

Cell Reports, Volume 42

Supplemental information

Genetic determinants of host tropism in *Klebsiella* phages

Beatriz Beamud, Neris García-González, Mar Gómez-Ortega, Fernando González-Candelas, Pilar Domingo-Calap, and Rafael Sanjuan

Tree scale: 10

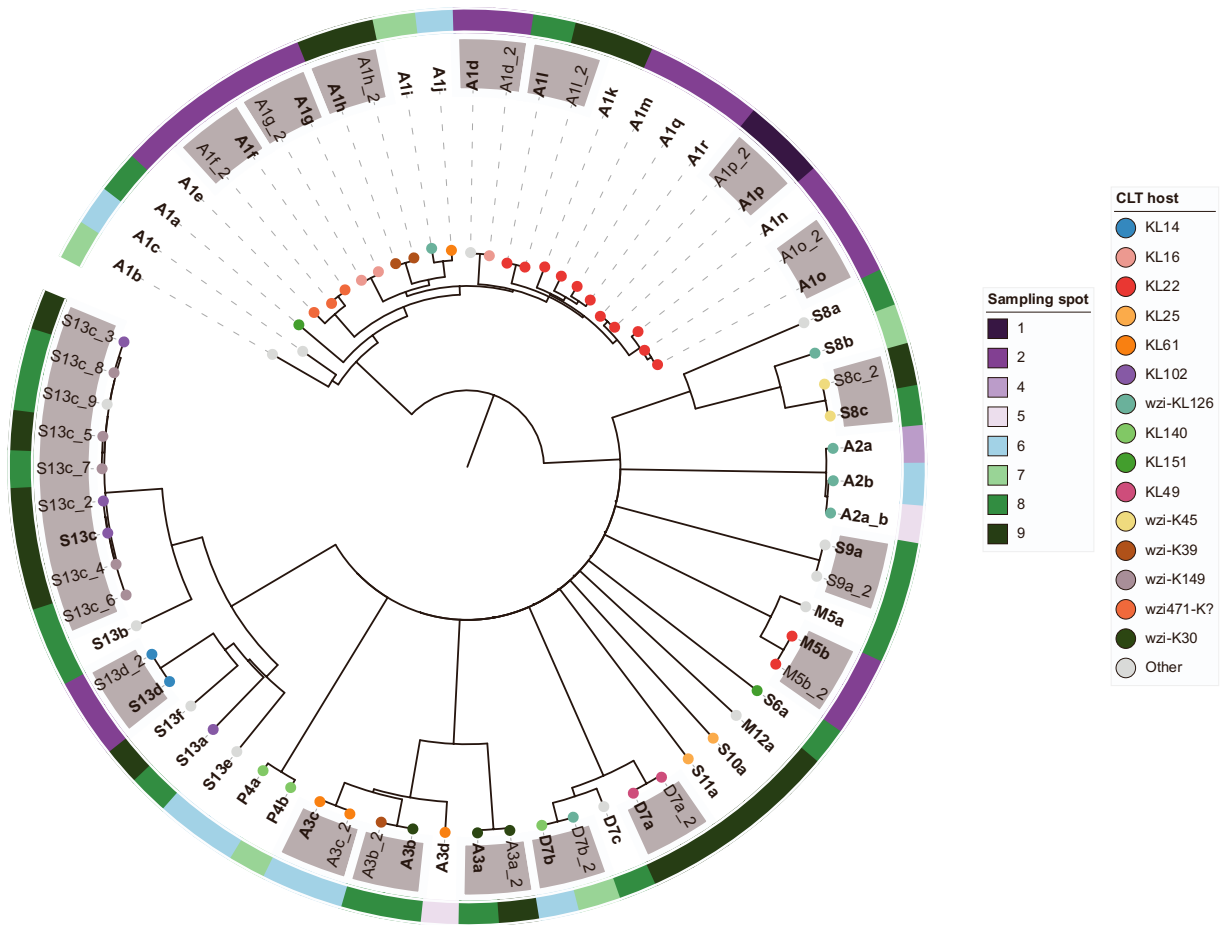


Figure S1. Dendrogram of the 70 phage isolates. A neighbor-joining tree was obtained from pairwise average nucleotide identity (ANI) values. Phages that were considered redundant are marked with gray boxes, and the selected representative of each phage strain is indicated in bold. This resulted in 46 distinct phages that were selected for further study. The capsular locus type (CLT) of the *Klebsiella* host used for phage isolation is indicated with colored circles at the tips of the dendrogram. CLTs with one occurrence are shown as 'Other'. ? indicates *wzi* alleles whose correspondence with a CLT is unknown. The different sampling spots from which phages were isolated are indicated with colors outside the dendrogram. Tree scale denotes ANI distance (100-ANI).

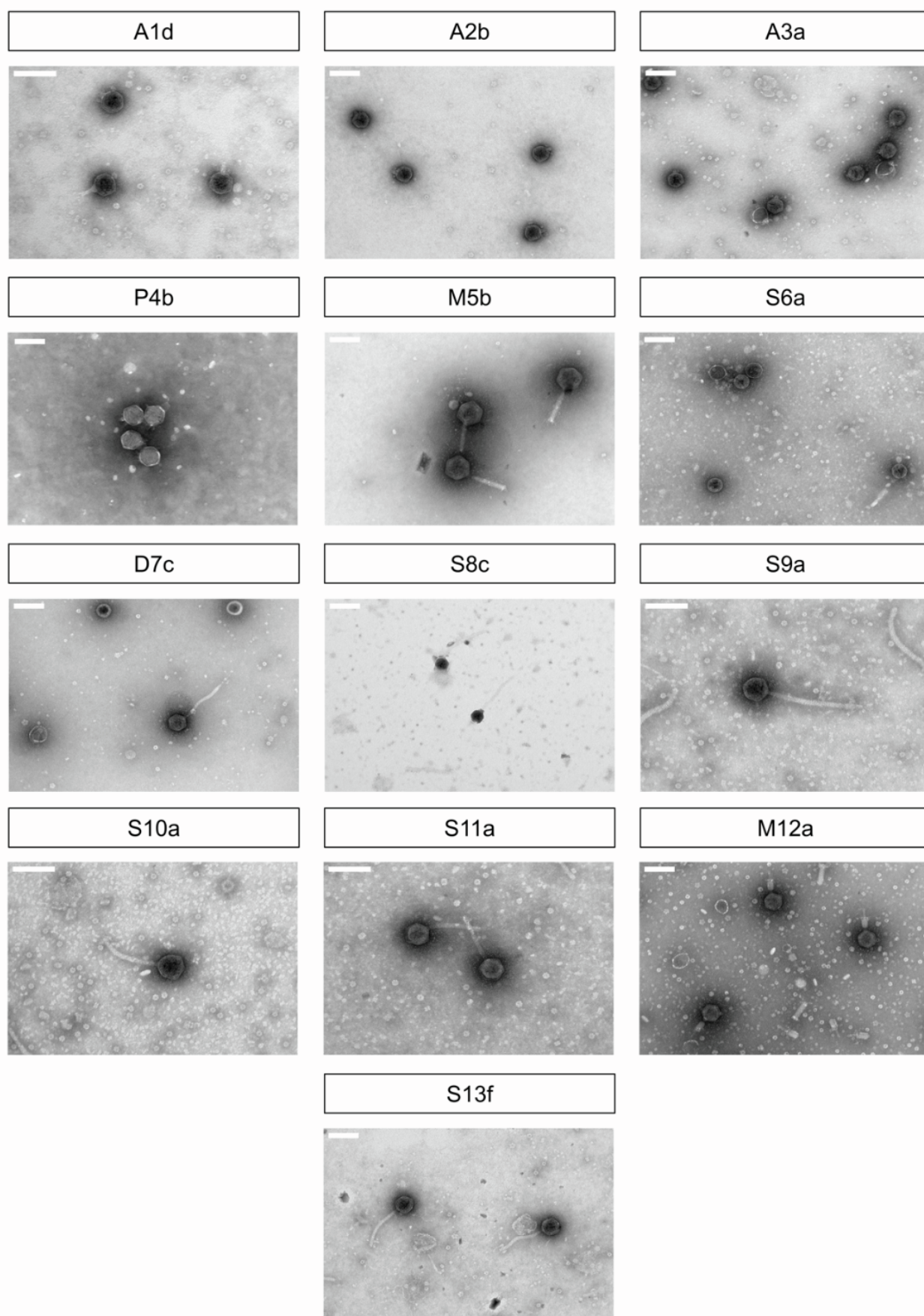


Figure S2. Transmission electron micrographs of representative phages from each group. Scale bar: 100 nm.

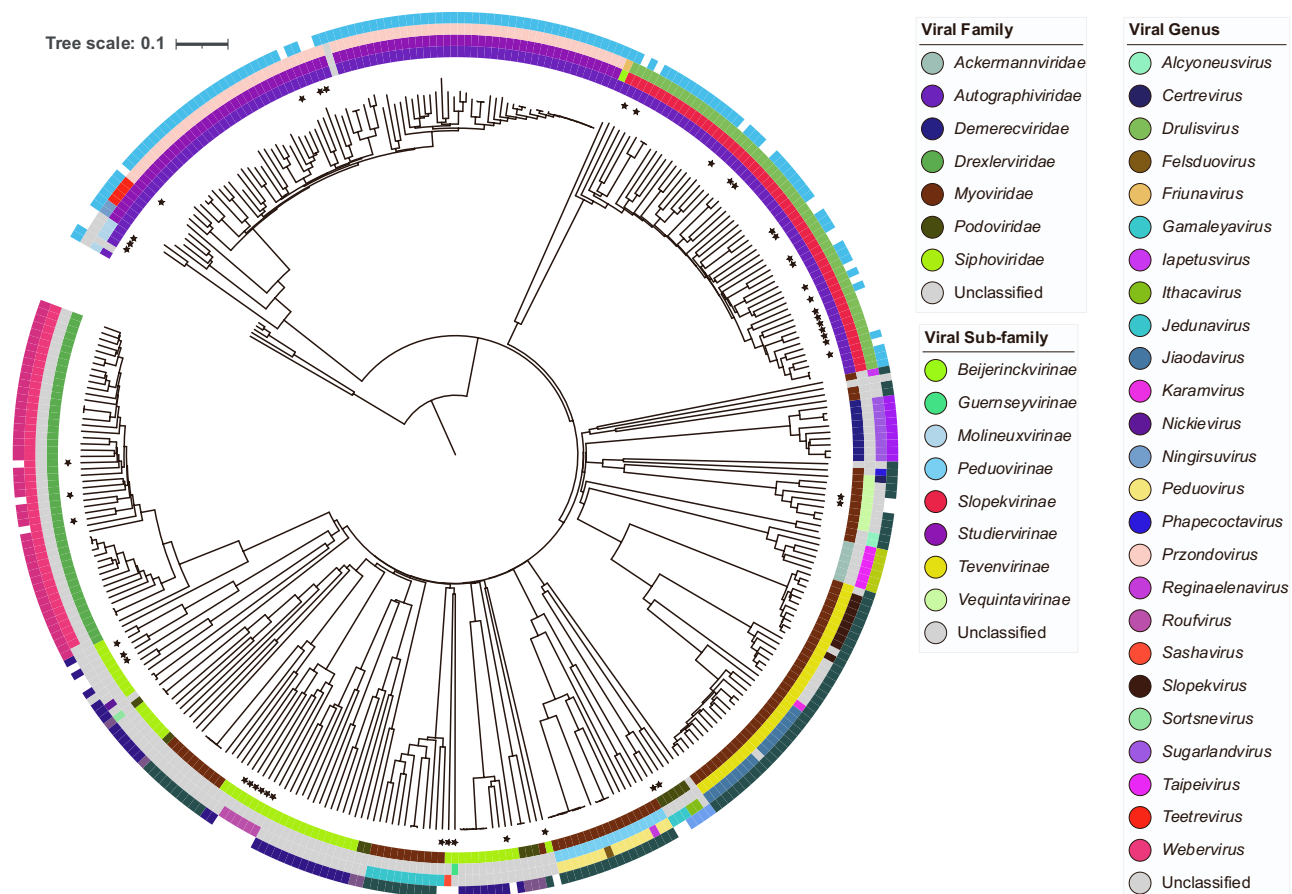


Figure S3. Dendrogram of the isolated phages in the context of previously described *Klebsiella* phages. Phage proteomic tree of all *Klebsiella* phages (February 2021), including the 46 sequenced phages (marked with a star) after computing genome-wide sequence similarities with tblastx. The tree scale shows global genomic distances. Clades of phages were colored based on available ICTV taxonomy.

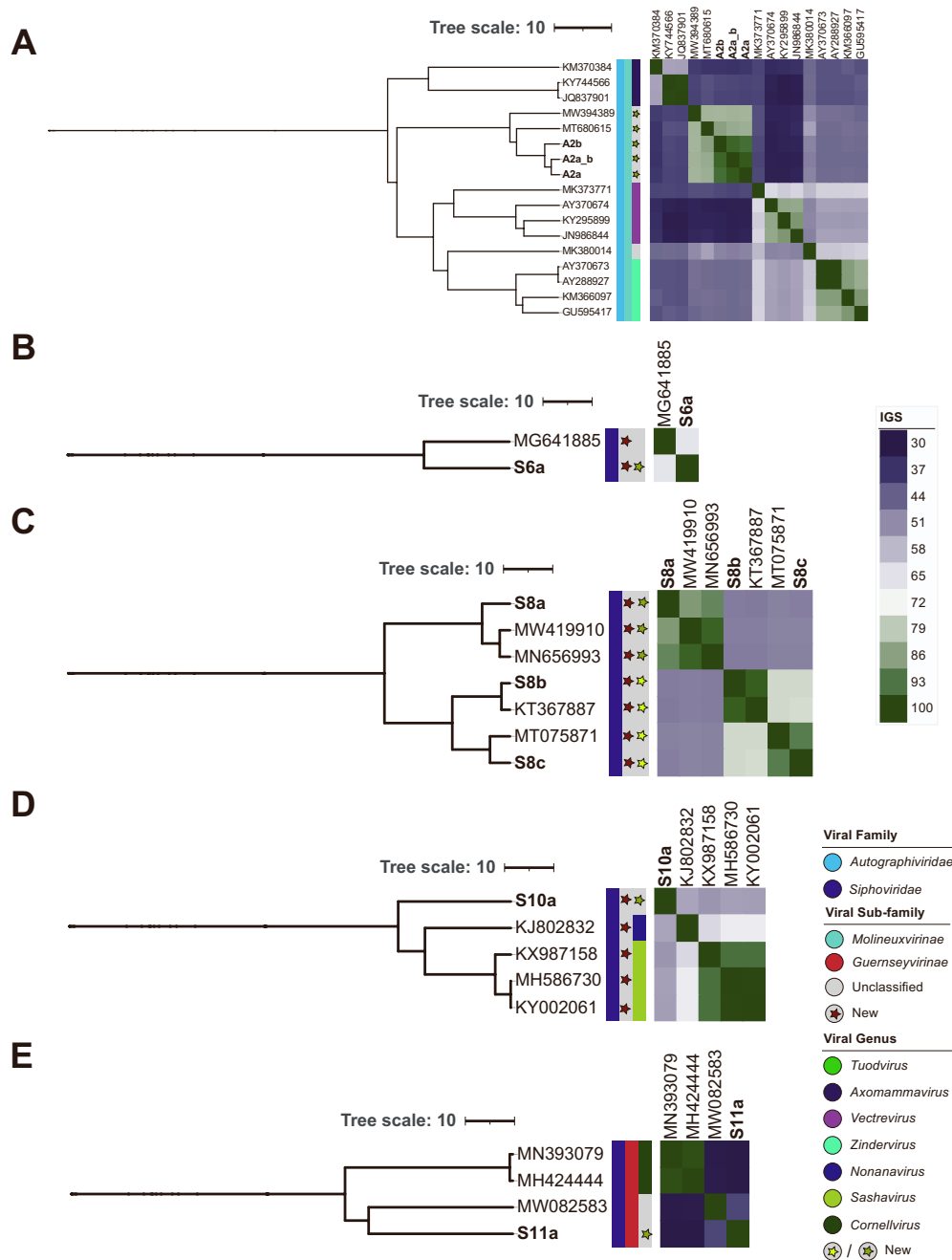


Figure S4. Dendrograms of unclassified viruses based on IGS values. Phages that were not classified at the subfamily and/or genus levels by similarity (IGS<45%) with any phage from the database were confronted against the NCBI (Accessed: 01/03/22) to exclude that additional related phages had been deposited and classified since the initial analysis. Colored strips denote family, subfamily, and viral genus. Subfamily and genus suggestions are marked with a star and are based on the ICTV 70% nucleotide identity threshold of the full genome length. Phages from this work are in bold. **A.** Phages of group 2 (A2a, A2b, and A2a_b) represent a genus, which includes *Proteus* phage PmP19 (MT680615) and *Klebsiella* phage vB_KpP_FBKp16 (MW394389), as their IGS values are >70%. **B.** Phage of group 6, S6a, is an orphan phage related at the subfamily level with *Salmonella* phage PMBT28 (MG641885) (IGS=65.3%). **C.** Phages of group 8 belong to the same subfamily (IGS > 45%) but to different genera: phage S8a belongs to the same genus as *Enterobacter* phages ATCEA85 (MN656993) and ATCEA23 (MW419910), whereas phages S8b and S8c are grouped with *Klebsiella* phages vB_Kp3 (same species as phage S8b) and vB_KleS-HSE3 (MT075871). **D.** Phage of group 10, S10a, is related at the subfamily level with *Sashavirus* and *Nonanavirus* phages, but represents another genus. **E.** Phage of group 11, S11a, is an orphan phage within the *Guernseyvirinae* subfamily and another proposed genus. Tree scale denotes IGS distance (100-IGS).

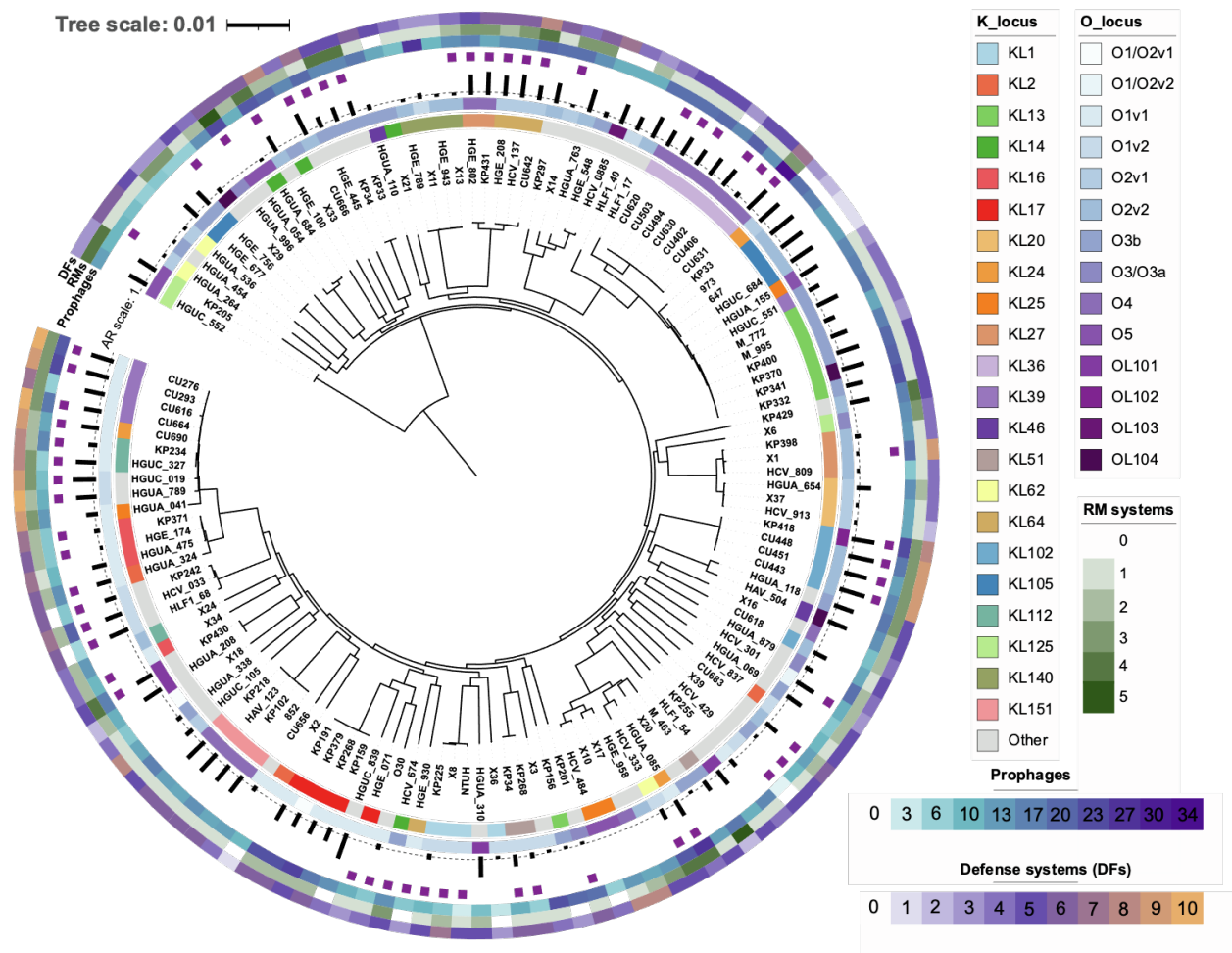


Figure S5. Core genome maximum likelihood phylogeny of *K. pneumoniae* strains used in this study. The innermost ring shows K-antigen prediction (capsular locus type, CLT) followed by O-antigen prediction (OLT). CLTs with fewer than three strains are shown as 'Other'. Black bars indicate the number of resistance genes to different antibiotic classes. Purple squares indicate the presence of CRISPR-Cas loci. The number of prophages, Restriction-Modification systems (RMs) and other defense systems (DFs) are also shown for each strain. Five strains that did not belong to *K. pneumoniae sensu stricto* (KP42, X22, CU371, KP205, HGUC_552) were omitted from this view.

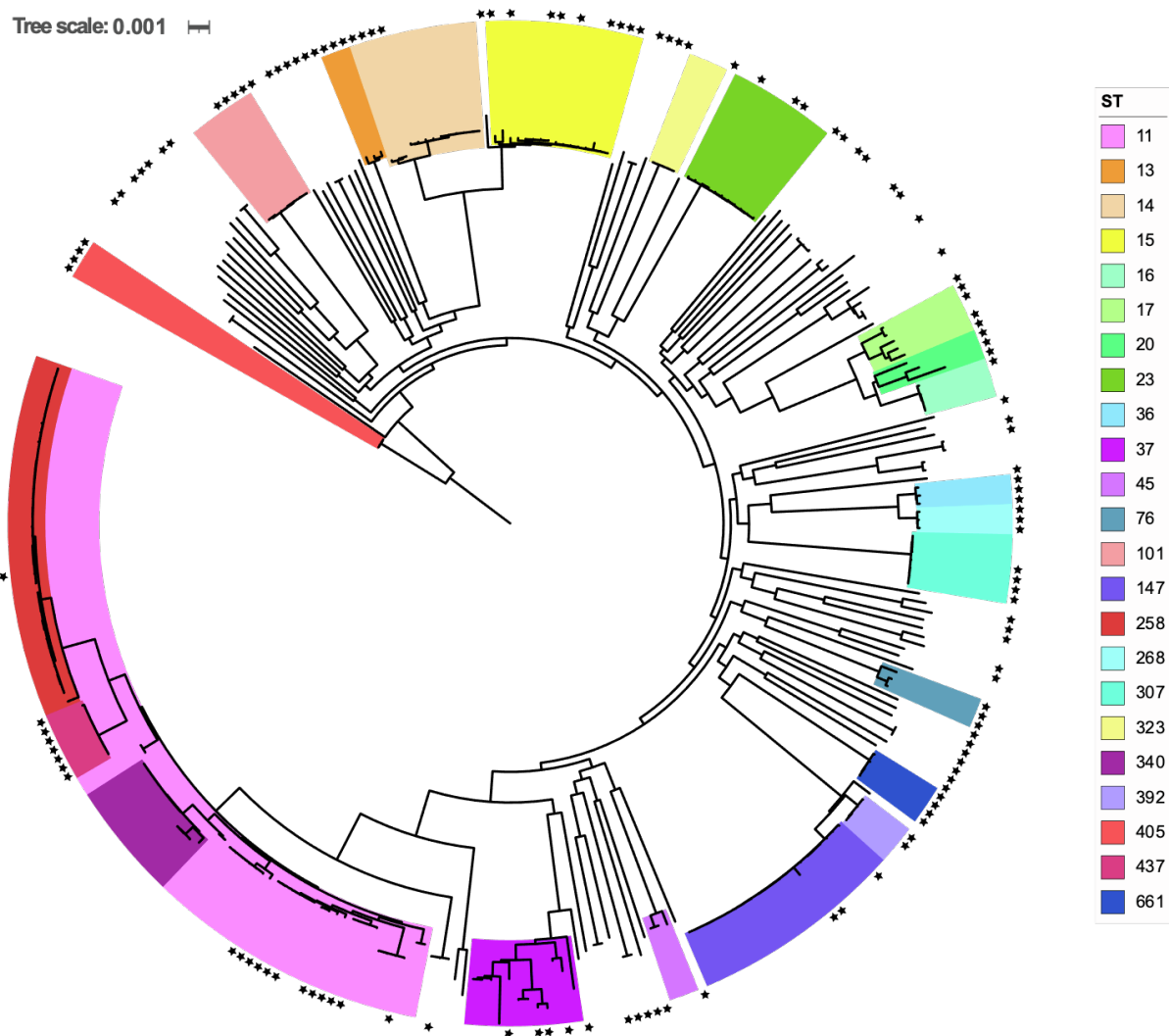
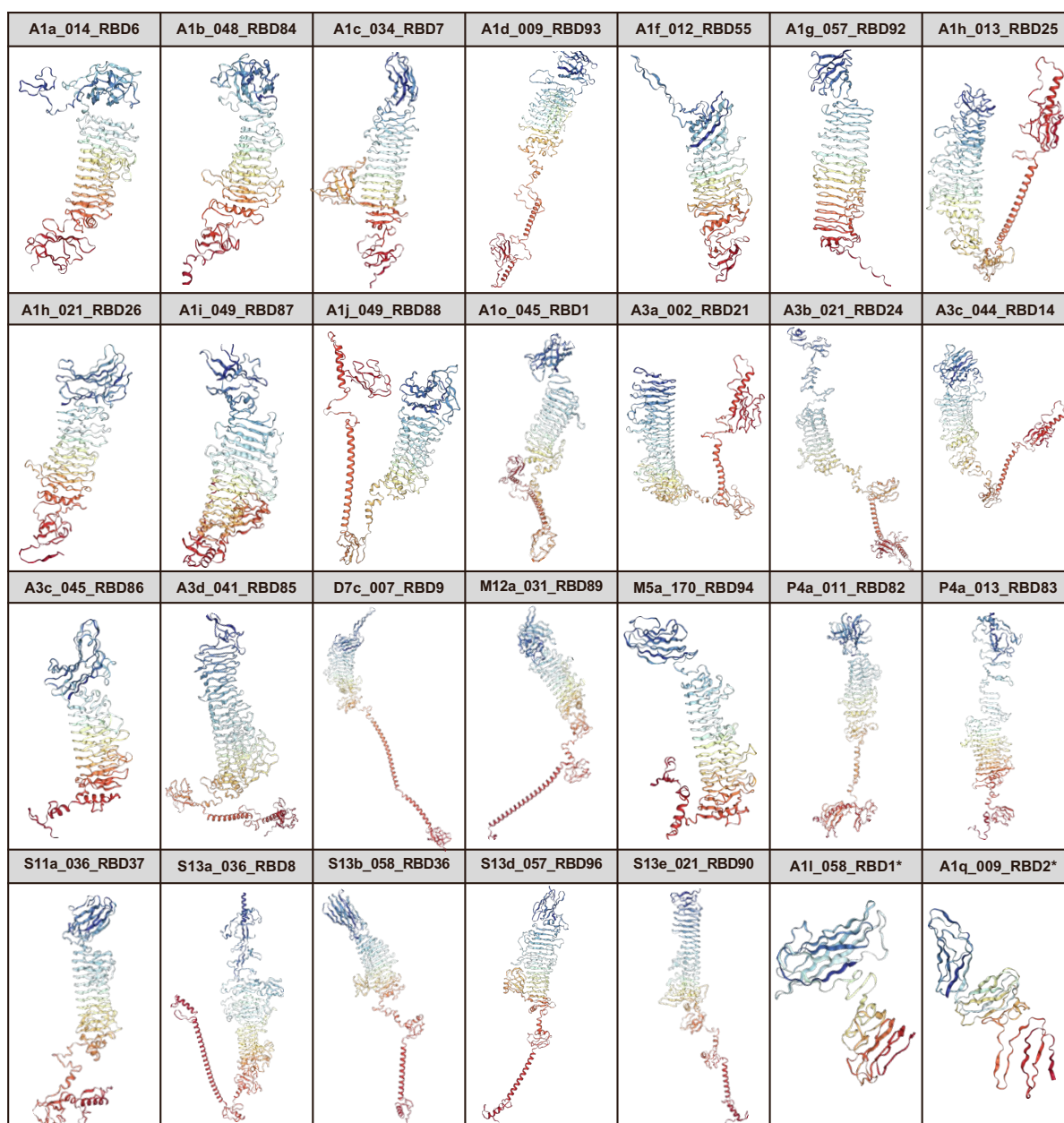


Figure S6. Phylogenetic context of the *K. pneumoniae* strains used in this study. A maximum-likelihood tree was constructed using the 138 strains included in this study (indicated with asterisks) and 172 representative sequences of each *K. pneumoniae* clonal complex. Colors represent major sequence types (STs). The five strains that did not belong to *K. pneumoniae sensu stricto* were excluded from this view.

A



B

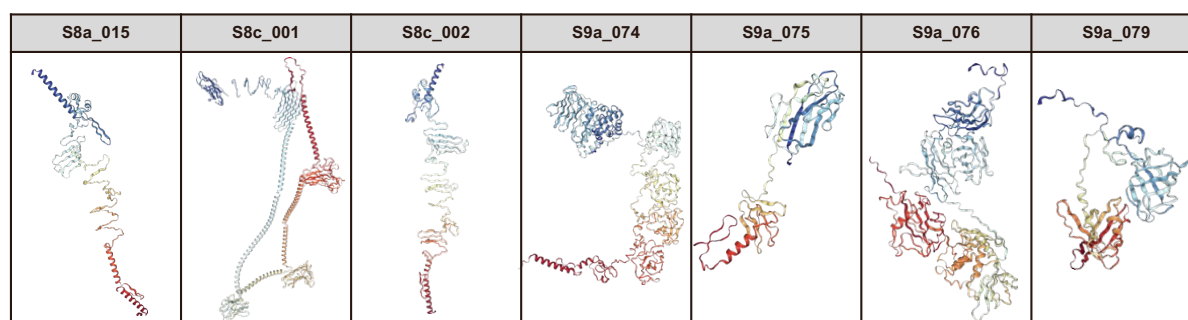


Figure S7. Predicted RBP structures of phages. A. TSPs with predicted depolymerase activity. **B.** RBPs structures of S8/S9 phages. Only structures of RBPs < 1500 aa are shown. Each structure is labelled with the phage name, the protein id and the assigned RBD cluster, if applicable. Asterisks denote putative soluble depolymerases.

Table S5. RBPs of S8/S9 phages and associated domains found with InterProScan.

Phage	Protein	start	end	hit	InterProScan	e-value
S8a	S8a_00018	340	500	Tip attachment protein J	IPR032876	2.00E-12
S8a	S8a_00018	621	714	Fibronectin type III	IPR036116	9.86E-06
S8a	S8a_00018	1265	1428	Galactose-binding domain-like	IPR008979	2.98E-05
S8a	S8a_00018	1532	1664	Carbohydrate binding domain	IPR003305	3.90E-12
S8a	S8a_00018	1916	2064	Galactose-binding domain-like	IPR008979	7.01E-05
S8a	S8a_00015	271	323	Chaperone of endosialidase	IPR030392	6.90E-10
S8b	S8b_00001	484	563	Fibronectin type III	IPR036116	2.10E-05
S8b	S8b_00001	1925	2062	Galactose-binding domain-like	IPR008979	7.82E-08
S8b	S8b_00001	2575	2719	Galactose-binding domain-like	IPR008979	3.23E-07
S8b	S8b_00002	324	371	Chaperone of endosialidase	IPR030392	1.80E-07
S8c	S8c_00001	84	246	Galactose-binding domain-like	IPR008979	4.76E-06
S8c	S8c_00001	544	688	Galactose-binding domain-like	IPR008979	1.26E-07
S8c	S8c_00001	997	1128	Fibronectin type III	PF12421	6.30E-09
S8c	S8c_00002	318	365	Chaperone of endosialidase	IPR030392	1.80E-07
S9a	S9a_00074	881	1016	SGNH hydrolase-type esterase domain	IPR013830	8.10E-16