

RESEARCH

Open Access



Isolation and genomic characterization of *Psychrobacillus* isolate L3 and bacteriophage Spoks: a new phage-host pair from Antarctic soil

Nikita Zrelavs¹, Karina Svanberga¹, Juris Jansons¹, Kristaps Lamsters², Janis Karuss², Maris Krievans², Davids Fridmanis¹, Andris Dislers^{1†} and Andris Kazaks^{1*†}

Abstract

Background Most habitats on Earth house unfathomable microbial diversity, yet much of it remains uncultured. The same applies to temperate phages, most of which documented to date are predicted purely in silico from the prophage-like genomic regions of the bacteria, lacking any experimental evidence of their functional integrity (e.g., the ability to undergo lytic replication). Hard-to-access parts of our planet with unique environments serve as especially promising places to collect samples for the isolation of novel microbes highly divergent from those isolated thus far. Antarctica, a continent mostly covered by a thick ice sheet, is one such area of our planet rife with novel microbiological entities. In this study, we aimed to isolate and characterize a novel culturable phage-host pair from Antarctic soils.

Results *Psychrobacillus* phage Spoks was retrieved alongside its host bacterial strain designated as “L3” from an ice-free soil sample collected at Waddington Bay, Graham Coast, Antarctica. Whole-genome sequencing of both the phage and the host revealed that they are divergent from, respectively, viruses and bacteria cultured and characterized thus far, and the intergenomic differences suggest that both might represent novel taxa. The genome of siphophage Spoks is a 36,472 bp long linear double-stranded DNA molecule with 11 base long 3' cohesive overhangs. Spoks can integrate into the chromosome of its isolation host strain in a site-specific fashion. Integration takes place in the genomic region of the host chromosome between the ORFs predicted to encode a DNA topoisomerase III and a Blal/Mecl/CopY family transcriptional regulator via recombination between attP and attB, which share a 19 bp “core” overlap sequence. L3 lysogens containing Spoks are not stable, with regular spontaneous induction occurring. Although the attachment site overlap sequence was found in the publicly available genomic sequences of several other *Psychrobacillus* spp. strains isolated from different habitats, none were found to contain a Spoks-like prophage.

[†]Andris Dislers and Andris Kazaks are co-senior authors.

*Correspondence:
Andris Kazaks
andris@biomed.lu.lv

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions The isolation and characterization of *Psychrobacillus* temperate phage Spoks and its host strain L3 from Antarctica highlight the potential for discovering novel biological entities divergent from their closest cultured relatives with relative ease, given access to such difficult-to-access undersampled environments, and are expected to encourage similar studies.

Keywords *Psychrobacillus*, *Caudoviricetes*, Antarctica, Temperate phage, Integration, Prophage, Lysogeny, Genomics, Sequencing

Background

Multiple environments are continuously shown to house microbial life different from what has been cultured thus far [1]. However, the diversity of culturable microbes that flourish in “distant” and “unusual” places, such as polar regions, general access to which remains limited by various factors, is especially undersampled [2]. One such place is Antarctica, which is a hard-to-access continent covered by a thick ice sheet that boasts some of the most extreme environments on Earth. Situated on the outskirts of human civilization, Antarctica has been isolated from other landmasses for millions of years [3].

Viruses infecting bacteria—(bacterio-)phages, which are estimated to be the most abundant biological entities on our planet—are omnipresent in any environment where bacteria reside. However, most of the recognized bacterial genera, even those with cultured representatives, have either no or very few phages that have been shown to infect them [4]. While infection by a lytic phage generally leads to the death of the host bacterium, temperate phages might become a part of the bacterial host genome and modulate host cell processes from their prophage form without causing immediate harm. Prophages were previously shown to impact such processes as metabolic capabilities and virulence of the host, which has consequences stretching far into the surrounding environment far beyond the individual lysogen cell [5].

While most temperate bacteriophages, such as arguably the most well-studied temperate bacteriophage—Lambda [6], can integrate themselves into the host chromosome, others may exist as circular (e.g., phage P1 [7]) or even linear (e.g., N15 [8]) episomes. Temperate phages that enter the chromosome of the host in a site-specific fashion encode an enzyme, integrase, which facilitates recombination between the relatively short attachment sites found within the phage (*attP*) and bacterial (*attB*) genomes (with the attachment site sequences being specific to particular temperate phages/their integrases [9]). Two evolutionarily distinct families of phage integrases are currently recognized on the basis of both evolutionary history and mechanism of action (catalytic residue), namely, tyrosine integrases and serine integrases. Serine integrases are usually thought of as more straightforward owing to the independence of the catalyzed reaction from the accessory factors encoded by the host. The reliance on accessory host factors, on the other

hand, is a hallmark for efficient integration reaction of most temperate phages encoding tyrosine integrases [10].

Psychrobacillus is a genus comprising Gram-positive rods, represented by at least nine different taxonomically recognized bacterial species (*P. insolitus* [11, 12], *P. psychrodurans* [11, 13], *P. psychrotolerans* [11, 13], *P. soli* [14], *P. lasiicapitis* [15], *P. vulpis* [16], *P. glaciei* [17], *P. faecigalinarum* [18], and *P. antarcticus* [19]), in addition to multiple other isolates yet not described in detail. Members of this genus were previously isolated from a plethora of different sources, such as soil [14, 20], red fox feces [16], ant heads [15], marine sediments [21], and Antarctic icebergs [17], to name a few. However, comparatively little is known about the representatives of this genus, and the reports on *Psychrobacillus* spp. are infrequent. As a consequence, even less is known about cultured bacteriophages infecting *Psychrobacillus* spp., characterizations of which are scarce in peer-reviewed scientific literature [21, 22].

In this study, we employ traditional micro- and molecular biology methods, as well as bioinformatics, to perform a baseline characterization and complete genome sequencing of a *Psychrobacillus* sp. L3 isolate and a bacteriophage infecting it, which we named “Spoks” (Latvian for “Ghost”, owing to its elusive microbiological behavior explainable by its temperate nature). The 19 bp long “core” overlap sequence used for the site-specific integration of phage Spoks into the chromosome of its host was determined *in silico* and validated. Both bacterial strain L3 and the associated phage Spoks were retrieved in the aftermath of the first Latvian Antarctic expedition from an ice-free soil sample collected at Waddington Bay, Graham Coast, Antarctica in 2018.

Methods

Soil sampling

Soil samples from the Antarctic Peninsula were collected on March 17, 2018, during the first Latvian expedition to Antarctica (February 10 - April 18, 2018). The soil was collected from a terraced area with peaty soil over rocky outcrops at coordinates $x = -64^{\circ} 5' 24''$ and $y = -65^{\circ} 15' 0''$. For soil sampling, three pits were dug, and soil samples from the depths of 0–4 cm, 4–10 cm, and 10–11.5 cm were collected using sterilized spoons and stored in airtight plastic bags.

Isolation of bacteria

Three grams of ice-free soil sample were mixed with 50 mL of liquid LB medium (Lysogeny Broth: 10 g/L tryptone (Sigma-Aldrich, St. Louis, MO, USA), 5 g/L yeast extract (Fluka, Buchs, Switzerland), 10 g/L NaCl (Sigma-Aldrich)) and shaken at 160 rpm at room temperature (RT) for 8 h. Subsequently, 50 μ L of the sample, diluted up to 10^{-6} , was spread on several Petri dishes containing solidified CR medium (Casamino-Yeast agar: 6 g/L casamino acids (Difco, Thermo Fisher Scientific, Waltham, MA, USA), 3 g/L yeast extract (Fluka), 3 g/L NaCl (Sigma-Aldrich), and 15 g/L agar (Sigma-Aldrich)). The plates were incubated in triplicate at three different temperatures (+37 °C, room temperature (RT), +4 °C) for a period of up to one month. Morphologically distinct colonies that could be observed with the eye were picked and subcultured three times.

Bacterial DNA extraction

Isolate L3 which was able to grow at +4 °C was one of the isolates selected for DNA extraction. Several such colonies were suspended in 400 μ L of 0.8% NaCl, and 1 μ L of proteinase K (18.7 mg/mL; Thermo Fisher Scientific) and 20 μ L of SDS (0.5% final concentration; Sigma-Aldrich) were mixed into the suspension. The mixture was incubated at +56 °C for 1 h with regular vortexing, cooled down to RT, and processed using the Genomic DNA Clean & Concentrator-10 Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol.

Amplification and Sanger-based sequencing of 16S rRNA

A PCR was carried out to amplify the 16S rRNA gene region of interest using approximately 100 ng of DNA from isolate L3 as a template using Taq DNA polymerase (Thermo Fisher Scientific). The thermocycler program was as follows: 3 min at +95 °C, 35 cycles of (30 s at +95 °C, 30 s at +55 °C, and 1.5 min at +72 °C), followed by 5 min at +72 °C and a hold at +4 °C. Universal 27F and 1492R primers ordered at Metabion (Steinkirchen, Germany) were used for amplification of the 16S rRNA gene. The PCR product was then subjected to native agarose gel electrophoresis. The gel was visualized under UV illumination, and the band corresponding to the amplified nearly complete 16S rRNA gene sequence of L3 (~1450 bp) was excised with a sterile scalpel. The product was extracted from the gel using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's protocol, including the additional optional step designated for further sequencing.

Sanger sequencing of the purified PCR product was performed in two separate reactions to generate forward and reverse reads using the 27F and 1492R primers respectively, on an ABI PRISM 3130xl sequencer (Thermo Fisher Scientific). The sequencing reactions

were prepared following the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Waltham, MA, USA) protocol. The resulting chromatograms were reviewed in GeneStudio (v. 2.2.0.0), and low-quality terminal sequences were trimmed before assembling the reads into a contig based on overlapping regions, using the highest-quality traces from either read to form the consensus sequence. The resulting near-complete 16S rRNA gene sequence of isolate L3 was queried against the EzBioCloud database [23] to identify the most closely related bacterial species.

Isolation and propagation of phage Spoks

Approximately five grams of soil were mixed with 100 mL of LB broth supplemented with 10 mM CaCl₂ and 10 mM MgCl₂ and then incubated at RT with shaking at 150 rpm for 16 h. The soil particles were sedimented via centrifugation at 2000 \times g for 30 min in an Eppendorf 5810R centrifuge (Eppendorf, Hamburg, Germany). The supernatant was filtered through a 0.45 μ m pore size syringe filter (Sarstedt, Nümbrecht, Germany) and concentrated using an Amicon Ultra-15 100 kDa MWCO filter (Merck, Kenilworth, NJ, USA) at 3000 \times g in an Eppendorf 5810R centrifuge to a final volume of 5 mL. The concentrated sample was then used for double agar overlay assays and spot tests. Luria 0.7% agar (10 g/L tryptone (Sigma-Aldrich), 5 g/L yeast extract (Fluka), 0.5 g/L NaCl (Sigma-Aldrich), and 7 g/L Bacto Agar (Sigma-Aldrich)) was used for the top layer, and 1.5% agar CR medium was used as the base. The top layer of agar was seeded with 100 μ L of an overnight culture of the bacterial isolate "L3" to screen for phages. Plaques with different morphologies were picked, suspended in 0.8% NaCl solution, and subcultured three times before propagation.

Phage Spoks was propagated by adding 1 mL of syringe-filtered (0.45 μ m pore size) phage plaque suspension to 200 mL of an exponentially growing L3 culture in LB medium at an OD₅₄₀ of ~0.1 and shaken at 140 rpm for 24 h. The phage-containing culture was sedimented by centrifugation at 2000 \times g for 30 min, and the supernatant was filtered through a 0.45 μ m pore size syringe filter. The lysate was then ultracentrifuged using an Optima TL-100 ultracentrifuge (Beckman Coulter, Brea, CA, USA) at 48,384 \times g for 1.5 h at 4 °C, after which the phage pellet was dissolved in 1.5 mL of LB liquid medium.

The dissolved pellet suspension (1 mL) was added on top of an 11.5 mL 0.7 g/mL CsCl solution in 20 mM Tris buffer in Ultra-Clear centrifuge tubes (Beckman Coulter), followed by centrifugation at 102,445 \times g for 20 h at 4 °C using an SW 40 Ti rotor (Beckman Coulter) in the Beckman Optima L-100XP ultracentrifuge. The resulting blue-colored band was collected and desalted twice with 5 mL of phosphate-buffered saline (PBS, pH 7.4) using an Amicon Ultra-15 100 kDa MWCO filter (Merck) by

centrifugation at $3000 \times g$ for 7 min in an Eppendorf 5810R centrifuge. The desalted phage sample was collected for DNA extraction, electron microscopy, and storage at -70°C with the addition of 40% glycerol.

Transmission electron microscopy

For both the phage (the desalted phage-containing band from CsCl gradient purification) and the host (exponentially-growing culture of isolate L3) sample TEM, 5 μL was adsorbed onto carbonized formvar-coated 300-mesh copper grids (Agar Scientific, Stansted, UK) for 5 min. The sample was then rinsed with 1 mL of 1 mM EDTA solution and negatively stained with 0.5% uranyl acetate solution for 1 min. After drying for several hours, the sample was examined using a JEM-1230 transmission electron microscope (JEOL, Akishima, Japan). Micrographs of the phage Spoks virions were captured with a Morada 11 MegaPixel TEM CCD camera (Olympus, Tokyo, Japan), and at least 10 different seemingly intact virions were measured using ImageJ 1.52v software, employing a scale bar for pixel-to-nanometer ratio calibration. The virion feature dimensions were summarized across seemingly intact virions in different fields of view and are presented as the average \pm standard deviation.

Short-read whole genome sequencing

Phage DNA was extracted from desalted CsCl bands using the same protocol as that used for host DNA extraction (see the “Bacterial DNA extraction” section). The quality and quantity of the extracted DNA were evaluated with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Pittsburgh, PA, USA) and a Qubit fluorometer (Invitrogen, Waltham, MA, USA) employing a high-sensitivity dsDNA quantification assay. Approximately 200 ng of DNA per sample was subjected to random shearing according to the 550 bp desired fragment size protocol using a Covaris S220 focused ultrasonicator (Covaris, Woburn, MA, USA). The fragmented DNA was utilized to create libraries compatible with the Illumina MiSeq platform for 250 bp paired-end sequencing. Libraries were prepared using the TruSeq DNA Nano Low-Throughput Library Prep Kit (Illumina, San Diego, CA, USA), following the recommended protocol and incorporating adapters from the TruSeq DNA Single Indexes Sets A and B (Illumina) to individually barcode the libraries for multiplexing. The quality and quantity of the prepared libraries were evaluated using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) with a High Sensitivity DNA Kit (Agilent) and a Qubit fluorometer (Invitrogen) employing a dsDNA high-sensitivity assay (Invitrogen). Libraries were pooled with other unrelated and uniquely barcoded libraries and sequenced on the Illumina MiSeq system (Illumina) with the 500-cycle MiSeq Reagent Kit v2 nano (Illumina).

Genomic DNA libraries for short-read sequencing of both the host and the phage were prepared and sequenced using the same protocol, with the only difference being in the index adapter used (although both the library preparations and sequencing of both were several years apart).

Long-read whole genome sequencing

Oxford nanopore was the selected technology for long-read sequencing of the isolate L3 genomic DNA. Library preparation was carried out following the ligation sequencing kit protocol (SQK-LSK109, ONT), without any deliberate fragmentation or size selection of the genomic L3 DNA used as an input ($\sim 1 \mu\text{g}$). The resulting library was loaded onto an R.9.4.1 flow cell (FLO-MIN106D, ONT) and sequenced using a MiniON Mk1C sequencer (ONT) with a fast basecalling model.

Bacteriophage Spoks genome assembly and functional annotation

Trimmomatic (v. 0.38 [24]) was used to trim the raw reads for an average quality of 30 in a sliding window of 5 bp, and the reads shorter than 100 bp were removed. SPAdes was used for *de novo* assembly of the phage Spoks (v. 3.11.1 [25]) trimmed reads. A pseudocircular contig of length 36,588 bp corresponding to the genome of phage Spoks was made nonredundant by manually removing the last 127 bp (an artifact corresponding to the length of the k-mers used for the resulting assembly). The resulting genome and the raw reads were used as inputs for PhageTerm (v. 1.0.12 [26]) to predict the physical termini type of the phage genome and reorient the genome accordingly. Custom primers were designed to hybridize upstream of the expected termini, ordered at Metabion, and used for Sanger-based sequencing to verify the predicted termini by sequencing phage genomic DNA directly, as well as from a PCR product amplified using the same primers from the ligated genomic phage DNA (Additional file 1).

Open reading frames (ORFs) were predicted using Glimmer [27] and GeneMark [28], considering ATG, GTG, CTG, and TTG as possible start codons and ORFs encoding a product longer than 40 aa. ARAGORN [29] and tRNAscan [30] were used to predict tRNA gene presence. These ORF and tRNA gene calling steps were performed using the aforementioned tool implementations within the DNA master sequence explorer (v. 5.23.6 [31]). The initial functional annotation of the predicted ORF products was carried out using a conserved domain database search [32], BLASTp [33], as well as HHpred [34] in early 2020. The annotated complete genome of phage Spoks was deposited to GenBank and became available starting from 27th May of the year 2020 under the following accession: MT410774.1.

In the early summer of 2024, the original 2020 annotation that was submitted to GenBank was compared with the results of Pharokka (v. 1.7.1 [35]) auto-annotation pipeline results, and no changes were deemed necessary. The presence of putative Shine–Dalgarno (SD) sequences complementary to the antiSD sequence of *Psychrobacillus* sp. L3 located at the 16S rRNA tail –3'-UUUCCUCC ACUAG-5' was inspected in the regions 20 bp upstream of the selected start codons for each Spoks ORF using `free_align.pl` script [36].

Selected phage Spoks protein phylogeny reconstructions

First, web-based BLASTp search against the non-redundant protein sequences was carried out under default settings for the terminase large subunit (TerL; QJT71634.1), major capsid protein (MCP; QJT71637.1), integrase (QJT71660.1) and helicase (QJT71675.1) of phage Spoks. All the hits to proteins of comparable length encoded by other phage or bacterial genomes from among the 100 top hits were selected and downloaded irrespective of their annotations. This dataset was deduplicated at 95% global identity using `cd-hit` (v. 4.8.1 [37]) to retain only representative sequences. Following this, another search querying the same proteins, but now restricted to cultured *Caudoviricetes* (taxid:2731619), was performed for the same proteins, and only the hits to proteins of comparable length encoded by cultured bacteriophages were downloaded. No deduplication of this dataset was performed.

Next, for each protein, both analogous protein amino-acid-sequence-containing datasets were aligned using MAFFT (v.7.525 [38]) and the resulting multiple sequence alignments were subjected to ML phylogeny reconstructions in IQ-TREE (v. 2.0.6 [39]), inferring the best-fitting substitution model using the built-in ModelFinder [40], allowing for polytomies and evaluating branch supports using 1000 ultrafast bootstrap [41] replicates.

The resulting trees were midpoint rooted and visualized in FigTree (v.1.4.4 [42]). Inkscape (v.1.0.1 [43]) was then used to combine, as well as annotate the trees.

Bacterial strain L3 genome assembly and functional annotation

The genome of *Psychrobacillus* sp. L3 was *de novo* assembled using a hybrid approach combining short Illumina+long Nanopore reads via Unicycler (v. 0.4.8 [44]) in a bold mode. Before the assembly, the short-read dataset was trimmed to remove any remaining adapters, and PhiX sequencing control spike-in reads, as well as to remove leading and trailing bases below a Phred quality score of 20 and drop any reads below 50 bp post-trimming using `bbduk` from the BBmap package (v. 38.69 [45]). The NCBI Prokaryotic Genome Annotation

Pipeline (`pgap_2024-04-27.build7426` [46]) was used to annotate the resulting *de novo* assembly, providing the closest relative on the basis of previous 16S rRNA gene sequencing data ("*Psychrobacillus glaciei*") as a species identifier.

The PHASTEST [47] webserver was used at the beginning of July 2024 to predict prophage regions in the genome of L3.

Determination of isolate L3 relatedness to other *Psychrobacillus* strains

All the RefSeq copies of the *Psychrobacillus* spp. assemblies available publicly on 14th of June 2024 were downloaded from the NCBI genome database ($n=37$) to be used as the context for the analysis alongside *Psychrobacillus* sp. L3 described in this study and *Psychrobacillus* sp. L4 (also isolated and sequenced in our laboratory from the same soil sample). The Anvio (v. 8 [48]) pangenomics workflow was run to obtain single-copy core genes shared among the *Psychrobacillus* spp. Single-copy core gene product aa sequences were then concatenated in the same order for each isolate and aligned using MUSCLE (v. 3.8.1551 [49]), and the resulting multiple sequence alignment was trimmed using `trimAl` (v. 1.4.rev15 [50]). The resulting clean MSA was used as an input for IQ-TREE (v. 2.3.4 [39]) using the WAG substitution model [51] and 1000 ultrafast bootstrap [41] replicates as a measure of branch support. The resulting tree was visualized in FigTree (v. 1.4.4 [42]).

The average nucleotide identities among the *Psychrobacillus* spp. isolates were calculated using `pyani` (v. 0.2.12 [52]) "ANIm" average nucleotide identity calculation method. The genomic data of the extended isolate dataset corresponding to the genus *Psychrobacillus* (taxid: 1221880) that was used for the ANI analysis was downloaded using "`genbank_get_genomes_by_taxon.py`" from the above-mentioned `pyani` package.

Bacterial growth and prophage induction

To evaluate bacterial growth and prophage induction in response to Mitomycin C (MMC; Sigma-Aldrich), the following procedure was employed: Three independent colonies of *Psychrobacillus* sp. isolate L3 were each grown separately in LB medium to the early exponential growth phase ($OD_{595} \sim 0.11\text{--}0.12$). From each culture, 190 μL were mixed with 10 μL of LB medium in four replicates for the "positive" control, and 10 μL of LB medium containing MMC for the different induction experiment conditions (final concentrations of 10, 1, 0.1, and 0.01 $\mu\text{g}/\text{mL}$ for the resulting 200 μL volume, respectively). Sterile LB medium and LB medium with MMC concentrations corresponding to those used for different experimental conditions (200 μL volumes) served as "negative" controls. Each experimental condition was performed in

four replicate wells of a 96-well plate, using a multichannel pipette. The optical density at 595 nm was measured using a VICTOR3V microplate reader (PerkinElmer, Waltham, MA, USA) at room temperature (~23 °C). Measurements were taken every 30 min for a total duration of 16.5 h post-induction.

Different attachment site conformation presence in individual L3 colonies

To demonstrate the spontaneous induction of Spoks virions from lysogenized cells and verify the presence of both prophage-free and lysogenized bacterial cells within a single colony - PCR amplification using primers specific to all four conformations of attachment sites (B-O-B', P-O-P', B-O-P', P-O-B') was performed for several randomly chosen colonies of L3 (individually). The resulting PCR products were subject to a native agarose gel electrophoresis, followed by excision of the expected zones from the gel using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). Sanger-based sequencing of the cleaned expected PCR products was performed using the same primers that were employed to amplify the respective region to investigate the homogeneity of the flanking regions around the attachment site overlap sequence (one primer per sequencing reaction). See Additional file 1 for used primer details.

Spoks induction quantification by droplet digital polymerase chain reaction

Droplet digital PCR (ddPCR) using custom primers ordered at Metabion (Additional file 1) was used to measure absolute concentrations and calculate the ratio of (pro)phage to bacterial genome copies as a proxy for Spoks induction on a QX200 system (Bio-Rad, Hercules, CA, USA). Both primer pairs were designed to bind to unique regions within either Spoks (MCP ORF) or L3 (*gyrB* gene) genomes and result in the same length product. Custom probes that bind to the sequence found between the primer pairs had a 5' 6-Fam (6-carboxyfluorescein, 495 and 520 nm excitation and emission maxima, respectively) and Hex (Hexachlorofluorescein, 533 and 559 nm excitation and emission maxima, respectively) fluorophore modifications for Spoks and L3, respectively. Both probes had a 3' BHQ^o-1 (Black Hole Quencher 1, 480–580 nm absorption range) modification.

Three colonies of the L3 strain were suspended in LB medium and grown to the early exponential phase (OD₅₉₅ ~0.12). The culture was then divided into seven 500 µL aliquots in Eppendorf tubes. MMC was added to two aliquots at a final concentration of 1 µg/mL and to another two at 0.1 µg/mL, as these concentrations were previously shown to induce a delayed but efficient bacterial lysis indicative of the likely increase in Spoks prophage induction. Three aliquots were used as controls.

After 15 h of growth, 300 µL from each sample was used for DNA extraction, following the protocol for bacterial genomic DNA isolation.

DNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific), diluted to ~10 ng/µL and measured precisely using the high-sensitivity dsDNA assay and Qubit fluorometer (Invitrogen). Samples were then diluted to a final concentration of 1 pg/µL as an input for ddPCR. For each reaction, a mastermix was prepared containing 1 µL of primers specific to the *gyrB* and MCP regions (final concentration: 900 nM), 1 µL of HEX and FAM probes (final concentration: 250 nM), 10 µL ddPCR Supermix for Probes (no dUTP) (Bio-Rad), and 5 µL nuclease-free water. To each reaction tube, 1 µL of diluted DNA or water (negative control) was added to a final volume of 22 µL. After gentle vortexing and brief centrifugation, 21 µL of the mixture was loaded into DG8 Cartridges (Bio-Rad), with 70 µL of droplet generation oil added to the designated wells. The cartridge was sealed with a DG8 Gasket (Bio-Rad) and processed in the QX200 Droplet Generator (Bio-Rad) to create droplets. Droplets (40 µL) were transferred to a ddPCR 96-well plate (Bio-Rad), heat-sealed with pierceable foil at 180 °C for 5 s using a PX1 PCR Plate Sealer (Bio-Rad).

PCR was performed using a T100 Thermal Cycler (Bio-Rad) with the following program: enzyme activation at 95 °C for 10 min, 40 cycles of denaturation at 94 °C for 30 s and annealing/extension at 56 °C for 1 min, followed by enzyme deactivation at 98 °C for 10 min and a final hold at 12 °C. A ramp rate of 2 °C/second was used. Post-amplification, the sealed plate was transferred to the QX200 Droplet Reader (Bio-Rad), supplemented with ddPCR Droplet Reader Oil (Bio-Rad).

Fluorescence signals (FAM/HEX) were quantified using QXManager Software v2.1 (Bio-Rad) with the absolute quantification mode. Concentrations were expressed as copies/µL. Spoks MCP ORF fragment concentration (FAM signal) divided by the L3 *gyrB* gene fragment concentration (HEX signal) represented a Spoks to L3 ratio. Positive droplet thresholds were set at 6,000 for MCP (FAM) and 4,000 for *gyrB* (HEX).

Phenotypic testing of isolate L3

Phenotypic testing of the isolate L3 was done using the Biolog GEN_{III} MicroPlate (Biolog, Hayward, CA, USA), which provides 94 phenotypic tests (71 carbon source utilization assays and 23 chemical sensitivity assays), following the manufacturer's instructions with some modifications. Several freshly grown L3 colonies (appearing after 48 h of incubation on agarized NBH medium: 10 g/L peptone (Fluka), 5 g/L yeast extract (Sigma-Aldrich), 5 g/L casein hydrolysate (Sigma-Aldrich), 5 g/L nutrient broth (Sigma-Aldrich), 5 g/L NaCl (Sigma-Aldrich), and

15 g/L Difco agar (Thermo Fisher Scientific) were picked and suspended in Biolog's inoculation fluid B, which was then inoculated into a GEN_{III} 96-well plate and incubated for 48+ hours at RT (~23 °C) in complete darkness. The plate was visually inspected once every ~12–16 h for the appearance of purple coloration, and the VICTOR3V microplate reader (PerkinElmer, Waltham, MA, USA) was used to measure the optical density at 490 nm wavelength after 48 h post-inoculation.

The experiment was repeated three times at an interval of several days, with different freshly grown inoculums used. Relative optical density values at 490 nm wavelength (OD₄₉₀) were calculated using the following formula:

$$\text{Relative OD}_{490} = \frac{\text{OD}_{490} \text{ for a given well} - \text{OD}_{490} \text{ of the negative control}}{\text{OD}_{490} \text{ of the positive control} - \text{OD}_{490} \text{ of the negative control}}$$

The A1 well served as a negative control, and the A10 well represented a positive control. The calculations were performed for each experiment independently and then expressed as percentages and averaged across three independent replicates. Average relative OD₄₉₀ percentages up to 33% were considered “negative” (no coloration of the well when inspected visually), values ranging from 33 to 66% were considered “intermediate (lilac/lavender-colored well contents)”, and values above 66% were considered “positive” (well contents turned purple).

Results

Isolation and physiology of the strain L3

The strain designated as “L3”, which was later identified as a genus *Psychrobacillus* representative, was isolated alongside other bacteria in the Summer of 2018 when looking for unusual hosts that would serve as indicator cultures for the potential isolation of novel bacteriophages in a hypothesis-free fashion from an ice-free Antarctic soil sample collected during the first Latvian Antarctic Expedition and brought to Latvia by the team [53].

The particular soil sample from which the bacteria described in this study originated was collected from a depth of 4–10 cm on a terraced surface, where peaty soil has formed over a cliff rock in Waddington Bay, Graham Coast, Antarctica.

An aliquot of the soil sample was mixed with liquid LB medium, and the supernatant was spread over a Casa-mino-Yeast agar-containing Petri dish. One of the colonies that formed after incubation of the aforementioned plate at +4 °C (white-yellowish, round, glossy) was subsequently designated as an isolate “L3” and immediately raised interest because of its psychrotolerant nature.

Isolate L3 can grow at temperatures of +4 - +30 °C on agar plates as well as in liquid media but has shown no

signs of growth at +37 °C. For routine experimentation, the isolate L3 cell lines were cultured at RT or +30 °C without noticeable differences. Transmission electron microscopy has revealed that L3 cells are rod-shaped and possess several pili or flagella attached to them (Additional file 2). The isolate tested positive for catalase activity, as evidenced by the formation of bubbles when mixed with 3% hydrogen peroxide.

When the isolate L3 was tested for metabolic activity using GEN_{III} MicroPlate (Biolog), it has convincingly demonstrated ability to grow at wells corresponding to the pH 6 (but not pH 5), 1% NaCl (but not 4% NaCl), as well as in wells containing Nalidixic acid and Aztreonam, respectively. Growth in the presence of 1% Sodium Lactate and Potassium Tellurite was observable, but both the fainter coloration of the well contents after incubating the inoculated plate and weaker OD₄₉₀ readings suggested scoring it as an “intermediate” result. The “intermediate” result was also observable in the well containing L-Glutamic Acid, whereas neither of the other 70 carbon sources seemed to show notable coloration compared with the negative control well. The relative OD₄₉₀ readings of intermediate wells were not very consistent across independent replicates, giving rise to standard deviations higher than those for convincing positives (Additional file 3). Faint light-lavender coloration was also observed in carbon assay utilization wells with L-Aspartic Acid, Glucuronamide, and Tween 40, as well as in chemical sensitivity wells containing Tetrazolium Violet and Tetrazolium Blue in at least one of the independent replicates, but the color was even weaker than that in the “intermediate” wells.

Complete *Psychrobacillus* sp. L3 genome assembly and its features

By utilizing a hybrid *de novo* assembly approach combining short-read Illumina (paired-end 250 bp read configuration) and long-read ONT reads performed using Unicycler, we were able to assemble what we believe to be a complete genome sequence of *Psychrobacillus* sp. L3 comprising several replicons.

The circular chromosome was determined to be 3,937,500 bp long and was accompanied by two more assembled circular replicons of 19,473 and 3,520 bp, respectively. Additionally, hybrid *de novo* assembly resulted in the bacteriophage Spoks genome being assembled as an additional circular replicon of 36,461 bp. The average GC% of the genome was determined to be 35.92%.

Standardized annotation via the latest NCBI PGAP pipeline available has resulted in the identification of 3,991 genes in total within the assembly comprising a chromosome and two plasmids (GCF_042945595.1), of which 3,806 were predicted to be protein-coding

genes, 121 RNA-coding genes (11 of each 5S, 16S, and 23S rRNA, 83 tRNA, and 5 ncRNA genes), and 64 pseudogenes.

No other putatively intact prophages were identified in the genome of the strain L3 by PHASTEST apart from the replicon corresponding to the circularized genome of bacteriophage Spoks we initially sequenced base-to-base and functionally annotated several years ago (MT410774.1).

Psychrobacillus sp. L3 within the context of other *Psychrobacillus* spp.

Within the context of psychrobacilli whose genomes were sequenced with sufficient quality, isolate L3 was found to be most closely related to the *Psychrobacillus* sp. L4, which was also isolated in our laboratory from the same material brought by the First Latvian Antarctic Expedition (an in-depth characterization of this strain

will be published elsewhere in the future). Both of these isolates formed a well-defined clade with the type strain of *Psychrobacillus glaciei*, namely *P. glaciei* strain PB01, which was previously isolated from an Antarctic iceberg sample [17].

In the concatenated single copy gene product aa sequence ML tree, strain L3 was ~0.041 and ~0.045 amino acid substitutions per site distant from strain L4 and *P. glaciei* strain PB01, respectively, (which shared the most recent common ancestor and were ~0.032 aa substitutions per site apart, Fig. 1). Similarly, ANIm identity-wise, isolate L3 was also shown to be most closely related to isolates L4 (~0.933 ANIm) and *P. glaciei* PB01 (~0.94 ANIm, Additional file 4).

Virion morphology of phage Spoks

Psychrobacillus phage Spoks virions exhibit characteristic siphophage features (Fig. 2), namely, an icosahedral

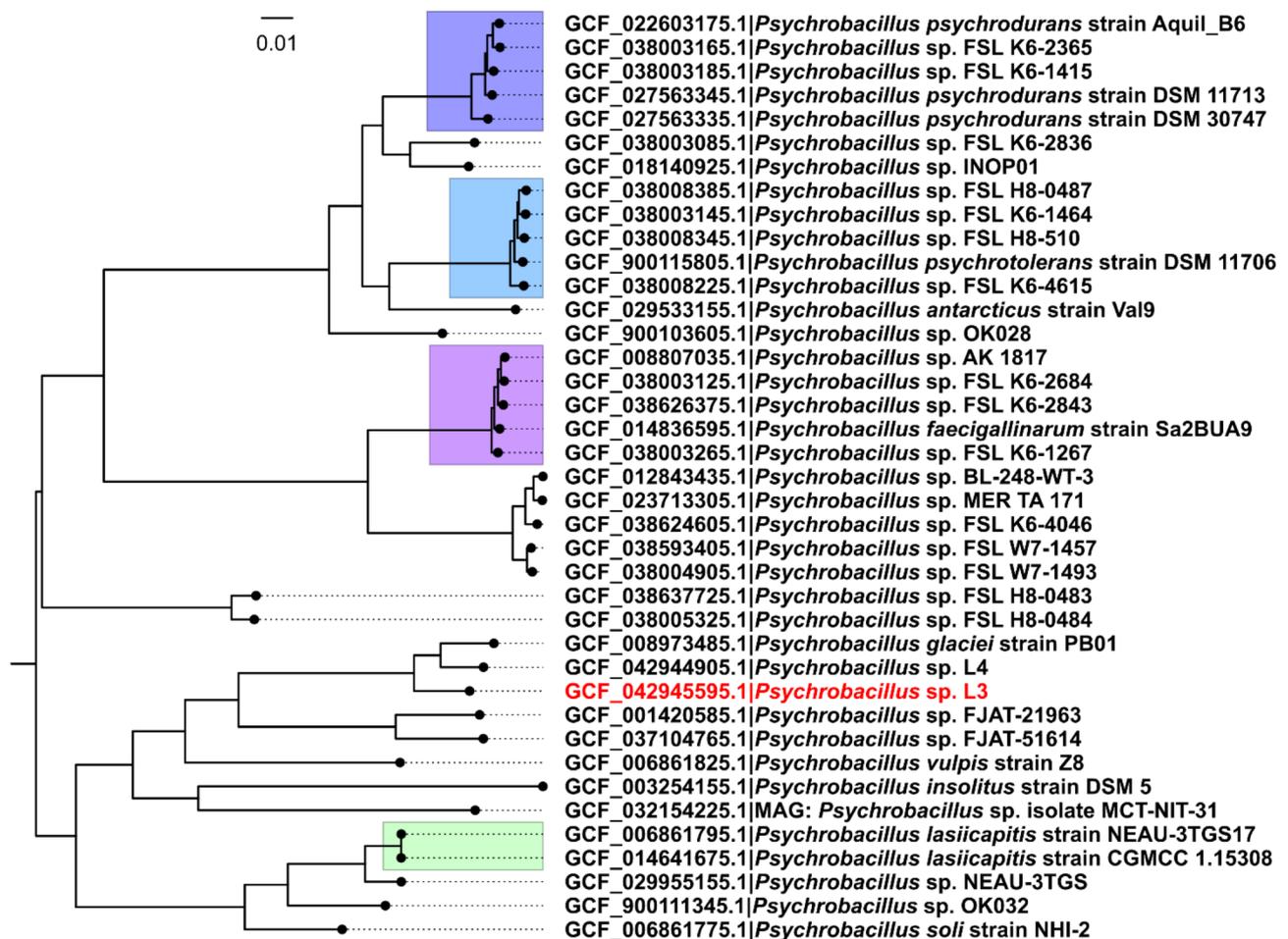


Fig. 1 Evolutionary relationships of *Psychrobacillus* sp. isolate L3 with other *Psychrobacillus* spp. bacterial isolates sequenced. A midpoint-rooted maximum-likelihood tree generated from concatenated core single copy gene product amino acid sequences ($n = 1267$) identified from the pangenome of *Psychrobacillus* spp. is shown. The scale bar corresponds to the number of amino acid substitutions per site. Clades comprising multiple closely related isolates showing a high degree of similarity to the recognized *Psychrobacillus* taxa are highlighted with arbitrarily colored boxes. Tip labels are in the form of “assembly accession|bacterial strain”. All branches longer than 0.001 substitutions per site had a UFBoot support of 100%

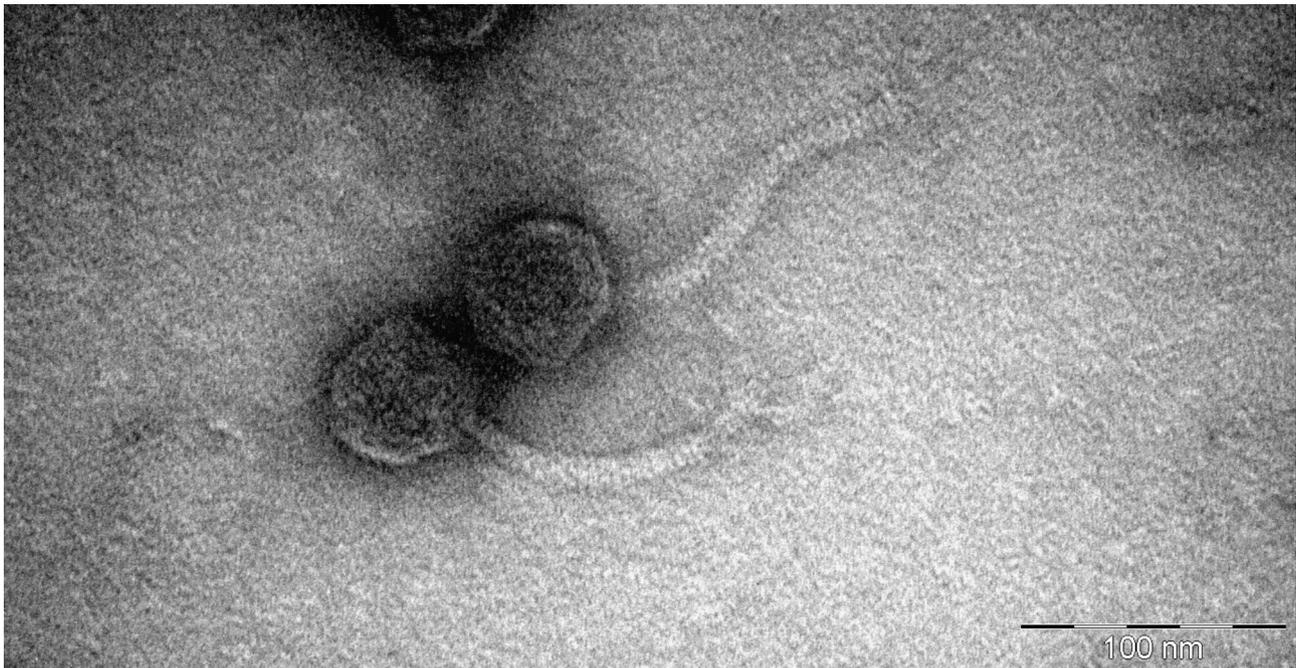


Fig. 2 *Psychrobacillus* bacteriophage Spoks virions. A representative transmission electron micrograph of two seemingly intact *Psychrobacillus* phage Spoks virions negatively stained with 0.5% uranyl acetate is shown; the scale bar represents 100 nm

capsid (diameter = 55 ± 2.5 nm) to which a long non-contractile tail is attached (length = 135 ± 6.5 nm, width = 11 ± 0.7 nm).

***Psychrobacillus* phage Spoks genomic features**

Inside the virions, the complete genome of phage Spoks is a linear 36,472 bp long dsDNA molecule with a GC% content of 35.83% and having an 11 base 3' cos overhangs (5'-CGGTAGGGGGA-3') at the termini.

The genome of Spoks was predicted to have a coding capacity of 92.77% and encode 52 proteins for 21 of which no putative function could be assigned reliably. The classical ATG is likely to serve as a start codon for 31 ORFs, TTG was predicted in 14 cases, and GTG— in seven cases. All but three of the Spoks ORFs were predicted to have at least a four-base upstream sequence match to the 16S rRNA tail of the host strain (5'-GAUCA CCUCCUUU-3').

At least twelve of the putative products for which we could make a functional prediction are thought to be involved in the virion morphogenesis, seven are participating in nucleic acid metabolism (in a broad sense), five are thought to serve as transcriptional regulators, and another four encode products with various other alleged functions not convincingly falling into other functional categories (namely, ORF4: Clp protease, ORF26: HTH-domain containing protein, ORF29: ImmA/IrrE family metallo-endopeptidase, ORF40: MBL fold metallo-hydrolase). Lysis proteins are represented by endolysin and holin, which are encoded by ORFs 23 and

22, respectively. ORF28 product was predicted to encode an integrase (Additional file 5).

***Psychrobacillus* phage Spoks within the context of the phage diversity uncovered thus far**

Querying the complete nucleotide sequence of the phage Spoks genome against the non-redundant nucleotide database limited to records of either bacteria (taxid:2) or *Caudoviricetes* (taxid:2731619) via BLASTn has revealed that bacteriophage Spoks has a highly divergent nucleotide sequence as its genome. The top-scoring hit was to the chromosome of *Psychrobacillus* sp. FSL K6-2843 isolated from raw milk (Accession: CP152021.1), however, this hit covered only 23% of the query with the average identity of 81.54% and resulted in the total score of less than 12% when scaled to the total score of Spoks self-to-self within the same comparison. This hit was followed by hits to other *Bacillaceae* family representatives from genera such as *Bacillus*, *Lysinibacillus*, and *Virgibacillus*.

A phage hit that had the highest score among the other phages was that of *Lysinibacillus fusiformis*-infecting phage SDFMU_Pfc (Accession: OQ884029.1, 21% query coverage of 73.60% identity) isolated from housefly larval intestines in November 2019 in China, but it scored merely a ~6.33% of the maximum observed self-vs-self score for Spoks. Among the top 25 hits, the only other cultured phage was *Virgibacillus* phage Mimir87 [54], which was also previously isolated and characterized in our laboratory at the Latvian Biomedical Research and Study Centre (Additional file 6).

Both the corresponding PHASTEST-predicted prophage region of *Psychrobacillus* sp. FSL K6-2843 (reannotated using Pharokka), demonstrating some similarity to Spoks, and the phage SDFMU_Pfc were reorganized to ensure genome representation collinearity with the genome of Spoks in its virions, for which the physical genome termini were determined experimentally, and used for the comparison of their genome organizations and proteome contents. Functionally-categorized protein-encoding gene synteny was revealed, with the recognizable functional module organization following the order of virion morphogenesis -> lysis -> lysogeny -> nucleic acid metabolism genes, when the strand encoding most of the Spoks ORF products is used as the direct strand and the 3' cohesive termini of Spoks is taken into the account for a linear representation of the genome. ORFs encoding products predicted to be involved in transcriptional regulation were not co-localized in a single module, although most of the ones recognized were located in the vicinity of proteins involved in lysogeny. Across these proteomes, the amino acid level similarity above 30% was confined largely to the ORFs encoding products involved in genome packaging and capsid morphogenesis, as well as nucleic acid metabolism and a likely regulatory protein encoded in a region between the lysis gene cassette and integrase (Fig. 3).

Phylogeny reconstructions for several selected proteins deemed to be functionally independent revealed the closest homologs of either Spoks TerL, MCP, Integrase, and Helicase are found in bacteria rather than phages. Not unexpectedly, the most recent common ancestor nodes were shared with the isolates of *Psychrobacillus* spp. in the case of each protein (Additional file 8). When only

the completely sequenced phages previously obtained in culture were considered, sufficiently close homologs (e.g., BLASTp identity of >70% over complete protein length) could only be found for TerL (UYL94139.1 and ANT39940.1 from *Geobacillus* phage vB_GthS_PK5.2 and *Bacillus* phage vB_BtS_BMBtp13, respectively) and Helicase (WNO29693.1 from *Bacillus* phage SDFMU_Pfc). Interestingly, no close homologs could be found among the cultured bacteriophages for Spoks MCP and Integrase (Fig. 4).

After failing to find any close phage Spoks relatives in the general nucleotide sequence database we have expanded our search to a more specific database focused on aggregating the uncultured virus diversity and providing rich metadata. An attempt to search for related phages in the IMG/VR v4 database of uncultivated virus genomes using the viral BLAST search associated with the resource has identified 97 hits, but neither of the alignments resulted in a query coverage of more than 8% (Additional file 7). Most of the top hits were restricted to the genomic region in Spoks encompassing ORFs encoding the predicted replication protein O (ORF42, QJT71674.1) and replicative DNA helicase (ORF43, QJT71675.1).

Phage Spoks as a prophage of *Psychrobacillus* sp. L3

The temperate bacteriophage Spoks was found to be capable of integrating into the chromosome of its host strain *Psychrobacillus* sp. L3 as well as spontaneously excising itself from it to undergo a lytic cycle.

Spoks integrates itself between the host ORFs encoding for DNA topoisomerase III (XHY47077.1; CP168493.1 complementary coordinates: 612,290–610,101) and

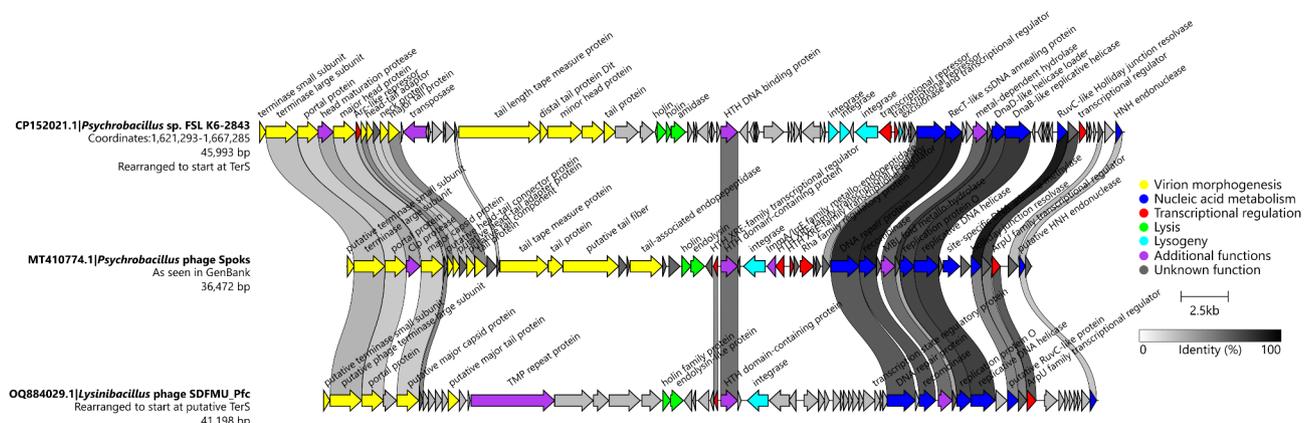


Fig. 3 Genome organization and annotated proteins of *Psychrobacillus* phage Spoks within the context of related (pro)phages. Genome organization and proteome content comparison of *Psychrobacillus* phage Spoks to the most similar prophage region (top) found in bacteria and the most similar cultured phage genome (bottom) is shown. The genomes are drawn to scale and the scale bar indicates 2500 base pairs. The predicted ORFs are shown as arrows pointing in the direction of the transcription. ORFs are colored on the basis of the functional group of their putative product according to the legend. Slanted labels above the arrows indicate the predicted function for the associated ORF product in the case where it had a function assigned (the prophage region was auto-annotated using Pharokka). Ribbons connect ORFs encoding proteins that share more than 30% amino acid sequence identity. Ribbons are shaded based on the identity of connected products in a gradient from lighter shades of gray to black according to the legend. The figure was generated using Clinker [55]

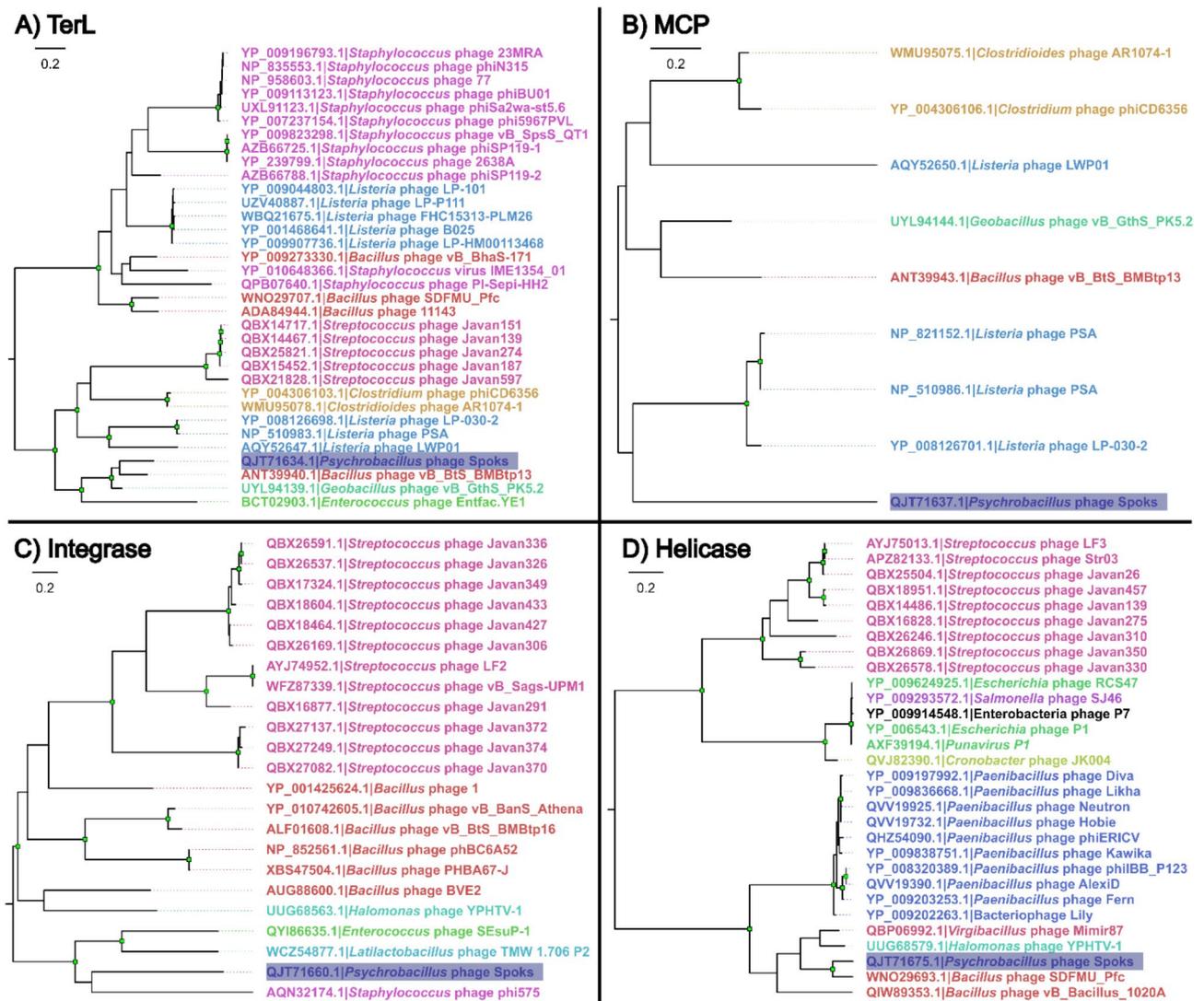


Fig. 4 Phylogenetic relationships of selected Spoks proteins within the context of counterparts from other cultured phages. Maximum-likelihood trees of the selected *Psychrobacillus* phage Spoks protein amino acid sequences and the most related non-redundant sequences originating from the proteomes of other completely sequenced phages are shown: **(A)** Terminase large subunit (TerL); **(B)** Major capsid protein (MCP); **(C)** Integrase; **(D)** Helicase. All the trees are drawn to their respective scales, and branch lengths correspond to the number of amino acid substitutions per site. Tips are labeled as “protein accession|originating phage” and are colored arbitrarily based on the given phage host genus. In all the trees, the tip containing the respective protein sequence of Spoks is highlighted in blue. The trees are midpoint-rooted. The distal nodes of branches with $\geq 95\%$ ultrafast bootstrap (UFBoot) support (out of 1000 replicates) are indicated by green squares. Descriptions of the selected phage marker protein amino acid sequence datasets and the generated MSAs, as well as features of the trees built, can be found in Additional file 9

a *BlaI/MecI/CopY* family transcriptional regulator (XHY47078.1; CP168493.1 direct strand coordinates: 612,480–612,893) in a site-specific and directional manner. The integrative recombination event takes place at the 19 bp 5'-ACACTACAATGAGTAATGT-3' (CP168493.1 coordinates: 612,447–612,465) sequence that serves as the Spoks attachment site overlap sequence, which gets doubled and flanks Spoks genome in its prophage form. The overlap of the attachment site is found between the phage ORFs encoding a hypothetical protein (ORF27 (QJT71659.1); MT410774.1 complementary strand coordinates: 20,920–20,786) and an integrase

(ORF28 (QJT71660.1); MT410774.1 complementary strand coordinates: 22,271–21,117) at the coordinates 21,021–21,039 of the phage Spoks genome as seen in its linear form found within the virions (MT410774.1, Fig. 5).

Inspection of a long-read dataset by seqkit grep revealed that more than half of the reads spanning the presumed Spoks attachment site overlap sequence had P-O-P' flanking sequence organization, followed by P-O-B' and B-O-P' in approximately equal amounts, summing up to an approximately 41% of the total reads in this region. Only a few reads had the corresponding genomic

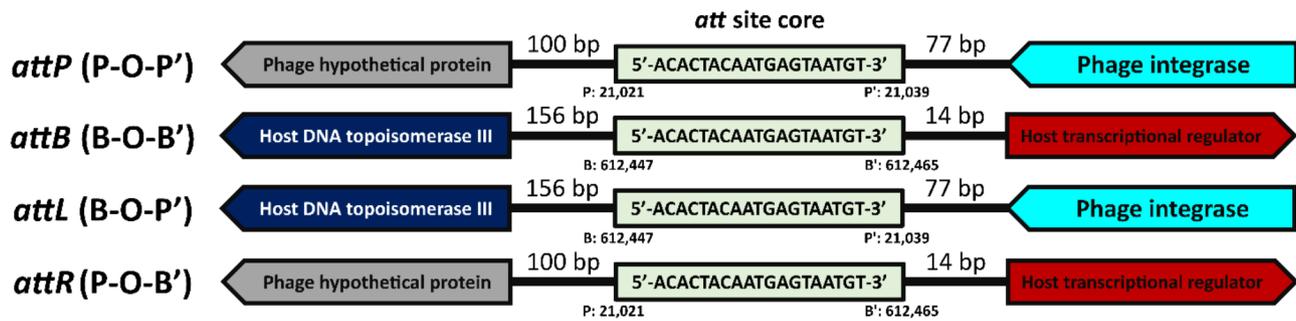


Fig. 5 Schematic representation of the genomic context of the Spoks and *Psychrobacillus* sp. L3 attachment sites. The arrows represent ORFs and point in the direction of transcription. Note that the figure is not drawn to scale

Table 1 Frequencies of different attachment site organizations within the sequencing reads

Attachment site organization	Part of	Sequence origin	Occurrences in the short-read dataset (SRR29989048)	Occurrences in the long-read dataset (SRR29989049)
P-O-P'	<i>attP</i>	Spoks	31	634
B-O-B'	<i>attB</i>	L3	1	3
B-O-P'	<i>attL</i>	L3+Spoks	25	212
P-O-B'	<i>attR</i>	Spoks+L3	18	229

During the search, the 19 bp core attachment site sequence was flanked by the corresponding upstream and downstream sequences of 21 and 20 bp, respectively, allowing up to six mismatches

region unscathed by Spoks integration as a B-O-B' conformation (Table 1).

Several individual colonies were picked and tested for the presence of all attachment site conformations (B-O-B' corresponding to the Spoks prophage-free host chromosome, P-O-P' corresponding to the unintegrated phage Spoks genome, and B-O-P' and P-O-B' corresponding to the Spoks integrated into the chromosome of isolate L3) via PCR reactions using specific primer pairs. The presence of all possible attachment site conformations was validated in all the colonies tested, although the relative amounts of the PCR products visually seemed to differ. The sequences of the products were verified using Sanger-based sequencing (Fig. 6).

When we tested whether Spoks can be further induced by outside stressors, 3.5 h after the addition of MMC at a concentration of 1 $\mu\text{g/mL}$, the growth of the L3

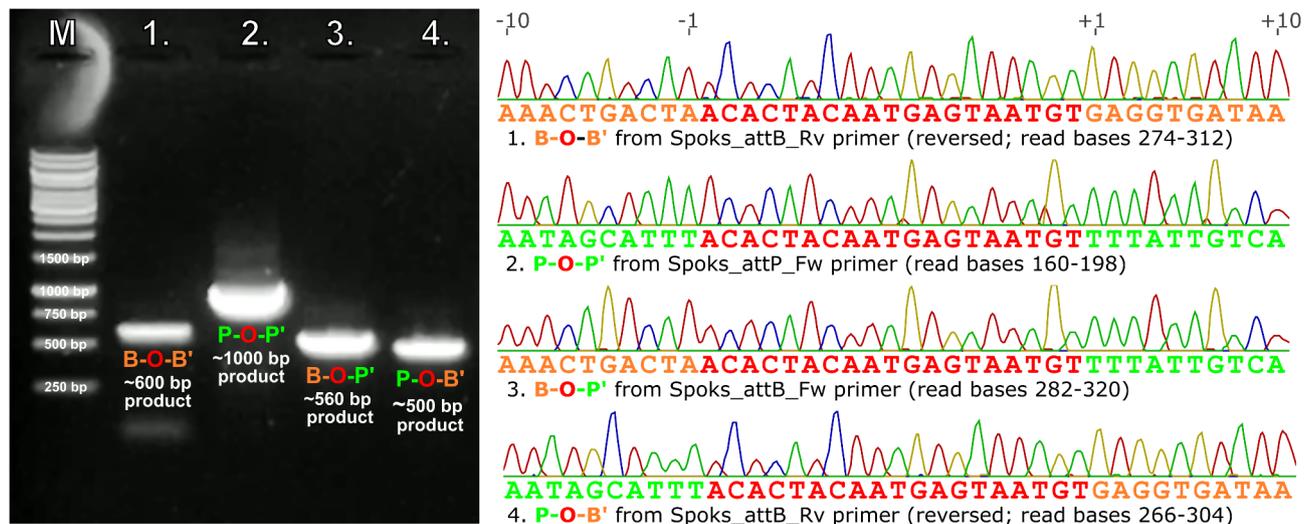


Fig. 6 Bacteriophage Spoks attachment site sequence variations in DNA isolated from *Psychrobacillus* sp. L3 colony. A representative example of the simultaneous presence of multiple phage Spoks attachment site conformations in DNA extracted from a single colony of L3 is shown. Left tile– Visualization of the PCR products corresponding to specific attachment site conformations. M– 1 kb DNA ladder, (1)– *attB*-containing PCR product (*attB*_Fw + *attB*_Rv amplification), (2)– *attP*-containing PCR product (*attP*_Fw + *attP*_Rv amplification), (3)– *attL*-containing PCR product (*attB*_Fw + *attP*_Rv amplification), (4) *attR*-containing PCR product (*attP*_Fw + *attB*_Rv amplification). Right tile– Sanger-based sequencing verification of the presence of multiple attachment site conformations. Chromatograms corresponding to the 19 bp long attachment site overlap sequence (red), as well as 10 flanking bases to each side are shown. The sequences of bacterial origin are colored in orange, and the sequences of phage origin– in green. For primer descriptions, see Additional file 1. The full-length original gel including the region shown in the figure is presented as an Additional file 11. For Fig. 6 the original was cropped to the regions of interest and annotated for clarity

lysogen containing prophage Spoks markedly decreased in comparison to that of the positive control. Similarly, a decrease in OD_{595} was observed after 4.5 h after the addition of MMC at the concentration of 0.1 $\mu\text{g}/\text{mL}$ to the exponentially-growing L3 population. Both these concentrations of MMC (1 and 0.1 $\mu\text{g}/\text{mL}$, respectively) eventually caused a drop in OD_{595} even below the levels demonstrated by the 10 $\mu\text{g}/\text{mL}$ MMC experimental group, which we assumed to be almost immediately lethal to L3 cell population (Fig. 7).

To verify the induction of the prophage at MMC concentrations sub-lethal to L3 *per se*, the lysate aliquots at the end of the experiment were sedimented on a tabletop centrifuge and the supernatant was subjected to a TEM analysis. The presence of seemingly intact siphovirus virions matching the dimensions of phage Spoks was observed across several randomly selected samples. The frequency of such virion occurrence in fields of view

was greater for induced samples than for positive control samples.

Additionally, droplet digital PCR was performed on either untreated *Psychrobacillus* sp. L3 control cultures or samples treated with MMC at the final concentration of 0.1–1 $\mu\text{g}/\text{mL}$ for 15 h. In each of the experimental groups, including the control, there was a significant difference between the number of genome copy numbers of phage Spoks and *Psychrobacillus* sp. L3 cells, with the ratio also being significantly different across the groups (Wilcoxon rank-sum test with Benjamini–Hochberg *p*-value adjustment for multiple comparisons resulted in the adjusted *p*-values of less than the $\alpha=0.05$) (Fig. 8).

Quantification has revealed that MMC-treated samples contained significantly more molecules corresponding to a region of phage Spoks major capsid protein ORF than the control. While there were only around 29 copies of

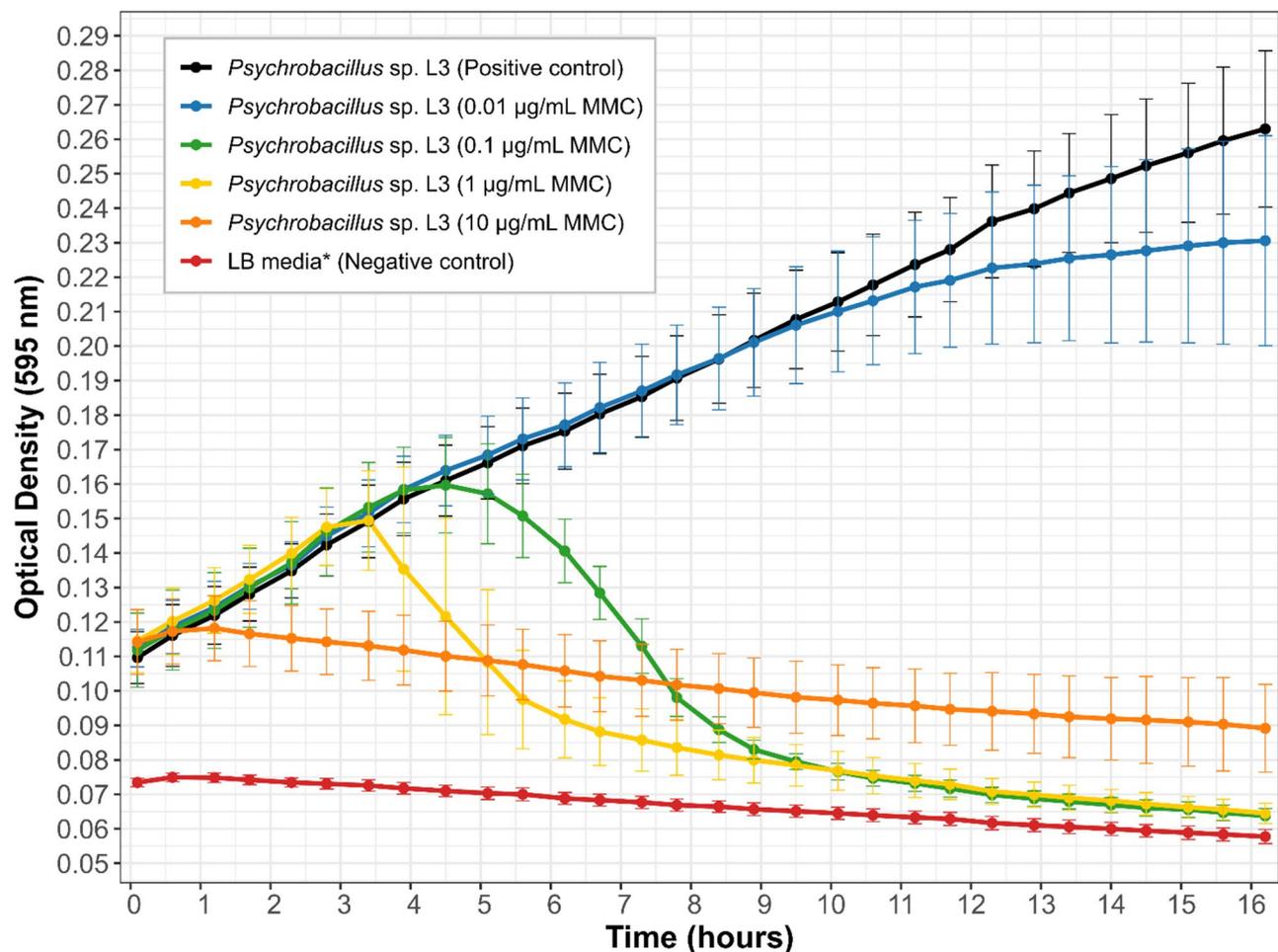


Fig. 7 Growth of *Psychrobacillus* sp. L3 Spoks lysogen with and without the addition of Mitomycin C. Samples were incubated at room temperature in a 96-well plate for 16.5 h. Points represent an average of 12 measurements per group and error bars indicate \pm one standard deviation. Three different isolate L3 colony line cultures were each measured in four wells per experimental group (with the exception of the negative control). The experimental groups are colored according to the legend. Note: The optical density of liquid LB medium at a wavelength of 595 nm did not significantly differ after the addition of MMC up to a final concentration of 10 $\mu\text{g}/\text{mL}$.

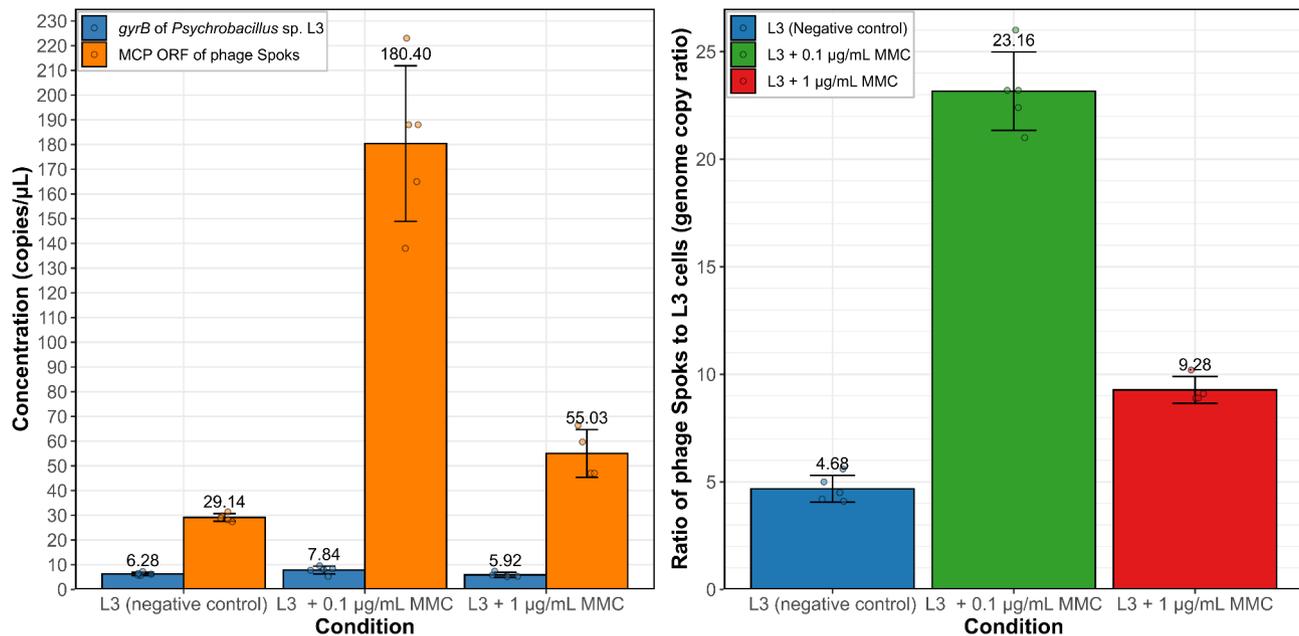


Fig. 8 Results of the ddPCR quantification of (pro)phage Spoks MMC induction from *Psychrobacillus* sp. L3. The left tile demonstrates concentrations of molecules harboring a sequence of phage Spoks major capsid protein or *Psychrobacillus* sp. L3 *gyrB* gene sequence fragment in copies per microliter of the extracted DNA diluted to 1 pg/microliter. The right tile shows the derived ratio of phage Spoks to host L3 cell genome. In both tiles, the results shown (both, bar heights and values above the bars) represent the mean of at least four replicates \pm one standard deviation (indicated by error bars)

MCP gene sequence harboring molecules in the microliter of control samples, on average, treatment with 1 $\mu\text{g}/\text{mL}$ MMC resulted in ~ 55 such copies per microliter, while treatment with 0.1 $\mu\text{g}/\text{mL}$ MMC has demonstrated more variation across the replicates, but averaged to 180 copies per microliter (range 138–223 copies/ μL). Interestingly, no significant differences were observed in the number of host chromosome copies as determined by the *gyrB*-specific primers (Kruskal-Wallis test p -value = 0.1053 > 0.05).

While the non-induced L3 cell population has around 4.7 phage Spoks genome copies per genome of the host cell after cultivation for 15 h, this ratio nearly doubles to an average of 9.3 when MMC is added at the concentration of 1 $\mu\text{g}/\text{mL}$. In the case of 0.1 $\mu\text{g}/\text{mL}$ MMC, however, the ratio reaches a whopping 23.2 phages to hosts on average, as approximated by their single-copy gene quantification.

Discussion

Following the notion of “whenever there is a bacterial host, there should be a phage” when working with less commonplace environmental bacteria opens up possibilities to obtain novel bacteriophages in culture that are highly divergent from their counterparts infecting the more “popular” isolation hosts. In this study, we demonstrate one such case by describing the isolation and providing a characterization of an interesting phage + host pair from an ice-free soil sample collected in Antarctica.

The continuous growth of the affordability of sequencing technologies continues to uncover the vast diversity of microbial life in a multitude of environments previously largely thought to be very limiting to the growth of microorganisms owing to “the great plate count anomaly” [56]. However, the rise of culture-free microbial diversity research going hand-in-hand with the sequencing data deluge, while undeniably allowing appreciation of the real diversity of microorganisms in various environments [57], provides very limited insights into interactions taking place in any given environment.

Bacteriophage Spoks is one of the first bacteriophages shown to infect a representative of the bacterial genus *Psychrobacillus* whose complete genome sequence has become publicly available (nearly simultaneously with that of phage Perkons [22], and, later on, phage PVJ1 [21]). All three of the cultured and characterized phages known to be associated with this host genus to date are temperate.

The isolate L3 that serves as the host for phage Spoks was initially identified based on the closest near full-length 16S rRNA gene sequencing to be a strain of *P. psychrodurans*, and later on internally reclassified as a strain of *P. glaciei* with the release of the 16S rRNA sequence of the latter [53]. Interestingly, the 16S rRNA sequences obtained for *Psychrobacillus* sp. isolates L3 and L4 isolated from the same material were the same; thus, the isolates were thought to be very closely related and naively assumed to differ only in their phage susceptibilities [22].

However, further whole-genome sequencing and phylogenetic analysis involving 1267 core single-copy genes of *Psychrobacillus* spp., surprisingly, revealed that both strains are very different genomically and differ greatly from their 16S rRNA sequence closest match represented by *P. glaciei* PB01. Yet, it is not currently clear whether L3 (and also L4) should be proposed as a representative for novel bacterial species, which would require further phenotypical characterization of these strains using methods employed by other researchers for other *Psychrobacillus* spp. strains previously to allow for direct comparisons of their phenotypical features, followed by their additional deposition to several larger internationally-recognized microbial collections.

This has once again highlighted the limited resolution 16S rRNA gene Sanger-based sequencing is able to provide among some bacterial taxa (such as several bacilli), that are known to have high similarity in their 16S rRNA gene sequences coupled with considerable overall genomic variability [58]. This observation raises the question of how many bacterial strains that might be different to the point of possibly representing novel species within particular genera might be preserved in comparatively modest institutional microbial collections across the world, especially the ones that have not yet adopted extensive WGS for their strain characterization.

While initially sequenced and genomically characterized in 2019, and then publicly released in the spring of 2020 as a part of the MSc. thesis of N. Zrelavs, the uniqueness of the genomic sequence phage Spoks demonstrated has warranted a wait for more context of microbial and phage diversity to appear in the public biological sequence repositories before presenting this phage to wider audiences. However, more than four years of wait have not resulted in the addition of any other relatives of Spoks to public biological sequence repositories, neither in the form of free viruses nor in the form of prophages residing within bacterial sequences released publicly since then.

Within the currently adopted rank-specific demarcation criteria framework used by the ICTV Bacterial Virus Subcommittee (e.g., <70% similarity and <95% similarity to any other bacteriophage over complete genome lengths for the ranks of genus and species, respectively [59]), the intergenomic distance of Spoks to any of the phages sequenced thus far warrants a proposal for the creation of a novel bacteriophage genus and species with Spoks as a sole isolate representing both currently.

Despite the uniqueness of the genomic nucleotide sequence, the comparative genomics and homology searching approaches utilized for predicted products encoded by the Spoks ORFs have allowed for a relatively straightforward and fruitful functional annotation of the proteome of Spoks. Multiple Spoks proteins involved in

nucleic acid metabolism were found to be the most conserved ones (e.g., with most homologs found in other biological objects). Unexpectedly, structural and other proteins involved in Spoks virion morphogenesis were largely unrelated to those associated with other cultured phages, and the top hits were mostly associated with prophages or prophage remnant regions from bacterial genomes, highlighting that many diverse phages have yet to be obtained in culture. Visually, the transmission electron micrograph appearance of phage Spoks was rather unremarkable and expectedly similar to a multitude of siphophages having long non-contractile tails attached to icosahedral capsids, which seems to be the most frequently cultured tailed phage morphotype (formerly comprising the paraphyletic family *Siphoviridae*, which was relatively recently abolished [4, 60, 61]).

The presence of a putative integrase-encoding gene has immediately been able to hint at the temperate nature of Spoks and explain why such frequent inconsistencies in the microbiological behavior during routine experimentation with the Spoks+L3 system were observed. The integrase of phage Spoks is encoded by the ORF28 (QJT71660.1) and is a 384 aa long protein with a molecular weight of ~45 kDa in which a site-specific recombinase XerD domain (COG4974) can be identified. By sequence similarity to tyrosine integrases from other phages, Tyr³⁶² is suggested as a residue that forms an intermediate with the DNA during a recombination event, and Spoks integrase can be classified as a tyrosine-type recombinase/integrase [62]. Interestingly, no recombination directionality factor/excisionase could be reliably identified in the genome of Spoks. However, several helix-turn-helix (HTH) domain-containing proteins have been identified, and the HTH domain is characteristic of at least several studied phage excisionases found in phages encoding a tyrosine integrase [63].

Reconstruction of individual selected protein phylogenies indicates that most homologs of Spoks proteins show greater similarity to their counterparts found in bacterial genomes across several different genera. Not unexpectedly, different bacilli seem to harbor proteins homologous to those found in Spoks. Such proteins are likely to originate from either active prophages or their cryptic remnant regions, that have lost their ability to be independently activated, residing in multiple bacterial strains. The reconstructed phylogenies, however, had non-coinciding branching patterns and contexts beyond the immediate vicinity of different *Psychrobacillus* spp. strains who shared the most recent common ancestor of these selected proteins with Spoks in each case. Such observations, we think, are indicative of a rather complex and likely independent evolutionary history of these selected proteins, globally suggesting a propensity of

Spoks to horizontal gene exchange and highlighting the expected mosaicism of its genome [64, 65].

The inability to perform a “classical” characterization of free-standing phage Spoks (e.g., in terms of its environmental stability, lifecycle length, and burst size) by the merit of Spoks plaque enumeration when grown with its host strain has prompted to shift efforts upon characterizing the phage-host interactions at the level of genomes themselves. Sequencing and the subsequent *de novo* assembly of the L3 host strain genome were partaken using both short-read Illumina and long-read Nanopore technologies [44].

Sequencing read mapping onto the *de novo* assembled chromosome of the strain L3 and manual inspection of the sequence alignment map in UGENE (v37.0 [66]) has revealed an unusual peak in base depth for a single 19 bp long region (5'-ACACTACAATGAGTAATGT-3'), with multiple mapped reads being soft-clipped at either side of the region (Additional file 10). Further investigation has revealed that the region is also present in the bacteriophage Spoks genome and that the region is flanked either by phage genome sequences from both sides or a combination of phage and bacteria sequences on either side, with a couple of reads corresponding to the bacterial flanking sequences only. Soft clipped sequences of the reads around the region could be, thus, put in several categories in which they did not differ, bar the long-read sequencing errors. The region was, therefore, determined to serve as the attachment site of phage Spoks, at which its integration into the chromosome of the host takes place. While the exact lengths of the attachment sites around the overlap sequence which are necessary for integrative recombination to occur were not investigated in detail at this point, the *attB* was found to be located in the intergenic region between the ORF596 (encoding a DNA topoisomerase III, XHY47077.1, from a complementary strand) and ORF597 (encoding a transcriptional regulator, XHY47078.1, from a direct strand) ORFs, whereas the *attP* is located between the Spoks ORF27 (encoding a hypothetical protein, QJT71659.1, from a complementary strand) and ORF28 (encoding an integrase itself, QJT71660.1, from a complementary strand), maybe even stemming beyond them.

Despite the bacterial DNA being isolated from individual colonies of L3, strikingly many reads corresponding to the Spoks genome in its non-integrated form were found (e.g., ~60% of the reads spanning the Spoks attachment site “core” overlap region of 19 bp corresponded to *attP*). This observation suggests that spontaneous induction and propagation of Spoks take place at relatively high frequencies. The presence of all four attachment site conformations in multiple individual single colonies was further investigated and validated via particular *att* site type-specific primers and PCR followed by Sanger-based

sequencing even without any particular stressors used to induce Spoks excision.

With the results on the integration of Spoks into the chromosome of *Psychrobacillus* sp. L3 presented within this study, we consider the case about the alleged lack of proof for the temperate nature of *Psychrobacillus* phage Spoks previously raised by Liu and colleagues [67] to be definitively resolved. Although the presence of a seemingly intact integrase gene annotated in the genome of a culturable phage (e.g., as well as Perkons, for that matter, characterization of which will be published elsewhere) would, in our opinion, suffice to make any phage researcher wary about the possibility of a temperate nature, the functional integrity of this integrase within the genomic context of phage Spoks has been demonstrated and raised additional questions for further studies.

The observations made so far suggest a highly site-specific nature of the phage Spoks integration, which takes place at a single site on the chromosome of isolate L3, likely precluding the possibility of prolonged (pro) phage Spoks existence as an extra-chromosomal episome without the ensuing lytic replication. The precise genome termini identification of bacteriophage Spoks revealed 11 base 3' cos overhangs (5'-CGGTAGGGG GA-3') at either terminus, which implies very likely the circularization of the genome as the first step upon its delivery into the host cytoplasm. While the exact course of events and protein actors involved in the lysis-lysogeny decision within the Spoks + L3 system are not known, it might be speculated that several transcriptional regulators encoded by Spoks might get simultaneously produced. The decision of whether “to lyse or not to lyse” might boil down to how fast and what concentrations of the competing regulatory proteins are reached, akin to the well-studied Lambda phage switch [68]. Yet, the Spoks prophage state seems to be very unstable in the L3 chromosome, and no Spoks proteins that might be beneficial to the Spoks-harboring lysogenic cell are readily identifiable.

Upon the addition of MMC to the final concentrations of 0.1 µg/mL up to 1 µg/mL, strain L3 cell culture experienced a notable drop in OD after a few hours when compared to the positive control, which we attribute to the induction of phage Spoks; however, quantitative PCR of the DNA extracted from the lysates at multiple times post-induction would be necessary to quantify the number of Spoks genome copies having an intact attachment site with a P-O-P' conformation. MMC at a final concentration of 0.01 µg/mL did not seem to impact the growth of isolate L3 Spoks lysogens for at least nine hours.

To validate this MMC-stimulated Spoks induction hypothesis, we have opted for a droplet digital PCR quantification of DNA molecules containing sequences

corresponding to the genomic regions within single-copy genes of the phage and the host, respectively. To this end, we have designed primers specific to the *gyrB* region of the L3 and MCP ORF region of Spoks that would result in a product of the same size and differently labeled probes for the region in between to allow simultaneous quantification of both the phage and the host in the same reaction by the proxy of their genome copy numbers and their relative ratio changes upon application of MMC.

In the case of no spontaneous induction taking place, with Spoks being firmly integrated into (nearly) every host cell, we would expect a ratio of MCP to *gyrB* to be close to one—each host cell has Spoks integrated into its chromosome in a single location. This, however, was not the case, as the phage genome copy number exceeded that of the host more than fourfold in the cultures that were not subjected to any recognized inducing agents from without the L3+Spoks system. When MMC was added as an outside inducing agent at various concentrations a shift in the ratio of phage to host cells was expected, with the increase indicating that Spoks starts to favour lytic replication more frequently. As it could be seen from the L3 growth curves, the lethal MMC dose of 10 µg/mL caused a near-immediate growth arrest and very gradual OD reduction, which we considered to be in line with the likely mostly the “individual” MMC action. However, when MMC concentrations sublethal to L3 cells (1 µg/mL or 0.1 µg/mL) were used, the OD reduction was greatly postponed but then demonstrated a quite steep rather than gradual drop, and, importantly, a drop below the levels achieved by the immediately cytotoxic MMC action at concentration of 10 µg/mL. Such a shift in the culture growth dynamics, we hypothesized, ideally coincided with a prophage response to an inducing agent demonstrating how induction in a population of lysogens would look like if we assume OD as a good proxy of cell viability.

Consistent with these expectations, we later observed that the ratio of phage to cell ratio increased to 23 when MMC was added at the minute concentration of 0.1 µg/mL resulting in the beginning of L3 population collapse after what seemed like an unhindered growth for at least 4.5 h (Figs. 7 and 8). The addition of MMC at the concentration of 1 µg/mL, however, was consistent with unimpaired growth dynamics of L3 for at least 3.5 h but resulted in the Spoks to L3 ratio of only 9.28 after 15 h of its presence.

Although we, obviously, should not expect that each genome copy has ended up as an infective intact Spoks virion, these results undeniably suggest a more frequent sway of Spoks lysis-lysogeny switch towards the former given that we see an increased Spoks replication taking place. In our scenario, we have shown that there consistently is a significant and notable difference in the yield

of the temperate phage Spoks from the aliquotes of the same lysogen L3 cell population when different doses of an inducing agent are applied. The difference in the ratio of phage to cells is likely explainable by the fact that the population size L3 manages to reach before the dawn of its demise due to Spoks is larger, and it collapses slower in the presence of 0.1 µg/mL MMC, when compared to 1 µg/mL. This, in turn, seems to support Spoks replication for a longer time, resulting in a significantly larger total phage progeny after 15 h. These results seem to be suggestive, that the yield of a temperate phage from its lysogen cell population can be increased by tuning the dosage of the inducing agent and optimal conditions can be sought to result in lysogenic cell growth dynamics that maximize the infective virion output if a need arises.

In the case of phage Spoks, consistent acquisition of phage plaques using isolate L3 was impossible, as is seemingly impossible to “cure” its host strain from the prophage. We attribute this stable coexistence of phage Spoks in both prophage and free form in association with L3 populations as a result of the high frequency of spontaneous induction taking place even in the absence of obvious inducing agents [69]. Such a feature of the L3+Spoks system that makes it very annoying to work with in the lab actually represents a very reasonable strategy in terms of the native ecological niche in which multiple other closely related bacteria some of which might be susceptible hosts are found. In such a scenario, Spoks ensures both its “safe” replication together with the host (strain L3 in this case) in its prophage form, as well as incessantly tends to scout for other susceptible hosts to infect and spread further by the means of other strains found nearby. Whereas during the solitary confinement to the L3 associated with the experimental setups in the lab, such scouting is set to fail and the future of progeny Spoks virions is bleak, as nearly exclusively any nearby cell to be encountered would either harbor Spoks in its prophage form - resulting in likely superinfection exclusion, or already have Spoks undergo a lytic replication within it already, which would lead to a failure of a particular virion to fulfill its sole purpose of multiplication. Although the frequency of host cell lysis, obviously, is expected to have some local fleeting perturbations due to cell-to-cell signaling and other possible chemical cues [70], they are unlikely to frequently sway the balance in favor of a sufficient local chain of Spoks lytic replication that would be required to consistently observe Spoks plaques on a lawn of L3 cells in double agar overlay assays.

Being so different from the other phages, disentangling the events taking place regarding phage Spoks within the L3 host seems interesting enough to warrant an RNA sequencing investigation of what phage transcripts are being produced within the L3 Spoks lysogens under

regular bacterial growth conditions, as well as what happens after the prophage induction via MMC. Such an experiment is also likely to narrow down the “suspect list” for a potential Spoks excisionase candidate, if any. For example, a tyrosine integrase IntB13 from an integrative conjugative element ICE*clc*, was shown to be the sole phage protein responsible for the directionality of ICE*clc* recombination (e.g., integration by recombination of its 18 bp overlap sequence with *attB* or excision by recombination of *attL* with *attR*), which is controlled by differential integrase expression [71]. An additional question of interest is what is/are host-encoded accessory factor/s (e.g., an analog of “integration host factor” from the *E. coli* + phage Lambda system), if any, that are necessary for the integration of Spoks into the chromosome of the host. Thus far, we have had no luck in definitively curing the strain L3 from the phage Spoks despite multiple (nearly a hundred) trials using various inducing agents, washing the cells, subculturing and screening random L3 colonies. The acquisition of Spoks-free L3 derivatives, however, would be of paramount importance for studying the impacts of multiple external factors on Spoks + L3 phage-host system interactions, as well as elucidating whether Spoks prophage gives any fitness advantage to L3 lysogens.

Conclusions

This study demonstrates successful isolation in culture, complete genome sequencing, and a follow-up characterization of a bacteria + temperate bacteriophage pair from Antarctic soil. Both, *Psychrobacillus* sp. isolate L3 and the associated phage Spoks, genomes demonstrate genomic distinctiveness and evolutionary divergence from the closest cultured relatives described to date, which necessitates their consideration for expansion of the bacterial and phage taxonomies, respectively. Phage Spoks integrates into the chromosome of L3 in a site-specific fashion via recombination at the 19-bp long attachment site core found in genomes of both. The lysogenic conversion of L3 by phage Spoks in unstable, spontaneous induction events seem to be common in the population of L3 lysogens. Application of Mitomycin C intensifies induction to the point of Spoks-prophage-containing *Psychrobacillus* sp. L3 cell population demise. Overall, this study expands the knowledge on the culturable bacteria and phage diversity associated with the Antarctic environments.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11425-z>.

Supplementary Material 1: Additional file 1. Description of custom primers used in this study

Supplementary Material 2: Additional file 2. Collage of transmission electron micrographs depicting *Psychrobacillus* sp. L3 cells that were negatively stained with 0.5% uranyl acetate. Scale bars represent 2 microns

Supplementary Material 3: Additional file 3. Relative optical densities (490 nm) of *Psychrobacillus* sp. L3 inoculated into Biolog GenIII 96-Well MicroPlate. Relative OD₄₉₀ values were calculated as “(OD₄₉₀ for a given well minus the OD₄₉₀ of the negative control (A1 well)) divided by the (positive control (A10 well) OD₄₉₀ minus the OD₄₉₀ of the negative control (A1 well))” for each experiment independently and then expressed as percentages and averaged across three independent replicates. Average relative OD₄₉₀ percentages up to 33% were considered “negative” (no coloration of the well when inspected visually), values ranging from 33 to 66% were considered “intermediate” (lilac/lavender-colored well contents); values above 66% were considered “positive” (well contents turned purple). For *Psychrobacillus* sp. L3, these values corresponded to the results of the visual inspection on the basis of the intensity of purple coloration observed

Supplementary Material 4: Additional file 4. Pairwise ANIm percentage identities and clustering of the *Psychrobacillus* spp. Labels corresponding to the genome of *Psychrobacillus* sp. L3 are in red. The green rounded rectangles in the matrix denote the row and column showing values corresponding to average nucleotide identities of L3 to other *Psychrobacillus* spp. isolates completely sequenced. Colors between the tips of the clustering tree and the matrix rows/columns represent different species-level *Psychrobacillus* taxa (if applicable)

Supplementary Material 5: Additional file 5. Genome annotation and predicted functions of *Psychrobacillus* phage Spoks ORFs. The DeltaG column indicates a change in the free energy required to bring the two strands of nucleotides (the region upstream of the respective Spoks ORF start codon putatively containing the SD sequence and antiSD sequence of *Psychrobacillus* sp. L3) together

Supplementary Material 6: Additional file 6. Results of a BLASTn search in which the complete genome nucleotide sequence of phage Spoks was queried against non-redundant sequences of bacterial and viral origins found in GenBank. The green row indicates a hit against oneself

Supplementary Material 7: Additional file 7. Results of a BLAST search against the IMG/VR v4 Viral Nucleotide Database with an E-value threshold of 1E-5

Supplementary Material 8: Additional file 8. Phylogenetic relationships of selected Spoks proteins within the context of counterparts from public biological sequence repositories (Top 100 BLASTp hits deduplicated at 95% global sequence identity by cd-hit). Maximum-likelihood trees of the selected *Psychrobacillus* phage Spoks protein amino acid sequences and the most related non-redundant sequences originating from the proteomes of other biological objects are shown: (A) Terminase large subunit (TerL); (B) Major capsid protein (MCP); (C) Integrase; (D) Helicase. All the trees are drawn to their respective scales, and branch lengths correspond to the number of amino acid substitutions per site. Tips are labeled as “protein accession|originating bacteria or phage” and are colored arbitrarily based on the given bacterial or phage host genus. In all the trees, the tip containing the respective protein sequence of Spoks is highlighted in blue. The trees are midpoint-rooted. The distal nodes of branches with $\geq 95\%$ ultrafast bootstrap (UFBoot) support (out of 1000 replicates) are indicated by green squares. Descriptions of the selected phage marker protein amino acid sequence datasets and the generated MSAs, as well as features of the trees built, can be found in Additional file 9

Supplementary Material 9: Additional file 9. Details of the input datasets, multiple sequence alignments, and phylogenies associated with the maximum-likelihood trees related to this study

Supplementary Material 10: Additional file 10. Representative view of the Nanopore dataset read mapping unto the *de novo* assembled Spoks-free chromosome of the isolate L3. A representative sequence alignment map region demonstrating reads exclusively mapping to either the actual B-O-P' or P-O-B' and clipped precisely before or after the integrated Spoks sequence is shown. Note a marked increase in sequencing depth of approximately twice the average depth specifically for the Spoks attachment site region spanning from 612,447 to 612,465bp

Supplementary Material 11: Additional file 11. The full-length original gel for the cropped and annotated region of the gel shown in Fig. 6. Tracks 1–5 are as described in Fig. 6 legend, other tracks - products resulting from

screening random L3 isolate colonies using only the primer pair amplifying P-O-P' Spoks attachment site conformation in hopes of finding a cell line which would not have a spontaneously inducible Spoks associated with it

Acknowledgements

We would like to acknowledge the personnel of the Latvian Biomedical Research and Study Centres' "Genome Center" core facility for the maintenance of the sequencing machines and their assistance in loading and starting the sequencing runs. The authors are especially thankful to MSc. Ance Roga for her invaluable help with the droplet digital polymerase chain reaction experiments.

Author contributions

N.Z., K.S., A.D., and A.K. wrote the initial version of the manuscript. Each co-author of this manuscript (N.Z., K.S., J.J., K.L., J.K., M.K., D.F., A.D., A.K.) has made substantial contributions to this study by either participating in the conception and design of the study, acquisition of the samples and resources, performing experimental work, analyzing and interpreting the data, preparation of visualizations, or a combination of these activities. All the authors participated in the preparation of the manuscript and agreed to the submitted version.

Funding

This work was supported by the LSC project Nr. Izp-2022/1-0111 ("Structural and functional studies of proteome encoded by three novel temperate DNA bacteriophages isolated from Antarctic soil") awarded to Dr. biol. Andris Kazaks. The funding agency had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Data availability

Relevant data pertaining to this manuscript is found as a part of an overarching BioProject accession PRJNA1130732. The short (Illumina 250 bp paired-end) and long (ONT) read datasets used to assemble the *Psychrobacillus* sp. L3 genome are available from the Sequencing Read Archive under accessions SRX25469097 and SRX25469096, respectively. The complete annotated *Psychrobacillus* sp. L3 genome comprising the chromosome and two plasmids is available from GenBank (CP168493.1, CP168494.1, CP168495.1) and RefSeq (NZ_CP168493.1, NZ_CP168494.1, NZ_CP168495.1). The complete annotated genome of *Psychrobacillus* phage Spoks is available from GenBank under accession number MT410774.1. All other data generated or analyzed during this study are included in this published article and its Additional files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Latvian Biomedical Research and Study Centre, Ratsupites 1 k-1, Riga LV-1067, Latvia

²Polar Research Center, Faculty of Science and Technology, University of Latvia, Jelgavas 1, Riga LV-1004, Latvia

Received: 21 November 2024 / Accepted: 28 February 2025

Published online: 18 April 2025

References

1. Nayfach S, Roux S, Seshadri R, Udwy D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol.* 2021;39(4):499–509.

- Heinrichs ME, Piedade GJ, Popa O, Sommers P, Trubl G, Weissenbach J, et al. Breaking the Ice: A review of phages in Polar ecosystems. *Methods Mol Biol.* 2024;2738:31–71.
- Vyverman W, Verleyen E, Wilmotte A, Hodgson DA, Willems A, Peeters K, et al. Evidence for widespread endemism among Antarctic micro-organisms. *Polar Sci.* 2010;4(2):103–13.
- Zrelavs N, Dislers A, Kazaks A. Motley crew: overview of the currently available phage diversity. *Front Microbiol.* 2020;11(October).
- Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence.* 2013;4(5):354–65.
- Casjens SR, Hendrix RW. Bacteriophage Lambda: early pioneer and still relevant. *Virology.* 2015;479–480:310–30.
- Łobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnerr H, et al. Genome of bacteriophage P1. *J Bacteriol.* 2004;186(21):7032–68.
- Ravin NV. N15: the linear phage-plasmid. *Plasmid.* 2011;65(2):102–9.
- Groth AC, Calos MP. Phage integrases: biology and applications. *J Mol Biol.* 2004;335(3):667–78.
- Fogg PCM, Colloms S, Rosser S, Stark M, Smith MCM. New applications for phage integrases. *J Mol Biol.* 2014;426(15):2703–16.
- Krishnamurthi S, Ruckmani A, Pukall R, Chakrabarti T. *Psychrobacillus* gen. nov. and proposal for reclassification of *Bacillus insolitus* Larkin & Stokes, 1967, *B. psychrotolerans* Abd-El Rahman et al., 2002 and *B. psychrodurans* Abd-El Rahman et al., 2002 as *Psychrobacillus insolitus* comb. nov., *Psychrobacillus*. *Syst Appl Microbiol.* 2010;33(7):367–73.
- Larkin JM, Stokes JL. Taxonomy of psychrophilic strains of *Bacillus*. *J Bacteriol.* 1967;94(4):889–95.
- Abd El-Rahman HA, Fritze D, Spröer C, Claus D. Two novel psychrotolerant species, *Bacillus psychrotolerans* sp. nov. and *Bacillus psychrodurans* sp. nov., which contain ornithine in their cell walls. *Int J Syst Evol Microbiol.* 2002;52(6):2127–33.
- Pham VHT, Jeong SW, Kim J. *Psychrobacillus soli* sp. nov., capable of degrading oil, isolated from oil-contaminated soil. *Int J Syst Evol Microbiol.* 2015;65(Pt9):3046–52.
- Shen Y, Fu Y, Yu Y, Zhao J, Li Y, Li Y, et al. *Psychrobacillus lasiacapitis* sp. nov., isolated from the head of an ant (*Lasius fuliginosus*). *Int J Syst Evol Microbiol.* 2017;67(11):4462–7.
- Rodríguez M, Reina JC, Béjar V, Llamas I. *Psychrobacillus vulpis* sp. nov., a new species isolated from faeces of a red fox in Spain. *Int J Syst Evol Microbiol.* 2020;70(2):882–8.
- Choi JY, Lee PC. *Psychrobacillus glaciei* sp. nov., a psychrotolerant species isolated from an Antarctic iceberg. *Int J Syst Evol Microbiol.* 2020;70(3):1947–52.
- Gilroy R, Ravi A, Getino M, Pursley I, Horton DL, Alikhan NF, et al. Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ.* 2021;9:e10941.
- da Silva MBF, da Mota FF, Cypriano J, Abreu F, Seldin L. *Psychrobacillus antarcticus* sp. nov., a psychrotolerant bioemulsifier producer isolated from King George Island, Antarctica. *Int J Syst Evol Microbiol.* 2023;73(11).
- Chiba A, Peine M, Kublik S, Baum C, Schlöter M, Schulz S. Complete genome sequence of *Psychrobacillus* sp. Strain INOP01, a Phosphate-Solubilizing bacterium isolated from an agricultural soil in Germany. *Microbiol Resource Announcements.* 2022;11(4):e00207–22.
- Liu W, Zheng X, Dai X, Zhang Z, Zhang W, Xiao T, et al. Isolation and characterization of the first temperate virus infecting *Psychrobacillus* from marine sediments. *Viruses.* 2022;14(1):108.
- Zrelavs N, Lamsters K, Karuss J, Krievans M, Dislers A, Kazaks A et al. PVJ1 Is Not the First Tailed Temperate Phage Infecting Bacteria from Genus *Psychrobacillus*. Comment on Liu Isolation and Characterization of the First Temperate Virus Infecting *Psychrobacillus* from Marine Sediments. *Viruses.* 2022;14(3):108.
- Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H et al. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol.* 2017.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using spades de novo assembler. *Curr Protocols Bioinf.* 2020;70(1):e102.
- Garneau JR, Depardieu F, Fortier LC, Bikard D, Monot M, PhageTerm. A tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci Rep.* 2017;7(1):1–10.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics.* 2007.

28. Besemer J, Borodovsky M. GeneMark. Web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* 2005;33:451–4.
29. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and TmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 2004;32(1):11–6.
30. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25(5):955–64.
31. Lawrence J. DNA Master [Internet]. Available from: https://phagesdb.org/DNA_Master/. Accessed 12 April 2022.
32. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 2011;39(SUPPL 1):225–9.
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
34. Södberg J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005.
35. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharaoh: a fast scalable bacteriophage annotation tool. *Bioinformatics.* 2023;39(1):btac776.
36. Starmer J, Stomp A, Vouk M, Bitzer D. Predicting Shine-Dalgarno sequence locations exposes genome annotation errors. *PLoS Comput Biol.* 2006;2(5):454–66.
37. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22(13):1658–9.
38. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013.
39. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
40. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017.
41. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 2013.
42. Rambaut A. FigTree v. 1.4.4 [Internet]. 2018. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 10 May 2021.
43. Inkscape Project. Inkscape [Internet]. 2020. Available from: <https://inkscape.org>. Accessed on 10 May 2021.
44. Wick RR, Judd LM, Gorrie CL, Holt KE, Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13(6):1–22.
45. Bushnell B, BBMap: A Fast, Accurate, Splice-Aware Aligner. Conference: 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, March 17–20, 2014 [Internet]. 2014; Available from: <https://sourceforge.net/projects/bbmap/>. Accessed 10 May 2021.
46. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44(14):6614–24.
47. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* 2023;51(W1):W443–50.
48. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for omics data. *PeerJ.* 2015;3:e1319.
49. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
50. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25(15):1972–3.
51. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*; 2001.
52. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2015;8(1):12–24.
53. Zrelavs N. Trīs no antarktiskās ledus-brīvas augšnes jaunizdalīto Caudovirales kārtas bakteriofāgu raksturošana [in Latvian; Characterization of three novel bacteriophages from order Caudovirales isolated from Antarctic ice-free soils] [Internet]. University of Latvia; 2020. Available from: <http://dspace.lu.lv/dspace/handle/7/50713>
54. Zrelavs N, Cernooka E, Dislers A, Kazaks A. Isolation and characterization of the novel Virgibacillus-infecting bacteriophage Mimir87. *Arch Virol.* 2020;165(3):737–41.
55. Gilchrist CLM, Chooi YH. Clinker & Clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics.* 2021;37(16):2473–5.
56. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol.* 1985;39(1):321–46.
57. Chen IMA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system V.7: content updates and new features. *Nucleic Acids Res.* 2023;51(D1):D723–32.
58. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol.* 2007;45(9):2761–4.
59. Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. *Viruses.* 2021.
60. Ackermann HW. 5500 Phages examined in the electron microscope. *Arch Virol.* 2007;152(2):227–43.
61. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, Alfenas-Zerbini P, et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch Virol.* 2023;168(2):1–9.
62. Smyshlyayev G, Bateman A, Barabas O. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol Syst Biol.* 2021;17(5):e9880.
63. Lewis JA, Hatfull GF. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.* 2001;29(11):2205–16.
64. Hendrix RW. Bacteriophage genomics. *Curr Opin Microbiol.* 2003;6(5):506–11.
65. Hatfull GF. Bacteriophage genomics. *Curr Opin Microbiol.* 2008;11(5):447–53.
66. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, et al. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics.* 2012;28(8):1166–7.
67. Liu W, Zheng X, Dai X, Zhang Z, Zhang W, Xiao T et al. Reply to Zrelavs PVJ1 Is Not the First Tailed Temperate Phage Infecting Bacteria from Genus Psychrobacillus. Comment on Liu Isolation and Characterization of the First Temperate Virus Infecting Psychrobacillus from Marine Sediments. *Viruses.* 2022 May;14(5):866.
68. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. Switches in bacteriophage lambda development. *Annu Rev Genet.* 2005;39(1):409–29.
69. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol.* 2015;197(3):410–9.
70. León-Félix J, Villicaña C. The impact of quorum sensing on the modulation of Phage-Host interactions. *J Bacteriol.* 2021;203(11):e00687–20.
71. Miyazaki R, van der Meer JR. A new Large-DNA-Fragment delivery system based on integrase activity from an integrative and conjugative element. *Appl Environ Microbiol.* 2013;79(14):4440–7.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.