A novel inducible prophage from *Burkholderia vietnamiensis* G4 is widely distributed across the species and has lytic activity against pathogenic *Burkholderia*

Rebecca Weiser 1*(https://orcid.org/0000-0003-3983-3272), Zhong Ling Yap1,2 (https://orcid.org/0000-0002-0288-7085), Ashley Otter1,3 (https://orcid.org/0000-0002-8317-9194), Brian V. Jones4 (https://orcid.org/0000-0002-1211-9186), Jonathan Salvage5 (https://orcid.org/0000-0002-0627-6249), Julian Parkhill6 (https://orcid.org/0000-0001-7069-5958) and Eshwar Mahenthiralingam1*

1Cardiff School of Biosciences, Cardiff University, Cardiff, UK
2Department of Microbiology, University of Manitoba, Winnipeg, Canada
3Public Health England, National Infection Service, Porton Down, UK
4Department of Biology and Biochemistry, University of Bath, Bath, UK
5School of Pharmacy and Biomolecular Sciences, University of Brighton, Brighton, UK
6Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

*Correspondence: Weiserr@cardiff.ac.uk (R.W.); MahenthiralingamE@cardiff.ac.uk (E.M.)

Received: date; Accepted: date; Published: date

**Abstract:** *Burkholderia* species have environmental, industrial and medical significance, and are important opportunistic pathogens in individuals with cystic fibrosis (CF). Using a combination of existing and newly determined genome sequences, this study investigated prophage carriage across the species *B. vietnamiensis*, and also isolated spontaneously inducible prophages from a reference strain, G4. Eighty-one *B. vietnamiensis* genomes were bioinformatically screened for prophages using PHASTER and prophage regions found to comprise up to 3.4% of total genetic material. Overall, 115 intact prophages were identified and there was evidence of polylysogeny in 32 strains. A novel, inducible Mu-like phage (vB_BvM-G4P1) was isolated from *B. vietnamiensis* G4 that had lytic activity against strains of five *Burkholderia* species prevalent in CF infections, including the Boston epidemic *B. dolosa* strain SLC6. The cognate prophage to vB_BvM-G4P1 was identified in the lysogen genome and was almost identical (>93.5% tblastx identity) to prophages found in 13 other *B. vietnamiensis* strains (17% of the strain collection). Phylogenomic analysis determined that the G4P1-like prophages were widely distributed across the population structure of *B. vietnamiensis*. This study highlights that genomic characterization of *Burkholderia* prophages can lead to the discovery of novel bacteriophages with potential therapeutic or biotechnological applications.

**Keywords:** *Burkholderia vietnamiensis*; Prophages; Phylogenomics; Induction; Lysogeny; Phage classification
1. Introduction

Bacteriophages are estimated to be the most abundant entities in the biosphere, and as the natural predators of bacteria, have important roles in bacteria ecology and evolution [1]. Lytic phages are capable only of infections ending in bacterial cell death and release of new phage progeny (lytic life cycle), whilst temperate phages can either behave as lytic phages, or integrate into the genome of a bacterial host as a prophage (lysogenic life cycle) [1]. Prophage carriage can confer a number of benefits to the lysogen including acquisition of novel traits through horizontal gene transfer, increased genetic variation and evolutionary innovation [2]. Resident prophages also present a risk for bacteria, as induction of the lytic cycle can occur spontaneously or in response to environmental cues (e.g. stress) and leads to cell lysis [2]. Prophages are abundant in many bacterial species and can represent up to 10-20% of the host’s genome [3]. Prophage carriage is very variable between bacteria, however, some bacteria have none, whilst others are polylysogenic and can carry over a dozen prophages [4]. Large-scale bacterial genome sequencing projects in parallel with the development of prophage finding software [5-7] have enhanced our ability to detect prophages and understand their distribution in a range of bacteria.

Burkholderia are a very diverse group of Gram-negative bacteria found in wide-ranging environments including soil, water and in association with plants, fungi, animals and humans [8]. The genus has undergone extensive taxonomic revision since its definition in 1992 [9], and has recently been split into six genera, Burkholderia sensu stricto, Paraburkholderia, Caballeronia, Trinickia, Mycetohabitas and Robbsia [10]. There are currently over 30 Burkholderia species, 22 of which belong to the Burkholderia cepacia complex (Bcc) [10]. Members of the Bcc are opportunistic pathogens linked with life-threatening lung infections in individuals with cystic fibrosis (CF) [11] and are problematic contaminants of industrial products [12]. Conversely, Bcc bacteria have numerous beneficial traits including antimicrobial production, plant growth promotion and bioremediation [8]. The large (6-9 Mb), multi-replicon genomes of Burkholderia have a huge coding capacity that underpins their genetic and metabolic versatility, and ability to thrive in different environments [13]. Whilst Bcc species are closely related by taxonomic analysis based on a single marker gene (>97.7% 16S rDNA sequence similarity) [14], recent comparative genomic studies have identified huge genomic diversity within the complex; there are approximately only 1000 genes shared between strains of the 22 Bcc species (the ‘core genome’), highlighting that a large proportion of the genome is highly variable [15].

Burkholderia genomes have also been shown to comprise up to 10% recently horizontally-acquired DNA [16] and potentially represent a rich source of prophages. Burkholderia phage literature is relatively limited, however, and although phage therapy for the treatment of Bcc infections is being explored [17], there have been few systematic studies looking at prophage carriage within Burkholderia genomes.

B. vietnamiensis is part of the Bcc and known to cause infections in individuals in CF. It is encountered less frequently than the two major Bcc CF pathogens, B. cenocepacia and B. multivorans [18], and the prevalence of B. vietnamiensis within overall Burkholderia CF lung infections is generally under 10% [11,18-22]. B. vietnamiensis has been associated with non-CF clinical infections [18,21], but also has strong links to the environment, being isolated from soil and plant roots, having nitrogen fixation abilities [23,24], and the capacity to degrade a range of xenobiotic compounds [25]. There are currently 48 B. vietnamiensis genomes available in Genbank (Accessed January 2020; https://www.ncbi.nlm.nih.govgenome/genomes/1136?) averaging 6.86 Mb in size (range: 5.73-8.39 Mb). The largest genome belongs to B. vietnamiensis strain G4 (ATCC 53716) whose 8.39 Mb genome is organised into 3 chromosomes and 5 plasmids. This strain was originally isolated from wastewater in 1987 and is well known for its ability to degrade trichloroethane when grown on toluene or phenol [26,27]. B. vietnamiensis genomes have not yet been investigated for prophage carriage and only one lysogenic phage has been previously isolated from B. vietnamiensis ATCC 29424 [28].

We have taken advantage of the increasing numbers of available B. vietnamiensis genomes, supplemented these with 35 sequenced for this study, and used them to characterise prophage regions within the species. We report that prophage material ranged from 0.1% to 3.4% of B. vietnamiensis genomes and polylysogeny was observed. In addition, a novel inducible and functional
prophage (G4P1) was discovered in *B. vietnamiensis* G4 that had lytic activity against strains of five Bcc species. Phage G4P1 was widely distributed across the population structure of the species and found in 17% of the 81 *B. vietnamiensis* strains examined. Our investigations linked genomic and phylogenomic approaches to characterize prophage carriage across the species, and facilitate the discovery of novel phages with potential therapeutic or biotechnological applications.

2. Materials and Methods

2.1 Bacterial genomic analysis and phylogenomics

A total of 35 *B. vietnamiensis* strains were genome sequenced using the Illumina HiSeq 2000 and HiSeq X Ten platforms to generate 125-nucleotide and 150-nucleotide paired end reads, respectively, as described by Mullins *et al.* (2019) [29]. The raw read data have been submitted to the European Nucleotide Archive (ENA; sequencing project PRJEB9765). Individual accession numbers are given in Table S1. Additional *B. vietnamiensis* genomes were downloaded directly as draft assemblies (FASTA sequences) from Genbank (n=46) (Table S1). There were two entries for strain FL_5_2_10_S1_D0 in Genbank which were designated FL_5_2_10_S1_D0 (GCA_001524025.1) and FL_5_2_10_S1_D0_Repeat (GCA_001524045.1) throughout.

All bioinformatic analyses were performed using the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) computing resource [30]. Raw reads were trimmed using Trim Galore v0.4.3 for paired end reads [31] and quality assessed with FastQC v0.11.5 [32]. Genome assembly was achieved using Unicycler v0.4.7 [33] with SPAdes v3.11.0 [34] and the option for short-read assembly. Assembly quality was visualised with Bandage v0.8.1 assembly graphs [35]. All genome assemblies were quality checked using the Quality Assessment Tool for Genome Assemblies (QUAST) v.4.6.3 to confirm expected genome size and key assembly statistics (Table S1). *B. vietnamiensis* species identity was confirmed by calculating the shared average nucleotide identify (ANI) between genomes using FastANI v1.1 [36] to ensure that the 95% species threshold was met. Prokka v1.12 [37] was used for gene annotation. Roary v3.6.0 [38] and double-precision FastTree v2.1.8 [39] were used for comparative core genome analysis of *B. vietnamiensis* strains and generation of phylogenomic trees as described previously [40]. The phylogenomic tree root position was determined using the outgroup *B. ambifaria* AMMD (GCF_000203915.1; Figure S1). Manual inspection of *Burkholderia* genomes was performed using Artemis v17.0.1 [41] with the FNA and GFF file outputs from Prokka.

2.2 Prophage identification and comparison

Prophage regions were identified in *B. vietnamiensis* genomes using the PHASTER (PHAge Search Tool Enhanced Release) online tool [5] [performed January 2019] which scores putative phage regions as either intact, questionable or incomplete. R statistical software [42] was used to visualise trends and perform linear regression analysis (lm function) to identify significant correlations (p ≤ 0.05 level) between the amount of prophage material carried and bacterial genome size. Comparative ANI analysis of prophages sequences identified as ‘intact’ was performed using pyani v0.2.7 [43] with the ANIb option. The command-line version of NCBI BLAST v2.7.1 [44] was used to create and search (blastn, tblastx) local BLAST databases of intact prophage sequences with a prophage sequence query. Prophage genome alignments were performed using ProgressiveMauve and default settings (Mauve; v20150226) [45]. The GBK files for the visualisation of genome comparisons in ProgressiveMauve were obtained by annotation of intact prophage sequences using Prokka with the Millardlab Caudovirales custom annotation database (http://millardlab.org/bioinformatics/lab_server/phage-genome-annotation/).

Transposase proteins from *B. vietnamiensis* prophages (this study), related sequences identified by blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (results restricted within Myoviridae, taxid 10662), and other Mu-like *Burkholderia* phages (BcepMu, NC_005882.1, 8E255, NC_009237.1), were compared by alignment using ClustalW [46] and Maximum Likelihood phylogenetic trees generated using MEGA7 (Jones-Taylor-Thornton model, with Neighbor-Join and BioNJ methods used to create initial trees for the heuristic search) [47,48].
2.3 Bacterial strains and culture media

Burkholderia strains were obtained from the BCC culture collection held at Cardiff University [49] (Table 1). Strains were routinely grown on Tryptone Soya Agar (TSA; Oxoid Ltd., Cambridge, UK) and overnight cultures prepared by inoculating 3 ml of Tryptone Soya Broth (TSB; Oxoid Ltd.) with fresh (<72 hours) growth material and incubated on an orbital shaker (150 rpm). All cultures were incubated at 30°C for 16-18 hours. Bacterial strains were stored frozen in TSB containing 8% (v/v) dimethylsulphoxide (DMSO; Sigma-Aldrich Company Ltd., Dorset, UK) at -80°C.

2.4 Isolation of spontaneously induced bacteriophages from lysogenic B. vietnamiensis G4

An overnight culture of B. vietnamiensis G4 prepared in TSB + 10 mM MgCl₂ was centrifuged at 2504 g for 10 minutes and the supernatant filter sterilised (pore size 0.2 µm; Sartorius Stedim Biotech, Cambridge, UK). Lytic phage activity was screened for against a panel of 24 strains representing 9 different Bcc species, 2 strains of non-Bcc Burkholderia species and 3 Paraburkholderia strains (Table 1) using the double agar overlay technique [50] and a ‘drop test’; a soft agar top layer (TSB + 0.3% [w/v] purified agar [Oxoid Ltd.] + 10 mM MgCl₂) seeded with approximately 10⁸ viable host cells was overlaid on TSA and 10 µl culture supernatant dropped onto the surface. After overnight incubation at 30°C for 16-18 hours, lytic phage activity was identified as a zone of clearing in the overlay layer. Zones were classified as strong activity (complete clearing), weak activity (partial clearing) or no activity. Positive supernatants were re-tested.

Plaque assays with appropriate host strains [51] were used to identify individual lytic bacteriophages in culture supernatants. The host strains used to produce the phage stocks were B. ambifaria BCC1212 (phage G4P1) and B. cenocepacia BCC1210 (phages G4P2 and G4P3). Low titer individual phage stocks were prepared by transferring a single plaque into 5 ml phage buffer (Phosphate Buffered Saline [PBS; Oxoid Ltd.] + 10 mM MgCl₂). The phage suspension was then vortex mixed for 1 minute, centrifuged at 2504 g for 10 minutes and filter sterilised (pore size 0.2 µm). To obtain high titer phage stocks (approximately 10⁸-10⁹ pfu/ml) plaque assays were performed with low titer phage stocks and agar overlays displaying semi-confluent lysis (≥ 4 agar plates) were transferred into 10 ml phage buffer, vortex mixed for 1 minute, centrifuged at 2054 g for 10 minutes then filter sterilised (pore size 0.2 µl). Storage of culture supernatants and phage stocks was at 4°C (short term) or at -80°C with 50% glycerol (Fisher Scientific UK Ltd., Loughborough, UK) (long term).

2.5 Bacteriophage host range determination of phage G4P1 isolated from B. vietnamiensis G4

A double layer agar overlay of each host strain (Table 1) was prepared and 10 µl of high titer G4P1 phage stock dropped onto the overlay surface. Following overnight incubation at 30°C for 16-18 hours, lytic phage activity was observed as a zone of clearing in the top agar layer. Positive results were confirmed by repeating the assay. Plaque assays were also performed with the G4P1 high titer stock and four of the host strains (B. ambifaria BCC1212, B. cenocepacia BCC0019, B. dolosa BCC1359 and B. vietnamiensis BCC0027) to identify individual phage plaques and confirm that the zone of clearing in the drop test was of phage origin rather than another inhibitory factor.

2.6 Transmission electron microscopy (TEM)

The morphology of phage virions was investigated by TEM as previously described [52]. Phage particles from high titer stocks were deposited onto 200-mesh Formvar/carbon copper grids (Agar Scientific, Essex, UK), and negatively stained with 2% phosphotungstic acid (pH 7.4) (Sigma-Aldrich Company Ltd.). Virions were imaged using a field emission gun-scanning transmission electron microscope (Zeiss Sigma FEG-STEM) at 20 kV accelerating voltage, a 20 µm aperture and a 2.7 mm working distance.

2.7 Bacteriophage genomic analysis

Bacteriophage DNA was extracted from high titer phage stocks using the Phage DNA isolation kit (Norgen Biotek Corp., Ontario, Canada) according to the manufacturer’s instructions and with
the recommended DNase I treatment step. A 16S rRNA gene PCR [53] was performed to confirm that phage DNA stocks were free from contamination with host bacterial DNA. DNA quantification was achieved using a Qubit fluorometer and the Qubit dsDNA BR assay kit (Invitrogen, Massachusetts, USA).

DNA sequencing library preparation and sequencing was performed by the Cardiff University Genomics Research Hub. Libraries were prepared with the Nextera XT kit (Illumina, California, USA) and sequencing was carried out on an Illumina NextSeq500 using a NextSeq 500/550 Mid Output v2 kit (2 x 150 cycles) to give on average 122 bp paired-end reads. Read trimming, genome assembly and quality control was performed as described for bacterial genomes to produce a single contig (36725 bp). Coverage was determined by mapping the sequence reads onto the assembly using Samtools v1.7 [54] and Bedtools v2.25.0 [55] and was on average 442X (Table S2; Figure S2). Gene annotation was achieved using a combination of Prokka with the Millardlab Caudovirales custom annotation database (as described for prophage annotation), PHASTER and the InterProScan tool of the online InterPro resource for further functional analysis of predicted protein sequences (https://www.ebi.ac.uk/interpro/search/sequence/) [56,57]. Isolated and genome sequenced bacteriophages were named according to recent recommendations [58].

To determine the genomic location of isolated phages as prophages within the lysogen genome, the command-line version of NCBI BLAST v2.7.1 [44] was used to create and search (blastn) a local BLAST database of the lysogen genome with a phage DNA sequence query. Easyfig v2.2.3 [59] was used to visualise blastn comparisons between phage and lysogen sequences, and the Burkholderia genome database [60] used for lysogen gene annotations.

2.8 Data summary

All de novo determined genome sequence data for B. vietnamiensis is available at the European Nucleotide Archive under sequencing project PRJEB9765.

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

3. Results

3.1 Prophage carriage was common in B. vietnamiensis

To gain an understanding of prophage carriage in B. vietnamiensis, PHASTER was used to examine the genomes of the strain collection. All 81 B. vietnamiensis strains had evidence of prophage material and 59 of these had at least one intact prophage (Table S1). Overall, 115 intact, 23 questionable and 81 incomplete prophages were discovered and the majority of strains carried one or two prophage regions of each type (Figure 1A). Intact prophage regions (average = 35.2 kb, range = 12.3-92.0 kb) were larger than questionable prophage regions (average =27.4 kb, range = 4.9-55.4 kb), which in turn were larger than incomplete regions (average = 17.8 kb, range = 5.6-48.6 kb) (Figure 1B). As genome size increased, the total prophage material carried also increased (Figure 1C; p<0.001), and on average strain genomes comprised 1.1% (range: 0.1-3.4%) prophage material (Table S1).

Polylysogeny with intact prophage regions was observed in 32 strains; 18 strains had two intact prophages, 10 strains had three, 3 strains had four, one strain (BCC1193) had five and one strain (BCC1172) had six (Table S1).
Figure 1. Characteristics of prophage regions identified in 81 *B. vietnamiensis* strains by PHASTER. (a) Distribution of intact (n=115), questionable (n=23) and incomplete (n=81) prophage regions in bacterial genomes. (b) Size ranges of prophage regions; boxplots display the mean (x), median, upper and lower quartile, maximum and minimum value, and any outliers. (c) Correlation between bacterial genome size and total prophage region content (significant positive correlation; \( p<0.001 \)).

3.2 Prophages were spontaneously induced from *B. vietnamiensis* strain G4

*B. vietnamiensis* G4 was historically isolated as a bioremediation strain and examined as a model strain in biotechnological studies [25]. Although it was one of the first members of the Bcc to be fully genome sequenced the prophage content of this strain has not been characterised. PHASTER predicted that the 8.39 Mb *B. vietnamiensis* G4 genome held two intact, two questionable and six incomplete prophage regions (Table S1) which totalled 276 kb, representing 3.29% of the G4 genome.

To determine if any of these regions were spontaneously inducible and able to form plaques, the supernatant of an overnight culture of *B. vietnamiensis* G4 was tested against a panel of 26 *Burkholderia* and 3 *Paraburkholderia* strains (Table 1). Lytic activity was observed against strains of 5 different Bcc species (*B. ambifaria*, *B. cenocepacia*, *B. contaminans*, *B. dolosa* and *B. vietnamiensis*) (Table 1) in a drop test assay.

To isolate individual phages, plaque assays were performed with the *B. vietnamiensis* G4 culture supernatant and the host strain *B. ambifaria* BCC1212. This resulted in clear plaques 0.5-1 mm in diameter (Figure 2a-b) and an individual plaque was isolated, the phage purified and designated G4P1. The host range of G4P1 was identical to that of the *B. vietnamiensis* G4 supernatant with the exception that it was not active against *B. cenocepacia* BCC1210 and had ‘extra’ activity against *B. vietnamiensis* BCC1304. This suggested the presence of more than one inducible prophage with a different host range to G4P1, and that individual phages may behave differently when purified from culture supernatants. Additional plaque assays with the *B. vietnamiensis* G4 supernatant and *B. cenocepacia* BCC1210 as the host isolated at least one other phage; two different plaque morphologies were consistently observed and designated G4P2 (clear plaques 0.5-1 mm in diameter; Figure 2c-d) and G4P3 (uniform but slightly cloudy plaques 0.5-1 mm in diameter; Figure 2e-f). Only G4P1 stored stably at 4°C without a loss in titer and was further characterised. TEM revealed that G4P1 was a tailed phage with an icosahedral head belonging to the order *Caudovirales* and was approximately 200 nm in length (Figure 2g). Extended host range testing with G4P1 determined that the phage had activity against 5 other strains of *B. dolosa* including the Boston epidemic strain SLC6 (Table 1).

<p>| Table 1. Host range of <em>B. vietnamiensis</em> G4 culture supernatant and purified bacteriophage G4P1 |
| --- | --- | --- | --- | --- | --- |
| Host Species | BCC# | Alternative strain name | Origin (code) | G4 Culture supernatant | G4P1 |
| <em>Burkholderia cepacia complex</em> | | | | | |</p>
<table>
<thead>
<tr>
<th><strong>B. ambifaria</strong></th>
<th>BCC0191</th>
<th>HI 2345; J82</th>
<th>Soil (ENV)</th>
<th>+/-</th>
<th>+/-</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC0192</td>
<td>Ral-3; R-8863</td>
<td>Rhizosphere (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC0197</td>
<td>ATCC 51671</td>
<td>Leaves (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC0203</td>
<td>BCF/HG1-A; LMG-P 24640</td>
<td>Environmental (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC0338</td>
<td>ATCC 53266; LMG 17828</td>
<td>Roots (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC0410</td>
<td>MVP/C1 64</td>
<td>Maize (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC1212</td>
<td>MC40-6</td>
<td>Rhizosphere (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>B. cenocepacia</strong></th>
<th>BCC0019</th>
<th>LMG 18829; PC184/NEH4</th>
<th>Cystic fibrosis (CF)</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC1202</td>
<td>AU1054</td>
<td>Cystic fibrosis (CF)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BCC1210</td>
<td>MC0-3</td>
<td>Rhizosphere (ENV)</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>B. cepacia</strong></th>
<th>BCC0001</th>
<th>ATCC 25416; LMG1222-T</th>
<th>Onion (ENV)</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC0362</td>
<td>R-9929; CEP0964</td>
<td>Cystic fibrosis (CF)</td>
<td>+/-</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>B. dolesa</strong></th>
<th>BCC1343</th>
<th>AU0794</th>
<th>Cystic fibrosis (CF)</th>
<th>Not tested</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC1356</td>
<td>AU3556; clinical isolate of the SLC6 epidemic strain</td>
<td>Cystic fibrosis (CF)</td>
<td>Not tested</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC1357</td>
<td>AU1568; clinical isolate of the SLC6 epidemic strain</td>
<td>Cystic fibrosis (CF)</td>
<td>Not tested</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC1359</td>
<td>AU3960</td>
<td>Cystic fibrosis (CF)</td>
<td>Not tested</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC1360</td>
<td>AU4298; clinical isolate of the SLC6 epidemic strain</td>
<td>Cystic fibrosis (CF)</td>
<td>Not tested</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC1361</td>
<td>AU2130; clinical isolate of the SLC6 epidemic strain</td>
<td>Cystic fibrosis (CF)</td>
<td>Not tested</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

| **B. lata** | BCC0803 | ATCC 17660; LMG 22485T; R-18194; 383 | Soil (ENV) | - | - |

<table>
<thead>
<tr>
<th><strong>B. multivorans</strong></th>
<th>BCC0005</th>
<th>MA; LMG 18822; C5393</th>
<th>Cystic fibrosis (CF)</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC0011</td>
<td>C1576</td>
<td>Cystic fibrosis (CF)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BCC1421</td>
<td>ATCC 17616; LMG17588</td>
<td>Soil (ENV)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

| **B. pyrocinia** | BCC0180 | LMG 14191-T | Soil (ENV) | - | - |

<table>
<thead>
<tr>
<th><strong>B. vietnamiensis</strong></th>
<th>BCC0027</th>
<th>LMG 18835; PC259; JCM-APRIL93; CEP0040</th>
<th>Cystic fibrosis (CF)</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC0030</td>
<td>LMG 10929; FC0369</td>
<td>Ricercot (ENV)</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BCC0324</td>
<td>J1742</td>
<td>Non-CF (CLIN)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BCC1162</td>
<td>CEP1224</td>
<td>Cystic fibrosis (CF)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BCC1304</td>
<td>-</td>
<td>Industry (IND)</td>
<td>-</td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>BCC1309</td>
<td>-</td>
<td>Industry (IND)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Non-Burkholderia cepacia complex**

| **B. gladioli** | BCC0238 | MA4 | Cystic fibrosis (CF) | - | - |

| **B. thailandensis** | BCC0779 | LMG 20219; ATCC 700388; E264 | Soil (ENV) | - | - |

**Paraburkholderia**

<table>
<thead>
<tr>
<th><strong>P. phytophthora</strong></th>
<th>BCC1607</th>
<th>LMG 22487; PsJN</th>
<th>Environmental (ENV)</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC1604</td>
<td>LMG 21445</td>
<td>Environmental (ENV)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>B. graminis</td>
<td>BCC0774</td>
<td>ATCC 700544; LMG 18924</td>
<td>Soil (ENV)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Footnotes: (+), strong lytic activity; (+/-), weak lytic activity; (-) no activity; ENV, environmental; CF, cystic fibrosis; IND, industry
Figure 2. Lytic activity of *Burkholderia vietnamiensis* G4 bacteriophages and morphology of G4P1.

G4P1 displayed strong lytic activity against the host strain *B. ambifaria* BCC1212 in a drop test (a) and a plaque assay (b). G4P2 displayed strong lytic activity against the host strain *B. cenocepacia* BCC1210 in a drop test (c) and a plaque assay (d). G4P3 displayed weak lytic activity against the host strain *B. cenocepacia* BCC1210 in a drop test (e) and a plaque assay (f). Transmission electron microscopy revealed that G4P1 was a tailed phage with an icosahedral head (order Caudovirales) approximately 200 nm in length (g).

3.3 G4P1 was localised to chromosome 1 of *B. vietnamiensis* G4

The DNA sequence obtained for G4P1 was 36725 bp in length and was compared to the *B. vietnamiensis* G4 genome to determine its location as a prophage. G4P1 shared 100% blastn identity with a 36695 bp region on chromosome 1 starting in the intergenic region between tRNA-Arg (R0040) and spanning Bcep1808_1284 to Bcep1808_1331 (Figure 3). G4P1 encoded 49 genes, mobility by transposition and had clusters of early (non-structural; n=17) and late (structural; n=32) genes characteristic of Mu-like phages, named on the basis of their different levels of transcription during the lytic lifecycle (Figure 3; Table S3). G4P1 was similar in organisation to the previously discovered *Mannheimia* Mu-like phages vB_MhM_3927AP2 and phiMhaMu2, and *Haemophilus* phage SuMu, having a late phage lysin gene located immediately downstream of the early transcription activator gene mor, and lacking middle genes such as C which are late gene transcription activators [61]. Approximately 45% of the G4P1 genes (n=21) were designated as hypothetical proteins due to annotation as such or lack of concordance between annotation methods (Table S3). These features, in addition to the morphological characteristics, placed the phage within the family Myoviridae and hence G4P1 was given the name vB_BvM-G4P1 (Virus of Bacteria, *B. vietnamiensis* lysogen, Myoviridae, phage G4P1) to account for these characteristics. PHASTER identified an intact prophage in this location as prophage region 10 (Figure 3; Table S1), but inclusion of the Bcep1808_1284 gene (beta-galactosidase) was not predicted by the software, perhaps because this gene was likely to be of bacterial origin.
Figure 3. Chromosomal location of the vB-BvM-G4P1 prophage region in *B. vietnamiensis* G4. The G4P1 sequence shared 100% blastn identity with a 36695 bp region on *B. vietnamiensis* G4 chromosome.

1. vB_BvM-G4P1 integrates next to a tRNA-Arg (R0040) and spans 49 genes, Bcep1808_1284 to Bcep1808_1331. Two sub-regions are highlighted which contain early (Bcep1808_1315-1331) and late (Bcep1808_1284-1314) genes. This region was identified by PHASTER as prophage region 10 (highlighted by the red box) although inclusion of Bcep1808_1284 (beta-galactosidase) was not predicted by the software. Genes for mobility by transposition are found in the early gene cluster and the ‘A transposase’ protein sequence (highlighted in grey) was used in molecular phylogenetic analysis in Figure 5. Colour-coded functional gene categories are displayed in the key at the bottom and detailed gene annotations are given in Table S4. Easyfig software was used to visualise the comparison.

### 3.4 vB_BvM-G4P1 was found in other *B. vietnamiensis* strains

G4 PHASTER prophage region 10 (G4P1 without the beta-galactosidase gene) was compared to the 114 other intact *B. vietnamiensis* prophage regions identified by PHASTER. Highly similar regions were found in 13 other strains, nine of which had ANI scores >95% (tblastx >99%), two had ANI scores between 90% and 95% (tblastx 94.5%) and two had ANI scores between 85% and 90% (tblastx 89-94%) (Table S4). The genome structure and organisation of the 14 prophages was visualised using EasyFig software (Figure 4). A conserved backbone was found of 34-35 kb that comprised genes organised into early (non-structural) and late (structural) regions as found in G4P1. Manual inspection of the lyosgen genomes found a beta-galactosidase gene directly upstream of the G4P1-like prophage insertion site in 11 strains (BCC1170, BCC1186, LA_5_5_30_S1_D2_1, BCC0587, PC082, HI13392, BCC0194, BCC1172, BCC1193, BCC1301 and FL_5_2_10_S1_D0_Repeat). As in G4P1, the beta-galactosidase gene was adjacent to the tail fiber assembly protein, suggesting inclusion in 11 of the G4P1-like prophage genomes.

By using a lower cut-off of 70-75% ANI to compare prophage regions against G4 PHASTER prophage region 10, an additional four G4P1-like variants were identified in strains WP2 (one region), FL_2_3_30_S1_D0 (one region) and BCC0194 (two regions). BCC0194 already had a G4P1-like region, and whether the multiple G4P1-like and G4P1-variant regions resulted from superinfection, prophage duplication, or a combination of both is unknown. When G4P1, G4P1-like and G4P1-variant prophages were aligned, variation in the 34-35 kb shared backbone between prophages could be clearly observed; there was increased diversity in sequence and gene content in the G4P1-variant prophages compared to the G4P1 and G4P1-like prophages (Figure S3; Table S4).
The transposase proteins identified in G4P1, the 13 G4P1-like and the 4 G4P1-variant prophage genomes clustered together in molecular phylogenetic analysis and were distinct from transposase proteins identified in the Mu-like Burkholderia phages BcepMu (NC_005882.1) and ΦE255 (NC_009237.1) (Figure 5). Interestingly, the G4P1 transposase clusters reflected the overall ANI shared between full length prophage sequences (Figure 5). Transposase proteins sharing the highest sequence similarity to the G4P1 transposases were from phages of Rhizobium, Klebsiella, Haemophilus, Mannheimia and another Mu-like Burkholderia phage, KS10 (Figure 5; Table S5). Overall, the protein sets in G4P1, G4P1-like and G4P1-variant prophages were similar, and PHASTER identified that they were more closely related to phages of Mannheimia, Ralstonia and Haemophilus species than Burkholderia phages (Table S1).

Figure 4. Similarities in the genome organisation of G4P1 and the G4P1-like prophage regions identified by PHASTER. Easyfig was used to visualize the comparison of G4 prophage region 10 (G4P1 without the beta-galactosidase gene) to the 13 G4P1-like prophage regions identified by PHASTER. The comparison revealed a shared backbone of 34-35 kb, starting with a tail fiber assembly protein (orange arrow) and ending with a putative phage cI repressor (green arrow). The region comprised early (grey bracket; purple arrows) and late (orange bracket; blue arrows) genes. Each prophage encoded a transposase gene of 2.2 kb in length within the early region (yellow arrow). Genes outside the shared region are shown by black arrows. Strains have been organized from highest (top) to lowest (bottom) %ANI similarity to G4 prophage region 10 and prophage region identifiers are shown to the left of the alignment. Genome length is indicated by the scale bar to the right of the alignment. An asterisk (*) next to the strain name indicates the presence of a beta-galactosidase gene upstream of the tail fiber assembly protein in the lysogen genome. The grey vertical blocks between prophage sequences display %blastn shared similarity and the degree of sequence similarity is indicated by the gradient scale.
Figure 5. Molecular phylogenetic analysis of prophage transposase proteins. A maximum-likelihood phylogenetic tree infers the evolutionary history of the transposase proteins found in G4P1 (G4_10), the G4P1-like prophages sharing >95% ANI (green) and 85-95% ANI (orange) with G4P1, and G4P1-variant prophages sharing 70-75% ANI G4P1, related transposase sequences in other Myoviridae (blue) and Mu-like Burkholderia phages (black). Burkholderia phage KS10 (blue) is a Mu-like Burkholderia phage identified as having a transposase sequence related to the G4P1-like prophages. The position of the B. vietnamiensis G4P1 (G4_10) transposase is shown with a black arrow. Bootstrap percentages are shown next to the branches and the scale bar displays the evolutionary distance as number of base substitutions per site.

3.5 vB_BvM-G4P1 was widely distributed across the population structure of B. vietnamiensis

G4P1 and the G4P1-like prophages were prevalent in B. vietnamiensis being found in 17% of the strain collection (14 of 81 strains), and the G4P1-variant found in a further 2 strains. To understand the distribution of the region across B. vietnamiensis, phylogenomics was used to determine the population structure of the species. A core-genome phylogeny constructed from 3460 genes revealed that the population split into 6 clades that encompassed 77 out of the 81 strains (Figure 6). Strains carrying the G4P1-like prophage were widely distributed across the phylogenomic tree. Pairs of highly related strains carried the G4P1-like prophage (BCC1193 and BCC1172, BCC1186 and BCC1170, FL_5_2_10_S1_D0_repeat and BCC0194, PC082 and BCC0587, and G4 and BCC1301) but these were found across clades 1, 2, 4 and 5 and were from different isolation sources (cystic fibrosis, environmental, industrial), geographic locations and sequencing projects (Table S1). The G4P1-
variant prophages were found in strains from environmental sources in clades 1, 4 and 6. Strains carrying the G4P1, G4P1-like and G4P1-variant prophages all had evidence of other prophage regions and on average had 3 intact prophages (range: 1-6) and genomes comprising 2.1% prophage material (range: 0.97-3.4%).

Figure 6. Core-gene phylogeny of 81 B. vietnamiensis strains aligned with the presence of G4P1, G4P1-like and G4P1-variant prophages. A Maximum-likelihood phylogenetic tree was generated based on the alignment of 3460 shared ‘core’ genes. The root was determined using a secondary tree containing an outgroup species, B. ambifaria AMMD (Figure S1). The scale bar displays the evolutionary distance as number of base substitutions per site. Six clades were identified (grey graded sections; numbers 1-6) that encompassed all but four of the strains (BCC0136, AU25004, AZ_3_30_S1_D7 and CBMB40). Strain isolation source is indicated by strain name colour: green is cystic fibrosis, red is clinical, blue is environmental, and black is industrial. The distribution of G4P1 prophages across the species is shown to the right of the core-gene phylogeny. The B. vietnamiensis G4 prophage region 10 identified by PHASTER is indicated with the black arrow, dark green represents the 10 G4P1-like regions sharing >95% ANI with this region, medium green represents the
4 G4P1-like regions sharing 85-95% ANI and light green shows the 4 G4P1-variant regions sharing 70-75% ANI.

4. Discussion

As *Burkholderia* are intrinsically antibiotic resistant and almost impossible to eradicate in chronic lung infections in people with CF, there is considerable interest in characterising phages as potential therapeutics [51]. In addition, given the biotechnological interest in *Burkholderia* [8], further genetic tools such as integrative vectors are needed, and understanding prophage biology will enhance the development of these tools [17]. Prophages have been found in the genomes of all *Burkholderia* species that have been investigated except for *B. mallei* [62], and bacteriophages have been isolated from lysogenic strains of *B. ambifaria* [51], *B. cepacia* [63,64], *B. cenocepacia* [51,63,65,66], *B. multivorans* [28,51,63], *B. pseudomallei* [62], *B. pyrocinia* [51], *B. stabilis* [63], *B. thailandensis* [62,67], *B. vietnamiensis* [28] and members of the closely related *Paraburkholderia* genus [68]. Whilst lytic phages are the optimal choice for phage therapy, genetic modification of temperate phages has been performed to produce non-lysogenic derivatives that have been successfully used therapeutically [69]. The novel phage vB_BvM-G4P1 had broad host range lytic activity against strains of five species of Bcc including *B. dolosa*. Although *B. dolosa* is not the most commonly encountered *Burkholderia* species in CF lung infections, it is one of the most antibiotic resistant and virulent [70]. In addition, *B. dolosa* strain SLC6, which G4P1 had activity against, was responsible for an outbreak at Boston’s Children’s hospital between 1998 and 2005 that infected over 40 patients and had a high mortality rate [71]. Therefore, further characterisation of vB_BvM-G4P1 is warranted from a therapeutic perspective and as a potential tool to genetically manipulate pathogenic *Burkholderia*.

Apart from vB_BvM-G4P1, the only other phage isolated from *B. vietnamiensis* is NS1 [28], induced from the lysogen *B. vietnamiensis* ATCC 29242 (BCC0587 in this study). In our study, PHASTER determined that BCC0587 harboured a G4P1-like prophage and one other intact prophage, both of which were approximately 35 kb in length (Supplementary Table 1). There is no complete genome sequence for NS1 but the phage genome was predicted to be 48 kb in length, 15 kb smaller than the regions predicted by PHASTER. Further sequence information for NS1 is therefore required to determine the origin of this phage. The novel phage G4P1 was approximately 37 kb in length, carried genes for mobility by transposition and had genomic characteristics observed in many Mu-like phages [72]. Mu-like phages have previously been found in the genomes of *Burkholderia*, including DK4 (BcepMu) [65,73] and KS10 [74] which lysogenize *B. cenocepacia* strains, and ΦE255, found in the genome of *B. thailandensis* E255 [62]. As identified by PHASTER, G4P1 and the G4P1-like prophages carry gene sets more closely related to *Mannheimia* (vB_MhM-3927AP2), *Ralstonia* (RS138) and *Haemophilus* (SuMu) phages than *Burkholderia* phages. A recent publication actually identified similarity between a 13 kb portion of *Ralstonia* phage RS138 and *B. vietnamiensis* G4 chromosome 1 [75] which is likely to be related to G4P1, and the gene content and organisation of G4P1 is highly similar to *Mannheimia* and *Haemophilus* Mu-like phages [61]. Whilst the phages of *B. cenocepacia* have been found to encode putative virulence factors, and transduction of antibiotic resistance genes by *Burkholderia* phages has been observed, the overall contributions of phages to virulence and fitness of *Burkholderia* remains largely unknown [28,76,77]. Similarly, the G4P1 prophage does not appear to encode any virulence or antibiotic resistance determinants but has a large number of hypothetical proteins (n=21) of unknown function.

The G4P1-like prophages were common in the genomes of *B. vietnamiensis* being found in 17% of the strain collection. Although phage genomes are renowned for being mosaic in nature [7], recent studies have used genomics to discover the presence of syntenous prophages across different species [78,79], and across different strains of the same species [80-83]. The present study is the first to report the population genomics of *B. vietnamiensis* and map prophage carriage onto the population. From this analysis we could identify that vB_BvM-G4P1 was not simply localised to a closely related group of strains or single clade, but was widely distributed across the species. As vB_BvM-G4P1 does not
appear to encode genes relating to fitness, the mechanism behind its extensive shared carriage is unclear; this occurrence could represent an ancestral integration that has been preserved, or phage transmission between strains in close contact [80,84]. Whether the G4P1-like prophages are inducible and functional remains to be investigated. However, as G4P1 has broad host range lytic activity, there is the possibility that retention of the functional prophage could confer a competitive advantage to the lysogenic population (2). There were also G4P1-variants in two other B. vietnamiensis strains and one strain carrying one G4P1-like plus two G4P1-variant prophages. Here, further work is needed to understand the infection processes, immunity systems and evolutionary history of G4P1, G4P1-like and G4P1-variant prophages within B. vietnamiensis.

Prophage detection in bacterial genomes remains challenging and the limitations of using only PHASTER for prophage identification are acknowledged [77]. However, PHASTER is an excellent screening tool for examining large numbers of genomes and was applied herein to putatively predict prophage carriage across 81 B. vietnamiensis strains, the largest collection of a Burkholderia species investigated to date for prophages. All strains had evidence of prophage material, with 59 strains harbouring at least one putative intact prophage. The percentage of genetic material within B. vietnamiensis genomes attributed to prophages ranged from 0.1% to 3.4%. This is lower than the percentages found in other bacterial genera (Escherichia, Streptococcus, Borrelia; 10-20%) [3,84], but close to others (Salmonella; 4.9%) [84], and similar to other Burkholderia and closely related Paraburkholderia. Prophage material was found to comprise up to 3.67% of B. cenocepaica genomes (16 strains) [77], up to 2.34% in B. pseudomallei (6 strains), 2.00% in B. thailandensis (1 strain), 0.55% in Paraburkholderia xenovorans (1 strain), up to 4.98% in B. multivorans (3 strains) [62] and up to 4% in Paraburkholderia genomes (36 strains) [80], although certain Paraburkholderia strains can carry up to 10% of their genome as prophage material [68].

5. Conclusions

This study exploited B. vietnamiensis genomics and phylogenomics for prophage identification, facilitating the discovery of a novel Mu-like phage and its widespread distribution across the species. G4P1 and G4P1-like prophages represent a highly syntenous group, distinct from previously described Mu-like phages found to infect other Burkholderia species. G4P1 is inducible from B. vietnamiensis G4 and has lytic activity against multiple Bcc species encountered in CF infections, therefore having potential for therapeutic development. Few systematic studies of prophage carriage have been performed for Burkholderia, which is likely to change as larger genome collections become available. There is certainly evidence of polylysogeny and considerable untapped potential for the discovery of novel prophages within Burkholderia genomes.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Core-gene phylogeny of 81 B. vietnamiensis strains rooted with the outgroup B. ambifaria AMMD, Figure S2: Depth of sequencing coverage plot for vB_BvM-G4P1, Figure S3: Similarities in the genome organisation of the G4P1, G4P1-like and G4P1-variant prophage regions identified by PHASTER, Table S1: B. vietnamiensis strains, genomes and prophage regions identified by PHASTER, Table S2: Summary of vB_BvM-G4P1 sequencing and genome assembly results, Table S3: Gene annotations for the chromosomal location of the vB-BvM-G4P1 prophage region in B. vietnamiensis G4, Table S4: Comparison of G4P1, the G4P1-like and G4P1-variant prophage regions, Table S5: Sequences with highest similarity to the vB_BvM sequences identified within Genbank by blastp.

Author Contributions: Conceptualisation: RW and EM; Data Curation: RW; Formal analysis: RW; Funding Acquisition: EM and JP; Investigation: RW, ZLY and AO; Methodology: RW, EM, BVJ, JS and JP; Project Administration: RW and EM; Resources: EM, BVJ, JS and JP; Software: RW; Supervision: RW and EM; Validation: RW, ZLY and AO; Visualisation: RW; Writing-Original Draft: RW; Writing-Review & Editing: RW, EM, ZLY, AO, BVJ, JS and JP.

Funding: Genome sequence analysis of 35 B. vietnamiensis strains within this study was performed at the Wellcome Sanger Institute as part of research funded by the Biotechnology and Biology Research Council (BBSRC; grant BB/L021692/1 to EM and JP). Bioinformatics was performed using the CLIMB resource which is funded by MRC research grant MR/L015080/1.
Acknowledgments: We would like to acknowledge Angela Marchbank and the services provided by the Cardiff University Genomics Research Hub for bacteriophage genome sequencing. We also acknowledge assistance by Dr. Matt Bull (Public Health Wales, University Hospital Wales) and Dr. Alex Mullins (Microbiomes, Microbes and Informatics group, Cardiff University) who provided advice on phylogenomic analysis.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References


Viruses 2020, 12, x FOR PEER REVIEW


720 76. Sumner, E.J.; Gill, J.J.; Upton, C.; Gonzalez, C.F.; Young, R. Role of phages in the pathogenesis of Burkholderia, or ‘Where are the toxin genes in Burkholderia phages?’. Current Opinion in Microbiology 2007, 10, 410-417, doi:https://doi.org/10.1016/j.mib.2007.05.016.


© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).