Unity in Variety – the Pan-Genome of the *Chlamydiae*

Astrid Collingro, et al.

Supplementary Material

Supplementary Text

The chlamydial core gene set

Among the *Chlamydiae*-specific genes are 18 hypothetical protein-coding genes and genes encoding the plasmid protein pGP6-D. Interestingly, three genes are involved in type III secretion underlining the importance of this system in chlamydial biology. The *Chlamydiae*-specific genes also encode a potential regulatory protein, the histone-like protein HctA, the Ser/Thr-protein kinase PknD, homologs of CT538, a gene up-regulated during Fe-depletion (Dill, Dessus-Babus, Raulston 2009) and CT043, which recently has been shown to be a major chlamydial antigen and a potential vaccine candidate (Meoni et al. 2009).

Orthologs shared by *Simkania*, *Waddlia*, and *Parachlamydiaceae*

Simkania, *Waddlia*, and the *Parachlamydiaceae* have retained many genes, which were described recently as absent in obligate intracellular bacteria (Merhej et al. 2009). Of the 100 COGs thought to be lost in a function-dependent, non-random way by obligate intracellular bacteria, 17 are present across all three groups, and an additional 35 COGs are present in at least one member of the three groups (supplementary table S4). This shows that despite a long evolutionary history as intracellular bacteria, *Simkania*, *Waddlia*, and the *Parachlamydiaceae*, have retained several functions generally absent in other intracellular mutualists or parasites studied so far.

The chlamydial type III secretion system

SctC, which forms the outer membrane ring, is present in two copies in *Parachlamydia* and *Simkania*. And also SctJ, a lipoprotein spanning the periplasmic space, is encoded twice in the *Parachlamydia* genome. In *Waddlia,* SctQ, a component of the basal body required for structural assembly of the T3S machinery, is split into two genes, a situation which is different from all other *Chlamydiae* but similar to a number of plant pathogens including *Pseudomonas* spp. (Fadouloglou et al. 2004). Recently, the needle forming protein SctF of the chlamydial T3S system was identified (Betts et al. 2008). Interestingly, a homolog of this protein is not found in the *Simkania* genome; it might have been replaced by another protein encoded at the same genomic locus (figure 4A).

Type III effector proteins

In addition to Inc proteins, several other T3S effectors of the *Chlamydiaceae* are known to be secreted into the host cell or the chlamydial inclusion (Beeckman, Vanrompay 2009). Some of these proteins are also encoded in the *Waddlia*, *Simkania*, *Parachlamydia* or *Protochlamydia* genomes (table 2, supplementary table S10). For example, the Ser/Thr kinase PknD, which phosphorylates SctD and seems to play a major role in chlamydial growth (Johnson et al. 2009), the Ser/Thr protease Pkn5, and the negative regulator of the T3S system SctW (CopN) are present. In addition, the CPn0827 (CT560) homolog, which was shown to be expressed throughout the developmental cycle and to be secreted from 20 h post infection on in *C. pneumoniae* (Herrmann et al. 2006), as well as the macrophage infectivity potentiator protein Mip, a surface exposed lipoprotein of EBs able to induce host immune response (Neff et al. 2007; Bas et al. 2008) were also present. The T3S translocator component CopB is present in all *Chlamydiae* except for *Simkania*. The methyltransferase (CPn0178), thought to be specific to some *Chlamydiaceae*, is present in *Parachlamydia* and *Simkania*

The chlamydial outer membrane

The *Chlamydiaceae* possess truncated pathways for LPS synthesis and are characterized by unique LPS containing a KDOp-(2-8)-KDO linkage (Brade, Brade, Nano 1987). The LPS pathway in *Simkania*, *Waddlia* and the *Parachlamydiaceae* differs from that of the *Chlamydiaceae*. All *Chlamydiae* encode a characteristic KDO-transferase (*kdtA*), but *Simkania*, *Waddlia* and the *Parachlamydiaceae* encode five additional proteins involved in LPS biosynthesis. The lipid A biosynthesis lauroyl acyltransferase (*htrB*) is present in the *Chlamydiaceae* and *Simkania* only. These differences in the LPS synthesis pathway are consistent with the observation that antibodies targeting the specific LPS epitope of the *Chlamydiaceae* do not recognize *Simkania*, *Waddlia* and the *Parachlamydiaceae* (Henning et al. 2002; Friedman, Dvoskin, Kahane 2003).

A conjugative plasmid in *Simkania*

Typically F-plasmids carry modules for plasmid stability, replication, adaptation and propagation (reviewed in (Norman, Hansen, Sorensen 2009)). pSn encodes a DNA helicase for plasmid replication. Usually the copy number of F-plasmids is controlled by Rep replicases, which are not present on pSn, but there is evidence for components of systems

for active partitioning and toxin/antitoxin maintenance (Norman, Hansen, Sorensen 2009; Van Melderen, Saavedra De Bast 2009) (supplementary table S13).

Several proteins encoded on pSn play a potential role in host adaptation. Some of these are involved in protection against heavy metals, while others are involved in metabolic processes like starch and glycerophospholipid metabolism. Some proteins are present that could play a role in pathogenicity or host cell interaction. Two proteins harbour ankyrin or TPR repeats, and two show weak sequence similarity to colicins. pSn also encodes a SWIB-domain protein, which is present in all *Chlamydiae*. It is likely to play a role in chromosome condensation-decondensation either in the chlamydial cell or its host (Bennett-Lovsey et al. 2002). Two proteins contain pectin lyase fold motifs; proteins with this motif are found in plant and human pathogens, e.g. the pertactin P69 of *Bordetella pertussis*, which mediates adhesion to mammalian cells (Jenkins, Mayans, Pickersgill 1998). In addition, one of three Mip homologs of *Simkania* is encoded on pSn.

Many F-plasmids contain mobility regions where proteins involved in transposition are encoded. Eleven proteins putatively involved in transposition and an RNA-directed DNA polymerase are present on pSn (supplementary table S13). Two integrases have homologs on the plasmids of the *Chlamydiaceae*.

Interestingly, a homolog of the pilin ($TraA_F$) forming the pilus structure could not be identified; however pilin protein sequences are usually not well conserved. Pilin proteins are typically short proteins with two transmembrane helices and a protein cleavage site, yielding a mature protein of about 70-90 amino acid residues (Lawley et al. 2003). No such protein exists in the genomic neighborhood of the *tra*-region in *Simkania*. This putative lack of a pilin protein might also explain why no *Simkania* pili have been reported so far. In other organisms, accumulating data show that efficient transfer of DNA can take place via direct cell-to-cell contact rather than through the F-pilus itself (Alvarez-Martinez, Christie 2009) and thus the production of pili might be dispensable.

One of these, $TraA_{Ti}$, is the relaxase/helicase of this system and could nick the DNA strand at *oriT* in *Simkania*, piloting DNA through the translocation channel and recircularizing the transferred DNA strand in the recipient cell. $TraD_{Ti}$ is likely to be an accessory protein for $TraA_{Ti}$ and has been shown to be important for efficient DNA nicking and transport (Cho, Winans 2005).

References

Alvarez-Martinez, CE, PJ Christie. 2009. Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev 73:775-808.

Bas, S, L Neff, M Vuillet, U Spenato, T Seya, M Matsumoto, C Gabay. 2008. The proinflammatory cytokine response to Chlamydia trachomatis elementary bodies in human macrophages is partly mediated by a lipoprotein, the macrophage infectivity potentiator, through TLR2/TLR1/TLR6 and CD14. J Immunol 180:1158-1168.

Beeckman, DS, DC Vanrompay. 2009. Bacterial Secretion Systems with an Emphasis on the Chlamydial Type III Secretion System. Curr Issues Mol Biol 12:17-42.

Bennett-Lovsey, R, SE Hart, H Shirai, K Mizuguchi. 2002. The SWIB and the MDM2 domains are homologous and share a common fold. Bioinformatics 18:626-630.

Betts, HJ, LE Twiggs, MS Sal, PB Wyrick, KA Fields. 2008. Bioinformatic and biochemical evidence for the identification of the type III secretion system needle protein of Chlamydia trachomatis. J Bacteriol 190:1680-1690.

Brade, H, L Brade, FE Nano. 1987. Chemical and serological investigations on the genus-specific lipopolysaccharide epitope of Chlamydia. Proc Natl Acad Sci U S A 84:2508-2512.

Cho, H, SC Winans. 2005. VirA and VirG activate the Ti plasmid repABC operon, elevating plasmid copy number in response to wound-released chemical signals. Proc Natl Acad Sci U S A 102:14843-14848.

Dill, BD, S Dessus-Babus, JE Raulston. 2009. Identification of iron-responsive proteins expressed by Chlamydia trachomatis reticulate bodies during intracellular growth. Microbiology 155:210-219.

Fadouloglou, VE, AP Tampakaki, NM Glykos, MN Bastaki, JM Hadden, SE Phillips, NJ Panopoulos, M Kokkinidis. 2004. Structure of HrcQB-C, a conserved component of the bacterial type III secretion systems. Proc Natl Acad Sci U S A 101:70-75.

Friedman, MG, B Dvoskin, S Kahane. 2003. Infections with the chlamydia-like microorganism *Simkania negevensis*, a possible emerging pathogen. Microbes Infect 5:1013-1021.

Henning, K, G Schares, H Granzow, U Polster, M Hartmann, H Hotzel, K Sachse, M Peters, M Rauser. 2002. *Neospora caninum* and *Waddlia chondrophila* strain 2032/99 in a septic stillborn calf. Vet Microbiol 85:285-292.

Herrmann, M, A Schuhmacher, I Muhldorfer, K Melchers, C Prothmann, S Dammeier. 2006. Identification and characterization of secreted effector proteins of *Chlamydophila pneumoniae* TW183. Res Microbiol 157:513-524.

Jenkins, J, O Mayans, R Pickersgill. 1998. Structure and evolution of parallel beta-helix proteins. J Struct Biol 122:236-246.

Johnson, DL, CB Stone, DC Bulir, BK Coombes, JB Mahony. 2009. A novel inhibitor of Chlamydophila pneumoniae protein kinase D (PknD) inhibits phosphorylation of CdsD and suppresses bacterial replication. BMC Microbiol 9:218.

Lawley, TD, WA Klimke, MJ Gubbins, LS Frost. 2003. F factor conjugation is a true type IV secretion system. FEMS Microbiol Lett 224:1-15.

Meoni, E, E Faenzi, E Frigimelica, et al. 2009. CT043, a protective antigen that induces a CD4+ Th1 response during Chlamydia trachomatis infection in mice and humans. Infect Immun 77:4168-4176.

Merhej, V, M Royer-Carenzi, P Pontarotti, D Raoult. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4:13.

Neff, L, S Daher, P Muzzin, U Spenato, F Gulacar, C Gabay, S Bas. 2007. Molecular characterization and subcellular localization of macrophage infectivity potentiator, a Chlamydia trachomatis lipoprotein. J Bacteriol 189:4739-4748.

Norman, A, LH Hansen, SJ Sorensen. 2009. Conjugative plasmids: vessels of the communal gene pool. Philos Trans R Soc Lond B Biol Sci 364:2275-2289.

Van Melderen, L, M Saavedra De Bast. 2009. Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet 5:e1000437.

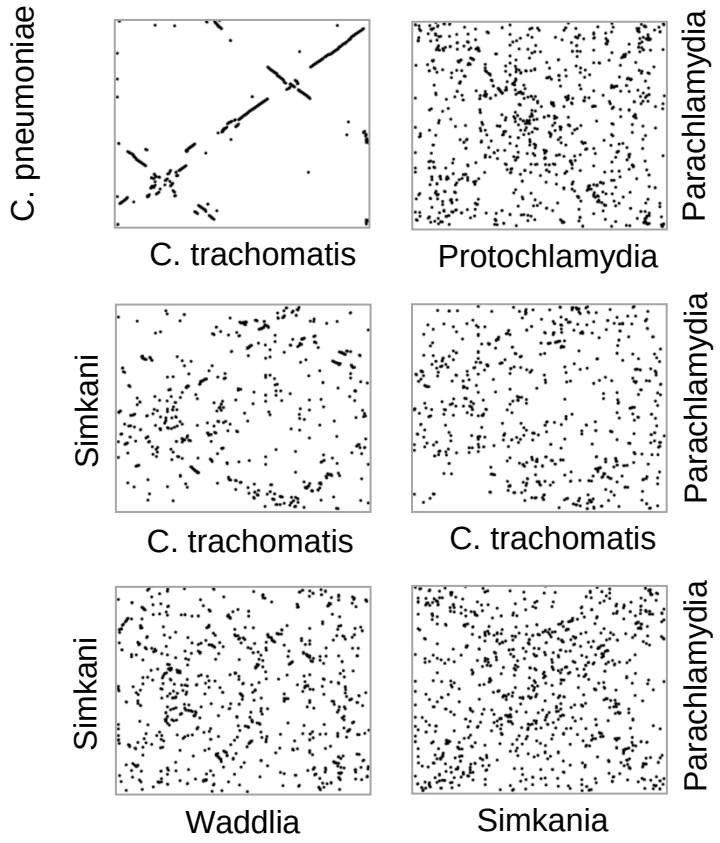

Figure S1: Synteny of chlamydial genomes. BLAST dot plots showing the high degree of conservation of gene order between the genomes of the *Chlamydiaceae* and the lack of synteny within in the *Parachlamydiaceae* and between the members of the different chlamydial families, respectively. Only complete genome sequences were used. Plots were generated using the RAST annotation platform (Aziz et al. 2008).

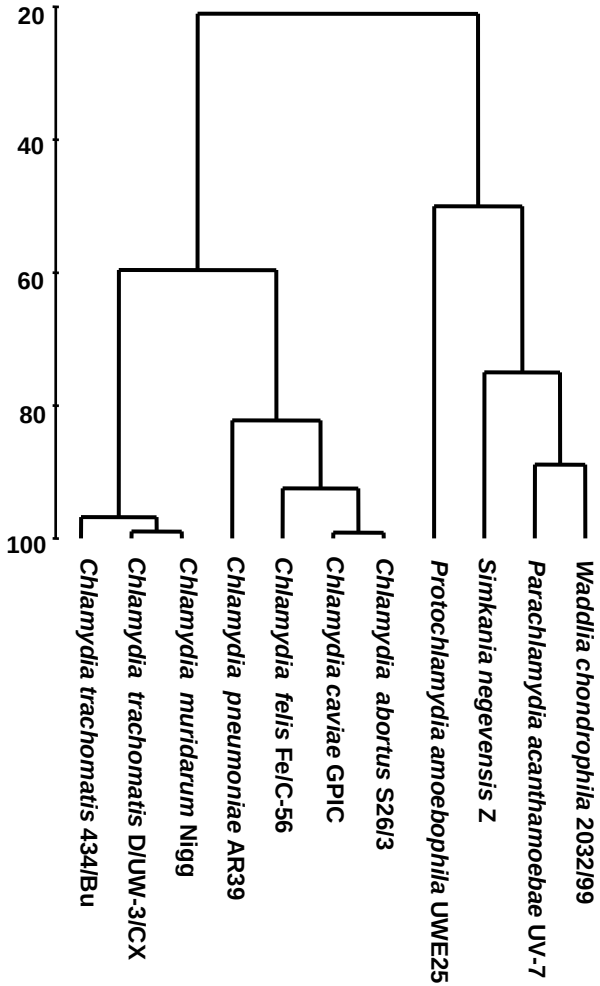

Figure S2: Cluster analysis of the presence/absence of Inc candidate proteins among the *Chlamydiae.* The presence or absence of Inc homologues was compared with a presence/absence based Bray-Curtis similarity matrix and clustered with complete linkage with the program PRIMER 5.0.

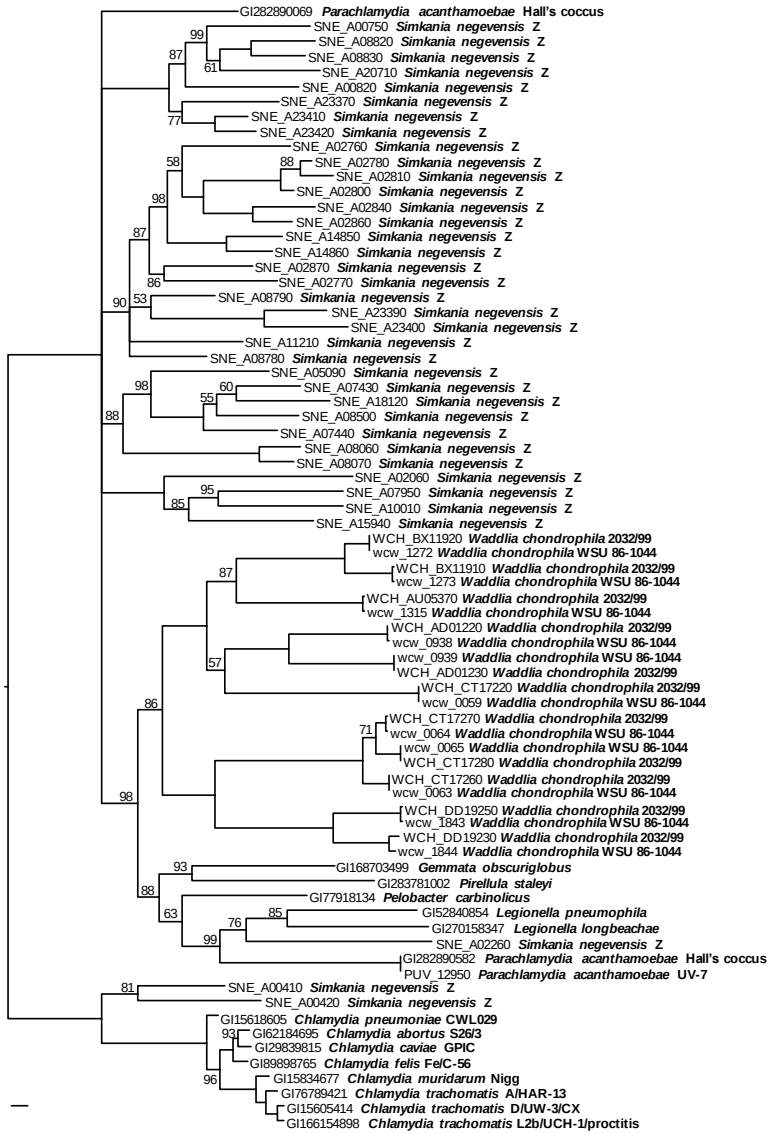

Fig. S2: Evolutionary history of MOMP and MOMP-like proteins. Bayesian inference analysis of MOMP and MOMP-like proteins. Scale bar represents 0.1 expected changes per site. If not indicated Bayesian posterior probabilities are 100%.

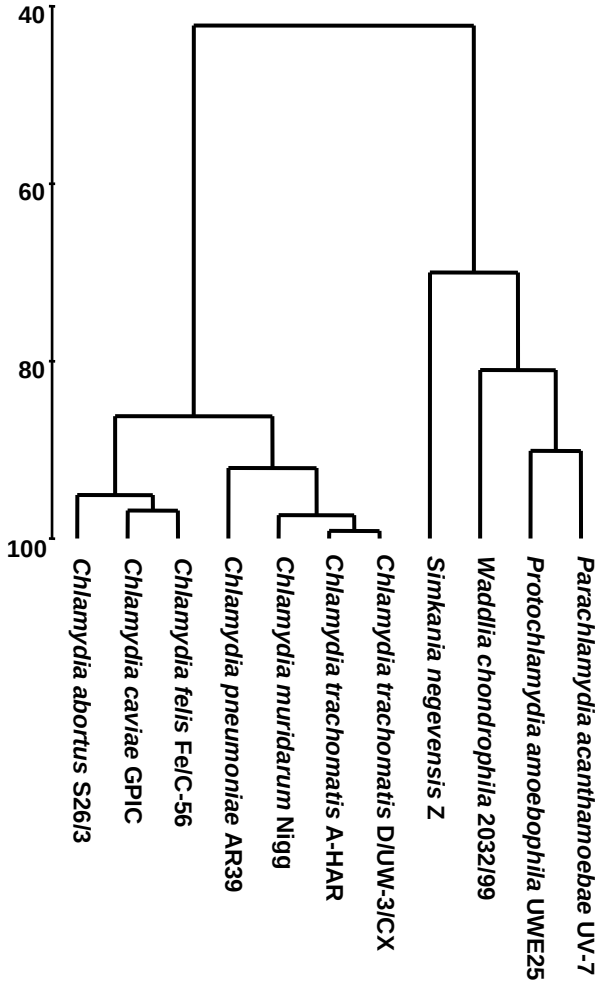

Figure S4: Cluster analysis of the presence/absence of predicted outer membrane proteins among the *Chlamydiae.* Clustering was performed based on a Bray-Curtis presence/absence similarity matrix by complete linkage with PRIMER 5.0 (Primer-e Ltd).

Table S1: Phylogenetic marker proteins used for reconstruction of the evolutionary history of the *Chlamydiae.*

Ribosomal proteins		Other phylogenetic marker proteins
Large subunit proteins	Small subunit proteins	
rl1	rs2	RpoB
rl2	rs3	RpoC
rl3	rs4	GyrB
rl4	rs5	RecA
rl5	rs8	EfTu
rl9	rs9	
rl10	rs10	
rl11	rs15	
rl13	rs16	
rl14	rs17	
rl16	rs18	
rl19	rs19	
rl20	rs20	
rl21		
rl22		
rl23		
rl24		
rl27		
rl31		