Research in Microbiology 169 (2018) 628–637

Contents lists available at ScienceDirect

Research in Microbiology

journal homepage: www.elsevier.com/locate/resmic

Occurrence, integrity and functionality of *Aca*ML1–like viruses infecting extreme acidophiles of the *Acidithiobacillus* species complex

Paulo C. Covarrubias [a, b, 1], Ana Moya-Beltrán [a, b], Joaquin Atavales [a], Francisco Moya-Flores [c], Pablo S. Tapia [a], Lillian G. Acuña [a, 2], Silvia Spinelli [d], Raquel Quatrini [a, 3, *]

[a] *Fundación Ciencia y Vida, Avenida Zañartu 1482, Ñuñoa, 7780272, Santiago, Chile*
[b] *Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile*
[c] *Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA*
[d] *Architecture et Fonction des Macromolécules Biologiques (AFMB), Aix-Marseille Univ-Centre National de la Recherche Scientifique (CNRS), UMR 7257, Campus de Luminy, Case 932, 13288, Marseille Cedex 09, France*

ARTICLE INFO

Article history:
Received 1 April 2018
Accepted 23 July 2018
Available online 20 August 2018

Keywords:
Acidithiobacilli
A. caldus
T. tepidarius
VLPs
Bacteriophage
PFGE

ABSTRACT

General knowledge on the diversity and biology of microbial viruses infecting bacterial hosts from extreme acidic environments lags behind most other econiches. In this study, we analyse the *Aca*ML1 virus occurrence in the taxon, its genetic composition and infective behaviour under standard acidic and SOS-inducing conditions to assess its integrity and functionality. Occurrence analysis in sequenced acidithiobacilli showed that *Aca*ML1-like proviruses are confined to the mesothermophiles *Acidithiobacillus caldus* and *Thermithiobacillus tepidarius*. Among *A. caldus* strains and isolates this provirus had a modest prevalence (30%). Comparative genomic analysis revealed a significant conservation with the *T. tepidarius Aca*ML1-like provirus, excepting the tail genes, and a high conservation of the virus across strains of the *A. caldus* species. Such conservation extends from the modules architecture to the gene level, suggesting that organization and composition of these viruses are preserved for functional reasons. Accordingly, the *Aca*ML1 proviruses were demonstrated to excise from their host genomes under DNA-damaging conditions triggering the SOS-response and to produce DNA-containing VLPs. Despite this fact, under the conditions evaluated (acidic) the VLPs obtained from *A. caldus* ATCC 51756 could not produce productive infections of a candidate sensitive strain (#6) nor trigger it lysis.

1. Introduction

General knowledge of the diversity and biology of microbial viruses from extreme environments lags behind most other econiches. Since the early 90s viruses infecting acidophilic archaea (*Sulfolobales* and *Thermoplasmales*) have been recovered from enrichment cultures of samples originating in acidic hot springs [1,2] and moderately hot acidic mine drainages [3]. Although the number of known acidophilic archaeal viruses is limited, their morphological diversity is exceptional, comprising spindles, filaments, bottles and turreted icosahedral morphotypes never observed among bacterial viruses [2,4]. The genes they encode are underrepresented in current databases and their proteins products carry rather unusual folds [1,5]. Also, they have utterly unique life cycles, with mechanistic innovations in virion egress [6]. Like their hosts, these viruses show adaptations to the extreme acidity (and high temperature) of the econiches they inhabit, being stable at pH values lower than 3 (and temperatures above 80 °C) [7]. Their uniqueness has raised many relevant mechanistic and evolutionary questions [4] and their properties, as naturally occurring nanoparticles, have promising biotechnologically applications [8].

In contrast, to their archaeal counterparts, very few viruses infecting bacterial acidophiles have been detected or described in

Abbreviations: VLPs, viral-like particles; MGEs, mobile genetic elements; ICEs, integrative conjugative elements; RISC, reduced inorganic sulfur compound; PFGE, pulse field gel electrophoresis.

* Corresponding author.
E-mail address: rquatrini@cienciavida.org (R. Quatrini).
[1] Current address: uBiome Chile, SpA, Santiago, Chile.
[2] Current address: Laboratorio de Genética y Patogénesis Bacteriana, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile.
[3] Millennium Nucleus in the Biology of the Intestinal Microbiota, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile.

https://doi.org/10.1016/j.resmic.2018.07.005

the literature. These include the ØAcl bacteriophage from *Acidiphilium* sp. ATCC 55305 [9], the ThGEI-J prophage from *Thiomonas* sp. 3As [10], the *Aca*ML1 provirus from *Acidithiobacillus caldus* ATCC 51756 [11] and a putative virus (AMDV1), reconstructed from metagenomic data from an acid mine drainage biofilm, presumably infecting group II and III *Leptospirillum* spp. [12]. Even if virions have been found to occur in cognate samples [9,10,13] these bacterial viruses, featuring temperate life cycles, remain poorly characterized.

All four aforementioned bacterial species have representatives in the biotechnological process of bioleaching. In particular, the acidithiobacilli are ubiquitous in industrial acidic biotopes, with several characterized species that play key roles in metal recovery [14]. Representative members of this group have recently been sequenced [15] and a wide variety of mobile genetic elements [16] have been recognized to populate the genomes of the different species through comparative genomic analyses. Despite this fact no temperate virus other than the one infecting the *A. caldus* type strain (*Aca*ML1) has been detected so far in this complex taxon [17], and little is known about the influence of this biological agent on the biology of its host and its performance in bioleaching. On the other hand, viruses are useful tools in genetic engineering of cognate hosts and, as such, their description and characterization holds the promise of developing reproducible genetic tools for these biotechnologically important acidophilic model microorganisms.

Herein, we explore the *Aca*ML1 virus occurrence, composition and behaviour under standard and SOS-inducing conditions, to evaluate its integrity and functionality. Given the absence of reproducible genetic modification tools for the acidithiobacilli, identification of a lysogenic bacteriophage capable of infecting and transducing genetic information into members of this genus could represent an important contribution to the field.

2. Materials and methods

2.1. Nucleotide sequence data analysis

Completely sequenced genomes of *Acidithiobacillus* spp. and *Thermithiobacillus tepidarius* (Table 1; Supplementary Table 1) were obtained from the NCBI website (http://www.ncbi.nlm.nih.gov). The genome of strain DSM 16786 was obtained form the site: http://biominingdb.cmm.uchile.cl/genomes/At_ferrooxidans_Wenelen. *Aca*ML1-like proviruses from *A. caldus* strains were sequenced from PCR amplicons spanning the different *Aca*ML1 modules, using primers detailed in Supplementary Table 2. ORFs were identified using Meta-GeneMark [18] and the predicted protein products derived were functionally annotated using well-established pipelines [11]. *Aca*ML1 (JX507079) CDSs were used as queries to search across the predicted proviral ORFs from *A. caldus* strains using BlastP [19]. Queries producing no matches were searched by TBlastN [19] against the reconstructed proviral genomes. An E-value of 10(-5) was used as cut-off in both of the alignments. Orthologs in predicted proteomes of each strain were cross-compared using bidirectional BlastP, a reciprocal 95% identity [E,10(-10)] cut-off and in-house BioPerl scripts. Global similarity between queries and ORFs with significant hits was calculated using the Smith-Waterman algorithm implemented in the SSEARCH [20]. Sequence nucleotide identity between proviruses was performed with the Python3 module pyani (https://github.com/widdowquinn/pyani). Gene clusters synteny analysis was performed using genoPlotR package [21].

2.2. Bacterial strains, growth conditions and collection

Bacterial strains used in this study are described in Table 2. *A. caldus* strains were grown in modified 9 K [22] supplemented with 5 gr/L elemental sulfur at 40 °C and pH 3.5 or in mineral salts

Table 1
Occurrence of *Aca*ML1 in genomes of *Acidithiobacillus* spp. and *T. tepidarius*.

Species	Strains	Genome ID	G + C %	Size (Gb)	Viral Genes	*Aca*ML1
T. tepidarius	DSM3134	AUIS01	66.8	2,96	34	+/−
A. caldus	ATCC51756	NZ_CP005986	61.4	2,99	72	+
	DX	LZYE01	61.0	3,12	7	−
	ZBY	LZYF01	61.0	3,16	6	−
	ZJ	LZYG01	61.0	3,14	6	−
	MTH-04	LXQG01	61.3	2,87	7	−
	SM-1	NC_015850	60.9	0,32	8	−
	S1	LZYH01	60.9	2,79	9	−
A. ferrooxidans	ATCC23270	NC_011761	58.8	0,30	9	−
	ATCC53993	NC_011206	58.9	2,89	9	−
	DSM 16786	AFWEN	58.4	3,7	11	−
	GGI-221	AEFB01	58.6	3,17	6	−
	Hel18	LQRJ01	58.6	3,11	8	−
	YQH-1	LJBT01	58.6	3,11	2	−
	DLC-5	JNNH01	57.6	4,18	10	−
A. ferriphilus	BY0502	LVXZ01	56.8	2,98	6	−
A. ferrivorans	SS3	NC_015942	56.6	3,21	7	−
	PQ33	LVZL01	56.6	3,30	9	−
	YL15	MASQ01	56.6	3,00	7	−
	CF27	CCCS02	56.4	3,43	12	−
A. thiooxidans	ATCC 19377	AFOH01	53.2	3,02	4	−
	ZBY	LZYI01	53.2	3,79	10	−
	A01	AZMO01	53.1	3,82	11	−
	DMC	LWSB01	53.1	3,85	10	−
	BY-02	LWRZ01	53.1	3,81	10	−
	JYC-17	LWSD01	53.1	3,8	10	−
	A02	LWSA01	53.0	3,72	10	−
	DXS-W	LWRY01	52.9	3,95	12	−
	GD1-3	LWSC01	52.9	3,95	12	−
	Licanantay	JMEB01	52.8	3,9	9	−
	CLST	LGYM01	52.4	3,97	11	−
A. albertensis	DSM14366	MOAD01	52.6	3,50	10	−

Table 2
A. caldus strains and isolates used in this study.

Strain	Origin	DOI
ATCC 51756	Kingsbury coal spoil U.K.	10.1111/j.1574–6968.1983.tb00426.x
BC13	Birch Coppice Colliery, U.K.	10.1111/j.1574–6968.1983.tb00426.x
MNG	Arsenopyrite pilot plant, South Africa	10.1128/AEM.00864-08
F	Nickel pilot plant, South Africa	10.1128/AEM.00864-08
#6	Fairview mine, South Africa	10.1128/AEM.00864-08
CSH12	Continuous bioreactor, Australia	10.1128/AEM.00864-08
MEL1	Copper bioleaching heap, Chile	10.1016/j.resmic.2014.07.014

medium (MSM) with trace elements [23] supplemented with 5 mM potassium tetrathionate at 40 °C and pH 2.5, under aerobic conditions (200 rpm). Cultures were scaled up to a final volume of 500 ml and brought to late exponential growth phase to obtain sufficient biomass for downstream procedures. Cell and viral concentrates were obtained by tangential flow filtration using Minimate filtering cassettes (Pall) with 1000 kDa or 100 KDa pore-size. The subcellular fraction ('viral fraction') was processed as in Covarrubias et al. [13].

2.3. General DNA techniques

A. caldus stationary phase cultures to be used for nucleic acid purification, and/or for PFGE, were centrifuged at 10,000 rpm for 15 min at 4 °C. The harvested cells were stored as cell pellets at −80 °C for DNA isolation or were resuspended in 1X MSM for PFGE applications. DNA isolation and routine manipulations were carried out following standard protocols [24].

Polymerase chain reaction (PCR) products were amplified with Dreamtaq (Thermo Fisher Scientific) or the high-fidelity Herculase II fusion DNA polymerase (Agilent Technologies) depending on the experiment. Oligonucleotide primers used in each case are listed in Supplementary Table 2. Each PCR reaction contained 10 ng of template DNA, 0.5 mM of required primers and 0.2 mM of each deoxyribonucleotide in a volume of 25 ml of 1X PCR buffer containing 1.5 mM MgCl2. PCR conditions were as follows: initial denaturing step at 95 °C for 5 min followed by 28–30 amplification cycles (denaturation at 95 °C for 20 s, annealing at the appropriate temperature depending on the specific primer pairs for 20 s and elongation at 72 °C) and a final elongation step at 72 °C for 10 min. Amplicons to be used in nested PCR reactions or as Southern blot probes were purified from the agarose gels with the Mini Elute Gel Extraction Kit (Qiagen) and the Purelink Purification Kit (Thermo Fisher Scientific), respectively. DNA sequencing was carried out at Macrogen Inc.

2.4. Pulsed field gel electrophoresis assays

Cellular suspensions of 10(8) cells/ml were used in the preparation of gel plugs in a 1:1 (v/v) proportion with 1% Pulse Field Certified Agarose (BioRad). Plugs were incubated for 1 h at 50 °C in cell lysis buffer (1X TE; 50 mM Tris, 50 mM EDTA, pH 8.0) with 1% Sarcosyl and 20 mg/ml proteinase K under constant agitation (200 rpm). Plugs were washed 4 times in preheated (50 °C) distilled water and 3 times with preheated 1X TE buffer. The plugs were stored in 1X TE at 4 °C until use. Samples were digested with the 10U/ul *Xba*I (Thermo Fisher Scientific) in the recommended reaction buffer at 37 °C for 4 h. DNA fragments were separated in a 1% Pulse Field Certified Agarose with 0.5X Tris-Borate EDTA (TBE) running buffer at 14 °C. The pulsed field gel electrophoresis was performed in a CHEF-DR III apparatus (BioRad). Orthogonal-field alternation was achieved by varying the pulse time between 1 and 30 s for 13 h, with a rotation angle of 120° and voltage of 6 V/cm. The 5 kb ladder CHEF DNA size standard (BioRad) was used to estimate the molecular weight of the resolved DNA samples. Gels were stained with SYBR Safe DNA gel Stain™ (Thermo Fisher Scientific) for 30 min under constant agitation and washed for 10 min in distilled water before being photographed. Images were stored as TIFF files.

2.5. DNA probe labelling and southern blot

An internal fragment of the *gp39* gene (0.3 kb in length) was labelled with Biotin-14-dATP using the Thermo Fisher Scientific Nick Translation System according to manufacturer's recommendations. PFGE-resolved DNA was transferred onto positively charged nylon membranes Hybond-XL (General Electric) by capillary neutral transfer using the ULTRAhyb hybridization buffer at ambient temperature for 2 h. Transferred DNA was immobilized by UV cross-linking in a UV oven (BioLink) at 1134 J/m2. The membranes were placed into glass hybridization bottles and prehybridized in ULTRAhyb buffer (Thermo Fisher Scientific) at 42 °C for 30 min in a hybridization oven. The denaturated (90 °C for 10 min), labelled DNA probe was added to the prehybridization solution and incubated for 16 h at 42 °C under slow-speed rotation. The membranes were washed at 42 °C, twice with the high-stringency buffer from the NorthernMax Kit (Thermo Fisher Scientific) for 5 min and twice with low-stringency buffer from the same Kit for 15 min. The washed membranes were incubated with Streptavidine (Biolegend) diluted 1:5000 (v/v) in high-stringency buffer at 42 °C for 1 h under constant shaking. To remove Streptavidine excess, a final wash with this buffer was carried on at room temperature for 5 min. The membrane was revealed using the Pierce ECL western blotting substrate (Thermo Fisher Scientific) and conventional autoradiographic methods.

2.6. DNA damaging treatments to assess proviral excision

To produce DNA damage, and induce proviral excision, 10(10) total cells concentrated 100 fold were UV irradiated (200 J/m2, 254 nm for 3 min) or incubated with iron (III) sulfate (200 mM for 45 min) or mitomycin C (1 μg/ml for 8 h). Induced cells were recovered for one or two generation times (8–16 h, respectively) in fresh 1X MSM media (with potassium tetrathionate) at 40 °C and 200 rpm. All experiments were performed in triplicate. After recovery, cells were harvested for excision analysis using either Nested PCR or PFGE combined with Southern blot as described above.

2.7. Infection assays to assess host lysis and virus like particles production

A. caldus ATCC 51756 cultures were processed as in 2.6. to induce the *Aca*ML1 provirus lytic cycle. Cells were centrifuged to recuperate the spent supernatants and the supernatants processed as in section 2.2 to obtain the virus-enriched concentrates. Equivalently treated strain #6 cultures (virus negative) and untreated *A. caldus* ATCC 51756 supernatant concentrates served as controls. Each

supernatant concentrate (100X) was added to an exponentially growing culture of *A. caldus* strain #6 (phage sensitive) 24 h after inoculation [10(5) cell/ml]. Cell counts were recovered every 20 h and plotted against time. The Live/Dead BacLight kit (Thermo Fisher Scientific) was used to assess the relative numbers of bacteria with intact plasma membranes from those with compromised membranes.

Virus enriched supernatants were filtered through a 0.022 μm Anopore Inorganic Membrane (Anodisc) and observed by epi-fluorescence microscopy after staining with SYBR Green I (Thermo Fisher Scientific), according to established procedures [24]. In selected samples the viral like particles (VLPs) were precipitated by the addition of NaCl (1 M) and polyethylene glycol 8000 (10%) [25], and analysed by electron microscopy using a Tecnai Spirit operated at 120 kV and a 2 K × 2 K CCD camera (Magnification 48,500 × −4,95 Å/pixel). Specimens were prepared for EM using the conventional negative staining procedure with 2% uranyl-acetate.

3. Results and discussion

3.1. Search for AcaML1-like viruses in sequenced acidithiobacilli

To investigate whether *Aca*ML1-like lysogenic proviruses were present in other strains of the species, their occurrence was evaluated in all publically available genome sequences of the taxon (Table 1; Supplementary Table 1). *Thermithiobacillus tepidarius* was also included in the analysis as the closest ancestor sequenced. BLASTp and HMM searches failed to reveal the presence of viral proteins related to *Aca*ML1 or other known bacteriophages in the genomes of the *Acidithiobacillus* spp. strains analysed. A few viral proteins distantly related to the *Aca*ML1 prophage, encoding integrases, excisionases and phage-type regulators were however widespread in all genomes investigated (Supplementary Table 1). These proteins typically conform viral gene modules of larger mobile genetic elements of the ICE type (Integrative Conjugative Elements), where they play a role in the control and regulation of integration and excision [26,27]. A DNA segment potentially encoding a full provirus was predicted in the genome of *T. tepidarius*. At the protein level, this provirus presents similarities with the *A. caldus Aca*ML1 provirus, with equivalent gene products related to the infectious cycle of the virus (excisionase, recombinases), lysis of the host (lysozyme), and a structural gene (capsid protein), but it lacks orthologous genes related to the morphogenesis of the base plate and tail (Supplementary Table 1). According to available genomic evidence, *Aca*ML1-like proviruses are confined to the mesothermophilic acidithiobacilli, *T. tepidarius* and *A. caldus*.

3.2. Prevalence of AcaML1-like proviruses in A. caldus strains and industrial isolates

To investigate whether lysogenic proviruses like *Aca*ML1 were present in other strains of *A. caldus*, occurrence of *Aca*ML1 signatures was evaluated in a collection of seven *A. caldus* isolates (Table 2) using pulsed-field gel electrophoresis (PFGE) combined with Southern blot. Hybridization of a biotinilated probe, directed to the viral gene *gp39* (encoding the viral capsid), to the PFGE fingerprints of all seven isolates (Fig. 1A) revealed the presence of lysogenic proviruses similar to *Aca*ML1 in four of the strains tested, namely the type strain (ATCC 51756), strain F, BC-3 and CSH12 (Fig. 1B). Under the conditions evaluated a single band was identified in the four positive lanes, implying that only the integrated form of the virus could be detected in the type strain and the other *Aca*ML1 positive isolates. PFGE-based size estimations indicate that all these proviruses are similar in size to *Aca*ML1 (aprox. 60 kb). This result was confirmed by PCR amplification of selected gene markers encompassing the different gene modules of the *Aca*ML1 provirus genome. All 19 gene markers tested could be amplified from total DNA preparations obtained from strains BC13, F and CSH-12 indicating, as well, global conservation of the gene content of the *Aca*ML1-like proviruses present in *A. caldus* strains (Fig. 2).

3.3. Comparative genomics of AcaML1-like proviruses from A. caldus strains

*Aca*ML1 proviruses present in *A. caldus* strains F, BC-13 and CSH-12 were PCR amplified using high-fidelity and high-processivity Herculase II fusion DNA polymerase and sequenced at Macrogen on a module-to-module basis using primers indicated in Supplementary Table 2. Sequences obtained (MH142219, MH142220) were aligned against the *Aca*ML1 genome (JX507079) and cross-compared to each other. The putative provirus identified in the whole genome sequence of the neutrophilic relative of *A. caldus*, *T. tepidarius*, was included in this analysis for comparative reasons. Results of the whole genome multiple alignments are shown in Fig. 3.

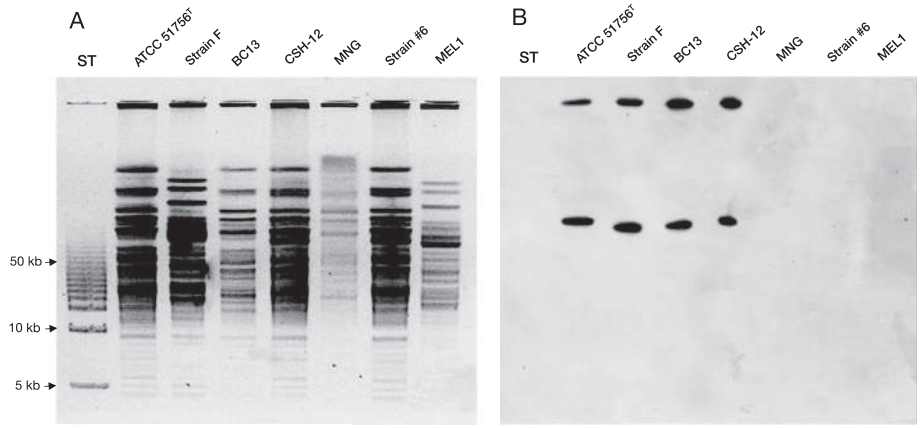

Fig. 1. Prevalence of *Aca*ML1-like prophages in *A. caldus* strains. **A**. PFGE profiles of *Xba*I digested genomic DNA of *A. caldus* strains. Staining was done with SYBR Green I. The molecular weight standard used is the 5 kb ladder DNA size standard (BioRad). **B**. Southern blot of the gel in panel A using a 300 bp long probe specific for the gene *gp39* encoding the viral capsid.

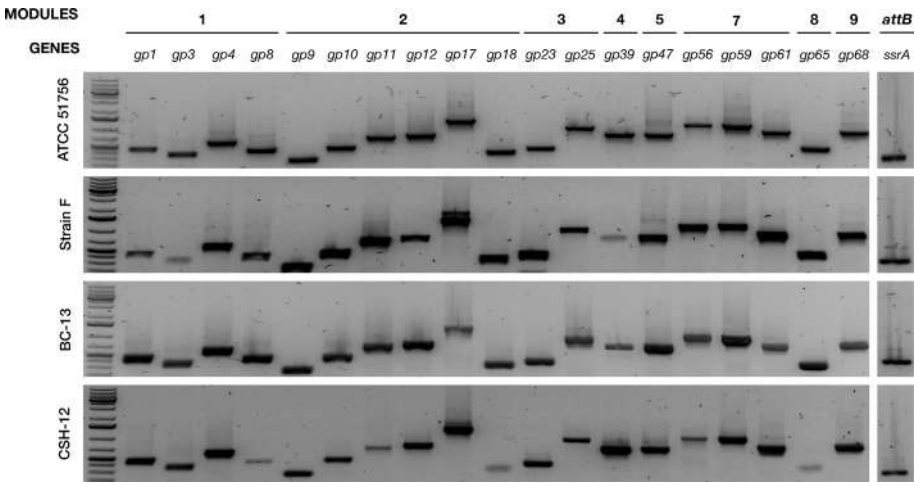

Fig. 2. Occurrence of *Aca*ML1 genes in *A. caldus* strains. Different predicted viral genes spanning gene modules 1 to 9 were amplified by end point PCR using primers indicated in Supplementary Table 1. The housekeeping gene *ssrA*, predicted integration site for the provirus (*attP* site), was used as PCR positive control. The molecular weight standard used is the Thermo Scientific GeneRuler 1 kb DNA Ladder.

The genomic divergence between *Aca*ML1-like proviruses and the predicted provirus integrated in the genome of *T. tepidarius* (26.1%) is slightly lower than the genomic divergence existing between the two species (31.1%, value calculated for both type strains), being thus coherent with the phylogenetic distance between hosts. Average nucleotide identity between the sequenced *A. caldus* proviruses is higher than 99.9%, highlighting the conservation of the provirus at DNA sequence level between acidophilic strains. Such conservation is puzzling, considering the geographical origin of the strains (Table 2). Recent evidence from studies on the aerosolization of sand-dust and organic aggregates in sea sprays, showing high fluxes and high deposition rates for viruses across the free atmosphere [28], could provide an explanation for this observation. However, in the absence of additional information on the genetic background of the host strains and/or genomic data of the *A. caldus* populations present at each site it is unfeasible to draw further conclusions on this respect.

The genomes of the *A. caldus* lysogenic proviruses had an average length of 59.4 kb, including the redundant ends and encoded 72 ± 2 protein coding ORFs. The *T. tepidarius* provirus was predicted to be significantly larger (84.6 kb) with 94 predicted protein coding genes. In both cases a ~10-bp direct repeat flanking the proviruses was identified (DR_{Aca}: CCACCATTTA; DR_{Ttp}: CCATCGACACG). The GC content of the *Aca*ML1-like proviruses averaged 65.5%, being higher than those of the host genomes (61.72%; [26]). A lower G + C content of phage genomes in comparison to those of their hosts is a widespread phenomenon, for example found for *Escherichia coli* phages T4 and JS98 [29]. In contrast, increased G + C content has seldom been reported (e.g. [30]).

Predicted *orfs* cover 96.7% of the sequence and are located on different strands. One group of genes points leftward (32 genes) while the rest of the prophage (40 genes) is oriented rigthward with respect to the predicted origin of replication of the host genome. The “lefthand” side of the genome groups candidate early genes, while the ORFs on the “righthand” side are predicted to be intermediate and late transcription genes. Results from both BlastP searches and curation using phage-dedicated resources (see Materials and Methods) indicated that 59.4%, of the predicted proteins in *Aca*ML1 have significant sequence similarity to proteins of known function or unknown function but of viral origin, including nucleotide metabolism, DNA replication/recombination, components of viral particles and others. The characteristics of these ORFs and their corresponding predicted proteins are described in Supplementary Table 1. Genes with functional assignments were named accordingly. Gene function assignment and

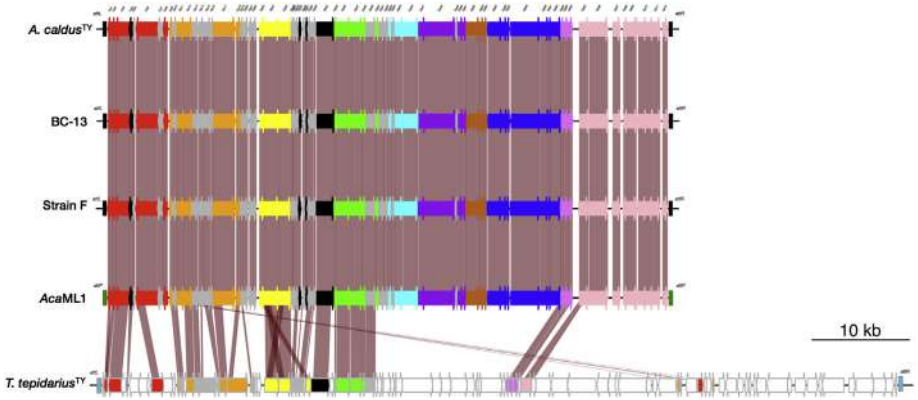

Fig. 3. Whole genome multiple alignment of *Aca*ML1-like proviruses present in the *Acidithiobacilaceae*. Reconstructed proviral genomes were aligned against the *Aca*ML1 genome (JX507079) and the *T. tepidarius* predicted provirus (conting AUIS01000006, coordenates 2854 to 86496). Similarity between predicted proteins was calculated using BlastP (Cut-off values: E−05; >30% identity) and displayed graphically with genoPlotR.

architecture analysis allowed a tentative subdivision of the *Aca*ML1 provirus genome into modules as follows: replication/regulation, packaging, structure/morphogenesis, lysis and lysogeny (Fig. 3). This modular organization is typical for temperate phages from *Myoviridae* family [31].

3.4. Gene products from the leftward region

Divergently oriented gene clusters 1 and 2 encode signature proteins implicated in lysogeny establishment and regulation and control of the lysogeny-lytic switch. Gene module/cluster 1 contains 8 ORFs (*gp1-gp8*) and includes genes coding for a putative transcriptional regulator (Gp1), the small (Gp2) and large (Gp3) subunits of a site-specific recombinase (pfam00239), similar to the integrases of P2 temperate phages [32] that catalyze unidirectional integration of the prophage into the host genome and an excisionase (pfam06806), necessary to enable excisive recombination by the phage integrase [33]. Genes *gp4-gp7* have no predicted function. Divergently oriented to *gp8* is another ORF (*gp9*) which encodes a putative transcriptional regulator, baring the HTH motif of MerR family regulators. This gene is the best candidate for the cI-like phage repressor and forms part of the 2nd cluster/module of *Aca*ML1 together with other 15 genes (*gp9 – gp26*) with ascribed functions to phage replication and regulation. The deduced amino acid sequence of ORF *gp11* bares significant similarity to a phage antirepressor protein (COG3617) with a conserved N-terminal BRO domain (pfam02498). Presumably, Gp11 is implicated in the maintenance of lysogeny by modulating the activity of the repressor [34], although other roles in replication and/or transcription have also been proposed for proteins carrying BRO domains [35]. Open reading frame *gp17* encodes for a phage primase (P4 family), which is predicted to synthesize the RNA primer at the origin of replication required for replication initiation and *gp18* an RNAse H endonuclease predicted to degrade the RNA primer during replication. Gp23 contains a CRO domain (pfam09048); members of this family of proteins typically act as early gene expression repressors and are required to reach the lytic growth stage [36]. Two *orfs gp24* and *gp25*, encoding two highly similar methyltransferases, are probably related to DNA modification associated to phage replication.

3.5. Gene products from the rightward region

The right part of the genome represents the “late region” and is made up of gene products playing a role in viral particle formation and assembly, DNA packaging and host lysis for viral particles release. The first gene module in this region of the genome (*gp34 – gp42*) encodes several functions directly related with the formation of the procapsid shell and the genome packaging of complex double-stranded DNA viruses into the viral head. Proteins Gp34 and Gp35 were identified as the small (GpNu) and the large subunits (GpA) of a terminase (pfam05876), respectively. The amino acid sequence which exhibits weak similarity to terminases of other phages, contains two partially conserved domains: one in the N-terminal half for a homing endonuclease and one in the C-terminal part for a terminase [37]. The terminase hetero-multimeric complex is responsible for packaging DNA in lamboid like phages such as P2, through recognition, cleavage and translocation of DNA concatemers into de viral head. Immediately adjacent to the terminase is *gp36* which encodes for the phage protein GpW (pfam02831), proven necessary for the stabilization of DNA within the phage head and for attachment of tails onto the head during morphogenesis [38]. As in other known prophages the gene encoding the portal protein (Gp37, pfam05136), the serine protease responsible for procapsid maturation (Gp38, pfam01343, [39]), the head decoration protein D (Gp39, pfam02924, [40]) and the major capsid protein E (Gp40, pfam03864) follow the terminase. Although portal proteins are not globally conserved, Gp37 is highly similar to the portal protein of *Comamonas aquatica* (91% similarity), an aquatic *Burkholderiales* bacterium. In other phages the portal protein serves as the entry and exit door for phage DNA during genome packaging and ejection, as a heads-tail connector and possibly as a nucleator for capsid assembly [40]. The major capsid protein belongs to the E superfamily, typically found in lamboid like phages like P2 [41]. P2 forms a 60 nm icosahedral capsid, with T = 7 symmetry, from 415 copies of capsid protein derived from gene product N (gpN) to package its 33.5 kb DNA [42]. A number of experimental studies have suggested that virion sizes are a function of genome sizes and that an allometric relationship between the genome length and the volumen of the virion has been recognized [43]. Since *Aca*ML1s genome is 26.1 kb longer than P2. *Aca*ML1 virion size is predicted to be somewhat larger.

Blast analysis of the predicted phage products revealed several proteins as being involved in phage tail morphogenesis, recognition of the host and genome delivery upon infection. Four gene products of *Aca*ML1 could be associated to the baseplate formation including orthologs for the genes encoding the spikes of the tails end (Gp43: protein V pfam04717, [44]), the wedge protein (Gp45: protein W, pfam04965, [45]), the receptor-binding protein (Gp46: protein J, pfam04865, [46]) and assembly protein I (Gp47: protein I, pfam09684) of P2-type phages. Gp45 has a conserved lysozyme domain, similar to that of T4-like phages protein GP5, predicted to aid in penetration of the peptidoglycan layer during the initial infection process [47]. ORFs *gp48-gp53* encode predicted tail fibers sharing some degree of similarity with P2 phages important for adsorption of phage to the outer membrane of the bacterial cell [48]. Gp49 is a phage-associated protein with predicted endosialidase activity. This enzymatic activity has been found in the tail spikes of certain phages where it catalyzes the degradation of the bacterial capsule [49]. Gp53 shares distant similarity in its N-terminal domain with proteins of the lectin C superfamily. Lectin-like domains have been found associated to receptor binding proteins in bacteriophages where they confer binding variability [50]. These enzymes may be tail components involved in adsorption of the phage or/and DNA injection. Next to the tail fibers encoding genes are two genes, *gp56* and *gp57*, encoding the major tail sheath FI (pfam04984) and the tail tube protein FII (pfam04985) of *Myoviridae* bacteriophages [51], respectively. Gp58 encodes a small tail protein (protein E, pfam06528) that stabilizes the tail in P2 bacteriophages [52]. Distal to this cluster is gene *gp59*, encoding the protein that determines tail size by working as a template for measuring tail length during assembly [53] better known as tape measure protein (pfam10145). It consists of 1026 residues and is one of the longest open reading frames of the *Aca*ML1 genome. The protein has a well-conserved core region that is often found in phage tail tape measure proteins. Assuming 1.5 Å per amino acid residue [53], the calculated length of the *Aca*ML1 tail would be 154 nm. The function of the distal genes in this cluster *gp60-gp62* is less clear. Gp60 corresponds to protein U (pfam06995) of P2-like phages and is likely to play a role in tail assembly [51]. Gp61 is similar to protein X of P2 (pfam05489) and bares a LysM domain (pfam01476) with predicted peptidoglycan degrading activity [54]. Gp62 shows similarity with protein D (pfam05954), another phage associated protein with unclear function.

The contiguous cluster of four genes encodes proteins involved in programmed lysis. As other dsDNA phages, *Aca*ML1 seems to make use a dual lysis system consisting of a holin and an endolysin [55,56]. Protein Gp63 encodes a low complexity, highly conserved, hydrophilic 94 amino acids long protein with a single transmembrane region (amino acid position 66–85) that places the N-

teminus in the cytoplasm and the C-terminal end of the protein in the periplasm, resembling other known class III holins [57]. Holins facilitate the release of phage particles from infected bacteria during viral-induced cell lysis [58]. A second small ORF (*gp64*), downstream of the putative holin gene, encodes a 158 aa long protein of unknown function. Orthologs of this gene frequently concur with *gp63* orthologs and the predicted protein products are extremely well conserved (e.g. in *T. tepidarius*). Predicted protein Gp65 is 160 amino acids long and contains the signature of known phage endolysins of the muramidase type (pfam00959) and lack predicted secretory signals. Muramidases cleave the glycosidic beta 1,4-bonds between the N-acetylmuramic acid and the N-acetylglucosamine of the peptidoglycan from outside the cytoplasm, thereby enabling progeny virions to be liberated [59]. Because of their unique ability to cleave peptidoglycan in a generally species-specific manner (shared homology is not found within the catalytic domain), endolysins provide a means of selective and rapid killing of bacteria with no effect on the rest of the native microflora. Access to peptidoglycan depends on the small hydrophobic holins, which enable endolysin molecules to cross the inner membrane. These properties confer a biotechnological potential to both types of proteins.

The distal gene module in *Aca*ML1 contains three genes (*gp66 – gp69*) encoding a DNA restriction-modification system and two insertion sequences. Gp66 (pfam00145) is a predicted site-specific DNA methylase that modifies cytosine to 5-methylcytosine. In turn, Gp67 and Gp68 encode the two subunits of a predicted McrBC (pfam10117) restriction enzyme cutting DNA in multiple positions within two distant locations (between 40 pb and 3 Kb apart) baring the 5-methylcytosine modification preceded by a purine (RmC[N]RmC) [60]. Presence of RM-systems in phages and other mobile genetic elements is frequent, playing a relevant role in the stabilization of the mobile element within the genome [61] or being target of frequent horizontal transfer between bacteria [62]. Similar to other cases appointed in the literature (e.g. [63]), the RM-system of *Aca*ML1 is flanked by transposases of the IS5 and IS21 families (*gp70-74*) (pfam05598 and pfam13751). These genes could facilitate transfer of the RM gene cassettes between the host chromosome and horizontal transfer vector such as phages [62].

3.6. Induction of replication intermediaries of AcaML1-like proviruses by DNA damage

Based on the above prevalence and genomic analyses, attempts to establish the capacity of the proviruses to produce functional viral progeny were made in all four *Aca*ML1 positive (*A. caldus* ATCC 51756, BC13, CSH12, F). As initial strategy mitomycin C, a DNA damaging agent that produces DNA-adducts and lesions that trigger the SOS response, was used to induce proviral excision. Standard mitomycin C concentrations (1 μg/ml) for this type of

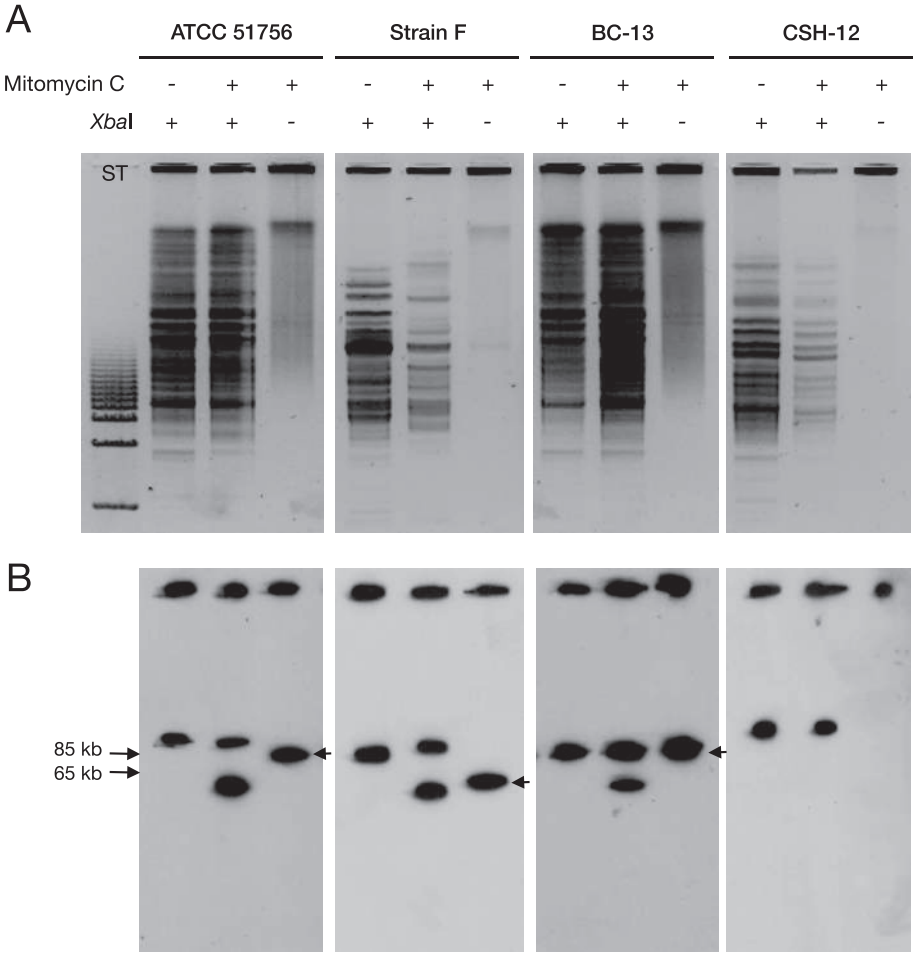

Fig. 4. Excision of *Aca*ML1-like proviruses from the chromosome of *A. caldus* strains upon mitomycin C treatment. **A**. Pulsed-field gel electrophoresis profiles of *Xba*I digested genomic DNA of *A. caldus* strains. Staining was done with SYBR Green I. Molecular weight standard used is the 5 kb ladder DNA size standard (BioRad). **B**. Southern blot of the gel in the left panel using a 300 bp long probe specific for the gene *gp39*, encoding the viral capsid. Excised forms of *Aca*ML1 are indicated with arrows in each panel.

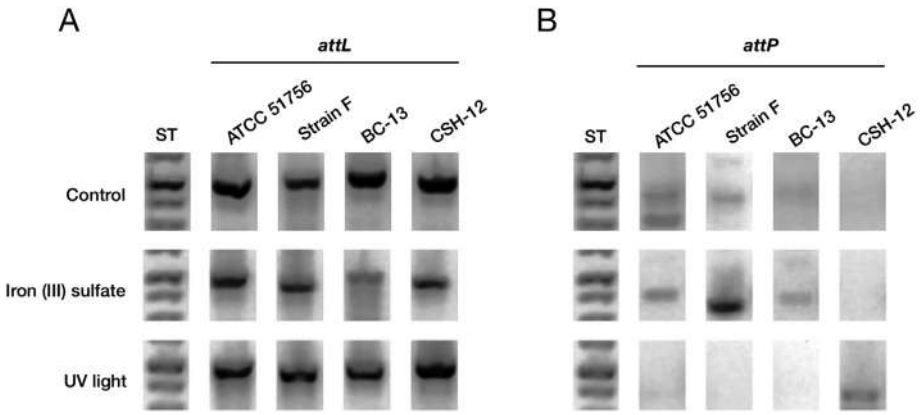

Fig. 5. Iron (III) sulfate and UV exposure as elicitors of excision of *Aca*ML1-like proviruses from the chromosome of *A. caldus* strains. The chromosomal junction formed to the left of the integrated provirus (*attL*) and the viral attachment site (*attP*) were amplified by nested PCR using primers indicated in Supplementary Table 2. DNA damage was induced with iron (III) sulfate (200 mM for 45 min) or UV light exposure (200 J/m^2, 254 nm for 3 min). The molecular weight standard used is the Thermo Scientific GeneRuler 1 kb DNA Ladder.

assays were used [64]. Excision was evaluated through PFGE and Southern blot (as in 3.2) using both *Xba*I digested and undigested DNA (Fig. 4). The results obtained indicate that the proviruses present in the chromosome of *A. caldus* ATCC 51756, strain F and BC13 respond to the induction trigger and produce an episomal form of the virus ranging in size between 65 and 85 Kb. No episomal form of *Aca*ML1 could be identified in strain CSH-12, despite of the fact that the integrated form was confirmed in both mitomycin C treated and untreated cells.

Other DNA damaging agents, iron (III) sulfate and UV exposure, were also tested as inducers of provirus excision. Capacity of *Aca*ML1 to excise out of the respective host genomes upon treatment with these DNA damage elicitors was assayed by nested PCR. Targeted regions (shown in Supplementary Figure 1) included the chromosomal junctions formed at both sides of the integrated provirus (*attL* and *attR*) and the viral and bacteria chromosomal attachment sites (*attP* and *attB* sites, respectively). Representative results are shown in Fig. 5 and full results are summarized in Table 3. Positive amplicons of the *attL* and *attP* products were obtained in 200 mM ferric iron treated cell of the type strain, strain F and BC-13. In turn, UV light at the exposure dose tested (200 J/m2, 254 nm for 3 min) only elicited excision in the CSH-12 strain (Fig. 5B; Table 3). These results demonstrate that all four *A. caldus* strains evaluated carry proviruses capable of producing the episomal replication intermediary upon excisive induction.

3.7. *Infection assays using AcaML1-like proviruses from* A. caldus *strains*

To evaluate if the induced viruses prompted host lysis and produced viral particles, infections assays were set up. No lawn-based plate assay is available for *A. caldus*, thus lytic capacity of *Aca*ML1 could not be tested using conventional plaque formation assays. To assess host killing, the effect of Viral-Like Particles (VLPs) on growth decline (with cell lysis) and/or loss of viability (without cell lysis) was evaluated using conventional liquid cultures and modified 9 K media supplemented with elemental sulfur as energy source. *A. caldus* ATCC 51756 was used as source of *Aca*ML1 VLPs. In order to avoid the possibility of superinfection related immunity issues, *A. caldus* strain #6 (which proved negative to the presence of lysogenic *Aca*ML1 in 3.2) was chosen as virus sensitive strain for the infection assays. Production of VLPs was triggered by exposing *A. caldus* ATCC 51756 mid log cultures to ferric iron for 8 h (as in 2.6). VLPs were recovered from TFF-concentrated culture supernatants, treated with NaCl (1 M) and polyethylene glycol 8000 (10%) to precipitate the virus, and visualized using epifluorescence microscopy after SYBR-Green staining (Fig. 6B). Mid log cultures of strain #6, grown for 24 hs (I) in fresh 9 K media (F9K) were challenged with the *Aca*ML1 VLPs concentrates or with spent media concentrates of control cultures (ferric iron-treated *A. caldus* strain #6). To uncover variations in growth and/or viability with respect to untreated and control samples, lysis of the sensitive strain was assessed by direct cell counts and phase contrast microscopy, and by Live-Dead staining and epifluorescence microscopy.

Spent culture supernatant from ferric iron-treated *A. caldus* ATCC 51756 cultures had very little effect on the cell counts of fresh cultures of strain #6 with respect to the controls at all times tested (Fig. 6A). Similarly, only minor effects were observed on the viability (V) of strain #6 (Fig. 6C; Supplementary Figure 2), with 38% of cell death at 48 h after treatment with virus-positive spent media concentrates (SC9KTY) and 30% death in similarly treated the cells with virus-negative spent media concentrates (SC9K6). Supporting statistics for these tests are provided in Supplementary Table 3. These results suggest that *Aca*ML1 preparations had no or very little lytic effect of the chosen sensitive strains. Similar results were obtained with other virus-negative *A. caldus* strains (MNG, MEL1, data not shown). Other forms of resistance to *Aca*ML1, other than superinfection, cannot presently be discarded. Seeking for an explanation to these results, electron microscopy was used to inspect the *Aca*ML1 concentrates. In all preparations evaluated icosahedral tail-less virus-like particles were observed (Fig. 6C). These results explain the neglectable levels of lysis observed in the

Table 3
PCR-based detection of the episomal *Aca*ML1 form in *A. caldus* strains upon treatment with DNA damage elicitors.

attL-attR/attB-attP Strain	Control	Mitomycin C	Iron(III) sulfate	UV light
ATCC 51756	+/*	+/*	+/+	+/−
F	+/*	+/*	+/+	+/−
BC-13	+/−	−/+	+/+	+/−
CHS-12	+/−	+/*	++/−	+/+

a. Mitomycin C (1 μg/ml; 8 h).
b. Iron (III) sulfate (200 mM; 45 min).
c. UV (200 j/m^2; 3 min).
(+) *attR* and *attL* present; (−) *attR* and *attL* absent; (*) one *att* site of the pair absent.

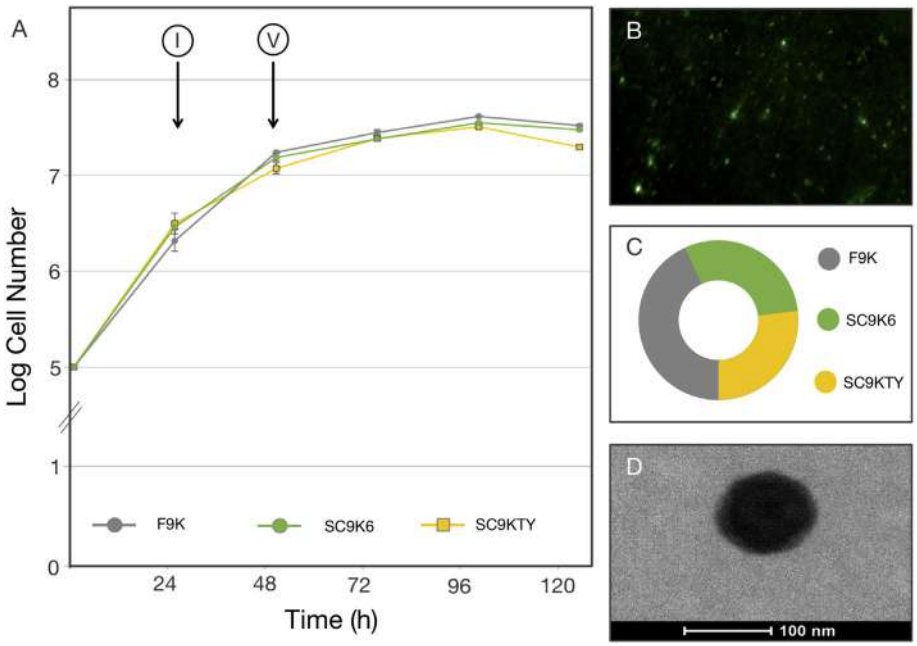

Fig. 6. *Aca*ML1 infection assays using *A. caldus* sensitive strain #6. **A**. Growth curve of strain #6 cultures challenged with spent media concentrates of *A. caldus* ATCC 51756 (SC9KTY) and # 6 (SC9K6) ferric iron-treated cultures or with spent media concentrates of strain # 6 untreated cultures (F9K). Infection (I) occurred 24 h after set up of the experiment. **B**. Fluorescence microscopy of SYBR Green I stained *A. caldus* ATCC 51756 spent media concentrates after iron (III) sulfate induction showing fluorescent viral-like particles (VLPs). **C**. Viability counts obtained 48 h after onset of the assay (V) using Live-Dead staining of treated (SC9KTY), untreated (F9K) and control (SC9K6) samples. **D**. Electron microscopy of VLP recovered form *Aca*ML1-enriched concentrates.

infection assays and cast doubt on the capacity of the *Aca*ML1 provirus to assemble and maintain functionally stable viral particles under the acidic conditions tested. A definite answer to this question will require additional research in the near future.

4. Conclusions

Mesothermophilic *Acidithiobacilliaceae* family members, *A. caldus* and *T. tepidarius*, host in their genomes lysogenic proviruses with redundant ends, ranging in size between 59 and 85 kb. Comparative genomic analyses performed herein revealed partial conservation of the gene content and gene modules architecture between *A. caldus* and *T. tepidarius* proviruses, suggesting that their organization and composition is preserved for functional reasons. Among *A. caldus* strains the integrated proviruses had only modest prevalence (30%), but were highly conserved at both the amino acidic and DNA sequence levels. In all four *Aca*ML1-provirus positive strains (ATCC 51756, BC13, F and CSH12) the proviral genome is organized in a leftward region that encodes signature proteins implicated in lysogeny establishment and regulation and control of the lysogeny-lytic switch, and a rightward region encoding genes predicted to play a role in viral particle formation and assembly, DNA packaging and host lysis for viral particles release.

Induction assays, performed under DNA-damaging conditions triggering the SOS-response, demonstrated that the *Aca*ML1-positive strains carry proviruses capable of producing the episomal replication intermediary upon excision, and that DNA-containing VLPs are produced. Despite this fact, under the conditions evaluated (acidic) the VLPs obtained from *A. caldus* ATCC 51756 could not produce productive infections of a candidate sensitive strain (#6) nor trigger it lysis. The neglectable levels of lysis observed in the infection assays correlate to the presence of icosahedral tail-less virus-like particles in all preparations evaluated by transmission electron microscopy. This aspect casts doubt on the capacity of the *Aca*ML1 provirus to assemble and maintain functionally stable viral particles under the conditions tested, which emulated the acidity of their usual habitats. Further studies are required to define if the lack of tails is an experimental artefact of the current experimental design or an adaptive biological response.

Conflicts of interest

None declared.

Acknowledgements

This work was supported by Basal AFB 170004 and by Fondecyt grants # 1181251 and 1140048. PCC and AM were recipients of graduate training CONICYT fellowships. We thankfully acknowledge Douglas Rawlings and Shelly Dean for providing the bacterial strains.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.resmic.2018.07.005.

References

[1] Dellas N, Snyder JC, Bolduc B, Young MJ. Archaeal viruses: diversity, replication, and structure. Annu Rev Virol 2014;1:399–426.
[2] Pina M, Bize A, Forterre P, Prangishvili D. The archeoviruses. FEMS Microbiol Rev 2011;35:1035–54.
[3] Dick G, Andersson A, Baker B, Simmons S, Thomas B, Yelton AP, et al. Community-wide analysis of microbial genome sequence signatures. Genome Biol 2009;10:R85.
[4] Prangishvili D. Archaeal viruses: living fossils of the ancient virosphere? Ann N Y Acad Sci 2015;1341:35–40.
[5] Krupovic M, White MF, Forterre P, Prangishvili D. Postcards from the edge: structural genomics of Archaeal viruses. Adv Virus Res 2012;2:51–81.
[6] Prangishvili D, Quax TE. Exceptional virion release mechanism: one more surprise from archaeal viruses. Curr Opin Microbiol 2011;14:315–20.

[7] Prangishvili D, Garrett RA, Koonin EV. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res 2006;117:52–67.
[8] Evans DJ. Exploitation of plant and archaeal viruses in bionanotechnology. Biochem Soc Trans 2009;37:665–70.
[9] Ward TE, Bruhn DF, Shean ML, Watkins CS, Bulmer D, Winston V. Characterization of a new bacteriophage which infects bacteria of the genus *Acidiphilium*. J Gen Virol 1993;74:2419–25.
[10] Arsène-Ploetze F, Koechler S, Marchal M, Coppée JY, Chandler M, Bonnefoy V, et al. Structure, function, and evolution of the *Thiomonas* spp. genome. PLoS Genet 2010;6:e1000859.
[11] Tapia P, Flores FM, Covarrubias PC, Acuña LG, Holmes DS, Quatrini R. Complete genome sequence of temperate bacteriophage AcaML1 from the extreme acidophile *Acidithiobacillus caldus* ATCC 51756. J Virol 2012;86:12452–3.
[12] Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 2008;320:1047–50.
[13] Covarrubias P, Muñoz R, Bobadilla-Fazzini R, Martinez P, Quatrini R. Are there viruses in industrial bioleaching econiches? Solid State Phenom 2017;262:521–5.
[14] Johnson DB. Biomining-biotechnologies for extracting and recovering metals from ores and waste materials. Curr Opin Biotechnol 2014;30:24–31.
[15] Cárdenas JP, Quatrini R, Holmes DS. Progress in acidophile genomics. In: Quatrini R, Johnson DB, editors. Acidophiles: life in extremely acidic environments. UK: Caister Academic Press; 2016. p. 179–97.
[16] Quatrini R, Ossandon FJ, Rawlings DE. The Flexible Genome of Acidophilic Prokaryotes. In: Quatriniand R, Johnson DB, editors. Acidophiles: life in extremely acidic environments. UK: Caister Academic Press; 2016. p. 199–220.
[17] Nuñez H, Moya-Beltrán A, Covarrubias PC, Issotta F, Cárdenas JP, González M, et al. Molecular systematics of the genus *Acidithiobacillus*: insights into the phylogenetic structure and diversification of the taxon. Front Microbiol 2017;8:30.
[18] Besemer J, Lomsadze A, Borodovsky M. GeneMark: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 2001;29:2607–18.
[19] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.
[20] Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. Genomics 1991;11:635–50.
[21] Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 2010;26:2334–5.
[22] Silverman MP, Lungren DG. Studies on the chemoautotrophic iron bacterium *Ferrobacillus ferrooxidans*. I. An improved medium and a harvesting procedure for securing high cell yields. J Bacteriol 1959;77:642–7.
[23] Dopson M, Lindström EB. Potential role of *Thiobacillus caldus* in arsenopyrite bioleaching. Appl Environ Microbiol 1999;65:36–40.
[24] Nieto PA, Covarrubias PC, Jedlicki E, Holmes DS, Quatrini R. Selection and evaluation of reference genes for improved interrogation of microbial transcriptomes: case study with the extremophile *Acidithiobacillus ferrooxidans*. BMC Mol Biol 2009;10:63.
[25] Noble RT, Fuhrman JA. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. Aquat Microb Ecol 1998;14:113–8.
[26] Acuña LG, Cárdenas JP, Covarrubias PC, Haristoy JJ, Flores R, Nuñez H, et al. Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus caldus*. PLoS One 2013;8:e78237.
[27] Bustamante P, Covarrubias P, Levicán G, Katz A, Tapia P, Bonnefoy V, et al. ICEAfe1, an actively excising integrative-conjugative element from the biomining bacterium *Acidithiobacillus ferrooxidans*. J Mol Microbiol Biotech 2012;22:399–407.
[28] Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA. Deposition rates of viruses and bacteria above the atmospheric boundary layer. ISME J 2018;12:1154–62.
[29] Zuber S, Ngom-Bru C, Barretto C, Bruttin A, Brüssow H, Denou E. Genome analysis of phage JS98 defines a fourth major subgroup of T4-like phages in *Escherichia coli*. J Bacteriol 2007;189:8206–14.
[30] Zheng Q, Zhang R, Xu Y, White III RA, Wang Y, Luo T, et al. A marine inducible prophage vB_CibM-P1 isolated from the aerobic anoxygenic phototrophic bacterium *Citromicrobium bathyomarinum* JL354. Sci Rep 2014;4:7118.
[31] Botstein D. A theory of modular evolution for bacteriophages. Ann N Y Acad Sci 1980;354:484–90.
[32] Thorpe HM, Smith MC. *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci USA 1998;95:5505–10.
[33] Cho EH, Gumport RI, Gardner JF. Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. J Bacteriol 2002;184:5200–3.
[34] Summer EJ, Enderle CJ, Ahern SJ, Gill JJ, Torres CP, Appel DN, et al. Genomic and biological analysis of phage Xfas53 and related prophages of *Xylella fastidiosa*. J Bacteriol 2010;192:179–90.
[35] Bideshi DK, Renault S, Stasiak K, Federici BA, Bigot Y. Phylogenetic analysis and possible function of bro-like genes, a multigene family widespread among large double-stranded DNA viruses of invertebrates and bacteria. J Gen Virol 2003;84:2531–44.
[36] Atsumi S, Little JW. Role of the lytic repressor in prophage induction of phage λ as analyzed by a module-replacement approach. Proc Natl Acad Sci USA 2006;103:4558–63.
[37] Rao VB, Feiss M. The bacteriophage DNA packaging motor. Annu Rev Genet 2008;42:647–81.
[38] Maxwell KL, Yee AA, Booth V, Arrowsmith CH, Gold M, Davidson AR. The solution structure of bacteriophage lambda protein W, a small morphogenetic protein possessing a novel fold. J Mol Biol 2001;308:9–14.
[39] Medina E, Wieczorek D, Medina EM, Yang Q, Feiss M, Catalano CE. Assembly and maturation of the bacteriophage lambda procapsid: gpC is the viral protease. J Mol Biol 2010;401:813–30.
[40] Doan DN, Dokland T. The gpQ portal protein of bacteriophage P2 forms dodecameric connectors in crystals. J Struct Biol 2007;157:432–6.
[41] Rajagopala SV, Casjens S, Uetz P. The protein interaction map of bacteriophage lambda. BMC Microbiol 2011;11:213.
[42] Dokland T, Lindqvist BH, Fuller SD. Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2-P4 bacteriophage system. EMBO J 1992;11:839–46.
[43] Cui J, Schlub TE, Holmes CE. An allometric relationship between the genome length and virion volume of viruses. J Virol 2014;88:6403–10.
[44] Haggard-Ljungquist E, Jacobsen E, Rishovd S, Six EW, Nilssen O, Sunshine MG, et al. Bacteriophage P2: genes involved in baseplate assembly. Virology 1995;213:109–21.
[45] Szewczyk B, Bienkowska-Szewczyk K, Kozloff LM. Identification of T4 gene 25 product, a component of the tail baseplate, as a 15K lysozyme. Mol Gen Genet 1986;202:363–7.
[46] Werts C, Michel V, Hofnung M, Charbit A. Adsorption of bacteriophage lambda on the LamB protein of E. coli K12: point mutations in gene J of lambda responsible for extended host range. J Bacteriol 1994;176:941–7.
[47] Ye N, Nemoto N. Processing of the tail lysozyme (gp5) of bacteriophage T4. J Bacteriol 2004;186:6335–9.
[48] Leiman PG, Arisaka F, van Raaij MJ, Kostyuchenko VA, Aksyuk AA, Kanamaru S, et al. Morphogenesis of the T4 tail and tail fibers. Virol J 2010;7:355.
[49] Stummeyer K, Dickmanns A, Mühlenhoff M, Gerardy-Schahn R, Ficner R. Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. Nat Struct Mol Biol 2005;12:90–6.
[50] McMahon SA, Miller JL, Lawton JA, Kerkow DE, Hodes A, Marti-Renom MA, et al. The C-type lectin fold as an evolutionary solution for massive sequence variation. Nat Struct Mol Biol 2005;12:886–92.
[51] Haggard-Ljungquist E, Halling C, Calendar R. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J Bacteriol 1992;174:1462–77.
[52] Christie GE, Temple LM, Bartlett BA, Goodwin TS. Programmed translational frameshift in the bacteriophage P2 FETUD tail gene operon. J Bacteriol 2002;184:6522–31.
[53] Katsura I, Hendrix RW. Length determination in bacteriophage lambda tails. Cell 1984;39:691–8.
[54] Bateman A, Bycroft M. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). J Mol Biol 2000;299:1113–9.
[55] Young R. Bacteriophage holins: deadly diversity. J Mol Microbiol Biotechnol 2002;4:21–36.
[56] Ziedaite G, Daugelavicius R, Bamford JK, Bamford DH. The holin protein of bacteriophage PRD1 forms a pore for small-molecule and endolysin translocation. J Bacteriol 2005;187:5397–405.
[57] Wang IN, Deaton J, Young R. Sizing the holin lesion with an endolysin-beta-galactosidase fusion. J Bacteriol 2003;185:779–87.
[58] Young R. Phage lysis: three steps, three choices, one outcome. J Microbiol 2014;52:243–58.
[59] Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. Bacteriophage T4 genome. Microbiol Mol Biol Rev 2003;67:86–156.
[60] Sutherland E, Coe L, Raleigh EA. McrBC: a multisubunit GTP-dependent restriction endonuclease. J Mol Biol 1992;225:327–48.
[61] Kobayashi I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res 2001;29:3742–56.
[62] Kita K, Kawakami H, Tanaka H. Evidence for horizontal transfer of the EcoT38I restriction-modification gene to chromosomal DNA by the P2 phage and diversity of defective P2 prophages in *Escherichia coli* TH38 strains. J Bacteriol 2003;185:2296–305.
[63] Brassard S, Paquet H, Roy PH. A transposon-like sequence adjacent to the AccI restriction-modification operon. Gene 1995;157:69–72.
[64] Bellanger X, Morel C, Decaris B, Guédon G. Derepression of excision of integrative and potentially conjugative elements from *Streptococcus thermophilus* by DNA damage response: implication of a cI-related repressor. J Bacteriol 2007;189:1478–81.