



Genetic Mining of Newly Isolated Salmophages for Phage Therapy

Julia Gendre, Mireille Ansaldi, David Olivenza, Yann Denis, Josep Casadesús,
Nicolas Ginet

► To cite this version:

Julia Gendre, Mireille Ansaldi, David Olivenza, Yann Denis, Josep Casadesús, et al.. Genetic Mining of Newly Isolated Salmophages for Phage Therapy. International Journal of Molecular Sciences, 2022, 23 (16), pp.8917. 10.3390/ijms23168917 . hal-03781524

HAL Id: hal-03781524

<https://amu.hal.science/hal-03781524>

Submitted on 14 Nov 2022

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



Article

Genetic Mining of Newly Isolated Salmophages for Phage Therapy

Julia Gendre ¹, Mireille Ansaldi ^{1,*} , David R. Olivenza ², Yann Denis ³, Josep Casadesús ² and Nicolas Ginet ^{1,*} 

¹ Laboratoire de Chimie Bactérienne, (UMR7283)-CNRS/Aix-Marseille Université, 13009 Marseille, France

² Departamento de Genética, Facultad de Biología, Universidad de Sevilla, 41012 Sevilla, Spain

³ Institut de Microbiologie de la Méditerranée, (FR3479)-CNRS/Aix-Marseille Université, 13009 Marseille, France

* Correspondence: ansaldi@imm.cnrs.fr (M.A.); nginet@imm.cnrs.fr (N.G.)

Abstract: *Salmonella enterica*, a Gram-negative zoonotic bacterium, is mainly a food-borne pathogen and the main cause of diarrhea in humans worldwide. The main reservoirs are found in poultry farms, but they are also found in wild birds. The development of antibiotic resistance in *S. enterica* species raises concerns about the future of efficient therapies against this pathogen and revives the interest in bacteriophages as a useful therapy against bacterial infections. Here, we aimed to decipher and functionally annotate 10 new *Salmonella* phage genomes isolated in Spain in the light of phage therapy. We designed a bioinformatic pipeline using available building blocks to de novo assemble genomes and perform syntactic annotation. We then used genome-wide analyses for taxonomic annotation enabled by vContact2 and VICTOR. We were also particularly interested in improving functional annotation using remote homologies detection and comparisons with the recently published phage-specific PHROG protein database. Finally, we searched for useful functions for phage therapy, such as systems encoded by the phage to circumvent cellular defenses with a particular focus on anti-CRISPR proteins. We, thus, were able to genetically characterize nine virulent phages and one temperate phage and identify putative functions relevant to the formulation of phage cocktails for *Salmonella* biocontrol.

Keywords: bacteriophages; *Salmonella*; phage therapy; biocontrol; genomic analysis; functional annotation; PHROG database



Citation: Gendre, J.; Ansaldi, M.; Olivenza, D.R.; Denis, Y.; Casadesús, J.; Ginet, N. Genetic Mining of Newly Isolated Salmophages for Phage Therapy. *Int. J. Mol. Sci.* **2022**, *23*, 8917. <https://doi.org/10.3390/ijms23168917>

Academic Editor: Alicja Węgrzyn

Received: 20 June 2022

Accepted: 7 August 2022

Published: 10 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The massive and often indiscriminate use of antibiotics for decades in human and animal healthcare as well as in agriculture gives rise to antibiotic multi-resistance in bacterial pathogens; this has become a major concern worldwide as it leads to an increasing number of therapeutic failures. According to current projections reported by the World Health Organization (WHO), infectious diseases could once again become one of the leading causes of death in the world by 2050 and cause dramatic damage to the world economy [1]. The “One Health” concept recognizes the interdependency between the environment, human health and animal health; it is promoted by various international organizations such as the WHO, the Organization for Animal Health (OIE), and the Food and Agriculture Organization (FAO) (see for a definition the FAO website at <https://www.fao.org/one-health/en>) (accessed on 14 June 2022). One Health aims, among other objectives, to promote the rational use of antibiotics and the development of alternative strategies to tackle bacterial infections.

Bacteriophages (or phages) are viruses that infect bacteria and thus represent, as anti-bacterial agents, a promising alternative (or complement) to antibiotics. A given bacteriophage specifically infects a more or less narrow range of bacteria, sometimes down to the species or serotype level. Furthermore, bacteriophages are very diverse and generally easy to isolate and produce. Bacteriophages were independently discovered during the

First World War by Frederick W. Twort in 1915 [2] and by Félix d'Hérelle in 1917 [3], who coined the term “bacteriophage” (see Félix d'Hérelle seeding article presented in English by Dr. Roux [4]). Very rapidly in 1918, when antibiotics did not even exist, Félix d'Hérelle foresaw how bacteriophages could be used against bacterial infections and successfully utilized them to treat dysentery [5]. Phage therapy was born and developed to treat both humans and animals. However, in the 1940s, phage therapy was supplanted by the emergence of antibiotics and was abandoned in the Western world, but it was kept in use until today in the former Eastern Bloc countries and former USSR member states [6]. Nowadays, there is a renewed interest worldwide for bacteriophages and phage therapy in the search for an alternative to antibiotics to tackle multi-resistant bacteria and develop targeted antibacterial therapies.

Salmonella enterica is a Gram-negative zoonotic bacterium and a major cause of diarrhea worldwide. Avian breeding farms are the main reservoirs for the *Salmonella* species, transmitted to humans by food, especially eggs. In the concept of “One Health”, treating these reservoirs would have a positive impact on human health. Isolation and characterization of phages targeting *Salmonella* are thus important to develop *S. enterica* biocontrol strategies and prevent massive antibiotic resistance in animal reservoirs. Phage therapy relies essentially on the formulation of a cocktail, regrouping several different phages in order to avoid the emergence of bacterial resistance, especially cross-resistance, and to eventually target as wide a range of bacterial serotypes as possible. Another beneficial property of bacteriophage cocktails is that they can be designed to target specific pathogens, leaving the rest of the microbiota unaffected, which is not the case with antibiotics and their indiscriminate action against a wide range of bacteria. A recent and relevant study by Nale et al. highlights the synergistic effects of a three-phage cocktail against prevalent *Salmonella* serotypes on poultry and pigs [7]. Using *Galleria mellonella* as a model that correlates to large-scale animal models, the authors proved that, in combination, their 3 phages could lyse about 99.97% of the 22 serotypes they investigated. They also showed in vivo in the *Galleria mellonella* model that their cocktail was very efficient as a prophylactic agent. Closer to human health issues is the recent report of the successful treatment of a patient infected by a multi-drug resistant *Mycobacterium abscessus* strain by a patient-tailored three-phage cocktail [8]. However, the authors cautiously warn that a generalized approach is far from being a reality.

Just like antibiotherapy, phage therapy collides with bacterial resistance to phage infection. Its success relies on the evaluation of such a probability, implying that we should not reproduce the errors we made with antibiotics. Indeed, the co-evolution of bacteria and their viruses for eons led to the adaptation and acquisition of a high diversity of bacterial anti-phage strategies. The “pan-immune system” of bacteria comprising anti-phage defenses has been nicely reviewed by Bernheim and Sorek [9]. These strategies evolved by bacteria can be as simple as selecting for mutations in the phage receptor and preventing phage adsorption by modifying and adapting its lipopolysaccharide (LPS). Of greater concern are mutations in non-receptor host factors that can lead to phage cross-resistance, as was shown recently in *S. enterica* serovar Typhimurium [10]. Bacteria can also resort to elegant, more complex and specialized defense systems once the phage genome has been injected, such as Restriction-Modification (RM) [11], CRISPR-Cas [12], abortive infection (Abi) [13] comprising the recently described cyclic-oligonucleotide-based anti-phage signaling system (CBASS) [14], and bacteriophage Exclusion (BREX) [15–17] systems to name but a few. These systems are often passed on between bacteria via horizontal gene transfers. As we keep discovering new anti-phage defense systems, the selection of phages and the formulation of a cocktail should take bacterial defense against phage infection into consideration. Undoubtedly, co-evolution works both ways, and phages, therefore, evolved their own arsenal to counter cellular defenses. One can cite anti-RM or anti-CRISPR proteins [18] as well as the newly discovered anti-CBASS and anti-Pycsar proteins [19,20]. Given the diversity of anti-phage defense systems, a matching diversity of yet unknown anti-cellular defense functions is probably buried in the phage genomic “dark-matter”

awaiting discovery. The phage “dark-matter” encompasses all those predicted genes in phage genomes and metaviromes with unknown functions [21]. It is important to collect as much information as possible on the phage genomes to select the best candidates when foreseeing phage therapy and to keep a balance between killing efficiency, susceptibility against cellular defenses, and potential antagonistic interactions between the therapeutical phages. A thorough genomic characterization of phage genomes, and especially functional annotation, are thus of crucial importance for the successful deployment of phage-based biocontrol strategies.

In this study, we sequenced and analyzed 10 new bacteriophage genomes infecting *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain ATCC 14028S. These phages were isolated from wastewater and fresh water ponds in the Sevilla area in Spain. Most phages originate from a study by Olivenza et al. [22] aimed at developing epigenetic phage biosensors used to identify and select phages from different natural environments. Others belong to the personal collections of Dr. M. Ansaldi and Prof. J. Casadesús. To improve phage gene detection and annotation, we benchmarked five different gene callers for syntactic annotation. We could ascribe a taxonomic affiliation for each phage at the genus level, thanks to genome-wide comparative analyses enabled by vContact2 [23] (based on the gene-sharing network) and VICTOR (based on genome-wide phylogeny) [24]. For functional annotation of the predicted ORFs, we combined complementary approaches based on remote homologies detection using Hidden Markov Model (HMM) protein profiles comparisons with PHROG [25], a newly published database of viral protein clusters, and the pdb70 database at the Protein Data Bank [26]. Thirdly, we specifically searched for anti-CRISPR proteins (Acr) using the integrated online platform AcrHub [27], relying on machine learning for Acr prediction. Acr are indeed valuable assets when foreseeing phage therapy. We further mined our annotated genomes for functions that may favor or, on the contrary, disfavor the selection of a phage for phage therapy.

The 10 newly isolated phage strains studied here all belong to the *Caudoviricetes* and are spread among four genera (*Kuttervirus*, *Chivirus*, *Jerseyvirus*, and *Lederbergvirus*) with a majority of six *Kuttervirus* phages isolated (*Ackermannviridae*). Eight of these ten strains are new species. One phage, *Salmonella* phage Salfasec13b (*Lederbergvirus*), was predicted as temperate, which is not a desired property for phage therapy. Among the most interesting features, we found phage-encoded proteins targeting the cellular defenses against phage infections, such as the Restriction-Modification, CRISPR-Cas, and Abortive infection systems. It is noteworthy that, *Salmonella* phage SeF3a (*Kuttervirus*) is likely to code for a novel anti-Abi system derived from the bacterial Phage Shock Protein A (PspA), making this phage an interesting candidate for phage therapy with two different anti-Abi systems, one found in all our *Kutterviruses* previously known to allow bacteriophage T4 to resist the Rex Abi system encoded in some prophages, and the second found in *Salmonella* phage SeF3a hypothesized in this study. We incidentally could propose functional annotations for 24 protein families in the PHROG database currently annotated “unknown function”. Altogether, this work highlights the need to design specific pipelines for phage genomic analyses to unravel the phage genomic “dark matter” and illustrates the importance of prior knowledge of functions encoded in phage genomes for an educated selection of phage candidates for therapeutic cocktails.

2. Results and Discussion

In this work, we used 14 phage isolates from wastewater and fresh water ponds in the Sevilla region in Spain, some of them detected and isolated by Olivenza et al. using an epigenetic phage biosensor [22].

2.1. Phage Genome de Novo Assembly

Genome de novo assembly with SPAdes led to the identification of a total of 18 meaningful contigs in the entire dataset with contig lengths between 40 and 160 kb, a range compatible with phage genome sizes (Appendix A, Table A1). Of note, SPAdes

assembled two distinct contigs from Salfasec_11, Salfasec_13, Se_F3, and Se_F6 purified DNA, possibly representing two different phages in the same sample. This is in accordance with the TEM images (Figure 1), where we can visualize two different phage morphologies in purified phage suspensions for Salfasec_11 (two siphovirus with different tail lengths, red arrows), Se_F3 (one siphovirus with a small head and a long tail, and one myovirus, red and blue arrows, respectively), and Se_F6 (one siphovirus with a small head and a long tail and one myovirus, red and blue arrows, respectively) phage suspensions. In Salfasec_13 phage suspension, we could only identify one myovirus morphology (blue arrow) despite the assembly of two distinct genomes with very different sizes (157,296 and 59,161 bp).

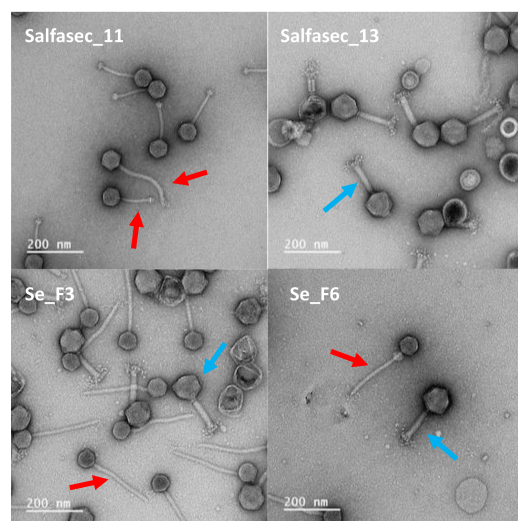


Figure 1. TEM images of negatively stained purified phage suspension (Salfasec_11, Salfasec_13, Se_F3, and Se_F6). Phages were purified by centrifugation of ATCC 14028S culture lysates and resuspended in TEM buffer. Grids were prepared and visualized as described in the Materials and Methods section. Red arrows: siphovirus morphology. Blue arrows: myovirus morphology. Scale bar: 200 nm.

When relevant, the two DNA sequences present in the same sample were distinguished by adding “a” and “b” to the sample isolate name. Experience in many labs dealing with phage isolation and sequencing demonstrated that the experimenters are not immune to sample cross-contaminations despite the important precautions taken, or that the same phage can be isolated several times from different samples. Resident prophage(s) can also be induced at a basal level in laboratory culture conditions or by the stress triggered by incoming phage infections. In such cases, one would expect more than one genome to be sequenced. As a reminder, our propagation strain *S. enterica* ATCC 14028S is lysogenic for at least three functional temperate bacteriophages, namely Gifsy-1, Gifsy-2, and Gifsy-3 [28], with a fourth one we predicted as complete and functional using PHASTER [29] in silico analysis. We chose to center our study on potential new phage species. We used VIRIDIC [30] to group the contigs summarized in Appendix A, Table A1 (species threshold similarity 95% and genus threshold similarity 70%). We thus could classify the 18 contigs into 4 different genera and 10 different species. Some species clusters contained 100% identical DNA sequences. For instance, all three contigs in species cluster 3 (Se_ML1, Salfasec_13a, and Se_F3a) are identical. We then selected 10 unique contigs representing the 10 different species clusters for downstream analyses (Table 1). The intergenomic similarity matrix, as well as the species and genus cluster, are shown in Table S1.

Table 1. The 10 contigs selected for our downstream genomic analyses with their closest homolog in the NCBI nr database.

Contig	Size (bp)	Genus Cluster	Blastn Best Hit ¹	Cov. ²	Evalue	Identity	Length (bp) ³
Se_AO1	157,543	3	Salmonella phage PhiSH19	97%	0.0	99.6%	157,785
Se_EM1	159,312	3	Salmonella phage S8	93%	0.0	98.2%	158,432
Se_EM2	156,948	3	Salmonella phage S8	89%	0.0	96.1%	158,432
Se_EM4	157,810	3	Salmonella phage vB_SalM-LPST94	89%	0.0	97.6%	156,548
Se_F1	42,784	2	Salmonella phage S101	90%	0.0	92.2%	42,621
Se_F2	59,084	1	Salmonella phage BPS1	99%	0.0	98.9%	58,852
Se_F3a	157,219	3	Salmonella phage vB_SentM_sal1	97%	0.0	99.2%	157,338
Se_F6a	157,240	3	Salmonella phage Chennai	91%	0.0	96.0%	157,462
Se_F6b	59,519	1	Salmonella phage Chi	98%	0.0	93.0%	59,578
Salfasec_13b	40,455	4	Salmonella phage S149	93%	0.0	99.8%	41,391

¹ We kept for Blastn best-hits only fully sequenced phage genomes (accession numbers in Appendix A, Table A3).

² Query coverage. ³ Query length.

Among these, only Salfasec_13b contig exhibits some small but significant DNA sequence identity with the propagation strain ATCC 14028S genome but only on 3 kb with 89.6% identity. This finding rules out that any of the sequences we obtained originate from resident ATCC 14028S prophages induced during phage propagation. Hereafter, these contigs will be referred to as Salmonella phage strains.

2.2. DNA Packaging Prediction

When DNA libraries were prepared from randomly fragmented DNA, as was the case for Se_F1, Se_F2, Se_F3a, Se_F6a, and Se_F6b, we used PhageTerm to predict putative genome termini and DNA packaging mode. Table 2 summarizes the prediction results.

Table 2. Summary of PhageTerm results.

Contig	Ends	Left	Right	Permutated	Class	Type	Sequence Cohesive
Se_F1	Redundant	Random	Random	Yes	-	-	-
Se_F2	Non-redundant	48,283	48,294	No	Cos (5')	Lambda	GGTGCGCAGAGC
Se_F3a	Redundant	Random	Random	Yes	-	-	-
Se_F6a	Redundant	Random	Random	Yes	-	-	-
Se_F6b *	Non-redundant	1	12	No	Cos (5')	Lambda	GGTGCGCAGAGC

* Result to be considered carefully as the mean genome coverage is 15, under the software recommended coverage threshold (50). Ends: describe whether the ends are unique or redundant. Left and Right: nucleotide position of the Left and the Right termini in the DNA sequence. Permutated: describe whether individual encapsidated DNA molecules are all the same (non-permutated, e.g., for cos packaging) or circularly permutated (e.g., head-full packaging). Class: predicted packaging mechanism. Type: model phage for the considered packaging mechanism.

According to PhageTerm predictions, Se_F1, Se_F3a, Se_F6a use headful (or *pac*) packaging. This is, for instance, the packaging mode of Salmonella phage P22 [31]. This mode of packaging implies that the first cut by the terminase occurs at a *pac* site on the DNA concatemer issued from the phage DNA replication cycle. The second cut occurs at a random site on the concatemer when the phage capsid is full (more than 1 genome unit length is packaged). The third cut occurs at a random site on the concatemer for the second capsid packaging, and so on. Accordingly, at the population level, we obtained circularly permutated genomes for Se_F1, Se_F3a, and Se_F6a. In contrast, Se_F2 and Se_F6b lay in the *cos* packaging group exemplified by Escherichia coli phage Lambda [32], although for Se_F6b, the low sequence coverage precludes valid termini prediction by PhageTerm. In this mode of packaging, the concatemer is cut at *cos* sites, and each phage capsid contains one identical genome unit length. We could not use PhageTerm for Se_AO1, Se_EM1,

Se_EM2, Se_EM4, and Salfasec_13b due to the enzymatic DNA fragmentation method used for the preparation of the DNA libraries that does not allow recovering genome extremities (see Materials and Methods). Nevertheless, we could gain some information on their DNA packaging strategies. Paired-reads mapping with Bowtie2 [33] and mapping visualization with IGV [34] (data not shown) revealed circularly permuted sequences, suggesting that DNA packaging for these phages was probably through the headful mechanism.

2.3. Syntactic Annotation of 10 Newly Isolated *Salmonella* Phage Genomes

It is common knowledge that predictions vary from one gene caller to another, resulting at the extreme in either missing “real” ORFs or predicting “false” ones, not to mention discrepancies in initiation codon predictions [35]. ORF prediction is rendered even more complex in bacteriophages due to their genome organization. Indeed, phage genomes are compact, with overlapping genes; in some cases, a gene can be entirely included in another gene, a situation called “overprinting”. This is the case, for example, for Rz-spanin-o in *E. coli* phage Lambda, which is entirely coded in the Rz-spanin-i gene [36]. Phage genomes are often organized in functional modules that can be acquired/modified by recombination with other viruses or mobile genetic elements or by recombination with their bacterial host and its eventual prophage(s); this mosaic organization resulting from many horizontal gene transfers makes it difficult for any gene caller to build uniform predictive models. The “phage language” is yet to be completely translated and is a current focus of many groups in the field, with great hopes put into deep learning techniques.

We, thus, compared the output of AMIGene, Glimmer, MetaGeneAnnotator, Phanotate, and Prodigal to identify the best algorithm (or combination of algorithms) to optimize the gene calling of our phage genomes. We used Se_AO1 as a case study before generalizing the method to the other nine phage genomes included in this study. In order to compare the different predictions, we built a table for each gene caller where an individual predicted ORF is identified by a stop codon and a coding strand (+ or −) without its predicted initiation codon. This strategy should avoid discrepancies in the predicted initiation codon between the different algorithms. We then compiled the five prediction sets in a single table for each phage with the following unique rule: a predicted ORF from a gene caller is deemed identical to an ORF predicted by another gene caller if they share the same stop codon on the same strand, regardless their respective predicted initiation codon that may be different between the two gene callers. For each phage, we indicated in the table for each predicted ORF whether it is predicted or not by each of the five gene callers. Results are listed in Table S2, and gene callers’ comparison is illustrated in the Venn diagrams shown in Figure 2 for Se_AO1.

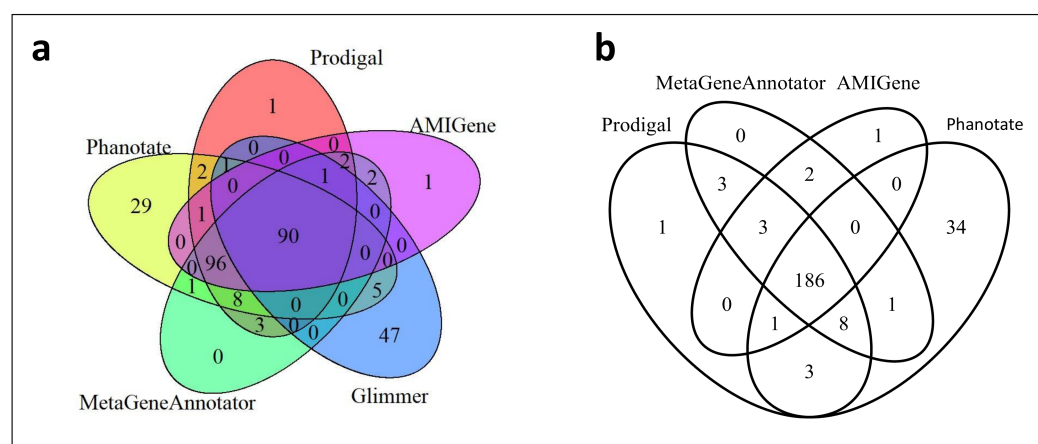


Figure 2. Comparison of the ORFs predictions for the Se_AO1 genome. (a) Venn diagram for AMIGene, Glimmer, MetaGeneAnnotator, Phanotate, and Prodigal predictions ($n = 290$ ORFs). (b) Venn diagram omitting Glimmer predictions ($n = 243$ ORFs).

Figure 2a compares the results obtained with all five gene callers. Glimmer displayed the lowest ($n = 144$) and Phanotate the highest ($n = 233$) number of predicted ORFs. A total of $n = 290$ ORFs were predicted, 90 of which were shared by all algorithms (31%). According to this analysis, 69% of the predicted ORFs would depend on the specificity of the gene caller algorithms, which is very improbable. Glimmer predicted $n = 144$ ORFs, 32.6% of which were unique to this software; therefore, Glimmer seems ill-fitted to our purpose. Thus, in Figure 2b, we drew another Venn diagram omitting Glimmer predictions. On the $n = 243$ predicted ORFs, 76.5% were common to all algorithms ($n = 186$), which is an expected range if all predictors ran with a similar level of accuracy. Of note, Phanotate predicted 34 ORFs overlooked by the 3 other gene callers. This is not surprising as Phanotate has been purposely designed to identify ORFs in phage genomes.

Setting a cut-off limit for the gene size was not an easy task, as we did not want to keep false predictions or discard ORFs that do code for small proteins. One has to keep in mind that small ORFs in bacteriophages can encode small proteins enabling phages to evade host anti-phage systems. Among them, we found anti-CRISPR proteins (Acr), such as ACR3112-12 (52 aa) from *Pseudomonas* phage D3112 [37]; proteins preventing superinfection, such as the immunity protein Imm (83 aa) of bacteriophage T4 [38] or the lipoprotein Llp (77 aa) of bacteriophage T5 [39]; anti-RM proteins, such as Ocr (117 aa) of bacteriophage T7 [40]; the newly discovered anti-CBASS Acb (about 94 aa) found in an increasing number of phages of various phylogenetic background, such as *Pseudomonas* phages PaMx33, 35, 41, and 43, as well as the enterophageT4 [19,20].

As the identification of anti-CRISPR and other types of anti-host defenses proteins was one of our goals, we chose a conservative approach as a first approximation and kept all the ORFs predicted by AMIGene, MetaGeneAnnotator, Phanotate, and Prodigal for the 10 phage genomes studied here. When several possibilities for the initiation codon were available, we kept the longest version of the ORF. Across our 10 genomes, about 72% of ORFs were predicted by all 4 gene callers (slightly less for Salfasec_13b) and about 14.5% by Phanotate only (Appendix A, Table A2).

2.4. Taxonomic Affiliation

Due to the remarkable diversity of phage nucleotide sequences and the pervasive mosaicism of phage genomes, phage phylogeny does not follow traditional hierarchical phylogeny, as noted by Dion et al. [41], and it has always been a challenge to define universal gene markers to classify phages such as the 16S RNA commonly used for bacteria. The current taxonomy of viruses is evolving rapidly using genome-wide analyses, particularly the concept of shared orthologous protein families to infer evolutionary relationships between viral genomes [42,43]. One important advantage of using such methods to classify phages is that they do not rely on prior knowledge of protein functions.

We thus resorted to genome-wide analyses at the protein level for taxonomic classification with vContact2 and VICTOR. Figure 3a displays a representation generated by Cytoscape of the interaction network between *Caudoviricetes* bacteriophages generated by vContact2 and including our 10 newly isolated phages. In this representation, the strength of the connection linking two phages depends on the number of protein clusters they have in common.

vContact2 also defines viral clusters (VC) as regrouping phages that share a significant number of protein clusters. Each VC comprises viruses taxonomically classified at various levels (genus, sub-family, or family). Our 10 newly isolated phages fall into 4 different viral clusters whose members are colored in Figure 3b. Se_F2 and Se_F6b belong to the VC_20 (orange), comprising 22 viruses defined at the genus level (*Chivirus*), Se_AO1, Se_EM1, Se_EM2, Se_EM4, Se_F3a, and Se_F6a to the VC_34 (blue) comprising 63 viruses defined at the family level (*Ackermannviridae*), Se_F1 to the VC_75 (red) comprising 25 viruses defined at the sub-family level (*Guernseyvirinae*), and finally, Salfasec_13b to the VC_243 (green) comprising 16 viruses defined at the genus level (*Lederbergvirus*). The complete list of the VC generated by vContact2 is available in Table S3.

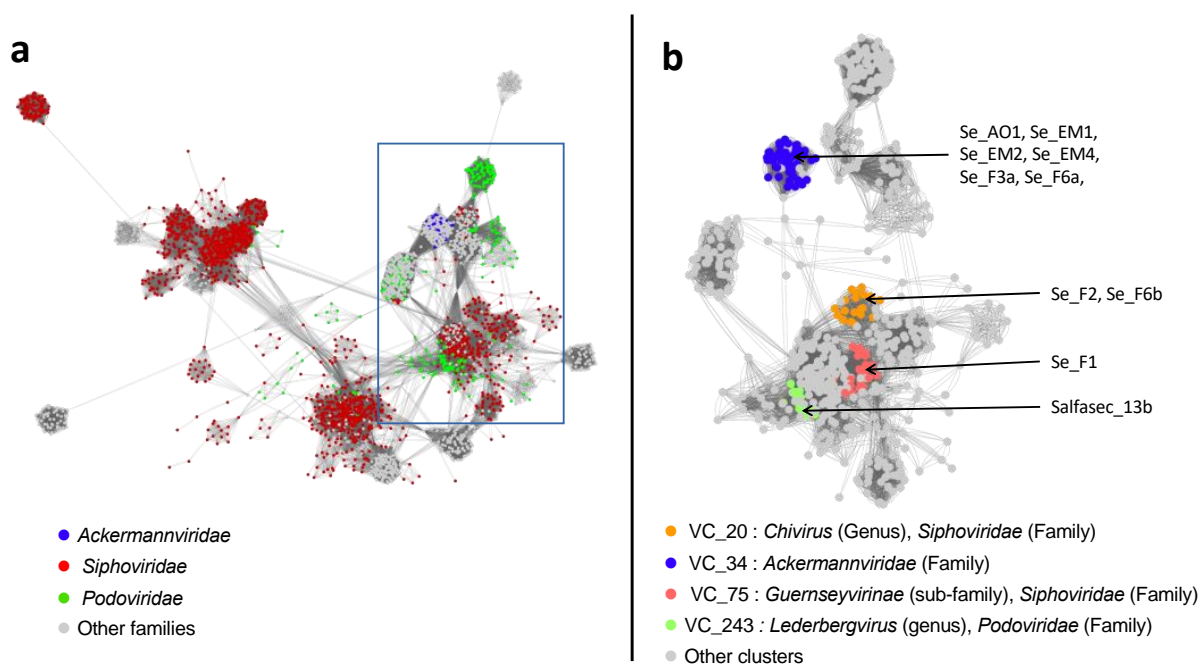


Figure 3. Interaction network between *Caudoviricetes* bacteriophages based on shared protein ortholog families. The interaction network was built using vContact2 and visualized with Cytoscape. Each dot corresponds to a single phage. **(a)** Network comprising all the *Caudoviricetes* viruses in the vContact2 database. Viruses belonging to the *Ackermannviridae*, *Siphoviridae*, and *Podoviridae* families are highlighted in blue, red, and green, respectively. The rectangle indicates the region in the network where the 10 phages studied here are positioned. **(b)** Sub-network of the region underlined in **(a)** restricted to *Ackermannviridae*, *Podoviridae*, and *Siphoviridae* surrounding our 10 new phages. Viruses belonging to the *Chivirus* genus (VC_20), the *Ackermannviridae* family (VC_34), the *Guernseyvirinae* sub-family (VC_75), and the *Lederbergvirus* genus (VC_243) are highlighted in orange, blue, red, and green, respectively. The positions in the sub-network of the 10 new phages are indicated by arrows. *Siphoviridae* and *Podoviridae* families have been removed from the latest ICTV taxonomic classification but still existed in the vContact2 database we used for our analyses.

For the phages that remained unclassified at the genus level after the previous analysis, we refined the classification with VICTOR using the “amino acid” option analysis to build phylogenetic trees for Se_F1 (Figure 4a) and Se_AO1, Se_EM1, Se_EM2, Se_EM4, Se_F3a, and Se_F6a (Figure 4b). In each case, a minimum of three genomes belonging to different families were used as outgroups.

Figure 4 unambiguously shows that Se_F1 belongs to the *Jerseyvirus* genus and that Se_AO1, Se_EM1, Se_EM2, Se_EM4, Se_F3a, and Se_F6a belong to the *Kuttervirus* genus.

To determine whether these 10 phage strains represent novel species or not, we first downloaded for each of the 4 genera (*Kuttervirus*, *Chivirus*, *Jerseyvirus*, and *Lederbergvirus*) all the species officially acknowledged in the latest version of the ICTV classification, then computed with VIRIDIC within each genus the intergenomic similarities table (Table S4). With a species threshold of 95%, we, thus, could define 8 new species. Table 3 summarizes the final taxonomic affiliation of the 10 new phage strains studied in this work, 8 of which are new species. All phages in this table belong to the *Duplodnaviria* realm, *Heunggongvirae* kingdom, *Uroviricota* phylum, and *Caudoviricetes* class. The eight new phage species are noted in blue in Table 3 with a proposed name in the recommended binomial format. The proposed strain name is given in the first column.

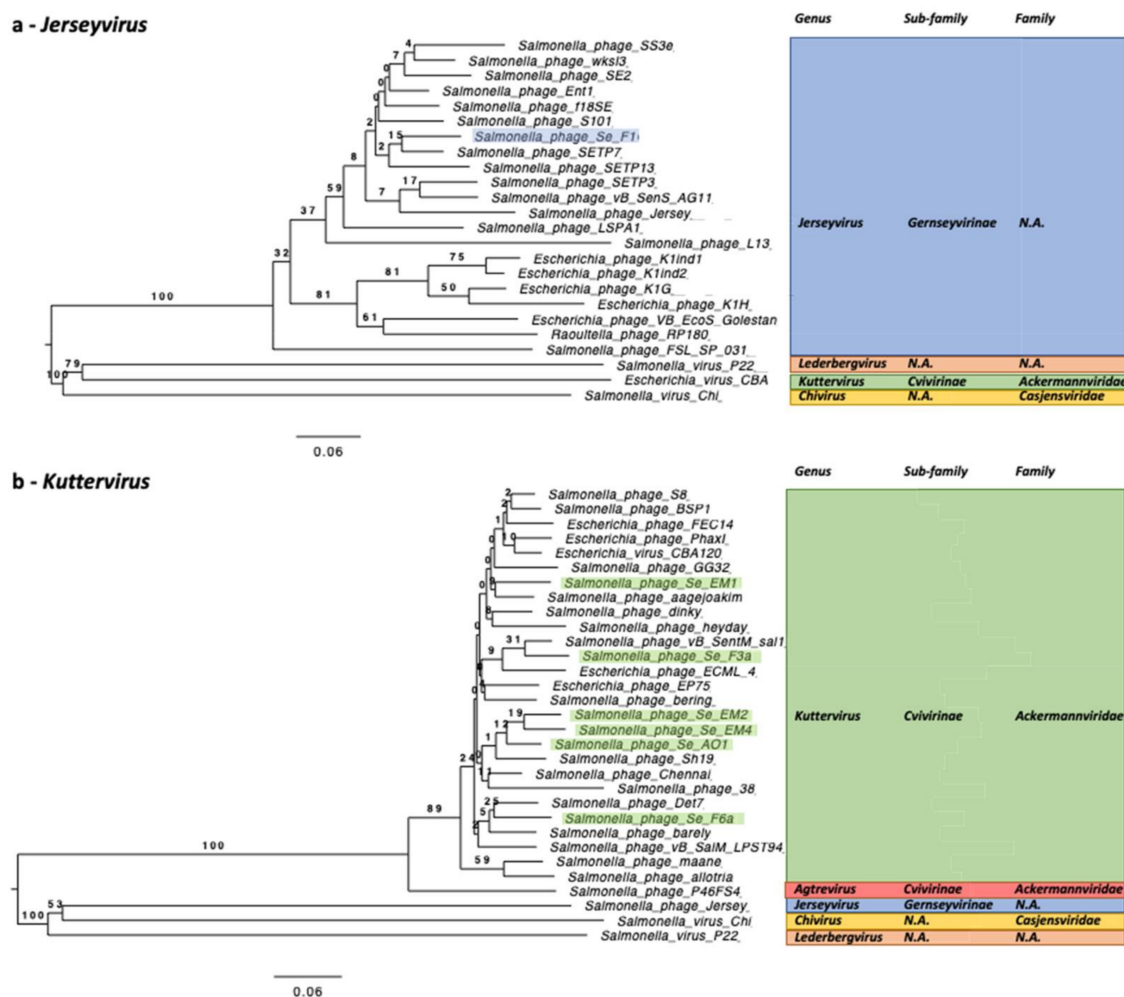


Figure 4. Taxonomic affiliation at the genus level based on genome-wide amino acid analysis by VICTOR. (a) *Jerseyvirus* genus. Twenty *Jerseyvirus* from the RefSeq database were included, as well as one *Kuttervirus* (*Escherichia virus CBA120*), one *Chivirus* (*Salmonella phage Chi*), and one *Lederbergvirus* (*Salmonella phage P22*) serving as outgroups. Se_F1 is highlighted in light blue. (b) *Kuttervirus* genus. Twenty-one *Kuttervirus* from the RefSeq database were included, as well as one *Agtrevirus* (*Salmonella phage P46FS4*), one *Lederbergvirus* (*Salmonella phage P22*), one *Chivirus* (*Salmonella phage Chi*), and one *Jerseyvirus* (*Salmonella phage Jersey*) serving as outgroups. Se_AO1, Se_EM1, Se_EM2, Se_EM4, Se_F3a, Se_F6a are highlighted in light green.

Table 3. Taxonomic classification of the 10 newly isolated *Salmonella* phages included in this study.

Strain	Family	Sub-Family	Genus	Species
Salmonella phage SeAO1	Ackermannviridae	Civirinae	Kuttervirus	Kuttervirus SH19
Salmonella phage SeEM1	Ackermannviridae	Civirinae	Kuttervirus	"Kuttervirus SeEM1"
Salmonella phage SeEM2	Ackermannviridae	Civirinae	Kuttervirus	"Kuttervirus SeEM2"
Salmonella phage SeEM4	Ackermannviridae	Civirinae	Kuttervirus	"Kuttervirus SeEM4"
Salmonella phage SeF3a	Ackermannviridae	Civirinae	Kuttervirus	"Kuttervirus SeF3a"
Salmonella phage SeF6a	Ackermannviridae	Civirinae	Kuttervirus	"Kuttervirus SeF6a"
Salmonella phage SeF1	N.A. #	Guernseyvirinae	Jerseyvirus	"Jerseyvirus SeF1"
Salmonella phage SeF2	Casjensviridae	N.A.	Chivirus	"Chivirus SeF"2
Salmonella phage SeF6b	Casjensviridae	N.A.	Chivirus	Chivirus FSLSP088
Salmonella phage Salfasec13b	N.A.	N.A.	Lederbergvirus	"Lederbergvirus Salfasec13b"

Non-Applicable. In blue with quotation marks: new phage species.

This taxonomic classification supports the PhageTerm predictions for Salmonella phage SeF2 and SeF6b (*Chivirus* genus) as their predicted termini sequences are the same as the one experimentally determined for the *Chivirus* Salmonella phage Chi (5'-GGTGCGCAGAGC-3') [44]. This gives credit to PhageTerm prediction for Salmonella phage SeF6b despite the low sequence coverage (Table 2).

2.5. Functional Annotation

To date, the most common and fastest method to tentatively ascribe a function to a newly predicted gene product is by comparing the ORF or gene product sequences with existing sequences deposited in various databases, such as the nr/nt databases at NCBI. This approach relies on the confidence one can have in previous annotations. It is sometimes tedious to trace back the reasons justifying current annotations. For proteins that finally end up with “unknown function” annotation, one has to keep a critical eye for several reasons: (i) the function can truly be a novel one (the phage “dark-matter”), (ii) the amino acid sequence is too divergent to detect homologies with proteins of known function using “simple” sequence alignment algorithms such as BlastP, or (iii) the gene product does not exist (an artifact due to the ORF prediction algorithms). Functional annotation can be improved thanks to methods developed to detect remote homologies, classify protein in orthologous groups, or fold primary sequences onto known secondary or tertiary structures. As noted by Chen et al. [45], these many different approaches can be combined to yield reliable functional predictions.

Here we chose to combine two complementary approaches for the functional annotation of our 10 phage genomes.

1. Comparison of the predicted gene products with the newly published PHROG database comprising protein sequences exclusively from viruses infecting bacteria and archaea as well as their prophages. In PHROG, viral proteins are clustered in protein orthologs families called “phrog” built on remote homology detection and functionally annotated.
2. Comparison of the predicted gene products with protein structures from the pdb70 database at the Protein Data Bank.

We, thus, first compared all predicted gene products for each phage genome with the PHROG database. Table S5 recapitulates the best hit in PHROG for each predicted ORF for each genome, as well as the comparison metrics. We then used the comparison metrics and the genomic environment to perform a manual curation of the predicted ORFs. In total, 59 ORFs were thus manually curated (highlighted in red in Tables S2 and S5), ranging from 1 curated ORF per genome (Salmonella phage SeF2 and SeF6b) to 10 curated ORFs per genome (Salmonella phage SeAO1). Of these, 89.8% of curated ORFs (53 ORFs) were predicted by Phanotate, suggesting that this gene caller designed for phage genomes tends to overpredict ORFs. For the remaining ORFs annotated “unknown function” with a weak affiliation to a phrog family, we replace the phrog number with “singleton”, meaning that this ORF does not have (yet) any ortholog in the PHROG database. This can be either a false positive ORF or a truly novel function.

For gene products still annotated with “unknown function” after comparison with the PHROG database, we tried to improve the annotation by comparison with the pdb70 database of the Protein Data Bank. The best hits for each gene product of each genome are listed in Table S6. After manual inspection of the results and the comparison metrics, we could propose a functional annotation for 24 phrog families annotated “unknown function”. These propositions are summarized in Table 4.

The proposed annotations cover 1120 protein sequences included in the PHROG database, 352 of which are structural proteins, and 196 participate in gene regulation according to our proposed annotations. Of note, phrog_6201 is predicted as a CI-like repressor, one of the usual markers of the temperate bacteriophages. This phrog contains 20 protein sequences from 20 distinct phages that are, thus, likely to be temperate phages. Of note, Salmonella phage Salfasec13b Gp_057 protein belongs to the phrog_6201, hinting

at a temperate lifestyle for this bacteriophage. This example illustrates that one can gain valuable functional insights into an entire set of bacteriophages by improving the functional annotation of a single phrog. Table 5 summarizes the final syntactic and functional annotation for each phage. Detailed functional annotations for each genome can be found in Table S7.

Table 4. Suggested improvement to the functional annotation of 24 phrog families currently annotated “unknown function” in the PHROG database. * Functional categories defined in PHROG.

Phrog	Sequences #	Proposed Phrog Annotation	Functional Category *
phrog_4858	27	head-tail_connector_protein	connector
phrog_12508	8	endopeptidase_domain-containing_protein	lysis
phrog_600	231	lysosyme	
phrog_3838	37	receptor_binding_protein	tail
phrog_1030	144	tail_needle_knob	
phrog_3833	37	base_plate_wedge_protein	
phrog_2448	60	carbohydrate-binding_protein	
phrog_3018	47	head_decoration_protein	head and packaging
phrog_33262	2	phosphoadenosine_phosphosulfate_reductase	moron, auxiliary metabolic gene and host takeover
phrog_1203	125	phage_shock_protein_A	
phrog_3614	40	complement_C1Q-like_protein	
phrog_4435	31	ryanodine_receptor	
phrog_4049	34	ABC-	
phrog_11528	9	type_bacteriocin_transporter_peptidase_domain	transcription regulation
phrog_2426	61	threonine_deaminase	
phrog_4600	29	magnesium_chelataase	
phrog_16119	6	DNA-binding_protein	
phrog_7401	16	transcription_initiation_factor	
phrog_3786	38	DNA-binding protein	transcription regulation
phrog_1934	79	transcription_initiation_factor	
phrog_12574	8	DNA-binding_protein	
phrog_6201	20	Anti-TRAP_protein	
phrog_37032	2	CI-like_repressor	other
phrog_4700	29	pentapeptide_repeat_protein	
phrog_4700	29	ADP-ribose_transferase	DNA, RNA, and nucleotide metabolism

Number of orthologous protein sequences in the phrog cluster. * Functional categories defined in PHROG.

Table 5. Summary of the final syntactic and functional annotations of the 10 phage genomes. “Unknown function” represents the percentage of predicted ORFs annotated “unknown function”. “Singleton” represents the percentage of predicted ORFs that do not belong to any phrog family.

Strain	Size (bp)	ORFs	tRNAs	Unknown Function #	Singleton &	Accession
Salmonella phage SeAO1	157,543	233	Asn, Gln, Met, Ser, Tyr, Val	57.5%	10.3%	SAMEA6862452
Salmonella phage SeEM1	159,312	236	Asn, Met, Pro, Ser, Tyr	57.2%	11.0%	SAMEA6862907
Salmonella phage SeEM2	156,948	233	Asn, Met, Pro, Ser, Tyr	55.4%	11.1%	SAMEA6862908
Salmonella phage SeEM4	157,810	231	Asn, Met, Pro, Ser, Val	55.0%	10.8%	SAMEA6862910
Salmonella phage SeF1	42,784	74	None	37.8%	8.1%	ON809761
Salmonella phage SeF2	59,084	86	None	51.2%	12.8%	ON809762
Salmonella phage SeF3a	157,219	236	Asn, Ile, Met, Tyr	55.5%	10.2%	ON809763
Salmonella phage SeF6a	15,7240	233	Asn, Ile, Pro, Ser, Tyr	53.6%	9.0%	ON809764
Salmonella phage SeF6b	59,519	84	None	50.0%	11.9%	ON809755
Salmonella phage Salfasec13b	40,455	74	Asn	41.9%	12.2%	ON809756

Percentage of predicted ORFs annotated “unknown function”. & Percentage of predicted ORFs that do not belong to any phrog family.

The files for submission were built using the PHROG annotation for each ORF; when pertinent, the predicted functions derived from the PDB comparisons were added together with its corresponding PDB entry. As the PHROG database is not yet recognized by the International Nucleotide Sequence Database Collaboration (INSDC), it was not possible to use the “db_xref” flag for the PHROG annotations in the GenBank or ENA file. As a consequence, we reported the phrog number and the corresponding functional annotation for each gene as a “note” in the GenBank or ENA file. The 10 bacteriophage genomes were submitted to GenBank (Salmonella phage SeF1, SeF2, SeF3a, SeF6a, SeF6b, Salfasec13b, BioProject PRJNA767534) or ENA (SeAO1, SeEM1, SeEM2, SeEM4, BioProject PRJEB37792) under the accession numbers listed in Table 5.

2.6. Phage Lifestyle

Consensually, the development of phage therapy strategies relies on strictly virulent bacteriophages to formulate cocktails. The exclusion of temperate phages aims to avoid lysogenization that (1) would not immediately kill the targeted cell, (2) could enable the acquisition of virulent genes or other genes, increasing the fitness of the lysogens. In a somehow iconoclast view, Monteiro et al. [46] recently argued that temperate phages could be interesting as they are abundantly present in bacterial genomes, ready to be used and eventually engineered, although current regulations towards genetically modified organisms do not yet favor such approaches. Hence, the prediction of phage lifestyle prior to therapeutic cocktail formulation still remains a prerequisite. We scanned each phage proteome for predicted functions related to the bacteriophage lifestyles (virulent or temperate). When no integrase together with other functions generally associated with temperate phages, such as a Recombination Directionality Factor (RDF, also called excisionase) or CI-like repressors, were found, we hypothesized a “virulent” lifestyle. Such was the case for all phages except Salmonella phage Salfasec13b.

We believe that Salmonella phage Salfasec13b (*Lederbergvirus* genus) is a temperate phage as it contains many functions associated with temperate phages, such as an integrase of the tyrosine recombinase family (Gp_026, phrog_216), a CIII anti-termination protein (Gp_041, phrog_550), a CI-like repressor (Gp_057, phrog_6201), and a CII-like regulator (Gp_060, phrog_725). However, we could not detect a Recombination Directionality Factor (RDF or excisionase) required for prophage excision. RDFs are difficult to predict as they are small proteins sharing very few sequence homologies (66 aa for TorI, the RDF for the defective prophage KpIE1 in *E. coli* K12 [47,48]). It is, thus, probable that Salmonella phage Salfasec13b RDF is located among the small ORFs of unknown function. RDF genes are commonly found in the vicinity of their cognate integrase genes. Of note, in Salmonella phage Salfasec13b, the integrase gene (*gp_026*) is surrounded by predicted ORFs encoding a small protein of unknown function (Gp_027, 56 aa) compatible with the size of an RDF.

Salmonella phage SeF1 (*Jerseyvirus* genus) is an interesting case as it contains a predicted RDF (Gp_053, 76 aa, phrog_66) whose modeled structure we found matches very well with TorI (PDB 1Z4H), the RDF of the defective prophage KpIE1 mentioned above. Nevertheless, no ORF coding for an integrase could be predicted, and other key functions associated with temperate phages are missing. The same findings were made for the closest annotated sequence homolog found with Blastn in the nt database, i.e., Salmonella phage S101 (Table 2). The PHROG database includes 22 *Jerseyvirus* genomes, 20 of them coding for RDF belonging to the phrog_66, making it a marker of this genus, although *Jerseyvirus* seem to be strictly virulent phages. Salmonella phage SeF1 and its relatives might have been temperate phages that have lost their integrase gene, or the RDF encoded in their genome has been acquired by horizontal gene transfer and may serve to induce resident prophage(s) in the host. The latter hypothesis is worth investigating, as induction of resident prophages is likely to ensure efficient lysis of the host.

2.7. Genetic Mining of 10 Newly Isolated *Salmonella* Phage Genomes in the Light of Phage Therapy

The primary aim of our isolation of new *Salmonella*-targeting phages is the development of phage therapy against *Salmonella* species, and we were particularly interested in two major kinds of protein functions.

1. Those that could give a selective advantage for host infection, successful virion productions, and cell lysis (mostly features allowing evasion from cellular defenses).
2. Those that could interfere with the formulation of a bacteriophage cocktail, such as functions allowing lysogeny or super immunity against other phage infections.

We first reviewed these two categories of functions by the bacteriophage genus before focusing on the prediction of anti-CRISPR proteins, which belong to the first category of proteins likely to confer a selective advantage.

2.7.1. *Jerseyvirus*: *Salmonella* Phage SeF1

We identified a DarB-like Type-I RM anti-restriction protein (Gp_039, phrog_10089). In bacteriophage P1 infecting *E. coli*, several anti-restriction proteins, including DarB, are packaged during virion assembly and released upon infection into the host cytoplasm, ready to act against host RM systems [49]. DarB alone is required for protection against type-I EcoB and EcoK restriction. The Gp_039 annotation (DarB-like anti-restriction) is sound because this protein belongs to phrog_10089, predicted to be highly similar (probability 100%, Evalue 3E-210) to phrog_1685, which does contain Bacteriophage P1 DarB. It, thus, seems that *Salmonella* phage SeF1 encodes at least a DarB-like anti-restriction protein. Of note, *Salmonella enterica* Typhimurium codes for SB, an enzyme similar to *E. coli* EcoA, EcoB, and EcoK type-I restriction enzymes [50]. The *Salmonella* phage SeF1 genome contains six restriction sites recognized by SB.

Gp_023 is predicted as an anti-repressor of the Ant type (phrog_130). In *Salmonella* prophages such as Gifsy-1 and Gifsy-2, prophage induction is repressed by the binding of the repressor Rep on the upstream promoter genes, allowing prophage induction [51]. The production of Ant derepresses these genes by titrating the Rep protein, thus triggering prophage(s) induction and eventually cell lysis. Gp_023 expression is then susceptible to inducing resident prophage(s) in the infected host and may then help host takeover. Together with the RDF mentioned previously, this is the second protein indicative of a temperate lifestyle, although *Salmonella* phage SeF1 lacks the other functions required for lysogenization.

Gp_027 is predicted as an immunity-to-superinfection protein (phrog_1039). In *E. coli* phage T4, the immunity protein Imm blocks the injection of T4 DNA, preventing superinfection by the same phage or another phage targeting the same bacterium and using the same infection mechanisms. Bacteriophage T4 Imm belongs to the same phrog as Gp_027, strengthening the proposed annotation for Gp_027. The production of an Imm-like protein in an infected cell will alter co-infection by another virion, either from the same strain or from a bacteriophage that resorts to the same DNA injection mechanism.

Salmonella phage SeF1 seems to harbor an extensive arsenal for host lysis with two Rz-like spanins (Gp_009 and Gp_010, phrog_2536 and phrog_5672, respectively), one endopeptidase predicted from structural comparison with the pdb70 (Gp_034, phrog_12508 currently annotated “unknown function” in the PHROG database), two holins (Gp_063 and Gp_064, phrog_2359 and phrog_2508, respectively) and one endolysin (Gp_065, phrog_33083).

2.7.2. *Chivirus*: *Salmonella* Phage SeF2 and SeF6b

Salmonella phage Chi and other Chi-like viruses are flagellotropic phages requiring a motile flagellum as its primary receptor; it is then very likely that *Salmonella* phage SeF2 and SeF6b share the same requirement. Both *Salmonella* phage SeF2 and SeF6b phages contain 3 gene products with predicted carbohydrate-binding domains that we predicted from structural comparison with the pdb70 (phrog_2448 currently annotated “unknown function” in the PHROG database) encoded on the same strand downstream

of the tail genes module, suggesting that these proteins may be part of the tail and the host recognition apparatus. These genes encoding carbohydrate-binding domains are shared among 36 other *Chivirus* present in the PHROG database. Sugars are components of the LPS, and it has also been shown that *S. enterica* Typhimurium flagella contain at least 16 different sugars [52]. The flagellotropic *Salmonella* phage Chi and its relatives may use these domains to recognize and bind the bacterial flagellum and its sugar moieties. It has been reported that *Salmonella* phage Chi can infect both *Salmonella* spp. and *E. coli* hosts and thus has been proposed as a good candidate for the development of phage-based applications (pathogens detection, remediation, phage therapy) [53]. Inclusion in a therapeutical cocktail of bacteriophages such as *Chivirus* that may target receptors other than the LPS would be a good option to avoid cross-resistance.

Both phages encode a predicted DNA methyltransferase (phrog_111) homologous to the DNA N-6-adenine methyltransferase according to PFAM prediction (PF05869) attached to this phrog. This enzyme may serve to modify phage DNA during replication, allowing the phages to escape host restriction–modification systems.

2.7.3. *Lederbergvirus*: *Salmonella* Phage Salfasec13b

As discussed above, *Salmonella* phage Salfasec13b is a temperate phage and a *Lederbergvirus* (as the archetypal *Salmonella* phage P22). This phage genome contains all the genes found in P22 that are important to promote lysogeny or confer a selective advantage to the host including *gtrA* (Gp_022, phrog_2224), *gtrB* (Gp_021, phrog_34335), and *gtrC* (Gp_019, phrog_4829). Together, the GtrABC complex modifies the host O-antigen at the host cell surface [54]. Changing the LPS prevents infection by other phages using it as a primary receptor. In the light of phage therapy, this is not the desired outcome.

2.7.4. *Kuttervirus*: *Salmonella* Phage SeAO1, SeEM1, SeEM2, SeEM4, SeF3a, SeF6a

We first pinpoint protein functions that seem to be a hallmark of the *Kuttervirus* genus as they are found both in our 6 *Kuttervirus* phage genomes and in almost all the 15 *Kuttervirus* genomes included in the PHROG database.

An interesting feature is the gene product belonging to phrog_1510 (superinfection exclusion) that includes the Gp17 protein of *Salmonella* phage P22. In the classical literature, Gp17 is described as a protein necessary for P22 to counteract a superinfection exclusion system encoded in the Fels-2 prophage found in many *S. enterica* Typhimurium strains [55]. A sensible hypothesis is that *Kuttervirus* phages have somehow acquired a P22 *gp17*-like gene that allows them to successfully infect and propagate in lysogenic *Salmonella* strains harboring Fels-2-like superinfection exclusion systems. This is an interesting feature for a phage cocktail, as *Salmonella* spp. harbor many diverse prophages [56].

Another hallmark is phrog_600 (unknown function), which we could predict as a lysozyme-like protein by comparison with the pdb70. This adds to the predicted Rz-spanin and endolysin in the arsenal of proteins, ensuring host lysis.

The RIIA (phrog_3661, RIIA lysis inhibition) and RIIB (phrog_3803, RIIB lysis inhibition) are also encoded in most *Kuttervirus* phages. The phage proteins RIIA and RIIB enable bacteriophage T4 to resist the Rex restriction system encoded by *E. coli* Lambda lysogens. Rex is an abortive infection defense system and relies on the action of two proteins, RexA and RexB [57]. However, the molecular mechanisms behind Rex still remain elusive. RIIA and RIIB could confer an advantage to *Kuttervirus* phages when infecting lysogens encoding a Rex-like system.

We were intrigued by the presence in all our six *Kuttervirus* genomes of a predicted second terminase small subunit (phrog_800) situated far away from the traditional pair of consecutive genes encoding the small and the large terminase subunits. To the best of our knowledge, this is unheard of in phage genomes. A closer inspection of phrog_800 revealed a probable misannotation. Indeed, by checking the list of PDB hits (Table S6), we found that phrog_800 proteins are predicted to be deoxynucleoside kinase proteins. In our genomes, the phrog_800-encoding gene is followed by genes encoding a thymidiate

synthase, a deoxynucleoside monophosphate kinase, and finally, a nucleoside triphosphate pyrophosphohydrolase. This genomic environment strongly suggests a gene cluster involved in DNA metabolism and strengthens our proposed new annotation for phrog_800 as a deoxynucleoside_kinase.

Finally, an interesting feature was found only in *Salmonella* phage SeF3a. Our comparison with the pdb70 strongly suggests that Gp_125 (phrog_1203, unknown function) is homologous to PspA, the Phage Shock Protein A. Four of the fifteen *Kuttervirus* phage genomes included in the PHROG database harbor such a gene. *pspA* belongs to the *psp* operon found in some bacterial genomes, noticeably in *E. coli* and *Salmonella* species. This system prevents the dissipation of the proton-motive force (PMF) triggered by various stresses, including the infection by filamentous phages or the mislocalization of secretins [58]. PspA plays a role in the regulation of the phage-shock-response *psp* operon but also associates with the inner membrane to prevent the loss of PMF, although the molecular mechanisms still remain unclear. Several Abi systems, such as Rex in Lambda lysogens mentioned above or AbiZ in *Lactococcus lactis* [13], lead to cell death or cellular growth arrest through membrane permeabilization and dissipation of the PMF or loss of ATP. A reasonable hypothesis could be that the phage-encoded PspA acts as a novel anti-Abi system, preventing cell suicide. This hypothesis is worth further investigation. This feature singles out *Salmonella* phage SeF3a from the other *Kuttervirus* phages isolated in this study, with a putative additional anti-host defense to its arsenal.

2.7.5. Prediction of Anti-CRISPR Proteins (Acr)

Salmonella enterica species possess a type I-E CRISPR-Cas system with a *cas* operon and two CRISPR arrays [59]. This adaptative immune defense system may interfere with phage infection; hence, searching for phage-encoded anti-CRISPR (Acr) proteins is interesting in the light of phage therapy. We performed Acr encoding genes prediction with AcrHub. Among all the Acr predicted in our 10 phage genomes, we retained the most probable predictions, summarized in Table 6.

Table 6. Predicted Acr proteins in the 10 phage genomes studied.

Salmonella Phage	Genus	Phrog	Protein	Size (aa)	PaCRISPR Score	AcRanker Score	AcrHub ID #	CRISPR Type &
SeAO1	<i>Kuttervirus</i>	4097	Gp_007	56	0.62	−0.75	Acr00210	II-A
		519	Gp_015	71	0.54	−2.08	Acr00037	I-D
		5934	Gp_049	105	0.69	−4.64	Acr00193	II-A
		3594	Gp_088	62	0.58	−3.57	Acr00054	I-D
SeEM1	<i>Kuttervirus</i>	4097	Gp_008	56	0.62	−0.75	Acr00210	II-A
		3594	Gp_080	62	0.55	−2.90	Acr00212	II-A
SeEM2	<i>Kuttervirus</i>	4097	Gp_008	56	0.62	−0.75	Acr00210	II-A
		519	Gp_016	71	0.54	−2.08	Acr00037	I-D
SeEM4	<i>Kuttervirus</i>	4097	Gp_007	56	0.62	−0.75	Acr00210	II-A
		519	Gp_015	71	0.54	−2.08	Acr00037	I-D
SeF3a	<i>Kuttervirus</i>	4097	Gp_008	56	0.62	−0.74	Acr00210	II-A
		519	Gp_014	78	0.48 *	−1.17	Acr00037	I-D
		3594	Gp_082	62	0.54	−2.77	Acr00212	II-A
SeF6a	<i>Kuttervirus</i>	4097	Gp_007	56	0.62	−0.74	Acr00210	II-A
		5934	Gp_042	110	0.68	−3.85	Acr00211	II-A
SeF1	<i>Jerseyvirus</i>	13,851	Gp_038	73	0.58	−4.11	Acr00209	II-A
		18,511	Gp_057	61	0.53	−1.96	Acr00311	V-A
		6094	Gp_059	173	0.83	−3.57	Acr00061	I-E
		2180	Gp_068	61	0.53	−4.87	Acr00128	I-F
		1402	Gp_072	69	0.64	−3.92	Acr00070	I-F
		Singleton	Gp_073	62	0.70	−4.91	Acr00264	II-C
Se_F2	<i>Chivirus</i>	7965	Gp_051	120	0.84	−4.46	Acr00254	II-C
		7734	Gp_052	111	0.54	−4.86	Acr00029	I-D
		4445	Gp_057	85	0.77	−5.17	Acr00159	II-A
Se_F6b	<i>Chivirus</i>	7965	Gp_051	120	0.84	−4.21	Acr00275	II-C
		7734	Gp_052	111	0.53	−4.80	Acr00029	I-D
		4445	Gp_058	86	0.69	−4.94	Acr00198	II-A
Salfasec_13b	<i>Lederbergvirus</i>	11,224	Gp_044	41	0.65	−4.56	Acr00064	I-E

Closest homolog in the AcrHub database. & Inhibited CRISPR type. * PaCRISPR score below 0.5, but this predicted Acr belongs to the phrog_519 that is found in the other five *Kuttervirus* genomes with reliable scores.

In all phage genomes studied here, we could retain at least one Acr encoding gene prediction. *Salmonella* phage SeF1 harbors the highest number of predicted Acr ($n = 6$). These predictions are generally consistent for bacteriophages belonging to the same genus. For *Kuttervirus* genomes, phrog_4097 is found in all six genomes studied, and phrog_519 in five out of six genomes. For *Chivirus* genomes, *Salmonella* phage SeF2 and SeF6b genomes code for the same three predicted Acr.

Interestingly, the transposable *Pseudomonas aeruginosa* Mu-like phage genomes contain an *acr* locus comprising two to three genes located among the genes involved in the phage head assembly [37]. In these phages, the *acr* locus is bracketed by genes involved in phage head morphogenesis, usually upstream of a protease/scaffold gene. The lower panel in Figure 5 describes the genetic context of AcrIE3, a type-IE Acr protein from *P. aeruginosa* phage DMS3 that has been functionally investigated [60]. In our six *Kuttervirus* phages, we found one small protein (56 aa) belonging to phrog_4097 (unknown function) predicted with good scores by both PaCRISPR and AcRanker as a type II-A anti-CRISPR protein. The gene encoding this protein in the *Kuttervirus* genomes is situated between the portal protein and head scaffolding protein-encoding genes, reminiscent of the genetic context of AcrIE3 (Figure 5, upper panel).

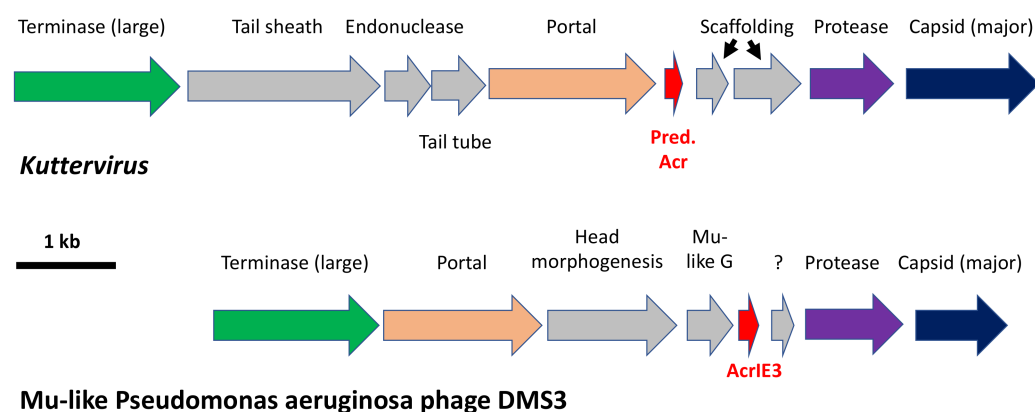


Figure 5. Comparison of the genomic surroundings of bona fide Acr-encoding genes (*acrIF2*) in Mu-like *Pseudomonas aeruginosa* phage DMS3 (accession NC_008717) and the predicted Acr-encoding ORF in our 6 *Kuttervirus*. For clarity, genes are drawn to scale but not the intergenic regions.

One can also find a similar gene coding for a phrog_4097 protein in the same genomic context in 14 of the 15 *Kuttervirus* genomes found in the PHROG database, making this predicted Acr another hallmark of *Kuttervirus* genomes. This predicted Acr is, thus, a serious candidate for anti-CRISPR functional studies and an interesting feature for therapeutical cocktails.

Salmonella phage SeF1 Gp_057 exhibits high scores for both PaCRISPR and AcRanker predictions. Although we could not link gp_057 genomic context to known contexts of experimentally validated Acr, we believe Gp_057 is also a serious candidate for anti-CRISPR functional studies.

3. Materials and Methods

3.1. Phage Lysates and Bacterial Strain

Se_AO1, Se_EM1, Se_EM2, Se_EM3, Se_EM4, Se_F1, Se_F2, Se_F3, Se_F6, Salfasec_9, Salfasec_10, Salfasec_11, and Salfasec_13 culture lysates were kindly provided by Dr. Olivenza (Departamento de Genética, Facultad de Biología, Universidad de Sevilla, 41012 Sevilla, Spain). All phages were isolated in the Sevilla area (Spain) from wastewater or freshwater pounds. When required, *Salmonella enterica* subsp. *enterica* serovar Typhimurium ATCC 14028S (ATCC 14028S) was used to propagate and titrate bacteriophages. Cells were grown aerobically in either Lysogeny Broth (LB) at 37 °C under shaking (180 rpm) for liquid cultures or on LB-1% agar plates incubated at 37 °C.

3.2. Phage Propagation, Purification, and Titration

To propagate phages from culture lysates, 200 mL Erlenmeyer flasks containing 25 mL of LB were inoculated with an overnight (ON) culture of ATCC 14028S at an optical density measured at 600 nm (OD_{600nm}) of 0.04. Cells were grown until OD_{600nm} reached 0.4, then inoculated with 100 μ L of 0.22 μ m-filtered and chloroform-treated culture lysate. The culture was further incubated for about 4 h. Cells were lysed by the addition of 10% chloroform (vortexed for 15 s and incubated for 5 min at room temperature, repeated twice). The aqueous supernatant (phage lysate) was recovered after centrifugation ($7500\times g$, 15 min) to pellet the cell debris and separate the aqueous phase from the solvent. The phage lysate was then 0.22 μ m-filtered and stored at 4 °C. Phage titration by double agar overlay plaque assay was performed according to Kropinski et al. [61]. Phage titers are expressed in Plaque Forming Unit per mL (PFU/mL). For further phage purification and concentration, all centrifugation steps were done at 4 °C, $20,800\times g$ for 1 h, and ice-cold buffers were used. A total of 4 mL of phage lysate (titer between 1.10^{10} and 3.10^{11} PFU/mL) were centrifuged, and the pellet was resuspended in 2 mL of 0.22 μ m-filtered Phage Buffer (PB: 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 25 mM $MgCl_2$, 1 mM $CaCl_2$). Phages were pelleted once again and resuspended in 200 μ L of PB. Purified phage suspensions were stored at 4 °C.

3.3. Phage Morphological Characterization by TEM

For Transmission Electronic Microscopy (TEM) exploration, the final phage pellet resuspension was performed in 2 mL of 0.22 μ m-filtered TEM buffer (0.1 M ammonium acetate pH 7). Phages were pelleted once again and resuspended in 20 μ L of TEM buffer. Formvar/carbon-coated copper grids were prepared at the Institut de la Méditerranée (IMM) Microscopy facility. A total of 5 μ L of the phage solution was pipetted onto the grid surface and allowed to sediment for 3 min at RT. Excess liquid was then blotted and grids negatively stained according to Ackermann's protocol with 2% uranyl acetate. Observations were made on an FEI Tecnai G2 20 TWIN (200KV), laB6, Gatan Oneview 4k \times 4k CMOS transmission electron microscope. Images were visualized using FIJI [62].

3.4. Phage DNA Purification

To remove non-encapsidated nucleic acids, 20 μ L of DNase I (6 mg/mL, EUROMEDEX, Souffelweyersheim, France), 2 μ L of RNase A (4 mg/mL, Promega), and 4 μ L of DpnI restriction enzyme (20,000 U/mL, New England Biolabs, Ipswich, MA, USA) were added to 200 μ L of purified phages in PB. The sample was then incubated for 1 h at 37 °C. DNase I was inactivated by incubating the sample for 20 min at 65 °C under shaking. Phage DNA was released from the capsid by adding 20 μ L of 10% SDS and 20 μ L of proteinase K (50 μ g/mL, EPICENTRE, LGC, Teddington, UK) and incubating the sample for 1 h at 56 °C. The sample volume was completed with PB buffer q.s. 600 μ L. Phage DNA was extracted by adding 600 μ L phenol/chloroform/isoamyl alcohol (25:24:1, Sigma-Aldrich, St. Louis, MO, USA) and then vortexing the sample for 30 s. The upper aqueous phase containing the nucleic acids was recovered after centrifugation ($10,000\times g$, 10 min, 4 °C). The extraction was repeated twice. The aqueous phase-containing DNA was then ethanol-precipitated. The DNA pellet was finally resuspended in 20 μ L DNase/RNase-free water and stored at −20 °C. Phage dsDNA was quantified with a QubitTM fluorometer in combination with the QubitTM dsDNA HS Assay kit (Invitrogen).

Phage genomic DNA purified from Se_F1, Se_F2, Se_F3, Se_F6, and Salfasec_11 phage lysates was mechanically fragmented in 50 μ L microtubes using a Covaris M220 sonifier with the following parameters: peak power 75W, duty factor 10%, Cycles/Burst 200.

Phage genomic DNA purified from Se_AO1, Se_ML1, Se_EM1, Se_EM2, Se_EM3, Se_EM4, Salfasec_9, Salfasec_10, and Salfasec_13 phage lysates was enzymatically fragmented by the tagmentase technology from the NEXTERA XT kit (Illumina[®], San Diego, CA, USA) according to the manufacturer's protocol.

DNA libraries for high throughput sequencing were prepared from the fragmented DNA at the IMM Transcriptomic and Genomic facility with the NEBNext® Ultra™ II DNA Library Prep kit for Illumina® (New England Biolabs, Ipswich, MA, USA) for Se_F1, Se_F2, Se_F3, Se_F6, Salfasec_11, and with the NEXTERA XT kit (Illumina®, San Diego, CA, USA) for Se_AO1, Se_ML1, Se_EM1, Se_EM2, Se_EM3, Se_EM4 according to the manufacturer's protocols.

Salfasec_9, Salfasec_10, and Salfasec_13 DNA libraries preparation with the NEXTERA XT kit (Illumina®, San Diego, CA, USA) was subcontracted to AllGenetics (A Coruña, Spain).

3.5. High Throughput DNA Sequencing

Prior to sequencing, Se_AO1, Se_EM1, Se_EM2, Se_EM3, Se_EM4, Se_ML1, Se_F1, Se_F2, Se_F3, Se_F6, and Salfasec_11 DNA libraries were quantified with a Qubit™ ds-DNA HS Assay kit (Invitrogen, Waltham, MA, USA) and their size distribution profiles recorded with the TapeStation 4200 System (Agilent) in combination with the D5000 DNA ScreenTape System (Agilent, Santa Clara, CA, USA). Libraries were then diluted at 4 nM in the appropriate buffer. Paired-end (2 × 150 bp) DNA sequencing was performed on the MiSeq sequencer (Illumina®, San Diego, CA, USA) hosted at the IMM Transcriptomic and Genomic facility with a MiSeq v2 (300-cycles) flow cell (Illumina®, San Diego, CA, USA) according to the manufacturer's protocol.

Salfasec_9, Salfasec_10, and Salfasec_13 DNA libraries sequencing was subcontracted to AllGenetics (Coruña, Spain).

Raw sequencing reads (FASTQ files trimmed from their Illumina adaptors) for each BioSample were submitted to the NCBI Sequence Read Archive under the BioProject accession number PRJNA767534. Raw sequencing reads quality was then improved with Trimmomatic [63] using the following parameters: SLIDINGWINDOW:4:25 MINLEN:75 for Se_AO1, Se_ML1, Se_EM1, Se_EM2, Se_EM3, and Se_EM4 or ILLUMINACLIP TruSeq-SE.fasta SLIDINGWINDOW:4:28 MINLEN:75 for all other phages. Only trimmed paired-reads were used for de novo genome assembly.

3.6. Genome de Novo Assembly

Genome de novo assembly was performed with SPAdes 3.14.1 with default parameters. Table A1 (Appendix A) summarizes the size of the 18 meaningful contig(s) obtained for the 14 BioSamples together with their mean genome coverage. By “meaningful”, we mean contigs with significant sequence coverage and length compatible with a phage genome (50 kb average size for dsDNA phages) and contigs that are not contaminants. Contaminants are usually small contigs with low sequencing coverage and whose DNA sequences match with unrelated organisms (Blastn analyses). For most contigs, we could identify identical 77 bp-DNA sequences at each contig extremity due to the SPAdes assembly algorithm. We observed this with circularly permuted genomes. We later removed the right-most copy to obtain the final contigs we considered as circular for downstream syntactic annotation, although we are well aware that the encapsidated biological DNA molecule is linear. Indeed, we found out that some genes were spanning the contig extremities. Such was the case for Se_F2 and Se_F6b genomes.

3.7. Genomes Clustering

We used VIRIDIC (Available online: <http://rhea.icbm.uni-oldenburg.de/VIRIDIC/>) (accessed on 27 July 2022) to cluster the 18 contigs with 95% and 70% for the species and genus threshold, respectively, and the following BLASTN parameters: ‘-word_size 7-reward 2-penalty-3-gapopen 5-gapextend 2’. Genomes from the *Kuttervirus*, *Chivirus*, *Jerseyvirus*, and *Lederbergvirus* sanctioned by the ICTV (Master Species List 2021_v2) were downloaded from NCBI.

3.8. Prediction of DNA Packaging

We used PhageTerm [64] to predict putative phage genome termini as well as the DNA packaging mode. PhageTerm is based on the statistical analysis of the sequence coverage after mapping the short sequencing reads onto the assembled contig. For a statistically sound result, PhageTerm requires a minimum sequence coverage of 50. Obviously, PhageTerm requires that the genome extremities are sequenced, and this depends on the technology used to construct the DNA libraries for sequencing. Genome extremities can be recovered when the phage DNA has been mechanically fragmented, as described above. This was the case for Se_F1, Se_F2, Se_F6, and Salfasec_11. When the tagmentation has been used to construct the libraries (Nextera XT kit from Illumina®, San Diego, CA, USA), one cannot recover the extremities of a linear dsDNA, which precludes termini identification by PhageTerm when DNA packaging is initiated by the terminase at fixed positions (*cos* or *pac* sites) and ends at a fixed position (*cos* site). Tagmentation was used for Se_AO1, Se_ML1, Se_EM1, Se_EM2, Se_EM3, Se_EM4, Salfasec_9, Salfasec_10 and Salfasec_13.

3.9. Syntactic Annotation

For tRNA prediction, we used ARAGORN [65]. For Open Reading Frames (ORF) prediction, we tested five different gene callers. Indeed, many gene callers are freely available to predict ORFs from nucleic acid sequences, but they rely on different algorithms, leading to variable predictions. We selected four of them optimized for bacteria (AMIGene, Glimmer, MetaGeneAnnotator, and Prodigal) and one optimized for phage genomes (Phanotate) [66–70].

Locus tag names were built with the same convention for each phage: an alphanumeric string defining the phage name coupled with an alphanumeric string defining the gene product numbering. For instance, Se_AO1_GP_001 refers to the first ORF predicted for *Salmonella* phage SeAO1. When present, tRNAs were labeled following the same convention: Se_AO1_tRNA1 refers to the first tRNA predicted for the *Salmonella* phage SeAO1 genome. To ease further genomic comparisons, we decided to start the ORFs numbering with the terminase small subunit-encoding ORF as *gp_001*. All DNA sequences were flipped and/or rotated accordingly.

3.10. Taxonomic Classification

For taxonomic classification, we used vContact2 0.9.19 (with the ProkaryoticViral-RefSeq201 database accessed on 27 September 2021) and VICTOR (Virus Classification and Tree Building Online Resource with the amino acid option selected, Available online: <https://ggdc.dsmz.de/victor.php>) (accessed on 15 November 2021) Network visualization was done with Cytoscape [71].

3.11. Detection of Protein Sequence Remote Homologies

We generated Hidden Markov Model (HMM) profiles for each predicted protein sequence of each phage with HHblits from the UniRef30_2020_06 database in order to detect sequence remote homologies [72,73].

3.12. PHROG-Based Functional Annotation

For functional annotation, we used the newly published PHROG database (accessed on 27 September 2021) dedicated to viral proteins. In this database, viral proteins are distributed among families called “phrog”; each phrog represents a cluster of viral proteins orthologs built using remote homology detection by HMM profile–profile comparisons. At the date of its publication, the PHROG database contains 17,473 (pro)viruses of prokaryotes or archaea for a total of 938,864 proteins. A total of 38,880 clusters were defined, and manual inspection led to the annotation of 5108 phrog clusters, representing 50.6% of the total protein dataset. Following the procedure available on the PHROG website (<https://phrogs.lmge.uca.fr/READMORE.php>), we used HH-suite [74] to compare each predicted gene product from the phages studied here with the PHROG database and ascribe a phrog

and its annotation whenever possible. When several hits were found, we kept the best one for the considered protein. The corresponding phrog number and its functional annotation were then transferred to our newly predicted protein. When no hit with the PHROG database was found, or the affiliation to an existing phrog was too distant (probability < 80%, $\text{Evalue} > 1 \times 10^{-4}$), the ORF was annotated as a singleton of unknown function according to the PHROG guidelines. The entire list of the best hit for each ORF for each phage before manual curation is supplied in Table S5.

3.13. PDB-Based Functional Annotation

In order to improve the PHROG-based functional annotation for protein still annotated “unknown function”, we performed with HHblits a comparison of each predicted protein sequence HMM profile with the Protein Data Bank pdb70 database (accessed on 28 September 2021) to detect homologies with known protein structures. We kept the first hit for each protein sequence for further consideration when the probability was greater than 80% and the Evalue less than 1×10^{-3} . When these criteria were met, we also manually checked the prediction coverage to avoid transferring a PDB annotation based on a predicted structural similarity between a small portion (a domain or a sub-domain) of the PDB hit and the phage protein sequences. After all these considerations, we manually checked for phage ORF ascribed to a phrog of unknown function if a reliable structural prediction from the pdb70 could be used instead. If so, a new annotation inferred from the structural prediction was transferred to the corresponding phrog of unknown function (see Table 4). A new annotation for this phrog cluster will be proposed to the PHROG database through a dedicated form on its website (<https://phrogs.lmge.uca.fr/suggestions.php>).

3.14. ORFs Manual Curation

Phage genes sometimes overlap or may even be entirely included in one another. Nevertheless, in some cases, these overlaps can reveal misprediction from the syntactic annotation. Thus, in an attempt to reduce false positive predictions, we manually scanned each genome to detect highly overlapping ORFs. The most common situation we encountered was two highly overlapping ORFs, with one ORF coded on one strand with a strong affiliation to a phrog family while the ORF encoded on the opposite strand had a weak affiliation to a phrog family (probability < 70%, $\text{Evalue} > 1 \times 10^{-4}$). If the latter did not show any strong homology with a known structure from the PDB (probability > 80%, $\text{Evalue} < 1 \times 10^{-3}$), we decided to discard the ORF. As an example of manual curation, for Salmonella phage SeAO1, *gp_014* and *gp_016* almost entirely overlap with *gp_015*, all 3 ORFs predicted on the same strand. According to the phrog affiliation metrics generated by HHblits (see Table S5), Gp_015 is strongly affiliated to phrog_519 (Prob. 99.5%, $\text{Evalue} 1.4 \times 10^{-18}$, 261 proteins sequences in this phrog family), whereas Gp_014 is weakly affiliated to phrog_32723 (Prob. 55%, $\text{Evalue} 1.1$, 2 proteins sequences in this phrog family) and Gp_016 is weakly affiliated to phrog_8445 (Prob. 23.5%, $\text{Evalue} 7.4$, 13 proteins sequences in this phrog family). In this example, we discarded *gp_014* and *gp_016*. In other examples, such as *gp_212*, *gp_213*, and *gp_214*, it was not possible to decide which ORF was significant based on the phrog affiliation metrics; all three ORFs, although overlapping, code for proteins strongly affiliated to a phrog family. In this case, we chose to keep all three ORFs. The last example of manual curation is *gp_222*. Based on phrog affiliation metrics, we classified Gp_222 as a “singleton” of unknown function. However, *gp_222* completely overlaps with the reliably predicted tRNA^{Val} on the same strand; we, thus, decided to discard *gp_222*. Table 5 summarizes the final gene content predicted for each phage after manual curation. This procedure was carried out for every genome included in this study.

3.15. Prediction of Anti-CRISPR Proteins (Acr)

Until very recently, searching for Acr was tedious, essentially relying on “guilt by association” with anti-CRISPR-associated (Aca) proteins containing “helix-turn-helix” motifs. However, not all Acr are associated with an Aca. Thanks to the new approaches

based on machine learning, it is now possible to predict Acr found in undescribed genomic environments. AcrHub is a platform that incorporates state-of-the-art Acr predictors and three analytical modules (similarity analysis, phylogenetic analysis, and homology network analysis) [27]. The AcrHub database contains 339 experimentally validated, and 71,728 predicted Acr proteins. Validated Acr are mostly short proteins (89% between 60 and 160 aa). We ran our phage proteomes in AcrHub on two different Acr prediction algorithms (PaCRISPR [75] and AcRanker [76]). Meaningful score thresholds are over 0.5 for PaCRISPR and over -5 for AcRanker. We then used the AcrHub “Similarity analysis” module to relate our predicted Acr with experimentally validated Acr proteins in the AcrHub database. We excluded overlong proteins (>200 aa) and those whose query cover was below 40%. When pertinent, we compared the genomic surrounding of our predicted *acr* gene with an experimentally validated *acr* gene product.

4. Conclusions

In this study, we present a genomic description of 10 phage strains representing 8 new species isolated in wastewater and freshwater ponds in the Sevilla area in Spain and infecting *S. enterica* serovar Typhimurium strain ATCC 14028S. Based on genome-wide analyses independent from prior knowledge on functional annotation, we could ascribe taxonomic classification down to the genus level. These phages all belong to the *Caudoviricetes* class (dsDNA-tailed bacteriophages with an HK97 major capsid fold) and to four different genera (Figures 3 and 4). *Kuttervirus* phages are overrepresented, with six strains identified, five of them being new species. Only Salmonella phage Salfasec13b, belonging to the *Lederbergvirus* genus, was predicted as a temperate phage.

We wish to highlight in our study two important methodological points. The first one is that by combining the results of four different gene callers, including Phanotate especially designed for phage genomes (but that nevertheless tends to over-predict ORFs), we believe we could predict as many ORFs as is reasonably possible. However, we also showed that manual inspection and curation are still needed to get rid of obvious false positive predictions. The second point concerns the improvement of functional annotation with PHROG, a database dedicated to viral proteins at large (bacterial and archaeal viruses and their prophages). We believe this is a promising approach for functional annotation as it relies on viral protein orthologs clustering based on remote homology detection. PHROG capitalizes on decades worth of past work of many research teams that have led to the experimental validations of many functions, and PHROG also benefits from the manual annotation of various experts in the field (we could propose in this study a functional annotation to 24 phrog clusters previously annotated with unknown function).

We mined our 10 annotated genomes for relevant functions in the context of future biocontrol application of *Salmonella* spp. We could, thus, identify phage-encoded proteins targeting bacterial anti-phage defenses such as Abortive infection, Restriction-Modification, and CRISPR-Cas systems. Our study also illustrates that in otherwise very similar genomes such as our six *Kuttervirus* (ANI values between 0.94 and 0.98), we could identify subtle variations in gene equipment that can single out one phage from its closest relatives. Indeed, we predicted that the Salmonella phage SeF3a genome alone codes for a new anti-Abi system based on a phage-encoded Phage Shock Protein A ortholog, potentially conferring Salmonella phage SeF3a a selective advantage over the other five viruses of the same *Kuttervirus* genus. This example emphasizes the need to extensively mine bacteriophage genomes in order to tailor therapeutic cocktails matching the targeted pathogen. Conversely, one has to obtain as much genomic information as possible about the pathogen itself in order to identify its anti-phage defenses and select appropriate phages to formulate a therapeutic cocktail that can overcome these cellular defenses. The recent publication by Tesson et al. is helpful in that respect with the description of DefenseFinder, a bioinformatic tool specially designed to systematically predict anti-phage systems in prokaryotic genomes [77].

The coevolution of bacteria and their viruses for ages beyond count is often described as an arms race that gifted both protagonists with highly diverse molecular weapons and

shields. To go beyond empiric approaches and rationally design efficient phage-based biocontrol of bacteria, we need to adopt a holistic approach to the bacteriophage/host binomen. In that respect, genomic and functional analyses of both partners are crucial to meeting the existing challenges in tackling bacterial infections.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23168917/s1>.

Author Contributions: Conceptualization, M.A. and N.G.; methodology, J.G. and N.G.; software, J.G.; validation, J.G. and N.G.; formal analysis, J.G. and N.G.; investigation, M.A. and N.G.; resources, D.R.O., M.A., J.C. and Y.D.; data curation, J.G. and N.G.; writing—original draft preparation, N.G.; writing—review and editing, M.A., N.G., J.G. and J.C.; visualization, J.G. and N.G.; supervision, N.G.; project administration, M.A.; funding acquisition, M.A. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agence Nationale de la Recherche of France, grant number ANR-16-CE12-0029, the Ministerio de Ciencia e Innovación of Spain (Agencia Estatal de Investigación), grant number BIO2016-75235-P, and the European Regional Fund, grant number PID2020-116995RB-I00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Note applicable.

Data Availability Statement: Raw sequencing reads and Genbank files for the 10 newly isolated salmonella phages studied here are openly available at NCBI under the BioProject accession numbers PRJNA767534 and PRJEB37792. Input and output files pertaining to vContact2 analyses are publicly available since 17 June 2022 at figshare (<https://doi.org/10.6084/m9.figshare.20088356>). Output files pertaining to VICTOR analyses are publicly available since 17 June 2022 at figshare (<https://doi.org/10.6084/m9.figshare.20088416> for *Kuttervirus* and <https://doi.org/10.6084/m9.figshare.20088473> for *Jersyvirus*).

Acknowledgments: We thank Artemis Kosta and Hugo Le Guenno from the Microscopy facility at the Institut de Microbiologie de la Méditerranée (IMM, FR 3479, CNRS–Aix-Marseille Université, Marseille, France) for TEM grids preparation and image acquisition. We are very thankful to Marie-Agnès Petit (INRAE, Jouy-en-Josas, France) for fruitful discussions on functional prediction and phage genome annotation, discussions we had personally but also during dedicated workshops organized by the French bacteriophage network (www.phages.fr, accessed on 17 June 2022). We thank François Enault and Eric Olo Ndela (LMGE, Université Clermont Auvergne, Clermont-Ferrand, France) for their help with the PHROG database. We are very grateful to the “Phage Cycle and Bacterial Metabolism” team members at the Laboratoire de Chimie Bactérienne (UMR 7283, CNRS–Aix-Marseille Université, Marseille, France) for their help and support. We are also grateful to Modesto Carballo, Laura Navarro, and Cristina Reyes from the Servicio de Biología, CITIUS, Universidad de Sevilla, for help with experiments performed at the facility.

Conflicts of Interest: The authors declare no conflict of interest.

Dedication: Our co-author Josep Casadesús—affectionately known as “Pepe” by his friends and colleagues—sadly passed away on 2 August 2022, a few days before this manuscript was formally accepted. As a scientist, Pepe will leave a lasting mark on the field of microbial genetics and epigenetics, particularly in *Salmonella*. He will be missed by many, and certainly by the co-authors of this manuscript, as a mentor and friend. We dedicate this work to Pepe, a joyful man who loved life and microbiology.

Appendix A

Table A1. Summary of the 18 meaningful contigs obtained after SPAdes de novo assembly of the trimmed sequencing paired reads.

BioSample	Contig Name	Size (bp)	Coverage *
Se_AO1	Se_AO1	157,620	58
Se_EM1	Se_EM1	159,389	42
Se_EM2	Se_EM2	157,025	56
Se_EM3	Se_EM3	157,176	87
Se_EM4	Se_EM4	157,887	116
Se_ML1	Se_ML1	157,296	67
Salfasec_9	Salfasec_9	43,186	504
Salfasec_10	Salfasec_10	43,187	269
Salfasec_11	Salfasec_11a	59,161	49
	Salfasec_11b	42,744	870
Salfasec_13	Salfasec_13a	157,296	79
	Salfasec_13b	40,532	259
Se_F1	Se_F1	42,861	796
Se_F2	Se_F2	59,161	685
Se_F3	Se_F3a	157,296	362
	Se_F3b	59,161	15
Se_F6	Se_F6a	157,317	354
	Se_F6b	59,596	15

* Three different sequencing runs were performed at different times and in different conditions (e.g., number of genomes loaded onto the Illumina flow cell), accounting in part for the differences observed in sequence coverage.

Table A2. Summary of ORFs predictions for each phage genome.

Strain	ORFs (Total)	Predicted by All 4 Gene Callers	Predicted by Phanotate Only
<i>Salmonella</i> phage SeAO1	243	76.5%	14.0%
<i>Salmonella</i> phage SeEM1	242	73.5%	14.5%
<i>Salmonella</i> phage SeEM2	239	73.6%	15.1%
<i>Salmonella</i> phage SeEM4	240	73.3%	14.6%
<i>Salmonella</i> phage SeF1	76	72.2%	11.8%
<i>Salmonella</i> phage SeF2	87	74.7%	14.9%
<i>Salmonella</i> phage SeF3a	243	70.8%	15.2%
<i>Salmonella</i> phage SeF6a	242	73.5%	12.8%
<i>Salmonella</i> phage SeF6b	85	75.3%	14.1%
<i>Salmonella</i> phage Salfasec13b	82	61.0%	18.3%

Table A3. Accession numbers of the publicly available phage genomes used in this study.

Sequence Name	Accession	Genus	Cited In [#]
Salmonella phage PhiSH19	JN126049.1	<i>Kutterovirus</i>	Table 3, Figure 4
Salmonella phage S8	KY630163.1	<i>Kutterovirus</i>	Figure 4
Salmonella phage vB_SalM-LPST94	MH523359.1	<i>Kutterovirus</i>	Figure 4
Salmonella phage S101	MH370359.1	<i>Jerseyvirus</i>	Figure 4
Salmonella phage BPS1	NC_052994.1	<i>Chiovirus</i>	Table 1
Salmonella phage vB_SentM_sal1	MT499896.1	<i>Kutterovirus</i>	Table 1, Figure 4
Salmonella phage Chennai	MN953776.1	<i>Kutterovirus</i>	Table 1, Figure 4
Salmonella phage Chi	KM458633.1	<i>Chiovirus</i>	Table 1, Figure 4
Salmonella phage S149	MH370387.1	<i>Lederbergvirus</i>	Table 1
Escherichia phage K1G	NC_027993	<i>Jerseyvirus</i>	Figure 4
Escherichia K1H	NC_027994	<i>Jerseyvirus</i>	Figure 4
Escherichia phage Kind1	NC_041897	<i>Jerseyvirus</i>	Figure 4
Escherichia phage Kind2	NC_041898	<i>Jerseyvirus</i>	Figure 4
Escherichia phage VB_EcoS-Golestan	NC_042084	<i>Jerseyvirus</i>	Figure 4
Raoultella phage RP180	NC_048181	<i>Jerseyvirus</i>	Figure 4
Salmonella phage f18SE	NC_028698	<i>Jerseyvirus</i>	Figure 4
Salmonella phage Jersey	NC_021777	<i>Jerseyvirus</i>	Figure 4
Salmonella phage L13	NC_021317	<i>Jerseyvirus</i>	Figure 4
Salmonella phage LSPA1	NC_026017	<i>Jerseyvirus</i>	Figure 4
Salmonella phage SE2	JQ007353	<i>Jerseyvirus</i>	Figure 4
Salmonella phage SETP3	NC_009232	<i>Jerseyvirus</i>	Figure 4
Salmonella phage SETP7	NC_022754	<i>Jerseyvirus</i>	Figure 4
Salmonella phage SETP13	NC_022752	<i>Jerseyvirus</i>	Figure 4
Salmonella phage SS3e	NC_006940	<i>Jerseyvirus</i>	Figure 4
Salmonella phage vB_SenS_AG11	NC_041991	<i>Jerseyvirus</i>	Figure 4
Salmonella phage vB_SenS_Ent1	NC_019539	<i>Jerseyvirus</i>	Figure 4
Salmonella phage wksl3	NC_041992	<i>Jerseyvirus</i>	Figure 4
Salmonella phage FSL_SP-031	NC_021775	<i>Jerseyvirus</i>	Figure 4
Salmonella phage FSL_SP-088	NC_021780	<i>Jerseyvirus</i>	Table 3
Escherichia phage ECML-4	NC_025446	<i>Kutterovirus</i>	Figure 4
Escherichia phage EP75	NC_049433	<i>Kutterovirus</i>	Figure 4
Escherichia phage FEC14	NC_049427	<i>Kutterovirus</i>	Figure 4
Escherichia phage PhaxI	NC_019452	<i>Kutterovirus</i>	Figure 4
Escherichia virus CBA120	JN593240	<i>Kutterovirus</i>	Figure 4
Salmonella phage 38	NC_029042	<i>Kutterovirus</i>	Figure 4
Salmonella phage aagejoakim	NC_049503	<i>Kutterovirus</i>	Figure 4
Salmonella phage Det7	NC_027119	<i>Kutterovirus</i>	Figure 4
Salmonella phage dinky	NC_049504	<i>Kutterovirus</i>	Figure 4
Salmonella phage GG32	NC_031045	<i>Kutterovirus</i>	Figure 4
Salmonella phage heyday	NC_049500	<i>Kutterovirus</i>	Figure 4
Salmonella phage maane	NC_049508	<i>Kutterovirus</i>	Figure 4
Salmonella virus allotria	NC_0495501	<i>Kutterovirus</i>	Figure 4
Salmonella virus barely	NC_049505	<i>Kutterovirus</i>	Figure 4
Salmonella virus bering	NC_049502	<i>Kutterovirus</i>	Figure 4
Salmonella virus BSP101	NC_049502	<i>Kutterovirus</i>	Figure 4
Salmonella phage P22	NC_002371	<i>Lederbergvirus</i>	Figure 4
Salmonella phage P46FS6	NC_049509	<i>Agtrevirus</i>	Figure 4

[#] Figure and/or Table where the phage sequence is cited in the main text.

References

- Murray, C.J.; Ikuta, K.S.; Sharara, F.; Swetschinski, L.; Aguilar, G.R.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; et al. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *Lancet* **2022**, *399*, 629–655. [\[CrossRef\]](#)
- Twort, F.W. An Investigation on the Nature of Ultra-Microscopic Viruses. *Lancet* **1915**, *186*, 1241–1243. [\[CrossRef\]](#)
- d’Herelle, F. Sur un Microbe Invisible Antagoniste des Bacilles Dysentériques. *C. R. Acad. Sci. Paris* **1917**, *165*, 373–375.
- D’Herelle, F. On an Invisible Microbe Antagonistic toward Dysenteric Bacilli: Brief Note by Mr. F. D’Herelle, Presented by Mr. Roux. 1917. *Res. Microbiol.* **2007**, *158*, 553–554. [\[CrossRef\]](#)
- D’Herelle, F. Sur le Rôle du Microbe Filtrant Bactériophage Dans la Dysenterie Bacillaire. *Comptes Rendus Acad. Sci.* **1918**, *167*, 970–972.
- Chanishvili, N. Phage Therapy—History from Twort and d’Herelle through Soviet Experience to Current Approaches. *Adv. Virus Res.* **2012**, *83*, 3–40. [\[CrossRef\]](#)
- Nale, J.Y.; Vinner, G.K.; Lopez, V.C.; Thanki, A.M.; Phothaworn, P.; Thiennimitr, P.; Garcia, A.; AbuOun, M.; Anjum, M.F.; Korbsrisate, S.; et al. An Optimized Bacteriophage Cocktail Can Effectively Control Salmonella in Vitro and in Galleria Mellonella. *Front. Microbiol.* **2021**, *11*, 609955. [\[CrossRef\]](#)
- Dedrick, R.M.; Guerrero-Bustamante, C.A.; Garlena, R.A.; Russell, D.A.; Ford, K.; Harris, K.; Gilmour, K.C.; Soothill, J.; Jacobs-Sera, D.; Schooley, R.T.; et al. Engineered Bacteriophages for Treatment of a Patient with a Disseminated Drug-Resistant Mycobacterium Abscessus. *Nat. Med.* **2019**, *25*, 730–733. [\[CrossRef\]](#)
- Bernheim, A.; Sorek, R. The Pan-Immune System of Bacteria: Antiviral Defence as a Community Resource. *Nat. Rev. Microbiol.* **2020**, *18*, 113–119. [\[CrossRef\]](#)
- Adler, B.A.; Kazakov, A.E.; Zhong, C.; Liu, H.; Kutter, E.; Lui, L.M.; Nielsen, T.N.; Carion, H.; Deutschbauer, A.M.; Mutalik, V.K.; et al. The Genetic Basis of Phage Susceptibility, Cross-Resistance and Host-Range in Salmonella. *Microbiology* **2021**, *167*, 001126. [\[CrossRef\]](#)
- Weigele, P.; Raleigh, E.A. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chem. Rev.* **2016**, *116*, 12655–12687. [\[CrossRef\]](#) [\[PubMed\]](#)
- Garneau, J.E.; Dupuis, M.-È.; Villion, M.; Romero, D.A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadán, A.H.; Moineau, S. The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA. *Nature* **2010**, *468*, 67–71. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lopatina, A.; Tal, N.; Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu. Rev. Virol.* **2020**, *7*, 371–384. [\[CrossRef\]](#) [\[PubMed\]](#)
- Millman, A.; Melamed, S.; Amitai, G.; Sorek, R. Diversity and Classification of Cyclic-Oligonucleotide-Based Anti-Phage Signalling Systems. *Nat. Microbiol.* **2020**, *5*, 1608–1615. [\[CrossRef\]](#) [\[PubMed\]](#)
- Goldfarb, T.; Sberro, H.; Weinstock, E.; Cohen, O.; Doron, S.; Charkap-Amikam, Y.; Afik, S.; Ofir, G.; Sorek, R. BREX Is a Novel Phage Resistance System Widespread in Microbial Genomes. *EMBO J.* **2015**, *34*, 169–183. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gordeeva, J.; Morozova, N.; Sierro, N.; Isaev, A.; Sinkunas, T.; Tsvetkova, K.; Matlashov, M.; Truncaite, L.; Morgan, R.D.; Ivanov, N.V.; et al. BREX System of *Escherichia coli* Distinguishes Self from Non-Self by Methylation of a Specific DNA Site. *Nucleic Acids Res.* **2019**, *47*, 253–265. [\[CrossRef\]](#)
- Picton, D.M.; Luyten, Y.A.; Morgan, R.D.; Nelson, A.; Smith, D.L.; Dryden, D.T.F.; Hinton, J.C.D.; Blower, T.R. The Phage Defence Island of a Multidrug Resistant Plasmid Uses Both BREX and Type IV Restriction for Complementary Protection from Viruses. *Nucleic Acids Res.* **2021**, *49*, 11257–11273. [\[CrossRef\]](#)
- Hampton, H.G.; Watson, B.N.J.; Fineran, P.C. The Arms Race between Bacteria and Their Phage Foes. *Nature* **2020**, *577*, 327–336. [\[CrossRef\]](#)
- Huiling, E.; Athukoralage, J.; Guan, J.; Silas, S.; Carion, H.; Bondy-Denomy, J. Bacteriophages Antagonize CGAS-like Immunity in Bacteria. *bioRxiv* **2022**. bioRxiv:2022.03.30.486325. [\[CrossRef\]](#)
- Hobbs, S.J.; Wein, T.; Lu, A.; Morehouse, B.R.; Schnabel, J.; Leavitt, A.; Yirmiya, E.; Sorek, R.; Kranzusch, P.J. Phage Anti-CBASS and Anti-Pycsar Nucleases Subvert Bacterial Immunity. *Nature* **2022**, *605*, 522–526. [\[CrossRef\]](#)
- Hatfull, G.F. Dark Matter of the Biosphere: The Amazing World of Bacteriophage Diversity. *J. Virol.* **2015**, *89*, 8107–8110. [\[CrossRef\]](#) [\[PubMed\]](#)
- Olivenza, D.R.; Casadesús, J.; Ansaldi, M. Epigenetic Biosensors for Bacteriophage Detection and Phage Receptor Discrimination. *Environ. Microbiol.* **2020**, *22*, 3126–3142. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bin Jang, H.; Bolduc, B.; Zablocki, O.; Kuhn, J.H.; Roux, S.; Adriaenssens, E.M.; Brister, J.R.; Kropinski, A.M.; Krupovic, M.; Lavigne, R.; et al. Taxonomic Assignment of Uncultivated Prokaryotic Virus Genomes Is Enabled by Gene-Sharing Networks. *Nat. Biotechnol.* **2019**, *37*, 632–639. [\[CrossRef\]](#) [\[PubMed\]](#)
- Meier-Kolthoff, J.P.; Göker, M. VICTOR: Genome-Based Phylogeny and Classification of Prokaryotic Viruses. *Bioinformatics* **2017**, *33*, 3396–3404. [\[CrossRef\]](#) [\[PubMed\]](#)
- Terzian, P.; Olo Ndela, E.; Galiez, C.; Lossouarn, J.; Pérez Bucio, R.E.; Mom, R.; Toussaint, A.; Petit, M.-A.; Enault, F. PHROG: Families of Prokaryotic Virus Proteins Clustered Using Remote Homology. *NAR Genom. Bioinform.* **2021**, *3*, lqab067. [\[CrossRef\]](#)
- Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [\[CrossRef\]](#)

27. Wang, J.; Dai, W.; Li, J.; Li, Q.; Xie, R.; Zhang, Y.; Stubenrauch, C.; Lithgow, T. AcrHub: An Integrative Hub for Investigating, Predicting and Mapping Anti-CRISPR Proteins. *Nucleic Acids Res.* **2021**, *49*, D630–D638. [\[CrossRef\]](#)
28. Figueroa-Bossi, N.; Uzzau, S.; Maloriol, D.; Bossi, L. Variable Assortment of Prophages Provides a Transferable Repertoire of Pathogenic Determinants in Salmonella. *Mol. Microbiol.* **2001**, *39*, 260–271. [\[CrossRef\]](#)
29. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A Better, Faster Version of the PHAST Phage Search Tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [\[CrossRef\]](#)
30. Moraru, C.; Varsani, A.; Kropinski, A.M. VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses* **2020**, *12*, 1268. [\[CrossRef\]](#)
31. Casjens, S.; Hayden, M. Analysis in Vivo of the Bacteriophage P22 Headful Nuclease. *J. Mol. Biol.* **1988**, *199*, 467–474. [\[CrossRef\]](#)
32. Feiss, M.; Widner, W.; Miller, G.; Johnson, G.; Christiansen, S. Structure of the Bacteriophage Lambda Cohesive End Site: Location of the Sites of Terminase Binding (CosB) and Nicking (CosN). *Gene* **1983**, *24*, 207–218. [\[CrossRef\]](#)
33. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative Genomics Viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [\[CrossRef\]](#)
35. Dimonaco, N.J.; Aubrey, W.; Kenobi, K.; Clare, A.; Creevey, C.J. No One Tool to Rule Them All: Prokaryotic Gene Prediction Tool Annotations Are Highly Dependent on the Organism of Study. *Bioinformatics* **2022**, *38*, 1198–1207. [\[CrossRef\]](#)
36. Kongari, R.; Rajaure, M.; Cahill, J.; Rasche, E.; Mijalis, E.; Berry, J.; Young, R. Phage Spanins: Diversity, Topological Dynamics and Gene Convergence. *BMC Bioinform.* **2018**, *19*, 326. [\[CrossRef\]](#)
37. Pawluk, A.; Bondy-Denomy, J.; Cheung, V.H.W.; Maxwell, K.L.; Davidson, A.R. A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of *Pseudomonas Aeruginosa*. *mBio* **2014**, *5*, e00896-14. [\[CrossRef\]](#)
38. Lu, M.J.; Henning, U. The Immunity (Imm) Gene of *Escherichia coli* Bacteriophage T4. *J. Virol.* **1989**, *63*, 3472–3478. [\[CrossRef\]](#)
39. Braun, V.; Killmann, H.; Herrmann, C. Inactivation of FhuA at the Cell Surface of *Escherichia coli* K-12 by a Phage T5 Lipoprotein at the Periplasmic Face of the Outer Membrane. *J. Bacteriol.* **1994**, *176*, 4710–4717. [\[CrossRef\]](#)
40. Isaev, A.; Drobiazko, A.; Sierro, N.; Gordeeva, J.; Yosef, I.; Qimron, U.; Ivanov, N.V.; Severinov, K. Phage T7 DNA Mimic Protein Ocr Is a Potent Inhibitor of BREX Defence. *Nucleic Acids Res.* **2020**, *48*, 5397–5406. [\[CrossRef\]](#)
41. Dion, M.B.; Oechslin, F.; Moineau, S. Phage Diversity, Genomics and Phylogeny. *Nat. Rev. Microbiol.* **2020**, *18*, 125–138. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Kristensen, D.M.; Waller, A.S.; Yamada, T.; Bork, P.; Mushegian, A.R.; Koonin, E.V. Orthologous Gene Clusters and Taxon Signature Genes for Viruses of Prokaryotes. *J. Bacteriol.* **2012**, *19*, 941–950. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Turner, D.; Kropinski, A.M.; Adriaenssens, E.M. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* **2021**, *13*, 506. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Hendrix, R.W.; Ko, C.-C.; Jacobs-Sera, D.; Hatfull, G.F.; Erhardt, M.; Hughes, K.T.; Casjens, S.R. Genome Sequence of Salmonella Phage χ . *Genome Announc.* **2015**, *3*, e01229-14. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Chen, J.; Long, R.; Wang, X.; Liu, B.; Chou, K.-C. DRHP-PseRA: Detecting Remote Homology Proteins Using Profile-Based Pseudo Protein Sequence and Rank Aggregation. *Sci. Rep.* **2016**, *6*, 32333. [\[CrossRef\]](#)
46. Monteiro, R.; Pires, D.P.; Costa, A.R.; Azeredo, J. Phage Therapy: Going Temperate? *Trends Microbiol.* **2019**, *27*, 368–378. [\[CrossRef\]](#)
47. ElAntak, L.; Ansaldi, M.; Guerlesquin, F.; Jean, V.M.; Morelli, X. Structural and Genetic Analyses Reveal a Key Role in Prophage Excision for the TorI Response Regulator Inhibitor. *J. Biol. Chem.* **2005**, *280*, 36802–36808. [\[CrossRef\]](#)
48. Panis, G.; Franche, N.; Méjean, V.; Ansaldi, M. Insights into the Functions of a Prophage Recombination Directionality Factor. *Viruses* **2012**, *4*, 2417–2431. [\[CrossRef\]](#)
49. Piya, D.; Vara, L.; Russell, W.K.; Young, R.; Gill, J.J. The Multicomponent Antirestriction System of Phage P1 Is Linked to Capsid Morphogenesis. *Mol. Microbiol.* **2017**, *105*, 399–412. [\[CrossRef\]](#)
50. Nagaraja, V.; Shepherd, J.C.W.; Pripfl, T.; Bickle, T.A. Two Type I Restriction Enzymes from Salmonella Species. Purification and DNA Recognition Sequences. *J. Mol. Biol.* **1985**, *182*, 579–587. [\[CrossRef\]](#)
51. Lemire, S.; Figueroa-Bossi, N.; Bossi, L. Bacteriophage Crosstalk: Coordination of Prophage Induction by Trans-Acting Antirepressors. *PLoS Genet.* **2011**, *7*, e1002149. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Seville, M.; Ikeda, T.; Hotani, H. The Effect of Sugars on the Morphology of the Bacterial Flagellum. *FEBS Lett.* **1993**, *332*, 260–262. [\[CrossRef\]](#)
53. Phothaworn, P.; Dunne, M.; Supokaivanich, R.; Ong, C.; Lim, J.; Taharnklaew, R.; Vesaratchavest, M.; Khumthong, R.; Pringsulaka, O.; Ajawatanawong, P.; et al. Characterization of Flagellotropic, Chi-Like Salmonella Phages Isolated from Thai Poultry Farms. *Viruses* **2019**, *11*, 520. [\[CrossRef\]](#)
54. Davies, M.R.; Broadbent, S.E.; Harris, S.R.; Thomson, N.R.; van der Woude, M.W. Horizontally Acquired Glycosyltransferase Operons Drive Salmonellae Lipopolysaccharide Diversity. *PLoS Genet.* **2013**, *9*, e1003568. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Susskind, M.M.; Botstein, D.; Wright, A. Superinfection Exclusion by P22 Prophage in Lysogens of Salmonella Typhimurium. III. Failure of Superinfecting Phage DNA to Enter SieA⁺ Lysogens. *Virology* **1974**, *62*, 350–366. [\[CrossRef\]](#)
56. Wahl, A.; Battesti, A.; Ansaldi, M. Prophages in Salmonella Enterica: A Driving Force in Reshaping the Genome and Physiology of Their Bacterial Host? *Mol. Microbiol.* **2019**, *111*, 303–316. [\[CrossRef\]](#)
57. Parma, D.H.; Snyder, M.; Sobolevski, S.; Nawroz, M.; Brody, E.; Gold, L. The Rex System of Bacteriophage Lambda: Tolerance and Altruistic Cell Death. *Genes Dev.* **1992**, *6*, 497–510. [\[CrossRef\]](#)

58. Darwin, A.J. The Phage-Shock-Protein Response. *Mol. Microbiol.* **2005**, *57*, 621–628. [\[CrossRef\]](#)
59. Kushwaha, S.K.; Bhavesh, N.L.S.; Abdella, B.; Lahiri, C.; Marathe, S.A. The Phylogenomics of CRISPR-Cas System and Revelation of Its Features in Salmonella. *Sci. Rep.* **2020**, *10*, 21156. [\[CrossRef\]](#)
60. Bondy-Denomy, J.; Pawluk, A.; Maxwell, K.L.; Davidson, A.R. Bacteriophage Genes that Inactivate the CRISPR/Cas Bacterial Immune System. *Nature* **2013**, *493*, 429–432. [\[CrossRef\]](#)
61. Kropinski, A.M.; Mazzocco, A.; Waddell, T.E.; Lingohr, E.; Johnson, R.P. Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. *Methods Mol. Biol.* **2009**, *501*, 69–76. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, *9*, 676–682. [\[CrossRef\]](#)
63. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [\[CrossRef\]](#)
64. Garneau, J.R.; Depardieu, F.; Fortier, L.-C.; Bikard, D.; Monot, M. PhageTerm: A Tool for Fast and Accurate Determination of Phage Termini and Packaging Mechanism Using next-Generation Sequencing Data. *Sci. Rep.* **2017**, *7*, 8292. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Laslett, D.; Canback, B. ARAGORN, a Program to Detect tRNA Genes and tmRNA Genes in Nucleotide Sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Bocs, S.; Cruveiller, S.; Vallenet, D.; Nuel, G.; Médigue, C. AMIGene: Annotation of Microbial Genes. *Nucleic Acids Res.* **2003**, *31*, 3723–3726. [\[CrossRef\]](#)
67. Delcher, A.L.; Bratke, K.A.; Powers, E.C.; Salzberg, S.L. Identifying Bacterial Genes and Endosymbiont DNA with Glimmer. *Bioinformatics* **2007**, *23*, 673–679. [\[CrossRef\]](#)
68. Noguchi, H.; Taniguchi, T.; Itoh, T. MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. *DNA Res.* **2008**, *15*, 387–396. [\[CrossRef\]](#)
69. Hyatt, D.; Chen, G.-L.; Locascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinform.* **2010**, *11*, 119. [\[CrossRef\]](#)
70. McNair, K.; Zhou, C.; Dinsdale, E.A.; Souza, B.; Edwards, R.A. PHANOTATE: A Novel Approach to Gene Identification in Phage Genomes. *Bioinformatics* **2019**, *35*, 4537–4542. [\[CrossRef\]](#)
71. Otasek, D.; Morris, J.H.; Bouças, J.; Pico, A.R.; Demchak, B. Cytoscape Automation: Empowering Workflow-Based Network Analysis. *Genome Biol.* **2019**, *20*, 185. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment. *Nat. Methods* **2012**, *9*, 173–175. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Suzek, B.E.; Wang, Y.; Huang, H.; McGarvey, P.B.; Wu, C.H. UniRef Clusters: A Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches. *Bioinformatics* **2015**, *31*, 926–932. [\[CrossRef\]](#)
74. Söding, J.; Biegert, A.; Lupas, A.N. The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res.* **2005**, *33*, W244–W248. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Wang, J.; Dai, W.; Li, J.; Xie, R.; Dunstan, R.A.; Stubenrauch, C.; Zhang, Y.; Lithgow, T. PaCRISPR: A Server for Predicting And Visualizing Anti-CRISPR Proteins. *Nucleic Acids Res.* **2020**, *48*, W348–W357. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Eitzinger, S.; Asif, A.; Watters, K.E.; Iavarone, A.T.; Knott, G.J.; Doudna, J.A.; Minhas, F.U.A.A. Machine Learning Predicts New Anti-CRISPR Proteins. *Nucleic Acids Res.* **2020**, *48*, 4698–4708. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Tesson, F.; Hervé, A.; Mordret, E.; Touchon, M.; d’Humières, C.; Cury, J.; Bernheim, A. Systematic and Quantitative View of the Antiviral Arsenal of Prokaryotes. *Nat. Commun.* **2022**, *13*, 2561. [\[CrossRef\]](#)