University of the Sunshine Coast
Queensland, Australia

USC RESEARCH BANK

http://research.usc.edu.au

This is the author accepted version of the following publication:

Myers, G S A, Mathews, S A, Eppinger, M, Mitchell, C M, O'Brien, K K, White, O R, Benahmed, F, Brunham, R C, Read, T D, Ravel, J, Bavoil, P M, Timms, Peter (2009) Evidence that human Chlamydia pneumoniae was zoonotically acquired, Journal of Bacteriology, 191:23, 7225-7233, DOI: 10.1128/JB.00746-09

PERMISSIONS

Permission has been granted by the copyright holder to deposit this author accepted version as Open Access in the USC Research Bank. Open Access research is digital, online and free of charge, and is made possible by the consent of the author or copyright holder.

This is the author's version of a work that was accepted for publication in Journal of Bacteriology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Bacteriology, 191:23, 7225-7233, DOI: 10.1128/JB.00746-09

Copyright © 2009 American Society for Microbiology

Evidence that human *Chlamydia pneumoniae* was zoonotically acquired

G. S. A. Myers[1*], S. A. Mathews[2], M. Eppinger[1], C. Mitchell[2], K. K. O'Brien[1], O. R. White[1], F. Benahmed[3], R. C. Brunham[4], T. D. Read[5], J. Ravel[1], P. M. Bavoil[6], P. Timms[2]

[1]Institute for Genome Sciences, University of Maryland, Baltimore, Maryland 21201, USA; [2]Institute of Health & Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Queensland 4059 Australia; [3]J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, MD 20850 USA; [4]University of British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver BC V5Z 4R4, Canada; [5]Division of Infectious Diseases & Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Atlanta Georgia 30322 USA; [6]Department of Microbial Pathogenesis, University of Maryland Baltimore, 650 West Baltimore Street, Baltimore Maryland 21201 USA

Running title: Human *Chlamydia pneumoniae* was zoonotically acquired

[*]**Corresponding Author**
G.S.A. Myers
Institute for Genome Sciences
University of Maryland School of Medicine
801 West Baltimore Street, Baltimore
Maryland 21201 USA
Phone: (410) 706-5678
Fax: (410) 706-1482
Email: gmyers@som.umaryland.edu

Abstract

Zoonotic infections are a growing threat to global health. *Chlamydia pneumoniae,* a major human pathogen, is widespread in human populations causing acute respiratory disease and is associated with chronic disease. *C. pneumoniae* was first identified solely in human populations; however its host range now includes other mammals, marsupials, amphibians and reptiles. Australian koalas (*Phascolarctos cinereus*) are widely infected with two species of *Chlamydia, C. pecorum* and *C. pneumoniae*. Transmission of *C. pneumoniae* between animals and humans has not been reported; however two other chlamydial species, *C. psittaci* and *C. abortus*, are known zoonotic pathogens. We have sequenced the 1,241,024bp chromosome and a 7.5kb cryptic chlamydial plasmid of the koala strain of *C. pneumoniae* (LPCoLN) using the whole genome shotgun method. Comparative genomic analysis, including pseudogene and SNP distribution, and phylogenetic analysis of conserved genes and SNPs against the human isolates of *C. pneumoniae* show that the LPCoLN isolate is basal to human isolates. Thus we propose based on compelling genomic and phylogenetic evidence that humans were originally infected zoonotically by animal isolate(s) of *C. pneumoniae* which adapted to humans primarily through the processes of gene decay and plasmid loss, to the point where the animal reservoir is no longer required for transmission.

Introduction

Zoonotic infections from wildlife were recently suggested to be the most significant growing threat to global health of all the emerging infectious diseases (16). *Chlamydia* comprises a group of obligate intracellular bacterial parasites responsible for a variety of diseases in humans and animals, including several zoonoses. In 1999, Everett *et al.* proposed a reassignment from the single genus *Chlamydia* into two genera, *Chlamydia* and *Chlamydophila*, based on apparent differential clustering of the 16S rRNA genes (10). This change has not been widely accepted by the chlamydial research community, thus reversion to the single genus *Chlamydia* was recently recommended (33). Accordingly, we use the *Chlamydia* nomenclature here.

Chlamydia pneumoniae (previously known as TWAR) was first recognized as a distinct species in 1988 (6) and is widespread in human populations causing acute respiratory disease, with effective human-to-human transmission by aerosol (30). It has also been associated with several human chronic diseases, including asthma (35), atherosclerosis (41), stroke (9) and late onset Alzheimer's disease (1). *C. pneumoniae* was initially identified solely in humans; however its host range is now the most cosmopolitan of all the chlamydiae, encompassing both warm and cold-blooded animals such as horses, koalas and other marsupials, amphibians and reptiles (3). Populations of the Australian koala (*Phascolarctos cinereus*) are widely infected with two species of *Chlamydia*: *C. pecorum* and *C. pneumoniae* (3). While *C. pecorum* infections are present at ocular and urogenital sites, *C. pneumoniae* infections are commonly found in the koala respiratory tract linked to symptoms of respiratory disease (40), which is consistent with acute human *C. pneumoniae* disease. Transmission of *C. pneumoniae* between animals and humans has

not been documented; however two other chlamydial species, *C. psittaci* and *C. abortus*, are well known zoonotic pathogens transmitted from birds and ruminants (21) that cause psittacosis, a life-threatening pneumonia, and abortion respectively. Here we propose on the basis of compelling genomic and phylogenetic evidence that *C. pneumoniae*, a major human pathogen that is essentially clonal, was originally derived from an animal source.

The genome sequences of four epidemiologically distinct human-derived *C. pneumoniae* isolates have previously been determined (18, 27, 31). These isolates are perceived as being genetically homogenous; this is supported by fewer than 300 single nucleotide polymorphisms (SNPs) scattered around the chromosome in no discernible pattern (Fig. 1; Table 1). Such a degree of similarity between temporally and geographically disparate isolates supports a relatively recent clonal expansion of human *C. pneumoniae* isolates (26) but is otherwise uninformative for deciphering the evolutionary origin(s) of this pathogen. Accordingly, we sequenced the complete genome of the koala *C. pneumoniae* isolate LPCoLN, seeking molecular insight into host specificity, evolutionary origin and pathogenicity.

Materials and Methods

The koala *C. pneumoniae* LPCoLN isolate was originally isolated from a nasal swab of a captive koala showing signs of respiratory illness. *C. pneumoniae* was detected by PCR and gene sequencing, and LPCoLN was grown *in vitro* in HEp-2 cell monolayers. No other bacterium or virus was recovered from the nasal swab.

The complete genome sequence of *C. pneumoniae* LPCoLN was determined using the whole-genome shotgun method (23). Physical and sequencing gaps were closed using a combination of primer walking, generation and sequencing of transposon-tagged libraries of large-insert clones, and multiplex PCR (36). Identification of putative protein-coding genes and annotation of the genome were performed as previously described (23). An initial set of coding sequences (CDSs) predicted to encode proteins was identified with GLIMMER (7). CDSs consisting of fewer than 30 codons were eliminated. Frame-shift and point mutations were corrected or designated "authentic", as previously described (23). Functional assignment, identification of membrane-spanning domains, determination of paralogous gene families, and identification of regions of unusual nucleotide composition were performed as previously described (23). Sequence alignments were generated using the methods described previously (23).

C. pneumoniae LPCoLN and the genomes of four previously sequenced human-derived *C. pneumoniae* isolates (18, 27, 31) (GenBank accession numbers: CWL029, AE001363; TW-183, AE009440; AR39, AE002161 and J138, BA000008) were compared at the nucleotide level by suffix tree analysis using MUMmer (8) and the data parsed by custom Perl scripts. Predicted *C. pneumoniae* genes were compared by BLAST against the complete set of genes from other chlamydial genomes using an E-value cut-off of 10^{-5}. Synteny and BLAST Score Ratio analyses were performed as previously described (25).

High quality synonymous SNPs (sSNPs) were identified by comparing the predicted genes on the closed genome of *C. pneumoniae* strain AR39 with the LPCoLN genome sequence using

MUMmer (8). A polymorphic site was considered high quality when its underlying sequence comprised at least three sequencing reads with an average Phred quality score greater than 30 (11). sSNPs in CWL029, TW183 and J138 were similarly identified although no assessment of quality could be made, as quality scores are not available for these genomes. Concatenated sSNPs for the individual *C. pneumoniae* isolates were further analyzed by the HKY85 method (13) with 200 bootstrap replicates, and the results used to generate an unrooted phylogenetic tree according to the PhyLM algorithms (12).

One hundred and eleven clusters of shared proteins, with a BLAST Score Ratio greater than, or equal to 0.8 (25), were identified between all *C. pneumoniae* isolates, *C. pecorum* E58 (Myers *et al.*, unpublished), *C. muridarum* Nigg (27) (GenBank: AE002160), *C. caviae* GPIC (28) (GenBank: AE015925), *C. psittaci* 6BC (Myers *et al.*, unpublished) and *C. abortus* S26/3 (39) (GenBank: CR848038). Protein clusters were aligned using ClustalX (37) and back translated into nucleotide alignments using TRANSALIGN, part of the EMBOSS software package (29). Concatenated aligned genes, spanning a total of 121,674 positions with a sequence similarity of 82.2% and identity of 58.8% were further analyzed by the HKY85 method (13) with 200 bootstrap replicates, and the results used to generate an unrooted phylogenetic tree according to the PhyLM algorithms (12).

Results

C. pneumoniae LPCoLN possesses a single, circular chromosome of 1,241,024 base pairs (bp), slightly larger (by approximately 10 kb) than the human-derived *C. pneumoniae* isolates. The

small cryptic chlamydial plasmid (7,655 bp) that is absent from all characterized human *C. pneumoniae* isolates, is present in the koala strain, and is highly conserved with previously published chlamydial plasmid sequences. LPCoLN has 1,095 predicted coding sequences (CDSs), with 988 (90.2%) CDSs conserved with human *C. pneumoniae* isolate AR39, 14 (1.3%) divergent and 93 unique relative to AR39 (Fig. 1, Table 2). Most unique CDSs encode hypothetical proteins with no currently discernible function (Supplementary Table 1).

Comparative genomic and proteomic analyses (25) show that the LPCoLN genome is highly similar and syntenic with the four sequenced human-derived isolates (Supplementary Fig. 1). However, unlike the small number of SNPs found between the human-derived isolates, 6213 SNPs (3298 synonymous and 2915 non-synonymous) separate the genomes of LPCoLN and human isolate AR39 (Tables 1 & 3). Phylogenetic analysis of all *C. pneumoniae* isolates based on SNPs (Fig. 2) and 111 highly conserved genes from across all sequenced animal chlamydial genomes (Fig. 3) indicate that LPCoLN is basal to the sequenced *C. pneumoniae* isolates from humans. Thus, while LPCoLN is a contemporary isolate, phylogeny places it closer to a presumptive ancestor of the *C. pneumoniae* isolates found in human populations.

The genome-wide SNP distribution observed in the koala isolate compared to the human-derived isolates provides further evidence for a zoonotic origin of *C. pneumoniae* recovered from humans. There are ten noteworthy regions of SNP accumulation (Fig. 1 & 4), representing genomic “hotspots” that are likely evolving at different rates in *C. pneumoniae* from koalas and humans. Notably, many of the human isolates’ CDSs within these hotspots are truncated or

fragmented relative to LPCoLN, suggesting ongoing gene decay processes, with presumed concomitant loss of function in human-derived *C. pneumoniae*. Several of these hotspots encode known virulence or metabolic factors that display sequence polymorphisms and are variably represented in other chlamydial strains and species, including the polymorphic membrane protein (Pmp) family, secreted type III secretion (T3S) effectors, and enzymes involved in the biosynthesis of chorismate, a precursor of aromatic amino acids (Fig. 1 & 4). Gene truncation and fragmentation is also evident at several of these loci within the human-derived isolates, suggesting that microevolutionary processes are also ongoing in human *C. pneumoniae*. Of the human isolates, CWL029 consistently exhibits a higher degree of gene truncation and fragmentation in several hotspots; AR39 shows the least, with TW-183 and J138 being intermediate.

The largest SNP hotspot corresponds to the "plasticity zone" (PZ), a region that encapsulates much of the sequence diversity in all chlamydial genomes (28). The PZ, which has been shown to contain host and/or tissue specific genes in other chlamydial species (27, 28), appears to be fully intact in the koala isolate but is highly fragmented in all sequenced human-derived *C. pneumoniae* isolates. While many small CDSs appear to be unique to human-derived *C. pneumoniae*, comparison to LPCoLN reveals that several of these are actual remnants of four larger genes (Fig. 1) that are part of a previously unknown 11-member gene family encoding predicted membrane-bound proteins with predicted membrane spanning domains.

There are only three genes of known function that are present in the human isolates and absent from LPCoLN: *guaBA* and *add*, required for the synthesis of guanosine 5'-monophosphate, a

precursor for the synthesis of guanine nucleoside triphosphates, located in the PZ of the human-derived *C. pneumoniae* isolates. However, in three of the four human isolates, *guaB* is fragmented, indicating that it is presumably not essential for human infection. Such gene fragmentation patterns are only observed in the genomes of human isolates, and are only discernible by comparison to the koala LPCoLN genome. This unidirectional pattern of gene fragmentation seen throughout the human-derived *C. pneumoniae* genomes not only supports the phylogenetic analyses (Fig. 2 & 3), suggesting that animal-derived *C. pneumoniae* predates human-derived *C. pneumoniae*, but also suggests that animals were the original hosts for *C. pneumoniae.*

Discussion

Prior to this study, *C. pneumoniae* genome sequences were only available for four isolates, all of human origin. These genomes showed surprisingly high similarity, with an approximate total of only 300 SNPs between them. Such a high degree of genomic conservation has been hypothesized to be evidence that *C. pneumoniae* was recently transmitted to humans followed by a rapid spread throughout human populations, giving little opportunity for genomic changes. This level of homology within *C. pneumoniae* is in contrast to the degree of genetic variability seen in the other chlamydial species, in particular *C. trachomatis*, which is thought to have infected humans throughout human evolution (4, 19, 32, 38). The host range of *C. pneumoniae* has been expanded significantly in the last 10 years with infections reported in horses (34), reptiles (3), amphibians (2, 3, 14) and several Australian marsupials, including koalas (40) and bandicoots (20). Previous DNA sequence comparisons have focused on 16S rRNA and *ompA* genes. While these analyses have revealed differences between strains of human and animal origins, these differences have been minimal and relatively uninformative with regard to

determinants of host specificity. Our whole-genome analysis of the koala LPCoLN isolate of *C. pneumoniae* has enabled insight into the genetic differences between animal and human-derived *C. pneumoniae*, and the putative evolutionary events that have governed the spread of this organism, and shows that human isolates of *C. pneumoniae* exhibit more heterogeneity than previously thought.

The chlamydial cryptic plasmid is present in some chlamydial species, including *C. pneumoniae* N16 isolated from horses (24) but is absent from others. The role of the plasmid in chlamydial biology is still largely unknown. All human-derived *C. pneumoniae* isolates studied to date lack the plasmid, however, the koala isolate carries a full-length chlamydial plasmid, encoding all eight CDSs described in other chlamydial cryptic plasmids. Mitchell *et al*. (22) reported a much faster growth rate *in vitro* for the koala LPCoLN isolate compared to the human isolate AR39 - it is possible that one or more of the genes present on the cryptic plasmid may account for this faster growth rate.

The most compelling evidence to support that LPCoLN is either ancestral, or closely related to an ancestral form of *C. pneumoniae* human isolates, is the presence of several putatively full-length CDSs in LPCoLN, which are fragmented in human-derived *C. pneumoniae*, forming clusters of pseudogenes (Fig. 1 & 4). The MAC/perforin gene, which has been associated with virulence in other intracellular pathogens including *Toxoplasma* (17) and *Plasmodium* (15), is a 2,457 bp CDS in LPCoLN, but is partially truncated in all four human-derived isolates due to an 840 bp deletion towards the 5' end. Although the function of chlamydial MAC/perforin is

currently unknown, we predict that it may be involved in host cell egression and invasion similar to *Toxoplasma*. The truncated version seen in the human isolates may then reflect adaptation to a specific niche within humans.

All *pmpE* and *pmpG* orthologs are intact in the koala strain but are fragmented in several of the human isolates (Fig. 1). The Pmp family of proteins is considered to represent the expansion of progenitor proteins proposed to be involved in key roles such as adherence, immune evasion and proinflammatory responses. In addition, orthologs of the Inc family of proteins are also extensively fragmented in the human isolates (Fig. 4). Inc proteins are a diverse family of chlamydial type III secreted effectors; *incA* has been localized to the outer face of the inclusion membrane and is involved in the homotypic fusion of multiple inclusions of *C. trachomatis*. The apparent ongoing loss of several functional *pmp* and *inc* genes in the human-derived isolates of *C. pneumoniae* again suggests an adaptation to the human host. It is conceivable that different Pmp and Inc profiles confer differential niche specificity in different hosts. The fragmentation of functional *pmp* and *inc* alleles in human-derived *C. pneumoniae* may therefore represent an example of convergent evolution of the two species in response to properties that are specific to humans (e.g. a more effective immune response against PmpE antigens in humans or the relative unavailability of cell surface receptors to PmpE in humans vs. koalas).

Taken together, our analysis of the koala *C. pneumoniae* LPCoLN genome sequence, combined with phylogenetic analyses of all *C. pneumoniae* SNPs and the conserved chlamydial CDSs, the patterns of CDS fragmentation and plasmid loss in human-derived *C. pneumoniae* isolates

provides strong evidence that human isolates of *C. pneumoniae* have derived from zoonotic *C. pneumoniae,* supporting the conclusion of the selected SNP analysis of Rattei *et al* (26). Thus we propose that *C. pneumoniae* was originally an animal pathogen that crossed the species barrier to humans through ongoing reductive evolutionary processes and has adapted to the point where human isolates of *C. pneumoniae* no longer require an animal reservoir for transmission.

A limitation of our study is that it is based on the genome sequence of only one animal-derived *C. pneumoniae* isolate, the koala LPCoLN strain and of four human-derived *C. pneumoniae* isolates. Hence, key questions such as how many times has this host species jump occurred before terminal adaptation or from which specific animal host the human-derived *C. pneumoniae* actually originated cannot be addressed with this dataset alone. In addition to the full genome sequence from the LPCoLN isolate of koala *C. pneumoniae*, we also obtained and analysed a second koala *C. pneumoniae* isolate, EBB. This isolate was obtained from a pharyngeal swab from a koala in a wild population from a geographically separate location than LPCoLN. Nine genes were sequenced from the EBB isolate and in all cases the sequences were 100% identical to the sequences obtained from the koala *C. pneumoniae* LPCoLN isolate (data not shown).

The koala *C. pneumoniae* LPCoLN isolate has been relatively well characterized previously in regards to morphological and *in vitro* growth characteristics. Coles *et al.* (5) reported that LPCoLN produced large inclusions in both human and koala monocytes and in HEp-2 cells. Koala *C. pneumoniae* was able to induce foam cell formation both with and without added low-density lipoprotein, in contrast to TW183, which produced increased foam cell formation only in

the presence of low-density lipoprotein. More recently, Mitchell *et al*. (22) compared the *in vitro* growth characteristics of LPCoLN with the human isolate AR39. LPCoLN displayed inclusions of size and morphology clearly distinct from those of the human isolate, and had a much faster doubling time (3.4-4.9 hr versus 5.9-8.7 hr) when grown in HEp2 cell monolayers. Rates of inclusion fusion were also much higher with LPCoLN (100%) than with AR39 (30-40%). These biological differences between koala and human-derived *C. pneumoniae* are consistent with the range of genomic differences that we have identified in this work. Such phenotypic studies demonstrate the compensatory power of comparative pathogenomics in a genetically intractable organism such as *C. pneumoniae*. Moreover the ability to compare genome sequences of organisms infecting different hosts provides "snapshots" of the evolutionary process as if frozen in time. The search is now on to find *C. pneumoniae* isolates from animals that are most closely related to human-derived isolates, as this will better indicate when this host species jump may have occurred.

Our findings indicate that the high prevalence and disease burden of *C. pneumoniae* in humans may represent a major evolutionary and public health corollary of zoonotic infections - the emergence of a full-fledged human pathogen, transmitted without the original animal vector, causing substantial acute and chronic disease sequelae.

References

1. **Balin, B. J., C. S. Little, C. J. Hammond, D. M. Appelt, J. A. Whittum-Hudson, H. C. Gerard, and A. P. Hudson.** 2008. *Chlamydophila pneumoniae* and the etiology of late-onset Alzheimer's disease. J Alzheimers Dis **13:**371-80.

2. **Berger, L., K. Volp, S. Mathews, R. Speare, and P. Timms.** 1999. *Chlamydia pneumoniae* in a free-ranging giant barred frog (*Mixophyes iteratus*) from Australia. J Clin Microbiol **37:**2378-80.
3. **Bodetti, T. J., E. Jacobson, C. Wan, L. Hafner, A. Pospischil, K. Rose, and P. Timms.** 2002. Molecular evidence to support the expansion of the hostrange of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. Syst Appl Microbiol **25:**146-52.
4. **Carlson, J. H., S. F. Porcella, G. McClarty, and H. D. Caldwell.** 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. Infect Immun **73:**6407-18.
5. **Coles, K. A., P. Timms, and D. W. Smith.** 2001. Koala biovar of *Chlamydia pneumoniae* infects human and koala monocytes and induces increased uptake of lipids *in vitro*. Infect Immun **69:**7894-7.
6. **Cox, R., C.-C. Kuo, T. Grayston, and L. A. Campbell.** 1988. Deoxyribonucleic acid relatedness of *Chlamydia* sp. strain TWAR to *Chlamydia trachomatis* and *Chlamydia psittaci*. Int J Syst Bacteriol **38:**265-268.
7. **Delcher, A. L., S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg.** 1999. Alignment of whole genomes. Nucleic Acids Res **27:**2369-76.
8. **Delcher, A. L., A. Phillippy, J. Carlton, and S. L. Salzberg.** 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res **30:**2478-83.
9. **Elkind, M. S., and J. W. Cole.** 2006. Do common infections cause stroke? Semin Neurol **26:**88-99.
10. **Everett, K. D., R. M. Bush, and A. A. Andersen.** 1999. Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol **49 Pt 2:**415-40.
11. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8:**175-85.
12. **Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel.** 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res **33:**W557-9.
13. **Hasegawa, M., H. Kishino, and T. Yano.** 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol **22:**160-74.
14. **Hotzel, H., E. Grossmann, F. Mutschmann, and K. Sachse.** 2001. Genetic characterization of a *Chlamydophila pneumoniae* isolate from an African frog and comparison to currently accepted biovars. Syst Appl Microbiol **24:**63-6.
15. **Ishino, T., Y. Chinzei, and M. Yuda.** 2005. A *Plasmodium* sporozoite protein with a membrane attack complex domain is required for breaching the liver sinusoidal cell layer prior to hepatocyte infection. Cell Microbiol **7:**199-208.
16. **Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak.** 2008. Global trends in emerging infectious diseases. Nature **451:**990-3.
17. **Kafsack, B. F., J. D. Pena, I. Coppens, S. Ravindran, J. C. Boothroyd, and V. B.**

Carruthers. 2009. Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. Science **323:**530-3.

18. **Kalman, S., W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis, and R. S. Stephens.** 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat Genet **21:**385-9.
19. **Kari, L., W. M. Whitmire, J. H. Carlson, D. D. Crane, N. Reveneau, D. E. Nelson, D. C. Mabey, R. L. Bailey, M. J. Holland, G. McClarty, and H. D. Caldwell.** 2008. Pathogenic diversity among *Chlamydia trachomatis* ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis **197:**449-56.
20. **Kutlin, A., P. M. Roblin, S. Kumar, S. Kohlhoff, T. Bodetti, P. Timms, and M. R. Hammerschlag.** 2007. Molecular characterization of *Chlamydophila pneumoniae* isolates from Western barred bandicoots. J Med Microbiol **56:**407-17.
21. **Longbottom, D., and L. J. Coulter.** 2003. Animal chlamydioses and zoonotic implications. J Comp Pathol **128:**217-44.
22. **Mitchell, C. M., S. A. Mathews, C. Theodoropoulos, and P. Timms.** 2009. In vitro characterisation of koala *Chlamydia pneumoniae*: morphology, inclusion development and doubling time. Vet Microbiol **136:**91-9.
23. **Myers, G. S., D. Parker, K. Al-Hasani, R. M. Kennan, T. Seemann, Q. Ren, J. H. Badger, J. D. Selengut, R. T. Deboy, H. Tettelin, J. D. Boyce, V. P. McCarl, X. Han, W. C. Nelson, R. Madupu, Y. Mohamoud, T. Holley, N. Fedorova, H. Khouri, S. P. Bottomley, R. J. Whittington, B. Adler, J. G. Songer, J. I. Rood, and I. T. Paulsen.** 2007. Genome sequence and identification of candidate vaccine antigens from the animal pathogen *Dichelobacter nodosus*. Nat Biotechnol **25:**569-75.
24. **Pickett, M. A., J. S. Everson, P. J. Pead, and I. N. Clarke.** 2005. The plasmids of *Chlamydia trachomatis* and *Chlamydophila pneumoniae* (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents. Microbiology **151:**893-903.
25. **Rasko, D. A., G. S. Myers, and J. Ravel.** 2005. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics **6:**2.
26. **Rattei, T., S. Ott, M. Gutacker, J. Rupp, M. Maass, S. Schreiber, W. Solbach, T. Wirth, and J. Gieffers.** 2007. Genetic diversity of the obligate intracellular bacterium *Chlamydophila pneumoniae* by genome-wide analysis of single nucleotide polymorphisms: evidence for highly clonal population structure. BMC Genomics **8:**355.
27. **Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser.** 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res **28:**1397-406.
28. **Read, T. D., G. S. Myers, R. C. Brunham, W. C. Nelson, I. T. Paulsen, J. Heidelberg, E. Holtzapple, H. Khouri, N. B. Federova, H. A. Carty, L. A. Umayam, D. H. Haft, J. Peterson, M. J. Beanan, O. White, S. L. Salzberg, R. C. Hsia, G. McClarty, R. G. Rank, P. M. Bavoil, and C. M. Fraser.** 2003. Genome sequence of *Chlamydophila*

caviae (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. Nucleic Acids Res **31:**2134-47.

29. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16:**276-7.
30. **Saikku, P.** 1992. The epidemiology and significance of *Chlamydia pneumoniae*. J Infect **25 Suppl 1:**27-34.
31. **Shirai, M., H. Hirakawa, M. Kimoto, M. Tabuchi, F. Kishi, K. Ouchi, T. Shiba, K. Ishii, M. Hattori, S. Kuhara, and T. Nakazawa.** 2000. Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. Nucleic Acids Res **28:**2311-4.
32. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis.** 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282:**754-9.
33. **Stephens, R. S., G. Myers, M. Eppinger, and P. M. Bavoil.** 2009. Divergence without difference: phylogenetics and taxonomy of *Chlamydia* resolved. FEMS Immunol Med Microbiol **55:**115-9.
34. **Storey, C., M. Lusher, P. Yates, and S. Richmond.** 1993. Evidence for *Chlamydia pneumoniae* of non-human origin. J Gen Microbiol **139:**2621-6.
35. **Sutherland, E. R., and R. J. Martin.** 2007. Asthma and atypical bacterial infection. Chest **132:**1962-6.
36. **Tettelin, H., D. Radune, S. Kasif, H. Khouri, and S. L. Salzberg.** 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. Genomics **62:**500-7.
37. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:**4876-82.
38. **Thomson, N. R., M. T. Holden, C. Carder, N. Lennard, S. J. Lockey, P. Marsh, P. Skipp, C. D. O'Connor, I. Goodhead, H. Norbertzcak, B. Harris, D. Ormond, R. Rance, M. A. Quail, J. Parkhill, R. S. Stephens, and I. N. Clarke.** 2008. *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res **18:**161-71.
39. **Thomson, N. R., C. Yeats, K. Bell, M. T. Holden, S. D. Bentley, M. Livingstone, A. M. Cerdeno-Tarraga, B. Harris, J. Doggett, D. Ormond, K. Mungall, K. Clarke, T. Feltwell, Z. Hance, M. Sanders, M. A. Quail, C. Price, B. G. Barrell, J. Parkhill, and D. Longbottom.** 2005. The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Res **15:**629-40.
40. **Wardrop, S., A. Fowler, P. O'Callaghan, P. Giffard, and P. Timms.** 1999. Characterization of the koala biovar of *Chlamydia pneumoniae* at four gene loci--ompAVD4, ompB, 16S rRNA, groESL spacer region. Syst Appl Microbiol **22:**22-7.
41. **Watson, C., and N. J. Alp.** 2008. Role of *Chlamydia* pneumoniae in atherosclerosis. Clin Sci (Lond) **114:**509-31.

Acknowledgements This work was supported by the National Institute of Allergy and Infectious Disease grant 1R01AI051472. We thank the former TIGR and current IGS Faculty, the TIGR/IGS Informatics group for expert advice and assistance, and the JCVI Sequencing Facility.

The sequences of the *C. pneumoniae* LPCoLN chromosome and plasmid have been deposited in GenBank with the accession numbers CP001713 and CP001714 respectively.

Figure 1. Comparative analysis of sequenced *C. pneumoniae* genomes. Circular representation of *C. pneumoniae* genomes and analyses. For each genome, data are from outermost circle to innermost. Circles 1 and 2: Tick marks represent predicted coding sequences on the plus strand of *C. pneumoniae* AR39 and minus strand respectively, colored by cellular role. Role categories and colors are as follows: amino acid biosynthesis, violet; biosynthesis of cofactors, prosthetic groups and carriers, light blue; cell envelope, light green; cellular processes, red; central intermediary metabolism, brown; DNA metabolism, gold; energy metabolism, light gray; fatty acid and phospholipid metabolism, magenta; protein synthesis and fate, pink; biosynthesis of purines, pyrimidines, nucleosides and nucleotides, orange; regulatory functions and signal transduction, olive; transcription, dark green; transport and binding proteins, blue-green; other categories, salmon; unknown function, gray; conserved hypothetical proteins, blue; hypothetical proteins, black; Circles 3 and 4: Histogram of cumulative SNP density on the plus and minus strands respectively of the LPCoLN genome compared to AR39, using a 5000bp window. Histogram coloring: <30 SNPs, red; 20-29 SNPs, orange; 10-19 SNPs, yellow; 0-9 SNPs, light blue. Circles 5 and 6: Tick marks represent SNP locations on the plus (green) and minus (red) strands respectively of the LPCoLN genome. Circles 7 and 8: Tick marks represent SNP locations on the plus (green) and minus (red) strands respectively of the TW183 genome. Circles 9 and 10: Tick marks represent SNP locations on the plus (green) and minus (red) strands respectively of the J138 genome. Circles 11 and 12: Tick marks represent SNP locations on the plus (green) and minus (red) strands respectively of the CWL-029 genome. Ten regions of high SNP accumulation in the LPCoLN genome versus AR39 are marked with light blue radial

sections and numbered. Callouts: Detail of high SNP accumulation in regions III and VI, showing SNP location and type (synonymous, green; non-synonymous, red) within PMP clusters (from left to right, LPCoLN gene region: ORF00989 to ORF00956; AR39 gene region: CP_0280 to CP_0309) and the PZ (LPCoLN gene region: ORF00689 to ORF00665; AR39 gene region: CP_0585 to CP_0622) respectively. Regions highlighted with gray boxes show SNP-associated CDS fragmentation.

Figure 2. Synonymous SNP phylogenetic tree using all sequenced *C. pneumoniae* genomes. The number of separating sSNPs is denoted on each branch.

Figure 3. Phylogeny of all animal chlamydiae and *C. pneumoniae*, using 111 highly conserved gene clusters. The host range for each species is noted in parentheses; the known chlamydial zoonotic agents are boxed. The bootstrap value at each branch point is 100% unless otherwise denoted.

Figure 4. Annotated detail of additional regions shown in Figure 1 with high SNP accumulation, showing SNP location and type (synonymous, green; non-synonymous, red). Colours are used to distinguish notable annotated genes in each region. Regions highlighted with gray boxes show SNP-associated CDS fragmentation. The icon to the left of each region denotes the approximate location of each region as shown in Figure 1.

Table 1. Total SNPs in sequenced *C. pneumoniae* genomes, using *C. pneumoniae* AR39 as reference.

Isolate	Host species	# synonymous SNPS	# non-synonymous SNPs	Total SNPs
TW-183	*Homo sapiens*	120	164	284
CWL029	*Homo sapiens*	120	154	274
J138	*Homo sapiens*	62	145	207
LPCoLN	*Phascolarctos cinereus*	3298	2915	6213

Table 2. Breakdown of predicted protein orthologs in all human-derived *C. pneumoniae* genomes, compared to *C. pneumoniae* LPCoLN using the Blast Score Ratio method (25).

	Conserved (≥0.5)	Unique (≤0.4)	Divergent (< 0.5 & > 0.4)	Total
LPCoLN vs. AR39	988	93	14	1095
LPCoLN vs. TW138	989	93	13	1095
LPCoLN vs. CWL029	955	127	13	1095
LPCoLN vs. J138	982	98	15	1095

Table 3. Total (a) Synonymous and (b) Nonsynonymous SNPs identified between *C. pneumoniae* genomes, showing shared (italics) and separating unique (bold) SNPs. SNPs were identified in three-way genome comparisons in respect to the CDS of the reference strain *C. pneumoniae* AR39. The individual branch lengths of the separating SNPs are given in parenthesis.

(a)

	AR39	LPCoLN	TW-183	CWL029	J138
LPCoLN	*2969*		**47**	**43**	**6**
TW-183	*105*	*2980 (2922/58)*		**41**	**3**
CWL029	*103*	*2986 (2926/60)*	*126 (64/62)*		**6**
J138	*58*	*3015 (2963/52)*	*157 (102/55)*	*149 (97/52)*	

(b)

	AR39	LPCoLN	TW-183	CWL029	J138
LPCoLN	*2435*		**80**	**70**	**18**
TW-183	*144*	*2419 (2355/64)*		**70**	**17**
CWL029	*125*	*2420 (2365/70)*	*123 (71/51)*		**17**
J138	*128*	*2527 (2417/1100*	*238 (127/111)*	*219 (108/111)*	