

RESEARCH ARTICLE

Open Access

# Comparative analysis of two phenotypically-similar but genomically-distinct *Burkholderia cenocepacia*-specific bacteriophages

Karlene H Lynch<sup>1</sup>, Paul Stothard<sup>2</sup> and Jonathan J Dennis<sup>1\*</sup>

## Abstract

**Background:** Genomic analysis of bacteriophages infecting the *Burkholderia cepacia* complex (BCC) is an important preliminary step in the development of a phage therapy protocol for these opportunistic pathogens. The objective of this study was to characterize KL1 (vB\_BceS\_KL1) and AH2 (vB\_BceS\_AH2), two novel *Burkholderia cenocepacia*-specific siphoviruses isolated from environmental samples.

**Results:** KL1 and AH2 exhibit several unique phenotypic similarities: they infect the same *B. cenocepacia* strains, they require prolonged incubation at 30°C for the formation of plaques at low titres, and they do not form plaques at similar titres following incubation at 37°C. However, despite these similarities, we have determined using whole-genome pyrosequencing that these phages show minimal relatedness to one another. The KL1 genome is 42,832 base pairs (bp) in length and is most closely related to *Pseudomonas* phage 73 (PA73). In contrast, the AH2 genome is 58,065 bp in length and is most closely related to *Burkholderia* phage BcepNazgul. Using both BLASTP and HHpred analysis, we have identified and analyzed the putative virion morphogenesis, lysis, DNA binding, and MazG proteins of these two phages. Notably, MazG homologs identified in cyanophages have been predicted to facilitate infection of stationary phase cells and may contribute to the unique plaque phenotype of KL1 and AH2.

**Conclusions:** The nearly indistinguishable phenotypes but distinct genomes of KL1 and AH2 provide further evidence of both vast diversity and convergent evolution in the BCC-specific phage population.

## Background

The clinical administration of bacteriophages, referred to as phage therapy, has now been used to treat bacterial infections for nearly a century. Although this type of therapy had been largely abandoned outside of Eastern Europe since antibiotics became available in the 1940s, the emergence of antibiotic-resistant pathogens has re-established phage therapy as a viable antibacterial treatment [1]. Recent studies have shown that phages and phage components are effective both in animal models (against species such as *Staphylococcus*, *Pseudomonas*, *Klebsiella*, *Escherichia*, *Salmonella*, and *Campylobacter*) and in human clinical trials [2-8]. Advances in phage delivery and storage (such as nebulization, lyophilization, and spray drying for

respiratory phage therapy) and genomic characterization (including high-throughput sequencing and annotation) have made phage therapy more feasible with respect to both logistics and safety [9-12].

One group of bacteria that is thought to be an excellent target for phage therapy is the *Burkholderia cepacia* complex (BCC). These bacterial species, which primarily infect patients with cystic fibrosis (CF), are problematic because they can cause serious illness (including, in up to 20% of cases, a fatal necrotizing pneumonia referred to as 'cepacia syndrome'), they are capable of patient-to-patient spread (particularly in settings such as CF centers), and, perhaps most importantly, they are highly antibiotic resistant [13-16]. Very few antibiotics are active against the BCC, even in combination: Zhou et al. [16] tested a panel of antibiotics against BCC clinical isolates and determined that less than half of the strains were susceptible to even the most effective drugs. Clinically,

\* Correspondence: jon.dennis@ualberta.ca

<sup>1</sup>6-008 Centennial Centre for Interdisciplinary Science, Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada  
Full list of author information is available at the end of the article

the most commonly isolated BCC species are *Burkholderia multivorans* and *Burkholderia cenocepacia*, with the latter thought to be the most pathogenic [17]. BCC phage therapy trials have focused on this species and, to date, phages have been shown to be effective against *B. cenocepacia* in both invertebrate and mammalian infection models [18-20].

As *B. cenocepacia* infections are some of the most problematic for the CF community, the isolation and characterization of novel phages that infect this species remains a priority. Many of these phages have been isolated in recent years, but only some have been fully sequenced (reviewed in [21,22]). Here, we describe the isolation and characterization of KL1 and AH2, two novel *B. cenocepacia*-specific phages with identical host ranges and unique growth characteristics, but strikingly dissimilar genomes.

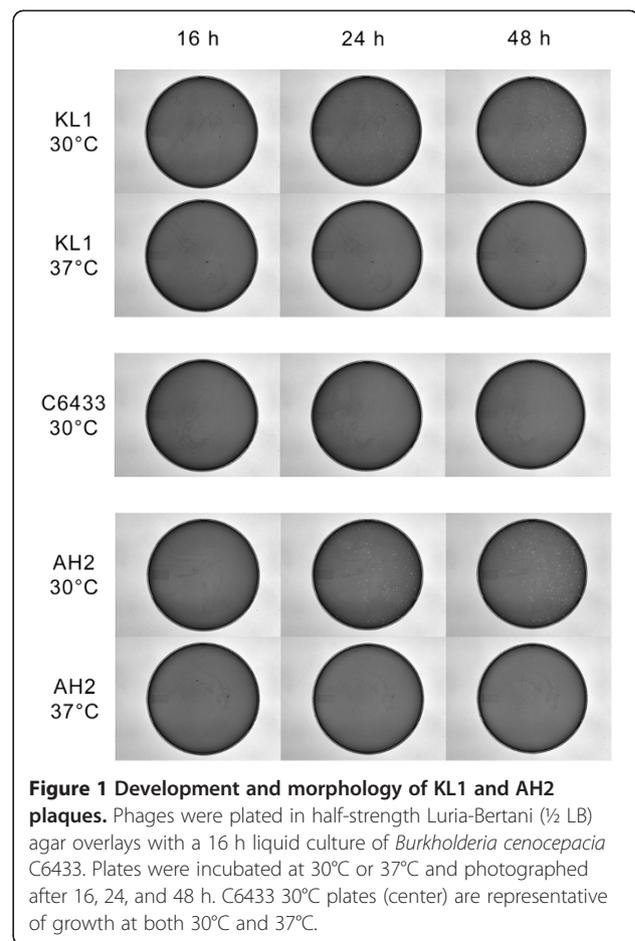
## Results and discussion

### Isolation, host range and morphology

KL1 was isolated from sewage using *B. cenocepacia* K56-2 as a host. In contrast to enterobacteria phages, which are commonly found in sewage [23], this is the first report of BCC phage isolation from this source. AH2 was isolated from *Nandina* sp. (also known as heavenly bamboo) soil using *B. cenocepacia* C6433. BCC phages have commonly been isolated from both rhizospheres and soil samples, including that of onion and *Dracaena* sp. [18,20,24-27].

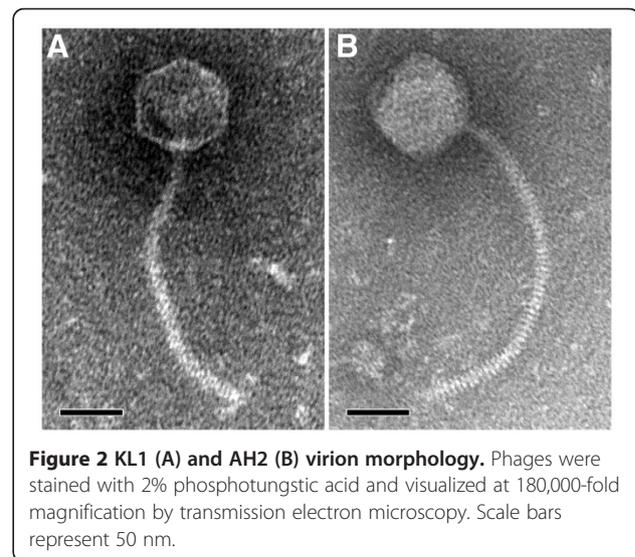
KL1 and AH2 are very similar with respect to both host range and growth characteristics. These phages have a relatively narrow tropism, infecting *B. cenocepacia* K56-2, C6433, 715], and K63-3. Both KL1 and AH2 exhibit a pattern of lysis that is unique in our collection of BCC-specific phages: although high titre stocks of these phages are very concentrated (up to  $10^{11}$  plaque forming units [PFU]/ml), these phages do not produce clear lysis in agar overlays after 16 h incubation like other phages that we have previously characterized [18,19,26,28-30]. Instead, turbid or no clearing is observed at high titres, with mottling or individual plaques observed at lower titres (approximately  $10^7$  PFU/ml or less). At low titres, incubation at 30°C for greater than 16 h is required for plaque formation, but plaques are not observed if incubation is at 37°C (Figure 1). Individual plaques are turbid with a diameter of 0.5-2 mm (larger plaques may have a punctate appearance). When tested with a panel of K56-2 mutants with progressive deficiencies in lipopolysaccharide (LPS) structure (from the O-antigen to the core) [31,32], both phages were able to infect each mutant, suggesting that neither KL1 nor AH2 uses LPS as a major receptor.

Both KL1 and AH2 belong to the order *Caudovirales* and family *Siphoviridae* as determined by electron



**Figure 1 Development and morphology of KL1 and AH2 plaques.** Phages were plated in half-strength Luria-Bertani (½ LB) agar overlays with a 16 h liquid culture of *Burkholderia cenocepacia* C6433. Plates were incubated at 30°C or 37°C and photographed after 16, 24, and 48 h. C6433 30°C plates (center) are representative of growth at both 30°C and 37°C.

microscopy. The KL1 virion has a non-contractile tail approximately 160 nm in length and a capsid approximately 55 nm in diameter (Figure 2A). The AH2 virion is slightly larger, with a non-contractile tail approximately 220 nm in length and a capsid approximately



**Figure 2 KL1 (A) and AH2 (B) virion morphology.** Phages were stained with 2% phosphotungstic acid and visualized at 180,000-fold magnification by transmission electron microscopy. Scale bars represent 50 nm.

60 nm in diameter (Figure 2B). The stacked rings comprising the tail structure are visible in the AH2 micrograph (Figure 2B).

### Genome characterization

Despite the similarities in phenotype between KL1 and AH2 with respect to host range and growth characteristics, the genomes of these two phages are dissimilar. Restriction fragment length polymorphism (RFLP) analysis shows distinct banding patterns of EcoRI-digested KL1 and AH2 genomic DNA, suggesting that their sequences are substantially different (Figure 3). This prediction is confirmed by the results of whole genome pyrosequencing (discussed below) and is illustrated in Figure 4A: in a Circos plot of a PROmer comparison of these two phages, no regions of similarity at the protein level are observed under the parameters used.

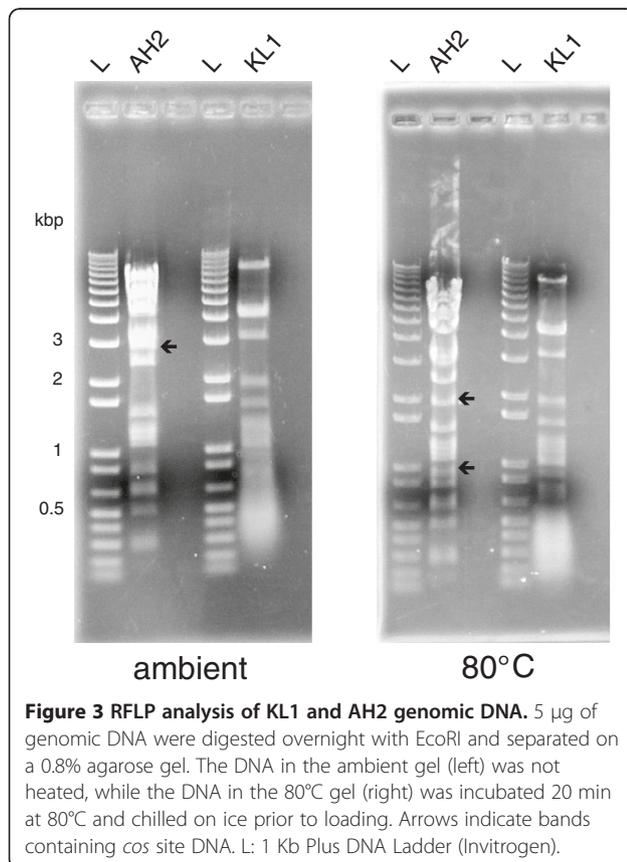
The KL1 genome is 42,832 base pairs (bp) in length and has a 54.6% GC content. This percentage is lower than that for most *Burkholderia*-specific phages, which tend to have GC contents between 60–65% (excluding phages such as BcepB1A [54.5%], BcepF1 [55.9%], and BcepGomr [56.3%]). We were unable to identify a KL1 *cos* site following incubation of the DNA at 80°C, as the RFLP profiles appeared identical both before and after

heating (Figure 3). KL1 is predicted to encode 55 proteins, all of which have an ATG start codon, except for gp2 which has a GTG codon (Figure 5, Table 1).

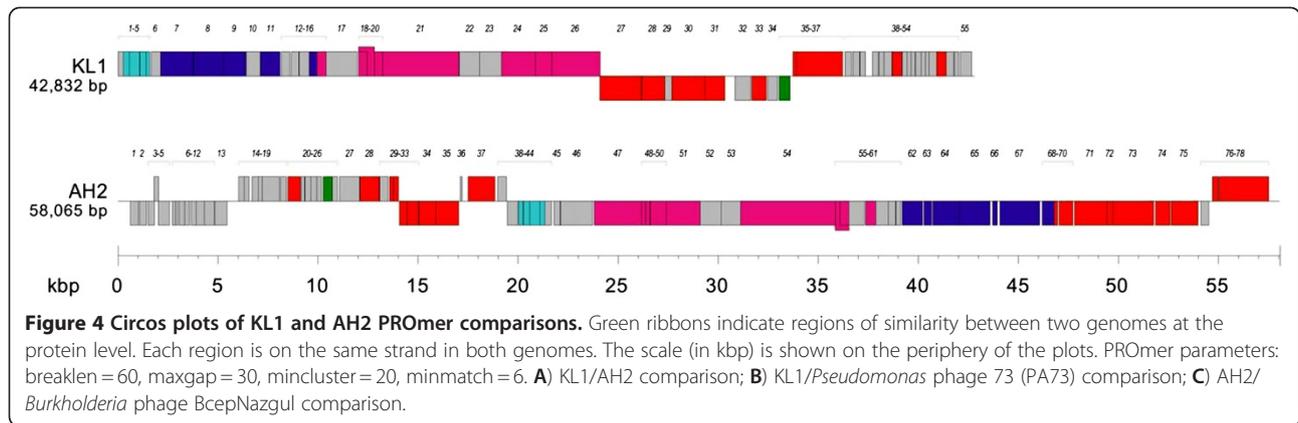
KL1 is most similar to *Pseudomonas* phage 73 (PA73; NC\_007806), a siphovirus that infects *Pseudomonas aeruginosa* [33]. These phages are similar with respect to genome length (42,999 bp for PA73 and 42,832 bp for KL1), GC content (53.6% for PA73 and 54.6% for KL1), and predicted number of proteins (52 for PA73 and 55 for KL1). BLASTN comparison of KL1 and PA73 indicates that these sequences are similar over 69% of the KL1 genome. KL1 encodes a protein most similar to each PA73 protein from ORF001–ORF052 (excluding 12 proteins) (Table 1). Most PA73 proteins show limited similarity to others in the NCBI database and have not been assigned a putative function [33]. Of the 9 PA73 proteins with predicted functions, all but one (peptidyl-tRNA hydrolase [peptide chain release factor]) is similar to a KL1 protein: holin, terminase large subunit, head morphogenesis protein, tail tape measure protein, DNA polymerase, superfamily II helicase/restriction enzyme, helicase (annotated here as recombinase), and dCMP deaminase (KL1 gp2, gp7, gp9, gp21, gp27, gp30, gp33, and gp52, respectively) (Table 1). Of the KL1 proteins most similar to a PA73 protein, the most similar is gp33 (91% identity with ORF032) and the least similar is gp24 (36% identity with ORF023) (Table 1). In a Circos plot of a PROmer comparison of these phages, the majority of the two genomes are similar at the protein level (Figure 4B).

The AH2 genome is 58,065 bp in length and has a 61.3% GC content. Incubation of the DNA at 80°C caused a shift in the RFLP profile (Figure 3), suggesting the presence of a *cos* site. Sequencing of the shifted fragments indicates that AH2 has a 12 bp 5' overhang *cos* site with a sequence almost identical (1 bp difference) to that of *Burkholderia* phage BcepNazgul (NC\_005091). AH2 is predicted to encode 78 proteins (Figure 5, Table 2). The majority of the start codons (70) are ATG, 6 are GTG and 2 are TTG (Table 2).

AH2 is most similar to BcepNazgul, a siphovirus isolated from soil that infects *Burkholderia ambifaria*. Like PA73 and KL1, these phages are similar with respect to genome length (57,455 bp for BcepNazgul and 58,065 bp for AH2), GC content (60.6% for BcepNazgul and 61.3% for AH2), and predicted number of proteins (73 for BcepNazgul and 78 for AH2). In contrast to KL1 (which is closely related to a single phage), AH2 encodes proteins similar to those from a variety of bacteria and phages (Table 2) and so is less closely related to BcepNazgul than KL1 is to PA73. BLASTN comparison of AH2 and BcepNazgul indicates that these sequences are similar over 16% of the AH2 genome. Twenty-one AH2 proteins are most similar to a BcepNazgul protein



**Figure 3 RFLP analysis of KL1 and AH2 genomic DNA.** 5 µg of genomic DNA were digested overnight with EcoRI and separated on a 0.8% agarose gel. The DNA in the ambient gel (left) was not heated, while the DNA in the 80°C gel (right) was incubated 20 min at 80°C and chilled on ice prior to loading. Arrows indicate bands containing *cos* site DNA. L: 1 Kb Plus DNA Ladder (Invitrogen).



(Table 2) and 39 show some similarity based on BLASTP analysis. Of the AH2 proteins most similar to a BcepNazgul protein, the most similar is gp12 (74% identity with Nazgul10) and the least similar is gp20 (24% identity with Nazgul21) (Table 2). In a Circos plot of a PROmer comparison of these phages, the most similar regions at the protein level correspond to AH2 gp12, gp71, gp78 (similar to BcepNazgul Nazgul10, helicase, and DR0530-like primase, respectively) and a portion of the putative capsid morphogenesis and DNA packaging module (Figure 4C).

## Module analysis

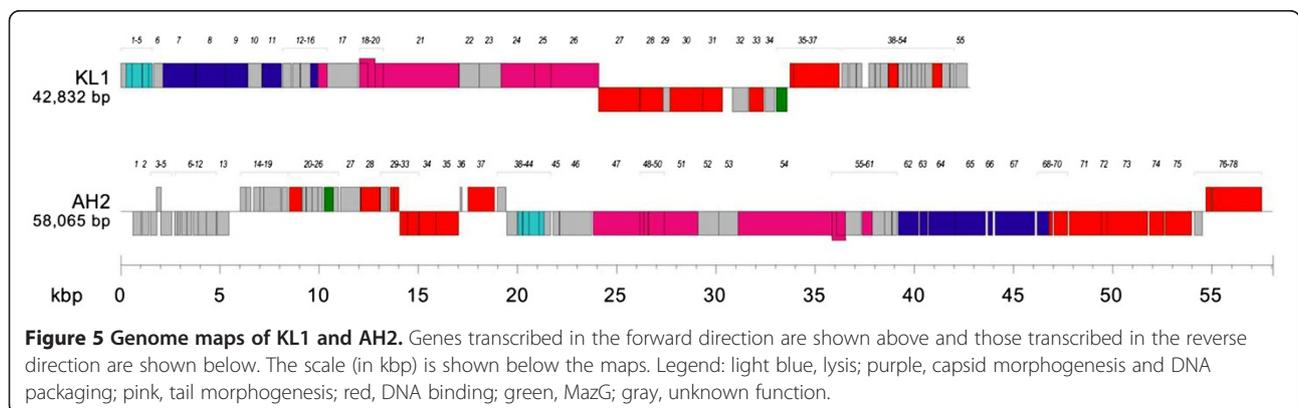
### Overview

We have identified the proteins encoded by KL1 and AH2 as belonging to four different functional categories: virion morphogenesis (including capsid morphogenesis/DNA packaging and tail morphogenesis), lysis, DNA binding (the largest and broadest category), and MazG (a pyrophosphohydrolase [34]). Although the proteins encoded by each phage perform many of the same functions (e.g. both KL1 gp11 and AH2 gp62 are predicted to be major capsid proteins) (Tables 1 and 2), the proteins themselves are dissimilar. As we discuss below, the finding that KL1 and AH2 can create nearly identical

phenotypes with two dissimilar sets of proteins may be compelling evidence for convergent evolution occurring in these BCC-specific phages.

### Virion morphogenesis

Although we have determined that KL1 is a siphovirus (Figure 2A), the identity of many of the structural genes remains unknown. As discussed above, KL1 is most closely related to PA73, a phage whose proteins have largely uncharacterized functions. Based on BLASTP analysis, we have been able to predict the identity of only eight KL1 structural proteins: three involved in capsid morphogenesis and DNA packaging and five involved in tail morphogenesis. Gp7 (terminase large subunit) and gp9 (head morphogenesis protein) are similar to PA73 ORF006 and ORF008, respectively, both of which have been assigned putative functions in the PA73 annotation (Table 1). Gp11 (major capsid protein) is similar to the major capsid proteins of *Escherichia* phage K1H and *Listonella* phage φHSIC. Gp20 is similar to tail proteins from multiple *Escherichia* phages including K1G, K1H, and K1ind1-K1ind3. Gp21 is predicted to be the tail tape measure as it is the largest protein encoded by KL1 (1272 amino acids [aa]) and it is similar to the predicted PA73 tape measure protein ORF020 (Table 1). Finally,



**Table 1 KL1 genome annotation**

| Gene | Start | End   | Putative function          | Strand | Predicted ribosome binding site and start codon | Length (amino acids) | Closest relative                  | Alignment region (amino acids) | Percent identity | Source                                  | GenBank accession number |
|------|-------|-------|----------------------------|--------|---|----------------------|-----------------------------------|--------------------------------|------------------|---|--------------------------|
| 1    | 1     | 267   | unknown                    | +      | AGGGGCGAActtctgATG                              | 88                   | hypothetical protein ORF001       | 1-84/84                        | 77               | <i>Pseudomonas</i> phage 73             | YP_001293408.1           |
| 2    | 264   | 560   | holin                      | +      | AAAGGGGCGGtaacGTG                               | 98                   | hypothetical protein ORF002       | 3-88/88                        | 42               | <i>Pseudomonas</i> phage 73             | YP_001293409.1           |
| 3    | 514   | 1080  | lysine                     | +      | AAAAGGGGttatcgaATG                              | 188                  | hypothetical protein bglu_1g27070 | 2-181/188                      | 47               | <i>Burkholderia glumae</i> BGR1         | YP_002912484.1           |
| 4    | 1091  | 1408  | Rz                         | +      | AAGTAAGGGGttcgaaATG                             | 105                  | hypothetical protein ORF004       | 1-101/101                      | 37               | <i>Pseudomonas</i> phage 73             | YP_001293411.1           |
| 5    | 1329  | 1592  | Rz1                        | +      | GAAAGGtgccgccgATG                               | 87                   | conserved hypothetical protein    | 1-79/86                        | 40               | <i>Burkholderia</i> sp. Ch1-1           | ZP_06842908.1            |
| 6    | 1647  | 2138  | unknown                    | +      | ACTAGGccgcgattATG                               | 163                  | hypothetical protein ORF005       | 1-162/162                      | 59               | <i>Pseudomonas</i> phage 73             | YP_001293412.1           |
| 7    | 2116  | 3756  | terminase large subunit    | +      | AACAGGAAAttgctaATG                              | 546                  | hypothetical protein ORF006       | 10-531/531                     | 84               | <i>Pseudomonas</i> phage 73             | YP_001293413.1           |
| 8    | 3770  | 5266  | portal protein             | +      | AAAGGAAAcgaaatcATG                              | 498                  | hypothetical protein ORF007       | 3-494/501                      | 85               | <i>Pseudomonas</i> phage 73             | YP_001293414.1           |
| 9    | 5269  | 6384  | head morphogenesis protein | +      | GGGGCGTAatcATG                                  | 371                  | hypothetical protein ORF008       | 1-364/364                      | 73               | <i>Pseudomonas</i> phage 73             | YP_001293415.1           |
| 10   | 6403  | 7110  | unknown                    | +      | AAGGAGtccttgaaATG                               | 235                  | hypothetical protein ORF009       | 1-235/239                      | 82               | <i>Pseudomonas</i> phage 73             | YP_001293416.1           |
| 11   | 7123  | 8097  | major capsid protein       | +      | AAGGAcactttatcATG                               | 324                  | hypothetical protein ORF010       | 1-325/325                      | 90               | <i>Pseudomonas</i> phage 73             | YP_001293417.1           |
| 12   | 8171  | 8587  | unknown                    | +      | AAGGAGtttcgaacATG                               | 138                  | hypothetical protein ORF011       | 1-134/134                      | 69               | <i>Pseudomonas</i> phage 73             | YP_001293418.1           |
| 13   | 8656  | 9033  | unknown                    | +      | AAAGGAGcgtcgaacATG                              | 125                  | hypothetical protein ORF012       | 1-123/123                      | 70               | <i>Pseudomonas</i> phage 73             | YP_001293419.1           |
| 14   | 9047  | 9565  | unknown                    | +      | AAGGGGcgcgcatcATG                               | 172                  | hypothetical protein ORF013       | 1-172/172                      | 83               | <i>Pseudomonas</i> phage 73             | YP_001293420.1           |
| 15   | 9570  | 9944  | head-tail joining protein  | +      | GATAAGGGTctaacgtATG                             | 124                  | hypothetical protein ORF014       | 1-124/126                      | 59               | <i>Pseudomonas</i> phage 73             | YP_001293421.1           |
| 16   | 9941  | 10399 | minor tail protein         | +      | ATACGGTAttgttcgacaATG                           | 152                  | hypothetical protein ORF015       | 5-151/151                      | 68               | <i>Pseudomonas</i> phage 73             | YP_001293422.1           |
| 17   | 10412 | 11965 | unknown                    | +      | AAGGAGttacgaaaATG                               | 517                  | hypothetical protein ORF016       | 3-511/511                      | 78               | <i>Pseudomonas</i> phage 73             | YP_001293423.1           |
| 18   | 12030 | 12458 | tail protein               | +      | GGAGTAAAccaaATG                                 | 142                  | hypothetical protein ORF017       | 1-142/142                      | 79               | <i>Pseudomonas</i> phage 73             | YP_001293424.1           |
| 19   | 12030 | 12823 | tail protein               | +      | GGAGTAAAccaaATG                                 | 264                  | hypothetical protein ORF017       | 1-142/142                      | 79               | <i>Pseudomonas</i> phage 73             | YP_001293424.1           |
|      |       |       |                            |        |   |                      | hypothetical protein ORF018       | 1-118/118                      | 78               | <i>Pseudomonas</i> phage 73             | YP_001293425.1           |
| 20   | 12792 | 13226 | tail protein               | +      | AAAAGGGCGcgcaacagaATG                           | 144                  | hypothetical protein ORF019       | 1-144/144                      | 80               | <i>Pseudomonas</i> phage 73             | YP_001293426.1           |
| 21   | 13232 | 17050 | tail tape measure          | +      | AAGGAttgcagaaATG                                | 1272                 | hypothetical protein ORF020       | 1-78, 131-1202/1204            | 61, 57           | <i>Pseudomonas</i> phage 73             | YP_001293427.1           |
| 22   | 17069 | 18067 | unknown                    | +      | AGGAAtacgaattATG                                | 332                  | hypothetical protein XALc_0225    | 1-295/307                      | 30               | <i>Xanthomonas albilineans</i> GPE PC73 | YP_003374757.1           |
| 23   | 18070 | 19179 | unknown                    | +      | GAGGAAAAactaatcATG                              | 369                  | hypothetical protein ORF033       | 1-332/333                      | 25               | <i>Pseudomonas</i> phage M6             | YP_001294541.1           |
| 24   | 19179 | 20870 | tail assembly protein      | +      | AAGAAGAtcgcataATG                               | 563                  | hypothetical protein ORF023       | 63-565/568                     | 36               | <i>Pseudomonas</i> phage 73             | YP_001293430.1           |
| 25   | 20867 | 21688 | tail assembly protein      | +      | AAGGAcgattccagaATG                              | 273                  | hypothetical protein ORF024       | 1-273/274                      | 49               | <i>Pseudomonas</i> phage 73             | YP_001293431.1           |
| 26   | 21689 | 24100 | tail assembly protein      | +      | AAGATGGGtcggttaaATG                             | 803                  | hypothetical protein ORF025       | 1-755/813                      | 49               | <i>Pseudomonas</i> phage 73             | YP_001293432.1           |

**Table 1 KL1 genome annotation (Continued)**

|    |       |       |  |   |                        |     |  |             |    |  |                |
|----|-------|-------|--|---|------------------------|-----|--|-------------|----|--|----------------|
| 27 | 24097 | 26166 | DNA polymerase                                   | - | AAGGAAttgcccgATG       | 689 | hypothetical protein ORF026  | 1-682/683   | 83 | <i>Pseudomonas</i> phage 73                  | YP_001293433.1 |
| 28 | 26179 | 27339 | DNA polymerase III<br>β subunit                  | - | AAGGGGttaaaaATG        | 386 | hypothetical protein ORF027  | 2-380/380   | 74 | <i>Pseudomonas</i> phage 73                  | YP_001293434.1 |
| 29 | 27323 | 27691 | unknown  | - | GAATGGtgaaattATG       | 122 | hypothetical protein Dole_2913   | 5-84/87     | 33 | <i>Desulfococcus oleovorans</i><br>Hxd3      | YP_001530793.1 |
| 30 | 27696 | 29351 | superfamily II<br>helicase/restriction<br>enzyme | - | AAGGGttacgaATG         | 551 | hypothetical protein ORF029  | 1-551/551   | 90 | <i>Pseudomonas</i> phage 73                  | YP_001293436.1 |
| 31 | 29344 | 30342 | exonuclease                                      | - | GGAAGGcgaagaacgATG     | 332 | hypothetical protein ORF030  | 1-365/365   | 65 | <i>Pseudomonas</i> phage 73                  | YP_001293437.1 |
| 32 | 30852 | 31637 | unknown  | - | GAAAGGtgaaacgaacATG    | 261 | hypothetical protein Isop_2441   | 1-118/151   | 37 | <i>Isosphaera pallida</i> ATCC<br>43644      | YP_004179564.1 |
| 33 | 31696 | 32412 | recombinase                                      | - | AGGTGAACgtATG          | 238 | hypothetical protein ORF032  | 1-238/238   | 91 | <i>Pseudomonas</i> phage 73                  | YP_001293439.1 |
| 34 | 32471 | 32980 | unknown  | - | AAGGAAcccaaaaATG       | 169 | hypothetical protein ORF033  | 7-146/146   | 49 | <i>Pseudomonas</i> phage 73                  | YP_001293440.1 |
| 35 | 33059 | 33598 | pyrophosphohydrolase                             | - | AGGGGcatcgtATG         | 179 | hypothetical protein ORF034  | 8-185/185   | 69 | <i>Pseudomonas</i> phage 73                  | YP_001293441.1 |
| 36 | 33746 | 33934 | transcriptional<br>regulator                     | + | GGGGcaagcATG           | 62  | hypothetical protein ORF035  | 1-61/62     | 51 | <i>Pseudomonas</i> phage 73                  | YP_001293442.1 |
| 37 | 33924 | 36233 | primase  | + | GAAGGcttgccaaatATG     | 769 | hypothetical protein ORF036  | 1-773/773   | 85 | <i>Pseudomonas</i> phage 73                  | YP_001293443.1 |
| 38 | 36366 | 36668 | unknown  | + | GAAGGAgttacgaacATG     | 100 | hypothetical protein   | 132-217/217 | 44 | <i>Deftia</i> phage φW-14                    | YP_003359005.1 |
| 39 | 36735 | 37091 | unknown  | + | GAAGGAGtacacgccATG     | 118 | unnamed protein product  | 262-336/404 | 32 | <i>Azospirillum lipoferum</i> 4B             | YP_004974060.1 |
| 40 | 37097 | 37360 | unknown  | + | AGAAGAAGGAGtaagccATG   | 87  | PREDICTED: photosystem II<br>reaction center PSB28<br>protein, chloroplastic | 22-86/179   | 32 | <i>Vitis vinifera</i>                        | XP_002271666.1 |
| 41 | 37728 | 38024 | unknown  | + | AAAGGAGcgcagccATG      | 98  | hypothetical protein ORF039  | 1-97/98     | 70 | <i>Pseudomonas</i> phage 73                  | YP_001293446.1 |
| 42 | 38060 | 38296 | unknown  | + | AAGGAAcccccgatcATG     | 78  | hypothetical protein ORF040  | 1-80/80     | 50 | <i>Pseudomonas</i> phage 73                  | YP_001293447.1 |
| 43 | 38302 | 38703 | unknown  | + | AAAGGGGtaattactATG     | 133 | hypothetical protein ORF042  | 1-120/124   | 40 | <i>Pseudomonas</i> phage 73                  | YP_001293449.1 |
| 44 | 38707 | 39195 | Vsr endonuclease                                 | + | GACGAAGttgcattaagccATG | 162 | hypothetical protein ORF043  | 1-176/179   | 61 | <i>Pseudomonas</i> phage 73                  | YP_001293450.1 |
| 45 | 39201 | 39458 | unknown  | + | GGAAGGAGtaaccxaaATG    | 85  | hypothetical protein<br>Astex_0306   | 3-81/183    | 44 | <i>Asticcacaulis excentricus</i><br>CB 48    | YP_004086155.1 |
| 46 | 39455 | 39655 | unknown  | + | GGCGAAGctgctgaATG      | 66  | monooxygenase, FAD-binding   | 385-445/546 | 38 | <i>Streptomyces griseoflavus</i><br>Tu4000   | ZP_07309792.1  |
| 47 | 39652 | 39840 | unknown  | + | AAGGAGtacgccaccATG     | 62  | hypothetical protein<br>METUNv1_00516  | 11-65/68    | 39 | <i>Methyloversatilis universalis</i><br>FAM5 | ZP_08503515.1  |
| 48 | 39882 | 40154 | unknown  | + | AAAAGGAGtaacgaacATG    | 90  | hypothetical protein<br>Cflav_PD2164   | 58-133/172  | 30 | bacterium Ellin514                           | ZP_03630603.1  |
| 49 | 40138 | 40374 | unknown  | + | GAACCGGAttacgattATG    | 78  | hypothetical protein<br>ORF047   | 2-77/77     | 67 | <i>Pseudomonas</i> phage 73                  | YP_001293454.1 |
| 50 | 40374 | 40550 | unknown  | + | GGGTTAcgaataATG        | 58  |  | 90-140/227  | 29 |  | YP_004435864.1 |

**Table 1 KL1 genome annotation (Continued)**

|    |       |       |                |   |                    |     |   |           |    |  |  |                |
|----|-------|-------|----------------|---|--------------------|-----|---|-----------|----|--|--|----------------|
|    |       |       |                |   |                    |     | hypothetical protein<br>Glaag_3667      |           |    |  | <i>Glaciecola</i> sp. 4 H-3-7 +<br>YE-5  |                |
| 51 | 40562 | 40933 | unknown        | + | GAAAGGtgaaatcATG   | 123 | hypothetical protein<br>BURMUCGD2M_4586 | 8-67/70   | 34 |  | <i>Burkholderia multivorans</i><br>CGD2M | ZP_03569237.1  |
| 52 | 40930 | 41415 | dCMP deaminase | + | GGAACGtccggcATG    | 161 | hypothetical protein ORF049             | 2-153/155 | 75 |  | <i>Pseudomonas</i> phage 73              | YP_001293456.1 |
| 53 | 41412 | 41786 | unknown        | + | AAAGGctgaatcATG    | 124 | hypothetical protein ORF050             | 4-125/127 | 43 |  | <i>Pseudomonas</i> phage 73              | YP_001293457.1 |
| 54 | 41826 | 42032 | unknown        | + | GGGGAtgcccacattATG | 68  | hypothetical protein ORF051             | 37-94/94  | 45 |  | <i>Pseudomonas</i> phage 73              | YP_001293458.1 |
| 55 | 42120 | 42674 | unknown        | + | AAGGAGttttacaaATG  | 184 | hypothetical protein ORF052             | 9-190/190 | 66 |  | <i>Pseudomonas</i> phage 73              | YP_001293459.1 |

**Table 2 AH2 genome annotation**

| Gene | Start | End  | Putative function            | Strand | Predicted ribosome binding site and start codon | Length (amino acids) | Closest relative                       | Alignment region (amino acids) | Percent identity | Source                                     | GenBank accession number |
|------|-------|------|------------------------------|--------|---|----------------------|--|--------------------------------|------------------|--|--------------------------|
| 1    | 619   | 1035 | unknown                      | -      | AAGGAAAcgacATG                                  | 138                  | hypothetical protein Nazgul32          | 12-130/130                     | 29               | <i>Burkholderia</i> phage BcepNazgul       | NP_918966.1              |
| 2    | 1073  | 1423 | unknown                      | -      | AGGGGGGAACggccATG                               | 116                  | conserved hypothetical protein         | 1-116/116                      | 72               | <i>Burkholderia multivorans</i> CGD1       | ZP_03586942.1            |
| 3    | 1501  | 1818 | unknown                      | -      | GGATTActgaccATG                                 | 105                  | family 2 glycosyl transferase          | 292-387/387                    | 32               | <i>Haloterrigena turkmenica</i> DSM 5511   | YP_003404522.1           |
| 4    | 1809  | 2024 | unknown                      | +      | GAGAAAtagagATG                                  | 71                   | mobilization protein mbeA              | 190-237/325                    | 37               | <i>Escherichia coli</i> E128010            | EFZ49597.1               |
| 5    | 2021  | 2578 | unknown                      | -      | AGGGGttacatcATG                                 | 185                  | hypothetical protein Nazgul06          | 88-158/330                     | 44               | <i>Burkholderia</i> phage BcepNazgul       | NP_919015.1              |
| 6    | 2728  | 2877 | unknown                      | -      | AGGTGcaaaaATG                                   | 49                   | hypothetical protein BokIE_20935       | 6-38/38                        | 48               | <i>Burkholderia oklahomensis</i> EO147     | ZP_02357945.1            |
| 7    | 2874  | 3002 | unknown                      | -      | AGGGGcgatcATG                                   | 42                   | polysaccharide deacetylase             | 21-60/287                      | 35               | <i>Bacillus mycoides</i> Rock3-17          | ZP_04156726.1            |
| 8    | 3071  | 3325 | unknown                      | -      | AAAGAgctATG                                     | 84                   | major facilitator superfamily MFS_1    | 131-209/467                    | 37               | <i>Burkholderia gladioli</i> BSR3          | YP_004349464.1           |
| 9    | 3322  | 3579 | unknown                      | -      | GGAGTAtccgccATG                                 | 85                   | hypothetical protein Plabr_1809        | 308-361/603                    | 31               | <i>Planctomyces brasiliensis</i> DSM 5305  | YP_004269441.1           |
| 10   | 3663  | 3911 | unknown                      | -      | GGGGGTatgacATG                                  | 82                   | HAD-superfamily hydrolase              | 70-119/268                     | 38               | <i>Methanosphaerula palustris</i> E1-9c    | YP_002465429.1           |
| 11   | 3913  | 4314 | unknown                      | -      | AGGGGGAGtaacggccATG                             | 133                  | hypothetical protein Nazgul09          | 1-129/141                      | 59               | <i>Burkholderia</i> phage BcepNazgul       | NP_919018.1              |
| 12   | 4320  | 4805 | unknown                      | -      | AGGGGttacatcATG                                 | 161                  | hypothetical protein Nazgul10          | 1-151/160                      | 74               | <i>Burkholderia</i> phage BcepNazgul       | NP_919019.2              |
| 13   | 4846  | 5454 | unknown                      | -      | AAAAAGGGGttttgacATG                             | 202                  | 194 gene product                       | 101-187/188                    | 43               | <i>Salmonella</i> phage PVP-SE1            | YP_004894001.1           |
| 14   | 6021  | 6302 | unknown                      | +      | AAGGAGcaatcATG                                  | 93                   | hypothetical protein Nazgul13          | 3-93/93                        | 41               | <i>Burkholderia</i> phage BcepNazgul       | NP_919022.1              |
| 15   | 6311  | 6550 | unknown                      | +      | AGGCGGtcgtATG                                   | 79                   | hypothetical protein BDB_mp60418       | 1-67/67                        | 45               | blood disease bacterium R229               | CCA83252.1               |
| 16   | 6707  | 7015 | unknown                      | +      | ACACGAcaccATG                                   | 102                  | hypothetical protein MC7420_4162       | 43-84/88                       | 45               | <i>Microcoleus chthonoplastes</i> PCC 7420 | ZP_05027813.1            |
| 17   | 7012  | 7218 | unknown                      | +      | GAAGGtccggcATG                                  | 68                   | hypothetical protein Cy51472DRAFT_4929 | 53-81/152                      | 45               | <i>Cyanothece</i> sp. ATCC 51472           | ZP_08976132.1            |
| 18   | 7215  | 8069 | unknown                      | +      | AGGAAAGgaaATG                                   | 284                  | hypothetical protein TK90_2682         | 5-175/177                      | 45               | <i>Thioalkalivibrio</i> sp. K90mix         | YP_003494636.1           |
| 19   | 8123  | 8407 | unknown                      | +      | GAGAAGGcacacacATG                               | 94                   | GTP-binding protein                    | 150-232/1016                   | 29               | <i>Gemmata</i> sp. Wa1-1                   | AAX07516.1               |
| 20   | 8499  | 9128 | DNA polymerase III β subunit | +      | GAACGGTGAGcttATG                                | 209                  | hypothetical protein Nazgul21          | 24-216/237                     | 24               | <i>Burkholderia</i> phage BcepNazgul       | NP_918955.1              |

**Table 2 AH2 genome annotation (Continued)**

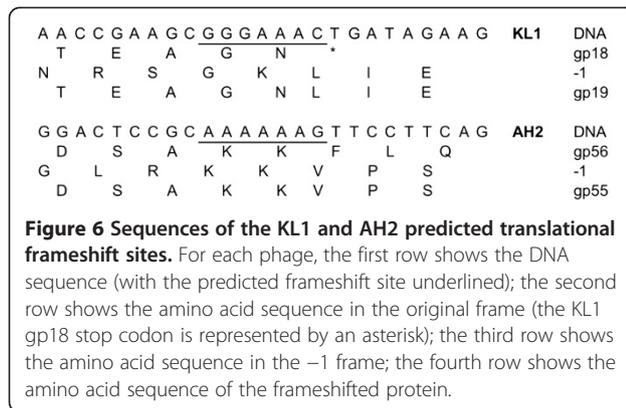
|    |       |       |                           |   |                          |     |   |              |    |   |                |
|----|-------|-------|---------------------------|---|--------------------------|-----|---|--------------|----|---|----------------|
| 21 | 9149  | 9343  | unknown                   | + | AGGAGAAAggagATG          | 64  | hypothetical protein R2APBS1DRAFT_0277                                | 9-63/344     | 31 | <i>Rhodanobacter</i> sp. 2APBS1                                   | ZP_08951135.1  |
| 22 | 9346  | 9645  | unknown                   | + | GGGGGTAtctgaccATG        | 99  | hypothetical protein PFL_2108   | 3-63/70      | 33 | <i>Pseudomonas fluorescens</i> Pf-5                               | YP_259216.1    |
| 23 | 9642  | 9938  | unknown                   | + | GGAGGGtcaTTG             | 98  | aspA gene product   | 38-122/317   | 32 | <i>Rhodospirillum centenum</i> SW                                 | YP_002297975.1 |
| 24 | 9935  | 10171 | unknown                   | + | GGGGcttgccgtATG          | 78  | hypothetical protein Nazgul19   | 18-97/97     | 39 | <i>Burkholderia</i> phage BcepNazgul                              | NP_919028.2    |
| 25 | 10256 | 10711 | pyrophosphohydrolase      | + | AAGGAAAggacATG           | 151 | hypothetical protein BCAS0549   | 15-139/140   | 60 | <i>Burkholderia cenocepacia</i> J2315                             | YP_002153936.1 |
| 26 | 10720 | 10977 | unknown                   | + | GAGGccgcccATG            | 85  | hypothetical protein AGRO_3677  | 208-273/300  | 41 | <i>Agrobacterium</i> sp. ATCC 31749                               | ZP_08529674.1  |
| 27 | 11082 | 12074 | unknown                   | + | AGGAGAAatcGTG            | 330 | hypothetical protein  | 8-95/113     | 48 | <i>Escherichia</i> phage vB_EcoM_ECO1230-10                       | ADE87960.1     |
| 28 | 12101 | 13075 | transcriptional regulator | + | AAGGAAccgacATG           | 324 | hypothetical protein Pnap_4317  | 25-252/342   | 45 | <i>Polaromonas naphthalenivorans</i> CJ2                          | YP_973341.1    |
| 29 | 13078 | 13497 | unknown                   | + | GCTGACGAtctctgaccATG     | 139 | hypothetical protein SCHCODRAFT_69044                                 | 549-631/848  | 33 | <i>Schizophyllum commune</i> H4-8                                 | XP_003030158.1 |
| 30 | 13574 | 13768 | transcriptional regulator | + | AGGGAttttcATG            | 64  | hypothetical protein APT_2164   | 9-65/75      | 53 | <i>Acetobacter pasteurianus</i> NBRC 101655                       | GAB28674.1     |
| 31 | 13768 | 14031 | transcriptional regulator | + | AAGCGGAGccgtctgATG       | 87  | hypothetical protein Bcep1808_2468                                    | 2-85/86      | 73 | <i>Burkholderia vietnamiensis</i> G4                              | YP_001120302.1 |
| 32 | 14064 | 14450 | Vsr endonuclease          | - | GGAGGAatgATG             | 128 | DNA mismatch endonuclease Vsr   | 15-141/141   | 65 | <i>Methylocella silvestris</i> BL2                                | YP_002360880.1 |
| 33 | 14450 | 15025 | excinuclease              | - | AACAGAGttgcagcGTG        | 191 | Excinuclease ABC C subunit domain protein                             | 3-183/192    | 58 | <i>Pseudomonas syringae</i> pv. <i>lachrymans</i> str. M301315    | EGH83133.1     |
| 34 | 15038 | 15892 | restriction endonuclease  | - | GGCAAAGGtcgccgacATG      | 284 | conserved hypothetical protein  | 1-285/285    | 70 | <i>Ralstonia solanacearum</i> CMR15                               | CBJ36134.1     |
| 35 | 15889 | 17031 | cytosine methylase        | - | AGGGGGttccgGTG           | 380 | DNA-cytosine methyltransferase  | 1-385/385    | 66 | <i>Ralstonia solanacearum</i> CMR15                               | CBJ36133.1     |
| 36 | 17107 | 17199 | unknown                   | + | ACGAAGccttgcttaATG       | 30  | resistance-nodulation-cell division acriflavin:proton (H+) antiporter | 850-868/1014 | 68 | <i>Bacillus pumilus</i> SAFR-032                                  | YP_001486844.1 |
| 37 | 17511 | 18842 | integrase                 | + | GAAGGAGGtcttgtagcactgATG | 443 | chorismate mutase family protein                                      | 1-362/386    | 62 | <i>Phaeobacter gallaeciensis</i> BS107                            | ZP_02147383.1  |
| 38 | 18990 | 19412 | unknown                   | + | AAGGAGGAatcATG           | 140 | hypothetical protein Dda3937_00584                                    | 60-163/163   | 40 | <i>Dickeya dadantii</i> 3937                                      | YP_003882998.1 |
| 39 | 19462 | 20001 | unknown                   | - | GGAGAttttcATG            | 179 | hypothetical protein PcarcW_20243                                     | 68-197/198   | 67 | <i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i> WPP14 | ZP_03833564.1  |
| 40 | 20034 | 20264 | Rz1                       | - | GGAGGAcgccATG            | 76  | hypothetical protein BURPS668_A2333                                   | 27-81/81     | 62 | <i>Burkholderia pseudomallei</i> 668                              | YP_001063327.1 |

**Table 2 AH2 genome annotation (Continued)**

|    |                                   |   |                     |      |   |                      |    |  |                |
|----|-----------------------------------|---|---------------------|------|---|----------------------|----|--|----------------|
| 41 | 20277 20588 Rz                    | - | AGGGGGccgtATG       | 103  | hypothetical protein ORF004             | 2-101/101            | 35 | <i>Pseudomonas</i> phage 73  | YP_001293411.1 |
| 42 | 20585 21091 lysin                 | - | AAGGAGAAGAcacGTG    | 168  | hypothetical protein HMPREF0005_02034   | 1-161/163            | 60 | <i>Achromobacter xylosoxidans</i> C54                                    | EFV83908.1     |
| 43 | 21088 21339 holin                 | - | GAAGGGGtgaccaccATG  | 83   | conserved exported hypothetical protein | 1-83/85              | 35 | blood disease bacterium R229   | CCA83792.1     |
| 44 | 21336 21665 unknown               | - | AAGGGGccagaagATG    | 109  | hypothetical protein HDEF_1702          | 3-87/92              | 31 | Candidatus <i>Hamiltonella defensa</i> 5AT ( <i>Acyrtosiphon pisum</i> ) | YP_002924457.1 |
| 45 | 21807 22121 unknown               | - | AAGGAGAAAcacATG     | 104  | hypothetical protein PPL19_05085        | 1-103/161            | 53 | <i>Pseudomonas psychrotolerans</i> L19                                   | ZP_09283635.1  |
| 46 | 22133 23731 tail fiber protein    | - | GGAACGtgacATG       | 532  | hypothetical protein Bpse112_32291      | 69-240/282           | 45 | <i>Burkholderia pseudomallei</i> 112                                     | ZP_02502292.1  |
| 47 | 23809 26178 tail assembly protein | - | AGAGGAAGAcacATG     | 789  | hypothetical protein HCH_05649          | 2-727/728            | 34 | <i>Hahella chejuensis</i> KCTC 2396                                      | YP_436732.1    |
| 48 | 26175 26375 tail assembly protein | - | GGGGGCAAgaaATG      | 66   | hypothetical protein HCH_05650          | 4-67/71              | 50 | <i>Hahella chejuensis</i> KCTC 2396                                      | YP_436733.1    |
| 49 | 26372 26608 tail assembly protein | - | GAGGActgatcATG      | 78   | putative transmembrane protein          | 7-82/82              | 47 | <i>Rhodobacter</i> sp. SW2   | ZP_05845047.1  |
| 50 | 26618 27418 tail assembly protein | - | AGGGGGAtcaacaATG    | 266  | hypothetical protein HCH_05652          | 1-268/269            | 39 | <i>Hahella chejuensis</i> KCTC 2396                                      | YP_436735.1    |
| 51 | 27415 29100 tail assembly protein | - | AAGAAGAtcacTTG      | 561  | hypothetical protein HCH_05654          | 35-560/563           | 32 | <i>Hahella chejuensis</i> KCTC 2396                                      | YP_436736.1    |
| 52 | 29097 30158 unknown               | - | GACGAGGtttgaaATG    | 353  | hypothetical protein D11S_2171          | 1-326/327            | 23 | <i>Aggregatibacter actinomycetemcomitans</i> D11S-1                      | YP_003256741.1 |
| 53 | 30160 31122 unknown               | - | GAGCGAGGcataacGTG   | 320  | hypothetical protein XALc_0225          | 1-194/307            | 35 | <i>Xanthomonas albilineans</i> GPE PC73                                  | YP_003374757.1 |
| 54 | 31124 35860 tail tape measure     | - | GGACTGAAcggaaATG    | 1578 | phage tape measure protein              | 1-109, 452-1680/1683 | 33 | <i>Sinorhizobium meliloti</i> AK83                                       | YP_004548730.1 |
| 55 | 35853 36538 tail protein          | - | AAGGGGGCGgacATG     | 228  | pre-tape measure frameshift protein G-T | 1-242/243            | 34 | <i>Burkholderia</i> phage BcepNazgul                                     | NP_918998.2    |
| 56 | 36098 36538 tail protein          | - | AAGGGGGCGgacATG     | 146  | hypothetical protein Sinme_1368         | 4-126/142            | 34 | <i>Sinorhizobium meliloti</i> AK83                                       | YP_004548729.1 |
| 57 | 36549 37337 unknown               | - | GAGGAAtcaatcATG     | 262  | hypothetical protein Sinme_1367         | 1-257/262            | 45 | <i>Sinorhizobium meliloti</i> AK83                                       | YP_004548728.1 |
| 58 | 37385 37897 minor tail protein    | - | GAGGAAAgataATG      | 170  | hypothetical protein Sinme_1366         | 7-177/177            | 50 | <i>Sinorhizobium meliloti</i> AK83                                       | YP_004548727.1 |
| 59 | 37897 38517 unknown               | - | GACGCAGGtttccgacATG | 206  | hypothetical protein Nazgul55           | 5-198/205            | 49 | <i>Burkholderia</i> phage BcepNazgul                                     | NP_918988.2    |
| 60 | 38514 38873 unknown               | - | GAGGCCgctgATG       | 119  | hypothetical protein Sinme_1364         | 3-120/125            | 38 | <i>Sinorhizobium meliloti</i> AK83                                       | YP_004548725.1 |

**Table 2 AH2 genome annotation (Continued)**

|    |             |                                     |   |                       |     |                                    |             |    |  |                |
|----|-------------|-------------------------------------|---|-----------------------|-----|------------------------------------|-------------|----|--|----------------|
| 61 | 38886 39134 | unknown                             | - | AAAGGAAccatcATG       | 82  | hypothetical protein Nazgul57      | 1-38/85     | 47 | <i>Burkholderia</i> phage BcepNazgul           | NP_918990.1    |
| 62 | 39205 40233 | major capsid protein                | - | AAGGAGAAAGcaaaATG     | 342 | capsid protein E                   | 2-343/346   | 50 | <i>Burkholderia</i> phage BcepNazgul           | NP_918991.1    |
| 63 | 40290 40688 | decorator protein                   | - | AGGAGAAccatcATG       | 132 | decorator protein D                | 4-123/131   | 49 | <i>Burkholderia</i> phage BcepNazgul           | NP_918992.1    |
| 64 | 40743 42071 | prohead protease                    | - | AGGACCAGAAccaATG      | 442 | prohead protease ClpP              | 4-427/434   | 53 | <i>Burkholderia</i> phage BcepNazgul           | NP_918994.2    |
| 65 | 42068 43591 | portal protein                      | - | GGAAccctcgATG         | 507 | phage portal protein               | 57-554/559  | 59 | <i>Staphylococcus</i> phage SA1                | ACZ55505.1     |
| 66 | 43736 43960 | head-tail joining protein           | - | GGACAAcactATG         | 74  | head-tail joining protein Lambda W | 13-76/76    | 56 | <i>Burkholderia</i> phage BcepNazgul           | NP_918996.1    |
| 67 | 44097 46076 | terminase large subunit             | - | AAGAcctcgATG          | 659 | terminase large subunit TerL       | 44-677/677  | 58 | <i>Burkholderia</i> phage BcepNazgul           | NP_918997.2    |
| 68 | 46210 46803 | terminase small subunit             | - | GAAGGTGAtagcgATG      | 197 | TerS                               | 9-179/222   | 49 | <i>Burkholderia</i> phage BcepNazgul           | NP_918999.1    |
| 69 | 46796 46990 | transcriptional regulator           | - | AGGAGTAcggtATG        | 64  | aminoglycoside phosphotransferase  | 423-473/487 | 29 | <i>Frankia</i> sp. EUN1f                       | ZP_06416368.1  |
| 70 | 47047 47736 | repressor                           | - | GAAAGGCAAGGcagcagcATG | 229 | hypothetical protein Rvan_1213     | 14-180/242  | 36 | <i>Rhodococcus</i> <i>vannielii</i> ATCC 17100 | YP_004011581.1 |
| 71 | 47833 49446 | helicase                            | - | ACGAcctctcgATG        | 537 | helicase                           | 11-507/522  | 52 | <i>Burkholderia</i> phage BcepNazgul           | NP_919000.2    |
| 72 | 49443 49745 | resolvase                           | - | GAAAGGAGGAttcactGTG   | 100 | conserved phage protein            | 15-103/108  | 55 | <i>Burkholderia</i> phage BcepNazgul           | NP_919001.2    |
| 73 | 49742 51796 | DNA polymerase                      | - | ACGTcaccATG           | 684 | hypothetical protein ORF026        | 48-670/683  | 45 | <i>Pseudomonas</i> phage 73                    | YP_001293433.1 |
| 74 | 51875 52609 | single-stranded DNA binding protein | - | AAAGGTGAcaaaaATG      | 244 | conserved phage protein            | 4-186/198   | 35 | <i>Staphylococcus</i> phage SA1                | ACZ55548.1     |
| 75 | 52655 53995 | Cas4 superfamily exonuclease        | - | GATCctctcgaccccATG    | 446 | conserved phage protein            | 8-448/454   | 48 | <i>Burkholderia</i> phage BcepNazgul           | NP_919005.2    |
| 76 | 54140 54538 | unknown                             | - | GGAGAAatcATG          | 132 | hypothetical protein RUMHYD_01446  | 1-120/122   | 26 | <i>Blautia hydrogenotrophica</i> DSM 10507     | ZP_03782010.1  |
| 77 | 54718 55017 | Cro                                 | + | AACGGAGAtcacaATG      | 99  | hypothetical protein Nazgul73      | 5-90/97     | 31 | <i>Burkholderia</i> phage BcepNazgul           | NP_919007.1    |
| 78 | 55054 57534 | primase                             | + | GGAGGGgcaATG          | 826 | DR0530-like primase                | 1-843/843   | 49 | <i>Burkholderia</i> phage BcepNazgul           | NP_919008.2    |



gp24-gp26 are similar to BcepNazgul tail assembly proteins. Using HHpred analysis, we were able to identify an additional three proteins at a probability threshold of 75%. Gp8 is similar to bacteriophage SPP1 portal protein (99.44% probability), gp15 is similar to  $\lambda$  gpFII head-tail joining protein (82.86% probability), and gp16 is similar to  $\lambda$  gpU minor tail protein (77.70% probability) (Additional file 1: Table S1).

In comparison with KL1, the structural proteins of AH2 are well defined. Genes 62-68 make up the capsid morphogenesis and DNA packaging module, containing genes encoding the major capsid protein, decorator protein, prohead protease, portal protein, head-tail joining protein, and terminase subunits (large and small) (Table 2). Each of these proteins is similar to a BcepNazgul protein, with percent identities between 49-58%. Several genes between 47 and 56 are similar to genes encoding BcepNazgul conserved tail assembly proteins, tape measure protein, and pre-tape measure frameshift protein G-T (with percent identities between 26-38%). Two additional AH2 tail proteins were identified using BLASTP (gp46, similar to *Pseudomonas psychrotolerans* L19 phage tail fiber protein) or HHpred (gp58, similar to  $\lambda$  gpU minor tail protein) analysis (Additional file 2: Table S2). Hypothetical proteins encoded in this region are likely to be involved in tail morphogenesis based on the proximity of their genes to this module.

Most tailed phages encode two tail proteins proximal to the tail tape measure gene by way of a -1 translational frameshift [35]. We have previously identified these frameshifted genes in the BCC-specific phages KS9, KS5, KS14, and KL3 [19,29]. Using FSFinder and manual scanning for XXXYYZ motifs, we predict that KL1 gp18/gp19 and AH2 gp55/gp56 are expressed using this mechanism. The predicted frameshift site in KL1 is GGGAAAC, immediately upstream of the gp18 TGA stop codon (Figure 6 and Additional file 3: Figure S3). A -1 ribosomal shift following the terminal C will allow for expression of the 264 aa gp19 and the 142 aa gp18 from the same start codon (Figure 6). Although most

phages encode their frameshifted proteins immediately upstream of the tail tape measure gene, KL1 encodes an intervening tail protein, gp20 (Table 1, Figure 5). This organization is similar to that of *Escherichia coli* phage HK97, *Bacillus subtilis* phage SPP1, *Methanobacterium thermoautotrophicum* phage  $\psi$ M2, *Methanothermobacter wolfei* phage  $\psi$ M100, *Lactococcus* phages c2 and BIL67, and *Natrialba magadii* phage  $\phi$ ch1 [35]. The predicted frameshift site in AH2 is AAAAAAG (Figure 6 and Additional file 3: Figure S3), the same sequence used by *E. coli* phage VT1-Sakai, *M. thermoautotrophicum* phage  $\psi$ M2, *Staphylococcus aureus* phages PVL and PV83, *Lactococcus lactis* phage ul36, and *Borrelia burgdorferi* prophage Borreliapro [35]. In the case of AH2, a -1 shift of the ribosome following the G in this sequence will allow for the 228 aa gp55 to be expressed instead of the 146 aa gp56 (Figure 6). Using BLASTP or HHpred searches, we were unable to identify the KL1 or AH2 major tail proteins. However, we predict that these proteins may be gp17 in KL1 and gp57 in AH2 as the major tail genes are generally positioned upstream of the frameshifted protein genes [35]. Although not present in all sequences, RNA secondary structures are often found downstream of frameshift sites [19,29,35,36]. Mfold analysis of the 35 bases downstream of the putative KL1 and AH2 sites suggests that stem-loop structures could form in both of these regions (Additional file 3: Figure S3).

### Lysis

In KL1, we have identified the genes putatively encoding the holin, lysin, Rz and Rz1 lysis proteins. In a BLASTP search, gp2 shows similarity to putative holin proteins of PA73 and BcepNazgul. TMHMM analysis of this protein indicates that it has two transmembrane domains, so gp2 is predicted to be a class II holin [37]. Gp3 is similar to the endolysin of *Erwinia* phage vB\_EamP-S6 (HQ728266) and contains lysozyme and peptidoglycan-binding conserved domains. Although gp4 does not show similarity to any Rz proteins in the NCBI database, it is predicted to contain a single N-terminal transmembrane domain, characteristic of Rz proteins [38]. Gp5 is predicted to be the KL1 Rz1 protein as it is similar to BcepNazgul Rz1 and LipoP analysis identifies a signal peptidase II cleavage site between positions 17 and 18 (resulting in a 70 aa protein with 4 proline residues [5.7% proline]). The proportion of prolines in the predicted Rz1 lipoprotein is low compared to previously identified Rz1 proteins in BCC phages [19,29,39].

The same lysis proteins were identified in AH2. Like KL1 gp2, the putative AH2 holin gp43 is similar to the BcepNazgul holin, has two transmembrane domains, and is predicted to be a class II holin. Although gp42 shows no similarity to endolysins in a BLASTP search,

HHpred analysis reveals similarity to both eukaryotic and prokaryotic lysozyme proteins. Gp41 is predicted to be the AH2 Rz protein as it has a single N-terminal transmembrane domain. Although manual annotation has been required for identification of the Rz1 gene in KL1 and in our previous studies [19,29], we predict that the GeneMark-assigned gp40 is the AH2 Rz1 protein. Gp40 is similar to BcepNazgul Rz1 and has a signal peptidase II cleavage site between amino acids 15 and 16. Similar to the predicted KL1 Rz1, the proportion of prolines present in this protein is relatively low (3/61 or 4.9%). It is unclear from this analysis what protein(s) may contribute to the unique plaque phenotype observed in both of these phages. Aside from the low proportion of proline found in the putative Rz1 proteins, KL1 and AH2 appear to have relatively standard lysis modules, suggesting that unique (and as yet unidentified) proteins may be responsible for controlling lysis timing in each phage.

#### DNA binding

Of the 8 KL1 proteins similar to a PA73 protein with an assigned function, half of these are DNA- or nucleotide-binding proteins: DNA polymerase (gp27), superfamily II helicase/restriction enzyme (gp30), helicase (annotated here as recombinase [gp33]), and dCMP deaminase (gp52) (Table 1). In addition, KL1 encodes a putative DNA polymerase III  $\beta$  subunit (gp28), exonuclease (gp31), transcriptional regulator (gp36), primase (gp37), and Vsr endonuclease (gp44) (Table 1 and Additional file 1: Table S1). In a multi-genome analysis performed by Lopes et al. [40], it was determined that PA73 ORF032 is distantly related to *Lactococcus* phage  $\phi$ 31 Sak4 recombinase. When this protein was expressed in *E. coli*, it exhibited recombinase activity, but was found to be less efficient than  $\lambda$  Red $\beta$  [40]. Furthermore, PA73 encodes an exonuclease, as is found in characterized phage recombinase pairs such as Red $\alpha\beta$  in  $\lambda$  and RecET in *rac* [40]. KL1 gp33 is most closely related to PA73 ORF032 and, with 91% identity, is the KL1 protein most similar to a PA73 protein. In addition, KL1 gp31 has 65% identity with PA73 ORF030 and both of these proteins are similar to  $\lambda$  Red $\alpha$  (99.21% probability for gp31 and 99.17% probability for ORF030) (Table 1 and Additional file 1: Table S1). It is interesting to note that, despite the relatively limited similarity between KL1 and other previously sequenced BCC-specific phages, both gp31 and gp33 are similar to proteins from *Burkholderia* phage BcepGomr (BcepGomrgp43 and BcepGomrgp45, respectively) [40]. Although further characterization of these proteins is required in both KL1 and BcepGomr, it is possible that these exonucleases and Sak4-like recombinases represent a conserved recombination system in certain BCC-specific phages.

AH2 encodes DNA replication, modification, and repair proteins including a putative DNA polymerase III  $\beta$  subunit (gp20), Vsr endonuclease (gp32), excinuclease (gp33), restriction endonuclease/methylase pair (gp34/gp35), integrase (gp37), helicase (gp71), resolvase (gp72), DNA polymerase (gp73), single-stranded DNA binding protein (gp74), Cas4 superfamily exonuclease (gp75), and primase (gp78) (Table 2). Other putative DNA binding proteins are predicted to be involved in transcriptional regulation. Gp28 is similar to partitioning and regulation proteins from *Thermus thermophilus* (100% probability) and *E. coli* (99.86% probability) (Additional file 2: Table S2). The gp30 and gp31 predicted proteins belong to the helix-turn-helix MerR superfamily and the pyocin activator superfamily, respectively. Both of these proteins, in addition to gp69, also show similarity to excisionase proteins (Additional file 2: Table S2). Gp70 and gp77 are similar to the lysogeny control proteins CI from enterobacteria phage 186 (99.87% probability) and Cro from *Xylella fastidiosa* Ann-1 (96.60% probability), respectively (Additional file 2: Table S2).

AH2 gp32-gp35 are predicted to be part of a DNA protection and repair module. Vsr (very short patch repair) endonucleases are involved in the repair of 5-methylcytosine to thymine deamination [41]. Previously, we identified a Vsr endonuclease in the BCC-specific phage KL3 that, along with an EcoRII-C endonuclease/methylase pair, was predicted to be part of a novel non-self DNA degradation and self DNA protection/repair module [29]. Our model proposed that non-KL3 DNA (i.e. that of the host or a superinfecting phage) would be degraded by the endonuclease (gp45), while KL3 DNA would be protected by the methylase (gp47) (converting cytosine to 5-methylcytosine). Vsr endonuclease (gp46) and very short patch repair would then prevent the accumulation of mutations caused by 5-methylcytosine deamination [29].

The DNA protection and repair system of AH2 is analogous to that of KL3. AH2 gp32 has 51% identity with the KL3 Vsr endonuclease and is similar to *E. coli* Vsr endonuclease (100% probability) (Additional file 2: Table S2). AH2 also encodes an endonuclease/methylase pair: gp34 is similar to *Kluyvera ascorbata* KasI (64% identity) while gp35 is similar to *K. ascorbata* M.KasI, *Brevundimonas diminuta* ATCC 11568 cytosine-specific methyltransferase NlaX, and *Acetobacter pomorum* DM001 modification methylase HpaII (63-66% identity). Gp35 also has several methylase conserved domains, including Dcm (an enzyme that produces 5-methylcytosine bases at sites recognized by Vsr endonuclease) [41]. Gp33 is similar to *Thermotoga maritima* UvrABC system protein C (98.35% probability) and could function together with UvrAB in nucleotide excision repair (Additional file 2: Table S2) [42]. Although further

experiments are required to identify the recognition sites of gp34 and gp35, we predict that this module may function as follows: gp34 cleaves non-self DNA, while self DNA is protected by gp35 methylation and subsequent gp32 repair (with gp33 participating in nucleotide excision repair). Although the identity and arrangement of genes in this module is different in AH2 than in KL3, the identification of a similar module in an unrelated BCC-specific phage suggests that these genes may be widely used for DNA protection and repair in this group of viruses.

### **MazG**

A notable protein encoded by both KL1 and AH2 is MazG. MazG is a pyrophosphohydrolase that acts on ppGpp, one of the signaling molecules in bacteria produced during the stringent response [43]. When bacterial cells are in an amino acid-limited environment, RelA synthesizes pppGpp, the precursor of ppGpp, and the latter activates the expression of genes required for cell survival (such as *rpoS*) and represses genes required for protein and DNA synthesis (reviewed in [44]). Recently, there has been a great deal of interest in marine phages (especially cyanophages) that encode MazG homologs, such as *Prochlorococcus* phages P-SSM2 and P-SSM4, *Synechococcus* phage S-PM2, *Prochlorococcus* and *Synechococcus* phage Syn9, *Roseobacter* phage SIO1, *Pseudalteromonas* phage H105/1, almost one-fifth of the cyanophages tested by Bryan et al. [45], and all of the cyanophages analyzed by Sullivan et al. [46] [47-51]. It has been suggested that these MazG-encoding phages are better able to infect and propagate within their hosts, which are found in nutrient-limited water. By inactivating ppGpp, these phages can promote the expression of genes that would usually be expressed by an exponential phase cell under nutrient-rich conditions, such as those required for protein and DNA synthesis [52]. There are few published reports of the *mazG* gene in non-marine phages, but it has been previously identified in *Mycobacterium* phage Mx8 and mycobacteriophage L5 [45].

The putative MazG proteins encoded by KL1 and AH2 are gp35 and gp25, respectively. KL1 gp35 is similar to putative MazG proteins from phages infecting *Synechococcus* (including S-CRM01, S-SM2, and S-ShM2), *Prochlorococcus* (including P-HM1, P-HM2, and P-SSM2), and *Bacillus* (0305φ8-36), as well as to PA73 hypothetical protein ORF034 (Table 1). AH2 gp25 is similar to putative *Clostridium* MazG proteins and to the *Burkholderia* phage proteins φE255 gp37, BcepMu gp06, and BcepB1A gp71. Both gp35 and gp25 are similar to *E. coli* MazG (100% and 99.76% probability, respectively) (Additional file 1: Table S1 and Additional file 2: Table S2). Because BCC bacteria found in soil and water are likely to be nutrient-limited (similar to

cyanobacteria), MazG proteins in BCC-specific phages may help to facilitate infection in the environment. This protein may also be involved in the unique plaque phenotype of these phages, as the appearance of plaques at low titre after >16 h incubation (at which time the bacterial lawn appears intact) (Figure 1) suggests that lysis of stationary phase cells may be occurring. Such a trait would be especially important for clinical use, as phage activity may be increased against stationary and/or biofilm cells found in the CF lung.

MazG may also have an effect with respect to BCC pathogenicity. Synthesis of ppGpp has been associated with virulence in species such as *Legionella*, *Listeria*, *Pseudomonas*, *Salmonella*, *Mycobacterium*, and *Vibrio* (although the association in this species has been controversial) [53-59]. In *P. aeruginosa*, *relA* mutants are less virulent than the wildtype when tested in the *Drosophila melanogaster* model [55] and *relA spoT* mutants have reduced antibiotic tolerance [60]. Because MazG activity may mimic the effects of these mutations, it is possible that phage-encoded MazG could modulate the virulence and/or antibiotic tolerance of a lysogen. Further experiments are required to determine if the putative KL1 and AH2 MazG proteins have pyrophosphohydrolase activity, if these genes are expressed in lysogens, and if MazG expression has an effect on pathogenicity.

### **Convergent evolution**

Although there have been relatively few papers published on the subject, the occurrence of convergent evolution in bacteriophages has been documented previously. Most studies examine the phenomenon at the molecular level by identifying identical base pair and amino acid changes that occur in different phage lineages under the same environmental conditions [61-64]. Structural examples of convergent evolution, such as the *Caudovirales* tail and the tectivirus pseudo-tail, have been reviewed previously [65]. Given the ever-increasing number of completed phage genome sequences, it is expected that many more examples remain to be identified (particularly at the whole genome level). Furthermore, there are likely many examples in the literature of phages with similar phenotypes but dissimilar genomes that have not explicitly been identified as examples of convergent evolution, perhaps because they exhibit what is considered to be a "standard" plaque phenotype.

We predict that KL1 and AH2 represent examples of phage convergent evolution at the whole genome level. As discussed above, these two phages exhibit a plaque phenotype that is both similar and unique in comparison to all other BCC-specific phages that we have characterized previously. Because of these characteristics, KL1

and AH2 were thought to be the same phage prior to RFLP and genomic analysis. However, these phages appear to have convergently evolved because, as discussed throughout, their genomes are almost entirely dissimilar (Figure 4A). The relative rarity of this phenotype among characterized phages of the BCC and other species may be at least partially explained by sampling bias. Standard phage isolation protocols most readily identify those phages that have easily visible plaques on multiple hosts after overnight incubation at a broad range of temperatures. Phages such as KL1 and AH2 may be missed because of poorly visible plaques, incompatible hosts, insufficient incubation times, incorrect temperatures, titres that are too high or too low, overgrowth of bacteria, and/or competition by more rapidly lysing phages. As novel phages continue to be isolated from environmental samples using diverse bacterial hosts, the prevalence, distribution, and genetic basis of this phenotype should become more apparent.

Several mechanisms could explain the delayed plaque formation observed here, including long latent periods or lysis inhibition (both with concomitantly large burst sizes) [66], preferential infection of stationary phase cells, or the gradual release of diffusible lytic enzymes from small plaques. In order to differentiate these possibilities, we performed one-step growth curves for both phages using either exponential or stationary phase C6433 as a host. Using a variation of a standard protocol (described in Methods), the phage titres unexpectedly remained stable (within one order of magnitude) over a 4 h period. Given the uninformative nature of these results, we have thus far been unable to identify the mechanism(s) responsible for the plaque phenotype. Taking into consideration the very specific conditions required for the observation of KL1 or AH2 plaques on solid medium, we predict that the infection kinetics in liquid culture may be highly dependent upon host (both strain and growth phase), incubation time, temperature, titre, and potentially other factors (such as medium) that are not accounted for using standard one-step growth curve protocols.

## Conclusions

A recent publication by Ceysens et al. [67] provides an interesting counterpoint to our study. While we identified KL1 and AH2 as phages that were phenotypically-similar but genomically-distinct, this group analyzed a set of *Pseudomonas* phages that were phenotypically-distinct but genomically-similar. They found that, among  $\phi$ KMV-like viruses with between 83-97% nucleotide identity, there were significant differences observed with respect to latent period, host range, and antibody reactivity [67]. We have made similar observations with our collection of BCC-specific phages: two phages can

have distinct phenotypes with respect to liquid clearing and host range while at the same time having almost identical genomes [22]. Taken together, the observations made by Ceysens et al. [67] and those discussed in this study provide a) novel examples of both divergent and convergent phage evolution and b) further evidence of the broad diversity of phages that infect Gram-negative opportunistic pathogens.

## Methods

### Bacterial strains and growth conditions

*Burkholderia cenocepacia* strains K56-2 and C6433, part of the *Burkholderia cepacia* complex experimental strain panel [68,69], were used for phage isolation and propagation. Strains used for host range analysis (also part of the panel) were acquired from the Belgium Coordinated Collection of Microorganisms LMG Bacteria Collection (Ghent, Belgium) and the Canadian *Burkholderia cepacia* complex Research and Referral Repository (Vancouver, BC). Strains were grown aerobically overnight at 30°C on half-strength Luria-Bertani ( $\frac{1}{2}$  LB) solid medium or in  $\frac{1}{2}$  LB broth with shaking. Lysates for DNA isolation were prepared from soft agar overlays made with  $\frac{1}{2}$  LB medium containing agarose instead of agar.

### Phage isolation and propagation

KL1 and AH2 were isolated from sewage and *Nandina* sp. soil, respectively, using standard extraction protocols [26]. Environmental samples were incubated with shaking at 30°C in a slurry of  $\frac{1}{2}$  LB broth, suspension medium (SM) (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM MgSO<sub>4</sub>, 0.01% gelatin solution), and BCC liquid culture (K56-2 for KL1 isolation and C6433 for AH2 isolation). Solids were pelleted by centrifugation and the supernatant was filter-sterilized, plated in soft agar overlays with the BCC strain used in the extraction, and incubated overnight at 30°C and >24 h at room temperature. Plaques were picked using a sterile Pasteur pipette and transferred into 1 ml SM. Phage propagation was performed using soft agar overlays: 100  $\mu$ l liquid culture and 100  $\mu$ l phage stock (diluted in SM if necessary) were incubated 20 min at room temperature, mixed with 3 ml 0.7%  $\frac{1}{2}$  LB top agar, overlaid on a plate of  $\frac{1}{2}$  LB solid medium, and incubated at 30°C and room temperature until plaque formation was complete. High titre stocks were made by transferring multiple plaques into SM or by overlaying plates with SM and incubating 4–8 h at 4°C on a platform rocker.

### Lysis characterization

Host ranges were performed using soft agar overlays (as described above) or by spotting 10  $\mu$ l aliquots of phage stock (at multiple dilutions) onto a freshly-plated soft agar overlay containing 100  $\mu$ l liquid culture. K56-2 LPS

mutant [31,32] host ranges were performed similarly using wildtype K56-2, RSF19 (*wbxE::pRF201*), XOA7 (*waal::pGP $\Omega$ Tp*), XOA15 (*wabR::pGP $\Omega$ Tp*), XOA17 (*wabS::pGPApTp*), XOA8 (*wabO::pGP $\Omega$ Tp*), and CCB1 (*waac::pGP $\Omega$ Tp*) (kindly provided by Miguel Valvano).

One-step growth curves were performed using a variation of a standard protocol [39]. One hundred microliters of diluted phage lysate containing  $10^6$  PFU of KL1 or AH2 was mixed with  $10^8$  colony forming units of C6433 (900  $\mu$ l 5 h liquid culture [for exponential phase curves] or 100  $\mu$ l 16 h liquid culture diluted in 800  $\mu$ l spent  $\frac{1}{2}$  LB broth [for stationary phase curves]). The suspension was incubated 15 minutes at 30°C, diluted 1:1000 into a flask containing  $\frac{1}{2}$  LB broth (exponential) or spent  $\frac{1}{2}$  LB broth (stationary), and incubated without shaking at 30°C. One milliliter samples were withdrawn at one hour intervals for 4 h. Two 100  $\mu$ l samples were plated immediately in soft agar overlays with C6433. One hundred microliters of chloroform was then added to the sample, mixed 5 s on a vortexer, and separated by centrifugation for 1 min at 13,000 rpm. Two 100  $\mu$ l chloroform-treated samples were then plated immediately in soft agar overlays with C6433. Plates were incubated 48 h at 30°C prior to plaque enumeration. Experiments were performed in triplicate for each condition (KL1 exponential or stationary phase, AH2 exponential or stationary phase).

#### Electron microscopy

Filter-sterilized high titre stocks of KL1 and AH2 were used for electron microscopy. 5–10  $\mu$ l of phage lysate was deposited onto a carbon-coated copper grid and incubated 5 min at room temperature. Following adsorption of excess lysate onto a filter paper, the grids were stained with 2% phosphotungstic acid for 2 min. Grids were viewed using a Philips/FEI (Morgagni) transmission electron microscope with charge-coupled device camera (University of Alberta Department of Biological Sciences Advanced Microscopy Facility).

#### DNA isolation, RFLP analysis, and sequencing

Phage DNA was isolated using polyethylene glycol precipitation and guanidine thiocyanate lysis. One hundred milliliters of phage lysate (propagated on C6433) was collected by overlaying turbid-clear or mottled  $\frac{1}{2}$  LB agarose plates with SM and incubating at 4°C 4–8 h on a platform rocker. Following the addition of chloroform, debris in the lysate was pelleted by centrifugation for 10 min at 10,000 rcf and 4°C and the supernatant was filter-sterilized with a Millex-HA 0.45  $\mu$ m syringe driven filter unit (Millipore, Billerica, MA). Fifty milliliter aliquots of the supernatant were incubated at 37°C  $\geq$ 40 min with 10  $\mu$ l DNase I, 10  $\mu$ l DNase I buffer and 6  $\mu$ l RNase (Fermentas, Burlington, ON) to degrade

contaminating bacterial nucleic acids. Following centrifugation for 10 min at 4000 rcf and 4°C, phages in the supernatant were precipitated in 1 M NaCl and 10% w/v PEG 8000 at 4°C. The precipitated phages were pelleted by centrifugation for 20 min at 10,000 rcf and 4°C and resuspended in 1.6 ml SM. To eliminate residual DNase I activity, the phage suspension was incubated at 37°C 10 min with 40  $\mu$ l 20 mg/ml proteinase K. Following extraction of the phages with an equal volume of chloroform and the addition of EDTA to 100 mM,  $\frac{1}{2}$  volume of 6 M guanidine thiocyanate was added to disrupt the capsids and release the phage DNA. DNA was then purified using the GENE CLEAN Turbo Kit (Qbiogene, Irvine, CA). Phage DNA was quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA).

RFLP analysis was performed using 5  $\mu$ g of phage DNA digested overnight at 37°C with EcoRI (Invitrogen, Carlsbad, CA). For *cos* site screening, 5  $\mu$ g EcoRI digests were incubated 20 min at 80°C, cooled on ice, and separated on 0.8% agarose gels in 1x TAE (pH 8.0). Bands present only in the heated sample were excised from the gel, purified using a GENE CLEAN III kit (Qbiogene), cloned into pJET1.2 (Fermentas), and sequenced to identify the *cos* site. Preliminary sequencing of EcoRI phage DNA fragments cloned into pUC19 was performed as described previously [19,29]. For complete genome sequencing, phage DNA was submitted to 454 Life Sciences (Branford, CT) for pyrosequencing. The genome sequences of KL1 and AH2 have been deposited in GenBank with the accession numbers JF939047 and JN564907. Sequence start sites for these files were chosen based on alignment with PA73 for KL1 and at the *cos* site for AH2.

#### Bioinformatics analysis

Annotation of the genome sequences and determination of GC contents were performed using GeneMark ([http://exon.biology.gatech.edu/gmhmm2\\_prok.cgi](http://exon.biology.gatech.edu/gmhmm2_prok.cgi)) [70]. Manual annotations were performed for KL1 5 (encoding Rz1) and KL1 19/AH2 55 (encoding translationally-frameshifted tail proteins). Homology searches and conserved domain searches were performed using HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>) [71] and NCBI's BLASTN/BLASTP (for full genomes and individual proteins, respectively) (<http://blast.ncbi.nlm.nih.gov>) [72] and Conserved Domain Search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [73]. FSFinder was used for translational frameshift identification (<http://wilab.inha.ac.kr/fsfinder>) [74]. Mfold was used for stem-loop structure identification (<http://mfold.rna.albany.edu/?q=mfold>) [75]. Sequence comparisons were visualized using Circos (<http://circos.ca>) [76] and PROmer (<http://mummer.sourceforge.net>) [77] with the following

parameters: breaklen = 60, maxgap = 30, mincluster = 20, minmatch = 6. Lysis protein analysis was performed using TMHMM for transmembrane region identification (<http://www.cbs.dtu.dk/services/TMHMM>) [78] and LipoP for signal peptidase II cleavage site identification (<http://www.cbs.dtu.dk/services/LipoP>) [79].

## Additional files

**Additional file 1: Table S1.** KL1 HHpred predictions.

**Additional file 2: Table S2.** AH2 HHpred predictions.

**Additional file 3: Figure S3.** Stem-loop structures predicted by mfold analysis of the KL1 (left) and AH2 (right) frameshift regions (including the putative frameshift sites and 35 downstream bases).

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

KHL isolated KL1, performed electron microscopy, sequenced, annotated, and analyzed the genomes, and drafted the manuscript. PS constructed Figure 4 and performed FSFinder and preliminary HHpred analysis. JJD devised the study and assisted with experimental design, data analysis, and the writing of the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

The authors would like to thank Amberlie Heaman for isolation and preliminary host range analysis of AH2, Miles Peterson for assistance with figure construction, Arlene Oatway (University of Alberta Department of Biological Sciences Advanced Microscopy Facility) for assistance with electron microscopy, and Miguel Valvano (University of Western Ontario) for providing K56-2 LPS mutants.

JJD gratefully acknowledges funding from the Canadian Institutes of Health Research (CIHR grant 200705XNE-170954), and Cystic Fibrosis Canada (CFC grant RES0001467), for operating grant support. KHL gratefully acknowledges funding from Cystic Fibrosis Canada, NSERC, the Killam Trusts, and Alberta Innovates – Health Solutions for scholarship support.

## Author details

<sup>1</sup>6-008 Centennial Centre for Interdisciplinary Science, Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada. <sup>2</sup>1400 College Plaza, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2C8, Canada.

Received: 17 October 2011 Accepted: 15 May 2012

Published: 07 June 2012

## References

- Merrill CR, Scholl D, Adhya SL: The prospect for bacteriophage therapy in Western medicine. *Nat Rev Drug Discov* 2003, **2**:489–497.
- Daniel A, Euler C, Collin M, Chahales P, Gorelick KJ, Fischetti VA: Synergism between a novel chimeric lysin and oxacillin protects against infection by methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2010, **54**:1603–1612.
- Morello E, Saussereau E, Maura D, Huerre M, Touqui L, Debarbieux L: Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: First steps towards treatment and prevention. *PLoS One* 2011, **6**:e16963.
- Kumari S, Harjai K, Chhibber S: Bacteriophage versus antimicrobial agents for the treatment of murine burn wound infection caused by *Klebsiella pneumoniae* B5055. *J Med Microbiol* 2011, **60**:205–210.
- Oliveira A, Sereno R, Azeredo J: *In vivo* efficiency evaluation of a phage cocktail in controlling severe colibacillosis in confined conditions and experimental poultry houses. *Vet Microbiol* 2010, **146**:303–308.
- Waseh S, Hanifi-Moghaddam P, Coleman R, Masotti M, Ryan S, Foss M, MacKenzie R, Henry M, Szymanski CM, Tanha J: Orally administered P22 phage tailspike protein reduces *Salmonella* colonization in chickens: Prospects of a novel therapy against bacterial infections. *PLoS One* 2010, **5**:e13904.
- Carvalho CM, Gannon BW, Halfhide DE, Santos SB, Hayes CM, Roe JM, Azeredo J: The *in vivo* efficacy of two administration routes of a phage cocktail to reduce numbers of *Campylobacter coli* and *Campylobacter jejuni* in chickens. *BMC Microbiol* 2010, **10**:232.
- Wright A, Hawkins CH, Ånggård EE, Harper DR: A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; A preliminary report of efficacy. *Clin Otolaryngol* 2009, **34**:349–357.
- Golshahi L, Seed KD, Dennis JJ, Finlay WH: Toward modern inhalational bacteriophage therapy: Nebulization of bacteriophages of *Burkholderia cepacia* complex. *J Aerosol Med Pulm Drug Deliv* 2008, **21**:351–359.
- Golshahi L, Lynch KH, Dennis JJ, Finlay WH: *In vitro* lung delivery of bacteriophages KS4-M and ΦKZ using dry powder inhalers for treatment of *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* infections in cystic fibrosis. *J Appl Microbiol* 2011, **110**:106–117.
- Matinkhoo S, Lynch KH, Dennis JJ, Finlay WH, Vehring R: Spray-dried respirable powders containing bacteriophages for the treatment of pulmonary infections. *J Pharm Sci* 2011, **100**:5197–5205.
- Henn MR, Sullivan MB, Stange-Thomann N, Osburne MS, Berlin AM, Kelly L, Yandava C, Kodira C, Zeng Q, Weiland M, Sparrow T, Saif S, Giannoukos G, Young SK, Nusbaum C, Birren BW, Chisholm SW: Analysis of high-throughput sequencing and annotation strategies for phage genomes. *PLoS One* 2010, **5**:e9083.
- Isles A, Macluskay I, Corey M: *Pseudomonas cepacia* infection in cystic fibrosis: An emerging problem. *J Pediatr* 1984, **104**:206–210.
- Ryley HC, Doull JM: *Burkholderia cepacia* complex infection in patients with cystic fibrosis: Laboratory investigations, epidemiology and clinical management. *Rev Med Microbiol* 2003, **14**:15–24.
- LiPuma JJ, Dasen SE, Nielson DW, Stern RC, Stull TL: Person-to-person transmission of *Pseudomonas cepacia* between patients with cystic fibrosis. *Lancet* 1990, **336**:1094–1096.
- Zhou J, Chen Y, Tabibi S, Alba L, Garber E, Saiman L: Antimicrobial susceptibility and synergy studies of *Burkholderia cepacia* complex isolated from patients with cystic fibrosis. *Antimicrob Agents Chemother* 2007, **51**:1085–1088.
- Baldwin A, Mahenthalingam E, Drevinek P, Pope C, Waine DJ, Henry DA, Speert DP, Carter P, Vandamme P, LiPuma JJ, Dowson CG: Elucidating global epidemiology of *Burkholderia multivorans* in cases of cystic fibrosis by multilocus sequence typing. *J Clin Microbiol* 2008, **46**:290–295.
- Seed KD, Dennis JJ: Experimental bacteriophage therapy increases survival of *Galleria mellonella* larvae infected with clinically relevant strains of the *Burkholderia cepacia* complex. *Antimicrob Agents Chemother* 2009, **53**:2205–2208.
- Lynch KH, Seed KD, Stothard P, Dennis JJ: Inactivation of *Burkholderia cepacia* complex phage KS9 gp41 identifies the phage repressor and generates lytic virions. *J Virol* 2010, **84**:1276–1288.
- Carmody LA, Gill JJ, Summer EJ, Sajjan US, Gonzalez CF, Young RF, LiPuma JJ: Efficacy of bacteriophage therapy in a model of *Burkholderia cenocepacia* pulmonary infection. *J Infect Dis* 2010, **201**:264–271.
- Semler DD, Lynch KH, Dennis JJ: The promise of bacteriophage therapy for *Burkholderia cepacia* complex respiratory infections. *Front Cell Inf Microbiol* 2012, **1**:27.
- Lynch KH, Dennis JJ: Cangene gold medal award lecture - genomic analysis and modification of *Burkholderia cepacia* complex bacteriophages. *Can J Microbiol* 2012, **58**:221–235.
- Imamovic L, Ballesté E, Jofre J, Muniesa M: Quantification of Shiga toxin-converting bacteriophages in wastewater and in fecal samples by real-time quantitative PCR. *Appl Environ Microbiol* 2010, **76**:5693–5701.
- Langley R, Kenna DT, Vandamme P, Ure R, Govan JRW: Lysogeny and bacteriophage host range within the *Burkholderia cepacia* complex. *J Med Microbiol* 2003, **52**:483–490.
- Summer EJ, Gonzalez CF, Bomer M, Carlile T, Embry A, Kucherka AM, Lee J, Mebane L, Morrison WC, Mark L, King MD, LiPuma JJ, Vidaver AK, Young R: Divergence and mosaicism among virulent soil phages of the *Burkholderia cepacia* complex. *J Bacteriol* 2006, **188**:255–268.
- Seed KD, Dennis JJ: Isolation and characterization of bacteriophages of the *Burkholderia cepacia* complex. *FEMS Microbiol Lett* 2005, **251**:273–280.
- Gill JJ, Summer EJ, Russell WK, Cologna SM, Carlile TM, Fuller AC, Kitsopoulos K, Mebane LM, Parkinson BN, Sullivan D, Carmody LA, Gonzalez

- CF, LiPuma JJ, Young R: Genomes and characterization of phages Bcep22 and BcepL02, founders of a novel phage type in *Burkholderia cenocepacia*. *J Bacteriol* 2011, **193**:5300–5313.
28. Goudie AD, Lynch KH, Seed KD, Stothard P, Shrivastava S, Wishart DS, Dennis JJ: Genomic sequence and activity of KS10, a transposable phage of the *Burkholderia cepacia* complex. *BMC Genomics* 2008, **9**:615.
29. Lynch KH, Stothard P, Dennis JJ: Genomic analysis and relatedness of P2-like phages of the *Burkholderia cepacia* complex. *BMC Genomics* 2010, **11**:599.
30. Lynch KH, Stothard P, Dennis JJ: Characterization of DC1, a broad-host-range Bcep22-like podovirus. *Appl Environ Microbiol* 2012, **78**:889–891.
31. Loutet SA, Flannagan RS, Kooi C, Sokol PA, Valvano MA: A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival in vivo. *J Bacteriol* 2006, **188**:2073–2080.
32. Ortega X, Sillipo A, Saldías MS, Bates CC, Molinaro A, Valvano MA: Biosynthesis and structure of the *Burkholderia cenocepacia* K56-2 lipopolysaccharide core oligosaccharide: Truncation of the core oligosaccharide leads to increased binding and sensitivity to polymyxin B. *J Biol Chem* 2009, **284**:21738–21751.
33. Kwan T, Liu J, DuBow M, Gros P, Pelletier J: Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J Bacteriol* 2006, **188**:1184–1187.
34. Zhang J, Inouye M: MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with Era, an essential GTPase in *Escherichia coli*. *J Bacteriol* 2002, **184**:5323–5329.
35. Xu J, Hendrix RW, Duda RL: Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell* 2004, **16**:11–21.
36. Alam SL, Atkins JF, Gesteland RF: Programmed ribosomal frameshifting: Much ado about knotting! *Proc Natl Acad Sci USA* 1999, **96**:14177–14179.
37. Young R, Wang I-N, Roof WD: Phages will out: Strategies of host cell lysis. *Trends Microbiol* 2000, **8**:120–128.
38. Summer EJ, Berry J, Tran TAT, Niu L, Struck DK, Young R: Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. *J Mol Biol* 2007, **373**:1098–1112.
39. Summer EJ, Gonzalez CF, Carlisle T, Mebane LM, Cass AM, Savva CG, LiPuma J, Young R: *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. *J Mol Biol* 2004, **340**:49–65.
40. Lopes A, Amarir-Bouhram J, Faure G, Petit M-A, Guerois R: Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res* 2010, **38**:3952–3962.
41. Hennecke F, Kolmar H, Brundl K, Fritz H-J: The *vsr* gene product of *E. coli* K-12 is a strand- and sequence-specific DNA mismatch endonuclease. *Nature* 1991, **353**:776–778.
42. Lin J-J, Sancar A: The C-terminal half of UvrC protein is sufficient to reconstitute (A)BC excinuclease. *Proc Natl Acad Sci USA* 1991, **88**:6824–6828.
43. Gross M, Marianovsky I, Glaser G: MazG - A regulator of programmed cell death in *Escherichia coli*. *Mol Microbiol* 2006, **59**:590–601.
44. Magnusson LU, Farewell A, Nyström T: ppGpp: A global regulator in *Escherichia coli*. *Trends Microbiol* 2005, **13**:236–242.
45. Bryan MJ, Burroughs NJ, Spence EM, Clokie MRJ, Mann NH, Bryan SJ: Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. *PLoS One* 2008, **3**:e2048.
46. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osborne MS, Henn MR, Chisholm SW: Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 2010, **12**:3035–3056.
47. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW: Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* 2005, **3**:e144.
48. Mann NH, Clokie MRJ, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM: The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol* 2005, **187**:3188–3200.
49. Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, King J, Hatfull GF, Lawrence JG, Hendrix RW: Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* 2007, **9**:1675–1695.
50. Angly F, Youle M, Nosrat B, Srinagesh S, Rodriguez-Brito B, McNairnie P, Deyanat-Yazdi G, Breitbart M, Rohwer F: Genomic analysis of multiple Roseophage SIO1 strains. *Environ Microbiol* 2009, **11**:2863–2873.
51. Duhaime MB, Wichels A, Waldmann J, Teeling H, Glöckner FO: Ecogenomics and genome landscapes of marine *Pseudoalteromonas* phage H105/1. *ISME J* 2011, **5**:107–121.
52. Clokie MRJ, Mann NH: Marine cyanophages and light. *Environ Microbiol* 2006, **8**:2074–2082.
53. Hammer BK, Swanson MS: Co-ordination of *Legionella pneumophila* virulence with entry into stationary phase by ppGpp. *Mol Microbiol* 1999, **33**:721–731.
54. Taylor CM, Beresford M, Epton HAS, Sigeo DC, Shama G, Andrew PW, Roberts IS: *Listeria monocytogenes relA* and *hpt* mutants are impaired in surface-attached growth and virulence. *J Bacteriol* 2002, **184**:621–628.
55. Erickson DL, Lines JL, Pesci EC, Venturi V, Storey DG: *Pseudomonas aeruginosa relA* contributes to virulence in *Drosophila melanogaster*. *Infect Immun* 2004, **72**:5638–5645.
56. Jeong J-H, Song M, Park S-I, Cho K-O, Joon HR, Choy HE: *Salmonella enterica* serovar Gallinarum requires ppGpp for internalization and survival in animal cells. *J Bacteriol* 2008, **190**:6340–6350.
57. Klinkenberg LG, Lee J-H, Bishai WR, Karakousis PC: The stringent response is required for full virulence of *Mycobacterium tuberculosis* in guinea pigs. *J Infect Dis* 2010, **202**:1397–1404.
58. Haralalka S, Nandi S, Bhadra RK: Mutation in the *relA* gene of *Vibrio cholerae* affects in vitro and in vivo expression of virulence factors. *J Bacteriol* 2003, **185**:4672–4682.
59. Silva AJ, Benitez JA: A *Vibrio cholerae* relaxed (*relA*) mutant expresses major virulence factors, exhibits biofilm formation and motility, and colonizes the suckling mouse intestine. *J Bacteriol* 2006, **188**:794–800.
60. Nguyen D, Joshi-Datar A, Lepine F, Bauerle E, Olakanmi O, Beer K, McKay G, Siehnel R, Schafhauser J, Wang Y, Britigan BE, Singh PK: Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science* 2011, **334**:982–986.
61. Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ: Exceptional convergent evolution in a virus. *Genetics* 1997, **147**:1497–1507.
62. Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ: Different trajectories of parallel evolution during viral adaptation. *Science* 1999, **285**:422–424.
63. Wichman HA, Scott LA, Yarber CD, Bull JJ: Experimental evolution recapitulates natural evolution. *Phil Trans R Soc Lond* 2000, **355**:1677–1684.
64. Bollback JP, Huelsenbeck JP: Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* 2009, **181**:225–234.
65. Ackermann H-W: 5500 Phages examined in the electron microscope. *Arch Virol* 2007, **152**:227–243.
66. Abedon ST: Phage evolution and ecology. *Adv Appl Microbiol* 2009, **67**:1–45.
67. Ceysens P-J, Glonti T, Kropinski AM, Lavigne R, Chanishvili N, Kulakov L, Lashkhi N, Tediashvili M, Merabishvili M: Phenotypic and genotypic variations within a single bacteriophage species. *Virol J* 2011, **8**:134.
68. Mahenthiralingam E, Coenye T, Chung JW, Speert DP, Govan JRW, Taylor P, Vandamme P: Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia* complex. *J Clin Microbiol* 2000, **38**:910–913.
69. Coenye T, Vandamme P, LiPuma JJ, Govan JRW, Mahenthiralingam E: Updated version of the *Burkholderia cepacia* complex experimental strain panel. *J Clin Microbiol* 2003, **41**:2797–2798.
70. Lukashin AV, Borodovsky M: GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res* 1998, **26**:1107–1115.
71. Söding J, Biegert A, Lupas AN: The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005, **33**:W244–W248.
72. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**:3389–3402.
73. Marchler-Bauer A, Bryant SH: CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res* 2004, **32**:W327–W331.
74. Moon S, Byun Y, Kim H-J, Jeong S, Han K: Predicting genes expressed via -1 and +1 frameshifts. *Nucleic Acids Res* 2004, **32**:4884–4892.

75. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**:3406–3415.
76. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: An information aesthetic for comparative genomics.** *Genome Res* 2009, **19**:1639–1645.
77. Delcher AL, Phillippy A, Carlton J, Salzberg SL: **Fast algorithms for large-scale genome alignment and comparison.** *Nucleic Acids Res* 2002, **30**:2478–2483.
78. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL: **Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes.** *J Mol Biol* 2001, **305**:567–580.
79. Juncker AS, Willenbrock H, von Heijne G, Brunak S, Nielsen H, Krogh A: **Prediction of lipoprotein signal peptides in Gram-negative bacteria.** *Protein Sci* 2003, **12**:1652–1662.

doi:10.1186/1471-2164-13-223

**Cite this article as:** Lynch *et al.*: Comparative analysis of two phenotypically-similar but genomically-distinct *Burkholderia cenocepacia*-specific bacteriophages. *BMC Genomics* 2012 **13**:223.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

