Arch Virol (2015) 160:2269–2282
DOI 10.1007/s00705-015-2497-9

ORIGINAL ARTICLE

Identification and characterization of a novel *Geobacillus thermoglucosidasius* bacteriophage, GVE3

Leonardo Joaquim van Zyl[1] · Falone Sunda[1] · Mark Paul Taylor[2] · Don Arthur Cowan[1,3] · Marla Iris Trindade[1]

Received: 11 March 2015 / Accepted: 12 June 2015 / Published online: 30 June 2015

Abstract The study of extremophilic phages may reveal new phage families as well as different mechanisms of infection, propagation and lysis to those found in phages from temperate environments. We describe a novel siphovirus, GVE3, which infects the thermophile *Geobacillus thermoglucosidasius*. The genome size is 141,298 bp (G+C 29.6 %), making it the largest *Geobacillus* spp-infecting phage known. GVE3 appears to be most closely related to the recently described *Bacillus anthracis* phage vB_BanS_Tsamsa, rather than *Geobacillus*-infecting phages described thus far. Tetranucleotide usage deviation analysis supports this relationship, showing that the GVE3 genome sequence correlates best with *B. anthracis* and *Bacillus cereus* genome sequences, rather than *Geobacillus* spp genome sequences.

Introduction

The ubiquity of bacteriophages (phages) in nature and their impact on various trophic levels is widely appreciated [58, 76]. As phages directly affect microbial communities that play a pivotal role in biogeochemical cycles, they in turn play a role in altering those cycles [18, 33, 74]. Phages are also known to be prevalent in many extreme environments, including soda lakes, terrestrial hot springs, deep-sea hydrothermal vents, hot/cold deserts and hypersaline systems, with some of the highest phage numbers being recorded in these habitats [40]. However, few studies have investigated the functional relationships between extremophiles and the phages that infect them, compared to the wealth of data that exist for phages and hosts in temperate environments.

Morphological and sequence-based characterization of phages from many temperate environments has shown the predominance of tailed viruses (order *Caudovirales*) with members of the families *Siphoviridae*, *Myoviridae* and *Podoviridae* most often recorded [3, 4, 68, 71]. Morphological characterization of extremophilic phages has led to the introduction of several new families, including *Lipothrixviridae*, *Rudiviridae* and *Fuselloviridae* [6]. The study of extremophilic phages has also revealed new mechanisms for host lysis, as in the case of the deep-sea thermophilic bacteriophage GVE2 [17], and have demonstrated interactions between phage and host proteins that are unlike those normally observed for mesophilic phages [32]. *Thermus thermophilus* phage ϕYS40 promoters are

Electronic supplementary material The online version of this article (doi:10.1007/s00705-015-2497-9) contains supplementary material, which is available to authorized users.

✉ Leonardo Joaquim van Zyl
vanzyllj@gmail.com

Falone Sunda
falone.sunda@gmail.com

Mark Paul Taylor
marktaylorimbm@gmail.com

Don Arthur Cowan
don.cowan@up.ac.za

Marla Iris Trindade
prof.marlatt@gmail.com

[1] Institute for Microbial Biotechnology and Metagenomics (IMBM), University of the Western Cape, Robert Sobukwe Road, Bellville, Cape Town, South Africa

[2] TMO Renewables Limited, 40 Alan Turing Road, The Surrey Research Park, Guildford, Surrey GU2 7YF, UK

[3] Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria 0002, South Africa

thought to be leaderless (i.e., contain no -10 or -35 elements), unlike those found in T4 and many other mesophilic phages, which require phage- and host-encoded sigma factors for transcription [70].

It is therefore likely that the further study of phages infecting extremophiles will reveal new phage families and alternate strategies for infection or the "decision" between lysis and lysogeny and will shed further light on the behaviour and the role of host organisms in their natural environments [44, 57, 62]. Extremophilic phages may also provide a source of novel enzymes that are adapted to extreme conditions and serve as the basis for the development of genetic systems by providing strong regulatable promoters, and as vehicles for the introduction of large DNA segments into bacterial hosts for which no genetic tools currently exist [53, 59].

Geobacillus thermoglucosidasius is a Gram-positive thermophile that has been isolated from soil, oil fields, compost heaps, deep-sea sediment and hot springs [54, 67]. This promising "platform" organism is capable of producing a range of useful metabolites, including ethanol, isobutanol and polylactic acid [19, 42, 79; http://tinyurl.com/po6a52q]. Several *Geobacillus* species phages have been described (GVE1, GVE2, GBSV1, GBK2, DE6 and ϕ OH2), sequenced and studied [20, 33, 43, 45, 72, 73, 82–84], although none infecting *G. thermoglucosidasius* have been reported. Here, we describe the first phage (GVE3) known to specifically infect *G. thermoglucosidasius*.

Materials and methods

Media, bacterial strains and plasmids

G. thermoglucosidasius strains were cultured in tryptone glycerol pyruvate (TGP) medium. One liter of TGP broth contains 17 g tryptone, 3 g soy peptone, 2.5 g K_2HPO_4 and 5 g NaCl. The pH was adjusted to 7.3 before autoclaving, after which 4 g Na-pyruvate and 4 mL glycerol (filter sterilized) were added. For solid media, 15 g/L agar was added before autoclaving. TGP was used for general maintenance of cultures. Cultures were incubated at 60 °C with vigorous aeration.

DNA manipulations and sequencing

Plasmid preparations, restriction endonuclease digestions, gel electrophoresis and ligations were performed using standard methods or following the manufacturers' recommendations. Total DNA from all bacterial strains was prepared as described [34]. Phage DNA was prepared by first preparing a phage lysate from 1 L of culture as described below. The phage was pelleted by centrifugation at 13000 × *g* for 30 min after addition of PEG8000 (7.5 ml of 20% PEG8000 per 30 ml lysate) and incubation at 4 °C overnight. The pellet was resuspended in 1 ml of SM buffer (5.8 g of NaCl per liter, 1.2 g of $MgSO_4$ per liter, 50 mL of 1 M Tris-HCl, pH 7.5, 0.1 g of gelatin per liter). The suspension was treated with DNaseI and RNaseA (Fermentas; final concentration, 0.1 μg/ml) at 37 °C for 1 hour. The presence of contaminating bacterial DNA was tested by amplifying the 16S rRNA gene. The suspension was treated with proteinase K (Fermentas; final concentration, 1 μg/ml) at 55 °C for 2 hours before addition of 70 μl of 20 % (wt/vol) SDS and incubation at 37 °C for 1 hour. An equal volume of phenol:chloroform:isoamylalcohol (P:C:I; 25:24:1) was added, the sample was centrifuged (15 ml Sterillin tube, Eppendorf 5810R centrifuge, 5000 rpm for 10 min) to separate the phases, and the top, aqueous phase was removed and transferred to a fresh tube. A second P:C:I extraction was performed. An equal volume of C:I (24:1) was added to the supernatant and re-centrifuged. The top phase was removed and transferred to a fresh tube, and a tenth volume of 3 M sodium acetate (pH 5.2) and two volumes of 100 % ethanol were added. This mixture was incubated at 4 °C to precipitate overnight. The sample was centrifuged at 13,000 rpm for ten minutes to pellet the DNA, and the pellet was resuspended in 40 μl of TE buffer. The phage DNA was electrophoresed on a 1 % low-melting-point agarose gel, excised and purified from the gel using standard agarase (Fermentas) treatment. The pellet was resuspended in 40 μl of TE buffer. The quality and integrity of the DNA was checked using a Bioanalyzer prior to library preparation. Sanger DNA sequencing was performed using an ABI PRISM 377 automated DNA sequencer (University of Stellenbosch Central Analytical Facility), and next-generation sequencing was performed using either a Roche GS Junior with a LibL library preparation kit or an Illumina MiSeq with a Nextera XT 150 bp library kit (Illumina). The raw reads were trimmed and de-multiplexed at the sequencing facility (the University of the Western Cape Next Generation Sequencing facility), resulting in two (2 × 150) paired fastq files. Sequences were analyzed with DNAMAN (version 4.1, Lynnon BioSoft), Newbler (Roche) or CLC Genomics Workbench version 6.5 (CLC Bio). Open reading frames were predicted using the built-in tools in the CLC Genomics workbench and confirmed by BLASTp search against the NCBInr database. Smaller ORFs not identified by the software were assigned through manual translation of DNA sequences and BLASTp analysis of putative ORFs [5]. The complete genome sequence of *G. thermoglucosidasius* bacteriophage GVE3 is available in the GenBank database under accession no. KP144388. RAST [http://rast.nmpdr.org/; 7] and PHAST (http://phast.wishartlab.com/) [87] were used to identify closely related phages. RADAR was used to identify protein repeat regions (http://www.ebi.ac.uk/Tools/pfa/radar/). Direct repeats were identified using REPFIND [http://zlab.bu.edu/repfind/form.html; 9] with a 15-bp minimum repeat length. Inverted repeats were identified using UGENE (http://ugene.unipro.ru/) with a 20-bp minimum and 80 % similarity as search parameters. tRNA genes were predicted using the tRNAscan-SE program [http://lowelab.ucsc.edu/tRNAscan-SE/; 46] and ARAGORN [http://mbio-serv2.mbioekol.lu.se/ARAGORN/; 39]. Transmembrane regions were predicted using the TMHMM server v2.0 [http://www.cbs.dtu.dk/services/TMHMM/; 36]. Intron prediction was done using the RNAweasel server [http://megasun.bch.umontreal.ca/RNAweasel/; 38]. For phylogenetic tree construction, the full-length amino acid sequences of selected terminase proteins were aligned using MEGA6, and the tree was constructed using the built-in program [24, 88].

Polymerase chain reaction

Polymerase chain reaction (PCR) was performed using Phusion DNA polymerase (New England Biolabs. Generally, 50 ng of DNA was used in a 50-µl reaction volume containing 2 mM $MgCl_2$, 0.125 µM each primer, 0.2 mM each deoxynucleoside triphosphate, and 1 U of DNA polymerase. Reactions were carried out in a Bio-Rad T-100 thermocycler, with an initial denaturation at 98 °C for 3 min, followed by 30 cycles of denaturation (30 s at 98 °C), annealing (30 s), and variable elongation times at 72 °C as required.

Phage purification, maintenance and characterization

Phage lysates were prepared by culturing *G. thermoglucosidasius* to an OD_{600nm} of 0.4 and addition of phage particles at a multiplicity of infection (MOI) of 10. Infected cultures were incubated until complete culture lysis was observed. A 1/10 volume of chloroform was added to lyse residual bacterial cells and release bacteriophage. Cell debris and chloroform were removed by centrifugation (5000 rpm for 10 min), and the supernatant was recovered as the phage stock.

The lysate was diluted in TGP broth and used in standard overlay plaque assays with sloppy agar (0.3 % wt/vol agar). Single plaques from these assays were picked using a cut pipette tip to stab into the agar and lift the plaques from the plate. Plaques were crushed and suspended in 1 ml of TGP broth and then used in subsequent rounds of plaque assays. Three rounds of plaque purification were performed, and the purified phages were used in all subsequent experiments.

Mass spectrometry

Samples were precipitated using five volumes of ice-cold acetone and incubated overnight at -20 °C. Precipitates were pelleted by centrifugation at 12 000 × *g* for 10 min. Supernatants were carefully removed, and pellets were air-dried prior to dissolution in 100 mM triethylammonium bicarbonate (TEAB) and determination of protein concentrations (A_{280nm}). Aliquots of 100 µg of solubilized proteins were reduced with 5 mM tris-carboxyethyl phosphine (TCEP; Fluka) for 30 minutes at room temperature. Cysteine residues were methylated by treatment with 10 mM methane methylthiosulfonate (MMTS; Sigma) for 15 minutes at room temperature. After methylation, samples were diluted to 95 µL with 50 mM TEAB before the addition of 5 µL of trypsin (Promega) at 1 mg/mL. Samples were incubated at 37 °C overnight, dried, and resuspended in 30 µL of 2 % acetonitrile:water/0.05 % TFA.

Residual digest reagents were removed using an in-house-manufactured C18 stage tip. The samples were loaded onto the stage tip after activating the C18 membrane with 30 µL of methanol (Sigma) and equilibration with 30 µL of 2% acetonitrile:water/0.05 % TFA. The bound sample was washed with 30 µL of 2 % acetonitrile:water/0.05 % TFA before elution with 30 µL 50 % acetonitrile:water/0.05 % TFA. The eluate was evaporated to dryness. The dried peptides were dissolved in 2 % acetonitrile:water and 0.1 % TFA for LC-MS analysis. Liquid chromatography was performed on a Thermo Scientific Ultimate 3000 RSLC equipped with a 2 cm × 100 µm C18 trap column and a 25 cm × 75 µm Pepmap C18 analytical column. The solvent system employed was as follows: loading, 2 % acetonitrile:water/0.1 TFA; solvent A, 2 % acetonitrile:water/0.1 % TFA; solvent B, 80 % acetonitrile:water. The samples were loaded onto the trap column using loading solvent at a flow rate of 5 µL/min from a temperature-controlled autosampler set at 7 °C. Loading was performed for 10 min before the sample was eluted onto the analytical column. The gradient was generated at 300 nL/min as follows: 0-4 min 2 % A, 4-6 min 6 % A, 6-95 min 6-35 % A (Chromeleon non-linear gradient 6); 95-100 min 35-50 % A. Chromatography was performed at 50°C, and the outflow was delivered to the mass spectrometer through a stainless steel nano-bore emitter. Mass spectrometry was performed on a Thermo Scientific Fusion mass spectrometer. Data were acquired in positive mode using a Nanospray Flex nano-ESI source (Thermo Scientific) with the spray voltage set to 1.7 kV and the ion transfer tube temperature set to 300 °C. MS1 scans were recorded in the Orbitrap mass analyser set to 12 000 resolution over the scan range m/z = 350-1650 with a fill time of 50 ms or until the adaptive gain control (AGC) target of 4e5 was reached. Ion filter criteria were set to mono-isotopic precursors only with charge state 2-6 and dynamic exclusion of 1 over 40 s with mass tolerance of 10 ppm. Precursor selection was performed in top-speed data-dependent mode with the most intense precursor selected first with a cutoff intensity higher than 50,000. Precursor selection was performed using the quadrupole mass analyser with an isolation window of m/z = 1.5 prior to HCD fragmentation. HCD collision energy was set to 35 %. Detection was performed in the ion trap mass analyser with ion injection time of 40 ms or until an AGC target of 1e4 was reached. The raw files generated by the mass spectrometer were imported into Proteome Discoverer v1.4 (Thermo Scientific) and processed using Sequest HT. Database interrogation was performed against GVE3-predicted ORF sequences with trypsin cleavage, allowing for two missed cleavages. Precursor mass tolerance was set to 10 ppm, and fragment mass tolerance set to 0.8 Da. Deamidation (NQ) and oxidation (M) were allowed as dynamic modifications, and thiomethylation of C as a static modification.

Electron microscopy

Phage suspensions were prepared as described previously [2]. Three microliters of each sample was pipetted onto carbon-coated 200-mesh copper grids and stained with 2 % aqueous uranyl acetate. The samples were viewed using a LEO 912 Omega TEM at 120 kV (Zeiss, Oberkochen, Germany) housed at the University of Cape Town Physics Department. Images were collected using a ProScan CCD camera.

Results and discussion

Isolation, morphology and host range testing

The phage was a donation from TMO Renewables. Transmission electron microscopy indicated that *G. thermoglucosidasius* phage GVE3 had morphological characteristics of the B1 morphotype group of the family *Siphoviridae* [1] with a non-contractile tail (± 210 nm long) and isometric head (90 nm–100 nm in diameter) (Fig. 1B). Despite attempts to image phage with nucleic acid in the head, no clear micrographs could be obtained. Fig. 1A, however, shows some phage particles that may have nucleic acid in the head attached to cell debris. GVE3 was tested for its ability to infect a range of *Geobacillus* species (Table 1), but was only capable of infected *G. thermoglucosidasius*.

The GVE3 genome

The GVE3 genome sequence was determined to be 141,298 bp in length and showed a much lower G+C content (29.6 %) than its *G. thermoglucosidasius* host (44 %), as is typical for most phage host pairs [64]. It has been shown that higher AT content results in lower relative entropy (D_{KL}) of a DNA molecule, which could be associated with structural changes in the molecule [12]. Perhaps the lower than average AT content of GVE3 plays a role in its adaptation to thermophily, or alternatively is a reflection of the energy cost of producing nucleotides for phage genome synthesis [64]. This genome size makes it the largest known *Geobacillus*-infecting phage. Overall, the GVE3 genome shares little nucleotide-level identity with any bacteriophage genome currently in the NCBI database (as of 03-03-2015). However, small sections of the genome share significant nucleotide sequence identity with other phage genomes (vB_BanS_Tsamsa, Spβc2, c-st) and *Geobacillus*, *Bacillus* and *Clostridium* genome sequences (Table S3).

A total of 202 putative open reading frames were identified, 62 of which could be assigned a function based on BLAST similarity to genes of known function. The GVE3 genome displays the classical modular arrangement seen in many other members of the family *Siphoviridae* (Fig. 2). GC skew analysis indicated that a replication terminus could be located between the putative holin/endolysin (ORF53) genes and recombinase (ORF54) [65; c-st], while the origin of replication was predicted to lie at ±3700 bp (Fig. 3). Repeat regions, often < 10 bp, are associated with regions where DNA replication is initiated, correspond to sites of gene regulation or transcription termination [10, 56, 61]. Depending on the search criteria, hundreds of inverted and direct repeats of < 10 bp could be identified in the GVE3 genome, although their functional importance, if any, remains to be determined. A search for direct and inverted repeats of > 7 bp and no more than 30 bp apart with 100 % nucleotide sequence identity gave a total of 582 repeats. Two of these invert repeats (TATTTTTT/TAATTAT) are located immediately downstream of ORF3 and in the region predicted to be the origin of replication and may play a role in the initiation of replication.

Although GVE3 does not appear to encode any tRNAs, it does encode a putative ADP-ribose-1-monophosphatase (ORF184; Appr-1-p), an enzyme typically involved in tRNA splicing and encoded in a wide variety of phage genomes, including vB_BanS-Tsamsa [25]. The exact role of this phage element is not clearly established [66], although the link with tRNA synthesis suggests that it could function to remove a rate-limiting step in tRNA

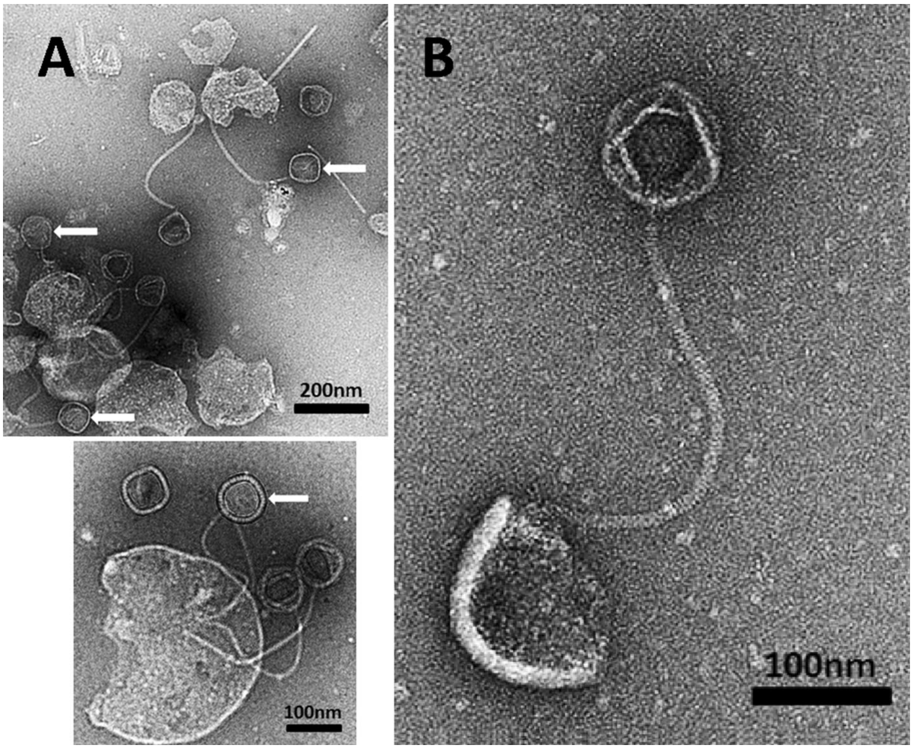

Fig. 1 Bright field TEM of phage GVE3. A) Lower- (top micrograph) and higher-magnification (bottom micrograph) images of several phage attached to cell debris, including some that may still contain nucleic acid in the head (white arrows). B) High magnification image of a single phage particle

Table 1 GVE3 host range

Bacterium	Strain	BGSC no.	Sensitivity to GVE3
Geobacillus stearothermophilus	ATCC 12980^{T}	9A20^{T}	-
Geobacillus thermoleovorans	DSM 5366T	96A1^{T}	-
Geobacillus thermoleovorans	DSM 7263	90A1	-
Geobacillus subterraneus	DSM 13552^{T}	91A1^{T}	-
Geobacillus subterraneus	SAM	91A2	-
Geobacillus thermodenitificans	DSM 465^{T}	94A1^{T}	-
Geobacillus thermoglucosidans	DSM 2542^{T}	95A1^{T}	+
Geobacillus toebii	DSM 14590^{T}	99A1^{T}	-
Geobacillus kaue	HU	105A1	-

processing in the host or to aid in recycling of nucleotides [37].

The closest relatives to GVE3, based on subsystems analysis using RAST, appear to be uncharacterized prophages from *Clostridium thermocellum* and *Bacillus* species. The phages predicted to be the most closely related to GVE3, using PHAST (Table S4; http://tinyurl.com/mtg3fbs), are those from *Bacillus* (Spβc2; vB_BanS_Tsamsa) and *Clostridium* (c-st) rather than the known *Geobacillus* phages, an observation that is consistent with an analysis of the terminase large subunit (Fig. 4). GVE3 thus appears to be most closely related to the recently described *B. anthracis*-infecting vB_BanS_Tsamsa [25].

Tetranucleotide usage deviation (TUD) analysis gave a Pearson's correlation coefficient of 0.665 when comparing GVE3 to the genome of *G. thermoglucosidasius*. Interestingly, when comparing the GVE3 sequence to those of *Bacillus anthracis* and *Bacillus cereus,* significantly higher correlation coefficients were obtained (0.796 and 0.797, respectively). TUD analysis using all available *Geobacillus* species genome sequences (*G. kaustophilus, G. toebii, G. thermodinitrificans, G. themoleovorans, G. thermoglucosidasius, G. thermoglucosidans, G. stearothermophilus, G. subterraneus and G. caldoxylosilyticus*) demonstrated that the TUD of GVE3 was most closely matched to that of *G. toebii* (0.705).

Assuming that TUD analysis provides a measure of the adaptation of phage genomes to that of their hosts over time [60], the GVE3 TUD value suggests that *G. thermoglucosidasius* may not be the prevalent host in nature.

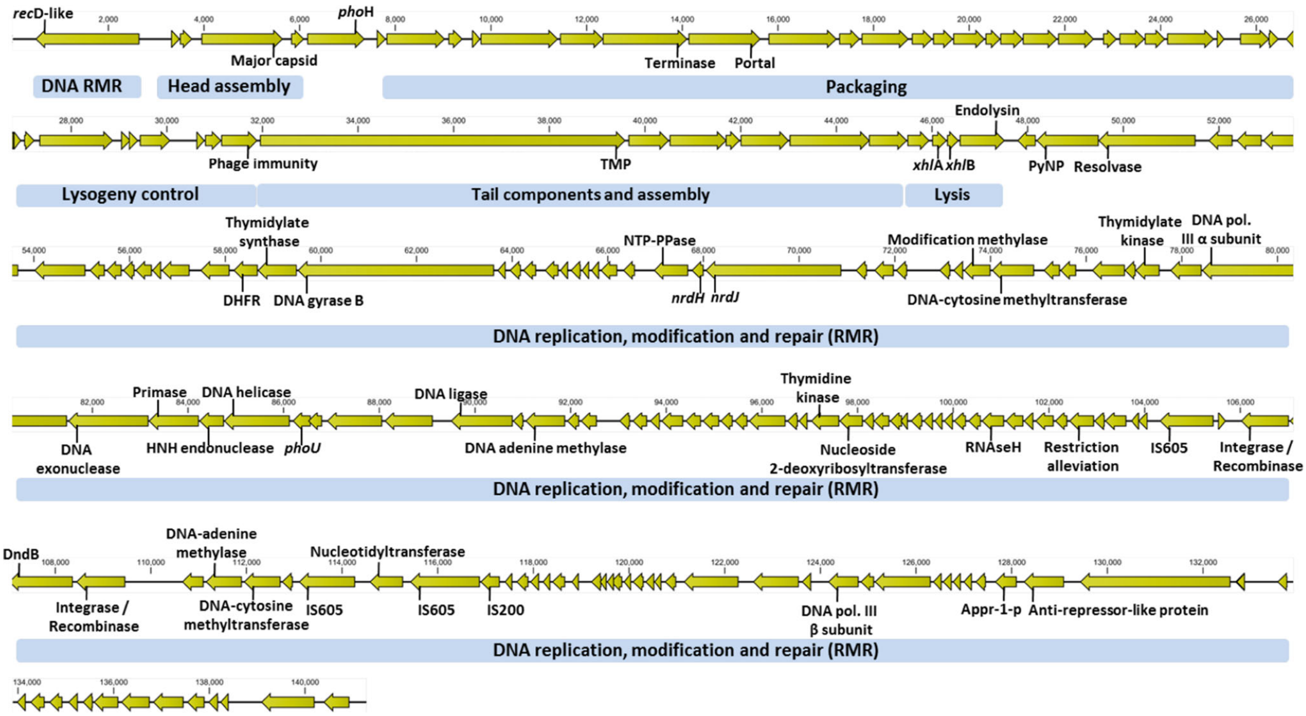

Fig. 2 GVE3 genomic arrangement. Blue boxes indicate modular areas

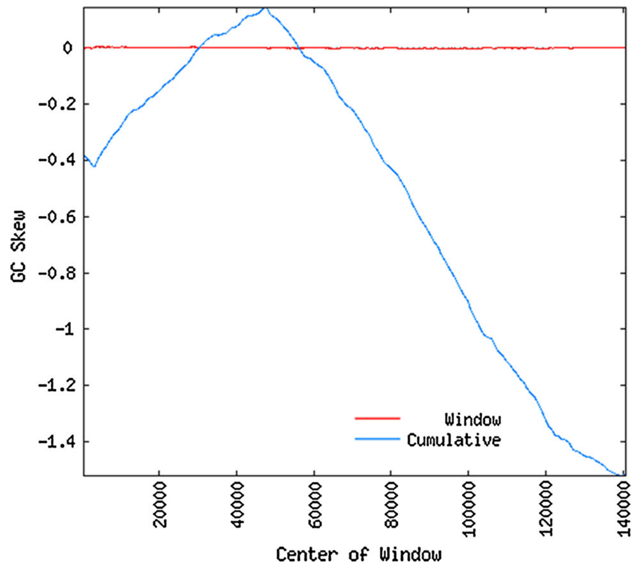

Fig. 3 GC skew analysis of the GVE3 genome showing putative replication origin (*ori*) and termination sites (*ter*) calculated using a window size of 1000 bp and a step size of 100 bp

The higher correlation coefficients of the GVE3 TUD when compared to *B. anthracis* and *B. cereus* (cf. *G. thermoglucosidasius*) suggest that there may be an as yet unidentified *Geobacillus* species with TUD patterns more similar to these two *Bacillus* species that could be the "natural" hosts for GVE3. Alternatively, these results could suggest that GVE3 has "recently" evolved from a mesophilic counterpart and that the high TUD correlation to mesophilic *Bacillus* species is a genuine indication of its evolutionary heritage. A similar relationship has been observed for GBK2, a *G. kaustophilus*-infecting phage that is most closely related to the *Bacillus subtilis* phage SPP1 [50].

Evolution from mesophily to thermophily should involve the adaptation (in both thermophily and thermostability) of phage proteins, and it is therefore unlikely that the thermophilic GVE3 phage would be capable of replicating effectively in a mesophilic host.

DNA metabolism and replication

GVE3 encodes several proteins associated with nucleotide metabolism, including pyrimidine nucleoside phosphorylase (ORF55; PyNP), thymidylate synthase (ORF69; TS), thymidine kinase (ORF123; TK), ribonucleotide reductase (ORF82/83; RNR), nucleoside triphosphate pyrophosphohydrolase (ORF81; NTP-PPase) and nucleoside-deoxyribosyltransferase (ORF124; ND). Although the ORF encoding the putative RNR is most closely related to class II RNRs, there is a small ORF directly downstream of this gene that shows high homology to a *nrdH*-like gene. Ribonucleotide reductases can be divided into several classes (Ia, Ib, Ic; II and III), of which class II RNRs are usually encoded by a single ORF (*nrdJ*), are oxygen

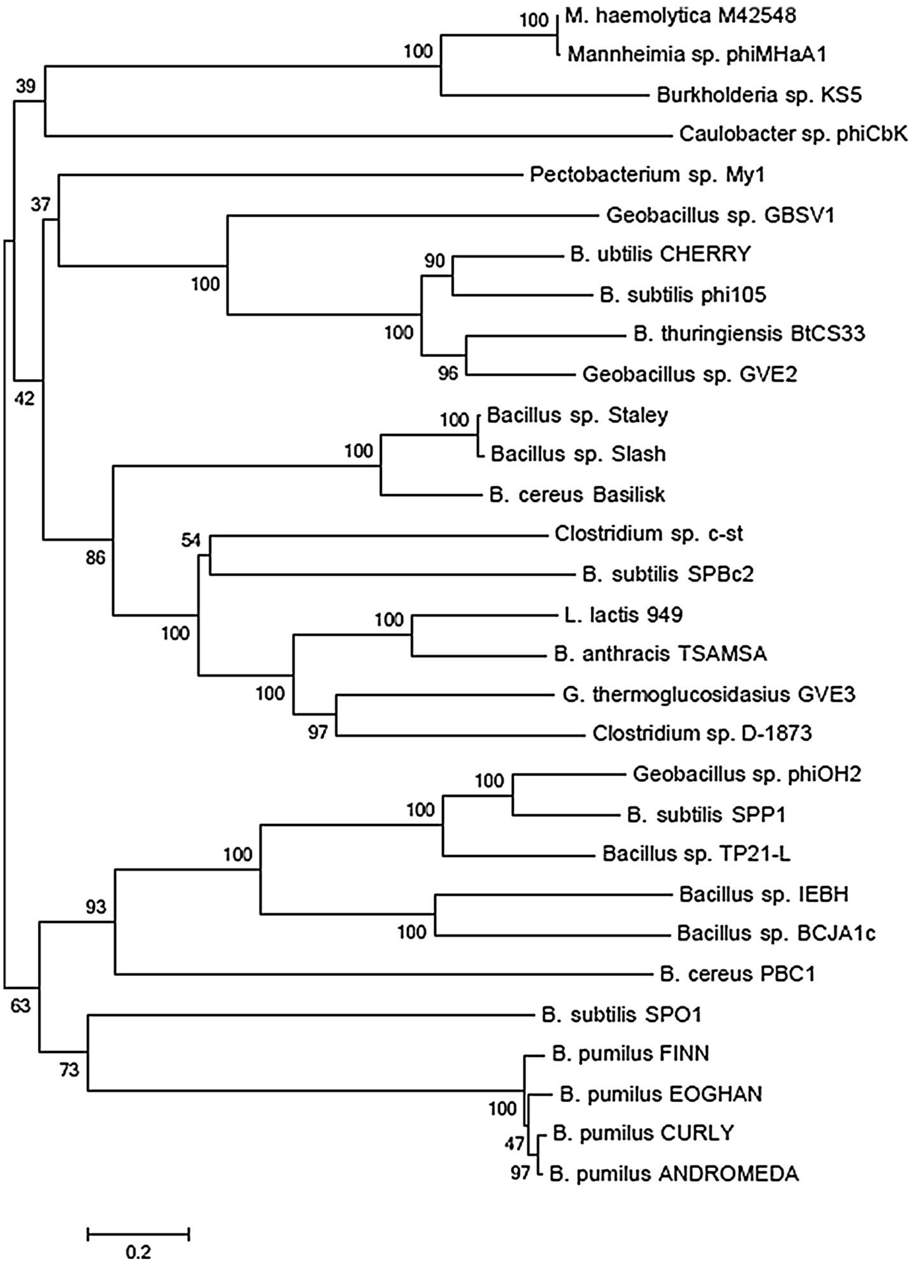

Fig. 4 Neighbor-joining tree comparing full length amino acid sequences of GVE3 terminase large subunit with related proteins. The optimal tree with the sum of branch length = 15.69592415 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in units of the number of amino acid substitutions per site (scale bar). The analysis involved 28 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 821 positions in the final dataset. GVE3, *G. thermoglucosidasius* (KP144388); IEBH, *Bacillus sp.* (NC_011167); BCJA1c, *Bacillus sp.* (NC_006557); TP21-L, *Bacillus sp.* (NC_011645); SPP1, *B. subtilis* (NC_004166); PBC1, *B. cereus* (NC_017976); phiOH2, *Geobacillus sp.* (NC_021784); D-1873, *Clostridium sp.* (ACSJ01000014); vB_BanS-Tsamsa, *B. anthracis* (NC_023007); 949, *L. lactis* (NC_015263); SPBc2, *B. subtilis* (AF020713); c-st. *Clostridium sp.* (D90210); Basilisk, *B. cereus* (KC595511); SPO1, *B. subtilis* (NC_011421); Slash, *Bacillus sp.* (NC_022774); Staley, *Bacillus sp.* (NC_022767); FINN, *B. pumilus* (NC_020480); EOGHAN, *B. pumilus* (NC_020477); ANDROMEDA, *B. pumilus* (NC_020478); CURLY, *B. pumilus* (NC_020479); BtCS33, *B. thuringiensis* (NC_018085); phi105, *B. subtilis* (NC_004167); CHERRY, *B. anthracis* (NC_007457); GBSV1, *Geobacillus sp.* (NC_008376); My1, *Pectobacterium sp.* (NC_018837); phiCbK, *Caulobacter sp.* (NC_019405); KS5, *Burkholderia sp.* (NC_015265); phiMHaA1, *Mannheimia sp.* (NC_008201); M2548, *M. haemolytica* (CP005383)

independent, and usually rely on vitamin B12 for generation of the tyrosyl radical *in vivo* [21]. Class Ib RNRs, encoded by *nrdHIEF,* rely on the glutaredoxin-like protein encoded by *nrdH* to generate the radical needed for catalysis [80]. GVE3 encodes a class II ribonucleotide reductase (*nrdJ*), as well as a component of a class Ib RNR (*nrdH*). The presence and spatial orientation of both *nrdJ*-like and *nrdH*-like ORFs would suggest that they function together. This unusual arrangement has been described for three *Mycobacterium* siphoviruses: Che12, D29 and L5 [21]. It has been argued that the *nrdH* homologue in these genomes was acquired through horizontal gene transfer. Phage genomes are, however, under strong selective pressure to remain within a strict size limit, and all retained genes are expected to confer some metabolic advantage to the host and the phage [23]. In the case of GVE3, the proximity and spatial arrangement of *nrdH* and *nrdJ* as well as the retention of only the *nrdH* homolog (as opposed to any of the *nrdIEF* genes or gene fragments) would argue that these genes confer an advantage to the phage, perhaps *via* interaction with host-encoded components.

The NTP-PPase contains a MazG domain. MazG belongs to the family of α-nucleoside triphosphate pyrophosphohydrolases, which are thought to be responsible for hydrolysis of all non-canonical nucleoside triphosphates produced as a by-product of metabolism and which are toxic to the host, into monophosphate derivatives, thus playing a house-cleaning role [13]. An alternative hypothesis is that, at least in *E. coli*, the NTP-PPase controls the levels of the global regulator ppGpp, redirecting transcription in favour of genes important for starvation survival [47]. Homologues of these proteins have been identified in many phage genomes [28]. In *E. coli, mazG* is co-transcribed with a toxin-antitoxin system (*mazFE*) [27]. It is worth noting that GVE3 encodes several MazF/PemK homologues (ORF38, 40 and 185), although no *mazE* homologues could be identified, and the GVE3 *mazG*-homologue is not co-transcribed with any of these. Whether or not the phage NTP-PPase fulfils multiple roles after host infection, such as regulating the levels of MazF-like toxin produced or eliciting a host survival response to steer its metabolism towards viral production and/or removing toxic nucleoside triphosphates, remains to be determined.

Three DNA-polymerase-like subunits are present on the GVE3 genome. Two of these (ORF97 and 176) are most closely related to the alpha- and beta-clamp subunits of the DNA polymerase III family, similar to those found in *Bacillus* phage vB_BanS_Tsamsa, *Clostridium* phages c-st, D-1873 and *Lactococcus* phage 949. The third subunit, ORF8, shows homology to DNA polymerase A. Other ORFs, the products of which may form part of the DNA Pol III holoenzyme, are a primase (ORF99) and a helicase (ORF101). It has been demonstrated for the *E. coli* DNA polymerase that only the alpha subunit is required for processive replication *in vitro*, although the authors conceded that other subunits, including subunit ε, may be required *in vivo* due to the polymerase encountering obstacles such as proteins bound to the DNA, and DNA lesions not taken into account in their *in vitro* assay system [49]. As not all DNA polymerase III holoenzyme components could be identified on the GVE3 genome, it is possible that the phage recruits host-encoded subunits to complete the polymerase holoenzyme assembly to enable the highly processive DNA replication required for fast and accurate replication of the phage genome [14, 69].

Structural proteins

A putative tail tape measure protein (TMP; ORF42/43) appears to be interrupted by a 310-bp insertion (bp 33,537-33,847), most likely a group I self-splicing intron, as predicted by RNAweasel. As for phage JCL1032 from *Lactobacillus delbruckeii* [63], the 3’ end of the ORF encoding the N-terminal protein sequence (bp 31,948-33,536) ends with a TAG stop codon followed by the intron. Over the length of the putative TMP, seven large imperfect amino acid repeats could be identified (≤ 102 aa). The presence of repeat regions in these proteins has been reported previously and is thought to be of structural significance in determining tail length [8].

Mobile elements

Four putative integrase/recombinase genes were identified (ORF28, 54, 147 and 149), none of which share significant amino acid similarity with each other, a feature noted with phage vB_BanS_Tsamsa [25]. The GVE3 phage genome carries three IS605 family OrfB genes (ORF145, 154 and 156). Insertion sequences of this family sometimes comprise two genes encoding an OrfA (IS200 family) and OrfB, together serving as the functional transposon [30]. One OrfB homologue in GVE3 (ORF156) does have an IS200 family gene directly upstream (ORF157), suggesting that they act co-ordinately. The arrangement of the genes is unusual in that they are transcribed in the same direction while most IS200 family transposons, when associated with an OrfB IS605 element, are divergently transcribed. Parts of GVE3 genome have been incorporated into *Geobacillus toebii* WCH70 CRISPR regions (Table S2). One of these spacers (36 bp) is located in the sequence directly downstream of ORF143 on GVE3. Currently, the incorporation of sequences into CRISPR spacer regions is thought to occur through the identification of bi- or trinucleotide sequences found adjacent to the protospacers, which are eventually incorporated in the CRISPR array, and it is now thought that all type I CRISPR systems target invading DNA for degradation [86]. Interestingly, an IS605/IS200 element (GWCH70_2010 and 2011) is situated directly upstream of the Cas6 (2068682bp-2069410bp) gene in WCH70, probably inactivating this CRISPR array. This CRISPR array also carries the 36-bp spacer, and it is tempting to speculate that a connection exists between these elements. The 36-bp sequence may be important in the ability of the ORF143 transposon to jump, and incorporation of this spacer into a CRISPR cassette may inactivate the transposon, preventing it from inactivating host defence systems.

Nucleotide modifications

Digestion with several restriction endonucleases, including the four-base cutter *RsaI*, for which there are 228 sites on the GVE3 genome, was not successful, whereas treatment with *Alu*I (335 sites) resulted in digestion of the DNA (Table S5). Examples where *Alu*I but not *Rsa*I would digest DNA have been reported and are thought to be due to substitution of thymine with deoxyuridine or substitution of guanine with deoxyinosine [11]. It has also been established that *Alu*I cannot digest the following modified sites: ^{m6}AGCT, AG^{m4}CT, AG^{m5}CT, AG^{hm5}CT [51], and these can probably be excluded as the modifications present in GVE3 DNA. The presence of putative methylases potentially targeting adenine and cytosine residues (ORFs 108, 151 and 152) as well as a DndB domain (ORF146) suggests that the phage DNA is modified to avoid digestion by host-encoded enzymes. For example, *E. coli* T-even phages contain hydroxymethylcytosine (HMC), and *B. subtilis* phage PBS1 contains uracil in place of thymine. The pyrimidine 5-hydroxymethyluracil (HMU) replaces thymine in *B. subtilis* phages SP8, SP5C, SPOl, SP82 and 4e [55]. GVE3 also encodes a putative restriction alleviation protein (bp 102,951-103,163), possibly part of a strategy to avoid host defences.

The presence of restriction endonucleases that inhibit genetic transformation of *Geobacillus* species, and in particular *Hae*III in *G. thermoglucosidasius*, has been reported (WO2006117536A1; [77]). Interestingly, all but one of the *Hae*III sites on GVE3, of which there are only 10, are located within the 3′-terminal 946 bp of the phage genome. They are irregularly spaced and do not appear to form part of conserved repeats. Digestion of phage DNA with *Hae*III could not be detected. The limited number of *Hae*III sites and their location may indicate that the phage genome is under selective pressure to remove such sites. We speculate that, as for phage P1, the 946-bp region containing *Hae*III sites constitutes a *pac* site and that *pac* site cleavage is controlled by the methylation state surrounding the cleavage site [75].

GVE3 proteome

To confirm the expression of predicted ORFs, the complete proteome of GVE3 was determined. The protein products of all predicted ORFs listed in Table S1, except ORF5, 60 and 169, could be identified by at least three unique peptides. The three segments of ORF60, which contains two frameshift mutations, are clearly similar to a hypothetical protein identified in a *Bacillus* species. (WP_028394443.1). However, no peptides similar to any of the three segments of the ORF could be identified, and we conclude that ORF60 is an un-translated region. A peptide corresponding to the PyNP protein was identified, suggesting that this enzyme may play a role in postinfection nucleotide metabolism (see below). No peptide sequences could be identified for the 310-bp region predicted to be a group I self-splicing intron (bp 33,537-33,847) indicating that this is likely to be an untranslated region. If the intron self-excises from this region once inside the host, it is reasonable to expect that a fusion protein, the functional TMP, would be formed by the N- and C-terminal regions of the predicted TMP interrupted by this intron. However, no evidence could be found for the formation of such a fusion between these two terminal regions, and it is likely that each ORF is expressed as a unique protein. The DNA sequence of ORF70 contains a stop codon (TAG) in the reading frame, which translates to VLD*EVK. The identification of a VLDEVK-containing peptide suggests that either readthrough translation or ribosome slippage occurs over this codon. GVE3 structural proteins were also analysed by SDS-PAGE (Fig. 5). Eleven proteins could be identified, of which band 6 corresponds to the size of the predicted major head protein (ORF4), while bands 7 and 8 are likely to correspond to the N-terminal portion of the tape measure protein (ORF42) and the portal protein (ORF14), respectively.

Lysis and lysogeny

There are at least two holin homologues located directly upstream of the endolysin-encoding ORF, the second of these having what appears to be a dual start motif (M-X$_n$-M) with a lysine being one of the two residues separating the methionines. The arrangement of the genes and homology to *xhlA*/*xhlB*/*xlyA* genes from *B. subtilis* phage PBSX suggest that lysis might occur in a manner similar to that system [35]. ORF51 has one predicted transmembrane region (aa 75-97), while ORF52 has two such regions (aa 9-31; aa 41-59).

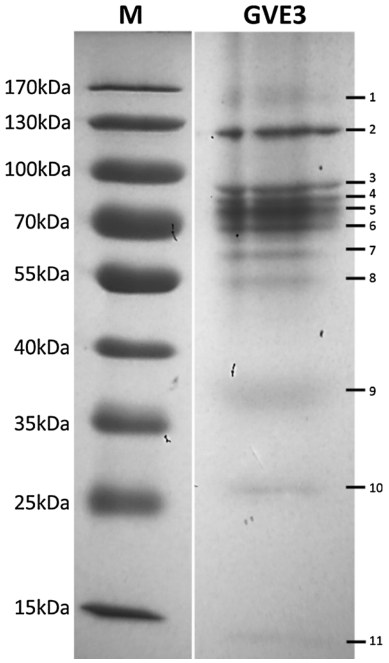

Fig. 5 SDS-PAGE of GVE3 structural proteins. M, Molecular mass marker

Initial plaque assays demonstrated "bulls-eye" plaque morphology, suggestive of host lysogeny [41]. Several bacterial colonies could be observed growing inside plaques. These were isolated and tested for their sensitivity to the phage and were found to be resistant to phage infection. The genome sequence of one of these isolates was determined (Illumina MiSeq; 55-fold coverage) and served as confirmation of the GVE3 genome sequence obtained by Roche 454 sequencing. This showed that the phage genome had inserted into the bacterial host genome and that the *attB* site, with a 23-bp sequence (GGTGGCGTCGGCGATACGACGAC) that was duplicated on insertion (Fig. 6). This sequence only occurs once in the *G. thermoglucosidasius* 11955 genome, located 247 bp from the start of the pyrimidine nucleoside phosphorylase gene (*deoA*), a region known to be interrupted by phage insertion in other genomes [16]. The phage encodes a putative PyNP, downstream of a resolvase, in which the *attP* site is situated. Incorporation of the GVE3 genome sequence in CRISPR spacer regions of the lysogen could not be detected, although two spacers with some nucleotide sequence similarity to regions of the GVE3 genome were identified in the *G. thermoglucosidasius* 11955 genome sequence (Table S2). Integration of the GVE3 genome sequence into that of its host is likely to inactivate the host-encoded PyNP. The presence of a phage-encoded PyNP could suggest an obligate requirement to retain this activity, and that once integrated, the host relies on the phage PyNP, making use of a promoter located in the C-terminal region of the integrase (ORF55) or of readthrough transcription from the promoter located upstream of the host-encoded PyNP (Fig. 6). The PyNP on GVE3 does not show 100 % amino acid sequence identity to the gene from *G. thermoglucosidasius,* or any genes in the NCBInr database. If not essential for either the phage or the host (mutation in PyNP is non-lethal), it may suggest that GVE3 is a specialized transducing phage. No *G. thermoglucosidasius* genomic sequences were observed in the GVE3 genome or in the NGS data, suggesting that GVE3 is unlikely to be a generalized transducing phage. A putative anti-repressor protein (ORF183), which contains an ORF6N domain and has amino acid similarity (50 % over 112 aa) to a truncated annotated anti-repressor protein in *Peptoclostridium difficile*, was identified (Table S1; [31]). In phage lambda, this serves as part of the regulatory mechanism to switch between lysis and lysogeny. Early evidence, based on its overexpression in the host prior to infection, suggests that it plays the same role as in phage lambda (van Zyl et al., unpublished data).

Auxiliary metabolic genes

The GVE3 gene carries the auxiliary metabolic gene *phoH*, and a putative regulator of *phoH* expression, *phoU*, is located upstream of the genes for DNA replication and distant (±79 kb apart) from the *phoH* homologue. The phage also encodes a putative ADP-ribose-1-monophosphatase, which catalyses the conversion of ADP-ribose-1-monophosphate to ADP-ribose as part of the tRNA splicing pathway [37]. The role of *phoH* has not been clearly defined, with some studies demonstrating upregulation under phosphate stress or phage infection [26] while others show downregulation or no change. Should the GVE3 *phoH* gene expression be upregulated, this might suggest that, as with other phages, DNA (and RNA) synthesis becomes rate limiting in the host once replication and transcription of the phage genome is initiated.

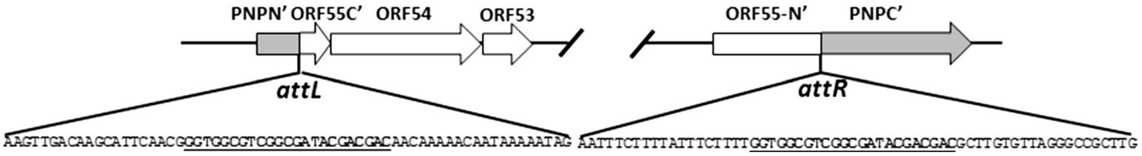
PNPN'
ORF55C'
ORF54
ORF53
ORF55-N'
PNPC'
attL
attR
AAGTTGACAAGCATTCAACGGGTGGCGTCGGCGATACGACGACAACAAAAACAATAAAAATAG
AATTTCTTTTATTTCTTTTGGTGGCGTCGGCGATACGACGACGCTTGTGTTAGGGCCGCTTG

Fig. 6 Layout of the integrated phage. The space between the diagonal lines denotes the rest of the phage genome. The grey box and arrow represent the N- and C-termini, respectively, of the *G. thermoglucosidasius* pyrimidine nucleoside phosphorylase

GVE3 signatures in *Geobacillus* genomes

Two regions of 100 % nucleotide sequence identity to CRISPR spacer regions were found in *G. toebii* WCH70 (Table S2). The presence of these nucleotide sequences suggests that GVE3 or a highly similar phage infected this strain in the past. PCR analysis using four primer pairs targeted to various areas of the GVE3 genome could not detect GVE3 in the chromosome of the *G. toebii* DSM 14590^{T} strain (Table 1), suggesting that superinfection immunity is unlikely to be the cause of failure to infect this strain. Several other putative GVE3-related sequences were identified in CRISPR repeats in a range of *G. thermoglucosidasius* genome sequences (Table S2).

Of the two spacers identified in the *G. toebii* WCH70 genome, one is located at the trailer end of the repeat region, and the other, located in a second CRISPR array, at the leader end in that array, suggesting that this strain has been repeatedly challenged with the same phage [29, 85]. The absence of evidence of lysogenic integration of the GVE3 genome in the WCH70 genome could be due to CRISPR-mediated killing of hosts containing an integrated phage or those that have been infected in the past [22, 48]. Imperfectly matched spacers similar to GVE3 in CRISPR arrays in the 11955 genome could suggest infection by a closely related phage, as seen in polyclonal phage populations during phage blooms or adaptation by the phage to circumvent CRISPR resistance [48, 85]. We suggest that GVE3 represents the latest iteration of a much older version of the phage not currently targeted by the CRISPR system in *G. thermoglucosidacius* 11955. Insertion of spacer sequences based on those identified in WCH70 could be used to engineer resistance by incorporating these into one of the 11955 CRISPR arrays [52].

Conclusion

GVE3, although a member of the well-known family *Siphoviridae* and unremarkable with respect to the overall layout of genes and the genes encoded, appears to have a unique genome sequence, with no close relatives in the current databases. Although there are indications that it may have had the capacity to infect other *Geobacillus* species in the past, the current specificity appears to be restricted to *G. thermoglucosidasius*. The relationships between the GVE3 genome and those of mesophilic phages and bacteria may be a consequence of the small number of thermophilic phage genome sequences in the databases but may reflect the evolutionary history of a phage in transition from mesophily to thermophily. GVE3 encodes a number of enzymes, including ATP-dependent DNA ligase, DNA polymerase III, RNaseH, PyNP, holin and endolysin [78], all of which should be thermostable. These could be of commercial value or employed as research tools, such as in the use of endolysin in the treatment of milk to kill *Geobacillus* species spoilage organisms [15, 81]. *G. thermoglucosidasius* has been engineered as a platform organism for ethanol production and other industrial products, but to date, there is no mechanism for the introduction of large DNA fragments (>12 kb), and GVE3 could potentially be developed as a system for introduction of novel or engineered metabolic and biosynthetic pathways.

Acknowledgments The authors wish to thank TMO Renewables for the gift of the GVE3 phage. This work was funded by the National Research Foundation (NRF) of South Africa. The authors declare no conflict of interest.

References

1. Ackermann HW (2007) 5500 Phages examined in the electron microscope. Arch Virol 152:227–243
2. Ackermann HW, Heldal M (2010) Basic electron microscopy of aquatic viruses. In: Manual of aquatic viral ecology, Chapter 18. American Society of Limnology and Oceanography, Inc., p 182–192
3. Adriaenssens EM, van Zyl LJ, de Maayer P, Rubagotti E, Rybicki E, Tuffin M, Cowan DA (2014) Metagenomic analysis of the viral community in Namib Desert hypoliths. Environ Microbiol. doi:10.1111/1462-2920.12528
4. Ahn D-G, Kim S-I, Rhee J-K, Kim KP, Pan J-G, Oh J-W (2006) TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote *Thermoproteus tenax*. Virol 351:280–290
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
6. Arnold HP, Zillig W, Ziese U, Holz I, Crosby M, Utterback T, Weidmann JF, Kristjanson JK, Klenk HP, Nelson KE, Fraser CM (2000) A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. Virol 267:252–266
7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST server: Rapid annotations using subsystems technology. BMC Genomics 9:75
8. Belcaid M, Bergeron A, Poisson G (2011) The evolution of the tape measure protein: units, duplications and losses. BMC Bioinform 12:S10
9. Betley JN, Frith MC, Graber JH, Choo S, Deshler JO (2002) A ubiquitous and conserved signal for RNA localization in chordates. Curr Biol 12:1756–1761
10. Blatny JM, Godager L, Lunde M, Nes IF (2004) Complete genome sequence of the *Lactococcus lactis* temperate phage φLC3: comparative analysis of φLC3 and its relatives in lactococci and streptococci. Virology 318:231–244
11. Bodnarz JW, Zempsky W, Warder D, Bergson C, Ward DC (1983) Effect of nucleotide analogs on the cleavage of DNA by the restriction enzymes *Alu*I, *Dde*I, *Hin*fI, *Rsa*I, and *Taq*I. J Biol Chem 258:15206–15213
12. Bohlin J, van Passel MWJ, Snipen L, Kristoffersen AB, Ussery D, Hardy SP (2012) Relative entropy differences in bacterial

chromosomes, plasmids, phages and genomic islands. BMC Genomics 13:66–78

13. Bryan MJ, Burroughs NJ, Spence EM, Clokie MRJ, Mann NH, Bryan SJ (2008) Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. PLOS One. doi:10.1371/journal.pone.0002048
14. Bullard JM, Williams JC, Acker WK, Jacobi C, Janjic N, McHenry CS (2002) DNA polymerase III holoenzyme from *Thermus thermophiles* identification, expression, purification of components, and use to reconstitute a processive replicase. J Biol Chem 277:13401–13408
15. Burgess SA, Lindsay D, Flint SH (2010) Thermophilic bacilli and their importance in dairy processing. Int J Food Microbiol 144:215–225
16. Buxton RS, Hammer-Jespersen K, Hansen TD (1978) Insertion of bacteriophage lambda into the *deo* operon of *Escherichia coli* K-12 and isolation of plaque-forming λdeo^+ transducing bacteriophages. J Bacteriol 136:668–681
17. Chen Y, Wei D, Wang Y, Zhang X (2013) The role of interactions between bacterial chaperone, aspartate aminotransferase, and viral protein during virus infection in high temperature environment: the interactions between bacterium and virus proteins. BMC Microbiol 13:48
18. Clokie MRJ, Millard AD, Letarov AV, Heaphy S (2011) Phages in nature. Bacteriophage 1:31–45
19. Cripps RE, Eley K, Leak DJ, Rudd B, Taylor M, Todd M, Boakes S, Martin S, Atkinson T (2009) Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. Metab Eng 11:398–408
20. Doi K, Mori K, Martono H, Nagayoshi Y, Fujino Y, Tashiro K, Kuhara S, Ohshima T (2013) Draft Genome Sequence of Geobacillus kaustophilus GBlys, a Lysogenic Strain with Bacteriophage OH2. Genome Announc 1:e00634–e00713
21. Dwivedi B, Xue B, Lundin D, Edwards RA, Breitbart M (2013) A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. BMC Evol Biol 13:33
22. Edgar R, Qimron U (2010) The *Escherichia coli* CRISPR system protects from lysogenization, lysogens, and prophage induction. J Bacteriol 192:6291–6294
23. Feiss M, Siegele DA (1979) Packaging of the bacteriophage lambda chromosome: dependence of *cos* cleavage on chromosome length. Virology 92:190–200
24. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
25. Ganz HH, Law C, Schmuki M, Eichenseher F, Calendar R, Loessner MJ, Getz WM, Korlach J, Beyer W, Klumpp J (2014) Novel giant Siphovirus from *Bacillus anthracis* features unusual genome characteristics. PLoS One 9:e85972
26. Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle CA, Weinbauer MG, Sandaa RA, Breitbart M (2011) Development of phoH as a novel signature gene for assessing marine phage diversity. Appl Environ Microbiol 77:7730–7739
27. Gross M, Marianovsky I, Glaser G (2006) MazG—a regulator of programmed cell death in *Escherichia coli*. Mol Microbiol 59:590–601
28. Hargreaves KR, Kropinski AM, Clokie MRJ (2014) Bacteriophage behavioral ecology: how phages alter their bacterial host's habits. Bacteriophage. doi:10.4161/bact.29866
29. Heler R, Marraffini LA, Bikard D (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol Microbiol. doi:10.1111/mmi.12640
30. Höök-Nikanne J, Berg DE, Peek RM Jr, Kersulyte D, Tummuru MKR, Blaser MJ (1999) DNA sequence conservation and diversity in transposable element IS605 of *Helicobacter pylori*. Helicobacter 3:79–85
31. Iyer LM, Koonin EV, Aravind L (2002) Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. Gen. Biol 3:research0012.1–research0012.11
32. Jin M, Ye T, Zhang X (2013) Roles of bacteriophage GVE2 endolysin in host lysis at high temperatures. Microbiology 159:1597–1605
33. Jin M, Chen Y, Xu C, Zhang X (2014) The effect of inhibition of host MreB on the infection of thermophilic phage GVE2 in high temperature environment. Sci Rep 4:4823
34. Kotze AA, Tuffin IM, Deane SM, Rawlings DE (2006) Cloning and characterization of the chromosomal arsenic resistance genes from *Acidithiobacillus caldus* and enhanced arsenic resistance on conjugal transfer of *ars* genes located on transposon TnAtcArs. Microbiology 152:3551–3560
35. Krogh S, Jørgensen ST, Devine KM (1998) Lysis genes of the *Bacillus subtilis* defective prophage PBSX. J Bacteriol 180:2110–2117
36. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580
37. Kumaran D, Eswaramoorthy S, Studier FW, Swaminathan S (2005) Structure and mechanism of ADP-ribose-1-monophosphatase (Appr-1-pase), a ubiquitous cellular processing enzyme. Prot Sci 14:719–726
38. Lang BF, Laforest MJ, Burger G (2007) Mitochondrial introns: a critical view. Trends Genet 23:119–125
39. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucl Acids Res 32:11–16
40. Le Romancer M, Gaillard M, Geslin C, Prieur D (2007) Viruses in extreme environments. In: Amils Ricardo, Ellis-Evans Cynan, Hinghofer-Szalkay Helmut (eds) Life in extreme environments. Springer, Netherlands, pp 99–113
41. Levine M, Truesdall S, Ramakrishan T, Bronson MJ (1975) Dual control of lysogeny by bacteriophage P22: an antirepressor locus and its controlling elements. J Mol Biol 91:421–438
42. Lin PP, Rabe KS, Takasumi JL, Kadisch M, Arnold FH, Liao JC (2014) Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. Metab Eng 24:1–8
43. Liu B, Wu S, Song Q, Zhang X, Xie L (2006) Two novel bacteriophages of thermophilic bacteria isolated from Deep-Sea hydrothermal fields. Curr Microbiol 53:163–166
44. Liu B, Zhang X (2008) Deep-sea thermophilic *Geobacillus* bacteriophage GVE2 transcriptional profile and proteomic characterization of virions. Appl Microbiol Biotechnol 80:697–707
45. Liu B, Zhou F, Wu S, Xu Y, Zhang X (2009) Genomic and proteomic characterization of a thermophilic *Geobacillus* bacteriophage GBSV1. Res Microbiol 160:166–170
46. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 25:955–964
47. Magnusson LU, Farewell A, Nystrom T (2005) ppGpp: a global regulator in *Escherichia coli*. Trends Microbiol 13:236–242
48. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11:181–190
49. Marians KJ, Hiasa H, Kim DR, McHenry C (1998) Role of the core DNA polymerase III subunits at the replication fork: ALPHA is the only subunit required for processive replication. J Biol Chem 273:2452–2457
50. Marks TJ, Hamilton PT (2014) Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*. Arch Virol 159:2771–2775

51. McClelland M, Nelson M, Raschke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucl Acids Res 22:3640–3659
52. Millen AM, Horvath P, Boyaval P, Romero DA (2012) Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. PLoS One 7:e51663
53. Moser MJ, DiFrancesco RA, Gowda K, Klingele AJ, Sugar DR, Stocki S, Mead DA, Schoenfeld TW (2012) Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. PLOS One. doi:10.1371/journal.pone.0038371
54. Nazina TN, Tourova TP, Poltaraus AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, Ivanov MV (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. IJSEM 51:433–446
55. Neubort S, Marmur J (1973) Synthesis of the unusual DNA of *Bacillus subtilis* bacteriophage SP-15. J Virol 12:1078–1084
56. Østergaard S, Brøndsted L, Vogensen FK (2001) Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. Appl Environ Microbiol 67:774–781
57. Payeta JP, Suttle CA (2013) To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status. Limnol Oceanogr 58:465–474
58. Peduzzi P, Gruber M, Gruber M, Schager M (2014) The virus's tooth: cyanophages affect an African flamingo population in a bottom-up cascade. ISME J 8:1346–1351
59. Plotka M, Kaczorowska A-K, Stefanska A, Morzywolek A, Fridjonsson OH, Dunin-Horkawicz S, Kozlowski L, Hreggvidsson GO, Kristjansson JK, Dabrowski S, Bujnicki JM, Kaczorowskia T (2013) Novel highly thermostable endolysin from *Thermus scotoductus* MAT2119 bacteriophage Ph2119 with amino acid sequence similarity to Eukaryotic peptidoglycan recognition proteins. Appl Environ Microbiol 80:886–895
60. Pride DT, Wassenaar TM, Ghose C, Blaser MJ (2006) Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. BMC Genomics. doi:10.1186/1471-2164-7-8
61. Quiles-Puchalt N, Tormo-Más MA, Campoy S, Toledo-Arana A, Monedero V, Lasa I, Novick RP, Christie GE, Penadés JR (2013) A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. Nucl Acids Res 41:7260–7275
62. Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S, McDermott T, Young MJ (2001) Viruses from extreme thermal environments. Proc Nat Acad Sci 98:13341–13345
63. Riipinen KA, Alatossava T (2004) Two self-splicing group I introns interrupt two late transcribed genes of prolate-headed *Lactobacillus delbrueckii* phage JCL1032. Arch Virol 149:2013–2024
64. Rocha EPC, Danchin A (2002) Base composition bias might result from competition for metabolic resources. Trends Genet 18:291–294
65. Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y, Ohnishi M, Ohtsubo E, Hattori M, Oguma K (2005) The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. Proc Nat Acad Sci 102:17472–17477
66. Savalia D, Westblade LF, Goel M, Florens L, Kemp P, Akulenko N, Pavlova O, Padovan JC, Chait BT, Washburn MP, Ackermann HW, Mushegian A, Gabisonia T, Molineux I, Severinov K (2008) Genomic and proteomic analysis of phiEco32, a novel *Escherichia coli* phage. J Mol Biol 377:774–789
67. Schmidt TR, Scott EJ II, Dyer DW (2011) Whole-genome phylogenies of the family Bacillaceae and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. BMC Genomics 12:430
68. Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, Mead D (2008) Assembly of viral metagenomes from yellowstone hot springs. Appl Environ Microbiol 74:4164–4174
69. Seco E, Zinder J, Manhart CM, Piano AL, McHenry C, Ayora S (2013) Bacteriophage SPP1 in vitro DNA replication strategies promote viral and disable host replication. Nucl Acid Res 41:1711–1721
70. Sevostyanova A, Djordjevic M, Kuznedelov K, Naryshkina T, Gelfand MS, Severinov K, Minakhin L (2007) Temporal regulation of viral transcription during development of *Thermus thermophilus* bacteriophage ϕYS40. J Mol Biol 366:420–435
71. Sime-Ngando ST, Lucas S, Robin A, Tucker KP, Colombet J, Bettarel Y, Desmond E, Gribaldo S, Forterre P, Breitbart M, Prangishvili D (2010) Diversity of virus–host systems in hypersaline Lake Retba. Environ Microbiol 8:1956–1972
72. Song Q, Zhang X (2008) Characterization of a novel non-specific nuclease from thermophilic bacteriophage GBSV1. BMC Biotechnol 8:43
73. Song Q, Ye T, Zhang X (2011) Proteins responsible for lysogeny of deep-sea thermophilic bacteriophage GVE2 at high temperature. Gene 479:1–9
74. Sorokin DY, Berben T, Melton ED, Overmars L, Vavourakis CD, Muyzer G (2014) Microbial diversity and biogeochemical cycling in soda lakes. Extremophiles 18:791–809
75. Sternberg N, Coulby J (1990) Cleavage of the bacteriophage P1 packaging site (*pac*) is regulated by adenine methylation. Proc Natl Acad. Sci 87:8070–8074
76. Suttle CA (2005) Viruses in the sea. Nature 437:356–361
77. Suzuki H, Yoshida K (2012) Genetic transformation of *Geobacillus kaustophilus* HTA426 by conjugative transfer of host-mimicking plasmids. J Microbiol Biotechnol 22:1279–1287
78. Szekera K, Zhou X, Schwab T, Casanueva A, Cowan D, Mikhailopulo IA, Neubauer P (2012) Comparative investigations on thermostable pyrimidine nucleoside phosphorylases from *Geobacillus thermoglucosidasius* and *Thermus thermophiles*. J Mol Cat B Enzymatic 84:27–34
79. Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA (2009) Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. Trends Biotechnol 27:398–405
80. Torrents E (2014) Ribonucleotide reductases: essential enzymes for bacterial life. Front Cell Infect Microbiol. doi:10.3389/fcimb.2014.00052
81. Viedma PM, Abriouel H, Omar NB, Lopez RL, Valdivia E, Gálvez A (2009) Inactivation of *Geobacillus stearothermophilus* in canned food and coconut milk samples by addition of enterocin AS-48. Food Microbiol 26:289–293
82. Wang Y, Zhang X (2008) Identification and characterization of a novel thymidylate synthase from deep-sea thermophilic bacteriophage *Geobacillus* virus E2. Virus Genes 37:218–224
83. Wang Y, Zhang X (2010) Genome analysis of deep-sea thermophilic phage D6E. Appl Environ Microbiol 76:7861–7866
84. Wei D, Zhang X (2010) Proteomic analysis of interactions between a deep-sea thermophilic bacteriophage and its host at high temperature. J Virol 84:2365–2373
85. Weinberger AD, Sun CL, Plucinski MM, Denef VJ, Thomas BC, Horvath P, Barrangou R, Gilmore MS, Getz WM, Banfield JF

(2012) Persisting viral sequences shape microbial CRISPR based immunity. PLoS One 8:e1002475

86. Westra ER, Swarts DC, Staals RHJ, Jore MM, Brouns SJJ, van der Oost J (2012) The CRISPRs, they are A-Changin': How prokaryotes generate adaptive immunity. Annual Rev Genet 46:311–339
87. Zhou Y, Liang Y, Lynch K, Dennis JJ, Wishart DS (2011) PHAST: a fast phage search tool. Nucl Acids Res 39:347–352
88. Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by Bryson V and Vogel HJ (eds). Academic Press, New York, pp. 97–166