, et al. Genome binning of viral entities from bulk metagenomics data. bioRxiv https://doi.org/10.1101/2021.07.07.451412 (2021).
52. Redgwell TA, Thorsen J, Petit MA, Deng L, Vestergaard GA, Russel J, et al. Prophages in the infant gut are largely induced, and may be functionally relevant to their hosts. bioRxiv https://doi.org/10.1101/2021.06.25.449885 (2021).
53. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27:1009–10.

ACKNOWLEDGEMENTS

We would like to acknowledge the Migale bio-informatics facility at INRAE for professional support, and Michi Waygood and Amanda Toperoff from Crayon-bleu for editing service. MBD is recipient of graduate scholarships from the FRQNT and Sentinel North. SM holds a T1 Canada Research Chair in Bacteriophages. QL-B and FT were recipient of a financial support from an INRAE MEM program.

AUTHOR CONTRIBUTIONS

MB, QL-B, JL, MD, SM, ELC, JE, OZ., contributed to the generation and analysis of data and revised the manuscript, FT, EM, CD, CA contributed to the generation of data, MAP analyzed data, supervised the work and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43705-021-00054-8.

Correspondence and requests for materials should be addressed to Marie-Agnès. Petit.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021