Author(s): Jalasvuori, Matti; Koskinen, Katariina

Title: Extending the hosts of Tectiviridae into four additional genera of Gram-positive bacteria and more diverse Bacillus species

Year: 2018

Version: Accepted version (Final draft)

Please cite the original version:

Jalasvuori, M., & Koskinen, K. (2018). Extending the hosts of Tectiviridae into four additional genera of Gram-positive bacteria and more diverse Bacillus species. Virology, 518, 136-142. https://doi.org/10.1016/j.virol.2018.02.014

Extending the hosts of *Tectiviridae* into four additional genera of Gram-positive bacteria and more diverse Bacillus species

Matti Jalasvuori[1,2*] and Katariina Koskinen[2]

[1]Department of Genetics, University of Cambridge, Cambridge, United Kingdom

[2]Department of Biological and Environmental Science, Nanoscience Center, University of Jyvaskyla, Finland

Corresponding author:
Matti Jalasvuori
E-mail: matti.jalasvuori@jyu.fi
Phone: +358 50 413 50 92
Postal address: Department of Biological and Environmental Science, POBox35, 40014 University of Jyvaskyla, Finland

*Present address: Department of Genetics, University of Cambridge, Cambridge, United Kindom

ABSTRACT

Tectiviridae are composed of tailless bacteriophages with an icosahedral capsid and an inner membrane enclosing a double-stranded 15 kb linear DNA genome. Five of the seven previously studied Tectivirus isolates infect bacteria from *Bacillus cereus* sensu lato group (Betatectivirus), one distantly related member (PRD1) infect *Enterobactericeae* (Alpatectivirus) and one recently discovered virus infect *Gluconobacter cerinus* (Gammatectivirus). Here we expand the host spectrum of Betatectivirus elements to four additional genera (*Streptococcus*, *Exiguobacterium*, *Clostridium* and *Brevibacillus*) and to more distantly related *Bacillus* species (*B. pumilus* and *B. flexus*) by studying the genomes of fourteen novel tectiviral elements. Overall, the genomes show significant conservation in gene synteny and in modules responsible for genome replication and formation of the virion core (including DNA packaging). Notable variation exists in regions encoding host attachment and lysis along with the surrounding area of a site in which mutations are known to alter phage life cycle.

Keywords: Tectiviridae, Tectivirus, bacteria, bacillus, bacteriophage, prophage, evolution,

INTRODUCTION

The virus family *Tectiviridae* contain a single genus, Tectivirus, that comprises tailless bacteriophages with an icosahedral protein capsid of approximately 70 nm in diameter. The protein coat encloses an inner membrane, which during infection extrudes from the capsid to penetrate the cell wall and membrane(s) of the host. The ~15kb double stranded linear DNA genome harbors around 30 genes and is replicated in protein-primed manner by a phage-encoded polymerase (Caldentey et al., 1992; Berjón-Otero et al., 2016). After genome replication, DNA is packed with an ATPase into the capsid and the host cell is lysed with an endolysin.

The type virus of the *Tectiviridae* is PRD1, Gram-negative bacteria infecting plasmid-dependent phage (Olsen et al., 1974). As PRD1 binds to a plasmid-encoded receptor on the host cell, its host-range aligns with that of the plasmid (Olsen et al., 1974). PRD1 has been studied intensively for several decades, making it one of the best characterized viruses in terms of genome, structure and function (Abrescia et al., 2004; Cockburn et al., 2004). Yet, PRD1 is the sole representative of the family in the Gram-negative clade (excluding a handful of other isolates that are genetically almost 100% identical to PRD1, Saren et al., 2005). The other previously known members of *Tectiviridae* infect Gram-positive bacteria of *Bacillus cereus* sensu lato group (*Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*, Gillis and Mahillion, 2014a). However, very recently, a novel Tectivirus GC1 infecting *Gluconobacter cerinus* was characterized and proposed to form a new genus, "Gammatectivirus" (Philippe et al., 2018). There are clear differences among the clades of the family. PRD1 is a strictly lytic virus whereas the *Bacillus* infecting representatives and GC1 can establish lysogeny (in case of tectiviruses, the genome is replicated within the host cell without integrating into the host genome). Lysogenic cycle of Gram-positive bacteria infecting tectiviruses, however, is prone to switch into strictly lytic cycle when the host-encoded LexA repressor binding site on the virus genome gets disrupted by mutations (Fornelos et al., 2011; Fornelos et al., 2015). Genetically the clades are very different and there is little homology on the level of DNA (Ravantti et al., 2003; Philippe et al., 2018). However, the products of the core genes of these viruses (namely the genes coding for major capsid protein and DNA packaging ATPase) share approximately 20-30% similarity on the level of amino acids and all tectiviruses are structurally very similar.

Currently identified members of *Bacillus cereus* sensu lato infecting *Tectiviridae* share 60% to 100% identity on DNA level. These phages are Bam35 (or almost identical Gil01 that differs by eleven nucleotides, Verheust et al., 2001; Ravantti et al., 2003), Gil16 (Verheust et al., 2005), AP50 (Sozhamannan et al., 2008), Wip1 (Kan et al., 2013) and pBClin15 (Stromsten et al., 2003).

One of early tectiviral isolates infecting *B. acidocaldarius*, phiNS11 (Sakaki et al., 1977), is no longer available in the laboratory and has not been sequenced. A large-scale PCR-based screening for Tectivirus-related elements among 2000 *B. cereus* isolates confirmed their presence in only 2.7% of the strains, indicating that they are relatively rare but also genetically conserved (Gillis and Mahillion, 2014a, Gillis and Mahillion, 2014b). It is also notable that only around half of the annotated open reading frames (ORFs) of these tectiviruses have identifiable homologs in databases (Sozhamannan et al., 2008; Berjón-Otero et al., 2017). Given that the currently identified phages infect relatively closely related hosts (and PRD1 and GC1 infect extremely distant hosts), it has not been possible to study the evolution of (Gram-positive bacteria infecting) tectiviruses beyond *B. cereus* sensu lato group. In this paper, we analyze the genomes of fourteen new Tectivirus-related elements from four new genera (*Streptococcus*, *Exiguobacterium*, *Clostridium* and *Brevibacillus*), from more distantly related *Bacillus* species (*B. pumilus* and *B. flexus*) in addition with elements from new strains of *Bacillus cereus* sensu lato group.

RESULTS AND DISCUSSION

Tectivirus genomes were used as a query to search homologous viral elements in databases. Genomes of twelve novel tectivirus-like elements were discovered from the GenBank. Four of these elements originate from novel genera, namely *Streptococcus*, *Clostridium*, *Brevibacillus* and *Exiguobacterium*, none of which have been previously known to host tectiviruses. Six additional elements were discovered from different *Bacillus* species: *B. cereus*, *B. thuringiensis*, *B. pumilus* and *B. licheniformis*. The latter two are not part of the *Bacillus cereus* sensu lato group, and thus provides examples of these elements in more distant *Bacillus* hosts. Further, there are two very incomplete environmental metagenomics sequences from Red Sea in which clear Tectivirus-like elements exists (GenBank ids KX984138 and KX984131). However, we left them out from

detailed analysis due to the lack of information on hosts and the incompleteness of the sequences. Nevertheless, the sequences studied here double the number of Tectivirus-like elements for which nearly-complete genome sequence is available (listed in Table 1. Note that the two elements in *Streptococcus* are the same and the element in *Bacillus cereus* strain FSL M8-0473 is identical to previously studied element pBClin15 because the same strain was sequenced twice, first in 2003 (Ivanova et al. 2003) and later in 2017 (direct deposit).

All of these elements were manually checked for their (automated) annotations. Unsurprisingly, many of the likely coding domains had not been identified. We curated the genomes by adopting the following criteria to locate putative genes: the open reading frames (ORFs) should be preceded by a potential Shine-Dalgarno site (AGGAGG sequence allowing some permutations) to which ribosomes bind, and the ORFs should fill the apparent "non-coding" regions in the genome (as tectiviruses are known to be tightly packed with coding areas and thus existence of long regions without genes is unlikely). Five to ten new putative genes were identified from each element and most had homologs in other tectiviral elements, suggesting that they are likely to be genes. Further, analysis of the deposited annotations revealed a short-coming of these automated tools as they generally failed to describe the genetic content accurately enough in order to identify them as potential proviruses without either utilizing specific tools (such as the improved Phage Search Tool, PHASTER, Arndt et al., 2016, which indeed successfully recognized the elements studied here as potential tectiviruses) or conducting targeted analysis of individual genes on sequence level. The problem in identification derives from the annotations of the known viruses in databases as most of them describe the coding domains to produce "hypothetical proteins" or "cytoplasmic protein". Yet, in reality, many of these hypothetical proteins have been demonstrated to produce structural components of the virions (see e.g. Strömsten et al., 2003; Sozhamannan et al., 2008) or to have interactions with other proteins in the genome (Berjón-Otero et al., 2017). Accurate identification of the elements appear relevant for deriving useful information from whole genome

sequencing projects given that members of *Tectiviridae* are known to influence growth rate, biofilm formation, swarming motility, sporulation (Gillis and Mahillion, 2014c) and they continuously produce viral particles in liquid cultures (that infect other hosts in their environment, Jalasvuori et al., 2009; Jalasvuori et al., 2013).

The evolutionary distances of the elements were inferred by studying the major capsid protein (MCP, Figure 1) along with the packaging ATPase (Supplementary Figure 1) as they have been repeatedly argued to define the identity of the virus and are considered to be evolutionarily tightly linked (and thus likely to be co-inherited to viral offspring from the same virus, Bamford et al., 2002; Krupovic and Bamford, 2010; Sinclair et al., 2017); the same approach has been utilized to group relatively distant elements in other viruses with icosahedral capsids and inner membranes (Jalasvuori et al., 2010; Pawlowski et al., 2014). Both proteins indeed produce comparable trees (Figure 1 and Supplementary Figure 1), although there are some uncertainties (low bootstrap values) with more distantly related elements. Yet, in the case of Gil16c and the elements in *B. cereus* strains VD166 and VD184, a clear recombination between the genes for ATPase and MCP has occurred. On amino acid level, Gil16 ATPase is 96% identical to that of VD184 but only 79% identical to VD166 while the Gil16c MCP is 98% identical to MCP of VD166 but only 72% identical to VD184. This reveals that even the areas between tightly linked "virus identifying" genes may sometimes serve as sites for recombination among *Tectiviridae*. Overall, the most divergent members of the group are the elements in the hosts that were previously not known to harbor tectiviruses, i.e. *Brevibacillus* sp., *Clostridium* sp., *Exiguobacterium antarcticum*, *Bacillus pumilus* and *Bacillus flexus*, all sharing around 32-40% similarity on DNA level with the others (Table 2).

Overall, the divergence of tectiviruses appears to follow that of their host organisms when the 16sRNA trees of Gram-positive bacteria is used as a reference (Onyenwoke et al., 2004). This is expected, given the host specificity of bacteriophages in general and tectiviruses in particular

(Jalasvuori et al., 2013; Gillis and Mahillion, 2014b; Mattila et al., 2015), making jumps to taxonomically distant hosts unlikely. However, there is single exception. The element in *Streptococcus* is closely related to *Bacillus anthracis* viruses of AP50 and Wip1 despite of the more distant relationship of their hosts. This suggest that a Tectivirus-like element of probably *Bacillus* origin has recently invaded *Streptococcus*.

We compared the genomes of all the elements to each other (Figure 2) and scored the number of (potential) homologs in other members of the group (including PRD1). Hits to identical elements were ignored (pBClin15 and element in *Bacillus cereus* strain FSL M8-0473; and both *Streptococcus pneumoniae* elements). Overall gene synteny among tectiviruses or Tectivirus-like element in Gram-positive bacteria is very conserved and no major shifting of the modules is observed, excluding the previously denoted localization of DNA polymerase in Wip1 in the 3' end of the genome (Kan et al., 2013). The genome replication machinery of Wip1 in general shows notable divergence in comparison to other elements. Otherwise, several genomic regions are conserved, including those for DNA polymerase and the genome terminal protein, both of which are necessary for the protein-primed replication, along with those genes encoding for key structural components of the virion and DNA packing. Most variability, on the other hand, is observed around the previously demonstrated host encoded LexA suppressor binding site, where mutations often result in virus to be unable to suppress its lytic life cycle (Fornelos et al., 2011). Such mutations can be favourable in certain ecological situations, such as when there is large susceptible host population for lytic mutants to proliferate in. Yet, in the absence of new hosts, reversion to temperate cycle can be necessary for the long-term survival of the phage (Jalasvuori and Koonin, 2015). These possible alterations in ecological conditions have been suggested to contribute in the genetic diversity of the region (Jalasvuori et al., 2014). However, while previously most of the genes in this region had homologs neither in databases nor among other tectiviruses, now we were able to identify potential homologs for majority of the genes in other

elements. This suggest that the region, while variable, is not evolving as rapidly as was formerly considered. Also, relatively long putatively non-coding sequences are more common within this region than elsewhere in the genome. Second region where variability abounds is the one in the 3' end of the genome encoding for host recognition features along with endolysins responsible of successful release of virions to the environment. Given that both functions act either on cell surface or murein components, both of which are in itself highly divergent (Scott and Barnett, 2006; Davis and Weiser, 2011), the variability is possibly best explained by host-parasite evolutionary arms race.

In most tectiviral elements, the polymerase is immediately followed by a short and conserved gene coding for a protein that binds to LexA (a host-encoded repressor activated upon bacterial stress, see e.g. Fornelos et al., 2016) enhancing its binding to the promoter that activates the operon for lytic functions. Further, this protein in sufficiently high concentrations can interfere with bacterial SOS response (Fornelos et al., 2015), suggesting a phage-adaptation to control normal bacterial regulation pathways under stress. *Brevibacillus* element lacks the gene for LexA binding protein, but at the same loci it harbors a gene encoding a putative protein with distant similarity to DNA unwinding helicases, DNA binding proteins and zinc finger proteins. Further, in both *B. flexus* and *B. pumilus* elements, the gene is similarly missing but has been replaced with two and three genes, respectively, without any similarity to proteins in databases. Given the genomic location, it is possible that these genes code for proteins that have functions in controlling normal gene expression and/or stress response of the host and may thus be of value for future research in these species.

Brevibacillus element in general has most genes without resemblance to other members of the group. The gene in the beginning of the virion assembly region encodes a possible TonB-domain (COG0810), which is known to link inner and outer membranes in Gram-negative bacteria. Given the inner membrane in Tectivirus virion, the protein can be speculated to be a structural

component of the capsid. None of the putative proteins encoded from genes in the 3' end of the element has significant similarity to proteins in databases suggesting that the element encodes a novel receptor binding spike. Further, the second last gene in *Exiguobacterium antarctica* element is dissimilar to other tectiviral elements but similar to peripherally T4-related phage Pf16 infecting *Pseudomonas putida*, indicating potential gene exchange between a Tectivirus and a member of *Caudovirales*. Previously similar recombination was identified in PRD1-like tectiviruses that apparently acquired a lysis cassette from *Caudovirales* (Krupovic et al., 2008). Also, its notable that around half of the elements contain putative genes of below 100 nucleotides in length (sometimes being as short as 63 nts), a limit utilized by many algorithms for detecting open reading frames. These putative genes are shared by multiple elements, suggesting that they are not annotation artefacts. In PRD1, short genes encode peptides with transmembrane helices and are known to be involved in linking DNA packing complex to the virion (Strömsten et al., 2003). Yet, most of the short peptides analyzed for this study have no predictable transmembrane regions (although some do) and possibly carry out different functions. Many of the genes indeed have identical counterparts in certain plasmids of the same or closely related species suggesting that they do provide fitness benefits for several types of extrachromosomal genetic elements that are not likely to form virions. Nevertheless, as short genes are likely to go unnoticed by automated tools and thus, even if potential homologs exist in sequences *per se*, their translations are likely to be absent from curated non-redundant protein databases that lack majority of directly deposited sequences (such as BLAST). As has been argued previously (Samoya et al., 2011; Andrews and Rothnagel, 2014), it is still a somewhat overlooked area of research to scan existing (bacterial and bacteriophage) genomes solely for putative short genes that show characteristics of being actually translated into peptides (such as having significant Shine-Dalgarno sites, encode structurally stable peptides and the ORFs having counterparts in whole genome alignments with closely related species) and then associate them with different phenotypic, ecological and environmental qualities.

To conclude, tectiviral elements in Gram-positive bacteria are conserved for their gene synteny and for genes necessary for building up the core of the virion. Genomic variability is observed for regions that are responsible of controlling the life cycle as well as mediating host attachment and lysis. All elements reside in bacteria known to be present in soil (along with various other environments). *Brevibacillus* and *Streptococus* are not known to form spores, thus implying that *Tectiviridae* in Gram-positive bacteria are not restricted to spore-formers. In the future, whole-genome sequencing projects are likely to produce more Tectivirus-like elements, thus expanding the known host spectrum even further. Yet the complete absence of new elements from Gram-negative hosts despite of the exponentially increasing amount of sequences from all possible samples suggests that they may indeed be very rare.

MATERIALS AND METHODS

Amino acid sequence of major capsid proteins of known Tectivirus members AP50, Gil16, Bam35 and pBClin15 (complete genome RefSeq ids NC_011523, NC_006945, NC_005258 and NC_00472, respectively), were used as seeds for BLAST (https://blast.ncbi.nlm.nih.gov/). Significant hits to unknown sequences were manually checked for the size and the presence of other Tectivirus-like genes. Fourteen sequences of likely tectiviral origin were obtained and thereafter handled with Geneious 11.0.2 software (www.geneious.com). PHASTER (Arndt et al., 2016; http://phaster.ca) was used to further investigate whether the elements can be recognized by automated tools as prophages by submitting the complete sequence for analysis (all of them being identified correctly as prophages related to Bam35). Whole genome alignments were conducted with ClustalW 2.1. (Larkin et al., 2007) with IUB cost matrix and gap open and extension costs of 7 and 4, respectively. The annotations of the sequences were checked manually for potential errors and unmarked putative genes, and were re-annotated accordingly (as described in Results and

Discussion). All putative gene products of the novel elements along with (putative) genes in all previous Tectivirus sequences were translated into a local protein database. Then each protein from each element was BLASTed against this database. Two proteins were considered potential homologs when the selected E- values were considered positive hits using the following criteria. The E - value to a hit was below 1.0e-1 or in the case of short proteins (i.e.< 60 amino acids) the pairwise align to a similar sized protein including identical or similar amino acids. Naturally, in some cases where the similarity is relatively low, potential homology cannot be inferred in terms of the genes sharing a common ancestral gene in an ancestor virus, but may rather implicate different types of relationships such as being originated by duplication of partial genes or transposition of partial sequences elsewhere in the genome. A total of approximately 550 proteins were manually checked for their hits to putative proteins in other members of the family. Short proteins of below around 35 amino acids in size were investigated for the existence of putative transmembrane regions with TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The existence of potential homologs in databases to putative gene products of Tectivirus-like elements was conducted with BLAST. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). Divergence of MCPs and ATPases were inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The trees in Figure 1 and in Supplementary Figure 1 are the ones with the highest log likelihoods (-7105.24 and -3586.80, respectively). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All positions containing gaps and missing data were eliminated.

ACKNOWLEDGEMENTS

The study was supported by the Academy of Finland (grants #252411 and #297049) and Emil Aaltonen Foundation. Authors thank C. Illingworth for the proofreading of the manuscript.

REFERENCES

1. Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martín C, Burnett RM, Stuart DI, Bamford DH, Bamford JK. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. Nature. 432:68-74.
2. Andrews SJ, Rothnagel JA. 2014. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet. 15:193-204.
3. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res.44:W16-21.
4. Bamford DH, Burnett RM, Stuart DI. 2002. Evolution of viral structure. Theor Popul Biol. 61:461-70.
5. Berjón-Otero M, Villar L, Salas M, Redrejo-Rodríguez M. 2016. Disclosing early steps of protein-primed genome replication of the Gram-positive tectivirus Bam35. Nucleic Acids Res. 44:9733-9744.
6. Berjón-Otero M, Lechuga A, Mehla J, Uetz P, Salas M, Redrejo-Rodríguez M. 2017. Bam35 tectivirus intraviral interaction map unveils new function and localization of phage ORFan proteins. J Virol. pii: JVI.00870-17. doi: 10.1128/JVI.00870-17.
7. Caldentey J., Blanco L., Savilahti H., Bamford D.H., Salas M. In vitro replication of bacteriophage PRD1 DNA. Metal activation of protein-primed initiation and DNA elongation. Nucleic Acids Res. 1992;20:3971–3976.
8. Cockburn JJ, Abrescia NG, Grimes JM, Sutton GC, Diprose JM, Benevides JM, Thomas GJ Jr, Bamford JK, Bamford DH, Stuart DI. 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature. 432:122-5.
9. Davis KM, Weiser JN. 2011. Modifications to the peptidoglycan backbone help bacteria to establish infection. Infect Immun. 79:562-70.
10. Fornelos N, Butala M, Hodnik V, Anderluh G, Bamford JK, Salas M. 2015. Bacteriophage GIL01 gp7 interacts with host LexA repressor to enhance DNA binding and inhibit RecA-mediated auto-cleavage. Nucleic Acids Res. 43:7315-29. doi: 10.1093/nar/gkv634.
11. Fornelos N, Bamford JK, Mahillon J. 2011. Phage-borne factors and host LexA regulate the lytic switch in phage GIL01. J Bacteriol. 193:6008-19. doi: 10.1128/JB.05618-11.
12. Gillis A, Mahillon J. 2014a. Phages preying on Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis: past, present and future. Viruses. 6:2623-72. doi: 10.3390/v6072623.

13. Gillis A, Mahillon J. 2014b. Prevalence, genetic diversity, and host range of tectiviruses among members of the Bacillus cereus group. Appl Environ Microbiol. 80:4138-52. 10.1128/AEM.00912-14.
14. Gillis A, Mahillon J. 2014c. Influence of Lysogeny of tectiviruses GIL01 and GIL16 on Bacillus thuringiensis Growth, Biofilm Formation, and Swarming Motility. Appl Environ Microbiol. 80: 7620–7630.
15. Ivanova N, Sorokin A, Anderson I, Galleron N, Candelon B, Kapatral V, Bhattacharyya A, Reznik G, Mikhailova N, Lapidus A, Chu L, Mazur M, Goltsman E, Larsen N, D'Souza M, Walunas T, Grechkin Y, Pusch G, Haselkorn R, Fonstein M, Ehrlich SD, Overbeek R, Kyrpides N. 2003. Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis. Nature. 423:87-91.
16. Jalasvuori M, Pawlowski A. and Bamford J.K.H (2010) A Unique Group of Virus-Related Genome Integrating Elements Found Solely in the Bacterial Family Thermaceae and the Archaeal Family Halobacteriaceae. J. Bact. 192:3231-3234.
17. Jalasvuori M, Koonin E. 2015. Classification of Prokaryotic Genetic Replicators: Between Selfishness and Altruism. Ann NY Acad Sci. 1341:96-105.
18. Jalasvuori M., Örmälä A-M., and Bamford, J.K.H. 2009. On the astrobiological relevance of viruses in extraterrestrial ecosystems. Int J Astrobio 8:95-100.
19. Jalasvuori M, Palmu S, Gillis A, Kokko H, Mahillon J, Bamford JK, Fornelos N. 2013. Identification of five novel tectiviruses in Bacillus strains: analysis of a highly variable region generating genetic diversity. Res Microbiol. 164:118-26. doi: 10.1016/j.resmic.2012.10.011.
20. Jones D.T., Taylor W.R., and Thornton J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comp. App. Biosci. 8: 275-282.
21. Kan S, Fornelos N, Schuch R, Fischetti VA. 2013. Identification of a ligand on the Wip1 bacteriophage highly specific for a receptor on Bacillus anthracis. J. Bacteriol. 195:4355–4364. 10.1128/JB.00655-13
22. Krupovic M, Bamford DH. 2010. Order to the viral universe. J Virol. 84:12476-9.
23. Krupovic M, Cvirkaite-Krupovic V, Bamford DH. 2008. Identification and functional analysis of the Rz/Rz1-like accessory lysis genes in the membrane-containing bacteriophage PRD1. Mol Microbiol. 68:492-503.
24. Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33:1870-1874.
25. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947-2948.
26. Mattila S, Ruotsalainen P, Jalasvuori M. 2015. On-demand isolation of bacteriophages against drug-resistant bacteria for personalized phage therapy. Frontiers in Microbiology. 6: 1271.
27. Olsen, R. H., J. Siak, and R. H. Gray. 1974. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. J. Virol. 14:689–699.
28. Onyenwoke RU, Brill JA, Farahi K, Wiegel J. 2004. Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). Arch Microbiol 182: 182-192

29. Philippe C, Krupovic M, Jaomanjaka F, Claisse O, Petrel M, le Marrec C. 2018. Bacteriophage GC1, a Novel Tectivirus Infecting Gluconobacter Cerinus, an Acetic Acid Bacterium Associated with Wine-Making. Viruses, 10:pii E39.
30. Ravantti JJ, Gaidelyte A, Bamford DH, Bamford JK. 2003. Comparative analysis of bacterial viruses Bam35, infecting a Gram-positive host, and PRD1, infecting Gram-negative hosts, demonstrates a viral lineage. Virology. 313:401-14
31. Samayoa J, Yildiz FH, Karplus K. 2011. Identification of prokaryotic small proteins using a comparative genomic approach. Bioinformatics. 27:1765-71.
32. Sakaki Y, Yamada K, Oshima M, Oshima T. 1977. Bacteriophage phiNS11: a lipid-containing phage of acidophilic thermophilic bacteria. II. Purification and some properties of the phage. J Biochem. 82:1451-6.
33. Saren AM, Ravantti JJ, Benson SD, Burnett RM, Paulin L, Bamford DH, Bamford JK. 2005. A snapshot of viral evolution from genome analysis of the tectiviridae family. J Mol Biol. 350:427-40.
34. Sinclair RM, Ravantti JJ, Bamford DH. 2017. Nucleic and Amino Acid Sequences Support Structure-Based Viral Classification. J Virol. 91. pii: e02275-16.
35. Scott JR, Barnett TC. 2006. Surface proteins of Gram-positive bacteria and how they get there. Annu Rev Microbiol. 60:397-423.
36. Stromsten N.J., Benson S.D., Burnett R.M., Bamford D.H., Bamford J.K. 2003. The Bacillus thuringiensis linear double-stranded DNA phage Bam35, which is highly similar to the Bacillus cereus linear plasmid pBClin15, has a prophage state. J. Bacteriol. 185:6985–6989.
37. Strömsten NJ, Bamford DH, Bamford JK. 2003. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. J. Virol. 77:6314-6321
38. Pawlowski A, Rissanen I, Bamford JKH, Krupovic M, Jalasvuori M. 2014. "Gammasphaerolipovirus", a newly proposed bacteriophage genus, unifies viruses of halophilic archaea and thermophilic bacteria within the novel family "Sphaerolipoviridae". Arch Virol. 159:1541-54.
39. Verheust C., Fornelos N., Mahillon J. GIL16, a new Gram-positive tectiviral phage related to the Bacillus thuringiensis GIL01 and the Bacillus cereus pBClin15 elements. J. Bacteriol. 2005;187:1966–1973.
40. Verheust C., Jensen G., Mahillon J. pGIL01, a linear tectiviral plasmid prophage originating from Bacillus thuringiensis serovar israelensis. Microbiology. 2003;149:2083–2092.

Figure and Table Legends

Table 1. *Tectiviridae* related elements in databases.

Tectiviridae element	Tectiviridae element in bacterium	Sequence length	Putative genes	GC%	RefSeq id	Putative genes in original annotation
BThuPhage1	Bacillus thuringiensis serovar sumiyoshiensis strain BGSC 4AO1	14329	31	39.9	NZ_NFCM01000003	23
BCerPhage1	Bacillus cereus strain MOD1_Bc143	14642	27	33.9	NZ_NBNG01000017	20
BCerPhage2	Bacillus cereus strain FSL M8-0473	15258	26	38.1	NZ_MUAP01000079	21
BCerPhage3	Bacillus cereus strain MOD1_Bc67	12062	25	37.4	NZ_MIFF01000080	19
BCerPhage4	Bacillus cereus VD184	15259	30	40.1	NZ_KB976851	25
BCerPhage5	Bacillus cereus VD166	14445	29	36.1	NZ_JH791864	22
BPumPhage1	Bacillus pumilus strain CB01	14220	29	42.2	NZ_LYXP01000014	24
BLicPhage1	Bacillus licheniformis strain B4092	13295	24	44.1	NZ_LQYK01000004	19
ClosPhage1	Clostridium sp. HMSC19B11	8527	18	38.6	NZ_KV785301	13
EAntPhage1	Exiguobacterium antarcticum DSM 14480	14829	28	42.2	NZ_JMKS01000002	23
BFlexPhage1	Bacillus flexus T6186-2	13112	27	36.0	NZ_JANV01000006	22
SPneuPhage1	Streptococcus pneumoniae strain 6B	14440	31	35.4	NZ_CWJH01000041	21
SpneuPhage2	Streptococcus pneumoniae strain 38	14558	31	38.9	NZ_CMPS01000060	21
BreviPhage1	Brevibacillus sp. CF112	13935	31	45.9	NZ_AKKB01000094	25

Table 2. DNA identity matrix of complete Tectivirus-elements.

Element/Tivirus	Bacillus pumilus strain CB01	Bacillus licheniformis strain B4092	Clostridium sp. HMSC19B11	Bacillus flexus T6186-2	Bacillus cereus strain FSL M8-0473	Bacillus cereus phage pBClin15	Bacillus cereus strain MOD1_Bc67	Bacillus cereus VD166	Bacillus thuringiensis serovar sumiyoshiensis strain BGSC 4AO1	Bacillus thuringiensis phage Bam35	Bacillus thuringiensis phage Gil16	Bacillus cereus strain MOD1_Bc143	Bacillus cereus VD184	Streptococcus pneumoniae strain 6B	Bacillus anthracis phage AP50	Bacillus anthracis phage Wip1	Exiguobacterium antarcticum DSM 14480	Brevibacillus sp. CF112
Bacillus pumilus strain CB01		55.8	48.0	45.5	37.2	37.2	37.4	36.4	36.6	37.3	36.8	36.8	36.1	37.5	36.8	36.3	35.7	33.1
Bacillus licheniformis strain B4092			46.9	44.3	36.3	36.3	37.4	35.9	36.3	36.3	36.1	35.9	35.1	36.2	35.8	35.1	34.4	32.2
Clostridium sp. HMSC19B11				44.3	37.5	37.5	37.9	36.1	36.1	36.8	36.3	36.5	35.4	37.1	36.7	35.3	34.7	34.2
Bacillus flexus T6186-2					38.4	38.4	40.1	38.1	38.9	38.0	37.8	37.4	37.2	38.1	37.8	37.5	36.6	35.3
Bacillus cereus strain FSL M8-0473						100.0	54.1	52.0	52.3	54.8	53.7	53.6	56.9	54.4	53.5	45.6	38.6	36.4
Bacillus cereus linear plasmid pBClin15							54.1	52.0	52.3	54.8	53.7	53.6	56.8	54.4	53.5	45.6	38.6	36.4
Bacillus cereus strain MOD1_Bc67								82.4	83.4	73.9	73.5	68.7	63.4	49.0	49.2	46.9	38.0	35.7
Bacillus cereus VD166									82.9	69.3	70.3	71.9	58.5	48.0	47.6	46.3	36.9	34.7
Bacillus thuringiensis serovar sumiyoshiensis strain BGSC 4AO1										79.6	69.8	66.3	63.9	47.3	47.4	46.6	37.3	34.7
Bacillus thuringiensis phage Bam35											82.3	79.9	67.2	50.6	50.7	45.4	39.0	35.5
Bacillus thuringiensis phage Gil16												80.4	63.3	50.4	50.7	44.7	38.3	35.4
Bacillus cereus strain MOD1_Bc143													64.3	50.0	50.2	43.9	38.6	35.2
Bacillus cereus VD184														52.5	52.4	43.2	39.1	34.7
Streptococcus pneumoniae strain 6B															83.5	60.4	40.3	37.0
Bacillus anthracis phage AP50																58.5	39.7	36.1
Bacillus anthracis phage Wip1																	37.3	33.9
Exiguobacterium antarcticum DSM 14480																		34.6

Figure 1. Molecular Phylogenetic analysis of the major capsid protein sequences of the Tectivirus-like elements. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The percentage of trees in which the associated taxa clustered together is shown next to the branches. There was a total of 348 positions in the final dataset.

Figure 2. Genomic comparison of known tectiviruses and Tectivirus-like elements. The translations of all annotated open reading frames (ORFs) were compared to the translations of annotated ORFs in other elements. The whole genome alignment was conducted via ClustalW. The stated functions for genes in the bottom of the figure are those determined or suggested for phage AP50, but the functions of genes in similar aligned positions in other phages may still differ.

Figure 1

Click here to download Figure: Virology_Fig1.pptx

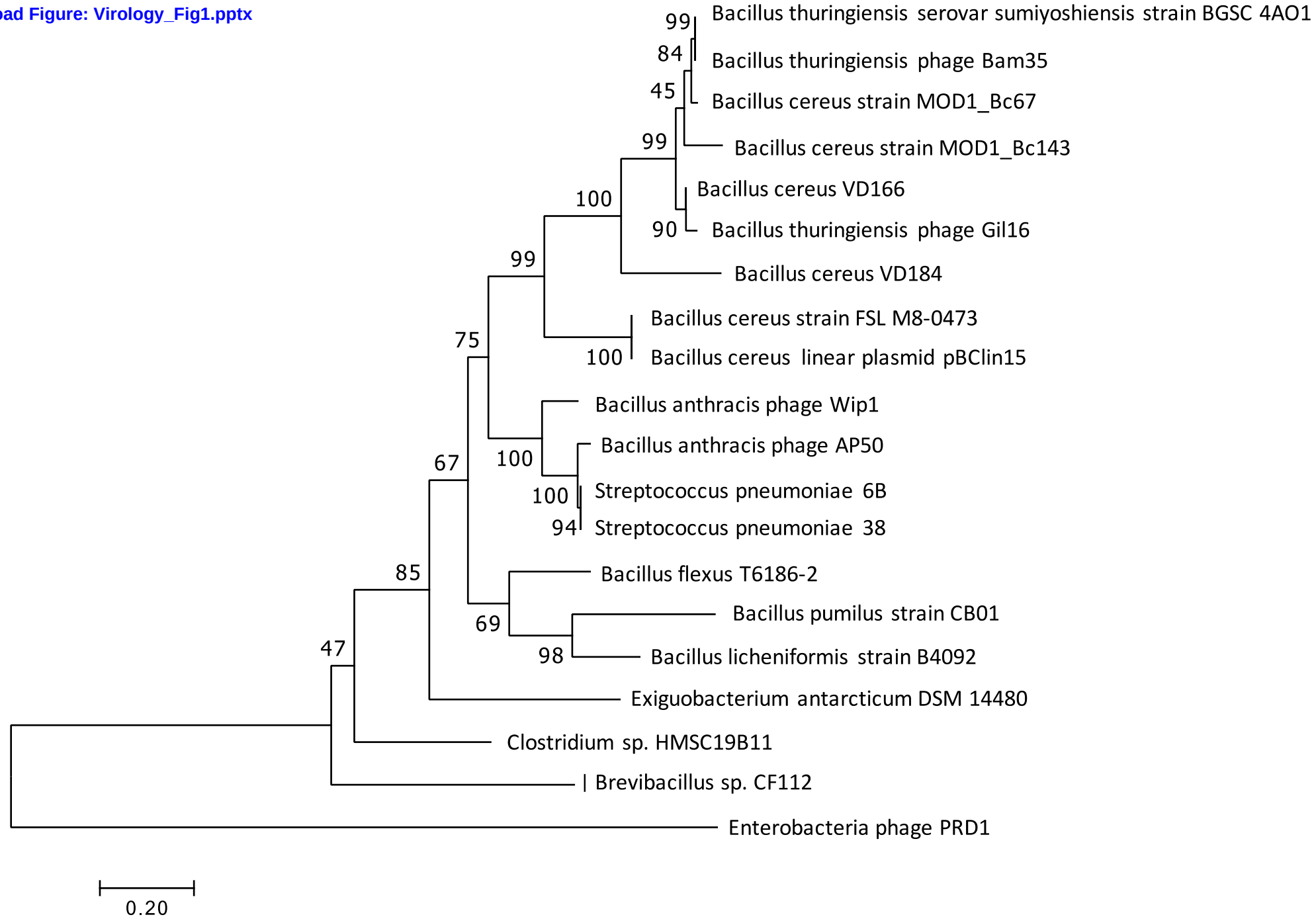

Figure 2
Click here to download Figure: Virology_Figure2.pdf

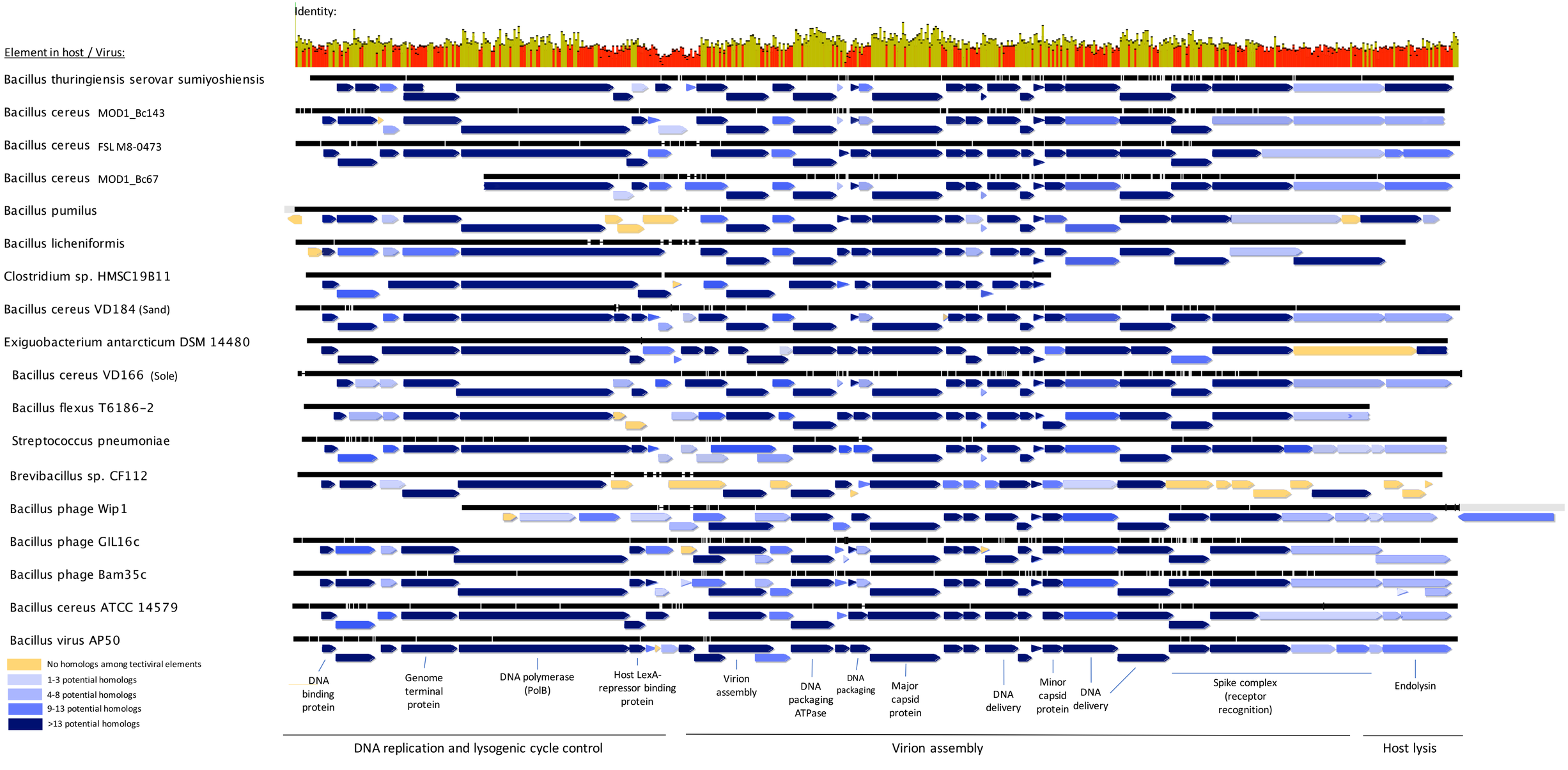

Table 1

Tectiviridae element	Tectiviridae element in bacterium
BThuPhage1	Bacillus thuringiensis serovar sumiyoshiensis strain BGSC 4AO1
BCerPhage1	Bacillus cereus strain MOD1_Bc143
BCerPhage2	Bacillus cereus strain FSL M8-0473
BCerPhage3	Bacillus cereus strain MOD1_Bc67
BCerPhage4	Bacillus cereus VD184
BCerPhage5	Bacillus cereus VD166
BPumPhage1	Bacillus pumilus strain CB01
BLicPhage1	Bacillus licheniformis strain B4092
ClosPhage1	Clostridium sp. HMSC19B11
EAntPhage1	Exiguobacterium antarcticum DSM 14480
BFlexPhage1	Bacillus flexus T6186-2
SPneuPhage1	Streptococcus pneumoniae strain 6B
SpneuPhage2	Streptococcus pneumoniae strain 38
BreviPhage1	Brevibacillus sp. CF112

Table 2

Element in / virus	Bacillus pumilus strain CB01	Bacillus licheniformis strain B4092
Bacillus pumilus strain CB01		55.8
Bacillus licheniformis strain B4092		
Clostridium sp. HMSC19B11		
Bacillus flexus T6186-2		
Bacillus cereus strain FSL M8-0473		
Bacillus cereus linear plasmid pBClin15		
Bacillus cereus strain MOD1_Bc67		
Bacillus cereus VD166		
Bacillus thuringiensis serovar sumiyoshiensis strain BGSC 4AO1		
Bacillus thuringiensis phage Bam35		
Bacillus thuringiensis phage Gil16		
Bacillus cereus strain MOD1_Bc143		
Bacillus cereus VD184		
Streptococcus pneumoniae strain 6B		
Bacillus anthracis phage AP50		
Bacillus anthracis phage Wip1		
Exiguobacterium antarcticum DSM 14480		
Brevibacillus sp. CF112		

Supplementary Material (To be Published)

Click here to download Supplementary Material (To be Published): Virology_SI.docx