Diversity of *Pseudomonas aeruginosa* Temperate Phages

Genevieve Johnson,ᵃ Swarnali Banerjee,ᵇ Catherine Putontiᵃ,ᶜ,ᵈ

ᵃBioinformatics Program, Loyola University Chicago, Chicago, Illinois, USA
ᵇDepartment of Mathematics and Statistics, Loyola University Chicago, Chicago, Illinois, USA
ᶜDepartment of Biology, Loyola University Chicago, Chicago, Illinois, USA
ᵈDepartment of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois, USA

ABSTRACT   Modern sequencing technologies have provided insight into the genetic diversity of numerous species, including the human pathogen *Pseudomonas aeruginosa*. Bacterial genomes often harbor bacteriophage genomes (prophages), which can account for upwards of 20% of the genome. Prior studies have found *P. aeruginosa* prophages that contribute to their host's pathogenicity and fitness. These advantages come in many different forms, including the production of toxins, promotion of biofilm formation, and displacement of other *P. aeruginosa* strains. While several different genera and species of *P. aeruginosa* prophages have been studied, there has not been a comprehensive study of the overall diversity of *P. aeruginosa*-infecting prophages. Here, we present the results of just such an analysis. A total of 6,852 high-confidence prophages were identified from 5,383 *P. aeruginosa* genomes from strains isolated from the human body and other environments. In total, 3,201 unique prophage sequences were identified. While 53.1% of these prophage sequences displayed sequence similarity to publicly available phage genomes, novel and highly mosaic prophages were discovered. Among these prophages, there is extensive diversity, including diversity within the functionally conserved integrase and C repressor coding regions, two genes responsible for prophage entering and persisting through the lysogenic life cycle. Analysis of integrase, C repressor, and terminase coding regions revealed extensive reassortment among *P. aeruginosa* prophages. This catalog of *P. aeruginosa* prophages provides a resource for future studies into the evolution of the species.

IMPORTANCE Prophages play a critical role in the evolution of their host species and can also contribute to the virulence and fitness of pathogenic species. Here, we conducted a comprehensive investigation of prophage sequences from 5,383 publicly available *Pseudomonas aeruginosa* genomes from human as well as environmental isolates. We identified a diverse population of prophages, including tailed phages, inoviruses, and microviruses; 46.9% of the prophage sequences found share no significant sequence similarity with characterized phages, representing a vast array of novel *P. aeruginosa*-infecting phages. Our investigation into these prophages found substantial evidence of reassortment. In producing this, the first catalog of *P. aeruginosa* prophages, we uncovered both novel prophages as well as genetic content that have yet to be explored.

KEYWORDS *Pseudomonas aeruginosa*, prophages, temperate phages

High-throughput short-read and long-read sequencing technologies have produced >300,000 prokaryotic genomes to date. Annotation and analysis of these genomes have found phage genes and prophages, which can be integrated into the bacterial genome or persist as an extrachromosomal plasmid. Early estimates found that prophages account for 10 to 20% of the host's genome (1). The presence of a prophage within the

TABLE 1 Summary statistics of VirSorter prophage prediction results

No. of prophages predicted for category:					
1	2	3	4	5	6
2,668	12,579	5,666	4,184	17,367	6,638

Categories 1 and 4 are the highest-confidence predictions. Categories 3 and 6 are the lowest-confidence predictions, e.g., typically partial phages or phage-like genes.

bacterial genome can give the host a selective advantage (2, 3). Prophages can also protect the bacterial cell from environmental stressors and confer antibiotic resistance to the bacterial host (4). While some prophage sequences may persist only through the lysogenic life cycle, others, temperate phages, are able to also display the lytic life cycle. These temperate phages can contribute greatly to the wide genomic diversity of bacteriophages through horizontal gene transfer and recombination (5).

Prophage and phage sequence prediction tools, including Phage Finder (6), PHAST (7), PHASTER (8), Prophinder (9), and VirSorter (10), have been used to catalog prophage sequences from microbial communities (11, 12) and publicly available genomes of several bacterial taxa, e.g., *Staphylococcus aureus* (13), *Salmonella*, and *Escherichia* species (6, 9, 14–16), *Mycobacterium smegmatis* (17), *Klebsiella pneumoniae* (18), and *Staphylococcus pneumoniae* (19). The use of phage sequence prediction tools throughout these studies shows the high prevalence of prophages within a multitude of different bacterial genera. These studies have also shown that while there are several conserved gene regions, e.g., integrases and terminases, between phages that infect the same bacterial species, there is also distinct mosaicism that results in broad phage diversity.

While numerous obligately lytic phages have been described for *Pseudomonas aeruginosa*, research into *P. aeruginosa* prophages has primarily focused on those associated with increased bacterial pathogenicity and fitness. Prophage genes carried by *P. aeruginosa* can have a profound effect on bacterial phenotype, competitiveness, and pathogenicity (20). The Liverpool epidemic strain (LES) of *P. aeruginosa*, which has been associated with infections in cystic fibrosis patients, harbors 5 prophages: LESϕ2, LESϕ3, LESϕ4, LESϕ5, and LESϕ6 (21). Not only do the prophages improve the invasiveness of LES but they also assist in the displacement of other *P. aeruginosa* strains by LES, allowing LES full competitive advantage in lung infections (21). The *P. aeruginosa* phage FIZ15 causes lysogenic conversion in *P. aeruginosa* PAO1, leading to PAO1's increased resistance to phagocytosis, increased resistance to normal human serum, and increased adhesion to human epithelial cells (22). *P. aeruginosa* prophages can also be responsible for producing cytotoxins, as is the case for the *Pseudomonas* φCTX phage (23). In yet another example, *Inoviridae* prophages, such as the Pf family of prophages, can promote biofilm formations in *P. aeruginosa* (24). The results of Knezevic et al. (25) suggest that Pf1-like prophages are abundant among *P. aeruginosa* strains.

While there have been a few studies focused on the diversity of specific types of *P. aeruginosa* phages, there has yet to be a comprehensive study on the diversity of prophages infectious to *P. aeruginosa*. Here, we conduct an investigation of all publicly available *P. aeruginosa* genomes, which now exceeds 5,383 genomes, with particular focus on temperate, tailed *P. aeruginosa*-infecting phages. Cataloging *P. aeruginosa* temperate phages serves two purposes. First, it provides insight into the diversity and putative roles that phages play in *P. aeruginosa* fitness and pathogenicity. Second, temperate phages can be viable candidates for phage therapy. Prior studies have explored *P. aeruginosa* phage use in animal models (26–30) as well as in humans (31–68).

RESULTS

Prophages of *P. aeruginosa*. VirSorter identified 49,102 putative prophage sequences in the 5,383 publicly available *P. aeruginosa* genomes examined here (Table 1). These genomes represent isolates from various human body sites, the environment, and industrial settings. The isolation sites were significantly associated with the number of predicted

prophage sequences for the strains of *P. aeruginosa* after normalization ($P < 2.2e-16$). Of the total predicted prophages, 6,852 prophages were of category 1 and category 4, the most confident categories for unintegrated and integrated prophages, respectively. Per VirSorter's documentation, sequences predicted within these two categories have significant enrichment of virus-like genes and/or non-*Caudovirales* genes over the entire predicted region and encode at least one hallmark viral gene (10). The predicted prophage sequences can be found at http://doi.org/10.5281/zenodo.5072377. Given that these high-confidence predictions likely represent viable temperate phages, further analysis was restricted to category 1 and 4 prophage sequences only. A total of 3,672 *P. aeruginosa* genomes encoded prophage sequences in these categories. While these genomes had on average 1 or 2 prophages, *P. aeruginosa* XDR-PA (GCA_900707735.1) harbored 15 predicted prophage sequences.

Genetic diversity of *P. aeruginosa* prophages. The 6,852 category 1 and 4 predicted prophages were clustered based on sequence homology in an effort to ascertain the number of unique prophages that were identified. In total, there are 3,201 unique clusters of prophages. The largest cluster contains 169 phages; 2,496 of these clusters contain only 1 phage, meaning that the predicted prophage sequence was not detected in any of the other *P. aeruginosa* genomes examined (see Fig. S1A in the supplemental material). Similar prophage sequences, i.e., prophage sequences belonging to the same cluster, can be the result of (i) pervasive phages or (ii) genomes from the same strain and/or clonal strains. The genomic diversity of *P. aeruginosa* genomes harboring prophages belonging to the same cluster were examined further using average nucleotide identity (ANI) as a means of assessing genomic similarity. For the largest cluster, the *P. aeruginosa* strains had an ANI of 99.86% (Fig. S1B). The second largest cluster, $n = 126$, had an ANI of 99.00% (Fig. S1C). Neither of these clusters included genome sequences with a pairwise ANI of 100%. Nevertheless, the lack of complete metadata for publicly available genomes limits our ability to distinguish between the two scenarios and, thus, the ability to consider pervasiveness of prophage species.

Next, the taxonomy of the 6,852 prophages was determined by querying each predicted prophage sequence against all publicly available characterized, sequenced phages. This resulted in 11 classified as *Inoviridae*, 106 as *Microviridae*, 672 as *Myoviridae*, 105 as *Podoviridae*, and 2,744 as *Siphoviridae*. The remaining 3,214 prophages did not share significant sequence similarity to any characterized, sequenced phages; thus, a taxonomic classification could not be made.

The diversity of *P. aeruginosa* prophages was visualized through a network consisting of the predicted prophages as nodes and the connecting edges representative of the number of genes shared between each prophage. Further investigation into the category 1 and 4 prophage sequences revealed high-confidence predictions of sequences too small to be a viable phage. In an effort to focus our investigation on complete (and likely viable) prophages, we introduced a threshold; only edges representative of 5 or more shared genes between prophages (nodes) were retained; nodes that were not connected to any other node were removed from further consideration. After this thresholding, 6,676 of the originally predicted 6,852 (97.43%) prophages remained (Fig. 1). The network contained a total of 3,814,212 edges representing the genes shared between the prophages. It contains 2 *Inoviridae* prophages, 105 *Microviridae* prophages, 635 *Myoviridae* prophages, 99 *Podoviridae* prophages, 2,662 *Siphoviridae* prophages, and 3,173 prophages for which no taxonomic classification could be made (unknown).

The prophages cluster within 5 connected components (CC), labeled CC1, CC2, CC3, CC4, and CC5 in Fig. 1. The majority (6,513 prophages; 97.56%) belonged to one large connected component, CC1, which includes tailed prophages (*Myoviridae*, *Podoviridae*, and *Siphoviridae*), the *Inoviridae* prophages, and unclassified (unknown) prophages. The 105 *Microviridae* prophages (green nodes in Fig. 1) belong to their own connected component, CC2, sharing genes only among other microviruses. CC4 in Fig. 1 contains 21 *Siphoviridae* prophages, which are distinctly different from other *Siphoviridae* prophage sequences within the largest connected component. These prophage sequences were
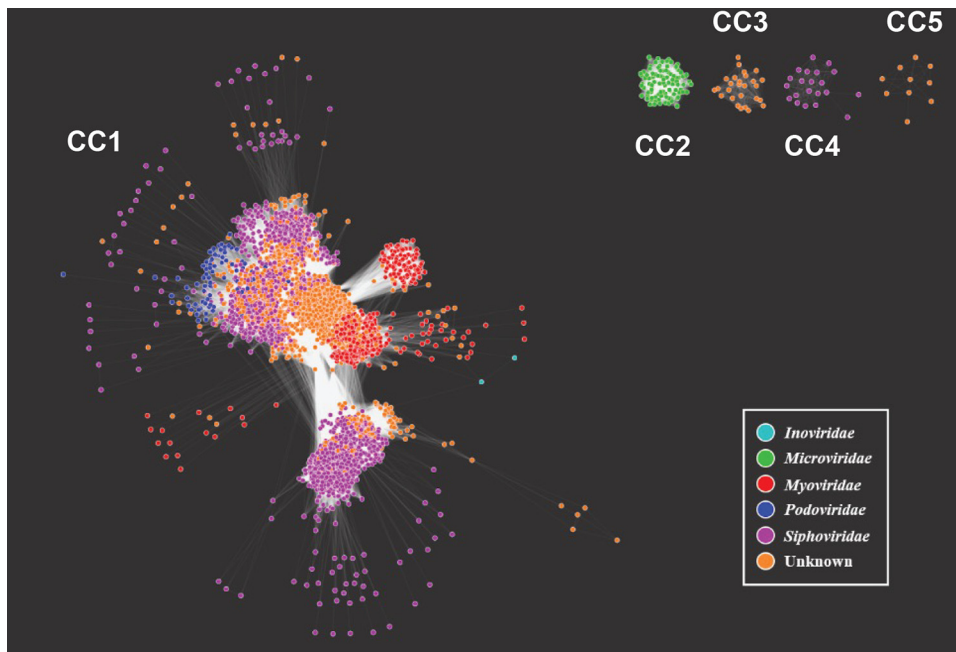
FIG 1 Network of *P. aeruginosa* predicted prophages. Prophages are the nodes color-coded by taxonomy: *Inoviridae*, *Microviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*, or unknown. Prophages are connected by the number of shared genes between each prophage. Five separate connected components (CC) represent these prophages, labeled CC1 through CC5.

identified from 18 different *P. aeruginosa* genomes, derived from 3 different studies. Further investigation of these sequences revealed genetic homology to strains of *Escherichia coli* as well as to the *E. coli* lambda phage (Table S1).

CC3 and CC5 contained prophages classified as unknown (orange nodes in Fig. 1). The first of the unknown clusters consisted of 27 prophages predicted from 27 different genomes of *P. aeruginosa*. While BLAST queries of these prophage sequences revealed homology to a pyocin gene fragment (L06240.1), BAGEL4 (69), a bacteriocin prediction tool, did not predict these sequences to be bacteriocins. The other cluster classified as unknown contained 10 prophages predicted from 7 different *P. aeruginosa* genomes. The prophage sequences were input to BLAST and all found to have high similarity (over 75%) to at least 1 of 2 uncultured *Caudovirales* phages (clone 3S_15, MF417945.1; clone 3S_19, MF417971.1), both identified through a skin metavirome project. Five of these predicted prophages were also similar (>70% identity and 8 to 24% query coverage) to the partial genomes of 2 *Myoviridae* isolates (isolate ctQcn1, BK028718.1; isolate ct2Fx3, BK037698.1). Further investigation of this cluster is required to make a taxonomic classification.

In a prior study, Pf1-like phages (*Inoviridae*) were found to be prevalent in *P. aeruginosa* strains, with ~60% of 241 strains screened via PCR to contain at least 1 Pf1-like genetic element (25). Thus, we expected to find many more inoviruses in the 5,383 genomes examined. Prior studies have documented that current prophage prediction tools, including VirSorter, frequently miss inoviruses (10, 70, 71). Pf1-like genetic elements were identified in an additional 93 predicted prophages of lower-confidence VirSorter categories. These sequences had a high sequence similarity (>75% identity), but identities were only between one or a few genes (query coverage averaged 20%). Thus, we specifically mined the 5,383 *P. aeruginosa* genomes for Pf1, identifying the prophage (sequence identity and query coverage of >99%) in 123 of the genomes (Table S2); *P. aeruginosa* strain PSE6684 (GenBank accession no. CP053917.1) was predicted to contain 2 complete Pf1 prophages. Thus, VirSorter missed most instances of this well-studied prophage. Given this result, we reframe our catalog to be a representation of the tailed phages within publicly available *P. aeruginosa* strains.
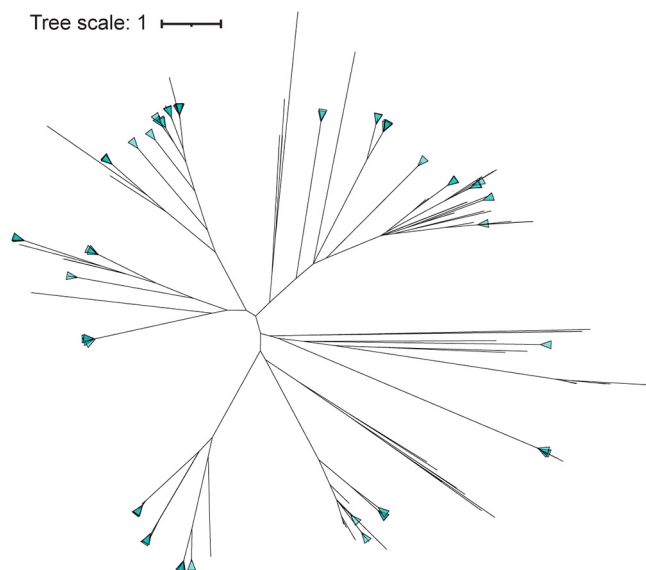
Tree scale: 1 ⊢——•——⊣

FIG 2 Phylogenetic tree of all annotated integrase coding regions.

Diversity of *P. aeruginosa* prophage integrases, C repressors, and terminase coding sequences. The associations between prophages that share homologous terminases and also share homologous integrases or C repressors were considered to investigate the genetic reassortment between prophage genomes that are likely to be temperate. Terminases are genes typically conserved among tailed phages. Prophage sequence annotation identified 283,493 coding regions within the 6,852 predicted prophage sequences; 14,172 of these coding regions are unique. Among these coding regions, 1,358 were annotated as integrases, 2,344 as C repressors, and 4,546 as terminases. To ensure that all integrase, C repressor, and terminase genes within these prophage sequences were included in our analysis, we conducted additional BLAST homology searches for representatives of these 3 genes. These queries identified an additional 12, 17, and 24 integrase, C repressor, and terminase coding regions, respectively. Thus, a total of 1,370 integrases, 2,361 C repressors, and 4,570 terminases are encoded within the predicted *P. aeruginosa* prophage sequences. These three marker genes were identified in 19.99%, 34.46%, and 66.34% of the predicted prophage sequences, respectively. A total of 850 prophages encoded both a terminase and an integrase, 391 prophages encoded both a terminase and a C repressor, and 713 prophages contained coding regions for all 3 of these genes (Fig. S2).

To investigate the diversity of integrase and C repressor sequences within the *P. aeruginosa* prophages, phylogenetic trees were derived. As Fig. 2 shows, there are distinct lineages of integrase genes among the prophages. Furthermore, these lineages were divergent, with an average amino acid pairwise identity of 16.1% ($n = 1,370$). We also examined the sequence diversity of the 2,361 C repressor coding sequences within the predicted *P. aeruginosa* prophages (Fig. 3). While distinct lineages were also observed, the set of C repressor sequences exhibited greater sequence similarity than the integrases; the average amino acid pairwise identity is 58.0%.

We next performed clustering of the integrase ($n = 1,370$), C repressor ($n = 2,361$), and terminase ($n = 4,570$) coding sequences. In total, 72 different clusters of integrases were identified, with the largest containing 110 integrase coding regions. The integrase sequences within this cluster exhibit little sequence variation (average pairwise identity, 98.3%). The C repressor sequences clustered into just 44 groups, with over half ($n = 1,604$; 67.94%) belonging to a single group. This largest cluster of C repressor coding regions contains an average pairwise identity of 99.7%. The 4,570 terminase coding regions clustered into 81 homologous groups. The largest of the terminase clusters contained 424 terminase coding regions with an average pairwise identity of 89.6%.
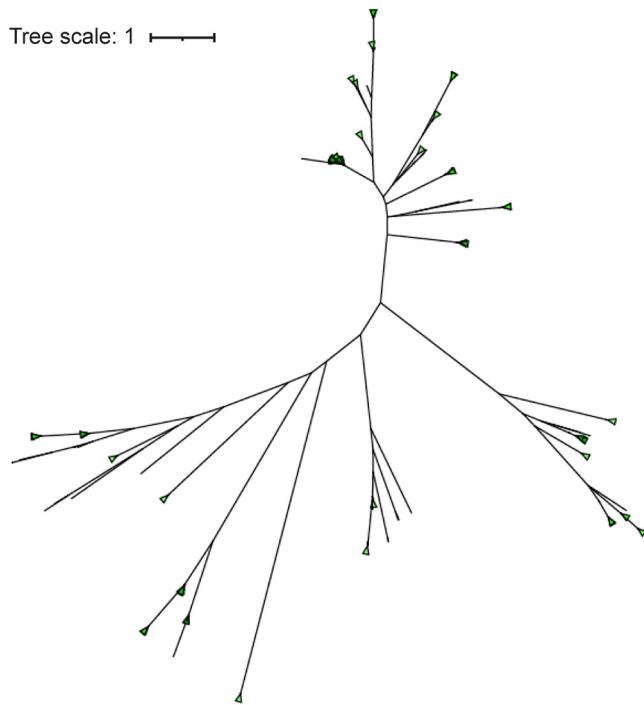
FIG 3 Phylogenetic tree of all annotated C repressor coding regions.

Each prophage was next associated with the integrase, C repressor, and terminase clusters identified in the previous section. Between all 72 integrase clusters and all 81 terminase clusters, 40 integrase clusters shared all of their prophages with individual terminase clusters and 3 terminase clusters shared all of their prophages with individual integrase clusters. Figure 4 displays the distribution of shared prophages between the integrase clusters and the terminase clusters, with the highest numbers of cooccur-
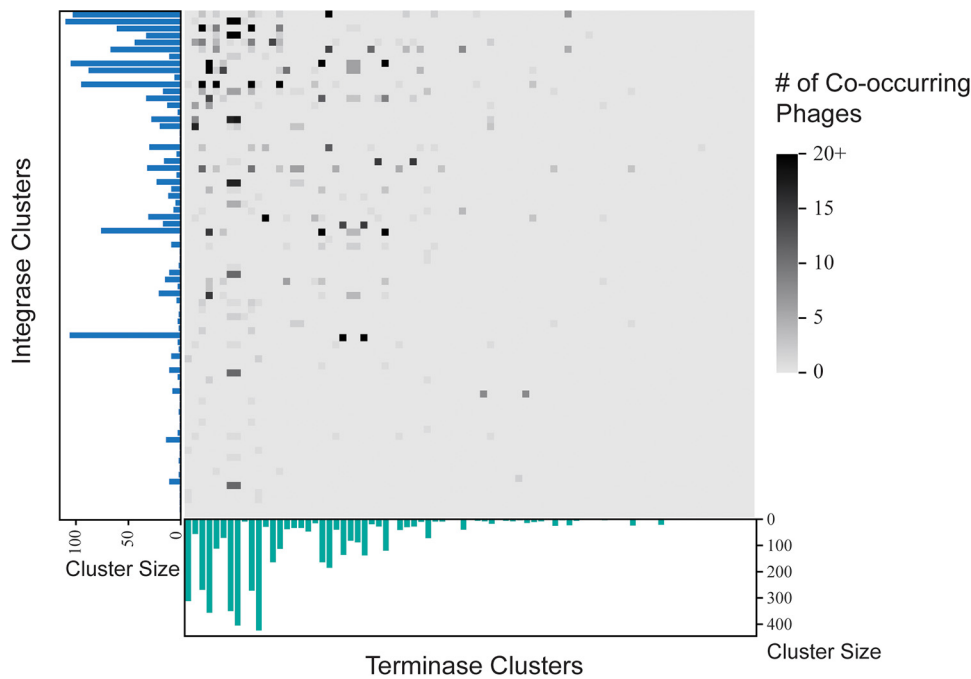

FIG 4 Prophage membership shared between integrase clusters and terminase clusters. The largest terminase clusters tend to share the most prophages with the largest integrase clusters.
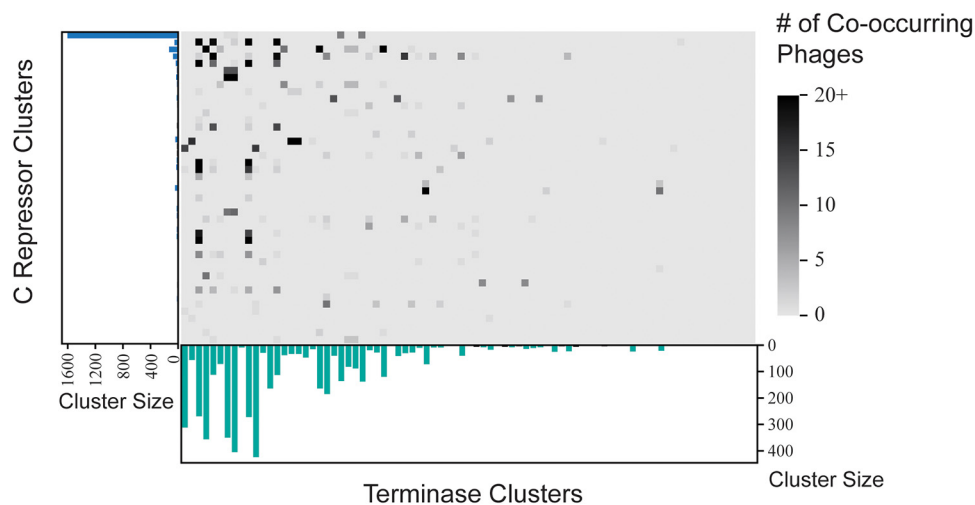
FIG 5 Prophage membership shared between C repressor clusters and terminase clusters. The largest terminase clusters tend to share the most prophages with the largest C repressor clusters.

ring prophages existing between the largest integrase clusters and the largest terminase clusters.

Of the 44 C repressor clusters and the 81 terminase clusters, 17 C repressor clusters shared all of their prophages with individual terminase clusters and 5 terminase clusters shared all of their prophages with individual C repressor clusters. The C repressor clusters and the terminase clusters in Fig. 5 display a deduction similar to that shown in Fig. 4, wherein the most shared prophages are between the largest C repressor clusters and the largest terminase clusters.

The integrase cluster membership of a phage is not associated with the terminase cluster membership of the phage; phages within terminase clusters contain integrases from significantly more than 1 integrase cluster type ($P = 8.275e-06$). Similarly, the C repressor cluster membership of a phage is not associated with the terminase cluster membership of the phage; phages within terminase clusters contain C repressors from significantly more than 1 C repressor cluster ($P = 8.079e-04$).

Antibiotic resistance, virulence factors, and metabolic genes encoded by prophages. All category 1 and 4 predicted prophage sequences were examined for antibiotic resistance genes; 71 predicted antibiotic resistance genes from 42 different prophage genomes were identified (Table S3). The prophage sequences were next screened for virulence factors; 502 prophages possessed a total of 515 predicted virulence factor sequences (Table S4). The 515 predicted sequences were composed of 72 unique virulence factor sequences for 63 unique virulence genes, and 41 prophages contained both predicted virulence factors and predicted antibiotic resistance genes.

Next, all category 1 and 4 predicted prophage sequences were examined for the functional categories of their gene content. Per COG classification, 2,686 unique prophage genes that are classified as metabolic genes and 331 COGs associated with metabolism were identified (Table S5). The most frequently detected putative auxiliary metabolite genes are associated with amino acid metabolism, membrane transport, methane metabolism, and carbohydrate metabolism. Auxiliary metabolite genes for metal resistance and transport were also identified (Table S5).

DISCUSSION

This is the first known report of the diversity of prophages among *P. aeruginosa* strains. We limited our analysis to only the highest category prophage predictions from VirSorter, categories 1 and 4. Other investigations of phages in genomic and metagenomic data sets have taken a less conservative approach, including VirSorter category 1, 2, 4, and 5 predictions (11, 70, 72). While others have noted that the lower

confidence categories (3 and 6) tend to include only partial prophage genomes (11, 70, 72), our own prior work predicting prophages in the urinary microbiome found that most (~71%) of the phage sequences in the category 2 and 5 predictions are not complete prophage sequences (11). While some nontailed phages were identified in the category 1 and 4 predictions, our direct search for the inovirus Pf1 (i) revealed many Pf1 prophages that were not detected, (ii) supports current literature acknowledging the limitations of current prophage prediction tools in identifying nontailed prophages, including inoviruses (10, 70, 71), and (iii) led to our focus on *P. aeruginosa*-infecting tailed prophages. Even with the stringent threshold applied, there are likely some false positives and prophage artifacts (nonfunctional prophages) included in these higher confidence predictions. For instance, 332 prophage predictions were less than 3,000 bp long, which is less than the smallest characterized *Pseudomonas* phage genome, PRR1 (NC_0082941), at 3,573 bp. These smaller predicted prophage sequences may be the result of erroneous calls by VirSorter or a result of fragmented bacterial assemblies, as most of the genomes examined here are draft assemblies.

The catalog of *P. aeruginosa* prophages includes 6,852 prophages from 5,383 genomes that were isolated from a variety of sources. We recognize that the genomes examined may include representatives of the same clonal strain. The lack of detailed metadata presents a challenge in ascertaining the source of many of these strains. Additionally, the core genome of the species is large and highly conserved. As prior comparative genomics studies of *P. aeruginosa* genomes found, strains belonging to the same group, which includes isolates from different sources (e.g., human and environmental), locations, and times, have ANI values of >99% (73–75). Thus, we cannot speculate as to the prevalence of the identified prophage species. By clustering the prophage sequences, however, we gain a glimpse into the diversity of prophages for the species, finding 3,201 unique prophage sequences. Because most of these clusters contain a single prophage sequence, we conclude that the genetic diversity of *P. aeruginosa* phages has yet to be fully explored.

In total, 68.21% of the genomes examined contained at least one category 1 or 4 prophage prediction. Lower confidence prophage sequences were far more abundant (Table 1). These regions may signify past phage infections. Prophages are believed to play a significant role in the evolution of this pathogen (76). On average, strains harbored only one or two prophage sequences. These prophages may inhibit additional prophage acquisition via superinfection exclusion. Prior studies of *P. aeruginosa* lysogens have found high rates of resistance to phage infection through prophage-mediated effects on structures required for phage attachment and/or adsorption (77, 78).

A level of 53.1% of the predicted prophages exhibited a query coverage greater than 50% and a percent identity of over 70% to publicly available phage genomes. These similarities were most frequent to tailed phages, which is expected given the tool used and the fact of the overrepresentation of tailed phages in sequence databases. Based upon sequence similarity, we have classified the majority of the predicted prophages as siphoviruses (n = 2,744). Most predicted prophage sequences (n = 3,214) did not meet our threshold of sequence similarity to characterized phages and are representative of either novel prophages or highly mosaic prophages infectious of *P. aeruginosa*. The query coverage threshold of 50% enabled taxonomic classification of mosaic prophages that share most of their genes with previously characterized phages. However, novel prophages or highly mosaic prophages were likely classified as unknown. Prior studies have observed mosaicism within *P. aeruginosa*-infecting siphoviruses (79). The network analysis performed here provides a means to predict the putative taxonomic classification of many of these unknown prophages based upon their proximity to classified prophages. As Fig. 1 shows, the majority of the unknown prophages clustered with siphoviruses or myoviruses. Nodes in the periphery of the large connected component CC1 share few genes with other predicted prophages and, thus, may represent families of phages yet to be characterized. The prophage sequences in CC4 include predictions from clinical isolates. Eight of these sequences are full-

length lambda phage sequences; the remaining predicted prophages in this connected component are small sequences, partial representatives of the lambda phage genome (see Table S1 in the supplemental material). A lambda-like phage has recently been described for *P. aeruginosa* (80).

Further evidence of reassortment among *P. aeruginosa* infecting phages can be seen from our analysis of the terminases, integrases, and C repressor gene sequences. No association between prophages that share homologous terminases with prophages that share homologous integrases or C repressors was observed (Fig. 4 and 5). On average, for a given terminase group, these prophage sequences span 3 different integrase clusters. The same applies to C repressors where, on average, phages with a terminase also contain C repressors from approximately 2 different C repressor clusters.

The integrase and C repressor coding regions that were identified in the predicted prophages exhibited extensive diversity. The integrase coding regions shared an average amino acid pairwise identity of only 16.1%. This is far less than that seen for other pathogenic species, e.g., *Staphylococcus aureus* integrase diversity is a minimum of 38% nucleotide identity (13). When clustered based on sequence similarity, the integrases formed 72 distinct groups with very high similarity (>95% average pairwise identity) in the largest ($n = 110$) of the groups. Integrase diversity has also been observed for other pathogenic species, e.g., *S. aureus* has 8 major prophage integrase types (13), group B *Streptococcus* has 16 prophage integrase types (81), and *Salmonella enterica* has 23 integrase types among food-associated strains (15). In contrast to the integrases, less sequence diversity was observed among the *P. aeruginosa* prophage C repressor coding regions (average pairwise identity of 58.0%). These C repressor sequences clustered into 44 separate groups with high similarity (>86% average pairwise identity) within the largest ($n = 1,604$) of the groups. While the integrase gene and the C repressor gene are widely conserved in temperate phages, they display substantial diversity between their genomic groups.

Prior studies have found that moron (accessory) prophage genes can affect, e.g., biofilm formation, motility, and virulence factor production (see the review in reference 82). Furthermore, within *P. aeruginosa* the effects of the same moron gene can vary between strains (20). The survey of *P. aeruginosa* prophages here found a large reservoir of genes; over 14,000 unique coding regions were encoded by these prophages. Prior pangenome analysis of 1,311 *P. aeruginosa* genomes found that 8% of all genes unique to the bacterial genomes were phage genes (75). While most of the prophage sequences examined here do not encode antibiotic resistance genes or common virulence factors, many auxiliary metabolic genes were identified. Phage-encoded metabolic genes have been best studied in marine phages (83, 84), and it was more recently explored in groundwater (85). Prior investigation of a few lytic *P. aeruginosa*-infecting phages found 3% of genes to be putative auxiliary metabolic genes, and phage-specific metabolic effects during infection have been observed (86). Further investigation is needed to explore their putative role as well as the function of the numerous hypothetical proteins identified within the prophage sequences.

As previously mentioned, the phages identified here may not be an exhaustive list of the prophages harbored by the *P. aeruginosa* strains examined. VirSorter v.1 is optimized to recognize double-stranded DNA tailed phages (*Caudovirales*). These are the prophages that have been focused on here through our subsequent analysis of integrase, C repressor, and terminase gene sequences. Several different tools have been developed specifically for prophage identification, e.g., ProphET (87), PHASTER (8), and PhiSpy (88). Other tools, such as MARVEL (89), VIBRANT (90), VirFinder (91), DeepVirFinder (92), and VironFinder (93), have been optimized for viral sequence detection in metagenomic data. More recently, although after we began the study presented here, VirSorter2 (94) was released, expanding its classifiers to include *Caudovirales* as well as other viral groups. These other viral groups are not as well studied as *Caudovirales*, and future experimental work will be pivotal in increasing not only our knowledge of these phages

but also genetic signatures that can be used to improve and/or validate bioinformatic predictions.

Beyond providing insight into the evolution of *P. aeruginosa*-infecting phages and *P. aeruginosa* strains, the prophages found in this study create a catalog of phages for potential use as phage therapies. Our analysis suggests that many of the high-confidence phages are temperate phages. While past phage therapies of *P. aeruginosa* infections have traditionally employed obligately lytic phages (31–68), temperate phages are also effective treatments when used in combination with other temperate phages as a cocktail. A 2015 study displayed successful results of temperate phage cocktails against strains of *P. aeruginosa* as well as against *Clostridium difficile* strains (95–97). Induction and characterization of temperate prophages provides additional resources in the face of growing antibiotic resistance. More importantly, the catalog presented here initiates exploration of the diversity of *P. aeruginosa*-infecting phages, the genetic content that they carry, and their impact on the fitness of the bacterial host.

MATERIALS AND METHODS

Cataloging *P. aeruginosa* prophages. Through the Genome Information by Organism section of the NCBI Genome database, *P. aeruginosa* bacterial genomic assemblies were downloaded (September 2020). Genome quality was assessed, using PATRIC's "Genome Quality" assessment (98) and CheckM (99) completeness and contamination scores through PATRIC. The following CheckM threshold was applied: completeness of >95% and contamination of <5%. In total, 5,383 genomes passed this threshold. All genomes were then entered into VirSorter v.1, a bioinformatic tool to predict phage sequences found inside bacterial genomes (10). VirSorter v.1 detects viral signals using both reference-based homology as well as reference-independent methods. It then predicts phage sequences with confidence levels ranging from 1 to 3 for extrachromosomal phages, e.g., lytic or plasmid phages, and 4 to 6 for prophages, where 1 and 4 are the highest confidence predictions and 3 and 6 are the lowest confidence predictions.

Identified prophage sequences were clustered by similarity using MeshClust (v.1.2.0) with a nucleotide sequence identity threshold of 0.95 (100). The largest clusters of prophages were further investigated. The bacterial sequences that the prophages were predicted from were compared using average nucleotide identity (ANI) to determine if the highly similar prophages were predicted from similar or dissimilar strains of *P. aeruginosa*. The software tool used to compute ANI is pyANI (v.0.2.11) (101).

Taxonomic classification of *P. aeruginosa* prophages. All category 1 and 4 sequences were compared to previously characterized phage genomes in an effort to determine their likely taxonomic family. Each sequence was queried against all complete and partial genome sequences in GenBank (organism "Virus," division "PHG") using local BLAST (102). This database includes 26,381 sequences, and the blastn algorithm was used. Homologous results with a query coverage greater than 50% and a percent identity over 70% were considered to be acceptable, and the taxonomies of the resulting similar phages were used to predict the taxonomy of the query phages.

Evaluating genomic relationships between *P. aeruginosa* prophages. Predicted phage sequences were then examined using Anvi'o v.6.2 to find the number of shared genes between each predicted phage (103). Anvi'o was used to identify homologs given the following parameters: MCL inflation value of 2 during cluster identification and a minbit heuristic score of 0.35 to remove weak gene matches. A Python script was used to produce an edgelist of each node's connections to other nodes and the number of genes shared between the connected nodes. The edgelist file was then pruned using a Python script to eliminate self-loop edges or duplicate edges. Phage gene similarity was visualized using Cytoscape v.3.8.2 (104). Different thresholds of edge weights (number of genes in common between phage genomes) were considered.

Diversity of *P. aeruginosa* prophage integrase, C repressor, and terminase coding sequences. The predicted phages were annotated using PATRIC v.3.6.9 with the Bacteriophage Domain, the Genetic Code for Bacteria and Archaea, and the Bacteriophage Annotation Recipe (98, 105). The resulting fasta descriptions for each gene contig were parsed to identify coding regions annotated as an integrase using the words "integrase" and "Integrase." The same process was repeated to identify C repressor coding regions using the word "repressor" in combination with the word "cI," "CI," "c1," or "C1." Finally, the terminase coding regions were identified using the words "terminase" and "Terminase." The annotated integrase, C repressor, and terminase coding regions were then compared for similarity to all other annotated coding regions from the predicted phages. A local BLAST (blastp) with a maximum of 1 target sequence was used with the non-integrase coding regions as the database and the integrase coding regions as the query sequences. Any gene that had a percent identity greater than 70% and a query coverage greater than 70% was then added to the respective list of integrases, and the same process was repeated for the C repressors and terminases. Each gene group was aligned using the MAFFT v.7.450 (106) multiple alignment with the automatic algorithm option with its default parameters. Using FastTree v.2.1.11 (107) with its default parameters, Newick trees were constructed and then visualized in iTOL v.6.1 (108). For easier visibility, clades with an average branch length less than 0.00056 were collapsed.

USEARCH v.11.0.667 (109) was used to create clusters of the integrase coding regions, clusters of the C repressor coding regions, and clusters of terminase coding regions. A threshold of "id = 0.7" was used for clustering. The clusters for the 3 genes were then uploaded to Geneious Prime 2020.1.2 to visualize the quality of the clusters. The phage memberships in the integrase, C repressor, and terminase clusters output from USEARCH were examined. A Python script using Biopython SeqIO, NumPy, and Pandas packages counted the number of phages shared between each of the 3 sets of coding region clusters (110–112). A one-sample *t* test was used to determine the statistical significance of the phage memberships. We examined if all members of one integrase/C repressor cluster shared the same terminase cluster and, thus, the same number of nonzero clusters.

Predicting antibiotic resistance, virulence factors, and metabolic genes encoded by prophages. ResFinder 4.1 (113) and its databases were installed and downloaded from the tool's bitbucket page (https://bitbucket.org/genomicepidemiology/resfinder/src/master/). All of the predicted prophages were separately run using the authors' suggestions of an 80% threshold and 60% minimum coverage for acquired antibiotic resistance genes. The virulence factors of the phages were then predicted using VFDB (114). The full data set that encompasses all virulence factor genes, both predicted and experimentally known, was downloaded from the VFDB website (http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi?func=VFanalyzer) and used as the BLAST database for local analysis via blastn with a threshold of greater than 70% identity. Each predicted phage was compared for sequence similarity to the virulence factor database entries.

To identify prophage coding regions associated with metabolism, the prophage FASTA format sequences were submitted to the MG-RAST webserver (115). No filtration of FASTA sequences was performed. Gene predictions assigned to a Metabolism COG category were retrieved from the webserver; these data include information about the prophage sequence containing the predicted gene, the COG number, the COG function, and the gene sequence. Python was used to parse this file to extract functional information.

Mining for Pf1 phage sequences. The Pf1 genome sequence (GenBank accession no. AY324828.1) was queried against the 5,383 *P. aeruginosa* genomes locally via blastn. Results were filtered, removing hits with a sequence identity or query coverage of <99%.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.

FIG S2, PDF file, 0.6 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.01 MB.

TABLE S3, XLSX file, 0.01 MB.

TABLE S4, XLSX file, 0.04 MB.

TABLE S5, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Casjens S, Hendrix RW. 2004. Bacteriophages and the bacterial genome, p 39–52. *In* The Bacterial Chromosome. John Wiley & Sons, Ltd, New York, NY.
2. Little JW. 2005. Lysogeny, prophage induction, and lysogenic conversion, p 37–54. *In* Phages. John Wiley & Sons, Ltd, New York, NY.
3. Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68:560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004.
4. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. Nat Commun 1:147. https://doi.org/10.1038/ncomms1146.
5. Dion MB, Oechslin F, Moineau S. 2020. Phage diversity, genomics and phylogeny. Nat Rev Microbiol 18:125–138. https://doi.org/10.1038/s41579-019-0311-5.
6. Fouts DE. 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res 34:5839–5851. https://doi.org/10.1093/nar/gkl732.
7. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.

8. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
9. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. Bioinformatics 24:863–865. https://doi.org/10.1093/bioinformatics/btn043.
10. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.
11. Miller-Ensminger T, Garretto A, Brenner J, Thomas-White K, Zambom A, Wolfe AJ, Putonti C. 2018. Bacteriophages of the urinary microbiome. J Bacteriol 200:e00738-17. https://doi.org/10.1128/JB.00738-17.
12. Kim M-S, Bae J-W. 2018. Lysogeny is prevalent and widely distributed in the murine gut microbiota. ISME J 12:1127–1141. https://doi.org/10.1038/s41396-018-0061-9.
13. Goerke C, Pantucek R, Holtfreter S, Schulte B, Zink M, Grumann D, Bröker BM, Doskar J, Wolz C. 2009. Diversity of prophages in dominant Staphylococcus aureus clonal lineages. J Bacteriol 191:3462–3468. https://doi.org/10.1128/JB.01804-08.

14. Bobay L-M, Rocha EPC, Touchon M. 2013. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol 30:737–751. https://doi.org/10.1093/molbev/mss279.

15. Colavecchio A, D'Souza Y, Tompkins E, Jeukens J, Freschi L, Emond-Rheault J-G, Kukavica-Ibrulj I, Boyle B, Bekal S, Tamber S, Levesque RC, Goodridge LD. 2017. Prophage integrase typing is a useful indicator of genomic diversity in salmonella enterica. Front Microbiol 8:1283. https://doi.org/10.3389/fmicb.2017.01283.

16. Mottawea W, Duceppe M-O, Dupras AA, Usongo V, Jeukens J, Freschi L, Emond-Rheault J-G, Hamel J, Kukavica-Ibrulj I, Boyle B, Gill A, Burnett E, Franz E, Arya G, Weadge JT, Gruenheid S, Wiedmann M, Huang H, Daigle F, Moineau S, Bekal S, Levesque RC, Goodridge LD, Ogunremi D. 2018. Salmonella enterica prophage sequence profiles reflect genome diversity and can be used for high discrimination subtyping. Front Microbiol 9:836. https://doi.org/10.3389/fmicb.2018.00836.

17. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Jr, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education, Mycobacterial Genetics Course. 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Elife 4:e06416. https://doi.org/10.7554/eLife.06416.

18. Wang F, Wang D, Hou W, Jin Q, Feng J, Zhou D. 2019. Evolutionary diversity of prophage DNA in Klebsiella pneumoniae chromosomes. Front Microbiol 10:2840. https://doi.org/10.3389/fmicb.2019.02840.

19. Brueggemann AB, Harrold CL, Rezaei Javan R, van Tonder AJ, McDonnell AJ, Edwards BA. 2017. Pneumococcal prophages are diverse, but not without structure or history. Sci Rep 7:42976. https://doi.org/10.1038/srep42976.

20. Tsao Y-F, Taylor VL, Kala S, Bondy-Denomy J, Khan AN, Bona D, Cattoir V, Lory S, Davidson AR, Maxwell KL. 2018. Phage morons play an important role in Pseudomonas aeruginosa phenotypes. J Bacteriol 200:e00189-18. https://doi.org/10.1128/JB.00189-18.

21. Davies EV, James CE, Kukavica-Ibrulj I, Levesque RC, Brockhurst MA, Winstanley C. 2016. Temperate phages enhance pathogen fitness in chronic lung infection. ISME J 10:2553–2555. https://doi.org/10.1038/ismej.2016.51.

22. Vaca-Pacheco S, Paniagua-Contreras GL, García-González O, de la Garza M. 1999. The clinically isolated FIZ15 bacteriophage causes lysogenic conversion in Pseudomonas aeruginosa PAO1. Curr Microbiol 38:239–243. https://doi.org/10.1007/pl00006794.

23. Hayashi T, Baba T, Matsumoto H, Terawaki Y. 1990. Phage-conversion of cytotoxin production in Pseudomonas aeruginosa. Mol Microbiol 4:1703–1709. https://doi.org/10.1111/j.1365-2958.1990.tb00547.x.

24. Ambroa A, Blasco L, López-Causapé C, Trastoy R, Fernandez-García L, Bleriot I, Ponce-Alonso M, Pacios O, López M, Cantón R, Kidd TJ, Bou G, Oliver A, Tomás M. 2020. Temperate bacteriophages (prophages) in Pseudomonas aeruginosa isolates belonging to the international cystic fibrosis clone (CC274). Front Microbiol 11:556706. https://doi.org/10.3389/fmicb.2020.556706.

25. Knezevic P, Voet M, Lavigne R. 2015. Prevalence of Pf1-like (pro)phage genetic elements among Pseudomonas aeruginosa isolates. Virology 483:64–71. https://doi.org/10.1016/j.virol.2015.04.008.

26. Wang J, Hu B, Xu M, Yan Q, Liu S, Zhu X, Sun Z, Reed E, Ding L, Gong J, Li QQ, Hu J. 2006. Use of bacteriophage in the treatment of experimental animal bacteremia from imipenem-resistant Pseudomonas aeruginosa. Int J Mol Med 17:309–317.

27. Watanabe R, Matsumoto T, Sano G, Ishii Y, Tateda K, Sumiyama Y, Uchiyama J, Sakurai S, Matsuzaki S, Imai S, Yamaguchi K. 2007. Efficacy of bacteriophage therapy against gut-derived sepsis caused by Pseudomonas aeruginosa in mice. Antimicrob Agents Chemother 51:446–452. https://doi.org/10.1128/AAC.00635-06.

28. Fukuda K, Ishida W, Uchiyama J, Rashel M, Kato S, Morita T, Muraoka A, Sumi T, Matsuzaki S, Daibata M, Fukushima A. 2012. Pseudomonas aeruginosa keratitis in mice: effects of topical bacteriophage KPP12 administration. PLoS One 7:e47742. https://doi.org/10.1371/journal.pone.0047742.

29. Pabary R, Singh C, Morales S, Bush A, Alshafi K, Bilton D, Alton EWFW, Smithyman A, Davies JC. 2016. Antipseudomonal bacteriophage reduces infective burden and inflammatory response in murine lung. Antimicrob Agents Chemother 60:744–751. https://doi.org/10.1128/AAC.01426-15.

30. Cafora M, Deflorian G, Forti F, Ferrari L, Binelli G, Briani F, Ghisotti D, Pistocchi A. 2019. Phage therapy against Pseudomonas aeruginosa infections in a cystic fibrosis zebrafish model. Sci Rep 9:1527. https://doi.org/10.1038/s41598-018-37636-x.

31. Jennes S, Merabishvili M, Soentjens P, Pang KW, Rose T, Keersebilck E, Soete O, François P-M, Teodorescu S, Verween G, Verbeken G, De Vos D, Pirnay J-P. 2017. Use of bacteriophages in the treatment of colistin-only-sensitive Pseudomonas aeruginosa septicaemia in a patient with acute kidney injury—a case report. Crit Care https://doi.org/10.1186/s13054-017-1709-y.

32. Chan BK, Turner PE, Kim S, Mojibian HR, Elefteriades JA, Narayan D. 2018. Phage treatment of an aortic graft infected with Pseudomonas aeruginosa. Evol Med Public Health 2018:60–66. https://doi.org/10.1093/emph/eoy005.

33. Aslam S, Courtwright AM, Koval C, Lehman SM, Morales S, Furr C-LL, Rosas F, Brownstein MJ, Fackler JR, Sisson BM, Biswas B, Henry M, Luu T, Bivens BN, Hamilton T, Duplessis C, Logan C, Law N, Yung G, Turowski J, Anesi J, Strathdee SA, Schooley RT. 2019. Early clinical experience of bacteriophage therapy in 3 lung transplant recipients. Am J Transplant 19:2631–2639. https://doi.org/10.1111/ajt.15503.

34. Aslam S, Pretorius V, Lehman SM, Morales S, Schooley RT. 2019. Novel bacteriophage therapy for treatment of left ventricular assist device infection. J Heart Lung Transplant 38:475–476. https://doi.org/10.1016/j.healun.2019.01.001.

35. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, Hatfull GF, Spencer H. 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nat Med 25:730–733. https://doi.org/10.1038/s41591-019-0437-z.

36. Jault P, Leclerc T, Jennes S, Pirnay JP, Que Y-A, Resch G, Rousseau AF, Ravat F, Carsin H, Le Floch R, Schaal JV, Soler C, Fevre C, Arnaud I, Bretaudeau L, Gabard J. 2019. Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by Pseudomonas aeruginosa (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. Lancet Infect Dis 19:35–45. https://doi.org/10.1016/S1473-3099(18)30482-1.

37. Kortright KE, Chan BK, Koff JL, Turner PE. 2019. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. Cell Host Microbe 25:219–232. https://doi.org/10.1016/j.chom.2019.01.014.

38. LaVergne S, Hamilton T, Biswas B, Kumaraswamy M, Schooley RT, Wooten D. 2018. Phage therapy for a multidrug-resistant Acinetobacter baumannii craniectomy site infection. Open Forum Infect Dis 5:ofy064. https://doi.org/10.1093/ofid/ofy064.

39. Łusiak-Szelachowska M, Zaczek M, Weber-Dąbrowska B, Międzybrodzki R, Kłak M, Fortuna W, Letkiewicz S, Rogóż P, Szufnarowski K, Jończyk-Matysiak E, Owczarek B, Górski A. 2014. Phage neutralization by sera of patients receiving phage therapy. Viral Immunol 27:295–304. https://doi.org/10.1089/vim.2013.0128.

40. Nir-Paz R, Gelman D, Khouri A, Sisson BM, Fackler J, Alkalay-Oren S, Khalifa L, Rimon A, Yerushalmy O, Bader R, Amit S, Coppenhagen-Glazer S, Henry M, Quinones J, Malagon F, Biswas B, Moses AE, Merril G, Schooley RT, Brownstein MJ, Weil YA, Hazan R. 2019. Successful treatment of antibiotic-resistant, poly-microbial bone infection with bacteriophages and antibiotics combination. Clin Infect Dis 69:2015–2018. https://doi.org/10.1093/cid/ciz222.

41. Petrovic Fabijan A, Lin RCY, Ho J, Maddocks S, Ben Zakour NL, Iredell JR, Westmead Bacteriophage Therapy Team. 2020. Safety of bacteriophage therapy in severe Staphylococcus aureus infection. Nat Microbiol 5:465–472. https://doi.org/10.1038/s41564-019-0634-z.

42. Reardon S. 2014. Phage therapy gets revitalized. Nat News 510:15–16. https://doi.org/10.1038/510015a.

43. Rhoads D, Wolcott R, Kuskowski M, Wolcott B, Ward L, Sulakvelidze A. 2009. Bacteriophage therapy of venous leg ulcers in humans: results of a phase I safety trial. J Wound Care 18:237–243. https://doi.org/10.12968/jowc.2009.18.6.42801.

44. Rose T, Verbeken G, Vos DD, Merabishvili M, Vaneechoutte M, Lavigne R, Jennes S, Zizi M, Pirnay J-P. 2014. Experimental phage therapy of burn wound infection: difficult first steps. Int J Burns Trauma 4:66–73.

45. Schooley RT, Biswas B, Gill JJ, Hernandez-Morales A, Lancaster J, Lessor L, Barr JJ, Reed SL, Rohwer F, Benler S, Segall AM, Taplitz R, Smith DM, Kerr K, Kumaraswamy M, Nizet V, Lin L, McCauley MD, Strathdee SA, Benson CA, Pope RK, Leroux BM, Picel AC, Mateczun AJ, Cilwa KE, Regeimbal JM, Estrella LA, Wolfe DM, Henry MS, Quinones J, Salka S, Bishop-Lilly KA, Young R, Hamilton T. 2017. Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant Acinetobacter baumannii infection. Antimicrob Agents Chemother 61:e00954-17. https://doi.org/10.1128/AAC.00954-17.

46. Armata Pharmaceuticals, Inc. 2016. Ascending dose study of the safety of AB-SA01 when topically applied to intact skin of healthy adults. ClinicalTrials registration no. NCT02757755.

47. Armata Pharmaceuticals, Inc. 2019. A phase 2, randomized, double-blind, placebo-controlled study to evaluate the safety and microbiology of C16G2 administered in multiple oral gel and/or mouth rinse doses to healthy adult subjects. ClinicalTrials registration no. NCT02044081.

48. Armata Pharmaceuticals, Inc. 2019. A phase 2, open-label study to evaluate the safety, microbiology and pharmacokinetics of C16G2 administered in single or multiple oral gel doses to healthy adult subjects. ClinicalTrials registration no. NCT02254993.

49. Armata Pharmaceuticals, Inc. 2019. A phase 2, randomized, double-blind, placebo-controlled study to evaluate the safety and microbiology of C16G2 administered in multiple oral gel doses to adolescent and adult dental subjects. ClinicalTrials registration no. NCT02509845.

50. Armata Pharmaceuticals, Inc. 2019. A phase 2, double-blind, placebo-controlled and open-label multi-center study to evaluate the microbiology and safety of C16G2 administered in multiple oral gel doses to adult and adolescent dental subjects. ClinicalTrials registration no. NCT02594254.

51. Armata Pharmaceuticals, Inc. 2019. A phase 2, single-blind, randomized, placebo-controlled study to evaluate the safety and microbiology of C16G2 varnish administered in multiple doses to adolescent and adult dental subjects. ClinicalTrials registration no. NCT03004365.

52. BiomX, Inc. 2021. A phase 1, randomized, single-blind, placebo-controlled study to evaluate the safety, tolerability, and fecal pharmacokinetics of orally administered BX002-A in healthy adult individuals. ClinicalTrials registration no. NCT04737876.

53. Bristol-Myers Squibb. 2011. Phase I study of BMS-188667 (CTLA4Ig) in patients with Psoriasis vulgaris. ClinicalTrials registration no. NCT00306878.

54. Leitner L, Sybesma W, Chanishvili N, Goderdzishvili M, Chkhotua A, Ujmajuridze A, Schneider MP, Sartori A, Mehnert U, Bachmann LM, Kessler TM. 2017. Bacteriophages for treating urinary tract infections in patients undergoing transurethral resection of the prostate: a randomized, placebo-controlled, double-blind clinical trial. BMC Urol 17:90. https://doi.org/10.1186/s12894-017-0283-6.

55. Locus Biosciences. 2021. A multi-center randomized, double-blind study to assess the safety, tolerability, pharmacokinetics and pharmacodynamics of LBP-EC01 in patients with lower urinary tract colonization caused by E. coli. ClinicalTrials registration no. NCT04191148.

56. National Cancer Institute. 2015. A phase IIb randomized, double-blind, placebo-controlled clinical trial of topical bacteriophage T4 endonuclease V in renal allograft recipients with a history of non-melanoma skin cancer. ClinicalTrials registration no. NCT00089180.

57. University Hospital Montpellier. 2013. Bacteriophages effects on Pseudomonas aeruginosa presents in sputum of cystic fibrosis (CF) patients. ClinicalTrials registration no. NCT01818206.

58. Weir T. 2019. PreforPro: a randomized, placebo controlled crossover study. ClinicalTrials registration no. NCT03269617.

59. Weir T. 2020. BacterioPHAGE for gastrointestinal health 2 study. ClinicalTrials registration no. NCT04511221.

60. Wolcott R. 2011. A prospective, randomized double-blind controlled study of WPP-201 for the safety and efficacy of treatment of venous leg ulcers. ClinicalTrials registration no. NCT00663091.

61. MicroPhage, Inc. 2010. Prospective study into the performance of the microphage S. Aureus/MSSA/MRSA test direct from blood culture positives. ClinicalTrials registration no. NCT01184339.

62. National Institute of Allergy and Infectious Diseases (NIAID.). 2008. Study of immune responses and safety of recombinant CD40 ligand in patients with X-linked hyper IgM syndrome. ClinicalTrials registration no. NCT00001145.

63. Cano EJ, Caflisch KM, Bollyky PL, Van Belleghem JD, Patel R, Fackler J, Brownstein MJ, Horne B, Biswas B, Henry M, Malagon F, Lewallen DG, Suh GA. 2020. Phage therapy for limb-threatening prosthetic knee Klebsiella pneumoniae infection: case report and in vitro characterization of anti-biofilm activity. Clin Infect Dis https://doi.org/10.1093/cid/ciaa705.

64. Adaptive Phage Therapeutics. 2021. Case Studies – APT's PhageBank™ Therapy. vol 2021. https://www.aphage.com.

65. Doub JB, Ng VY, Johnson AJ, Slomka M, Fackler J, Horne B, Brownstein MJ, Henry M, Malagon F, Biswas B. 2020. Salvage bacteriophage therapy for a chronic MRSA prosthetic joint infection. Antibiotics 9:241. https://doi.org/10.3390/antibiotics9050241.

66. Duplessis C, Biswas B, Hanisch B, Perkins M, Henry M, Quinones J, Wolfe D, Estrella L, Hamilton T. 2018. Refractory Pseudomonas bacteremia in a 2-year-old sterilized by bacteriophage therapy. J Pediatric Infect Dis Soc 7:253–256. https://doi.org/10.1093/jpids/pix056.

67. Duplessis C, Stockelman M, Hamilton T, Merril GL, Brownstein MJ, Bishop-Lilly KA, Schooley RT, Henry MS, Horne B, Sisson BM, Quinones J, Aslam S, LaVergne SM, Nir-Paz R, & Biswas B. 2019. A Case Series of Emergency Investigational New Drug Applications for Bacteriophages Treating Recalcitrant Multi-drug Resistant Bacterial Infections: Confirmed Safety and a Signal of Efficacy. Journal of Intensive and Critical Care 5:11.

68. Gainey AB, Burch A-K, Brownstein MJ, Brown DE, Fackler J, Horne B, Biswas B, Bivens BN, Malagon F, Daniels R. 2020. Combining bacteriophages with cefiderocol and meropenem/vaborbactam to treat a pandrug resistant Achromobacter species infection in a pediatric cystic fibrosis patient. Pediatr Pulmonol 55:2990–2994. https://doi.org/10.1002/ppul.24945.

69. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. 2018. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 46:W278–W281. https://doi.org/10.1093/nar/gky383.

70. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. Nature 536:425–430. https://doi.org/10.1038/nature19094.

71. Paez-Espino D, Pavlopoulos GA, Ivanova NN, Kyrpides NC. 2017. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. Nat Protoc 12:1673–1682. https://doi.org/10.1038/nprot.2017.063.

72. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. 2021. Massive expansion of human gut bacteriophage diversity. Cell 184:1098–1109. https://doi.org/10.1016/j.cell.2021.01.029.

73. Subedi D, Vijay AK, Kohli GS, Rice SA, Willcox M. 2018. Comparative genomics of clinical strains of Pseudomonas aeruginosa strains isolated from different geographic sites. Sci Rep 8:15668. https://doi.org/10.1038/s41598-018-34020-7.

74. Ozer EA, Nnah E, Didelot X, Whitaker RJ, Hauser AR. 2019. The population structure of Pseudomonas aeruginosa is characterized by genetic isolation of exoU+ and exoS+ lineages. Genome Biol Evol 11:1780–1796. https://doi.org/10.1093/gbe/evz119.

75. Freschi L, Vincent AT, Jeukens J, Emond-Rheault J-G, Kukavica-Ibrulj I, Dupont M-J, Charette SJ, Boyle B, Levesque RC. 2019. The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. Genome Biol Evol 11: 109–120. https://doi.org/10.1093/gbe/evy259.

76. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage genomics. 2 Microbiol Mol Biol Rev 67:238–276. https://doi.org/10.1128/MMBR.67.2.238-276.2003.

77. Chung I-Y, Jang H-J, Bae H-W, Cho Y-H. 2014. A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. Proc Natl Acad Sci U S A 111:11503–11508. https://doi.org/10.1073/pnas.1403537111.

78. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. ISME J 10:2854–2866. https://doi.org/10.1038/ismej.2016.79.

79. Flores V, Sepúlveda-Robles O, Cazares A, Kameyama L, Guarneros G. 2017. Comparative genomic analysis of Pseudomonas aeruginosa phage PaMx25 reveals a novel siphovirus group related to phages infecting hosts of different taxonomic classes. Arch Virol 162:2345–2355. https://doi.org/10.1007/s00705-017-3366-5.

80. Yu X, Xu J, Gu Y, Zhang R, Zhu Y, Liu X. 2020. Molecular characterization and comparative genomic analysis of vB_PaeP_YA3, a novel temperate bacteriophage of Pseudomonas aeruginosa. Front Microbiol 11:947. https://doi.org/10.3389/fmicb.2020.00947.

81. Crestani C, Forde TL, Zadoks RN. 2020. Development and application of a prophage integrase typing scheme for group B Streptococcus. Front Microbiol 11:1993. https://doi.org/10.3389/fmicb.2020.01993.

82. Taylor VL, Fitzpatrick AD, Islam Z, Maxwell KL. 2019. The diverse impacts of phage morons on bacterial fitness and virulence. Adv Virus Res 103: 1–31. https://doi.org/10.1016/bs.aivir.2018.08.001.

83. Hurwitz BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and taxonomic niche specialization in the "core" and "flexible" Pacific Ocean virome. ISME J 9:472–484. https://doi.org/10.1038/ismej.2014.143.

84. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB, Tara Oceans Coordinators. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537:689–693. https://doi.org/10.1038/nature19366.

85. Kothari A, Roux S, Zhang H, Prieto A, Soneja D, Chandonia J-M, Spencer S, Wu X, Altenburg S, Fields MW, Deutschbauer AM, Arkin AP, Alm EJ, Chakraborty R, Mukhopadhyay A. 2021. Ecogenomics of groundwater phages suggests niche differentiation linked to specific environmental tolerance. mSystems 6:e0053721. https://doi.org/10.1128/mSystems .00537-21.

86. De Smet J, Zimmermann M, Kogadeeva M, Ceyssens P-J, Vermaelen W, Blasdel B, Bin Jang H, Sauer U, Lavigne R. 2016. High coverage metabolomics analysis reveals phage-specific alterations to Pseudomonas aeruginosa physiology during infection. ISME J 10:1823–1835. https://doi.org/10.1038/ismej.2016.3.

87. Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. 2019. ProphET, prophage estimation tool: a stand-alone prophage sequence prediction tool with self-updating reference database. PLoS One 14:e0223364. https://doi.org/10.1371/journal.pone.0223364.

88. Akhter S, Aziz RK, Edwards RA. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res 40:e126. https://doi.org/10.1093/nar/gks406.

89. Amgarten D, Braga LPP, da Silva AM, Setubal JC. 2018. MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. Front Genet 9:304. https://doi.org/10.3389/fgene.2018.00304.

90. Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8:90. https://doi.org/10.1186/s40168-020-00867-0.

91. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5:69. https://doi.org/10.1186/s40168-017-0283-5.

92. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, Xie X, Poplin R, Sun F. 2020. Identifying viruses from metagenomic data using deep learning. Quant Biol 8:64–77. https://doi.org/10.1007/s40484-019-0187-4.

93. Fang Z, Zhou H. 2021. VirionFinder: identification of complete and partial prokaryote virus virion protein from virome data using the sequence and biochemical properties of amino acids. Front Microbiol 12:615711. https://doi.org/10.3389/fmicb.2021.615711.

94. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37. https://doi.org/10.1186/s40168-020-00990-y.

95. Burkal'tseva MV, Krylov SV, Kropinskii± AM, Pletneva EA, Shaburova OV, Krylov VN. 2011. Bacteriophage phi297–the new species of temperate phages Pseudomonas aeruginosa with a mosaic genome: potential use in phagotherapy. Genetika 47:900–904.

96. Nale JY, Redgwell TA, Millard A, Clokie MRJ. 2018. Efficacy of an optimised bacteriophage cocktail to clear Clostridium difficile in a batch fermentation model. Antibiotics 7:13. https://doi.org/10.3390/antibiotics7010013.

97. Nale JY, Spencer J, Hargreaves KR, Buckley AM, Trzepiński P, Douce GR, Clokie MRJ. 2016. Bacteriophage combinations significantly reduce clostridium difficile growth in vitro and proliferation in vivo. Antimicrob Agents Chemother 60:968–981. https://doi.org/10.1128/AAC.01774-15.

98. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 48:D606–D612. https://doi.org/10.1093/nar/gkz943.

99. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

100. James BT, Luczak BB, Girgis HZ. 2018. MeShClust: an intelligent tool for clustering DNA sequences. Nucleic Acids Res 46:e83. https://doi.org/10.1093/nar/gky315.

101. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H.

102. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

103. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.

104. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.

105. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.

106. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.

107. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

108. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128. https://doi.org/10.1093/bioinformatics/btl529.

109. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

110. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

111. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. 2020. Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2.

112. Reback W, Van den Bossche T, Augspurger P, Cloud S, Hawkins G, Young Sinhrks M, Roeschke A, Klein T, Petersen J, Tratner C, She W, Ayd S, Naveh PM, Garcia J, Schendel A, Hayden D, Saxton V, Jancauskas M, Gorelli R, Shadrach A, McMaster P, Battiston S, Seabold S. 2021. Pandasdev/pandas: Pandas 1.2.4. Zenodo. https://doi.org/10.5281/zenodo.3509134.

113. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75:3491–3500. https://doi.org/10.1093/jac/dkaa345.

114. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325–D328. https://doi.org/10.1093/nar/gki008.

115. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386. https://doi.org/10.1186/1471-2105-9-386.