Genomics 106 (2015) 373–383

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparative genomic analysis of human *Chlamydia pneumoniae* isolates from respiratory, brain and cardiac tissues

Eileen Roulis [a], Nathan L. Bachmann [b], Garry S.A. Myers [c], Wilhelmina Huston [a], James Summersgill [d], Alan Hudson [e], Ute Dreses-Werringloer [e], Adam Polkinghorne [a,b], Peter Timms [a,b,*]

[a] Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Queensland, Australia
[b] Faculty of Science, Health, Education & Engineering, University of the Sunshine Coast, Sippy Downs, Queensland, Australia
[c] i3 Institute, University of Technology, Sydney, New South Wales, Australia
[d] University of Louisville, Division of Infectious Diseases, Louisville, KY, United States
[e] Department of Immunology and Microbiology, Wayne State University, Detroit, MI, United States

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form 14 August 2015
Accepted 24 September 2015
Available online 28 September 2015

Keywords:
Chlamydia pneumoniae
Comparative genomics
Evolution
Recombination
Cardiovascular
Alzheimer's disease

ABSTRACT

Chlamydia pneumoniae is an obligate intracellular bacterium implicated in a wide range of human diseases including atherosclerosis and Alzheimer's disease. Efforts to understand the relationships between *C. pneumoniae* detected in these diseases have been hindered by the availability of sequence data for non-respiratory strains. In this study, we sequenced the whole genomes for *C. pneumoniae* isolates from atherosclerosis and Alzheimer's disease, and compared these to previously published *C. pneumoniae* genomes. Phylogenetic analyses of these new *C. pneumoniae* strains indicate two sub-groups within human *C. pneumoniae*, and suggest that both recombination and mutation events have driven the evolution of human *C. pneumoniae*. Further fine-detailed analyses of these new *C. pneumoniae* sequences show several genetically variable loci. This suggests that similar strains of *C. pneumoniae* are found in the brain, lungs and cardiovascular system and that only minor genetic differences may contribute to the adaptation of particular strains in human disease.

1. Introduction

The *Chlamydiae* are a phylum of obligate intracellular bacterial parasites infecting a broad range of eukaryotic hosts including amoebae, invertebrates such as insects, and vertebrates including humans [1,2]. All members of this phylum share a biphasic developmental cycle that is unique in the bacterial kingdom, consisting of an extracellular, infectious elementary body (EB) and the intracellular, metabolically active non-infectious reticulate body (RB), which replicates within a host-derived parasitophorous vacuole termed "inclusion" [3]. Members of the family *Chlamydiaceae* are well-described eukaryotic pathogens. These include the human pathogen *Chlamydia trachomatis* [4], the zoonotic pathogen *Chlamydia psittaci* [5], and respiratory pathogen *Chlamydia pneumoniae* [6,7].

C. pneumoniae was first identified as the cause of two human pneumonia epidemics in 1985, and later associated with a range of chronic diseases such as asthma, cardiovascular disease, arthritis and Alzheimer's disease (reviewed in [8]). The first report of *C. pneumoniae* infection in a non-human host came with its isolation from horses in 1993 [9], and it was subsequently identified as one of the two chlamydial species infecting koalas in 1994 [10]. Subsequently, *C. pneumoniae* has been identified worldwide in a wide range of cold and warm-blooded animals, making it the most cosmopolitan of all chlamydial species [9,11–13].

The *Chlamydiaceae* are characterized by their compact, highly conserved genomes, and in spite of the wide range of hosts and diseases caused, all members of this phylum share a highly conserved core genome of over 500 genes, equating to half the gene set of a member of the *Chlamydiaceae* [3]. Species and strain-specific differences amongst the *Chlamydiaceae* consist of nucleotide differences in various metabolic genes including tryptophan [14] and purine [15] biosynthesis pathways and potential virulence factors such as polymorphic membrane protein families [16,17], inclusion membrane proteins [18–22], type three secretion effectors [23–25] as well as functionally uncharacterized hypothetical proteins, commonly located within a distinct genetic region termed the plasticity zone.

Human strains of *C. pneumoniae* have been described as highly conserved in both their gene and nucleotide content [26], with only around 300 single nucleotide polymorphisms (SNPs) identified that separate the four human *C. pneumoniae* respiratory strains with whole genome sequences available to date [27]. The genomes of koala [27] and bandicoot [28] *C. pneumoniae* were recently sequenced revealing that *C. pneumoniae* strains between humans and animals are much more variable – although it is interesting to note that like the human strains, the animal strains are themselves virtually identical.

* Corresponding author at: Faculty of Science, Health, Education & Engineering, University of the Sunshine Coast, Sippy Downs, Queensland, Australia.

http://dx.doi.org/10.1016/j.ygeno.2015.09.008
0888-7543/

Whilst there are limited genetic differences reported between human *C. pneumoniae* strains, there have nevertheless been several attempts to correlate genetic differences to tropism for particular tissues. One such difference between human strains has been attributed to the tyrosine permease (*tyrP*) copy number; strains with a single copy of this gene have been suggested to be associated with vascular tropism [29]. The other "major" genetic differences observed to date between respiratory *C. pneumoniae* strains are mostly limited to genes encoding (i) the polymorphic membrane proteins (*pmp*), outer membrane proteins [30,31]; and (ii) the inclusion membrane protein family (*inc*) [22]. Sequences encoding the *pmp* genes in *C. pneumoniae* makes up to 22% of the difference in extra coding capacity when comparing human *C. pneumoniae* and *C. trachomatis* [15]. In comparing *pmp*s across human *C. pneumoniae* strains, distinct differences have been noted at both the nucleotide [17,32] and gene expression level [33–35]. The *incA* family is also considerably expanded in *C. pneumoniae* compared to other members of the *Chlamydiae*, with over one hundred characterised and putative *incs* identified in human *C. pneumoniae* strains [18,19,21]. However, the ability to correlate any of these genetic changes to potential phenotypes has been otherwise limited due to the lack of genome sequences for *C. pneumoniae* strains isolated from other human tissues.

In this study, we determined the whole genome sequences for two *C. pneumoniae* strains of non-respiratory origin (A03 and TOR-1) [36,37], and compared them to previously sequenced *C. pneumoniae* strains, thereby providing a broader insight into the genetic diversity of strains isolated from different human tissues.

2. Materials and methods

2.1. Description of *C. pneumoniae* strains, cell culturing and DNA purification

Two previously described *C. pneumoniae* strains isolated from either human heart (*C. pneumoniae* A03 [36]) or brain tissue (*C. pneumoniae* TOR-1 [37]) were used for comparative genomics analysis. The human *C. pneumoniae* A03 cardiovascular strain was obtained from the American Type Culture Collection (ATCC® VR-1452). This strain was originally isolated from the coronary artery of a patient undergoing heart transplantation in 1995 [36]. *C. pneumoniae* strain TOR-1 was originally isolated from post-mortem brain tissue of an Alzheimer's disease patient in 2003 [37].

Isolate A03 was propagated on HEp-2 cells in T75 flasks for five passages, based on a previously described method [38]. Infected cells were pooled and semi-purified using a probe sonication and centrifugation method prior to passage. The final semi-purified product was stored in an equal volume of SPG media [39]. 100 μl of this semi-purified material was used for DNA extraction. 100 μl of ultra-purified TOR-1 culture material was used for DNA extraction, its isolation and propagation has been previously described in Dreses-Werringloer et al. [37].

DNA extraction for both samples was performed using the QIAGEN PureGene kit, as per the manufacturer's instructions, with final hydration of the DNA pellet into 50 μl of hydration buffer. 500 ng of extracted DNA was used to perform pan-*Chlamydiales* 16S *rRNA* [40] and *C. pneumoniae* specific *rpoB* [41] PCR to confirm the presence of *C. pneumoniae* DNA, and 500 ng of stock DNA was electrophoresed on a 0.8% TBE agarose gel to confirm high molecular weight DNA. Each extraction yielded greater than 2 μg of high molecular weight genomic DNA, which was used for sequence capture and Illumina whole genome sequencing at the Institute for Genome Sciences, Baltimore, Maryland.

2.2. Sequence capture, whole genome sequencing and assembly

Sequence capture was performed on total DNA extracted from A03 and TOR-1 with Agilent probes designed to *C. pneumoniae* reference strain AR39, using a DNA hybridisation capture and amplification process based on the methods of Gnirke et al. that has been recently adapted for *Chlamydia* [42–44]. Captured and amplified products were sequenced using the Illumina HiSeq 2500 platform, resulting in paired-end 100 base-pair reads. Read quality analyses were performed using FastQC 0.10.1 (Babraham Bioinformatics group [http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc]) to determine read quality cut-offs. *De novo* assembly and readmapping of paired-end data from A03 and TOR-1 was performed using CLC genomics workbench (CLC – Denmark), after trimming for size and quality. Contig sequences were mapped to reference strain AR39 using the CONTIGuator web server [45], and stop gap spacers manually inserted before concatenation in Geneious 6.18 [46]. Concatenated genome sequences were annotated using the RAST pipeline [47] and manually curated using ARTEMIS [48].

The A03 and TOR-1 whole genome sequencing projects can be found on National Center for Biotechnology Information (NCBI) BioProject under accession numbers PRJNA239383 and PRJNA239384 with reads deposited in the Short Reads Archive under accession numbers SRR1954962 and SRR1954966 respectively.

2.3. Whole genome comparative analyses and phylogeny

Consensus sequences for A03 and TOR-1 extracted from readmapped assemblies were aligned to the existing human *C. pneumoniae* whole genome sequences of AR39, CWL029, J138 and TW-183 [15,26,49,50] in Geneious using the MAFFT plugin implementation [51]. Coverage analyses for readmapped sequences and manual curation of annotated genomes was performed using ARTEMIS [48]. Coding sequences identified as polymorphic in A03 or TOR-1 were translated and functional changes were predicted following analysis of encoded sequences using Simple Modular Architecture Research Tool (SMART) [52].

Bayesian phylogenetic analyses were performed on *C. pneumoniae* whole genome alignments, with the LPCoLN koala *C. pneumoniae* strain [27] included as an outlier. Whole genome alignments were also filtered for poorly aligned and gap regions using Gblocks 0.91b [53]. Mid-point rooted trees were constructed with the MrBayes plugin [54] in Geneious, utilising a Jukes–Cantor substitution model with four Markov Chain Monte Carlo (MCMC) chains and 1.1 million cycles, sampled every 1000 generations and the first 10,000 trees discarded as burn-in. Estimates of strain evolution over time were performed on whole genome alignments using the BEAST package [55]. Respiratory and non-respiratory isolates were defined in separate taxon sets and a GTR nucleotide substitution model was employed. Tip dates and taxon designations are outlined in Table 1. MRCA priors were set at a normal distribution with a mean of 95.2 ± 7.4 given the estimates of the last common ancestor of *C. pneumoniae* and its transmission to humans at 150,000 years [56]. MCMC chain length was set to 5×10^7 to ensure effective sample sizes were sufficient for strong posterior distribution statistics. ClonalFrame [57] was used to determine homologous recombination within human *C. pneumoniae* genomes, and progressive MAUVE [58] was used to generate the input alignments. Three successive runs of ClonalFrame were performed on the whole genome alignment, with 20,000 iterations and 10,000 of these discarded as burn-in. The three runs were checked for convergence and their trees combined for analysis.

The GenBank accession numbers for the following *C. pneumoniae* whole genome sequences were used in the comparative analyses and

Table 1
Tip dates and taxon designations for *C. pneumoniae* strains used in BEAST analyses.

Strain	Tip date	Taxon	Reference
A03	1995	Non-respiratory	[36]
AR39	1983	Respiratory	[49]
CWL029	1987	Respiratory	[15]
J138	1994	Respiratory	[26]
TOR-1	2003	Non-respiratory	[37]
TW-183	1965	Non-respiratory	[50]