Fig. S1

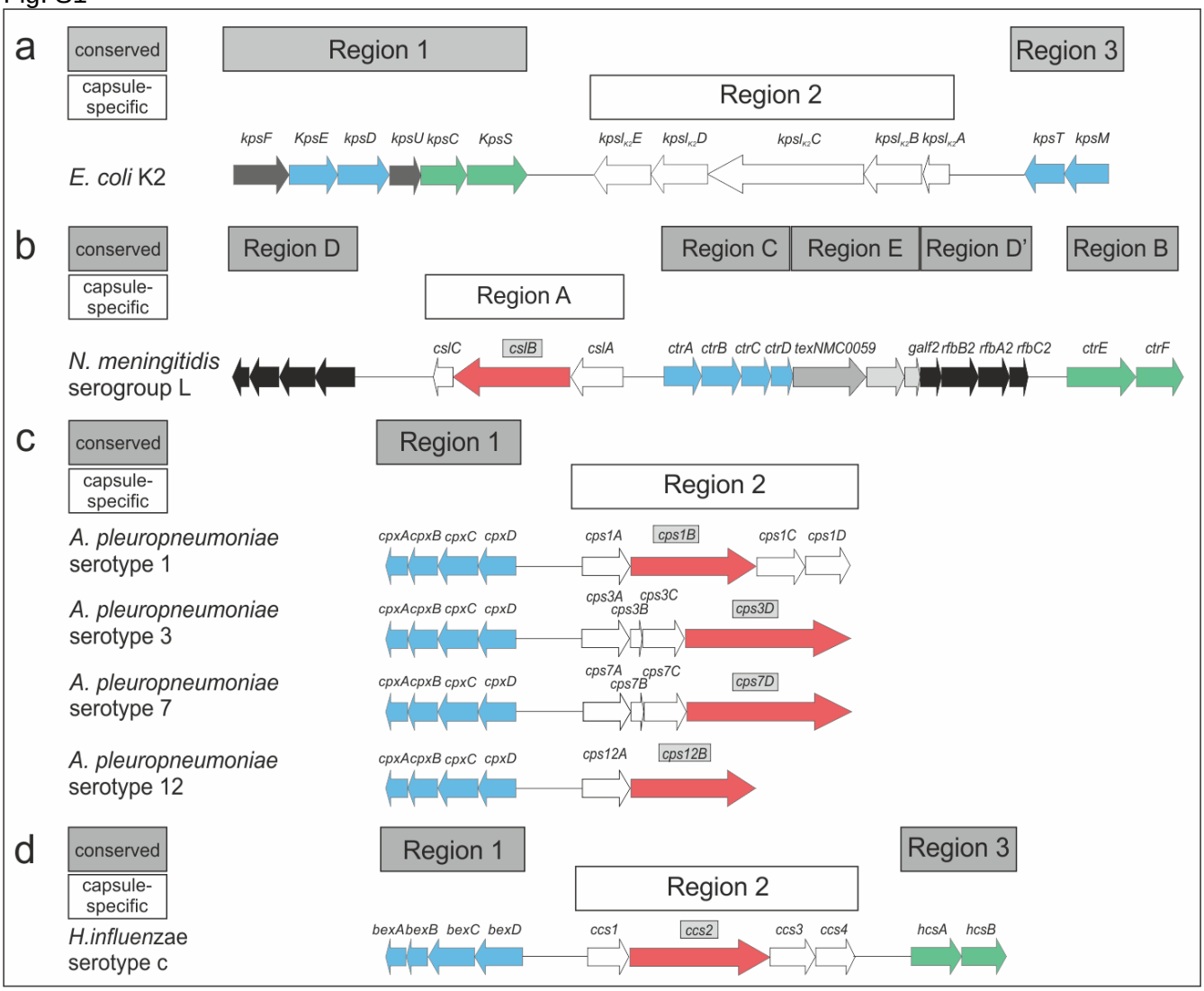


Fig. S1: Schematic overview of the capsule gene clusters of **a** *E. coli* K2 (modified from E. L. Buckles, X. Wang, M. C. Lane, C. V. Lockatell, D. E. Johnson, D. A. Rasko, H. L. T. Mobley, M. S. Donnenberg, J Infect Dis 199:1689–97, 2009), **b** *N. meningitidis* serogroup L (modified from O. B. Harrison, H. Claus, Y. Jiang, J. S. Bennett, H. B. Bratcher, K. A. Jolley, C. Corton, R. Care, J. T. Poolman, W. D. Zollinger, et al., Emerg Infect Dis 19:566–73, 2013), **c** *A. pleuropneumoniae* serotypes 1, 3, 7 and 12 (modified from H. Ito, J Vet Med Sci 77:583–6, 2015; S. G. Jessing, P. Ahrens, T. J. Inzana, Ø. Angen, Vet Microbiol 129:350–9, 2008; Z. Xu, X. Chen, L. Li, T. Li, S. Wang, H. Chen, R. Zhou, J Bacteriol 192:5625–36, 2010) and **d** *H. influenzae* serotype c (modified from S. W. Satola, P. L. Schirmer, M. M. Farley, Infect Immun 71:3639–3644, 2003; T.-T. Lâm, H. Claus, M. Frosch, U. Vogel, Res Microbiol 162:483–7, 2011; S. Sukupolvi-Petty, S. Grass, J. W. St Geme, III, J Bacteriol 188:3870–7, 2006). The gene clusters are divided into conserved regions (grey boxes) and capsule-specific regions (white boxes). The conserved regions encode proteins necessary for translocation (green) and export (blue) of the capsule polymer to the cell surface. Genes encoding capsule polymerases are highlighted in red and localized in the capsule specific region. Genes and interspaces in this scheme are not drawn to scale. Graphical representation in the style of B. F. Cress, J. A. Englaender, W. He, D. Kasper, R. J. Linhardt, M. A. G. Koffas, FEMS Microbiol Rev 38:660–697, 2014.

Fig. S2

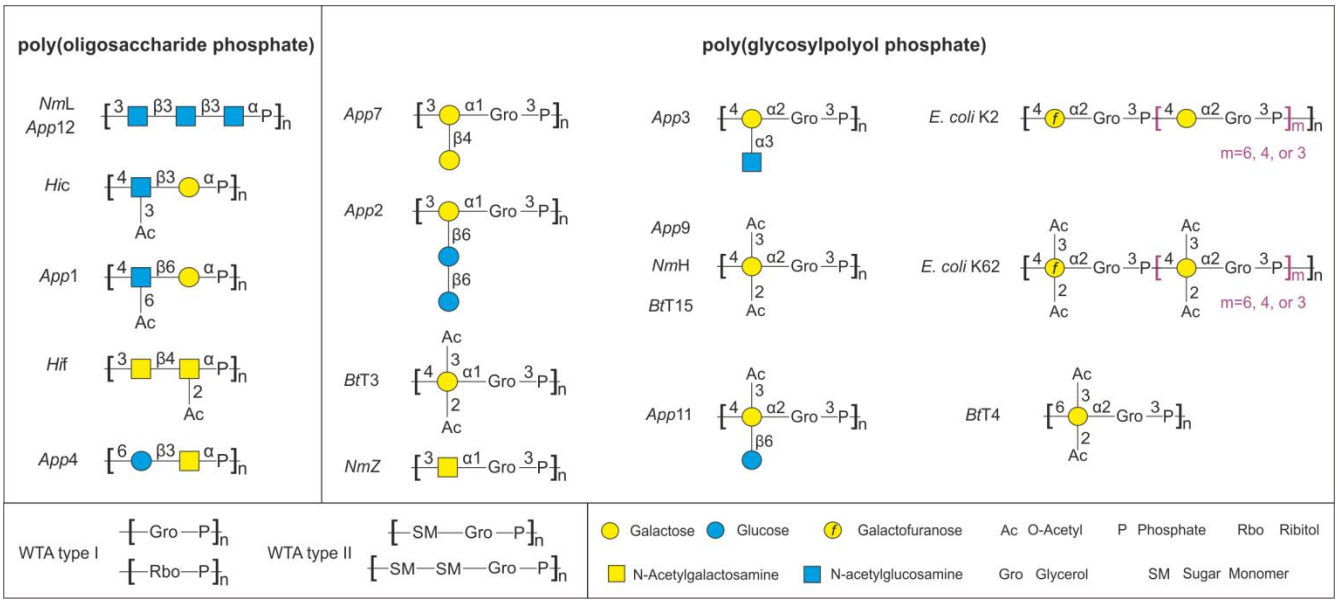


Fig. S2: Capsule structures of group 2 capsule expressing bacteria that encode TagF-like polymerases. Branching mono- and oligosaccharides as well as O-acetyl groups are usually introduced by separate enzymes. Schematics of wall teichoic acid (WTA) type I and II are depicted for comparison and displayed according to I. B. Naumova, A. S. Shashkov, E. M. Tul'skaya, G. M. Streshinskaya, Y. I. Kozlova, N. V Potekhina, L. I. Evtushenko, E. Stackebrandt, *FEMS Microbiol Rev 25:*269–84, 2001. To allow a concise display, bacterial species are abbreviated in italics and serogroup/serotype classification is added in regular font. Abbreviations used are: *App*, *Actinobacillus pleuropneumoniae*; *Bt*, *Bibersteinia trehalosi*; *Hi*, *Haemophilus influenza*; *Nm*, *Neisseria meningitidis*.

Fig. S3

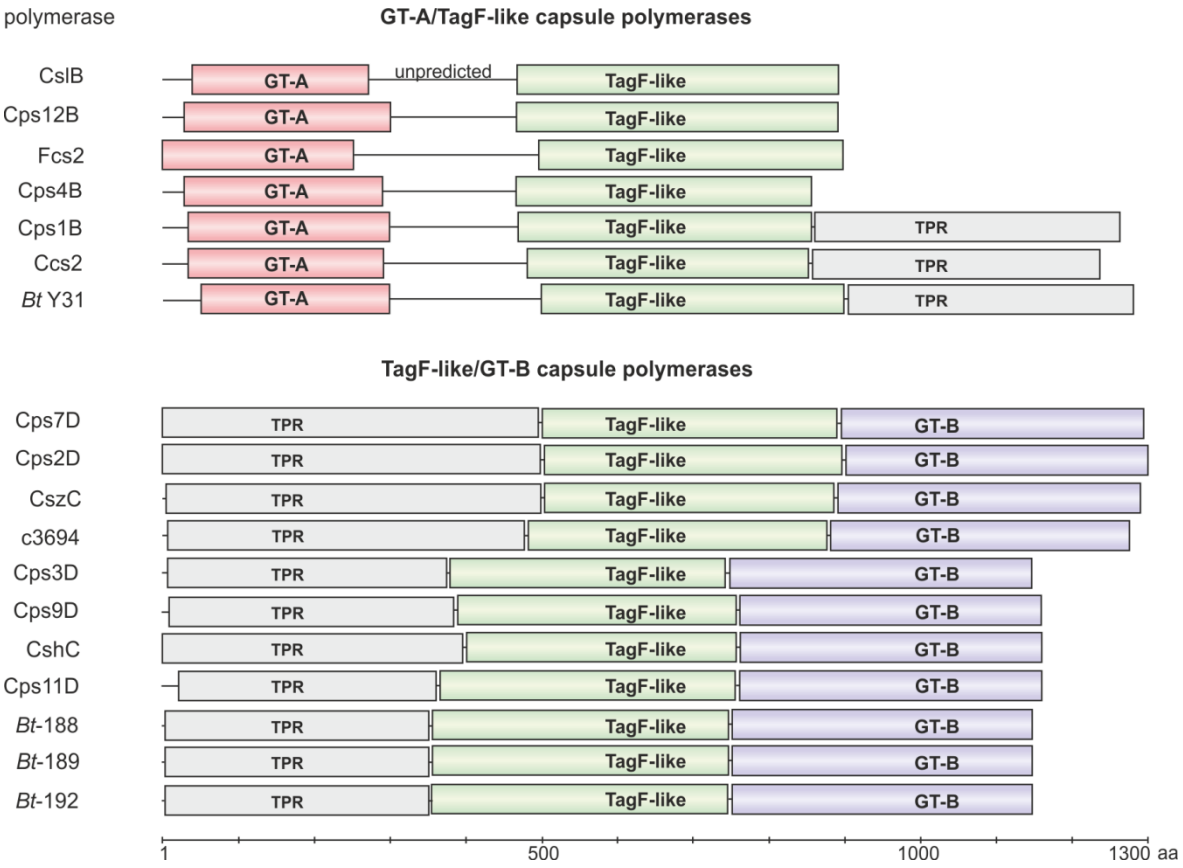


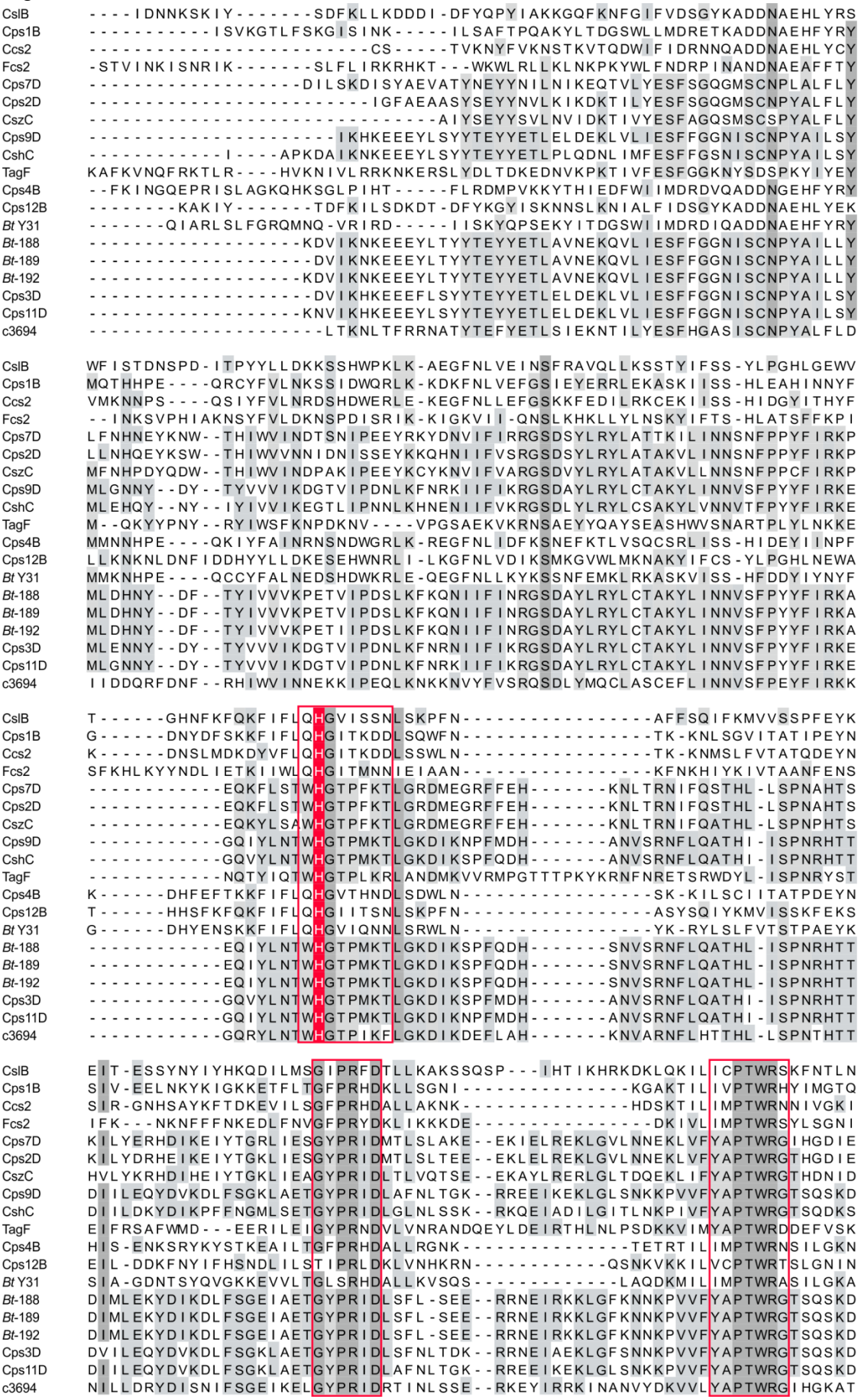
Fig. S3: Overview of the predicted architecture of all TagF-like polymerases analyzed in this study. Homology modeling was performed using the structure prediction tool PHYRE2 (L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, Nat Protoc 10: 845–58, 2015.). The name of the polymerase is displayed in front of each model. A ruler (bottom) indicates the length of each polypeptide as well as the sequence coverage of each modelled domain. The following protein sequences were submitted to PHYRE2: CslB of *N. meningitidis* serogroup L (uniprot: Q9RGQ9), Cps1B of *A. pleuropneumoniae* serotype 1 (uniprot: E0EA77), Cps12B of *A. pleuropneumoniae* serotype 12 (uniprot: Q69AA8), Ccs2 of *H. influencae* serotype c (GenBank: AEC50903.1), Fcs2 of *H. influencae* serotype f (GenBank: AAQ12660.1), Cps4B of *A. pleuropneumoniae* serotype 4 (uniprot: F4YBG0), *Bt* Y31 of the non-serotyped *Bibersteinia trehalosi* strain Y31 (GenBank: OAQ14264.1), Cps7D of *A. pleuropneumoniae* serotype 7 (GenBank: ACE62291.1), Cps2D of *A. pleuropneumoniae* serotype 2 (uniprot: Q6UYC4), CszC of *N. meningitidis* serogroup Z (uniprot: Q5QRV6), Cps3D of *A. pleuropneumoniae* serotype 3 (GenBank: KY807157), Cps9D of *A. pleuropneumoniae* serotype 9 (uniprot: E0F019), CshC of *N. meningitidis* serogroup H (uniprot: H6T5X6), Cps11D of *A. pleuropneumoniae* serotype 11 (uniprot: E0FCQ3), *Bt*-188 of the non-serotyped *Bibersteinia trehalosi* strain USDA-ARS-USMARC-188 (GenBank: AHG82487.1), *Bt*-189 of the non-serotyped *Bibersteinia trehalosi* strain USDA-ARS-USMARC-189 (GenBank: AHG84818.1), *Bt*-192 of the non-serotyped *Bibersteinia trehalosi* strain USDA-ARS-USMARC-192 (GenBank: AGH37704.1) and c3694 of *E. coli* K2 strain CFT073 (GenBank: AAN82142.1).

Fig. S4

	TagF	CshC	Bt189	Bt188	Bt192	Cps3D	Cps9D	Cps11D	c3694	CszC	Cps7D	Cps2D	Fcs2	Cps1B	BtY31	Ccs2	Cps4B	Csl B	Cps12B
TagF	100.00	28.57	27.96	28.46	28.15	27.64	28.30	28.61	29.14	28.93	27.73	28.53	16.62	18.10	19.51	17.59	18.60	18.45	17.42
CshC	28.57	100.00	68.32	68.78	67.89	71.16	70.87	70.57	47.23	44.29	41.58	44.77	16.06	18.77	17.13	17.90	17.68	15.52	17.72
Bt189	27.96	68.32	100.00	98.67	97.91	77.45	77.37	77.23	48.54	44.81	44.83	46.90	17.68	21.98	17.85	19.57	18.40	14.71	17.22
Bt188	28.46	68.78	98.67	100.00	99.21	77.45	77.87	77.78	48.40	45.08	44.71	46.36	18.04	21.98	18.15	19.88	18.71	15.32	17.18
Bt192	28.15	67.89	97.91	99.21	100.00	77.45	77.37	77.28	47.75	45.08	44.71	46.36	17.99	21.98	17.85	19.57	18.40	15.02	17.22
Cps3D	27.64	71.16	77.45	77.45	77.45	100.00	92.29	92.06	44.27	44.14	42.86	44.35	17.07	20.99	16.87	18.27	16.51	14.67	18.65
Cps9D	28.30	70.87	77.37	77.87	77.37	92.29	100.00	100.00	44.83	43.05	42.55	43.82	17.33	20.99	16.87	18.89	16.82	15.27	18.67
Cps11D	28.61	70.57	77.23	77.78	77.28	92.06	100.00	100.00	44.71	43.05	42.22	43.82	17.33	20.99	16.87	18.89	16.82	15.27	18.67
c3694	29.14	47.23	48.54	48.40	47.75	44.27	44.83	44.71	100.00	44.99	48.55	48.40	17.82	19.63	17.38	20.00	17.33	18.99	18.92
CszC	28.93	44.29	44.81	45.08	45.08	44.14	43.05	43.05	44.99	100.00	69.89	70.62	16.27	17.99	16.67	18.35	16.92	16.52	17.47
Cps7D	27.73	41.58	44.83	44.71	44.71	42.86	42.55	42.22	48.55	69.89	100.00	87.53	17.77	18.60	16.36	17.13	20.24	17.99	18.67
Cps2D	28.53	44.77	46.90	46.36	46.36	44.35	43.82	43.82	48.40	70.62	87.53	100.00	16.92	19.82	17.88	19.27	19.94	17.99	19.34
Fcs2	16.62	16.06	17.68	18.04	17.99	17.07	17.33	17.33	17.82	16.27	17.77	16.92	100.00	30.13	26.91	30.58	26.77	30.18	32.55
Cps1B	18.10	18.77	21.98	21.98	21.98	20.99	20.99	20.99	19.63	17.99	18.60	19.82	30.13	100.00	51.91	50.66	50.13	31.17	31.61
BtY31	19.51	17.13	17.85	18.15	17.85	16.87	16.87	16.87	17.38	16.67	16.36	17.88	26.91	51.91	100.00	50.53	54.43	31.12	31.63
Ccs2	17.59	17.90	19.57	19.88	19.57	18.27	18.89	18.89	20.00	18.35	17.13	19.27	30.58	50.66	50.53	100.00	52.91	30.67	32.71
Cps4B	18.60	17.68	18.40	18.71	18.40	16.51	16.82	16.82	17.33	16.92	20.24	19.94	26.77	50.13	54.43	52.91	100.00	30.36	29.74
Csl B	18.45	15.52	14.71	15.32	15.02	14.67	15.27	15.27	18.99	16.52	17.99	17.99	30.18	31.17	31.12	30.67	30.36	100.00	51.36
Cps12B	17.42	17.72	17.22	17.18	17.22	18.65	18.67	18.67	18.92	17.47	18.67	19.34	32.55	31.61	31.63	32.71	29.74	51.36	100.00

Fig. S4: Sequence identity matrix (in %) based on a Clustal Omega multiple sequence alignment of all predicted TagF-like domains (as they are shown in Fig. S3) and the modelling template TagF.

Fig. S5


CslB
Cps1B
Ccs2
Fcs2
Cps7D
Cps2D
CszC
Cps9D
CshC
TagF
Cps4B
Cps12B
Bt Y31
Bt-188
Bt-189
Bt-192
Cps3D
Cps11D
c3694

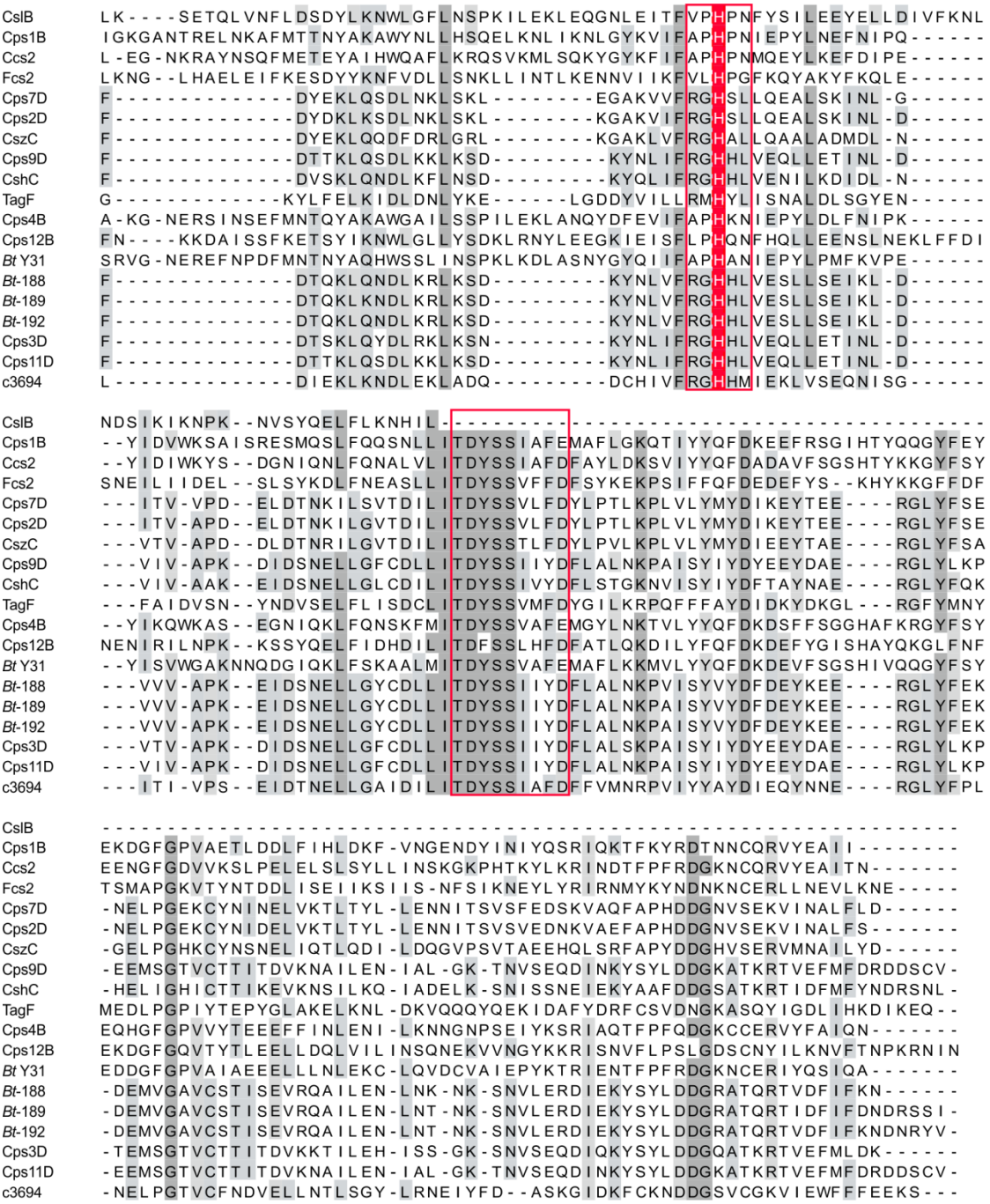


Fig. S5: Sequence alignment of all predicted TagF-like domains analyzed in this study including the sequence of the template TagF of *Staphylococcus epidermidis* (uniprot: Q5HLM5) used for PHYRE2 modeling. Database references for all TagF-like polymerase sequences are indicated in the figure legend of Supplementary Figure 3. Identical amino acids are shown in grey boxes and the conserved histidine residues are highlighted in red. The conserved active site motifs reported for TagF are shown in boxed sections (red). The sequence alignment was performed with Clustal Omega (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, Mol Syst Biol 7:539, 2011) on the uniprot website (http://www.uniprot.org/align/) (E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, A. Bairoch, Methods Mol Biol 406:89–112, 2007) and annotated with the Jalview software (A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Bioinformatics 25:1189–91, 2009).

Fig. S6

a

	K4CP	Fcs2	CslB	Cps12B	BtY31	Cps1B	Ccs2	Cps4B
K4CP	100.00	19.55	18.84	20.82	20.38	17.96	17.47	18.10
Fcs2	19.55	100.00	32.09	32.22	27.31	31.25	30.04	30.38
CslB	18.84	32.09	100.00	69.47	41.59	47.11	45.29	46.22
Cps12B	20.82	32.22	69.47	100.00	42.55	44.87	38.96	41.57
BtY31	20.38	27.31	41.59	42.55	100.00	54.11	47.37	49.79
Cps1B	17.96	31.25	47.11	44.87	54.11	100.00	54.84	54.37
Ccs2	17.47	30.04	45.29	38.96	47.37	54.84	100.00	58.63
Cps4B	18.10	30.38	46.22	41.57	49.79	54.37	58.63	100.00

b

	TarM	CshC	Cps3D	Cps9D	Cps11D	Bt189	Bt188	Bt192	c3694	CszC	Cps7D	Cps2D
TarM	100.00	22.04	23.03	23.19	23.10	27.05	25.76	25.99	21.95	22.02	22.66	21.93
CshC	22.04	100.00	70.26	70.66	70.95	70.69	64.62	65.63	46.09	40.95	41.28	40.33
Cps3D	23.03	70.26	100.00	92.03	92.03	75.84	68.89	69.51	45.29	40.87	41.49	41.49
Cps9D	23.19	70.66	92.03	100.00	100.00	77.12	68.89	69.51	45.31	40.82	42.05	41.43
Cps11D	23.10	70.95	92.03	100.00	100.00	77.12	69.07	69.51	45.29	40.87	41.75	41.49
Bt189	27.05	70.69	75.84	77.12	77.12	100.00	80.15	79.84	46.60	41.65	40.98	41.49
Bt188	25.76	64.62	68.89	68.89	69.07	80.15	100.00	98.71	46.34	39.85	40.72	40.72
Bt192	25.99	65.63	69.51	69.51	69.51	79.84	98.71	100.00	46.46	40.57	41.09	40.83
c3694	21.95	46.09	45.29	45.31	45.29	46.60	46.34	46.46	100.00	46.89	49.09	49.09
CszC	22.02	40.95	40.87	40.82	40.87	41.65	39.85	40.57	46.89	100.00	72.12	67.65
Cps7D	22.66	41.28	41.49	42.05	41.75	40.98	40.72	41.09	49.09	72.12	100.00	83.12
Cps2D	21.93	40.33	41.49	41.43	41.49	41.49	40.72	40.83	49.09	67.65	83.12	100.00

Fig. S6: Sequence identity matrices (in %) based on a Clustal Omega multiple sequence alignment of **a** all predicted GT-A domains (as they are shown in Fig. S3) and the C-terminal domain of the modelling template K4CP (*E. coli* K4 polymerase) and **b** all predicted GT-B domains (as they are shown in Fig. S3) and the modelling template TarM.

Fig. S7

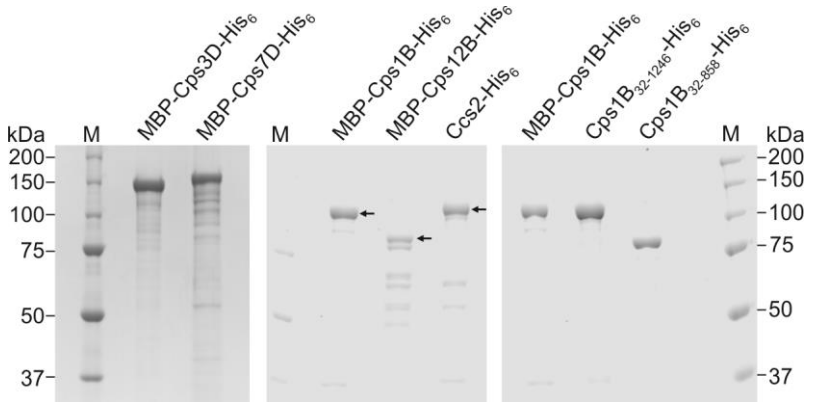


Fig. S7: Coomassie-stained SDS-polyacrylamide gel of TagF-like polymerases purified by affinity chromatography via their C-terminal His_6-tag. N-terminal fusion to maltose binding protein is indicated with 'MBP' in the construct name. 1.5-3 µg of protein were loaded per lane. MBP-Cps3D-His_6 (177 kDa) and MBP-Cps7D-His_6 (192 kDa) could be enriched as full-length constructs. Western blot analysis with an α-MBP antibody (data not shown) demonstrated that MBP-Cps1B-His_6 (190 kDa), MBP-Cps12B-His_6 (146 kDa) and Ccs2-His_6 (144 kDa) were exclusively purified as N-terminal degradation products (indicated by arrows), lacking the MBP-tag. N-terminal degradation is common for group 2 polymerases and usually does not interfere with activity (T. Fiebig, F. Freiberger, V. Pinto, M. R. Romano, A. Black, C. Litschko, A. Bethe, D. Yashunsky, R. Adamo, A. Nikolaev, F. Berti, R. Gerardy-Schahn, *J Biol Chem 289:*19395–407, 2014; C. Litschko, M. R. Romano, V. Pinto, H. Claus, U. Vogel, F. Berti, R. Gerardy-Schahn, T. Fiebig, *J Biol Chem 290:*24355–66, 2015). The dominant protein band purified from the MBP-Cps1B-His_6 expression culture was N-terminally sequenced and could be identified as ΔN31 truncation of Cps1B. The corresponding construct $Cps1B_{32\text{-}1246}$-His_6 (144kDa) was cloned and purified as well as the N- and C-terminally truncated construct $Cps1B_{32\text{-}858}$-His_6 (99 kDa) lacking the TPR domain. M, marker; His_6, hexa-histidine tag.

Fig. S8

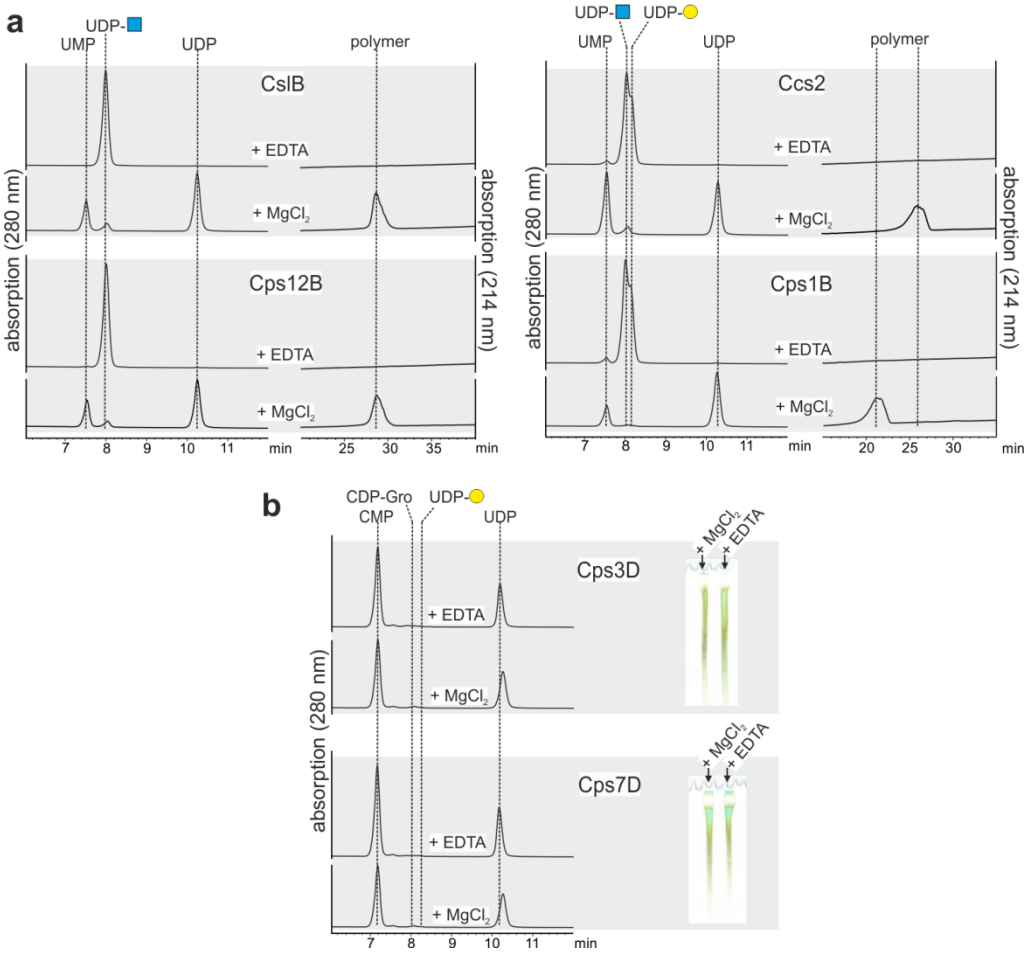


Fig. S8: HPLC-AEC assay in the presence (+$MgCl_2$) and absence (+EDTA) of magnesium chloride. a Polymerases containing GT-A-folded domains depend on Mg^{2+}, most likely to stabilize the negative charge of the diphosphate of their donor substrate (C. Breton, L. Snajdrová, C. Jeanneau, J. Koca, A. Imberty, *Glycobiology 16*:29R–37R, 2006). **b** Consistent with the fact that there is no evidence of a bound metal ion associated with catalysis in GT-B folded enzymes (C. Breton, L. Snajdrová, C. Jeanneau, J. Koca, A. Imberty, *Glycobiology 16*:29R–37R, 2006), polymerases adopting the TagF-like/GT-B architecture also work in the presence of the chelating agent EDTA.

Fig. S9

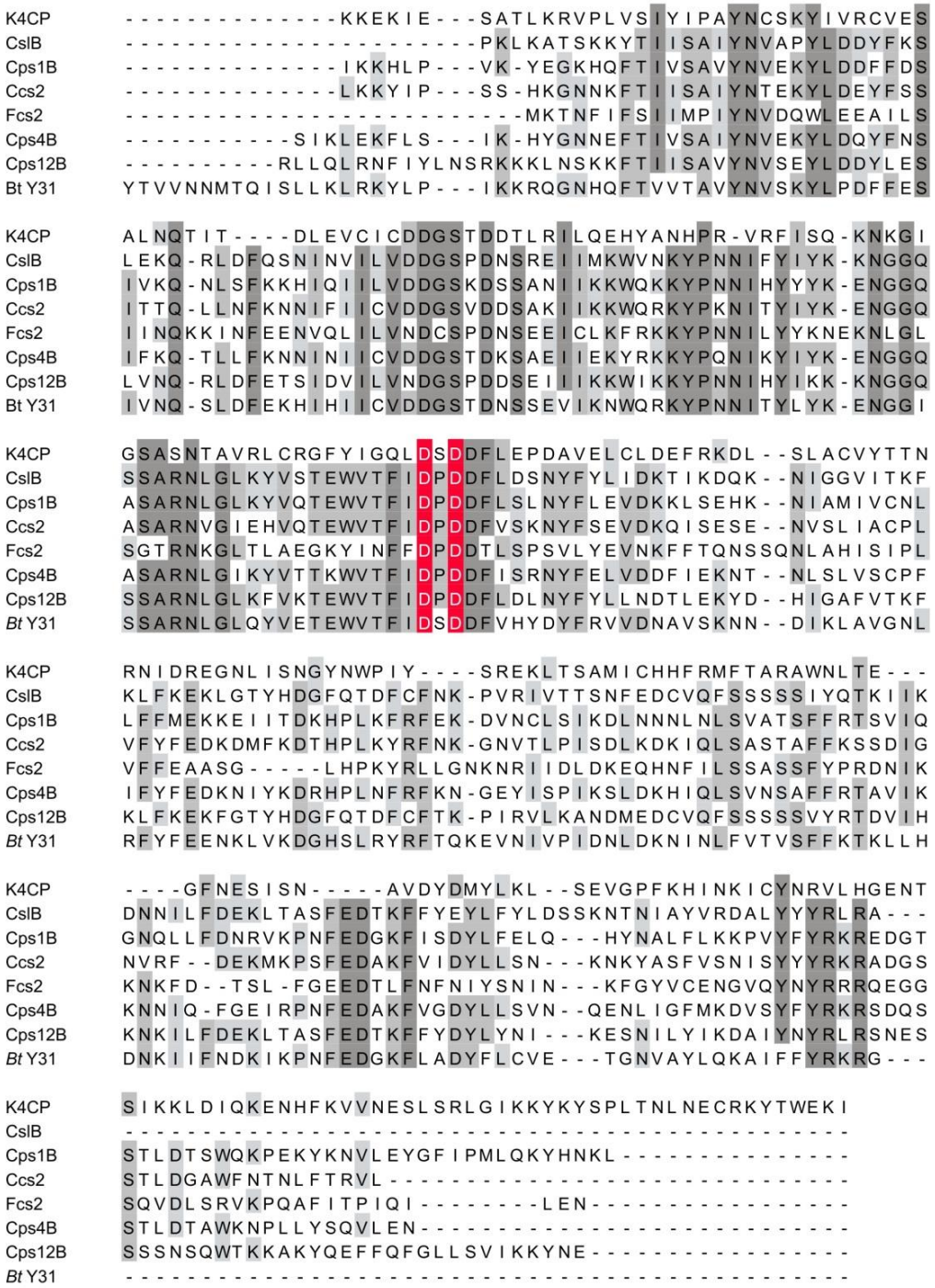


Fig. S9: Sequence alignment of all predicted N-terminal GT-A domains analyzed in this study including the sequence of the template K4CP (uniprot: Q8L0V4) used for PHYRE2 modeling. Database references for all TagF-like polmyerase sequences are indicated in the legend of Fig. S3. Identical amino acids are shown in grey boxes and aspartate residues of the conserved DxD motif are highlighted in red. The sequence alignment was performed with Clustal Omega (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, Mol Syst Biol 7:539, 2011) on the uniprot website (http://www.uniprot.org/align/) (E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, A. Bairoch, Methods Mol Biol 406:89–112, 2007) and annotated with the Jalview software (A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Bioinformatics 25:1189–91, 2009).

Fig. S10

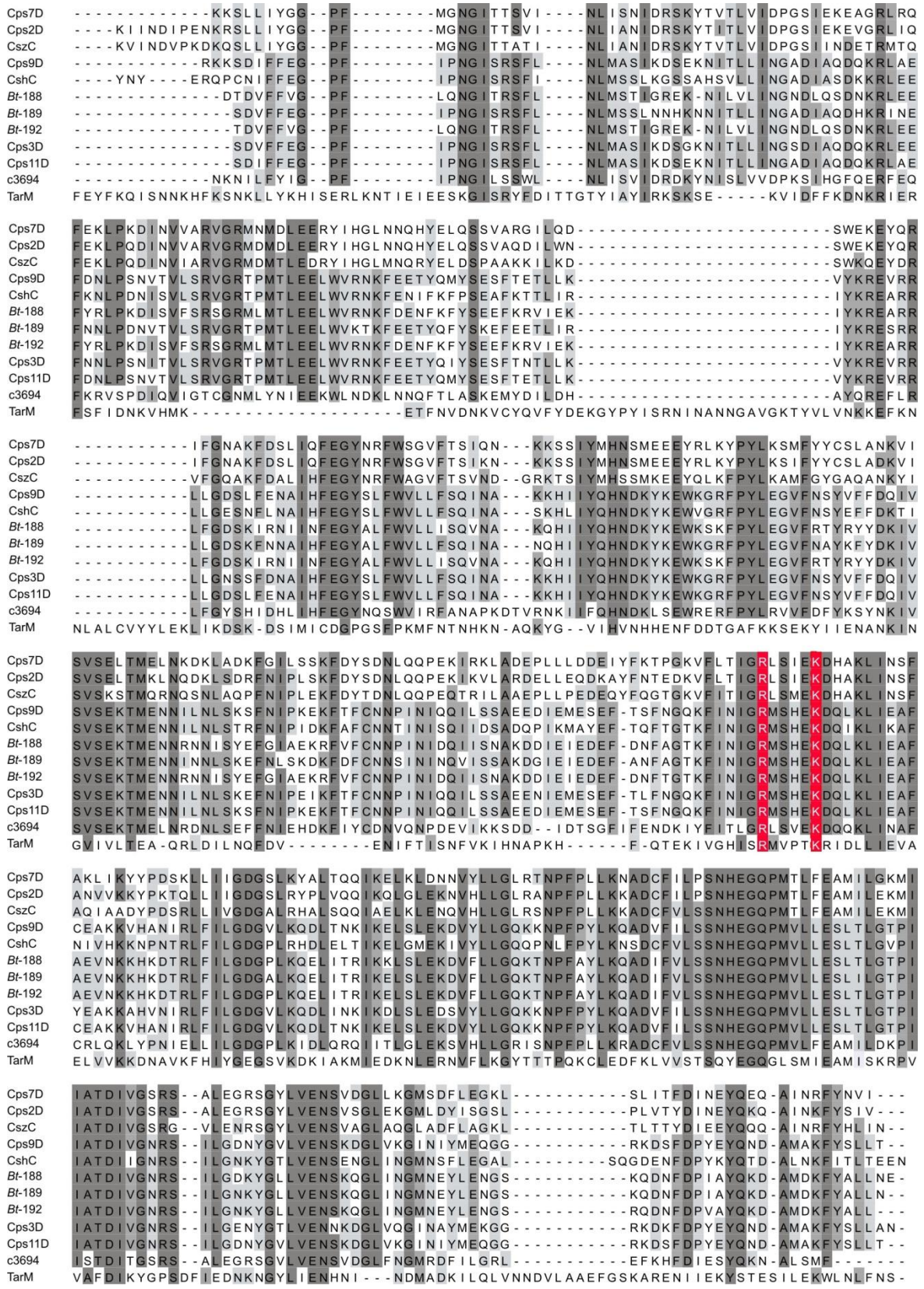


Fig. S10: Sequence alignment of all predicted C-terminal GT-B domains analyzed in this study including the sequence of the template TarM of *Staphylococcus aureus* (uniprot: A0A0J9X257)

used for PHYRE2 modeling. Database references for all TagF-like polymerase sequences are indicated in the legend of Supplementary Figure 3. Identical amino acids are shown in grey boxes and the conserved arginine and lysine residues are highlighted in red. The sequence alignment was performed with Clustal Omega (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, Mol Syst Biol 7:539, 2011) on the uniprot website (http://www.uniprot.org/align/) (E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, A. Bairoch, Methods Mol Biol 406:89–112, 2007) and annotated with the Jalview software (A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Bioinformatics 25:1189–91, 2009).