Molecular evolution of the Chlamydiaceae

Robin M. Bush1 and Karin D. E. Everett2

Author for correspondence: Karin D. E. Everett. Tel: +1 706 583 0237. Fax: +1 706 542 5771.
e-mail: keverett@calc.vet.uga.edu or kdeeverett@hotmail.com

Phylogenetic analyses of surface antigens and other chlamydial proteins were used to reconstruct the evolution of the Chlamydiaceae. Trees for all five coding genes (the major outer-membrane protein (MOMP), GroEL chaperonin, KDO-transferase, small cysteine-rich lipoprotein and 60 kDa cysteine-rich protein) supported the current organization of the family Chlamydiaceae, which is based on ribosomal, biochemical, serological, ecological and DNA–DNA hybridization data. Genetic distances between some species were quite large, so phylogenies were evaluated for robustness by comparing analyses of both nucleotide and protein sequences using a variety of algorithms (neighbour-joining, maximum-likelihood, maximum-parsimony with bootstrapping, and quartet puzzling). Saturation plots identified areas of the trees in which factors other than relatedness may have determined branch attachments. All nine species were clearly differentiated by distinctness ratios calculated for each gene. The distribution of virulence traits such as host and tissue tropism were mapped onto the consensus phylogeny. Closely related species were no more likely to share virulence characters than were more distantly related species. This phylogenetically disjunct distribution of virulence traits could not be explained by lateral transfer of the genes we studied, since we found no evidence for lateral gene transfer above the species level. One interpretation of this observation is that when chlamydiae gain access to a new niche, such as a new host or tissue, significant adaptation ensues and the virulence phenotype of the new species reflects adaptation to its environment more strongly than it reflects its ancestry.

Keywords: Chlamydia, Chlamydophila, Chlamydiales, phylogeny, intracellular bacteria

INTRODUCTION

Bacteria in the family Chlamydiaceae are obligately intracellular parasites that infect a diverse array of vertebrates. Chlamydiae cause a wide variety of health problems, including spontaneous abortion in livestock, systemic disease in birds and both endemic and zoonotic infection of humans (Herring, 1993; Everett & Andersen, 1997). In humans, chlamydiae are the leading cause of preventable blindness and sexually transmitted disease and have suspected links to cardiovascular disease (Campbell et al., 1998; Schachter, 1999). The evolutionary processes responsible for this diverse array of virulence phenotypes have long been of interest. Previous evolutionary studies of the Chlamydiaceae involved phylogenetic reconstruction using rRNA genes (Fig. 1a) (Everett et al., 1999a) and the gene for the major outer-membrane protein (MOMP) (Fitch et al., 1993). Examinations of chlamydial mechanisms of disease, virulence determinants and genome structure have produced a rich set of sequence data for additional coding genes (reviewed by Stephens, 1999; also Kalman et al., 1999; Read et al., 2000). We have used the available data to study the extent to which the evolutionary history of genes in the family Chlamydiaceae explains the distribution of virulence phenotypes among chlamydial species.

This paper is dedicated to the late Jan Ursing, Associate Editor of IJSB, who was extraordinarily helpful with our paper on reclassification of Chlamydiaceae.

Abbreviations: INDELS, insertions and deletions; LGT, lateral gene transfer; ML, maximum-likelihood; MOMP, major outer-membrane protein; MP, maximum-parsimony; NJ, neighbour-joining; QP, quartet puzzling.

The GenBank accession numbers for the scanned and new data produced in this study are AF269256–AF269282 and AF240773.
Prior to the availability of molecular and genetic phylogenies, our ability to study virulence within an evolutionary context was severely limited. The intracellular lifestyle of chlamydiae obscures phenotypic characters and causes difficulties in isolation, culture and identification. In the absence of genetic analysis, only a small set of biochemical, physiological, morphological, serological and DNA–DNA hybridization data can be used to distinguish chlamydial species (reviewed by Everett et al., 1999a). Those characters allowed the identification of only four chlamydial species (Fig. 1b). However, several of the four species clearly encompassed clusters of biologically and ecologically differentiated strains (Herring, 1993), a problem corrected by our recent taxonomic revision (Fig. 1a, Table 1) (Everett et al., 1999a).

Previous efforts to correlate genetic sequence variation with virulence phenotype among serotypically distinguishable chlamydiae were primarily undertaken through analysis of the MOMP gene, a surface antigen (Baehr et al., 1988; Carter et al., 1991; Fitch et al., 1993; Kaltenboeck et al., 1993; Stephens et al., 1987; Zhang et al., 1993). According to Stothard et al. (1998), MOMP sequence variation appears to be a microevolutionary response to immune pressure. Variation in a few Chlamydia trachomatis MOMP sequences has also been explained by invoking lateral gene transfer (LGT) (Fitch et al., 1993; Hayes et al., 1994). Recent studies showing that a large number of sequences in the genomes of both Chlamydia trachomatis and Chlamydophila pneumoniae appear to be distantly related to genes in plants and other organisms has encouraged such hypotheses (Stephens et al., 1998; Wolf et al., 1999). Doolittle (1999) suggests that lateral transfer is a frequently observed method for transferring virulence traits among free-living bacterial species. However, Stiller & Hall (1999) caution that apparent LGT can also be readily inferred from an erroneous phylogeny. The potential for making such an error is great in the chlamydiae, as homoplasy and the lack of appropriate outgroups make chlamydial phylogenetics problematic (Everett et al., 1999a; Herrmann et al., 2000; Pettersson et al., 1997). Thus, one goal of this work is to examine the hypothesis of lateral transfer within the Chlamydiaceae.

The molecular genetic analyses that contributed to the phylogenetic revision illustrated in Fig. 1(a) were limited to the 16S and 23S full-length ribosomal genes. Our current investigation focuses on coding genes for which multiple sequences are available from diverse chlamydial species. These include MOMP, the small cysteine-rich lipoprotein, and the 60 kDa cysteine-rich protein, all components of the bacterial envelope. We have also studied the GroEL chaperonin, which is involved in the stimulation of host inflammatory responses and scarring (LaVerda et al., 1999), KDO-transferase, which is required for synthesis of LPS, a surface-exposed lipopolysaccharide, and an expanded ribosomal dataset. We examine the evolution of these six genetic loci with respect to the distribution of virulence traits in the Chlamydiaceae. We examine the genetic isolation of the newly described chlamydial species.
Table 1. Virulence traits associated with natural infections by *Chlamydiaceae*


<table>
<thead>
<tr>
<th>Chlamydia</th>
<th>Chlamydophila</th>
</tr>
</thead>
<tbody>
<tr>
<td>trachomatis</td>
<td>suis</td>
</tr>
<tr>
<td>Route of entry</td>
<td>Pharynx, eye, genital, rectal</td>
</tr>
<tr>
<td>Disseminated in bloodstream</td>
<td>+</td>
</tr>
<tr>
<td>Asymptomatic persistence in host</td>
<td>+</td>
</tr>
<tr>
<td>Typical host (s)</td>
<td>Humans</td>
</tr>
<tr>
<td>Infected tissues</td>
<td>Eye†, genital, joints, neonate lung, prostate</td>
</tr>
</tbody>
</table>

* Strains or biovars of these chlamydial species are occasionally found in alternate hosts. For example, *Chlamydiophila psittaci*, *Chlamydiophila abortus* and *Chlamydiophila felis* have been found in humans (Schachter *et al.*, 1969; CDC, 1998; Herring, 1993; Jorgensen, 1997), with mortality particularly associated with *Chlamydiophila psittaci* pneumonitis and *Chlamydiophila abortus* abortion. *Chlamydiophila psittaci* has been found in dying tortoises (Vanrompay *et al.*, 1994), a cat (Lipman *et al.*, 1994) and cattle (Cox *et al.*, 1998; Page, 1967); *Chlamydiophila pneumoniae* was associated with disease in a dying giant barred frog in Australia (Berger *et al.*, 1999) and in frogs imported to the USA from Africa (Reed *et al.*, 2000).

† Conjunctiva of the eye.
species and look for evidence that LGT might have influenced the distribution of virulence traits among chlamydial species. Our ultimate goal is to use studies of the molecular evolution of genes involved in immunity and disease to better understand how virulence and survival strategies have evolved among chlamydial species.

**METHODS**

**Sequence acquisition and alignment.** Several sequences were provided prior to release by R. J. Birtles, B. Herrmann, B. Kaltenboeck and F. R. Ruringsawa. Sequences available in GenBank as of December 1999 were obtained for the ribosomal operon, GroEL chaperonin, KDO-transferase, the cysteine-rich lipoprotein, the 60 kDa cysteine-rich protein and MOMP. Use of *Chlamydia trachomatis* MOMP sequences (from the 55 in GenBank) was limited to selected serotypes. Additional MOMP sequences for FP Baker, FP Cello, GPIC and MN Zhang were scanned from the original publications (May et al., 1996; Zhang et al., 1989). Resequencing of the FP Cello and FP Baker MOMP genes showed them to be identical to one another. Portions of the 16S rRNA gene from N16 and all of the 16S rRNA gene from SPFd were resequenced. The lipoprotein genes from B/TW-5/OT and L2/434/BU were also resequenced and minor corrections were made as necessary (see Figures). New MOMP sequences were determined by W. A. Hambly; new lipoprotein, 60 kDa, and some additional MOMP sequences, were determined by K. D. E. Everett. Sequences were assembled using Sequencer data analysis software (Gene Codes). All new sequences were double-stranded data obtained directly from PCR products using automated DNA cycle sequencing and fluorescent dye terminators (Rozenblum et al., 1997). Sequencing and oligonucleotide primer synthesis were performed by the Iowa State University DNA Sequencing and Synthesis Facility, Ames, IA, USA. Bacterial strains and GenBank accession numbers (including multiple strains with identical sequences) are noted in the Figures and documented in GenBank and in original publications. Scanned and new data produced in this study have GenBank accession numbers AF269256–AF269282 and AF240773.

Sequences were aligned using CLUSTAL W (Thompson et al., 1994) and corrected by eye. Outgroups were identified by using BLAST searches (http://www.ncbi.nlm.nih.gov). Saturation plots (see below) were used to identify appropriate outgroup sequences. Excessive homoplasy prevented the use of outgroups, except for rRNA and GroEL. The analyses of three genes (the GroEL chaperonin, the 60 kDa cysteine-rich protein and KDO-transferase) each included a small number of sequences that were substantially shorter than the rest of the sequences in those datasets. These short sequences were included because they were representative of poorly sampled species. Short sequences were used only if their inclusion did not alter the structure of the tree produced using full-length sequences. Sequence alignments can be obtained from R. M. Bush (rmbush@uci.edu) or as supplementary data in IJSEM Online (http://ijsem.sgmjournals.org/).

**MOMP (ompA, ompA**). MOMP is a surface-exposed porin that, upon reduction in the host cell, permits the entry of essential molecules and nutrients and may permit the exit of virulence factors. Genetic variation in the four hypervariable MOMP segments suggests that they are under intense selective pressure by host immune systems (Brunham et al., 1994; Stothard et al., 1998). The aligned *ompA* dataset of 1236 nucleotides (412 residues) contained 58 sequences representing all nine species in the *Chlamydiaceae*. We analysed sequences starting with the N-terminus of the post-translationally processed protein because the 22/23-residue signal sequence data can sometimes include PCR amplification primer sequences. N-terminal data were missing from LW508 (14 residues), LW613 (16 residues), and 66p130 (22 residues), and 12 residues were missing at their C-termini. Skua R54 was missing 30 N-terminal residues and 11 C-terminal residues. *Chlamydia suis* and *Chlamydiophila abortus* sequences numbered A7004873–A7005617 were missing 37 C-terminal residues. The aligned protein dataset was analysed both with and without the four hypervariable segments (Fitch et al., 1993), which were difficult to align. In our aligned dataset these are residues 88–109, 165–190, 254–269 and 322–356.

**GroEL chaperonin (encoded by the gene known as groEL or hypB**). GroEL chaperonin triggers host inflammation and subsequent scarring (Ward, 1995). The aligned *groEL* chaperonin dataset of 1632 nucleotide positions (544 residues) contained 11 sequences representing seven *Chlamydiaceae* species. Three sequences, B577, AR-388 and FCStr, lacked 22 residues on the 5’ end and 25 residues on the 3’ end. The ‘pigeon’ sequence lacked the first 128 residues. We included one outgroup sequence, from *Rhodothermus marinus*, in the analysis.

**KDO-transferase (kdtA; previously gseA**). KDO-transferase catalyses the addition of three 3-deoxy-d-manno-2-octulosonic acid molecules onto lipid A precursors in the synthesis of chlamydial LPS. LPSs are generally endotoxins in Gram-negative bacteria, and chlamydial LPS is a mitogen. However, there is little evidence for chlamydial LPS induced endotoxic shock ( Kosma, 1999), despite evidence for transport of most chlamydial species in the bloodstream (Table 1). The aligned *kdtA* dataset of 1314 nucleotide positions (438 residues) contained 13 sequences representing five species from the *Chlamydiaceae*. Sequences from strains B577, FCStr and AR-388 each lacked 69 residues on the 5’ end and 108 residues on the 3’ end.

**Small cysteine-rich lipoprotein (omp3, envA, omcA, omlA**). The cysteine-rich lipoprotein is an important structural constituent of the chlamydial outer envelope (Hatch, 1996). The aligned small cysteine-rich lipoprotein dataset of 279 aligned nucleotide positions (93 residues) contained seven sequences representing five species from the *Chlamydiaceae*.

**Large cysteine-rich 60 kDa protein (omp2, ompB, envB, omcB, cnbB**). The 60 kDa cysteine-rich protein is a periplasmic structural constituent of the chlamydial outer envelope (Hatch, 1996). The aligned 60 kDa cysteine-rich protein dataset at 1677 nucleotides (559 residues) contained 14 sequences representing six species from the *Chlamydiaceae*. Four sequences, KC, N16, E. and Koala lacked 380 of the 559 residues. These sequences began at residue 12 and ended around residue 191.

**Ribosomal operon and ribosomal construct.** Full-length sequence data for the entire ribosomal operon are available for at least two strains in each of six species. Data for only one full-length operon is available from *Chlamydia caviae*, *Chlamydiophila felis* and *Chlamydiophila abortus*. The sequence data for isolate N16 are not full-length, beginning at position 53. Two different ribosomal datasets were analysed. Each was a subset of the full-length, contiguous 16S and 23S rRNA data. The dataset we termed the ‘ribosomal operon’ was a contiguous dataset containing 1236 nucleotides in the 16S rRNA and 1240 nucleotides in the 23S rRNA.
Molecular evolution of the Chlamydiaceae

2708 aligned nucleotide positions that started at position 8 of the 1566 bp 16S rRNA gene (Escherichia coli numbering), continued through the 16S/23S intergenic spacer, and included the most diverse segment (the first 851 bases) of the approximately 3000 bp 23S rRNA gene. We also analysed a dataset called the ‘ribosomal construct’, which was the ‘ribosomal operon’ with the genetically heterogeneous (and thus difficult to align) intergenic spacer removed. (There was almost no sequence identity between the Chlamydiaceae and outgroup families in the intergenic spacer.) This permitted an analysis of 20 chlamydial isolates representing all nine Chlamydiaceae species. Outgroup sequences were obtained and used for three recently described families in the Chlamydiales: Parachlamydiaceae, Simkaniaaceae and Waddliaceae.

Phylogenetic analysis. Phylogenetic trees were constructed using PAUP* version 4.0b4 (Swofford, 2000). A variety of algorithms were used: quartet puzzling (QP), maximum-likelihood (ML), maximum-parsimony (MP) and neighbour-joining (NJ). MP and NJ analyses were run 10 times, randomizing in each run the order in which the sequences were input. All analyses except MP were run using both nucleotide sequences and the corresponding amino acids for coding genes. The QP nucleotide trees are shown in the Figures, and variation among trees produced in different analyses is described in the results. Reliability values from the QP analysis are shown for each node when greater than or equal to 75%. These values are the percentage of times a node was seen in 1000 of the intermediate trees produced in the QP routine (Strimmer & von Haeseler, 1996). For MP analyses, the tree bisection–reconnection branch swapping option of the heuristic search routine was used. MP bootstrap values are shown parenthetically with corresponding reliability values when the bootstrap values were less than or equal to 75%. The phylogenetic distribution of insertions and deletions (INDELs) was examined by eye (Gupta, 1998).

Genetic differentiation of chlamydial species. Distinctness ratios contrast the genetic distances between groups of isolates with respect to the variation within those groups (Palys et al., 1997). When a distinctness ratio calculated using one sequence or gene differs greatly from ratios calculated using other sequences or genes, the data should be examined for recombination events, strain contamination or misidentification of isolates. A distinctness ratio is the ratio of the mean between-taxon genetic divergence to mean within-taxon divergence. The mean within-taxon divergence for a species pair is the mean of the two within-taxon divergence values. Mean distinctness ratios were calculated across all pairwise comparisons of nucleotide sequences for all six loci. Gapped positions were not included in these calculations. According to the criteria of Palys et al. (1997), distinctness ratios of two or greater are sufficient to define genetically distinct species. Distinctness ratios were calculated for the MOMP gene and the ribosomal operon with and without the four variable segments and the intergenic spacer, respectively.

Saturation analysis. Saturation plots were used to evaluate the degree of homoplasy in each dataset. A saturation plot compares the phylectic and percentage pairwise genetic distances between pairs of isolates (Vuillaumier et al., 1997). The percentage pairwise genetic distance between two sequences was calculated as the percentage of non-gapped nucleotide positions at which the two sequences differed. The phylectic distance used was the ML estimate of genetic distance between two isolates produced in the QP analyses, in nucleotide substitutions per site. When pairwise and phylectic distances are identical, the slope of a regression line drawn through the points on a saturation plot is linear and has a slope of 1. Phylectic distances increase more rapidly than pairwise distances when multiple nucleotide changes per position occur over time. As the frequency of multiple mutations per position increases, the correlation between the phylectic and pairwise distance declines. When no correlation remains (as the slope approaches zero), mutational saturation is said to have occurred, and processes such as long-branch attraction (Felsenstein, 1978) rather than relatedness might be determining the attachment of branches between distantly related clades. Saturation plots were useful for evaluating the appropriateness of potential outgroups for the Chlamydiaceae and for evaluating homoplasy within the Chlamydiaceae as well.

RESULTS

MOMP

The unrooted MOMP QP tree contained sequences from all nine species in the Chlamydiaceae (Fig. 2). The reliability value from the QP analysis provided 100% support for the separation of Chlamydia and Chlamydophila. All analyses showed Chlamydophilus abortus evolving from, rather than as a sister clade to, Chlamydophilus psittaci. All of the nodes within the Chlamydiaceae depicted by curved lines in Fig. 1(a) were difficult to resolve in the MOMP analysis. Different topologies were produced, depending on the algorithm, on whether full-length data were used or variable segments were removed, and on whether nucleotide or protein data were used. This was most likely due to excessive homoplasy, as the phylectic distances were substantially larger than pairwise genetic distances between the genera and even for some of the between-species comparisons (Fig. 3, shown with the MOMP variable segments removed). Evidence that homoplasy might be interfering with phylogenetic reconstruction through long branch attraction was most apparent under maximum-parsimony. Using MP, the three Chlamydia species (Chlamydia suis, Chlamydia trachomatis and Chlamydia muridarum) did not form discrete clades when the variable segments of the MOMP genes were included in the analysis. When separate analysis of each genus was performed under MP, Chlamydia suis, Chlamydia trachomatis and Chlamydia muridarum were clearly differentiated.

GroEL chaperonin

The GroEL chaperonin QP tree (Fig. 4) was congruent with the consensus tree in Fig. 1(a) except that it lacked sequences for Chlamydia suis and Chlamydophilus felis. The GroEL chaperonin was the least genetically differentiated of the coding genes. The outgroup, Rhodothermus marinus, was as similar to the Chlamydiaceae in analysis of GroEL (58% at the nucleotide level) as Chlamydia and Chlamydophilus were to each other in analysis of KDO-transferase (Fig. 3). Phylogenetic reconstruction of the relationships between species was probably not affected by homoplasy, as
Fig. 2. MOMP phylogeny. QP tree constructed using the ompA gene, which encodes the MOMP. The four variable segments were excluded from this analysis. ATCC reference numbers and GenBank accession numbers are provided for each strain. QP reliability values are shown on each branch when greater than or equal to 75%, MP bootstrap values at these nodes are shown in parentheses if less than or equal to 75%. † = type strain. * A22 data were a personal communication from Alan Herring; † = Dl/Cal-8 is identical to Da/TW-448 (X62921) and D/B-120 (X62928); ‡ the reliability value for this node was only 58%, while the bootstrap value was 91% (this likely reflects long-branch attraction, as QP and MP analyses using only isolates from the genus Chlamydia showed 100% reliability and bootstrap values for nodes separating the three Chlamydia species); § MoPn sequence M64171 differs from the other MoPn sequences by 2 bases.
without the outgroup all analyses produced the same tree structure, and the phyletic distances were comparable to pairwise genetic distances (Fig. 3). However, analyses including the *R. marinus* outgroup varied according to the algorithm or type of data used. The QP analysis and the MP analysis of amino acids rooted the *Chlamydiaceae* so that the two genera were monophyletic lineages. However, MP analysis of nucleotide data and NJ analysis gave a variety of other results. Saturation analysis suggested that *R. marinus* was probably too distantly related to serve as a reliable outgroup (Fig. 3).

**KDO-transferase**

The unrooted KDO-transferase QP tree (Fig. 4) included only five *Chlamydiaceae* species but was otherwise congruent with the consensus tree in Fig. 1(a). KDO-transferase sequences were the most genetically differentiated data in this study. Saturation analysis suggested that homoplasy might affect phylogenetic reconstruction of the family even without the presence of outgroups in the analysis (Fig. 3). The relationships among species did not vary when different phylogenetic algorithms were used. However, the use of different algorithms affected apparent relationships among *Chlamydia trachomatis* strains and caused variation in which the *Chlamydia trachomatis* isolate was apparently closest to the root of the *Chlamydia trachomatis* clade, a result typical of long-branch attraction.

**Small cysteine-rich lipoprotein**

The unrooted QP tree for the small cysteine-rich lipoprotein was congruent with the consensus tree in Fig. 1(a), except that it contained only five species (Fig. 4). The tree showed *Chlamydophila caviae* as a sister taxon to *Chlamydophila pneumoniae*. This probably not have occurred had a representative of *Chlamydophila pecorum* been present in the dataset (this configuration had zero bootstrap support in MP analysis). Results did not vary when different phylogenetic algorithms were used.
Large cysteine-rich 60 kDa protein

The unrooted QP tree constructed using the gene for the large cysteine-rich 60 kDa protein contained six species (Fig. 4) and was otherwise congruent with the consensus tree (Fig. 1a). The differentiation of species and the order in which species diverged were consistent using the various algorithms. The saturation plot indicated some homoplasy in the dataset for this gene (Fig. 3). The *Chlamydia trachomatis* isolate that was closest to the root of the *Chlamydia trachomatis* clade, and the relationships among the *Chlamydia trachomatis* strains, varied among the different analyses. This pattern, also seen with KDO-transferase, is typical of long-branch attraction.

Ribosomal operon

The ribosomal operon was the only genetic locus for which sequences from all nine *Chlamydiaceae* species and closely related outgroups were available (Fig. 5). The QP ribosomal tree was congruent with the consensus tree previously determined using other algorithms (Fig. 1a). Support was present but not strong in any analysis for pairing *Chlamydophila pecorum* and *Chlamydophila pneumoniae* as sister taxa. QP did not resolve the order of divergence of *Chlamydia pneumoniae* and *Chlamydia felis*, but otherwise provided a branching order that had good statistical support regardless of whether the intergenic spacer was included (Fig. 5 shows the tree constructed...
Molecular evolution of the *Chlamydiaceae*

**Fig. 5.** Ribosomal phylogeny. QP tree using a ribosomal construct containing the 16S rRNA and the first 851 bases of the 23S rRNA gene. ATCC reference numbers and GenBank accession numbers are provided for each strain. QP reliability values are shown on each branch when greater than or equal to 75%. MP bootstrap values at these nodes are shown in parentheses if less than or equal to 75%. † = type strain; *including new or resequenced data.

Without the spacer. The three outgroup sequences were as genetically distant from one another as they were from the family *Chlamydiaceae*. *Waddlia* and *Parachlamydia* were sister taxa in all analyses. The saturation plot for this ribosomal segment showed a close correspondence of pairwise and phyletic
Table 2. Distinctness ratios for species in the Chlamydiaceae

The ratio of between-taxon to within-taxon divergence (Palys et al., 1997) for the ribosomal construct and for ompA without the four variable segments.

<table>
<thead>
<tr>
<th>Species 1</th>
<th>Species 2</th>
<th>Ribosomal construct</th>
<th>ompA (MOMP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean divergence</td>
<td>Mean divergence</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>(±SE) within</td>
<td>(±SE) within</td>
<td>divergence</td>
</tr>
<tr>
<td></td>
<td>species 1</td>
<td>species 2</td>
<td>between/within</td>
</tr>
<tr>
<td>Chlamydophila</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abortus</td>
<td>0.003</td>
<td>0.005 ± 0.002</td>
<td>0.009 ± 0.002</td>
</tr>
<tr>
<td>psittaci</td>
<td>0.003</td>
<td>0.003</td>
<td>0.023 ± 0.001</td>
</tr>
<tr>
<td>felis</td>
<td>0.005 ± 0.002</td>
<td>0.003</td>
<td>0.021 ± 0.002</td>
</tr>
<tr>
<td>abortus</td>
<td>0.003</td>
<td>0.003</td>
<td>0.020 ± 0.001</td>
</tr>
<tr>
<td>psittaci</td>
<td>0.005 ± 0.002</td>
<td>0.003</td>
<td>0.020 ± 0.002</td>
</tr>
<tr>
<td>caviae</td>
<td>0.003</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>abortus</td>
<td>0.003</td>
<td>0.001</td>
<td>0.043 ± 0.001</td>
</tr>
<tr>
<td>psittaci</td>
<td>0.005 ± 0.002</td>
<td>0.001</td>
<td>0.043 ± 0.001</td>
</tr>
<tr>
<td>felis</td>
<td>0.003</td>
<td>0.001</td>
<td>0.049 ± 0.001</td>
</tr>
<tr>
<td>caviae</td>
<td>0.001</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>abortus</td>
<td>0.003</td>
<td>0.010</td>
<td>0.045 ± 0.002</td>
</tr>
<tr>
<td>psittaci</td>
<td>0.005 ± 0.002</td>
<td>0.010</td>
<td>0.043 ± 0.003</td>
</tr>
<tr>
<td>felis</td>
<td>0.003</td>
<td>0.010</td>
<td>0.048 ± 0.001</td>
</tr>
<tr>
<td>caviae</td>
<td>0.010</td>
<td>0.050 ± 0.001</td>
<td></td>
</tr>
<tr>
<td>pneumonieae</td>
<td>0.001</td>
<td>0.046 ± 0.001</td>
<td></td>
</tr>
<tr>
<td>Chlamydia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trachomatis</td>
<td>0.002 ± 0.001</td>
<td>0.002</td>
<td>0.025 ± 0.001</td>
</tr>
<tr>
<td>muridarum</td>
<td>0.002 ± 0.001</td>
<td>0.014</td>
<td>0.029 ± 0.001</td>
</tr>
<tr>
<td>trachomatis</td>
<td>0.002 ± 0.001</td>
<td>0.014</td>
<td>0.030 ± 0.000</td>
</tr>
</tbody>
</table>
Table 3. INDEL analysis of the Chlamydiaceae

Full-length gene alignments were included, with the exception of the ribosomal intergenic spacer and variable segments of MOMP. There were no INDELs in the GroEL chaperonin.

<table>
<thead>
<tr>
<th>Groups above the species level:</th>
<th>Total</th>
<th>60 kDa protein</th>
<th>Small lipoprotein</th>
<th>KDO-transferase</th>
<th>MOMP construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differentiated the two genera</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouped Chlamydia psittaci and</td>
<td>5</td>
<td>1</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Chlamydia pneumoniae</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouped Chlamydia abortus and</td>
<td>3</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia psittaci</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouped Chlamydia abortus and</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia psittaci and Chlamydia caviae</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouped Chlamydia abortus and</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia psittaci and Chlamydia felis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal:</strong></td>
<td><strong>12</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiated species:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia caviae</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia pecorum</td>
<td>3</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Chlamydia pneumoniae</td>
<td>2</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia suis</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia muridarum</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Subtotal:</strong></td>
<td><strong>19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiated isolate(s) within species:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia pneumoniae (Koala)</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia abortus (OCLH196)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Chlamydia pecorum (several isolates)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Subtotal:</strong></td>
<td><strong>3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconsistent with phylogeny:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia, without Chlamydia felis and</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia caviae</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed isolates from Chlamydia trachomatis and</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Chlamydia suis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal:</strong></td>
<td><strong>2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>36</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

distances, even between Chlamydiaceae and these outgroups (Fig. 3). Different algorithms produced the same tree structure except for the order of divergence of Chlamydia caviae and Chlamydia felis.

Genetic differentiation of species

Distinctness ratios calculated for nearly all genes and species were greater than or equal to 2.0, the value suggested as appropriate for species-level differentiation (Palys et al., 1997). The only exception was the Chlamydiaceae species with variable segments removed, which was 1.9. Table 2 shows these ratios for MOMP and for the ribosomal construct, the two datasets that included all nine species. When variable segments or intergenic spacer segments were included in the MOMP or ribosomal analyses, respectively, the results in Table 2 changed very little. Distinctness ratios were, on average, smaller for MOMP than for the ribosomal construct. This was because within-species divergence (the denominator of the distinctness ratio) was about ten times greater for MOMP than for the ribosomal construct, while variation between species (the numerator) using MOMP was only about five times greater than for the ribosomal construct.

INDEL analysis

INDEL analysis (Table 3) suggested an evolutionary pattern that was almost entirely consistent with the reconstruction of chlamydial evolution shown in Fig. 1(a). The GroEL chaperonin lacked INDELs. The other five loci contained a total of 36 INDELs, 34 of which were consistent with the phylogeny in Fig. 1(a). Two INDELs in the ribosomal construct differentiated the two genera. Nineteen INDELs in four genes differentiated six of the nine species. The clade containing Chlamydia pecorum and Chlamydia caviae...
pneumoniae was differentiated by five INDELs in three genes. The Chlamydia abortus/Chlamydia psittaci clade shared one INDEL in KDO-transferase and two in the small lipoprotein, and these two species shared an INDEL with Chlamydia felis in the MOMP gene and with Chlamydia caviae in the small lipoprotein. The additional three INDELs were strain specific. The two INDELs that were not concordant with Fig. 1(a) were both in the ribosomal construct. One INDEL grouped subsets of strains from Chlamydia trachomatis and Chlamydia suis. The other grouped Chlamydia felis and Chlamydia caviae with the genus Chlamydia. The ribosomal intergenic spacer and the variable segments of the MOMP gene contained large numbers of INDELs, but these segments were so difficult to align that we could not confidently draw inference from their INDELs. The INDELs in these hypervariable segments often differentiated isolates into clusters that were inconsistent with any reasonable phylogeny.

**DISCUSSION**

This study of the evolutionary history of five coding genes in the Chlamydiaceae (Figs 2 and 4) showed all genes evolving in concert with the ribosomal genes (Fig. 5), albeit at very different rates. The evolutionary structure of the family consisted of two monophyletic lineages, each containing deeply divergent clusters of sequences. Distinctness ratios calculated for each cluster met or exceeded the recommended values for species (Table 2) (Palys et al., 1997). The congruence of the individual gene trees, the support of these trees by analysis of INDELs, and the clear genetic separation of similarity clusters shown by distinctness ratios argue against the presence of interspecies recombination and incorrect strain or species identification in these data. We could not discern the order of divergence in a few places on our trees (nodes shown using curved lines on Fig. 1a), due to homoplasy; however, it was relatively easy to validate which nodes were robust by comparing results obtained by varying the phylogenetic algorithms, by saturation analysis, and by using both nucleotide and protein sequence data. Clustering of isolates by genetic relatedness was consistent with a recent taxonomic revision of the family Chlamydiaceae (Everett et al., 1999a). This revision was multi-faceted, being based on biochemical, ecological and DNA-DNA hybridization data; however, the molecular phylogenetic analyses contributing to the revision were limited to ribosomal sequence data (Everett et al., 1999a). This ribosomal phylogeny, rooted with sequences from three new families in the Chlamydiaceae for both the full-length 16S rDNA and the full-length 23S rDNA for all species, showed the Chlamydiaceae radiating into two genera and a total of nine species. Thus, these results corroborate the recent phylogenetic revision, illustrated in Fig. 1(a), and are consistent with the proposed standard of 95%, 90% and 80% ribosomal sequence identity for separating genera, families and the order Chlamydiales, respectively (Everett et al., 1999a). This phylogeny can now be used as a foundation upon which to examine the evolution of additional genes, investigate evolutionary processes responsible for differences in virulence phenotype among species, and study the rate and direction of evolution of the important coding genes analysed here. The five coding genes varied greatly in their rates of evolution, as can be seen by comparing the genetic distances separating the two genera (Fig. 6). The most rapidly evolving gene encodes KDO-transferase, an enzyme that synthesizes LPS. The most slowly evolving coding gene expresses GroEL chaperonin, which triggers host inflammation and subsequent scarring (Ward, 1995). The genes evolving at intermediate rates code for MOMP, cysteine-rich lipoprotein and cysteine-rich 60 kDa protein. These genes, which are important structural constituents of the chlamydial envelope, have no known homologues in other bacterial species.

**The evolution of virulence**

One of the most striking observations about evolution within the Chlamydiaceae has been the reported lack of correlation between virulence phenotype and phylogenetic relatedness (Fitch et al., 1993; Stothard et al., 1998). However, those studies were based on a phylogeny that was not yet well resolved or on only a limited set of species, respectively. To determine if an improved phylogeny with more appropriate species designations explains the evolution of chlamydial virulence, we mapped the most commonly surveyed virulence characters from the literature onto the new chlamydial phylogeny (Table 1). This is also illustrated in Fig. 1(a) using host specificity traits. Our goal was to infer character states for the ancestors of the nine extant species using parsimony analysis, hoping to learn something about the processes driving the diversification within this group. Could chlamydial access to a host have occurred at the time of host speciation? The molecular phylogeny of Chlamydiaceae hosts was not at all congruent with the molecular phylogeny of Chlamydiaceae pathogens. Furthermore, there is not an appropriate fossil record that would provide a standard of measure for the divergence of chlamydiae relative to their hosts (Haag et al., 1998; Clark et al., 1999). Analysis of other virulence traits in Table 1 showed only a modest degree of phylogenetic pattern. Chlamyphila pneumoniae, Chlamyphila pecorum and Chlamyphila psittaci, for example, have all been isolated or identified from brain tissue of infected hosts, and both Chlamyphila pneumoniae and Chlamyphila psittaci cause systemic infections. Recent data for Chlamyphila pneumoniae in amphibia suggests that it may eventually be found to be as promiscuous as Chlamyphila psittaci. There is also increasing evidence that some chlamydiae reside in amoebae and may be widely distributed in the soil and water, certainly a confounding element in the analysis (Fritsche et al., 1998; Horn et al., 2000).
Our results suggest that closely related *Chlamydiaceae* species are no more likely to share a host, or any of the other virulence traits listed in Table 1, than are distantly related chlamydial species. One explanation that has been offered for such observations is possible lateral transfer of genes encoding virulence traits. There is evidence for LGT between the lymphogranuloma venereum and trachoma biovars of *Chlamydia trachomatis*, as can be seen by comparing the placement of L1, L2 and L3 in the MOMP tree (Fig. 2) with their placement in the cysteine-rich 60 kDa protein tree (Fig. 5). However, we found no evidence for LGT above the species level in any gene. Thus, LGT would only explain the lack of congruence between host and pathogen phylogenies if it occurred at genetic loci other than those studied here. LGT above the species level in any gene.

Despite these genetic differences, the only known exchange could occur. Some evidence suggests that gene transfer can occur within amoebae, but experimental co-infection of amoebae with different species of chlamydiae has not been attempted, to our knowledge.

An alternative explanation to LGT and co-evolution that explains the disjunct phylogenetic distribution of host specificity is that considerable adaptation occurs in chlamydiae which successfully occupy new niches (host or host tissue), and that this has led to species formation. The virulence phenotype of the new species might then bear little similarity to those of the ancestors or nearest relatives. This is the same conclusion drawn by Fitch *et al.* (1993), who noted that the picture that resulted from their study, which was similar to ours but used only MOMP, the 60 kDa protein, and a limited collection of virulence traits, "is more one of pathogen versatility than coevolutionary constraint".

**Differences between the two genera**

While there is no single virulence trait that distinguishes the two chlamydial genera and the processes driving the initial split are unknown, there is nonetheless a clear genetic separation of the genera. The monophyly of the two genera is supported by all eight genes that have been studied, as shown in our results and in those of previous studies (Everett & Andersen, 1997; Everett *et al.*, 1999a; Fitch *et al.*, 1993; Herrmann *et al.*, 2000; Kaltenboeck *et al.*, 1993; Pudiatmoko *et al.*, 1997; Takahashi *et al.*, 1997; Tanner, 1999; Zhang & Perlman, 1995). However, the conditions required for one chlamydial species to take up DNA from another seem fairly prohibitive, requiring transport across one, two or three membranes. DNA transfer could occur only by a host cell being infected by both species, or by the host cell first taking up DNA from one species (e.g. by pinocytosis) while it was concurrently infected with a second, or by viral (phage) transfer of genomic sequences. Chlamydophages have been identified in *Chlamydophila psittaci*, *Chlamyphila abortus*, *Chlamyphilis caviae* and *Chlamyphilis pneumoniae* species (Bavoil *et al.*, 2000; Everson *et al.*, 2000), but these do not have incorporated genomic sequences. Researchers who have attempted to co-infect host cells with two different species of chlamydiae found that intracellular inclusions did not merge (Matsumoto *et al.*, 1991), and hence there was no common vacuole in which DNA
biochemical marker that separates the genera is glycogen particles, which can be detected only in *Chlamydia trachomatis, Chlamydia suis* and *Chlamydia muridarum* (Moulder et al., 1984; Rogers et al., 1996; Rogers & Andersen, 1996). In over 60 years of research, reliable biological markers that fulfil the plus/minus criteria of numerical classification methods have not been found for *Chlamydiaceae* species or genera. Thus, our ability to discern the evolutionary processes behind their divergence is limited to genetic data. One feature that does emerge from these results is that the genus *Chlamydia*, with its many serotypes and nearly a dozen genetically distinct isolates of *Chlamydia suis*, is less genetically diverse than *Chlamydophila* and appears to be less variable with respect to host range. This may well be an artifact of sampling, as until recently it was believed that humans were the almost exclusive host to this lineage. The prevalence of *Chlamydia* in swine is disturbing and suggestive of potential health risks to humans. Why *Chlamydia suis, Chlamydia trachomatis* and *Chlamydia muridarum* should have a common ancestry is not obvious, based on their apparent host specificities. However, swine, birds and humans exchange zoonotic viruses that play clear roles in influenza evolution (reviewed by Scholtissek, 1995).

**Future directions**

Thus far, characterization of the *Chlamydiaceae* has been primarily confined to humans, food animals and companion animals. With improved identification and characterization of chlamydiae and specific targeting using PCR and other DNA-based assays, a wider host range can be tested for the presence of these pathogens (e.g. Everett & Andersen, 1997; Everett et al., 1999a, b). At the present time there is genetic evidence for over 98 new lineages of chlamydia-like bacteria (Ossewaarde & Meijer, 1999; Meijer et al., 2000). There is morphological evidence for chlamydia-like bacteria in alligators, bivalves, chameleons and fish (Cajaraville & Angulo, 1991; Groff et al., 1996; Homer et al., 1994; Huchzermeyer, 1997; Jacobson & Telford, 1990; Szakolczai et al., 1999). Further study will show whether traits such as species specificity are artifacts resulting from difficulty in laboratory isolation, from test specificity, or from limitation of studies to particular hosts or diseases. Insight into the evolution of virulence and pathogenesis of these species will help us plan strategies for countering their infectivity. This dataset provided an unusual opportunity to examine the parallel evolution of functionally distinct and evolutionarily important genes within a genetically diverse lineage. Understanding the evolutionary history of these genes is the first step toward advanced molecular evolution studies which will examine factors affecting relative rates of evolution, identify amino acid substitutions that alter protein structure or function, characterize covariation between sequence positions, and seek out evidence of positive selection on individual codons (Bush et al., 1999a, b).

**ACKNOWLEDGEMENTS**

We thank Walter M. Fitch for assistance with some of the sequence alignments, Wendy A. Hambly for sequence analyses, James W. Moulder and Arthur A. Andersen for comments on the manuscript, and Arthur A. Andersen for his support.

**REFERENCES**


