Chlamydia psittaci comparative genomics reveals intraspecies variations in the putative outer membrane and type III secretion system genes

Bernard J. Wolff, 1,2 Shatavia S. Morrison, 2 Denise Pesti, 1 Satishkumar Ranganathan Ganakammal, 3 Ganesh Srinivasamoorthy, 3 Shankar Changayil, 3 M. Ryan Weil, 3 Duncan MacCannell, 3 Lori Rowe, 3 Michael Frace, 3 Branson W. Ritchie, 1 Deborah Dean 4,5 and Jonas M. Winchell 2

1Infectious Diseases Laboratory, College of Veterinary Medicine, University of Georgia, Athens, GA, USA
2Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA
3The National Center for Emerging and Zoonotic Infectious Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA
4Children’s Hospital Oakland Research Institute, Oakland, CA
5UCSF and UC Berkeley Joint Graduate Program in Bioengineering, Oakland, CA

Chlamydia psittaci is an obligate intracellular bacterium that can cause significant disease among a broad range of hosts. In humans, this organism may cause psittacosis, a respiratory disease that can spread to involve multiple organs, and in rare untreated cases may be fatal. There are ten known genotypes based on sequencing the major outer-membrane protein gene, ompA, of C. psittaci. Each genotype has overlapping host preferences and virulence characteristics. Recent studies have compared C. psittaci among other members of the Chlamydiaceae family and showed that this species frequently switches hosts and has undergone multiple genomic rearrangements. In this study, we sequenced five genomes of C. psittaci strains representing four genotypes, A, B, D and E. Due to the known association of the type III secretion system (T3SS) and polymorphic outer-membrane proteins (Pmps) with host tropism and virulence potential, we performed a comparative analysis of these elements among these five strains along with a representative genome from each of the remaining six genotypes previously sequenced. We found significant genetic variation in the Pmps and T3SS genes that may partially explain differences noted in C. psittaci host infection and disease.

INTRODUCTION

The Chlamydiaceae family of bacteria comprises nine distinct species, namely Chlamydia trachomatis, C. suis, C. muridarum, C. psittaci, C. pneumoniae, C. abortus, C. felis, C. pecorum and C. caviae (Read et al., 2013). This diverse group of obligate intracellular Gram-negative bacteria have adapted to their host cell niche by evolving to use a complex biphasic life cycle, which consists of a metabolically inactive, infectious form known as an elementary body (EB) and a metabolically active and non-infectious reticulate body (RB) (Abdelrahman & Belland, 2005). Chlamydiaceae infect a broad range of hosts, and evidence of infection has been found in nearly every phylogenetic group of animals (Kaleta & Taday, 2003; Read et al., 2003).

C. psittaci is primarily a zoonotic pathogen that is normally transmitted through close contact with infected birds and some mammals including cattle, pigs, sheep, swine, goats, cats and horses, not to mention feral animals (Hotzel et al., 2004; Read et al., 2003). The bacteria are often found in secretions and faecal droppings, remaining infectious for up to 30 days (Beeckman & Vanrompay, 2009; Haag-Wackernagel & Moeh, 2004; Harkinezhad et al., 2009; Heddeam et al., 2006a). Infection in humans may

| Abbreviations: | EB, elementary body; Pmp, polymorphic membrane protein; RB, reticulate body; T3SS, type III secretion system |
|---------------|--------------------------------------------------------------------------------------------------|
| The GenBank/EMBL/DDBJ accession numbers for the Chlamydia psittaci genome sequences generated are provided in Table 1. |
| Three supplementary figures and one supplementary table are available with the online Supplementary Material. |
lead to psittacosis, a severe respiratory illness often associated with multiorgan involvement causing significant morbidity and mortality (Smith et al., 2011). *C. psittaci* is also a major cause of economic loss in the poultry industry in the US and abroad, and poses a significant risk to farm workers as well as having potential for laboratory-acquired infection (Gaede et al., 2008; Miller et al., 1987; Smith et al., 2011). Because of the potential public health risk of *C. psittaci*, the National Association of State Public Health Veterinarians published a compendium on testing and strategies for managing the disease in birds and humans (Smith et al., 2011).

*C. psittaci* has the widest documented host range within the *Chlamydiaceae* family including avian, mammalian, reptilian and human hosts (Harkinezhad et al., 2009). *C. psittaci* is classified into ten genotypes, designated A–G, E/B, M56 and WC (Read et al., 2013; Van Lent et al., 2012). Recently, a real-time PCR assay targeting differences in the *ompA* gene was developed to differentiate the genotypes (Mitchell et al., 2009). Biological differences in host preference and virulence can be noted between the different *C. psittaci* genotypes. For instance, genotype A is endemic in psittacine birds and is theorized to be a common cause of respiratory disease and/or flu-like symptoms in exposed humans, while genotypes C and D have primarily been associated with waterfowl and poultry, respectively (Heddema et al., 2006a; c; Smith et al., 2011). Genotype E has been shown to infect a diverse group of avian species, including pittacines, pigeons, waterfowl and turkeys, and was first described to infect humans in the late 1920s following exposure to diseased parrots (Harkinezhad et al., 2007). Although all genotypes may infect humans, including genotypes from both avian and mammalian species, genotype A has been referenced as the most common cause of human disease. There is substantial variation in the likelihood of infection and disease following exposure to infected birds, even within what is currently classified as the same genotype (Heddema et al., 2006b, c; Wreghitt & Taylor, 1988).

Genetic manipulation techniques cannot reliably be used on *Chlamydiaceae* due to their intracellular requirement for replication. As a result, research on the specific interactions and contributions of genes to virulence, infectivity and replication are limited. Comparisons between the various genotypes and between species are limited despite some recent advancements with better characterized species and the use of surrogate systems (Peters et al., 2007). While some members of the *Chlamydiaceae* family have been sequenced and studied for their host and tissue preferences and virulence mechanisms, only 16 have been completely characterized to date (Kalman et al., 1999; Read et al., 2000, 2003, 2013; Stephens et al., 1998; Thomson et al., 2005; Van Lent et al., 2012; Voigt et al., 2012). Next-generation sequencing bypasses some of the limitations of bench work by allowing more meaningful investigation of the genotypes at the genomic level and their potential association with host specificity and virulence.

Here, we report on the comparative genomics of all ten genotypes of *C. psittaci* of which genotypes A, B, D and E were sequenced in this study. Our research focused on the genes encoding polymorphic membrane proteins (Pmps) and the type III secretion system (T3SS) proteins because of their known involvement in pathogenicity (Voigt et al., 2012). The Pmps are a large family of proteins unique to *Chlamydiaceae*, which are highly variable in numbers and homology among the family, and are thought to be involved in niche adaptation based on adherence to the host cell, molecular transport and cell wall associated functions (Rockey et al., 2000). The T3SS transports effector proteins into the host cytoplasm using a needle-like apparatus, similar to other Gram-negative bacteria (Hueck, 1998). While the structural genes are well conserved, the secreted effector proteins, although difficult to identify, are widely diverse and have many unique functions (Valdivia, 2008). Some secreted effectors have been identified and fairly well characterized. Specifically, the *tarp* gene in *C. trachomatis* was well characterized by Somboonna et al. (2011) as a virulence factor and has been associated with actin recruitment and inclusion development. We found numerous insertions, deletions and single nucleotide polymorphisms among the *pmp* and T3SS genes that may account for host preferences and virulence characteristics for the *C. psittaci* genotypes.

**METHODS**

**C. psittaci strains.** *C. psittaci* strains DD-34 (ATCC VR-854, genotype A), CP3 (ATCC VR-574, genotype B), NJ1 (genotype D), Frances (ATCC VR-122, genotype E) and a genotype A strain (UGA) recovered from a cockatiel, were sequenced in this study. The DD-34 strain was originally isolated from a parrot in 1949 (Davis, 1949). The CP3 strain was isolated from a pigeon in 1958 (Page, 1966). The Frances strain was isolated from a ferret inoculated with human material in 1934 (Francis & Magill, 1938). *C. psittaci* CP3 and NJ1 genomic sequences are also available elsewhere (Van Lent et al., 2012), while Frances, DD-34 and UGA are newly described in the current study. This study represents comprehensive analysis of these genomes using the whole genome assembly and annotation methods describe below.

**C. psittaci culture.** *C. psittaci* culture was performed as previously described (Mitchell et al., 2009). Briefly, *C. psittaci* reference strains were propagated in Vero cell monolayers grown in 150 cm$^2$ culture flasks in Eagle’s minimum essential medium (MEM) supplemented with MEM nonessential amino acids, 2 μM l-glutamine, 20 μM HEPES buffer, 10 % FCS, 20 μg streptomycin ml$^{-1}$ and 25 μg vancomycin ml$^{-1}$. Confluent cell monolayers were inoculated by replacing the growth medium with 5 ml of stock *C. psittaci* culture diluted 1:10 in MEM containing 1 μg cycloheximide ml$^{-1}$. The inoculated monolayers were placed at 37°C and 5 % CO$_2$ for 2 h before an additional 50 ml of MEM containing cycloheximide was added to each flask. Cultures were incubated for 7 days at 37°C or until the monolayers demonstrated nearly 100 % cytopathic effects. The cell culture was transferred to 50 ml tubes and stored at −80°C prior to EB isolation.

EB isolation was performed by density-gradient centrifugation as previously described (Mukhopadhyay et al., 2004). DNA was extracted from the pellet using a QiAamp DNA minikit (Qiagen).
Whole genome sequencing and assembly. The genomic DNA for *C. psittaci* strains DD-34, CP3, NJ1, Frances and UGA was prepared for whole genome paired-end sequencing on an Illumina GAIIx DNA sequencer using standard protocols and reagents from Illumina. Approximately 1 μg of genomic DNA was sheared using a Covaris S2 sonicator (Covaris) to a mean size of 350 bp. DNA sequencing libraries were then prepared using Illumina TrueSeq chemistry and size selected using double Ampure (Beckman Coulter) selection. Paired-end flowcells underwent cluster formation using an Illumina cBot, followed by 100 × 100 bp cycle sequencing using SBS cycle sequencing V5 kits. Sequence data were processed using CASAVA (v1.8.2) into paired FASTQ read sets. Read quality checks were performed using a combination of publicly available tools and in-house scripts. *C. psittaci* 6BC (NC_015740) was used as reference genome for all analysis. Trimming of reads based on quality, mapping of reads to a reference genome, and *de novo* assembly were performed using CLC Genomics Workbench 5.5.1. Since CLC Genomics 5.5.1 was no longer supported on the computing environment, CLC Genomics Workbench 5.5.1 was no longer supported on the computing environment, CLC Genomics Workbench 7.0.4 was used to assembly the RTX sequence reads.

Whole genome annotation. All five genomes were submitted to NCBI and annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). The GenBank accession numbers for each genome are listed in Table 1.

Whole genome comparison. The *de novo* assembled contigs for each sample were ordered against the 6BC reference genome using abacas tool (Assefa et al., 2009). The ordered samples’ superscaffolded genomes and reference genome were compared using Mauve to identify conserved and rearranged regions (Darling et al., 2010). The above input was used for Blast Ring Image Generator (BRIG) analysis to identify differences (Alikhan et al., 2011). Using the whole genome SNPs, identified by the kSNP version 2.1.2 (Gardner & Hall, 2013) software application, a maximum-likelihood phylogenetic tree was reconstructed with the RAxML version 7.3.0-PTHREAD (Stamatakis, 2014). The tree was visualized with Figtree (Andrew Rambaut, 2015). The phylogenetic tree represents the evolutionary relationship of *C. psittaci* genotypes with *C. abortus* S26/3 as an outgroup.

### Table 1. Whole genome sequencing and annotation results

| Characteristic               | DD34    | UGA     | CP3     | NJ1     | Frances |
|-----------------------------|---------|---------|---------|---------|---------|
| Genotype                    | A       | A       | B       | D       | E       |
| GenBank accession no.       | AFVL000000000 | AWXQ000000000 | AFVN000000000 | AFVK000000000 | AFVM000000000 |
| Total assembled size (bp)   | 1 163 748 | 1 164 948 | 1 163 075 | 1 160 660 | 1 162 120 |
| Plasmid size (bp)           | 7553    | 7553    | 7553    | 7532    | 7545    |
| Total reads                 | 13 181 088 | 12 562 928 | 12 695 762 | 7 803 252 | 13 291 438 |
| Total bases                 | 1 317 642 394 | 1 256 351 619 | 1 268 662 067 | 780 739 971 | 1 327 907 989 |
| Mean read length            | 101     | 101     | 101     | 101     | 101     |
| Mean coverage               | 576.2   | 548     | 556.1   | 340.4   | 581.1   |
| Number of contigs           | 4       | 6       | 4       | 3       | 2       |
| N50                         | 778 174 | 779 321 | 778 149 | 775 399 | 1 159 687 |
| Genes predicted             | 1057    | 1059    | 1054    | 1057    | 1044    |
| tRNAs predicted             | 38      | 38      | 38      | 38      | 38      |
| DNA G + C content (mol%)    | 39.03   | 39.02   | 39.01   | 38.95   | 39.03   |

Annotated *pmp* and T3SS genes described by Voigt et al. (2012) from the 6BC genome (CP002549) were compared by BLAST analysis to the genomes sequenced in this study and six reference genomes available in GenBank [WC (NC_018624.1), M56 (NC_018623.1), VS225 (NC_018621.1), WS-RT-E3 (NC_018622.1), GR9 (NC_018620.1), RTH (SRA061571)] (Read et al., 2013; Van Lent et al., 2012; Voigt et al., 2012). A formatted report was generated using an in-house gene search reporter pipeline (blast-based). The gene sequences that reported a pairwise identity below 75 % were marked as low quality and were considered absent.

A heatmap was constructed using R version 3.0.1 using the ‘gplots’ and ‘RColorBrewer’ packages (Neuvirth, 2007; Warnes et al., 2014). The *pmp* genes pairwise identities between all *C. psittaci* genomes included this study against *C. psittaci* 6BC were shown.

A maximum-likelihood tree was reconstructed with the *pmp* genes to show a similar evolutionary relationship in Voigt et al. (2012). We selected genes based on a protein pairwise identity to 6BC that met the following criteria: 75 % sequence identity and above 60 % gene coverage. ClustalOmega was used to perform independent multiple sequence alignments before sequence analysis to minimize gene rearrangement (Sievers et al., 2011). Once each individual protein alignment was built, the independent alignments were concatenated. RAxML was used to generate a phylogenetic tree with 1000 replicates for bootstrapping (Stamatakis, 2014). The tree was visualized with Figtree (Andrew Rambaut, 2015).

Protein domain identification. The gene sequences for CPSIT_0757 (dihydrolipicolinate reductase) and CPSIT_0192 (putative TARP) for 6BC genome (CP002549) were used as reference sequences to identify the orthologous gene sequences in the genomes included in this study. Three of these genes were selected based upon their critical role in virulence or metabolism and genetic variability between strains (Voigt et al., 2012). Orthologous gene sequences were identified with a shared sequence identity greater than 75 %. Genomes GR9 (NC_018620.1), M56 (NC_018623.1) and WS-RT-E30 (NC_018622.1) did not contain genes that met the criteria, thus excluding them from the protein domain identification analysis. The amino acid sequences for each gene were submitted to the European Bioinformatics Institute InterPro web service (Jones et al., 2014). InterPro provides a functional analysis of protein sequences by predicting protein domains based on domain signatures found other protein family and domain databases (Jones et al., 2014). A detailed
description of the databases that make up InterProScan is given by Hunter et al. (2012).

RESULTS

Whole genome sequencing, assembly and annotation

Table 1 summarizes the sequencing statistics and genomic characteristics for each newly sequenced genome included in this study. The sequencing resulted in assemblies covering 99% or greater of the entire genome for each strain. All five strains yielded an approximately 1.16 Mb chromosome and a fully sequenced 7.5 kb plasmid. The plasmids were remarkably conserved with a pairwise identity of 99%. The DNA G+C content for all five genomes was approximately 39%, which is consistent with other sequenced Chlamydia genomes (Reference 6BC). The genome coverage ranged from 340.4X (NJ1) to 581.1X (Frances), with a mean read length for all five genomes of 101 bp, and total reads generated were between 7,803,252 (NJ1) and 13,291,438 (Frances). The N50 values were between 775,399 (NJ1) and 1,159,687 (Frances). The numbers of genes predicted for each genome were similar. The lowest number of genes (1,044) was predicted in the Frances genome compared to a high of 1,059 genes in UGA. A total of 38 tRNAs were found in all five genomes.

Pan genome comparison

The 6BC strain was chosen as the reference with which to compare all other strains used in this study, because it was the only completely sequenced genome available at the start of this study. Fig. 1 is a circular map representing the nucleic acid sequence similarity of the five strains sequenced for this study and the genomes sequenced by Read et al. (2013) compared with the 6BC reference genome generated using the BRIG. The comparison shows a high degree of similarity (≥99%) across all genotypes of C. psittaci. As expected, nearly identical sequence similarity was observed between the three genotype A genomes (6BC, UGA and DD34). While all the genomes appear to be very closely related, two regions demonstrated significant sequence divergence among the different genotypes. These regions are highlighted, and the genes encoded in each region are listed (Table S1, available in the online Supplementary Material). Region 1 contains approximately 20 kb; region 2 spans 17 kb. The majority of the genes contained in these regions encode Pmps belonging to the pmpG group. Several of the genomes sequenced by Read et al. (2013) had large deletions when compared with the reference 6BC.

A whole genome SNP tree rooted to C. abortus strain S26/3 was reconstructed using previously reported C. psittaci genomes currently in the NCBI database and the genomes sequenced for this project (Fig. 2). The tree demonstrates the closely related nature of the C. psittaci genotypes. Nine genotype A strains are represented in the tree and formed the first clade. Genotype F was the next closest relative to genotype A, followed by the WC clade. A fourth clade composed of genotypes C and WS/RT/E30 was present, followed by genotypes B and E representing the fifth clade. Genotypes D, M56 and RTH formed their own separate branches, with strain RTH, genotype G, the most distantly related.

T3SS and effector genes

Table 2 is a summary of the pairwise identity values (≥75%) of the T3SS apparatus, chaperone and effector genes (n=39) among all C. psittaci genotypes using the 6BC genome from Voigt et al. (2012) as the reference strain. There is a large degree of conservation amongst the genes comprising the T3SS. However, the effector genes show greater sequence divergence compared with the apparatus genes (Fig. 3a, b). Strain RTH (genotype G) was the most divergent, with only two genes sharing 100% pairwise identity with 6BC. A large number of genes (16 out of 39) shared 100% pairwise identity compared to 6BC for nine of the ten genotypes, with the RTH (genotype G) being the only divergent strain. However, substantial differences do exist among many of the genes. Two genes (CPSIT_0844 and CPSIT_0846) were absent in three or four strains (Table 2, yellow highlight). CPSIT_0844 and CPSIT_0846 encode IncA family proteins and were absent in GR9 (genotype C), RTH (genotype G) and WS/RT/E30 (genotype E/B); CPSIT_0846 was also absent in Frances (genotype E). A putative inner-membrane protein (CPSIT_0463) was absent in RTH and WC (genotypes G and WC, respectively).

The effector protein dihydrodipicolinate reductase (CPSIT_0757) was conserved among DD34 and UGA (genotypes A), VS225 (genotype F) and RTH (genotype G) but displayed differences in the other seven genotypes (Table 2, blue highlight). While the InterPro analysis revealed two domains for a NAD(P) binding domain present in all genotypes, a multiple sequence alignment of those domains showed that DD34, UGA, VS 225 and RTH clustered together (Fig. S1, magenta highlight) and WS/RT/E30, GR9, WC, Frances, CP3, M56 and NJ1 formed a second cluster (Fig. S1, underlined). In the first domain, eight of ten substitutions were clustered in this fashion, while the second domain showed nine of thirteen substitutions sharing this pattern. Other sporadic changes within the catalytic domains were also observed requiring further study. M56 had a single substitution compared to the other strains in the first domain (Fig. S1, blue highlight), and NJ1 had a unique substitution in the C-terminal domain (Fig. S1, red highlight).

Other effector genes such as the tarp gene (CPSIT_0192) were 100% conserved among DD34, UGA and NJ1 (genotypes A and D, respectively), but there were differences in the other strains, as the pairwise identity was 92% or greater in each. To identify functional regions of the tarp
Fig. 1. BRIG analysis. A BRIG comparing the *C. psittaci* genomes sequenced in this study and *C. psittaci* genomes sequenced by Read et al. (2013). Genotype 6BC was used as a reference for comparison. The innermost ring represents the base position along the map. The second ring indicates the GC content along the length of the genome. The colour rings correspond to the genomes with each genome indicated by a unique colour as indicated in the figure legend. A change in colour corresponds to a decrease in pairwise identity compared to the reference genome. Four regions with sequence divergence are noted in red circles with gene identities listed in orange boxes.
gene, a nucleotide and amino acid sequence alignment (Fig. S2, S3) were performed using the genomes sequenced in this study, those sequenced by Read et al. (2013), and C. trachomatis L2 and L3b strains (accession numbers AM884176 and AM884177) characterized by Somboonna et al. (2011). The tarp genes in the C. psittaci genomes had very little sequence similarity with the two C. trachomatis genomes (Fig. S2). The C. psittaci amino acid sequences of Tarp were examined in InterPro and three catalytic domains of unknown function were returned that were present in all strains. A multiple sequence alignment of those domains revealed five amino acid differences in each of the three domains (Fig. S3, green highlight). Of these fifteen substitutions, four were specific to the RTH strain, two were only present in the M56 strain, eight were shared between the M56 and RTH strains and one amino acid substitution was only present in the NJ1 strain (Fig. S3, green highlight).

The IncA (CPSIT_0594) and IncB (CPSIT_0532) were largely conserved. The incA gene was 100 % conserved in every genome with the exception of RTH and M56, whereby the pairwise identity was 77 % and > 95 % respectively. The incB gene was 100 % conserved in 10 genomes compared to 6BC, with RTH being the only exception with a pairwise identity of 83 %.

Pmp genes

The 30 Pmps characterized by Voigt et al. (2012) were compared with the strains sequenced in this study as well as the other strains with whole genome data now available (Table 3). The BRIG analysis (Fig. 1, regions 1 and 2) revealed that the G group of proteins had the highest degree of divergence among the different genotypes of C. psittaci. Fig. 4 shows the heatmap of the Pmps showing that the pmpG group is clearly the most diverged. Of the 14 genes of this group, only seven were present in all 12 genomes. Also notable is that, among the three genotype A strains (6BC, DD34 and UGA), three genes in the G family were present only in the 6BC strain. Two genes (CPSIT_0310 and CPSIT_0311) were only present in 6BC, CP3, M56 and WC, and a weak match was also returned for CPSIT_0310 in NJ1. Strain 6BC and WC had the largest conservation of this group of genes. We found 13 of the 14 pmpG genes in the 6BC had significant pairwise identity (> 85 %) with those in WC.

The other Pmp gene groups had significantly higher conservation among the genotypes. Genes of the pmpA, pmpB/C, pmpD and pmpH groups were present in all 12 genomes, with a small degree of sequence divergence present among some of the genes. The most divergent gene was CPSIT_0231 of the pmpB/C family. Its corresponding gene in the M56 genome had a 90 % pairwise identity and an 89 % identity in the RTH genome. All the other pairwise identities for those families were 88 % or greater among all 12 genomes. The pmpE/F genes were the only ones that were absent in any of the genomes. Both genes were not present in the RTH and M56 genomes. A heat map demonstrates the variability of the pmpG group of genes among the strains compared to the other Pmp groups (Fig. 4).

In reconstructing a phylogenetic tree of the Pmps for each strain with Pmps, the phylogeny is similar to the whole-genome phylogeny of Fig. 2 (Fig. 5).
### Table 2. T3SS genes

| 6BC locus tag (CP002549) | Name | Strain Genotype | DD34 | UGA | CP3 | GR9 | NJI | Frances | VS225 | RTH | WS/RT/E30 | M56 | MS6 | WC |
|---------------------------|------|-----------------|------|-----|-----|-----|-----|---------|-------|-----|----------|-----|------|----|
| **Apparatus genes**       |      |                 |      |     |     |     |     |         |        |     |          |     |      |    |
| CPSIT_0074                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 88     | >95  | >95 | 100      |     |      |    |
| CPSIT_0245                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | >95  | >95 | 100      |     |      |    |
| CPSIT_0313                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | >95  | >95 | 100      |     |      |    |
| CPSIT_0357                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 96     | >95  | >95 | 100      |     |      |    |
| CPSIT_0421                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 88     | >95  | >95 | 100      |     |      |    |
| CPSIT_0429                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | –      |      |     |          |     |      |    |
| CPSIT_0431                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 93     | 100  | 100 | 100      |     |      |    |
| CPSIT_0490                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 94     | 100  | 100 | 100      |     |      |    |
| CPSIT_0594                | IncA  | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 77     | >95  | >95 | 100      |     |      |    |
| CPSIT_0602                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 91     | >95  | >95 | 100      |     |      |    |
| CPSIT_0656                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 90     | 100  | 100 | 100      |     |      |    |
| CPSIT_0749                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | >80    | >80  | >80 | 100      |     |      |    |
| CPSIT_0767                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | >95    | >95  | >95 | 100      |     |      |    |
| CPSIT_0785                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 93     | 100  | 100 | 100      |     |      |    |
| CPSIT_0828                | DNA recombination protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | 100  | 100 | 100      |     |      |    |
| CPSIT_0844                | IncA family protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | 100  | 100 | 100      |     |      |    |
| CPSIT_0930                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 77     | >95  | >95 | 100      |     |      |    |
| CPSIT_0933                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | 100  | 100 | 100      |     |      |    |
| CPSIT_0997                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | 100  | 100 | 100      |     |      |    |
| CPSIT_1054                | 5-formyltetrahydrofolate-cyclo-ligase | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 92     | 100  | 100 | 100      |     |      |    |
| **Effector genes**        |      |                 |      |     |     |     |     |         |        |     |          |     |      |    |
| CPSIT_0192                | tarp  | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 92     | >95  | >95 | >95      |     |      |    |
| CPSIT_0220                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | 100  | 100 | 100      |     |      |    |
| CPSIT_0314                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | >95  | >95 | >95      |     |      |    |
| CPSIT_0422                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 84     | 100  | 100 | >95      |     |      |    |
| CPSIT_0461                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 91     | 100  | 99  | 100      |     |      |    |
| CPSIT_0463                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 92     | 100  | 100 | 100      |     |      |    |
| CPSIT_0532                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 83     | 100  | 100 | 100      |     |      |    |
| CPSIT_0555                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 76     | 100  | 100 | 100      |     |      |    |
| CPSIT_1042                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 86     | 100  | 100 | 100      |     |      |    |
| CPSIT_0856                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 94     | >95  | >95 | 100      |     |      |    |
| CPSIT_0962                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 97     | 100  | 100 | 100      |     |      |    |
| CPSIT_0974                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 96     | 100  | 100 | 100      |     |      |    |
| CPSIT_1042                | Hypothetical protein | 100 | 100 | 100 | >95 | 100 | 100 | 100 | 95     | 100  | 100 | 100      |     |      |    |

Locus tags highlighted in yellow are absent in one or more strains, and those highlighted in blue are discussed in the manuscript. –, The gene was not present.
Fig. 3. Heatmap analysis of T3SS genes. The nucleotide pairwise identities for (a) the apparatus genes and (b) the effector genes are shown for the strains in this study compared with 6BC T3SS reference genes. Pairwise identities of ≥75 % are shown. Pairwise identifies are coded with the following colour schema: red: 0–59 %, orange: 60–69 %, yellow: 70–79 %, green: 80–89 % and blue: 90–100 %.
### Table 3. Polymorphic outer-membrane protein genes

| PMP family | 6 BC locus tag (CP002549) | Strain: | DD34 | UGA | CP3 | GR9 | NJ1 | Frances | VS225 | RTH | WS/RT/E30 | M56 | WC |
|------------|--------------------------|---------|------|-----|-----|-----|-----|--------|-------|-----|-----------|-----|----|
| A          | CPSIT_0232               | 100     | 100  | 100 | 100 | 100 | 100 | 100    | 95    | 100 | >95       | 100 |    |
| B/C        | CPSIT_0231               | 100     | 100  | 100 | >95 | >95 | 100 | 100    | 94    | >95 | >90       | 100 |    |
| D          | CPSIT_0856               | 100     | 100  | 100 | >95 | 100 | 100 | 100    | 84    | >95 | >90       | 100 |    |
| E/F        | CPSIT_0297               | 100     | 100  | >95 | >95 | >95 | >95 | 100    | –     | >95 | –         | 100 |    |
| G          | CPSIT_0302               | 100     | 100  | >95 | 100 | >75 | >95 | >95    | 86    | 100 | >84       | >95 |    |
|            | CPSIT_0304               | >95     | >95  | >94 | >95 | 81  | >95 | >95    | 78    | >95 | >79       | >95 |    |
|            | CPSIT_0305               | 100     | 100  | >95 | 100 | 100 | 100 | >95    | 93    | >95 | >90       | 100 |    |
|            | CPSIT_0306               | 100     | 100  | >95 | 100 | 100 | 100 | >95    | 93    | >95 | 100       | 100 |    |
|            | CPSIT_0307               | 100     | 100  | >95 | 100 | 100 | 100 | >95    | 93    | 100 | 100       | 100 |    |
|            | CPSIT_0308               | >75     | >96  | –   | >85 | –   | >95 | >95    | 91    | –   | >85       | >95 |    |
|            | CPSIT_0310               | –       | –    | 80  | –   | >76 | –   | –      | –     | –   | –         | >79 | >95 |
|            | CPSIT_0311               | –       | –    | >87 | –   | –   | –   | –      | –     | –   | –         | 75  | >95 |
|            | CPSIT_0312               | >76     | >83  | >76 | >95 | –   | >84 | >95    | 89    | –   | >90       | >85 |    |
|            | CPSIT_0313               | 100     | 100  | >95 | 100 | 100 | 100 | 89     | 100   | 100 | 100       | 100 |    |
|            | CPSIT_0314               | 100     | 100  | >95 | >95 | >95 | >95 | >95    | 89    | >95 | >90       | >95 | >95 |
|            | CPSIT_0666               | –       | –    | >77 | 100 | –   | >95 | –      | >95   | –   | –         | –   | >95 |
|            | CPSIT_0667               | >78     | >78  | –   | –   | 84  | >95 | >95    | >95   | –   | >85       | >95 |    |
|            | CPSIT_0668               | 75      | >81  | >95 | 80  | >95 | 86  | >95    | 86    | >95 | >85       | >95 |    |
| H          | CPSIT_0301               | 100     | 100  | >95 | 100 | >95 | >95 | 88     | >95   | >95 | 100       | >95 |    |
| Other      | CPSIT_0057               | 100     | 100  | 100 | –   | –   | 100 | 100    | 95    | 100 | 100       | –   |    |
|            | CPSIT_0207               | 100     | 100  | 100 | 100 | 100 | 100 | 100    | 97    | 100 | >95       | 100 |    |
|            | CPSIT_0300               | 100     | 100  | 100 | 100 | 100 | 100 | 94     | 100   | 100 | >95       | 100 |    |
|            | CPSIT_0329               | 100     | 100  | >95 | >95 | >95 | >95 | >95    | 95    | >95 | >90       | >95 | >95 |
|            | CPSIT_0330               | 100     | 100  | >95 | >95 | >95 | >95 | >95    | 95    | >95 | >90       | >95 | >95 |
|            | CPSIT_0345               | 100     | 100  | 100 | 100 | 100 | 100 | 100    | 95    | 100 | >90       | 100 |    |
|            | CPSIT_0523               | 100     | 100  | 100 | 100 | 100 | 100 | 95     | 100   | 100 | 100       | 100 |    |
|            | CPSIT_0967               | 100     | 100  | 100 | 100 | 100 | 100 | 100    | 95    | 100 | 100       | 100 |    |
|            | CPSIT_1035               | 100     | 100  | 100 | 100 | 100 | 100 | 100    | 95    | 100 | 100       | 100 |    |

"-" indicates the gene was not present.
**DISCUSSION**

*Chlamydiaceae* has one of the largest host ranges and varied virulence characteristics of any bacterial family currently known (Kaleta & Taday, 2003). Application of next-generation sequencing technologies has yielded novel information about *C. psittaci* genomes and provided insight into the differing virulence mechanisms, tissue tropism, evolutionary dynamics and host range of this species (Read et al., 2013; Voigt et al., 2012). Voigt et al. (2012) compared the genes of other *Chlamydia* species with those of *C. psittaci* strain 6BC and found a number of *psittaci*-specific Pmps of the G family, a high degree of genome rearrangement, and numerous differences in the T3SS. Read et al. (2013) demonstrated *C. psittaci* has undergone many recombination events and has the ability to switch hosts frequently. Our phylogenetic analysis revealed similar genotypic clade relationships seen by Read et al. (2013) and Voigt et al. (2012). The current study further enhances our understanding of this species by identifying a variety of genetic variations among the *pmp* and the T3SS genes that may, in part, be responsible for the variability in tissue tropism, host preferences, and virulence observed among genotypes.

---

**Fig. 4.** Heatmap analysis of Pmp genes. The nucleotide pairwise identities for all 30 *C. psittaci* Pmp genes are shown for the strains in this study compared with 6BC Pmp reference genes. Pairwise identities of ≥75 % are shown. Pairwise identities are coded with the following colour schema: red: 0–59 %, orange: 60–69 %, yellow: 70–79 %, green: 80–89 % and blue: 90–100 %.

http://mic.sgmjournals.org 1387
*Chlamydiaceae* contains a functional T3SS that serves as a major virulence factor by secreting effectors that recruit actin to facilitate entry into the cell (TARP), manipulate the host cell cytoplasm (CdsF), and alter the inclusion body (IncG and IncA) (Betts *et al.*, 2008; Fields *et al.*, 2003; Hsia *et al.*, 1997; Voigt *et al.*, 2012). Studies have identified numerous apparatus proteins along with multiple chaperone proteins and secreted effector molecules (Betts *et al.*, 2008; Herrmann *et al.*, 2006; Peters *et al.*, 2007; Stone *et al.*, 2008; Voigt *et al.*, 2012). The type III apparatus proteins are highly conserved among many bacterial species including *Shigella*, *Yersinia* and *Salmonella* (Hueck, 1998; Mota & Cornelis, 2005). This conservation is so extensive that small molecule inhibitors such as INP0400 designed for one species have efficacy against chlamydial species secretion systems, including IncG and IncA, and can inhibit the development of *Chlamydia* after entry into host cells (Muschiol *et al.*, 2009).

Intraspecies comparisons of the T3SS genes (Voigt *et al.*, 2012) showed that the majority of the genes were largely conserved in all 12 *C. psittaci* genomes with a pairwise identity of 95%. However, we found significant differences for some genes among the different genotypes of *C. psittaci*. For example, seven of the 39 genes were missing in one or more of the genotypes compared with 6BC (Table 2, yellow highlight). Of these seven, three are hypothetical proteins, two are associated with the IncA protein family, and two are putative membrane proteins. For the other 31 genes, there was little sequence divergence with pairwise identities ranging from 77% to 100%. However, all apparatus structure genes had at least a 95% pairwise identity except for the RTH strain (genotype G) where the identities were more variable.

Differences in the TARP, dihydrodipicolinate reductase and the adherence factor genes appear to impact virulence and host specific characteristics within the different *C. psittaci* genotypes. *tarp* was less conserved among the strains, which is consistent with a report showing that virulent strains of *C. psittaci* such as genotypes A and D recruit actin more efficiently than those associated with less virulence strains (Beeckman & Vanrompay, 2010). The increase in actin recruitment may provide a more efficient mechanism for EB attachment, invasion and inclusion formation. The amino acid sequences of the TARP suggest that this particular gene may be responsible for differences in the M56 (genotype M56) and RTH (genotype G) strains as the majority of the substitutions in the catalytic domains were associated with those two strains. This is not surprising as these two strains have widely different host preferences compared to the other genotypes. Further experiments are needed to better elucidate the functional properties of the *C. psittaci* tarp gene and protein among the different strains and genotypes, and to identify other possible genes involved in virulence.

Sequence divergence and amino acid changes in the dihydrodipicolinate reductase gene (Cpsit_0757) suggest a potential
difference in lysine biosynthesis for 6BC, DD34 and UGA (genotypes A), VS225 (genotype F) and RTH (genotype G) compared with the other C. psittaci genotypes. These substitutions could have important consequences for cell wall synthesis as lysine has been shown to be an important amino acid in the synthesis of peptidoglycan in other bacteria (Pavelka & Jacobs, 1996). This, combined with the recent report providing strong evidence for the presence of peptidoglycan in the cell wall of Chlamydia, warrants further investigation of this protein (Liechti et al., 2014).

The incA (CPSIT_0594) and incB (CPSIT_0532) genes were remarkably conserved among all the genotypes displaying 100 % pairwise identity, except for the RTH strain (genotype G) that had 77 % and 83 % pairwise identities, respectively. These genes were highly variable among the different species within the Chlamydiaceae family, but this divergence is not observed past the species level (Voigt et al., 2012). These data suggest these genes operate in the same manner among the majority of C. psittaci strains, but further in vitro studies will help to determine how their function compares to their orthologues in the RTH strain and other Chlamydia species.

The Pmps are another major source of diversity among the different species of the Chlamydiaceae family as shown for C. abortus and C. psittaci in prior studies (Thomson et al., 2005; Voigt et al., 2012). We compared the Pmps identified by Voigt et al. (2012) in addition to nine genes identified as membrane proteins from our annotation analysis. Among the different subsets of outer-membrane protein genes, the pmpG group has been demonstrated to be the most divergent. This family of proteins has previously been described as the most rapidly evolving group of proteins, exhibiting numerous deletion and duplication events among other Chlamydiaceae (Thomson et al., 2005). Voigt et al. (2012) characterized several new pmpG genes in C. psittaci that were not present in any other Chlamydia species. This study was able to conclusively determine that this diversity extends to the different genotypes of C. psittaci. The number of Pmps missing in some genomes, and the large sequence divergence observed among strains compared to 6BC, even within the same genotype A, suggests that these genes may be used to rapidly adapt to different environments (Fig. 4). While some sequence divergence was seen among other species, it is possible that the pmpG group plays a major role in tissue tropism and host preferences of the different strains of C. psittaci because of their diverse nature and ability to rapidly evolve (Read et al., 2013; Voigt et al., 2012). The Pmp tree (Fig. 5) is similar to the whole genome phylogeny suggesting that these genes may significantly contribute to strain and genotype diversity. A follow up study sequencing a number of different strains of the same genotype from diverse geographical regions and animal species would be helpful in determining if this family of genes is responsible for phenotypic differences. This will also allow a comprehensive evaluation to determine if variant genotypes are emerging due to the rapidly evolving nature of Pmps.

ACKNOWLEDGEMENTS

This work was supported in part by NIH grant ROI AI098843 to D. Dean, the UGARF Health Technology Development Fund and UGA Infectious Disease Laboratory, and by the Advanced Molecular Detection (AMD) initiative at the Centers for Disease Control and Prevention.

REFERENCES

Abdelrahman, Y. M. & Belland, R. J. (2005). The chlamydial developmental cycle. FEMS Microbiol Rev 29, 949–959.

Ali Khan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. (2011). BLAST Ring Image Generator (BIRG): simple prokaryote genome comparisons. BMC Genomics 12, 402–412.

Assefa, S., Keane, T. M., Otto, T. D., Newbold, C. & Berriman, M. (2009). ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25, 1968–1969.

Beeckman, D. S. A. & Vanrompay, D. C. G. (2009). Zoonotic Chlamydiaphila psittaci infections from a clinical perspective. Clin Microbiol Infect 15, 11–17.

Beeckman, D. S. A. & Vanrompay, D. C. G. (2010). Biology and intracellular pathogenesis of high or low virulent Chlamydiaphila psittaci strains in chicken macrophages. Vet Microbiol 141, 342–353.

Betts, H. J., Twiggs, L. E., Sal, M. S., Wyrick, P. B. & Fields, K. A. (2008). Bioinformatic and biochemical evidence for the identification of the type III secretion system needle protein of Chlamydia trachomatis. J Bacteriol 190, 1680–1690.

Darling, A. E., Mau, B. & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5, e11147.

Davis, D. J. (1949). The use of phenolized allantoic fluid antigen in the complement fixation test for psittacosis. J Immunol 62, 193–200.

Fields, K. A., Mead, D. J., Dooley, C. A. & Hackstadt, T. (2003). Chlamydia trachomatis type III secretion: evidence for a functional apparatus during early-cycle development. Mol Microbiol 48, 671–683.

Francis, T. & Magill, P. T. (1938). An unidentified virus producing acute meningitis and pneumonitis in experimental animals. J Exp Med 68, 147–160.

Gaede, W., Reckling, K. F., Drensekamp, B., Kenklies, S., Schubert, E., Noack, U., Irmscher, H. M., Ludwig, C., Hotzel, H. & Sachse, K. (2008). Chlamydiaphila psittaci infections in humans during an outbreak of psittacosis from poultry in Germany. Zoonoses Public Health 55, 184–188.

Gardner, S. N. & Hall, B. G. (2013). When whole-genome alignments just won’t work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. PLoS One 8, e87160.

Haag-Wackernagel, D. & Moch, H. (2004). Health hazards posed by feral pigeons. J Infect 48, 307–313.

Harkinezhad, T., Verminnen, K., Van Droogenbroeck, C. & Vanrompay, D. (2007). Chlamydiaphila psittaci genotype E/B transmission from African grey parrots to humans. J Med Microbiol 56, 1097–1100.

Harkinezhad, T., Geens, T. & Vanrompay, D. (2009). Chlamydiaphila psittaci infections in birds: a review with emphasis on zoonotic consequences. Vet Microbiol 135, 68–77.

Hedema, E. R., Ter Sluis, S., Buys, J. A., Vandenbroucke-Grauls, C. M. J. E., van Wijnen, J. H. & Visser, C. E. (2006a). Prevalence of
Chlamydia psittaci in fecal droppings from feral pigeons in Amsterdam, The Netherlands. Appl Environ Microbiol 72, 4423–4425.

Heddem, E. R., van Hannen, E. J., Duim, B., de Jongh, B. M., Kaan, J. A., van Kessel, R., Lumeij, J. T., Visser, C. E. & Vandenbroucke-Grauls, C. M. J. E. (2006b). An outbreak of psittacosis due to Chlamydia psittaci genotype A in a veterinary teaching hospital. J Med Microbiol 55, 1571–1575.

Heddem, E. R., van Hannen, E. J., Duim, B., Vandenbroucke-Grauls, C. M. J. E. & Pannekoek, Y. (2006c). Genotyping of Chlamyphila psittaci in human specimens. Emerg Infect Dis 12, 1989–1990.

Herrmann, M., Schuhmacher, A., Mühldorfer, I., Melchers, K., Prothmann, C. & Dammeler, S. (2006). Identification and characterization of secreted effector proteins of Chlamydiella pneumoniae TW183. Res Microbiol 157, 513–524.

Hotzel, H., Berndt, A., Melzer, F. & Sachse, K. (2004). Occurrence of Chlamydiaceae spp. in a wild boar (Sus scrofa L.) population in Thuringia (Germany). Vet Microbiol 103, 121–126.

Hsia, R. C., Pannekoek, Y., Ingerowski, E. & Bavoil, P. M. (1997). Type III secretion genes identify a putative virulence locus of Chlamydia. Mol Microbiol 25, 351–359.

Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev 62, 379–433.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., & other authors (2012). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40 (D1), D306–D312.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A. & other authors (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kalata, E. F. & Taday, E. M. A. (2003). Avian host range of Chlamydiella spp. based on isolation, antigen detection and serology. Avian Pathol 32, 435–461.

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. & Stephens, R. S. (1999). Comparative genomes of Chlamydia pneumoniae and C. trachomatis. Nat Genet 21, 385–389.

Liechti, G. W., Kuru, E., Hall, E., Kalinda, A., Brun, Y. V., VanNieuwenhze, M. & Maurelli, A. T. (2014). A new metabolic cell-wall labelling method reveals peptidoglycan in Chlamydiella trachomatis. Nature 506, 507–510.

Miller, C. D., Songer, J. R. & Sullivan, J. F. (1987). A twenty-five year review of laboratory-acquired human infections at the National Animal Disease Center. Am Ind Hyg Assoc J 48, 271–275.

Mitchell, S. L., Wolff, B. J., Thacker, W. L., Ciembor, P. G., Gregory, C. R., Everett, K. D. E., Ritchie, B. W. & Winchell, J. M. (2009). Genotyping of Chlamydia psittaci by real-time PCR and high-resolution melt analysis. J Clin Microbiol 47, 175–181.

Mota, L. J. & Cornelis, G. R. (2005). The bacterial infection kit: type III secretion systems. Ann Med 37, 234–249.

Mukhopadhyay, S., Clark, A. P., Sullivan, E. D., Miller, R. D. & Summersgill, J. T. (2004). Detailed protocol for purification of Chlamydia pneumoniae elementary bodies. J Clin Microbiol 42, 3288–3290.

Muschiol, S., Normark, S., Henriques-Normark, B. & Subtil, A. (2009). Small molecule inhibitors of the Yersinia type III secretion system impair the development of Chlamydia after entry into host cells. BMC Microbiol 9, 75.

Neuwirth, E. (2007) RColorBrewer: ColorBrewer palettes. R package version 1.0–2.

Page, L. A. (1966). Interspecies transfer of psittacosis-LGV-trachoma agents: pathogenicity of two avian and two mammalian strains for eight species of birds and mammals. Am J Vet Res 27, 397–407.

Pavelka, M. S. Jr & Jacobs, W. R. Jr (1996). Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of Mycobacterium smegmatis. J Bacteriol 178, 6496–6507.

Peters, J., Wilson, D. P., Myers, G., Timms, P. & Bavoil, P. M. (2007). Type III secretion to Chlamydia. Trends Microbiol 15, 241–251.

Read, T. D., Brunham, R. C., Shen, C., Gill, S. R., Heidelberg, J. F., White, O., Hickey, E. K., Peterson, J., Utterback, T. & other authors (2000). Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res 28, 1397–1406.

Read, T. D., Myers, G. S. A., Brunham, R. C., Nelson, W. C., Paulsen, I. T., Heidelberg, J., Holtzapple, E., Khouri, H., Federova, N. B. & other authors (2003). Genome sequence of Chlamydiella caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. Nucleic Acids Res 31, 2134–2147.

Read, T. D., Joseph, S. J., Didelot, X., Liang, B., Patel, L. & Dean, D. (2013). Comparative analysis of Chlamydia psittaci genomes reveals the recent emergence of a pathogenic lineage with a broad host range. MBio 4, e00604–e00612.

Rockey, D. D., Lenart, J. & Stephens, R. S. (2000). Genome sequencing and our understanding of Chlamydiaceae. Infect Immun 68, 5473–5479.

Rambaut, A. (2015) FigTree. 1.4.2 edn. http://tree.bio.ed.ac.uk/software/figtree/

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & other authors (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using CLUSTAL Omega. Mol Syst Biol 7, 539.

Smith, K. A., Campbell, C. T., Murphy, J., Stobierski, M. G. & Tengelsen, L. A. (2011). Compendium of Measures to Control Chlamydia psittaci Infection Among Humans (Psittacosis) and Pet Birds (Avian Chlamydioidosis), 2010 National Association of State Public Health Veterinarians (NASPHV). J Exot Pet Med 20, 32–45.

Somboonna, N., Wan, R., Ojcius, D. M., Pettengill, M. A., Joseph, S. J., Chang, A., Hsu, R., Read, T. D. & Dean, D. (2011). Hypervirulent Chlamydia trachomatis clinical strain is a recombinant between lymphogranuloma venereum (L2(2)) and D lineages. MBio 2, e00045–e11.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Arvind, L., Mitchell, W., Olinger, L., Tatusov, R. L. & other authors (1998). Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science 282, 754–759.

Stone, C. B., Johnson, D. L., Bulir, D. C., Gilchrist, J. D. & Mahony, J. B. (2008). Characterization of the putative type III secretion ATPase CdsN (Cput0707) of Chlamydophila pneumoniae. J Bacteriol 190, 6580–6588.

Thomson, N. R., Yeats, C., Bell, K., Holden, M. T. G., Bentley, S. D., Livingstone, M., Cerden˜o-Tarraga, A. M., Harris, B., Doggett, J. & other authors (2005). The Chlamydia abortus abortus genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Res 15, 629–640.

Valdivia, R. H. (2008). Chlamydia effector proteins and new insights into chlamydial cellular microbiology. Curr Opin Microbiol 11, 53–59.
Van Lent, S., Piet, J. R., Beeckman, D., van der Ende, A., Van Nieuwerburgh, F., Bavoil, P., Myers, G., Vanrompay, D. & Pannekoek, Y. (2012). Full genome sequences of all nine Chlamydia psittaci genotype reference strains. J Bacteriol 194, 6930–6931.

Voigt, A., Schöfl, G. & Saluz, H. P. (2012). The Chlamydia psittaci genome: a comparative analysis of intracellular pathogens. PLoS One 7, e35097.

Warnes, G. R., Bolker, B. & Lumley, T. (2014) Gplots: Various R programming tools for plotting data. R package version 2.6.0.

Wreghitt, T. G. & Taylor, C. E. D. (1988). Incidence of respiratory tract chlamydial infections and importation of psittacine birds. Lancet 331, 582.

Edited by: R. Lan