Biogenesis of a Bacteriophage Long Non-Contractile Tail

Anait Seul, Sandrine Brasilès, Isabelle Petitpas, Rudi Lurz, Valérie Campanacci, Christian Cambillau, Frank Weise, Mohamed Zairi, Paulo Tavares, Isabelle Auzat

Biogenesis of a Bacteriophage Long Non-Contractile Tail

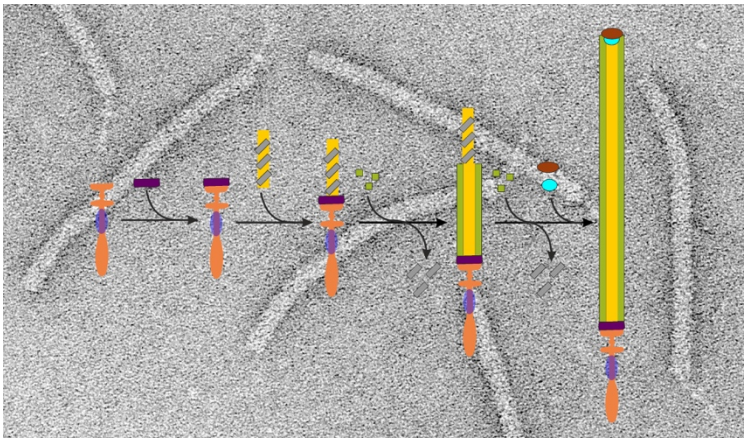
Anait Seul[1], Sandrine Brasilès[1,2], Isabelle Petitpas[1], Rudi Lurz[3], Valérie Campanacci[4, #], Christian Cambillau[4], Frank Weise[3,#], Mohamed Zairi[1], Paulo Tavares[1,2,*], and Isabelle Auzat[1,2,*]

***1 - Unité de Virologie Moléculaire et Structurale**, Centre de Recherche de Gif, CNRS UPR 3296 and IFR115, CNRS, Gif-sur-Yvette, France*
***2 - Institute for Integrative Biology of the Cell**, Université Paris-Saclay, CEA, CNRS, 91198, Gif-sur-Yvette, France*
***3 - Max Planck Institute for Molecular Genetics**, D-14195 Berlin, Germany*
***4 - Architecture et Fonction des Macromolécules Biologiques**, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, Marseille, France.*

[*]Correspondence to: Paulo Tavares, paulo.tavares@i2bc.paris-saclay.fr (*P. Tavares*) or to Isabelle Auzat, isabelle.auzat@i2bc.paris-saclay.fr (*I. Auzat*). Tel. +33 1 6982 3860

[#]Present addresses: Valérie Campanacci, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France; Frank Weise, Department of Molecular Biology, NMI, Reutlingen, Germany.

Graphical Abstract



Research Highlights

- Long bacteriophage tails are nanotube devices for delivery of viral DNA to the host cell cytoplasm
- Sequential program of protein interactions leading to assembly of phage SPP1 tail is uncovered
- Tail chaperones are essential to stabilize the TMP structure and for tail tube assembly
- Gp16.1, a highly conserved phage protein, acts during assembly of the tail interface for capsid binding
- SPP1 uses a minimal protein set to build long phage tails

Biogenesis of a Bacteriophage Long Non-Contractile Tail

Anait Seul[1], Sandrine Brasilès[1,2], Isabelle Petitpas[1], Rudi Lurz[3], Valérie Campanacci[4,#], Christian Cambillau[4], Frank Weise[3,#], Mohamed Zairi[1], Paulo Tavares[1,2,*], and Isabelle Auzat[1,2,*]

***1 - Unité de Virologie Moléculaire et Structurale**, Centre de Recherche de Gif, CNRS UPR 3296 and IFR115, CNRS, Gif-sur-Yvette, France*
***2 - Institute for Integrative Biology of the Cell**, Université Paris-Saclay, CEA, CNRS, 91198, Gif-sur-Yvette, France*
***3 - Max Planck Institute for Molecular Genetics**, D-14195 Berlin, Germany*
***4 - Architecture et Fonction des Macromolécules Biologiques**, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, Marseille, France.*

[*]Correspondence to: Paulo Tavares, paulo.tavares@i2bc.paris-saclay.fr (*P. Tavares*) or to Isabelle Auzat, isabelle.auzat@i2bc.paris-saclay.fr (*I. Auzat*). Tel. +33 1 6982 3860

[#]Present addresses: Valérie Campanacci, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France; Frank Weise, Department of Molecular Biology, NMI, Reutlingen, Germany.

Abstract

Siphoviruses are main killers of bacteria. They use a long non-contractile tail to recognize the host cell and to deliver the genome from the viral capsid to the bacterial cytoplasm. Here, we define the molecular organization of the *Bacillus subtilis* bacteriophage SPP1 ~6.8 MDa tail and uncover its biogenesis mechanisms. A complex between gp21 and the tail distal protein (Dit) gp19.1 is assembled first to build the tail cap (gp19.1-gp21Nter) connected by a flexible hinge to the tail fiber (gp21Cter). The tip of the gp21Cter fiber is loosely associated to gp22. The cap provides a platform where tail tube proteins (TTPs) initiate polymerization around the tape measure protein gp18 (TMP), a reaction dependent on the non-structural chaperones gp17.5 and gp17.5* (TACs). Gp17.5 is essential for stability of gp18 in the cell. Helical polymerization stops at a precise tube length followed by binding of proteins gp16.1 and gp17 (THJP) to build the tail interface for attachment to the capsid portal system. This finding uncovers the function of the extensively conserved gp16.1-homologs in the assembly of long tails. All SPP1 tail components, apart from gp22, share homology to conserved proteins whose coding genes' synteny is broadly maintained in siphoviruses. They conceivably represent the minimal essential protein set necessary to build functional long tails. Proteins homologous to SPP1 tail building blocks feature a variety of add-on modules that diversify extensively the tail core structure, expanding its capability to bind host cells and to deliver the viral genome to the bacterial cytoplasm.

KEYWORDS siphoviruses; macromolecular assembly; chaperone; tape measure; delivery device

Introduction

Tail tubes are machines designed to deliver macromolecules to target cells. They were first identified in bacterial viruses (bacteriophages or phages). The function of phage tails is to recognize specifically host bacteria and to deliver a surgical attack of the cell envelope, opening the way for viral genome transfer from the phage capsid to the bacterial cytoplasm [1–9]. A diverse group of structurally and functionally related long contractile tails was later found to be built by bacteria for delivery of protein effectors to other bacteria or eukaryotic cells [10–15]. More than 95 % of the currently known bacterial viruses have a

tail device associated to an icosahedral capsid that shields the viral genome (order *Caudovirales* [16]). This structural organization was retained throughout evolution for efficient infection of Eubacteria and several Archaea [17]. Phage tails can be short (family *Podoviridae*), long but non-contractile (family *Siphoviridae*) or long and contractile (family *Myoviridae*). The tails of siphoviruses and myoviruses are composed of a long tube and of a host adsorption apparatus. Their architectural components are built by structurally conserved proteins. In addition, myoviruses have a distinctive sheath around the tail tube, absent in siphoviruses, which contracts at the onset of infection. Tails are attached to the unique portal vertex of the icosahedral capsid through which DNA enters and exits the capsid [18]. The end of the tail distal from the capsid has the adsorption apparatus that is responsible for interaction with the host cell surface. This structure varies significantly in complexity from a single fiber to elaborated baseplates [2,19–24]. The sequential nature of the siphoviruses' long tail assembly pathways was established in pioneer studies of phage lambda and supported by more recent studies on lactococcal phages Tuc2009 [25] and TP901-1 [26], but the underlying molecular assembly mechanisms remain largely unknown.

Bacteriophage SPP1 is a well-characterized siphovirus that infects the Gram-positive bacterium *Bacillus subtilis*. Its long non-contractile tail follows the conserved building plan of long-tailed phages featuring an adsorption apparatus at the end distal from the capsid. Adsorption of SPP1 to *B. subtilis* is mediated by the gp21 (Receptor Binding Protein (RBP) of SPP1) carboxyl terminus (Cter) region (Receptor Binding Protein (RBP) of SPP1) of the tail fiber [4]. The fiber is attached to the tail cap [19] that is composed of the gp21 trimer amino terminus (Nter) and a gp19.1 hexamer (Distal Tail Protein (Dit) of SPP1) [27,28]. This structure closes the 160 nm-long tail tube that is built by a helical array of the Tail Tube Proteins (TTPs) gp17.1 and gp17.1* [19,29] surrounding a core probably formed by the Tape Measure Protein (TMP) gp18. Purified gp17.1 has the intrinsic property to polymerise into nanotubes of variable length that have a helical lattice similar to the one found in phage tails [30,31]. The tail tube is tapered by the Tail-to-Head Joining Protein (THJP) gp17 that creates the interface for attachment to the capsid connector structure [32,33]. Although SPP1 tail protein components and their organization are partially known, we still lack crucial information on the complete set of effectors necessary for tail assembly and on the program of their sequential reactions to build the complete structure. Here we describe the composition of the SPP1 tail, its architectural

organization, and the set of sequential reactions leading to its assembly. Bioinformatics and comparative biology reveal that the SPP1 tail is representative of a minimal core to build a long tail nano-device.

Results

Morphology and composition of bacteriophage SPP1 tails

In order to determine the SPP1 tail composition we established methods to obtain highly purified structures (Figure 1A and Materials & Methods). Tails were purified from lysates of the non-permissive *B. subtilis* strain YB886 infected with a conditional lethal mutant defective for production of the major capsid protein gp13 (SPP1*sus31*, hereafter abbreviated SPP1*gp13*$^{-}$) [34]. A fraction of the tails was precipitated by 10 % (w/v) polyethylene glycol 6000-8000 (pellet P2 in Figure 1A) while a second fraction required addition of 1 M NaCl to be pelleted (P3) from the supernatant of S2 (Figure 1A). The resuspended tail pellets were fractionated by anion exchange chromatography (AEC). Tails from P2 and P3 exhibited a slightly different chromatographic behaviour: they elute between ~230 and ~520 mM NaCl in the case of P2 and between ~280 and ~580 mM NaCl in the case of P3 (Figure 1A). The procedure yielded highly pure tails in case of P3 (best fraction P3f8) while elimination of minor contaminants from P2 (pool of P2f4 to P2f10 of the AEC) required sedimentation through a 5-20 % (w/v) sucrose gradient (best fraction P2f6S) (Figure 1A). The purified tails were identical to the ones found in phage particles. They consisted of a 174 nm-long flexible tube terminating at the head proximal end in a ring narrower than the tail tube diameter [29] and in a tail fiber on the other end (Figure 1B). The fiber (or tip) that is anchored at the tail cap [19,27] is a thin structure with a central ellipsoid-like element and a broader, somewhat diffuse, extremity (Figure 1B).

Protein amino-terminal sequencing (gp19.1 and gp21), mass spectrometry (gp16.1, gp17, gp17.1, gp17.1*, gp18, gp19.1 and gp21), and western blotting (gp17, gp17.1, gp17.1*, gp18, gp19.1, gp21 and gp22) allowed to assign gp17.1, gp17.1*, gp18, gp19.1, and gp21 to specific bands in the Coomassie Blue pattern of purified tails (Figure 1C,D, Figure S1 and Table S1). Gp18 was detected in three independent bands by mass spectrometry (Figure 1C, Figure S1, Table S1). The species with low electrophoretic mobility (>175 kDa) corresponds probably to gp18 aggregates or SDS-resistant multimers

($gp18_{mult}$). The apparent molecular masses of the two other bands are approximately 145 kDa ($gp18_{145kDa}$) and 73 kDa ($gp18_{73kDa}$). Independent mass spectrometry experiments showed that both species have a good coverage of most of the gp18 sequence (900 residues of the 1032 total protein sequence) (Figure S1 and Table S1). There is no straightforward explanation for this behavior. It may be speculated that the faster migrating band corresponds to differently truncated versions of gp18. Alternatively, the full-length gp18 adopts two conformations that bind SDS differently leading to an abnormal faster migration of one of them ($gp18_{73kDa}$) as found for some proteins with transmembrane segments [35]. An atypical electrophoretic behavior of gp18 can result of its basic theoretical pI (10.4), its putative transmembrane segments (Figure S2A), and its extended conformation inside the tail tube [19]. The three gp18 forms were recognized by a polyclonal antibody raised against its N-terminal region ($gp18_{181-435}$) (Figure 1D) although the band corresponding to the full-length protein $gp18_{145kDa}$ was frequently less intense or not detected (e.g. phage particles in Figure 1D). We have thus used the $gp18_{mult}$ and $gp18_{73kDa}$ bands as the most faithful markers for wild type gp18 detection. Gp16.1 and gp17 co-migrated in a single band in SDS-PAGE (Figure 1C).

Interestingly, gp22 was present in P3f8 tails but was found neither in P2f6S tails not in phage particles (Figure 1D). This feature did not correlate with any detectable difference of tail morphology (Figure 1B) but presence of gp22 could eventually account for the different precipitation and chromatographic behaviour of P3f8 and P2f6S tails (Figure 1A). Immunoelectron microscopy with anti-gp22 IgG showed no detectable labeling of the two tail preparations, probably because the antibody stripped gp22 weakly bound to the P3f8 tails. This hypothesis was confirmed by the observation that when tails were pre-fixed with glutaraldehyde, the tail fiber region of P3f8 tails was cross-linked by IgG through the region between their ellipsoid element and the distal extremity (Figure 2A,B). Anti-rabbit IgG colloidal gold labelling confirmed the specificity of gp22 immunolabeling (Figure 2C). Gp22 thus associates to the tail tip structure during SPP1 assembly but the interaction is weak after release from infected cells leading to dissociation of gp22 from the tip structure. To investigate if the association with gp22 is a widespread feature of gp21 homologous proteins we carried pBlast to identify homologs of gp22. The resulting 43 Blast hits identified proteins encoded immediately downstream of the gene coding for a gp21 homolog (Figure S3A). Gp21-homologs co-encoded with gp22 present homology to the complete SPP1 gp21 sequence (Figure S3B), consistent with an interaction of gp22

with the gp21 Cter fiber domain. Both gp22 and the sub-set of gp21 proteins are very closely related to their homologs (Figure S3C,D). However, more than 100 genomes code proteins highly homologous to gp21 (e-value < 2e-150) throughout their complete length but no detectable gp22-homologs (Figure S3E). We conclude that the gp21-gp22 interaction, although conserved in a sub-set of phages, is not an essential feature for gp21 fiber function.

Identification of SPP1 tail conditional lethal mutants

The library of SPP1 conditional lethal mutants [36] was screened for phages unable to produce tail structures or that produced DNA-filled capsids without tails in non-permissive infections. Mutations were subsequently mapped to specific SPP1 genes using a plasmid complementation assay. All putative tail genes were localized between the SPP1 capsid and lysis gene clusters (Figure 3). Previously uncharacterized *sus* mutations were mapped in genes *16.1* (*sus666*), *17.5*/*17.5** (*sus9*), *17.5** (*sus222*) and *18* (*sus76*) (Figure 3 and Table S2). Thermo sensitive (*ts*) mutations impaired in multiplication of SPP1 at 46 °C were assigned by complementation to genes *19.1* (*tsJ20*) and *21* (*tsJ5*, *tsJ15*, and *tsA58*) (Figure 3 and Table S2). DNA sequencing showed that mutants in gene *21* carried multiple mutations leading to several single amino acid substitutions in gp21 (Figure 3). However, the three mutants were complemented by a plasmid coding for gp21 assigning the temperature sensitivity phenotype to this protein. Other conditional lethal mutations mapped in genes coding for tail proteins but that were not further characterized are compiled in Figure 3 and in Table S2. All mutants are hereafter named according to the gene they affect for simplicity (e.g. SPP1*sus666* is named SPP1*gp16.1*$^{-}$; in the case that several mutants of a single gene are studied the original mutant name is included as, for example, SPP1*gp21*$^{-}$*(tsJ5)*).

The tail composition analysis combined with the mutants mapping and sequencing identified essential genes for tail assembly (Tables S1 and S2). They all coded for SPP1 tail structural proteins with the exception of gp17.5 and gp17.5* (Figure 1D). Gp17.5* was proposed to be generated by a -1 frameshift during translation of gene *17.5* at a TTTTTT slippery sequence based on bioinformatics [29,37]. Mutation *sus222* in SPP1*gp17.5**$^{-}$ introduced a stop codon downstream of this putative frameshift position disrupting specifically the gene *17.5** coding frame (Figure 3 and Table S2). This finding

demonstrated that gp17.5* is generated by a recoding event and that it is an essential protein for SPP1 multiplication.

No conditional lethal mutations were identified in gene *22* and in open reading frames *orf23*, *orf23.1*, and *orf24* localized between gene *21* and the lysis genes *24.1*, *25* and *26* [38,39].

Stability of SPP1 tail protein components

Synthesis of SPP1 proteins required for tail assembly was detected 12-15 min (10 min in case of gp17.5) after infection of *B. subtilis* at 37 °C and their amount increased until lysis (Figure 4A). Gp17.1, gp17.5, gp19.1, gp21 and gp22 appear as single bands in western blots. In the case of gp18, the two or three different bands found in purified phage particles and tails (see above; Figure 1D) were detected specifically with the gp18 Nter antibody also in extracts of infected bacteria (white arrowheads in Figure 4). Furthermore, those bands were absent in infections with SPP1*gp18*$^{-}$ (Figure 4B). This feature and the increase of intensity of bands during infection were used as criteria to distinguish gp18 species from several non-specific bands recognized by the antibody (Figure 4). In contrast to other tail components that were stable upon freezing and thawing of extracts, the gp18 species were labile, being detected best in freshly prepared extracts, and their ratios varied somehow between preparations.

The presence of different tail proteins was analyzed in cells infected with mutants defective for the production of functional forms of each individual tail protein. Non-permissive infections with suppressor-sensitive mutants showed that the corresponding gene products were absent in infected cells (Figure 4B). Gp17.1, gp17.5, gp19.1, gp21 and gp22 were detected as single, stable bands, independently of the presence of other tail proteins (Figure 4B,C). Gp21 was present in non-permissive infections of SPP1*gp21*$^{-}$*(tsJ15)*, SPP1*gp21*$^{-}$*(tsA58)* and SPP1*gp21*$^{-}$*(tsJ5)* (Figure 4C) revealing that the mutant proteins were produced at non-permissive temperatures but were not-functional. Analysis of SPP1*gp19.1*$^{-}$ that codes gp19.1$_{S244N}$ revealed that the mutant protein appears less stable in non-permissive infections (Figure 4C). The amino acid substitution S244N disrupts hydrogen bonding of S244 with Y175 and Q242 in the galectin domain of gp19.1 (Figure S4A-D) [27] that holds the gp21Nter in the SPP1 cap [28], providing a plausible

explanation for gp19.1$_{S244N}$ lower stability and non-functionality at high temperature. Gp17.1 of SPP1*gp19.1*$^{-}$ exhibited reproducibly a reduced electrophoretic mobility (Figure 4C) due to presence of an additional mutation (Table S2). The resulting gp17.1$_{I135M}$ amino acid substitution did not impair tail assembly because the temperature sensitive phenotype of this phage was complemented by gene *19.1* expressed from a plasmid.

Contrary to other tail components, detection of the mature gp18 pattern of bands found in assembled tails required the production of several other tail proteins. No gp18 was detected in absence of SPP1*gp17.5*$^{-}$, while an additional gp18 species with an apparent mass of ~83 kDa (gp18$_{83kDa}$) was found when gp17.1, gp17.5*, gp19.1 or gp21 were absent (diamonds in Figures 4B,C and 5). In contrast, infections with SPP1*gp13*$^{-}$, which is defective in capsid assembly but forms normal tails [34], yielded a gp18 band pattern identical to the one found in cells infected with wild type phages and in SPP1 virions. Collectively, these results show that gp18 relies on interactions with gp17.5 for stability and on assembly of the tail tube to achieve the protection pattern observed in infectious virions.

Tail proteins complexes and definition of the SPP1 tail assembly pathway

In order to define the network of interactions leading to tail assembly, we produced extracts of cells infected with mutants defective for production of each tail protein. Stable tail protein complexes formed at different steps of the assembly pathway were enriched by centrifugation through a sucrose cushion and then sedimented through a 10-30 % (w/v) glycerol gradient. Different running times followed by fractionation and composition analyses of both sedimentation steps were tested to obtain complete tail structures or its assembly intermediates (Figure S5).

Under conditions in which complete tails are assembled (SPP1*gp13*$^{-}$ infection; Figure 1), all tail proteins peaked as one fast sedimenting species corresponding to complete tails (Figure 6A, top panel). These structures were rather pure and morphologically similar to the tails in Figure 1B. High purity was achieved, when necessary, by an additional anion exchange step as in Figure 1A. When the major tail proteins gp17.1/gp17.1*, the tape measure protein gp18, or the non-structural proteins gp17.5/gp17.5* were absent, gp19.1 and gp21 co-sedimented as a ~19 *S* form (Figure 6A). The sedimentation coefficient was

estimated by comparison to the behavior of standards of known sedimentation coefficient run in parallel (Figure 6B, bottom panel). Silver staining showed that the corresponding fractions were significantly contaminated by other proteins (not shown) hampering electron microscopy observation of the putative gp19.1-gp21 complexes. Slower sedimenting species corresponding to full length gp21 (~11 *S*) and a shorter gp21 form (8.5 *S*) were detected. They corresponded to species with apparent molecular masses of ~142 kDa and ~81 kDa in SDS-PAGE, respectively. The shorter gp21 form, detectable with antibodies raised against both the protein Nter and Cter regions, was not found in extracts of infected cells (Figure 4), showing that it was generated during the partial purification procedure. The two gp21 forms co-sedimented partially in a sample from non-permissive infection with SPP1*gp19.1*$^{-}$ that codes for the non-functional mutant gp19.1$_{S244N}$ (Figure 6B). In contrast, gp19.1 was not found in the upper fractions of the glycerol gradient of a sample from infection with SPP1*gp21*$^{-}$(*tsJ5)* (Figure 6B), probably because gp19.1 was not pelleted in the sucrose cushion step that precedes the glycerol gradient (Figure S5).

Assembly of tails with wild type length require presence of gp21, gp19.1, gp18, gp17.5, gp17.5*, and the TTP gp17.1 (the minor TTP gp17.1* species is not essential for phage assembly [29]) (Figure 6). Analysis of the composition of purified tails identified two other tail structural proteins, gp17 and gp16.1 (Figure 1C and Table S1). We have previously shown that gp17 is the SPP1 tail-to-head joining protein found at the tail end distal from the adsorption apparatus [33]. In order to investigate the role and topology of gp16.1 in the tail structure, we used phage SPP1*gp16.1*$^{-}$ (Figure 3). During infection under non-permissive conditions tail proteins, including gp17, are produced but not gp16.1 (Figure 7A). Tails purified from those lysates had a sedimentation behavior similar to wild type tails (Figures S6 and 6A, respectively). They lacked gp16.1, as anticipated, but also gp17 (Figure 7B). Electron microscopy observation showed that the tails had a normal adsorption apparatus and tube length (Figure 7C). However, they did not exhibit the gp17 ring that caps the tail tube in wild type tails (white arrows in Figure 7C) [33]. Therefore, gp16.1 is a structural component required for normal tail construction. It binds to the assembled tail tube extremity preceding attachment of gp17 that forms the head-to-tail joining ring found in isolated tails.

Plasticity of the tail components building plans

We then carried out homology searches of the SPP1 tail proteins to investigate the landscape of their diversity and modular organization. Amino acid sequence homology was found predominantly to prophage-encoded proteins. Homologous proteins were detected in Firmicutes, mostly in phages and prophages of the *Bacillus* genus. Their genes synteny is highly conserved, following the arrangement found in the genome of SPP1 (Figure 3) [29,40,41]. A previous study showed that the major TTP of SPP1 gp17.1 can be decorated with a Cter FN3 domain that is exposed to the tube outside [29]. The alignment of the TMP gp18 homologous proteins reveals a pattern of homology throughout the complete gp18 sequence punctuated with internal deletions and insertions (Figure S2B) which likely play a role to determine tail length as shown experimentally for phages λ [42], T4 [43], and TP901-1 [44]. Most gp18-homologs have N- and C-termini longer than gp18 which might further extend the tail tube length and/or play a role in specific interactions with proteins at the tail tube ends.

Alignment of the 253 amino acids-long gp19.1 Dit amino acid sequence with its homologs revealed a modular organization (Figure S4E). Its Nter, the most conserved sequence, has a TTP-like fold in the gp19.1 hexamer (Figure S4A,B) that extends the tail tube [28]. The gp19.1 Cter galectin domain sequence is less conserved albeit homology is detected (Figure S4E). This domain irradiates outwards from the tail tube hexamer-like structure (Figure S4A,B) to clamp, together with the Nter ring, the gp21 Nter trimer in the tail cap structure [28]. A clear disruption of homology (Figure S4E) is found in a loop protruding to the outside of the galectin domain (Figure S4A). This marks an increase in sequence variability and length of gp19.1 homologs (Figure S4E). HHpred predicts that their Nter has a gp19.1 TTP-like fold, as expected, but also that the far Cter end is structurally related to the Cter of gp19.1 galectin domain. The two regions are separated in numerous cases by either an internal module of unknown function, by a carbohydrate binding module previously reported for "evolved Dits" of phages infecting other Gram-positive bacteria (PDB 6LY8) [45,46], or by combination of several modules. This is exemplified by HHPred structural homology analysis of the 783 amino acids-long Dit WP_077616833.1 that distinguishes three modules integrated in the galectin domain: a module of unknown function, a module with several structural homologs including bacterial cell envelope binders (PDB 6LVV, 3FLP), and the carbohydrate binding module of "evolved Dits" (PDB 6LY8). Collectively, these results highlight the remarkable structural plasticity of gp19.1 homologous proteins.

Gp21 homologs exhibit high sequence conservation on their Nter module (Figure S3B,E). In SPP1, $gp21_{1-522}$ was shown to form a trimer that closes the tail cap [28]. HHpred reveals also structural relatedness of $gp21_{1-399}$ to the structure of trimeric proteins from phage tails like a *Listeria monocytogenes* EGD-e prophage tail component (PDB 3GS9), *Staphylococcus* phage 80α (PDB 6V8I) [47] and the *Escherichia coli* myophage Mu tail baseplate (PDB 1WRU) but also to tail-related cellular structures like pyocin R2 (PDB 6U5H), the baseplate of *Photorhabdus asymbiotica* PVC system (PDB 6J0M), or the anti-feeding secretion system of *Serratia entomophila* (PDB 6RBK). The gp21 Nter cap module is thus a highly conserved trimeric building block found in tails of siphovirus and myophages of evolutionarily distant hosts that is also used by different tail tube-like cellular weapons [48]. These are frequently individual proteins. They can close reversibly the tail tube end [9,20,28] or extend the tube that is plugged by the tail cell puncturing device (PDB 1K28 [49], 1WRU). Their common function appears to be acting as a Tail Adaptor Protein (TAP) that builds the 6 to 3-fold symmetry transition at the tail end [9].

The gp21 Nter is followed by a sequence of weak structural homology to glycoside hydrolases consistent with a role of gp21 in cell wall degradation (Tal function) although this activity was not showed experimentally. The Cter of gp21 is significantly more diverse in sequence and no significant structure-based homology hits were found with HHpred. The gp21 closest homologs align through their complete length (Figure S3E) and their genes are frequently linked to genes coding gp22-homologs (Figure S3A,B). More distant gp21-homologs show homology within a shorter Cter region and then have an unrelated sequence at their Cter end or alternate segments of homology with unrelated sequences (Figure S3E). The less related gp21-homologs align only through their Nter with a clear homology breakpoint around residue 522 of gp21 (Figure S3E), in excellent correlation with the domain that closes the tail cap. Importantly, all these proteins have a Cter domain which can be longer than 1000 residues, irrespectively of the length of homology with gp21. We conclude that gp21-homologs represent a group of proteins that build long tail fibers covalently attached to the tail cap trimeric module that closes reversibly the tail tube. We propose that they are an all-in-one surgical device that binds the bacterial receptor, most likely perforates the cell envelope, and opens to deliver the TMP and viral DNA to the host cell. They differ from shorter trimers (<650 residues-long) whose complete structure is related to the gp21Nter (individual TAP proteins). Those are found in siphoviruses (e.g. phages p2 ORF16, *L monocytogenes* EGD-e prophage gp18,

staphylococcal phage 80α gp59), and in myoviruses (e.g. phages T4 gp27, Mu gp44) featuring complex baseplates with numerous RBPs hooked in the Dit protein hub and/or other proteins that form an axial fiber extending from the gp21 Nter-like cap region [9,20,22,47,49].

Discussion

In this work we established the bacteriophage SPP1 tail composition and characterized its assembly intermediates to obtain a model of the complete set of sequential protein-protein interaction program leading to the functional tail (Figure 8). The SPP1 tail has a simple architecture when compared to other siphoviruses representing the minimal core of essential elements to build a functional tail. The two SPP1 tail originalities are its transient gp22, found in a limited number of phages with SPP1 gp21 homologs (Figure S3), and the optional FN3 domain of the TTP gp17.1* [29].

Gp21 is a multi-functional trimeric protein whose Nter closes reversibly the tail cap (TAP function) [28] and the Cter has a receptor-binding activity (RBP) [4]. These functions are associated to two distinct domains connected by a flexible hinge underneath the cap that allows free rotation and tilting of the fiber RBP relative to the tail tube (Figure 2A,B) for three-dimensional search of receptor YueB in the *B. subtilis* surface. Like SPP1, the tail central fiber is the essential tool for host recognition and receptor binding in phages lambda [50] and T5 [9,51] which are the other well-characterized siphoviruses that use a bacterial protein receptor for host recognition. In these phages the tail cap plug and the RBP are different proteins while SPP1 gp21 typifies a single-protein minimal system to carry out both functions (TAP/RBP). This modular organization is highlighted by sequence alignment of gp21 homologous proteins (Figure S3E). Higher divergence of their RBP Cter module is consistent with increased plasticity to recognize different bacterial receptors while other homologous proteins lack the Cter module probably because the RBP function is accomplished by a different phage protein. A gp21 ~11 *S* species that is stably formed in absence of the cap gp19.1 (Dit) protein was the first precursor of tail assembly detected in our study (Figure 6B). This species is likely a trimer, as found in viral particles [19], but its elongated shape hampered molecular mass confirmation from the sedimentation data. The $gp21_{1\text{-}522}$ Nter cap plug could be produced soluble in heterologous systems [28] but full-length gp21 and its Cter were insoluble in spite of extensive optimisation work (25; our

observations). It is possible that correct folding and trimerization of the gp21 Cter fiber domain occurs at low gp21 concentration to prevent aggregation, or that it requires a dedicated, yet unidentified, chaperone during infection. The transient binding of gp22 to the fiber (Figure 2) renders it a potential candidate for this role. Although all gp22 homologous proteins are encoded adjacently to genes coding full-length gp21 homologs, they are found only on a minority of the large number of gp21 homologous proteins presently identified (Figure S3). We thus favour the hypothesis that gp22 does not play a critical role in tail fiber assembly but might have another function like shielding the tail fiber from adsorbing to cell debris when infected cells lyse.

When gp21 and gp19.1 are produced during SPP1 infection, they form the 19 *S* cap-tail fiber complex (Figure 6). The direct interaction between gp19.1 and the $gp21_{1\text{-}522}$ Nter purified proteins was previously demonstrated to build an SPP1 tail cap-like structure [28] similar to the one found in viral particles [19]. The 19 *S* complex accumulates in bacteria infected with phages that do not produce the major tail tube protein gp17.1, the tape measure gp18 or its non-structural chaperones gp17.5 and gp17.5* (Figure 6). Thus, in the case of SPP1, no stable intermediate is formed between the gp19.1-gp21 complex and tail tube assembly effectors in contrast to phage lambda, where a complex of the tail fiber-cap with the TMP gpH and possibly the chaperone gpG was detected [52,53]. We hypothesize that such complex is also assembled to initiate SPP1 tail tube formation (bracketed in Figure 8) but is unstable when the TTP does not polymerize around the TMP scaffold to build the tail tube. Anchoring of the TMP C-terminus trimer on the Dit central cavity of phage 80α [47] further supports the notion that this interaction is central to initiate tube assembly.

Detection of SPP1 TMP gp18 in infected cells depends strictly on the presence of the non-structural protein gp17.5 (Figures 4B,C and 5). Gp17.5 chaperones most likely gp18 folding and/or shields gp18 from protease attack, thus maintaining it in an extended conformation competent for tail tube assembly before protection inside the tube. In mutants defective for assembly of the cap-tail fiber complex (gp19.1 and gp21) or of the tail tube (gp17.1 and gp17.1*), there is a $gp18_{83kDa}$ species that is reproducibly found in infected cells but is absent if tail tube construction is achieved. We conclude that p17.5 is necessary for $gp18_{83kDa}$ stability. $Gp18_{83kDa}$ could be a precursor of $gp18_{73kDa}$ during assembly as the latter species is very poorly detected when the tail tube is not assembled

(Figure 4). However, the low amount of gp18 in infected bacteria and the lability of its forms hampered investigation of this hypothesis. Gp17.5*, whose first 112 residues are similar to gp17.5 (Figure 4A) [29], is not necessary to chaperone $gp18_{83kDa}$ (Figure 5) but is essential for tail tube assembly. This is consistent with a role of the gp17.5* Nter to cross-talk with the gp17.5-gp18 complex and its unique Cter to have a different function. The Cter could act as a tail cap-gp18-gp17.5 initiator complex that primes gp17.1 helical polymerization for building-up the tail tube, as proposed for phage lambda [54,55]. There are, however, noteworthy differences between the two phage systems, such as the simpler two-protein cap-tail fiber system that primes tail tube assembly and the requirement of gp17.5 for SPP1 TMP gp18 stability in the infected cell. In contrast, lambda full-length gpH is stably produced, yet in a non-functional form [55], and associates to the tail tip when the chaperone gpG is not present in the cell [52,53].

Another major finding of our work was the assignment of SPP1 gp16.1 to a specific role at the stage of tail assembly termination. Bioinformatics showed that gp16.1 homologous proteins encoded by a gene adjacent to the THJP gene are widespread in siphoviruses and myoviruses consistent with a conserved function [41]. Mutants of phages lambda (gpZ) [52], P2 (gpS) [56], Mu (gpG) [57], and TP901-1 ($Tap_{TP901-1}$) [26] related their gp16.1 homologs to phage infectivity or to tail to capsid joining. Furthermore, their relative of phage T5 is a tail component (T5 (p143) [51]. However, their precise function in phage assembly remained unknown. We have now filled this gap showing that gp16.1 is a component of the tail that is incorporated on its structure after tube assembly. The TMP gp18 remains stable at these assembly steps between the halt of tail tube polymerization and the final viral particle state (Figure 7B). Gp16.1 thus acts on the assembled tail tube and its presence correlates with recruitment of gp17 to form a ring at the extremity of the tail (arrow in Figure 7C). This ring is a hexameric structure that binds to the capsid connector gp16 protein [58], finalizing assembly of the infective virion [33].

Collectively, our work provides a comprehensive model of the complete SPP1 tail assembly pathway and elects it as a minimal model system for tail build-up. Its essential eight core protein components are gp21 and gp19.1 constructing the cap-tail fiber complex, the tail tube gp18 and gp17.1 building blocks assisted by non-structural chaperones gp17.5 and gp17.5*, and the tail-capsid joint effectors gp16.1 and gp17. Functional equivalents of these proteins are found in the vast majority of long-tailed phages (Figures S2-S4).

Remarkably, all components of the fiber and tube can fuse with additional domains (e.g. the insertions in the galectin domain of gp19.1-homologs [Figure S4], additions to the gp17.1 Cter [29], and the elaboration of gp21 Cter-homologs [Figure S3]) or recruit other protein elements (e.g. the hub role of gp19.1-homologs to build elaborated baseplates [59]). Such plasticity delivers a large panoply of functional add-ons that diversify the phage particle capabilities to adsorb to surfaces (48 and references therein), recognize different host receptors [15,45,46,59], to attack the cell envelope [61], and to initiate host cell takeover [62]. These large combinatorial possibilities offer phages with immense opportunities to change and adapt. The SPP1 tail offers a gadget-free system, apart from the dispensable gp17.1* Cter FN3 domain [29] and possibly gp22 (Figure 2), to characterize the limited number of essential sequential reactions to build a long flexible tail. The molecular principles driving the process apply very likely to the general building of tail tube-like delivery devices.

Materials & Methods

Bacterial and phage strains

E. coli DH5α, was used for cloning procedures. Bacteriophages SPP1 wild type [63], *sus* and *ts* mutants were from the MPIMG collection [36]. The permissive strains *B. subtilis* HA101B [64] was used to amplify all *sus* mutants with the exception of SPP1*gp17.5/gp17.5**- (mutant SPP1*sus9* [34]) that was amplified in *B. subtilis* BG295 [65]. Mutants *ts* were multiplied in strain YB886 [66] at 30°C. Non-permissive infections were carried in strain YB886 for *sus* mutants and in the same strain at 46°C for *ts* mutants.

Cloning procedures

Genes *16.1* to *17.1* were amplified from SPP1 phage DNA by PCR using primers 60102BZ and 60102AZ (Table S3). The PCR fragment cleaved with EcoRI-PstI was cloned into vector pBluescript pSK- (Stratagene) cut with the same enzymes generating pSK17. A BamHI-MslI fragment carrying gene *16.1* was excised from pSK17 and cloned in plasmid pRSETA (Invitrogen) cleaved with BamHI-PvuII to generate pBT361.

Genes *17.5* to *22* were amplified from SPP1 phage DNA using primers 17.5CS1 and

448 23NCS1 (Table S3) in a Expand^{TM} Long Template PCR System (Roche) reaction. The
449 PCR fragment cleaved with BamHI-PstI was cloned into plasmid pHP13 [67] cut with the
450 same enzymes to generate pPT220. Plasmid pPT220 was then cleaved with EcoRI and re-
451 ligated to obtain pPT221 that codes genes *21* and *22*.

452 Genes *17.5* to *17.5** were amplified from SPP1 phage DNA by PCR using primers
453 gene 17.5 CS and gene 17.5* NCS. The PCR fragment was cleaved with BamHI-PstI and
454 cloned into plasmid pRSETA cleaved with the same enzymes to generate pPT118. Genes
455 *17.5* to *17.5** of pPT118 were then amplified by PCR using primers prset1 and prset2
456 (Table S3). The PCR fragment cleaved with SmaI-PstI was cloned into SwaI-PstI sites
457 under the control of promoter PN25/0 in a pHP13-derived shuttle vector [68] generating
458 pPT200. This plasmid codes gp17.5 and gp17.5* fused to the pRSETA Nter tag featuring a
459 hexahistidine sequence (Invitrogen). Plasmid pPT200 was used as template for site-
460 directed mutagenesis with the QuikChange kit (Stratagene) using primers gene 17.5 to
461 17.5* CS and gene 17.5 to 17.5* NCS (Table S3) to insert a C nucleotide in the putative
462 slippery sequence (TTTTTT) proposed to generate the -1 frameshift for translation of
463 gp17.5* [29]. The resulting plasmid pPT209 codes tagged gp17.5*. Site-directed
464 mutagenesis of pPT200 with primers gene 17.5 only CS 2mut and gene 17.5 only NCS
465 2mut (Table S3) was used to degenerate the genes 17.5-17.5* putative slippery sequence
466 for production of gp17.5 only. A by-product clone (pPT212) carrying deletion of gene *17.5*
467 first phenylalanine codon and a nucleotide substitution changing the second phenyalanine
468 codon to leucine within the slippery sequence (TTTTTT→CTT) proved to be the most
469 effective way to disrupt the -1 frameshift. Plasmid pPT212 leads to exclusive
470 overproduction of tagged gp17.5 carrying mutation $F_{111}F \rightarrow L_{111}$.

471 Overproduction of several gp18 polypeptides was attempted due to the known
472 difficulty for stable production of the tape measure protein. The criteria to design
473 polypeptides for production were based on sequence homology conservation and to avoid
474 (i) putative transmembrane segments, (ii) breaking putative α-helices, (iii) predicted
475 unstructured regions, and (iv) low sequence complexity regions. Fragments coding $gp18_{181-435}$
476 (primers gene 18 181 CS and gene 18 435 NCS (Table S3)), $gp18_{347-435}$ (primers gene
477 18 347 CS and gene 18 435 NCS), and $gp18_{672-919}$ (primers gene 18Cter CS and gene 18
478 919 NCS) were generated by PCR. They were then cleaved with BamHI-PstI and cloned
479 into pRSETA (Invitrogen) cut with the same enzymes. The resulting plasmids pPT122,

480 pPT124 and pPT123, respectively, encode the gp18 polypeptides with a pRSETA Nter tag
481 containing a hexahistidine sequence.

482 A 756 bp fragment covering the gene *19.1* sequence lacking the initiation and stop
483 codons (position from 15365 to 16120 of the SPP1 nucleotide sequence; access code
484 X97918.3) was amplified by PCR from SPP1*wt* DNA with two flanking sequences coding
485 for 3′-PstI and 5′-AgeI restriction sites. Oligonucleotides F-gp19.1 and R-gp19.1 (Table
486 S3) were used as forward and reverse primer, respectively. The purified PCR product was
487 totally sequenced at GATC Biotech (Konstanz, Germany), cleaved with PstI and AgeI, and
488 cloned into the pIA2 shuttle vector [33] digested with the same restriction enzymes. The
489 resulting plasmid was named pIA29 (6His-gp19.1).

490 Genes *23* and *23.1* were amplified by PCR and cloned by Gateway recombination into
491 the pETG-20A vector using an established strategy [69,70]. The resulting plasmids
492 encoded His_6-tagged thioredoxin followed by a tobacco etch virus (TEV) protease cleavage
493 site fused to the Nter of gp23 (this work) and gp23.1 [70].

494 A PCR fragment bearing gene 24 was amplified with oligonoucleotides F-gp24 and R-
495 gp24 (Table S3) followed by cloning in the pIA2 vector [33] as described above for gene
496 *19.1*.

497

498 **Mapping and sequencing of SPP1 conditional lethal mutations**

499 The SPP1 conditional lethal mutants collection [36] was screened against individual clones
500 of the tail genes ORFeome in *B. subtilis* using a complementation spot assay in 96-well
501 plates. The screening led to the identification of *sus* mutants in genes *17.5*, *17.5**, and *18*.
502 An alternative large scale mapping approach used plasmids covering clusters of SPP1 late
503 genes (pPT25 [29], pPT220 and pPT221) followed by fine mapping to individual genes.
504 This approach led to identification of a supressor-sensitive mutant in gene *16.1* and of *ts*
505 mutants in genes *17.5*, *19.1* and *21*. Identification of mutations was carried out by DNA
506 sequencing of the gene complementing the conditional lethal mutant. The observation that
507 the gene *19.1*-defective mutant SPP1*ts20* codes for gp17.1 with a retarded electrophoretic
508 mobility (Figure 4C) prompted also sequencing of its gene *17.1* which carries a missense
509 mutation in addition to the one in gene *19.1* (Figure 3).

510

511 **Purification of SPP1 phage tails**

512 Growth of the bacterial host strain, lysis and enzymatic treatment of the lysate were carried
513 out at 37 °C and the following steps at 4 °C, unless stated otherwise. Figure 1A illustrates
514 the method. *B. subtilis* strain YB886 was grown in 1 L of LB medium to an OD_{600} of 0.8,
515 then supplemented with 10 mM $CaCl_2$ and infected with SPP1*gp13*$^-$ mutant phages at an
516 input multiplicity (i.m.) of 0.5. At 2 h post-infection, Roche complete protease inhibitors
517 (without EDTA) were added, and the lysate was treated with 200 µg mL^{-1} of lysozyme, 5
518 µg mL^{-1} of DNase and 80 µg mL^{-1} of RNase for 15 min. After overnight centrifugation at
519 12785 *g*, the supernatant was stirred with 10 % w/v of PEG 6000-8000 for 1 h, and
520 centrifuged for 1 h at 11325 *g*. The pellet P2 was resuspended in a total volume of 14 mL
521 of buffer TBT (100 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$) and dialysed
522 against 100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$. The P2 dialysate was centrifuged for
523 15 min at 13776 *g* and the supernatant was consecutively filtered through a 0.45 µm, and a
524 0.20 µm syringe filter (Sartorius). The sample was applied to a Ressource^{TM}Q anion
525 exchange column (GE Healthcare) and eluted with consecutive gradients of 0 M to 0.35 M
526 NaCl (6 column volumes (CV)) and 0.35M to 1M NaCl (0.5 CV) in 100 mM Tris pH 7.5,
527 10 mM $MgCl_2$ at room temperature and at a flow rate of 6 mL min^{-1}. Tail containing
528 fractions identified by SDS-PAGE and Coomassie Blue staining were pooled and
529 concentrated in a Vivaspin 20 device (MWCO 10,000, Sartorius), the buffer being
530 concomitantly exchanged for TBT. Aliquots of 200 µL each were loaded onto 5 to 20 %
531 w/v sucrose gradients prepared in TBT, which were spun in an SW41 rotor (Beckman) at
532 35000 rpm, 4°C, for 4 h and 15 min. Fractions of 1 mL each were collected starting from
533 the bottom of each tube, and analyzed by SDS-PAGE and Coomassie staining.

534 The supernatant S2 of the PEG precipitation step (Figure 1A) was precipitated by
535 addition of NaCl to 1 M over a period of 20 min. After stirring for another 1 h, the sample
536 was centrifuged for 1 h at 11325 *g*. The pellet P3 was resuspended in a volume of 14 mL of
537 TBT and dialysed against 100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$. The dialysate was
538 filtered through a 0.20 µm syringe filter and subjected to anion exchange chromatography,
539 as described for the resuspended P2 sample. Fractions containing purified SPP1*sus31* tails
540 were stored at 4 °C.

541 A lower scale phage tail purification starting from 130 mL culture infected either with
542 SPP1*gp13*$^-$ or with SPP1*gp16.1*$^-$ used the method developed for partial purification of tail

complexes from *B. subtilis* lysates followed by sedimentation of the semi-purified tail protein complexes through a 10 % to 30 % w/v glycerol gradient prepared in TBT buffer, as detailed in the sections with those titles below. Tail-enriched fractions were pooled, dialysed against 100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and subjected to an anion exchange chromatography step under the conditions described above to obtain pure tails.

Partial purification of tail protein complexes from *B. subtilis* lysates

Complexes of tail proteins were enriched and partially purified from lysates of *B. subtilis* YB886 infected with the SPP1*sus* or SPP1*ts* mutants of interest (Table S2). During growth, infection and lysis, cultures were maintained at 37 °C when working with *sus* mutants, and at 46 °C for the *ts* mutants. A 130 mL LB culture of the host strain, grown to OD_{600} = 0.8 and supplemented with 10 mM $CaCl_2$, was split for parallel infection of 40 mL cultures, each with the SPP1 mutant(s) under analysis, at an i.m. of 5 pfu cfu^{-1}. At 37 °C, lysis was allowed to proceed for 1 h after infection. The lysates were supplemented with Roche complete protease inhibitors and treated with 200 µg mL^{-1} of lysozyme, 10 µg mL^{-1} DNAse and 100 µg mL^{-1} RNAse for 15 min. At 46 °C, the time spans were reduced to 48 min post-infection and 12 min for enzymatic treatment. All the following steps were carried out on ice or at 4 °C. Lysates were pre-cleared by centrifugation for 15 min at 39,281 *g*, splitted in four aliquots of ~10 mL that were overlayed on top of 3 mL cushions of 20 % w/v sucrose in buffer TBT, and run at 32,000 rpm in a SW41 rotor (Beckman). The sedimentation time was 15 h or prolonged to 19 h in case of *ts* mutant lysates and of the *sus76* lysate run in parallel with those. After centrifugation, three fractions of approximately 8 mL, 2.1 mL and 1.6 mL were collected by pipetting, starting from the top of each tube, leaving the bottom 0.6 to 0.8 mL of the sucrose cushion in each tube. Pellets were resuspended in this remaining liquid, and material from the four tubes was pooled for each mutant.

Sedimentation of semi-purified tail protein complexes

The pooled samples from the sucrose cushions were concentrated to 200 µL each, and the buffer was concomitantly exchanged for sucrose-free TBT using Vivaspin 20 devices (MWCO 10,000). Samples were loaded onto 10 % to 30 % w/v glycerol gradients prepared

in buffer TBT and centrifuged at 35000 rpm, 4 °C in an SW41 rotor. Sedimentation times were 7 h in the case of SPP1*gp13*$^{-}$ and SPP1*gp16.1*$^{-}$, 12 h for the other SPP1*sus* mutants and 15 h for the *ts* mutants and the SPP1*gp18*$^{-}$ control sedimented in parallel with those (Figure 6). Twelve fractions of 1 mL each were collected with a syringe and peristaltic pump, starting from the bottom of each tube. Pellets were resuspended in the roughly 0.8 mL remaining in each tube after fractionation.

Mass spectrometry of tail proteins

Samples were separated on SDS-PAGE gels and stained with Coomassie Blue. Excised bands were digested with trypsin and peptides were analysed by MALDI-TOF (instrument: Voyager DE-STR, Applied Biosystems) and ESI-MS/MS (nanoAcquity /Q-TOF Premier, Waters). Detected peptides were searched against a database containing known and putative gene products of SPP1 and against the nrNCBI database, using the MASCOT (www.matrixscience.com [71]) and FindPept (http://expasy.org/tools/findpept-ref.html [72]) softwares. All mass spectrometry analyses were performed by D. Cornu at the Service d'Identification et de Caractérisation des Protéines par Spectrométrie de masse (SICaPS) d'IMAGIF (CNRS, Gif-sur-Yvette).

N-terminal sequencing of tail proteins

Samples were separated on 16 % SDS-PAGE gels and transferred as described above into a PVDF membrane, which had been briefly pre-incubated with methanol. After the transfer, the membrane was abundantly rinsed with water and Coomassie stained (0.1 % w/v Coomassie Blue, 40 % v/v ethanol, 1 % v/v acetic acid). After several steps of destaining with 50 % v/v ethanol, the membrane was dried on Whatman paper before cutting out the bands of interest. Amino-terminus sequencing was carried out by J. d'Alayer at the Plateforme d'Analyse et de Microséquençage des Protéines (Institut Pasteur, Paris), using a Procise ABI 494 sequencer (Applied Biosystems).

Identification of tail components by western blotting

Aliquots of fractions P3f8 and P2f6S (Figure 1) were concentrated 1.7-fold and 5-fold, respectively. 10 µL each of the concentrated samples and of disrupted phage particles ($2x10^{12}$ pfu mL^{-1}) were mixed with 5 x sample buffer (45 % v/v glycerol, 0.023 % w/v bromophenol blue, 14 % w/v SDS, 0.18 M Tris-HCl pH 6.8, 1.27 M β-mercaptoethanol) and heated to 100 °C for 5 min. Multiples of these three samples were separated on 15 % SDS-PAGE gels and transferred onto nitrocellulose membranes at 100 V during 2 h 45 min in western blot transfer buffer. Membranes were Ponceau stained, then blocked with 5 % low fat milk powder in PBS with 0.1% Tween 20 (PBS-T) buffer for 1 h. After washing twice for 10 min each with PBS-T, the membrane was cut into stripes, which were incubated for 40 min with a suitable dilution of the respective primary antibodies in PBS-T / 5 % milk. After washing as before, the membranes were incubated with a 10,000-fold dilution of the secondary antibody (goat anti-rabbit, peroxidase-coupled) in PBS-T for 30 min. After 6 final washes of 5 min in PBS-T, proteins were detected using ECL western Blotting Reagents (GE Healthcare) and medical X-ray films. DNA-filled capsids and phage particles were disrupted as described [73] to be used as references in Coomassie Blue-stained gels and western blot.

Detection of tail proteins in extracts of infected cells

Bacterial cultures of *B. subtilis* YB886 were grown at 37 °C for infection with SPP1*wt* or SPP1*sus* mutants, and at 46 °C for non-permissive infections with SPP1*ts* mutants. For direct comparison, SPP1*wt* and SPP1*gp17.1/gp17.1** control infections were also carried out at 46 °C. Cultures were grown in LB medium to an OD_{600} of 0.8 and supplemented with 10 mM $CaCl_2$. An aliquot of 2 mL of the non-infected culture was centrifuged for 2.5 min at 13,200 rpm and room temperature in an eppendorf 5415D centrifuge. The pellet was briefly re-spun to completely remove all liquid by pipetting and frozen at -20 °C. Meanwhile, the culture was infected with SPP1*wt* or with the mutant phage of interest at an i.m. of 3 pfu cfu^{-1}. For kinetics of tail protein production, aliquots of 2 mL were taken at various time points after infection and treated as described for the non-infected sample. For comparison of protein production in different mutants, samples were taken shortly before lysis, i.e. 25 min after infection at 37 °C and 20 min after infection at 46 °C. Frozen pellets were transferred to ice and resuspended in 100 µL of resuspension buffer (50 mM Tris-HCl, pH 8, 50 mM glucose, 1 mM EDTA, 1 mg mL^{-1} lysozyme freshly prepared, and

protease inhibitors (Roche)), followed by 5 min of incubation at room temperature. Samples were then mixed with 100 µL each of ice-cold lysis buffer (150 mM NaCl, 1% NP-40, 50 mM Tris-Cl pH 7.5, 5 mM $MgCl_2$, 1 µg mL^{-1} DNAse, 100 µg mL^{-1} RNAse, and protease inhibitors (Roche)), incubated on ice for 30 min and centrifuged for 30 min at 14000 *g* and 4 °C to remove debris. Extracts were either stored at -80 °C or used directly for SDS-PAGE and western Blot, which were done as described for the purified tail samples.

Proteins production, purification and antibody production

One liter of BL21 (DE3) (pLysS) (pBT361) was grown in Luria-Bertani broth (LB) supplemented with ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL) at 37 °C to an $OD_{560\ nm}$ of 0.8. The culture was induced with 5 mM IPTG and harvested 2 h later. The cell pellet was lysed in K-phosphate buffer pH 7.8, 500 mM KCl, 6 M guanidine hydrochloride and denatured tagged gp16.1 was purified by Ni-NTA affinity chromatography following the protocols established for protein (pC)CAT [74].

One liter of *E. coli* BL21DE3 (pLysS) (pPT212) that overproduces gp17.5 was grown in Luria-Bertani broth (LB) supplemented with ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL) at 30 °C to an $OD_{560\ nm}$ of 0.6 and induced with 1 mM IPTG. After 3 h of growth at 30 °C, bacteria were sedimented at 10,000 *g* for 30 min, pellets were resuspended in 30 mL LB, transferred to a 50 mL Falcon tube, re-sedimented, and the cell pellet was stored frozen at -80°C. The pellet was thawed on ice and resuspended in ice-cold 50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 25 mM imidazole supplemented with a mixture of antiproteases (Complete EDTA-free, Roche). Bacteria were lysed by 4 cycles of 3 min sonication with pauses of 1 min (amplitude 60, 30–40 W in a Vibra Cell sonicator) and centrifuged at 17,000 *g* for 1 h at 4° C. The supernatant containing soluble gp17.5 was filtered through a 0.22 µm filter and applied to a 5 mL HisTrap™ column coupled to an Akta system (GE Healthcare) equilibrated with 50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 25 mM imidazole. After washing with 25 mL of the same buffer, the column was eluted with a 20 mL linear gradient to 50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 500 mM imidazole. Fractions of the gp17.5 peak were collected in tubes containing EDTA to reach a final concentration of 1 mM, pooled, applied to a Superdex™ 75 26/60 size-exclusion chromatography column equilibrated with 50 mM Hepes-NaOH pH 7.5, 500

mM NaCl, 1 mM EDTA and run with the same buffer. Eluted samples containing gp17.5 were pooled, supplemented with a cocktail of antiproteases (Complete EDTA-free, Roche), and stored in aliquots at -80 °C.

Gp17.5* was produced in *E. coli* BL21DE3 (pLysS) (pPT209) and purified as described above for gp17.5 with the exception that no size exclusion chromatography step was included in the purification procedure.

E. coli BL21DE3 (pLysS) transformed with pRSETA plasmid derivatives coding different gp18 polypeptides were grown in Luria-Bertani broth (LB) supplemented with ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL) at 37 °C to an $OD_{560\ nm}$ of 0.6 followed by induction with 1 mM IPTG. After 2 h growth at 37°C, a 1 mL of culture was concentrated 10-fold by centrifugation and 5 µL were loaded on SDS-PAGE gels for Coomassie Blue staining or western blot developed with anti-hexahistidine antibodies and polyclonal serum generated against SPP1 viral particles. Detectable gp18 protein production was found only in *E. coli* BL21DE3 (pLysS) (pPT122) producing $gp18_{181-435}$. This strain was grown at 37°C in 1.5 L of LB broth supplemented with ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL) to an OD_{600} of 0.8 and induced with 1 mM IPTG. After 3 h of growth at 37°C bacteria were sedimented at 10,000 *g* for 30 min, resuspended in 30 mL LB, transferred to 50 mL Falcon tubes, re-sedimented, and the cell pellet was stored frozen at -80° C. The pellet was thawed on ice and resuspended in ice-cold 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole supplemented with a mixture of antiproteases (Complete EDTA-free, Roche). Bacteria were lysed by 5 cycles of 3 min sonication with pauses of 1 min (amplitude 60, 30–40 W in a Vibra Cell sonicator) and centrifuged at 17,000 *g* for 45 min at 4°C. The pellet containing insoluble $gp18_{181-435}$ was resuspended in 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole, 6 M guanidinium hydrochloride and stirred at room temperature for 1 h to solubilize $gp18_{181-435}$. The suspension was centrifuged at 17,000 *g* for 45 min at 20°C and the supernatant filtered sequentially through 0.45 and 0.22 µm filters. The sample was applied to a 1 mL HisTrap™ column coupled to an Akta system (GE Healthcare) equilibrated with 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole, 6 M guanidinium hydrochloride and washed with 10 mL of the same buffer at 16° C. The buffer was exchanged with 10 mL in 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole, 8 M urea and the protein was eluted with a 20 mL linear gradient to 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 500

699 mM imidazole, 8 M urea. Fractions containing gp18$_{181-435}$ were pooled, concentrated by
700 ultrafiltration in a Vivaspin device (10 kDa cut-off) and dialysed sequentially against 50
701 mM Hepes-KOH pH 8.0, 500 mM KCl, 6 M urea and against the same solution with 5 M
702 urea. Aliquots of denatured gp18$_{181-435}$ were stored at -80°C.

703 *E. coli* JS218 freshly transformed with plasmid pIA29 was grown at 37°C in LB
704 supplemented with erythromycin (30 µg/mL) and chloramphenicol (10 µg/mL) to an OD_{600}
705 of 0.35-0.40. Cultures were transferred to 27°C and induced with 2 mM IPTG when the
706 OD_{600} reached 0.50-0.55 and incubation was continued overnight. Pelleted bacteria were
707 resuspended in cold lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton
708 X100, 10 mM $MgSO_4$) supplemented with a mixture of antiproteases (Complete EDTA-
709 free, Roche) and disrupted by performing three bursts of 1 min (amplitude 60, pulse 3, 30–
710 40 W) in a sonicator (Vibra Cell) with pauses of 1 min. The crude extract was centrifuged
711 at 4°C twice for 20 minutes at 27000 x *g*. Imidazole was then added to a final
712 concentration of 20 mM and recombinant hexahistidine gp19.1 protein was purified by
713 metal-chelating chromatography using a HiTrap column (GE Healthcare) following the
714 supplier conditions (NaH_2PO_4 20 mM pH 7.4, NaCl 500 mM). The column was washed
715 with increasing concentrations of imidazole. Gp19.1 was eluted at 100 mM imidazole,
716 concentrated and further purified by size exclusion chromatography on a Superdex 75
717 column (GE Healthcare) run in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5.

718 Gp22, gp23 and gp23.1 fused to a TEV-cleavable hexahistidine-thioredoxin tag were
719 overproduced in *E. coli* and purified by nickel-affinity chromatography, TEV proteolysis
720 to remove the tag, and size exclusion chromatography as previously described [27,69,70].

721 Gp24 was produced from pIA39 following the protocol used for gp19.1 with some
722 modifications. Induction was performed at 37°C instead of 27°C for 3 hours with 2 mM
723 IPTG when the OD_{600} reached 0.6. After an identical treatment of the bacterial pellet, the
724 crude extract was directly loaded on a HiTrap column, without addition of imidazole.
725 Gp24 was eluted at 500 mM, concentrated and loaded on a Superdex 75 column (GE
726 Healthcare) run in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5.

727 Polyclonal antibodies were produced by immunisation of rabbits with purified soluble
728 (gp17.5, gp17.5*, gp19.1, gp22, gp23, gp23.1 and gp24) and denatured (gp16.1 and
729 fragment gp18$_{181-435}$) proteins as described [75] at the INAF platform (Gif-sur-Yvette,
730 France) and at FZB Biotechnik GmbH (Berlin, Germany). Sera against gp17, gp17.1Nter,

731 gp21Nter and gp21Cter were previously described [4,29,33]. The IgG fraction of anti-
732 gp22 serum was purified by affinity chromatography in a Hi-Trap Protein A HP (GE
733 Healthcare) as described by the suppliers.

734

735 **Electron microscopy and immunolabeling**

736 Tail samples were negatively stained with 2 % uranyl acetate before imaging by electron
737 microscopy [76].

738 Tails P2f6 and P3f8 in TBT buffer were split in two identical samples of 25 µL. One
739 set of samples was fixed with glutaraldehyde (Sigma-Aldrich) at a final concentration of
740 0.2 % for 10 min at room temperature. The reaction was quenched with 0.1 M
741 glycylglycine for 30 min at room temperature. Control samples were processed identically
742 with the exception that the fixative was replaced by an identical volume of water. All
743 samples were then labeled with purified anti-gp22 IgG overnight at room temperature. The
744 volume was adjusted to 60 µL and loaded on a microspin column filled with 400 µL
745 Sephacryl 500 slurry equilibrated with buffer TBT. The column was centrifuged for 5 min
746 at 800 g to recover the tails and immune-complexes in the flow through while unbound
747 IgG was retained in the size-exclusion matrix. An aliquot of each sample was used for
748 negative staining with 2% uranyl acetate and visualized by electron microscopy. The
749 remaining sample was incubated overnight with 1 µL of goat anti-rabbit IgG complexed
750 with 5 nm colloidal gold (BritishBioCell) and prepared for electron microscopy
751 observation as above.

752

753 **Bioinformatics**

754 SPP1 tail sequence-related proteins were detected by pBlast [77] (e-value ≤0.0001) and
755 filtered through CD-Hit [78] to identify proteins with less than 80 % identity through 80%
756 of their amino acid sequence relative to the shortest sequence. Sequences were then aligned
757 using Clustal Omega set to default parameters [79]. Residues conservation was color-
758 coded using the UGENE program default function [80]. Gp21 and gp22 pBlast results were
759 used unfiltered for the analysis in Figure S3. Neighbourhood of their coding genes was
760 investigated by manual inspection of their coding genomes. Guide trees were computed
761 from pBlast sequence alignments using Clustal Omega and displayed with UGENE [80].

Homology search combined with structure prediction was carried out using HHpred [81]. Prediction of transmembrane segments was carried using TMHMM.2 set to default parameters [82].

Protein structures were analysed and displayed with Pymol.

CRediT authorship contribution statement

Anait Seul: Writing - original draft, Conceptualization, Methodology - tails and tail sub-structures purification methods/tails characterization/tail proteins fate during infection, Investigation, Visualization. **Sandrine Brasilès**: Methodology – protein purification methods, Resources, Investigation. **Isabelle Petitpas**: Methodology – protein purification methods, Resources, Investigation. **Rudi Lurz**: Investigation, Visualization. **Valérie Campanacci**: Methodology – protein purification methods, Resources. **Christian Cambillau**: Writing – review & editing, Supervision, Project administration, Funding acquisition. **Frank Weise**: Writing – review & editing, Methodology – protein purification methods, Resources, Investigation. **Mohamed Zairi**: Investigation, Formal analysis, Visualization. **Paulo Tavares**: Writing - original draft, Conceptualization, Resources, Investigation, Visualization, Supervision, Project administration, Funding acquisition. **Isabelle Auzat**: Writing - original draft, Conceptualization, Methodology – protein purification methods/tail proteins fate during infection, Resources, Investigation, Visualization, Supervision, Project administration.

Data and Materials Availability Statement

Experimental data of this manuscript have been contained within this manuscript or are available upon request from P.T. (paulo.tavares@i2bc.paris-saclay.fr) or I.A. (isabelle.auzat@i2bc.paris-saclay.fr). Biological materials are available upon request from P.T or I.A.

Acknowledgments

We thank Marie-Christine Vaney (Institut Pasteur, Paris) for Figure S4A-D and useful discussions. David Cornu (SICaPS, I2BC) is acknowledged for mass spectrometry measurements. This work has benefited from the facilities and expertise of the I2BC proteomic platform (Proteomic-Gif, SICaPS), supported by IBiSA, Ile de France Region, Plan Cancer, CNRS and Paris-Saclay University. We thank Jacques d'Alayer (Plate-forme d'Analyse et de Microséquençage des Protéines, Institut Pasteur, Paris) for proteins N-terminal sequencing.

Funding

The work was supported by institutional funding from CNRS and by ANR (grant SiphoPhages to C.C. and P.T.).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplemental data

Supplemental data to this article can be found online at xxx.

Table S1, Fingerprint of SPP1 proteins studied in this work, PDF file, 0.1 MB

Table S2, Protein coding genes studied in this work and their conditional lethal mutants. See Figure 4 for a graphical representation, PDF file, 0.1 MB

Table S3, Primers used in this work, PDF file, 0.1 MB

Figure S1, Mass spectrometry of tryptic peptides from SPP1 tail proteins. PDF file, 0.2 MB

Figure S2, SPP1 gp18 putative transmembrane segments and sequence alignment with homologous proteins. PNG file, 2.2 MB

Figure S3, SPP1 gp22 homologs are co-encoded with gp21 homologs in a sub-set of genomes that code for gp21 homologs. PNG file, 8.5 MB

Figure S4, SPP1 gp19.1 domain structure and its homologous proteins. PNG file, 4 MB

Figure S5, Partial purification of tail protein complexes. PNG file, 0.4 MB

Figure S6, Sedimentation of SPP1 tails produced during non-permissive infection with SPP1*gp16.1*$^{-}$. PDF file, 0.4 MB

References

[1] C.A. Roessner, D.K. Struck, G.M. Ihler, Injection of DNA into liposomes by bacteriophage λ, J. Biol. Chem. 258 (1983) 643–648.

[2] J. Böhm, O. Lambert, A.S. Frangakis, L. Letellier, W. Baumeister, J.L. Rigaud, FhuA-mediated phage genome transfer into liposomes, Curr. Biol. 11 (2001) 1168–1175. https://doi.org/10.1016/S0960-9822(01)00349-9.

[3] C. São-José, S. Lhuillier, R. Lurz, R. Melki, J. Lepault, M.A. Santos, P. Tavares, The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA, J. Biol. Chem. 281 (2006) 11464–11470. https://doi.org/10.1074/jbc.M513625200.

[4] I. Vinga, C. Baptista, I. Auzat, I. Petipas, R. Lurz, P. Tavares, M.A. Santos, C. São-José, Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection, Mol. Microbiol. 83 (2012) 289–303. https://doi.org/10.1111/j.1365-2958.2011.07931.x.

[5] A.R. Davidson, L. Cardarelli, L.G. Pell, D.R. Radford, K.L. Maxwell, Long Noncontractile Tail Machines of Bacteriophages, Adv Exp Med Biol. 726 (2012) 115–142. https://doi.org/10.1007/978-1-4614-0980-9_6.

[6] B. Hu, W. Margolin, I.J. Molineux, J. Liu, The Bacteriophage T7 Virion Undergoes Extensive Structural Remodeling During Infection, Science (80-.). 339 (2013) 576–579. https://doi.org/10.1126/science.1231887.

[7] B. Hu, W. Margolin, I.J. Molineux, J. Liu, Structural remodeling of bacteriophage T4 and host membranes during infection initiation, Proc. Natl. Acad. Sci. 112 (2015) E4919–E4928. https://doi.org/10.1073/pnas.1501064112.

[8] C.-A. Arnaud, G. Effantin, C. Vivès, S. Engilberge, M. Bacia, P. Boulanger, E. Girard, G. Schoehn, C. Breyton, Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection, Nat. Commun. 8 (2017) 1953. https://doi.org/10.1038/s41467-017-02049-3.

[9] R. Linares, C.-A. Arnaud, S. Degroux, G. Schoehn, C. Breyton, Structure, function and assembly of the long, flexible tail of siphophages, Curr. Opin. Virol. 45 (2020) 34–42. https://doi.org/10.1016/j.coviro.2020.06.010.

[10] P.G. Leiman, M. Basler, U.A. Ramagopal, J.B. Bonanno, J.M. Sauder, S. Pukatzki, S.K. Burley, S.C. Almo, J.J. Mekalanos, Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin, Proc. Natl. Acad. Sci. 106 (2009) 4154–4159. https://doi.org/10.1073/pnas.0813360106.

[11] L.G. Pell, V. Kanelis, L.W. Donaldson, P. Lynne Howell, A.R. Davidson, The phage major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system, Proc. Natl. Acad. Sci. 106 (2009) 4160–4165. https://doi.org/10.1073/pnas.0900044106.

[12] G. Bönemann, A. Pietrosiuk, A. Mogk, Tubules and donuts: a type VI secretion story, Mol. Microbiol. 76 (2010) 815–821. https://doi.org/10.1111/j.1365-2958.2010.07171.x.

[13] A. Zoued, E. Durand, Y.R. Brunet, S. Spinelli, B. Douzi, M. Guzzo, N. Flaugnatti, P. Legrand, L. Journet, R. Fronzes, T. Mignot, C. Cambillau, E. Cascales, Priming and

polymerization of a bacterial contractile tail structure, Nature. 531 (2016) 59–63. https://doi.org/10.1038/nature17182.

[14] D. Böck, J.M. Medeiros, H.-F. Tsao, T. Penz, G.L. Weiss, K. Aistleitner, M. Horn, M. Pilhofer, In situ architecture, function, and evolution of a contractile injection system, Science (80-.). 357 (2017) 713–717. https://doi.org/10.1126/science.aan7904.

[15] A. Goulet, S. Spinelli, J. Mahony, C. Cambillau, Conserved and Diverse Traits of Adhesion Devices from Siphoviridae Recognizing Proteinaceous or Saccharidic Receptors, Viruses. 12 (2020) 512. https://doi.org/10.3390/v12050512.

[16] H. Ackermann, Phage Classification and Characterization, in: Methods Mol. Biol., 2009: pp. 127–140. https://doi.org/10.1007/978-1-60327-164-6_13.

[17] M. Krupovic, V. Cvirkaite-Krupovic, J. Iranzo, D. Prangishvili, E. V. Koonin, Viruses of archaea: Structural, functional, environmental and evolutionary genomics, Virus Res. 244 (2018) 181–193. https://doi.org/10.1016/j.virusres.2017.11.025.

[18] P. Tavares, The Bacteriophage Head-to-Tail Interface, in: J.R. Harris, D. Bhella (Eds.), Subcell. Biochem., Springer, 2018: pp. 305–328. https://doi.org/10.1007/978-981-10-8456-0_14.

[19] C. Plisson, H.E. White, I. Auzat, A. Zafarani, C. São-José, S. Lhuillier, P. Tavares, E. V. Orlova, Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection, EMBO J. 26 (2007) 3720–3728. https://doi.org/10.1038/sj.emboj.7601786.

[20] G. Sciara, C. Bebeacua, P. Bron, D. Tremblay, M. Ortiz-Lombardia, J. Lichiere, M. van Heel, V. Campanacci, S. Moineau, C. Cambillau, Structure of lactococcal phage p2 baseplate and its mechanism of activation, Proc. Natl. Acad. Sci. 107 (2010) 6852–6857. https://doi.org/10.1073/pnas.1000232107.

[21] C. Bebeacua, P. Bron, L. Lai, C.S. Vegge, L. Brøndsted, S. Spinelli, V. Campanacci, D. Veesler, M. van Heel, C. Cambillau, Structure and Molecular Assignment of Lactococcal Phage TP901-1 Baseplate, J. Biol. Chem. 285 (2010) 39079–39086. https://doi.org/10.1074/jbc.M110.175646.

[22] D. Veesler, C. Cambillau, A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries, Microbiol. Mol. Biol. Rev. 75 (2011) 423–433. https://doi.org/10.1128/MMBR.00014-11.

[23] D. Veesler, S. Spinelli, J. Mahony, J. Lichiere, S. Blangy, G. Bricogne, P. Legrand, M. Ortiz-Lombardia, V. Campanacci, D. van Sinderen, C. Cambillau, Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism, Proc. Natl. Acad. Sci. 109 (2012) 8954–8958. https://doi.org/10.1073/pnas.1200966109.

[24] N.M.I. Taylor, N.S. Prokhorov, R.C. Guerrero-Ferreira, M.M. Shneider, C. Browning, K.N. Goldie, H. Stahlberg, P.G. Leiman, Structure of the T4 baseplate and its function in triggering sheath contraction, Nature. 533 (2016) 346–352. https://doi.org/10.1038/nature17971.

[25] S. Mc Grath, H. Neve, J.F.M.L. Seegers, R. Eijlander, C.S. Vegge, L. Brøndsted, K.J. Heller, G.F. Fitzgerald, F.K. Vogensen, D. van Sinderen, Anatomy of a Lactococcal Phage Tail, J. Bacteriol. 188 (2006) 3972–3982. https://doi.org/10.1128/JB.00024-06.

[26] S.R. Stockdale, B. Collins, S. Spinelli, F.P. Douillard, J. Mahony, C. Cambillau, D. van Sinderen, Structure and Assembly of TP901-1 Virion Unveiled by Mutagenesis, PLoS

One. 10 (2015) e0131676. https://doi.org/10.1371/journal.pone.0131676.

[27] D. Veesler, G. Robin, J. Lichière, I. Auzat, P. Tavares, P. Bron, V. Campanacci, C. Cambillau, Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): A baseplate hub paradigm in gram-positive infecting phages, J. Biol. Chem. 285 (2010) 36666–36673. https://doi.org/10.1074/jbc.M110.157529.

[28] A. Goulet, J. Lai-Kee-Him, D. Veesler, I. Auzat, G. Robin, D.A. Shepherd, A.E. Ashcroft, E. Richard, J. Lichière, P. Tavares, C. Cambillau, P. Bron, The Opening of the SPP1 Bacteriophage Tail, a Prevalent Mechanism in Gram-positive-infecting Siphophages, J. Biol. Chem. 286 (2011) 25397–25405. https://doi.org/10.1074/jbc.M111.243360.

[29] I. Auzat, A. Dröge, F. Weise, R. Lurz, P. Tavares, Origin and function of the two major tail proteins of bacteriophage SPP1, Mol. Microbiol. 70 (2008) 557–569. https://doi.org/10.1111/j.1365-2958.2008.06435.x.

[30] C. Langlois, S. Ramboarina, A. Cukkemane, I. Auzat, B. Chagot, B. Gilquin, A. Ignatiou, I. Petitpas, E. Kasotakis, M. Paternostre, H.E. White, E. V. Orlova, M. Baldus, P. Tavares, S. Zinn-Justin, Bacteriophage SPP1 tail tube protein self-assembles into β-structure-rich tubes, J. Biol. Chem. 290 (2015) 3836–3849. https://doi.org/10.1074/jbc.M114.613166.

[31] M. Zinke, K.A.A. Sachowsky, C. Öster, S. Zinn-Justin, R. Ravelli, G.F. Schröder, M. Habeck, A. Lange, Architecture of the flexible tail tube of bacteriophage SPP1, Nat. Commun. 11 (2020) 5759. https://doi.org/10.1038/s41467-020-19611-1.

[32] B. Chagot, I. Auzat, M. Gallopin, I. Petitpas, B. Gilquin, P. Tavares, S. Zinn-Justin, Solution structure of gp17 from the Siphoviridae bacteriophage SPP1: Insights into its role in virion assembly, Proteins Struct. Funct. Bioinforma. 80 (2012) 319–326. https://doi.org/10.1002/prot.23191.

[33] I. Auzat, I. Petitpas, R. Lurz, F. Weise, P. Tavares, A touch of glue to complete bacteriophage assembly: The tail-to-head joining protein (THJP) family, Mol. Microbiol. 91 (2014) 1164–1178. https://doi.org/10.1111/mmi.12526.

[34] B. Becker, N. De la Fuente, M. Gassel, D. Günther, P. Tavares, R. Lurz, T.A. Trautner, J.C. Alonso, Head morphogenesis genes of the Bacillus subtilis bacteriophage SPP1, J. Mol. Biol. 268 (1997) 822–839. https://doi.org/10.1006/jmbi.1997.0997.

[35] A. Rath, M. Glibowicka, V.G. Nadeau, G. Chen, C.M. Deber, Detergent binding explains anomalous SDS-PAGE migration of membrane proteins, Proc. Natl. Acad. Sci. 106 (2009) 1760–1765. https://doi.org/10.1073/pnas.0813167106.

[36] B. Behrens, G. Lüder, M. Behncke, T.A. Trautner, A.T. Ganesan, The genome of B. subtilis phage SPP1, Mol. Gen. Genet. MGG. 175 (1979) 351–357. https://doi.org/10.1007/BF00397235.

[37] J. Xu, R.W. Hendrix, R.L. Duda, Conserved Translational Frameshift in dsDNA Bacteriophage Tail Assembly Genes, Mol. Cell. 16 (2004) 11–21. https://doi.org/10.1016/j.molcel.2004.09.006.

[38] S. Fernandes, C. São-José, Probing the function of the two holin-like proteins of bacteriophage SPP1, Virology. 500 (2017) 184–189. https://doi.org/10.1016/j.virol.2016.10.030.

[39] L.M. Godinho, M.E.S. Fadel, C. Monniot, L. Jakutyte, I. Auzat, A. Labarde, K. Djacem, L. Oliveira, R. Carballido-Lopez, S. Ayora, P. Tavares, The revisited genome of bacillus subtilis bacteriophage SPP1, Viruses. 10 (2018). https://doi.org/10.3390/v10120705.

[40] D. Veesler, G. Robin, J. Lichière, I. Auzat, P. Tavares, P. Bron, V. Campanacci, C. Cambillau, Crystal Structure of Bacteriophage SPP1 Distal Tail Protein (gp19.1), J. Biol. Chem. 285 (2010) 36666–36673. https://doi.org/10.1074/jbc.M110.157529.

[41] A. Lopes, P. Tavares, M.-A. Petit, R. Guérois, S. Zinn-Justin, Automated classification of tailed bacteriophages according to their neck organization, BMC Genomics. 15 (2014) 1027. https://doi.org/10.1186/1471-2164-15-1027.

[42] I. Katsura, Determination of bacteriophage λ tail length by a protein ruler, Nature. 327 (1987) 73–75. https://doi.org/10.1038/327073a0.

[43] N.K. Abuladze, M. Gingery, J. Tsai, F.A. Eiserling, Tail Length Determination in Bacteriophage T4, Virology. 199 (1994) 301–310. https://doi.org/10.1006/viro.1994.1128.

[44] J. Mahony, M. Alqarni, S. Stockdale, S. Spinelli, M. Feyereisen, C. Cambillau, D. van Sinderen, Functional and structural dissection of the tape measure protein of lactococcal phage TP901-1, Sci. Rep. 6 (2016) 36667. https://doi.org/10.1038/srep36667.

[45] S. Hayes, R. Vincentelli, J. Mahony, A. Nauta, L. Ramond, G.A. Lugli, M. Ventura, D. van Sinderen, C. Cambillau, Functional carbohydrate binding modules identified in evolved dits from siphophages infecting various Gram-positive bacteria, Mol. Microbiol. 110 (2018) 777–795. https://doi.org/10.1111/mmi.14124.

[46] M.-E. Dieterle, S. Spinelli, I. Sadovskaya, M. Piuri, C. Cambillau, Evolved distal tail carbohydrate binding modules of L actobacillus phage J-1: a novel type of anti-receptor widespread among lactic acid bacteria phages, Mol. Microbiol. 104 (2017) 608–620. https://doi.org/10.1111/mmi.13649.

[47] J.L. Kizziah, K.A. Manning, A.D. Dearborn, T. Dokland, Structure of the host cell recognition and penetration machinery of a Staphylococcus aureus bacteriophage, PLOS Pathog. 16 (2020) e1008314. https://doi.org/10.1371/journal.ppat.1008314.

[48] E. Cascales, Microbiology: And Amoebophilus Invented the Machine Gun!, Curr. Biol. 27 (2017) R1170–R1173. https://doi.org/10.1016/j.cub.2017.09.025.

[49] S. Kanamaru, P.G. Leiman, V.A. Kostyuchenko, P.R. Chipman, V. V. Mesyanzhinov, F. Arisaka, M.G. Rossmann, Structure of the cell-puncturing device of bacteriophage T4, Nature. 415 (2002) 553–557. https://doi.org/10.1038/415553a.

[50] R. Hendrix, R. Duda, Bacteriophage lambda PaPa: not the mother of all lambda phages, Science (80-.). 258 (1992) 1145–1148. https://doi.org/10.1126/science.1439823.

[51] Y. Zivanovic, F. Confalonieri, L. Ponchon, R. Lurz, M. Chami, A. Flayhan, M. Renouard, A. Huet, P. Decottignies, A.R. Davidson, C. Breyton, P. Boulanger, Insights into Bacteriophage T5 Structure from Analysis of Its Morphogenesis Genes and Protein Components, J. Virol. 88 (2014) 1162–1174. https://doi.org/10.1128/JVI.02262-13.

[52] I. Katsura, P.W. Kühl, Morphogenesis of the tail of bacteriophage lambda, J. Mol. Biol. 91 (1975) 257–273. https://doi.org/10.1016/0022-2836(75)90379-4.

[53] L.-C. Tsui, R.W. Hendrix, Proteolytic processing of phage λ tail protein gpH: timing of the cleavage, Virology. 125 (1983) 257–264. https://doi.org/10.1016/0042-

997 6822(83)90199-X.

998 [54] J. Xu, R.W. Hendrix, R.L. Duda, A Balanced Ratio of Proteins from Gene G and
999 Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly, J. Mol.
1000 Biol. 425 (2013) 3476–3487. https://doi.org/10.1016/j.jmb.2013.07.002.

1001 [55] J. Xu, R.W. Hendrix, R.L. Duda, Chaperone–Protein Interactions That Mediate
1002 Assembly of the Bacteriophage Lambda Tail to the Correct Length, J. Mol. Biol. 426
1003 (2014) 1004–1018. https://doi.org/10.1016/j.jmb.2013.06.040.

1004 [56] J.A. Lengyel, R.N. Goldstein, M. Marsh, R. Calendar, Structure of the bacteriophage P2
1005 tail, Virology. 62 (1974) 161–174. https://doi.org/10.1016/0042-6822(74)90312-2.

1006 [57] F.J. Grundy, M.M. Howe, Morphogenetic structures present in lysates of amber mutants
1007 of bacteriophage Mu, Virology. 143 (1985) 485–504. https://doi.org/10.1016/0042-
1008 6822(85)90388-5.

1009 [58] Y. Chaban, R. Lurz, S. Brasilès, C. Cornilleau, M. Karreman, S. Zinn-Justin, P. Tavares,
1010 E. V. Orlova, Structural rearrangements in the phage head-to-tail interface during
1011 assembly and infection, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 7009–7014.
1012 https://doi.org/10.1073/pnas.1504039112.

1013 [59] S. Spinelli, D. Veesler, C. Bebeacua, C. Cambillau, Structures and host-adhesion
1014 mechanisms of lactococcal siphophages, Front. Microbiol. 5 (2014).
1015 https://doi.org/10.3389/fmicb.2014.00003.

1016 [60] J.S. Fraser, K.L. Maxwell, A.R. Davidson, Immunoglobulin-like domains on
1017 bacteriophage: weapons of modest damage?, Curr. Opin. Microbiol. 10 (2007) 382–387.
1018 https://doi.org/10.1016/j.mib.2007.05.018.

1019 [61] P. Boulanger, P. Jacquot, L. Plançon, M. Chami, A. Engel, C. Parquet, C. Herbeuval, L.
1020 Letellier, Phage T5 Straight Tail Fiber Is a Multifunctional Protein Acting as a Tape
1021 Measure and Carrying Fusogenic and Muralytic Activities, J. Biol. Chem. 283 (2008)
1022 13556–13564. https://doi.org/10.1074/jbc.M800052200.

1023 [62] M. Piuri, G.F. Hatfull, A peptidoglycan hydrolase motif within the mycobacteriophage
1024 TM4 tape measure protein promotes efficient infection of stationary phase cells, Mol.
1025 Microbiol. 62 (2006) 1569–1585. https://doi.org/10.1111/j.1365-2958.2006.05473.x.

1026 [63] S. Riva, M. Polsinelli, A. Falaschi, A new phage of Bacillus subtilis with infectious
1027 DNA having separable strands, J. Mol. Biol. 35 (1968) 347–356.
1028 https://doi.org/10.1016/S0022-2836(68)80029-4.

1029 [64] S. Okubo, T. Yanagida, Isolation of a suppressor mutant in Bacillus subtilis., J.
1030 Bacteriol. 95 (1968) 1187–1188. https://doi.org/10.1128/JB.95.3.1187-1188.1968.

1031 [65] A. Dröge, P. Tavares, In vitro Packaging of DNA of the Bacillus subtilis bacteriophage
1032 SPP1 1 1Edited by J. Karn, J. Mol. Biol. 296 (2000) 103–115.
1033 https://doi.org/10.1006/jmbi.1999.3449.

1034 [66] R.E. Yasbin, P.I. Fields, B.J. Andersen, Properties of Bacillus subtilis 168 derivatives
1035 freed of their natural prophages, Gene. 12 (1980) 155–159. https://doi.org/10.1016/0378-
1036 1119(80)90026-8.

1037 [67] P. Haima, S. Bron, G. Venema, The effect of restriction on shotgun cloning and plasmid
1038 stability in Bacillus subtilis Marburg, Mol. Gen. Genet. MGG. 209 (1987) 335–342.
1039 https://doi.org/10.1007/BF00329663.

[68] A. Isidro, A.O. Henriques, P. Tavares, The portal protein plays essential roles at different steps of the SPP1 DNA packaging process, Virology. 322 (2004) 253–263. https://doi.org/10.1016/j.virol.2004.02.012.

[69] D. Veesler, S. Blangy, S. Spinelli, P. Tavares, V. Campanacci, C. Cambillau, Crystal structure of Bacillus subtilis SPP1 phage gp22 shares fold similarity with a domain of lactococcal phage p2 RBP, Protein Sci. 19 (2010) 1439–1443. https://doi.org/10.1002/pro.416.

[70] D. Veesler, S. Blangy, J. Lichière, M. Ortiz-Lombardía, P. Tavares, V. Campanacci, C. Cambillau, Crystal structure of Bacillus subtilis SPP1 phage gp23.1, a putative chaperone, Protein Sci. 19 (2010) 1812–1816. https://doi.org/10.1002/pro.464.

[71] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis. 20 (1999) 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

[72] A. Gattiker, W. V. Bienvenut, A. Bairoch, E. Gasteiger, FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification, Proteomics. 2 (2002) 1435–1444. https://doi.org/10.1002/1615-9861(200210)2:10<1435::AID-PROT1435>3.0.CO;2-9.

[73] I. Vinga, A. Dröge, A.C. Stiege, R. Lurz, M.A. Santos, R. Daugelavičius, P. Tavares, The minor capsid protein gp7 of bacteriophage SPP1 is required for efficient infection of Bacillus subtilis, Mol. Microbiol. 61 (2006) 1609–1621. https://doi.org/10.1111/j.1365-2958.2006.05327.x.

[74] A. Isidro, M.A. Santos, A.O. Henriques, P. Tavares, The high-resolution functional map of bacteriophage SPP1 portal protein, Mol. Microbiol. 51 (2004) 949–962. https://doi.org/10.1046/j.1365-2958.2003.03880.x.

[75] E. Harlow, D. Lane, Antibodies: a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1988.

[76] A.C. Steven, B.L. Trus, J.V. Maizel, M. Unser, D.A.D. Parry, J.S. Wall, J.F. Hainfeld, F.W. Studier, Molecular substructure of a viral receptor-recognition protein, J. Mol. Biol. 200 (1988) 351–365. https://doi.org/10.1016/0022-2836(88)90246-X.

[77] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

[78] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data, Bioinformatics. 28 (2012) 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

[79] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539. https://doi.org/10.1038/msb.2011.75.

[80] K. Okonechnikov, O. Golosova, M. Fursov, Unipro UGENE: a unified bioinformatics toolkit, Bioinformatics. 28 (2012) 1166–1167. https://doi.org/10.1093/bioinformatics/bts091.

[81] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A.N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core, J. Mol. Biol. 430 (2018) 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

[82] A. Krogh, B. Larsson, G. von Heijne, E.L.. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen, J. Mol. Biol. 305 (2001) 567–580. https://doi.org/10.1006/jmbi.2000.4315.

[83] B. Behrens, G. Lüder, M. Behncke, T.A. Trautner, A.T. Ganesan, The genome of B. subtilis phage SPP1 - Physical arrangement of phage genes, MGG Mol. Gen. Genet. 175 (1979) 351–357. https://doi.org/10.1007/BF00397235.

Figure Legends

Figure 1 Purification of SPP1 phage tails and identification of their protein components. (A) Purification of SPP1 tails from *B. subtilis* YB886 infected with SPP1*gp13*$^{-}$, using differential precipitation, anion exchange chromatography (AEC) and zonal centrifugation. The lanes of the Coomassie-stained SDS-PAGE gels are aligned with the corresponding fractions in the AEC chromatograms. For pellet P2, AEC fractions pooled for further purification by zonal centrifugation are shaded. Purified tails peak in P2 sucrose gradient fraction 6 (P2f6S) and in P3 AEC fraction 8 (P2f8), respectively (both shaded). (B) Electron micrographs of negatively stained phage tails from fractions P2f6S and P3f8. Morphological elements of the tail and average measurements of their length and diameter. (n=20) are depicted in the scheme on the right. (C) Assignment of bands detected on a Coomassie gel (15 % (w/v) SDS-PAGE) of a P3f8 sample, based on mass spectrometry (MS) and N-terminal (N-ter) sequencing (Table S1) as labeled. The portal protein gp6 and the major capsid protein gp13, present only in the phage particle control, are identified based on their apparent molecular mass. (D) Coomassie-stained SDS-PAGE gel (15 % (w/v)) of 10-fold concentrated samples P2f6S and P3f8 and western blots of the same samples, re-diluted 2-fold (P2f6S) and 6-fold (P3f8), respectively. Disrupted phage particles are run as controls. Note that anti-gp17.1N detects both major tail proteins, gp17.1 ($M_{r,\ app}$ ~21 kDa) and gp17.1* (~29 kDa) [29] while anti-gp18N recognizes three different species of gp18.

Figure 2 Immunolabeling of gp22 in P3f8 tails. Overview (A) and detail (B) of P3f8 tails cross-linked by purified anti-gp22 IgG through their tip region. The cartoons are visual aids to identify the position of IgG relative to the tail structure whose elements are color-coded as in Figure 3. Red lines in (A) show the relative position of the tail fiber and tail tube central symmetry axis to highlight the free-tilting joint between the cap and the fiber. Immuno-labelling required pre-fixation of the P3f8 tails with glutaraldehyde. (C) Labelling of the anti-gp22 IgG position with goat IgG anti-rabbit IgG complexed with colloidal gold.

Figure 3 Schematic representation of the SPP1 virion's structural organization and of the genetic map of the two SPP1 tail gene operons. The region between nucleotides 7803 and 20825 of the SPP1 genome (NCBI accession: X97918.3 [39]) corresponds to the 3' portion of the gene *11* operon (promotor *P*L4) and to the operon controlled by promoter *P*L5 (figure bottom). It comprises open reading frames *13* to *24.1*. Coding regions for structural proteins previously identified [19,33] or characterized in this work are depicted as colored arrows. Genes coding for unassigned (gp14, gp23, gp23.1, and gp24) and non-structural proteins (gp17.5 and gp17.5*) are represented in grey. The arrangement of protein components within the SPP1 virion established in this and in previous work [19,29,33,58] is schematized on top of the figure. Proteins are represented in the same colors as the corresponding coding genes below. Their abbreviation name according to function is displayed on top: THJP – Tail to Head Joining Protein, TCP – Tail Completion Protein; TTP – Tail Tube Protein; TMP – Tape Measure Protein; Dit – Distal Tail protein; TAP – Tail Adaptor Protein; RBP – Receptor Binding Protein. The tail tube is depicted cut-open to show the TMP gp18 inside the tail tube central channel. Gp22 is rendered semi-transparent to indicate its transient association (Figure 1D) to the tail fiber region identified in Figure 2. Conditional lethal mutants that disrupt individual protein function investigated in this study are shown in the figure middle labelled with their original names [83]. They were renamed in the manuscript text according to the gene they knock-out for simplicity. SPP1*sus31* (named here SPP1*gp13*$^{-}$) [34] and SPP1*sus45* (named SPP1*gp17.1*$^{-}$) [29] were previously described while all other mutants were mapped by complementation and sequenced in this study (Table S2). Designations of suppressor sensitive (*sus*) mutants are shown in blue-framed boxes, featuring also the position of the stop codon preventing the production of the respective protein. Temperature sensitive (*ts*) mutants and their mutation sites are framed in green. The black bars at the bottom of the figure identify the coding region of each tail gene that was expressed to purify proteins for raising polyclonal rabbit antibodies. The lowercase "d" indicates that denatured polypeptides were used for immunization. Anti-gp17.1 [29] and anti-gp21 [4] antibodies were previously described.

Figure 4 Production and stability of SPP1 tail proteins in infected cells. (A-C) Production of SPP1 proteins in cells infected with SPP1 wild type (kinetics in (A)), suppressor sensitive (B) and temperature sensitive (C) mutants defective in tail genes (Figure 3).

Infections of the non-permissive host *B. subtilis* YB886 were carried out at 37°C (A,B) and 46°C (C). Bacteria infected with mutants defective in the proteins shown above the gel lanes were harvested at the time indicated. Tail proteins in cell extracts were detected by western blot. The three gp18-specific bands detectable in SPP1*wt* infections are marked by arrowheads and the gp18 band observed only after infection with SPP1*gp17.1*$^{-}$ (B,C) and SPP1*ts* mutants (C) with diamonds to distinguish them from nonspecific background. The signal of the capsid portal protein gp6 was used to compare the level of overall phage protein production in infections with different mutants. * - the capsids run in this lane carry a gp6 form with a hexahistidine tag at its Cter leading to its lower electrophoretic mobility when compared to wild type gp6.

Figure 5 Function of tail chaperones gp17.5 and gp17.5* on the TMP gp18 stability. The experimental conditions and symbols in the western blots are as in Figure 4.

Figure 6 Identification of SPP1 tail protein complexes from bacteria infected with supressor sensitive and temperature sensitive mutants (Figure 3). (A) Structures partially purified from lysates of the non-permissive strain *B. subtilis* YB886 infected with SPP1*sus* mutant phages defective in the production of the major capsid protein gp13, TTPs gp17.1/gp17.1*, tail assembly chaperones gp17.5/gp17.5*, and the TMP gp18. Structures concentrated through a sucrose cushion were applied to glycerol gradients (10 % to 30 % (w/v)) and spun at 35000 rpm in a SW41 rotor (Figure S5), for 7 h in the case of SPP1*gp13*$^{-}$ infection and 12 h for the others. (B) Structures produced in infections a 46°C by SPP1*ts* mutants defective in the production of gp19.1 and gp21. Gradients of concentrated structures were spun in parallel with a SPP1*gp18*$^{-}$ infection sample and a mixture of proteins with known sedimentation coefficients (bottom). The sedimentation conditions were as above except that the run was for 15 h. Tail proteins were detected by western blot. Polyclonal antibodies raised against SPP1 particles, which give a strong reaction with the Cter of gp17.1* [29], were used to detect the two SPP1 major tail proteins. Calibration proteins were detected by Coomassie staining (bottom).

Figure 7 Function of gp16.1 in tail assembly. (A) Western blot of extracts of the non-permissive B subtilis strain YB886 infected with SPP1*gp13*$^{-}$ and SPP1*gp16.1*$^{-}$. Tail proteins gp16.1, gp17 and gp19.1 were detected with specific antibodies. (B) Composition of tails produced in non-permissive infections with SPP1*gp13*$^{-}$ (tails) and SPP1*gp16.1*$^{-}$ (fraction 4 of the gradient in Figure S6). Western blots were developed with antibodies against tail structural proteins and the capsid portal protein gp6 (control). Phage particles and capsids are used as controls. (C) EM of negatively stained purified tails. White arrows indicate a narrow ring on the SPP1*gp13*$^{-}$ tail region that binds to tails. This ring is absent in SPP1*gp16.1*$^{-}$ tails.

Figure 8 Assembly pathway model for the SPP1 phage tail. The protein complexes formed sequentially and mutants blocking individual assembly steps are depicted according to the data of this study. Brackets indicate a putative labile complex of gp18 bound to the cap gp19.1-gp21 complex. "?" indicates gp22 and gp17.5* putative interaction steps during assembly. See text for details.

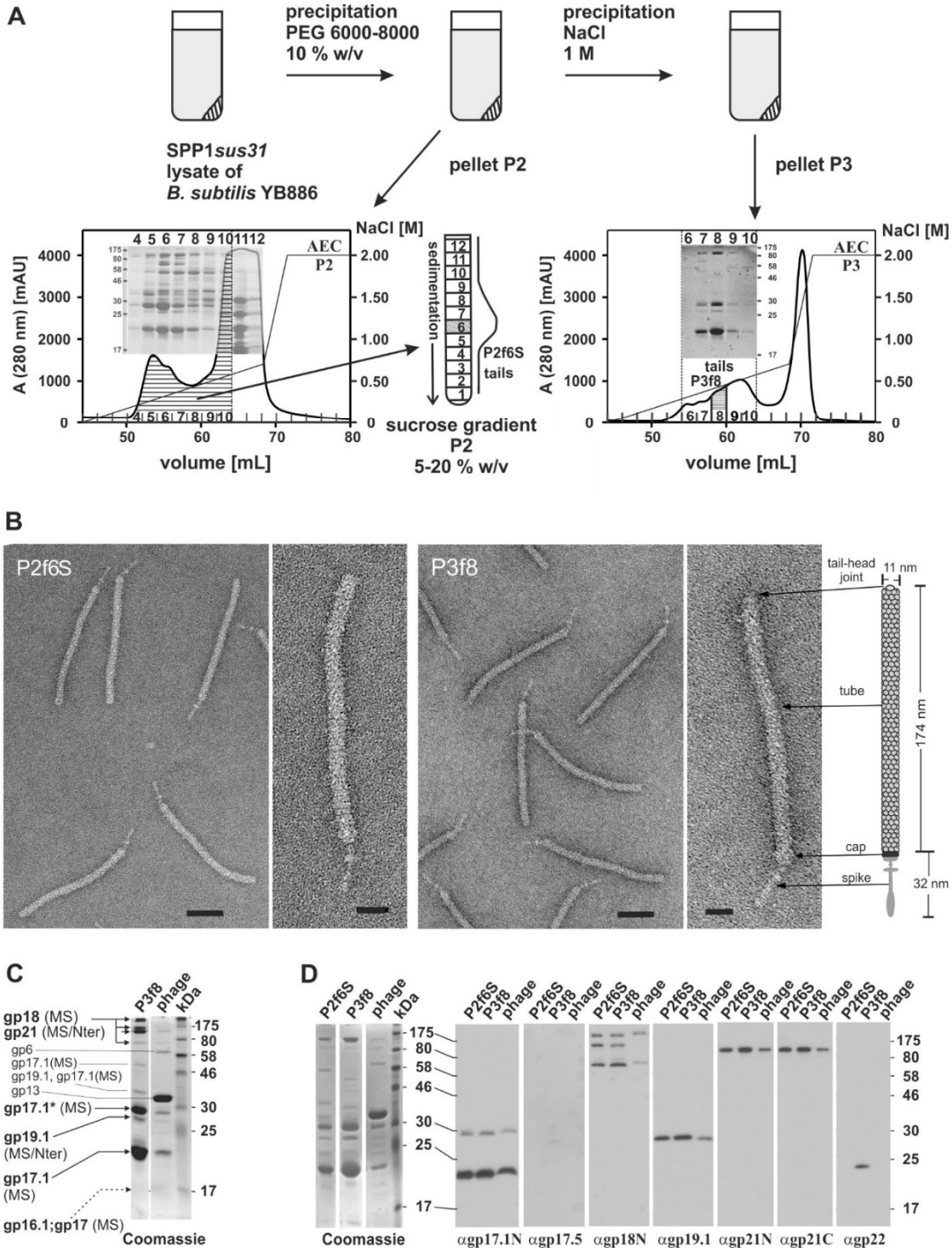
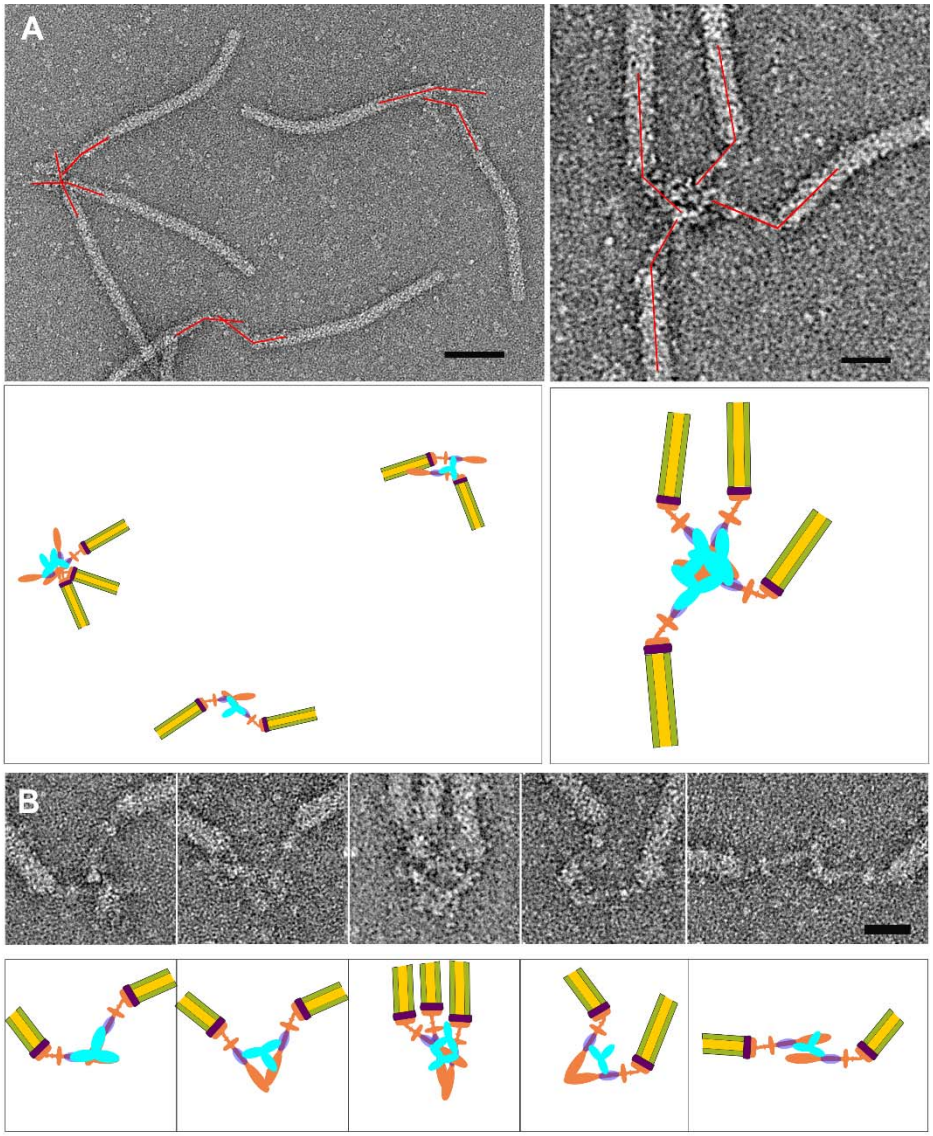


Figure 1 Purification of SPP1 phage tails and identification of their protein components. (A) Purification of SPP1 tails from *B. subtilis* YB886 infected with SPP1*gp13*$^{-}$, using differential precipitation, anion exchange chromatography (AEC) and zonal centrifugation. The lanes of the Coomassie-stained SDS-PAGE gels are aligned with the corresponding fractions in the AEC chromatograms. For pellet P2, AEC fractions pooled for further purification by zonal centrifugation are shaded. Purified tails peak in P2 sucrose gradient fraction 6 (P2f6S) and in P3 AEC fraction 8 (P2f8), respectively (both shaded). (B) Electron micrographs of negatively stained phage tails from fractions P2f6S and P3f8. Morphological elements of the tail and average measurements of their length and diameter. (n=20) are depicted in the scheme on the right. (C) Assignment of bands detected on a Coomassie gel (15 % (w/v) SDS-PAGE) of a P3f8 sample, based on mass spectrometry (MS) and N-terminal (N-ter) sequencing (Table S1) as labeled. The portal protein gp6 and the major capsid protein gp13, present only in the phage particle control, are identified based on their apparent molecular mass. (D) Coomassie-stained SDS-PAGE gel (15 % (w/v)) of 10-fold concentrated samples P2f6S and P3f8 and western blots of the same samples, re-diluted 2-fold (P2f6S) and 6-fold (P3f8), respectively. Disrupted phage particles are run as controls. Note that anti-gp17.1N detects both major tail proteins, gp17.1 ($M_{r,\,app}$ ~21 kDa) and gp17.1* (~29 kDa) [29] while anti-gp18N recognizes three different species of gp18.



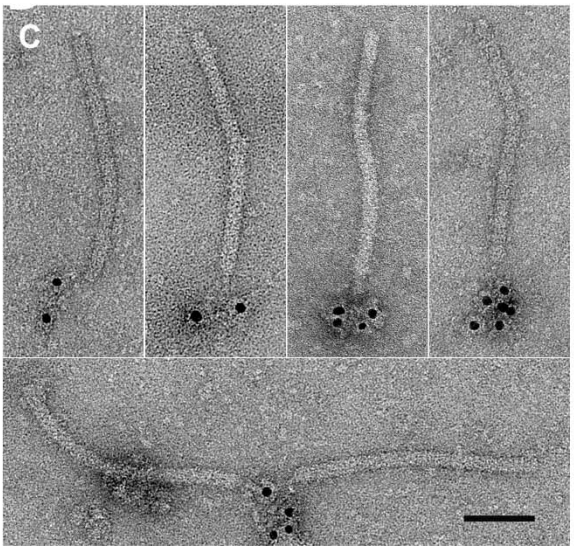


Figure 2 Immunolabeling of gp22 in P3f8 tails. Overview (A) and detail (B) of P3f8 tails cross-linked by purified anti-gp22 IgG through their tip region. The cartoons are visual aids to identify the position of IgG relative to the tail structure whose elements are color-coded as in Figure 3. Red lines in (A) show the relative position of the tail fiber and tail tube central symmetry axis to highlight the free-tilting joint between the cap and the fiber. Immuno-labelling required pre-fixation of the P3f8 tails with glutaraldehyde. (C) Labelling of the anti-gp22 IgG position with goat IgG anti-rabbit IgG complexed with colloidal gold.

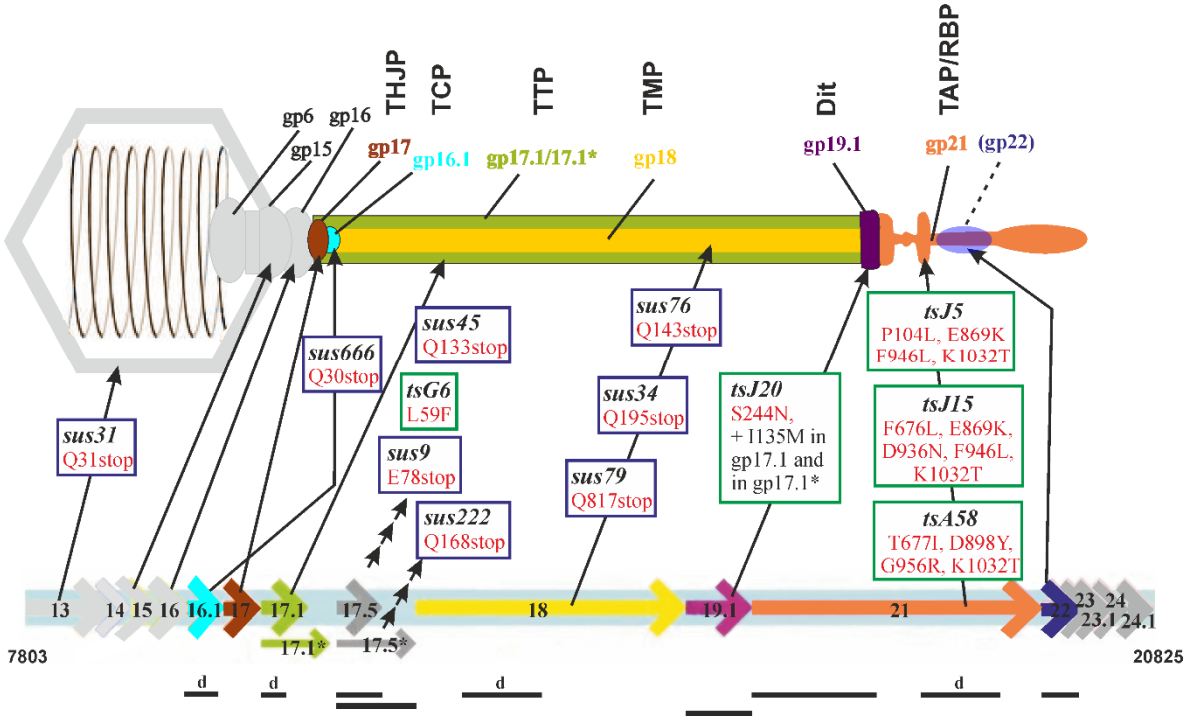


Figure 3 Schematic representation of the SPP1 virion's structural organization and of the genetic map of the two SPP1 tail gene operons. The region between nucleotides 7803 and 20825 of the SPP1 genome (NCBI accession: X97918.3 [39]) corresponds to the 3' portion of the gene *11* operon (promotor *P*L4) and to the operon controlled by promoter *P*L5 (figure bottom). It comprises open reading frames *13* to *24.1*. Coding regions for structural proteins previously identified [19,33] or characterized in this work are depicted as colored arrows. Genes coding for unassigned (gp14, gp23, gp23.1, and gp24) and non-structural proteins (gp17.5 and gp17.5*) are represented in grey. The arrangement of protein components within the SPP1 virion established in this and in previous work [19,29,33,58] is schematized on top of the figure. Proteins are represented in the same colors as the corresponding coding genes below. Their abbreviation name according to function is displayed on top: THJP – Tail to Head Joining Protein, TCP – Tail Completion Protein; TTP – Tail Tube Protein; TMP – Tape Measure Protein; Dit – Distal Tail protein; TAP – Tail Adaptor Protein; RBP – Receptor Binding Protein. The tail tube is depicted cut-open to show the TMP gp18 inside the tail tube central channel. Gp22 is rendered semi-transparent to indicate its transient association (Figure 1D) to the tail fiber region identified in Figure 2. Conditional lethal mutants that disrupt individual protein function investigated in this study are shown in the figure middle labelled with their original names [83]. They were renamed in the manuscript text according to the gene they knock-out for simplicity. SPP1*sus31* (named here SPP1*gp13*$^{-}$) [34] and SPP1*sus45* (named SPP1*gp17.1*$^{-}$) [29] were previously described while all other mutants were mapped by complementation and sequenced in this study (Table S2). Designations of suppressor sensitive (*sus*) mutants are shown in blue-framed boxes, featuring also the position of the stop codon preventing the production of the respective protein. Temperature sensitive (*ts*) mutants and their mutation sites are framed in green. The black bars at the bottom of the figure identify the coding region of each tail gene that was expressed to purify proteins for raising polyclonal rabbit antibodies. The lowercase "d" indicates that denatured polypeptides were used for immunization. Anti-gp17.1 [29] and anti-gp21 [4] antibodies were previously described.

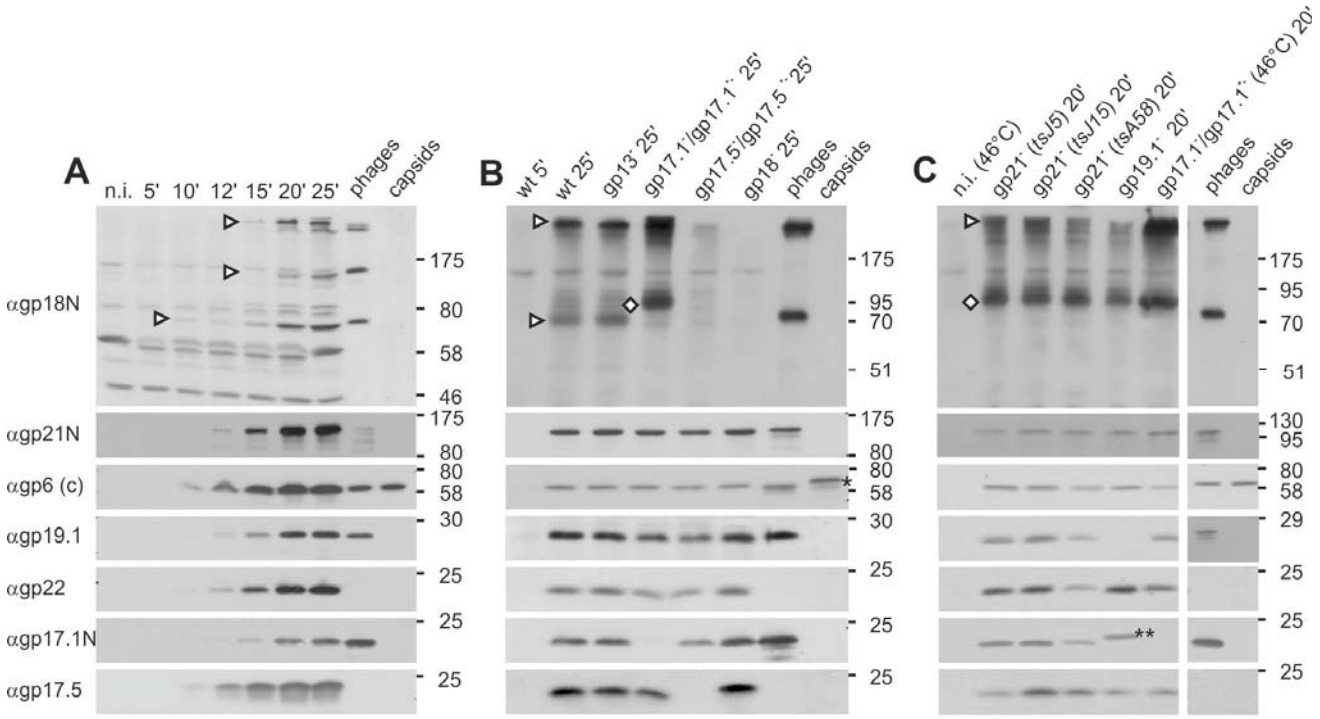


Figure 4 Production and stability of SPP1 tail proteins in infected cells. (A-C) Production of SPP1 proteins in cells infected with SPP1 wild type (kinetics in (A)), suppressor sensitive (B) and temperature sensitive (C) mutants defective in tail genes (Figure 3). Infections of the non-permissive host *B. subtilis* YB886 were carried out at 37°C (A,B) and 46°C (C). Bacteria infected with mutants defective in the proteins shown above the gel lanes were harvested at the time indicated. Tail proteins in cell extracts were detected by western blot. The three gp18-specific bands detectable in SPP1*wt* infections are marked by arrowheads and the gp18 band observed only after infection with SPP1*gp17.1*⁻ (B,C) and SPP1*ts* mutants (C) with diamonds to distinguish them from nonspecific background. The signal of the capsid portal protein gp6 was used to compare the level of overall phage protein production in infections with different mutants. * - the capsids run in this lane carry a gp6 form with a hexahistidine tag at its Cter leading to its lower electrophoretic mobility when compared to wild type gp6.

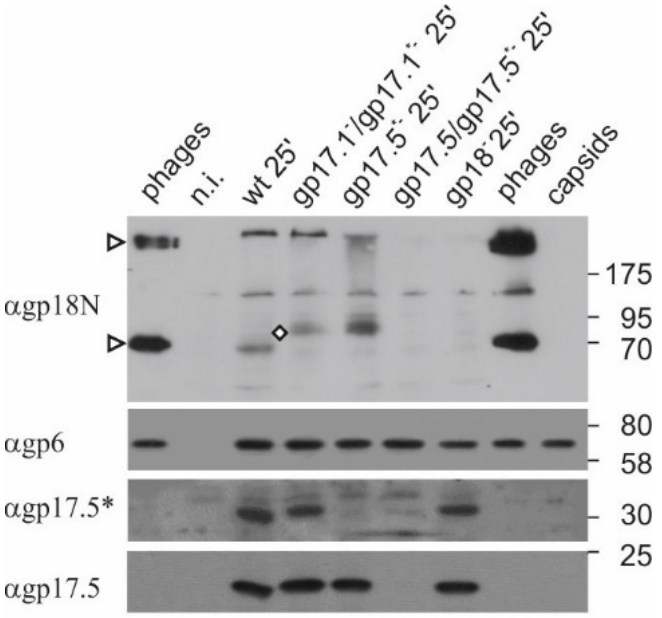


Figure 5 Function of tail chaperones gp17.5 and gp17.5* on the TMP gp18 stability. The experimental conditions and symbols in the western blots are as in Figure 4.

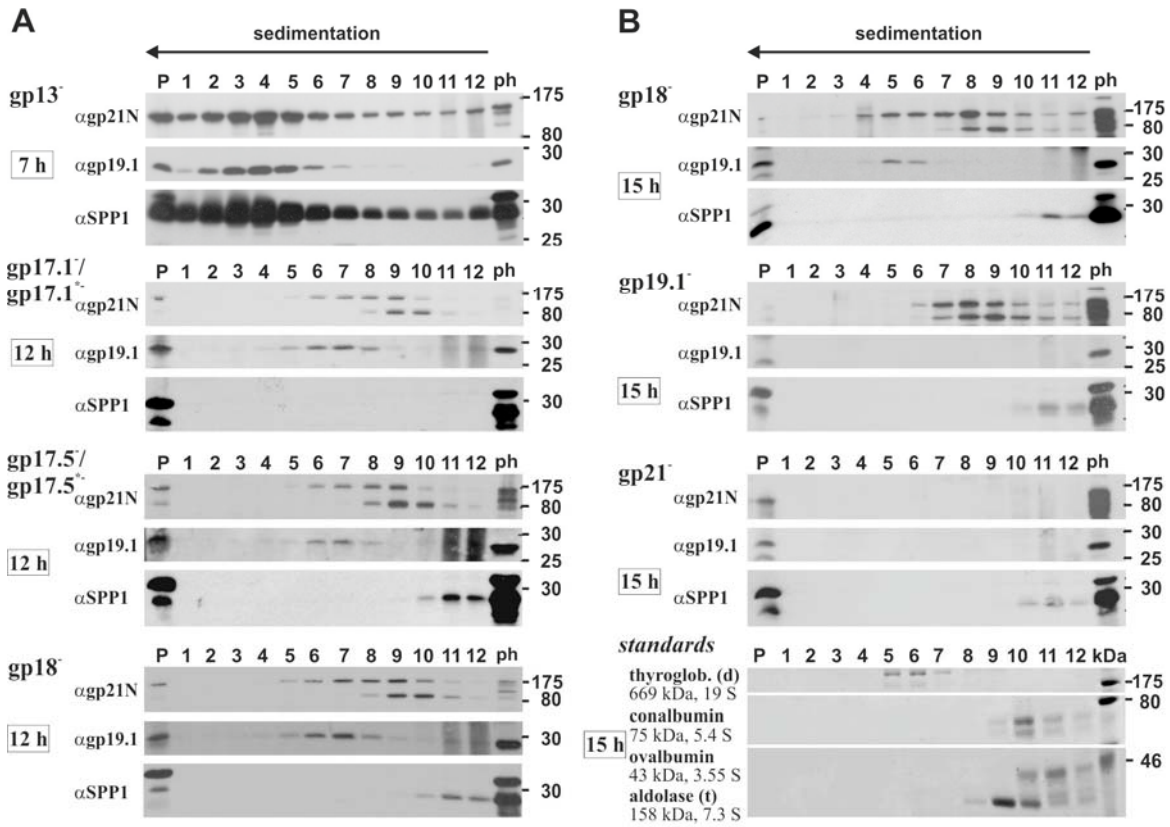


Figure 6 Identification of SPP1 tail protein complexes from bacteria infected with supressor sensitive and temperature sensitive mutants (Figure 3). (A) Structures partially purified from lysates of the non-permissive strain *B. subtilis* YB886 infected with SPP1*sus* mutant phages defective in the production of the major capsid protein gp13, TTPs gp17.1/gp17.1*, tail assembly chaperones gp17.5/gp17.5*, and the TMP gp18. Structures concentrated through a sucrose cushion were applied to glycerol gradients (10 % to 30 % (w/v)) and spun at 35000 rpm in a SW41 rotor (Figure S5), for 7 h in the case of SPP1*gp13*- infection and 12 h for the others. (B) Structures produced in infections a 46°C by SPP1*ts* mutants defective in the production of gp19.1 and gp21. Gradients of concentrated structures were spun in parallel with a SPP1*gp18*- infection sample and a mixture of proteins with known sedimentation coefficients (bottom). The sedimentation conditions were as above except that the run was for 15 h. Tail proteins were detected by western blot. Polyclonal antibodies raised against SPP1 particles, which give a strong reaction with the Cter of gp17.1* [29], were used to detect the two SPP1 major tail proteins. Calibration proteins were detected by Coomassie staining (bottom).

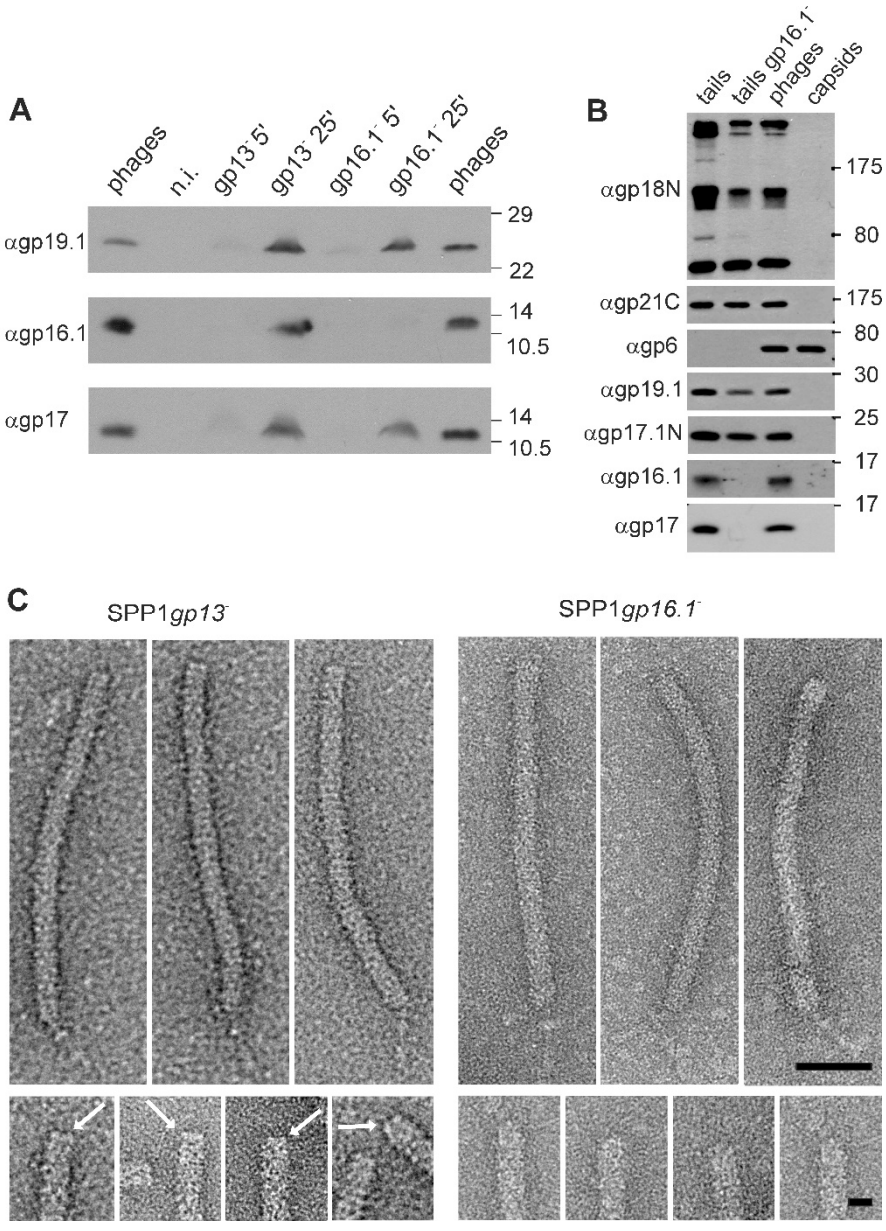


Figure 7 Function of gp16.1 in tail assembly. (A) Western blot of extracts of the non-permissive B subtilis strain YB886 infected with SPP1*gp13*⁻ and SPP1*gp16.1*⁻. Tail proteins gp16.1, gp17 and gp19.1 were detected with specific antibodies. (B) Composition of tails produced in non-permissive infections with SPP1*gp13*⁻ (tails) and SPP1*gp16.1*⁻ (fraction 4 of the gradient in Figure S6). Western blots were developed with antibodies against tail structural proteins and the capsid portal protein gp6 (control). Phage particles and capsids are used as controls. (C) EM of negatively stained purified tails. White arrows indicate a narrow ring on the SPP1*gp13*⁻ tail region that binds to tails. This ring is absent in SPP1*gp16.1*⁻ tails.

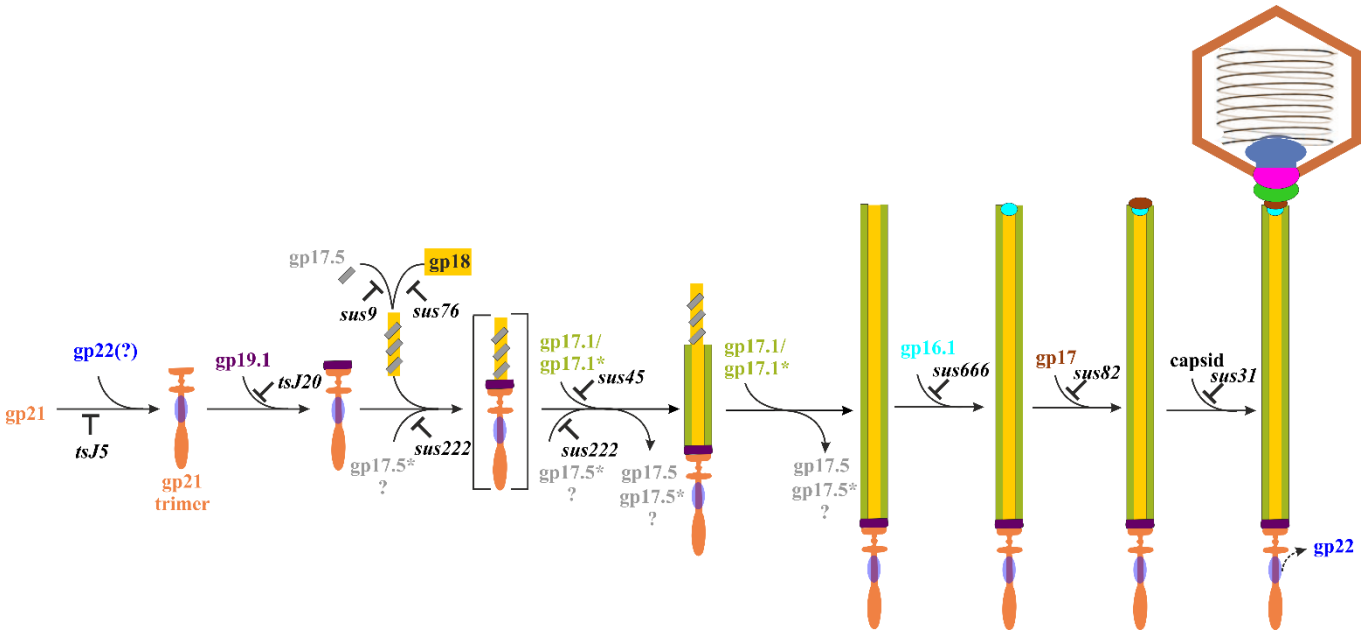


Figure 8 Assembly pathway model for the SPP1 phage tail. The protein complexes formed sequentially and mutants blocking individual assembly steps are depicted according to the data of this study. Brackets indicate a putative labile complex of gp18 bound to the cap gp19.1-gp21 complex. “?” indicates gp22 and gp17.5* putative interaction steps during assembly. See text for details.

Table S1 Fingerprint of SPP1 proteins studied in this work

protein	$M_{r,\,th}$ [kDa][a]	$M_{r,\,app}$ [kDa] [b]	Number of amino acids	Sequence coverage [c] set 1 (MALDI)	set 2a (LC-MS/MS)	set 2b (MALDI)	N-terminal sequence [d]	pI_{calc} [e]	protein detection[f] and references to previous work[g]
gp16.1	15.9	17	141	---	19-105 (36 %)	5-105 (40 %)	---	9.18	tail component (this work)
gp17	15.0	17	134	---	13-121 (59 %)	13-121 (59 %)	---	5.34	tail component [1,2]
gp17.1	19.2	21	177	---	1-177 (54 %)	---	---	4.31	tail component [3–5]
gp17.1*	28.2	29	264	---	1-209 (47 %)	---	---	4.36	tail component [3,4]
gp17.5	20.2	18	173	---	---	---	---	5.13	nonstructural protein (this work) [3,4]
gp17.5*	34.0	30	297	---	---	---	---	5.45	nonstructural protein (this work) [3,4]
gp18	110.9	>175 145 73	1032	6-916 (47 %) --- 6-1010 (44 %)	6-1010 (39 %) 6-902 (48 %) 6-902 (34 %)	6-909 (49 %) 6-909 (41 %) 21-909 (14 %)	--- --- ---	10.45	tail component; the gp18 species enriched in extracts of, e.g., *sus45* infected cells migrates at about 83 kDa (Fig. 4B,C) (this work) [3]
gp19.1	28.6	27	253	1-251 (80 %)	1-251 (55%)	---	MNIYDI	4.72	tail component (this work) [3,6,7]
gp21	123.6	138	1108	65-1107 (55 %)	8-1094 (53 %)	65-956 (47 %)	SKNIWIM	5.06	tail component (this work) [3,7,8]
gp22	16.7	23	146	1-146 (75 %)	---	---	---	4.32	tail component (this work) [9]
gp23	6.1	---	54	---	---	---	---	5.32	not detected in infected cells (this work)
gp23.1	5.8	---	51	---	---	---	---	3.73	not detected in infected cells (this work) [10]
gp24	8.4	---	76	---	---	---	---	4.41	not detected in infected cells (this work)

[a]molecular mass calculated from the protein amino acid sequence using the ExPASy ProtParam tool
[b]molecular mass estimated from the migration behaviour in SDS-PA gels. For the gp18 species and gp21 the estimation was made from 10 % w/v Coomassie stained gels (or, more accurately, from western blots) and for other proteins from 15 % w/v Coomassie gel

[c]Mass spectrometry of SPP1 tail structural proteins. “set 1”: set of 5 bands cut from two SDS-PA gels (15 % w/v and 20 % w/v) and analysed by peptide mapping using MALDI-TOF. “set 2”: set of 13 bands cut from two SDS-PA gels (15 % w/v both) and analysed by peptide mapping using (a) LC-MS/MS (all bands analysed) and (b) MALDI-TOF (only selected bands from set 2).
[d]Nter sequence determined by Edman degradation.
[e]isoelectric points estimated from the sequences using the ExPASy ProtParam tool.
[f]western blot detection of tail proteins in SPP1 phage particles, tails and in extracts of infected cells.
[g]references:

[1] B. Chagot, I. Auzat, M. Gallopin, I. Petitpas, B. Gilquin, P. Tavares, S. Zinn-Justin, Solution structure of gp17 from the Siphoviridae bacteriophage SPP1: Insights into its role in virion assembly, Proteins Struct. Funct. Bioinforma. 80 (2012) 319–326. https://doi.org/10.1002/prot.23191.

[2] I. Auzat, I. Petitpas, R. Lurz, F. Weise, P. Tavares, A touch of glue to complete bacteriophage assembly: The tail-to-head joining protein (THJP) family, Mol. Microbiol. 91 (2014) 1164–1178. https://doi.org/10.1111/mmi.12526.

[3] C. Plisson, H.E. White, I. Auzat, A. Zafarani, C. São-José, S. Lhuillier, P. Tavares, E. V. Orlova, Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection, EMBO J. 26 (2007) 3720–3728. https://doi.org/10.1038/sj.emboj.7601786.

[4] I. Auzat, A. Dröge, F. Weise, R. Lurz, P. Tavares, Origin and function of the two major tail proteins of bacteriophage SPP1, Mol. Microbiol. 70 (2008) 557–569. https://doi.org/10.1111/j.1365-2958.2008.06435.x.

[5] C. Langlois, S. Ramboarina, A. Cukkemane, I. Auzat, B. Chagot, B. Gilquin, A. Ignatiou, I. Petitpas, E. Kasotakis, M. Paternostre, H.E. White, E. V. Orlova, M. Baldus, P. Tavares, S. Zinn-Justin, Bacteriophage SPP1 tail tube protein self-assembles into β-structure-rich tubes, J. Biol. Chem. 290 (2015) 3836–3849. https://doi.org/10.1074/jbc.M114.613166.

[6] D. Veesler, G. Robin, J. Lichière, I. Auzat, P. Tavares, P. Bron, V. Campanacci, C. Cambillau, Crystal Structure of Bacteriophage SPP1 Distal Tail Protein (gp19.1), J. Biol. Chem. 285 (2010) 36666–36673. https://doi.org/10.1074/jbc.M110.157529.

[7] A. Goulet, J. Lai-Kee-Him, D. Veesler, I. Auzat, G. Robin, D.A. Shepherd, A.E. Ashcroft, E. Richard, J. Lichière, P. Tavares, C. Cambillau, P. Bron, The Opening of the SPP1 Bacteriophage Tail, a Prevalent Mechanism in Gram-positive-infecting Siphophages, J. Biol. Chem. 286 (2011) 25397–25405. https://doi.org/10.1074/jbc.M111.243360.

[8] I. Vinga, C. Baptista, I. Auzat, I. Petipas, R. Lurz, P. Tavares, M.A. Santos, C. São-José, Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection, Mol. Microbiol. 83 (2012) 289–303. https://doi.org/10.1111/j.1365-2958.2011.07931.x.

[9] D. Veesler, S. Blangy, S. Spinelli, P. Tavares, V. Campanacci, C. Cambillau, Crystal structure of Bacillus subtilis SPP1 phage gp22 shares fold similarity with a domain of lactococcal phage p2 RBP, Protein Sci. 19 (2010) 1439–1443. https://doi.org/10.1002/pro.416.

[10] D. Veesler, S. Blangy, J. Lichière, M. Ortiz-Lombardía, P. Tavares, V. Campanacci, C. Cambillau, Crystal structure of Bacillus subtilis SPP1 phage gp23.1, a putative chaperone, Protein Sci. 19 (2010) 1812–1816. https://doi.org/10.1002/pro.464.

Table S2 Protein coding genes studied in this work and their conditional lethal mutants. See Figure 3 for a graphical representation.

gene product[a]	**coding region**[b]	***sus* mutants**[c]	**nucleotide substitution (position)**	**stop codon**	***ts* mutants**[c]	**nucleotide substitution**	**amino acid substitution**	**reference (mapping / sequencing)** [e]
gp13	7803-8777	***sus31***	CAG→TAG (7893)	Q31*stop*	---			[1]
gp16.1	9644-10069	***sus666***	CAA→TAA (7922)	Q30*stop*	---			this work
gp17	10066-10470	*sus82*	CAA→TAA (19399)	Q111*stop*	---			[2]
gp17.1	10484-11017	***sus45***	CAA→TAA (10880)	Q133*stop*	---			[3]
gp17.1*	10484-11279	***sus45***	CAA→TAA (10880)	Q133*stop*	---			[3]
gp17.5	11363-11884	***sus9***	GAG→TAG (11594)	E78*stop*	*tsG6*	CTT→TTT (11537)	L59F	this work
gp17.5*	11363-12255	***sus9*** ***sus222***	GAG→TAG (11594) CAA→TAA (11863)	E78*stop* Q168*stop*	*tsG6*	CTT→TTT (11537)	L59F	this work
gp18	12267-15365	***sus76*** *sus34* *sus79*	CAA→TAA (12693) CAA→ TAA (12849) CAA→ TAA (14715)	Q143*stop* Q195stop Q817*stop*	---			this work
gp19.1	15362-16123	---			***tsJ20***	AGT→AAT, (16092) *(and ATT→ ATG (10889))* [d]	S244N (*and I135M in gp17.1 and gp 17.1**)* [d]	this work

gp21	16137-19463	---	***tsJ5***	CCT→CTT (16447), GAA→AAA (18741), TTC→TTA (18974), AAA→ACA (19231)	P104L, E869K, F946L, K1032T	this work
			tsJ15	TTT→CTT (18162), GAA→AAA (18741), GAC→AAC (18942), TTC→TTA (18974), AAA→ACA (19231)	F676L, E869K, D936N, F946L, K1032T	
			tsA5	ACA→ATA (18166), GAT→TAT (18828), GGG→AGG (19002), AAA→ACA (19231)	T677I, D898Y, G956R, K1032T	
gp22	19476-19916	---	---			
gp23	19932-20096	---	---			
gp23.1	20089-20244	---	---			
gp24	20237-20467	---	---			

[a]gene products shown to be either SPP1 structural, tail proteins or otherwise involved in tail morphogenesis are displayed in bold
[b]coordinates in the SPP1 nucleotide sequence, NCBI access code X97918.3 [4]
[c]SPP1*sus* and *ts* mutants are from our collection [5]. Mutants in bold were functionally characterized in this study.
[d]SPP1*tsJ20* carries a *ts* mutation in gene *19.1*, as assessed by complementation assays, and a missense mutation in genes *17.1*/*17.1** that leads to amino acid change I135M in gp17.1/gp17.1*
[g]references:

[1] B. Becker, N. De la Fuente, M. Gassel, D. Günther, P. Tavares, R. Lurz, T.A. Trautner, J.C. Alonso, Head morphogenesis genes of the Bacillus subtilis bacteriophage SPP1, J. Mol. Biol. 268 (1997) 822–839. https://doi.org/10.1006/jmbi.1997.0997.

[2] I. Auzat, I. Petitpas, R. Lurz, F. Weise, P. Tavares, A touch of glue to complete bacteriophage assembly: The tail-to-head joining protein (THJP) family, Mol. Microbiol. 91 (2014) 1164–1178. https://doi.org/10.1111/mmi.12526.

[3] I. Auzat, A. Dröge, F. Weise, R. Lurz, P. Tavares, Origin and function of the two major tail proteins of bacteriophage SPP1, Mol. Microbiol. 70 (2008) 557–569. https://doi.org/10.1111/j.1365-2958.2008.06435.x.

[4] L.M. Godinho, M.E.S. Fadel, C. Monniot, L. Jakutyte, I. Auzat, A. Labarde, K. Djacem, L. Oliveira, R. Carballido-Lopez, S. Ayora, P. Tavares, The revisited genome of bacillus subtilis bacteriophage SPP1, Viruses. 10 (2018). https://doi.org/10.3390/v10120705.

[5] B. Behrens, G. Lüder, M. Behncke, T.A. Trautner, A.T. Ganesan, The genome of B. subtilis phage SPP1 - Physical arrangement of phage genes, MGG Mol. Gen. Genet. 175 (1979) 351–357. https://doi.org/10.1007/BF00397235.

Table S3 Primers used in this work

Primer	Sequence[a]	Restriction sites
60102BZ	GG*GAATTC***AAAGGAG**AA*GGATCC***ATG**TCGGTTAGAATCGACCC	EcoRI, BamHI
60102AZ	CC*GGTACCCTGCAG*CATGTGTTTTACCCCTCTC	KpnI, PstI
17.5CS1	CCTCCA*GGATCC*GTC*ATCGAT*GAGTGTAATAAAAATGAGAGGGG	BamHI, ClaI
23NCS1	CCGGTACC*CTGCAG*CCGGGTTGATAATTTCCGG	PstI
gene 17.5 CS	CGG*CCCGGG***AAAGGAG**AA*GGATCC***ATG**AGAGGGGTAAAACACATG	SmaI, BamHI
gene 17.5* NCS	GAA*CTGCAG***CTA**TTTAGGTGTGAATTTTAGTCC	PstI
prset1	CGG*CCCGGG***AAAGGAG**AAATA*CATATG*CGGGGTTCTC	SmaI, NdeI
prset2	ATAC*AAGCTT*CGATTCCATGGTACC	HindIII
gene 17.5 to 17.5* CS	GCATTGGATAACTCAGGTTT**C**TTCCAAACGAGCTGTGCG	
gene 17.5 to 17.5* NCS	CGCACAGCTCGTTTGGAA**G**AAACCTGAGTTATCCAATGC	
gene 17.5 only CS 2mut	GCATTGGATAACTCAGGATTCTTCAAACGAGCTGTGCG	
gene 17.5 only NCS 2mut	CGCACAGCTCGTTTGAAGAATCCTGAGTTATCCAATGC	
gene 18Cter CS	GAA*GGATCC*GACAAAATAAAAGAATATACGG	BamH1
gene 18Cter NCS	GAA*CTGCAG***CTA**TCGTCTTTTCCTCTTCC	PstI

gene 18 181 CS	GAA*GGATCC*GAGGACTTCAAAAGCAATTGG	BamHI
gene 18 347 CS	GAA*GGATCC*GCGGATATGATGACATCTTTAG	BamHI
gene 18 435 NCS	TGAA*CTCGAGCTGCAG***CTA**CTTACCGATTTCCGGGTGC	XhoI, PstI
gene 18 919 NCS	TGAA*CTCGAGCTGCAG***CTA**CTTGATTGTTTTCGGAATGGC	XhoI, PstI
F-gp19.1	ATG*CTGCAG*AATATATATGACATTCTG	PstI
R-gp19.1	TCA*ACCGGT*TAAAAACTTTTCTGTAAA	AgeI
F-gp24	ATG*CTGCAG*ACTAAGCTAACTGAAAAT	PstI
R-gp24	TCA*ACCGGT*CTCCGGTTGTGGTTCTCC	AgeI

[a]Primers sequence identical to SPP1 is underlined. Point mutations are non-underlined within SPP1 sequences and a nucleotide insertion is in red, non-underlined. Translation ribosome binding sites (complementary to *B. subtilis* 16 *S* rRNA 3'), start and stop codons are in bold type. Restriction sites are in italics and listed in the column on the right.