Comparative genomics of swine pathogen *Chlamydia suis* with human- and mouse-adapted *Chlamydia* spp

Zoe E Dimond

University of Kansas

P. Scott Hefty (✉ pshefty@ku.edu)

University of Kansas <https://orcid.org/0000-0002-2303-2465>

Research article

Keywords: *Chlamydia suis*, pig, comparative genomics, chlamydiosis

Posted Date: July 31st, 2019

DOI: <https://doi.org/10.21203/rs.2.12226/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background The phylum, Chlamydiae, contains a unique set of obligately intracellular bacteria with specific host ranges, despite high genomic relatedness between species. *Chlamydia suis*, a ubiquitous swine pathogen, has the potential for zoonotic transmission to humans and often encodes for resistance to the primary treatment antibiotic, tetracycline. Because of this emerging threat, and to gain a better understanding of the basic biology, comparative genomics for a tetracycline-resistant swine isolate (R19) with inter- and intra-species genomes was performed. **Results** A 1.094Mb genome was determined through de novo assembly of Illumina high throughput sequencing reads. Annotation and subsystem analyses were conducted using novel bioinformatic platform, PATRIC, revealing 986 putative genes (Chls_###) that are predominantly orthologs to other known *Chlamydia* genes. Subsequent comparative genomics using *C. trachomatis* and *C. muridarum*, human and mouse pathogens, respectively, revealed a high level of genomic synteny and overall sequence identity. Around 857 genes were found to be conserved between these three species while 92 unique *C. suis* components were annotated. Direct comparison of chlamydia-specific gene families; inclusion membrane proteins, polymorphic membrane proteins and the major outer membrane protein as well as the plasticity zone, demonstrates the high gene content identity with *C. trachomatis* and *C. muridarum* as well as highlights putative host-specificity factors. **Conclusions** This study constitutes the first genome-wide comparative analysis for *C. suis*, generating a fully annotated reference genome using NIAID bioinformatic tool, PATRIC. Overall, these analyses build upon previous *C. suis* analyses to compare the composition of key genetic components with two closely related *Chlamydia* species. These studies will enable focused efforts on factors that provide key species specificity and adaptation to cognate hosts that are attributed to chlamydial infections, including humans. Continued analysis of the *C. suis* genome will facilitate studies into the complexities of the chlamydial proteome and discovery of basic biology.

Background

Chlamydia is a genus of biologically distinctive, obligate intracellular bacterial pathogens that cause disease in both humans and agriculturally important livestock. *Chlamydia suis* is a near-ubiquitous swine pathogen causing reproductive disorders, conjunctivitis, enteritis, rhinitis and pneumonia (1–4). Prevalence of *C. suis* disease within regions throughout the world has been investigated and those reports each found high incidence in farmed pigs. Providing a major challenge for porcine treatment, many *C. suis* strains have acquired a tetracycline-resistance cassette with clinical resistance confirmed in over seven countries (5–9). *In vitro* evidence has also demonstrated the ability for tetracycline resistance to be conveyed to *C. trachomatis* through horizontal gene transfer under conditions of co-infection (10, 11). Though traditionally restricted to its swine host, there is some evidence that *C. suis* may be zoonotically transmitted, although active human infection and clinical symptoms have not been observed (12–14).

The emerging potential threat of *C. suis* and ongoing concerns to the agricultural industry highlight a need for a comprehensive comparison of the *Chlamydia suis* genome to the already prominent human

pathogen, *Chlamydia trachomatis*. *C. trachomatis* is the most commonly reported bacterial infection worldwide and is the causative agent behind several important diseases including blinding trachoma and the sexually transmitted disease, *Chlamydia*. Importantly, genome sequencing and comparative studies between *Chlamydia* have enabled the identification and analysis of gene products that are diverse and may participate in host specific adaptations (15) (16) (17) (18). While there have been many intraspecies analyses of *C. suis* with intense focus on the tetracycline-resistance genes, including those by Donati et al (19), Joseph et al (20), Marti et al (21), and Seth-smith et al (22), interspecies comparisons, particularly with *C. trachomatis* and *C. muridarum*, the species most closely related to *C. suis*, have been limited. One key observation within *Chlamydia* is that despite the genetic similarity between species, there appear to be many host-restricting factors that are responsible for the variation in disease between species. We expect that by studying the *C. suis* genome in comparison to other *Chlamydia*, we can begin to tease out some of these key host-specificity factors.

The *C. suis* R19 strain is a tetracycline resistant strain isolated from a Nebraska farm (23) and has been widely studied for diagnostic and antibiotic resistance transfer capabilities. Previously, this strain was sequenced by Joseph et al with 14x coverage, but because this genome remains incomplete and with lower quality reads, we have used alternative methods to independently assemble a *C. suis* R19 genome and have used novel NIAID annotation pipeline, PATRIC, for genome annotation which was able to identify several new open reading frames and possible coded protein functions. These analyses have revealed that *C. suis* has a genome that is highly similar in content and organization to both *C. trachomatis* and *C. muridarum* and phylogenetic analysis confirms previous reports that *C. suis* is most closely related to these species. Accordingly, *C. suis* encodes for many genes similar to other *Chlamydia* but with unique combinations that are both distinct from as well as similar to *C. trachomatis* and *C. muridarum* which may play a role in the host-adaptation of *Chlamydia*.

Results And Discussion

The *Chlamydia suis* genome architecture and composition.

The *C. suis* R19 strain genome was sequenced using Illumina HiSeq with over 1000X coverage (Supplementary Figure 1). This enabled an unbiased, *de novo* assembly of the genome. The completed genome was 1,094,719 bp in length and a plasmid of 7,496 bp. Read density analysis of the plasmid relative to the chromosome supported approximately 2 copies of the plasmid per chromosome. The genome was annotated using both direct ORF prediction and BLAST homology-based annotation through Geneious and the PATRIC annotation platform revealing 986 coding regions (*Chlamydia suis*; Chls_###), with 6 rRNAs and 37 tRNAs, and an overall 41.7% G+C content. Relative features of the R19 genome are displayed in Figure 1A and Supplementary Table 1 with comparative data in Table 1.

The origin of replication for this genome was assigned according to the reference genome *C. trachomatis* L2 434/Bu (NC_010287) (24). Of the 986 coding sequences identified in *C. suis*, 686 were assigned through comparisons to other *Chlamydia* and through inferred functional assignment of known motifs

with a 75% identity threshold. The assigned genes were divided into 9 broad categories or defined as “other” using the subsystems approach and the RAST toolkit (25, 26). Overall, the largest division, at 165 genes, were predicted to be involved with protein processing followed closely by metabolism-related genes with 163 (Figure 1B). Additional subsystems include those involved with stress response, defense or virulence (50 genes), DNA processing (51 genes), energy (62 genes), RNA processing (31 genes), cellular processes (28 genes), cell envelope (28 genes) or membrane transport (3 genes). The remaining 282 genes had no homologues or identified motifs in Genbank and were annotated as hypothetical.

For comparison, *Chlamydia trachomatis* L2/434/Bu contains a predicted 962 genes while the *C. muridarum* Nigg genome consists of 983 genes, as re-annotated under the PATRIC annotation platform. *C. suis* contains 83 genes that do not appear to correlate with direct homologs in either *C. trachomatis* or *C. muridarum*, while most of these are hypothetical proteins, several are known members of the tetracycline resistance island. There is a high level of synteny, approximately 90%, in the genome organization for these three genomes as well as to other members of the phylum (Figure 2A). A complete genetic identity table comparing the *C. suis* R19 gene content to single representations from the chlamydial genomes in Figure 2A is provided in Supplementary Table 2. Both this genetic comparison, and genome-wide sequence phylogenies (Figure 2B) corroborate previous findings using 16S sequences that *C. suis* is in a phylogenetic clade with *C. trachomatis* and *C. muridarum* (27). Upon closer inspection of the genetic make-up of the genome, the “20 most phylogenetically informative proteins” as reported by Pillonel et al are present, intact, and highly similar to the other *Chlamydiales* (28). As in the other species in the genus, there is one plasmid with 8 genes encoding replicative machinery and putative virulence factors (29).

The *C. suis* R19 tetracycline resistance cassette, likely integrated by the IS605 transposase and horizontal gene transfer, is thought to be spreading among *C. suis* strains. Two separate analyses of the *tet* islands were conducted by Seth-Smith et al and Marti et al in 2017, both of which included R19 for comparison (16,17). Results demonstrated that the island was plastic, but that the key *tet* (C) gene was always intact, as it is for R19 (22) and that the cassette is transferable between *C. suis* strains by homologous recombination without the need for the transposase (21).

The *C. suis* plasticity zone

Arguably the most interesting genomic disparity between *Chlamydia* is the plasticity zone (PZ). This is a variable region present in all species of the genus that is hypothesized to contain genes responsible for host-specific virulence factors. The PZ of *C. muridarum* contains a series of phospholipases (PLDs), three genes often annotated as adherence factors that have cytotoxic domains similar to those in *Clostridium* spp, as well as a purine interconversion operon (30). In contrast, the significantly reduced PZ of *C. trachomatis* contains several of the phospholipases, but only contains a truncated version of one of the cytotoxin-resembling genes in specific serovars and has replaced the purine interconversion operon with a tryptophan biosynthesis pathway (Figure 3A) (31). Other *Chlamydia* contain permutations of these

plasticity zones where some contain no phospholipases, variable cytotoxin numbers with two present in *C. pecorum* and one ortholog in *C. caviae*, *C. felis*, and *C. psittaci*. The presence or absence of the tryptophan operon and close investigation of the PZ has provided insight into the immunological interactions between host and bacteria (32–34).

The *C. suis* PZ contains two adherence factors, both of which are similar to two of the *C. muridarum* cytotoxin-like genes as also determined by Manuela et al for *C. suis* MD56 (Figure 3) (35). To investigate the presence of these cytotoxins in other strains of *C. suis*, full or partial genomes were analyzed for 29 deposited SRA samples using the *C. suis* R19 genome as a reference (22). Interestingly, when these strains of *C. suis* were examined, seven were discovered to contain only one of the adherence factors, indicating that both copies are likely non-essential for the *in vivo* success of the organism, as also noted by Seth-Smith et al (22). It is unclear as of yet, whether the two cytotoxins have distinct biological functions *in vivo* or whether they play a role in host or biovar selectivity.

Also found in the *C. suis* PZ is an intact tryptophan biosynthesis pathway, homologous to the one in *C. trachomatis* genital strains. Complete tryptophan biosynthesis operons are also found in *C. pecorum* and *C. felis*. Tryptophan levels are known to play a key role in the inhibiting pathogenesis of intracellular organisms, including *Chlamydia*. Human (36) and pig (37) innate immune responses are able to limit the available tryptophan through the indoleamine 2,3-dioxygenase (38) response pathway mediated by interferon gamma (IFN- γ). Briefly, IFN- γ , activated by a chlamydial infection, will induce IDO to catalyze the breakdown of L-tryptophan into N-formylkynurenine effectively depleting available tryptophan available for the bacteria. Indole producers, like many members of the human vaginal microbiome, provide the necessary input for salvage by the *trp* operon. Importantly, ocular *C. trachomatis* do not encode the *trp* operon, potentially due to the absence of indole-producing microbes. Sherchand et al demonstrate that the *trp* operon can be deleterious to *Chlamydia* in the absence of indole-producers (39). It would follow that perhaps a strong negative selection on the *trp* operon due to a lack of indole-producers in *C. muridarum*-infected tissues would explain the absence of the *trp* operon in *C. muridarum*. There is no current evidence to support an IFN- γ -mediated IDO response in mice which further supports the negative selection of the *trp* operon in mouse-infecting *Chlamydia*. Overall, this pathway could provide a hypothesis as to why there are no incidences of *C. muridarum* infection in humans, but there have been reports of *C. suis* present in human samples while detection of active human infection or symptomology has not been reported.

As is seen in *C. muridarum* and *C. trachomatis*, the PZ of *C. suis* contains a putative operon encoding phospholipases (PLDs) and a conserved MAC/perforin-like protein. The operon in *C. suis*, however, seems to contain an increased number of PLDs relative to both *C. trachomatis* and *C. muridarum* (Fig. 3A). Eukaryotic phospholipase D has roles in lipid metabolism and vesicle regulation and this family of proteins has been exploited by pathogens and used to increase virulence (40). While the *in vivo* role of these genes in *Chlamydia* remains relatively unknown, a study done with *C. trachomatis* in HeLa cells provides evidence that the PZ phospholipases (pzPLDs) may be important for inclusion formation (41). Other bacterial phospholipase D genes have been characterized, however, and could provide insight into

their role in *Chlamydia*. A study in *Neisseria gonorrhea* showed that a phospholipase D homolog acts to increase adherence and invasion to cervical cells by stimulating complement receptor type 3-mediated endocytosis (42). This effect was species specific, as other bacterial phospholipase D proteins were not able to rescue a knock out mutant. In *Yersinia pestis*, a plasmid-encoded phospholipase D allows the pathogen to survive in its arthropod host further suggesting that these proteins may play a role in host-specificity (43).

The make-up of *Chlamydia* PLD family genes, therefore, may affect the ability of each species to infect its respective host(s). As noted, *C. suis* contains more PLD genes within its PZ than almost any other chlamydial species. In fact, several species including *C. caviae*, *C. pneumoniae* and many parachlamydiae are missing any PLD family genes in the plasticity zone, but retain ancestral chromosomal PLD outside of the PZ (41). The variability in number and sequence of this PLD operon, and the putative roles in virulence and host specificity from other bacterial species suggests that these enzymes could be providing an essential function in the manipulation of the unique host cell allowing the species to infect and survive within its specific host.

Phylogenetic analysis of the repertoire of PLD enzymes, as identified through their HxKx₄Dx₆G domains, revealed a distinct clade of chromosomal PLD which mirror classic phylogenetic relationships for the *Chlamydia* (Figure 3B). This indicates that this chromosomal PLD is likely truly ancestral and has evolved with the speciation of *Chlamydia*. The pzPLDs, however, are not so simply categorized. There is a great diversity in these sequences, with a subset of proteins more closely related the chromosomal PLD, perhaps representing duplication events. The clade more distantly related the chromosomal PLDs include the greatest number of pzPLDs, potentially pointing to a unique role for this subset of PLDs. The division between clades could suggest differential selection for these PLDs which would support a distinct biological function.

***C. suis* membrane proteins**

One key protein differentiating members of *Chlamydiaceae* is the major outer membrane protein (Momp, *ompA*). This porin-like structure is integrated into the outer membrane of all *Chlamydia* species and has been shown to make up approximately 60% of the chlamydial outer membrane (44). Both the completed structure and function of this protein remain a mystery, but its conserved and highly present nature indicates that this gene is crucial to bacterial survival in the host. In 2016, a study was conducted which looked at variation in *ompA* among strains of *C. muridarum* collected from wild mice and found through PCR analysis, a 99% identity between *ompA* genes (45). Similarly, several strains of *Chlamydia pecorum*, *Chlamydia abortus*, *Chlamydia pneumoniae*, and *Chlamydia psittaci* (46), have been deposited and close examination of these strains shows little *ompA* diversity. In contrast, *C. trachomatis ompA* has increased variability between strains when compared with *C. muridarum* and has allowed classification into unique serovars or clades based solely on *ompA*-typing.

While it has been shown that genotyping by *ompA* does not indicate genomic relatedness (47), these divisions do appear to correlate with specific pathobiovars. In *C. trachomatis*, *ompA* appears to be evolving at a faster rate than the rest of the genome, likely due to unique selective pressures, (48) and is characterized by four variable domains (VD) which are used to distinguish serovars. These variable domains are generally regions of lower hydrophobicity and potentially indicate outward facing domains, leading to the hypothesis for the biovar-specificity attributed to this gene. The variability in *ompA* could confer an evolutionary advantage specific to the host organism and the specific tissues involved in infection.

A recent analysis by Chahota et al investigated the potential for serotyping by Momp in *C. suis* using PCR analysis of VD2 and VD3 (49). When the deposited contigs were investigated for completed genes for *ompA*, and compared to R19, similar results were found. *C. suis* Momp appears to also be divided into four variable domains which directly correspond to those seen in *C. trachomatis*, though are divergent (Figure 4). Phylogenetic analysis of the *C. suis ompA* divides the strains into distinct clades, corroborating the observation by Chahota et al, that *C. suis* strains could be serotyped and a larger study could be performed to investigate the potential for distinctions between the clades.

There are several other important gene families that show some variability between the *Chlamydiales*, including a family of inclusion membrane proteins (Incs), and polymorphic membrane proteins (Pmps). Both families have been shown to be correlated with tissue tropism and therefore, may play a role in speciation (50). The chlamydial inclusion is a modified vacuole in which *chlamydia* are able to grow and divide. Relatively little is known about the modifications made to this vacuole that allow for the bacteria's survival, but the Inc family of proteins are secreted into the inclusion membrane and are known to interact with the host cytosol. Inc proteins make up approximately 7–10% of the chlamydial proteome and have a high degree of diversity (51). Despite this diversity, most Incs are defined by a characteristic bilobed hydrophobicity motif. This domain is conserved among predicted Inc proteins and seems to be a *Chlamydia*-specific motif.

There are varying predictions for the numbers of Inc proteins across the genus with 23 conserved *incs* encoded in 5 examined species (52). The *C. suis* R19 genome encodes each of these conserved *incs* as well as 36 non-conserved predicted Incs for a total of 58 predicted Incs. According to Lutter et al, *C. trachomatis* also encodes for 55 *incs* with 6 that differ from *C. muridarum*, which has 53 Incs. Analysis of the 58 putative inclusion membrane proteins in *C. suis* that were identified and aligned with *C. trachomatis* and *C. muridarum* revealed an additional 2 putative Incs for each species. A Venn diagram and phylogeny were constructed to display Incs shared between or specific to the three species and their evolutionary relationship (Figure 5). Briefly, 51 Incs are conserved within the three species of the clade. One Inc is unique to *C. muridarum* (TC0011, Figure 5B). Several Incs are shared between species, with 5 Incs found in *C. trachomatis* and *C. suis* only; CT529 (Chls_876), CT222 (Chls_534), CT224 (Chls_538), CT225 (Chls_539) and CT164 (Chls_474) (Figure 5C). Three of these are found in an operon of inclusion membrane proteins and possibly function together as a complex. Additionally, three Incs are unique to *C. muridarum* and *C. suis*, TC0239 (Chls_262), TC0171 (Chls_829) and TC0573 (Chls_617). These are not

phylogenetically similar and likely stem from an ancestral duplication event before the evolutionary division of the two species. Interestingly, and unexpectedly, due to the placement of *C. suis* as an intermediate between *C. trachomatis* and *C. muridarum*, there appears to be one Inc that is absent in *C. suis*, but found in the other two species (CT227/TC0498) (Figure 5C, panel 6). This Inc is closely related to the neighboring Inc CT226/TC0497 which is also present in *C. suis*. In general, the conservation of IncS between *C. suis*, *C. muridarum* and *C. trachomatis* suggests that the protein functions may be retained in all three species and that they may share cognate host interactions. Non-conserved IncS may indicate key differences in host interactions or the host environment in general (Figure 5C). Functional analysis of host-binding or -interacting partners for each of these IncS, could reveal their specific roles in their host species.

Like IncS, polymorphic membrane proteins (Pmps) share in varying numbers across species, though the specific roles for each Pmp are unknown. Pmps are outer membrane proteins and are immunogenic for humans. These genes make up approximately 1–2% of the chlamydial gene content and each contain a C-terminal phenylalanine and as well as multiple GGAI motifs which are associated with host cell adhesion (53). Nine putative *pmps* were annotated in *C. suis* R19 through a genome wide search for multiple GGAI motifs. Each of these correspond directly to a Pmp in *C. trachomatis* and *C. muridarum*. These Pmp primary sequences range from high identity between the three species to low identity: PmpA (79.8%), PmpI (78.8%), PmpH (77.0%), PmpB (74.9%), PmpG (74.0%), PmpE (73.9%), PmpD (72.9%), PmpF (65.2%) and PmpC (46.3%). Seven out of nine *C. suis* R19 Pmps are more similar to their *C. muridarum* orthologs while two (PmpH and PmpF) are more similar to *C. trachomatis*. Interestingly, Pmps have been shown to be involved in cellular tropism as six of the nine Pmps in *C. trachomatis* analyzed from different serovars were able to be phylogenetically clustered based on disease properties, suggesting a role for these membrane proteins in adhesion or differential biovar tropism (53).

Conclusions

The *C. suis* R19 genome was determined through whole genome sequencing and fully annotated using PATRIC. One key aspect of this study was the use of NIAID annotation service, PATRIC, for independent annotation. As indicated in Table 1, the PATRIC platform has annotated several genes not previously annotated by other methods in both *C. trachomatis* and *C. muridarum*. It is likely that these new annotations correspond with known open reading frames that were not originally included in the first depositions of these genomes. These could correspond with pseudogenes or untranscribed genes, as many are small <300bp open reading frames, which could be determined through further transcriptional analysis. Nevertheless, this emphasizes the need for continual evaluation of genomic content and locus tag numbering system.

Direct comparisons of *C. suis* R19 with other canonical *Chlamydia*, and specifically with *C. trachomatis* and *C. muridarum*, the closest relatives to *C. suis*, has allowed for the discovery of the unique characteristics of the *C. suis* genome. The high degree of synteny with *C. trachomatis* and *C. muridarum*

likely stems from a relatively recent evolutionary divergence and speaks to the similarities between their respective hosts; human, mouse and pig.

Several distinct aspects of *C. suis* emerge, however, likely due to adaptations to the porcine host. These include the those in the plasticity zone: a tryptophan biosynthesis operon absent in *C. muridarum*, adherence factors lost in *C. trachomatis* that show variable representation in *C. suis* and a unique subset of phospholipase D-like genes. Outside of the plasticity zone, the presence of variable regions in *ompA* suggest a similar role to *C. trachomatis*, likely involved with tissue tropism, as *C. suis* is known to inhabit several distinct tissues within pigs including eyes, gastrointestinal tract, respiratory tract and reproductive tracts. Other membrane proteins, *incs* and *pmps*, have a similar repertoire as found in *C. trachomatis* and *C. muridarum*. Additional studies on the roles of the non-conserved genes may provide insight into the biological role they may play as well as to provide insight into the host environment. One limitation of this study is in the use of few or single representative genomes from each species. To fully investigate the roles for these key families of proteins, a global analysis of *C. suis* strains, as well as other clinical *Chlamydia* isolates, collected from individual tissues or geographic locations, may be necessary. As with the small-scale studies performed by Chahota et al and Seth-Smith et al on *ompA* and the *tet* island, respectively, a larger-scale study could reveal the dynamics between host and gene content (22, 49).

Based on the evidence for *C. suis* evolution with *C. trachomatis* and *C. muridarum* and the presence of several genes present in *C. trachomatis* and absent in *C. muridarum*, *C. suis* may provide a better model for human chlamydial infection than *C. muridarum*. In order to evaluate this, several questions would need to be answered. Primarily, does *C. suis* infect, ascend and cause pathology in a mouse? One study has provided some evidence that *C. suis* may result in a more robust infection in mice than *C. trachomatis* (54), but the absence of direct infection analyses and comparison with *C. muridarum* leaves some uncertainty related to this hypothesis.

Comparative analysis of the *C. suis* genome opens the door for further studies into the complexities of the chlamydial genetic repertoire, proteome and host specificity adaptations. Interestingly, individual *Chlamydia* can display diverse but limited ranges of mammalian hosts. Livestock pathogens like *C. pecorum*, which cause chlamydiosis in a variety of animals particularly ruminants and swine, and is notably, the leading cause of infectious disease in koalas and *C. abortus*, which also infects ruminants and has been associated with spontaneous abortions in swine and sheep, are among the few able to infect and cause disease in a multiple ruminants species, while the majority of *Chlamydia*, including *C. suis*, *C. trachomatis* and *C. muridarum* appear to be largely restricted to a single host. Comparative genomic studies, like this report, will continue to identify molecular candidates potentially involved in restricting host range and enable direct genetic analyses to investigate the role of these host specific factors.

Methods

Whole Genome Sequencing and De Novo Assembly

Using the DNeasy Blood and Tissue Kit (Qiagen), DNA was isolated and purified from *C. suis* strain R19, provided by Dr. Daniel D Rockey (8). Library generation was performed at the University of Kansas Genome Sequencing Core using the NEBNext Ultra II DNA Library kit. Libraries were run on the Illumina Miseq PE100. Over 90 million reads, of 151bp, were obtained. Reads were trimmed and quality filtered using BBDuk. Subsequent de novo assembly was performed in the Geneious (version 9.1.8) software suite using Velvet v1.2.10 with the Velvet optimizer to determine an optimal Kmer (55). When needed, to resolve larger gaps and to verify final assembly, PCR and Sanger sequencing was used. Predicted coverage to the assembled R19 genome was over 1000x. The origin of replication and the first nucleotide was assigned using *C. trachomatis* L2 434/Bu (NC_010287) as a reference and genomes were adjusted accordingly.

Annotation

Annotation was performed using the RAST tool kit (RASTtk) through PATRIC web resources (56) and using open reading frame prediction through Geneious. Functional assignments of Geneious-identified open reading frames were verified through BLAST analysis against protein sequences in the NCBI database. Subsystems were determined through PATRIC automatic annotation.

Phylogenetic Analyses

Comparison of genetic content was performed using the following reference genomes: *C. suis* SWA-2 (NZ_LT821323), *C. suis* SWA-14 (NZLT860207), *C. suis* SWA-86 (NZ_LT860209), *C. suis* 2-26b (NZ_LT999997), *C. suis* 3-25b (NZ_LT999998), *C. trachomatis* D/UW-3/CX, *C. trachomatis* L2 434/Bu, *C. trachomatis* A/HAR-13 (NC_007429.1), *C. trachomatis* B/TZ1A828/OT (NC_012687.1), *C. trachomatis* E/Bour (NC_020971.1), *C. trachomatis* F/SW4 (NC_017951.1), *C. trachomatis* G/SotonG1 (NC_020941.1), *C. trachomatis* Ia/SotonIA1 (NC_020970.1), *C. trachomatis* K/SotonK1 (NC_020965.1), *C. muridarum* Nigg (NC_002620), *C. muridarum* MopnTet14 (NZ_ACUJ01000001.3), *C. muridarum* Nigg3 (NZ_CP009760.1), *C. muridarum* Weiss (ACOW01000004.1), *C. pneumoniae* strain Wien 1 (NZ_LN846980), *C. pecorum* strain strain E58 (NC_015408), *C. psittaci* GR9 (NC_018620), *C. abortus* strain GIMC 2006:CabB577 (NZ_CP024084), *C. caviae* GPIC (NC_003361), *C. felis* Fe/C-56 (NC_007899), and *C. avium* 10DC88 (NZ_CP006571). Genomic alignments were performed in Geneious using MAFFT (v7.309) alignment tools and coding sequence alignments were performed with the PATRIC proteome comparison web tool which uses bidirectional BLASTP with the following parameters: minimum coverage of 30%, minimum identity of 10%, BLAST E-value of 1e-5. Phylogenetic trees comparing single genes were constructed using the Geneious Tree Builder for a global alignment with free end gaps and a cost matrix set to Blosum65. Distances were obtained from pairwise alignments of all sequence pairs. To assemble the trees, a Jukes-Cantor genetic distance model was selected and UPGMA tree-building was performed. Nucleotide-based alignments and analyses were also performed in all cases and resulting phylogenies showed similar relationships to protein sequences (Data not shown). Whole genome alignments and subsequent phylogenies to determine genomic evolutionary relationships were performed

using the progressiveMauve algorithm with a match seed weight of 15 and a minimum LCB score of 30,000. Gaps were aligned using Muscle 3.6 and a phylogenetic tree was determined using the neighbor-joining method with bootstrapping.

Declarations

Ethics approval and consent to participate

Not Applicable

Consent for Publication

Not Applicable

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the NCBI repository, www.ncbi.nlm.nih.gov under the accession numbers CP034310 and CP034311.

Competing Interests

The authors declare that they have no competing interests

Funding

ZED was supported by NIH AI126785. PSH was supported by NIH AI126785 and P20GM113117. Genome sequencing was supported by NIH P20GM103638. NIH or associated administrative personnel had no direct influence on study design, analysis, interpretation of the data, or writing this manuscript.

Author's Contributions

ZED assembled and annotated the genome, performed all subsequent comparative analyses and wrote the manuscript. PSH was a major contributor of analytical direction and in revising the manuscript. All

authors read and approved the final manuscript.

Acknowledgements

Special appreciation to Daniel Rockey (Oregon State University) for manuscript review. We would like to acknowledge Rebecca Wattam and the American Society for Microbiology for providing a workshop on the PATRIC suite of tools.

References

1. Rogers DG AA, Hunsaker BD. Lung and nasal lesions caused by a swine chlamydial isolate in gnotobiotic pigs. . J Vet Diagn Invest. 1996;8(1):45-55.
2. Sachse K, Grossmann E, Berndt A, Schutt C, Henning K, Theegarten D, et al. Respiratory chlamydial infection based on experimental aerosol challenge of pigs with *Chlamydia suis*. Comp Immunol Microbiol Infect Dis. 2004;27(1):7-23.
3. Schiller I KR, Weilenmann R, Thoma R, Kaltenboeck B, Heitz P, Pospischil A. Mixed infections with porcine *Chlamydia trachomatis/pecorum* and infections with ruminant *Chlamydia psittaci* serovar 1 associated with abortions in swine. Vet Microbiol. 1997;58(2-4):251-60.
4. Rogers DG AA. Conjunctivitis caused by a swine *Chlamydia trachomatis*-like organism in gnotobiotic pigs. J Vet Diagn Invest. 1999;11(4):341-4.
5. Borel N RN, Di Francesco A, Donati M, Markov J, Masserey Y, Pospischil A. Selection for tetracycline-resistant *Chlamydia suis* in treated pigs. Vet Microbiol. 2012;156(1-2):143-6.
6. DiFrancesco A DM, Rossi M, Pignanelli S, Shurdhi A, Baldelli R, Cevenini R. Tetracycline-resistant *Chlamydia suis* isolates in Italy. Vet Rec. 2008;163(8):251-2.
7. Kaltenboeck B, Kousoulas KG, Storz J. Two-step polymerase chain reactions and restriction endonuclease analyses detect and differentiate ompA DNA of *Chlamydia* spp. J Clin Microbiol. 1992;30(5):1098-104.
8. J. Lenart AAA, D.D Rockey. Growth and Development of Tetracycline-Resistant *Chlamydia suis*. Antimicrob Agents Chemother. 2001;45(8):2198-203.
9. Schautteet K DCE, Miry C, Van Groenweghe F, Delava P, Kalmar I, Vanrompay D. Tetracycline-resistant *Chlamydia suis* in cases of reproductive failure on Belgian, Cypriote and Israeli pig production farms. J Med Microbiol. 2013;62(Pt 2):331-4.
10. Jeffrey BM, Suchland RJ, Eriksen SG, Sandoz KM, Rockey DD. Genomic and phenotypic characterization of in vitro-generated *Chlamydia trachomatis* recombinants. BMC Microbiol. 2013;13:142.
11. Suchland RJ, Sandoz KM, Jeffrey BM, Stamm WE, Rockey DD. Horizontal transfer of tetracycline resistance among *Chlamydia* spp. in vitro. Antimicrob Agents Chemother. 2009;53(11):4604-11.

12. Dean D RJ, Ruetttger A, Kandel RP, Sachse K. Zoonotic Chlamydiaceae Species Associated with Trachoma, Nepal. *Emerg Infect Dis.* 2013;19(12):1948-55.
13. De Puyssseleyr L DPK, Braeckman L, Morre SA, Cox E, Vanrompay D. Assessment of *Chlamydia suis* Infection in Pig Farmers. *Transbound Emerg Dis.* 2017;64(3):826-33.
14. De Puyssseleyr K, De Puyssseleyr L, Dhondt H, Geens T, Braeckman L, Morre SA, et al. Evaluation of the presence and zoonotic transmission of *Chlamydia suis* in a pig slaughterhouse. *BMC Infect Dis.* 2014;14(14):560.
15. Xie G, Bonner CA, Jensen RA. Dynamic diversity of the tryptophan pathway in chlamydiae: reductive evolution and a novel operon for tryptophan recapture. *Genome Biol.* 2002;3(9):research0051.
16. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* 2000;28(6):1397-406.
17. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, et al. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nat Genet.* 1999;21(4):385-9.
18. Girjes AA, Carrick FN, Lavin MF. Remarkable sequence relatedness in the DNA encoding the major outer membrane protein of *Chlamydia psittaci* (koala type I) and *Chlamydia pneumoniae*. *Gene.* 1994;138(1-2):139-42.
19. Donati M H-CH, Humphrys M, Di Paolo M, Di Francesco A, Myers GS. Genome Sequence of *Chlamydia suis* MD56, Isolated from the Conjunctiva of a Weaned Piglet. *Genome Announc.* 2014.
20. Joseph SJ, Marti H, Didelot X, Read TD, Dean D. Tetracycline Selective Pressure and Homologous Recombination Shape the Evolution of *Chlamydia suis*: A Recently Identified Zoonotic Pathogen. *Genome Biol Evol.* 2016;8(8):2613-23.
21. Marti H, Kim H, Joseph SJ, Dojiri S, Read TD, Dean D. Tet(C) Gene Transfer between *Chlamydia suis* Strains Occurs by Homologous Recombination after Co-infection: Implications for Spread of Tetracycline-Resistance among Chlamydiaceae. *Front Microbiol.* 2017;8(156):156.
22. Seth-Smith HM WS, Bachmann N, Marti H, Qi W, Donati M, di Francesco A, Polkinghorne A, Borel N. The *Chlamydia suis* Genome Exhibits High Levels of Diversity, Plasticity, and Mobile Antibiotic Resistance: Comparative Genomics of a Recent Livestock Cohort Shows Influence of Treatment Regimes. *Genome Biol Evol.* 2017;9(3):750-60.
23. Andersen AA RD. Resistance to tetracycline and sulfadiazine in swine *C. trachomatis* isolates. *Chlamydial Infections - Proceedings of the Ninth International Symposium on Human Chlamydial Infection.* 1998 313-6.
24. Thomson NR HM, Carder C, Lennard N, Lockey SJ, Marsh P, Skipp P, O'Connor CD, Goodhead I, Norbertzcak H, Harris B, Ormond D, Rance R, Quail MA, Parkhill J, Stephens RS, Clarke IN. *Chlamydia trachomatis*: Genome sequence analysis of lymphogranuloma venereum isolates. *Renome Res.* 2008;18(1):161-71.
25. Overbeek R BT, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen

- R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweyer H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. *Nucleic Acids Res.* 2005;33(17):5691-702.
26. Overbeek R OR, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014.
 27. Pudjiatmoko FH, Ochiai Y, Yamaguchi T, Hirai K. Phylogenetic Analysis of the Genus *Chlamydia* Based on 16s rRNA Gene Sequences. *Int J Syst Bacteriol.* 1997;47(2):425-31.
 28. Pillonel T, Bertelli C, Salamin N, Greub G. Taxogenomics of the order Chlamydiales. *Int J Syst Evol Microbiol.* 2015;65(Pt 4):1381-93.
 29. DD R. Unraveling the basic biology and clinical significance of the chlamydial plasmid. *J Exp Med.* 2011;208(11):2159-62.
 30. Read TD BR, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM Genome Sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* 2000;28(6):1397-406.
 31. Stephens RS KS, Lammel C, Fan J, Marathe R, Aravind L, Mitchcel W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. Genome Sequence of an Obligate Intracellular Pathogen of Humans: *Chlamydia trachomatis*. *Science.* 1998;282(5389):754-9.
 32. Rajaram K, Giebel AM, Toh E, Hu S, Newman JH, Morrison SG, et al. Mutational Analysis of the *Chlamydia muridarum* Plasticity Zone. *Infect Immun.* 2015;83(7):2870-81.
 33. Sait M LM, Clark EM, Wheelhouse N, Spalding L, Markey B, Magnino S, Lainson FA, Myers GS, Longbottom D. Genome sequencing and comparative analysis of three *Chlamydia pecorum* strains associated with different pathogenic outcomes. *BMC Genomics.* 2014.
 34. Jelocnik M BN, Kaltenboeck B, Waugh C, Woolford L, Speight KN, Gillett A, Higgins DP, Flanagan C, Myers GS, Timms P, Polkinghorne A. Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas. *BMC Genomics.* 2015;16:893.
 35. Manuela D H-CH, Humphrys M, Di Paolo M, Di Francesco A, Myers GSA. Genome Sequence of *Chlamydia suis* MD56, Isolated from the Conjunctiva of a Weaned Piglet. *Genome Announc.* 2014;2(3):e00425-14.
 36. Roshick C, Wood H, Caldwell HD, McClarty G. Comparison of gamma interferon-mediated antichlamydial defense mechanisms in human and mouse cells. *Infect Immun.* 2006;74(1):225-38.
 37. Meurens F, Summerfield A, Nauwynck H, Saif L, Gerdt V. The pig: a model for human infectious diseases. *Trends Microbiol.* 2012;20(1):50-7.

38. Bailey M CZ, Lewis MC. The evolutionary basis for differences between the immune systems of man, mouse, pig and ruminants. *Vet Immunol Immunopathol.* 2013;152(1-2):13-9.
39. Sherchand SP, Aiyar A. Ammonia generation by tryptophan synthase drives a key genetic difference between genital and ocular *Chlamydia trachomatis* isolates. *Proc Natl Acad Sci U S A.* 2019.
40. Selvy PE LR, Lindsley CW, Brown HA. Phospholipase D - enzymology, functionality, and chemical modulation. *Chem Rev.* 2011;111(10):6064-119.
41. Nelson DE CD, Taylor LD, Dorward DW, Goheen MM, Caldwell HD. Inhibition of *Chlamydiae* by Primary Alcohols Correlates with the Strain-Specific Complement of Plasticity Zone Phospholipase D Genes. *Infect Immun.* 2006;74(1):73-80.
42. Edwards JL ED, Apicella MA. Gonococcal Phospholipase D Modulates the Expression and Function of Complement Receptor 3 in Primary Cervical Epithelial Cells. *Infect Immun.* 2003;71(11):6381-91.
43. Hinnebusch BJ RA, Cherepanov P, Dixon JE, Schwan TG, Forsberg A. Role of *Yersinia* Murine Toxin in Survival of *Yersinia pestis* in the Midgut of the Flea Vector. *Science.* 2002;296(5568):733-5.
44. Caldwell HD, Kromhout J, Schachter J. Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*. *Infect Immun.* 1981;31(3):1161-76.
45. Ramsey KH SI, Sripsema JH, Denman CJ, Bowlin AK, Myers GA, Rank RG. Strain and Virulence Diversity in the Mouse Pathogen *Chlamydia muridarum*. *Infect Immun.* 2009;77(8):3284-93.
46. Schofl G VA, Litsche K, Sachse K, Saluz HP. Complete Genome Sequences of Four Mammalian Isolates of *Chlamydia psittaci*. *J Bacteriol.* 2011;193(16):4258.
47. Suchland RJ DZ, Putman TE, Rockey DD. Demonstration of Persistent Infections and Genome Stability by Whole-Genome Sequencing of Repeat-Positive, Same-Serovar *Chlamydia trachomatis* Collected From the Female Genital Tract. *J Infect Dis.* 2017;215(11):1657-65.
48. Nunes A, Borrego MJ, Nunes B, Florindo C, Gomes JP. Evolutionary dynamics of *ompA*, the gene encoding the *Chlamydia trachomatis* key antigen. *J Bacteriol.* 2009;191(23):7182-92.
49. Chahota R, Ogawa H, Ohya K, Yamaguchi T, Everett KD, Fukushi H. Involvement of multiple *Chlamydia suis* genotypes in porcine conjunctivitis. *Transbound Emerg Dis.* 2017.
50. Almeida F BV, Ferreira R, Borrego MJ, Gomes JP, Mota LJ. Polymorphisms in Inc Proteins and Differential Expression of *inc* Genes among *Chlamydia trachomatis* Strains Correlate with Invasiveness and Tropism of Lymphogranuloma Venereum Isolates. *J Bacteriol.* 2012;194(23):6574-85.
51. Dehoux P, Flores R, Dauga C, Zhong G, Subtil A. Multi-genome identification and characterization of *chlamydiae*-specific type III secretion substrates: the Inc proteins. *BMC Genomics.* 2011;12:109.
52. Lutter EI, Martens C, Hackstadt T. Evolution and conservation of predicted inclusion membrane proteins in *chlamydiae*. *Comp Funct Genomics.* 2012;2012:362104.
53. Gomes JP NA, Bruno WJ, Borrego MJ, Florindo C, Dean D. Polymorphisms in the Nine Polymorphic Membrane Proteins of *Chlamydia trachomatis* across All Serovars: Evidence for Serovar Da Recombination and Correlation with Tissue Tropism. *J Bacteriol.* 2006;118(1):275-86.

54. Donati M, Di Paolo M, Favaroni A, Aldini R, Di Francesco A, Ostanello F, et al. A mouse model for *Chlamydia suis* genital infection. *Pathog Dis.* 2015;73(1):1-3.
55. Kearse M MR, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-9.
56. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 2014;42(Database issue):D581-91.

Table

Table 1. Comparative Characteristics of the <i>C. suis</i> R19 Genome			
	<i>Chlamydia suis</i> R19	<i>Chlamydia trachomatis</i> 434/Bu	<i>Chlamydia muridarum</i> Nigg
Chromosome Length (bp)	1,094,719	1,038,842	1,072,950
Chromosomal genes (PATRIC)	986	962	983
NCBI reported chromosomal genes	N/A	937	952
Plasmid length (bp)	7,496	7,499	7,501
Plasmid genes	8	8	8
GC Content (%)	41.7	41.3	40.3

Figures

A.

- Chromosome
- FWD CDS
- REV CDS
- Ribosomal RNA
- Virulence Factor Genes
- Transporter Genes
- GC Content
- GC Skew
- Membrane-associated Genes

B.

Subsystem (Subsystems, Genes)

- Protein processing (35, 165)
- Metabolism (30, 163)
- Stress response, defense, virulence (14, 50)
- DNA processing (12, 51)
- Energy (12, 62)
- RNA processing (11, 31)
- Cellular processes (7, 28)
- Cell envelope (3, 11)
- Membrane transport (3, 3)

Figure 1

A) Annotated Genome for *Chlamydia suis* R19. Rings, from outside, show: Chromosome (navy), all annotated coding regions in the forward direction (green), annotated coding regions in the reverse direction (violet), non-coding annotations including ribosomal RNAs (red), known virulence factors including type 3 secretion components, as automatically annotated through PATRIC annotation platforms (gold), functionally annotated transporters (blue), relative GC content (lavender ring) and GC skew (yellow

ring), annotated membrane-associated proteins assigned through structural prediction and relatedness to reference Chlamydia and inclusion membrane proteins determined through comparative annotation and canonical hydrophobicity plots (cyan). B) Subsystems of functionally assigned genes excluding the 282 hypothetical genes. 122 assigned genes were also not categorized into one of these subsystems and were left out of the chart. The remaining genes were divided into 9 subsystems including Protein processing, Metabolism, Stress response/defense/virulence, DNA processing, Energy, RNA processing, Cellular processes, Cell envelope and Membrane transport.

Figure 2

A) Circular alignment of *C. suis* R19 coding regions (outer ring) with canonical Chlamydia. Protein sequence identity is indicated by colorimetric scale where purple/blue represents higher identity than orange/red. White or blank areas represent an absence of coding regions annotated at that site. There are four key regions with a notable absence of coding regions. The first of these (around 140-160kb) correlates with the ribosomal RNA operons. The second region (around 310kb) has diversity which can be attributed to a polymorphic membrane protein (pmp) operon where not all genes are conserved between Chlamydia. The third region (around 530kb), represents the plasticity zone which is where the greatest diversity is expected between species. The final region (around 850kb) is an additional pmp operon. This alignment was performed using the PATRIC proteome comparison tool. B) Phylogenetic analysis of Chlamydia using whole-genome sequences aligned through progressiveMauve demonstrates the positioning of representative *C. suis* genomes compared with selected *C. trachomatis* and *C. muridarum* reference genomes.

Figure 4

Figure 4

A) Alignment of *C. suis* Momp primary sequences with differences are marked in black. Regions of the greatest variation appear are labelled as variable domains (VD). Mean hydrophobicity is plotted in red and blue along the top. B) Phylogenetic tree from A), with the addition of select *C. trachomatis* Momp sequences (blue). R19 is indicated in red. Scale bar is in substitutions per site.

Figure 5

A) Venn diagram displaying the shared or specific inclusion membrane proteins (Inc) between *C. trachomatis* D/UW-3 (blue), *C. suis* R19 (red) and *C. muridarum* Nigg (green). B) Phylogenetic analysis of putative Inc protein sequences. C) Highlighted in boxes 1-8 are the Incs that are not conserved between *C. trachomatis* (blue), *C. muridarum* (green) and *C. suis* (red). Tree is transformed with ordered branching.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supfig1.jpg](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable1.xlsx](#)