Comparative genomics and phylogeny unveil lineage diversification of *Citrobacter rodentium* polyvalent bacteriophages

Carolina M. Mizuno, Mart Krupovic, Laurent Debarbieux* and Dwayne R. Roach*

Affiliations: Department of Microbiology, Institut Pasteur, Paris, 75015, France
*Co-corresponding author: dwayne.roach@pasteur.fr
*Co-corresponding author: laurent.debarbieux@pasteur.fr

ABSTRACT

Citrobacter rodentium is a mouse-restricted pathogen that has long been used as an *in vivo* model for two important human intestinal pathogen enteropathogenic *Escherichia coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC). And yet, in contrast to *E. coli*, little is known about the bacteriophages (phages) – bacterial viruses – that infect *C. rodentium*, reflecting in part a need to isolate and compare phages associated with this bacterial species. Here, we isolated two novel virulent phages CrRp3 and CrRp10 that infect *C. rodentium* and conduct *in vitro* and comparative genomic studies with other, related phages. We show that CrRp3 should be considered as a new species within the genus *Sp6virus* in the family *Podoviridae*, having relatively close relations to *E. coli* phages K1-E and K1-5 in both gene content and genome organization. By contrast, CrRp10 genome is 98% identical to that of *E. coli* phage Ime09 and thus should be considered as a new strain of the latter phage. We provide evidence that genomes of CrRp3 and CrRp10 have been shaped by both horizontal genetic exchange and genetic drift. Furthermore, phylogenetic analyses showed that these phages are more closely related to *E. coli* phages than to those infecting members of the *Citrobacter* genus, suggesting that CrRp3 and CrRp10 may have evolved from *E. coli* phages following a host-switch event.

INTRODUCTION

The genus *Citrobacter* belongs to the family of Enterobacteriaceae and comprises eleven different species of facultative anaerobic Gram-negative bacilli, widely distributed in water, soil, food and intestinal tract of humans and animals. Previously recognized as colonizers with low virulence or environmental contaminants, they are now known to account for up to 6% of all nosocomially acquired life threatening *Enterobacteriaceae* infections, such as urinary tract, respiratory, wound, bone, bloodstream, and central nervous system infections (1, 2). While the majority of human infections are caused by *C. freundii* and *C. koseri*, *C. youngae*, *C. braakii*, and *C. amalonaticus* are also important human pathogens, all of which are increasingly difficult to treat due to the rise of multidrug resistance (MDR) (1-3). By contrast, *C. rodentium* is a naturally occurring mouse-restricted pathogen, which is genetically highly similar to the important human pathogens enteropathogenic *Escherichia coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC) (4-6). Thus, *C. rodentium* has become an important *in vivo* model for several human intestinal diseases and disorders (7, 8).

The prominent role of bacteriophages (phages) – viruses that prey on bacteria – in the functioning of the biosphere has been firmly established across different ecosystems (9). Nevertheless, we have little knowledge on the role phages play in shaping bacterial population phenotypes in perhaps the most important and clinically relevant microbial ecosystem – the human microbiome. Despite knowing that antagonistic interactions between phages and bacteria play a key role in driving and maintaining microbial diversity, eco-evolutionary processes in humans and animals have not received much research attention (10, 11). As both obligate parasites and vectors of horizontal gene transfer, a better understanding of phage strain diversification and how viral diversity might impact bacterial diversity and populations is required. Furthermore, there is renewed interest in the use of phages to eliminate or modulate bacterial population, namely phage therapy, partly due to their specificity for host bacterial species and ability to kill MDR pathogens (12, 13).

Phages are the most abundant biological entities

on the planet, with an estimated 10^{31} present in the biosphere (9), suggesting there is an untapped biodiversity of *Citrobacter* phages. And yet, only a limited number of *Citrobacter* phages have been previously described, most of which only at the genome level. Furthermore, most previously characterized phages infect *C. freundii*, including putative members of the genera *T4virus* (Merlin, Miller, Moon) (14-16), *FelixO1virus* (Michonne, Mordin, Moogle) (17-19), *T1virus* (Stevie) (20), and *T7virus* (phiCFP-1, SH1, SH2, SH3, SH4, SH5) (21, 22). Genome sequences are also publically available for the *T7virus* related phages CR8 and CR44b that specifically infect *C. rodentium* (23). Among all these phages, only the *C. freundii T7virus* related phages have been experimentally studied.

In this study, we isolated and characterized two novel virulent phages, CrRp3 and CrRp10, which could infect and lyse the mouse-restricted intestinal pathogen *C. rodentium*, as well as several human intestinal pathogenic *E. coli* strains, suggesting these phages could be effective for their control. CrRp10 is, to the best of our knowledge, the first sequenced *C. rodentium* phage with a contractile tail, a hallmark of the family *Myoviridae*. Furthermore, comparative genomic analyses provide evidence that these phages may have evolved from *E. coli* Sp6-like and T4-like phages, respectively.

RESULTS

Biological characteristics of two novel *C. rodentium* phages.

We isolated phage strains vB_CroP_CrRp3 (CrRp3) and vB_CroM_CrRp10 (CrRp10) from different wastewater samples from Paris, France after they were found to form distinct clear plaques on the *C. rodentium* strain ICC180. Figure 1A shows characteristic transmission electron micrographs that revealed that the two phages have different *Caudovirales* morphotypes (24). CrRp3 has an isometric (most likely icosahedral) head and a short tail, which implies this virus belongs to the *Podoviridae* family. By contrast, the elongated, likely icosahedral, head of CrRp10 is connected to a long tail covered with a clearly discernable sheath, characteristic of the *Myoviridae* family.

CrRp3 and CrRp10 both have the ability to reduce bacterial population growth in liquid culture; the more potent virus appears to be the *Podoviridae* CrRp3 (Fig. 1B). That is, CrRp3 exhibited early lytic activity dynamics at the lowest multiplicity of infection (MOI; 0.001) tested that required ~60 min to show signs of reversing bacterial population growth, whereas CrRp10 required ~135 min. Increasing the phage concentration (i.e. higher MOIs of 0.1 and 10) abolished any differences in early lytic activity to initiate bacterial density reduction (Fig. 1B). However, late lytic activity dynamics again revealed differences in lytic potency. CrRp3 was able to nearly eliminate bacterial density by ~2.2 h at each MOI tested, while CrRp10 required at least a further 30 min to show bacterial elimination (Fig. 1C). Notably, however, whereas *C. rodentium* was able to gain resistance to CrRp3 infection within 4 h, no CrRp10-resistance cells have appeared for the duration of the study (Fig. 1C).

Genome structure and general features.

Table 1 shows the general genomic features for CrRp3, CrRp10, as well as for all other *Citrobacter* phage genomes available in the public databases. CrRp3 has a genome size of 44.3 kb and CrRp10 a size of 171.5 kb and displayed average GC contents of 45.1% and 35.5%, respectively. Notably, CrRp10 displays the lowest GC% content among all sequenced *Citrobacter* phages to date. Furthermore, it appears that all *C. rodentium* phages, including CrRp3 and CrRp10, exhibit GC contents significantly lower than their host *C. rodentium* (54.5% GC content). Likewise, *C. freundii* displays a higher GC content (~51.5%) than phages infecting it, with the exception of SH4 (52.6% GC content) (Table 1).

The annotated features of the CrRp3 and CrRp10 genomes are listed in Table S1 and Table S2, respectively. CrRp3 has a terminally repetitive dsDNA genome that consists of 54 coding sequences (CDSs) with 35% having putative functions. CrRp10 has a circularly permuted dsDNA genome that consists of 267 CDSs with 50% having putative functions and 10 tRNAs. CrRp10 to the best of our knowledge is the first member of the *Myoviridae* infecting *C. rodentium* to have its genome completely sequenced. Neither phage exhibited gene similarities to known bacterial virulence-associated genes or lysogeny-associated genes. Similarly, antibiotic resistance genes were not detected in either genome, which is consistent with previous findings of these genes being rare in phages (25).

Phylogenetic analysis

Using the Genome-BLAST Distance Phylogeny (GBDP) method (26), we show that CrRp3 is more closely related to *Podoviridae* phages infecting *Escherichia* rather than those infecting any *Citrobacter* species. That is, *Citrobacter* (Table 1) and *Escherichia* (Table S3) phages display a heterogeneous clustering of closely related podoviruses into three distinct clades, each with a maximal branch support (Fig. 2A). CrRp3 clusters only with *Escherichia* phages in clade 1 (C1), whereas

phages that infect the human pathogen *C. freundii* (phiCFP1, SH1, and SH2) cluster in clade 2 (C2) and (SH3 and SH4) clade 3 (C3). Interestingly, the only previously described *C. rodentium* phages, CR8 and CR44b, have genomes that also cluster in C3 despite, as previously mentioned, these phages having little protein sequence homology (Fig S1). Lastly, phage CVT2 branches separately. This is not unexpected because CVT2 was isolated from the gut of termites on an uncharacterized *Citrobacter* species (27).

The relationship of the CrRp10 genome with those of other *Citrobacter* and *Escherichia* myoviruses also shows a heterogeneous clustering (Fig. 3A). CrRp10 groups with several closely-related (as judged from the short branch lengths) *E. coli* phages (Ime09, vB_EcoM_UFV13, slur02, slur07 and slur14) in C1. Other known *Myoviridae* phages, which infect *C. freundii*, are only distantly related to CrRp10 and other C1 phages, but also form mixed clades (C2-4) with *Escherichia* phages. Interestingly, C3 is composed almost exclusively of *C. freundii* phages (IME, CF2, Miller, CfP1, and Margaery); with the exception of the *E. coli* phage Lw1.

Comparative genomic analysis with other, related phages

To elucidate the potential mechanisms underlying the different host specificities of phylogenetically close phages, comparative genomic analysis was conducted using BLAST genome alignment. The genomic comparison at the amino acid level of CrRp3 and CrRp10 against all *Citrobacter* phages showed that they are unique among other sequenced *Citrobacter* phages (Fig. 4). CrRp10 is most closely related with phage Moon, with an average amino acid identity of 70% over 50% coverage.

Nearly 91% of the CrRp3 genes had a best match among phages from the *Sp6virus* genus in the subfamily *Autographivirinae* (family *Podoviridae*), pointing towards a taxonomic relationship (Table S4). CrRp3 has the highest genome nucleotide and structure similarity to the *E. coli* phages K1-5 and K1-E, both of which belong to the *Sp6virus* genus (Fig. 2B). Almost half of CrRp3 genes products have the best hits either in K1-5 or K1-E genomes (Table S4). These, include the DNA and RNA polymerases, DNA ligase and major capsid protein as well as many hypothetical proteins. CrRp3 gene products that differ from K1-5 include the head tail connector protein, endolysin, tailspike protein, lyase, minor structural protein and several proteins with unknown function (Fig. 2B). Interestingly, CrRp3 lyase and minor structural protein are the only gene products with similarity to those in *Citrobacter* phage CR8 (Table S1). By contrast, CrRp10 genome shares significant synteny with the genome of *E. coli* phage Ime09 belonging to the *T4virus* genus in the subfamily *Tevenvirinae* (family *Myoviridae*) (98% nucleotide identity over the complete length) (Table S4, Fig. 3A). There are few genome features unique to CrRp10 compared to Ime09, including considerable divergence within the tail fiber gene (3% dissimilarly over 80% of the length, at the protein level). Members of the *T4virus* genus are known to have an extended host range, largely due to their tail fiber proteins having a unique ability to bind to several outer membrane proteins or lipopolysaccharide (LPS) receptors (28). Another striking feature of the CrRp10 genome is the recombination event, which resulted in the gain a dUTPase with high sequence similarity to that encoded by phage e11/2. This latter phage has been shown to infect the EHEC (29), which suggests that CrRp10 might be a good candidate for development of a new therapeutic agent to inhibit important *E. coli* O157:H7 strains. Other recombination events in CrRp10 have added several putative endonucleases with high similarity to homologs in other related Enterobacteriaceae phages (Table S2).

Lytic spectrum

Next, we tested the lytic spectrum of the virulent phages CrRp3 and CrRp10 along with other representative phages against various bacterial strains (Table 2). While, in addition to their isolation strain of *C. rodentium*, CrRp3 could infect the *E. coli* strain K-12, while CrRp10 displays a much broader host range, including K-12 and several pathotype strains of *E. coli*, as well as the *Erwinia carotovora* strain CFBP2141. Although the *E. coli* phage LF82_P10 also exhibits a relatively broad host range (11), it cannot infect *C. rodentium*. Moreover, most of the *E. coli* as well as *Pseudomonas aeruginosa* and *Serratia marcescens* strains tested were resistant to both CrRp3 and CrRp10.

DISCUSSION

In this study, we report the genomic and phenotypic characterization of two novel virulent phages, CrRp3 and CrRp10, which infect the mouse-restricted pathogen *C. rodentium*. In addition to doubling the available *C. rodentium* phage genomes in the public databases, these phages provide new evolutionary relationships with the expanding group of viruses belonging to the *Sp6virus* and *T4virus* genera, respectively. CrRp10 is the first reported virulent *Myoviridae* family phage with complete genome sequence, as the genome sequence of the previously isolated and characterized *C. rodentium Myoviridae*

phage phiCr1 remains unavailable (30). CrRp3 and CrRp10 appear to be quite distantly related to the previously sequenced *C. rodentium* phages CR8 and CR44b (both members of the genus *T7virus*) (23), as well as to other phages that infect the human pathogen *C. freundii*.

Comparative genomics analysis revealed that CrRp3 and CrRp10 may have evolved independently from closely related *E. coli* phages, presumably because it was advantageous to gain new specificities to infect *C. rodentium* and, as a result, occupy a new niche. The evolution of phages is a multifaceted and complex process, strongly influenced by genetic drift and horizontal acquisition of new genetic elements from various sources, including other phages (11). Interestingly, the tail associated genes, in particular those encoding for the tail fibers responsible for host recognition, and endolysin gene from CrRp3 appear to have evolved the most from the presumably ancestral genes of *E. coli* phage K1-5 (Fig. 2B). Phage K1-5 has been shown to exhibit two tail fiber genes, one carrying a lyase domain and the second one an endosialidase domain, which allow it to infect both *E. coli* K1 and K5 with different polysaccharide capsule types (31). Interestingly, CrRp3 lyase is more closely related to that from the *Citrobacter* phages CR8 and CR44b (Table S1), also known to infect *C. rodentium*, which suggests a mosaic genome structure that may be driven by recombination of modules from varied species. This is consistent with other phages of the genus *Sp6virus*, which exhibit a high genetic identity and structure (and highly specific RNA polymerase) (32), with the modest differences observed in gene products implicated in adaptation to host constraints.

For most bacterial virus genera, 95% DNA sequence identity is used by the International Committee on Taxonomy of Viruses (ICTV) as the species demarcation criterion (33, 34). CrRp3 is relatively closely related to *E. coli* Sp6-like phages K1-E and K1-5 in both gene content and genome organization, with which it shares 90% (over 76% of the genome) and 77% (over 75% of the genome) pairwise nucleotide identity, respectively. Accordingly, CrRp3 should be considered as a new species within the genus *Sp6virus* in the family *Podoviridae* (subfamily *Autographivirinae*). In contrast, CrRp10 genome is 98% identical to that of *E. coli* phage Ime09 and thus should be considered as a new strain of the latter phage. Notably, information about Ime09 is confined to the genomic analysis, whereas the details of its infection cycle and the host range (other than the isolation host) remain uncharacterized (35).

C. rodentium and *E. coli* pathotypes EPEC and EHEC are attaching and effacing (A/E) pathogens derived from a common ancestral origin (6). However, the genome of *C. rodentium* exhibits several features typical of a bacterium that has recently passed through an evolutionary bottleneck, including several large-scale genomic rearrangements and functional gene loss in the core genomic regions (5, 6, 36, 37). This has led researchers to postulate that *C. rodentium* may have emerged alongside the development of laboratory mice model of human *E. coli* infections (6, 37). Several studies have demonstrated the reciprocal selection of phages on bacterial populations and bacteria on phage populations. There is also increasing evidence that this process can maintain phage diversity, influence phage virulence, and increase phage evolvability (11). The genomic characteristics of CrRp3 and CrRp10 strengthen the hypothesis that their host *C. rodentium* has recently evolved from *E. coli*. Our work does not exclude the potential for isolation of other *C. rodentium* phages more closely related to phages infecting other *Citrobacter* species, such as *C. freundii* phages Merlin and Moon.

Phages CrRp3 and CrRp10 exhibit polyvalence, infecting strain across several genera within the gram-negative *Enterobacteriaceae*, including *Citrobacter*, *Escherichia*, and *Erwinia*. By contrast, the previously characterized *C. rodentium* phage phiCR1, which targets lipopolysaccharide (LPS) as a receptor, was shown to be highly species-specific infecting only *C. rodentium* and not closely related *Enterobacteriaceae,* including *C. freundii* and *E. coli* (30). Although, phiCR1 host range supports the general notion that phages are largely species-specific (38, 39), several other phages infecting various *Enterobacteriaceae* (38, 40, 41) and staphylococci (42) have also been demonstrated to be polyvalent. Polyvalent phages infecting strains across genera raise a question as to their so called “optimal” host and whether a phage’s host should be inferred by genomic relationships rather than bacterial strain plaquing *in vitro* (43). This question becomes increasingly relevant for uncultivated phages discovered through metagenomic surveys.

Moreover, virulent phages similar to CrRp3 and CrRp10 are being re-investigated as potential antimicrobial agents to both combat bacterial diseases and the dissemination of MDR bacteria (12). Because mice are resistant to EPEC and EHEC infections, *C. rodentium* is widely used as an *in vivo* model system for several important human gastrointestinal diseases (44, 45). However, to the best of our knowledge, phages that infect *C. rodentium* have not yet been explored for antibacterial potential in preclinical animal models. The newly characterized C. rodentium phages, in particular the polyvalent *Myoviridae* phage CrRp10 resilient against

resistance development (Fig. 1C), may lead studies into innovative antimicrobial agents for food safety, veterinary and clinical use.

MATERIAL AND METHODS

Bacterial strains, phage isolation, and culture conditions

The *C. rodentium* strain ICC180 (46) was used isolate phages from waste water. *C. rodentium* and other gram-negatives were grown at 37°C in Luria-Bertani (LB) medium on an orbital shaker. LB mixed with 1.5% agar provided a solid medium on which bacteria were cultured. Citrobacter phages were collected from Paris France municipal wastewater and isolated on lawns of early log phase seeded agar of *C. rodentium*. Select plaque lysates were serial passaged five times by spiking liquid ICC180 cultures grown to an optical density (OD) OD_{600} 0.25 with phages titer at a multiplicity of infection (MOI) of 0.1. When required, phage strains were further purified by cesium chloride density gradient (1.3, 1.5 and 1.6 g/ml) using ultracentrifugation at 140,000 g for 3 h. Single bands were dialyzed in cold H_2O once and cold Tris buffer twice.

Phage characterizations

Phage inhibition of bacterial population growth was measured in changes in culture optical densitometry. Microtiter plate wells were filled with 100 µl of 2x LB concentrate spiked with 2 $x10^6$ CFU of *C. rodentium* strain ICC180. Phages were diluted in PBS and added to wells at different MOIs and PBS was added to achieve a total assay volume of 200 µL. A Promega GloMax plate reader was used to measure OD_{600} nm at 15-min intervals for 18 h, while being incubated at 37°C and orbital shaken for 30 secs prior to each read. Phage host ranges were determined by spotting 4 µL of 10^7 PFU phages onto air-dried lawns of mid-log growing test bacteria strains on agar and grown overnight (Table 2).

Aliquots of 10 µl of cesium chloride purified phages dialyzed against Tris buffer were applied to carbon-coated copper grids, negatively stained with 2% uranyl acetate for 30 s and observed under a transmission electron microscope Tecnai Biotwin 120 FEI- 1 (FEI Company, USA) operating at 120kV.

Genome sequencing and bioinformatics analyses

Phage DNA was extracted from sterile DNase and RNase pretreated cesium chloride purified phages by a phenol-chloroform extraction as previously described (47). DNA samples were sequenced using an Illumina MiSeq (Illumina Inc., San Diego, CA) with 2x250 bp read length. For phage sequence analysis, the quality of Illumina reads was visualized by FastQC v0.10.1 Brabraham Bioinformatics (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality controlled trimmed reads were assembled to a single contig using CLC Assembler (Galaxy Version 4.4.2). Protein-coding genes in the assembled contigs were predicted using Prodigal (48), and tRNAs were predicted using tRNAscan-SE (49). Additional annotation of genes was done by comparing against the NCBI NR, COG (50), and TIGRfam (51) databases. In addition, genomes were also manually annotated using HHPRED server (52). Genomic comparisons among related viral genomes and reference genomes were performed using tBLASTx or BLASTN (53).

Phylogenetic trees

The phylogenies of phages CrRp3 and CrRp10 were constructed using VICTOR (54). All pairwise comparisons of the nucleotide sequences were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method (26) under settings recommended for prokaryotic viruses (54). GBDP approach was used for phylogenetic inference from all publically available *Podoviridae Citrobacter* phages, including phages CR8 and CR44b that infect *C. rodentium*, and 37 of the closest related *E. coli* phages (Table S3). All available *E. coli* and *Citrobacter* phage genomes were downloaded from https://www.ncbi.nlm.nih.gov/genome/browse/. The resulting intergenomic distances (100 replicates each) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR post processing (55). The trees were rooted at the outgroup (*Synechococcus* phages**)** and visualized with FigTree V1.4.3 (http://tree.bio.ed.ac.uk).

Nucleotide sequence accession numbers.

The complete genome sequence and annotations of phages vB_CroP_CrRp3 and vB_CroM_CrRp10 have been deposited GenBank under accession numbers MG775042 and MG775043, respectively.

Acknowledgements

We thank Elena Resen, Luisa De Sordi, and Marta Mansos Lourenço of Institut Pasteur for their assistance. CMM was supported by the European Molecular Biology Organization (ALTF 1562-2015) and Marie Curie Actions program from the European Commission (LTFCOFUND2013, GA-2013-609409), while DRR was supported by an European Respiratory Society

Fellowship (RESPIRE2–2015–8416). The funders had no role in study design, data analyses, or manuscript preparation.

REFERENCES

1. **Lavigne JP, Defez C, Bouziges N, Mahamat A, Sotto A.** 2007. Clinical and molecular epidemiology of multidrug-resistant *Citrobacter* spp. infections in a French university hospital. Eur J Clin Microbiol Infect Dis **26:**439-441.
2. **Deveci A, Coban AY.** 2014. Optimum management of *Citrobacter koseri* infection. Expert Rev Anti Infect Ther **12:**1137-1142.
3. **Janda JM, Abbott SL, Cheung WK, Hanson DF.** 1994. Biochemical identification of citrobacteria in the clinical laboratory. J Clin Microbiol **32:**1850-1854.
4. **Barthold SW, Coleman GL, Bhatt PN, Osbaldiston GW, Jonas AM.** 1976. The etiology of transmissible murine colonic hyperplasia. Lab Anim Sci **26:**889-894.
5. **Schauer DB, Zabel BA, Pedraza IF, O'Hara CM, Steigerwalt AG, Brenner DJ.** 1995. Genetic and biochemical characterization of *Citrobacter rodentium* sp. nov. J Clin Microbiol **33:**2064-2068.
6. **Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, Quail MA, Lennard N, Corton C, Barron A, Clark L, Toribio AL, Parkhill J, Dougan G, Frankel G, Thomson NR.** 2010. The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. J Bacteriol **192:**525-538.
7. **Collins JW, Keeney KM, Crepin VF, Rathinam VA, Fitzgerald KA, Finlay BB, Frankel G.** 2014. Citrobacter rodentium: infection, inflammation and the microbiota. Nat Rev Microbiol **12:**612-623.
8. **Bouladoux N, Harrison OJ, Belkaid Y.** 2017. The mouse model of infection with *Citrobacter rodentium*. Curr Protoc Immunol **119:**19 15 11-19 15 25.
9. **Suttle CA.** 2007. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol **5:**801-812.
10. **Scanlan PD.** 2017. Bacteria-bacteriophage coevolution in the human gut: implications for microbial diversity and functionality. Trends Microbiol **25:**614-623.
11. **De Sordi L, Khanna V, Debarbieux L.** 2017. The gut microbiota facilitates drifts in the genetic diversity and infectivity of bacterial viruses. Cell Host Microbe **22:**801-808 e803.
12. **Roach DR, Debarbieux L.** 2017. Phage therapy: awakening a sleeping giant. Emerging Topics in Life Sciences **1:**93.
13. **Roach DR, Leung CY, Henry M, Morello E, Singh D, Di Santo JP, Weitz JS, Debarbieux L.** 2017. Synergy between the host immune system and bacteriophage is essential for successful phage therapy against an acute respiratory pathogen. Cell Host Microbe **22:**38-47 e34.
14. **LeSage KC, Hargrove EC, Cahill JL, Rasche ES, Kuty Everett GF.** 2015. Complete genome sequence of *Citrobacter freundii* myophage Merlin. Genome Announc **3**.
15. **Hwang K, Luna AJ, Hernandez AC, Kuty Everett GF.** 2015. Complete genome sequence of *Citrobacter freundii* myophage Miller. Genome Announc **3**.
16. **Edwards GB, Luna AJ, Hernandez AC, Kuty Everett GF.** 2015. Complete genome sequence of *Citrobacter freundii* myophage Moon. Genome Announc **3**.
17. **Bernal CL, Berkowitz VE, Cahill JL, Rasche ES, Kuty Everett GF.** 2015. Complete genome sequence of *Citrobacter freundii* myophage Michonne. Genome Announc **3**.
18. **Nguyen QT, Luna AJ, Hernandez AC, Kuty Everett GF.** 2015. Complete genome sequence of *Citrobacter freundii* myophage Moogle. Genome Announc **3**.
19. **Guan J, Snowden JD, Cahill JL, Rasche ES, Kuty Everett GF.** 2015. Complete genome sequence of *Citrobacter freundii* myophage Mordin. Genome Announc **3**.
20. **Shaw JP, Aviles Medina CA, Chen Y, Luna AJ, Hernandez AC, Kuty Everett GF.** 2015. Complete genome of *Citrobacter freundii* siphophage Stevie. Genome Announc **3**.
21. **Zhao X, Huang S, Zhao J, He X, Li E, Li H, Liu W, Zou D, Wei X, Wang X, Dong D, Yang Z, Yan X, Shen Z, Yuan J.** 2016. Characterization of phiCFP-1, a virulent bacteriophage specific for *Citrobacter freundii*. J Med Virol **88:**895-905.
22. **Hamdi S, Rousseau GM, Labrie SJ, Kourda RS, Tremblay DM, Moineau S, Slama KB.** 2016. Characterization of five podoviridae phages infecting *Citrobacter freundii*. Front Microbiol **7:**1023.
23. **Toribio AL, Pickard D, Cerdeno-Tarraga AM, Petty NK, Thomson N, Salmond G, Dougan G.** 2014. Complete genome sequences of two *Citrobacter rodentium* bacteriophages, CR8 and CR44b. Genome Announc **2:**e00146-00114.
24. **Ackermann HW.** 2003. Bacteriophage observations and evolution. Res Microbiol **154:**245-251.
25. **Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit MA.** 2017. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J **11:**237-247.
26. **Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M.** 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics **14:**60.
27. **Tikhe CV, Martin TM, Gissendanner CR, Husseneder C.** 2015. Complete Genome Sequence of Citrobacter phage CVT22 isolated from the gut of the formosan subterranean termite, Coptotermes formosanus Shiraki. Genome Announc **3**.
28. **Montag D, Hashemolhosseini S, Henning U.** 1990. Receptor-recognizing proteins of T-even type bacteriophages. The receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. J Mol Biol **216:**327-334.
29. **Rivas L, Coffey B, McAuliffe O, McDonnell MJ, Burgess CM, Coffey A, Ross RP, Duffy G.** 2010. In vivo and ex vivo evaluations of bacteriophages e11/2 and e4/1c for use in the control of *Escherichia coli* O157:H7. Appl Environ Microbiol **76:**7210-7216.
30. **Petty NK, Toribio AL, Goulding D, Foulds I, Thomson N, Dougan G, Salmond GP.** 2007. A generalized transducing phage for the murine pathogen *Citrobacter rodentium*. Microbiology **153:**2984-2988.
31. **Scholl D, Adhya S, Merril CR.** 2002. Bacteriophage SP6 is closely related to phages K1-5, K5, and K1E but encodes a tail protein very similar to that of the distantly related P22. J Bacteriol **184:**2833-2836.
32. **Scholl D, Kieleczawa J, Kemp P, Rush J, Richardson CC, Merril C, Adhya S, Molineux IJ.** 2004. Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup. J Mol Biol **335:**1151-1171.
33. **Krupovic M, Dutilh BE, Adriaenssens EM, Wittmann J,**

Vogensen FK, Sullivan MB, Rumnieks J, Prangishvili D, Lavigne R, Kropinski AM, Klumpp J, Gillis A, Enault F, Edwards RA, Duffy S, Clokie MR, Barylski J, Ackermann HW, Kuhn JH. 2016. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. Arch Virol **161:**1095-1099.

34. **Adriaenssens EM, Krupovic M, Knezevic P, Ackermann HW, Barylski J, Brister JR, Clokie MR, Duffy S, Dutilh BE, Edwards RA, Enault F, Jang HB, Klumpp J, Kropinski AM, Lavigne R, Poranen MM, Prangishvili D, Rumnieks J, Sullivan MB, Wittmann J, Oksanen HM, Gillis A, Kuhn JH.** 2017. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. Arch Virol **162:**1153-1157.
35. **Li S, Fan H, An X, Fan H, Jiang H, Chen Y, Tong Y.** 2014. Scrutinizing virus genome termini by high-throughput sequencing. PLoS One **9:**e85806.
36. **Mundy R, MacDonald TT, Dougan G, Frankel G, Wiles S.** 2005. *Citrobacter rodentium* of mice and man. Cell Microbiol **7:**1697-1706.
37. **Petty NK, Feltwell T, Pickard D, Clare S, Toribio AL, Fookes M, Roberts K, Monson R, Nair S, Kingsley RA, Bulgin R, Wiles S, Goulding D, Keane T, Corton C, Lennard N, Harris D, Willey D, Rance R, Yu L, Choudhary JS, Churcher C, Quail MA, Parkhill J, Frankel G, Dougan G, Salmond GP, Thomson NR.** 2011. *Citrobacter rodentium* is an unstable pathogen showing evidence of significant genomic flux. PLoS Pathog **7:**e1002018.
38. **Hyman P, Abedon ST.** 2010. Bacteriophage host range and bacterial resistance. Adv Appl Microbiol **70:**217-248.
39. **Dufour N, Clermont O, La Combe B, Messika J, Dion S, Khanna V, Denamur E, Ricard JD, Debarbieux L, ColoColi g.** 2016. Bacteriophage LM33_P1, a fast-acting weapon against the pandemic ST131-O25b:H4 *Escherichia coli* clonal complex. J Antimicrob Chemother **71:**3072-3080.
40. **Hamdi S, Rousseau GM, Labrie SJ, Tremblay DM, Kourda RS, Ben Slama K, Moineau S.** 2017. Characterization of two polyvalent phages infecting Enterobacteriaceae. Sci Rep **7:**40349.
41. **Choi IY, Lee JH, Kim HJ, Park MK.** 2017. Isolation and characterization of a novel broad-host-range bacteriophage infecting *Salmonella enterica* subsp. *enterica* for biocontrol and rapid detection. J Microbiol Biotechnol **27:**2151-2155.
42. **Hsieh SE, Lo HH, Chen ST, Lee MC, Tseng YH.** 2011. Wide host range and strong lytic activity of *Staphylococcus aureus* lytic phage Stau2. Appl Environ Microbiol **77:**756-761.
43. **Roach DR, Sjaarda DR, Castle AJ, Svircev AM.** 2013. Host exopolysaccharide quantity and composition impact *Erwinia amylovora* bacteriophage pathogenesis. Appl Environ Microbiol **79:**3249-3256.
44. **Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB.** 2013. Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev **26:**822-880.
45. **Crepin VF, Collins JW, Habibzay M, Frankel G.** 2016. *Citrobacter rodentium* mouse model of bacterial infection. Nat Protoc **11:**1851-1876.
46. **Wiles S, Clare S, Harker J, Huett A, Young D, Dougan G, Frankel G.** 2004. Organ specificity, colonization and clearance dynamics *in vivo* following oral challenges with the murine pathogen *Citrobacter rodentium*. Cell Microbiol **6:**963-972.
47. **Pickard DJ.** 2009. Preparation of bacteriophage lysates and pure DNA. Methods Mol Biol **502:**3-9.
48. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.
49. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955-964.
50. **Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
51. **Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O.** 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res **29:**41-43.
52. **Soding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33:**W244-248.
53. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460-2461.
54. **Meier-Kolthoff JP, Goker M.** 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics **33:**3396-3404.
55. **Lefort V, Desper R, Gascuel O.** 2015. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol **32:**2798-2800.

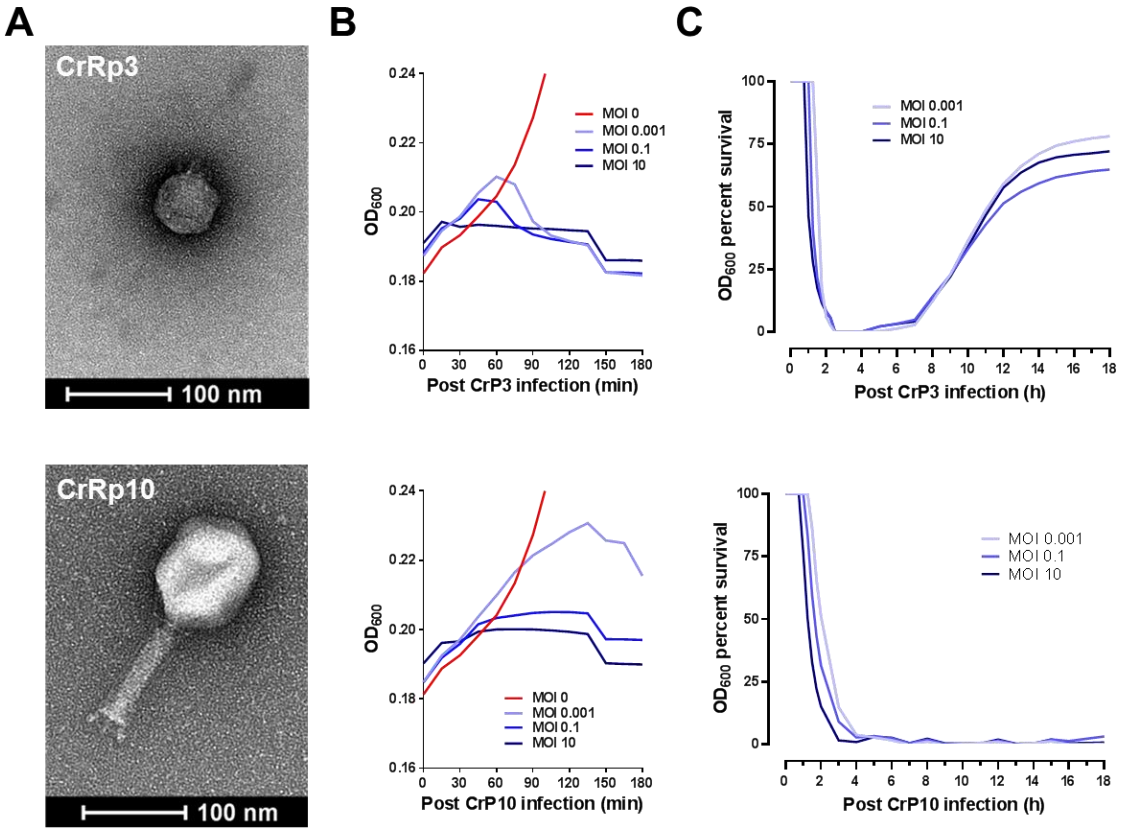

Figure 1. Morphological and Biological characterization of CrRp3 and CrRp10. A) Electron micrographs of *Citrobacter rodentium* phages negatively stained uranyl acetate. **B)** Early phage lysis dynamics and **C)** percent survival of *C. rodentium* cell populations at different initial multiplicity of infection (MOI) of CrRp3 (top) and CrRp10 (bottom) compared to growth in uninfected cultures (n = 3).

Table 1. Citrobacter bacteriophages and their genome features.

Phage	Host	Source	Phage family*	Size (kb)	GC%	Accession no.	Ref.
CrRp3	*C. rodentium*	municipal wastewater	P	44.3	45.1	MG775042	This study
CR44b	*C. rodentium*	sewage effluent	P	39.2	50.5	NC_023576	(23)
CR8	*C. rodentium*	sewage effluent	P	39.7	49.7	NC_023548	(23)
CVT22	*Citrobacter sp.*	termite gut	P	47.6	41.6	NC_027988	(27)
phiCFP-1	*C. freundii*	seawater	P	38.6	50.3	NC_028880	N/A
SH1	*C. freundii*	"	P	39.4	51	NC_031066	N/A
SH2	*C. freundii*	"	P	39.2	50.7	NC_031092	N/A
SH3	*C. freundii*	"	P	39.4	50.6	NC_031123	N/A
SH4	*C. freundii*	"	P	39.3	52.6	NC_031018	N/A
CrRp10	*C. rodentium*	municipal wastewater	M	171.5	35.5	MG775043	This study
IME-CF2	*C. freundii*	hospital wastewater	M	177.7	43.2	NC_029013	N/A
Margaery	*C. freundii*	"	M	178.2	44.9	NC_028755	N/A
Merlin	*C. freundii*	"	M	172.7	38.8	NC_028857	(14)
Michonne	*C. freundii*	"	M	90.0	38.8	NC_028247	(17)
Miller	*C. freundii*	"	M	178.2	43.1	NC_025414	(15)
Moogle	*C. freundii*	"	M	88.0	39	NC_027293	(18)
Moon	*C. freundii*	"	M	170.3	38.9	NC_027331	(16)
CfP1	*C. freundii*	sewage effluent	M	180.2	43.1	NC_031057	N/A
Stevie	*C. freundii*	soil	S	49.8	42.8	NC_027350	(20)

*Podoviridae (P), Myoviridae (M), Syphoviridae (S)

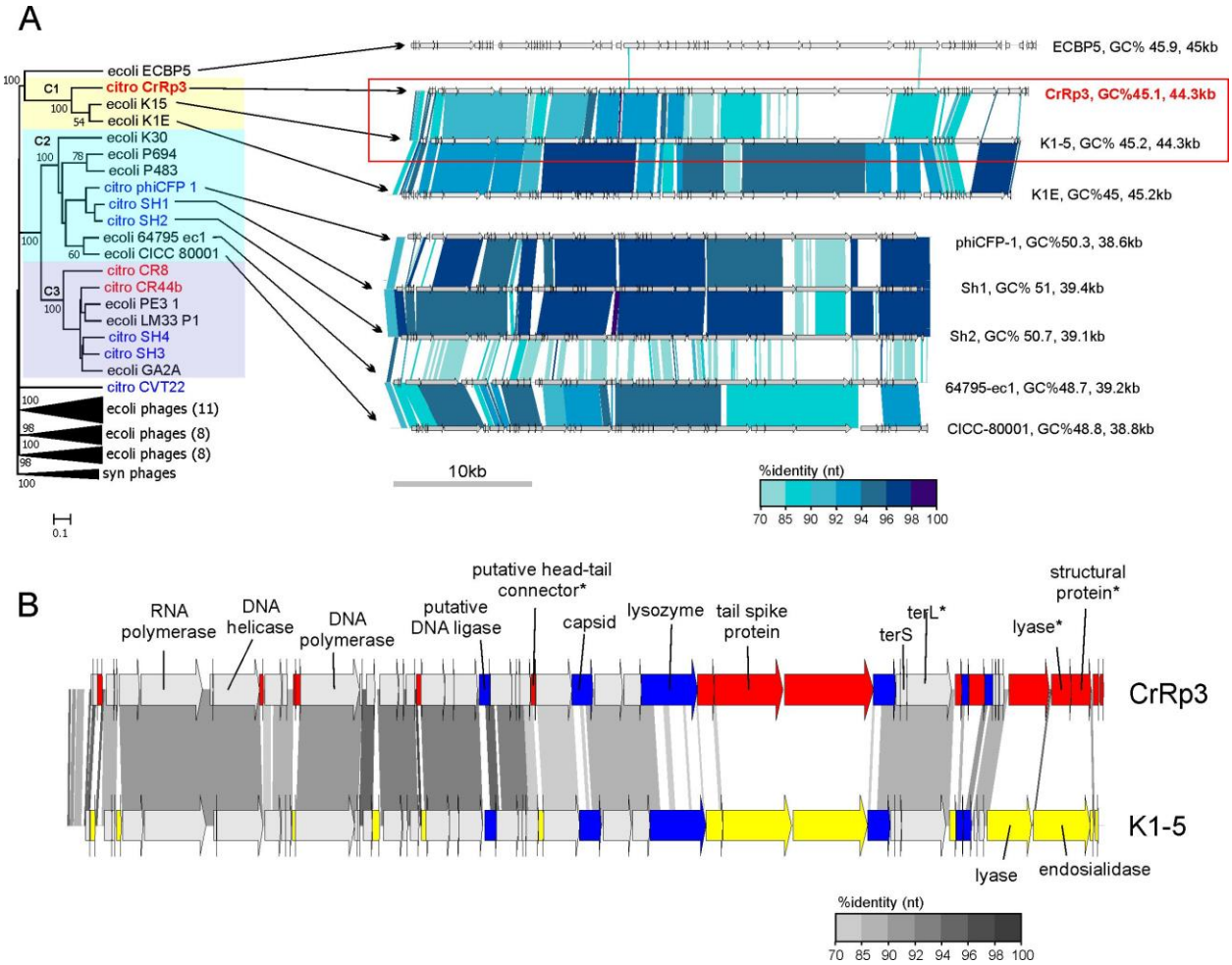

Figure 2. Genome structure and phylogeny of *Citrobacter rodentium* phage CrRp3. A) Genomic relationship of CrRp3 with other *Citrobacter* and *Escherichia* phages at nucleotide level. The tree shows bootstrap values (percentages of 100 replicates) below the branches and was rooted using *Synechococcus* phages as outgroup. Phages reported to infect *C. rodentium* are labeled in red and those that infect *C. freundii* and phage CVT22 are labeled in blue. **B)** Gene functional comparison of CrRp3 and *E. coli* phage K1-5. Genes are colored according to the relationship between CrRp3 and K1-5, with red labels being exclusive to CrRp3, yellow labels being exclusive to K1-5, while blue labels are homologous but highly variable. Gene products marked with (*) are those with some similarity to other *Citrobacter* phages (see Fig. 4).

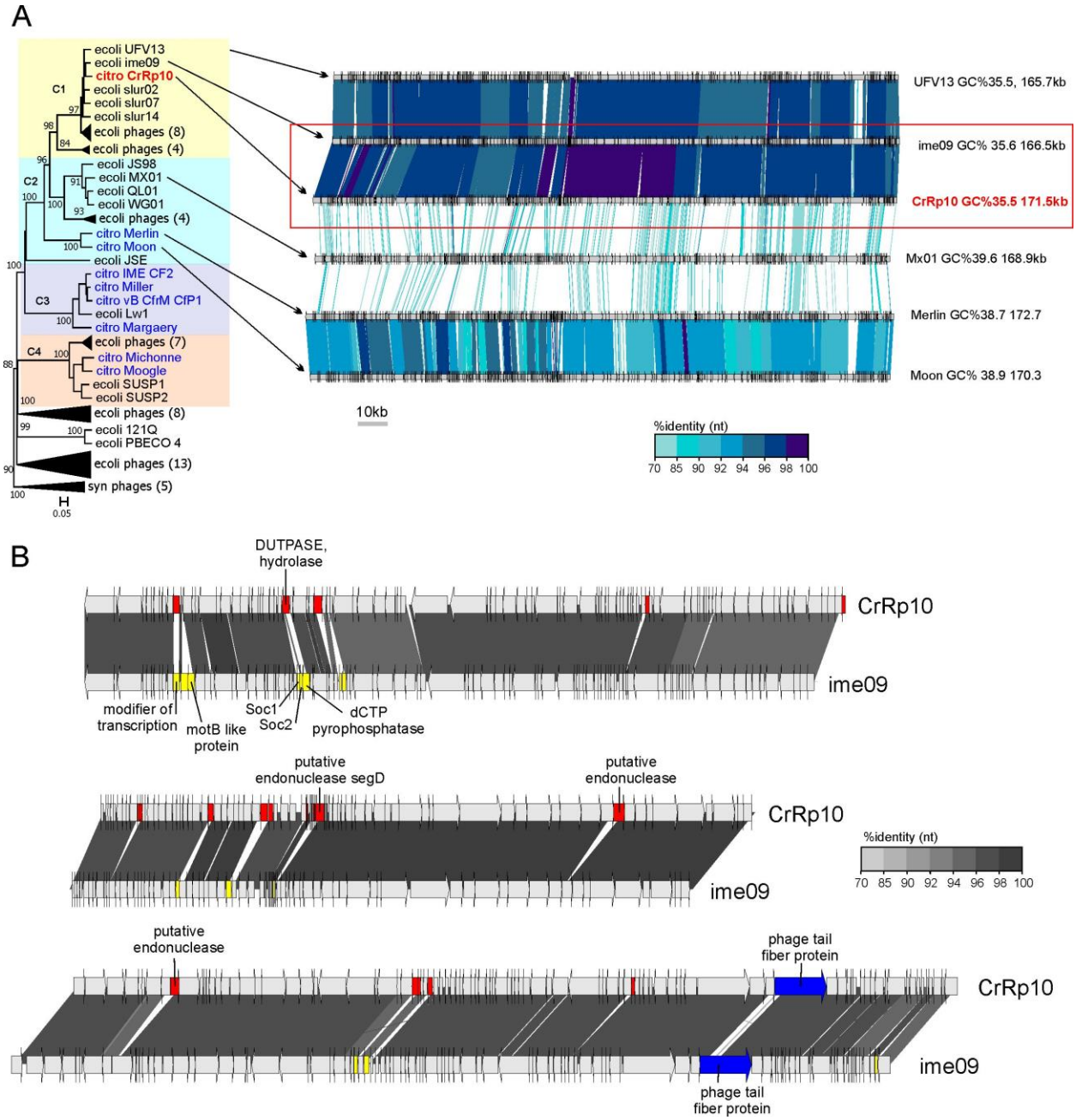

Figure 3. Genome structure and phylogeny of *Citrobacter rodentium* phage CrRp10. A) Genomic relationship of CrRp10 with other *Citrobacter* and *Escherichia* phages at nucleotide level. The tree shows bootstrap values (percentages of 100 replicates) below the branches and was rooted using *Synechococcus* phages as outgroup. Phages reported to infect *C. rodentium* are labeled in red and those that infect *C. freundii* are labeled in blue. **B)** Gene functional comparison of CrRp10 and *E. coli* phage ime09. Genes are colored according to the relationship between CrRp10 and ime09, with red labels being exclusive to CrRp10, yellow labels being exclusive to ime09, while blue labels are homologous but highly variable.

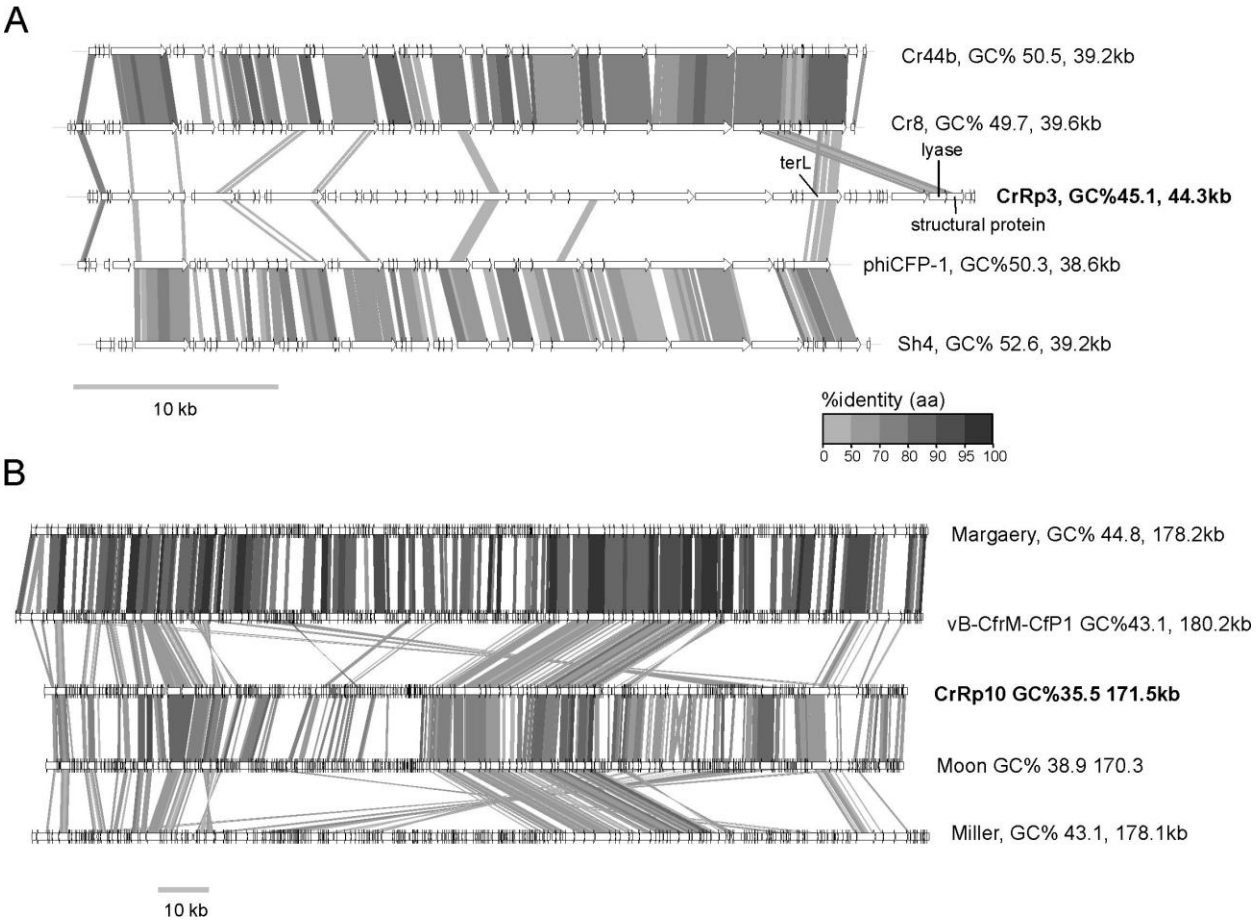

Figure 4. Whole genome alignment of *C. rodentium* phages at the amino acid level. Virulent *Podoviridae* **A)** and *Myoviridae* **B)**.

Table 2. Bacteriophage host ranges.

			Bacteriophage Strain[b]							
Species	**Strain**	**Type[a]**	**CrRp3**	**CrRp10**	**LF82_P10**	**LM33_P1**	**536_P7**	**AL505_P2**	**55989_P2**	**PAK_P1**
Citrobacter rodentium	ICC180		●	●	○	○	○	○	○	○
Erwinia carotovora	CFBL2141		○	●	●					○
Escherichia coli	LF31	AIEC	○	◗	●	○	○	○	◗	○
	LF50	AIEC	○	◗	○	○	●	○	○	○
	LF73	AIEC	○	●	○	○	○	○	○	○
	LF82	AIEC	○	○	●	○	○	◗	○	○
	LF110	AIEC	○	○	●	○	○	○	◗	○
	LF06075	AIEC	○	●	●	○	●	○	●	○
	LF07081	AIEC	○	○	○	○	○	○	○	○
	NRG	AIEC	○	●	●	○	○	○	○	○
	55989	EAEC	○	○	○	○	○	●	●	○
	E22	EPEC	○	○	○	○	○	○	○	○
	LM33	ExPEC	○	○	○	●	○	○	○	○
	Sp15	ExPEC	○	○	○	○	○	○	○	○
	536	UPEC	○	○	○	○	○	○	○	○
	AL505	UPEC	○	○	○	○	○	○	○	○
	MG1655	K-12	○	○	○		●		●	○
	BW25113	K-12	◗	◗	◗					○
	CR63	K-12	○	○	●					○
	OP_{50}	B	○	○	○					○
	BE	B	○	○	◗			●	●	○
	Nissle 1917	probiotic	○	●	●	○	◗	○	○	○
	M1/5		○	○	●					○
	SE15		○	●	○		●			○
	LM40		○	○	○	○	○	○	○	○
	H04		○	○	○	○	○	○	○	○
	BDX03		○	○	○	○	○	○	○	○
	LM02		○	○	○	○	○	○	○	○
	LM08		○	○	○	●	○	○	○	○
Pseudomonas aeruginosa	PAK		○	○	○	○	○	○	○	●
Rouxiella chamberiensis	nov		○	○	○					○
Serratia marcescens	Db11		○	○	○					○
	SM365		○	○	○					○

[a] *Adherent invasive E. coli (AIEC), enteroaggregative E. coli (EAEC), enteropathogenic E. coli (EPEC), extraintestinal pathogenic E. coli (ExPEC), and uropathogenic E. coli (UPEC)*

[b] Bacteriophages applied at 10^7 plaque forming units; ● clear plaque, ◗ partial clearing, ○ resistant, (blank) not tested