Chapter 6

A novel inducible prophage from the mycosphere inhabitant *Paraburkholderia terrae* BS437

Akbar Adjie Pratama, Jan Dirk van Elsas

Published in *Sci. Rep*. (2017) 7: 9156
doi: 10.1038/s41598-017-09317-8

Abstract

Bacteriophages constitute key gene transfer agents in many bacteria. Specifically, they may confer gene mobility to *Paraburkholderia* spp. that dwells in soil and the mycosphere. In this study, we first screened mycosphere and bulk soils for phages able to produce plaques, however found these to be below detection. Then, prophage identification methods were applied to the genome sequences of the mycosphere-derived *Paraburkholderia terrae* strains BS001, BS007, BS110 and BS437, next to *P.phytofirmans* strains BS455, BIFAS53, J1U5 and PsJN. These analyses revealed all bacterial genomes to contain considerable amounts [up to 13.3%] of prophage-like sequences. One sequence predicted to encode a complete phage was found in the genome of *P. terrae* BS437. Using the inducing agent mitomycin C, we produced high-titered phage suspensions. These indeed encompassed the progeny of the identified prophage (denoted ϕ437), as evidenced using phage major capsid gene molecular detection. We obtained the full sequence of phage ϕ437, which, remarkably, had undergone a reshuffling of two large gene blocks. One predicted moron gene was found, and it is currently analyzed to understand the extent of its ecological significance for the host.

Key words: Prophage, *Paraburkholderia terrae*, Fungal-interactive, Mycosphere, Predicted morons

Background

Viruses that infect bacteria - bacteriophages (phages) - play significant roles in the evolution of bacteria, at both the individual and community levels As agents of horizontal gene transfer (HGT), phages can enhance the fitness of their host cells in the form of lysogenic conversion and/moron genes, for instance by providing so-called auxiliary metabolic genes (AMGs) (Breitbart, 2012) as well as virulence or pathogenicity traits (Brüssow et al., 2004). Moreover, phages function in the biological 'warfare' among neighboring bacterial cells and can modulate the formation of bacterial biofilms at the population level (Secor et al., 2015).

Prophages - temperate phages that occur in an integrated form in the bacterial genome - are often present in considerable amounts in bacterial genomes. For example, a recent study (Bobay et al., 2013) of 69 *Escherichia* and *Salmonella* genomes revealed prophages to occupy up to 13.5% of the genome of *Escherichia coli* O157:H7 [strain EC4115] and up to 4.9% of that of *Salmonella* Newport strain SL254. Such prophages, when intact, may be induced from the host genome, yielding phage progeny in lysates. This may occur as a response to stress, for instance resulting from exposure to UV (Bloch et al., 2015; Lamont et al., 1989), hydrogen peroxide(Linn and Imlay, 1987) or mitomycin C (MMC) (Fortier and Moineau, 2007). Moreover, prophages can be 'spontaneously' induced, which implies that cues of unknown nature may have been at the basis of induction (Nanda et al., 2015). However, many potential prophages are, to different extents, defective or 'cryptic', as they have been subjected to genetic erosion (degradation and deletion) processes (Bobay et al., 2014; Canchaya et al., 2004; Casjens, 2003). Such defective prophages may endow their hosts with gene repertoires that allow survival in harsh environments (Wang et al., 2010).

The extant abundance of phages, as compared to their bacterial hosts, is often astounding (Breitbart and Rohwer, 2005; Canchaya et al., 2004). However, we have so far only just scratched the 'tip of the phage iceberg'. Moreover, whereas most studies on phages have been made in aquatic ecosystems (Breitbart, 2012; Brum and Sullivan, 2015; Hurwitz and U'Ren, 2016; Roux et al., 2016), those in soil have been lower in number or have only just emerged (Herron and Wellington, 1990; van Elsas and Pereira, 1987).

Members of the genus *Burkholderia* exhibit a tremendous phenotypic diversity and they inhabit diverse ecological settings (Estrada-De Los Santos et al., 2013), ranging from soil (Nazir et al., 2012a; Salles et al., 2002) to plants and humans (Sahl et al., 2015). A recent study divides *Burkholderia* into two clades, in which clade I contain all pathogenic *Burkholderia* species and clade II mainly so-called "environmental" bacteria. Clade II was renamed *Paraburkholderia* (Estrada-De Los Santos et al., 2013;

Sawana et al., 2014). This genus encompasses members with the largest genomes among all known bacteria. Such genomes may have resulted from frequent HGT events and potential selection (Haq et al., 2014). Zhang et al., (2014) recently provided arguments for the tenet that the mycosphere, in the light of the bacterium-'feeding' fungus and the multitude of active bacteria occurring there, constitutes a true arena that fosters HGT. Hence, there is great interest in digging deeper into the genetic legacies of such events in mycosphere dwellers. Nazir et al., (2012b) described a suite of truly fungal-interactive *Paraburkholderia* strains, including *P. terrae* strains BS001, BS007, BS110 and BS437, and *P. phytofirmans* BS455. Analysis of the 11.5 Mb large genome of the then selected *P. terrae* BS001 - in comparison with other similar genomes - revealed 96% of it to belong to the non-core [variable] part (Haq et al., 2014). Some evidence was presented for the presence of phage-typical integrases, along with other phage-related genes, raising the question whether phages could facilitate HGT in this organism.

In this study, we hypothesized that prophage sequences present in some of the aforementioned fungal-interactive *Paraburkholderia* strains can give rise to phage populations that foster adaptive processes in *Paraburkholderia* in the mycosphere. We thus first screened the mycosphere (and corresponding bulk soil) for free phages and then – in a search for prophages – examined the genomes of mycosphere-derived *Paraburkholderia* strains. Indeed, evidence was found for the presence of putative prophage and phage-like elements in several genomes. We then focused on a predicted full-phage sequence found in *P. terrae* strain BS437, the data of which are presented here. To the best of our knowledge, this is the first study that isolates induced prophage from *Paraburkholderia* isolated from the mycosphere.

Materials and Methods

Phage isolation from soil and mycosphere samples

Replicate soil and mycosphere samples (*Scleroderma citrinum* and *Galerina* spp.) were obtained from a forest in Noordlaren in autumn 2015, and processed as in Zhang et al., (2014). Attempts to isolate phage from these samples were made using two methods. First, 0.5 g of each mycosphere sample was added to 5 ml of sterile water, after which the mixtures were vortexed vigorously. After one minute still, centrifugation at 100 *xg* (30 s) was done to sediment course soil particles. The collected supernatant was then spun at maximal speed (7,000 *xg*) for 15 min, to remove fine soil particles. Following this, 100 µL was filtered over Whatman 0.22 µm cellulose acetate filter (GE Healthcare Life Sciences, Pittsburgh, PA, USA); the suspension was then added to 20 mL of LB (Sigma-Aldrich, St. Louis, Mo, USA), with 200 µL of overnight grown

'indicator' bacteria (**Supplementary Table 6.1**). The suspensions were incubated overnight at 28 °C.

Method 2 consisted of directly adding 0.5 g soil or mycosphere sample to 20 mL LB broth and incubating overnight at 28 °C, to foster bacterial growth and potential phage development. Following incubation, the cultures were centrifuged at maximal speed (7,000 *xg*) for 10 min at 4 °C to pellet bacterial cells, and supernatants filtered over Whatman 0.22 µm cellulose acetate filter (GE Healthcare Life Sciences, Pittsburgh, PA, USA). One mL of each filtered supernatant was then added to 3 mL indicator bacteria (**Supplementary Table 6.1**) in LB medium, and incubated overnight at 28 °C. The resulting cultures were then centrifuged at maximum speed for 30 min at 4 °C and the filtered supernatants used for later cultures. The procedure was repeated five times, ultimately yielding a suspension that presumably contains phage particles (Santamaria et al., 2014).

Prophage identifications across genomes

The genomes of the selected *Paraburkholderia* strains were screened for the presence of prophages by using PHAST (Zhou et al., 2011) - version October 2015, Prophinder/ACLAME (Leplae, 2004) - version 04, October 2015 and PhiSpy [PhiSpyNov11_3.2] (Akhter et al., 2012). PHAST and Prophinder identify prophage regions by using a database of known phage genes, sequence identification, tRNA identification (as phages often use tRNAs as target sites for integration), attachment site recognition and gene clustering density measurements (prophage regions can be identified as clusters of phage-like genes within a bacterial genome) (Leplae, 2004; Zhou et al., 2011). PhiSpy uses several distinct characteristics of prophages, as outlined in the following. First, the median length of predicted proteins; as the median protein lengths in phage regions is much higher than that of proteins in the bacterial genome. Additionally, the directionality of the transcription strand and the GC skew. Both directionality of the transcription strands and GC skew are correlated with the direction of replication. Most consecutive genes in phage genome tend to be encoded on the same strand, in contrast to bacterial consecutive genes. Any observed changes in GC skew might result from the insertion of foreign DNA. Also, the abundance of unique phage words is used, next to the phage insertion site (*attP*) and the similarity to known phage proteins (Akhter et al., 2012). We here also applied other criteria to define putative prophage-like (PP) regions: (1) PP of sizes below 10 Kb were discarded (Bobay et al., 2013; Casjens, 2003) and (2) when a region consistently appeared in all three independent analyses, we used the PHAST results, as PhiSpy was reported to give less consistent results (Popa et al., 2017).

Bacterial growth and MMC-mediated prophage induction

Paraburkholderia terrae strain BS437 became the focus of this study. It was isolated from the mycosphere of *Lyophyllum* sp. strain Karsten (Nazir et al., 2012b) and is a current reference strain in our laboratory. The strain was grown in LB broth at 28 °C with shaking (180 *rpm*). Induction with MMC (Sigma-Aldrich, St. Louis, Mo, USA) was conducted according to Fortier and Moineau (Fortier and Moineau, 2007), with modifications. Briefly, bacterial cells were introduced into 5 ml of LB medium and incubated overnight at 28 °C (shaking at 180 *rpm*). The resulting cultures were then transferred (1:100) into replicate Erlenmeyer flasks containing 40 ml of fresh LB medium and growth was monitored until the exponential growth phase (about 10 h incubation). Thereafter, all cultures were split into two 20 ml cultures. MMC was added to the cultures, at final concentrations of either 4 or 10 µg/mL (MMC-4, MMC-10, respectively), with the 'twin' culture serving as the control. The cultures were incubated and the OD600 was monitored for 24 h. Decreases of the cell density were taken as indications of progressive cell lysis and prophage release. The experiments were done with three biological replicates. The resulting crude lysates were finally filtered over Whatman 0.22 µm cellulose acetate filter (GE Healthcare Life Sciences, Pittsburgh, PA, USA) and stored at −20 °C until further analysis.

Assessment of host range and indicator bacterial strains

For all phage activity tests, the double agar layer (DAL) method, next to a spot test, was used according to (Adams, 1959), with some modifications. In one effort, we used the extracted mycosphere and bulk soil directly with selected indicator *Paraburkholderia* strains (**Supplementary Table 6.1**). Suspensions resulting from the fivefold enrichment with the same indicator bacteria were also used. Spot or "lysis from without" assays were also used on the induced lysates. Briefly, overnight cultures of the indicator bacteria (**Supplementary Table 6.1**) were poured onto R2A (Becton Dickinson, NJ, USA) plate agar. Then 5 µL (10−2, 10−3, 10−4, 10−5) diluted induced lysates were spotted onto the plate and the plates incubated overnight at 28 °C.

Quantitative PCR (qPCR)

Specific primer sets for detecting phage genes were developed as the indicator gene to verify the presence of phage ϕ437 in the induced lysate. We selected one phage ϕ437-specific gene: a major capsid protein using the *P. terrae* BS437 draft sequence (Haq et al., 2014). Major capsid genes have been used to assess viral diversity (see review by Adriaenssens and Cowan, (2014)). This method followed the path taken to quantify ten closely related lambdoid phages of *Escherichia coli* strain K-12 (Edelman and Barletta, 2003; Refardt, 2012).

Here, we treated the induced lysates and the control (not treated with MMC) with DNase to remove any host genomic DNA (confirmed by host-specific PCR). Using the ϕ437 specific primer set, a 198 bp band was produced from *P. terrae* BS437 DNA, whereas no bands were amplified from genomic DNA of *P. terrae* strains BS001, BS007, BS110, 17804T or *P. hospita* DSMZ 17164T and *P. caribensis* DSMZ 1323T (**Supplementary Figure 6.1A**). Then, these strains were used to detect and quantify phage progeny in the induced lysates as described (Edelman and Barletta, 2003; Refardt, 2012).

Briefly, induced cultures were centrifuged and filtered over Whatman 0.22 µm puradisc syringe cellulose acetate filters (GE Healthcare Life Sciences, Pittsburgh, PA, USA) to remove bacterial cells and debris. A drop of chloroform was added to 10-fold diluted filtrates. These were then centrifuged at 2700 *xg* for 10 min at 4 °C. Then, 2 units of DNaseI endonuclease (Sigma-Aldrich, St. Louis, Mo, USA) with 1.3 µL 10x reaction buffer (Sigma-Aldrich, St. Louis, Mo, USA) was added to 10 µL lysate and the mixture was kept at 37 °C for 1 h. Later 1.5 µL of stop solution (Sigma-Aldrich, St. Louis, Mo, USA) was added and the mixture incubated at 95 °C for 30 min to inactivate DNaseI and also to open up phage capsids. The resulting suspensions were then diluted 10 fold and stored at −20 °C for later analysis. Primers specific for the ϕ437 gene for major capsid protein were used (PP1.437_ca1F: 5′-CACGATGACACGATCCACAC-3′; PP1.437_ca1R: 5′-GAGAACCATGCCCTGAACC-3′). The qPCR reaction mixtures consisted of 12.5 µL SYBR Green (Applied Biosystems, CA, USA), 0.75 µL each primer (Eurogentec, Liège, Belgium), 10 µL ultrapure water and 1 µL sample, for a total 25 µL reaction volume. Amplification and detection of ϕ437 product were performed using ABI 7300 (ThermoFisher Scientific, Waltham, Mass, USA) with qPCR reaction conditions: denaturation at 95 °C for 30 sec, annealing at 60 °C for 1 min and elongation at 72 °C for 60 sec. The qPCR efficiency was 106%.

The examination of the presence of prophage within indicator hosts

Experiments were performed to test the potential integration of ϕ437 (**Supplementary Figure 6.1**) using spot tests with ϕ437 containing suspensions (titer estimated at 108 per ml) on several *Paraburkholderia* strains (*P. terrae BS001*, *BS007*, *BS110*, 17804T, *P. hospital* DSMZ 17164^T and *P. caribensis* DSMZ 1323^T) as previously explained. The top and bottom parts of each spots were later streaked onto the new R2A medium and incubated overnight at 28 °C. Colony PCR-based test using specific ϕ437 gene for major capsid protein (198 bp) were used and 20 single-colonies from each strains were tested. The isolated DNA of ϕ437 and the phage suspension produced from strain BS437 were used as positive controls, whereas the unspotted strains and *E.*

coli K-12 were used as negative controls. The test was applied to potential host strain BS007, with 50 more single-colonies.

Phage particle concentration by polyethylene glycol (PEG) 8000

The induced phage particles were purified according to the PEG method of Sambrook and Russell, (2001) with the following modifications. Induced phage lysate was centrifuged at 11,000 *xg* for 15 min at 4 °C, and then supernatants were filtered over a Whatman 0.22 µm puradisc syringe filter- cellulose acetate (GE Healthcare Life Sciences, Pittsburgh, PA, USA). NaCl (29.2 g) was dissolved into 500 mL lysates to final concentration 1 M, which was then stored on ice for 1 h. Solid polyethylene glycol (PEG) 8000 was added to the supernatant to a final concentration of 10% (w/v) and the mixture stored overnight at 4 °C to allow phage particles to precipitate. The PEG-precipitated lysate was then centrifuged at 11,000 *xg* for 10 min at 4 °C (Sorvall SLA-1500 rotor). The supernatants were discarded to 20 mL and 10x SM buffer (10 mM NaCl, 50 mM Tris, 10 mM MgSO4, and 0.1% gelatin) was added for storage and later analysis.

Phage DNA extraction and sequencing

Phage DNA extraction was performed with a Phage DNA Isolation Kit (Norgen, Biotek Corp, ON, Canada) using manufacturer's protocols, with slight modification, i.e. DNase I inactivation temperature was 80 °C for 10 min. In addition, 16S rRNA PCR amplification using 16SFP/16SRP universal 16S rRNA gene primer set (Pereira e Silva et al., 2012) was performed to confirm the absence of genomic DNA in the phage DNA extracts. Aliquots of amplification products were electrophoresed in 1% agarose gels stained with ethidium bromide and visualized under UV illumination.

Phage DNA was sequenced on the Illumina HiSeq. 2500 paired-end by BaseClear (Leiden, Netherlands). The libraries for the strains were prepared using Illumina genomic Nextera XT libraries. The quality analyses of FASTQ sequence reads were done using the Illumina Casava pipeline version 1.8.3. The Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing PhiX control signal were removed using an in-house filtering protocol. In addition, reads containing (partial) adapters were clipped (up to minimum read length of 50 bp). The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.10.0. The final quality scores per sample yielded 707,8049 reads, or 166 MB, at 37.45 average quality. Reads were then aligned and successfully assembled using the CLC genomics workbench 9 (Aarhus, Denmark) with the default parameters: mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5 and similarity 0.9.

RAST (Rapid Annotation using Subsystem Technology) was subsequently used to annotate the sequenced Genome (Brettin et al., 2015). Predicted hypothetical proteins were checked with PSI-BLASTP and Phyre^2 program (Kelley et al., 2015). Predicted amino acid sequences of genes with assigned function [and of those without] were analyzed against the non-redundant (nr) NCBI database and the tailed phages database by PSI-BLASTP. Phyre^2 was used to predict secondary and tertiary structures (**Supplementary Table 6.2**). To predict the lifestyle, PHACTS (uses a novel similarity algorithm to create a training set from known phage lifestyles and a random forest that classify a multitude of decision trees McNair et al., (2012)) was used. Phage-bound σ70 promoters were predicted using predicted promoter tool (http://www.fruitfly.org/ seq_tools/promoter.html) and *ρ*-independent terminators were identified using the Arnold terminator-finding program (Gautheret and Lambert, 2001). The analysis of tRNA in the phage genome was done using tRNAscan-SE (Lowe and Eddy, 1996). The attachment (*att*) sites were analyzed using motif-finding tools MEME (Hu et al., 2013). The PROBIUS prediction tool (Käll et al., 2004) was used to predict transmembrane and signal peptide of genome ϕ437.

Transmission electron microscopy (TEM)

Viral particles were detected, and viral morphology examined by TEM (PHILIPS CM10). The phage stocks were directly applied onto carbon-coated nitrocellulose grids, and let it set for about a minute. The excess of liquid was drained with filter paper before negative staining with 1% uranyl acetate followed by washing and drying, before immediate observation in the TEM.

Genome comparison and phylogenetic trees

Known *Burkholderia* phages such as, *Burkholderia cepacia* phage Bcep22 (AY349011), *B. cenocepacia* phage BcepM (AY539836), *B. cenocepacia* phage BcepB1A (NC_005886), *B. pseudomallei* phage 1026b (AY453853), *Burkholderia* virus E125 (AF447491), *Burkholderia* phage BcepIL02 (FJ937737), *Burkholderia* phage 52237 (NC_007145), *Burkholderia* phage E202 (NC_009234), *Burkholderia* phage E255 (NC_009237), *Burkholderia* phage 644-2 (NC_009235), *Burkholderia* phage E12-2 (NC_009236), *Burkholderia* phage Bcep1 (NC_005263), *Burkholderia* phage Bcep43 (NC_005342), *Burkholderia* phage Bcep781 (NC_004333) and *Burkholderia* phage BcepNY4 (0096001), including *Enterobacteria* phage T4 (NC_00086), *Enterobacteria* phage Mu (NC_000929), *Enterobacteria* phage sfV (NC_003444), *Enterobacteria* phage P2 (AF063097), and *Enterobacteria* phage lambda (NC_001416), coupled with the PSI-BLASTP best hits for hallmark genes (phage lysozyme, major capsid, portal, tail sheath, tail length tape measure and phage terminase large subunit gene) were

used to generated phylogenetic trees and molecular evolutionary analysis. Trees were analyzed using MEGA7 (Kumar et al., 2016). The comparisons were performed with three different approaches, such as ProgressiveMauve (Darling et al., 2010), pairwise comparison (Sullivan et al., 2011) and dot-plot analysis (Kumar et al., 2016). Pairwise analysis generated by BLAST + 2.4.0 (tBLASTx with cutoff value 10−3) and map comparison Figures were created with EasyFigure (Sullivan et al., 2011). Dot-plot analysis was done using Gepard with default parameters (Krumsiek et al., 2007).

Results

Screening of mycosphere and bulk soil samples for free *Paraburkholderia* phages

Given the fact that previous studies (Nazir et al., 2012b; Warmink et al., 2011) revealed a prevalence of *Paraburkholderia* types (in particular *P. terrae*) in the mycospheres of different soil fungi, we first screened two freshly-sampled mycospheres (*Scleroderma citrinum* and *Galerina* spp.) for the presence of phage particles able to produce plaques on selected strains of *Paraburkholderia* spp. including *P. terrae*, *P. phytofirmans*, *P. caribensis*, *P. hospita* and *P. terricola* (for details of the strains, see **Supplementary Table 6.1**). Both direct extracts and fivefold phage-enriched ones (See Materials and Methods) were tested. This first attempt to detect phages that, in a lytic or temperate manner, productively interact with any of the selected *Paraburkholderia* species was done using the classical double-agar-layer [DAL] method (Adams, 1959) and spot tests. Unfortunately, neither the crude phage preparations from the mycosphere as well as bulk soil samples nor the phage enrichments showed any single plaques or lysis zones across all assays that were performed. This indicated an insufficiently low titer of virions in the extracts that were able to produce detectable clear or turbid plaques on the lawns of indicator bacteria used (**Supplementary Table 6.1**).

Analysis of putative prophage regions across *Paraburkholderia* genomes

In the light of the presumed low prevalence of free phage particles in the mycosphere as well as bulk soil samples, we then examined the putative presence of integrated phage. For that, we analyzed the genomes of the mycosphere-derived *P. terrae* strains BS001, BS007, BS110 and BS437, as well as of *P. phytofirmans* strains BS455, BIFAS53, J1U5 and PsJN, for the presence of putative prophage-like (PP) elements (**Supplementary Table 6.2**). For this, we used the phage identification programs PHAST (Zhou et al., 2011), Prophinder/ACLAME(Leplae, 2004) and PhiSpy(Akhter et al., 2012). By applying the criteria (see Material and Methods), we identified a total

of 209 PP regions across the eight *Paraburkholderia* genomes. Following curation, 127 of the regions remained for further analyses (**Tables 6.1** and **Supplementary Table 6.2**). Most of these predicted prophage regions (**Supplementary Table 6.2**) were interpreted as putative legacies of previous phage insertions, as they appeared to have lost essential phage core genes (Bobay et al., 2014).

Across the *P. terrae* strains, *P. terrae* BS007 had the largest (11.8%), and *P. terrae* BS001 the lowest (8.17%) total amount of PP region. *P. terrae* strain BS007 also harbored the largest PP (encoded ϕ007-5), of about 205.2 Kb. For the *P. phytofirmans* strains examined, *P. phytofirmans* J1U5 had the largest (13.4%) and *P. phytofirmans* BIFAS53 the lowest (2.7%) total amount of PP region. *P. phytofirmans* strain J1U5 harbored the highest PP number, i.e. 27. In contrast, *P. phytofirmans* PsJN only carried two identifiable PP regions, i.e. (encoded by us) ϕPsJN-2 (63.1 Kb) and ϕPsJN-3 (15.7 Kb). *P. phytofirmans* strain BIFAS53 contained the smallest identifiable PP (ϕBIFAS53-4), of about 10.5 Kb (**Supplementary Table 6.2**).

For the next phase of this study, (*i*) only complete phage regions that could be predicted to form phage progeny, and (*ii*) were consistently detected by all three programs, were further analyzed. It should be noted here that both PHAST and PhiSpy indicated the presence of one complete prophage in each of *P. phytofirmans* BS455 and PsJN. These regions however were excluded, as we placed a focus on the fungal-interactive *Paraburkholderia terrae*. Very convincingly, all three programs indicated that one full PP region was present in *P. terrae* BS437, with size of about 43.6 Kb (positions 6,888,478 to 6,932,098); this prophage, tentatively denoted as ϕ437, thus formed the focus of the next parts of this study.

Thus, high levels of MMC induced lysis of BS437 cells, albeit partially, which occurred concomitantly with the release of TEM-detectable phage particles (**Figure 6.2**). We then tested the potential infectivity of the released phage particles using the DAL method and spot test with diverse indicator hosts (**Supplementary Table 6.1**), including *P. terrae* BS437. In several attempts (adding different concentrations of helper salts MgCl2, MnCl2 and CaCl2), the phage lysates did not give rise to any plaque on the different hosts tested. We also examined whether any integration event had taken place on selected hosts, using suites of 20 host clones taken from the areas where lysates were spotted (**Supplementary Figure 6.1**). The clones were PCR-screened using phage ϕ437 major capsid specific primers (see Material and Methods). The results showed that any integration event that might have occurred was below the detection limit of the applied method.

Table 6.1. Genome ϕ437 assignment.

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
1	-	1372	779	198	Hypothetical protein	Hypothetical protein AcaML1_0023 [*Acidithiobacillus* phage AcaML1]	38	0,00005	33	AFU62868
2	-	1553	1410	48	Hypothetical protein	Phage protein gp26 [*Burkholderia* phage BcepB1A]	59	4.4	32	YP_024873
3	-	2233	1550	228	Hypothetical protein	Minor tail protein [*Rhodobacter* phage RcRhea]	94	1E-22	33	YP_009213512
4	-	2505	2233	91	Hypothetical protein	Hypothetical protein TAEYOUNG_67 [*Arthrobacter* phage TaeYoung]	72	1.4	29	ALY10524
5	-	2874	2749	42	Hypothetical protein	Virion encapsidated RNAP [*Erwinia* phage vB_EamP-S6]	78	0.77	41	YP_007005815
6	+	3126	3335	70	Hypothetical protein	Endolysin [*Erwinia* phage vB_EamM-Y2]	73	7.0	25	YP_007004738
7	+	3658	3909	84	Hypothetical protein	Hypothetical protein FV3_00119 [*Escherichia* phage FV3]	51	0.84	37	YP_007006290.
8	+	3899	4318	140	Hypothetical protein	Plasmid stability protein [*Synechococcus* phage S-SSM5]	23	4.5	47	YP_004324760
9	-	4991	4383	203	Hypothetical protein	Hypothetical protein SEA_VINCENZO_40 [*Mycobacterium* phage Vincenzo]	22	1.5	31	YP_009210896
10	-	5242	4988	85	Phage protein	Hypothetical protein Bcep22_gp19 [*Burkholderia* virus Bcep22]	89	4E-21	49	NP_944247
11	-	5592	5239	118	Hypothetical protein	Hypothetical protein BcepF1.080 [*Burkholderia* virus BcepF1]	36	1.8	35	YP_001039764
12	-	5855	5592	88	Phage protein	Phage protein gp3 [*Burkholderia* phage Bcep176]	100	3E-22	45	YP_355338
13	-	6325	5852	158	Hypothetical protein	Phage conserved protein gp66 [*Burkholderia* virus phi1026b]	81	1E-38	56	NP_945097
14	-	6639	6322	106	Hypothetical protein	Hypothetical protein DM_180 [*Erwinia* phage vB_EamM_Deimos-Minion]	45	0.42	35	ANH52278

Table 6.1. Genome ϕ437 assignment.(Continued)

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
15	-	6993	6658	112	Phage protein	Hypothetical protein Ac42p014 [*Acinetobacter* phage Ac42]	100	3E-20	45	YP_004009376
16	-	7462	6977	162	Hypothetical protein	Phage protein gp74 [*Burkholderia* virus phi1026b]	86	7E-52	58	NP_945105
17	-	7904	7542	121	Phage protein	Unnamed protein product [*Pseudomonas* phage phi297]	86	2E-29	47	YP_005098034
18	-	8551	7901	217	Hypothetical protein	Hypothetical protein DIBBI_075 [*Xanthomonas* phage vB_XveM_DIBBI]	17	0.55	34	YP_006383682
19	-	9150	8548	201	Phage Holliday junction resolvase	Putative endodeoxyribonuclease RusA [*Burkholderia* phage Bups phi1]	73	2E-45	54	ABY40522
20	-	10331	9147	395	Replication protein O	DNA replication protein [*Salmonella* phage vB_SemP_Emek]	41	3E-14	31	YP_006560599
21	-	10586	10332	85	Hypothetical protein	DNA-binding protein [*Caulobacter* phage Sansa]	50	0.079	38	AKU43488
22	-	10894	10583	104	Hypothetical protein	Hypothetical protein QHH_02 [*Halomonas* phage QHHSV-1]	33	0.027	47	APC45914
23	-	11123	10911	71	Hypothetical protein	Tail component protein gp17 [*Burkholderia* phage KS9]	88	0.009	35	YP_003090193
24	-	11450	11259	64	Hypothetical protein	Putative HNH endonuclease [*Brucella* phage 02_19]	28	3.1	56	AKO58996
25	-	11847	11452	132	Hypothetical protein	Transcriptional regulator [*Staphylococcus* phage IME-SA4]	22	0.24	47	YP_009219655
26	-	12239	11952	96	Hypothetical protein	Hypothetical protein [*Moraxella* phage Mcat7]	74	0,00000005	32	AKI27330
27	+	12457	13317	287	Phage repressor	Phage CI repressor [*Bacteriophage* APSE-2]	48	6E-28	45	YP_002308514
28	+	13702	13857	52	Hypothetical protein	Putative tape measure protein [*Gordonia* phage GMA3]	45	0.45	48	YP_009188584

Table 6.1. Genome ϕ437 assignment.(Continued)

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
29	+	13857	14108	84	Hypothetical protein	Hypothetical protein BPS10C_040 [*Bacillus* phage BPS10C]	22	4.6	37	YP_009002926
30	+	14105	14254	50	Hypothetical protein	Phage protein gp41 [*Burkholderia* phage Bcep176]	100	0,0002	42	YP_355376
31	+	14254	14379	42	Hypothetical protein	Tail fibers protein [*Escherichia* phage 64795_ec1]	43	0.31	61	YP_009291518
32	+	14582	14914	111	Phage protein	Hypothetical protein BcepIL02_gp10 [*Burkholderia* virus Bcepil02]	90	9E-14	38	YP_002922682.
33	+	14961	15794	278	Hypothetical protein	Hypothetical protein BcepIL02_gp11 [*Burkholderia* virus Bcepil02]	96	6E-50	36	YP_002922683
34	+	15870	16580	237	Phage-related protein	Hypothetical protein F116p07 [*Pseudomonas* phage F116]	99	5E-47	45	YP_164271
35	+	16577	16912	112	Hypothetical protein	Hypothetical protein DC1_00025 [*Burkholderia* virus DC1]	74	0.001	30	YP_006589955
36	+	16991	17842	284	Hypothetical protein	Hypothetical protein BcepF1.035 [*Burkholderia* virus BcepF1]	21	0.89	40	YP_001039719
37	+	17873	18127	85	Hypothetical protein	Hypothetical protein [*Enterobacteria* phage P2-EC31]	89	0,0001	32	CAJ43161
38	+	18129	18605	159	Hypothetical protein	Phage protein gp42 [*Burkholderia* virus phi1026b]	29	2.1	30	NP_945073
39	-	18802	18608	65	Hypothetical protein	Endolysin [*Arthrobacter* phage Gordon]	46	8.2	37	ALY08979
40	-	19078	18821	86	Hypothetical protein	hypothetical protein PBI_ZAPNER_53 [*Mycobacterium* phage Zapner]	31	0.97	41	AHZ95507
41	-	19393	19205	63	Hypothetical protein	Hypothetical protein SPN3US_0221 [*Salmonella* phage SPN3US]	38	1.2	48	YP_009153515
42	+	19465	19938	158	Hypothetical protein	-	-	-	-	-

Table 6.1. Genome ϕ437 assignment.(Continued)

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
43	+	19976	20923	316	Bacteriophage protein gp37	Hypothetical protein gp38 [*Burkholderia* virus phi1026b]	100	4E-137	62	NP_945069
44	-	21318	20968	117	Hypothetical protein	Hypothetical protein [EBPR siphovirus 1]	55	0.11	30	AEI71224
45	+	21353	21742	130	Hypothetical protein	Hypothetical protein Bcep22_gp48 [*Burkholderia* virus Bcep22]	81	6E-29	52	NP_944277
46	-	21956	21696	87	Hypothetical protein	Unnamed protein product [*Bacillus* phage SPP1]	30	9.2	35	NP_690702
47	+	22248	23240	331	phage integrase family protein	Integrase [*Pseudomonas* phage D3]	40	2E-12	33	NP_061531
48	-	23786	23520	89	Hypothetical protein	Major capsid protein [uncultured *Myoviridae*]	60	0.76	32	ACT78915
49	-	24491	23805	229	protein of unknown function DUF159	Hypothetical protein gp28 [*Burkholderia* phage KS9]	90	8E-57	45	YP_003090205
50	-	25032	24541	164	Hypothetical protein	Hypothetical protein PBI_JAY2JAY_59 [*Streptomyces* phage Jay2Jay]	29	0.45	33	YP_009225784
51	-	25607	25047	187	Hypothetical protein	Baseplate hub subunit and tail lysozyme protein [*Escherichia* phage Lw1]	31	0.29	26	YP_008060715
52	-	26014	25625	130	Hypothetical protein	Hypothetical protein fHeYen901_253 [Yersinia phage fHe-Yen9-01]	30	0.21	41	ARB06026
53	+	26247	26951	235	RecA/RadA recombinase	Baseplate wedge subunit [*Synechococcus* phage S-RSM4]	24	5.0	31	YP_003097386
54	+	26999	27214	72	Hypothetical protein	Hinge connector of long tail fiber proximal connector [*Citrobacter* phage Merlin]	85	3.4	28	YP_009203991
55	+	27186	27833	216	LigD, ATP-dependent DNA ligase	ATP-dependent DNA ligase [*Bacillus* phage phi3T]	93	2E-20	30	APD21266

Table 6.1. Genome ϕ437 assignment.(Continued)

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
56	-	28268	27888	127	Tail fiber assembly protein	Hypothetical protein [*Salmonella* phage IME207]	75	2E-21	45	YP_009322735
57	-	28702	28268	145	Hypothetical protein	Tail protein [*Bacillus* phage BigBertha]	49	3.2	32	YP_008771129
58	-	29029	28745	95	Hypothetical protein	TreK [*Staphylococcus* phage phiIPLA-C1C]	32	4.5	26	YP_009214605
59	-	29469	29026	148	Hypothetical protein	HNH nuclease [*Bacillus* phage AR9]	95	2E-27	37	YP_009282937
60	-	29973	29629	115	Hypothetical protein	Hypothetical protein RcapMu34 [*Rhodobacter* phage RcapMu]	85	7E-30	57	YP_004934677
61	-	30422	29970	151	Chain A, D20c mutant of T4 lysozyme	Phage putative lysozyme [*Idiomarinaceae* phage Phi1M2-2]	94	2E-25	40	YP_009104271
62	-	30672	30424	83	Hypothetical protein	Minor tail protein Z [*Enterobacteria* phage mEp237]	31	0.79	42	YP_009224009
63	-	31444	30974	157	Hypothetical protein	Arc domain-containing protein [*Pseudomonas* phage PaBG]	30	2E-10	54	YP_008433620
64	+	31579	31758	60	Hypothetical protein	Putative Arc protein [*Pseudomonas* phage SM1]	79	5E-10	55	ALT58107
65	+	31813	32682	290	Phage antirepressor protein	Putative antirepressor protein Ant [*Edwardsiella* phage GF-2]	71	1E-39	39	YP_009126626
66	+	32682	33413	244	Phage DNA binding protein Roi	Putative DNA binding protein Roi [*Pseudomonas* phage PAN70]	35	8E-37	71	AIX12494
67	+	33589	34266	226	Hypothetical protein	Hypothetical protein CL2_12 [*Lactobacillus* phage CL2]	9	6.7	57	YP_009201807
68	-	35093	34269	275	Conserved domain protein	Glycosyl transferase [*Synechococcus* phage S-CRM01]	16	0.16	40	YP_004508523

Table 6.1. Genome ϕ437 assignment.(Continued)

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
69	-	35344	35156	63	Hypothetical protein	Hypothetical protein Syn7803US105_79 [*Synechococcus* phage ACG-2014g]	43	1.3	44	YP_009133639.
70	-	36414	35356	353	Prophage long tail fiber protein	Putative tail protein [*Burkholderia* phage Bups phi1]	76	8E-60	48	ABY40547
71	-	37019	36423	199	Prophage tail protein	Tail protein [*Shigella* phage SfIV]	97	1E-25	35	YP_008766883
72	-	38189	37026	388	Phage FluMu protein gp47	Baseplate protein [*Shigella* phage SfIV]	89	7E-34	32	YP_009147467
73	-	38637	38191	149	Bacteriophage protein GP46	Putative tail protein [*Salmonella* phage ST64B]	80	2e-21	42	NP_700393
74	-	39159	38641	173	Prophage baseplate assembly protein V	Putative base plate assembly protein [*Salmonella* phage ST64B]	87	9E-36	41	NP_700392
75	-	40355	39204	384	Prophage tail protein	Putative tail protein [*Escherichia* virus Mu]	89	2E-30	27	NP_050648
76	-	41917	40355	521	Phage tail length tape-measure protein	Phage protein gp14 T [*Burkholderia* phage BcepB1A]	28	0,000003	28	YP_291174
77	-	43357	41933	475	Phage tail/DNA circulation protein	Tail/DNA circulation protein [*Shigella* phage SfIV]	92	2E-36	28	YP_008766878
78	-	44081	43521	187	Putative phage protein	Hypothetical protein AcaML1_0057 [*Acidithiobacillus* phage AcaML1]	34	0.34	30	AFU62902
79	-	44459	44085	125	Phage tail tube protein	Tail tube protein [*Salmonella* phage ST64B]	91	0,00003	24	NP_700387
80	-	46013	44523	497	Bacteriophage tail sheath protein	Tail sheath protein [*Enterobacteria* phage SfI]	100	3E-131	42	YP_009147459
81	-	46198	46010	63	Mu-like prophage FluMu protein GP38	Hypothetical protein [*Escherichia* phage D108]	85	0,000007	43	YP_003335786

Table 6.1. Genome ϕ437 assignment.(Continued)

ORF	-/+	start	stop	aa	RAST annotation function	PSI-BLASTp best hit (gene)[Taxa]	Cov. (%)	E-value	Id. (%)	Acc. Blast hit
82	-	46808	46209	200	Hypothetical protein	Terminase [*Mycobacterium* phage DarthPhader]	30	7.4	28	AOZ61253
83	-	47142	46801	114	Putative phage protein	hypothetical protein AcaML1_0040 [*Acidithiobacillus* phage AcaML1]	84	5e-06	33	AFU62885
84	-	48206	47142	355	Phage-related functions and prophages	Major capsid [*Aurantimonas* phage AmM-1]	99	4E-65	38	YP_009146944
85	-	49229	48300	310	Hypothetical protein	Head decoration protein D [*Aurantimonas* phage AmM-1]	16	0,00004	44	YP_009146943
86	-	49876	49259	206	Putative phage protein	Internal virion protein D [*Pseudomonas* phage phiPsa17]	52	2.6	30	AKG94384
87	-	50781	49903	293	Head-tail preconnector protein GP5	Prohead protease; 36K type signal peptide peptidase SppA [*Achromobacter* phage phiAxp-2]	76	2E-53	43	YP_009226433
88	-	52466	50778	563	Phage portal protein	Portal protein [*Xylella* phage Sano]	90	5E-86	35	AHB12085
89	-	52711	52466	82	Hypothetical protein	Phi92_gp071 [*Enterobacteria* phage phi92]	66	2.1	24	YP_009012402
90	-	53969	52719	417	Phage terminase, large subunit	Packaging terminase large subunit gpA [*Acidithiobacillus* phage AcaML1]	81	9E-79	44	AFU62879

Bacteriophage induction in *P. terrae* BS437

Given the finding of the ϕ437 encoding sequence in the *P. terrae* BS437 genome, cultures of this organism were screened for the presence of virions, using induction with different levels of MMC, in comparison to a control (to address spontaneous release; **Figure 6.1**). We took a significant decrease of the OD600 in the BS437 cultures, following addition of MMC, as an indication that prophage had been induced to excise from the host genome, resulting in production of enhanced levels of phage progeny. Indeed, MMC had a population-reducing effect, as measured by the OD600 of the cultures, with higher levels of MMC resulting in stronger decreases of the OD600. Specifically, mid-log-phase cultures - upon treatment with 10 µg/mL MMC - showed significant decreases (ANOVA $n = 3$, $P < 0.05$) of the OD600 as compared to the untreated control up to 14 h. In the control, at 10 h, exponential growth was found, with the stationary phase at 18 h being followed by a slow decrease of optical density at 24 h (**Figure 6.1A**).

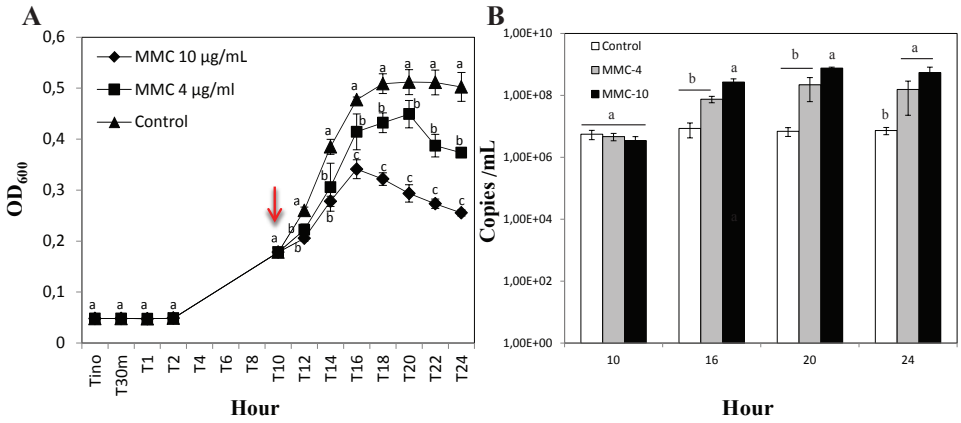


Figure 6.1. (A) Prophage induction and **(B)** quantitative PCR of the progeny. The MMC was added with a different concentration, MMC-4 (4 µg/mL) (■), MMC-10 (10 µg/mL) (♦) and without MMC/control (▲) to exponential-growing cell (10 hour; indicated with red arrow) of incubation in LB medium at 28 °C. Sample from 10 hour, 16 hour, 20 hour and 24 hour were used for quantitative PCR, error bars indicated s.d. values ($n = 3$). Significant of the treatments are marked with letter (**A**,**B**) for $P < 0.05$.

TEM was then used to observe phage progeny in the MMC-induced lysates as well as controls, and to determine the morphology of the phage particles (**Figure 6.2A**). First, phage particles were not observed in the controls, even after extensive screens. However, in the MMC-induced suspensions, homogeneous populations of virions were found. These particles were composed of isometric heads of ~50 nm in

diameter, and contractile tails with base plates of about ~75 nm length. Two to three long tail fibers were also distinguishable. According to morphological classification criteria [ICTV - International committee on taxonomy of viruses], the phage can be classified as belonging to the *Myoviridae*, with typical contractile tail features.

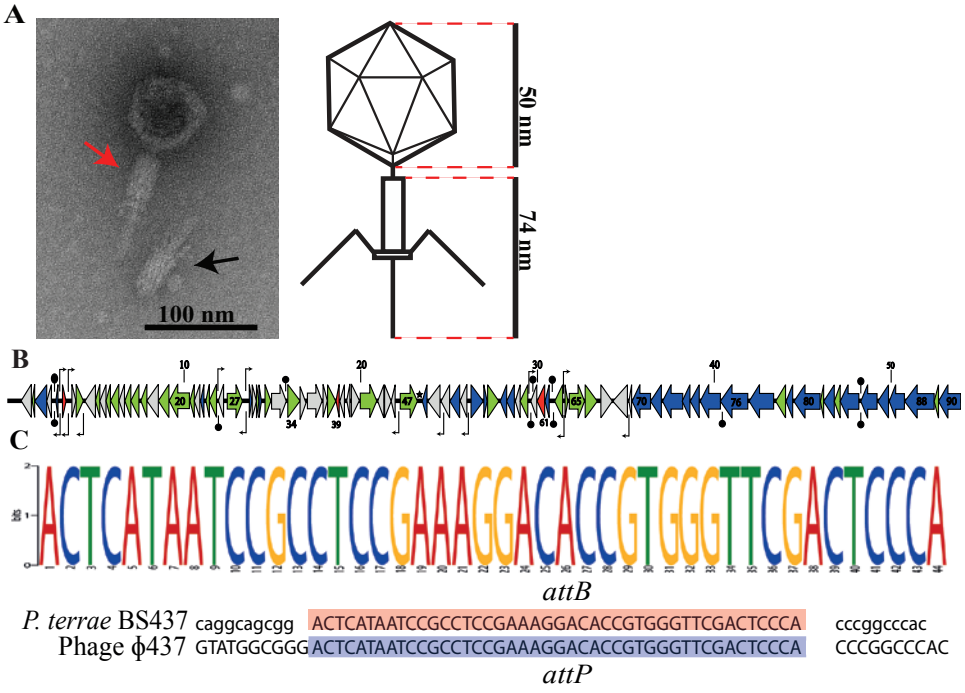


Figure 6.2. (A) The TEM image and approximate induced prophage measurement. Crude induced lysate was filtered with 0.22-µm-pore-size filter and centrifuged to pellet the cell derbies, then store in −20 °C for one night prior imaging. Image shows a typical *Myoviridae* family, the image also shows induced ϕ437 (red arrow) and ϕ437 that has lost its capsid structure (black arrow). The bar represents 100 nm. **(B)** genome sequence of ϕ437. Red arrows indicate phage lysis and lysogenic genes; blue arrows indicate phage structural genes (tail, capsid and fiber); green arrows indicate replication, recombination, repressor and phage related genes; gray arrows indicate hypothetical proteins. The black knobs indicate ρ-independent terminator and the bent arrows indicate putative promoters. The star indicates phage tRNA. **(C)** the attachment sites *attP* and *attB* of ϕ437. The *att* sites were analyzed using motif-finding tools MEME(Hu et al., 2013). The attachment sites on the tRNA *P. terrae* BS437 (*attB*) and ϕ437 (*attP*) are shown.

Linking the phage particle population to prophage ϕ437 specific genes using qPCR. We estimated that the phage lysates, estimated to have raised number of phage particles per ml (about 108 in the MMC-10 induction), contained dominant phage ϕ437 particles. To examine this tenet, we thus developed and performed phage ϕ437

based real-time quantitative PCR (Edelman and Barletta, 2003; Refardt, 2012), on extracts prepared from the control and the MMC (4 μg/mL and 10 μg/mL) induced phage lysates.

The results confirmed that the phage ϕ437 progeny levels increased over time in correspondence with the MMC concentration, with the highest gene copy number being 7.60 × 108 per ml at 20 h with MMC-10 induction (ANOVA significant $n = 3$, $P < 0.05$). On the other hand, in the control (no MMC induction), the copy numbers were consistently low, i.e. about 6.88 × 106 per ml at 20 h (**Figure 6.1B**). This result indicated that (*i*) phage ϕ437 - upon MMC induction - is indeed induced from the BS437 genome by MMC to form progeny, and (*ii*) it most likely concurs with the phage particles visualized by TEM, as described in the foregoing. Furthermore, we found a consistent presence of about 106 to 107 copies of the gene for the phage ϕ437 major coat protein in the control, indicating spontaneous release of phage particles; as yet, we still do not understand what type of 'cue', e.g. partial/incidental stress, may have caused such release.

Detailed analysis of the genome of phage ϕ437

The genome of phage ϕ437, as evidenced from virion population sequencing, was approximately 54 Kb in size, with GC-content of about 60.31%. This is slightly below the GC content of the host bacterium *P. terrae* BS437, of about 61.78%. Based on RAST annotation, the phage ϕ437 genome was found to contain 90 predicted open reading frames (ORFs), with 63 ORFs having more than 100 bp, 83 ORFs having start codon ATG (92%), four GTG (4%) and three TTG (3%). The identified PP region in BS437 (using our criteria, see materials and methods) was smaller than the sequenced genome of ϕ437. However, we did find that the PHAST-identified PP region had about 54 Kb in recent analysis. The comparison of the initially-identified smaller region with the sequenced ϕ437 genome is shown in **SupplementaryFigure 6.2**.

The biggest predicted gene in the genome of phage ϕ437 was *orf88*, of 1,688 bp (563 amino acids - aa). The predicted gene product was identified as a portal protein, which enables DNA passage during ejection and virion assembly. The predicted protein had 35% homology [90% coverage] to a similar one from *Xylella* phage Sano (AHB12085). The smallest gene (*orf31*) had only 126 bp (41 aa), and the predicted protein had 61% homology [43% coverage] to a tail fiber protein of *Escherichia* phage 64795_ec1 (YP_009291518). Interestingly, more than half of the genes of the ϕ437 genome (53 genes, 59%) were predicted to encode hypothetical proteins (**Table 6.1**.), with no designated phage sequences. This indicates the phage is indeed a repertoire of novel genes. To assign functions to the hypothetical gene products, PSI-BLASTP and Phyre[2] were used (see Materials and Methods), as detailed in the following.

Predicted genes encoding proteins that determine phage lifestyle. Phage ϕ437 was predicted to have a predominantly temperate lifestyle in its natural setting, as first evidenced by the fact that it was detected as a complete prophage. This tenet was also supported by PHACTS-supported and genomic analyses that showed the presence of typical genes involved in lysogeny. First, the phage ϕ437 genome encodes a predicted integrase (*orf47*), with 33% homology [40% coverage] to *Pseudomonas* phage D3 integrase (NP_061531). This integrase belongs to the tyrosine recombinase family, and a typical family representative is the phage lambda integrase (Williams, 2002). We also found a tRNA sequence in the intergenic region adjacent to the integrase-encoding gene (**Figure 6.2B,C**). We predict this site to be the phage integration site (Canchaya et al., 2004; Williams, 2002). A second piece of evidence for the prophage lifestyle of ϕ437 was the presence of phage lambda-like repressor genes (*orf27*), next to an antirepressor (*orf65*), indicating the presence of a system designed to 'hold'/'release' the integrated form.

Tail component and DNA packaging genes. As shown in **Table 6.1**, 28 phage ϕ437 morphogenesis genes were found, i.e. *orf3*, *5*, *23*, *28*, *31*, *48*, *51*, *54*, *56*, *57*, *62*, *70–77*, *79*, 80, *82*, *84–88* and *90*. PSI-BLASTP analyses of these genes showed homologies with database entries at 24–61% similarity and at coverages of 16–100%. PSI-BLASTP and Phyre[2] analyses revealed that some ORFs encoded hypothetical morphological proteins. Thus, predicted tail fiber protein (*orf56*) showed 45% homology [75% coverage] with a gene of *Salmonella* phage IME207 (YP_009322735). Phage ϕ437 also contained ORFs predicted to encode several baseplate proteins (*orf51*, *53*, *72 and 74*). Thus *orf51* and *orf72* may encode baseplate assembly proteins, as they showed 31% [26% coverage] and 32% homologies (89% coverage) with such ORFs from *Escherichia* phage Lw1 (YP_008060715) and *Shigella* phage SfIV (YP_009147467), respectively. Fifteen ORFs were predicted to encode a suite of tail proteins (*orf3*, *23*, *31*, *54*, *56*, *57*, *62*, *70*, *71*, *73*, *75*, *76*, *77*, *79* and *80*), next to a tail sheath protein (*orf80*). The latter showed 42% homology [100% coverage] to a putative tail protein in *Enterobacteria* phage SfI (YP_009147459). The products of *orf84* through *orf90*, next to *orf48* and *orf5*, were predicted to be involved in the packaging of DNA and in capsid formation (**Table 6.1**); the major capsid protein (*orf84*) showed 38% homology [99% coverage] to a similar protein from *Aurantimonas* phage AmM-1 (YP_009146944). The phage ϕ437 genome also contained a putative ORF encoding a portal protein (*orf40*) as well as an ORF for a large terminase subunit protein (*orf42*). These proteins showed 35% [90% coverage] and 28% homology [30% coverage] to their database counterparts, respectively. These proteins are all essential in phage DNA packaging processes.

The phage ϕ437 genome - comparison to related sequences and phylogenetic tree

In this analysis, a holistic approach was used, in which phylogenetic and overall DNA and protein sequence identities were used as the criteria. First, BLASTN analyses of the ϕ437 genome showed no similarity of the whole sequence to sequences present in the viral (tailed-phage) database. Subsequent PSI-BLASTP analyses revealed that proteins encoded by 19 of the 90 genes of the ϕ437 genome showed best hits to proteins encoded by other *Burkholderia* phages (**Table 6.1**). We thus compared the ϕ437 genome sequence to those of known *Burkholderia* phages (see Materials and Methods) using progressiveMauve (Darling et al., 2010), pairwise comparisons and nucleotide dot-plot analyses. The progressiveMauve analyses showed non-colinear synteny of the *P. terrae* phage ϕ437 sequence with those of other *Burkholderia* phages (**Figure 6.3**). Then, pairwise comparisons of the phage ϕ437 sequence to those of *Burkholderia* virus E125 (AF447491) and *B. pseudomallei* 1026b (AY453853) [both with similar genome sizes, i.e. 53.4 Kb and 54.8 Kb] confirmed the similarity, at a very low level, of phage ϕ437 with other *Burkholderia* phages (**Figure 6.4**). Finally, the dot-plot analyses also showed low similarities among the compared sequences (**Supplementary Figure 6.3**). Collectively, these results supporting the BLASTN and PSI-BLASTP analyses.

Phylogenetic analyses were then performed on the basis of selected proteins encoded by ϕ437, using MEGA7 [see Materials and Methods]. We thus analyzed phage hallmark genes, i.e. those encoding (1) lysozyme, (2) the major head capsid protein, (3) the portal, (4) the tail sheath protein, (5) the tail length tape measure protein and (6) the phage terminase large subunit. The closest hits to these proteins were most often proteins predicted from other phages (**Figure 6.5**). The trees thus consistently pointed to a relatedness of ϕ437 to other phages. However, the phage ϕ437 proteins were phylogenetically quite distantly related to similar proteins from other phages. Specifically, the phage ϕ437 encoded 150-aa lysozyme had 40% homology [94% coverage] to similar proteins encoded by *Idiomarinaceae* phage Phi1M2-2 (YP_009104271), classified to the family *Siphoviridae*. Moreover, the 354-aa major capsid protein showed 38% homology [99% coverage] to a similar protein encoded by *Aurantimonas* sp. phage AmM-1 (YP_009146944), which was classified to the family *Caudoviridae*. The 562-aa portal protein had 35% homology [90% coverage] to a similar protein encoded by *Xylella* phage Sano (AHB12085), classified to family *Siphoviridae*. The 496-aa tail sheath protein had 42% homology [100% coverage] to a tail sheath protein from *Enterobacteria* phage SfI (YP_009147459), classified to the family *Myoviridae*. The 520-aa tail length tape measure had 28% homology [28% coverage] to a similar protein from *Burkholderia* phage BcepB1A (YP_291174),

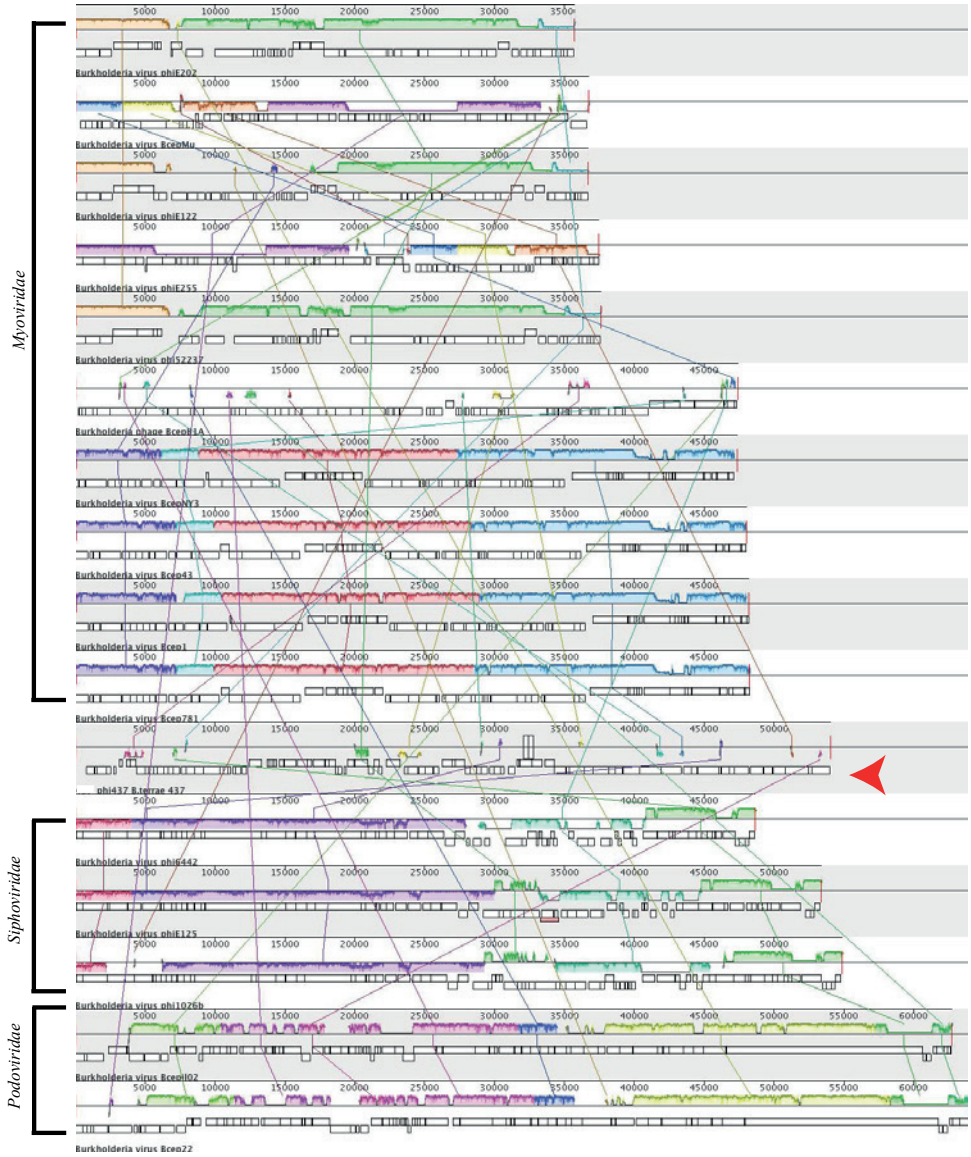


Figure 6.3. The Multiple genome alignment of *Burkholderia* phages. Genome were compare using progressiveMauve software, the genome homologous indicates by the local coliner blocks (LCB) and connected with lines. The analysis included known *Burkholderia* phages from *Myoviridae*, *Siphoviridae* and *Podoviridae*. The ϕ437 is indicated by red arrow.

classified to the family *Myoviridae*. Finally, the 416-aa phage terminase large subunit had 44% homology [81% coverage] to a similar protein from *Acidithiobacillus* phage AcaML1 (AFU62879), classified to the family *Myoviridae*. These results show an overall consistent yet low level of similarity to proteins from known phages, indicating (1) phage ϕ437 predicted proteins are related to similar ones from phages, and (2) overall, phage ϕ437 is only remotely related to any known phage.

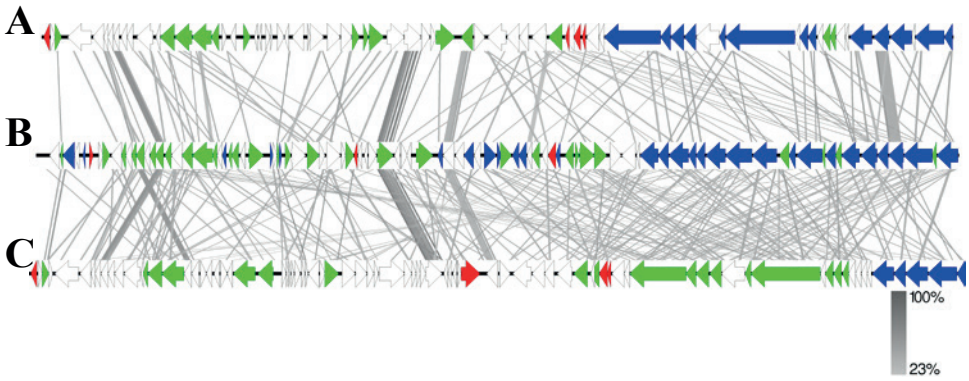


Figure 6.4. Comparison of (**A**) *Burkholderia* virus E125 (AF447491), (**B**) *Paraburkholderia terrae* phage ϕ437,and (**C**) *B. pseudomallei* phage 1026b (AY453853). Color boxes are indicated as previous Figure with additional. Comparison percentage was generated using BLAST + 2.4.0 (tBLASTx with cutoff value 10^{-3}) and map comparison Figures were created with EasyFigure as indicated in material and methods. Gene similarity percentage is indicated in gray scale bar.

Phage core genes versus predicted morons

Given the large genetic distance of most ϕ437 genes to genes of known phages (**Figure 6.3**), it was difficult to identify morons in the ϕ437 genome sequence. However, some genes with features that were strongly suggestive of morons were found (**Figure 6.2B**). In this study, we applied strict criteria for protein-encoding regions to be considered to constitute a moron: (1) they potentially give fitness advantages to the host and do not constitute phage core genes, (2) they are flanked by an upstream σ^{70} promoter and a downstream ρ-independent transcriptional terminator, allowing autonomous transcription Genes meeting criterion (2) were found in several putative intergenic regions (**Figure 6.2B**). As a third criterion (criterion 3), we used the fact that morons often have GC-contents different from those of neighboring sequences (Hendrix et al., 2000). Thus, *orf64* was singled out as a potential moron; the region was identified as a so-called *amrZ* (alginate and motility regulator Z)/Arc domain. PSI-BLASTP analysis showed *orf64* has 55% homology [79% coverage] with a similar protein present in *Pseudomonas* phage SM1 (ALT58107). This result was supported

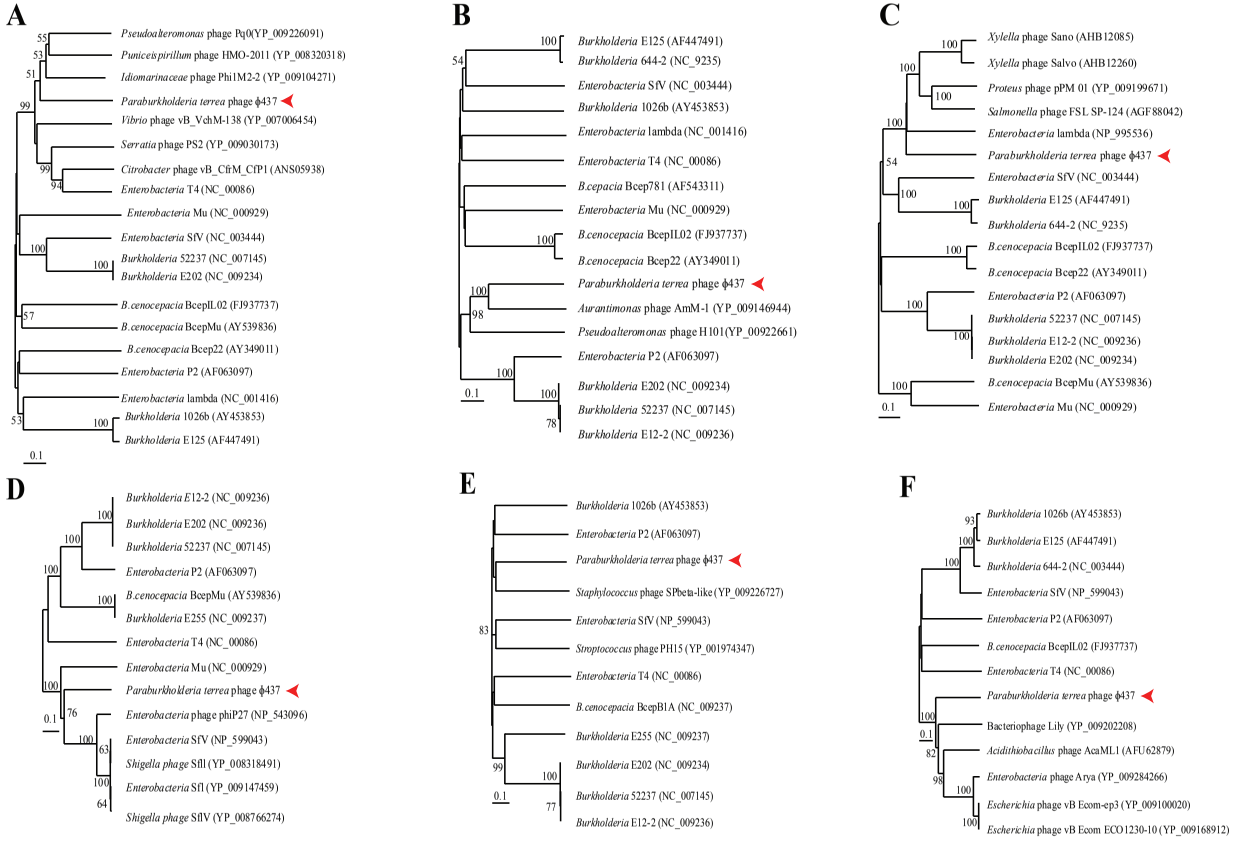


Figure 6.5. Phylogenetic trees of phage ϕ437 for (**A**) lysozyme, (**B**) major capsid, (**C**) portal, (**D**) tail sheath, (**E**) tail length tape measure and (**F**) phage terminase large subunit. Phylogenetic tree were generated with neighbor-joining tree Mega version 7 with 1,000 boothstrap method and *p*-distance methods. Red arrows indicate ϕ437. PSI-BLASTP best hits, coupled with other known *Burkholderia* phages were used in the analysis.

by Phyre[2] analysis (**Supplementary Table 6.3**). Furthermore, 55% homology [80% coverage] - with 100% confidence – was found with 'alginate and motility regulator Z' found in *Pseudomonas aeruginosa*. The *orf64* encoding transcriptional factor *AmrZ* was homologous to the *Pseudomonas* phage SM1 (ALT58107) Arc domain which had been shown to regulate virulence during infection (Xu et al., 2016). This factor is also essential for biofilm formation in *Pseudomonas aeruginosa*.

Discussion

In spite of the apparent selection and outgrowth in mycosphere soils of the *Paraburkholderia* types used as phage hosts, to our surprise we could not detect any phage that was productive (including highly lytic to temperate modes of action) on these. This indicated that such phage populations, if present, were very low in number, so that they were not detectable by the classical DAL or related spot tests. Alternatively, our indicator bacteria (**Supplementary Table 6.1**) may have had effective defense systems against the extant phage populations, which may have included R-M, CRISPR or BREX systems (Obeng et al., 2016; Weitz et al., 2013). Finally, the conditions that allow such phages to proliferate on DAL plates may not have been established in our screens. We thus set out to analyze the genomes of several selected mycosphere-isolated *Paraburkholderia* strains for predicted prophage sequences using currently accepted bioinformatics tools.

The analysis of the genomes of our *Paraburkholderia* strains to identify prophages/phage-like elements (PP) showed evidence for the contention that all of the analyzed sequences contain substantial amounts of prophage regions. Most of the identified PP regions turned out to be remnants of a phage 'history', as previously discussed (Bobay et al., 2014; Canchaya et al., 2004). These regions have probably been subjected to (stochastically acting) selective deletion pressures from the host cell, which may indicate their infrequent (re)selection. When phage structural machinery genes get eroded, prophages lose their abilities to produce progeny. Such prophages might still be coding and remain functional as they offer lysogenic conversion to host cell (Casjens, 2003) or they increasingly might represent 'passive genetic cargo' that is not transcribed (Canchaya et al., 2004). With respect to the identified phages, such hypotheses surely need experimental evidence.

A certain prevalence of prophages in the *Paraburkholderia* genomes was expected considering the fact that these *Paraburkholderia* species can inhabit the mycosphere, an environment that has been depicted as a hot spot for HGT processes in soil (Zhang et al., 2014). So far, only few studies have successfully described phages from *Burkholderia* (and/or *Paraburkholderia*) spp.(Gill et al., 2011; Lynch et al., 2010; Ronning et al., 2010; Seed and Dennis, 2005; Summer et al., 2004). However, most

phages described were from pathogenic strains isolated from clinical environments, i.e. *B*. *cepacia* complex isolates. To the best of our knowledge, no previous studies have as yet focused on *Paraburkholderia* phages in environmental isolates, especially from the mycosphere. We here singled out the *P*. *terrae* strain BS437 phage ϕ437, on the basis of the experimental and computational analysis, as outlined in the foregoing. Phage ϕ437 was apparently 'spontaneously' released in strain BS437 populations growing in liquid medium, whereas its particle numbers were raised by successful induction with MMC (**Figure 6.1A**). These observations were supported by the concomitant phage coat gene based qPCR analyses and TEM observations (**Figures 6.1**B and **6.2A**). However, we did not detect any infective phage particles by the DAL or spot tests applied to phage lysates, which may be due to (1) the absence of infectivity in our phage lysate, or (2) an intrinsic resistance or insusceptibility of host cells to released phages, as previously observed in other study. Notably, 45 strains of *Clostridium difficile* also failed to show infective phage production using the DAL method (Fortier and Moineau, 2007). The isolation, propagation and downstream analysis of phages from natural samples remain a challenge The absence of detectable phage activity in the spot tests clearly excluded a lysis-from-without scenario under these conditions.

The spontaneous prophage induction that was observed in the liquid controls used [non-MMC induction] (**Figure 6.1A**), if occurring in natural settings, might have an impact on host fitness (Nanda et al., 2015). We hypothesized that ϕ437 might modulate the formation of *P*. *terrae* BS437 biofilms on its fungal host strain, which we presume to be akin to *P*. *terrae* strain BS001 forming biofilms on *Lyophyllum* sp. strain Karsten (Haq et al., 2016). However, experimental work still needs to be done to prove this theory. Collectively, the significant decrease of the OD_{600} in strain BS437 cultures upon MMC induction, the phage progeny observed by TEM, and the increased gene copy number of the ϕ437 major capsid gene strongly indicate that phage ϕ437 was the major, if not only, phage that was released from the genome of *P*. *terrae* BS437.

The genomic architecture of ϕ437, compared to *Burkholderia* virus E125 (AF447491) and *B*. *pseudomallei* 1026b (AY453853) indicated a strong conservation of a cluster of functional genes (phage core genes) in the same relative spatial position. Tail (*orf70-orf80*) and head (*orf84-orf90*) morphogenesis genes were among the most conserved genes in the ϕ437 genome. This is consistent with data by Morgan et al., (2002) and Summer et al., (2004), indicating that such conserved genes as well as gene order represents a phage gene repertoire that is fine-tuned to effectively execute key phage functions (as shaped by evolution). Moreover, the key functional genes may be better interchanged in the continuous flux of gene acquisition and recombination

in the bacterial host genome. The analyses applied to assign the taxonomic class of ϕ437 show no large sequence similarity to any known phage sequences in the public database. However, the phylogenetic analysis of the selected phage hallmark genes (phage lysozyme, major capsid, portal, tail sheath, tail length tape measure and phage large terminase subunit) revealed ϕ437 to be most related to phages from the *Myoviridae* family. Moreover, the morphology of ϕ437 placed it in the *Myoviridae*. We thus propose ϕ437 as a new member of this family, with unique sequence features that do not relate to any of the currently ICTV-recognized subfamilies or genera.

The integration of phage ϕ437 is not well understood and does not fit classical integration mechanisms. We found the site/region of integration in the host bacterium and phage genome showed interrupted blocks, regardless of sequence identity. It is noteworthy that comparative studies of lambdoid bacteriophage genomes (Casjens, 2003) also revealed mosaicisms as a consequence of HGTs involving homologous and non-homologous recombinations (Hendrix, 2008; Juhala et al., 2000). Additionally, moron genes have been reported to be common in *Burkholderia* phages (Ronning et al., 2010). Our analyses found one moron (*orf64*) that potentially endows the host with a superinfection defense mechanism against other phage infection, enhance host fitness and enhance biofilm formation. Considering this line of evidence, we hypothesize that the gene product potentially plays a role in the *P. terrae* strain BS437 interaction with a host fungus in the mycosphere, including biofilm formation. Although the significance of this potential moron still remains enigmatic at this point, this analysis gives direction for future experiments.

Acknowledgements

This work was supported by a scholarship of the Indonesia Endowment Fund for Education (LPDP - Lembaga Pengelolaan Dana Pendidikan, Departemen Keuangan, Republik Indonesia) to A.A.P. Furthermore, we would like to thank Dr. Andrew Kropinski, chair of the Bacterial virus Subcommittee of ICTV for suggestions on phage classification. Thanks are also due to Dr. Pilar Eliana Puentez Tellez for helpful comments on an early draft. We also thank Dr. M.C.A. Stuart, Dr. K. Knoops, Dr. Maria Julia de Lima Brossi, Juliana Eschholz and Dr. Diego Javier Jimenez Avella for their technical help; Dr. Miaozhi Zhang, Dr. Ovidiu Popa, Dr. Irshad Ul Haq, Maryam Chaib De Mares, and Dr. Miao Wang for helpful discussions. Furthermore, we would like to thank the editor of *Sci. Rep.* and, importantly, two anonymous reviewers for their critical and helpful comments.

Author Contributions

Conceived, designed, performed the experiments and collected data: A.A.P. Performed the bioinformatics analyses and drafted the manuscript: A.A.P. Conceived of the study, and participated in its design, coordination and helped to draft the manuscript: J.D.V.E. All authors read and approved the final manuscript.

Competing Interests

The authors declare that they have no competing interests.

SUPPLEMENTARY INFORMATION CHAPTER 6

A novel inducible prophage from the mycosphere inhabitant *Paraburkholderia terrae* BS437

Akbar Adjie Pratama and Jan Dirk van Elsas

Published in *Sci. Rep*. (2017) 7: 9156
doi: 10.1038/s41598-017-09317-8

Supplementary table 6.1. *Paraburkholderia* strain used in double agar layer method (DAL) and spot-test for screening and host-range test[a].

Bacteria	Strain	Acc. No[a]	Reference	Bacteria genome (bp)	No. of PP[b]	Total PP regions genome (bp)
P.terrae	BS001	This study	22, 24	11294072	17	923462
	BS007	This study	26	11024679	23	1299512
	BS110	This study	87	11176041	17	1114486
	BS437	This study	26	11301206	25	1274709
P.phytofirmans	BS455	This study	26	8859905	8	390292
	BIFAS53	This study	88	8267758	8	224069
	J1U5	This study	88	10330795	27	1387966
	PsJN	NC_010681	89	4467537	2	78829

[a]DAL: Double agar layer method; N: no plaque forming observed

Supplementary table 6.2. General result of putative prophage regions analysis across *Paraburkholderia* genomes.

Bacteria	Strains	DAL[a]	Spot-test[a]
Paraburkholderia terrae	BS001	N	N
	BS007	N	N
	BS110	N	N
	BS437	N	N
	DMSZ 17804T	N	N
P.phytofirmans	BS455	N	N
	BIFAS53	N	N
	JU15	N	N
	PsJN	N	N
	BS413	N	N
	BS410	N	N
	BS420	N	N
	BS421	N	N
	BS425	N	N
P.caribensis	DSMZ 1323T	N	N
P.hospita	DSMZ 17164T	N	N
P.tericolla	BS430	N	N
	BS454	N	N

[a]Acc.No: accession number; [b] PP: putative prophage-like elements

Supplementary table 6.3. Phyre[2] analysis of genome ϕ437[a].

ORF	aa length	GC%	kDa	pI	PDB molecule	Con (%)	Cov (%)	Id (%)	Motif (TMD, SP)
1	197	60.1	21,67	6.31	Putative DNA-binding domain	97	29	21	N/N
2	47	59.03	5,17	11.79	LuxS/MPP-like metallohydrolase	17	85	23	N/N
3	227	60.08	24,97	10.14	Alpha-catenin/vinculin	72	21	26	N/N
4	90	62.27	9,9	8.06	Putative nucleoprotein	16	9	75	N/N
5	41	53.17	4,51	4.08	Inhibitory polypeptide	11	15	83	N/Y
6	69	50.47	7,59	7.81	Cell wall binding repeat	12	16	27	N/Y
7	83	55.16	9,13	9.10	Coil-vald	61	31	46	N/Y
8	139	49.05	15,29	4.94	Voltage-gated potassium channels	84	23	34	Y/N
9	202	58.29	22,22	5.50	Homeodomain-like	39	7	27	N/N
10	84	65.09	9,24	10.02	Binding protein glucocorticoid receptor dna-binding factor 1	62	36	20	N/N
11	117	54.23	12,87	5.83	Hydrolase	21	37	19	N/N
12	87	61.36	9,57	5.66	HAD-like	59	28	21	N/N
13	157	62.02	17,27	8.95	Acyl-CoA N-acyltransferases (Nat)	97	74	15	N/N
14	105	60.37	11,55	9.29	Immunoglobulin-like beta-sandwich	38	24	20	N/N
15	111	60.11	12,21	5.75	Uncharacterized protein drra	45	20	32	N/N
16	161	62.34	17,71	9.79	Prophage-derived uncharacterized protein ybco	100	47	28	N/N
17	120	62.80	13,2	4.56	CMP/hydroxymethyl cmp hydrolase	100	98	17	N/N
18	216	61.29	23,76	10.94	Pre-mrna-splicing factor snu114	100	44	21	N/N
19	200	63.18	22	9.16	Holliday junction resolvase RusA	100	57	30	N/N
20	394	60.16	43,34	8.87	DNA replication protein dnad	99	22	15	N/N
21	184	61.17	20,24	9.41	-	-	-	-	N/N
22	103	63.78	11,33	9.89	DNA-binding domain	92	42	33	N/N
23	70	63.38	7,7	11.53	Signaling protein t-lymphoma invasion and metastasis-inducing	10	10	57	N/Y
24	63	52.60	6,93	5.52	PABC (PABP) domain	24	41	19	N/N
25	131	61.61	14,41	9.29	Virulence-associated protein I	95	50	15	N/N
26	95	60.76	10,45	9	Transcription cro protein	96	34	47	N/N
27	286	57.14	31,46	5.64	Lambda repressor	100	75	20	N/N
28	51	61.53	5,61	7.83	Surfactant-associated protein d	12	62	30	N/N

Supplementary table 6.3. Phyre[2] analysis of genome ϕ437[a]. (Continued)

ORF	aa length	GC%	kDa	pI	PDB molecule	Con (%)	Cov (%)	Id (%)	Motif (TMD, SP)
29	83	57.53	9,13	10.55	40s ribosomal protein s10, putative	26	49	27	N/N
30	49	64.66	5,39	7.82	Motor protein,protein transport myosin-vi	21	27	54	Y/N
31	41	64.28	4,51	7.85	Single transmembrane helix	11	54	27	N/Y
32	110	62.16	62.16	5.86	Signal recognition particle alu RNA binding heterodimer, SRP9/14	24	25	25	N/N
33	277	61.39	61.39	5.25	Viral protein	59	31	18	N/N
34	236	63.71	25,96	5.56	PUA domain-like	96	27	20	N/N
35	111	61.01	12,21	3.21	Hydrolase, alpha-xylosidase bogh31a	27	33	24	N/N
36	283	68.19	31,13	9.32	Lambda repressor-like DNA-binding domains	27	7	29	Y/N
37	84	60	9,24	6.16	Activating signal cointegrator, RNA binding protein	100	90	37	N/N
38	158	57.86	17,38	4.83	Sulfite reductase [ferredoxin], chloroplastic	25	29	20	N/N
39	64	55.89	7,04	4.45	Transferase effector protein hopab2	30	30	32	N/N
40	85	57.36	9,35	8.89	Calcium-binding protein	14	12	70	N/N
41	62	57.14	6,82	4.54	Isomerase	27	34	29	N/N
42	157	59.91	17,27	9.17	Uncharacterized protein	25	27	29	N/N
43	315	62.44	34,65	6.03	Spore photoproduct lyase	98	39	12	N/N
44	116	54.41	12,76	4.87	Protein parc with unknown function	29	16	33	N/N
45	129	58.97	14,19	9.44	Succinate dehydrogenase/ fumarate reductase flavoprotein, catalytic domain	69	26	18	N/N
46	86	49.04	9,46	5.76	Peptidase family u32 with unknown function	28	36	26	N/N
47	330	61.02	36,3	10.33	Tyrosine recombinase XerH	100	77	18	N/N
48	88	51.68	9,68	4.94	Protein p31	33	23	40	Y/N
49	228	59.24	25,08	6.96	DNA binding protein	100	95	20	N/N
50	163	52.64	17,93	6.65	TerB-like	37	25	44	N/N
51	186	52.94	20,46	4.54	Chaperone	45	22	32	N/N
52	129	56.15	14,19	5.62	Uncharacterized protein	24	22	18	N/N
53	234	63.40	25,74	9.61	Hypothetical protein pa3008	100	47	23	N/N
54	71	50	7,81	4.73	Hydrolase	37	76	22	N/N
55	215	58.64	23,65	9.39	DNA ligase	100	97	24	N/N

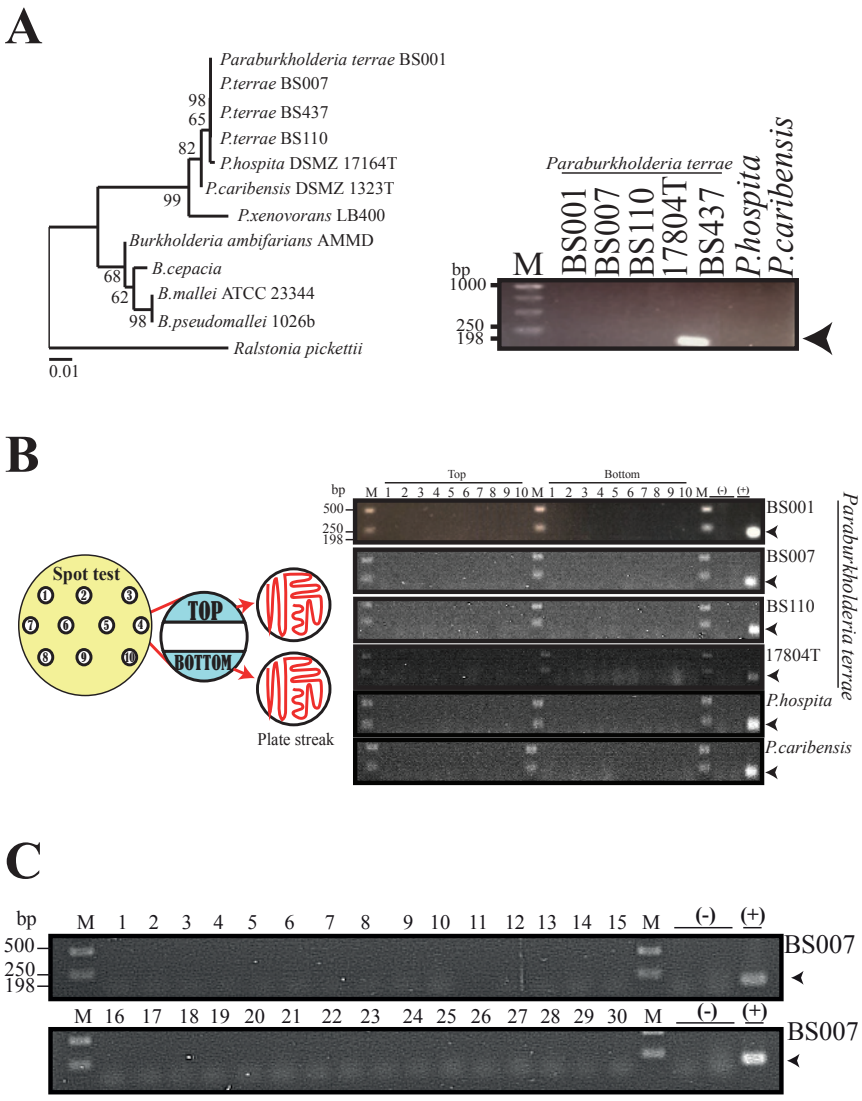
Supplementary table 6.3. Phyre[2] analysis of genome ϕ437[a]. (Continued)

ORF	aa length	GC%	kDa	pI	PDB molecule	Con (%)	Cov (%)	Id (%)	Motif (TMD, SP)
56	126	65.09	13,86	5.46	Hydrolase	97	54	19	N/N
57	144	61.37	15,84	5.07	Minor ampullate spidroin	56	55	14	N/Y
58	94	63.85	10,34	8.54	TIR domain	11	53	22	Y/Y
59	147	58.33	16,17	5.82	HNH endonuclease	92	41	21	N/N
60	114	62.31	12,54	4.93	Ligase	33	25	21	N/N
61	150	61.81	16,5	8.60	Endolysin,claudin-4	100	98	28	N/N
62	82	61.4	9,02	8.16	Signaling protein tnfaip3-interacting protein 2	71	44	31	Y/N
63	156	53.9	17,16	8.80	Alginate and motility regulator z	100	31	44	N/N
64	59	56.1	6,49	6.67	Alginate and motility regulator z	100	80	55	N/N
65	289	60.22	31,79	8.77	"Winged helix" DNA-binding domain	33	8	21	N/N
66	243	60.92	26,73	6.53	Iron-dependent repressor IdeR	94	26	16	N/N
67	225	61.06	24,75	5.52	Sulfite reductase	61	22	28	N/N
68	274	54.54	30,14	5.57	d-mycarose 3-c-methyltransferase	100	73	16	N/N
69	62	44.44	6,82	3.96	Transcriptional activator tipa-s	19	40	20	N/N
70	352	57.03	38,72	6.57	Ribonuclease Rh-like	66	8	28	N/N
71	189	61.13	20,79	5.57	Iron binding protein lipoprotein	36	55	8	N/N
72	387	63.48	42,57	4.55	Baseplate wedge protein gp6	100	90	13	N/N
73	148	61.96	16,28	3.96	gpW/gp25-like, Contains phage tail lysozyme	88	53	18	N/N
74	172	59.53	18,92	5.07	Baseplate assembly protein v	100	74	16	N/N
75	383	60.85	42,13	4.90	Tail protein	100	82	17	N/N
76	520	60.46	57,2	9.03	Hydrolase morphogenesis protein 1	100	25	20	N/N
77	474	62.80	52,14	5.02	Viral hypothetical protein	97	22	17	N/N
78	186	60.42	20,46	5	Chaperone	48	11	19	N/N
79	124	57.33	13,64	4.75	2Fe-2S ferredoxin-like	36	15	11	N/N
80	496	63.58	54,56	5.04	Uncharacterized protein dsy3957	100	75	16	N/N
81	62	61.90	6,82	5.61	Carboxypeptidase inhibitor	26	39	39	N/N
82	199	62	21,89	4.59	Phage tail protein-like	95	60	16	N/N
83	113	58.47	12,43	4.49	Phage tail proteins	100	85	28	N/N
84	354	60.75	38,94	5.15	Putative capsid protein of prophage	100	92	11	N/N

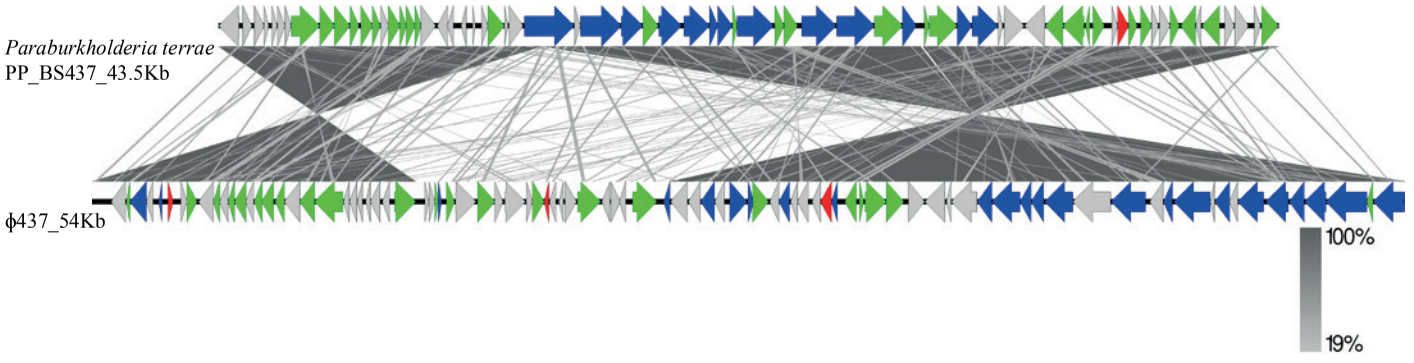
Supplementary table 6.3. Phyre[2] analysis of genome ϕ437[a]. (Continued)

ORF	aa length	GC%	kDa	pI	PDB molecule	Con (%)	Cov (%)	Id (%)	Motif (TMD, SP)
85	309	66.88	33,99	4.99	Head decoration protein D (gpD, major capsid protein D)	95	20	11	N/N
86	205	64.40	22,55	5.50	Lambda repressor-like DNA-binding domains	2	9	>5	N/Y
87	292	59.95	32,12	4.72	Protease 4	100	90	19	N/N
88	562	60.98	61,82	5.21	Portal protein	99	65	9	N/N
89	81	57.72	8,91	9.77	Head-to-tail joining protein W, gpW	98	65	32	N/N
90	416	62.90	45,76	8.68	Formate dehydrogenase protein FdhE-like	94	12	24	N/N

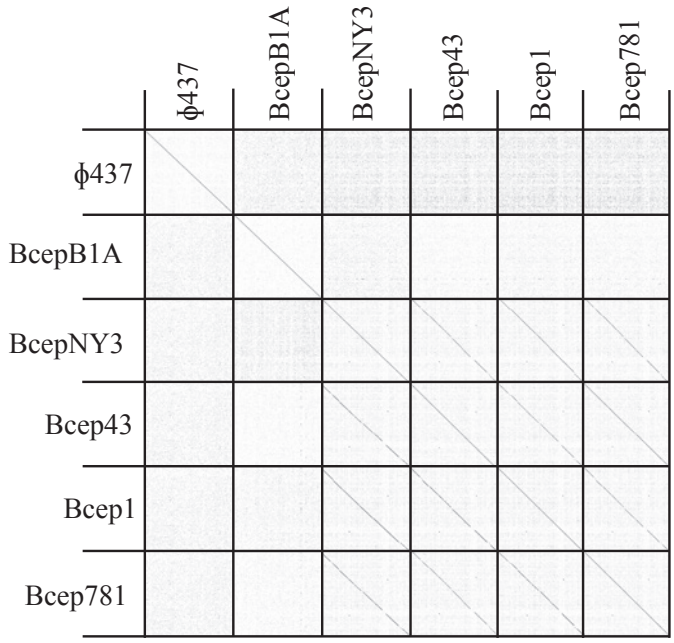
[a]kDa were calculated based on aa sequence; PDB, protein data bank; pI, isoelectric point; Con, confidence; Cov, coverage, Id, identity;TMD, Transmembrane domain; SP, signal Peptide; Y, present of TMD/SP; N, not present of TMD/SP.



Supplementary figure 6.1. The examination of the presence of prophage within indicator hosts. To address the integration events, the PCR-based tests were conducted. **(a)** Using the specific primer sets for phage ϕ437 major capsid protein. **(b)** The induced lysate were spotted onto the plates to later streak onto new plates and tested for the present of phage ϕ437 major capsid protein. Total of 20 colonies from each spots and indicator bacteria were tested. For negative control, colonies from pure culture of corresponding indicator bacteria strains were use, along with *E.coli* E12. For positive control, phage DNA and lysate were used. **(c)** The total of 30 more colonies from of *P.terrae* were further investigated.



Supplementary figure 6.2. Comparison of initially-identified PP region in host BS437 genome (PP_BS437) and sequenced of induced ϕ437. Comparison percentage was generated using BLAST+ 2.4.0 (tBLASTx with cutoff value 10^{-3}) and map comparison figures were created with Easyfig as indicated in material and methods. Gene similarity percentage is indicated in gray scale bar. Red arrows indicate phage lysis and lysogenic genes; blue arrows indicate phage structural genes (tail, capsid and fiber); green arrows indicate replication, recombination, repressor and phage related genes; gray arrows indicate hypothetical proteins.



Supplementary figure 6.3. Nucleotide dot-plot analysis. The fasta nucleotide comparisons of ϕ437 and other *Paraburkholderia* phages were compared using Gepard program (Krumsiek et al., 2007) with default parameters (see Materials and Methods), indicated by the diagonal lines.


Texel, 2016