

Research article

Open Access

Intrastrain and interstrain genetic variation within a paralogous gene family in *Chlamydia pneumoniae*

Wasna Viratyosin^{1,3}, Lee Ann Campbell², Cho-Chou Kuo² and Daniel D Rockey*¹

Address: ¹Department of Microbiology, Oregon State University, Corvallis, 97331-3804, USA, ²Department of Pathobiology, University of Washington, Seattle, 98195, USA and ³Present Address: National Center for Genetic Engineering and Biotechnology, 113 Paholyothin Rd., Klong 1, Klong Luang, Pathumthani 12120, Thailand

Email: Wasna Viratyosin - viratyow@onid.orst.edu; Lee Ann Campbell - lacamp@u.washington.edu; Cho-Chou Kuo - cckuo@u.washington.edu; Daniel D Rockey* - rockyd@orst.edu

* Corresponding author

Published: 2 December 2002

Received: 27 August 2002

BMC Microbiology 2002, 2:38

Accepted: 2 December 2002

This article is available from: <http://www.biomedcentral.com/1471-2180/2/38>

© 2002 Viratyosin et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: *Chlamydia pneumoniae* causes human respiratory diseases and has recently been associated with atherosclerosis. Analysis of the three recently published *C. pneumoniae* genomes has led to the identification of a new gene family (the Cpn 1054 family) that consists of 11 predicted genes and gene fragments. Each member encodes a polypeptide with a hydrophobic domain characteristic of proteins localized to the inclusion membrane.

Results: Comparative analysis of this gene family within the published genome sequences provided evidence that multiple levels of genetic variation are evident within this single collection of paralogous genes. Frameshift mutations are found that result in both truncated gene products and pseudogenes that vary among isolates. Several genes in this family contain polycytosine (polyC) tracts either upstream or within the terminal 5' end of the predicted coding sequence. The length of the polyC stretch varies between paralogous genes and within single genes in the three genomes. Sequence analysis of genomic DNA from a collection of 12 *C. pneumoniae* clinical isolates was used to determine the extent of the variation in the Cpn 1054 gene family.

Conclusions: These studies demonstrate that sequence variability is present both among strains and within strains at several of the loci. In particular, changes in the length of the polyC tract associated with the different Cpn 1054 gene family members are common within each tested *C. pneumoniae* isolate. The variability identified within this newly described gene family may modulate either phase or antigenic variation and subsequent physiologic diversity within a *C. pneumoniae* population.

Background

Chlamydia pneumoniae is an obligate intracellular bacterium that infects and causes disease in the respiratory tract [1,2] and has recently been associated with heart disease [3]. Approximately 10% of pneumoniae cases and 5% of bronchitis and sinusitis cases in the U.S. are attributed to

C. pneumoniae infection. Pathogenic mechanisms utilized by *C. pneumoniae* to replicate and disseminate within hosts remain unclear.

Little is known about strain-specific determinants of *C. pneumoniae*. Isolates of *C. pneumoniae* are virtually indis-

tinguishable using 16s rRNA [4], restriction fragment length polymorphism [5], and amplification fragment length polymorphism analysis [6]. Unlike *C. trachomatis*, only a single serotype or genotype of *C. pneumoniae* has been identified by any of the above methods.

Recently, three genomes of *C. pneumoniae* have been completed and published. These include CWL029 [http://chlamydia-www.berkeley.edu:4231/\[7\]](http://chlamydia-www.berkeley.edu:4231/[7]), AR39 [http://www.tigr.org/\[8\]](http://www.tigr.org/[8]), and J138 [http://w3.grt.kyushu-u.ac.jp/J138/\[9\]](http://w3.grt.kyushu-u.ac.jp/J138/[9]). Comparative analysis suggests that overall genomic organization and gene order in each *C. pneumoniae* genome is highly conserved [8,9]. Given this conservation, the study of individual regions of sequence variation will provide insight into strain-specific virulence, genetic diversity, and adaptive responses within and among *C. pneumoniae* populations.

Genomic analyses have recently revealed a large gene family of 21 polymorphic outer membrane proteins (Pmps) with predicted outer membrane localization in *C. pneumoniae* [7,10,11]. The function of this gene family in chlamydial growth and development remains unknown. Several studies have examined genetic variation and strain differentiation of Pmp proteins, which may be important for genetic flexibility and adaptive response. Recently, it has been reported that interstrain and intrastrain variation of gene expression and protein productions of *pmpG6* and *pmpG10* are modulated by deletion of tandem repeats in *pmpG6* [9,11] and variation in the length of polyguanosine tract in *pmpG10* [12,13]. This evidence suggests that variation may be an important requisite for the function of this gene family in the biology of *Chlamydiae*.

Examination of the *C. pneumoniae* genome sequences by Daugaard et al. [14] demonstrated that a unique and related family of genes is found within the *C. pneumoniae* genome, and that variation among strains leads to differences within several members of the gene family. Gene products of these paralogous genes contained a unique bi-lobed hydrophobic domain, which is a predictive marker for localization to the inclusion membrane [15]. In this study, we further characterize this family by examining variation in sequence of the family members both within and among different *C. pneumoniae* isolates.

Results

Bioinformatic analysis of the Cpn 1054 gene family

The Cpn 1054 gene family consists of 11 *C. pneumoniae*-specific genes or gene pairs: Cpn 07, 08/09, 010/010.1, 011/012, 041/042, 043/044, 045/046, 0124/0125, 0126,1054 and 1055/1056. Family members indicated with slashes have deletions or internal shifts in reading frame that interrupt otherwise related coding sequences (Figure 1). The size of the paralogous repeat unit ranges

from approximately 1,500 bp (1055/1056) to approximately 2,800 bp (Cpn 007). Family members are located in four separate regions on the chromosome including contig 1.0–1.1 (Cpn 007–Cpn 011–012), contig 1.5–1.6 (Cpn 041–Cpn 046), contig 2.5 (Cpn 0124–126) and contig 13.0–13.1 (Cpn 1054–Cpn 1056). The nucleotide sequence identity of the Cpn 1054 gene family members ranged between 20–99% (Figure 1, Table 1). Although in some cases the overall sequence similarity among family members is low, many share extensive regions of identity at the predicted 5' end (Figure 2). This region of identity includes a polyC tract within each paralogous locus. The CWL029 genome sequence has predicted GTG codons as a translational initiation site for several family members (Cpn 08, 010, 041, 045 and 1055; Figure 2).

The conserved hydrophobic domain of the Cpn 1054 gene family

In a previous study, we demonstrated that a bi-lobed (50–60 amino acid) hydrophobic domain is a predictive marker for protein localization to the inclusion membrane (Inc proteins; [15]). Consequently, 68 *C. pneumoniae* proteins, including several members of the Cpn 1054 family (Figure 3), were identified as putative inclusion membrane-associated proteins. The nucleotide sequence found at the 5' end of each paralogous repeat unit represents the region of highest identity among most family members (Figure 2). This includes the sequences that encode the major hydrophobic domain of each repeat unit (Figure 4A). In some cases (Cpn 007, 124, 126) the hydrophobic domain represents the primary reason the gene is included in the Cpn1054 gene family (Figure 4A,4B; Figure 5). Three nearly identical copies of this domain are found within Cpn 007 (Figure 3), a protein that is otherwise not similar to other family members.

Homopolymeric cytosine (poly C) tract and the variations of the length of poly C tracts in the Cpn 1054 gene family

Genomic analysis of the Cpn 1054 gene family revealed a repeat sequence of 6 to 15 homopolymeric cytosine residues, positioned either upstream or immediately within the predicted 5' end of Cpn 008, 010, 041, 043, 045, 1054, and 1055 (Figure 2). Analysis of each sequenced genomes indicated that three Cpn 1054 gene family members had identical polyC tracts in each of the respective genomes: Cpn 041 (CCCCCTCCCC), Cpn 043 (12C), and Cpn 045 (CCCCCC). In each of the other family members however, the length of the polyC tract varies in the published genomes. These analyses were expanded to examine the polyC tracts in several clinical isolates. Sequence data collected directly from amplification products of several Cpn 1054 family members showed that polymorphisms within the polyC tract were present in all tested family members except Cpn 041 and Cpn 045 (not shown). Therefore, the polyC tract in Cpn 041 and Cpn

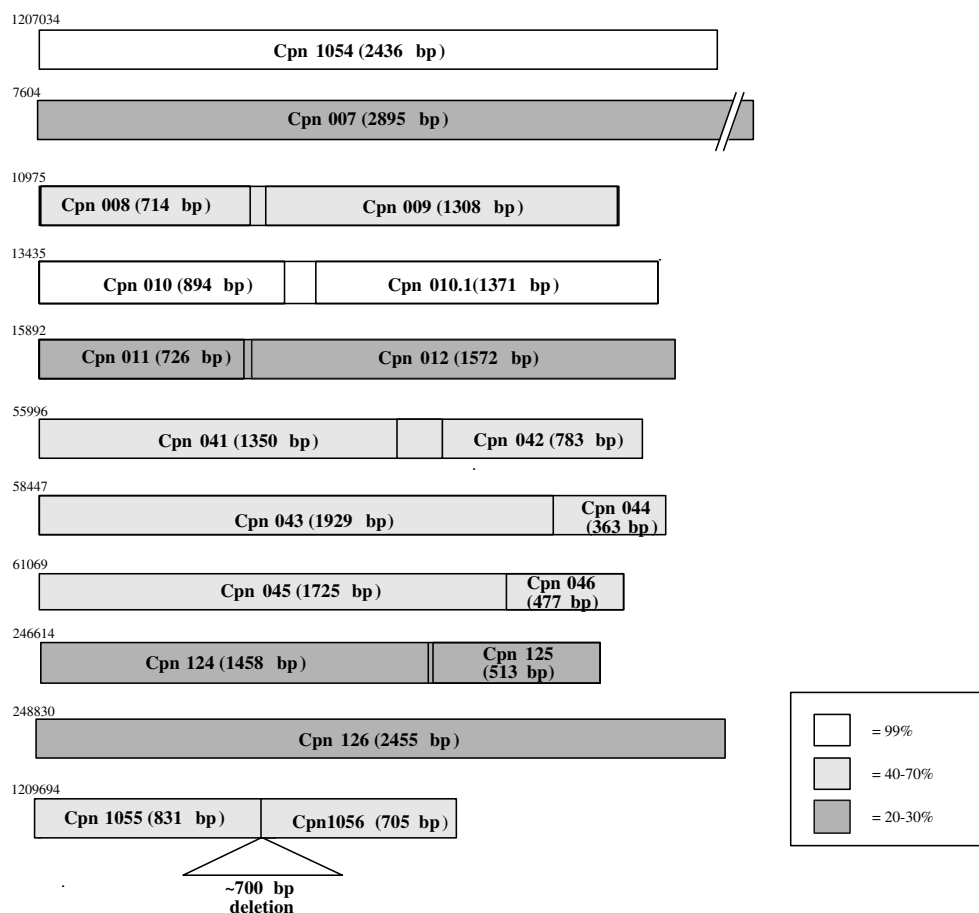


Figure 1

A comparison of the paralogous genes of the Cpn 1054 gene family in *Chlamydia pneumoniae* CWL029. The 5' end of Cpn 1054 is similar to Cpn 008, 010, 011, 043, 045, and 1055 while the 3' end was similar to the Cpn 009, 010.1, 012, 042, 044, 046, and 1056. The range of predicted amino acid similarity is 20–99%, as indicated by shading. Vertical bars indicate termination codons that interrupt the reading frame. The position of the beginning of each family member in the CWL029 genome is indicated above the sequence. All numbering is based on the CWL029 genome sequence [7].

045 appear to be conserved among isolates, a result consistent with the data in the genome sequences. In order to examine the clonal variation within other Cpn 1054 family members, sequences surrounding the polyC tracts were amplified from selected genes and the products cloned into plasmids. The polyC tract was then examined through nucleotide sequence analysis of a selection of independent plasmid constructs. Sequence variation within the polyC tract was first examined for Cpn 043, Cpn 1054, and Cpn 1055 of *C. pneumoniae* AR39. Sequence analysis of independent recombinant plasmids showed variability in the length of the polyC tract of each tested gene (Figure 6A). We next examined variation within a single gene, Cpn 1055, using genomic DNAs from strains AR39, AR 458 and PS 32 as template. Variation in the length of

polyC was observed in this gene from each tested isolate (Figure 6B).

Two approaches were used to demonstrate that the observed variation in length of the polyC tract was not a function of PCR errors during the analysis. First, two different thermostable polymerases (*Taq* and *Pwo* polymerase) were used to generate the primary amplification products for cloning and subsequent sequence analysis. Amplifications with each enzyme resulted in clones with variation in the length of the polyC tract (Figure 6C). A second approach for examination of the possibility of PCR errors was to reamplify the polyC tract from a single plasmid template, and examine the sequence of the polyC tract directly in these amplification products. No variation

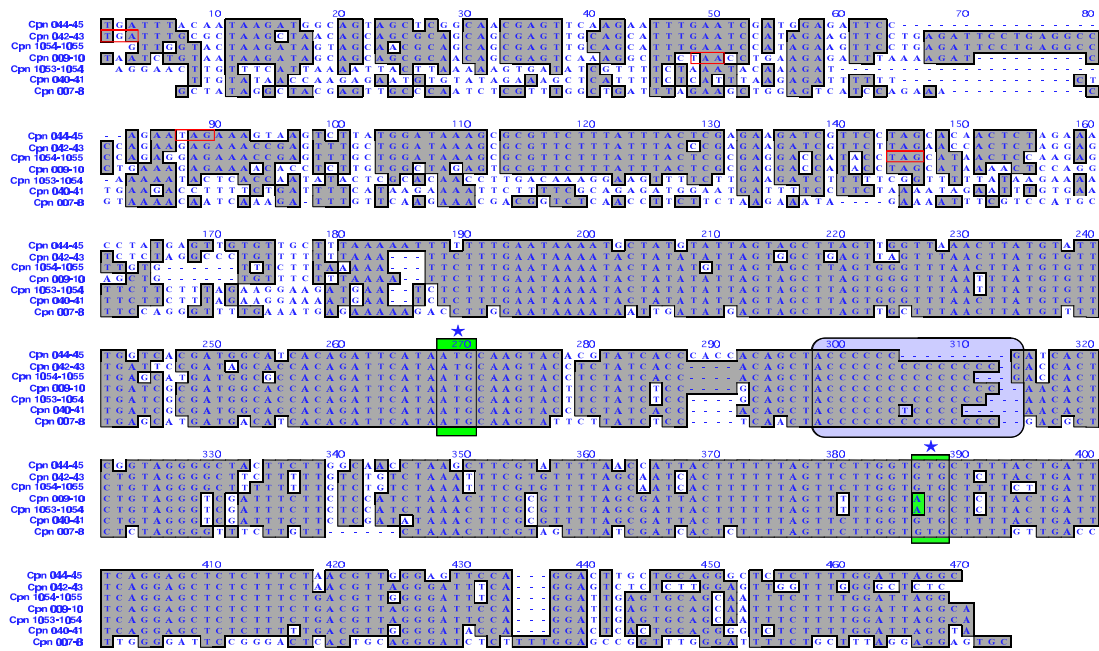


Figure 2

Multiple sequence alignments of the 5' end and upstream regions of selected Cpn 1054 family members. Three genes have been left out of this figure (Cpn 007, 124, 126) because they lack significant sequence identity at the 5' end. All data were obtained from the CWL029 genome sequence [7]. The designation for each sequence includes the gene number of the 1054 family member and the gene directly upstream. The position in the genomic contig is also indicated for each sequence. Where appropriate, the predicted stop codon of upstream genes (Cpn 009, 042, 044, 1054; red boxes) is indicated. Two regions encompassing candidate start codons for each gene are indicated with green boxes. In some cases this includes the codon GTG as a candidate translational start site, as predicted through the Chlamydia Genome Website. PolyC tracts, boxed in blue, are upstream of the predicted start site or are within the gene near the 5' end of coding region.

Table 1: Percent DNA sequence similarity of selected paralogous genes in the Cpn 1054 gene family^a.

010/10.1	81					
041/42	78	82				
043/44	76	76	77			
045/46	85	81	76	88		
1054	81	99	81	76	81	
1055/56	82	89	84	86	85	85
Gene	008/9	010/10.1	041/42	043/44	045/46	1054

^aPairwise analysis of the nucleic acid sequence performed using the BLAST 2 program (20). Family members Cpn 007, Cpn 111/12, Cpn 124/125, or Cpn 126 were not included in this table because the overall alignment of paralogs was less than 20%.

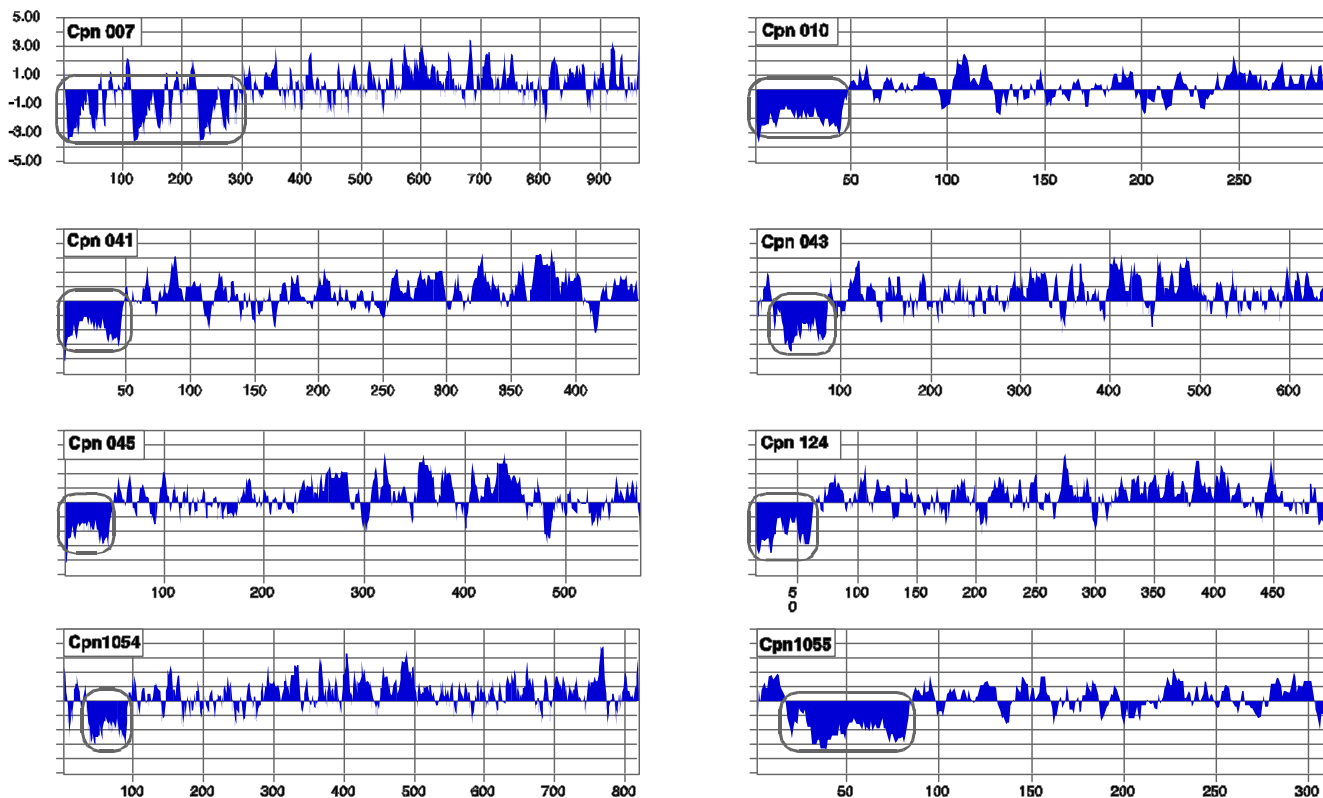


Figure 3

Hydrophilicity profiles of selected members of the Cpn 1054 gene family using the Kyte-Doolittle method with a sliding window of 7. The scale of hydrophilicity is indicated for Cpn 007 and the scale is identical in all panels. The bi-lobed hydrophobic domain is boxed in each profile. Note that the scale on the horizontal axis varies for each protein.

of the length of the polyC tract of Cpn 043, 1054, and 1055 was identified in these PCR products (not shown). These results support the conclusion that the variability in the length of the polyC tracts is not an artifact of the amplification process, and thus the observed variability reflects differences at these loci within individual isolates.

Allelic differences within Cpn 010–010.1

Analysis of published genomic sequence of several 1054 family members identified polymorphisms within several genes. Two examples of these polymorphisms can be found in Cpn 010/10.1 and Cpn 1054, family members that are over 98% identical at the nucleotide level [14]. Analysis of the published genome sequences indicates

that these two genes likely duplicated through nonreciprocal exchanges or by gene conversion [16]. Comparative analysis of the three genomes showed that the short sequence polymorphisms are present in Cpn 010/10.1, Cpn 043, and Cpn 1054 [14]. Cpn 10/10.1 and Cpn 1054 are especially interesting, as the two genes differ by only a few nucleotides over a 2,400 base pair coding sequence. The primary differences between these genes are 1) a single nucleotide polymorphism (SNP) that alters the reading frame in 10/10.1 [14], and, 2) a region of diversity at the 3' end of each gene (Figure 7). Dugaard et al. [14] identified an RFLP marker that exploits the SNP, and other candidate RFLPs can be identified in the 3' region of diversity (Figure 7). Short sequence polymorphisms were ex-

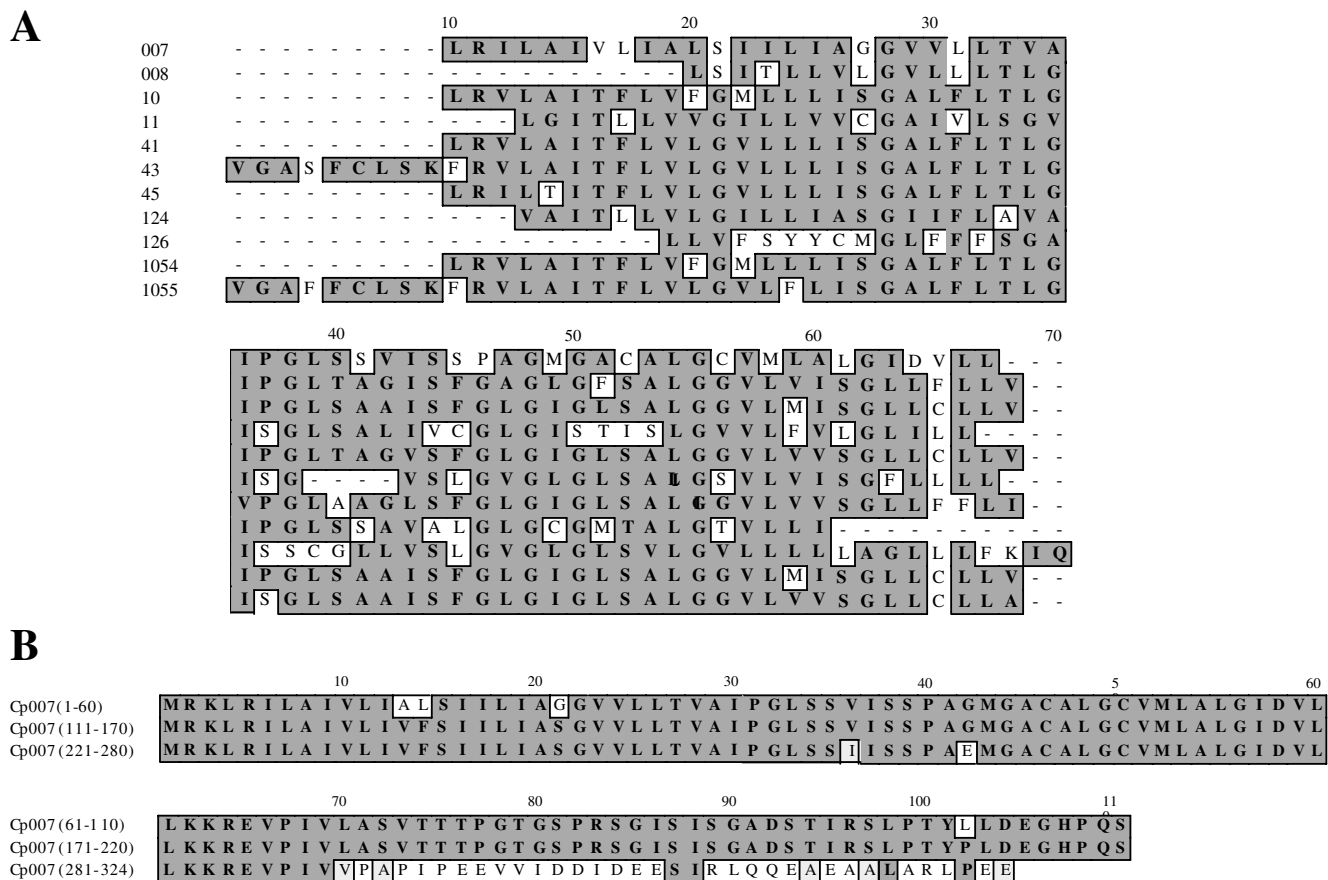


Figure 4

The conserved hydrophobic domain in the Cpn 1054 gene family. Panel A: Multiple sequence alignments of the hydrophobic domains of the Cpn 1054 gene family using CLUSTALW analysis. Similarity among the amino acid sequence is indicated by the grey regions. Panel B: Alignment of the three tandem hydrophobic domains at the amino terminus of Cpn 007. The sequences in grey are identical among this set of three tandem hydrophobic domains.

amined in Cpn 010/10.1 by PCR amplification and nucleotide sequencing in 12 *C. pneumoniae* isolates. Variability at both polymorphisms was observed between strains. The frameshift mutation leading to the truncated sequence of Cpn 010 was identified in CWL029 and 4 other isolates: AR388, AR231, KA5C and KA66 (not shown). The 3' polymorphism that distinguishes Cpn 10/10.1 from Cpn 1054 was found within Cpn 10.1 in two isolates: AR39 and AC43 (Figure 7). There was no apparent linkage between the central frameshift mutation and the 3' polymorphism, as the two regions varied independent of one another. There was also no evidence that these sequences vary within individual *C. pneumoniae* isolates.

Discussion

Relatively little is known about molecular pathogenesis, genetic diversity and adaptive strategy of *C. pneumoniae*. Although the genomic organization of these independent strains is very similar (over 99.9% identical), there are regions of variation within each isolate [8]. In the present study, we have identified a paralogous gene family within *C. pneumoniae*, designated as the Cpn 1054 gene family. This family consists of eleven paralogous loci, with single repeat elements consisting of single ORFs or ORF pairs. The identity of the predicted polypeptide sequences shared among family members ranges from 20–99%. It is likely that the diversity of these genes arose through gene

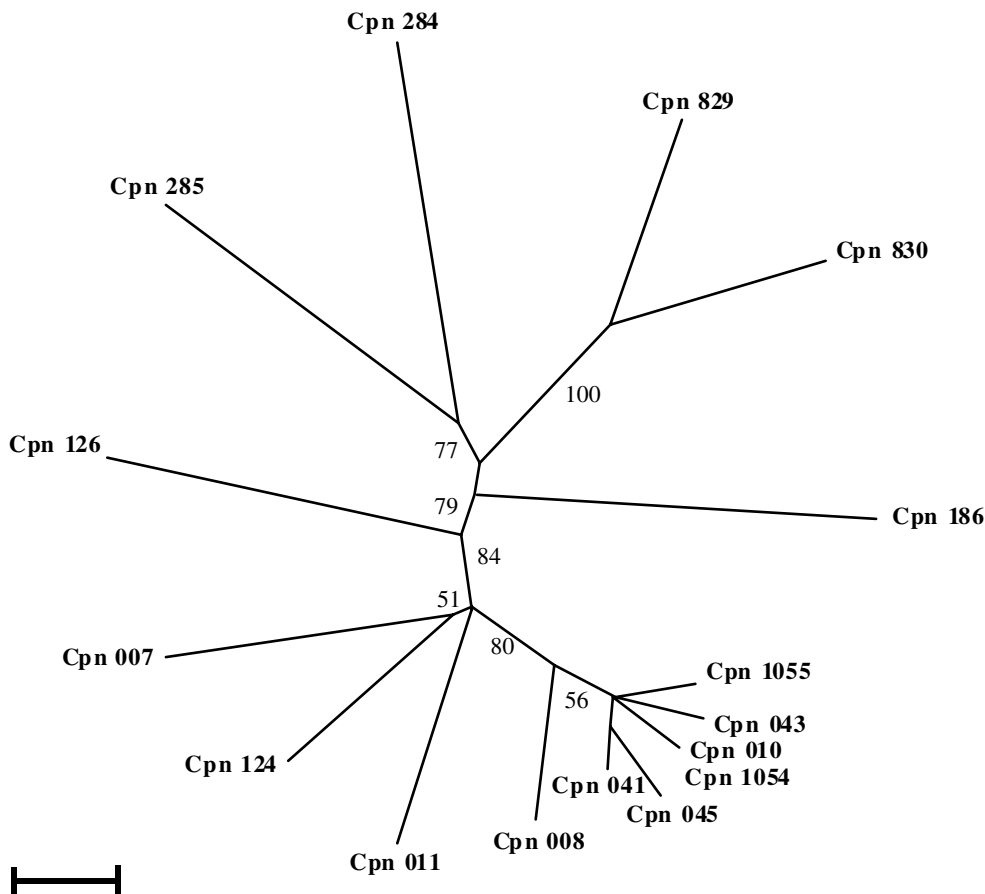


Figure 5

Phylogenetic analysis of the amino acid sequence of the hydrophobic domains of the Cpn 1054 gene family obtained from *C. pneumoniae* CWL029. Hydrophobic domains of Cpn 186, 284, 285, 829, and 830 are included in this figure as outliers. The outliers represent candidate Inc proteins that contain a bilobed hydrophobic domain similar to that encoded by Cpn 1054 gene family members, but that share no obvious amino acid similarity with family members. Multiple sequence alignments were performed and analyzed using CLUSTALW and PAUP* programs. Distance matrices were inferred by neighbor-joining to estimate evolutionary distances. Bootstrap values (based on 1000 replicates) are represented at each node. Scale bar indicates 0.1 change per nucleotide.

duplication and subsequent diversification. It appears that certain duplications were relatively recent, as at least two of the repeated loci- Cpn 010/10.1 and Cpn 1054- are nearly identical. Analysis of the three genomes also demonstrates that apparent gene conversion has occurred between 10/10.1 and 1054 in strain AR39 [16], and that an intact 1054 ORF is found within each sequenced genome [14]. However, its location varies between the two loci. The redundant nature of the Cpn 1054 family members is somewhat unusual against the generally reductive evolutionary strategy of the chlamydiae [17]. There is no evidence that the Cpn 1054 gene family is found outside of

C. pneumoniae, and thus the members of the family may be important in the unique biological traits of this species.

Gene duplication and subsequent genetic drift are the likely means by which variation is manifested between members of the Cpn 1054 gene family. Variation is also observed within individual family members, both between strains [14] and, as shown in this report, within individual isolates. Several gene family members, including Cpn 008, 010, 043, 1054, and 1055, contain homopolymeric cytosine repeats either upstream or at the predicted 5' end of the coding region. In *C. pneumoniae*, variation of the short repeat of homopolymeric nucleotides was first

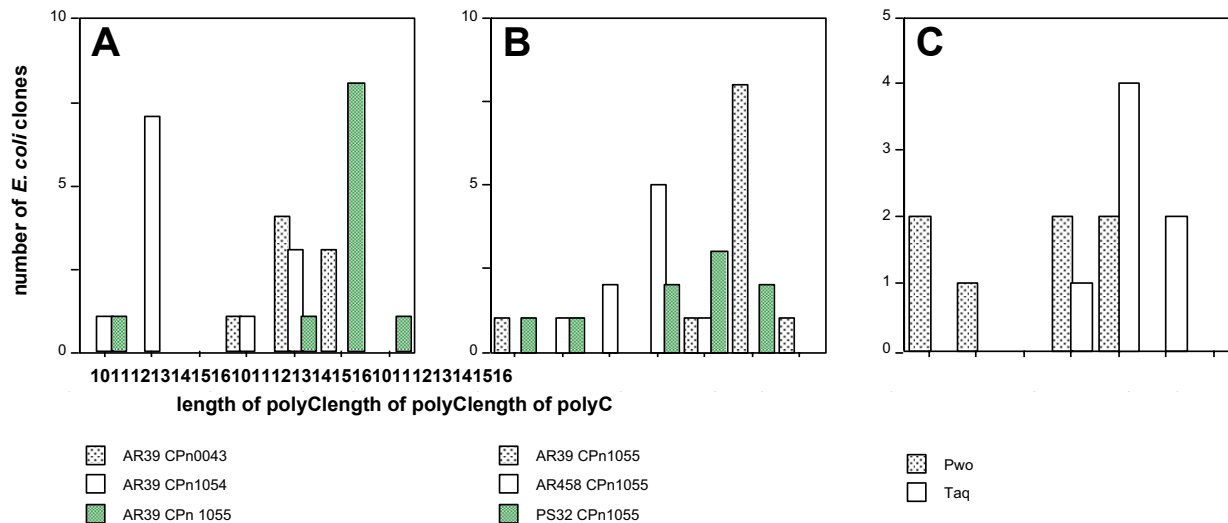


Figure 6

Variation of the homopolymeric cytosine (polyC) tract in the Cpn 1054 gene family in *C. pneumoniae*. In each panel the vertical axis indicates the number of individual clones with a specific length of polyC. The horizontal axis indicates the number of cytosines within an individual polyC tract. (A) Intrastrain variation of the length of the polyC tracts of Cpn 043, 1054, 1055 within strain AR39. (B) Variation of the length of the polyC tract in Cpn 1055 within AR39, AR 458 and PS32. (C) The variation of the length of the poly C tract within Cpn 1055 of AR39 identified in recombinant clones using *Taq* and *Pwo* DNA polymerases.

identified in the *pmp* family. Comparative genomic analysis and cloning expression showed that the length of the polyG tract of *pmpG* 10 varies between strains and within an isolate [12,13]. Furthermore, variation of the length of polyG has been demonstrated that it plays a role in the differential expression of *PmpG* 10 [12]. Variability in short nucleotide repeats generated via slipped strand mispairing are key elements in the generation of phenotypic diversity within many pathogenic microorganisms [18]. Further investigation will be required to determine if the expression of members of the Cpn 1054 gene family is affected by the observed variability in length of the polyC tract.

Although the proteins in the Cpn 1054 gene family are classified as candidate inclusion membrane proteins [15], their subcellular locations and role in infection and disease remain to be identified. It is also not yet known whether the Cpn 1054 gene family is expressed individually or coordinately, or to what extent each gene is expressed during the course of an infection. However, the variation, both within and between strains, is a potential requisite for this gene family that may contribute to the unique biology of *C. pneumoniae*.

Conclusions

The *C. pneumoniae* genome contains a gene family (the Cpn 1054 gene family) consisting of 18 different genes in 11 paralogous loci. Variation is observed both within and among isolates. This variation may be useful for the biotyping of *C. pneumoniae* clinical isolates, and may be important in phenotypic diversity within the species.

Methods

Bacterial strains, plasmids and *C. pneumoniae* genomes

All experiments were conducted using a collection of independent clinical isolates from a strain library (Table 1). Genomic DNA of *C. pneumoniae* was isolated from purified EBs using the methods of Campbell et al. [19]. Extracted DNA was stored at -20C. Genomics analyses were conducted using sequences from the genome websites listed in the introduction. Open reading frames, nucleotide positions and contig numbering were annotated based upon the *C. pneumoniae* CWL029 genome [7].

Bioinformatics analysis of the Cpn 1054 gene family of *C. pneumoniae* genomes

DNA and polypeptide sequences were aligned using CLUSTALW analysis (Mac Vector™ 6.0; Oxford Molecular, Genetics Computer Group, Inc. Madison, WI). Each gene

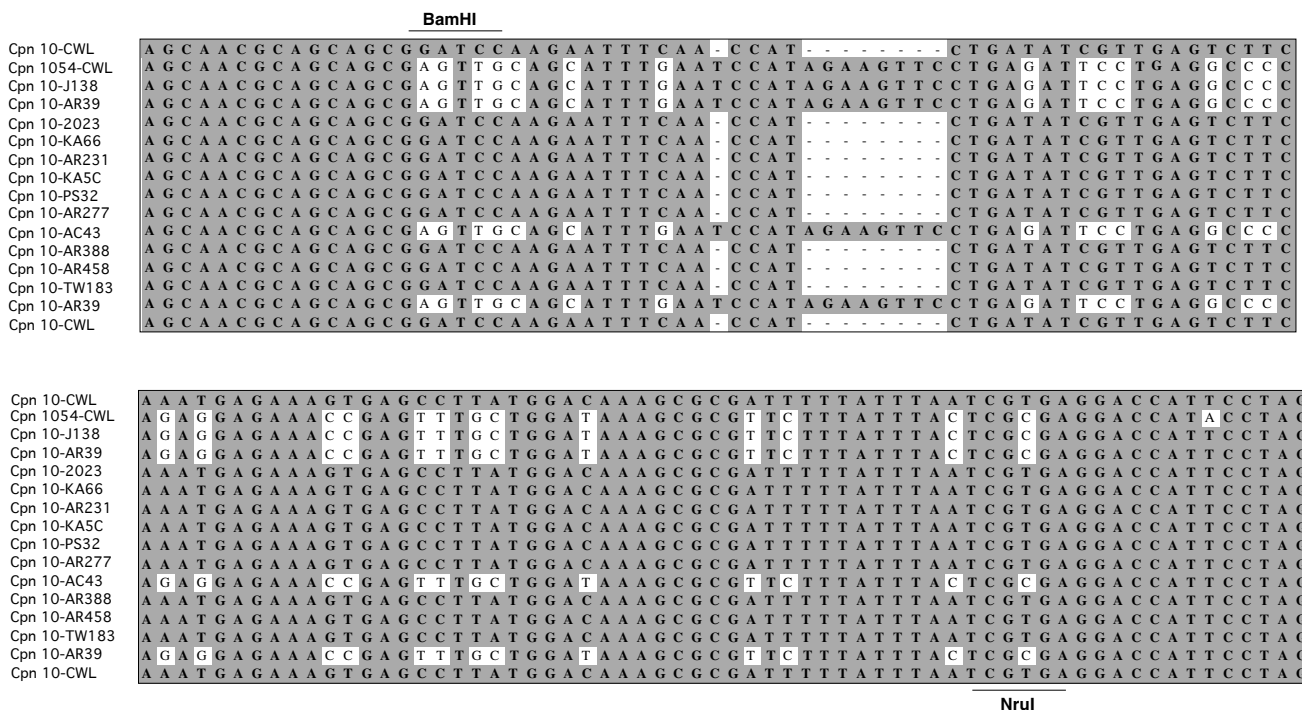


Figure 7
 Sequence analysis of Cpn 10.1 and Cpn 1054 from the genome sequence databases (First four sequences) and from nucleotide sequence analysis conducted in this study. The different strains are indicated to the left of each sequence. The data are continuous from the top line to the bottom, and represent the terminal 130–140 nucleotides of each open reading frame. Each sequence ends at the termination codon for the indicated gene. For reference purposes, the A at the start of each sequence represents position 1254 of Cpn10.1 and position 2310 of Cpn 1054 within the CWLO29 sequence accessible through the Chlamydia Genome Website [7]. Candidate restriction enzyme sites that may be used for RFLP analysis are indicated.

and each predicted gene product was also subjected to gap BLASTX and BLASTN respectively <http://www.ncbi.nlm.nih.gov/BLAST/>[20]. The similarity between two different DNA sequences was determined using the BLAST 2 sequence program from <http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>[21]. Hydrophilicity profiles of the gene product of each Cpn 1054 family member was determined using hydropathy plot analysis ([22] MacVector™).

Phylogenetic analyses of both DNA and amino acid sequences were performed using PAUP* [23]. In this study Cpn 0186 (IncA), and four additional candidate inclusion membrane proteins, Cpn 0284, Cpn 0285, Cpn 0829 and Cpn 0830, were selected as members of an outgroup for analysis. Phylogenetic trees were inferred by neighbor-joining to estimate evolutionary distances. Bootstrap values were obtained from a consensus of 100 neighbor-joining trees.

DNA amplification, and sequence analysis of the CPI054 gene family

Purified genomic DNA of selected isolates (Table 1) were amplified using specific oligonucleotide primers flanking the DNA region of interest (Table 2). The amplification products were purified using a Qiaquick PCR purification spin column kit (Qiagen, Valencia CA). Purified PCR products were sequenced using an ABI PRISM 377 (Perkin-Elmer, Norwalk, CT) through the Oregon State University Center for Gene Research and Biotechnology.

Examination of variation within isolates through cloning of PCR products

The variation of the length of the polyC tract within Cpn 043, 1054 and 1055 was determined through sequence analysis of purified amplification products and through sequencing of amplification products following cloning into plasmids. Both *Taq* polymerase (Promega, Madison, WI) and *Pwo* polymerase (Roche Diagnostic Corporation, Indianapolis, IN) were used in these studies. Amplifica-

Table 2: Description of the *C. pneumoniae* isolates used in this study

Isolate	Location(year) isolated	Clinical presentation	Ref.
CWL029	Atlanta, USA/1987	oropharyngeal swab/pneumonia	[24]
AR39	Seattle, USA/1983	throat swab/pharyngitis	[25]
TW183	Taiwan/1965	conjunctiva	[25]
AR458	Seattle/1986	acute respiratory infection	[25]
AR388	Seattle/1985	acute respiratory infection	[25]
AR231	Seattle/1984	acute respiratory infection	[25]
AR277	Seattle/1984	pneumonia	[25]
PS32	Seattle/1995	carotid endarterectomy	[26]
AC43	Japan/1989	nasopharynx/pneumonia	[27]
KA5C	Finland/1987	pneumonia	[28]
KA66	Finland/1987	pneumonia	[28]
2023	Brooklyn, NY/1987-89	acute respiratory infection	[29]

Table 3: Oligonucleotide primer pairs used for PCR and sequencing in this study.

ORF ^a	Region amplified ^a	5'-3' sequence of primers
Cpn 007	7488-8694	AAGAGTTTTCGCTAT TGACGACTCTTAGATGTTTCG
Cpn 010	13152-13502	GAAAACACCTCTTGTGGCTAG AGACATACCCACGGAATAATAAA
Cpn 010	13175-14431	TACTCGCGAGGACCATACCTAGC AACTCATGGAACGTTATCTCGTCC
Cpn 010.1	15178-16324	GCTTGGATGATTTACGACGAG CTAAGATTTTTGGGTGGCAGC
Cpn 041	55728-56987	GAAAATTCTTTTCGACAGAGATGGA TCCTGTGCTTTCATAACGCTCAC
Cpn 043	58069-59173	TGCTTTTCAGGCATTCCAGG GCGAGCTACGACCTCTTTAGCGA
Cpn 045	60755-61856	ATTTGAATCGATGGAGATTCCAGA TATCCTGCAGCCATTCTACATCTT
Cpn 1054	1206885-1208142	AAAGGAAGTTTTCTTGAAGATCTT ACTCATGGAACGTTATCTCGTCC
Cpn 1055	1209382-1210530	AGTTCCTGAGATTCTGAGGC TATCTCCTCAACTACTCGAGATAAC

Open Reading Frame (ORF) designations and region amplified are relative to the numbering in the CWL029 genome sequence available through reference [7].

tion products generated with *Pwo* polymerase were cloned using the Zeroblunt system, while products generated with *Taq* were cloned into pCRII (Invitrogen, Carlsbad, CA). Both cloning systems were used according to the manufacturer's instructions. Plasmid DNA of 8-10 different positive recombinant clones was isolated and sequenced. Variations of the length of the polyC tract of Cpn 043, Cpn 1054 and Cpn 1055 within *C. pneumoniae* AR39 were determined for 8-10 recombinant clones.

Nucleotide sequence accession numbers

The nucleotide sequences of variants within Cpn 010 identified from independent clinical isolates were deposited in GenBank under following accession numbers : AF474017 through 474026, and AF 461543 through 461552.

Authors' contributions

W.V. was responsible for the initiation of this project, carried out most of the technical work, and drafted the man-

uscript. L.A.C. was responsible for the isolate library and the production of genomic DNAs, and editing of the manuscript. C.C.K. participated in analysis of data and manuscript editing. D.D.R. supervised the work and produced the manuscript.

Authors' note

These data were presented at the American Society for Microbiology Meeting, Salt Lake City, May 2002

Acknowledgments

This work was supported by U.S.P.H.S. Awards # AI42869 and AI48769, and the Oregon State University Department of Microbiology N.L. Tartar Award Program. W.V. was sponsored through the Royal Thai Scholarship program from the Thailand National Center of Genetic Engineering and Biotechnology. We acknowledge John Bannantine and members of the Rocky laboratory for critical reading of the manuscript.

References

1. Grayston JT, Aldous MB, Easton A, Wang SP, Kuo CC, Campbell LA and Altman J **Evidence that *Chlamydia pneumoniae* causes pneumonia and bronchitis.** *J Infect Dis* 1993, **168**:1231-1235

2. Kuo CC, Jackson LA, Campbell LA and Grayston JT **Chlamydia pneumoniae (TWAR)**. *Clin Microbiol Rev* 1995, **8**:451-461
3. Saikku P **Chlamydia pneumoniae and atherosclerosis – an update**. *Scand J Infect Dis* 1997, **104**:53-56
4. Pettersson B, Andersson A, Leitner T, Olsvik O, Uhlen M, Storey C and Black CM **Evolutionary relationships among members of the genus Chlamydia based on 16S ribosomal DNA analysis**. *J Bacteriol* 1997, **179**:4195-4205
5. Meijer A, Kwakkel GJ, de Vries A, Schouls LM and Ossewaarde JM **Species identification of Chlamydia isolates by analyzing restriction fragment length polymorphism of the 16S-23S rRNA spacer region**. *J Clin Microbiol* 1997, **35**:1179-1183
6. Meijer A, Morre SA, van den Brule AJ, Savelkoul PH and Ossewaarde JM **Genomic relatedness of Chlamydia isolates determined by amplified fragment length polymorphism analysis**. *J Bacteriol* 1999, **181**:4469-4475
7. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW and Stephens RS **Comparative genomes of Chlamydia pneumoniae and C. trachomatis**. *Nat Genet* 1999, **21**:385-389
8. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J and Fraser CM **Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39**. *Nucleic Acids Res* 2000, **28**:1397-1406
9. Shirai M, Hirakawa H, Kimoto M, Tabuchi M, Kishi F, Ouchi K, Shiba T, Ishii K, Hattori M, Kuhara S and Nakazawa T **Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA**. *Nucleic Acids Res* 2000, **28**:2311-2314
10. Grimwood J and Stephens RS **Computational analysis of the polymorphic membrane protein superfamily of Chlamydia trachomatis and Chlamydia pneumoniae**. *Microb Comp Genomics* 1999, **4**:187-201
11. Grimwood J, Olinger L and Stephens RS **Expression of Chlamydia pneumoniae polymorphic membrane protein family genes**. *Infect Immun* 2001, **69**:2383-2389
12. Pedersen AS, Christiansen G and Birkelund S **Differential expression of Pmp10 in cell culture infected with Chlamydia pneumoniae CWL029**. *FEMS Microbiol Lett* 2001, **203**:153-159
13. Stephens RS and Lammel CJ **Chlamydia outer membrane protein discovery using genomics**. *Curr Opin Microbiol* 2001, **4**:16-20
14. Daugaard L, Christiansen G and Birkelund S **Characterization of a hypervariable region in the genome of Chlamydia pneumoniae**. *FEMS Microbiol Lett* 2001, **203**:241-248
15. Bannantine JP, Griffiths RS, Viratyosin W, Brown WJ and Rockey DD **A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane**. *Cell Microbiol* 2000, **2**:35-47
16. Jordan IK, Makarova KS, Wolf YI and Koonin EV **Gene conversions in genes encoding outer-membrane proteins in H. pylori and C. pneumoniae**. *Trends Genet* 2001, **17**:7-10
17. Zomorodipour A and Andersson SG **Obligate intracellular parasites: Rickettsia prowazekii and Chlamydia trachomatis**. *FEBS Lett* 1999, **452**:11-15
18. Deitsch KW, Moxon ER and Wellem TE **Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections**. *Microbiol Mol Biol Rev* 1997, **61**:281-293
19. Campbell LA, Kuo CC and Grayston JT **Characterization of the new Chlamydia agent (TWAR) as a unique organism by restriction endonuclease analysis and DNA:DNA hybridization**. *J Clin Microbiol* 1987, **25**:1911-1916
20. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**:3389-3402
21. Tatusova TA and Madden TL **Blast 2 sequences – a new tool for comparing protein and nucleotide sequences**. *FEMS Microbiol Lett* 1999, **174**:247-250
22. Kyte J and Doolittle RF **A simple method for displaying the hydropathic character of a protein**. *J Mol Biol* 1982, **157**:105-132
23. Swofford DL **PAUP*: Phylogenetic analysis using parsimony and other methods version 4.0b6**. *Sinauer Associates Inc., Sunderland, MA*
24. Campbell JF, Barnes RC, Kozarsky PE and Spika JS **Culture-confirmed pneumonia due to Chlamydia pneumoniae**. *J Infect Dis* 1991, **164**:411-413
25. Kuo CC, Chen HH, Wang SP and Grayston JT **Identification of a new group of Chlamydia psittaci strains Called TWAR**. *J Clin Microbiol* 1986, **24**:1034-1037
26. Jackson LA, Campbell LA, Kuo CC, Rodriguez DI, Lee A and Grayston JT **Isolation of Chlamydia pneumoniae from a carotid endarterectomy specimen**. *J Infect Dis* 1997, **176**:292-295
27. Yamazaki T, Nakada H, Sakurai N, Kuo CC, Wang SP and Grayston JT **Transmission of Chlamydia pneumoniae in young children in a Japanese family**. *J Infect Dis* 1990, **162**:1390-1392
28. Ekman MR, Grayston JT, Visakorpi R, Kleemola M, Kuo C-C and Saikku P **An epidemic of infections due to Chlamydia pneumoniae in military conscripts**. *Clin Infect Dis* 1993, **17**:420-425
29. Chirgwin K, Roblin PM, Gelling M, Hammerschlag MR and Schachter J **Infection with Chlamydia pneumoniae in Brooklyn**. *J Infect Dis* 1991, **163**:757-761

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

