Isolation and Characterization of vB_ArS-ArV2 – First Arthrobacter sp. Infecting Bacteriophage with Completely Sequenced Genome

Eugenijus Šimoliūnas1*, Laura Kaliniene1, Miroslav Stasilo1, Lidija Truncaitė1, Aurelij Zajančkauskaitė1, Juozas Staniulis2, Juozas Nainys3, Algirdas Kaupinis4, Mindaugas Valius4, Rolandas Meškys1

1 Department of Molecular Microbiology and Biotechnology, Institute of Biochemistry, Vilnius University, Vilnius, Lithuania, 2 Laboratory of Plant Viruses, Institute of Botany, Nature Research Centre, Vilnius, Lithuania, 3 Department of Eukaryote Gene Engineering, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania, 4 Proteomics Centre, Institute of Biochemistry, Vilnius University, Vilnius, Lithuania

Abstract

This is the first report on a complete genome sequence and biological characterization of the phage that infects Arthrobacter. A novel virus vB_ArS-ArV2 (ArV2) was isolated from soil using Arthrobacter sp. 68b strain for phage propagation. Based on transmission electron microscopy, ArV2 belongs to the family Siphoviridae and has an isometric head (~63 nm in diameter) with a non-contractile flexible tail (~194×10 nm) and six short tail fibers. ArV2 possesses a linear, double-stranded DNA genome (37,372 bp) with a G+C content of 62.73%. The genome contains 68 ORFs yet encodes no tRNA genes. A total of 28 ArV2 ORFs have no known functions and lack any reliable database matches. Proteomic analysis led to the experimental identification of 14 virion proteins, including 9 that were predicted by bioinformatics approaches. Comparative phylogenetic analysis, based on the amino acid sequence alignment of conserved proteins, set ArV2 apart from other siphoviruses. The data presented here will help to advance our understanding of Arthrobacter phage population and will extend our knowledge about the interaction between this particular host and its phages.

Introduction

Arthrobacter sp. strains are widely distributed in the environment and have been found to be among the predominant members of culturable aerobic soil bacteria [1]. The genus Arthrobacter includes a group of catalase-positive, strictly aerobic, sporogenous rod-shaped coryneform bacteria with a high mol% GC DNA composition (generally ranging from 59 to 66%) and A-type (A3a or A4a) peptidoglycans with L-lysine as the dibasic amino acid [2]. The environmental prevalence of Arthrobacter sp. strains is considered to be due to their nutritional versatility and their pronounced resistance to desiccation, long-term starvation and environmental stress [1]. It is unsurprising that strains of the genus Arthrobacter are phenotypically heterogeneous and have been isolated from distinct sources, such as soil [3], wastewater sediments [4], clinical specimens [5], animals [6], phyllosphere [7], paintings [8], cheese [9] and air [10]. Moreover, Arthrobacter sp. have been found in extreme environments, such as the Arctic/Antarctic waters and sediments [11], chemically contaminated sites [12] and radioactive environments [13]. It was reported that certain species of the genus Arthrobacter have the capacity to degrade various difficult-to-degrade chemical substrates [14] and, in a few cases, exhibit denitrification activity [15]. To summarize, bacteria of the genus Arthrobacter are thought to play a significant role in many ecosystems and affect human welfare.

Although different Arthrobacter strains have been the subject of extensive studies, relatively little is known about their predators in nature – bacteriophages, and only a limited number of reports on Arthrobacter phages have been published thus far. Robinson and Corke [16] observed plaques when soil peritrichs were plated with Arthrobacter. In addition, some authors [17,18] investigated the use of bacterial viruses in the phage typing of soil Arthrobacter. Most of publications, however, described only the isolation and/or partial characterization of bacteriophages active on laboratory strains and soil isolates of Arthrobacter [18,19,20,21,22,23,24], and only a few studies provided more detailed characterizations of Arthrobacter phages [25,26,27]. The majority of Arthrobacter phages described to date belong to the family Siphoviridae [28]. To our knowledge, only phages AN25S-1 and AN29R-2 [18] showed morphological characteristics of podoviruses. Interestingly, while most reports about Arthrobacter phages were published during the seventies or eighties of the last century, none of the published Arthrobacter phages was tested regarding genome sequence/organisation or similarity to already known viruses (see Table S1).
This study represents the first detailed characterization of an *Arthrobacter* bacteriophage with a complete DNA sequence and annotation. Phage vB_ArS-ArV2 (ArV2) was isolated from soil and characterized with respect to morphology and biological properties. The data of genomic, proteomic and phylogenetic analyses indicate that ArV2 has no close relatives within the family *Siphoviridae* of tailed bacteriophages. Given that arthrobacterial viruses have never been investigated at the genomic level to date, the results provided in this report finally offer a glimpse into the biology of bacteriophages that infect *Arthrobacter*.

### Materials and Methods

#### Phages and bacterial strains

This study does not require an ethics statement (N/A). Phage ArV2 was isolated from soil samples collected on private land in Vilnius, Lithuania (54.839573, 25.369245). The person, who should be contacted for future permissions is Dr. Rolandas Meskys (rolandas.meskys@bchi.vu.lt). *Arthrobacter* sp. strain 68b was used as the host for phage propagation and phage growth experiments. Bacterial strains used in this study for host range determination are described in Table 1. For phage experiments, bacteria were cultivated in Luria-Bertani broth (LB) or LB agar at 30°C. Bacterial growth was monitored turbidimetrically by reading OD_{600}. An OD_{600} of 1.0 corresponded to 9 × 10^8 *Arthrobacter* sp. 68b cells/ml.

#### Phage techniques

Bacteriophages active on *Arthrobacter* spp. were searched for in various samples of soil (collected from the different sites in Lithuania). Soil samples (1–5 g) were shaken for 1 h in 10 ml of LB broth followed by low-speed centrifugation at 5000 rpm for 15 min. The supernatant fluid was sequentially filtered through sterile 0.45 and 0.2 μm membrane filters and assayed for plaque-forming units by the soft agar overlay method described by Adams [34] with minor modifications. Briefly, 0.1 ml of diluted phage suspension or clarified environmental sample was mixed with 0.5 ml of indicator cells (OD_{600} – 1). The mixture then was added to 2.5 ml of 0.5% (w/v) soft agar and poured over the 1.2% LB agar plate as a uniform layer. The plates were incubated 24–48 h at 30°C before the enumeration of plaques. Bacteriophage culture was purified by performing five consecutive transfers of phage from individual plaques to new bacterial cell lawns. Phage ArV2 propagation was performed using a standard protocol with few modifications using *Arthrobacter* strain 68b as a host. Briefly, phage particles were subsequently collected by adding 5 ml of LB broth to the surface of each plate. The top agar was scraped off and the suspension recovered. After 30 min. of incubation at 4°C with mild stirring the mixture was centrifuged at 6000 rpm for 15 min. The phage-containing supernatant was decanted and the phage was concentrated by high-speed centrifugation at 30 000 rpm for 3 h. The resulting pellets were suspended in PB buffer (70 mM NaCl, 10 mM MgSO4, 50 mM Na2HPO4, 30 mM KH2PO4). To avoid bacterial DNA contamination, DNase I was added to the phage suspension, and the sample was incubated 1 h at 37°C. Further purification was performed using CsCl density gradient-purification. High-titer phage stocks were diluted and plated in duplicate. Plates incubated at 18, 21, 22, 24, 26, 27, 28, 30, 32, 34 and 37°C were read after 18–96 hours of incubation. The temperature at which the largest number of plaques was formed was taken as the standard for the e.o.p. calculation.

#### Transmission electron microscopy (TEM)

CsCl density gradient-purified phage particles were diluted to approximately 10^11 PFU/ml with distilled water, 5 μl of the sample was directly applied on the carbon-coated nitrocellulose grid, excess liquid was drained with filter paper before staining with two successive drops of 2% uranyl acetate (pH 4.5), dried and examined in Morgagni 268(D) transmission electron microscope (FEI, Oregon, USA).

#### DNA isolation and restriction analysis

Aliquots of phage suspension (10^{11}–10^{12} PFU/ml) were subjected to phenol/chloroform extraction and ethanol precipitation as described by Carlson and Miller [38]. Isolated phage DNA was subsequently used in restriction analysis, for PCR or was subjected to genome sequencing. Restriction digestion was performed with BamHI, EcoRI, EcoRII, EcoRV, HindIII, KpnI, MboI, NheI, NotI, PstI, PvuI, SmaI, SspI, VspI and XbaI restriction endonucleases (Fermentas) according to the supplier's recommendations. DNA fragments were separated by electrophoresis in a 0.8% agarose gel containing ethidium bromide. Restriction analysis was performed in triplicate to confirm the results.

#### Genome sequencing and analysis

The complete genome sequence of ArV2 was determined using Illumina DNA sequencing technology at BaseClear, the Netherlands. Open reading frames (ORFs) were predicted with Glimmer v2.02 ([http://nbc11.biologie.uni-kl.de](http://nbc11.biologie.uni-kl.de)) and Geneious v3.5.6. (Société Geneious, Colombes, France). Analysis of the genome sequence was performed using the Fasta-Protein, Fasta-Nucleotide, Fasta- Genome, BLAST2, PSI-Search, Transeq, ClustalW2, ([http://www.ebi.ac.uk](http://www.ebi.ac.uk)), HHsearch1.6 ([http://toolkit.tuebingen.mpg.de/hhpred]), COMA ([http://bioinformatics.itbh.fh:8085/coma](http://bioinformatics.itbh.fh:8085/coma)), Sequence editor ([http://www.fr33.net/seqedit.php](http://www.fr33.net/seqedit.php)) programs and tRNAscan-SE 1.21 ([http://lowelab.ucsc.edu/tRNAscan-SE/](http://lowelab.ucsc.edu/tRNAscan-SE/)) was used to search for tRNAs. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 [39]. The physical location and nature of the genome ends were determined by combining the restriction analysis and subsequent sequencing directly from the genome ends. The sequencing reactions were carried out using a CycleReader DNA sequencing kit (Fermentas, Lithuania). The oligonucleotide primers (ArV2_F 5’– GACTAGGGCAGCGCATTGGAAC and ArV2_R 5’– GCATGACGCAAGGACATC) for the sequencing reactions were 5’–end labeled using T4 polynucleotide kinase (Thermo Fisher Scientific, Lithuania) with [γ32P]ATP (PerkinElmer, USA).

#### Analysis of Structural Proteins

**SDS-PAGE.** CsCl-purified phage particles (10^{11} pfu/ml) were resuspended in a buffer containing 60 mM Tris–HCl (pH 6.8), 1% SDS (w/v), 1% 2-mercaptoethanol (v/v), 10% glycerol (v/v) and 0.01% bromophenol blue (w/v). The samples were boiled for 3 min and separated on 12% SDS PAGE following the method described by Laemmli [40]. Protein bands were visualized by staining the gel with PageBlue Protein Staining Solution (Thermo Fisher Scientific, Vilnius, Lithuania).

**Filter-aided protein sample preparation (FASP) for mass spectrometry analysis.** CsCl-purified phage particles were concentrated on Amicon Ultra-0.5 mL 30 kDa centrifugal filter...
unit and were denatured in 8 M urea, 100 mM DTT solution with continuous rotation at 800 rpm in the temperature controlled shaker for 3 hours at 37°C.

Trypsin digestion was performed according to a modified FASP protocol as described by Wisniewski et al. [41]. Briefly, phage particles were washed with buffer containing 8 M urea. The

| Strain | Relevant characteristics | Source or reference |
|--------|--------------------------|---------------------|
| Acinetobacter baumannii | type strain | Prof. E. Suziedeliene |
| Acinetobacter gen. sp. 13 | type strain | Prof. E. Suziedeliene |
| Arthrobacter chlorophenolicus DSM 12829 | type strain | DSMZ |
| Arthrobacter crystallotoepes DSM 20117 | type strain | DSMZ |
| Arthrobacter deliuvi DSM 18782 | type strain | DSMZ |
| Arthrobacter globiformis DSM 20124 | type strain | DSMZ |
| Arthrobacter histidinolovorans DSM 2015 | type strain | DSMZ |
| Arthrobacter ilicis DSM 20138 | type strain | DSMZ |
| Arthrobacter methylotrophus DSM 14008 | type strain | DSMZ |
| Arthrobacter nicotinovorans DSM 420 | type strain | DSMZ |
| Arthrobacter nitroguajacolicus DSM 15232 | type strain | DSMZ |
| Arthrobacter oxydans DSM 20119 | type strain | DSMZ |
| Arthrobacter sp. 68b | environmental isolate | [29] |
| Arthrobacter sp. 68m | environmental isolate | laboratory collection |
| Arthrobacter sp. 83b | environmental isolate | laboratory collection |
| Arthrobacter sp. 85 | environmental isolate | laboratory collection |
| Arthrobacter sp. 94 | environmental isolate | laboratory collection |
| Arthrobacter sp. 96 | environmental isolate | laboratory collection |
| Arthrobacter sp. 25DMP1 | environmental isolate | [14] |
| Arthrobacter sp. 25DOT1 | environmental isolate | [14] |
| Arthrobacter sp. PY11 | environmental isolate | [30] |
| Arthrobacter sp. PY21 | environmental isolate | [31] |
| Arthrobacter sp. PY22 | environmental isolate | [31] |
| Arthrobacter sp. VP3 | environmental isolate | [30] |
| Arthrobacter sp. PRH1 | environmental isolate | [31] |
| Arthrobacter sp. VM22 | environmental isolate | [30] |
| Arthrobacter sp. IN13 | environmental isolate | [32] |
| Arthrobacter ureafaciens DSM 20126 | type strain | DSMZ |
| Citrobacter freundii | | Prof. E. Suziedeliene |
| Enterobacter cloacae | | Prof. E. Suziedeliene |
| Erwinia carotovora 8982 | | Prof. E. Suziedeliene |
| Escherichia coli B³ | sup³ | Dr. L. W. Black |
| Escherichia coli BL21 (DE3) | F− dcm ompT hsdS(rB− mB−) gal λ(DE3) | Avidis |
| Escherichia coli DH5α | F− endA1 gltIV44 thi-1 recA1 relA1 gyrA96 deoR nupG-Φ00lacZAM15 ΔlacZYA-argF(U169, hsdR17(k− m−)) λ− | Pharmacia |
| Escherichia coli DH10B | F− endA1 recA1 galE15 galK16 nupG-Φ(aMC1895)ΔlacX74 Φ00lacZAM15 araD139 Δara,leu)7697 mcrA Δmrr-λhsdR515-mcrBC λ− | Invitrogen |
| Escherichia coli DH5α | F− endA1 gltIV44 thi-1 recA1 relA1 gyrA96 deoR nupG-Φ00lacZAM15 ΔlacZYA-argF(U169, hsdR17(k− m−)) λ− | Pharmacia |
| Klebsiella pneumoniae 279 | | Prof. E. Suziedeliene |
| Klebsiella sp. KV-3 | Phage host, veterinary isolate, Amp’, Str’, Tef’, Kan’, Gm’, Nc’, Cl’, Im’, Cm’ | [33] |
| Pseudomonas aeruginosa PA01 | | Prof. E. Suziedeliene |
| Salmonella enterica ser. Typhimurium 292 | | Prof. E. Suziedeliene |

doi:10.1371/journal.pone.0111230.t001
proteins were alkylated using iodoacetamide. Buffer was exchanged by washing twice with 50 mM NH₄HCO₃, and proteins digested overnight with TPCK Trypsin 20233 (Thermo Scientific, USA). Then peptides were recovered by centrifugation and washed twice with 50% CH₃CN. Afterwards, samples were combined, acidified, lyophilized, redissolved in 0.1% formic acid and analysed by mass spectrometry.

In-gel protein digestion for mass spectrometry analysis. In-gel trypsin digestion was done according to a protocol described by Hellman et al. [42] with minor modifications. Briefly, gel slices were destained with 200 μl of 50 mM ammonium bicarbonate in 50% CH₃CN, vacuum-dried, rehydrated in 50 μl (20 μg ml⁻¹) of trypsin TPCK Trypsin 20233 (Thermo Scientific, USA) containing 25 mM NH₄HCO₃ and incubated overnight at 37°C. The peptides were extracted from the gel using 100 μl of CH₃CN for 30 min. Next, gel pieces were washed with (100 μl) of 1% formic acid for 10 min. Extraction procedure was finished by adding 100 μl of CH₃CN. The peptides from all extractions were combined, acidified, concentrated by vacuum drying, resuspended in 40 μl 0.1% formic acid and then used for mass spectrometry analysis.

Liquid chromatography and mass spectrometry. Liquid chromatography (LC) separation of trypsin cleaved peptides was performed with nanoAcquity UPLC system (Waters Corporation, UK). Peptides were loaded on a reversed-phase trap column PST C18, 100A˚, 5 μm, 180 μm × 20 mm (Waters Corporation, UK) at a flow-rate of 15 μl/min using loading buffer of 0.1% formic acid and subsequently separated on HSS-T3 C18 1.8 μm, 75 μm × 250 mm analytical column (Waters Corporation, UK) in 30 min linear gradient (A: 0.1% formic acid, B: 100% CH3CN and 0.1% formic acid) for in-gel protein trypsin digested samples or 60 min for FASP material at a flow rate of 300 nl per min. The analytical column temperature was kept at 40°C.

Figure 1. Electron micrographs of *Arthrobacter* phage ArV2. doi:10.1371/journal.pone.0111230.g001

Figure 2. Plaques formed by bacteriophage ArV2 on a lawn of *Arthrobacter* sp. 68b. doi:10.1371/journal.pone.0111230.g002
The nano-LC was coupled online through a nanoESI 7 cm length, 10 mm tip emitter (New Objective, USA) with HDMS Synapt G2 mass spectrometer (Waters Corporation, UK). Data were acquired using Masslynx version 4.1 software (Waters Corporation, UK) in positive ion mode. LC-MS data were collected using data independent acquisition (DIA) mode MS² (for in-gel digested proteins) or MS³ in combination with online ion mobility separations (for FASP material).

The trap collision energy of mass spectrometer was ramped from 18 to 40 eV for high-energy scans in MS² mode. The trap and transfer collision energy for high-energy scans in HDMS mode was ramped from 4 to 5 eV and from 27 to 50 eV. For both analyses, the mass range was set to 50–2,000 Da with a scan time set to 0.9 second. A reference compound [Glu1]-Fibrinopeptide B (Waters Corporation, UK) was infused continuously (500 fmol/μl at flow rate 500 nl per min) and scanned every 30 seconds for online mass spectrometer calibration purpose.

Data Processing, Searching and Analysis. Raw data files were processed and searched using ProteinLynx Global SERVER (PLGS) version 2.5.2 (Waters Corporation, UK). The following parameters were used to generate peak lists: (i) minimum intensity for precursors was set to 100 counts, (ii) minimum intensity for fragment ions was set to 30 counts, (iii) intensity was set to 500 counts. Processed data was analysed using trypsin as the cleavage protease, one missed cleavage was allowed and fixed modification was set to carbamidomethylation of cysteines, variable modification...
tion was set to oxidation of methionine. Minimum identification criteria included 2 fragment ions per peptide, 5 fragment ions per protein and minimum of 2 peptides per protein. The false discovery rate (FDR) for peptide and protein identification was determined based on the search of a reversed database, which was generated automatically using PLGS when global false discovery rate was set to 4%.

Nucleotide sequence accession numbers
The complete genome sequence of *Arthrobacter* bacteriophage ArV2 was deposited in the EMBL nucleotide sequence database under accession number KF692088.

Results and Discussion

**Virion morphology**

TEM observations of ArV2 (Fig. 1) revealed a particle that fits B1 morphotype in Bradley’s classification [43]. Based on morphological characteristics, phage ArV2 belongs to the family *Siphoviridae* [44] and is characterized by an isometric head

### Table 2. Mass spectrometry analysis of phage ArV2 proteins.

| Gene     | Putative Function          | MW (KDa) | Peptide count | Sequence coverage (%) |
|----------|----------------------------|----------|---------------|-----------------------|
| ORF15*   | tape measure protein       | 127.953  | 74            | 35.2                  |
| ORF24*   | tail fiber protein         | 52.293   | 11            | 38.0                  |
| ORF03*   | phage portal protein       | 50.447   | 60            | 79.3                  |
| ORF17*   | tail protein               | 40.386   | 42            | 70.1                  |
| ORF16*   | tail protein               | 35.644   | 26            | 58.3                  |
| ORF18*   | tail protein               | 33.604   | 39            | 84.9                  |
| ORF04†   | capsid maturation protease | 31.057   | 19            | 63.3                  |
| ORF06*   | major capsid protein       | 30.937   | 21            | 78.9                  |
| ORF11*   | major tail protein         | 23.801   | 72            | 95.9                  |
| ORF23*   | hypothetical protein       | 20.621   | 16            | 95.4                  |
| ORF07    | hypothetical protein       | 14.566   | 13            | 91.2                  |
| ORF10    | hypothetical protein       | 14.563   | 9             | 70.4                  |
| ORF08    | hypothetical protein       | 12.709   | 7             | 78.8                  |
| ORF09†   | hypothetical protein       | 10.230   | 10            | 76.2                  |

* – putative structural ArV2 proteins identified by bioinformatics approaches;
† – proteins detected using the FASP method only;
† – ArV2-specific proteins that have no reliable identity to database entries.

Proteins were identified as a result of whole-phage shotgun analysis (WSA) using LC MS/MS analysis and Mascot search against an in-house database.

doi:10.1371/journal.pone.0111230.t002

**Figure 5. SDS-PAGE of virion proteins of ArV2.** Lanes: 1– molecular mass marker Page Ruler™ prestained Protein Ladder Plus (Thermofisher), 2– phage ArV2 structural proteins. Relative migrations of MW marker proteins are indicated on the left. Proteins identified by MS/MS are indicated on the right.

doi:10.1371/journal.pone.0111230.g005

**Figure 6. Neighbor-joining tree analysis based on the alignment of the amino acid sequences of the major capsid proteins from various tailed viruses.** The bootstrap values are indicated.

doi:10.1371/journal.pone.0111230.g006
(diameter, 62.8±4.9 nm; n = 30) and an apparently non-contractile, flexible tail (194.4±9.6 nm in length and 11.8±1.1 nm in width; n = 30). A baseplate was observed, although its diameter was not clearly distinguishable from the contractile sheath. Tail fibers are not obviously visible but upon closer inspection, six baseplate-associated short tail fibers (7.5±1.2 nm in length) can be seen in the vicinity of the tail tips (Fig. 1D, Fig. 1E).

The host range and physiological characteristics

In total, 40 bacterial strains (Table 1) were used to explore the host range of ArV2. With the exception of Arthrobacter sp. strain 68b, all of 2 Acinetobacter sp., 26 Arthrobacter sp., 1 Citrobacter sp., 1 Enterobacter sp., 1 Erwinia sp., 5 Escherichia sp., 2 Pseudomonas sp., 2 Klebsiella sp. and 1 Salmonella sp. strains were found to be resistant to ArV2. Arthrobacter sp. strain 68b was isolated from soil. Although a number of strains used in this study is not large, the diversity of bacteria tested allow presuming that the host range of ArV2 is limited to Arthrobacter only.

To determine the optimal conditions for phage propagation, the efficiency of plating (e.o.p.) test was performed. The e.o.p. of ArV2 was examined in the temperature range of 15–37°C, and the test revealed that the phage has an optimum temperature for plating around 30°C (Fig. S1). After 48 h of incubation at an optimum temperature of 30°C, bacteriophage ArV2 formed clear plaques of 1.5±0.5 mm in diameter (Fig. 2). While ArV2 plaques can be visible after one day (24 h) of incubation at 30°C, an accurate evaluation of plaque morphology as well as the plaque enumeration is best performed after 36–48 h of incubation.

Attempts to obtain a one-step growth curve of ArV2 on strain 68b were unsuccessful because of the slow adsorption kinetics. Despite varying the inorganic environments (MgCl₂ or CaCl₂ solutions were added to the medium to reach the concentration of 10 mM) and temperature, only about 50% of the PFU adsorb in 5 min, and after 15 min as many as 25% of the original PFU remain unattached (data not shown). The slow adsorption kinetics of ArV2, taken together with extremely narrow host range, suggests that perhaps other bacteria, rather than Arthrobacter sp. 68b, is the real host of ArV2. On the other hand, these results may reflect the inherent nature of the phage, or the failure to provide the most appropriate environment for adsorption.

Restriction analysis

To protect the genomic DNA from restriction endonucleases, phages employ various strategies, including adenine and cytosine methylation as well as hydroxymethylation of cytosine (HMC) and subsequent glucosylation of HMC derivatives [45,46]. Investigation of ArV2 DNA modifications was performed using EcoRI, NotI and MboI restriction endonucleases that do not cleave Dcm, CpG and Dam methylated DNA, respectively. The sensitivity of phage DNA to the restriction enzymes listed above suggests that the DNA of ArV2 possesses no significant amounts of bases with Dcm, CpG and Dam methylation (Fig. 3).

Overview of the phage ArV2 genome

Phage ArV2 has a linear, double stranded DNA genome consisting of 37,372 bp with a G+C content of 62.73%, which is similar to that observed for Arthrobacter [9]. Since the restriction digestion profiles of phage ArV2 DNA matched in silico predictions of a linear DNA molecule, the locations of the terminase-generated ends (cos sites) were readily determined from a restriction map. By sequencing directly from the genome ends (data not shown), the cos sites of ArV2 were identified as 9-bp 3′ overhang sequences (5′-CCTCCGGCA-3′).

The genome of ArV2 is close-packed: with an average ORF size of 535 bp, ~96% of the genome is coding. It is not surprising since in tailed viruses protein-coding genes are generally tightly packed and typically occupy >90% of the genome [47]. The genome sequence analysis revealed that ArV2 has a total of 68 probable protein-encoding genes and no genes for tRNA (Fig. 4). While most of the ArV2 genes were found to initiate from ATG (37 out
of 68 ORFs, 24 ORFs were found to initiate with GTG and 7 with TTG. A marked asymmetry in the distribution of the genes on the two phage ArV2 DNA strands was observed. In total, 62 (91%) ORFs were predicted to be transcribed from the same DNA strand, while 6 (9%) ORFs (most of these were detected in the lysogeny-related module) were found on the opposite strand.

The genome analysis revealed that 41% of the ArV2 genes (28 out of 68 ORFs) encode unique proteins that have no reliable identity (e-values >0.001) to database entries (Fig. 4). Among the ArV2 gene products with detectable homologs in other sequenced genomes, 34 were most similar to proteins from bacteria, such as Arthrobacter, Microbacterium, Rhodococcus, Mobiluncus, Actino-

myces, Lactococcus, etc. (Fig. S2A, Table S2). In addition, 24 of the aforementioned ArV2 ORFs showed amino acid sequence similarity, although to a lesser extent, to proteins from bacteriophages that infect Mycobacterium, Propionibacterium, Streptomy-

ces, Gordonia, Rhodococcus, Streptococcus and other bacteria from the actinobacterial group (Fig. S2B). Only 6 ArV2 gene products showed sequence similarity exclusively to the homologous viral proteins (Fig. S2A, Table S2). Through the examination of homology search results, a putative function was assigned to 32 ArV2 ORFs that are distributed across several functional categories, such as head and tail morphogenesis, DNA packaging, lysogeny, replication and regulation (Fig. 4).

DNA replication/recombination and modification enzymes

Several ArV2 gene products are potentially involved in nucleic acid replication or recombination processes. ORF54 is predicted to encode a ssDNA binding protein, which builds the scaffold of DNA replication and is reported to be essential both for DNA replication and for genetic recombination processes [48]. The protein encoded by ArV2 ORF61 shared 42% amino acid identity with the DNA helicase from Streptomyces sp. C, while the predicted product of ORF59 was found to be homologous to a DNA replication protein of Lactobacillus phage LF1 and showed weak homology to a DNA replication protein O from Sxt2 converting phage vB_Ecop_24B. The DNA helicases are known to unwind the duplex DNA and are involved in replication, repair and recombination processes [49,50], while the protein O is been reported to be necessary for the initiation of bacteriophage DNA replication [51]. Interestingly, although most of the phages encode their own conserved DNA polymerases [52], there is no DNA polymerase gene identified in the ArV2 genome suggesting that this phage takes advantage of host replication machinery.

Other ArV2 ORFs possibly involved in DNA recombination include a RecE exonuclease VIII (ORF52), a Holliday junction resolvase RusA [53] encoded by ORF65, two free-standing HHN endonucleases (ORF68 and ORF26) and one GIY-YIG nuclease (ORF058). Phage RecE 5’-3’ exonucleases act on dsDNA by production of a protruding 3’ ssDNA for strand annealing or invasion in homologous recombination [54,55], while homing endonucleases are a distinctive class of site-specific DNA endonucleases that promote the lateral transfer of their own coding region and flanking DNA between genomes by a recombination-dependent process termed homing [56,57].

Based on the amino acid sequence similarity, only one gene product of ArV2 was assigned as a protein involved in the DNA modification process. The derived protein product of ORF64 showed 58% amino acid sequence identity to the DNA-(N6-
 adenine)-methyltransferase from Burkholderia phage Bcep176 and was found to have a conserved Dam domain. According to the literature, adenine-specific DNA methyltransferases are often involved in chromosomal site-specific DNA modification systems, and it has been speculated that some bacteriophages express their own methylases to protect the genomic DNA against restriction enzymes of the host [58,59]. However, as it was mentioned above, the restriction analysis revealed that the DNA of ArV2 seems to possess no significant amounts of bases with Dcm, CpG or Dam methylation (Fig. 3) suggesting that the predicted DNA-(N6-
 adenine)-methyltransferase from ArV2 is either not functional or modifies just a negligible amounts of nucleotides.

DNA packaging

Packaging of phage genomes into empty procapsids is powered by the terminase holozyme that is generally composed of two subunits, large and small. In tailed phages, the small subunit of the terminase is responsible for specific DNA binding and holozyme formation, while the large subunit of the terminase mediates cleavage of the phage DNA into genome size units to be packaged into the prohead [60]. Bioinformatics analysis revealed that the large terminase subunit of ArV2 may be encoded by ORF02, as it shares 46% amino acid sequence identity with the terminase large subunit from Mycobacterium phage Validus, while the gene for a small terminase subunit seems to absent in ArV2. Nevertheless, in tailed phages, the gene encoding terminase small subunit is generally located upstream of the terminase large subunit gene, and both terminase genes are transcribed together in an operon-like structure [61,62]. In ArV2, the gene position of ORF01 is upstream of the terminase large subunit gene suggesting that the product derived from this ORF is a potential candidate to be a novel small terminase. However, experimental data are needed to confirm this suggestion.

Besides terminase, another phage protein involved in packaging is the head-tail connector (or portal protein), which is the key functional component of the capsid for DNA packaging and is involved in the signaling for packaging termination [63,64]. Comparative sequence analysis revealed that the portal protein of ArV2 is most likely encoded by the gene ArV2 ORF03, since it exhibited a moderate amino acid sequence identity (33%) to the portal protein from Mycobacterium phage BPs.

Structural proteins

As was observed in other siphoviruses [61,65], the packaging module in ArV2 is followed by a large genome cluster (~18 kb) that contains genes encoding structural components of the virion. Bioinformatics analysis allowed us to identify 9 structural proteins of ArV2, including those coding for head (ORF03, ORF06), tail (ORF11, ORF15–ORF18), tail fiber (ORF20, ORF24), as well as a capsid maturation protein (ORF04) and a scaffolding protein (ORF05).

Reversed-phase nano-liquid chromatography directly coupled with LC-MS/MS analysis of the structural ArV2 proteins separated by SDS PAGE, and filter-aided protein sample preparation (FASP) led to the experimental identification of 14 virion proteins (Table 2), including 9 that were predicted by bioinformatics approaches as well as 5 proteins, which either have no detectable homology to any entries in the public databases (ORF09, ORF23) or are similar to hypothetical proteins from other phages (ORF07) and bacteria (ORF08, ORF10). Unexpectedly, a putative capsid maturation protein encoded by ArV2 ORF04, which has poor (23% identity) but significant amino acid sequence similarity to a capsid maturation protein of Mycobacterium phage Job2, was identified by MS/MS. In most phages, such proteins as morphogenesis-associated proteases are thought to be lost from the structure during capsid maturation [66], however, in the case of certain bacterial viruses, such as the
coliphage P2 or mycobacteriophage Marvin, the protease is retained in mature virions [66,67].

The most abundant protein in the ArV2 virion (Fig. 5) was the major tail subunit (gp11) that shared 29% identity with the putative major tail protein from Corynebacterium phage P1201. In contrast to what has been observed in other tailed viruses, the predicted major capsid protein (gp06) of ArV2 was not the most abundant protein in the virion. The ArV2 ORF06 was predicted to encode a 30.9 kDa protein (290 residues) that shares 51% identity with the major capsid protein from Mycobacterium phage Severus. As seen from Fig. 5, gp06 was found to migrate with a molecular mass much larger than predicted, suggesting the covalent self-linking of the capsid proteins, a structural feature observed with other phage capsid proteins [68,62,69]. The cross-linking of the capsid proteins may explain a relatively low amount of gp06 detected by LC-MS/MS, presumably because cross-linking leads to the failure of the protein to enter the gel. The largest gene product (1209 aa) identified in the ArV2 genome was the tape measure protein (gp15) that showed 35% identity to the tape measure protein (TMP) from Rhodococcus phage RegiPInc5. The TMP usually functions as a template for measuring tail length during tail assembly [70]. As it can be seen in Fig. 5, gp15 was found to migrate with a molecular mass slightly lower than predicted (100 kDa vs 128 kDa). According to the literature, it is quite common for TMPs to be proteolytically processed during phage assembly [71,72,73]. Hence, the size reduction of ArV2 gp15 observed by SDS-PAGE suggests that this protein may be subjected to proteolytic cleavage.

At least five more tail proteins (gp16–18, gp20 and gp24) were detected in the virion of ArV2. The products of ORF16, ORF17 and ORF18 shared 30%, 33% and 32% amino acid sequence identity with putative minor tail proteins gp15, gp16 and gp17 from Propionibacterium phage P100-1, respectively. Meanwhile gp20 and gp24 of ArV2 were similar to the gp7 from Mycobacterium phage Trixie (47% aa identity) and to the hypothetical protein from Rhodococcus opacus B4 (32% aa identity), respectively.

Based on the results of proteomics analysis, the proteins encoded by ORF03, ORF05 and ORF21 were also present in ArV2 virion. A comparative sequence analysis revealed that ORF03 coded for portal protein (ORF03) and ORF21 were also present in ArV2 virion. A comparative sequence analysis revealed that ORF03 coded for portal protein (ORF03) and ORF21 predicted to encode a 30.9 kDa protein (290 residues) that shares 51% identity with the major capsid protein from Mycobacterium phage Severus. As seen from Fig. 5, gp06 was found to migrate with a molecular mass much larger than predicted, suggesting the covalent self-linking of the capsid proteins, a structural feature observed with other phage capsid proteins [68,62,69]. The cross-linking of the capsid proteins may explain a relatively low amount of gp06 detected by LC-MS/MS, presumably because cross-linking leads to the failure of the protein to enter the gel. The largest gene product (1209 aa) identified in the ArV2 genome was the tape measure protein (gp15) that showed 35% identity to the tape measure protein (TMP) from Rhodococcus phage RegiPInc5. The TMP usually functions as a template for measuring tail length during tail assembly [70]. As it can be seen in Fig. 5, gp15 was found to migrate with a molecular mass slightly lower than predicted (100 kDa vs 128 kDa). According to the literature, it is quite common for TMPs to be proteolytically processed during phage assembly [71,72,73]. Hence, the size reduction of ArV2 gp15 observed by SDS-PAGE suggests that this protein may be subjected to proteolytic cleavage.

Lysogeny

The lysogeny module of ArV2 was found downstream of the structural module. Unsurprisingly, since lysogeny-related genes usually are in the opposite orientation on the complementary strand relative to the other functional modules [80], two of the lysogeny-related genes of ArV2, namely integrase (ORF29c) and phage repressor (ORF34c), were found on the complementary strand (Fig. 4). The predicted phage repressor protein, which is critical to preservation of lysogeny [81], was encoded by ORF34c and shared 42% amino acid sequence identity with cI-like repressor from Streptococcus phage SII21. The product of ORF29c, which showed consistent homology to phage integrases (Fig. S3), was predicted to belong to the integrase family of tyrosine recombinases (Int family) [82]. Phage integrases are site-specific recombinases that mediate recombination between the phage genome and the bacterial chromosome [83]. According to the literature, Int integrases, such as lambda integrase, utilize a catalytic tyrosine to mediate strand cleavage and require cofactors encoded by the phage or the host bacteria [83]. Nevertheless, no phage encoded cofactors were found in the genome of the phage ArV2, suggesting that this phage either utilizes host-encoded cofactors or harbors peptides with negligible identity to known Cron-like proteins. On the other hand, there is also a possibility that ArV2 is a lysogenic phage that begins the lytic cycle spontaneously.

To determine whether lysogens could be recovered from ArV2 infections, cells from a spot where ArV2 particles had infected a lawn of Arthrobacter sp. 68b were recovered and grown on solid media. Bacterial growth was observed, and two independent colonies were restreaked twice more and then patched onto Arthrobacter sp. 68b lawns to test for phage release; none of the colonies recovered showed phage release (data not shown). Thus, although lysogeny-related genes are typically observed in temperate phages [84,85], there is no evidence that ArV2 is capable of lysogenizing Arthrobacter sp. 68b.

Phylogenetic relatedness

As observed with other bacterial viruses, gene content and genetic identity are highly heterogenous between phages and, more often than not, prevent the application of traditional phylogenetic methods using whole genome sequences [86]. A comparative sequence analysis revealed that the genome of Arthrobacter sp. infecting bacteriophage ArV2 shares no identity with any other sequenced genome. Moreover, at the amino acid sequence level, the majority of functionally annotated proteins of ArV2 are much more closely related to different bacterial proteins than to those from other sequenced viral genomes. Hence, the
single-gene comparison approach was undertaken, and phylo-
genetic analysis of the major capsid protein (Fig. 6) as well as four other ubiquitous proteins, including terminase, integrase, tail tape measure protein and portal protein, was carried out to better understand the evolutionary relationships between ArV2 and other tailed viruses (Fig. 7). All five phylogenetic trees showed that ArV2 is phylogenetically distant from other phages and, most likely, represents an evolutionarily distinct branch within the family Siphoviridae.

Conclusions
To our knowledge, this study represents the first complete genome sequence and genetic characterization of an Arthrobacter sp. infecting bacteriophage. Bioinformatics analysis revealed that the genome of ArV2 is significantly divergent from the siphovirus genomes sequenced to date, so much so that even structural proteins required experimental validation to annotate while a number of functionally important proteins (e.g. DNA polymerase, holin and endolysin) remained undetected. Undoubtedly, further genome sequencing and bioinformatic analysis should be performed to overcome the lack of genome annotation information and to draw a more detailed view of this particular group of bacterial viruses. However, the results of this study may provide new insights that deepen our understanding of Arthrobacter phage genetics and phage-host interactions in dynamic ecosystems, such as soil.

References
1. Cacciari I, Lippi D (1987) Arthrobacters: successful arid soil bacteria. A review. Arid Soil Res Rehabil 1: 1–30.
2. Jones D, Kerkis RM (2006) The genus Arthrobacter. In: Dworkin M, Falkowska S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. Springer, New York, USA. 945–960.
3. Hagedorn C, Holt JG (1975) A nutritional and taxonomic survey of Arthrobacter soil isolates. Can J Microbiol 21: 353–361.
4. Koh SW, Sung Y, Nam YD, Chang HW, Kim KH, et al. (2008) Arthrobacter soli sp. nov., a novel bacterium isolated from wastewater reservoir sediment. J Microbiol 46: 40–44.
5. Huang Y, Zhao N, He L, Wang L, Liu Z, et al. (2005) Arthrobacter aerenchym sp. nov. isolated from human clinical specimens. J Clin Microbiol 43: 1451–1455.
6. Collins MD, Hoyles L, Foster G, Falsen E, Weiss N (2002) Arthrobacter nasophorum sp. nov., from the common soil (Phocoid vitulina). In J Syst Evol Microbiol 52: 369–371.
7. Schelin-BR, Leveau JH (2013) Isolation of Arthrobacter species from the phyllosphere and demonstration of their epiphytic fitness. Microbiologyopen 2: 205–213.
8. Heyman J, Verheeren J, Schumann P, Swings J, Paul DV (2005) Six novel Arthrobacter species isolated from deteriorated mural paintings. Int J Syst Evol Microbiol 55: 1457–1464.
9. Monnet C, Lous V, Ghevat JF, Spinnler E, Barbe V, et al. (2010) The Arthrobacter arilaitensis Re117 genome sequence reveals its genetic adaptation to the surface of cheese. PLoS ONE 5: e13409.
10. Li Y, Kawamura Y, Fujiwara N, Naka T, Liu H, et al. (2004) Rothia aeria sp. nov., Rhusodens baikauensis sp. nov. and Arthrobacter russicus sp. nov., isolated from air in the Russian space laboratory Mir. Int J Syst Evol Microbiol 54: 2263–2266.
11. Pindi PK, Manorama R, Begum Z, Shivaji S (2010) Arthrobacter antarcticus sp. nov., isolated from an Antarctic marine sediment. Int J Syst Evol Microbiol 60: 2263–2266.
12. Hanbo Z, Changqun D, Qiyong S, Weimin R, Tao S, et al. (2004) Genetic and physiological diversity of phylogenetically and geographically distinct groups of Arthrobacter isolated from lead-zinc mine tailings. FEMS Microbiol Ecol 49: 333–341.
13. Fredrickson JK, Zachara JM, Bakkoll DL, Kennedy D, Li SM, et al. (2004) Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the Hanford site, Washington state. Appl Environ Microbiol 70: 4230–4241.
14. Kutunovas S, Rutkienė R, Ubelis G, Tauraitė D, Stankavičiūtė J, et al. (2013) Bionconversion of methylpyrazines and pyridines using novel pyrazines-degrading microorganisms. Chemija 24: 67–73.
15. Marshall SJ, White GF (2001) Complete denitrification of nitroglycerin by bacteria isolated from a wastewater soakaway. Appl Environ Microbiol 67: 2622–2626.

Supporting Information
Figure S1 Effect of temperature on the efficiency of plating of phage ArV2. (TIFF)

Figure S2 Taxonomy of the best BlastP hits of ArV2 ORFs. (A) The taxonomic distribution of homologues in phages that were grouped according to host specificity (B). (TIFF)

Figure S3 Sequence alignment of the Int family integrase from Arthrobacter phage ArV2 and other organisms. Conserved amino acid positions demonstrated by Nunes-Duby et al. [82] are shaded. (TIFF)

Table S1 The diversity of Arthrobacter bacteriophages described to date. S – Siphoviridae, P – Podoviridae, nd – not determined. (DOC)

Table S2 ArV2 ORFs with homologues in other viral or cellular organisms. (DOC)

Author Contributions
Conceived and designed the experiments: ES LK RM MV. Performed the experiments: ES MS LT AZ JS JN AK. Analyzed the data: ES LK AK. Contributed reagents/materials/analysis tools: JS MV RM. Wrote the paper: ES LK RM.
