

The *Chlamydia psittaci* Genome: A Comparative Analysis of Intracellular Pathogens

Anja Voigt¹*, Gerhard Schöfl¹*, Hans Peter Saluz^{1,2*}

1 Leibniz-Institute for Natural Product Research and Infection Biology, Jena, Germany, **2** Friedrich Schiller University, Jena, Germany

Abstract

Background: *Chlamydiaceae* are a family of obligate intracellular pathogens causing a wide range of diseases in animals and humans, and facing unique evolutionary constraints not encountered by free-living prokaryotes. To investigate genomic aspects of infection, virulence and host preference we have sequenced *Chlamydia psittaci*, the pathogenic agent of ornithosis.

Results: A comparison of the genome of the avian *Chlamydia psittaci* isolate 6BC with the genomes of other chlamydial species, *C. trachomatis*, *C. muridarum*, *C. pneumoniae*, *C. abortus*, *C. felis* and *C. caviae*, revealed a high level of sequence conservation and synteny across taxa, with the major exception of the human pathogen *C. trachomatis*. Important differences manifest in the polymorphic membrane protein family specific for the *Chlamydiae* and in the highly variable chlamydial plasticity zone. We identified a number of *psittaci*-specific polymorphic membrane proteins of the G family that may be related to differences in host-range and/or virulence as compared to closely related *Chlamydiaceae*. We calculated non-synonymous to synonymous substitution rate ratios for pairs of orthologous genes to identify putative targets of adaptive evolution and predicted type III secreted effector proteins.

Conclusions: This study is the first detailed analysis of the *Chlamydia psittaci* genome sequence. It provides insights in the genome architecture of *C. psittaci* and proposes a number of novel candidate genes mostly of yet unknown function that may be important for pathogen-host interactions.

Citation: Voigt A, Schöfl G, Saluz HP (2012) The *Chlamydia psittaci* Genome: A Comparative Analysis of Intracellular Pathogens. PLoS ONE 7(4): e35097. doi:10.1371/journal.pone.0035097

Editor: David M. Ojcius, University of California Merced, United States of America

Received: December 1, 2011; **Accepted:** March 8, 2012; **Published:** April 10, 2012

Copyright: © 2012 Voigt et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Jena School of Microbial Communication (JSCM), the Bundesministerium für Bildung und Forschung (BMBF), and the Thüringer Aufbaubank (TAB). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: This work was supported by the Thüringer Aufbaubank. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: hanspeter.saluz@hki-jena.de

☞ These authors contributed equally to this work.

Introduction

Chlamydia psittaci is the pathogenic agent of ornithosis or psittacosis, a primarily avian respiratory disease with sizeable impact on poultry farming and bird breeding economic returns [1]. It belongs to the family *Chlamydiaceae*, a small group of extremely successful obligate intracellular pathogens that efficiently colonize mucosal surfaces and thrive within a wide variety of animal hosts, including humans [1].

Although birds are the primary targets of *C. psittaci* infections [2], transmissions from birds to humans have been reported, especially where humans come into close contact with infected birds on a regular basis, as in the case of veterinarians, poultry farmers, or bird breeders [3–5]. Moreover, *C. psittaci* has been isolated from a variety of other mammalian hosts, including cattle and other ruminants, horses, and pigs [1,6].

Other medically important members of the family *Chlamydiaceae* are the human-specific *C. trachomatis* and the wide-host-range *C. pneumoniae*. Worldwide, *C. trachomatis* is a leading cause of sexually transmitted bacterial diseases and ocular infections (trachoma), potentially leading to blindness [7]. *C. pneumoniae* is transmitted by

respiratory droplets and the causative agent of an atypical pneumonia and other acute respiratory illnesses [8].

Currently, within *Chlamydiaceae* a total of nine species are organized in the single genus *Chlamydia*: *C. trachomatis*, *C. muridarum*, *C. pneumoniae*, *C. pecorum*, *C. suis*, *C. abortus*, *C. felis*, *C. caviae*, and *C. psittaci*. All *Chlamydiaceae* need to infect eukaryote host cells for replication. Despite major differences in host range, tissue tropism, and disease pathology they all share a characteristic biphasic developmental cycle unique among prokaryotes [9,10]. The infectious chlamydial form, the elementary body (EB), enters the eukaryotic cell and becomes internalized in a vacuole in the cytoplasm of the host cell. In this so called inclusion, the EB differentiates into a non-infectious, metabolically active form, the reticulate body (RB). The RB multiplies by binary fission [1,10]. Depending on the strain, two to three days after infection the RBs transform back into EBs, which get released by lysis of the host cell or exocytosis [9].

As a consequence of this life style, the arguably most important driving force for the evolution of these pathogens is the interaction with their peculiar ecological niche, the inclusion within the cytoplasm of a eukaryotic host. Specific genomic manifestations of

Table 1. Summary of *Chlamydiaceae* genome features.

	<i>C. psittaci</i> 6BC	<i>C. abortus</i> S26/3	<i>C. felis</i> Fe/C-56	<i>C. caviae</i> GPIC	<i>C. pneumoniae</i> LPCoLN Bu	<i>C. trachomatis</i> L2/434/
Genome size [bp]	1,171,660	1,144,377	1,166,239	1,173,390	1,241,020	1,038,842
No. of CDSs	967	932	1005	998	1097	874
Coding density [%]	89.39	87.62	91.22	89.42	89.44	89.09
Average gene size [bp]	1083	1075	1059	1051	1012	1059
% G+C content	39.06	39.87	39.38	39.22	40.55	41.33
% G+C of CDSs	39.46	40.24	39.95	39.82	41.26	41.66
tRNA	38	38	38	38	38	37
rRNA operons	1	1	1	1	1	2

doi:10.1371/journal.pone.0035097.t001

the obligate intracellular lifestyle have long been recognized, e.g. a strong trend towards reduced genome sizes, a dramatical reduction of effective population size relative to environmental bacteria making purifying selection comparatively less efficient, extreme sequence divergence in proteins that mediate the interaction with the host environment such as outer surface proteins and secretion systems (reviewed in [11]). Hence, a comparison of genes involved either directly or indirectly in interactions with the host cell is most likely to shed light on the evolution of the intracellular life style of the *Chlamydiaceae*, and their adaptation to different eukaryote hosts.

Currently, relatively little is known regarding the chlamydial factors involved in virulence, host interaction, or host specificity. Genes for which functions in relation to niche adaptation have been implicated are mainly (i) the polymorphic membrane proteins (pmps), a large family of proteins probably unique to the phylum *Chlamydiae* [12,13], and considered to be important in adhesion of the EB to the host cell, molecular transport, and cell wall associated functions [14]; (ii) genes located in the chlamydial “plasticity zone”, a distinct region close to the terminus of replication, characterized by unusually high levels of inter- and intraspecific polymorphism [15]; and (iii) the so called type III effector proteins, a highly variable class of proteins potentially secreted into the host cell by the molecular machinery of the type III secretion (T3S) apparatus.

The T3S apparatus and the effectors translocated by it form a sophisticated mechanism of bacterial pathogenesis found in a number of Gram-negative bacteria. It is characterized by the direct translocation of effector proteins into the host cytoplasm to mediate colonization and parasitization of susceptible hosts [16,17]. T3S effector proteins display little sequence homology across species. Although the N-terminal regions of T3S effectors show unusual amino acid compositions (e.g. [18]) no unambiguous common motif among different T3S signal sequences has been established, making the computational prediction of putative T3S effectors a difficult challenge [18,19].

Whole-genome comparison between phylogenetically distant chlamydial species that parasitize a range of host species, vary in their host specificity and pathogenicity can provide a foundation from which to comprehend factors involved in chlamydial niche adaptation. For this study we sequenced the genome of the pathogenic avian type strain *Chlamydia psittaci* 6BC. Meanwhile two additional *C. psittaci* genome sequences have become available [20,21]. This study represents the first detailed analysis of the *C. psittaci* genome sequence, however. Our results show a typical chlamydial genome with a coding capacity of 967 CDSs. The

comparative analysis of the *C. psittaci* genome with those of other *Chlamydiaceae* confirms the exceptional roles of the family of polymorphic membrane proteins and the chlamydial plasticity zone as source of most interspecies variation. The prediction of putative type III secreted effectors and a genome-wide analyses of non-synonymous to synonymous substitution rate ratios yield a number of novel candidate genes likely involved in host-pathogen interactions and adaptive divergence between *C. psittaci* and their relatives.

Results and Discussion

Genome sequence of the avian isolate 6BC of *C. psittaci*

Chlamydia psittaci 6BC possesses a single circular chromosome of 1.172 Mb and a plasmid of 7553 bp. The bacterial chromosome is predicted to contain 967 coding sequences (CDSs) and the plasmid is predicted to harbour eight coding sequences. 26% of the CDSs are annotated as encoding hypothetical products. The general features of the *C. psittaci* genome in comparison to other sequenced chlamydial genomes are summarized in Table 1.

A phylogenetic tree was reconstructed for all species within *Chlamydiaceae* for which full-length genomic sequences have become available, including intraspecific variants (Figure 1). The inferred topology is consistent with previous phylogenies of the *Chlamydiaceae* (e.g. [22,23]), and shows the close relationship of *C. psittaci* with the three chlamydial species originally considered as the “mammalian” *Chlamydia psittaci* abortion, feline, and Guinea pig strains (i.e., *C. abortus*, *C. felis*, and *C. caviae*; here referred to as “*psittaci*-group”).

A comparison of the genomes of the *psittaci*-group shows that all four species are highly conserved in terms of size, in the number of coding sequences and in nucleotide composition (Table 1). The degree of conservation is illustrated by a comparison of the coding capacity of these genomes: 872 of the predicted CDSs of *C. psittaci* are shared among all four species of the *psittaci*-group (with a total CDS count ranging from 967 to 1005), 82 are divergent, and only 13 genes are unique to *C. psittaci* (Figure 2A). Notably, six of the 13 *C. psittaci* genes identified as having no significant homology to any other chlamydial species encode polymorphic membrane proteins (pmps) of the G family, most others encode hypothetical proteins with unknown functions (Table 2). Four of these unique genes (CPSIT_0306, CPSIT_0429, CPSIT_0605, and CPSIT_0846) are predicted type III secreted effector proteins (Table 3 and Table S1). Three of these genes are located in the hypervariable chlamydial plasticity zone (CPSIT_603, CPSIT_605, and CPSIT_607, Figure 3), a region at the terminus of replication

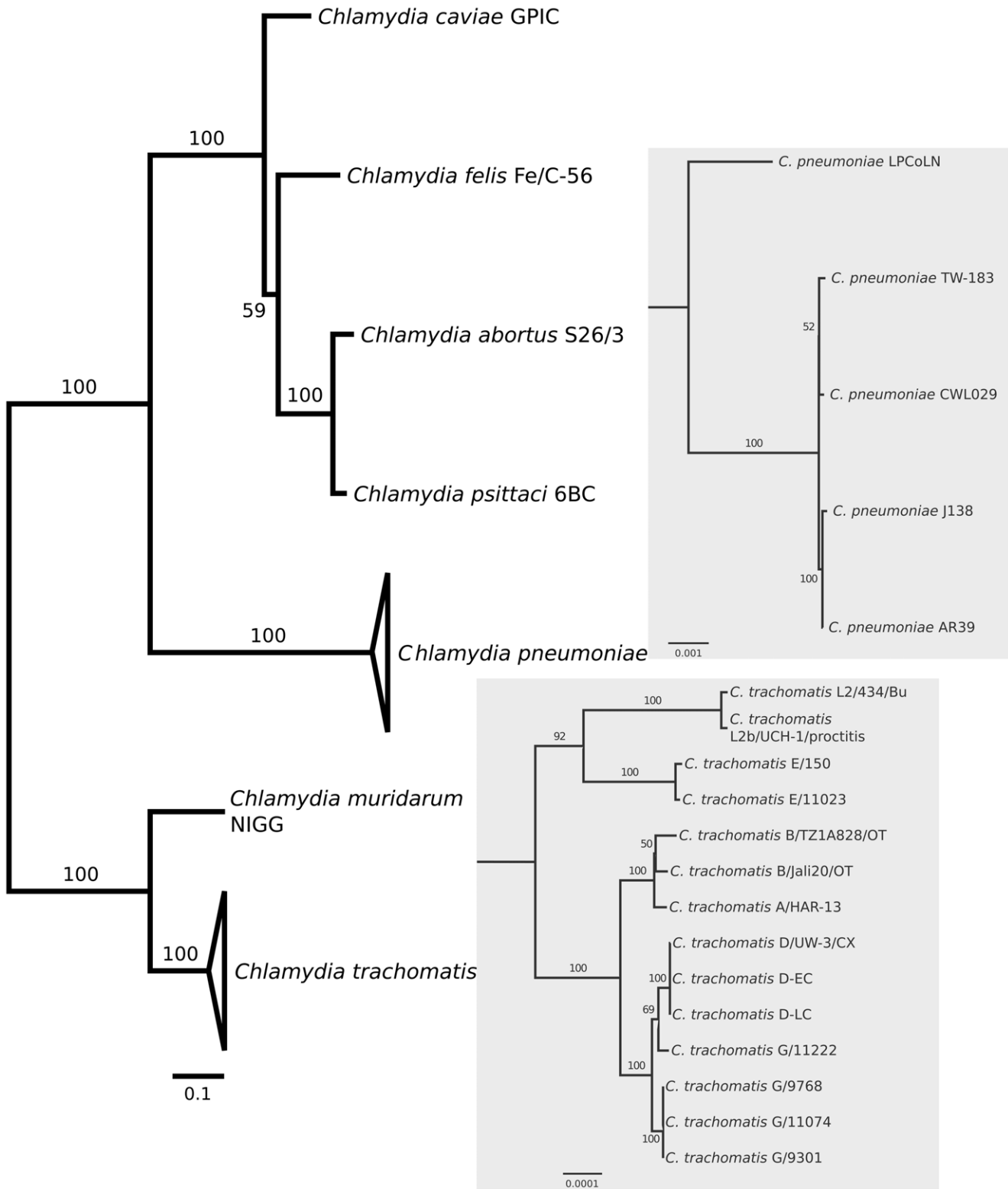


Figure 1. Phylogenomic relationships among sequenced chlamydial genomes. The maximum-likelihood tree is based on 100 randomly chosen conserved orthologous genes. Bootstrap values are displayed at the branches. The panels show the within-species phylogenetic relationships of the sequenced genomes of *Chlamydia pneumoniae* (upper panel) and *C. trachomatis* (lower panel). doi:10.1371/journal.pone.0035097.g001

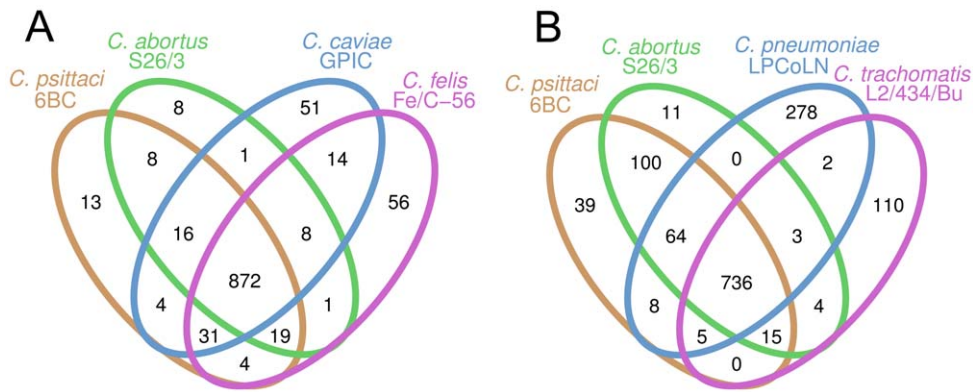


Figure 2. Venn diagrams showing the numbers of predicted CDSs that are unique or shared among two or more taxa. A. Number of shared genes among the closely related *C. psittaci*, *C. abortus*, *C. caviae*, and *C. felis*; **B.** number of shared genes among the wider range of *Chlamydiaceae*. Pseudogenes were scored as absent in this analysis. doi:10.1371/journal.pone.0035097.g002

that contains an array of chlamydial niche-specific genes putatively important in host-specific interactions [24].

When the more divergent species *C. pneumoniae* and *C. trachomatis* were included in the comparison, 736 of the predicted CDSs (ranging from 874 to 1097) were common to all *Chlamydiaceae* (Figure 2B). 64 genes were shared exclusively by *C. psittaci*, *C. abortus*, and *C. pneumoniae*, all members of the recently deprecated genus *Chlamydophila* [23,25] (Figure 2B).

Whole-genome comparisons with the other published *Chlamydiaceae* genomes show that the *C. psittaci* 6BC genome is essentially syntenic to sequences from *C. abortus*, *C. felis*, and *C. pneumoniae* (Figure 4).

A high degree of genomic rearrangement, however, becomes apparent with respect to *C. trachomatis* (Figure 4). Major deviations from synteny and/or sequence conservation are found in a hypervariable region near the replication terminus that has been termed “plasticity zone” (PZ) [15] and among the family of polymorphic membrane protein (pmp) genes [26] (Figure 4). These regions are suspected to harbour key genomic correlates to species-specific adaptation to different environmental niches, differential tissue tropism and differences in virulence and

pathogenicity [27]. Additionally, putative virulence factors, e.g. proteins mediating the chlamydial attachment to the host cell, or those related to the chlamydial inclusion membrane and development, may play a crucial role in niche adaptation, such as members of Inc/Tmh protein family (inclusion membrane proteins and transmembrane head proteins) and type III secreted effector proteins (T3SE) [15,28].

It is interesting to note, that a recent revision of the taxonomy of the *Chlamydiaceae* saw the genus *Chlamydia* retained as the sole genus in the family [25], thus deprecating the usage of the genus *Chlamydophila* introduced by Everett et al. in 1999 [29]. However, the phylogenetic analysis based on concatenated protein sequences (Figure 1), the numbers of shared orthologous genes (Figure 2B), and the level of synteny between genomes (Figure 4) all clearly do support the separation of *Chlamydiaceae* into two groups that correspond to the former genera, *Chlamydia* and *Chlamydophila*. Thus, irrespective of whether by any formal criteria *Chlamydophila* can be considered a genus, the label does remain useful as a moniker for an evolutionarily distinct branch within the *Chlamydiaceae*.

Plasticity zone and polymorphic membrane proteins

The size and organisation of the plasticity zone differs substantially among *Chlamydiaceae* (Figure 3). This high level of genetic diversity is thought to correspond to rapid evolution of a set of putative virulence factors which have accumulated in chlamydial PZs. Thus, a number of the genes contained in the chlamydial PZs have been linked to host-pathogen interactions/pathogenesis, e.g., a MAC/perforin domain gene [30], a cytotoxin gene similar to the EHEC adherence factor and clostridial large cytotoxins [24,31], tryptophan biosynthesis genes [31,32], or phospholipase D family enzymes [33].

In *C. psittaci* 6BC the PZ spans about 29 kb and encodes 16 genes. It has less gene content than the respective plasticity zones of *C. caviae* and *C. felis* (22 and 29 genes) but is larger than the PZ of *C. abortus* (11 genes). These differences arise because *C. psittaci* and *C. abortus* lack the complete tryptophan biosynthesis operon (*trpABFCDR*, *kynU*, *prsA*) present in the *C. felis* and *C. caviae* genomes. In contrast to *C. abortus*, *C. psittaci* shares a putatively functional 10074 bp EHEC-like adherence factor (CPSIT_0606), a 2466 bp MAC/perforin domain gene (CPSIT_0608), and a *guaAB-add* cluster serving purine nucleotide interconversion with *C. caviae* and *C. felis* (Figure 3). Three of the hypothetical proteins in

Table 2. Predicted CDSs unique to *C. psittaci* 6BC.

CDS	Product description
CPSIT_0309	polymorphic membrane protein, G family
CPSIT_0310	polymorphic membrane protein, G family
CPSIT_0311 ¹	polymorphic membrane protein, G family
CPSIT_0312 ¹	polymorphic membrane protein, G family
CPSIT_0316 ¹	polymorphic membrane protein, G family
CPSIT_0330	putative outer membrane protein
CPSIT_0429 ¹	hypothetical protein
CPSIT_0603	conserved hypothetical protein
CPSIT_0605	hypothetical protein
CPSIT_0607	hypothetical protein
CPSIT_0661	conserved domain protein
CPSIT_0668	polymorphic membrane protein, G family
CPSIT_0846 ¹	putative TMH-family membrane protein

¹predicted T3S effector protein.

doi:10.1371/journal.pone.0035097.t002

Table 3. Comparison of predicted type III secreted proteins of *C. psittaci* 6BC with orthologs in other *Chlamydiaceae*.

<i>C. psittaci</i> 6BC							
ORF	SVM value	Annotation	<i>C. abortus</i> S26/3	<i>C. felis</i> FeJ/C-56	<i>C. caviae</i> GPIC	<i>C. pneumoniae</i> LPCoLN	<i>C. trachomatis</i> L2/434/Bu
CPSIT0357	1.957	hypothetical protein	-	678*	-	325*	-
CPSIT0192	1.655	orthologous to <i>C. trachomatis</i> TARP	167*	837*	1089*	170*	0716*
CPSIT0429	1.309	hypothetical protein	-	-	-	-	-
CPSIT0580	1.224	putative inner membrane protein	522*	472*	-	536*	-
CPSIT0074	1.221	conserved hypothetical serine rich protein	063*	942*	-	062*	-
CPSIT0422	1.152	conserved hypothetical protein	-	618*	-	390	-
CPSIT0397	1.146	conserved hypothetical protein	357*	640*	0938*	367*	0528
CPSIT0463	1.109	putative inner membrane protein	412*	582	-	426	-
CPSIT0421	1.087	conserved hypothetical protein	376*	619*	-	389*	-
CPSIT0532	1.087	inclusion membrane protein B	477*	516*	0799*	491	0484*
CPSIT0997	1.080	putative inner membrane protein	910*	073*	0230*	941*	0828*
CPSIT0606	1.061	adherence factor	550	442	-	558*	-
CPSIT0245	1.054	carbohydrate isomerase	215*	788*	1041*	219*	0656
CPSIT0757	1.050	dihydropicolinate reductase	680*	303*	0474*	715*	0618
CPSIT0594	1.005	inclusion membrane protein A	536*	458*	-	550*	-
CPSIT0220	0.971	cyclodiphosphate synthase	191*	812*	01062	195*	0693
CPSIT0431	0.952	putative membrane protein	-	610	-	397*	-
CPSIT0846	0.896	putative TMH-family/lncA-family protein	766	-	-	-	-
CPSIT0844	0.864	putative TMH-family/lncA-family protein	764*	218	-	797*	-
CPSIT0689	0.858	conserved hypothetical protein	618	-	-	647	-
CPSIT0933	0.846	putative membrane protein	852	129*	-	-	0291*
CPSIT0314	0.829	polymorphic membrane protein, G family	283*	719*	-	-	-
CPSIT0749	0.769	conserved hypothetical protein	673*	-	-	707	-
CPSIT0296	0.746	hypothetical serine rich protein	264*	738*	-	270*	-
CPSIT0785	0.737	conserved hypothetical serine rich protein	706*	277*	0448*	739*	0238*
CPSIT0853	0.707	putative membrane protein	773*	214*	-	803*	-
CPSIT0602	0.706	conserved hypothetical protein	546	448	-	-	-
CPSIT0962	0.704	flagellar biosynthesis/type III secretory pathway	876*	106*	0265*	908*	0087
CPSIT0490	0.677	conserved hypothetical serine rich protein	437*	556*	0841*	451*	0338*
CPSIT1042	0.648	deoxyribonucleotide triphosphate pyrophosphatase	952*	031*	0180	982*	0869
CPSIT0828	0.635	DNA recombination protein	747*	234*	0407	779*	0197
CPSIT0760	0.573	hypothetical membrane protein	683*	300*	-	718	-
CPSIT0656	0.552	putative integral membrane protein	588*	388*	0664*	616	402*
CPSIT0974	0.549	trigger factor	887*	095*	0254*	919*	0076
CPSIT0313	0.545	polymorphic membrane protein, G family	282*	721*	0958	284*	-

Table 3. Cont.

<i>C. psittaci</i> 6BC		<i>C. abortus</i> S26/3	<i>C. felis</i> Fe/C-56	<i>C. caviae</i> GPIC	<i>C. pneumoniae</i> LPCoLN	<i>C. trachomatis</i> L2/434/Bu
ORF	SVM value	Annotation				
CPSIT1054	0.540	5-formyltetrahydrofolate cyclo-ligase	964*	019*	994*	0169
CPSIT0555	0.528	putative inner membrane protein	500*	494*	513	-
CPSIT0930	0.517	tRNA(Uracil-5-)-methyltransferase	849*	132*	883	0294*
CPSIT0461	0.513	hypothetical protein	411	-	425*	-
CPSIT0767	0.506	3-phosphoshikimate 1-carboxyvinyltransferase	690*	293*	723*	0466*

Orthologs predicted to be type three secreted are represented with *.

Only genes with an SVM decision value >0.5 are reported as candidate type III secreted proteins [18].

ORFs highlighted in gray have also been predicted as T3SEs by EffectiveT3 (<http://effectors.org>).

doi:10.1371/journal.pone.0035097.t003

the *C. psittaci* PZ (CPSIT_0603, CPSIT_0605, CPSIT_0607) are predicted to be type III secreted effector proteins.

Among the recognizable putative toxin genes of the PZ, a full-length version of a chlamydial MAC/perforin domain gene was present in the *C. psittaci*, *C. felis*, *C. pneumoniae* koala LPCoLN and *C. trachomatis* isolates. A MAC/perforin domain protein was absent from the *C. abortus* genome [27], and showed frame disruptions in *C. caviae* and the *C. pneumoniae* human isolates [34] (Figure 3). The *C. psittaci* MAC/perforin also shows a MIR (protein nanosyltransferase, IP3R and RyR) domain indicative of a possible ligand transferase function. It shares this feature only with the more distant orthologs in *C. pneumoniae* koala LPCoLN and *C. trachomatis*, but not *C. felis* [34].

Many eukaryotic MAC/perforin proteins function as membrane perforation proteins [35], and are known to play important roles in plant and animal immune response to bacterial infections [36]. The functional role of the MAC/perforin domain genes in *Chlamydiae* is unclear. It has been suggested that proteins with MAC/perforin structures might help in evading the host immune response through structural mimicry, e.g. if the assembly of a host MAC/perforin complex is prevented by pathogen-derived MAC/perforin molecules present on the pathogen's cell surface [37,38]. It is likely that the chlamydial MAC/perforin genes were obtained by horizontal gene transfer from an eukaryotic source [39,40].

A full-length lymphostatin/EHEC adherence factor is present only in *C. psittaci*, *C. felis*, and *C. caviae* (Figure 3) as well as in *C. muridarum* (three orthologs [15]) and *C. pecorum* [41]. This cytotoxin is generally absent from the human-specific pathogens *C. trachomatis* (except for gene fragments [42,43]) and *C. pneumoniae* [15,34].

The second major source of diversity among chlamydial genomes is the group of polymorphic membrane proteins. The pmps are a large protein family likely unique to the *Chlamydiae* [12,13]. Pmps are characterized by an unusually high level of mutational change within and across species, suggesting relatively fast evolutionary rates and high selective pressures potentially associated with adaptation to different hosts or immune responses [34,44,45]. They are present in varying numbers ranging from nine in *C. trachomatis* and *C. muridarum* [46] to 21 putative CDSs in *C. pneumoniae* [34,47] and *C. psittaci* (this study; Figure 5). The pmps group phylogenetically into six basic subfamilies (A, B/C, D, E/F, G/I, and H; Figure 6; [46]). Of these subfamilies, family G/I is the largest and the most rapidly evolving with numerous evolutionary recent independent events of gene duplication and loss in the various chlamydial lineages (Figure 6). The tendency to a proliferation of G/I family pmps is especially pronounced among the species belonging to the former genus *Chlamydothila* (i.e. the *psittaci*-group, *C. pneumoniae*, and *C. pecorum*). While there are only two G/I pmps present in *C. trachomatis* and *C. muridarum*, there are 14 pmp G/I family genes present in the *C. psittaci* genome (Figure 7).

In *C. psittaci* 6BC, one pmp gene is predicted to be truncated on the N-terminal side (CPSIT_0314). Similar to other *Chlamydiaceae*, a number of pmps harbour long poly-G tracts. Interestingly, while these poly-G stretches appeared to be in frame in the sequence generated by us, in a parallel sequencing effort on *C. psittaci* 6BC [21] the three pmp genes corresponding to CPSIT_0305, CPSIT_0312, and CPSIT_0666 are found to contain frameshifts in these long homopolymeric tracts. Whether this is a sequencing artefact or represents rapid change due to slippage mutations is unclear.

Despite their overall low amino acid and nucleotide similarities, all pmps share a unique domain structure. They contain a C-terminal autotransporter-like domain, a central pmp middle

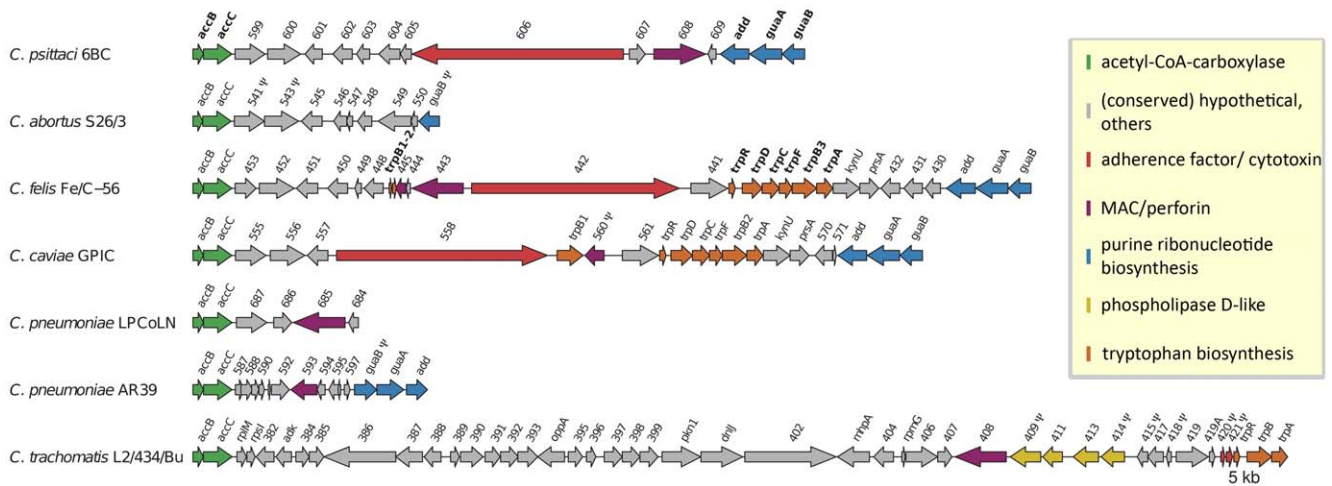


Figure 3. Comparison of the plasticity zone of *C. psittaci* 6BC, *C. abortus* S26/3, *C. felis* Fe/C-56, *C. caviae* GPIC, *C. pneumoniae* LPCoLN and AR39, and *C. trachomatis* L2/434/Bu. Genes are labelled with the published locus tag numbers. Colour-coded genes are discussed in the text. Pseudogenes are marked by Ψ. doi:10.1371/journal.pone.0035097.g003

domain and a varying number of the *Chlamydia*-specific short tetrapeptide motifs GGA(I, L, V) and FxxN on the N-terminal side [14,44,48] (compare Figure 5). In *C. psittaci* the numbers of conserved tetrapeptide motifs range from two to 18 for GGA(I, L, V), and from four up to 23 for FxxN. On average 9 FxxN and 4.8 GGA(I, L, V) motifs are found per *pmp* gene. These numbers are similar to other chlamydial species: *C. trachomatis* (13.6 and 6.5) and *C. pneumoniae* (11.3 and 5.0) [44]. Importantly, it has recently been shown that at least two copies of these repetitive tetrapeptide motifs are essential for chlamydial adhesion to the host cell [49].

Putative type III secreted effector proteins

Like a variety of other Gram-negative pathogens, *C. psittaci* uses a conserved type III secretion machinery as a basic mechanism of virulence determination, that allows transporting specific proteins known as type III secreted effectors (T3SE) into the cytoplasm of their host cells [17,50,51]. Many of the chlamydial T3S effectors are targeted to the inclusion membrane that encapsulates the pathogen inside their host cell [52]. Hence, these effectors are thought to play a crucial role for the modification of the chlamydial environment and for the survival of *Chlamydiae* in their inclusion vacuole [15].

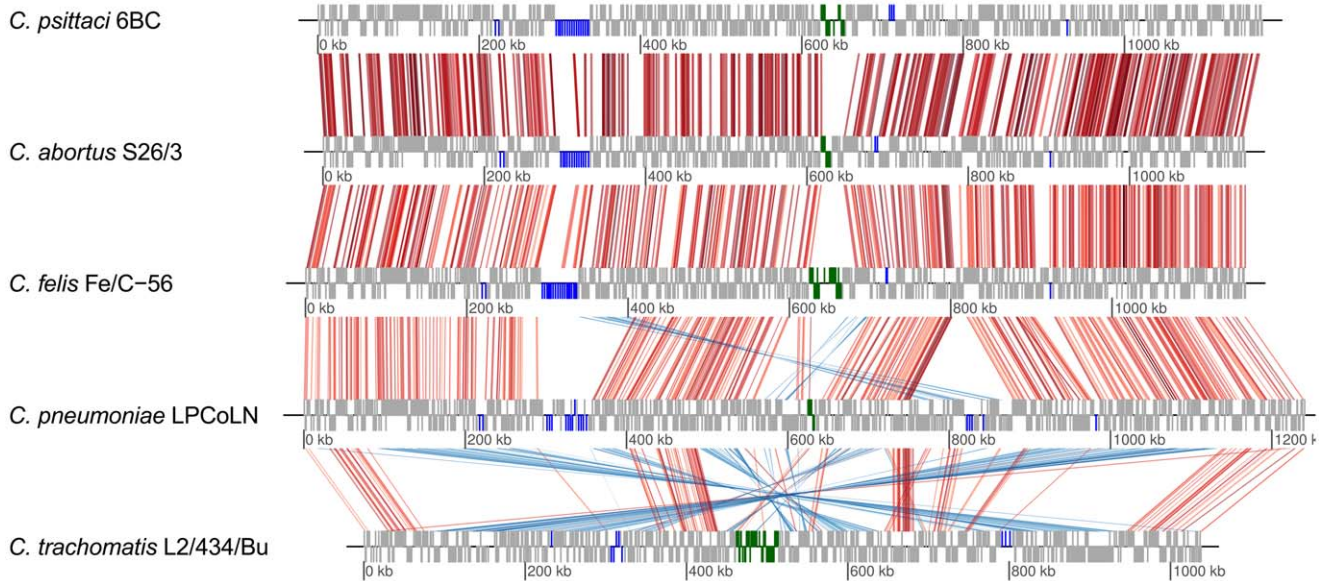


Figure 4. Global genome comparison between the *C. psittaci* 6BC, *C. abortus* S26/3, *C. felis* Fe/C-56, *C. pneumoniae* LPCoLN, and *C. trachomatis* L2/434/Bu genomes. The figure shows orthologous matches visualized using genoPlotR (compare Methods). The grey tick marks above and below the sequence lines represent the predicted CDSs on the plus strand and the minus strand of the genomes, respectively. Colour-marked are (blue) members of the polymorphic membrane protein family (*pmp*) and (green) the position of the plasticity zone (PZ). The red lines connecting genome lines represent direct orthologous matches. The blue lines represent reversed matches. Darker colours correspond to a higher bit scores. doi:10.1371/journal.pone.0035097.g004

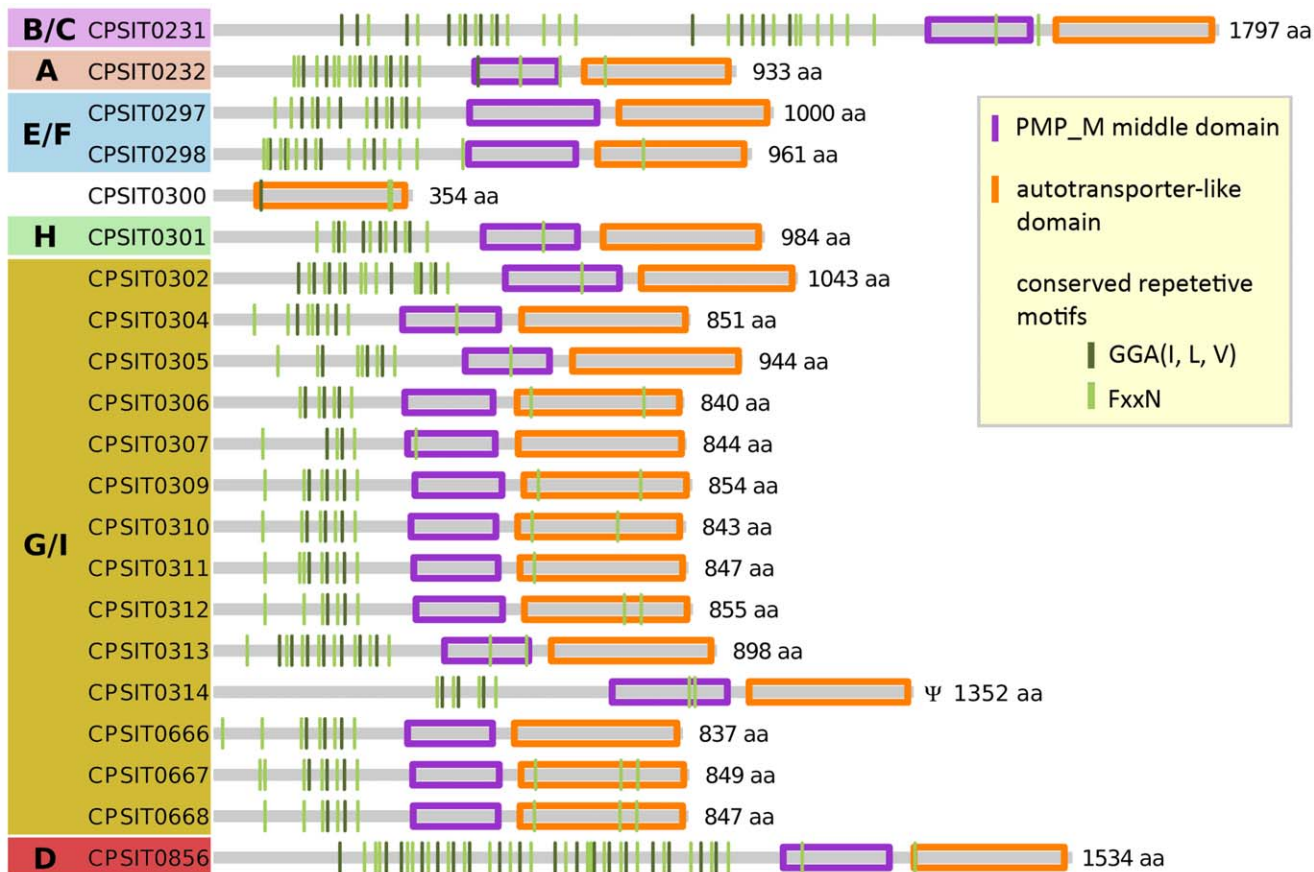


Figure 5. Structure of the pmp-family proteins of *C. psittaci* 6BC. Proteins are ordered by their position in the genome. The letter codes A to G/I indicate the pmp protein subfamilies as previously assigned by [44]. CPSIT_0314 has been reconstructed *in silico*. CPSIT_0300 is a gene remnant. doi:10.1371/journal.pone.0035097.g005

To predict potential T3S effector proteins in the genomes of *C. psittaci* 6BC, *C. abortus* S26/3, *C. felis* Fe/C-56, *C. caviae* GPIC, *C. pneumoniae* LPCoLN, and *C. trachomatis* L2/434/Bu we used a support vector machine (SVM) classifier developed by Wang et al. [18] that is based on T3S-specific features extracted from the N-terminal amino acid composition profile of proteins and the prediction software EffectiveT3 [19] (see Methods section).

Chlamydia psittaci CDSs predicted by these methodologies to encode T3S effector proteins and their orthologs in the other investigated species are presented in Table 3 (and Tables S1, S2, S3, S4, S5, S6, S7). Using a decision threshold of 0.5, 40 CDSs are classified as T3S effectors by the SVM algorithm (Table 3) and 68 CDSs are identified by EffectiveT3 (Table S7). 15 CDSs are identified by both approaches (Table 3).

As can be expected, many of the proteins classified as T3S effectors in *C. psittaci* are homologs to experimentally verified effector proteins from other species. CPSIT_0192, for instance, is orthologous to the important *C. trachomatis* T3S effector Tarp (translocated actin-recruiting protein) [53]. Tarp orthologs are present in all examined chlamydial species (Table 3). CPSIT_0192 has 100% query coverage and 91% sequence identity to *C. abortus* CAB167 and possesses three *Chlamydia*-specific domains of unknown function (DUF1547) and an actin-binding I/LWEQ domain. Query coverage and sequence identity to *C. trachomatis* Tarp CTL0716 are 62% and 37%, respectively. This highlights the high degree of variability in these genes among the *Chlamydiaceae*. Even within *C. trachomatis* variation in the Tarp

sequence has been reported [43]. Despite significant sequence differences to *C. trachomatis*, both, the *C. psittaci* and the *C. trachomatis* Tarp are expressed late in the developmental cycle and may have the same function [54,55].

An important family of T3S effectors tightly associated with the inclusion membrane are the Inc proteins. Members of this family show little general sequence similarity, but share a conspicuous bilobed hydrophobic domain of 60–80 amino acid residues [56]. An enrichment for coiled-coil regions typical for eukaryotic organisms has recently been described for putative Incs [12].

The *C. trachomatis* genome contains seven characterized Inc proteins (Inc A to G). The high sequence diversity in the Inc protein family makes the occurrence of most Inc proteins largely strain-specific. Thus, in *C. psittaci*, of three characterized Inc proteins A, B, and C only Inc B (CPSIT_0532) is conserved enough to be identified as an ortholog to *C. trachomatis* Inc B (CTL0484) by reciprocal BLAST.

In *C. psittaci* both Inc A and B were classified as T3S effectors by both prediction approaches (Table 3). For both proteins, type III secretion has also been experimentally confirmed in *C. psittaci* and *C. pneumoniae* [54,57]. Based on high immunological activity Inc A was the first Inc protein identified [58]. Inc B modulates host immune responses and might be involved in inclusion development and prevention of early lysosomal fusion [59]. The *C. psittaci* Inc C (CPSIT_0531) has not been classified as a T3SE by the SVM approach (but it is recognized as a T3SE by EffectiveT3; Table S7). This is likely explained by the complete lack of

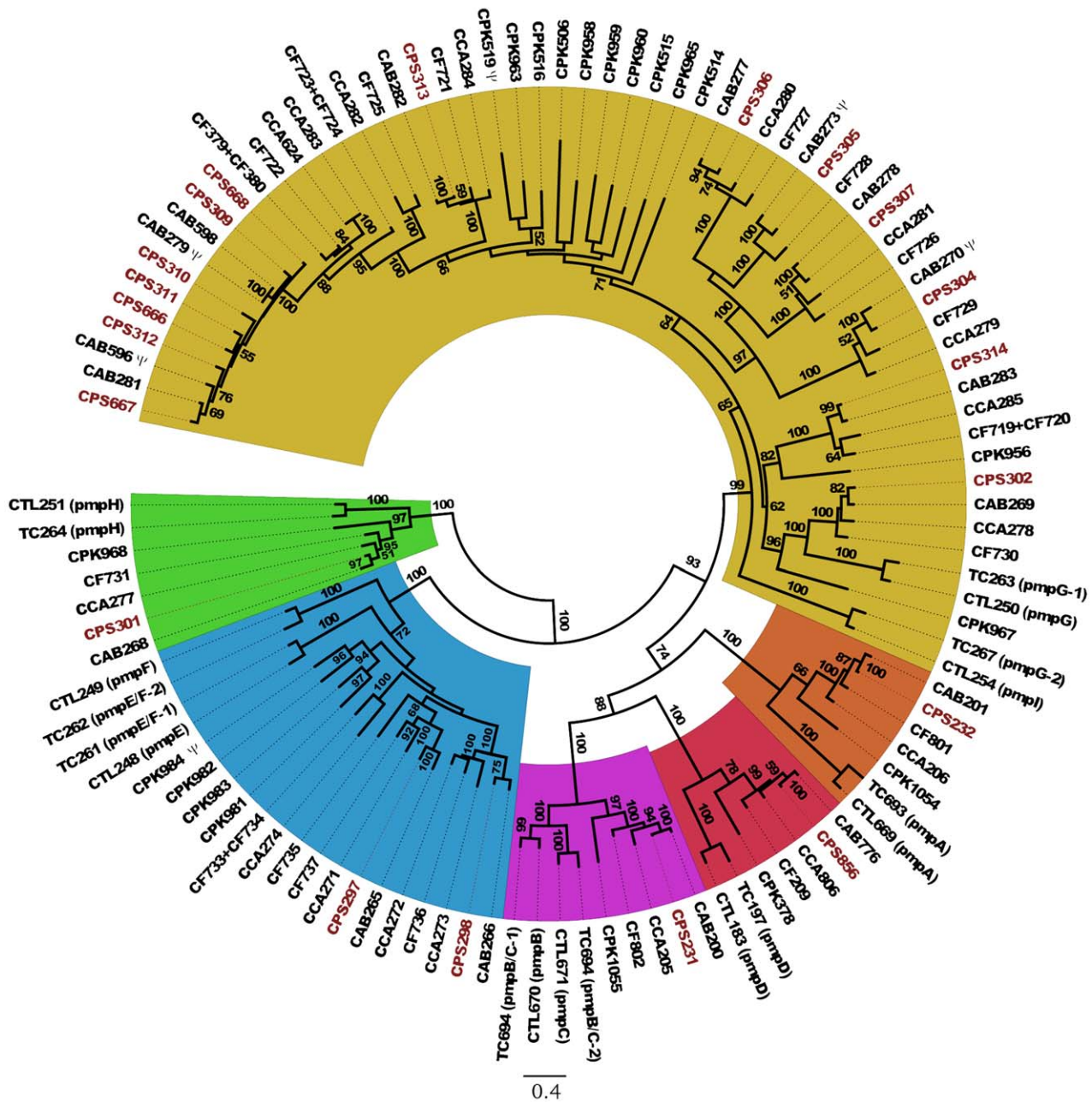


Figure 6. Phylogenetic relationship of chlamydial pmp-family proteins. The maximum-likelihood tree is based on alignments of the conserved PMP_M middle domain and autotransporter domain. Species included in the tree are *C. psittaci* 6BC (CPS), *C. abortus* 526/3 (CAB), *C. felis* Fe/C-56 (CF), *C.aviae* GPIC (CCA), *C. pneumoniae* LPCoLN (CPK), *C. muridarum* Nigg (TC) and *C. trachomatis* L2/434/Bu (CTL). Pmps cluster into 6 major subfamilies previously designated A (orange), B/C (purple), D (red), E/F (blue), G/I (yellow), and H (green) [44]. Bootstrap values are displayed at the branches. Pseudogenes are marked by Ψ.
doi:10.1371/journal.pone.0035097.g006

sequence homology in the N-terminal region with respect to the *C. trachomatis* Inc C (CTL0485; query coverage 53%, sequence identity 50%), which is an experimentally verified T3S effector.

Another cluster of genes putatively belonging to the larger Inc protein family and presumably playing similar roles, are the transmembrane head proteins (TMH) [27]. Transmembrane head proteins are characterized by a paired N-terminal transmembrane domain (IncA) followed by alpha-helical coiled-coil domains and show levels of sequence similarity significantly lower than the genome average [27].

In *C. psittaci* the *tmh* locus encodes 8 CDSs (CPSIT_0841, CPSIT_0842, CPSIT_0843, CPSIT_0844, CPSIT_0846, CPSIT_0848, CPSIT_0850, CPSIT_0851), all of which harbour an N-terminal IncA domain. The TMH proteins CPSIT_0844 and CPSIT_0846 were classified as possible T3S effectors by both prediction approaches and are orthologous to *C. abortus* CAB764 and CAB766, respectively (Table 3)

The comparison with *C. felis* and *C.aviae* suggest that the genes encoding the above proteins have arisen from a duplication event in the common ancestor of *C. psittaci* and *C. abortus*. The feline

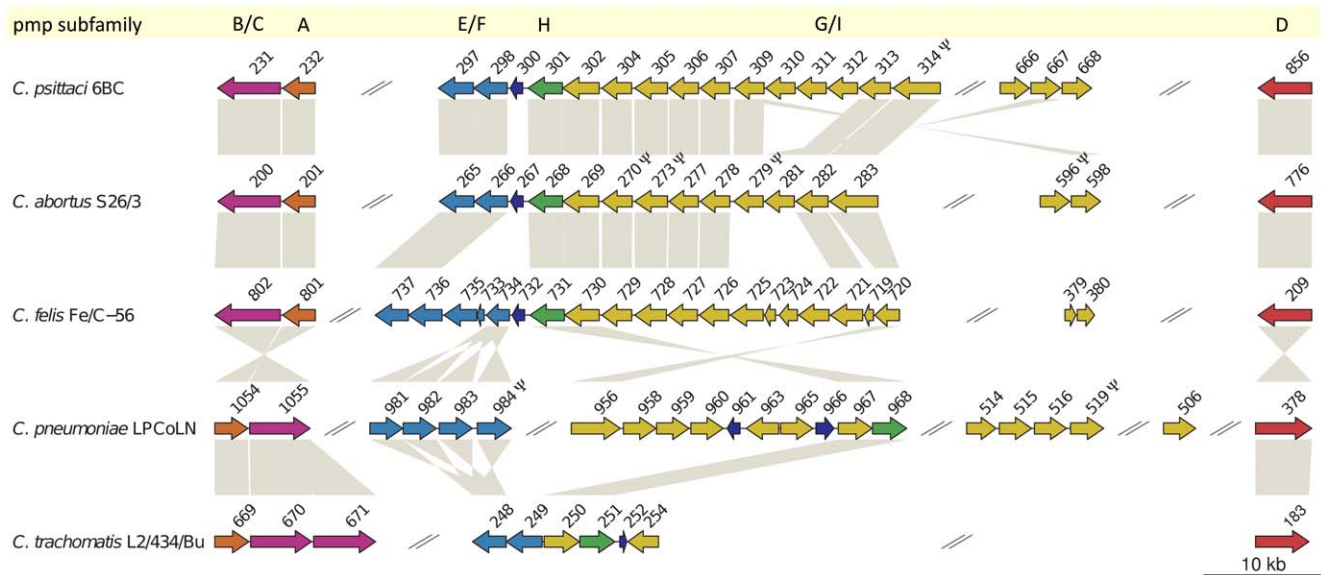


Figure 7. Organization of the pmp-family proteins compared between *C. psittaci* 6BC, *C. abortus* S26/3, *C. felis* Fe/C-56, *C. pneumoniae* LPCoLN, and *C. trachomatis* L2/434/Bu. Arrows indicate the gene orientation. The colour code denotes the membership to a pmp-subfamily. CDSs are designated by the numeric part of the published locus tags. Orthologous genes as inferred from the phylogenetic analysis shown in Figure 6, are connected by grey bars. Pseudogenes are marked by Ψ . doi:10.1371/journal.pone.0035097.g007

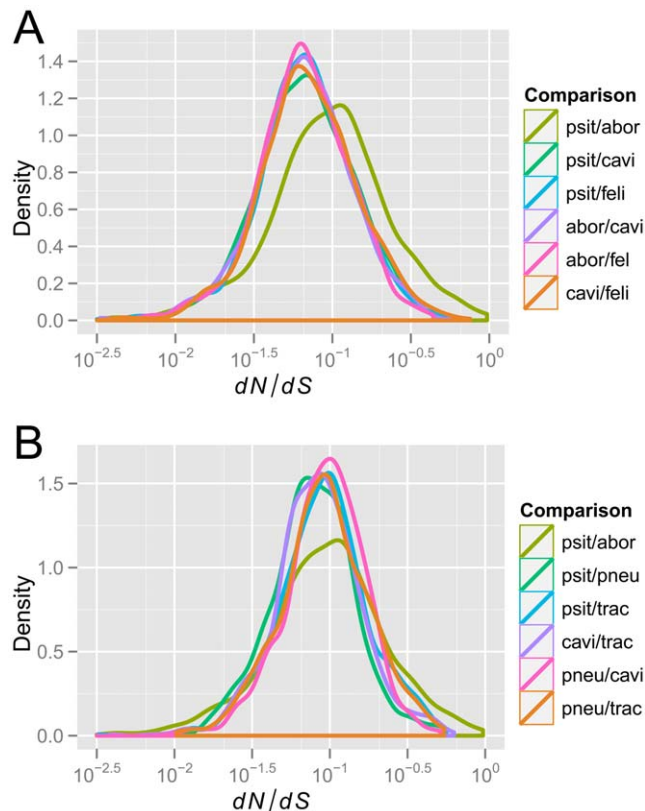


Figure 8. Distributions of d_N/d_S for orthologous genes from pairs of chlamydial genomes. **A.** Distributions for comparisons among the closely related species *C. psittaci*, *C. abortus*, *C. felis*, and *C. caviae*. **B.** Distributions for comparisons among the more distantly related species *C. psittaci* and *C. caviae* vs. *C. pneumoniae* and *C. trachomatis*, respectively. As a point of reference the comparison *C. psittaci* vs. *C. abortus* is included in both plots. Probability density curves were estimated by Gaussian-kernel smoothing. doi:10.1371/journal.pone.0035097.g008

ortholog CF0218 was shown to be distributed throughout the chlamydial inclusion bodies and confirmed to be immunogenic [60], but has not been classified as a T3S effector by our approach.

Besides a number of experimentally confirmed T3S effectors, some proteins with functional annotations that suggest a role in host-pathogen interactions and/or pathogenicity have been classified as T3S effectors. Among this group are a number of genes belonging to the pmp G family (CPSIT_0313, CPSIT_0314 [predicted by SVM], CPSIT_0311, CPSIT_0312, CPSIT_0316 [predicted by EffectiveT3]) and four of the 16 genes located in the plasticity zone. Thus, the adherence factor (CPSIT_0606) located in the PZ (Figure 3) is predicted as T3SE with a high SVM score and by EffectiveT3. Although the adherence factor has orthologs in *C. caviae* GPIC, *C. abortus* S26/3 (only a small gene remnant showing 2% query coverage, but 93% sequence identity) and *C. felis* Fe/C-56 (Table 3), the adherence factor is only predicted to be type III secreted for *C. psittaci* and *C. caviae* orthologs. The adherence factors of *C. felis* (CF0442) and *C. caviae* (CCA_00558) show in comparison to the psittacine adherence factor a query coverage of 90%, and a DNA sequence identity of 45% and 44%, respectively, suggesting a high evolutionary turn over.

Selection pressure on chlamydial genomes

The basic measure of selective pressure acting on protein coding sequences is the d_N/d_S -ratio. Generally, low values of d_N/d_S (i.e., values <1) are indicative of purifying selection acting on a given protein coding gene, while values >1 are usually interpreted as evidence for positive selection. Theoretically, the strength of purifying selection depends on the effective population size and the specific mutation and recombination rates of the compared lineages. Smaller effective population sizes and less recombination will lead to relatively larger d_N/d_S -values [61].

To characterize differences in selective pressure among chlamydial lineages on a genome-wide scale, we constructed the distributions of d_N/d_S for pairs of chlamydial species over their respective sets of orthologous genes. In agreement with previous findings [62,63], the shapes of these distributions were highly similar and best fitted by a log-normal distribution (Figure 8).

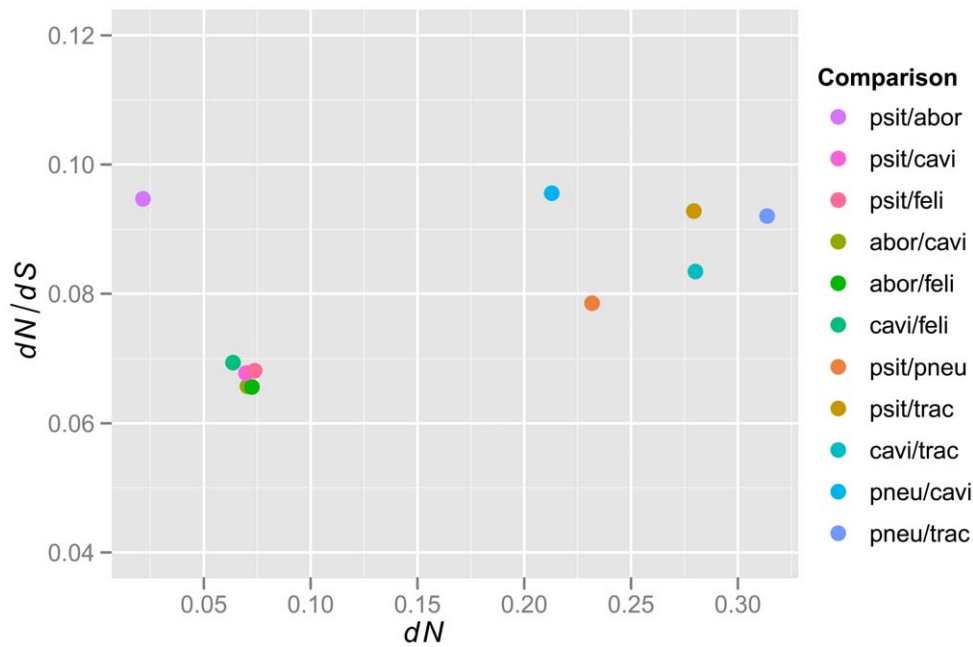


Figure 9. Dependence of median d_N/d_S on the genetic distance between genomes estimated as the median non-synonymous substitution rate, d_N .

doi:10.1371/journal.pone.0035097.g009

Median d_N/d_S -values fell in the range between 0.06 and 0.1 (Figure 9) and are indicative of the strong evolutionary constraints typical for the compact genomes of prokaryotes [63].

It has been observed that d_N/d_S correlates negatively with evolutionary distance, i.e. the smaller the distance between genomes the larger the estimates of d_N/d_S , thus leading to an overestimation of positive selection [64]. Such a pattern is not apparent among the chlamydial lineages compared here (Figure 9). In fact, with the notable exception of *C. psittaci* vs. *C. abortus*, pairwise comparisons between the closely related lineages within the “*C. psittaci*-group” show d_N/d_S -distributions shifted towards lower values, indicative of higher levels genomic conservation, than comparisons across larger genomic distances (compare Figures 8A vs. 8B and Figure 9). In addition to exhibiting less genomic constraint than other comparisons within the same group of lineages, the *C. psittaci*/*C. abortus* comparison also shows the highest overall variance in d_N/d_S -ratios (interquartile range = 0.106; interquartile ranges for all other comparisons range from 0.057 to 0.080).

The median d_N/d_S -values for pairwise comparisons among chlamydial lineages range between 0.066 and 0.096, and fall thus in the upper third of the range for global median d_N/d_S -values typically reported for prokaryotes (0.01–0.1 [63]). This is in line with a trend that the weakest purifying selection pressures are seen in obligate parasites, and is probably explained by their relatively small effective population sizes, frequent bottlenecks and low recombination rates [65].

Although no clear phenotypic correlates are apparent, the variability in purifying selection pressure affecting the evolution of different chlamydial lineages may thus also be a reflection of differences in their effective population sizes and/or the frequency of bottlenecks associated with differences in their life styles (e.g. host preferences or differences in pathogenicity may influence the numbers of infected carriers and thus the effective population sizes of the pathogens).

Using an (arbitrary) cut-off value of a gene-wide d_N/d_S -ratio greater than 0.75 for at least one of the nucleotide substitution models, the *C. psittaci*/*C. abortus* comparison is the only one to give a list of potential candidate genes under positive selection (Table 4).

Metabolic pathways

The *C. psittaci* 6BC genome encodes for all central metabolic pathways such as the glycolytic pathway and the tricarboxylic acid (TCA) cycle. The TCA cycle of obligate intracellular pathogens varies from complete, in e.g. *Coxiella burnetii* [66] and *Rickettsia prowazekii* [67] to absent, in e.g. *Mycoplasma* [68]. Like all other chlamydial species [69], *C. psittaci* 6BC lacks a number of core components of the tricarboxylic acid cycle, namely citrate synthase, aconitase, and isocitrate dehydrogenase. How *Chlamydiaceae* compensate for these deficiencies is not clear.

Chlamydiaceae also vary in the completeness of the biotin pathway. Like *C. abortus* S26/3, *C. felis* Fe/C56, and *C. pneumoniae* LPCoLN, the *C. psittaci* 6BC genome contains all genes necessary for the production of biotin from pimeloyl-CoA (Figure S1). In contrast, *C. trachomatis* L2/434/Bu has lost several genes from this pathway such as adenosylmethionine-8-amino-7-oxononanoate aminotransferase *bioA*, dethiobiotin synthetase *bioD*, and biotin synthase *bioB* (Figure S1). Also *C. muridarum* and *C. caviae* exhibit an incomplete biotin gene cluster [27].

Biotin is an essential cofactor involved in many pathways [70]. The phylogenetic distribution of the deficiencies in the biotin biosynthesis pathway within *Chlamydiaceae* suggests at least two independent events of gene loss (one in the common ancestor of *C. trachomatis* and *C. muridarum*, and one in *C. caviae*). This correlates with differences in host specificity. While for *C. trachomatis*, *C. muridarum*, and *C. caviae* only one (or, in the case of *C. muridarum* two closely related) host species has been reported, the host range is markedly broader for the other species [71]. Intracellular organisms generally are prone to loss of function of metabolic genes due to a relaxation of selective constraints in their metabolite-rich environment [72]. This trend may, however, be

Table 4. Genes potentially under positive selection between *C. psittaci* and *C. abortus*.

CCCL CDS1	CDS2	mean d_N/d_S ratio	product description
CPSIT_0350	CAB314	0.974233	putative serine-rich exported protein
CPSIT_0844	CAB764	0.957362	putative TMH-family/IncA-family protein
CPSIT_0161	CAB139	0.911909	putative lipoprotein
CPSIT_0604	CAB549	0.904697	conserved hypothetical protein (plasticity zone)
CPSIT_0469	CAB416	0.821031	putative exported protein
CPSIT_0036	CAB032	0.741458	conserved hypothetical protein
CPSIT_0034	CAB030	0.697462	conserved hypothetical protein
CPSIT_0336	CAB302	0.680659	conserved hypothetical protein
CPSIT_0390	CAB351	0.671977	putative inner membrane protein
CPSIT_0876	CAB793	0.669905	hypothetical protein
CPSIT_0841	CAB760	0.598583	putative TMH-family/IncA-family protein
CPSIT_0685	CAB614	0.595976	co-chaperonin GroES
CPSIT_0545	CAB491	0.582204	hypothetical protein
CPSIT_0207	CAB180	0.537288	small cystein-rich outer membrane protein
CPSIT_0513	CAB460	0.484486	putative lipoprotein

doi:10.1371/journal.pone.0035097.t004

exacerbated if a restricted host range leads to a reduced effective population size, and thus to a less efficient selection against deleterious mutations.

Differences between the chlamydial species also exist in the purine and pyrimidine pathways (Figures S2 and S3). The genome of *C. psittaci* 6BC contains a gene cluster consisting of IMP dehydrogenase *guaB*, GMP synthase *guaA*, and adenosine deaminase *add* relevant for purine interconversion (Figure S2). These genes are also present in *C. felis*, *C. caviae*, *C. pneumoniae* AR39 (with a *guaB* pseudogene, however), and *C. muridarum* [24], but lack from *C. abortus* (only a *guaB* pseudogene), *C. pneumoniae* LPCoLN, and *C. trachomatis* L2/434/Bu [27,34]. With respect to pyrimidine interconversion, the genomes of *C. psittaci* 6BC and all other *Chlamydiaceae* encode the conversion of UMP to CTP (uridylylate kinase *pyrH*, nucleoside diphosphate kinase *ndk*, and CTP synthase *pyrG*, Figure S3). With the exception of *C. trachomatis* L2/434/Bu all *Chlamydiaceae* examined here also encode orotate phosphoribosyltransferase *pyrE* (Figure S3). Only *C. pneumoniae* encodes uridine kinase *udk*, responsible for the conversion of uridine or cytidine into uridine monophosphate or cytidine monophosphate (UMP/CMP). Also these patterns suggest multiple independent events of loss of function, possibly due to a reduction of selective constraints on metabolic genes, but lack any clear correlation to known differences in host adaptation.

Several studies maintain that *Chlamydiaceae* do not import host dNTPs for DNA synthesis, but convert NTPs to dNTPs [73–75]. Both *C. psittaci* 6BC and *C. trachomatis* L2 can obtain all NTPs from the host cell [75–77]. In *C. trachomatis* nucleoside phosphate transporters Npt1 and Npt2 are present [74,78]. Npt1 mediates the exchange of host ATP and bacterial ADP, and Npt2 transports NTPs into the bacterium. An ATP/ADP translocase that enables the RBs to supply themselves with ATP from the host cell, has also been reported for *C. psittaci* 6BC [79]. Our genomic data support this finding. The *C. psittaci* 6BC genome encodes for an ATP/ADP translocase (CPSIT_0474) with 79% sequence identity and 93% query coverage compared to *C. trachomatis* Npt1.

Chlamydial species differ, however, in their requirements with respect to the availability of external sources of NTPs or precursors. While all chlamydial species investigated here are able

to synthesize CTP from UTP, only *C. psittaci* and *C. felis* can potentially also interconvert ATP and GTP, because only these two species encode a complete *guaAB-add* cluster. In other words, all *Chlamydiaceae* deficient in the *guaAB-add* cluster have to import ATP, GTP, and UTP or precursors from the host cell [78,80]. *C. psittaci* and *C. felis* crucially only depend on an external source of UTP and either ATP or GTP or the respective precursors.

Uniquely among *Chlamydiaceae*, *C. pneumoniae* possesses a uridine kinase *udk* (EC 2.7.1.48), converting uridine or cytidine to UMP or CMP [34]. This potentially makes *C. pneumoniae* independent from an external source of UTP if it can take up its precursors, i.e. uridine or cytidine.

Interestingly, it has recently been shown, that a cytosolic 5'-nucleotidase can have phosphotransferase activity in addition to hydrolase activity [81,82]. For the chlamydial isolates examined in this study, a 5'-nucleotidase (EC 3.1.3.5) has been predicted. If the phosphotransferase activity extends to the chlamydial 5'-nucleotidases, it potentially allows a conversion of (deoxy)guanosine, xanthosine, inosine, (deoxy)adenosine, uridine, and cytidine into their respective monophosphates (Figure S2 and S3). If these precursor molecules can be obtained from the host cell the *guaAB-add* cluster is rendered redundant, decreasing the selection pressure to maintain functional copies of these genes. However, whether *Chlamydiaceae* have the ability to obtain NMP precursors from external sources at all is contentious. While Tribby and Moulder [83] asserted that *C. psittaci* Cal10 incorporates adenine, guanine, their (deoxy)ribonucleosides, later studies [75,76,84] found that only precursors which the host cell has converted to nucleotides can successfully be incorporated.

Conclusions

With the sequencing of the genome of *Chlamydia psittaci* the complete genomic sequences for all species but one (*C. suis*) of the *Chlamydiaceae* has become available. The comparative study of these genomes provides important insights into evolutionary history of this group of closely related intracellular pathogens and allows the identification of genomic differences that may account for the observed variation in virulence, pathogenicity, and host specificity among the species. In this study we made use of the

C. psittaci genome to focus on the most prominent genomic regions outside of the well-conserved chlamydial core genome: the polymorphic membrane proteins, the chlamydial plasticity zone, and the type III secreted effector proteins. We have shown that the genetic differences of *C. psittaci* with respect to other *Chlamydiaceae* includes an array of unique pmp genes of the G/I subfamily, the lack of a tryptophan operon in the plasticity zone (similar to its sister taxon *C. abortus*), the presence of an uninterrupted adherence factor and a MAC/perforin in the plasticity zone, and a number of candidate type III secreted effectors some of which are not present in all *Chlamydiaceae* or have not been classified as T3SEs in all species. In addition, a number of genes with functional annotations indicative of a role in host-pathogen interactions show some indication of recent positive selection after the split of the *C. psittaci* and *C. abortus* lineages. Further investigation of these genes may provide insights in what enables some species to exploit a wide range of hosts while others seem restricted few closely related host species, or whether these genes indeed may account for differences in virulence and pathogenicity.

Methods

Chlamydial genomes

The avian *Chlamydia psittaci* isolate 6BC (GenBank accession number CP002549) was sequenced *de novo* by a combination of Roche 454 pyrosequencing, Illumina and Sanger sequencing to, on average, 487-fold sequence coverage, assembled and annotated as described in [85].

So far complete genomic sequences of seven other chlamydial species have been published: *C. trachomatis* [46], *C. muridarum* [15], *C. pecorum* [41], *C. pneumoniae* [15,47,86], *C. caviae* [24], *C. abortus* [27], and *C. felis* [87]. For a phylogenetic analysis based on complete genomes, sequences and annotations for the following publicly available chlamydial species were obtained from NCBI: *Chlamydia muridarum* Nigg (GenBank: AE002160), *Chlamydia trachomatis* (14 strains: AM884176, CP000051, FM872308, FM872307, CP002052, CP002054, AE001273, CP001886, CP001890, CP001930, CP001887, CP001889, CP001888, AM884177), *Chlamydia abortus* S26/3 (CR848038), *Chlamydia caviae* GPIC (AE015925), *Chlamydia felis* Fe/C-56 (AP006861), and *Chlamydia pneumoniae* (5 strains: AE002161, AE001363, BA000008, AE009440, CP001713).

Comparative analyses were mostly restricted to the following subset of the above genomes: *C. trachomatis* L2/434/Bu (AM884176), *C. pneumoniae* LPCoLN (CP001713), *C. felis* Fe/C-56, *C. abortus* S26/3, *C. caviae* GPIC, and *C. psittaci* 6BC.

Comparative analyses of genome content

For the identification of species- and genus-specific orthologous genes, an all-vs.-all comparison of the translated coding sequences (CDSs) of 6 chlamydial genomes (see above) was performed using BLAT v34 [88]. The BLAT-identified bidirectional hits were filtered, keeping only those with an expect score less than 10^{-3} and a cumulative match size of at least one-third of the query sequence length. Where query sequences yielded multiple matches, the match with the highest bit score was retained. Best reciprocal hits by these criteria were considered orthologs for the purpose of this study. A custom R script was written to construct multi-genome match tables and to generate four-way Venn diagrams. Discrepancies from mismatches between putative orthologs in the multi-genome comparison arose in 13 cases and were resolved manually by checking local synteny. To visualize the conservation of genomic context among the *Chlamydiaceae*, a whole-

genome synteny plot based on best reciprocal BLAT hits was constructed using the genoPlotR package [89].

Metabolic pathway reconstruction for *C. psittaci* was performed with ASGARD v1.5.3 [90], using the KEGG database [91] as a source for pathway definitions.

For constructing the global phylogeny of the *Chlamydiaceae*, we retrieved a set of 478 orthologous genes conserved across all 24 available chlamydial genomes by all-vs.-all BLAT-comparisons of the CDSs as implemented in the orthology mapping software mercator (<http://www.biostat.wisc.edu/~cdewey/mercator/>). From this set of orthologs, we aligned a random sample of 100 genes with MAFFT v6.717b [92] using the L-INS-i option. The concatenated alignment, spanning 121,258 positions with a total of 58,523 informative sites was employed to reconstruct an unrooted phylogeny by maximum likelihood inference, using PHYML v3.0 [93] under a general-time reversible (GTR) model with six rate categories. To avoid long-branch attraction, intra- and interspecies phylogenies were estimated separately. Base frequencies, transition/transversion ratios, and the gamma distribution parameter (α) were estimated from the data. Topological robustness was assessed by 100 non-parametric bootstrap replicates.

Comparative analysis of the polymorphic membrane protein family

Comparative genomics and phylogenetic estimation were used to characterize evolutionary changes affecting the chlamydial polymorphic membrane protein (pmp) family. Predicted pmp sequences were extracted from *C. psittaci* 6BC, *C. abortus* S26/3, *C. caviae* GPIC, *C. felis* Fe/C-56, *C. pneumoniae* LPCoLN, *C. trachomatis* L2/434/Bu, and *C. muridarum* Nigg by searching all translated putative genes for the pmp-specific C-terminal autotransporter β -barrel domain and the conserved PMP_M middle domain [27] motifs using the Pfam HMM database. Interrupted pmp genes (in *C. felis*) and annotated pseudogenes (*C. abortus* and *C. pneumoniae*) were reconstructed *in silico* for phylogenetic and comparative analyses.

Due to the inter- and intraspecific divergence of pmp-family proteins and following [27], the phylogenetic analysis of pmp genes was based on alignments of the conserved PMP_M middle domain and the C-terminal autotransporter domain alone. Multiple protein alignments were constructed with MAFFT v6.717b [92] using the L-INS-i option and the BLOSUM80 substitution matrix. A maximum likelihood tree was reconstructed using PHYML [93] under the WAG [94] model of protein evolution. Amino-acid frequencies and the gamma distribution parameter α were estimated from the data.

Test for positive selection and type III secreted proteins

To characterize the nature and strength of selective pressures affecting protein sequences, pairs of orthologous genes between the closely related genomes of *C. psittaci*, *C. abortus*, *C. felis*, and *C. caviae* as well as between the more distantly related genomes of *C. psittaci*, *C. caviae* and *C. pneumoniae*, *C. trachomatis* were identified as bidirectional best hits in an all-against-all BLAT search as described above. Amino acid sequences were aligned using the Needleman-Wunsch global alignment algorithm and the BLOSUM62 substitution matrix as implemented in R, and subsequently translated back to the corresponding DNA sequences. Ratios of the rates of non-synonymous to synonymous nucleotide substitutions per site (d_N/d_S), averaged over the entire alignment, were estimated using KaKs_Calculator 2.0 [95]. We calculated d_N/d_S -ratios under four of the candidate models of codon substitutions implemented in the software (γ -NG, γ -LWL, γ -MLWL, and γ -YN), and used d_N/d_S -values averaged over all models as an estimate of the selective

pressure that affect the compared genomes after their divergence from their most recent common ancestor.

In silico prediction of type III secreted (T3S) effector proteins was performed using BPBAac [18]. Briefly, BPBAac uses a Bi-profile Bayes (BPB) approach to feature extraction from training datasets [96] to extract T3S effector features from the position-specific N-terminal amino acid composition (Aac) profile of sets of validated T3S proteins and non-T3S proteins. Bi-Profile Bayes allows representing both the positive and the negative information contained in each peptide sequence in a single posterior probability vector. The posterior probability vectors derived from the training data sets are then used to train a machine learning algorithm, called support vector machine (SVM). Fundamentally, the SVM is a binary classifier that, given two datasets, learns to distinguish between them and to predict the classification of previously unseen samples. The robustness of the classification is expressed by SVM decision values (scores), which indicate the distance in feature space of data points to the nearest point on the decision boundary. Following the practice of [18], we shift the decision threshold from 0 to 0.5 to minimize the number of false positives reported as candidate type III secreted proteins. Additionally, T3S effector proteins were predicted using EffectiveT3 (<http://effectors.org>) [19]. This approach relies on a taxonomically universal and conserved type III secretion signal sequence in the N-terminus [19]. For prediction we used the standard EffectiveT3 classification module and a cut-off score of 0.9999.

Supporting Information

Figure S1 Biotin biosynthesis pathways in *C. psittaci* 6BC. The genomes of *C. psittaci* 6BC, *C. abortus* S26/3, *C. felis* Fe/C-56, and *C. pneumoniae* LPCoLN encode for all enzymes needed to convert pimeloyl-CoA to biotin. These include 8-amino-7-oxononanoate synthase *bioF*, adenosylmethionine-8-amino-7-oxononanoate aminotransferase *bioA*, dethiobiotin synthetase *bioD*, and biotin synthase *bioB*. The genome of *C. trachomatis* L2/434/Bu encodes only the first step.
(TIF)

Figure S2 Purine biosynthesis pathway of *C. psittaci* 6BC. The genomes of *C. psittaci* 6BC and *C. felis* Fe/C-56 encode a *guaB/A*-add cluster (dehydrogenase *guaB*, GMP synthase *guaA*, adenosine deaminase *add*) for the conversion of AMP, IMP, and GMP, while *C. abortus* S26/3, *C. pneumoniae* LPCoLN, and *C. trachomatis* L2/434/Bu lack this gene cluster. The scheme has been modified after the KEGG PATHWAY database (www.genome.jp/kegg/pathway.html). Dashed arrows indicate predicted reaction directions supported by enzyme profiles available from the KEGG ENZYME database (<http://www.genome.jp/kegg/kegg3.html>).
(TIF)

References

- Longbottom D, Coulter IJ (2003) Animal chlamydioses and zoonotic implications. *Journal of Comparative Pathology* 128: 217–244.
- Harkinezhad T, Geens T, Vanrompay D (2009) *Chlamydia psittaci* infections in birds: a review with emphasis on zoonotic consequences. *Veterinary Microbiology* 135: 68–77.
- Moroney JF, Guevara R, Iverson C, Chen FM, Skelton SK, et al. (1998) Detection of chlamydiosis in a shipment of pet birds, leading to recognition of an outbreak of clinically mild psittacosis in humans. *Clinical Infectious Diseases* 26: 1425–1429.
- Van Droogenbroeck C, Beekman DSA, Verminnen K, Marien M, Nauwynck H, et al. (2009) Simultaneous zoonotic transmission of *Chlamydia psittaci* genotypes D, F and E/B to a veterinary scientist. *Veterinary Microbiology* 135: 78–81.
- Petrovay F, Balla E (2008) Two fatal cases of psittacosis caused by *Chlamydia psittaci*. *Journal of Medical Microbiology* 57: 1296–1298.
- Henning K, Sachse K, Stüng R (2000) Nachweis von Chlamydien bei einem Stutenabort. *Deutsche tierärztliche Wochenschrift* 107: 49–52.
- Peeling RW, Brunham RC (1996) *Chlamydiae* as pathogens: new species and new issues. *Emerging Infectious Diseases* 2: 307–319.
- Saikku P (1992) The epidemiology and significance of *Chlamydia pneumoniae*. *The Journal of Infection* 25 Suppl 1: 27–34.
- Rockey DD, Matsumoto A (2000) The chlamydial developmental cycle. In: Brun YV, Shimkets LJ, eds. *Prokaryotic Development*. Washington D. C.: AMS Press. pp 403–426.
- Abdelrahman YM, Belland RJ (2005) The chlamydial developmental cycle. *FEMS Microbiology Reviews* 29: 949–59.

Figure S3 Pyrimidine biosynthesis pathway of *C. psittaci* 6BC. (A) A scheme of the pyrimidine biosynthesis pathway of *C. psittaci* 6BC, *C. abortus* S26/3, *C. felis* Fe/C56, *C. pneumoniae* LPCoLN, and *C. trachomatis* L2/434/Bu modified after the KEGG PATHWAY database (www.genome.jp/kegg/pathway.html). Only *C. pneumoniae* encodes uridine kinase *udk* (EC 2.7.1.48). Only *C. trachomatis* L2/434/Bu lacks orotate phosphoribosyltransferase *pyrE* (EC 2.4.2.10). (B) Partial view of the pyrimidine biosynthesis pathway including gene designations: *pyrB*, aspartate carbamoyltransferase; *pyrC*, dihydroorotase; *pyrD*, dihydroorotate dehydrogenase; *pyrE*, orotate phosphoribosyltransferase; *pyrF*, orotidine 5-phosphate decarboxylase; *pyrH*, uridylylase kinase; *ndk*, nucleoside diphosphate kinase; *pyrG*, CTP synthase. Dashed arrows indicate predicted reaction directions supported by enzyme profiles available from the KEGG ENZYME database (<http://www.genome.jp/kegg/kegg3.html>).
(TIF)

Table S1 Predicted type III secreted effectors in *Chlamydia psittaci* 6BC.
(DOC)

Table S2 Predicted type III secreted effectors in *Chlamydia trachomatis* L2/434/Bu.
(DOC)

Table S3 Predicted type III secreted effectors in *Chlamydia abortus* S26/3.
(DOC)

Table S4 Predicted type III secreted effectors in *Chlamydia felis* Fe/C-56.
(DOC)

Table S5 Predicted type III secreted effectors in *Chlamydia pneumoniae* LPCoLN.
(DOC)

Table S6 Predicted type III secreted effectors in *Chlamydia caviae* GPIC.
(DOC)

Table S7 Type III secreted effectors in *Chlamydia psittaci* 6BC predicted by EffectiveT3 (<http://www.effectors.org/>)
(DOC)

Acknowledgments

We are grateful to Konrad Sachse for providing the strain material and Frank Hänel for critical discussions.

Author Contributions

Conceived and designed the experiments: AV GS HPS. Performed the experiments: AV GS. Analyzed the data: AV GS. Contributed reagents/materials/analysis tools: AV GS HPS. Wrote the paper: AV GS HPS.

11. Toft C, Andersson SGE (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Reviews Genetics* 11: 465–75.
12. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, et al. (2011) Unity in Variety – the Pan-Genome of the Chlamydiae. *Molecular Biology and Evolution* 28: 3253–3270.
13. Bertelli C, Collyn F, Croxatto A, Rückert C, Polkinghorne A, et al. (2010) The Waddlia genome: a window into chlamydial biology. *PLoS One* 5: e10890.
14. Rockey DD, Lenart J, Stephens RS (2000) Genome Sequencing and Our Understanding of Chlamydiae. *Infection and Immunity* 68: 5473–5479.
15. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, et al. (2000) Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Research* 28: 1397–406.
16. Valdivia RH (2008) *Chlamydia* effector proteins and new insights into chlamydial cellular microbiology. *Current Opinion in Microbiology* 11: 53–9.
17. Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews* 62: 379–433.
18. Wang Y, Zhang Q, Sun MA, Guo D (2011) High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics* 27: 777–84.
19. Arnold R, Brandmaier S, Kleine F, Tischler P, Heinz E, et al. (2009) Sequence-based prediction of type III secreted proteins. *PLoS Pathogens* 5: e1000376.
20. Seth-Smith HMB, Harris SR, Rance R, West AP, Severin JA, et al. (2011) Complete genome sequence of the zoonotic pathogen *Chlamydia psittaci*. *Journal of Bacteriology* 193: 1282–1283.
21. Grinblat-Huse V, Drabek EF, Creasy HH, Daugherty SC, Jones KM, et al. (2011) Genome Sequences of the Zoonotic Pathogens *Chlamydia psittaci* 6BC and Call0. *Journal of Bacteriology* 193: 4039–4040.
22. Pannekoek Y, Dickx V, Beekman DSA, Jolley KA, Keijzers WC, et al. (2010) Multi locus sequence typing of *Chlamydia* reveals an association between *Chlamydia psittaci* genotypes and host species. *PLoS One* 5: e14179.
23. Stephens RS, Myers G, Eppinger M, Bavoil PM (2009) Divergence without difference: phylogenetics and taxonomy of *Chlamydia* resolved. *FEMS Immunology and Medical Microbiology* 55: 115–9.
24. Read TD, Myers SR, Brunham RC, Nelson WC, Paulsen IT, et al. (2003) Genome sequence of *Chlamydia caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. *Nucleic Acids Research* 31: 2134–2147.
25. Kuo CC, Stephens RS, Bavoil PM, Kaltenboeck B (2011) Genus I, *Chlamydia* Jones, Rake and Stearns 1945, 55. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, et al., editor. 22 editors *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume Four. New York: Springer. 2 edition. pp 846–865.
26. Stephens RS, Lammel CJ (2001) *Chlamydia* outer membrane protein discovery using genomics. *Current Opinion in Microbiology* 4: 16–20.
27. Thomson NR, Yeats C, Bell K, Holden MTG, Bentley SD, et al. (2005) The *Chlamydia abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. *Genome Research* 15: 629–640.
28. Hackstadt T, Scidmore-Carlson MA, Shaw EI, Fischer ER (1999) The *Chlamydia trachomatis* InCA protein is required for homotypic vesicle fusion. *Cellular Microbiology* 1: 119–130.
29. Everett KD, Bush RM, Andersen AA (1999) Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards. *International Journal of Systematic Bacteriology* 49 Pt 2: 415–440.
30. Taylor LD, Nelson DE, Dorward DW, Whitmore WM, Caldwell HD (2010) Biological characterization of *Chlamydia trachomatis* plasticity zone MACPF domain family protein CT153. *Infection and Immunity* 78: 2691–9.
31. McClarty G, Caldwell HD, Nelson DE (2007) Chlamydial interferon gamma immune evasion influences infection tropism. *Current Opinion in Microbiology* 10: 47–51.
32. Fehner-Gardiner C, Roshick C, Carlson JH, Hughes S, Belland RJ, et al. (2002) Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. *The Journal of Biological Chemistry* 277: 26893–903.
33. Nelson DE, Crane DD, Taylor LD, Dorward DW, Goheen MM, et al. (2006) Inhibition of chlamydiae by primary alcohols correlates with the strain-specific complement of plasticity zone phospholipase D genes. *Infection and Immunity* 74: 73–80.
34. Mitchell CM, Hovis KM, Bavoil PM, Myers GSA, Carrasco JA, et al. (2010) Comparison of koala LPCoLN and human strains of *Chlamydia pneumoniae* highlights extended genetic diversity in the species. *BMC Genomics* 11: 442.
35. Peitsch MC, Tschopp J (1991) Assembly of macromolecular pores by immune defense systems. *Current Opinion in Cell Biology* 3: 710–716.
36. Rosado CJ, Kondos S, Bull TE, Kuiper MJ, Law RHP, et al. (2008) The MACPF/CDC family of pore-forming toxins. *Cellular Microbiology* 10: 1765–74.
37. Stebbins CE, Galán JE (2001) Structural mimicry in bacterial virulence. *Nature* 412: 701–5.
38. Xu Q, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, et al. (2010) Structure of a membrane-attack complex/perforin (MACPF) family protein from the human gut symbiont *Bacteroides thetaiotaomicron*. *Acta Crystallographica Section F, Structural Biology and Crystalization Communications* 66: 1297–305.
39. Wolf YI, Aravind L, Koonin EV (1999) Rickettsiae and Chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends in Genetics* 15: 173–5.
40. Ponting CP (1999) Chlamydial homologues of the MACPF (MAC/perforin) domain. *Current Biology* 9: 911–913.
41. Mojica S, Huot Creasy H, Daugherty S, Read TD, Kim T, et al. (2011) Genome Sequence of the Obligate Intracellular Animal Pathogen *Chlamydia pecorum* E58. *Journal of Bacteriology* 193: 3690.
42. Seth-Smith HMB, Harris SR, Persson K, Marsh P, Barron A, et al. (2009) Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. *BMC Genomics* 10: 239.
43. Thomson NR, Holden MTG, Carder C, Lennard N, Lockey SJ, et al. (2008) *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Research* 18: 161–71.
44. Grimwood J, Stephens RS (1999) Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. *Microbial & Comparative Genomics* 4: 187–201.
45. Gomes JaP, Nunes A, Bruno WJ, Borrego MJ, Florindo C, et al. (2006) Polymorphisms in the nine polymorphic membrane proteins of *Chlamydia trachomatis* across all serovars: evidence for serovar Da recombination and correlation with tissue tropism. *Journal of Bacteriology* 188: 275–86.
46. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, et al. (1998) Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* 282: 754–759.
47. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, et al. (1999) Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nature Genetics* 21: 385–9.
48. Henderson IR, Lam AC (2001) Polymorphic proteins of *Chlamydia* spp. – autotransporters beyond the Proteobacteria. *Trends in Microbiology* 9: 573–578.
49. Mølleken K, Schmidt E, Hegemann JH (2010) Members of the Pmp protein family of *Chlamydia pneumoniae* mediate adhesion to human cells via short repetitive peptide motifs. *Molecular Microbiology* 78: 1004–1017.
50. Peters J, Wilson DP, Myers G, Timms P, Bavoil PM (2007) Type III secretion à la *Chlamydia*. *Trends in Microbiology* 15: 241–251.
51. Beekman DSA, Vanrompay DCG (2010) Bacterial secretion systems with an emphasis on the chlamydial Type III secretion system. *Current Issues in Molecular Biology* 12: 17–41.
52. Rockey DD, Scidmore MA, Bannantine JP, Brown WJ (2002) Proteins in the chlamydial inclusion membrane. *Microbes and Infection* 4: 333–40.
53. Clifton DR, Fields KA, Grieshaber SS, Dooley CA, Fischer ER, et al. (2004) A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. *Proceedings of the National Academy of Sciences of the United States of America* 101: 10166–71.
54. Beekman DSA, Geens T, Timmermans JP, Van Oostveldt P, Vanrompay DCG (2008) Identification and characterization of a type III secretion system in *Chlamydia psittaci*. *Veterinary Research* 39: 27.
55. Wang J, Chen L, Chen F, Zhang X, Zhang Y, et al. (2009) A chlamydial type III-secreted effector protein (Tarp) is predominantly recognized by antibodies from humans infected with *Chlamydia trachomatis* and induces protective immunity against upper genital tract pathologies in mice. *Vaccine* 27: 2967–80.
56. Bannantine JP, Griffiths RS, Viratytosin W, Brown WJ, Rockey DD (2000) A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. *Cellular Microbiology* 2: 35–47.
57. Subtil A, Parsot C, Dautry-Varsat A (2001) Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery. *Molecular Microbiology* 39: 792–800.
58. Rockey DD, Heinzen RA, Hackstadt T (1995) Cloning and characterization of a *Chlamydia psittaci* gene coding for a protein localized in the inclusion membrane of infected cells. *Molecular Microbiology* 15: 617–26.
59. Gupta R, Srivastava P, Vardhan H, Salhan S, Mittal A (2009) Host immune responses to chlamydial inclusion membrane proteins B and C in *Chlamydia trachomatis* infected women with or without fertility disorders. *Reproductive Biology and Endocrinology* 7: 38.
60. Ohya K, Takahara Y, Kuroda E, Koyasu S, Hagiwara S, et al. (2008) *Chlamydia felis* CF0218 is a novel TMH family protein with potential as a diagnostic antigen for diagnosis of *C. felis* infection. *Clinical and Vaccine Immunology* 15: 1606–15.
61. Kimura M (1985) *The Neutral Theory of Molecular Evolution* Cambridge University Press. 384 p.
62. Novichkov PS, Wolf YI, Dubchak I, Koonin EV (2009) Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. *Journal of Bacteriology* 191: 65–73.
63. Koonin EV, Wolf YI (2010) Constraints and plasticity in genome and molecular-phenome evolution. *Nature Reviews Genetics* 11: 487–498.
64. Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, et al. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of Theoretical Biology* 239: 226–235.
65. Mamirova L, Popadin K, Gelfand MS (2007) Purifying selection in mitochondria, free-living and obligate intracellular proteobacteria. *BMC Evolutionary Biology* 7: 17.
66. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, et al. (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proceedings of the National Academy of Sciences of the United States of America* 100: 5455–60.

67. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-ponte T, Eriksson AS, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396: 133–143.
68. Razin S, Yogeve D, Naot Y (1998) Molecular biology and pathogenicity of mycoplasmas. *Microbiology and Molecular Biology Reviews* 62: 1094–156.
69. Vandahl BBS, Birkelund S, Christiansen G (2004) Genome and proteome analysis of *Chlamydia*. *Proteomics* 4: 2831–42.
70. Streit WR, Entcheva P (2003) Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Applied Microbiology and Biotechnology* 61: 21–31.
71. Horn M (2008) *Chlamydiae* as symbionts in eukaryotes. *Annual Review of Microbiology* 62: 113–131.
72. Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* 323: 379–82.
73. Tipples G, McClarty G (1991) Isolation and initial characterization of a series of *Chlamydia trachomatis* isolates selected for hydroxyurea resistance by a stepwise procedure. *Journal of Bacteriology* 173: 4932–40.
74. Case EDR, Akers JC, Tan M (2011) CT406 Encodes a Chlamydial Ortholog of NrdR, a Repressor of Ribonucleotide Reductase. *Journal of Bacteriology* 193: 4396–404.
75. McClarty G, Tipples G (1991) In situ studies on incorporation of nucleic acid precursors into *Chlamydia trachomatis* DNA. *Journal of Bacteriology* 173: 4922–31.
76. Hatch TP (1975) Utilization of L-Cell Nucleoside Triphosphates by *Chlamydia psittaci* for Ribonucleic Acid Synthesis. *Journal of Bacteriology* 122: 393–400.
77. Ceballos MM, Hatch TP (1979) Use of HeLa cell guanine nucleotides by *Chlamydia psittaci*. *Infection and Immunity* July: 98–102.
78. Tjaden J, Winkler HH, Schwöppe C, Van Der Laan M, Möhlmann T, et al. (1999) Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for transport of energy. *Journal of Bacteriology* 181: 1196–202.
79. Hatch TP, Al-Hossainy E, Silverman JA (1982) Adenine nucleotide and Lysine transport in *Chlamydia psittaci*. *Journal of Bacteriology* 150: 662–670.
80. Tipples G, McClarty G (1993) The obligate intracellular bacterium *Chlamydia trachomatis* is auxotrophic for three of the four ribonucleoside triphosphates. *Molecular Microbiology* 8: 1105–14.
81. Amici A, Emanuelli M, Magni G, Raffaelli N, Ruggieri S (1997) Pyrimidine nucleotidases from human erythrocyte possess phosphotransferase activities specific for pyrimidine nucleotides. *FEBS Letters* 419: 263–267.
82. Pesi R, Allegrini S, Careddu MG, Filoni DN, Camici M, et al. (2010) Active and regulatory sites of cytosolic 5-nucleotidase. *FEBS Journal* 277: 4863–4872.
83. Tribby IIE, Moulder JW (1966) Availability of bases and nucleosides as precursors of nucleic acids in L cells and in the agent of meningopneumonitis. *Journal of Bacteriology* 91: 2362–7.
84. McClarty G, Fan H (1993) Purine metabolism by intracellular *Chlamydia psittaci*. *Journal of Bacteriology* 175: 4662–9.
85. Voigt A, Schö G, Heidrich A, Sachse K, Saluz HP (2011) Full-Length De Novo Sequence of the *Chlamydia psittaci* Type Strain 6BC. *Journal of Bacteriology* 193: 2662–2663.
86. Shirai M, Hirakawa H, Kimoto M, Tabuchi M, Kishi F, et al. (2000) Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. *Nucleic Acids Research* 28: 2311–4.
87. Azuma Y, Hirakawa H, Yamashita A, Cai Y, Rahman MA, et al. (2006) Genome sequence of the cat pathogen, *Chlamydia felis*. *DNA Research* 13: 15–23.
88. Kent WJ (2002) BLAT—The BLAST-Like Alignment Tool. *Genome Research* 12: 656–664.
89. Guy L, Roat Kultima J, Andersson SGE (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26: 2334–2335.
90. Alves JaMP, Buck GA (2007) Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. *Chemistry & Biodiversity* 4: 2593–2602.
91. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28: 27–30.
92. Katoh K, Kuma Ki, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 51151–8.
93. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59: 307–321.
94. Whelan S, Goldman N (2001) A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution* 18: 691–699.
95. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics* 8: 77–80.
96. Shao J, Xu D, Tsai SN, Wang Y, Ngai SM (2009) Computational identification of protein methylation sites through bi-profile Bayes feature extraction. *PLoS One* 4: e4920.