JOURNAL OF BACTERIOLOGY, Dec. 2009, p. 7225–7233 Vol. 191, No. 23
0021-9193/09/$12.00 doi:10.1128/JB.00746-09


# 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] and P. Timms[2]

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

Received 9 June 2009/Accepted 2 September 2009

**Zoonotic infections are a growing threat to global health. *Chlamydia pneumoniae* is a major human pathogen that is widespread in human populations, causing acute respiratory disease, and has been 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 1,241,024-bp chromosome and a 7.5-kb cryptic chlamydial plasmid of the koala strain of *C. pneumoniae* (LPCoLN) using the whole-genome shotgun method. Comparative genomic analysis, including pseudogene and single-nucleotide polymorphism (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 an 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.**

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 and 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 and are 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 were previously determined (18, 27, 31). These isolates are perceived as being genetically homogenous; this is supported by the fact that there are 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,

* Corresponding author. Mailing address: Institute for Genome Sciences, University of Maryland School of Medicine, 801 West Baltimore Street, Baltimore, MD 21201. Phone: (410) 706-5678. Fax: (410) 706-1482. E-mail: gmyers@som.umaryland.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 11 September 2009.

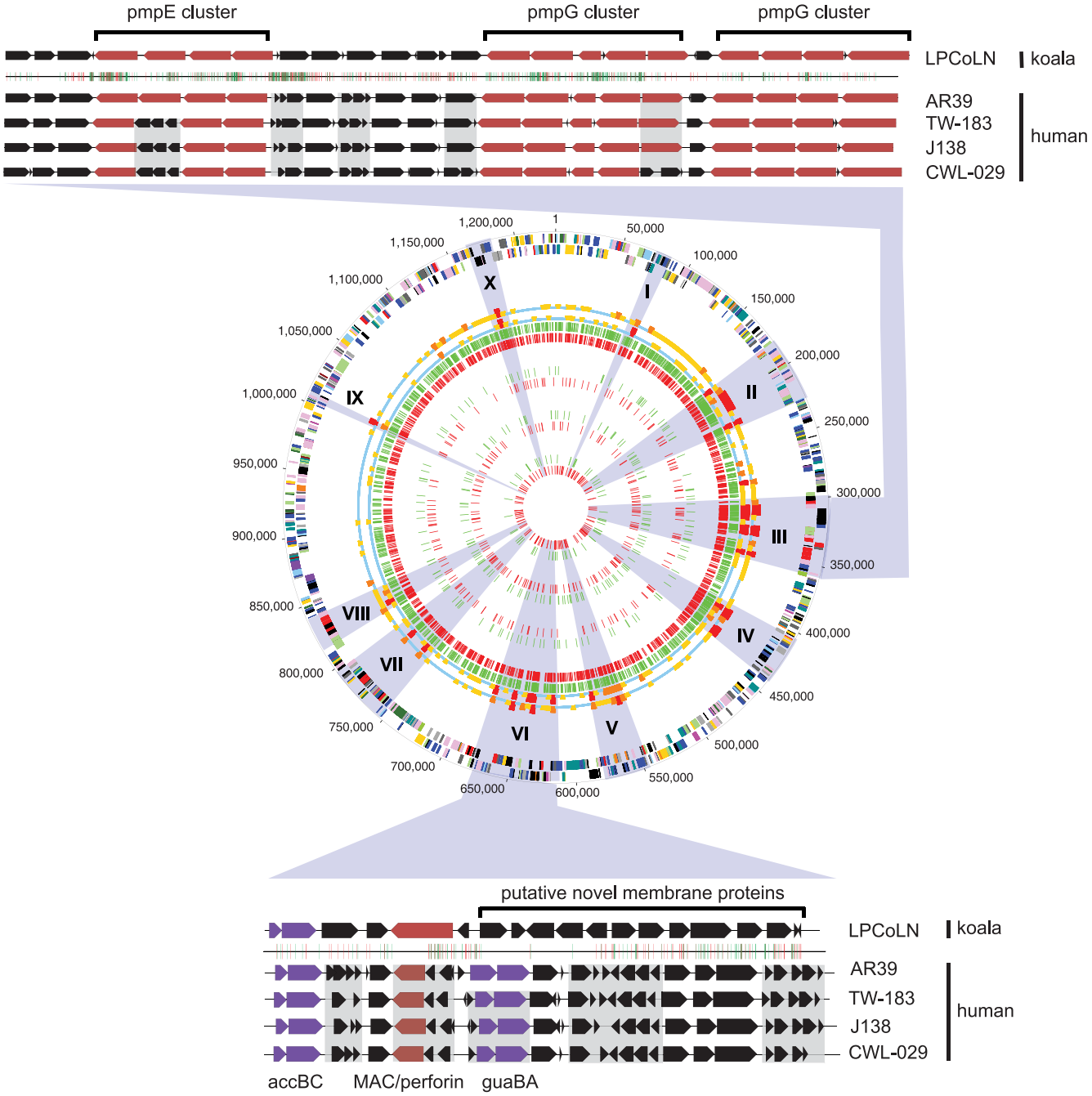


FIG. 1. Circular representation of *C. pneumoniae* genomes and comparative analyses. For each genome, data are from outermost circle to innermost. In circles 1 and 2, tick marks represent predicted CDSs on the plus strand of *C. pneumoniae* AR39 and minus strand, respectively, colored by cellular role 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 show a histogram of cumulative SNP density on the plus and minus strands, respectively, of the LPCoLN genome compared to AR39, using a 5,000-bp window. Histogram coloring: <30 SNPs, red; 20 to 29 SNPs, orange; 10 to 19 SNPs, yellow; 0 to 9 SNPs, light blue. In circles 5 and 6, tick marks represent SNP locations on the plus (green) and minus (red) strands, respectively, of the LPCoLN genome. In circles 7 to 12, tick marks represent SNP locations on the plus (green) and minus (red) strands of the TW183 (circles 7 and 8), J138 (circles 9 and 10), and CWL-029 (circles 11 and 12) genomes. Ten regions of high SNP accumulation in the LPCoLN genome versus AR39 are marked with light blue radial sections and numbered. The top and bottom panels show detail of high SNP accumulation in regions III and VI, with SNP location and type (synonymous, green; nonsynonymous, red) within PMP clusters ([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. Grey highlighting shows SNP-associated CDS fragmentation.

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

| Isolate | Host species | No. of SNPs | | |
|---|---|---|---|---|
| | | Synonymous | Nonsynonymous | Total |
| TW-183 | *Homo sapiens* | 120 | 164 | 284 |
| CWL029 | *Homo sapiens* | 120 | 154 | 274 |
| J138 | *Homo sapiens* | 62 | 145 | 207 |
| LPCoLN | *Phascolarctos cinereus* | 3,298 | 2,915 | 6,213 |

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)

| Strain used for comparison | No. of orthologs (score) | | | |
|---|---|---|---|---|
| | Conserved (≥0.5) | Unique (≤0.4) | Divergent (<0.5 and >0.4) | Total |
| AR39 | 988 | 93 | 14 | 1,095 |
| TW138 | 989 | 93 | 13 | 1,095 |
| CWL029 | 955 | 127 | 13 | 1,095 |
| J138 | 982 | 98 | 15 | 1,095 |

seeking molecular insight into host specificity, evolutionary origin, and pathogenicity.

## MATERIALS AND METHODS

The koala *C. pneumoniae* isolate LPCoLN 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. Frameshift 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 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 cutoff 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 were used to generate an unrooted phylogenetic tree according to the PhyLM algorithms (12).

One hundred eleven clusters of shared proteins, with a BLAST score ratio greater than or equal to 0.8 (25), were identified between the *C. pneumoniae* isolates *C. pecorum* E58 (G. S. A. Myers, unpublished data), *C. muridarum* Nigg (27) (GenBank no., AE002160), *C. caviae* GPIC (28) (GenBank no., AE015925), *C. psittaci* 6BC (Myers, unpublished), and *C. abortus* S26/3 (39) (GenBank no., 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 were used to generate an unrooted phylogenetic tree according to the PhyLM algorithms (12).

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

## RESULTS

*C. pneumoniae* LPCoLN possesses a single, circular chromosome of 1,241,024 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 relative to published chlamydial plasmid sequences. LPCoLN has 1,095 predicted CDSs, with 988 (90.2%) CDSs conserved, 14 (1.3%) divergent, and 93 unique relative to the human *C. pneumoniae* isolate AR39 (Fig. 1; Table 2). Most unique CDSs encode hypothetical proteins with no currently discernible function (see Table S1 in the supplemental material).

Comparative genomic and proteomic analyses (25) show that the LPCoLN genome is highly similar to and syntenic with the four sequenced human-derived isolates (see Fig. S1 in the supplemental material). However, unlike the small number of SNPs found between the human-derived isolates, 6,213 SNPs (3,298 synonymous and 2,915 nonsynonymous) separate the genomes of LPCoLN and human isolate AR39 (Tables 1 and 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 10 noteworthy regions of SNP accumulation (Fig. 1 and 4), representing genomic hot spots that are likely evolving at different rates in *C. pneumoniae* from koalas and humans. Notably, many of the human isolates' CDSs within these hot spots 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 hot spots 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 effectors, and enzymes involved in the biosynthesis of chorismate, a precursor of aromatic amino acids (Fig. 1 and 4). Gene truncation and fragmentation are 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

TABLE 3. SNPs identified between *C. pneumoniae* genomes[a]

| SNP type and strain | No. of SNPs relative to[b]: | | | | |
|---|---|---|---|---|---|
| | AR39 | LPCoLN | TW-183 | CWL029 | J138 |
| Synonymous | | | | | |
| LPCoLN | 2,969 | | **47** | **43** | **6** |
| TW-183 | 105 | 2,980 (2,922/58) | | **41** | **3** |
| CWL029 | 103 | 2,986 (2,926/60) | 126 (64/62) | | **6** |
| J138 | 58 | 3,015 (2,963/52) | 157 (102/55) | 149 (97/52) | |
| Nonsynonymous | | | | | |
| LPCoLN | 2,435 | | **80** | **70** | **18** |
| TW-183 | 144 | 2,419 (2,355/64) | | **70** | **17** |
| CWL029 | 125 | 2,420 (2,365/70) | 123 (71/51) | | **17** |
| J138 | 128 | 2,527 (2,417/1100) | 238 (127/111) | 219 (108/111) | |

[a] SNPs were identified in three-way genome comparisons with respect to the CDS of the reference strain, *C. pneumoniae* AR39.
[b] Numbers of shared (lighface) and separating unique (boldface) SNPs are shown. The individual branch lengths of the separating SNPs are given in parentheses.

human isolates, CWL029 consistently exhibits a higher degree of gene truncation and fragmentation in several hot spots; AR39 shows the least, with TW-183 and J138 being intermediate.

The largest SNP hot spot 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.

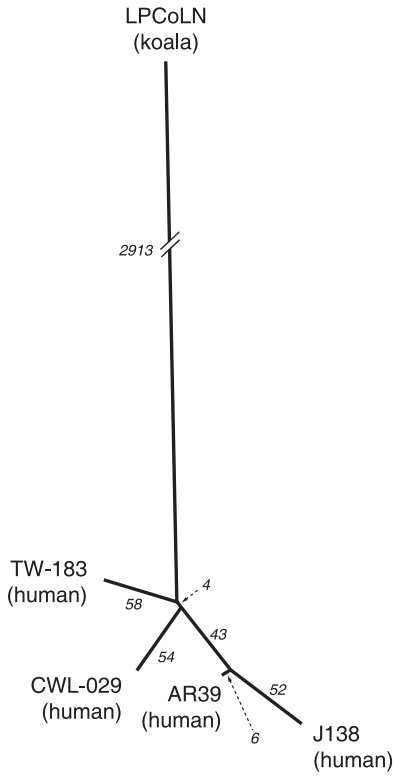


FIG. 2. sSNP phylogenetic tree using all sequenced *C. pneumoniae* genomes. The number of separating sSNPs is give on each branch.

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 GMP, 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 observed only in the genomes of human isolates and are discernible only 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 and 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 available for only 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

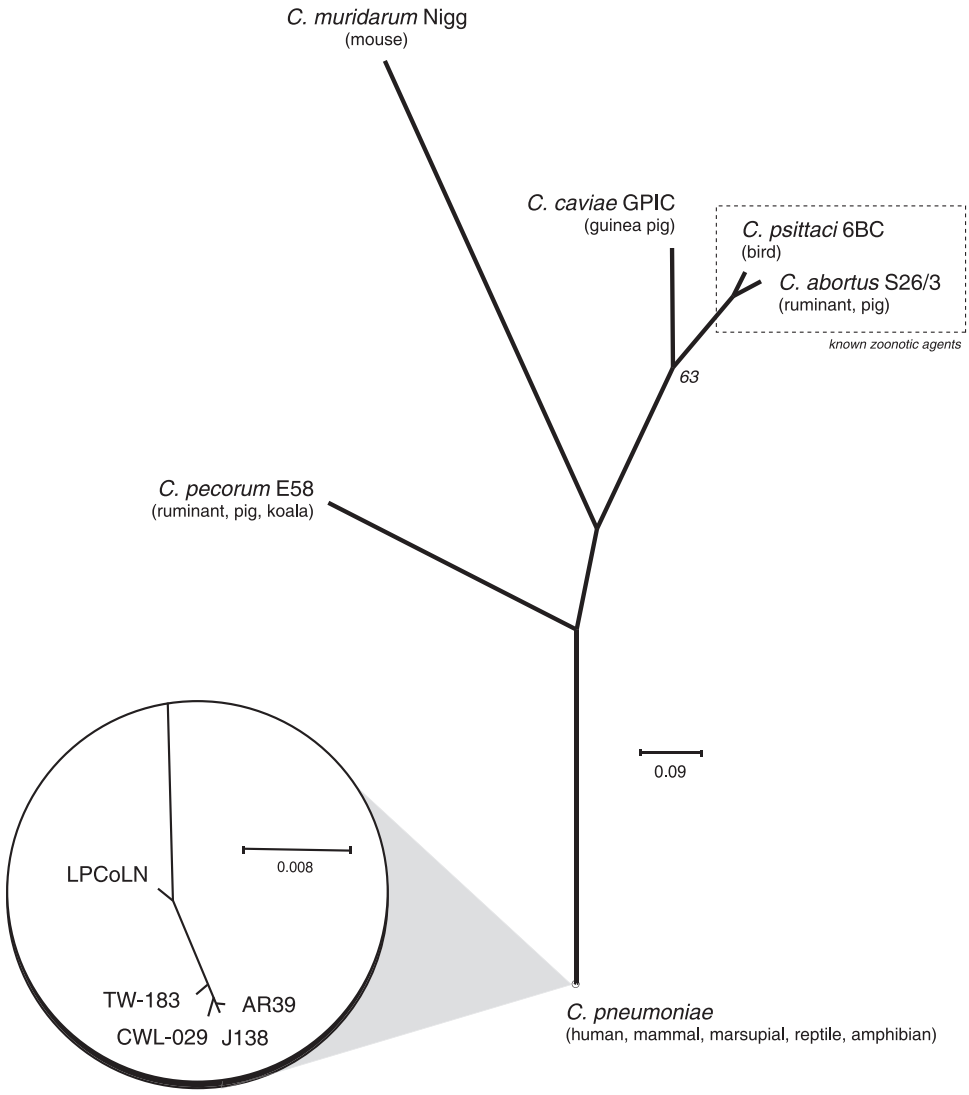


FIG. 3. Phylogeny of all animal chlamydiae and *C. pneumoniae*, using 111 highly conserved gene clusters. The host range for each species is noted in parentheses; the known chlamydial zoonotic agents are boxed. The bootstrap value at each branch point is 100% unless otherwise indicated.

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 provided 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 containing all eight CDSs found in other chlamydial cryptic plasmids. Mitchell et al. (22) reported a much higher growth rate in vitro for the koala LPCoLN isolate than for the human isolate AR39—it is possible that one or more of the genes present on the cryptic plasmid may account for this higher growth rate.

The most compelling evidence to support the idea 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 and 4). The membrane attack complex (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 toward 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.

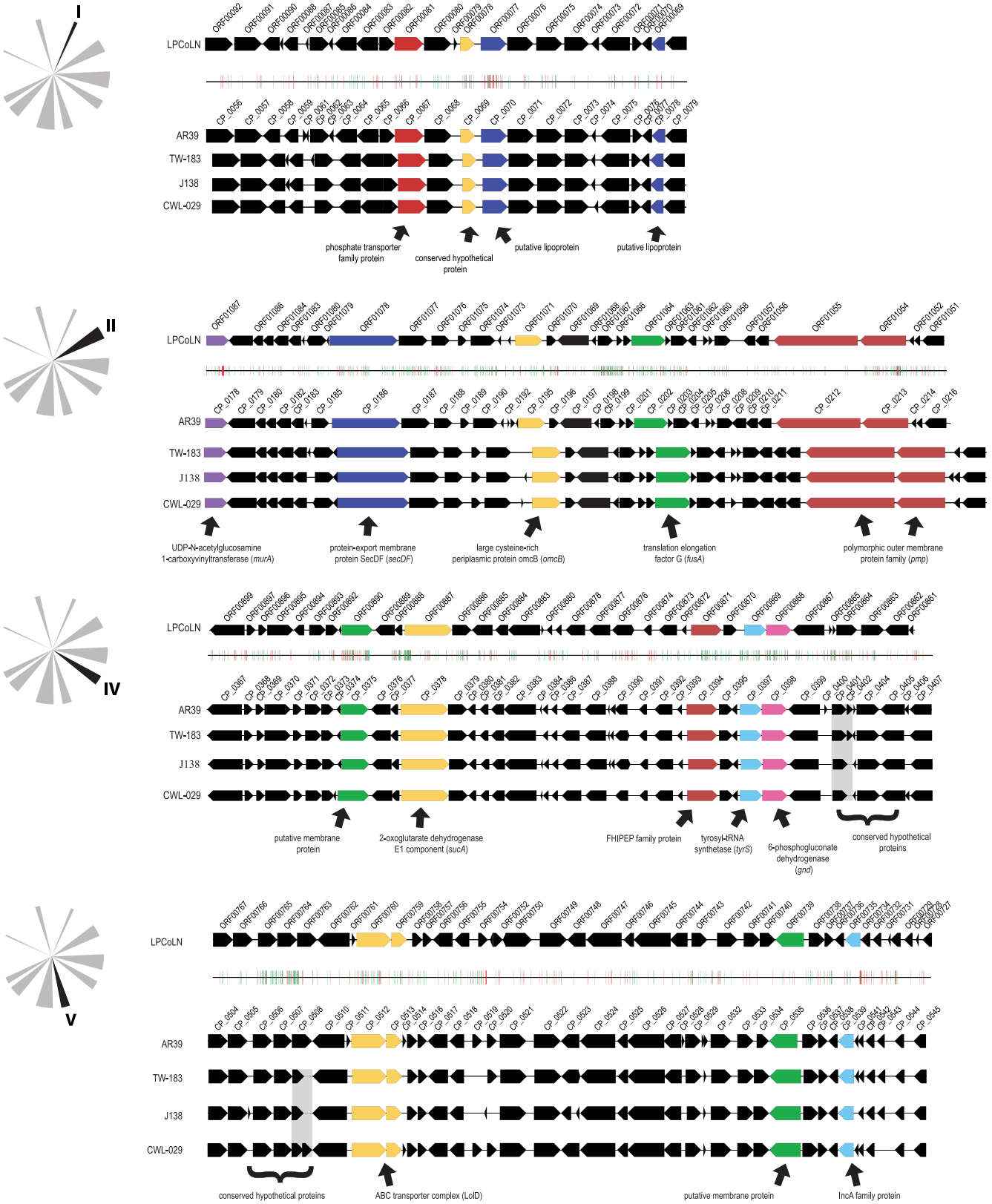


FIG. 4. Annotated detail of additional regions shown in Fig. 1 with high SNP accumulation, showing SNP location and type (synonymous, green; nonsynonymous, red). Colors are used to distinguish notable annotated genes in each region. Gray highlighting shows SNP-associated CDS fragmentation. The icon to the left of each region denotes the approximate location of each region in Fig. 1.

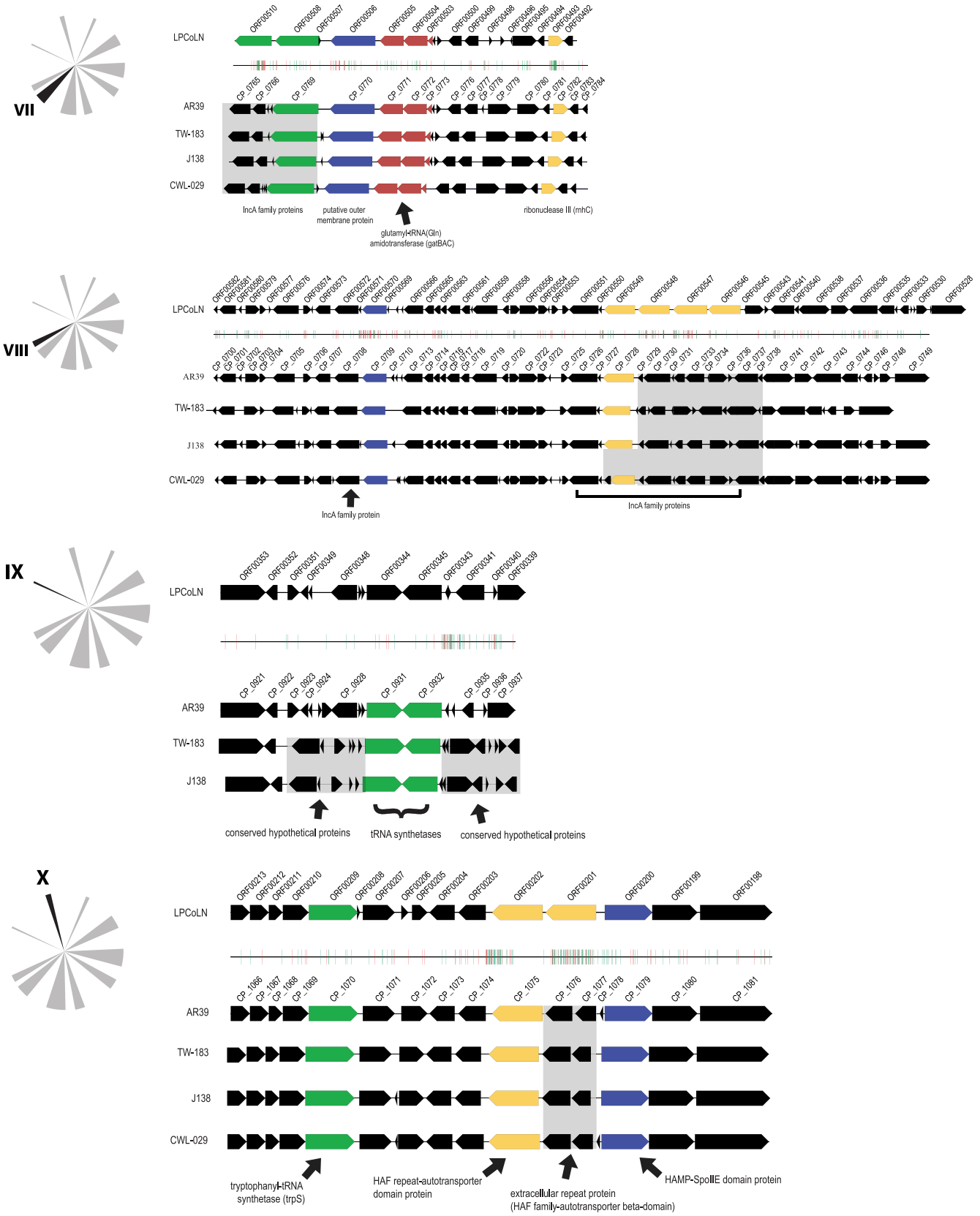


FIG. 4—*Continued*.

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 versus koalas).

Our analysis of the koala *C. pneumoniae* LPCoLN genome sequence, combined with phylogenetic analyses of all *C. pneumoniae* SNPs and the conserved chlamydial CDSs and 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 sequences 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 this host species jump occurred before terminal adaptation and which specific animal host the human-derived *C. pneumoniae* actually originated from cannot be addressed with this set alone. In addition to the full genome sequence from the LPCoLN isolate of koala *C. pneumoniae*, we also obtained and analyzed a second koala *C. pneumoniae* isolate, EBB. This isolate was obtained from a pharyngeal swab from a koala in a wild population from a location geographically separate from that of LPCoLN. Nine genes were sequenced from the EBB isolate, and in all cases the sequences were 100% identical to the sequences obtained from LPCoLN (data not shown).

The koala *C. pneumoniae* LPCoLN isolate has been relatively well characterized with regard 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 shorter doubling time (3.4 to 4.9 h versus 5.9 to 8.7 h) when grown in HEp-2 cell monolayers. Rates of inclusion fusion were also much higher with LPCoLN (100%) than with AR39 (30 to 40%). These biological differences between koala- and human-derived *C. pneumoniae* are consistent with the range of genomic differences that we 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.

## ACKNOWLEDGMENTS

This work was supported by the National Institute of Allergy and Infectious Disease grant 1R01AI051472.

We thank former TIGR and current IGS faculty, the TIGR/IGS Informatics group for expert advice and assistance, and the JCVI Sequencing Facility.

## 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:** 371–380.
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:**2378–2380.
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 host range of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. Syst. Appl. Microbiol. **25:**146–152.
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:**6407–6418.
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:**7894–7897.
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:**2369–2376.
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:**2478–2483.
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:**415–440.
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:**175–185.
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–W559.

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:**160–174.
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:**63–66.
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:**990–993.
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:**530–533.
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:**385–389.
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:**449–456.
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:**407–417.
21. **Longbottom, D., and L. J. Coulter.** 2003. Animal chlamydioses and zoonotic implications. J. Comp. Pathol. **128:**217–244.
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:**91–99.
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:**569–575.
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:**1397–1406.
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:**2134–2147.
29. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16:**276–277.
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:**2311–2314.
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:**754–759.
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:**115–119.
34. **Storey, C., M. Lusher, P. Yates, and S. Richmond.** 1993. Evidence for *Chlamydia pneumoniae* of non-human origin. J. Gen. Microbiol. **139:**2621–2626.
35. **Sutherland, E. R., and R. J. Martin.** 2007. Asthma and atypical bacterial infection. Chest **132:**1962–1966.
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:**500–507.
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:**4876–4882.
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:**161–171.
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:**629–640.
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:**22–27.
41. **Watson, C., and N. J. Alp.** 2008. Role of *Chlamydia* pneumoniae in atherosclerosis. Clin. Sci. (London) **114:**509–531.