GENOME WATCH

Vive la différence

Nicholas R. Thomson

The first *Chlamydia trachomatis* genome was sequenced in 1998, but now, with the recent publication of the *C. trachomatis* lymphogranuloma venereum (LGV) genome sequence, the genome catalogue of the different disease variants of this *Chlamydia* species is complete, and is discussed in this month's Genome Watch.

Chlamydia spp. are among the oldest described human pathogens, and have long fascinated researchers because they are obligate intracellular bacteria that have a unique developmental cycle (FIG. 1) and can infect virtually all cell types. The chlamydial developmental cycle is biphasic, consisting of two cell types: the infectious electron-dense elementary body, which is 0.2–0.6 µm in diameter and metabolically inactive, and a metabolically active reticulate body. Elementary bodies enter eukaryotic cells by endocytosis and remain sequestered in special non-exocytotic vesicles called inclusions. The elementary bodies that are present in inclusions differentiate into the larger (~1.5 µm diameter) reticulate bodies, which are the metabolically active replicating form of *Chlamydia*. After a period of approximately 2 days, the reticulate bodies differentiate back into elementary bodies, which are released from the cell either by membrane fusion or through mechanical disruption of the host cell.

Since 1999, the Chlamydiaceae family has been split into two genera, the *Chlamydia* and the *Chlamydophila* (*Chlamydia*-like). Although this Genome Watch focuses on *Chlamydia trachomatis*, one of the two main human chlamydial pathogens, representative human, animal and environmental chlamydial genomes have been sequenced. The large number of sequenced chlamydial genomes is also a reflection of their small genome size, their interesting biology and the breadth of diseases that they cause (see Further

information for a link to a comprehensive reference and education site ([Chlamydia and the chlamydiae](#)) that is hosted by the University of Southampton).

This Genome watch focuses on *C. trachomatis* variants that can cause 3 radically different disease pathologies in humans and have been divided, on the basis of serotype, into 15 different serovariants: serovars A–C cause trachoma, a serious infection of the mucous membranes of the eyes that, if untreated, can result in blindness; serovars D–K and L1, L2 and L3 are associated with sexually transmitted infections (STIs); and serovars D–K can cause cervicitis in women and urogenital infections in men. Of these, L1, L2 and L3 are the only serovars that are associated

with lymphogranuloma venereum (LGV). Of the sexually transmitted *C. trachomatis* serovars, the LGV strains are more invasive than serovars A–K, and cause systemic infections that disseminate to the local lymph nodes, which can result in large swellings, or buboes, that are characteristic of bubonic diseases. LGV serovars are rarely isolated in western countries, but are common in parts of Africa, South East Asia, South America and the Caribbean. However, this human demographic has changed recently with a European outbreak of LGV that was reported in men who have sex with men, although these individuals presented with proctitis rather than genital ulceration and the buboes that are regarded as the characteristic presentations of a classical LGV infection.

Figure 1 | **Diagram of an idealized chlamydial developmental cycle.** The small, infectious elementary bodies (EBs) are shown in blue and the larger, replicating reticulate bodies (RBs) are shown in pink. Figure courtesy of K.D. Everett, University of Georgia, Georgia, USA.

It has been proposed that the serovar which causes this variant form of LGV is a recently evolved variant of the classical LGV strain that is named L2b and represents an emergent infection.

Because no genetic tools or cell-free methods of culture are available to manipulate the chlamydial genome, genomics has provided important insights into chlamydial biology. Complete genome sequences are now available for three of the *C. trachomatis* pathotypes: *Chlamydia trachomatis* A/HAR-13 (serovar A), a trachoma strain; *Chlamydia trachomatis* D/UW-3/CX (serovar D), an STI strain; *Chlamydia trachomatis* 434/Bu (serovar L2), a classical LGV strain; and *Chlamydia trachomatis* L2b/UCH-1/proctitis (serovar L2b), a proctitis LGV strain^{1,2,3}. Whole-genome analysis has shown that the size of these 4 genomes varies by a maximum of 5,000 bp, with all 4 genomes comprised of ~1.04 Mbp. The four serovars also share a high level of synteny and genome-sequence conservation. This conservation is best illustrated by comparing the coding capacities of the three genomes. A comparison of the gene sets of single trachoma, STI and LGV strains revealed that 846 coding sequences (CDSs) were common to all 3 genomes (the total CDS count for these genomes range from 889–920). Of the remaining genes, most differences that were noted represented differences in *in silico* gene prediction rather than real differences in gene content. The authors concluded that horizontal gene transfer has probably not had a significant impact on the changes in disease pathology that are caused by this group of organisms³.

Although there were many common features among the *C. trachomatis* genomes, there were also some surprises. It was thought that Chlamydiae lack peptidoglycan, but all the genomes contained most of the genes that are required for the synthesis and assembly of this polymer. The presence of these genes might provide the answer to the puzzling observation that Chlamydiae are susceptible to inhibitors of peptidoglycan synthesis, yet seem to lack this polymer². This finding aside, the overriding theme of these chlamydial genomes is metabolic streamlining and gene loss, which is consistent with many other pathogens that exploit an intracellular lifestyle⁴.

Analysis of the metabolic capacity of *C. trachomatis* has shown that several important metabolic pathways are incomplete. One pertinent example is the tricarboxylic acid (TCA) cycle, in which the genes for citrate synthetase, aconitase and isocitrate dehydrogenase are lacking. There is also evidence for ongoing gene loss by point mutations of

other genes in this pathway: *C. trachomatis* 434/Bu (the classical LGV serovar L2 strain) has point mutations in the TCA-cycle genes *sdhC* (succinate dehydrogenase) and *fumC* (fumarase), which might indicate that 2-oxoglutarate is used to synthesize succinate and succinyl CoA^{2,3}.

There is also evidence in these genomes for large-scale deletion of genes, most of which are confined to a region at the terminus of replication that has been named the plasticity zone⁵. The plasticity zone varies in size between the different *C. trachomatis* genomes from approximately 20 to 25 Kbp^{1,2,3}. This variation in size is largely due to differential deletion of the central portion of the plasticity zone, which is thought to have originally encoded a cytotoxin (or cytotoxins) that is similar to those found in other chlamydial species. Interestingly, although all of the cytotoxin genes that are found in *C. trachomatis* isolates are remnants, or pseudogenes, one of the defining characteristics of the STI strain *C. trachomatis* D/UW-3/CX is that a cytotoxin fragment is expressed and produces a cytotoxic protein⁶. This is a salutary lesson to us all not to disregard apparently broken genes.

Other genetic features that correlate with a particular phenotype include the presence of tryptophan synthetase. As part of the human innate immune response, interferon- γ (IFN- γ) induces the expression of an enzyme, indoleamine 2,3-dioxygenase (IDO), that catabolizes the essential amino acid tryptophan and inhibits the growth of *C. trachomatis*, which is a natural tryptophan auxotroph. However, tryptophan synthetase can catalyse the conversion of indole into tryptophan. The gene that encodes tryptophan synthetase is intact in the STI strains, but truncated in the trachoma strains. Because indole is available in the genital tract as a waste product of metabolism by other bacteria, this may be a way that the genital STI strains mitigate the impact of IFN- γ -induced IDO. Why the trachoma strains have not retained this ability is not clear^{2,7,8}.

The genome sequence of the recently identified proctitis strain *C. trachomatis* L2b/UCH-1/proctitis, which causes a variant form of LGV, was equally revealing. A comparison with the classical LGV strain that was isolated ~40 years previously showed that its genome was 27 bp larger, that there were no whole-gene or pseudogene differences and that less than 600 single-base polymorphisms (SNPs) distinguish the 2 LGV isolates³. Thus, it was clear that there are no gross differences between the classical LGV strain and the LGV proctitis strain. Interestingly, the 600 SNP differences clustered in the genome. This feature was used to identify genes that were under

positive selection, and that therefore might represent new virulence genes or useful targets for vaccine development.

C. trachomatis has the ability to infect different niches and cause a range of distinct pathologies. However, whole-genome sequencing has shown us that the *C. trachomatis* pan-genome (that is, the core plus all the variable regions from all the *C. trachomatis* genomes combined) is tiny compared with free-living bacteria, such as pseudomonads⁹, and that the evolutionary history of this intracellular pathogen has involved loss of functions and genome reduction. For bacteria that are so recalcitrant to laboratory manipulation, these genomes have provided much-needed insights into their biology, which has explained some experimental observations and identified possible diagnostic and therapeutic targets. The new proctitis strain is likely to be “an old strain causing a new disease” (REF. 3), which highlights how, even though we have more genetic information on *C. trachomatis* than ever before, this bacterium still harbours many more secrets: a challenge indeed.

1. Carlson, J. H., Porcella, S. F., McClarty, G. & Caldwell, H. D. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect. Immun.* **73**, 6407–6418 (2005).
2. Stephens, R. S. *et al.* Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**, 754–759 (1998).
3. Thomson, N. R. *et al.* *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res.* **18**, 161–171 (2008).
4. Wernegreen, J. J. Genome evolution in bacterial endosymbionts of insects. *Nature Rev. Genet.* **3**, 850–861 (2002).
5. Read, T. D. *et al.* Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaeaceae. *Nucleic Acids Res.* **31**, 2134–2147 (2003).
6. Carlson, J. H. *et al.* Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. *Infect. Immun.* **72**, 7063–7072 (2004).
7. Caldwell, H. D. *et al.* Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. *J. Clin. Invest.* **111**, 1757–1769 (2003).
8. McClarty, G., Caldwell, H. D. & Nelson, D. E. Chlamydial interferon gamma immune evasion influences infection tropism. *Curr. Opin. Microbiol.* **10**, 47–51 (2007).
9. Mathee, K. *et al.* Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl Acad. Sci. USA* **105**, 3100–3105 (2008).

Nicholas R. Thomson is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro1929

Published online 9 June 2008

DATABASES

Entrez Genome: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genome>
Chlamydia trachomatis 434/Bu | *Chlamydia trachomatis* A/HAR-13 | *Chlamydia trachomatis* D/UW-3/CX | *Chlamydia trachomatis* L2b/UCH-1/proctitis

FURTHER INFORMATION

The comprehensive reference and education site for Chlamydia and the chlamydiae: <http://www.chlamydiae.com/>
 ALL LINKS ARE ACTIVE IN THE ONLINE PDF