Table of contents

Table S1
Enterobacteriaceae tailed phage MCP matches in bacterial genomes

Cluster	Prototype phage	Prototype phage host	Bacterial genera / authentic phage with >50% MCP matches outside the *Enterobacteriaceae*[1]
Lytic 1	T1	*Escherichia*	None / none
Lytic 2	T4	*Escherichia*	None / *Acinetobacter* phage Acj9-75%
Lytic 3	Vi01	*Salmonella*	None / *Delftia* phage øW-14-54%
Lytic 4	T5	*Escherichia*	None / *Vibrio* phage pVp-1-60%
Supercluster	T7		
Lytic 5	T7	*Escherichia*	None / *Stenotrophomonas* phage IME15 82%
Lytic 6	SP6	*Salmonella*	None / none
Lytic 7	KP34	*Klebsiella*	None / *Vibrio* phage VP93-74%
Lytic 8	LIMEzero	*Pantoea*	None / *Ralstonia* phage RSB3-51%
Lytic 9	øKT	*Escherichia*	None / none
Lytic 10	GAP227	*Cronobacter*	None / *Aeromonas* phage øAS7-82%
Lytic 11	N4	*Escherichia*	*Mezorhizobium-75%* / *Delftia* phage RG-2014-62%
Lytic 12	9NA	*Salmonella*	None / none
Lytic 13	Chi	*Salmonella*	None / none
Lytic 14	øEco32	*Escherichia*	None / none
Lytic 15	Felix-O1	*Salmonella*	None / none
Supercluster	SETP3		
Lytic 16	SETP3	*Salmonella*	*Acinetobacter-64% / none*
	K1-dep(1)	*Escherichia*	*Sinorhizobium-61%* / *Synechococcus* phage S-CBS4-56%
Lytic 17	SO-1	*Sodalis*	None / none
Lytic 18	ECO1230-10	*Escherichia*	*Marinobacter-60%* / *Pseudomonas* phage PPpW-3-68%
Lytic 19	Gj1	*Escherichia*	*Aeromonas-60%* / *Shewanella* phage Spp001-56%
Lytic 20	PY100	*Yersinia*	None / none
Supercluster	rV5		
Lytic 21	Ø92	*Escherichia*	None / none
Lytic 22	rV5	*Escherichia*	None / none
Lytic 23	SPN3US	*Salmonella*	None / none
Lytic 24	RaK2	*Klebsiella*	None / none
Lytic 25	øR1-37	*Yersinia*	None / none
Lytic 26	E1	*Salmonella*	None / *Vibrio* phage pYD38-A-79%
Lytic 27	EMCL-117	*Escherichia*	*Burkholderia-52%* / *Pseudomonas* phage DL60-67%[2]
Lytic 28	KF-1	*Edwardsiella*	None / *Vibrio* phage VPMS1-69%
Lytic 29	MSW-3	*Edwardsiella*	None / *Iodobacter* phage øPLPE-68%
Lytic 30	Ea35-70	*Erwinia*	None / none
Lytic 31	ϕEaH1	*Erwinia*	None / none
Lytic 32	9g	*Escherichia*	None / none
Lytic 33	IME-EC2	*Escherichia*	None / none
Lytic 34	Ss1	*Cronobacter*	None / none
Lytic 35	CVT22	*Citrobacter*	None / i phage CW02-63%
Supercluster	Lambda		
Temperate 1	Lambda	*Escherichia*	None / none
Temperate 2 & 3	ø80 & N15	*Escherichia*	None / none
	PY54	*Yersinia*	*Vibirio-76%* / none
Temperate 4	HK97	*Escherichia*	*Pseudomonas* 78% / none
	mEp235	*Escherichia*	*Pseudomonas-80%* / none
Temperate 5	ES18	*Salmonella*	*Pseudomonas*-69% / *Pseudomonas* phage ø297-69%
Temperate 6	Gifsy-2	*Salmonella*	*Photobacterium*-57% / *Pseudomonas* phage F10-55%
Temperate 7	BP-4795	*Escherichia*	None, none
Temperate 8	SfV	*Shigella*	*Aggregatibacter*-64% / none
	SfI	*Shigella*	*Vibrio-84%* / none
Temperate 9	P22	*Salmonella*	None / none
	Sf6	*Shigella*	None / none
Temperate 10	APSE-1	*Hamiltonella*	None / none
Temperate 11	933W	*Escherichia*	*Photobacterium*-63%, none

Table S1 cont.

Cluster	Prototype phage	Prototype phage host	Bacterial genera / phage with >50% MCP matches[1] outside *Enterobacteriaceae*
Temperate 12	HK639	*Escherichia*	*Castellaniella*-55% / none
Temperate 13	øES15	*Cronobacter*	None / none
Temperate 14	HS2*	*Sodalis*	*Acinetobacter*-77% / none
Temperate 15	ENT47670	*Cronobacter*	*Vibrio*-77% / none
Temperate 16	ZF40	*Pectobacterium*	None / none
Temperate 17	øEt88	*Erwinia*	None / *Xanthomonas* phage DIBBI-53%
Temperate 18	ε15	*Salmonella*	*Pseudomonas*-65% / none
Temperate 19	P1	*Escherichia*	None / none
Supercluster	P2		
Temperate 20	P2	*Escherichia*	*Burkholderia*-55% / *Burkholderia* phage-55%
Temperate 21	ESSI-2	*Cronobacter*	*Vibrio*-59% / *Vibrio* phage-53%
Temperate 22	Mu	*Escherichia*	*Marinomonas*-62% / *Pseudomonas* phage DMS3-51%
Temperate 23	SSU5	*Salmonella*	*Acinetobacter*-60% / none
Temperate 24	PEp14	*Erwinia*	*Rhizobium*-66% / *Sinorhizobium* phage PBC5 phage-66%
Temperate 25	GF-2	*Edwardsiella*	None / none
Temperate 26	SEN34	*Salmonella*	*Pseudomonas*-59% / none

1. Percent MCP identity as reported by BLASTp
   Red text highlights prophages in other bacterial families with MCPs >75% identical to MCPs of a temperate *Enterobacteriaceae* phage cluster.
   Blue text highlights prophages in other bacterial families with MCPs between 50% and 70% identical to MCPs of a lytic *Enterobacteriaceae* phage cluster.

2. Several phages in the EMCL-117-like phage group, for example phage øVader, have recently been reported to grow lytically on both *E. coli* and *Pseudomonas aeruginosa* (Malki *et al.*, 2015).

Table S2
P2 Supercluster Phages and Prophages

A *Enterobacteriaceae* P2-like phages in this study

Cluster	Phage	Host species	Accession number	Genome orientation	Genome sequence publication
Temperate 20	**P2**	*Escherichia coli*	AF063097	P2	numerous publications
	Wø	*Escherichia coli*	AY135739	P2	–
	L-413C	*Yersinia pestis*	AY251033	P2	Virology 372:85
	øAA91-ss	*Escherichia coli*	KF322032	P2	JVirol 87:13665
	pro147	*Escherichia coli*	KR073660	P2	–
	pro483	*Escherichia coli*	KR073661	P2	–
	186	*Escherichia coli*	U32222	P2	numerous publications
	PsP3	*Salmonella enterica*	NC_005340	P2	
	FSL-SP-004	*Salmonella enterica*	KC139521	P2	BMCgenomics 14:481
	ENT90	*Erwinia amylovora*	HQ110084	P2	–
	SopEø	*Salmonella enterica*	AY319521	P2	JBact 185:5182
	Fels-2	*Salmonella enterica*	NC_010463	P2	Nature 413:852
	RE-2010 (ELøS)	*Salmonella enterica*	HM770079	P2	ApplEnvMicro 78:1785
	SEN1	*Salmonella enterica*	KT630644	P2	–
	SEN4	*Salmonella enterica*	KT630645	P2	–
	SEN5	*Salmonella enterica*	KT630646	P2	–
	SEN8	*Salmonella enterica*	KT630647	P2	–
	P88	*Escherichia coli*	NC_026014	P2	–
Temperate 21	**ESSI-2**	*Cronobacter sakazakii*	HQ110083	HP1	Arc hVirol 156:721

Table S2
P2 Supercluster Phages and Prophages

B P2-like *Enterobacteriaceae* prophage panel used in this study

	Prophage	Host species and strain	MCP locus_tag
	AgonaP2-A	*Salmonella enterica* (Agona73.H.09)	SA73_2208
	CedP2-1	*Cedecea davisae* DSM 4568	HMPREF0201_03602
	CronoP2-B	*Cronobacter sakazakii* SP291	CSSP291_11880
	DickP2-B	*Dickeya zeae* EC1	W909_12715
	Eco541P2-A	*Escherichia coli* 541-15	EC54115_23648
	EcoC3P2-A*	*Escherichia coli* 2-011-08_S4_C3	AD24_4183
	EcoP2-CVM-A	*Escherichia coli* CVM N36225PS	PU46_04800
	EcoP2-E22-A	*Escherichia coli* E22	EcE22_1442
	EcoP2-KTE10-A	*Escherichia coli* KTE10	WCM_01332
	EcoP2-KTE234-A	*Eschericha coli* KTE234	A193_03379
	EdwardP2-A	*Edwardsiella tarda* ATCC 15947	ET1_14_00560
	EdwardP2-B	*Edwardsiella piscicida* C07-87	ETAC_02100
	EntP2-A	*Enterobacter cloacae* ATCC 13047	ECL_02851
	ErwinP2-A	*Erwinia billingiae Eb661*	EbC_01060
	HafP2-A	*Hafnia alvei* FB1	AT03_17965
	KlebP2-A	*Klebsiella pneumoniae* UHKPC57	H237_1271
	MorgP2-A	*Morganella morganii* KT	MU9_2040
	PantP2-A	*Pantoea rwandensis* ND04	LH22_RS05065
	PectoP2-C	*Pectobacterium caratovorum* PC1	PC1_2643
	ProdP2-A	*Pantoea rodasii* ND03	QU24_21505
	ProvP2-B	*Providencia rustigianii* DSM 4541	PROVRUST_04967
	RahnP2-B	*Rahnella aquatilis* HX2	Q7S_07135
	RauolP2-A	*Raoultella ornithinolytica* TNT	DF41_23110
	SalP2-C	*Salmonella enterica* Typhimurim YUHS 05-78	SE14_01921
	SerrP2-C	*Serratia odorifer* DSM 4582	HMPREF0758_1209
	ThompP2-A	*Salmonella enterica* (Thompson CFSAN000736)	SEETH391_21320
	XenoP2-A	*Xenorhabdus doucetiae* FRM16	XDD1_2854
	YersP2-A	*Yersinia pseudotuberculosis* YPIII	YPK_0899
	YersP2-B	*Yersinia enterocolitica* Y11	Y11_25391
	YokP2-A	*Yokenella regensburgei* ATCC 43003	HMPREF0880_02763

Table S2
P2 Supercluster Phages and Prophages

C P2-like phages outside the *Enterobacteriaceae*

	Phage	Host species	Host Family	Class	MCP locus_tag	Genome sequence publication
	BEK	*Burkholderia pseudomallei*	*Burkholderiaceae*	Beta-Proteobacteria	CP008753	–
	KL3	*Burkholderia cepacia*	*Burkholderiaceae*	Beta-Proteobacteria	GU911304	BMCGenomics 11:599
	KS14	*Burkholderia cenocepacia*	*Burkholderiaceae*	Beta-Proteobacteria	NC_015273	BMCGenomics 11:599
	KS5	*Burkholderia cenocepacia*	*Burkholderiaceae*	Beta-Proteobacteria	NC_015265	BMCGenomics 11:599
	ø52237	*Burkholderia pseudomallei*	*Burkholderiaceae*	Beta-Proteobacteria	NC_007145	–
	øE12-2	*Burkholderia pseudomallei*	*Burkholderiaceae*	Beta-Proteobacteria	NC_009236	–
	øE202	*Burkholderia thailandensis*	*Burkholderiaceae*	Beta-Proteobacteria	CP000623	–
	øRSA1	*Ralstonia solanacearum*	*Ralstoniaceae*	Beta-Proteobacteria	AB276040	JBact 190:143
	øX216	*Burkholderia pseudomallei*	*Burkholderiaceae*	Beta-Proteobacteria	JX681814	BMC Microbiol 12:289
	RSY1	*Ralstonia solanacearum*	*Ralstoniaceae*	Beta-Proteobacteria	AB981169	–
	ST79	*Burkholderia pseudomallei*	*Burkholderiaceae*	Beta-Proteobacteria	NC_021343	–
	535AP1	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	KP137432	BMCGenomeics15:175
	587AP1	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	KP137434	BMCGenomeics15:175
	1127AP1	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	KP137436	BMCGenomeics15:175
	1152AP1	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	NC_021778	BMCGenomeics15:175
	2256AP1	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	KP137438	BMCGenomeics15:175
	BAA410	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	DQ426905	Virology 350:79
	F108	*Pasteurella multocida*	*Pasteurellae*	Gamma-Proteobacteria	DQ114220	AEM 72:3154
	HP1	*Haemophilus influenzae*	*Pasteurellae*	Gamma-Proteobacteria	NC_001697	NucliecAcidRes 24:685
	HP2	*Haemophilus influenzae*	*Pasteurellae*	Gamma-Proteobacteria	AY027935	JBact 184:6893
	K139	*Vibrio cholerae*	*Vibrionaceae*	Gamma-Proteobacteria	NC_003313	JBact 184:6592
	Kappa	*Vibrio cholerae*	*Vibrionaceae*	Gamma-Proteobacteria	NC_010275	–
	øCTX	*Psuedomonas aerugunosa*	*Pseudomonadaceae*	Gamma-Proteobacteria	AB008550	MolMicro 31:399
	ø3	*Psuedomonas aerugunosa*	*Pseudomonadaceae*	Gamma-Proteobacteria	KT887559	–
	øMHaA1-PHL101	*Mannheimia haemolytica*	*Pasteurellae*	Gamma-Proteobacteria	NC_008201	Virology 350:79
	øO18P	*Aeromonas media*	*Aeromonadaceae*	Gamma-Proteobacteria	NC_009542	Virology 373:25
	PV94	*Vibrio vulnificus*	*Vibrionaceae*	Gamma-Proteobacteria	HG803181	PLoS One 9:e94707
	Smp131	*Stenotrophomonas maltophilia*	*Xanthomonaceae*	Gamma-Proteobacteria	JQ809663	BMC Microbiol 14:17
	VPUSM 8	*Vibrio cholerae*	*Vibrionaceae*	Gamma-Proteobacteria	NC_022747	–

Table S2
P2 Supercluster Phages and Prophages

D P2-like prophages outside the *Enterobacteriaceae*

Prophage	Host species and strain	Host Family	Class	MCP locus_tag
BrevuP2-A	*Brevundimonas diminuta* ATCC 11568	*Caudobactereacae*	Alpha-Proteobacteria	BDIM_14580
NovoP2-A	*Novosphingobium pentaromativorans* US6-1	*Sphingomonadaceae*	Alpha-Proteobacteria	JI59_04365
ZymoP2-A	*Zymomonas mobilis* ZM4	*Sphingomonadaceae*	Alpha-Proteobacteria	ZZM4_0034
AchromoP2-A	*Achromobacter sp.* LC458	*Alcaligenacae*	Beta-Proteobacteria	YH64_11480
BurkP2-A	*Burkholderia thailandensis* MSMB121	*Burkholderiaceae*	Beta-Proteobacteria	BTI_1970
BurkP2-B	*Burkholderia cenocepacia* MC0-3	*Burkholderiaceae*	Beta-Proteobacteria	Bcenmc03_0205
ColliP2-A	*Collimonas fungivorans* Ter331	*Oxalobacteracae*	Beta-Proteobacteria	CFU_0996
ComaP2-A	*Comamonas testosteroni* TK102	*Comamonadaceae*	Beta-Proteobacteria	O987_RS22905
CupP2-A	*Cupriavidus taiwanensis* LMG19424	*Burkholderiaceae*	Beta-Proteobacteria	RALTA_A0929
KingP2-A	*Kingella oralis* ATCC 51147	*Neisseriaceae*	Beta-Proteobacteria	GCWU000324_01254
NeissP2-A	*Neisseria meningitidis* alpha710	*Neisseriaceae*	Beta-Proteobacteria	NMBB_1121
PolarP2-A	*Polaromonas naphthalenivorans* CJ2	*Comamonadaceae*	Beta-Proteobacteria	Pnap_3291
DesulfP2-A	*Desulfovibrio desulfuricans* ATCC 27774	*Desulfovibrionaceae*	Delta-Proteobacteria	Ddes_0241
DesulfP2-B	*Desulfovibrio vulgaris* DP4	*Desulfovibrionaceae*	Delta-Proteobacteria	Dvul_2765
AcinP2-A	*Acinetobacter baumannii* AYE	*Pseudomonales; Moraxellaceae*	Gamma-Proteobacteria	ABAYE0571
AcinP2-B	*Acinetobacter ursingii* NIPH 706	*Pseudomonales; Moraxellaceae*	Gamma-Proteobacteria	F943_01500
AeroP2-A	*Aeromonas hydrophila* AL06-06	*Aeromonacea*	Gamma-Proteobacteria	RY45_19020
AeroP2-B	*Aeromonas salmonicida* 2009-144K3	*Aeromonacea*	Gamma-Proteobacteria	NX85_17375
AzoP2-A	*Azotobacter vinelandii* CA	*Pseudomonales; Pseudomonadaceae*	Gamma-Proteobacteria	AvCA_37590
CardP2-A	*Cardiobacterium hominis* ATCC 15826	*Cardiobacteriaceae*	Gamma-Proteobacteria	HMPREF0198_2128
ChromoP2-A	*Chromohalobacter salexigens* DSM 3043	*Halomonadaceae*	Gamma-Proteobacteria	Csal_1397
EndoP2-A	*Endozoicomonas montiporae* LMG 24815	*Hahellaceae*	Gamma-Proteobacteria	GZ77_07230
GalliP2-A	*Gallibacterium anatis* UMN179	*Pasteurellae*	Gamma-Proteobacteria	UMN179_00236
HahP2-A	*Hahella chejuensis* KCTC 2396	*Hahellaceae*	Gamma-Proteobacteria	HCH_00394
HaloP2-A	*Halomonas sp.* PBN3	*Halomonadaceae*	Gamma-Proteobacteria	Q671_08135
MannP2-A	*Mannheimia haemolytica* PKL10	*Pasteurellae*	Gamma-Proteobacteria	AK33_08830
MarmoP2-A	*Marinomonas sp.* MWYL1	*Oceanospirillaceae*	Gamma-Proteobacteria	Mmwyl1_0585
MarinP2-A	*Marinobacter nanhaiticus* D15-8W	*Alteromonadaceae; Alteromonadales*	Gamma-Proteobacteria	J057_04576
MortP2-A	*Moritella viscosa* MVIS1	*Alteromonadaceae; Alteromonadales*	Gamma-Proteobacteria	MVIS_2005
OleiP2-A	*Oleispira antarctica* RB-8	*Oceanospirillaceae*	Gamma-Proteobacteria	OLEAN_C08680
PastP2-A	*Pasteurella multocida* P1062	*Pasteurellae*	Gamma-Proteobacteria	P1062_0211455
PseudoalteroP2-A	*Pseudoalteromonas arctica* A 37-1-2	*Pseudoalteromonadaceae; Alteromonadales*	Gamma-Proteobacteria	PARC_05148
PseudoalteroP2-B	*Pseudoalteromonas atlantica* A 37-1-2	*Pseudoalteromonadaceae; Alteromonadales*	Gamma-Proteobacteria	Patl_0697
PseudoP2	*Pseudomonas aeruginosa* BL13	*Pseudomonales; Pseudomonadaceae*	Gamma-Proteobacteria	Q067_06185
SheP2-B	*Shewanella baltica* OS195	*Shewanellaceae; Alteromonadales*	Gamma-Proteobacteria	Sbal195_2914
StentoP2-A	*Stenotrophomonas maltophilia* D457	*Xanthomonaceae*	Gamma-Proteobacteria	SMD_3273
ViirioP2-B	*Vibrio cholerae* AM-19226	*Vibrionaceae*	Gamma-Proteobacteria	A33_2125
VibrioP2-A	*Vibrio fischeri* ES114	*Vibrionaceae*	Gamma-Proteobacteria	VF2012
XanthoP2-A	*Xanthomonas campestris*	*Xanthomonaceae*	Gamma-Proteobacteria	XCCB1459_1200

Table S3
P22-like and P2-like prophages in Salmonella genome sequences

Salmonella enterica subsp. *enterica*	O Group†	Serovar	Number of genome sequences	P22 MCP	CUS-3 MCP	Sf6 MCP	Percentage of genomes with P22-like prophage	P22 Tailspike type††	Number of tailspike genes†	P2-like Fels-2 MCP >50%	P2-like P88 MCP >50%	P2-like ESSI-2 MCP >45%
	O:1,3,19	Senftenberg	10	0	0	0	0.0%		0	5	0	2
	O:2	Paratyphi A	139	130	0	0	93.5%	Ia	130	107	32	124
	O:3,10	Anatum	5	2	0	0	40.0%	III / IIb	1 / 1	1	0	1
	O:3,10	Give	2	1	0	0	50.0%	Ia	1	1	0	2
	O:3,10	Muenster	4	2	0	0	50.0%	III / Vb	1 / 1	8	0	0
	O:3,10	Nchanga	2	2	0	0	100.0%	III	2	0	0	2
	O:3,10	Uganda	1	1	0	0	100.0%	III	1	0	0	0
	O:3,10	Weltevreden	2	1	0	0	50.0%	III	1	2	0	1
	O:3.10	London	1	0	0	0	0.0%		0	1	0	1
	O:3.10	Meleagridis	1	0	0	0	0.0%		0	0	0	1
	O:4	Abony	1	1	0	0	100.0%	Ia	1	0	0	0
	O:4	Agona	114	11	1	0	10.5%	Ia / Ib	8 / 4	111	1	83
	O:4	Brandenburg	1	1	0	0	100.0%	Ia	1	1	0	1
	O:4	Bredeney	1	0	0	0	0.0%		0	0	0	0
	O:4	Chester	1	0	0	0	0.0%		0	1	0	0
	O:4	Derby	2	0	0	0	0.0%		0	0	0	0
	O:4	Heidelberg	103	103	0	0	100.0%	Ia	103	38	0	102
	O:4	Indiana	1	0	0	0	0.0%		0	1	0	0
	O:4	Paratyphi B	10	1	0	0	10.0%	Ia	1	4	2	3
	O:4	Saintpaul	6	1	0	0	16.7%	Ia	1	1	0	2
	O:4	Schwarzengrund	2	0	0	2	100.0%	XI	1	1	0	1
	O:4	Sloterdijk	1	0	0	0	0.0%		0	0	0	0
	O:4	Stanley	1	0	0	0	0.0%		0	1	0	1
	O:4	Stanleyville	1	0	0	0	0.0%		0	0	0	0
	O:4	Typhimurium	445	388	0	0	87.2%	Ia / Ib / Vc	381 / 1 / 1	30	3	26
	O:6,7	Bareilly	103	12	0	0	11.7%	Vb	12	6	0	10
	O:6,7	Braenderup	2	0	0	0	0.0%		0	4	0	0
	O:6,7	Choleraesuis	5	5	0	0	100.0%	Va	5	0	0	0
	O:6,7	Hartford	1	0	0	0	0.0%		0	1	0	0
	O:6,7	Infantis	5	1	0	0	20.0%	Vb	1	2	0	5
	O:6,7	Livingstone	1	1	0	0	100.0%	Vc	1	1	0	0
	O:6,7	Lubbock	2	2	0	0	100.0%	Va	2	5	0	2
	O:6,7	Mbandaka	1	0	0	0	0.0%	Vc	1	1	0	0
	O:6,7/O:54*	Montevideo	55	4	0	0	7.3%	Vb / Ia	3 / 1	51	0	4
	O:6,7	Norwich	1	0	0	0	0.0%		0	1	0	0

Table S3
P22-like and P2-like prophages in Salmonella genome sequences

Salmonella enterica subsp. *enterica*	O Group†	Serovar	Number of Salmonella genome sequences	P22 MCP	CUS-3 MCP	Sf6 MCP	Percentage of genomes with P22-like prophage	P22 Tailspike type††	Number of tailspike genes†	Fels-2 MCP >50%	P88 MCP >50%	ESSI-2 MCP >45%
	O:6,7	Ohio	1	1	0	0	100.0%	Vc	1	0	0	0
	O:6,7	Oranienburg	3	0	0	0	0.0%		0	3	0	0
	O:6,7	Paratyphi C	1	0	1	0	100.0%	Vb	1	0	0	1
	O:6,7	Rissen	1	1	0	0	100.0%	Vc	1	2	0	0
	O:6,7	Tennessee	3	2	0	0	66.7%	Va / Vc	1 / 1	1	0	2
	O:6,7	Thompson	2	0	0	0	0.0%		0	0	0	0
	O:6,7	Virchow	3	0	0	0	0.0%		0	1	0	1
	O:8	Albany	1	1	0	0	100.0%	Vlia	1	0	0	0
	O:8	Bovismorbificans	2	0	0	0	0.0%		0	0	0	1
	O:8	Hadar	5	2	1	0	60.0%	VIIc	3	0	0	0
	O:8	Kentucky	15	0	0	0	0.0%		0	0	0	0
	O:8	Litchfield	1	0	0	0	0.0%		0	0	0	0
	O:8	Manhattan	2	0	0	0	0.0%		0	1	0	0
	O:8	Muenchen	4	0	0	0	0.0%		0	3	0	2
	O:8	Newport	47	5	3	2	21.3%	VIIa / VIIb / VIId	7 / 3 / 6	23	2	10
	O:8	Tallahassee	1	0	1	0	100.0%	VIIc	1	0	0	0
	O:9	Berta	3	0	0	0	0.0%		0	3	0	0
	O:9	Dublin	10	10	0	0	100.0%	Ia	10	10	0	0
	O:9	Eastbourne	1	0	0	0	0.0%		0	1	0	0
	O:9	Enteritidis	318	1	0	0	0.3%	III	1	208	0	22
	O:9	Gallinarum/Pullorum	12	1	0	0	8.3%	Ia	1	5	0	0
	O:9	Javiana	5	1	0	0	20.0%	Ia	1	5	0	0
	O:9	Miami	1	0	0	0	0.0%	Ia	1	0	0	0
	O:9	Napoli	1	0	0	0	0.0%		0	0	1	0
	O:9	Panama	1	0	0	0	0.0%		0	1	0	0
	O:9	Typhi	1755	0	0	0	0.0%		0	3526	68	24
	O:9	9,12:l,v:	1	1	0	0	100.0%	Ib	1	0	0	2
	O:9,46	Baildon	1	0	0	0	0.0%		0	0	0	0
	O:9,46	Ouakam	1	1	0	0	100.0%		0	0	0	0
	O:11	Abaetetuba	1	0	0	0	0.0%		0	0	0	0
	O:11	Rubislaw	2	1	0	0	50.0%	XIV	1	0	2	0
	O:13	Cubana	4	0	0	0	0.0%		0	2	0	0
	O:13	Havana	1	0	0	0	0.0%		0	1	0	1
	O:13	Mishmarhaemek	1	0	0	0	0.0%		0	1	1	0

Table S3
P22-like and P2-like prophages in Salmonella genome sequences

Salmonella enterica subsp. *enterica*	O Group†	Serovar	Number of genome sequences	P22 MCP	CUS-3 MCP	Sf6 MCP	Percentage of genomes with P22-like prophage	P22 Tailspike type††	Number of tailspike genes†	Fels-2 MCP >50%	P88 MCP >50%	ESSI-2 MCP >45%
	O:13	Mississippi	1	1	0	0	100.0%	VIII	1	0	0	0
	O:13	Poona	1	1	0	0	100.0%	VIII	1	0	0	1
	O:13	Worthington	1	0	0	0	0.0%		0	0	0	0
	O:16	Gaminara	2	1	0	0	50.0%	VI	1	2	1	0
	O:18	Cerro	21	21	0	0	100.0%	IV	21	0	0	0
	O:21	Minnesota	2	1	0	0	50.0%	X	1	0	0	0
	O:28	Pomona	1	0	0	0	0.0%		0	0	0	0
	O:30	Aqua	1	1	0	0	100.0%	Ia	1	0	0	0
	O:30	Soerenga	1	0	0	1	100.0%	XII	1	1	0	0
	O:30	Urbana	2	0	0	0	0.0%		0	2	0	1
	O:35	Adelaide	1	0	0	0	0.0%		0	1	0	0
	O:35	Alachua	1	0	0	0	0.0%		0	1	0	0
	O:38	Inverness	2	0	0	0	0.0%		0	1	0	0
	O:39	Namur	1	0	0	0	0.0%		0	0	0	0
	O:39	Wandsworth	1	1	0	0	100.0%	XIII	1	0	1	0
	O:40	Johannesburg	1	1	0	0	100.0%	IX	1	1	0	1
Salmonella enterica subsp. Arizonae	O:62		1	1	0	0	100.0%	XV	1	0	0	0
Salmonella enterica subsp. Diarizonae			2	0	0	0	0.0%		0	1	0	0
Salmonella enterica subsp. Houtenae			4	1	0	0	25.0%	XVI / VI	1 / 1	0	0	2
Salmonella enterica subsp. indica			2	0	0	0	0.0%		0	2	0	0
Salmonella enterica subsp. Salamae			2	1	0	0	50.0%	XVII	1	0	0	0
Salmonella bongeri			3	0	0	0	0.0%		0	2	1	1
Total			**3298**	**729**	**7**	**5**			**743****	**4197**	**115**	**446**
		Total P22-like MCPs		**741**				**Total P2-like prophages**		**4758**		
		Total P22-like prophages #		**744**								

Table S3 footnotes

The table was completed on July 31, 2015. GenBank entries with only "WP_xxxxxx" accession numbers are not included since the strain names and nucleic acid context cannot be determined.

† O antigen groups are from Grimont, P., Weill, F. 2007. Antigenic Formulae of the *Salmonella* Serovars, (9th ed.) Paris: WHO Collaborating Center for Reference and Research on *Salmonella*, Institut Pasteur. (Available at: http://www.pasteur.fr/sante/clre/cadrecnr/salmoms/ WKLM_En.pdf)

†† Where several types of P22-like tailspikes are present in different prophages they are separated by slashes (/). Red text highlights the anomalous tailspikes discussed in the text.

* O:6,7 antigen is expressed in Montevideo serotype cells when plasmid controlling O:54 antigen is not present.

** The total numbers of MCPs and tailspikes are not the same. This is because MCP or tailspike genes are occasionally missing from the translated annotation due to their falling between contigs, having been deleted from the prophage or not being correctly annotated. In addition several Newport prophages have two tailspike genes.

\# 741 MCP genes and 3 tailspike genes whose associated MCP genes are missing or not annotated.

Table S4
P22-like phages and prophages used this study

A	P22-like authentic phages			
Phage	**Host**	**Accession number**	**genome orientation**	**MCP type**
22	*Salmonella enterica (Typhimurium)*	KR296686	P22	P22
25	*Salmonella enterica (Typhimurium)*	KR296687	P22	P22
34	*Salmonella enterica (Typhimurium)*	KR296688	P22	P22
APSE-2	*Hamiltonella defensa**	EU794049	APSE-1	Sf6
ASPE-1	*Hamiltonella defensa**	AF157835	APSE-1	Sf6
CUS-3	*Escherichia coli*	CP000711	Sf6	CUS-3
Emek	*Salmonella enterica (Haardt)*	JQ806763	P22	P22
epsilon34	*Salmonella enterica (Anatum)*	EU570103	P22	P22
g341	*Salmonella enterica (Anatum)*	NC_013059	P22	P22
HK620	*Escherichia coli*	AF335538	Sf6	Sf6
IME10	*Escherichia coli*	JN600960	P22	P22
L	*Salmonella enterica (Typhimurium)*	AY795968†	P22	P22
ø20	*Salmonella enterica (Typhimurium)*	GQ422450	P22	P22
ø75	*Salmonella enterica (Typhimurium)*	GQ422451	P22	P22
P22	*Salmonella enterica (Typhimurium)*	BK000583	P22	P22
SE1	*Salmonella enterica (Typhimurium)*	NC_011802	P22	P22
SEN22	*Salmonella enterica (Diarizonae)*	KT630648	P22	P22
SETP1	*Salmonella enterica (Eneritidis)*	EF151184#	–	–
Sf101	*Shigella flexneri*	KJ832078	Sf6	CUS-3
Sf6	*Shigella flexneri*	AF547987	Sf6	Sf6
SPC-P1	*Salmonella enterica (TyphiC)*	CP000857	Sf6	CUS-3
SPN9CC	*Salmonella enterica (Typhimurium)*	JF900176	P22	P22
ST104	*Salmonella enterica (Typhimurium)*	AB102868	P22	P22
ST160	*Salmonella enterica (Typhimurium)*	NC_014900	P22	P22
ST64T	*Salmonella enterica (Typhimurium)*	AY052766	P22	P22
	* virion-like particles not yet shown to be infectious † our unpublished sequence # partial sequence			

Table S4
P22-like phages and prophages used this study

B	*Enterobacteriaceae* P22-like prophages			
	Prophage Host	**MCP locus_tag**	**genome orientation**	**MCP type**
	*Arsenophonus nasoniae**	ARN_26010	–	Sf6
	Citrobacter sp. BIDMC108	SK40_00812	Sf6	P22
	Cronobacter sakazakii 696	BN128_2797	P22	P22
	Cronobacter malonaticus†	WP_032971372*	–	–
	Enterobacter sp.GN02548*	ABR33_05860	–	CUS-3
	Enterobacter aerogenes UCI47	P850_03405	APSE-1	P22
	Erwinia tracheiphila BufGfH	SY86_19110	APSE-1 early right??	P22
	Escherichia coli B7A	EcB7A_4179	Sf6	Sf6
	Escherichia coli CVM N38428PS	PU66_07355	Sf6	CUS-3
	Escherichia coli W	EschWDRAFT_1163	Sf6	P22
	Klebsiella pneumoniae SB3432	KPR_2793	Sf6	CUS-3
	Kluyvera intermedia CAV1151	AB182_19130	P22	P22
	Leclercia adecarboxylata†	WP_039031990*	–	–
	Morganella morganii KT	MU9_2902	P22	P22
	Pantoea sp. PSNIH2	PSNIH2_13380	P22	P22
	Pectobacterium carotovorum BCS2	RC83_13195	P22	P22
	Pectobacterium carotovorum LMG21371	KS44_06225	P22	CUS-3
	Pluralibacter gergoviae DL84A27	ABW09_23385	–	P22
	Proteus mirabilis ATCC7002	DR94_520	Prot1	P22
	Providencia rettgeri DSM 1131	PROVRETT_06010	P22	Sf6
	Providencia rettgeri †	WP_048606460*	–	–
	Salmonella enterica CVM19536	SEEN536_00900	–	Sf6
	Serratia marcescens VGH107	F518_22685	P22	Sf6
	Serratia plymuthica AS9	SerAS9_2688	APSE-1 (ø80 int org)	P22
	Shigella flexneri 2002017	SFxv_0342	P22	P22
	Shimwella blattae DSM 4481	EBL_c21750	P22	Sf6
	Tatumella morbirosei LMG23360	HA49_08570	P22	P22
	Xenorhabdus bovienii CS03	XBW1_3034	Prot1	P22
	Yersinia aleksiciae 159	ACZ76_14855	P22	Sf6
	*Yersinia kristenseni** ATCC33639	DJ57_2548	P22	P22
	* Rather short contig so not included in genome plot Figure 6	* Accession number (Locus_tag not available)	"–" inidicates contig too short to determine oreintation	
	† Strain names and whole prophage sequence not available			

Table S4
P22-like phages and prophages used this study

C	*Salmonella* P22-like prophages			
	Prophage Host	**MCP locus_tag**		
	Salmonella enterica serovar Paratyphi AY09620	YNSPA_2311		
	Salmonella enterica serovar Heidelberg SL476	SeHA_C0403		
	Salmonella enterica serovar Agona 3432613	SEEA2613_05333		
	Salmonella enterica serovar Typhimurium U288	STU288_12845		
	Salmonella enterica serovar Derby CVMN51289	AEX05_09315		
	Salmonella enterica serovar Albany ATCC51960	SEEA1960_00245		
	Salmonella enterica serovar Hadar 05P066	SeH_A0879		
	Salmonella enterica serovar Newport CVM19536	SEEN536_00900		
	Salmonella enterica serovar Choleraesuis SCB-67	SCH_0357		
	Salmonella enterica serovar Bareilly CFSAN00219	CFSAN00219_		
	Salmonella enterica serovar Montevideo 8387	SEEM8387_07975		
	Salmonella enterica serovar Anatum USMARC-1735	AW50_6230		
	Salmonella enterica serovar Cerro CFSAN001588	CFSAN001588_015920		
	Salmonella enterica serovar Dublin SL1438	SEEDSL_005061		
	Salmonella enterica subspecies *Arizonae* RSK2980	SARI_02647		
D	Paratyphi A P22-like			
	Prophage Host	**MCP locus_tag**	**Isolation location**	
	Salmonella enterica serovar Paratyphi A (no strain name reported)	PA045_3010	–	
	Salmonella enterica serovar Paratyphi A AYN09620	YNSPA_2311	Yunnan, China	
	Salmonella enterica serovar Paratyphi A AKU_12601	SSPA2232e	Pakistan	
	Salmonella enterica serovar Paratyphi A CMCC50093	ACN89_12045	Beijing, China	
	Salmonella enterica serovar Paratyphi A GZ9A00052	GZSPA_2297	Guizhou, China	
	Salmonella enterica serovar Paratyphi A JX05-19	JXSPA_2296	Jiangxi, China	
	Salmonella enterica serovar Paratyphi A CMCC50503	LW89_14285	Jaingsu, China	
	Salmonella enterica serovar Paratyphi A (no strain reported)	PA136_4071	–	
	Salmonella enterica serovar Paratyphi ACMCC50973	IT63_14940	Jaingsu, China	

Table S4
P22-like phages and prophages used this study

E	Heidelberg P22-like prophages			
	Prophage Host	**MCP locus_tag**	**Isolation location**	
	Salmonella enterica serovar Heidelberg 82-2052	SEEH2052_04713	Maine, USA	
	Salmonella enterica serovar Heidelberg ABB07-SB3031	TJ50_03995	British Columbia, CA	
	Salmonella enterica serovar Heidelberg B182	SU5_01004	France	
	Salmonella enterica serovar Heidelberg CVM N45397	AEU76_20940	Minnesota, USA	
	Salmonella enterica serovar Heidelberg N26457	SEEH6457_20825	Colorado, USA	
	Salmonella enterica serovar Heidelberg SL476	SeHA_C0403	Minnesota, USA	
	Salmonella enterica serovar Heidelberg SL486	SeHB_A0358	Georgia, USA	
F	Newport P22-like prophages			
	Prophage Host	**MCP locus_tag**	**Isolation location**	
	Salmonella enterica serovar Newport CVM 19470	SEEN470_21064	Vietnam	
	Salmonella enterica serovar Newport CVM 19536	SEEN536_00900	Vietnam	
	Salmonella enterica serovar Newport CVM 22425	SEEN425_002235	Arizona, USA	
	Salmonella enterica serovar Newport CVM 22462	SEEN462_0001185	Arizona, USA	
	Salmonella enterica serovar Newport CVM 4176	SEEN176_15035	California, USA	
	Salmonella enterica serovar Newport Henan_3	G208_21144	Henan, China	
	Salmonella enterica serovar Newport JS09102	G209_24390	Jiangsu, China	
	Salmonella enterica serovar Newport SH111077	G206_16997	Shanghai, China	
	Salmonella enterica serovar Newport Shandong_3	G207_20579	Shandong, China	
	Salmonella enterica serovar Newport USMARC-S3124.1	SN31241_18260	USA	
G	Typhimurium prophages			
	Prophage Host	**MCP locus_tag**	**Isolation location**	
	Salmonella enterica serovar Typhimujrium 34502	SEET4502_15275	–	
	Salmonella enterica serovar Typhimujrium ABB1162-2	TJ39_05580	British Columbia, CA	
	Salmonella enterica serovar Typhimujrium CFSAN001921	CFSAN001921_15475	New York, USA	
	Salmonella enterica serovar Typhimujrium CVM N51254	AEW79_08460	Connecticut, USA	
	Salmonella enterica serovar Typhimujrium D23580	STMMW_03811	Malawi	
	Salmonella enterica serovar Typhimujrium DT104	DT104_03541	–	
	Salmonella enterica serovar Typhimujrium DT104 EWH6	ERS198263_02378	United Kingdom	
	Salmonella enterica serovar Typhimujrium FORC_015	ZV79_2330	South Korea	
	Salmonella enterica serovar Typhimujrium L-3553	STL3553_c03780	–	
	Salmonella enterica serovar Typhimujrium TN061786	SEE_01911	–	
	Salmonella enterica serovar Typhimujrium U288	STU288_12845	–	
	Salmonella enterica serovar Typhimujrium var.5 CVM N43832	AEU38_10780	New York, USA	

Table S4
P22-like phages and prophages used this study

H	Prophage P22-like tailspikes			
	***Salmonella enterica* serovar and strain**	**Tailspike locus_tag**	**Tailspike type**	
	9,12:l,v:- str. 94293	N925_10475	Ib	
	Abony 0014	SEEA0014_11145	Ia	
	Agona 648586-1	SEEA5861_16245	Ib	
	Albany_ATCC 51960(A)	SEEA1960_00180	VIIa	
	Anatum USDA-ARS-USMARC-1735	AW50_6260	Iib	
	AnatumCDC06-0532	AW57_13240	III	
	Aqua NVSL2001	I654_10070	Ia	
	Arizonae serovar 62:z4,z23:- str. RSK2980	SARI_02638	XIV	
	Bareilly CFSAN000183	SEEB0183_09067	Vb	
	Brandenburg CFSAN024765	QS22_03720	Ia	
	Cerro 87	NI35_4331	IV	
	Choleraesuis SC-B67	SCH_0367	Va	
	Dublin SL1438	SEEDSL_005126	Ia	
	Enteritidis 3402	SEEE3402_02210	III	
	Gaminara ATCC BAA-711	SEEGA711_14315	VI	
	Give var. 15 CFSAN004343	CFSAN004343_03735	Ia	
	Hadar ABB1048-1	TJ45_04405	VIIc	
	Heidelberg SL476	SeHA_C0413	Ia	
	Houtenae ATCC BAA-1581	SEHO0A_00503	XV	
	Houtenae serovar 16:z4,z32:-- str. RKS3027	D088_680001	VI	
	Infantis CFSAN024780	QS05_16285	Vb	
	Javiana PRS_2010_0720	SEEJ0720_04071	Ia	
	Johannesburg S5-703	LTSEJOH_2243	VIII	
	Livingstone CKY-S4	VN13_12135	Vc	
	Lubbock 10TTU468	TR76_09235	Va	
	Miami 1923	SEEM1923_08240	Ia	
	Minnesota ATCC 49284	SEEM9284_03916	IX	
	Mississippi A4-633	LTSEMIS_1115	X	
	Montevideo 4441 H	SEEM41H_02992	Vb	
	Montevideo CFSAN004346	CFSAN004346_01470	Ia	
	Muenster 0315	SEEM0315_05630	III	
	Muenster CFSAN004344	CFSAN004344_03295	Vb	
	Nchanga CFSAN001091	CFSAN001091_05799	III	
	Newport CVM 4176	SEEN176_15095	VIIa	
	Newport CVM 4176	SEEN176_15100	VIId	
	Newport CVM2242	SEEN425_002285	VIIb	
	Ohio CFSAN001079	CFSAN001079_21649	Vc	
	Paratyphi A ATCC 9150	SPA2388	Ia	
	Paratyphi B ATCC 8759	SEEPB759_13629	Ia	
	Poona ATCC BAA-1673	SEEP1673_06839	X	
	Pullorum S06004	I137_12070	Ia	
	Rissen 150	SEER_02967	Vc	
	Rubislaw A4-653	LTSERUB_1231	XIII	

Table S4
P22-like phages and prophages used this study

H cont.	Tailspikes in tree			
	Salmonella enterica **serovar and strain**	**Tailspike locus_tag**	**Tailspike type**	
	Saint Paul SARA26	SES26_19449	Ia	
	Salamae serovar 58:l,z13,z28:z6 str. 00-0163	SES60163_06541	XVI	
	Schwarzengrund SL480	SeSB_A0555	XI	
	Soerenga 695	SEES0695_12821	XII	
	Tallahassee 0012	SEET0012_06891	VIIc	
	Tennessee TXSC_TXSC08-19	SEET0819_19095	Va	
	Typhimurium CFSAN001921	CFSAN001921_15420	Ic	
	Typhimurium D23580	STMMW_03901	Ib	
	Typhimurium FORC_015	ZV79_2330	Vc	
	Typhimurium U228	STU288_12795	Ia	
	Uganda R8-3404	LTSEUGA_0549	III	
	Wandsworth A4-580	LTSEWAN_1004	XII	
	Weltevreden 2007-60-3289-1	SENTW_2521	III	

SUPPLEMENTARY MATERIAL FIGURES

Figure S1

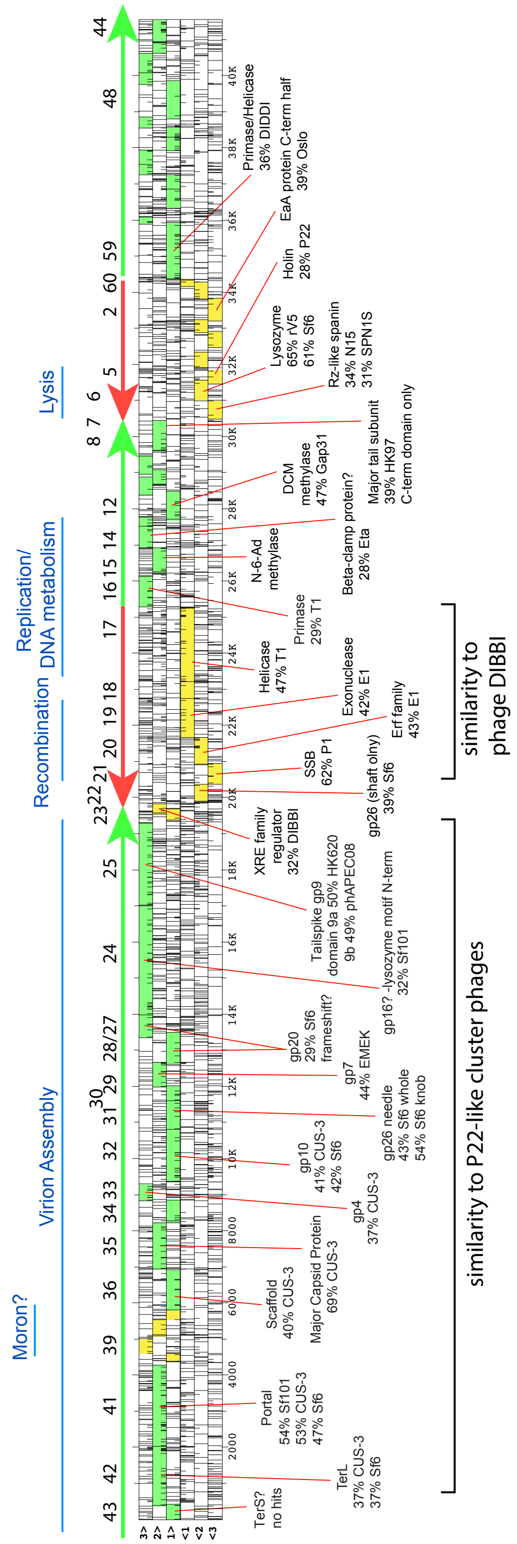


Figure S1. Phage IME-EC2 map and gene database matches.

The phage IME-EC2 genome is shown with the six possible reading frames (indicated at left) with stop codons marked as vertical lines that span the frame rectangle, and AUG codons are indicated by short vertical lines. Open reading frames called as genes are indicated by green and yellow rectangles (transcribed rightward and leftward, respectively). Our gene calls between genes *1* and *5* and between *44* and *59* do not precisely match those of Hua *et al.* (2014); we favor our calls since they leave very little intergenic space as is typical of phage genomes. The opened circular map has been re-oriented from the annotation GenBank accession number KF591601, so that the putative virion assembly genes form the left half of the genome. Above several functional clusters are indicated in blue, selected IME-EC2 locus_tag (gene) numbers of Hua *et al.* (2014) are shown in black, and transcriptional direction (with no operon connotation) is indicated by red and green arrows. Below the ORF map red lines point to the most similar phage proteins found in our analysis. In each graphic annotation the predicted function is given above and similarity to a protein encoded by the indicated phage is given below; no red line indicates there are no good matches in the current sequence database.

In this annotation it is clear that the IME-EC2 virion assembly genes are homologous to and syntenic with those of the P22-like cluster of phages that includes phages Sf6 and CUS-3 (Casjens et al., 2004; King et al., 2007), and “gp#” in the figure refers to the phage P22 gene product names. We note the following additional facts about the IME-EC2 virion assembly genes: (1) genes *43* and *42* overlap by 7 codons, which is typical of terS-terL overlaps in the P22-like phage cluster, suggesting that *43* may encode a TerS protein (TerS proteins are extremely variable, so its lack of homology to known TerS’s is not surprising). (2) Genes *37-40* have no homologues in the P22-like phages (or in other organisms), and we suggest that they are likely a late operon moron (Hendrix et al., 2000; Juhala et al., 2000; Cumby, Davidson, and Maxwell, 2012) that is not involved in virion assembly. (3) Two non-essential genes in this region of the P22-like phages are Orf186 and gene *14*. There are syntenic but not recognizably homologous genes, *34* and *30* respectively, in IME-EC2 that could have analogous functions to those poorly understood P22-like phage genes; more experimental work will be required to determine if this is true. Curiously, IME-EC2 genes *21* and *8* are homologous to particular domains of phage Sf6 (a P22-like phage) tail needle shaft, a trimeric coiled-coil (Olia, Casjens, and Cingolani, 2007), and the C-terminal domain of the HK97 tail shaft protein (HK97 has a long non-contractile tail that is similar to the phage lambda tail). The latter is thought not to be necessary the formation of the actual tail structure in such lambda-like tails; both of these IME-EC2 genes lie in what are likely regions that are expressed early and so may have functions other than virion assembly. (4) The two adjacent IME-EC2 genes *27* and *28* have homology to the 3′- and 5′-sections of P22 gene *20*, respectively. This could be the result of an IME-EC2 sequencing frameshift error, or it may indicate a separation of P22 gp20 into two parts so that its function(s) is carried out by two IME-EC2 proteins.

Figure S2

Enterobacteriaceae P2 supercluster MCP dot plot

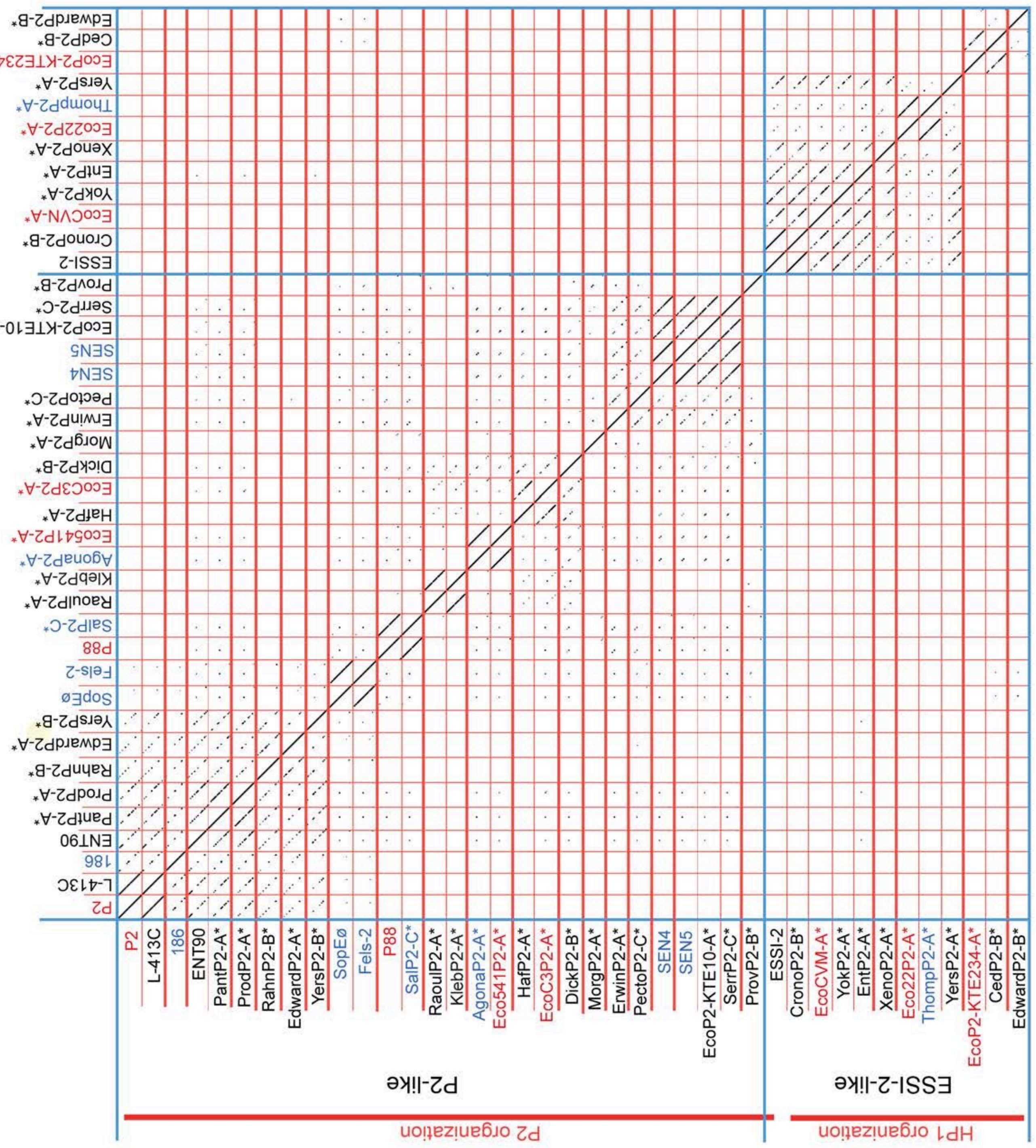


Figure S2. Dot plot analysis of P2-like prophage MCPs from the *Enterobacteriaceae*.

On the left phage and prophage names are shown (red text, *E. coli* host; blue, *Salmonella* host; black, other hosts) and asterisks (*) denote prophage sequences. Blue lines separate the two phage clusters, and thick red lines separate subclusters that are defined in the Figure 2 genome dot plot. Phages and prophages in the plot are listed with host species, genome accession numbers, and MCP locus tags in Table S2A and S2B, respectively. The dot plot was produced by Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 9.

Figure S3

Genome dot plot of authentic P2 supercluster phages

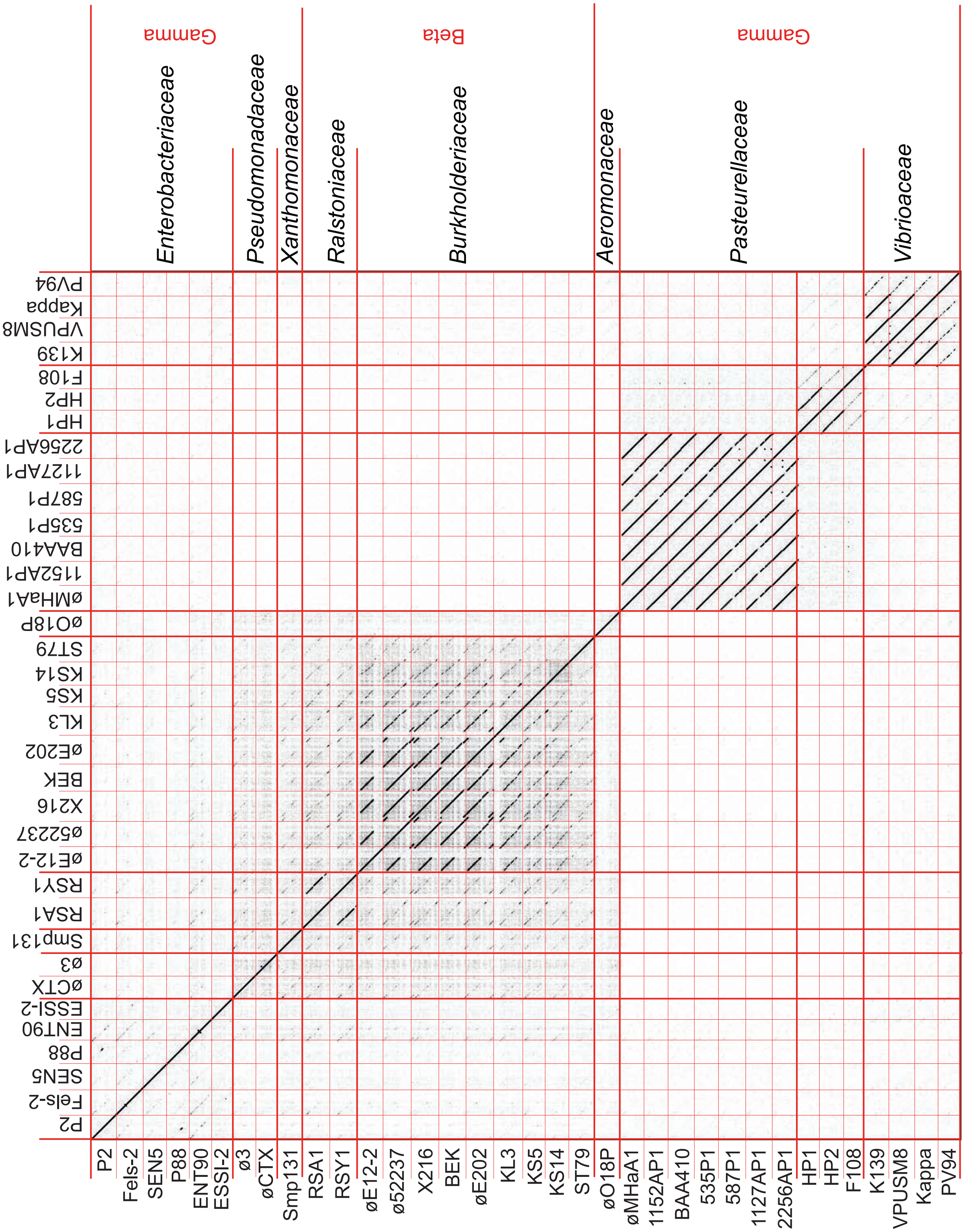


Figure S3. Genome dot plot of authentic P2 supercluster phages.

Phage names are given on the left and top; only representative authentic *Enterobacteriaceae* phage genomes are included. Thin red lines separate individual phage and prophage genomes and thick red line separate the bacterial host taxa shown at the right. Phages in the plot are listed with host species and genome accession numbers in Table S2A and S2C. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 11.

Figure S4

P2-like MCP plot

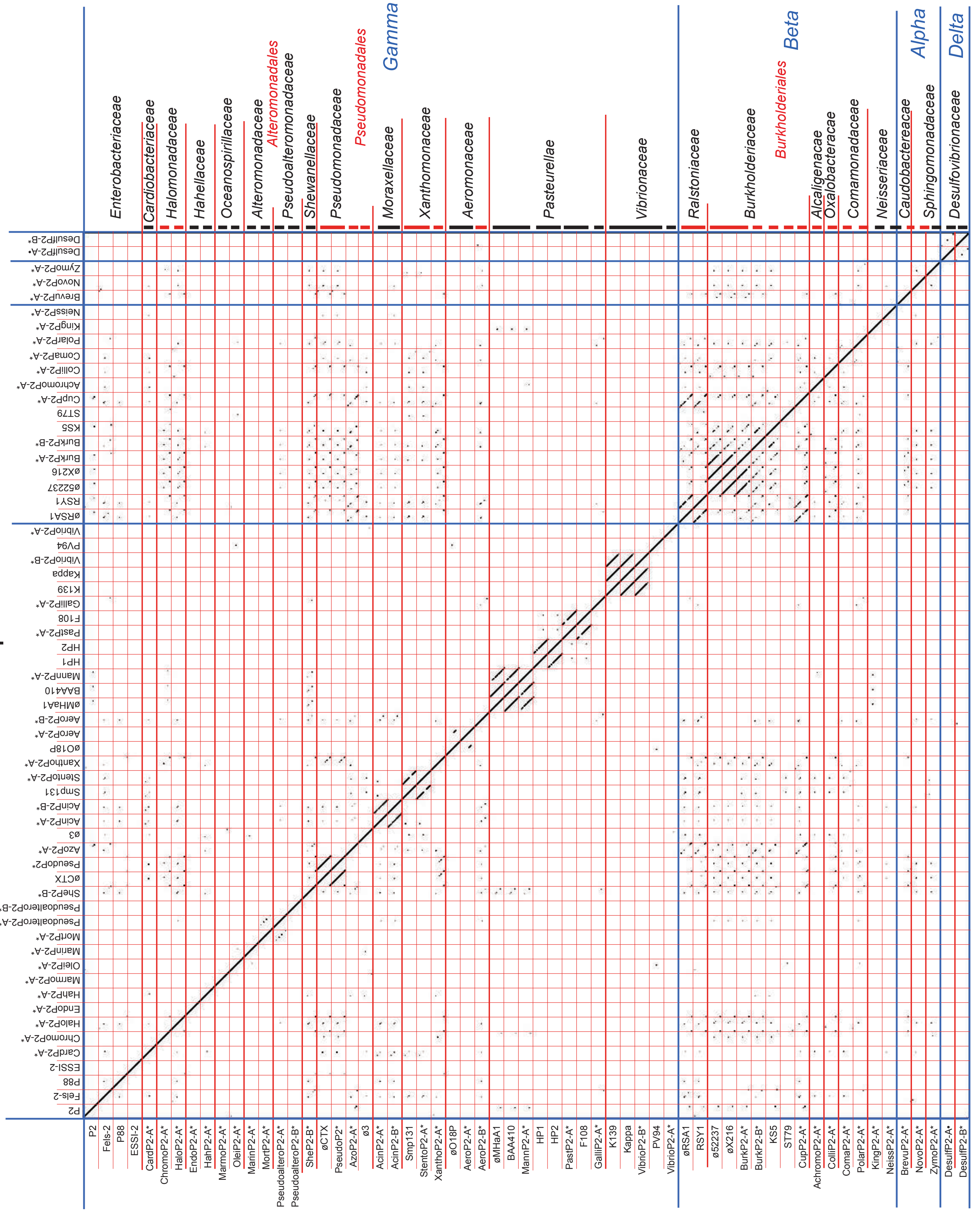


Figure S4. Dot plot analysis of MCPs of P2-like phages and prophages from the Proteobacteria.

Thin red lines separate individual phage and prophage genomes whose names are shown at the top and left; short and longer horizontal red lines separate bacterial host families (black text) and orders (red text), respectively, on the right. Thick blue lines separate Proteobacteria host classes (blue text on the right). Asterisks (*) indicate prophages, and only representative authentic phage genomes are included. Phages and prophages in the plot are listed with host species, genome accession numbers, and MCP locus tags in Table S2. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 9.

Figure S5

P2-like prophages with distantly related hosts

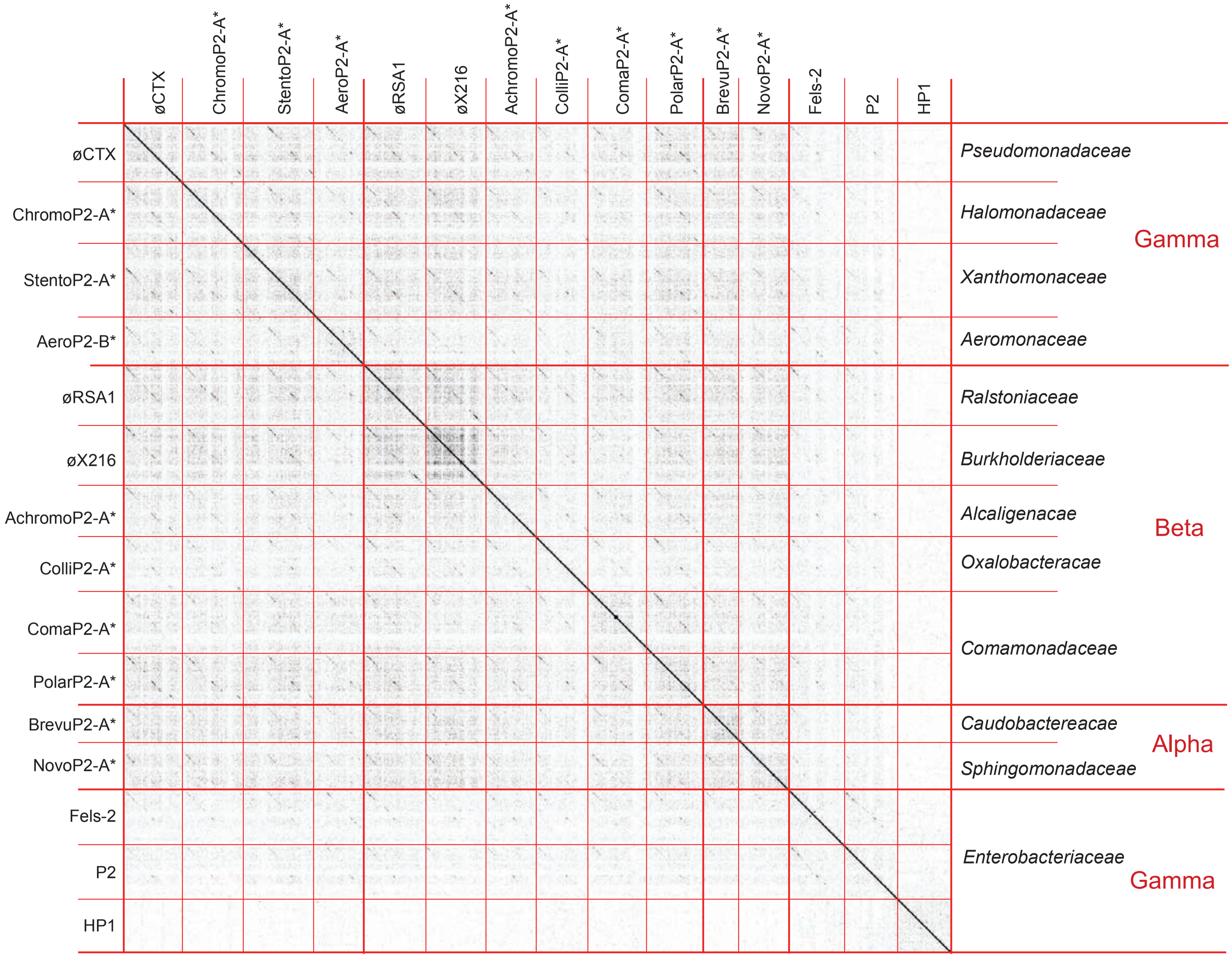


Figure S5. Expanded genome dot plot of P2-like phages and prophages with distantly related hosts.

Red lines separate phage and prophage genomes. Phage and prophage names are given on the left and top; asterisks (*) indicate prophages, and only representative authentic phage genomes are included. Host families (black text) and Proteobacteria classes (red text) are shown at the right. Phages and prophages in the plot are listed with host species, genome accession numbers and MCP locus tags in Table S2. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 10.

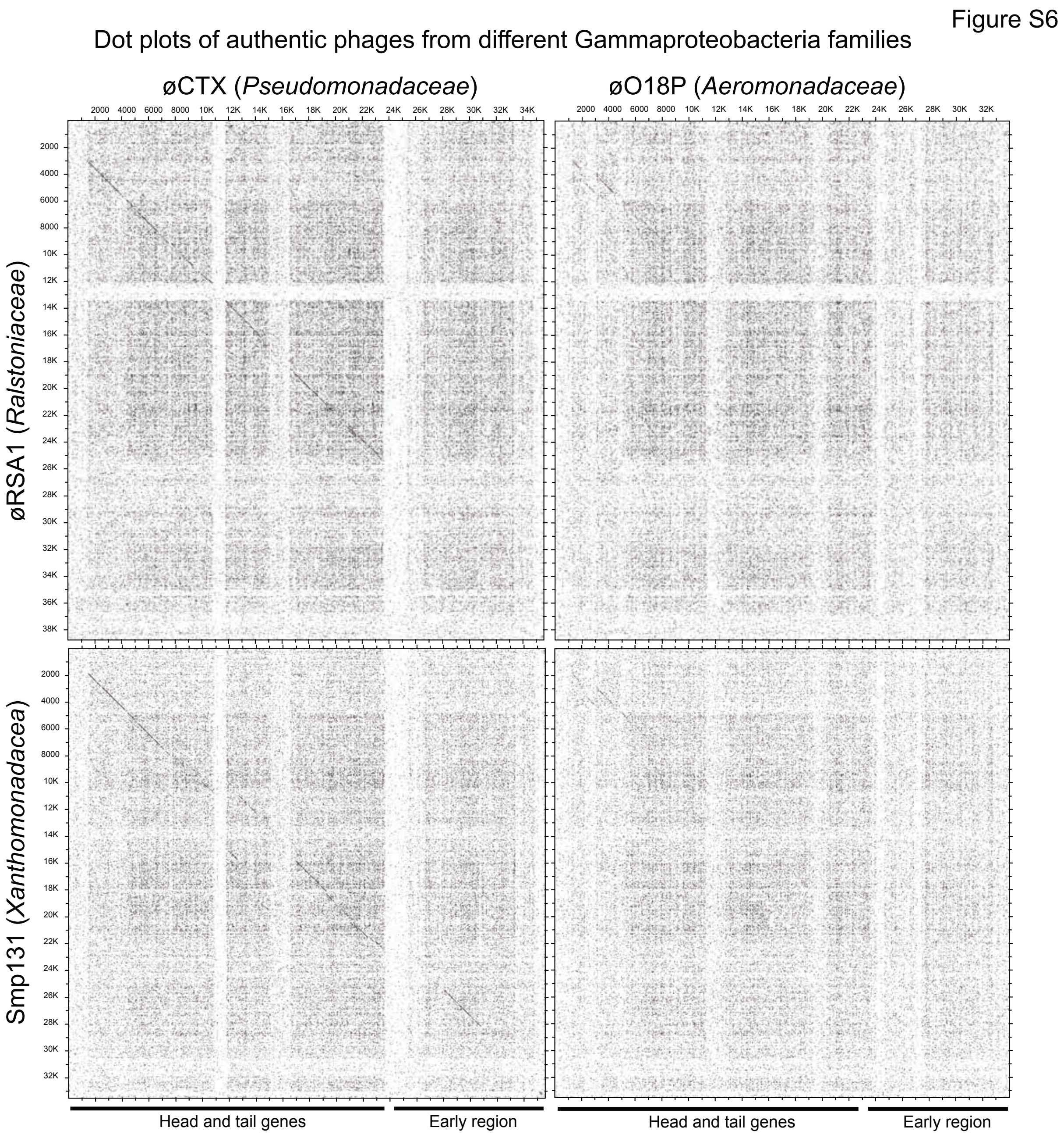


Figure S6. Expanded genome dot plot analysis of authentic P2-like phage examples from different Gammaproteobacteria families.

Phages indicated at the top and left of the plot are listed with host species and genome accession numbers in Table S2C. The dot plots were produced by DNA Strider (Douglas, 1994) with scan window 15 of 23 identities.

Figure S7

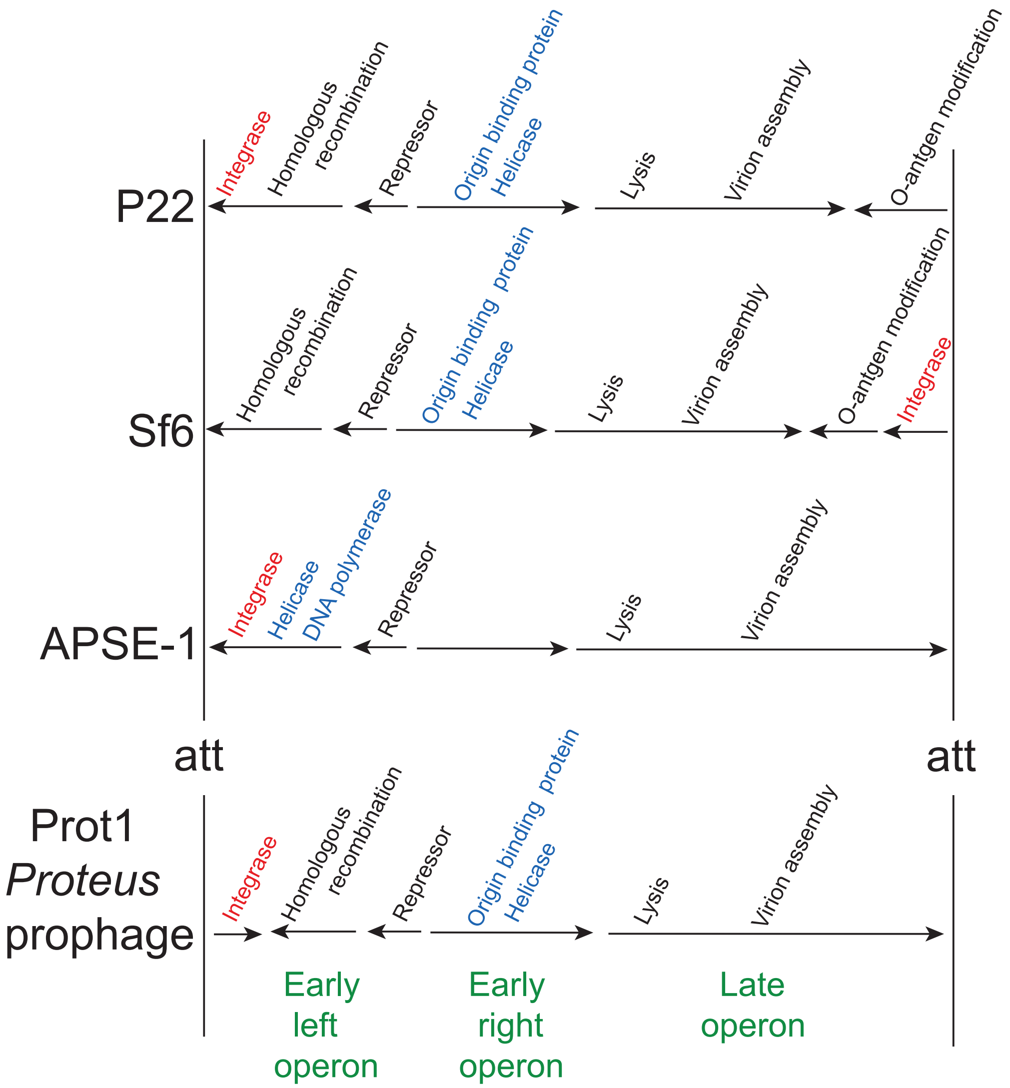


Figure S7. Genome organizational types of the P22-like phages.

Diagrammatic maps of the genomes of four different P22-like phages are shown. Arrows indicate transcriptional direction and the functions of genes in the various regions are shown above each map. The Prot1 prophage is in *Proteus mirabilis* strain ATCC7002 whose MCP locus_tag is DR94_520.

Figure S8

Examples of P22-like prophages

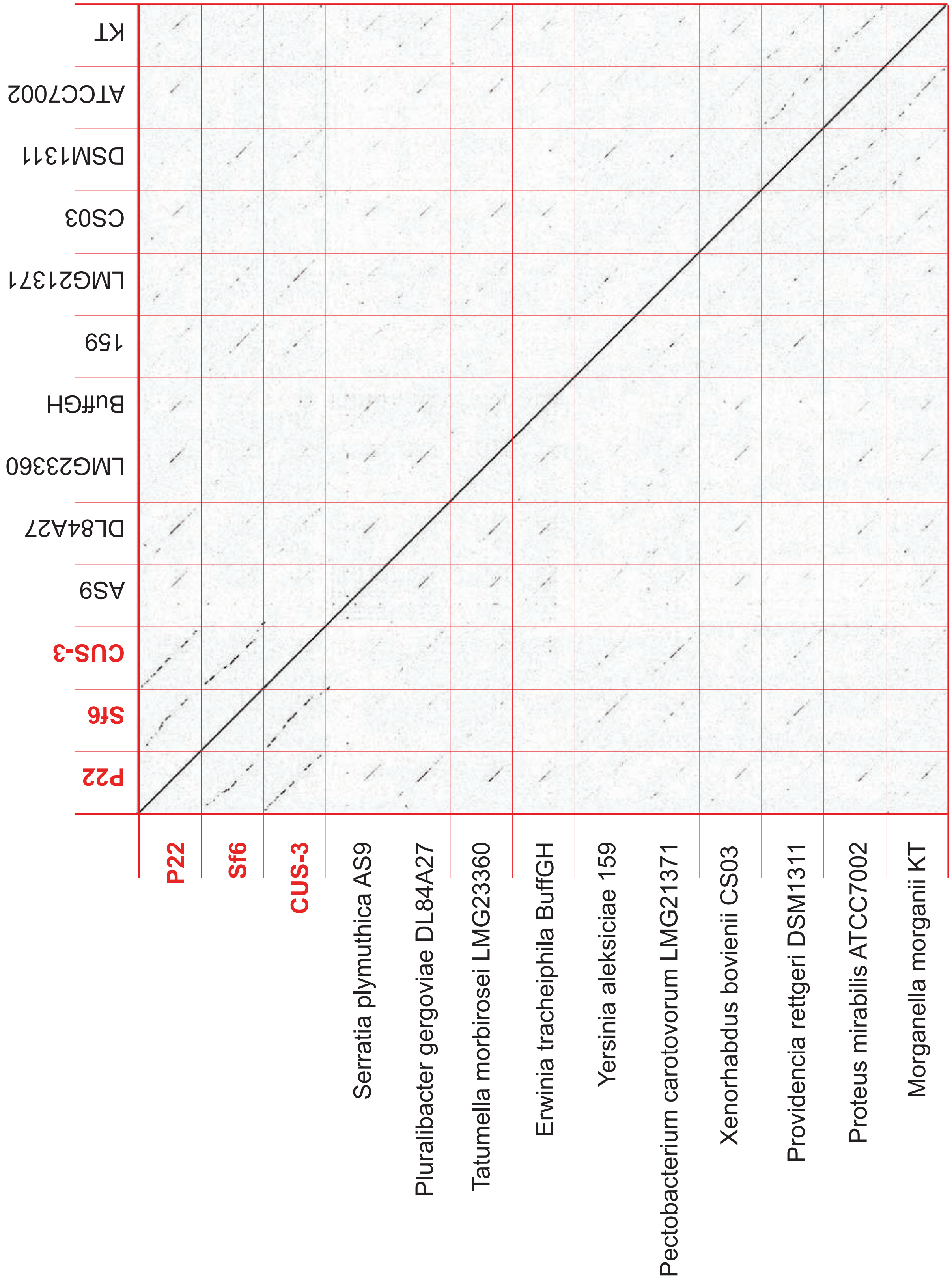


Figure S8. Dot plot analysis of a sample of diverse *Enterobacteriaceae* P22-like prophage genomes.

P22-like prophages in ten different *Enterobacteriaceae* species are shown with three authentic P22-ike phages (bold red text) for comparison. The prophages are indicated by the host strain that they inhabit, and their MCP locus_tags are listed in Table S4. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 10.

Figure S9

P22-like MCP plot

Figure S9. Dot plot analysis of P22-like phage and prophage MCPs.

Red lines separate phage and prophage MCPs. Authentic phage names are in bold red text, and prophage names are indicated by the host species and strain that carries them (listed in Table S4 with their MCP locus_tags). The dot plot was produced by Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 7.

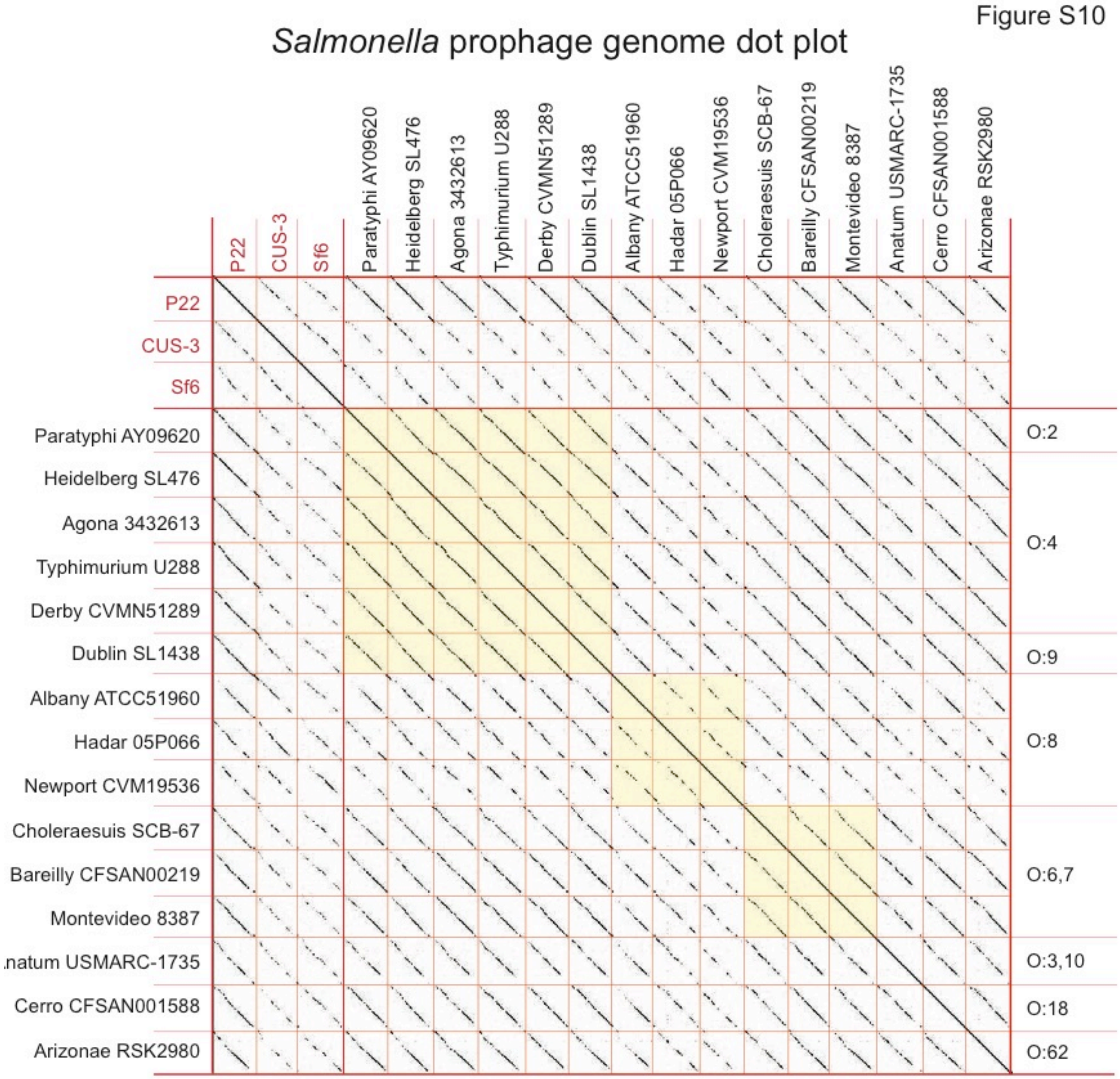


Figure S10. Dot plot analysis of P22-like prophage genomes in *Salmonella enterica*.

P22-like prophages from fifteen different *Salmonella enterica* serovars are shown with three authentic P22-like phages (red text) shown for comparison. The yellow highlight boxes enclose groups of prophages in serovars with the same or closely related O-antigen structures (see text of report). The prophages are indicated by the host strain that they inhabit, and their MCP locus tags are listed in Table S4. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 11.

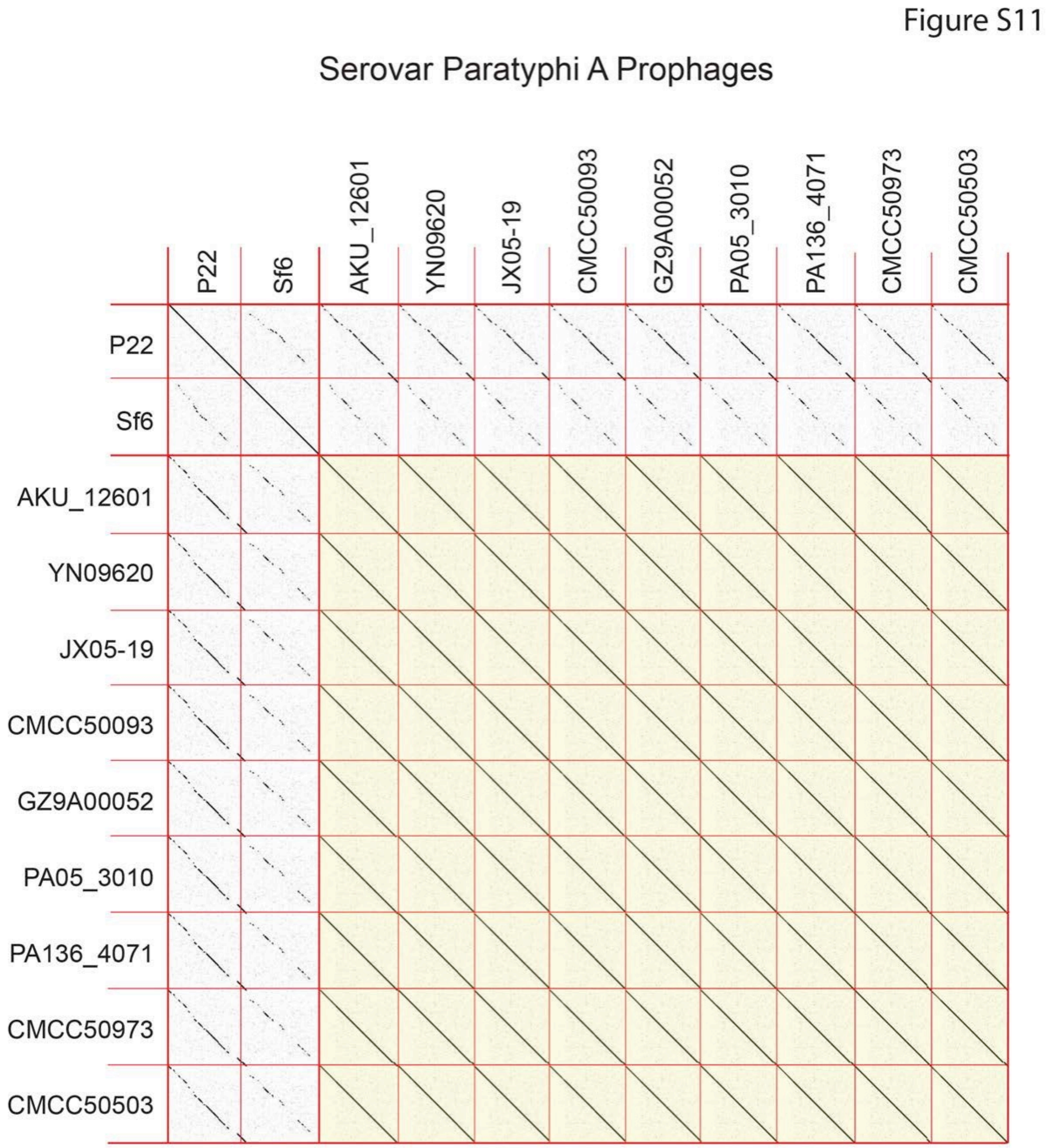


Figure S11. Dot plot analysis of genomes of P22-like prophages in *Salmonella enterica* serovar Paratyphi A isolates.

P22-like prophages from nine randomly chosen Paratyphi A isolates are shown with two authentic P22-like phages (P22 and Sf6) for comparison. The yellow highlight box encloses the nearly identical Paratyphi A prophages. The prophages are indicated by their host strain names (except two whose MCP locus_tags are given since no strain name for the isolate is given in its GenBank entry), and phages and prophages in the plot are listed with host species and MCP locus tags in Table S4D. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 10.

Figure S12

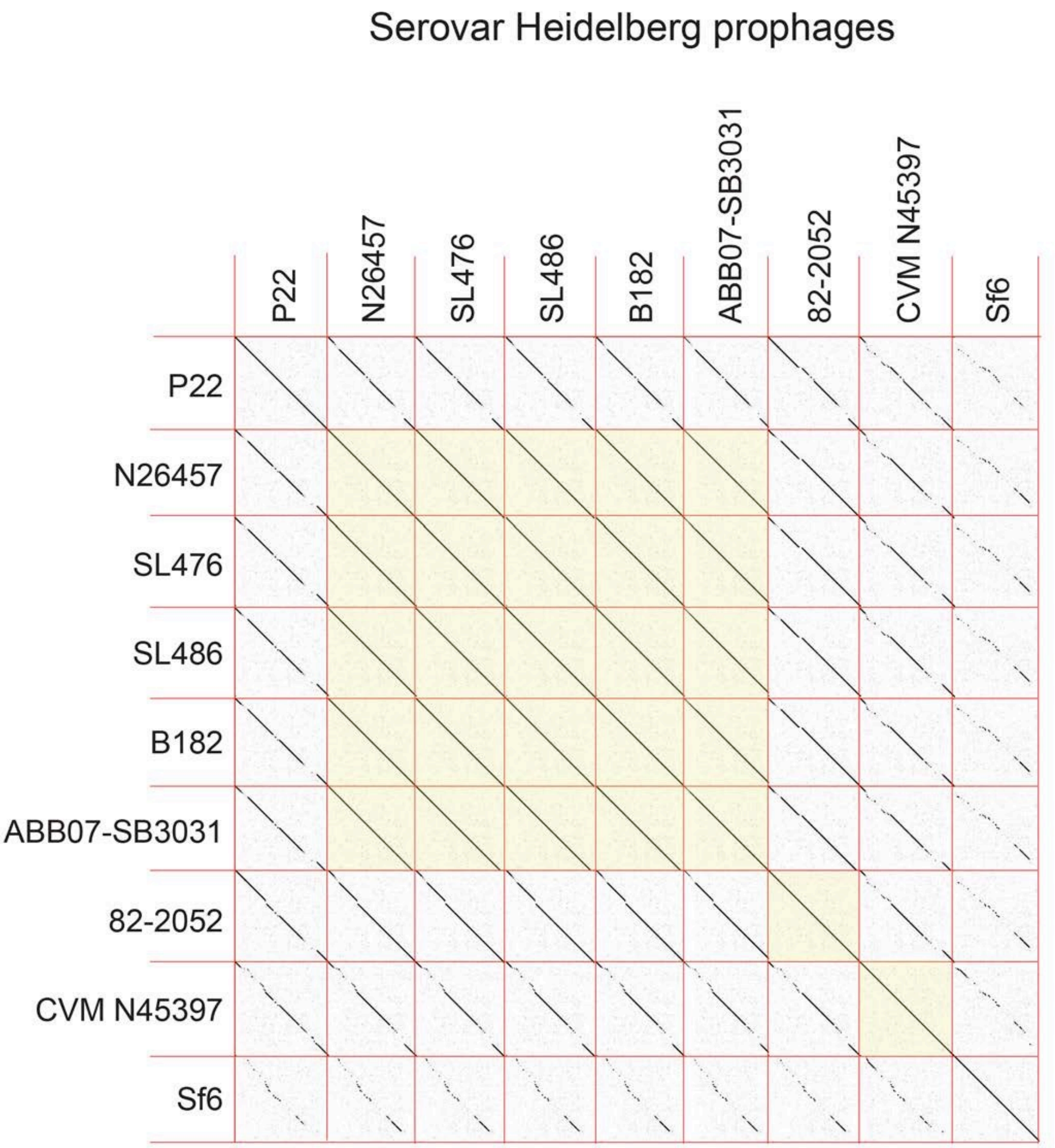


Figure S12. Dot plot analysis of genomes of P22-like prophages in *Salmonella enterica* serovar Heidelberg isolates.

P22-like prophages from seven Heidelberg isolates are shown with two authentic P22-like phages (P22 and Sf6) for comparison. The yellow highlight boxes indicate three different prophage types. The Heidelberg prophages are indicated by their host strain names. Prophages in the plot are listed with host species and MCP locus tags in Table S4E. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 10.

Figure S13

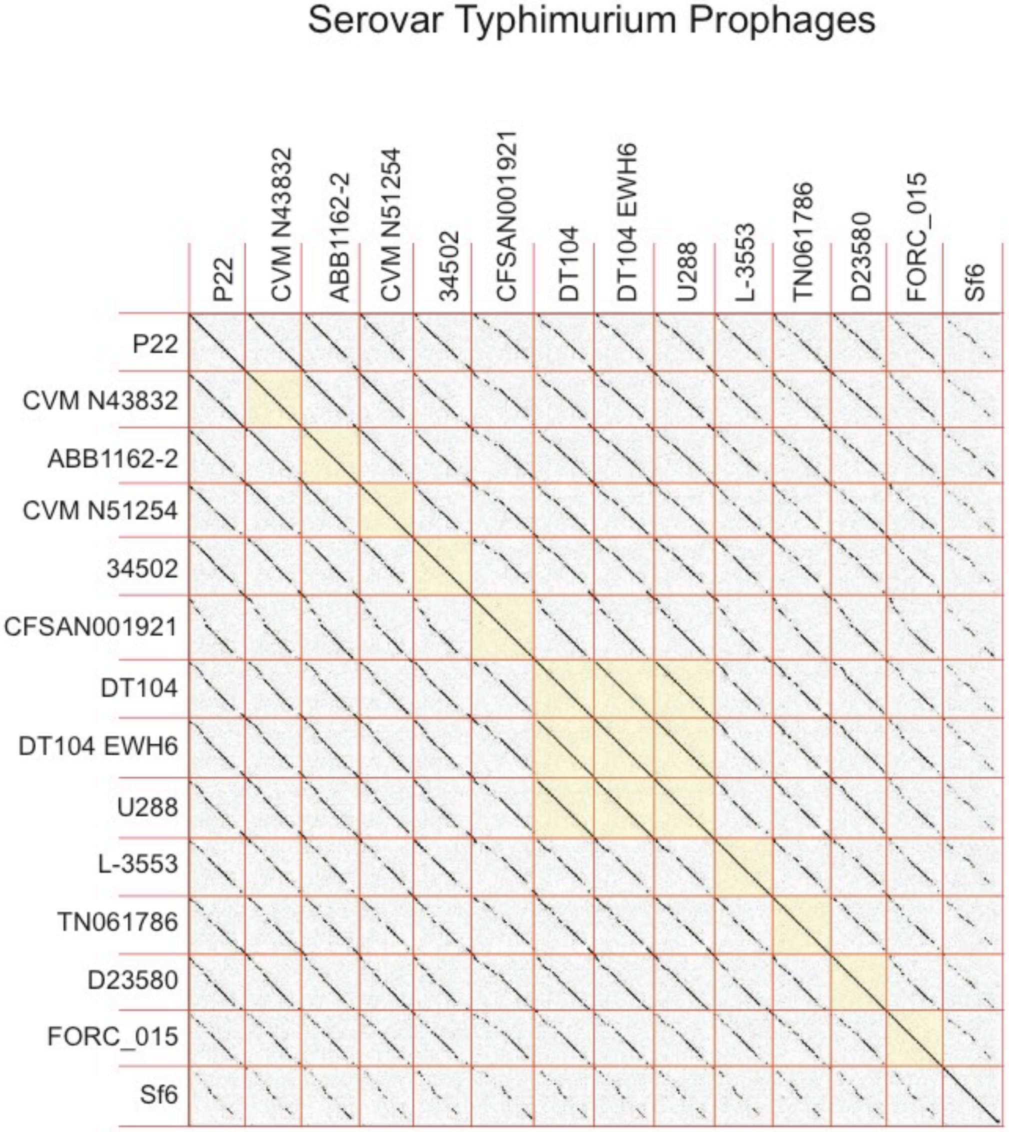


Figure S13. Dot plot analysis of genomes of P22-like prophages in *Salmonella enterica* serovar Typhimurium isolates.

P22-like prophages from twelve Typhimurium isolates are shown with two authentic P22-like phages (P22 and Sf6) for comparison. The Typhimurium sequences are indicated by their host strain names. The yellow highlight boxes indicate significantly different prophage types. Note that the serovar of strain FORC_015 prophage was probably determined incorrectly; it is most likely an O:6,7 strain (see text of report). Phages and prophages in the plot are listed with host species and MCP locus tags in Table S4G. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 10.

Figure S14

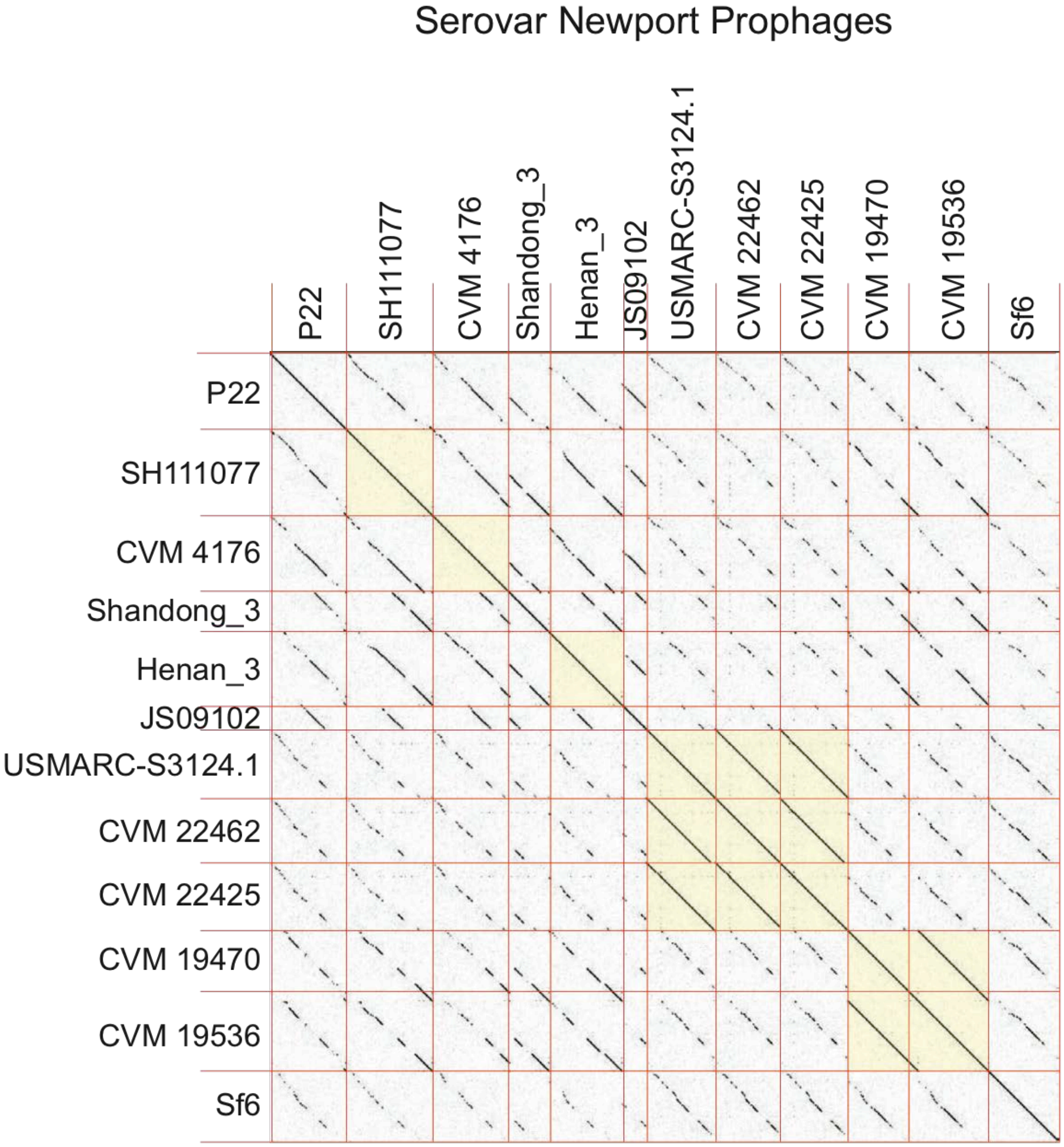


Figure S14. Dot plot analysis of genomes of P22-like prophages in *Salmonella enterica* serovar Newport isolates.

P22-like prophages from ten Newport isolates are shown with two authentic P22-like phages (P22 and Sf6) for comparison. The Newport sequences are indicated by their host strain names. The yellow highlight boxes indicate significantly different prophage types. Two of the ten Newport prophages, in strains Shandong_3 and JS9102, are on short contigs that were not included in the subsequent analysis (see text). Prophages in the plot are listed with host species and MCP locus tags in Table S4F. The dot plot was produced using Gepard (Krumsiek, Arnold, and Rattei, 2007) at a word size setting of 10.

Figure S15

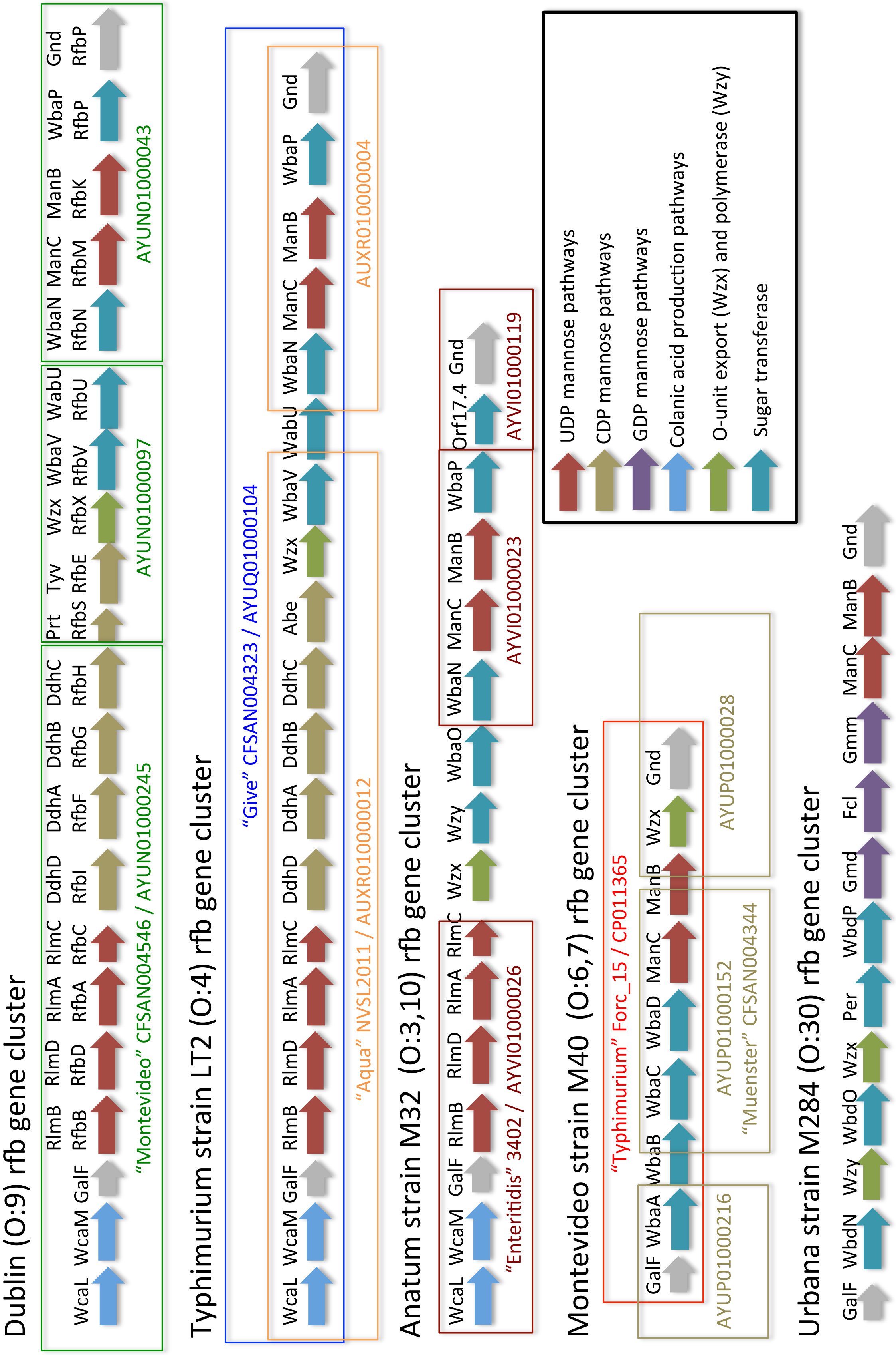

Dublin (O:9) rfb gene cluster
WcaL
WcaM
GalF
RlmB
RfbB
RlmD
RfbD
RlmA
RfbA
RlmC
RfbC
DdhD
RfbI
DdhA
RfbF
DdhB
RfbG
DdhC
RfbH
Prt
RfbS
Tyv
RfbE
Wzx
RfbX
WbaV
RfbV
WabU
RfbU
WbaN
RfbN
ManC
RfbM
ManB
RfbK
WbaP
RfbP
Gnd
RfbP
"Montevideo" CFSAN004546 / AYUN01000245
AYUN01000097
AYUN01000043
Typhimurium strain LT2 (O:4) rfb gene cluster
"Give" CFSAN004323 / AYUQ01000104
WcaL
WcaM
GalF
RlmB
RlmD
RlmA
RlmC
DdhD
DdhA
DdhB
DdhC
Abe
Wzx
WbaV
WabU
WbaN
ManC
ManB
WbaP
Gnd
"Aqua" NVSL2011 / AUXR01000012
AUXR01000004
Anatum strain M32 (O:3,10) rfb gene cluster
WcaL
WcaM
GalF
RlmB
RlmD
RlmA
RlmC
Wzx
Wzy
WbaO
WbaN
ManC
ManB
WbaP
Orf17.4
Gnd
"Enteritidis" 3402 / AYVI01000026
AYVI01000023
AYVI01000119
Montevideo strain M40 (O:6,7) rfb gene cluster
"Typhimurium" Forc_15 / CP011365
GalF
WbaA
WbaB
WbaC
WbaD
ManC
ManB
Wzx
Gnd
AYUP01000216
AYUP01000152
"Muenster" CFSAN004344
AYUP01000028
Urbana strain M284 (O:30) rfb gene cluster
GalF
WbdN
Wzy
WbdO
Wzx
Per
WbdP
Gmd
Fcl
Gmm
ManC
ManB
Gnd
UDP mannose pathways
CDP mannose pathways
GDP mannose pathways
Colanic acid production pathways
O-unit export (Wzx) and polymerase (Wzy)
Sugar transferase

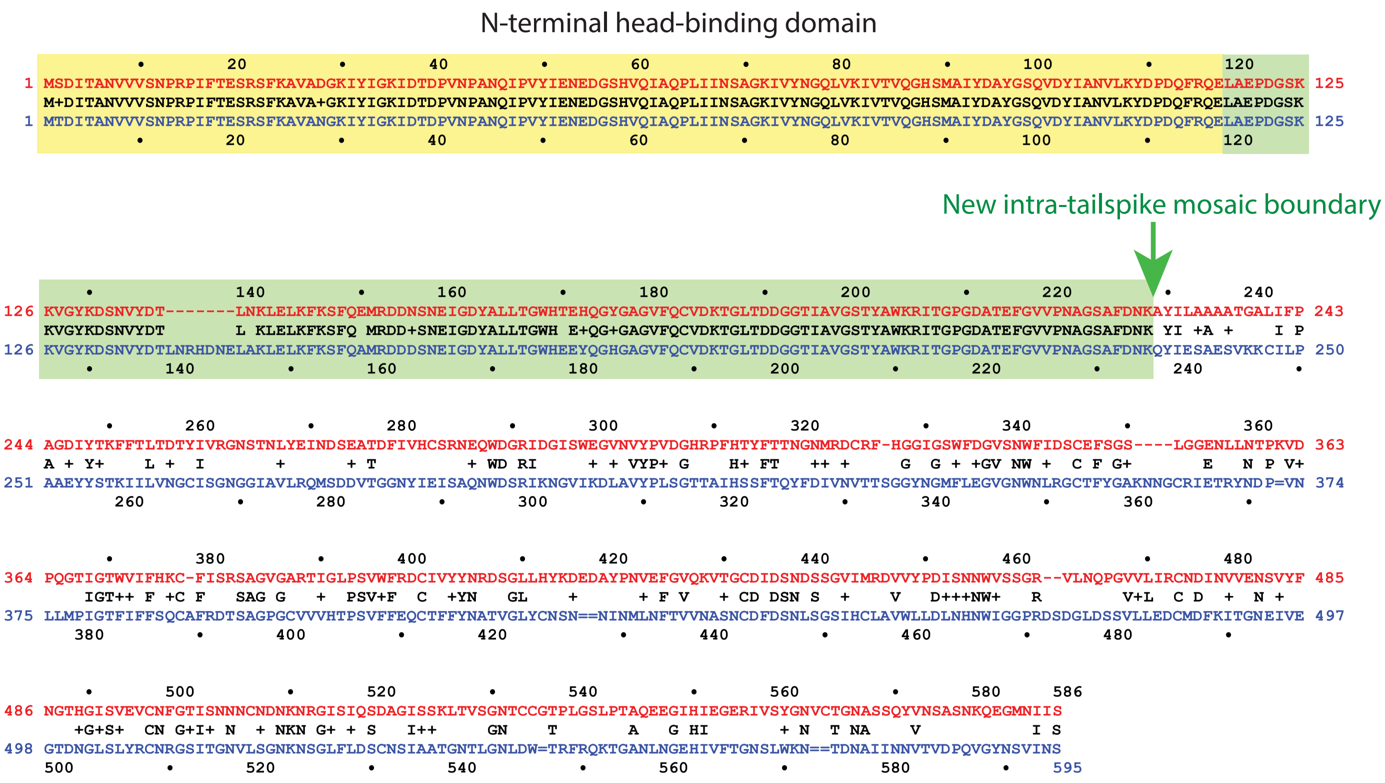
Figure S15. O-antigen synthesis (*rfb* gene) regions of the genomes of the *Salmonella* strains that carry "anomalous" tailspike genes.

Colored arrows represent genes in selected *rfb* regions from the genomes of the indicated *S. enterica* reference strains whose names are given above in large black text; gene names are shown above each map. Colored boxes and similarly colored text indicate the strains and sequence contig accession numbers of the six genomes with "anomalous" tailspike gene discussed in the text of this report. The gene order and content of each of these strains uniquely matches one of the reference regions. The reference *S. enterica rfb* gene regions shown are from the following sources: O:9 serovar Dublin, bp 2019907–1996618 of accession No. LK931502; O:4 serovar Typhimurium strain LT2, X56793 (Wyk and Reeves, 1989; Jiang et al., 1991; Liu et al., 1993; Reeves, 1993); O:3,10 serovar Anatum strain M32, X60665 (Liu et al., 1993) and ATCC BAA-1592, CP007531 (Timme et al., 2013); O:6,7 serovar Montevideo, M84642 (Reeves, 1993); and O:30 serovar Urbana strain M284, AY730593 (Jensen and Reeves, 2001; Samuel et al., 2004).

Figure S16

Novel mosaic boundary inside the tailspike gene

Red = serovar Choleraesuis str. SC-B67 locus_tag="SCH_0367"
Blue = serovar Ohio str. CFSAN001079 locus_tag="CFSAN001079_21649"



```
             •        20         •        40         •        60         •        80         •       100         •       120
  1 MSDITANVVVSNPRPIFTESRSFKAVADGKIYIGKIDTDPVNPANQIPVYIENEDGSHVQIAQPLIINSAGKIVYNGQLVKIVTVQGHSMAIYDAYGSQVDYIANVLKYDPDQFRQELAEPDGSK 125
    M+DITANVVVSNPRPIFTESRSFKAVA+GKIYIGKIDTDPVNPANQIPVYIENEDGSHVQIAQPLIINSAGKIVYNGQLVKIVTVQGHSMAIYDAYGSQVDYIANVLKYDPDQFRQELAEPDGSK
  1 MTDITANVVVSNPRPIFTESRSFKAVANGKIYIGKIDTDPVNPANQIPVYIENEDGSHVQIAQPLIINSAGKIVYNGQLVKIVTVQGHSMAIYDAYGSQVDYIANVLKYDPDQFRQELAEPDGSK 125
             •        20         •        40         •        60         •        80         •       100         •       120

        •              140         •       160         •       180         •       200         •       220         •       240
126 KVGYKDSNVYDT-------LNKLELKFKSFQEMRDDNSNEIGDYALLTGWHTEHQGYGAGVFQCVDKTGLTDDGGTIAVGSTYAWKRITGPGDATEFGVVPNAGSAFDNKAYILAAAATGALIFP 243
    KVGYKDSNVYDT       L KLELKFKSFQ MRDD+SNEIGDYALLTGWH E+QG+GAGVFQCVDKTGLTDDGGTIAVGSTYAWKRITGPGDATEFGVVPNAGSAFDNK YI +A +    I P
126 KVGYKDSNVYDTLNRHDNELAKLELKFKSFQAMRDDDSNEIGDYALLTGWHEEYQGHGAGVFQCVDKTGLTDDGGTIAVGSTYAWKRITGPGDATEFGVVPNAGSAFDNKQYIESAESVKKCILP 250
        •       140         •       160         •       180         •       200         •       220         •       240         •

          •       260         •       280         •       300         •       320          •       340         •           360
244 AGDIYTKFFTLTDTYIVRGNSTNLYEINDSEATDFIVHCSRNEQWDGRIDGISWEGVNVYPVDGHRPFHTYFTTNGNMRDCRF-HGGIGSWFDGVSNWFIDSCEFSGS----LGGENLLNTPKVD 363
    A + Y+    L +  I       +      + T         + WD RI     + + VYP+ G    H+ FT   ++ +     G  G + +GV NW +  C F G+       E   N P V+
251 AAEYYSTKIILVNGCISGNGGIAVLRQMSDDVTGGNYIEISAQNWDSRIKNGVIKDLAVYPLSGTTAIHSSFTQYFDIVNVTTSGGYNGMFLEGVGNWNLRGCTFYGAKNNGCRIETRYNDP=VN 374
           260         •       280         •       300         •       320         •       340         •       360         •

          •        380         •       400         •       420         •       440         •       460           •       480
364 PQGTIGTWVIFHKC-FISRSAGVGARTIGLPSVWFRDCIVYYNRDSGLLHYKDEDAYPNVEFGVQKVTGCDIDSNDSSGVIMRDVVYPDISNNWVSSGR--VLNQPGVVLIRCNDINVVENSVYF 485
        IGT++ F +C F   SAG G   +  PSV+F  C  +YN   GL    +      + F V   + CD DSN S  +    V   D+++NW+   R        V+L  C D + N +
375 LLMPIGTFIFFSQCAFRDTSAGPGCVVVHTPSVFFEQCTFFYNATVGLYCNSN==NINMLNFTVVNASNCDFDSNLSGSIHCLAVWLLDLNHNWIGGPRDSDGLDSSVLLEDCMDFKITGNEIVE 497
       380         •       400         •       420            •       440         •       460         •       480         •

        •       500         •       520         •       540         •       560         •       580   586
486 NGTHGISVEVCNFGTISNNNCNDNKNRGISIQSDAGISSKLTVSGNTCCGTPLGSLPTAQEEGIHIEGERIVSYGNVCTGNASSQYVNSASNKQEGMNIIS
       +G+S+  CN G+I+ N  + NKN G+ + S   I++     GN    T       A   G HI       + N  T NA    V           I S
498 GTDNGLSLYRCNRGSITGNVLSGNKNSGLFLDSCNSIAATGNTLGNLDW=TRFRQKTGANLNGEHIVFTGNSLWKN==TDNAIINNVTVDPQVGYNSVINS
    500         •       520         •       540          •       560         •         580         •  595
```

Figure S16. New mosaic boundary in the P22-like tailspike gene.

The newly discovered mosaic P22-like phage boundary in the tailspike gene is shown by comparing the amino acid sequences of type Va (*S. enterica* serovar Choleraesuis strain SC-B67) and type Vc (*S. enterica* serovar Ohio strain CFSAN001079) tailspikes as aligned by DNA Strider (Douglas, 1994). The N-terminal head-binding domain is indicated by the yellow box, the green box marks the N-terminal common region of the type V tailspike polysaccharide-binding domain, and the green vertical arrow marks the position of the newly discovered mosaic boundary.

References

Casjens, S., Winn-Stapley, D., Gilcrease, E., Moreno, R., Kühlewein, C., Chua, J. E., Manning, P. A., Inwood, W., and Clark, A. J. (2004). The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J. Mol. Biol.* **339,** 379-394.

Cumby, N., Davidson, A. R., and Maxwell, K. L. (2012). The moron comes of age. *Bacteriophage* **2,** 225-228.

Douglas, S. E. (1994). DNA Strider. A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol. Biol.* **25,** 181-194.

Hendrix, R. W., Lawrence, J. G., Hatfull, G. F., and Casjens, S. (2000). The origins and ongoing evolution of viruses. *Trends Microbiol.* **8,** 504-508.

Hua, Y., An, X., Pei, G., Li, S., Wang, W., Xu, X., Fan, H., Huang, Y., Zhang, Z., Mi, Z., Chen, J., Li, J., Zhang, F., and Tong, Y. (2014). Characterization of the morphology and genome of an *Escherichia coli* podovirus. *Arch. Virol.* **159,** 3249-3256.

Jensen, S. O., and Reeves, P. R. (2001). Molecular evolution of the GDP-mannose pathway genes (*manB* and *manC*) in *Salmonella enterica*. *Microbiology* **147** (Pt 3)**,** 599-610.

Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K., and Reeves, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Molec. Microbiol.* **5,** 695-713.

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299,** 27-51.

King, M. R., Vimr, R. P., Steenbergen, S. M., Spanjaard, L., Plunkett, G., 3rd, Blattner, F. R., and Vimr, E. R. (2007). *Escherichia coli* K1-specific bacteriophage CUS-3 distribution and function in phase-variable capsular polysialic acid O acetylation. *J. Bacteriol.* **189,** 6447-6456.

Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23,** 1026-1028.

Liu, D., Haase, A. M., Lindqvist, L., Lindberg, A. A., and Reeves, P. R. (1993). Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica:*

identification and characterization of transferase genes of groups B, C2, and E1. *J. Bacteriol.* **175,** 3408-3413.

Malki, K., Kula, A., Bruder, K., Sible, E., Hatzopoulos, T., Steidel, S., Watkins, S., and Putonti, C. (2015). Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phlya. *Virol. J.* **12,** 164.

Olia, A. S., Casjens, S., and Cingolani, G. (2007). Structure of phage P22 cell envelope-penetrating needle. *Nat. Struct. Mol. Biol.* **14,** 1221-1226.

Reeves, P. (1993). Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet.* **9,** 17-22.

Samuel, G., Hogbin, J. P., Wang, L., and Reeves, P. R. (2004). Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii,* which express identical O antigens. *J. Bacteriol.* **186,** 6536-6543.

Timme, R. E., Pettengill, J. B., Allard, M. W., Strain, E., Barrangou, R., Wehnes, C., Van Kessel, J. S., Karns, J. S., Musser, S. M., and Brown, E. W. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. enterica inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* **5,** 2109-2123.

Wyk, P., and Reeves, P. (1989). Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B salmonellae: homology with galactose epimerase. *J. Bacteriol.* **171,** 5687-5693.