Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae

Astrid Collingro, Stephan Köstlbacher, Marc Mussmann,

Ramunas Stepanauskas, Steven J. Hallam, Matthias Horn

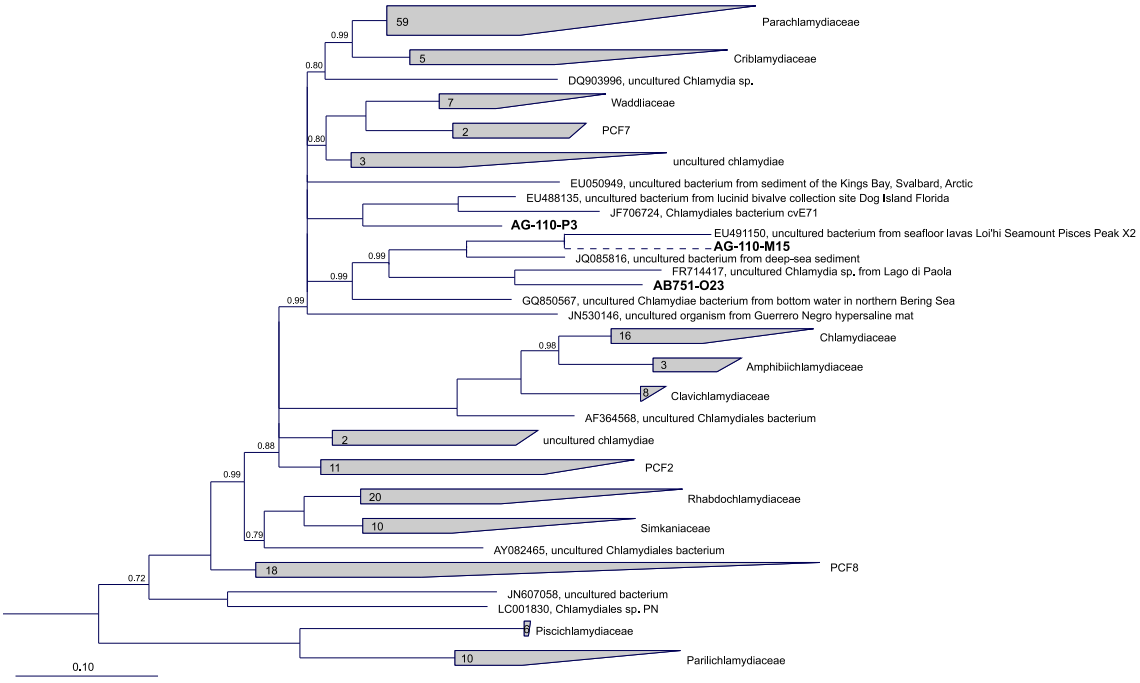
Supplementary Information

Contents

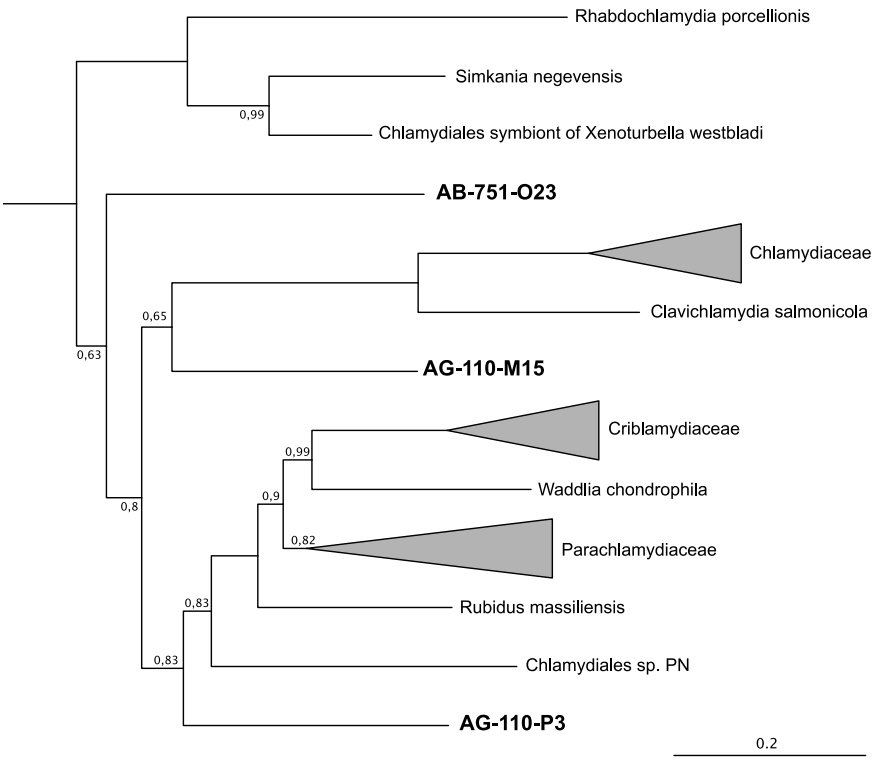
Phylogeny of marine SAGs

Supplementary Text 1

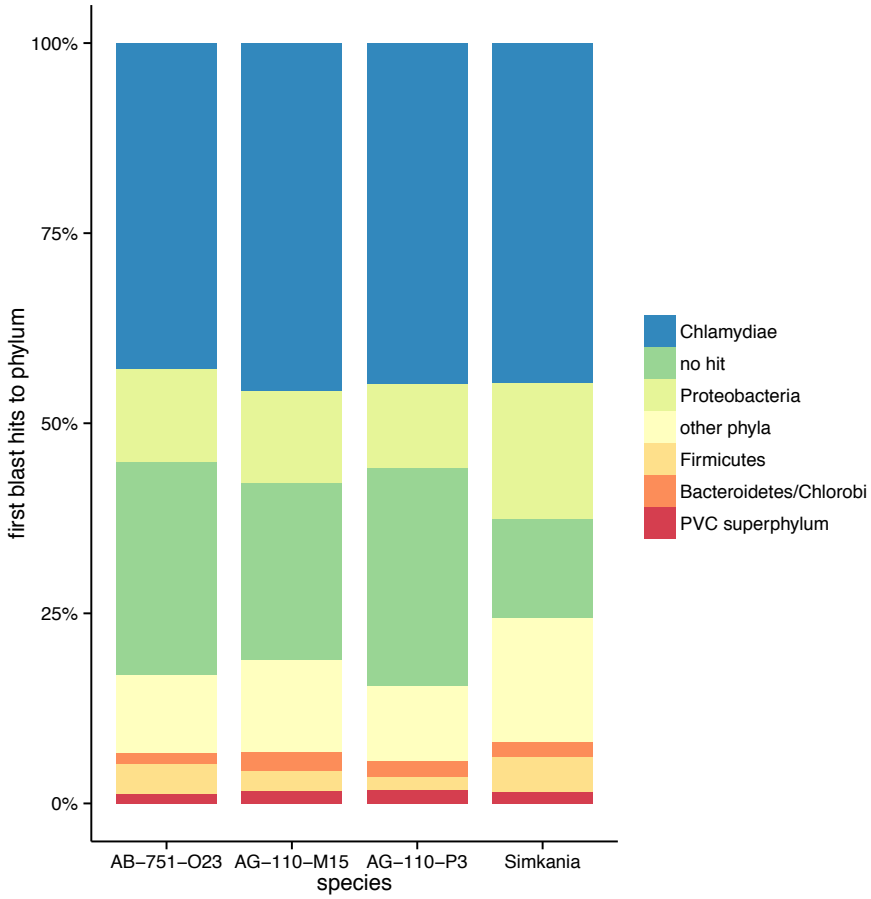
Phylogenetic analysis using publically available near-full length 16S rRNA sequences of chlamydiae demonstrated the affiliation of one SAG (AB-751-O23) with a lineage of deeply branching and yet uncultivated chlamydiae of marine origin represented so far only by a single clone sequence detected in the marine coastal lake Lago di Paola, Italy (FR714417, Pizzetti *et al.*, 2012, 90.44% similarity to AB-751-O23) (Figure 1, Supplementary Figure 1). The partial 16S rRNA sequence of SAG AG-110-M15 was added to the full-length gene tree and formed a divergent potential sister lineage to SAG AB-751-O23 (81% identity over 480 nucleotides). The sequence closest related to the SAG AG-110-M15 sequence is again of marine origin and belongs to a clone detected in a sample from seafloor lavas in Hawai, USA (EU491150, Santelli *et al.*, 2008, 88% similarity to AG-110-M15). The 16S rRNA of SAG AG-110-P3 clusters with members of the deeply branching family PCF5 identified earlier (Lagkouvardos *et al.*, 2014) and is most similar (90.37%) to a sequence derived from a lucinid bivalve collection site at Dog Island Florida, USA (EU488135).



Supplementary Figure 1. 16S rRNA-based phylogeny of SAGs indicating their affiliation to deeply branching chlamydial lineages.
Bayesian inference of near full-length 16S rRNA sequences. Posterior probabilities are indicated only if <1. The partial 16S rRNA sequence of AG-110-M15 was added to the tree using the Quick-Add Parsimony option in ARB.



Supplementary Figure 2. Phylogeny of marker genes present in all three SAGs confirms their affiliation to the chlamydiae.
Bayesian inference of seven concatenated single copy marker genes including the ribosomal proteins rs7, rs12, rl6, rl18, rs5, rl15, rs10, and rs8. Posterior probabilities are indicated only if <1.



Supplementary Figure 3. The SAGs have a similar taxonomic distribution of first blast hits as other chlamydial genomes.
The relative taxonomic distribution on phylum level of the first blast hit of all proteins encoded in the SAGs and for comparison in a complete chlamydial genome (*Simkania negevensis*) is shown.

Presence of chlamydial hallmark genes and new protein families in the marine SAGs

Supplementary Text 2

Due to their obligate intracellular lifestyle chlamydiae encode many genes for interference with and exploitation of their eukaryotic host cells (Elwell *et al.*, 2016). Some of them, including chlamydia-specific genes, could also be detected in at least one of the SAGs (Supplementary Table 1). Among those are genes encoding proteins involved in host cell manipulation such as the tail-specific protease Tsp and HtrA, a protease important in the replicative stage of the developmental cycle (Gloeckl *et al.*, 2013; Patel *et al.*, 2014). Structural components of the NF-T3S apparatus are encoded in the SAGs together with a large number of predicted effector proteins containing eukaryotic-domains and potentially involved in protein-protein interactions (like ankyrin repeats, TPR repeats, and F-box proteins), carbohydrate-binding (cadherin-like domains and RHS repeats) or the interference with the host ubiquitination pathway (ubiquitin carboxyl-terminal hydrolase and RING-finger domains)(Domman *et al.*, 2014).

In addition, a number of very large proteins potentially involved in host interaction are encoded in the SAGs. SAG AB-751-O23 for example hosts two protein families with three and four members, respectively. Proteins within the first family (BQ_00010, CA_00010, CT_00010) show weak homology to the ice-nucleation protein of *Caulobacter crescentus* or other alphaproteobacterial large exoproteins that are potentially involved in adhesion. Members of the second protein family (AI_00010, AV_00060, AZ_00040, BM_00010) show homology to self-associating autotransporters (SAATs)(Klemm *et al.*, 2006). Such SAATs have been described in a number of bacterial pathogens, in which they function in adhesion and invasion but also in auto-aggregation and even biofilm formation (Charbonneau *et al.*, 2006; Klemm *et al.*, 2006).

Supplementary Table 1. Presence of chlamydial virulence associated genes in the SAGs.

	AB-751-O23	AG-110-M15	AG-110-P3
Structural components of the NF-T3S apparatus	sctL (BV_00010), sctR (BV_00020), sctS (BV_00030), sctT (BV_00040), sctW (AF_00160), sctV (AF_00170), sctU (AF_00180)	sctV(AG_00110), sctU (AG_00120)	sctQ (AN_00020), sctN (AN_00100), sctG (AN_00120), sctF (AN_00130), sctE (AN_00140), sctD (AN_00150)
Nucleotide transporter	AF_00140, AG_00130	AA_00360, AD_00330, BA_00030	AJ_00150, AU_00190, AU_00200
Ser/Thr-proteases	pknD (AB_00310), tsp (AB_00220), Do-like (AA_00440)		pkn5 (AN_00010), tsp (AU_00070), htrA (AY_00070)
Transcriptional regulator early upstream open reading frame (euo)	euo (AE_00020)		
Chlamydia protease-like activity factor (CPAF)		AM_00080	AQ_00070
Macrophage infectivity potentiator (mip)		AW_00050	
Chlamydia virulence plasmid homologs		pGP6-D (AE_00420)	pGP8-D (AR_00180)

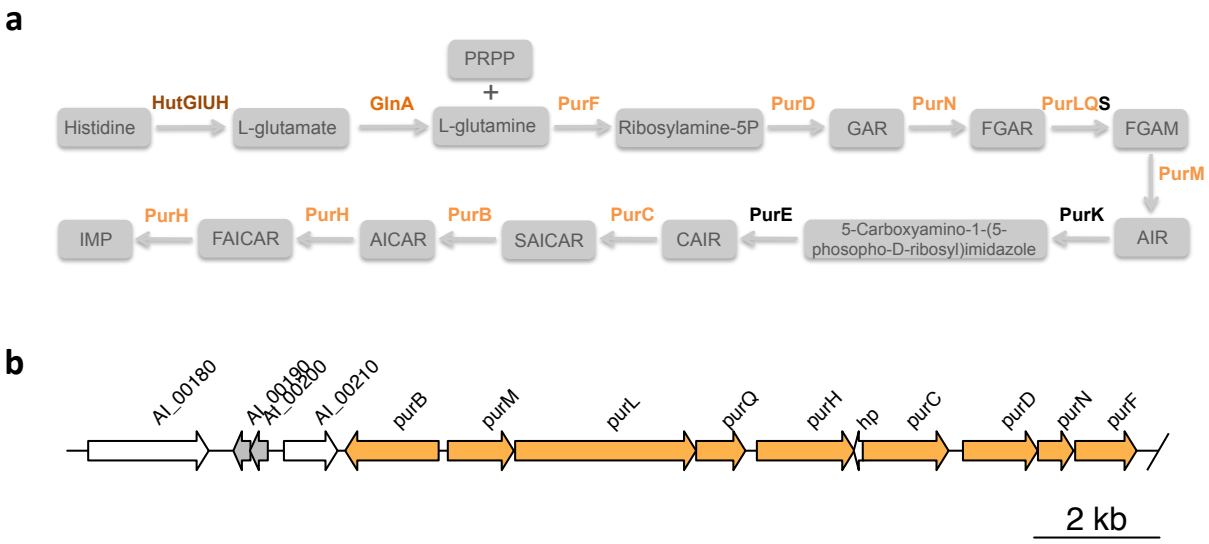
Metabolic features of SAGs absent in other chlamydiae

Supplementary Text 3

We identified a transporter for taurine in SAG AG-110-P3; this organic compound can serve as nutrient but is also known as cyroprotectant. Interestingly, a number of large proteins encoded on SAG AB-751-O23 and representing potential effector proteins (BQ_00010, CA_00010, CT_00010, AI_00010, AV_00060, AZ_00040, BM_00010; also see Supplementary Text 2) have also been implicated in cryoprotection in marine microorganisms. Thus, the identification of different proteins potentially involved in cryoprotection in different chlamydial SAGs suggests that marine chlamydiae have adopted different mechanisms to cope with cold temperatures in their natural environment.

Supplementary Text 4

Only two genes (*purKE*) are missing to complete the pathway for purine de novo synthesis in the SAG AG-110-M15 (Supplementary Figure 4a). Since all other genes of the purine biosynthesis pathway are encoded on a single genomic locus in the SAG, these genes might be present in the genome and only missing in the assembly, because the contig ends there (Supplementary Figure 4b). Most of the genes within the purine biosynthesis pathway show homology to genes from members of the *Marinimicrobia*. With the exception of two genes the remaining genes on this contig either are unique to AG-110-M15 (n=9) or are related to those of other chlamydiae (n=11). Whether this pathway is present in the SAG due to lateral gene transfer or was commonly present in ancient chlamydiae and lost in all other species analyzed to date is thus currently unclear.



Supplementary Figure 4. A nearly complete purine biosynthesis pathway is present in AG-110-M15.

(a) Representation of the purine biosynthesis pathway (adapted from KEGG). Colored proteins are encoded, black proteins are missing in the assembly of AG-110-M15. The same color indicates presence at the same genomic locus. PRPP – phosphoribosylpyrophosphate, GAR – glycinamide ribonucleotide, FGAR – N-formylglycinamide ribonucleotide, FGAM – N-formylglycinamidine ribonucleotide, AIR – Aminoimidazole ribonucleotide, CAIR – Carboxyaminoimidazole ribonucleotide, SAICAR – N-succinocarboxyamide-5-aminoimidazole ribonucleotide, AICAR – aminoimidazole-4-carboxamide ribonucleotide, FAICAR – 5-formamido-4-imidazolecarboxamide ribonucleotide, IMP – Inosine monophosphate **(b)** Gene organization at the purine locus in the assembly of AG-110-M15. The contig ends after *purF*. Orange genes belong to the purine pathway, grey genes are present in other chlamydial genomes, white genes encode hypothetical proteins.

Flagellar genes in chlamydiae

Supplementary Text 5

Non-flagellar T3SSs are present in all chlamydiae. This secretion system originally evolved from the flagellar system, and both systems still share many homologous proteins (Figure 3a)(Abby and Rocha, 2012). In addition to the NF-T3SS, members of the *Chlamydiaceae* also encode few homologs of flagellar genes (*fliA, F, H, I, flhA*), whose functions are not well understood (Peters *et al.*, 2007; Ferrell and Fields, 2016). Because in members of the *Chlamydiaceae* the flagellar genes are transcribed early in the developmental cycle, and because interactions between flagellar proteins and NF-T3SS components could be detected (Stone *et al.*, 2010), the current hypothesis is, that *Chlamydiaceae* use their flagellar genes to assemble a second slightly different NF-T3S apparatus, in which some NF-T3SS proteins are replaced by their respective flagellar homologs. This alternative T3SS might be fully assembled at the middle of the developmental cycle and might be employed for interaction with the host cell during this developmental stage (Peters *et al.*, 2007; Ferrell and Fields, 2016).
Another flagellar component found in all chlamydial genomes including the SAGs, but lacking in *C. trachomatis,* is the gene for the flagellar cytoplasmic transmembrane component FliO. FliO is the least conserved protein among the membrane-bound components of the flagellar export apparatus (Tsang and Hoover, 2014) and has no homolog in NF-T3SSs. Indeed it is not annotated in many genomes, because of its weak homology to recognized FliO proteins. FliO has been shown to regulate FliP, and to be necessary for full motility of the flagellum (Barker *et al.*, 2010). What FliO would contribute in chlamydial genomes without flagella is unclear, but it could be involved in regulation of the NF-T3SS.

Flagellar genes in marine SAGs

Supplementary Text 6

The flagellar gene set found in SAG AB-751-O23 is the most complete flagellar machinery ever detected in chlamydiae. Only genes for the formation of the inner rod (*flgB, C, F, G*), the P- and L-ring components (*flgH, I*), *fliE* and *flhE* are absent in AB-751-O23 (Figure 3a)(Chevance and Hughes, 2008). With the exception of *flhE* these genes can be found in the related SAG AG-110-M15. In this SAG the *flgBC* and *fliE* genes are located at one genomic region, whereas *flgFGAHIJN* form a cluster with *flgKL and fliWC*. The latter part of this gene cluster containing *flgKL* and *fliWC* is also present in SAG AB-751-O23, but the contig ends after *flgK* (Figure 3b). The synteny between the two SAGs at this locus indicates that the missing membrane anchoring parts of the flagellar basal body are indeed encoded in SAG AB-751-O23, but just missed due to contig break point at this location.
The flagellar genes are dispersed at various loci in each of the three SAGs. In AB-751-O23 six contigs harboring flagellar genes were detected. The arrangement of the genes at the different loci resembles that of *Thermotoga maritima* (Nelson *et al.*, 2001) and is not unusual in bacterial genomes. Flagellar genes detected in the assemblies of AG-110-M15 and AG-110-P3 can be found at four and two loci, respectively. The presence and patchy distribution of flagellar and chemotaxis components in chlamydial genomes is somewhat puzzling and might be best explained via differential gene loss.

Phylogeny of flagellar proteins encoded in the SAGs

Supplementary Text 7

Flagellar proteins encoded at different genomic loci in the SAGs were used for single protein phylogenies in order to further confirm that the respective genes indeed do not represent contaminations in the assemblies (i.e. are not closely related with known flagellar genes from other taxonomic groups), and to investigate whether the flagellar genes of the different SAGs are related with each other.

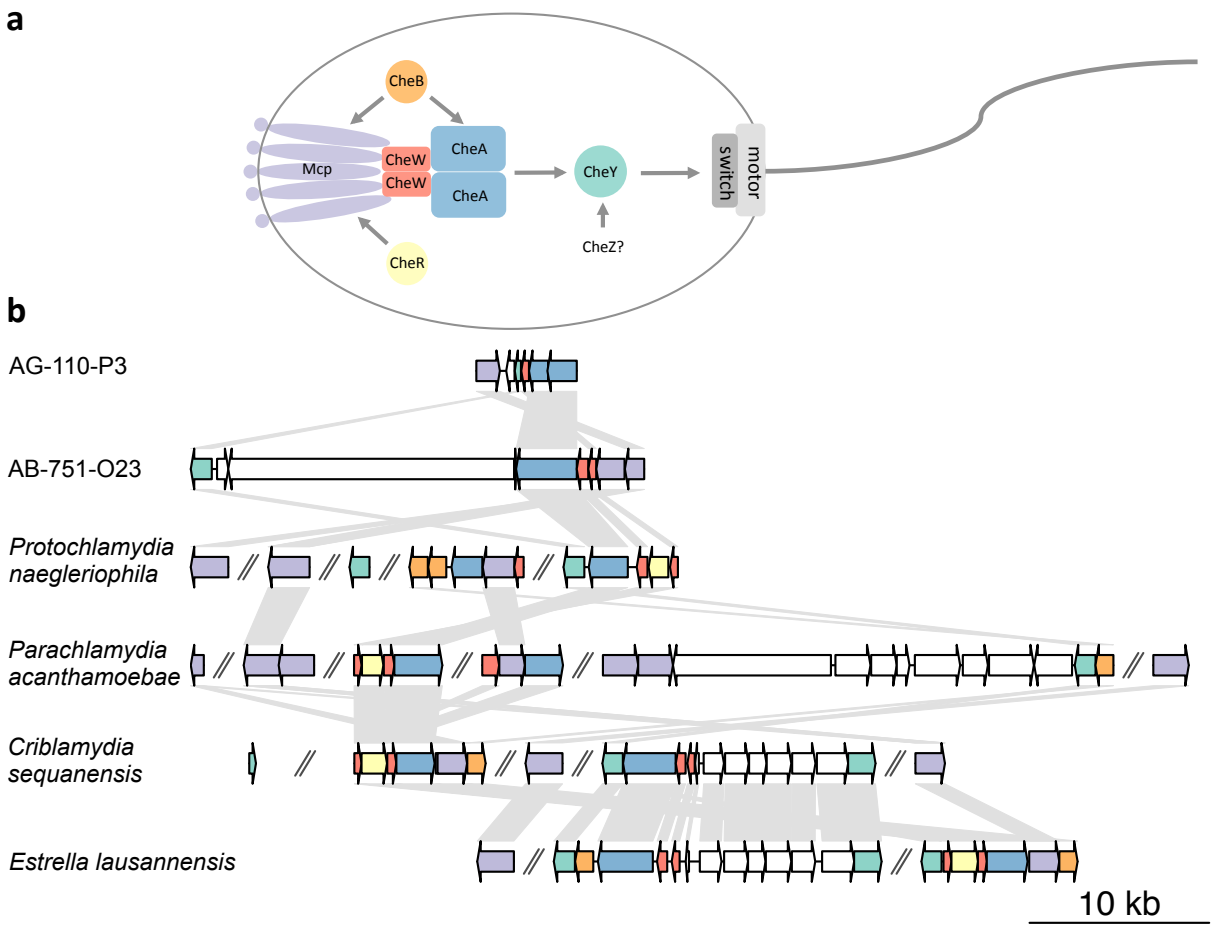
Single protein phylogenies were conducted for FlgL, FliS, FliQ/SctS, and FlhA/SctV with RAxML and Bayesian Inference. In addition, a concatenated dataset including FliPQR and FlhA was used (Figure 4, Supplementary Table 2). Where possible also homologous proteins present in NF-T3SSs were included in the datasets. As expected, none of the proteins analyzed showed a close relationship with known flagellar proteins from other taxonomic groups. With the exception of the phylogeny of FlhA, which is present only in AB-751-O23, and FlgL all trees demonstrated the monophyly for the SAG sequences (Figure 4, Supplementary Data). For FlgL, the flagellar hook-associated protein type 3, AB-751-O23 and AG-110-M15 were closest related without a sister lineage (Figure 4d), whereas the AG-110-P3 sequence grouped together with members of the *Verrucomicrobia*. FlgL is located extracellularly in the flagellum and therefore more prone to higher rates of evolution (Nogueira *et al.*, 2012), which might be the reason for the lack of monophyly for chlamydial proteins and the generally observed inconsistency of the branching pattern with 16S rRNA-based organismal phylogeny.

In the phylogenetic trees for FliQ/SctS (Figure 4c) and FlhA/SctV (Figure 4b) the SAGs formed individual deep branches, whereas in the cases of FliS (Figure 4e) and FliPQR-FlhA (Figure 4a) they formed sister clades to the *Verrucomicrobia* and the *Acidobacteria*, respectively. The monophyly of the SAGs was always highly supported independent of the treeing method applied, whereas the affiliation to the sister lineages showed only weak support due to the ancient origin of flagella and the lack of resolution close to the root of the tree (Figure 4, Supplementary Table 2, Supplementary Data).

Supplementary Table 2. Summary of phylogenetic analyses of selected flagellar proteins encoded in the SAGs.

Protein	Dataset	Presence in SAGs	Method	Monophyly of SAGs*	Closest related phyla in dataset*
FliQ/SctS	Flagellar and NF-T3SS proteins	All	MrBayes	Yes (1)	None
			RAxML	Yes (99)	Unclear
FlhA/SctV	Flagellar and NF-T3SS proteins	AB-751-O23	MrBayes	Single branch	Firmicutes and Spirochaetes (0.5)
			RAxML	Single branch	Firmicutes (22)
FliS	Flagellar proteins	AB-751-O23, AG-110-M15	MrBayes	Yes (1)	Verrucomicrobia (0.72)
			RAxML	Yes (85)	Verrucomicrobia (15)
FlgL	Flagellar proteins	All	MrBayes	AB-751-O23 and AG-110-M15: Yes (1); AG-110-P3: No	AB-751-O23 and AG-110-M15: none; AG-110-P3: Verrucomicrobia (0.97)
			RAxML	AB-751-O23 and AG-110-M15: Yes (100); AG-110-P3: No	AB-751-O23 and AG-110-M15: Thermotogae (26); AG-110-P3: Verrucomicrobia (51)
FliPQR, FlhA/ SctRST, SctV	Concatenated flagellar and chlamydial NF-T3SS proteins	All	MrBayes	Yes (1)	Acidobacteria (0.99)
			RAxML	Yes (100)	Acidobacteria (29)

*Numbers in brackets indicate posterior probability values (MrBayes) or bootstrap support values [%](RAxML).



Supplementary Figure 5. Presence and synteny of chemotaxis genes in chlamydial genomes.
a) Schematic overview of the arrangement of proteins involved in chemotaxis. Mcp – Methyl-accepting chemotaxis protein, CheW – coupling protein, CheA – histidine kinase, CheY – response regulator, CheR – methyltransferase, CheB methylesterase, CheZ –phosphatase.
b) Synteny of chemotaxis genes in chlamydial genomes. Color coding as in a).

Supplementary Table 3. Genome sequences used for the extraction of proteins used for phylogenetic analyses.

Replicon/Strain name	Abbreviation in tree files	NCBI Reference Sequence	System Type	NF-T3SS family	Replicon Type	Class	Order
Acetohalobium arabaticum DSM 5501	acear	NC_014378.1	Flg		chromosome	Firmicutes	Clostridia
Acidimicrobium ferrooxidans DSM 10331	AFER	NC_013124.1	Flg		chromosome	Actinobacteria	Acidimicrobidae
Acidobacterium capsulatum ATCC 51196	ACP	NC_012483.1	Flg		chromosome	Acidobacteria	Acidobacteriales
Actinosynnema mirum DSM 43827	AMIR	NC_013093.1	Flg		chromosome	Actinobacteria	Actinobacteridae
Aeromonas salmonicida subsp. salmonicida A449 plasmid 5	ASA	NC_009350.1	NF-T3SS	ysc	plasmid	Proteobacteria	Gammaproteobacteria
Agrobacterium tumefaciens strain Ach5	Ach5	NZ_CP011246.1	Flg		chromosome	Proteobacteria	Alphaproteobacteria
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446	AACI	NC_013205.1	Flg		chromosome	Firmicutes	Bacillales
Alkalilimnicola ehrlichii MLHE-1	MLG	NC_008340.1	Flg		chromosome	Proteobacteria	Gammaproteobacteria
Alkaliphilus metalliredigens QYMF	AMET	NC_009633.1	Flg		chromosome	Firmicutes	Clostridia
Alteromonas macleodii str. 'Deep ecotype'	MADE	NC_011138.3	Flg		chromosome	Proteobacteria	Gammaproteobacteria
Anaeromyxobacter dehalogenans 2CP-C	ADEH	NC_007760.1	NF-T3SS, Flg	Myxo	chromosome	Proteobacteria	Deltaproteobacteria
Anaeromyxobacter sp. Fw109-5	ANAE109	NC_009675.1	NF-T3SS	Myxo	chromosome	Proteobacteria	Deltaproteobacteria
Anaeromyxobacter sp. K	ANAEK	NC_011145.1	NF-T3SS	Myxo	chromosome	Proteobacteria	Deltaproteobacteria
Anoxybacillus flavithermus WK1	AFLV	NC_011567.1	Flg		chromosome	Firmicutes	Bacillales
Aquifex aeolicus VF5	aq	NC_000918.1	Flg		chromosome	Aquificae	Aquificales
Arthrobacter chlorophenolicus A6	ACHL	NC_011886.1	Flg		chromosome	Actinobacteria	Actinobacteridae
Azorhizobium caulinodans ORS 571 DNA	AZC	NC_009937.1	Flg		chromosome	Proteobacteria	Alphaproteobacteria
Bacillus clausii KSM-K16 DNA	ABC	NC_006582.1	Flg		chromosome	Firmicutes	Bacillales
Bacillus halodurans C-125 DNA	BH	NC_002570.2	Flg		chromosome	Firmicutes	Bacillales
Bacillus pumilus SAFR-032	BPUM	NC_009848.1	Flg		chromosome	Firmicutes	Bacillales
Bacillus subtilis subsp. subtilis str. 168	BSU	NC_000964.3	Flg		chromosome	Firmicutes	Bacillales
Blastopirellula marina DSM 3645	Blastopirellula	NZ_AANZ00000000.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Bordetella bronchiseptica strain RB50	BB_2, BB_RS	NC_002927.3	NF-T3SS	ysc	chromosome	Proteobacteria	Betaproteobacteria
Bordetella parapertussis Bpp5	BN117	NC_018828.1	NF-T3SS	ysc	chromosome	Proteobacteria	Betaproteobacteria
Bordetella pertussis Tohama I	BP	NC_002929.2	NF-T3SS	ysc	chromosome	Proteobacteria	Betaproteobacteria
Borrelia burgdorferi B31	BB_1, BB_	NC_001318.1	Flg		chromosome	Spirochaetes	Spirochaetales
Borrelia hermsii MTW	BHW	NZ_CP005680.1	Flg		chromosome	Spirochaetes	Spirochaetales
Brachyspira hyodysenteriae WA1	BHWA1	NC_012225.1	Flg		chromosome	Spirochaetes	Spirochaetales
Bradyrhizobium japonicum USDA 6 DNA	BJ6T	NC_017249.1	NF-T3SS	Rhizob	chromosome	Proteobacteria	Alphaproteobacteria
Buchnera aphidicola str. 5A (Acyrthosiphon pisum)	BUAP5A	NC_011833.1	Flg		plasmid	Proteobacteria	Gammaproteobacteria
Burkholderia ambifaria MC40-6 chromosome 1	BAMMC406	NC_010551.1	Flg		chromosome	Proteobacteria	Betaproteobacteria
Burkholderia ambifaria MC40-6 chromosome 2	BAMMC406	NC_010552.1	NF-T3SS	Hrp2	chromosome	Proteobacteria	Betaproteobacteria
Burkholderia ambifaria MC40-6 chromosome 3	BAMMC406	NC_010557.1	NF-T3SS	Hrp2	chromosome	Proteobacteria	Betaproteobacteria
Burkholderia mallei ATCC 23344 chromosome 1	BMA	NC_006348.1	Flg		chromosome	Proteobacteria	Betaproteobacteria
Burkholderia mallei ATCC 23344 chromosome 2	BMA	NC_006349.2	NF-T3SS	SPI-1, Hrp2	chromosome	Proteobacteria	Betaproteobacteria
Caldicellulosiruptor bescii DSM 6725	ATHE	NC_012034.1	Flg		chromosome	Firmicutes	Clostridia
Caldicellulosiruptor saccharolyticus DSM 8903	CSAC	NC_009437.1	Flg		chromosome	Firmicutes	Clostridia
Calditerrivibrio nitroreducens DSM 19672	CALNI	NC_014758.1	Flg		chromosome	Deferribacteres	Deferribacterales
Campylobacter jejuni RM1221	CJE	NC_003912.7	Flg		chromosome	Proteobacteria	Epsilonproteobacteria
Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)	HDEF	NC_012751.1	NF-T3SS	SPI-1, SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Candidatus Koribacter versatilis Ellin345	ACID345	NC_008009.1	Flg		chromosome	Acidobacteria	Candidatus Koribacter
Candidatus Nitrospira defluvii chromosome	NIDE	NC_014355.1	Flg		chromosome	Nitrospirae	Nitrospirales
Candidatus Protochlamydia amoebophila UWE25	PC	NC_005861.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Candidatus Rubidus massiliensis	Rubis, BN1013	CCSC01000001.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Caulobacter crescentus CB15	CC	NC_002696.2	Flg		chromosome	Proteobacteria	Alphaproteobacteria
Chlamydia abortus S26/3	CAB	NC_004552.2	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia avium 10DC88	RT28	NZ_CP006571.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia caviae GPIC	CCA	NC_003361.3	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales

Chlamydia felis Fe/C-56	CF	NC_007899.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia muridarum Nigg	TC	NC_002620.2	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia pecorum E58	G5S	NC_015408.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia pneumoniae CWL029	CPn	NC_000922.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia pneumoniae LPKoLN	CPK	NC_017287.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia psittaci 6BC	G5O	NC_017287.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia trachomatis A/HAR-13	CTA	NC_007429.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydia trachomatis D/UW-3/CX	CT	NC_000117.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydiae bacterium SM23_39	SM23_39	LJUH00000000.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydiales bacterium SCGC AB-751-O23	AB_751_O23	FLYF01000000.1	Flg, NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydiales bacterium SCGC AG-110-P3	AG_110_P3	FLYP01000000.1	NF-T3SS, Flg	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chlamydiales bacterium SCGC AG-110-M15	AG_110_M15	FLYO01000000.1	NF-T3SS, Flg	Chlamy	chromosome	Chlamydiae	Chlamydiales
Chromobacterium violaceum ATCC 12472	CV	NC_005085.1	Flg, NF-T3SS	SPI-1, SPI-2	chromosome	Proteobacteria	Betaproteobacteria
Citrobacter rodentium ICC168	ROD	NC_013716.1	NF-T3SS	SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Clostridium botulinum A str. Hall	CLC	NC_009698.1	Flg		chromosome	Firmicutes	Clostridia
Clostridium cellulolyticum H10	CCEL	NC_011898.1	Flg		chromosome	Firmicutes	Clostridia
Colwellia psychrerythraea 34H	CPS	NC_003910.7	Flg		chromosome	Chlamydiae	Chlamydiales
Criblamydia sequanensis	CSEC	NZ_CCEJ00000000.1	NF-T3SS	Chlamy	chromosome	Proteobacteria	Gammaproteobacteria
Cupriavidus taiwanensis str. LMG19424	RALTA	NC_010530.1	NF-T3SS	Hrp2	chromosome	Proteobacteria	Betaproteobacteria
Denitrovibrio acetiphilus DSM 12809	DACET	NC_013943.1	Flg		chromosome	Deferribacteres	Deferribacterales
Desulfovibrio vulgaris str. Hildenborough chromosome	DVU	NC_002937.3	Flg		chromosome	Proteobacteria	Deltaproteobacteria
Desulfovibrio vulgaris str. Hildenborough plasmid pDV	DVUA	NC_005863.1	NF-T3SS	ysc	plasmid	Proteobacteria	Deltaproteobacteria
Dickeya zeae Ech1591	DD1591	NC_012912.1	Flg, NF-T3SS	Hrp1	chromosome	Proteobacteria	Gammaproteobacteria
Diplosphaera colitermitum TAV2	Diploshpaera	NZ_ABEA00000000.3	Flg		chromosome	Verrucomicrobia	Opitutae
Erwinia amylovora ATCC 49946	EAM	NC_013971.1	NF-T3SS, Flg	SPI-1, SPI-1, Hrp	chromosome	Proteobacteria	Gammaproteobacteria
Erwinia pyrifoliae strain Ep1/96	EPC	NC_012214.1	NF-T3SS, Flg	SPI-1, Hrp1	chromosome	Proteobacteria	Gammaproteobacteria
Escherichia coli O157:H7 str. Sakai chromosome	ECs	NC_002695.1	NF-T3SS, Flg	SPI-1, SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Estrella lausannensis CRIB-30	Elau	CWGJ01000000	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Exiguobacterium sibiricum 255-15	EXIG	NC_010556.1	Flg		chromosome	Firmicutes	Bacillales
Gemmata obscuriglobus UQM2246	Gemmata	NZ_ABGO00000000.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Gemmatimonas aurantiaca T-27 DNA	GAU	NC_012489.1	Flg		chromosome	Gemmatimonadet	Gemmatimonadales
Geobacillus kaustophilus HTA426 DNA	GK	NC_006510.1	Flg		chromosome	Firmicutes	Bacillales
Gimesia maris DSM8797	Gimesia	NZ_ABCE00000000.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Hahella chejuensis KCTC 2396	HCH	NC_007645.1	NF-T3SS, Flg	ysc, ysc	chromosome	Proteobacteria	Gammaproteobacteria
Herbaspirillum seropedicae SmR1	HSERO	NC_014323.1	NF-T3SS	Hrp1	chromosome	Proteobacteria	Betaproteobacteria
Hydrogenobaculum sp. Y04AAS1	HY04AAS1	NC_011126.1	Flg		chromosome	Aquificae	Aquificales
Isosphaera pallida ATCC43644	Isosphaera	NC_014962.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Legionella pneumophila str. Paris	LPP	NC_006368.1	Flg		chromosome	Proteobacteria	Gammaproteobacteria
Leptospira interrogans serovar Lai str. 56601 chromosome I	LA	NC_004342.2	Flg		chromosome	Spirochaetes	Spirochaetales
Marinomonas mediterranea MMB-1	MARME	NC_015276.1	NF-T3SS	Hrp1	chromosome	Proteobacteria	Gammaproteobacteria
Mesorhizobium loti MAFF303099 DNA	MAFF	NC_002678.2	Flg, NF-T3SS	Rhizob	chromosome	Proteobacteria	Alphaproteobacteria
Myxococcus fulvus HW-1	LILAB	NC_015711.1	NF-T3SS	Myxo, Myxo	chromosome	Proteobacteria	Deltaproteobacteria
Myxococcus xanthus DK 1622	MXAN	NC_008095.1	NF-T3SS	Myxo, Myxo	chromosome	Proteobacteria	Deltaproteobacteria
Neochlamydia sp. EPS4	EPS4, DB42	NZ_JSDQ00000000.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Neochlamydia sp. S13	NeoS13	NZ_BASK00000000.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Neochlamydia sp. TUME1	TUME1	NZ_JRXI00000000.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Opitutaceae bacterium TAV1	Opitutaceae	NZ_AHKS00000000.1	Flg		chromosome	Verrucomicrobia	Opitutae
Opitutus terrae PB90-1	Opitutus, OTER	NC_010571.1	Flg		chromosome	Verrucomicrobia	Opitutae
Parachlamydia acanthamoebae UV-7	PUV	NC_015702.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Parachlamydiaceae bacterium HS-T3	HS_T3	NZ_BBPT00000000.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Petrotoga mobilis SJ95	PMOB	NC_010003.1	Flg		chromosome	Thermotogae	Thermotogales
Photorhabdus luminescens subsp. laumondii TTO1	PLU	NC_005126.1	Flg, NF-T3SS	ysc	chromosome	Proteobacteria	Gammaproteobacteria

Phycisphaera mikurensis NBRC 102666	Phycisphaera	NC_017080.1	Flg		chromosome	Planctomycetes	Phycisphaera
Pirellula staleyi DSM 6068	PSTA, Pirellula	NC_013720.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Planctopirus limnophila	Planctopirus	NC_014148.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Proteus mirabilis strain HI4320	PMI	NC_010554.1	NF-T3SS, Flg	SPI-1	chromosome	Proteobacteria	Gammaproteobacteria
Protochlamydia naegleriophila	ProNaeg	NZ_LN879502.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Pseudomonas aeruginosa PAO1	PA	NC_002516.2	NF-T3SS	ysc	chromosome	Proteobacteria	Gammaproteobacteria
Pseudomonas syringae pv. syringae HS191	PSYRH	NZ_CP006256.1	Flg, NF-T3SS	Hrp1	chromosome	Proteobacteria	Gammaproteobacteria
Ralstonia solanacearum CMR15 chromosome	CMR15	NC_017559.1	Flg		chromosome	Proteobacteria	Betaproteobacteria
Ralstonia solanacearum CMR15 plasmid CMR15mp	CMR15	NC_017589.1	NF-T3SS	Hrp2	plasmid	Proteobacteria	Betaproteobacteria
Rhabdochlamydia porcellionis 15C	Rhab	single proteins	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Rhodopirellula baltica SH1	Rhodopirellula	NC_005027.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Rubinisphaera brasiliensis DSM5305	Rubinisphaera	NC_015174.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Salmonella enterica subsp. enterica serovar Typhi str. CT18	STY	NC_003198.1	Flg, NF-T3SS	SPI-1, SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	STM	NC_003197.1	Flg, NF-T3SS	SPI-1, SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Schlesneria paludicola DSM18645	Schlesneria	NZ_AHZR00000000.1	Flg		chromosome	Planctomycetes	Planctomycetacia
Shewanella violacea DSS12 DNA	SVI	NC_014012.1	NF-T3SS	SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Simkania negevensis Z	SNE	NC_015713.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Sinorhizobium fredii NGR234 chromosome	NGR	NC_012587.1	Flg		chromosome	Proteobacteria	Alphaproteobacteria
Sinorhizobium fredii NGR234 plasmid pNGR234a	NGR	NC_000914.2	NF-T3SS	Rhizob	plasmid	Proteobacteria	Alphaproteobacteria
Sinorhizobium fredii NGR234 plasmid pNGR234b	NGR	NC_012586.1	NF-T3SS	Rhizob	plasmid	Proteobacteria	Alphaproteobacteria
Sodalis glossinidius str. 'morsitans' DNA	SGP1	NC_007712.1	Flg, NF-T3SS	SPI-1, SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Spirochaeta thermophila DSM 6578	SPITH	NC_017583.1	Flg		chromosome	Spirochaetes	Spirochaetales
Stigmatella aurantiaca DW4/3-1	STAUR	NC_014623.1	NF-T3SS	Myxo, Myxo	chromosome	Proteobacteria	Deltaproteobacteria
Sulfurihydrogenibium azorense Az-Fu1	SULAZ	NC_012438.1	Flg		chromosome	Aquificae	Aquificales
Thermoanaerobacter tengcongensis MB4	TTE	NC_003869.1	Flg		chromosome	Firmicutes	Clostridia
Thermosipho melanesiensis BI429	TMEL	NC_009616.1	Flg		chromosome	Thermotogae	Thermotogales
Thermotoga maritima MSB8	TM	NC_000853.1	Flg		chromosome	Thermotogae	Thermotogales
Thioalkalivibrio sulfidophilus HL-EbGr7	TGR7	NC_011901.1	Flg		chromosome	Proteobacteria	Gammaproteobacteria
Treponema denticola ATCC 35405	TDE	NC_002967.9	Flg		chromosome	Spirochaetes	Spirochaetales
Variovorax paradoxus EPS	VARPA	NC_014931.1	NF-T3SS	Hrp2	chromosome	Proteobacteria	Betaproteobacteria
Verrucomicrobium spinosum DSM 4136	Verrucomicrobium, VSP	NZ_ABIZ01000001.1	NF-T3SS		chromosome	Verrucomicrobia	Verrucomicrobiae
Vibrio parahaemolyticus RIMD 2210633 chromosome 1	VP	NC_004603.1	Flg, NF-T3SS	ysc	chromosome	Proteobacteria	Gammaproteobacteria
Vibrio parahaemolyticus RIMD 2210633 chromosome 2	VPA	NC_004605.1	Flg		chromosome	Proteobacteria	Gammaproteobacteria
Waddlia chondrophila WSU 86-1044	WCW	NC_014225.1	NF-T3SS	Chlamy	chromosome	Chlamydiae	Chlamydiales
Xanthomonas campestris pv. campestris str. ATCC 33913	XCC	NC_003902.1	Flg, NF-T3SS	Hrp2	chromosome	Proteobacteria	Gammaproteobacteria
Yersinia enterocolitica subsp. enterocolitica 8081 chromosome	YE	NC_008800.1	Flg, NF-T3SS	SPI-1	chromosome	Proteobacteria	Gammaproteobacteria
Yersinia enterocolitica subsp. enterocolitica 8081 plasmid pYVe8081	YEP	NC_008791.1	NF-T3SS	ysc	plasmid	Proteobacteria	Gammaproteobacteria
Yersinia pestis CO92 chromosome	YPO	NC_003143.1	Flg, NF-T3SS	SPI-2	chromosome	Proteobacteria	Gammaproteobacteria
Yersinia pestis CO92 plasmid pCD1	YPCD1	NC_003131.1	NF-T3SS	ysc	plasmid	Proteobacteria	Gammaproteobacteria

Sequences of marked species were extracted from protein fasta files downloaded from NCBI instead of genbank files.

Supplementary Table 4. Metagenomes harboring FlhA sequences closely related to AB-751-O23

Gene ID	Locus Tag	Gene Product Name	Genome
3300002835	B570J40625_10000123015	Flagellar biosynthesis pathway, component FlhA	Freshwater microbial communities from Lake Mendota, WI - 13AUG2008 deep hole epilimnion (Lake Mendota Combined Ray assembly, ASSEMBLY_DATE=20140605) (Lake Mendota Combined Ray assembly, ASSEMBLY_DATE=20140605) (*) (MER-FS) (assembled)
3300002152	C687J26660_10047071	Flagellar biosynthesis pathway, component FlhA	Groundwater microbial communities from Rifle, Colorado - Rifle Oxygen_injection A3 (Rifle Oxygen_injection A3, ASSEMBLY_DATE=20131127) (*) (MER-FS) (assembled)
3300002503	C687J35164_100025531	Flagellar biosynthesis pathway, component FlhA	Soil microbial communities from Rifle, Colorado - Rifle CSP2_sed 13_3 (Uranium-contaminated soil microbial communities from Rifle, Colorado - Rifle CSP2_sed 13_3, ASSEMBLY_DATE=20140303) (*) (MER-FS) (assembled)
3300004274	Ga0066607_10054914	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI075_LV_DNA_120m (*) (MER-FS) (assembled)
3300004277	Ga0066611_100192532	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI075_LV_DNA_200m (*) (MER-FS) (assembled)
3300005351	Ga0074239_11440264	Flagellar biosynthesis pathway, component FlhA	Bioreactor Anammox bacterial community from Nijmegen, The Netherlands - Scalindua species enirchment (re-annotation) (*) (MER-FS) (assembled)
3300001380	JGI1356J14229_1000445110	Flagellar biosynthesis pathway, component FlhA	Subsurface groundwater microbial communities from S. Glens Falls, New York, USA - GMW37 contaminated, 5.8 m (Subsurface groundwater monitoring well GMW37 contaminated, 5.8 m, ASSEMBLY_DATE=20130411) (*) (MER-FS) (assembled)
3300001380	JGI1356J14229_100197444	Flagellar biosynthesis pathway, component FlhA	Subsurface groundwater microbial communities from S. Glens Falls, New York, USA - GMW37 contaminated, 5.8 m (Subsurface groundwater monitoring well GMW37 contaminated, 5.8 m, ASSEMBLY_DATE=20130411) (*) (MER-FS) (assembled)
3300001380	JGI1356J14229_100495682	Flagellar biosynthesis pathway, component FlhA	Subsurface groundwater microbial communities from S. Glens Falls, New York, USA - GMW37 contaminated, 5.8 m (Subsurface groundwater monitoring well GMW37 contaminated, 5.8 m, ASSEMBLY_DATE=20130411) (*) (MER-FS) (assembled)
3300003494	JGI26240J51127_10029472	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S3LV_150m_DNA (Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S3LV_150m_DNA, ASSEMBLY_DATE=20140926) (*) (MER-FS) (assembled)
3300003492	JGI26245J51145_10065432	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S4LV_200m_DNA (Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S4LV_200m_DNA, ASSEMBLY_DATE=20140930) (*) (MER-FS) (assembled)
3300003582	JGI26252J51714_10008226	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI073_LV_10m_DNA (Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI073_LV_10m_DNA, ASSEMBLY_DATE=20141005) (*) (MER-FS) (assembled)
3300003599	JGI26270J51728_10101871	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S3LV_10m_DNA (Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S3LV_10m_DNA, ASSEMBLY_DATE=20141005) (*) (MER-FS) (assembled)
3300003600	JGI26272J51733_10026096	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S3LV_120m_DNA (Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - SI037_S3LV_120m_DNA, ASSEMBLY_DATE=20141004) (*) (MER-FS) (assembled)
3300000255	LP_F_10_SI03_135DRAFT_10022538	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in Line P, North Pacific Ocean - sample_F_10_SI03_135 (Line P sample_F_10_SI03_135, March 2012 Assem) (*) (MER-FS) (assembled)
3300000237	SI34jun09_150mDRAFT_10045362	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 34 06/16/09 150m (Saanich Inlet 34 06/16/09 150m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000172	SI34jun09_200mDRAFT_10088612	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 34 06/16/09 200m (Saanich Inlet 34 06/16/09 200m, May 2012 Assem) (*) (MER-FS) (assembled)
3300000172	SI34jun09_200mDRAFT_10116882	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 34 06/16/09 200m (Saanich Inlet 34 06/16/09 200m, May 2012 Assem) (*) (MER-FS) (assembled)
3300000170	SI36aug09_135mDRAFT_10079322	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 36 08/11/09 135m (Saanich Inlet 36 08/11/09 135m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000167	SI39nov09_120mDRAFT_10136222	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 39 11/10/09 120m (Saanich Inlet 39 11/10/09 120m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000212	SI47jul10_120mDRAFT_10025297	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 47 07/07/10 120m (Saanich Inlet 47 07/07/10 120m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000193	SI47jul10_135mDRAFT_10037382	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 47 07/07/10 135m (Saanich Inlet 47 07/07/10 135m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000154	SI47jul10_150mDRAFT_10029228	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 47 07/07/10 150m (Saanich Inlet 47 07/07/10 150m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000171	SI47jul10_200mDRAFT_10063161	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 47 07/07/10 200m (Saanich Inlet 47 07/07/10 200m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000150	SI48aug10_120mDRAFT_10021921	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 48 08/11/10 120m (Saanich Inlet 48 08/11/10 120m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000215	SI53jan11_120mDRAFT_10015725	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 53 01/11/11 120m (Saanich Inlet 53 01/11/11 120m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000211	SI53jan11_135mDRAFT_10008994	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 53 01/11/11 135m (Saanich Inlet 53 01/11/11 135m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000216	SI53jan11_150mDRAFT_10069052	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 53 01/11/11 150m (Saanich Inlet 53 01/11/11 150m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000201	SI54feb11_135mDRAFT_100045811	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 54 02/08/11 135m (Saanich Inlet 54 02/08/11 135m, March 2012 Assem) (*) (MER-FS) (assembled)
3300000214	SI54feb11_200mDRAFT_10013984	Flagellar biosynthesis pathway, component FlhA	Marine microbial communities from expanding oxygen minimum zones in the Saanich Inlet - 54 02/08/11 200m (Saanich Inlet 54 02/08/11 200m, March 2012 Assem) (*) (MER-FS) (assembled)

Supplementary References

Abby SS, Rocha EPC. (2012). The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet* **8**: e1002983.

Barker CS, Meshcheryakova IV, Kostyukova AS, Samatey FA. (2010). FliO regulation of FliP in the formation of the Salmonella enterica flagellum. *PLoS Genet* **6**: e1001143.

Charbonneau M-E, Berthiaume F, Mourez M. (2006). Proteolytic processing is not essential for multiple functions of the Escherichia coli autotransporter adhesin involved in diffuse adherence (AIDA-I). *J Bacteriol* **188**: 8504–8512.

Chevance FFV, Hughes KT. (2008). Coordinating assembly of a bacterial macromolecular machine. *Nat Rev Microbiol* **6**: 455–465.

Domman D, Collingro A, Lagkouvardos I, Gehre L, Weinmaier T, Rattei T, *et al.* (2014). Massive expansion of Ubiquitination-related gene families within the Chlamydiae. *Mol Biol Evol* **31**: 2890–2904.

Elwell C, Mirrashidi K, Engel J. (2016). Chlamydia cell biology and pathogenesis. *Nat Rev Microbiol* **14**: 385–400.

Ferrell JC, Fields KA. (2016). A working model for the type III secretion mechanism in Chlamydia. *Microbes Infect Inst Pasteur* **18**: 84–92.

Gloeckl S, Ong VA, Patel P, Tyndall JDA, Timms P, Beagley KW, *et al.* (2013). Identification of a serine protease inhibitor which causes inclusion vacuole reduction and is lethal to Chlamydia trachomatis. *Mol Microbiol* **89**: 676–689.

Klemm P, Vejborg RM, Sherlock O. (2006). Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol IJMM* **296**: 187–195.

Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M. (2014). Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the Chlamydiae. *ISME J* **8**: 115–125.

Nelson KE, Eisen JA, Fraser CM. (2001). Genome of Thermotoga maritima MSB8. *Methods Enzymol* **330**: 169–180.

Nogueira T, Touchon M, Rocha EPC. (2012). Rapid evolution of the sequences and gene repertoires of secreted proteins in bacteria. *PloS One* **7**: e49403.

Patel P, De Boer L, Timms P, Huston WM. (2014). Evidence of a conserved role for Chlamydia HtrA in the replication phase of the chlamydial developmental cycle. *Microbes Infect Inst Pasteur* **16**: 690–694.

Peters J, Wilson DP, Myers G, Timms P, Bavoil PM. (2007). Type III secretion à la Chlamydia. *Trends Microbiol* **15**: 241–251.

Pizzetti I, Fazi S, Fuchs BM, Amann R. (2012). High abundance of novel environmental chlamydiae in a Tyrrhenian coastal lake (Lago di Paola, Italy). *Environ Microbiol Rep* **4**: 446–452.

Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML, *et al.* (2008). Abundance and diversity of microbial life in ocean crust. *Nature* **453**: 653–656.

Stone CB, Bulir DC, Gilchrist JD, Toor RK, Mahony JB. (2010). Interactions between flagellar and type III secretion proteins in Chlamydia pneumoniae. *BMC Microbiol* **10**: 18.

Tsang J, Hoover TR. (2014). Requirement of the flagellar protein export apparatus component FliO for optimal expression of flagellar genes in Helicobacter pylori. *J Bacteriol* **196**: 2709–2717.