Figure S1
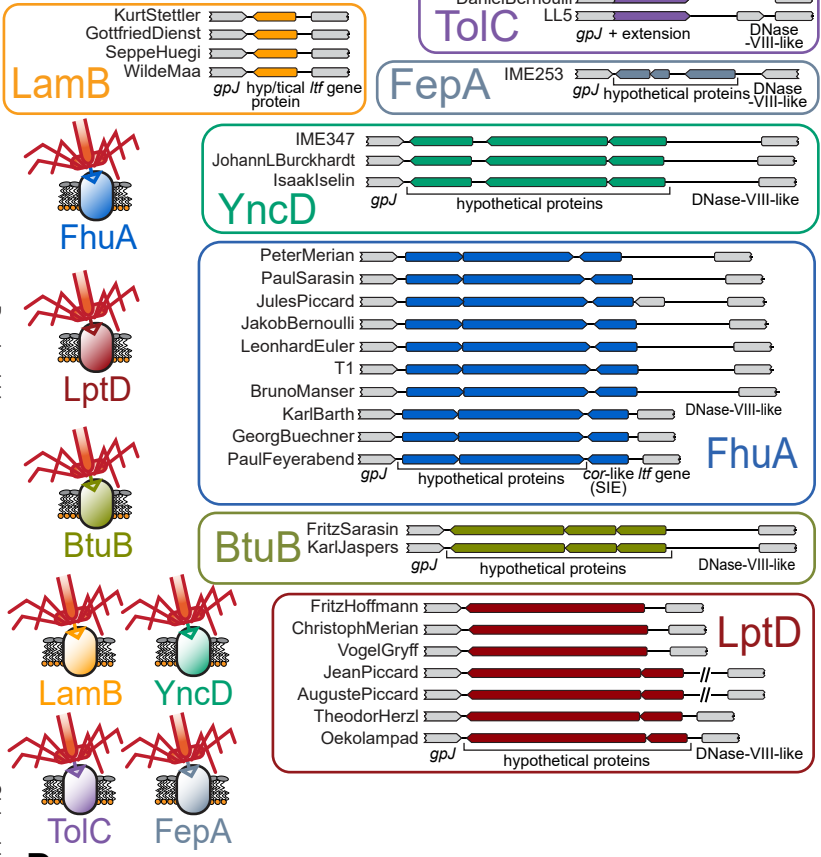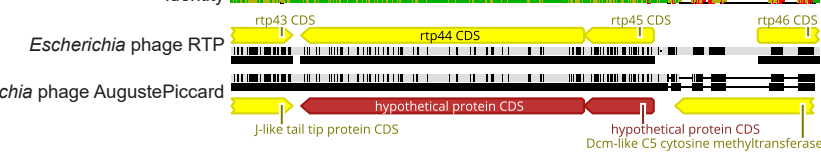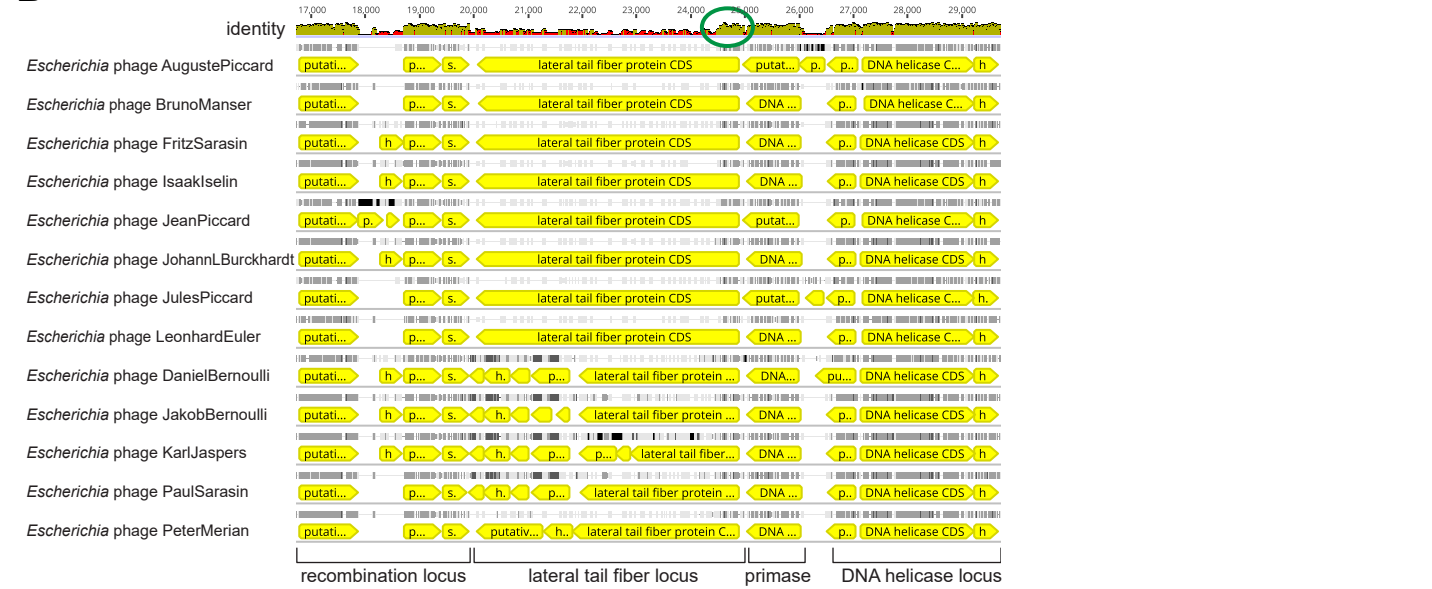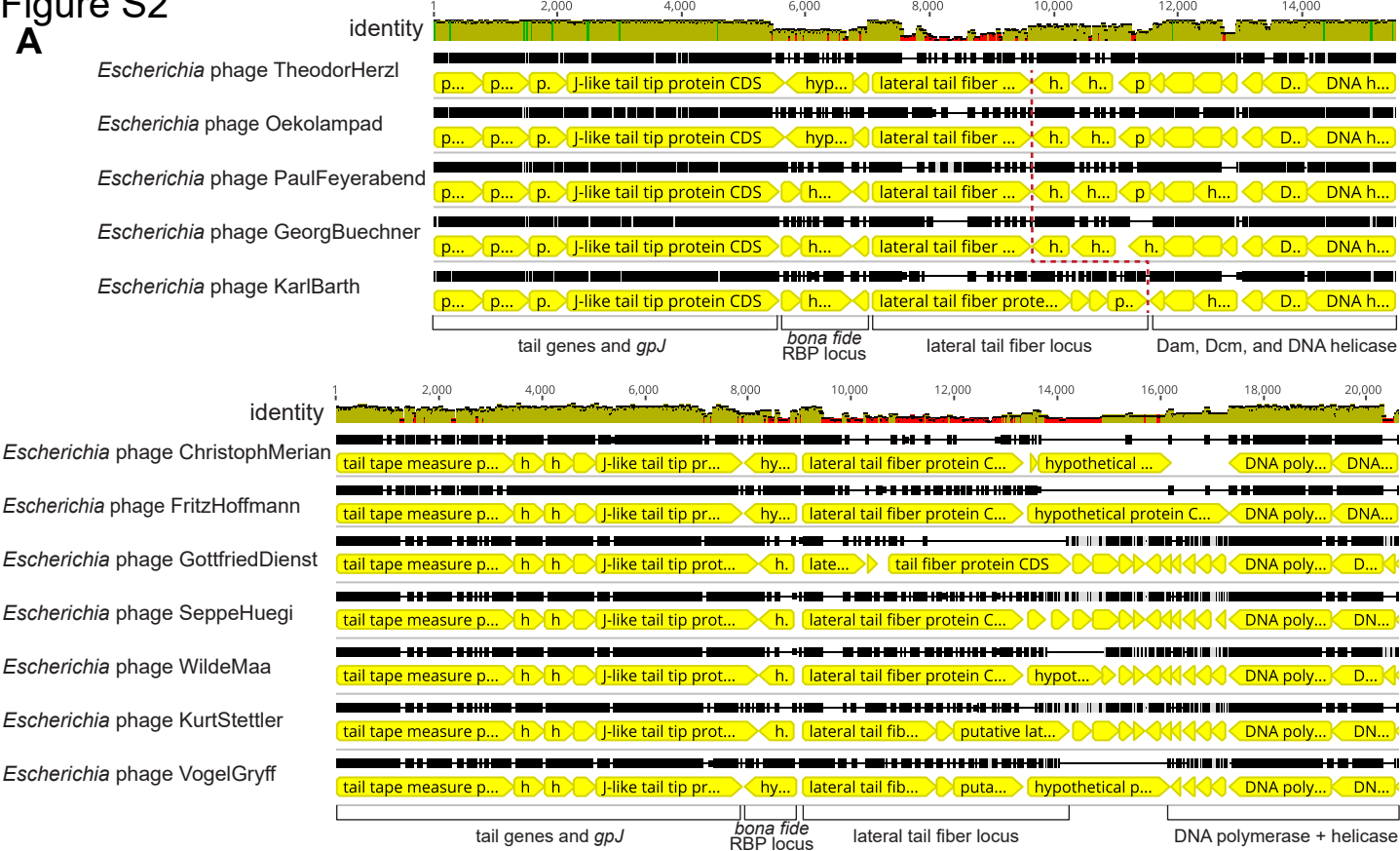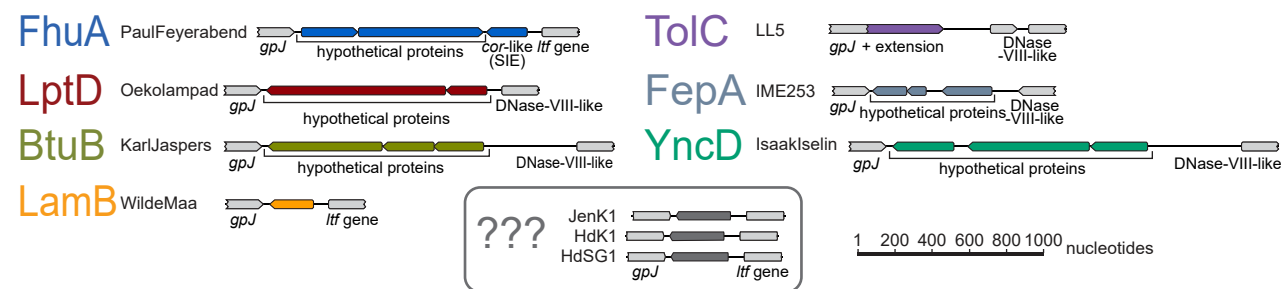
S1 Figure. Supplemental data for Figs 4 and 5.

(A) Maximum-Likelihood phylogeny of *Drexlerviridae* and the *Dhillonvirus* genus of *Siphoviridae* based on several core genes with bootstrap support of branches shown if > 70/100. It is clearly apparent that *Drexlerviridae* are split into two major clades, one formed by *Braunvirinae* and *Rogunavirinae* and another one formed by *Tempevirinae*, *Tunavirinae*, plus a few other groups. Given that the phylogenies strongly agree on all major branches, the root of the *Drexlerviridae* phylogeny shown in Fig 4B was placed between these two major clades. **(B)** The locus encoding lateral tail fibers was analyzed in a sequence alignment of the thirteen *Drexlerviridae* phage genomes of the BASEL collection (see *Materials and Methods*). It is clearly visible that the upstream and downstream regions (encoding genes involved in recombination as well as primase / helicase proteins for genome replication) are highly conserved and fully syntenic, with exception of small insertions in a few sequences. Conversely, only the most 5' end of the largest lateral tail fiber protein gene is very similar among all analyzed genomes (green circle), while the rest shows neither synteny nor clear homology across all genomes. **(C)** The *bona fide* RBP loci downstream of the *gpJ* homolog are shown for all small siphoviruses (*Drexlerviridae* and *Siphoviridae* of *Dhillonvirus*, *Nonagvirus*, and *Seuratvirus* genera) where we had experimentally determined the terminal receptor (together with selected representatives where previous work had determined the receptor specificity). **(D)** The *bona fide* RBP locus of *E. coli* phage RTP was aligned to the homologous locus of phage AugustePiccard (Bas01) as described in *Materials and Methods*. For the region comprising *rtp44* and *rtp45* of phage RTP, the pairwise identity of the two nucleotide sequences is ca. 93%.
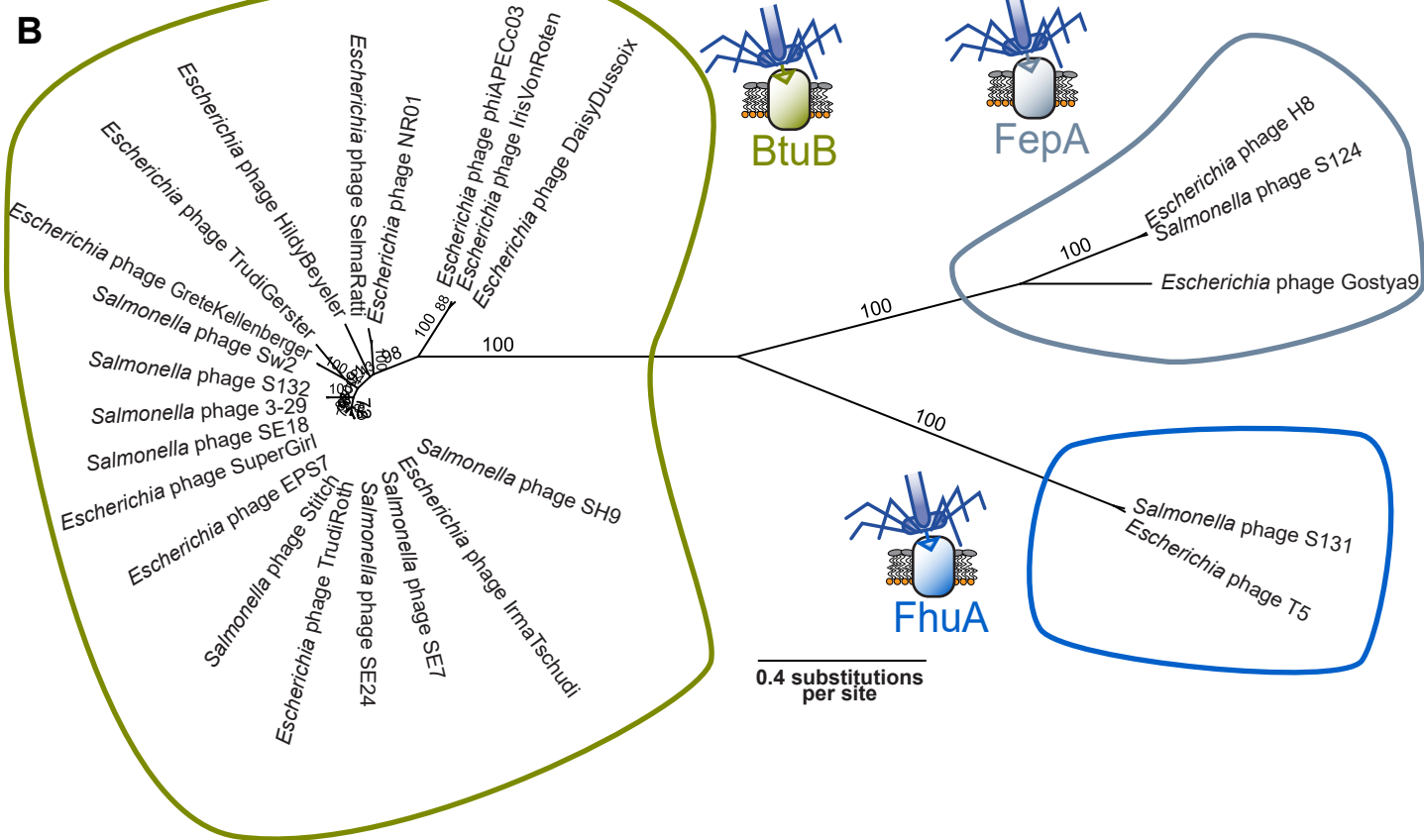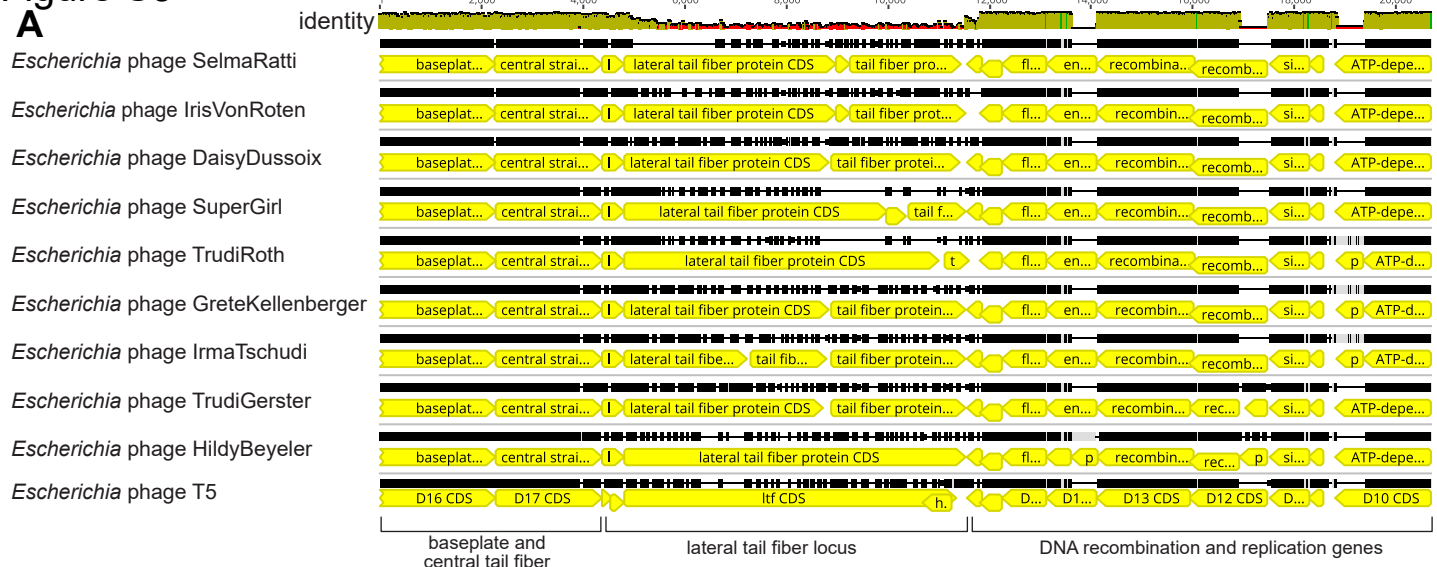
Figure S2

A

identity					
Escherichia phage TheodorHerzl	p... p... p.	J-like tail tip protein CDS	hyp...	lateral tail fiber ...	h. h.. p	D.. DNA h...
Escherichia phage Oekolampad	p... p... p.	J-like tail tip protein CDS	hyp...	lateral tail fiber ...	h. h.. p	D.. DNA h...
Escherichia phage PaulFeyerabend	p... p... p.	J-like tail tip protein CDS	h...	lateral tail fiber ...	h. h... p h...	D.. DNA h...
Escherichia phage GeorgBuechner	p... p... p.	J-like tail tip protein CDS	h...	lateral tail fiber ...	h. h.. h.	D.. DNA h...
Escherichia phage KarlBarth	p... p... p.	J-like tail tip protein CDS	h...	lateral tail fiber prote... p..	h... D.. DNA h...

tail genes and *gpJ* — *bona fide* RBP locus — lateral tail fiber locus — Dam, Dcm, and DNA helicase

identity

Escherichia phage ChristophMerian	tail tape measure p... h h	J-like tail tip pr...	hy...	lateral tail fiber protein C...	hypothetical ...	DNA poly... DNA...
Escherichia phage FritzHoffmann	tail tape measure p... h h	J-like tail tip pr...	hy...	lateral tail fiber protein C...	hypothetical protein C...	DNA poly... DNA...
Escherichia phage GottfriedDienst	tail tape measure p... h h	J-like tail tip prot...	h.	late...	tail fiber protein CDS	DNA poly... D...
Escherichia phage SeppeHuegi	tail tape measure p... h h	J-like tail tip prot...	h.	lateral tail fiber protein C...		DNA poly... DN...
Escherichia phage WildeMaa	tail tape measure p... h h	J-like tail tip prot...	h.	lateral tail fiber protein C...	hypot...	DNA poly... D...
Escherichia phage KurtStettler	tail tape measure p... h h	J-like tail tip prot...	h.	lateral tail fib... putative lat...		DNA poly... DN...
Escherichia phage VogelGryff	tail tape measure p... h h	J-like tail tip pr...	hy...	lateral tail fib... puta... hypothetical p...		DNA poly... DN...

tail genes and *gpJ* — *bona fide* RBP locus — lateral tail fiber locus — DNA polymerase + helicase

B

FhuA PaulFeyerabend
gpJ — hypothetical proteins — *cor*-like *ltf* gene (SIE)

LptD Oekolampad
gpJ — hypothetical proteins — DNase-VIII-like

BtuB KarlJaspers
gpJ — hypothetical proteins — DNase-VIII-like

LamB WildeMaa
gpJ — *ltf* gene

TolC LL5
gpJ + extension — DNase-VIII-like

FepA IME253
gpJ — hypothetical proteins — DNase-VIII-like

YncD IsaakIselin
gpJ — hypothetical proteins — DNase-VIII-like

???
JenK1
HdK1
HdSG1
gpJ — *ltf* gene

1 200 400 600 800 1000 nucleotides

S2 Figure. Supplemental data for Fig 6.

(A) The locus encoding lateral tail fibers was analyzed in sequence alignments of the five *Dhillonvirus* phage genomes of the BASEL collection (top) and the seven *Nonagvirus + Seuratvirus* phage genomes of the BASEL collection (bottom) as described in *Materials and Methods*. In both cases two clear dips in overall sequence similarity are obvious, once at the *bona fide* RBP locus and then at the lateral tail fiber locus downstream of the far 5' end of its first gene. **(B)** Schematic comparison of representative *bona fide* RBP loci as shown in S1C Fig to the corresponding allele of *E. coli* phages JenK1, HdK1, and HdsG1 that does not match clearly match any of them.
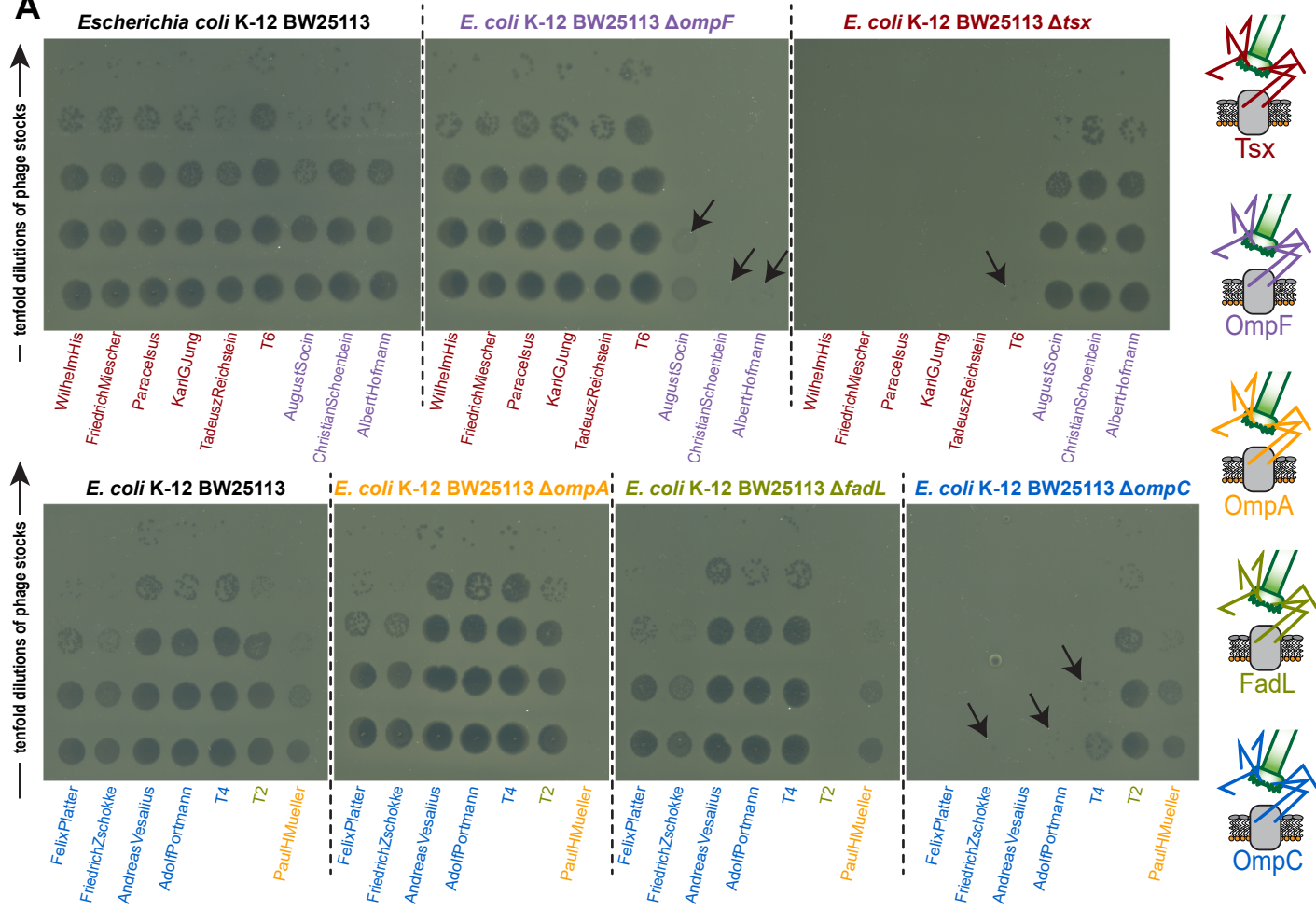
Figure S3

A

identity

Escherichia phage SelmaRatti
Escherichia phage IrisVonRoten
Escherichia phage DaisyDussoix
Escherichia phage SuperGirl
Escherichia phage TrudiRoth
Escherichia phage GreteKellenberger
Escherichia phage IrmaTschudi
Escherichia phage TrudiGerster
Escherichia phage HildyBeyeler
Escherichia phage T5

baseplate and central tail fiber

lateral tail fiber locus

DNA recombination and replication genes

B

BtuB   FepA

FhuA

S3 Figure. Supplemental data for Fig 7.

(A) The locus encoding lateral tail fibers was analyzed in a sequence alignment of the *Demerecviridae: Markadamsvirinae* phage genomes of the BASEL collection as described in *Materials and Methods*. Sequence identity is high upstream and downstream of the lateral tail fiber locus (with exception of presence / absence of a few putative homing endonucleases) but drops considerably at the lateral tail fiber genes. Note that, as described previously, the lateral tail fibers can either be composed of a single large polypeptide or by two (or more) separate proteins [51, 60]. The same diversity in architecture of lateral tail fibers can also be seen at the co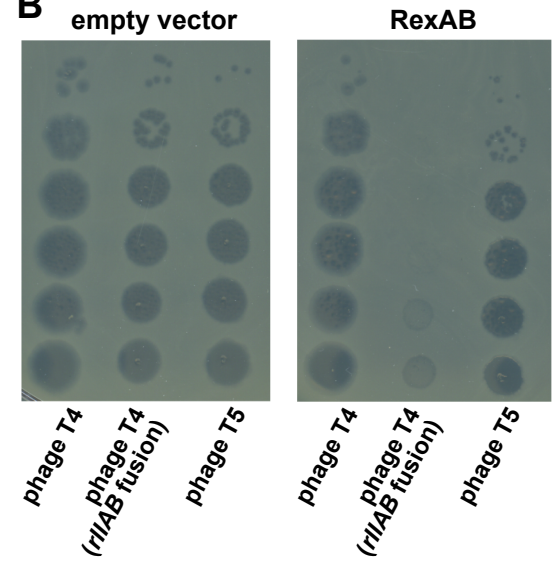rresponding loci of small siphoviruses (S1B and S2A Figs). **(B)** The illustration shows a phylogeny of the RBPs of all *Markadamsvirinae* phages shown in Fig 7B. Briefly, the RBP genes of all genomes (invariably encoded directly upstream of the terminase genes) were translated, aligned, and then used to generate a phylogeny as described in *Materials and Methods*. Three clear clusters emerge, one including all phages known to bind BtuB (left), one including all phages known to bind FepA (top right), and one including all phages known to bind FhuA (bottom right). Based on similar analyses by others [62], we conclude that the position of RBPs in these clusters is predictive of terminal receptor specificity of the phages encoding them.
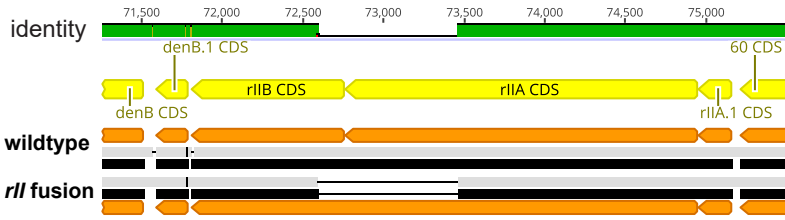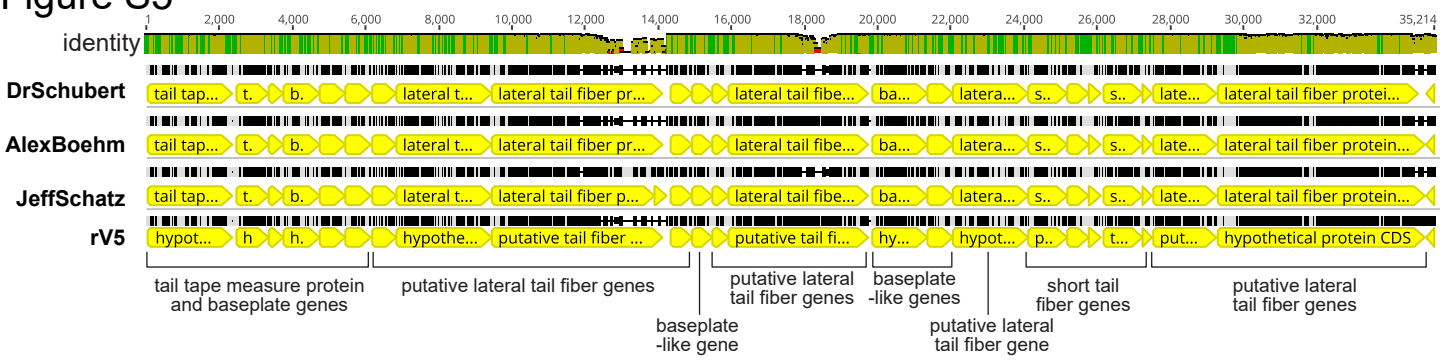
Figure S4

A

Escherichia coli K-12 BW25113

WilhelmHis, FriedrichMiescher, Paracelsus, KarlGJung, TadeuszReichstein, T6, AugustSocin, ChristianSchoenbein, AlbertHofmann

E. coli K-12 BW25113 ΔompF

WilhelmHis, FriedrichMiescher, Paracelsus, KarlGJung, TadeuszReichstein, T6, AugustSocin, ChristianSchoenbein, AlbertHofmann

E. coli K-12 BW25113 Δtsx

WilhelmHis, FriedrichMiescher, Paracelsus, KarlGJung, TadeuszReichstein, T6, AugustSocin, ChristianSchoenbein, AlbertHofmann

tenfold dilutions of phage stocks

Tsx, OmpF, OmpA, FadL, OmpC

E. coli K-12 BW25113

FelixPlatter, FriedrichZzschokke, AndreasVesalius, AdolfPortmann, T4, T2, PaulHMueller

E. coli K-12 BW25113 ΔompA

FelixPlatter, FriedrichZzschokke, AndreasVesalius, AdolfPortmann, T4, T2, PaulHMueller

E. coli K-12 BW25113 ΔfadL

FelixPlatter, FriedrichZzschokke, AndreasVesalius, AdolfPortmann, T4, T2, PaulHMueller

E. coli K-12 BW25113 ΔompC

FelixPlatter, FriedrichZzschokke, AndreasVesalius, AdolfPortmann, T4, T2, PaulHMueller

B

empty vector | RexAB

phage T4, phage T4 (rIIAB fusion), phage T5

C

identity

denB.1 CDS, denB CDS, rIIB CDS, rIIA CDS, rIIA.1 CDS, 60 CDS

71,500 72,000 72,500 73,000 73,500 74,000 74,500 75,000

T4 wildtype

T4 rII fusion

S4 Figure. Supplemental data for Fig 8.

(A) Top agar assays with different surface protein mutants of the KEIO collection in comparison to the ancestral *E. coli* K-12 BW25113 strain were performed with serial tenfold dilutions of all *Tevenvirinae* phages used in this study (undiluted high-titer stocks at the bottom and increasingly diluted samples towards the top). The phages show strongly or totally abolished growth on each one of the mutants which identifies the primary receptor of each phage (also indicated by the color code highlighted on the right). **(B)** Top agar assays of reference strain *E. coli* K-12 ΔRM carrying empty vector pBR322_ΔP*tet* or pAH213_rexAB were performed with serial tenfold dilutions of phage T4 wildtype, a T4 variant encoding an apparently hypomorphic *rIIAB* fusion (see (C)), and phage T5 (as control). The *rIIAB* mutant of phage T4 is unable to form plaques on the *rexAB*-expressing host and shows only "lysis from without" [98], while the T4 wildtype and phage T5 are not affected. **(D)** A T4 phage mutant was erroneously obtained from a culture collection instead of the wildtype and encoded a peculiar *rII* allele with fusion of the *rIIA* and *rIIB* open reading frames (as revealed by whole-genome sequencing). Since such a mutant seems unlikely to arise spontaneously during shipping, we find it likely that this phage strain is related to the *rIIAB* fusion mutants employed for discovery of the triplet nature of the genetic code that were once commonly used [concisely reviewed in reference 149, 150]. Notably, position and size of the deletion fusing *rIIA* and *rIIB* are indistinguishable from the sketch drawn by Benzer and Champe for the *rIIAB* fusion mutant *r1589* which was used in the aforementioned work [151]. Unlike T4 wildtype, the *rIIAB* fusion mutant is susceptible to *rexAB* when ectopically expressed from a plasmid vector (see (C)) and therefore validates functionality of the *rexAB* construct.

Figure S5

S5 Figure. Supplemental data for Fig 9.

The illustration shows a sequence alignment of the locus encoding lateral tail fiber genes in phage rV5 and new isolates DrSchubert, AlexBoehm, and JeffSchatz that broadly cover the phylogenetic range of this genus (Fig 9B). It extends from the tail tape measure protein of phage rV5 (*gp49*) to the last large lateral tail fiber gene (*gp27*) [70, 71]. Note that most genes are highly conserved including the lateral tail fiber component with sugar deacetylase domain (compare Fig 9D; around position 8'000 in this alignment) or the short tail fiber locus (compare Fig 9E). Only the three largest lateral tail fiber genes show considerable allelic variation that is very strong for two of them (positions 10'000-14'000 and 16'000-19'000) and moderate for another one (positions 29'000 to 34'000).