AN ABSTRACT OF THE DISSERTATION OF

Timothy E. Putman for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on June 3, 2015.

Title: Using Culture-independent Genomics as a Means for Understanding Chlamydial Biology.

Abstract approved: _____

Daniel D. Rockey

The chlamydiae are a family of obligate intracellular bacteria that have a unique bi-phasic developmental cycle, unique cellular properties, and a unique set of challenges to studying its biology. While genetic manipulation is becoming routine in *Chlamydia*, there are significant challenges to working with this intracellular parasite. The body of this work focuses on the chlamydiae in context to comparative analysis of genome structure, organization, population dynamics, and recombination. This approach presents its own challenges, and this work includes the development of tools and methods for improving the comparative genomic research.

Next generation sequencing approaches have led to completion of several dozen chlamydial genome sequences, most of which are from *Chlamydia trachomatis*. Analysis of these genomes has shown that chlamydiae, like other obligate intracellular bacteria, have a much reduced genome structure that implies dependence on the host for much of its metabolic capability. Certain groups of genes, including those encoding inclusion membrane proteins and the family of polymorphic membrane proteins, have been

significantly expanded against this general reductive evolutionary strategy. Pregenomic and postgenomic sequence analysis of *C. trachomatis* has led to considerable understanding of nucleotide polymorphisms, insertions and deletions that are associated with certain clinical presentations.

While widespread recombination in the human pathogen, *Chlamydia trachomatis,* is revealed in previous studies, the development of novel culture-independent sequencing tools in this work have added and expanded upon our knowledge of the extent and diversity of recombination in *Chlamydia spp..* These methods are then adapted and applied to a large-scale comparative study of the ruminant pathogen, *Chlamydia abortus,* illuminating a closely related group of genomes using an interesting form of pre-transcriptional gene regulation for diversifying the permutations of membrane proteins on its cell surface. Finally, the protocols are modified to vastly improve methods for genome sequencing clonal isolates that have been grown in minimal cell culture.

Future research will address chlamydial genome structure in the context of the system in which they live, and will include data on the host microbiome and host genetic background. We anticipate that integrating these areas of research will lead to significant progress in our understanding of the nature of chlamydial infection and disease.

Using Culture-independent Genomics as a Means for Understanding Chlamydial Biology.

by
Timothy E. Putman

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 3, 2015
Commencement June 2015

Doctor of Philosophy dissertation of Timothy E. Putman presented on June 3, 2015

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Timothy E. Putman, Author

ACKNOWLEDGEMENTS

My graduate studies have been influenced and supported by many people. First and foremost I would like to express my sincere appreciation to Dr. Dan Rockey, my advisor, for positive and supportive guidance throughout my time at Oregon State University. I would like to acknowledge all of the members, past and present, of the Rockey lab that have given me guidance, support, meaningful scientific conversation, and fun place to work. I would like to acknowledge Mark Dasenko of the Center for Genomic Research and Biocomputing Core Laboratories for his tireless efforts and unequaled skill in Illumina based sequencing. All the members of my committee: Dr.s Jeff Chang, Brian Dolan, Steve Giovannoni, and Frank Chaplen, have given me priceless technical and personal guidance and deserve recognition.

I would like to thank my family for unending support. My parents, brothers, and grandmother, have been encouraging and supportive through out this arduous process. Last, and most importantly, I would like to thank my supportive wife Meg, and my two wonderful sons Fisher and Milo, for working together with me to accomplish this goal, and always putting a smile on my face, no matter how busy I was.

CONTRIBUTION OF AUTHORS

Chapter 1

John Ivanovitch was a major contributor to the development of the culture-independent sequencing methods, Bob Suchland was responsible for all specimen procurement, and Daniel D. Rockey contributed the majority of the intellectual content, as well as funding and laboratory materials.

Chapter 2

Kathy O'Reilly and Beth Mamer were involved in clinical specimen procurement, diagnostics, and quantification. Bernhard Kaltenboeck provided intellectual analysis, writing, and the historical repository strains that he acquired from Johannes Storz. Daniel D. Rockey contributed the majority of the intellectual content, as well as funding and laboratory materials.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

Using Culture-independent Genomics as a Means for Understanding Chlamydial Biology.

Chlamydia trachomatis Genome Structure

Timothy E. Putman and Daniel D. Rockey

Chlamydial Infection: A Clinical and Public Health Perspective.
Issues Infect Dis. Basel, Karger, 2013,
Volume 7, pp 61-77
(DOI:10.1159/000348752)

Introduction

Chlamydia species fulfill an obligate intracellular niche so successful in longevity and range of hosts that there is not a vertebrate on the planet that has escaped its parasitism (1). Its bi-phasic developmental cycle is unique to its phylum, and produces two forms of the chlamydial cell; the infectious elementary body (EB) and replicative reticulate body (RB). The spore-like EB protects the pathogen as it is searching for a host cell, and RB utilizes host resources, which allows it to undergo binary fission and produce more *Chlamydia*. The obligate intracellular nature of *Chlamydia* presents unique challenges to the pathogen itself, the host, and the researcher trying to understand its biology. All 3 of these components of the chlamydial research are addressed in this document with the goal of understanding a pathogen that is the leading cause of blindness in the developing world (2) and the most reported bacterial infection in the United States (3).

There have been many technological revolutions in the study of pathogenic bacteriology. These revolutions include the Gram stain and the Petri dish techniques and tools first described in the 1880s that have routine utility to this day. This era also saw the first successful efforts to culture bacteria and the association between infection and disease- critical studies that have changed the way we address issues of health. Much more recently, the invention of PCR revolutionized many aspects of science, and allowed specific genetic regions from trace amounts of template to be amplified into a workable amount of genetic material. PCR and other amplification technologies also set

the table for the latest revolution in biological analysis, which has allowed a completely different set of questions to be answered about infectious disease and almost every other aspect of biological investigation. This revolution centers on the incredible advances in next generation sequencing, proteomics and computer-aided analysis of data (i.e. bioinformatics). Advances in these areas are truly spectacular; new tools in genome sequencing allow vast amounts of DNA to be analyzed both by individuals and at large sequencing centers. For example, the Department of Energy Joint Genomics Institute, one of several very large-scale sequencing centers worldwide, generated thirty trillion quality bases of nucleotide sequence in the first half of 2012 (updated quarterly statistics available at: http://www.jgi.doe.gov/sequencing/statistics. html). Genome sequencing has taken incredible strides in the last few years, with genomes such as *Chlamydia* being sequenceable in just a few days with very little starting material and rapidly decreasing costs. Advances in proteomics parallel these gains, and the computer-based analysis tools have also kept pace. This has led to an unprecedented availability for researchers and, perhaps soon, the general public of reasonably priced global genetic informatic tools of discovery and diagnosis. These technologies facilitate a truly global analysis of the biology surrounding a host and pathogen, against the background of genetic variability and the myriad unrelated organisms that also exist in this interaction. These advances are touching every aspect of biomedical investigations such as tailoring drug efficacy to individual patient needs (pharmacogenomics), through analysis of participants in complicated disease etiologies, for example, bacterial vaginosis (4), through a systems biology approach to understanding and addressing

global aspects of known infectious conditions. As discussed in the previous chapters, chlamydial disease pathogenesis is a function of the intimate interactions between the specific infecting agent and the host immune response. Many of these interactions remain poorly elucidated, and understanding these interactions will be critical to the development of predictive and protective clinical interactions. While the Petri dish and the Gram stain were clearly revolutionary and allowed physicians to rethink the fundamental nature of infectious disease, the current revolution in genomics will facilitate a detailed understanding of disease that until recently could not have been anticipated. The purpose of this chapter will be to address the nature of chlamydial genomics and how the study of genomics has allowed a clearer understanding of the organism and its mechanisms of pathogenesis.

Chlamydia Genome Structure and Evolution

The 'obligate' in obligate intracellular bacteria is a product of requirements by the bacterium to acquire nutrients or anabolic precursors from the host cell. This is commonly reflected in a reduction in both metabolic capability encoded by the parasite, and a concomitant decrease in genome size (5). This is reflected in the chlamydial genome, where evolution through different levels of animal groups has produced highly syntenous (i.e. similar and similarly ordered) and consistent genomes within the genus *Chlamydia* that are between 1 and 1.2 million base pairs in size (6-11). This is roughly one fifth to one sixth the size of the *Escherichia coli* and *Pseudomonas aeruginosa* genomes, respectively (12, 13). This reduction in size reflects a much more limited

metabolic diversity and flexibility. A bioinformatic analysis of the chlamydial metabolome indicates that many biosynthetic pathways present in organisms that live outside of cells are absent or severely truncated in the chlamydiae. This includes gene sets associated with amino acid synthesis, nucleotide assembly and other processes involved in free-living growth. In contrast, the chlamydiae have expanded certain lineage-specific gene sets, including those encoding inclusion membrane proteins (Incs) (14), and the polymorphic membrane proteins (Pmps (15)), which, collectively, represent between 6 and 10% of the different genomes. Other chlamydiae-specific proteins that likely function in the intimate interactions between host cell and chlamydiae are the major outer membrane protein (MOMP) (15) the chlamydial proteasome-like activity factor and other proteases(16), the translocated actin-recruiting protein (TARP) (17, 18) and possible mediators of cellular survival and immunity (19-21)some of which are described elsewhere in this dissertation.

The unique intracellular niche occupied by chlamydiae has been exploited by *Chlamydia*-like bacteria since before vertebrates evolved. Analysis of genome sequences from parachlamydiae and protochlamydiae demonstrate that this lineage of bacteria has continued to manicure and reduce its genomic capability as the chlamydiae have become more completely dependent on the host, and have discarded genes that become ‘extra baggage’ in this environment (5). This model of chlamydial evolution is consistent with genomic structures of many different bacteria, both intracellular and extracellular, that have evolved to an intimate and dependent interaction with a

particular host. For example, the obligatory intracellular rickettsiae have genome sizes similar to chlamydiae (22, 23), while *Treponema pallidum* (1.1 megabase genome) (24) and *Mycoplasma* spp. (0.6–1.3 megabase genomes) (25), each of which are fastidious extracellular bacteria that are intimately intertwined with host mucosal environments, have similarly-sized or smaller genomes.

Many chlamydiae also carry a remarkable small plasmid that has recently been correlated with virulence (26, 27). The presence of the plasmid is noteworthy for several reasons. This highly conserved genetic element in many ways is quite different to plasmids in other systems, where these mosaic and ever changing genetic elements are used to shuttle genetic capability among strains and species, variably integrating into the chromosome or being maintained extrachromosomally, adding overall variability to the genetic capability of a species [reviewed in (28)]. In *Chlamydia trachomatis*, however, the plasmid is another example of consistency among genomes, with differences in plasmid presence and structure being the rare exception. The conservation of the chlamydial plasmid calls into question whether or not it could still be considered a plasmid, considering it bares many properties of an extra chromosome.

Nevertheless, until its possible misnomer is addressed, the chlamydial 'plasmid' encodes eight open reading frames, at least one of which is important to virulence. Variation in chlamydial plasmid structure has recently led to problems in diagnosis of chlamydial infections, as the target of a routinely used commercial amplification-based assay was

specific to a region of the plasmid. This region was deleted in a *C. trachomatis* strain circulating in Sweden, leading to false negatives when patients were tested(29). This resulted in expansion of this particular strain in an interesting phenomenon of artificial selection being carried out by a diagnostic test. Strains that included the diagnostic marker on the plasmid were treated with antibiotics and selected against; those that had deleted it had obtained a selective advantage and were able to expand via transmission among hosts that were unaware of their infection.

Clues to the evolution of the *Chlamydia* genus can be observed in the analysis of genome structure of organisms related to the chlamydial pathogens of humans; this includes the members of the genera *Parachlamydia*, *Protochlamydia*, *Simkania* and *Waddlia* (Fig. 1.1).

Most of these related organisms also encode certain proteins that are considered essential to the chlamydial lineage. A highly conserved type III secretion machine is encoded by each of these groups, and it is postulated that this important and widely distributed tool for interacting directly with host cells might have evolved originally in an organism of this lineage (30). While different members of the lineage will lack one or another of the proteins described in the above paragraphs, evolution or acquisition of genes encoding these proteins appear to be among the most important players in the exploitation of the chlamydial intracellular niche by the evolving *Chlamydia*-like bacteria (1).

The considerable synteny of the chlamydial genome makes it both straightforward and challenging to identify regions of sequence variability that are important to differences in pathogenicity. A comparison of each of the *C. trachomatis* genomes shown in Figure 1.1 demonstrates that there are very few loci in the genome that are highly variable across the species (Table 1.1), and there are no examples within *C. trachomatis* of genuine genomic islands as seen in many other pathogenic species. This is in contrast to many pathogenic species that cause diseases of humans or animals, but is consistent with highly evolved intracellular pathogens. Variability in *E. coli*, for example, is spread across the genome and there are large genomic islands (i.e. pathogenicity islands) that clearly define differences in tropism, disease capacity, or antibiotic resistance among strains (31). There is a single genomic island found in *Chlamydia* spp. – the '*tet*(C) island' of *C. suis* (32), a classic antibiotic resistance element that was acquired from unrelated bacteria (33). This island allows these organisms to survive in an animal husbandry environment that has historically included sub-therapeutic administration of tetracycline as a growth promoter. These strains are found worldwide (34) and the resistance allele can be transferred to *C. trachomatis* in the laboratory (35), demonstrating that human pathogenic chlamydiae can become resistant to antibiotics that are commonly used to treat such infections. Outside of this genomic island, *Chlamydia* spp. tend to vary by differentially inactivating, deleting, duplicating or modifying individual genes from within the lineage. The following sections will describe individual coding sequences and proteins that vary among chlamydial strains, with a goal of addressing how these proteins might affect differences in pathogenesis.

Variation in Genome Structure within C. trachomatis

ompA (CT681)

Identification of regions of variability within the *C. trachomatis* genome began before genome sequences in this system were possible to obtain. Initial examples of sequence variation centered on *ompA*, the gene encoding MOMP, which is the major serovariant antigen in *C. trachomatis* and other chlamydiae (15). Serotype differences among different chlamydiae were first characterized with antibodies to MOMP, and these differences were associated with differences in disease spectrum among strains. Strains of serovars, classified with the letters *A, B* and *C* ,were associated primarily with Trachoma, serovars *D-K* with classical urogenital chlamydial disease, and serovars *L1, L2* and *L3* associated with lymphogranuloma venereum (LGV), a more aggressive and invasive condition. Sequence analysis of MOMP identified four major hypervariable regions, which represent the primary regions of the protein that are exposed to the surface (36). These primary sequencing experiments also revealed that recombination occurred among chlamydial strains, leading to MOMP proteins that are mosaics of different serovars (37).

The Chlamydial Plasticity Zone (CT152–176)

The first four chlamydial genome sequences to be completed included *C. trachomatis* serovar D, *C. pneumoniae* AR39 and CWL029, and *C. muridarum* Nigg. A comparative

genomics analysis by one of the groups conducting this sequencing revealed a 20–50 Kb region of the genome that varies considerably among these four strains, against the described background of considerable sequence similarity and synteny (10). Continued exploration of chlamydial genome sequences illustrates a region that is variable across the genus *Chlamydia* (10). This region of the chromosome contains several genes that encode putative virulence factors that may play a role in why these pathogens target different species (i.e. mouse or human) and/or different tissues within species (i.e. lung, genital tract or eye). Several consistent components of the plasticity zone in *C. trachomatis* and *C. muridarum* highlight the possible role of some of these differences, and are described in the next sections.

Tryptophan Synthesis

The amino acid tryptophan has long held a curious place in chlamydial biology. It has been known that starving cells of tryptophan in vitro, or of one of several other amino acids, leads to interruption of the classical developmental cycle and the formation of division-incompetent, aberrant reticulate bodies that do not mature to elementary bodies (38).

These aberrant forms likely have a unique place in disease pathogenesis associated with chlamydial infection. The role of tryptophan synthesis in *C. trachomatis*-host interactions coincides with the up-regulation of exogenous interferon gamma in human cells, resulting in the production of indoleamine deoxygenase, a protein that depletes

intracellular tryptophan abundance (39). Thus, the host cell is working to reduce the ability of the intracellular pathogen to grow, by removing this important building block from the nutrient pool.

This story becomes even more interesting when variation in tryptophan biosynthetic machinery is compared among the different chlamydial strains. The genes *trpA* (CT171) and *trpB* (CT 170) are present in the plasticity zone of *C. trachomatis* strains, while these genes (and any other fragment of the Trp biosynthetic machinery) are absent in the closely related *C. muridarum* (7). Expression of intact chlamydial t*rpA/B* in *E. coli* demonstrated that they were functional, and that they led to the ability of strains to metabolize tryptophan from indole. Analysis of a large collection of strains in the University of Washington Chlamydia Repository demonstrated that *trp*A/B in ocular strains of *C. trachomatis* have frame shifts due to deletions and are predicted to encode non-functional proteins, while these genes are intact in strains and serovars that grow in the genital tract (40). This distinction becomes more apparent when serovar B is considered. Strains of this serovar can be divided into those that cause blinding trachoma and those that cause classic urogenital disease. In each strain, the genotype at *trpA/B* is consistent with the target tissue of the infection (i.e. Serovar B ocular infection have functional tryptophan synthesis pathway while the same serovar in a genital tract infection will have deactivated this pathway). There is considerable variation in the *trp* loci among the different chlamydial species, likely reflecting their need to differently

exploit the host-based nutritional condition and stresses in each of their target hosts (10).

Differences in the Trp operon among different *C. trachomatis* strains represent a fascinating example of how genome structure allows these organisms to exploit a particular host niche. Why would a pathogen of the genital tract be selected for an ability to metabolize indole to tryptophan, while a strain that grows in the conjunctivae specifically lacks this ability? The answer could come from the nature of the microbiota in each environment. The lower genital tract is colonized by a large variety of organisms that likely can provide some level of indole to the community. This is not the case in the conjunctivae, where the abundance and diversity of the microbiota is much lower. Therefore, chlamydiae infecting the genital tract have the opportunity to exploit a nutritional source that allows them to grow in a hostile environment that may be replete with IFN-γ-secreting T cells, which are working to starve the intracellular bacteria for tryptophan. While the T cells are likely present as well in the conjunctivae, the source of indole might be lacking, leading to a reduced need for the ability to metabolize this product.

C. trachomatis and *C. muridarum* also encode partial fragments or complete proteins sharing identity with cellular cytotoxins from other species (41, 42). Treatment of host cells with homologous toxins from other species (e.g. *Clostridium difficile* and *E. coli*) leads to cytoskeletal rearrangement resulting from disruption of host cell GTPase

activity. Within the chlamydiae, the number and structure of these toxins varies among and within species. For example, *C. muridarum* encodes three genes encoding different variations of this toxin, while different *C. trachomatis* strains carry different fragments or frameshift mutations of the toxin open reading frame. *C. trachomatis* serovars L1–L3 encode very short fragments of the toxin, and almost certainly express no toxin activity; one exception being a recombinant L2 strain discussed later in this chapter. Hypotheses surrounding the role of these toxins in the differential tissue tropism and pathogenesis of different *C. trachomatis* strains have been challenging.

Other plasticity zone genes vary among strains (e.g. CT149–151) (43), but it is not clear how the presence or absence of these coding sequences functions in the different tropisms and clinical presentations observed within *C. trachomatis*. Considering the conservation across the chlamydial pan genome, the plasticity zone appears to be one region of the chlamydial chromosome in which reductive evolution is still in progress, and changes in this region have likely been selected for optimal growth in the different environments occupied by these otherwise closely related strains and species.

Pmps, Incs and TARP

There are several other genes that vary in structure across the *C. trachomatis* genome, including some that might have a role in host and tissue specificity. The two large families of genes mentioned earlier in the review – the Pmps and the Incs – are present across the chlamydial lineage, and vary significantly across the species in number and

expression(44, 45). For example, chlamydial Pmps, which are autotransporter (i.e. type V) secreted proteins with coding source sequences scattered across all *Chlamydia* spp. genomes. The effector domains of these proteins are secreted out of the chlamydial developmental forms and *C. trachomatis* has over 10 different *pmp* or *pmp*-like genes, while *C. pneumoniae* has expanded this family of proteins to over 20. The encoded proteins have a variety of functions in chlamydial biology, and there are differences in the number of Pmps between species and differences in the sequence of Pmps within species. This pattern is also found in the chlamydial Inc proteins. The number and structure of Incs vary both across and within species, with significant differences observed among the different human pathogenic species. Although there are excellent examples where gene function has been elucidated for Incs and Pmps, the study of these proteins remains in its infancy. It is possible that differences in number and sequence of Pmps and Incs play a substantial role in the interactions of the different chlamydiae with different hosts and host cells.

The chlamydial TARP (CT 456) functions in early interactions between chlamydiae and the host cell, by causing major rearrangements to the actin cytoskeleton of the host cell, facilitating primary uptake of the pathogen (17, 18). TARP from each species contains actin-binding domains and a proline-rich domain; TARP from strains within *C. trachomatis* also have a tyrosine-rich phosphorylation domain. It is proposed that the proline-rich domain facilitates TARP-TARP interactions, which focus the actin nucleation events, facilitating uptake. It is not clear how that phosphorylation domain participates in this process, as it is absent from the *C. trachomatis* strains examined to date.

Structural differences are present in TARP both within and between species (46), and differences in TARP structure correlate with the differences in invasive characteristics of *C. trachomatis* (47).

Smaller Scale Genetic Variation and Disease Pathogenesis

We have discussed the overall synteny and conservation of genome structure in *C. trachomatis*, and how it is likely that genetic differences such as *trpA/B* structure likely participate in the variety of diseases that are caused by these pathogens. There is also evidence that *C. trachomatis* genetic variation within an individual patient might lead to variation in different disease outcomes in that patient, or perhaps in patient partners following transmission of the pathogen. Single nucleotide polymorphisms exist in strains from clinical samples that may significantly affect the biology of the pathogen in vivo. An example of this was first uncovered in patients persistently infected with a serovar I strain (48). These individuals were positive for the same serovar of *C. trachomatis* for many months or even years, and a subsequent study showed that the infecting chlamydiae expressed a variable phenotype in culture over the course of their infections. Isolated and cloned *C. trachomatis* from these individuals expressed either a fusogenic (multiple inclusions in a single cell will fuse to form a single inclusion) or nonfusogenic inclusion phenotype, which is correlated with the presence or absence of an inclusion membrane protein, IncA (CT 119) (49). A retrospective analysis of patients infected with IncA-negative (i.e. nonfusogenic) strains demonstrated that this genotype was associated with lower infectious organisms and clinical symptoms in the patient

(50). Subsequent work demonstrated that individual strains infecting patients can be a mixture of IncA-negative and IncA-positive clones, and that genomic variation at this loci likely leads to switching of this phenotype in vivo (51).

A second gene that varied in the patients infected with a mixture of IncA-positive and IncA-negative, highly related strains was CT135, a hypothetical gene that was variably intact or interrupted in different infecting strains. While the function of CT135 remains unclear, its role as a virulence factor was demonstrated by Sturdevant et al. (52), who showed that laboratory strains of *C. trachomatis* contain subpopulations that have deletions in this open reading frame, and that these deletions lead to differences in the ability of the pathogen to colonize the upper genital tract in a murine model system. These studies support a hypothesis that patients are likely infected with pools of strains that vary at different positions in the genome, and that these differences affect tissue tropism and other aspects of disease. Other phenotypes might also vary in vivo (53)and there is good evidence that many routinely used *C. trachomatis* strains exist as pools of organisms that differ slightly at many positions in the genome. This is also supported by primate studies that show subtle genetic differences can lead to significant differences in pathology in a primate model of Trachoma (54). The study of additional areas of genomic variation continues in many laboratories (42, 54, 55), and it is likely that elucidation of the functions associated with such differences will greatly enhance our ability to understand differences in pathogenesis among strains.

Efforts to Develop Molecular Typing Strategies

Historically, *Chlamydia* infections of many species were diagnosed through the use of specific antibodies to either the group-specific lipopolysaccharide molecule, or the serovariant major outer membrane protein. Highly specific antibodies were generated for these studies, and such analyses remain useful in some diagnostic and research settings. Nucleic Acid Amplification Test (NAATS) based approaches for diagnosis of infection followed the advent of PCR, leading to urine-based diagnostics that are minimally invasive. While there have been examples of problems with these assays, in general they are excellent tools for diagnostic analysis and epidemiology. However, these analyses remain limited to plus/minus-type infection studies, where very little of the nature of the infecting organism can be assessed. As the body of knowledge of individual *Chlamydia* genome sequences has expanded, different strategies have emerged to facilitate both identification of *Chlamydia*-infected individuals and to work toward mechanisms to correlate individual chlamydial genotype with a specific pathologic process. The following paragraphs will address these new developments.

Multilocus Sequence Typing

The multilocus sequence typing (MLST) approach evaluates genetic polymorphisms in a relatively limited number of loci, with a goal of evaluating as many group-specific differences as possible. A variety of multilocus sequence typing profiles for chlamydiae have been developed by different laboratories, each based on a set of between 5 and 7 different genetic sequences in the *Chlamydia* genome (14, 56-58) (Table 1.2). A

different set of assays are based on tandem repeat sequences (e.g. multilocus variable number tandem repeat analysis) (59, 60) which can also be used as epidemiologic markers for different strains. While these strategies are clearly useful in the understanding of epidemiologic patterns in chlamydial infections, their utility in understanding differences in disease progression in patients remains to be elucidated. It is likely that continued advances in next-gen sequencing technologies will significantly affect our ability to conduct these analyses on a larger number of loci, leading to clearer definitions of strains for epidemiologic purposes.

Strain Evolution and Emergence

As mentioned earlier, original understanding of *C. trachomatis* strain variation was limited to the analysis of the different serotypes, which led to ~18 canonical serovars. As sequencing technologies improved, variation at many loci (e.g. *trp*, plasticity zone, the Pmps) led to understanding that variation clearly extended beyond serotype, and that many individual genetic variations might correlate with differences in disease and tissue tropism. Whole genome sequencing has greatly expanded our understanding of this issue. Sequence analysis has led to the identification of chlamydial genomes that are clear mosaics of different canonical serotypes, with variation not being observed at loci that have been previously associated with differences in pathogenesis or serovariation. The first of these were uncovered by Jeffrey *et al.* in 2010 (55), who determined that an IncA-negative variant strain (Ds/2923) contains a D *ompA* gene within a genetic background more closely related to E and F strains (Fig. 1.1). This manuscript also shows

evidence of recombination within an F variant strain, in regions of the chromosome that are different to those shown in Ds/2923. Work by Dean and colleagues also addresses this issue – these individuals have identified a unique recombinant strain that is a hybrid between strains of serovars D and L2, resulting in a strain with unusual growth characteristics and an apparently 'hyper virulent' phenotype. Notably, this recombinant L2 strain has an intact toxin gene (CT166; Table 1.1), which is absent in other LGV strains as discussed earlier in this chapter, and is associated with severe hemorrhagic proctitis in vivo, and an in vitro cytotoxicity not seen in other L2 strains (61).

Examples of apparently random, in context to mechanism, recombination across the genome of clinical chlamydial isolates was confirmed and greatly expanded by Harris *et al.* in 2012 (62), who added a large and significant list of sequenced strains to the genome database. These studies support a model in which recombination is widespread in *C. trachomatis* strains, likely occurring in patients, and perhaps resulting in phenotypic switching among these generally closely related strains. Issues that remain in this area include the mechanism of DNA exchange by chlamydiae, the actual role of this recombination in chlamydial biology and disease, and the possibility that recombination hotspots are used during chlamydial lateral gene transfer (63).

While much of the discussion here centers on gene-level or genome-level changes that affect pathogenesis of individual chlamydial strains, the role of small variations in genome structure, or simply the emergence of otherwise highly related strains in a

susceptible host population, remains important in this system. For example, a unique LGV isolate (L2b) has been found recently in Europe and Canada (64). The L2b *ompA* sequence of the L2b strains was identical to that found in strains archived in San Francisco in the 1980's, indicating that this strain lineage has been expanding into susceptible populations for at least 25 years. Harris and colleagues sequenced 12 contemporary isolates of the L2B lineage when they conducted their large-scale genome sequencing study (62), and showed that there was a maximum of 19 sequence differences between the members of that lineage. The precise genetic or social changes that led to the expansion of this clone remain unclear. However, there are differences in the *ompA* gene that likely reflect differences in MOMP antigenicity. Also, as discussed above, previous work in a primate system suggests that such differences can be critical to immune avoidance *C. trachomatis* infections (54). The L2b story likely represents a very good model of how *C. trachomatis* strains ebb and flow among patient populations, and the associated slow sequence changes that occur is complemented by recombination events between strains in multiply-infected individuals, leading to perhaps a larger reshuffling of genetic and phenotypic differences among *C. trachomatis* in human populations.

Analysis of host genes that participate in chlamydial disease pathogenesis plus the radical technological advances in sequence analysis has allowed a novel field to develop, the field of 'public health genomics'. The driving concept in this area is the integration of individual genome sequences and human genetic polymorphisms into the therapeutic

options that might be most applicable in a single infected patient. This approach is a developing area in chlamydial biology, but researchers still have a long way to go. Several different laboratories have worked in this area, with a goal of identifying host genetic variants that are associated with either enhanced pathology or reduced disease in the patient population (Table 1.3). Much of these data can be compared to experiments with murine model systems, either through analysis of knockout mice or analysis of microarray data. These are particularly challenging experiments, as the readout needs to be clearly defined and complicating comorbidity determinants have to be carefully addressed. The accumulating data suggests that variation in several different genes encoding proteins participating in the immune response are associated with disease severity, including genes encoding HLA molecules, cytokines and Toll-like receptors. As technologies for accumulating and correlating data mature, and as individuals in the field develop standardized metrics for integrating these data, increased clarity in these areas will emerge. It is anticipated that this field will move toward a clinically oriented integration of the host gene structure, chlamydial genes and genomes, the host microbiota and the health history of the patient, and that this integration will lead to more successful prediction of the patient prognosis following infection by *C. trachomatis* in the genital tract.

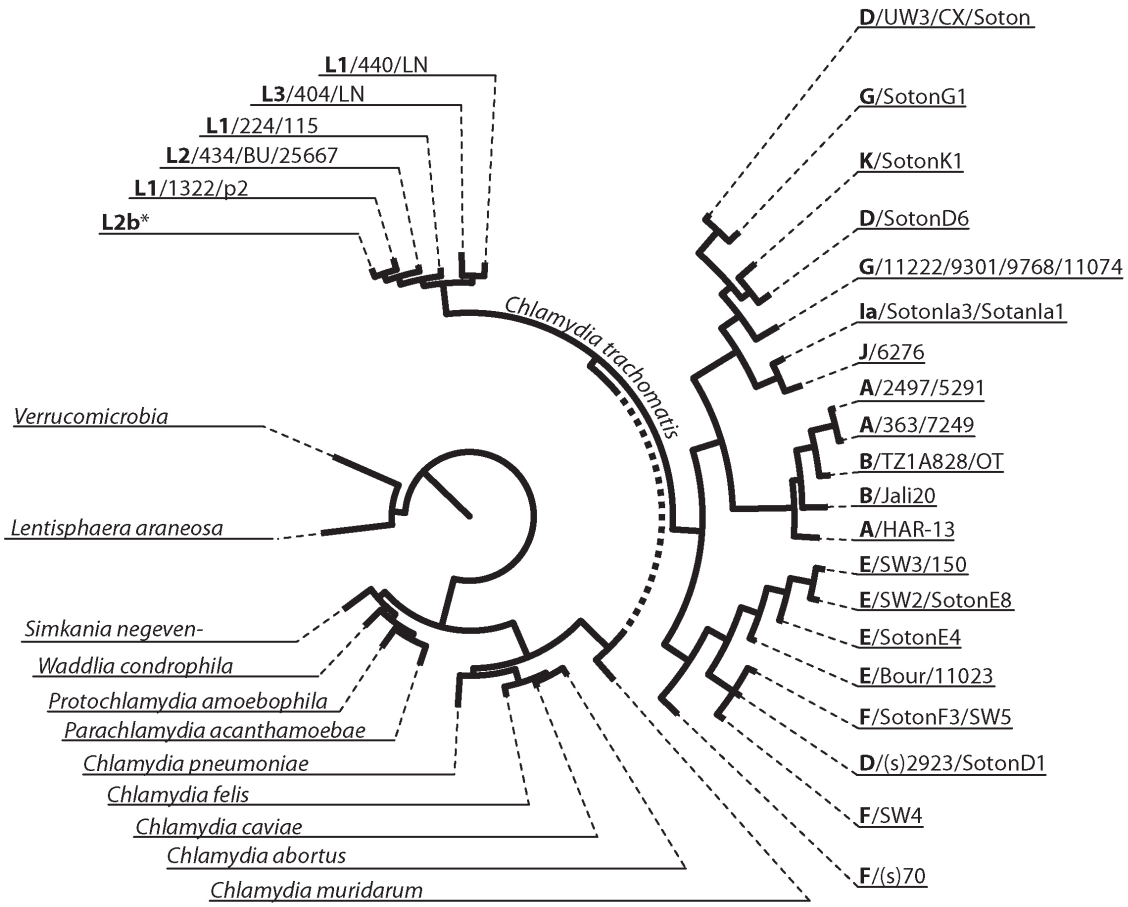

Figure 1.1. Phylogeny of *Chlamydia* and closely related species.

Whole genome phylogeny of all chlamydial species, and closely related species, with strain resolution for *Chlamydia trachomatis.*

Table 1.1. Identification of C. trachomatis open reading frames with differences of over 2% across all sequenced genomes

ORF	Protein	DNA %diff	AA %diff	Gene Name	Category/Function	Reference
CT046	NP_219549.1	2.44	1.34	hctB	histone-like protein 2	(65)
CT049	NP_219552.1	6.56	10.17		hypothetical protein	(66)
CT050	NP_219553.1	6.39	18.73		hypothetical protein	
CT051	NP_219554.1	5.21	10.38		hypothetical protein	
CT144	NP_219647.1	2.57	3.99		hypothetical protein	(67)
CT166*	NP_219669.1	19.73	15.06		glucosyltransferase	(68)
CT173*	NP_219677.1	20.02	21.25		hypothetical protein	
CT413	NP_219923.1	6.4	1.45	pmpB	membrane protein	(69)
CT414	NP_219924.1	5.3	1.91	pmpC	membrane protein	(69)
CT442	NP_219954.1	2.14	4.15		hypothetical protein	(70)
CT456	NP_219969.1	3.08	8.38		hypothetical protein	(67)
CT619	NP_220136.1	3.24	5.22		hypothetical protein	(67)
CT649	NP_220167.1	2.18	3	ygfA	synthetase	
CT651	NP_220169.1	2.21	2.37		hypothetical protein	
CT652	NP_220170.1	2.37	1.83	recD_2	exodeoxyribonuclease V	
CT675	NP_220194.1	2.09	1.04	karG	ATP phosphotransferase	(67)
CT677	NP_220196.1	3.29	1.14	frr	ribosomal release factor	
CT679	NP_220198.1	3.54	2.59	tsf	elongation factor	(67)
CT680	NP_220199.1	4.03	3.59	rpsB	30S-S2 ribosomal	
CT681	NP_220200.1	5.98	8.56	ompA	membrane protein	
CT748	NP_220267.1	2.04	0.6	mfd	transcription-repair	
CT852	NP_220374.1	2.43	3.23	yhgN	YhgN family	(67)
CT869	NP_220391.1	3.97	2.37	pmpE	membrane protein	(69)
CT870	NP_220392.1	5.73	4.23	pmpF	membrane protein	(69)
CT872	NP_220394.1	2.55	2.05	pmpH	membrane protein	(69)

Table 1.2. Summary of contemporary sequence typing systems for genotyping different C. trachomatis strains

Group [reference]	Method	Loci	# Strains	# Sequence Types	Description
Klint et al. (56)	MLST	CT046, CT058, CT144, CT172, CT682	47	32	5 ORFS with hypervariable regions, including 3 hypothetical genes
Pannekoek et al. (14)	MLST	CT587, CT855, CT0003, CT498, CT742, CT371, CT198	26	15	7 housekeeping genes
Dean et al. (71)	MLST	CT432, CT376, CT245, CT653, CT332, CT781, CT209	87	44	7 housekeeping genes
Pannekoek et al. (72)	omp1-VNTR	CT642-CT643 (intergenic), CT259-CT260 (Intergenic), CT172, CT681	93	87	4 loci containing variable tandem repeats, including intergenic regions and ompA
Pedersen et al. (59)	MLVA-5	CT046, CT456, CT632-CT633 (intergenic), CT868, CT872	43	15	5 loci containing variable tandem repeats, including intergenic regions

Table 1.3. Host genes associated with variation in clinical presentation following chlamydial infection of humans

Gene	Correlation shown; system	Reference
Haptoglobin type E	active trachoma	(73)
sickle cell trait HbAS	none	(73)
NLRP3	abdominal pain following Ct infection	(74)
miRNA-146A	none; genital tract	(74)
TLR-4 mutation	none; genital tract	(75)
MBL2	antibody response; C. pneumoniae infection	(76)
multiple SNPs in PRR's	trend toward relationship; tubal pathology	(77)
IL1-B, IL-1R	no association; tubal pathology	(78)
IL-8, CSF2, MMP9	Assoc. w/ trachomatous scarring	(79)
MMP9	scarring; ocular infection	(80)
TLR-4	pathogen accumulation; tubal factor infertility	(81)
MBL-low	susceptibility; tubal factor infertility	(81)
CCR5, TLR-2	none; tubal factor infertility	(81)
TLR-2 haplotype 1	reduced susceptibility; tubal pathology	(82)
MICA	host susceptibility; genital tract infection	(83)
IL-10, Ifn-g	immune response variation; tubal factor infertility	(84)
TNF-a, IL-10	severity of tubal damage; tubal factor infertility	(85)
HLA class I and II	chlamydial genital tract infection	(3)
HLA variants	recurrent infection	(86)

Culture-independent Sequence Analysis of *Chlamydia trachomatis* in Urogenital Specimens Identifies Regions of Recombination and In-patient Sequence Mutations

Timothy E. Putman, Robert J. Suchland, John D. Ivanovitch, and Daniel D. Rockey

Microbiology , Society for General Microbiology

12 Roger Street

London, UK

Volume 159, pages 2109–2117.

ABSTRACT

A culture-independent genome sequencing approach was developed and used to examine genomic variability in *Chlamydia trachomatis*-positive specimens that were collected from patients in the Seattle, WA, USA, area. The procedure is based on an immunomagnetic separation approach with chlamydial LPS-specific mAbs, followed by DNA purification and total DNA amplification, and subsequent Illumina-based sequence analysis. Quality of genome sequencing was independent of the total number of inclusion-forming units determined for the sample and the amount of non-chlamydial DNA in the Illumina libraries. A geographically and temporally linked clade of isolates was identified with evidence of several different regions of recombination and variable ompA sequence types, suggesting that recombination is common within outbreaks. Culture-independent sequence analysis revealed a linkage pattern at two nucleotide positions that was unique to the genomes of isolates from patients, but not in C. trachomatis recombinants generated in vitro. These data demonstrated that culture-independent sequence analysis can be used to rapidly and inexpensively collect genome data from patients infected by C. trachomatis, and that this approach can be used to examine genomic variation within this species.

INTRODUCTION

Chlamydia trachomatis is an obligate intracellular bacterium and the most prevalent sexually transmitted bacterial infection of humans (Centers for Disease Control and Prevention, 2011). Studying these genetically difficult organisms has become less

arduous, with the recent advent of high throughput sequencing technologies and workable genetic systems (87). While acquisition by chlamydiae of genetic material from other species is considered rare (32), recombination within C. trachomatis, both in vitro and in vivo, appears to be relatively common (32, 61, 62, 67, 88-90). Recombination has been documented between the lymphogranuloma venereum and urogenital biovars, fusogenic and nonfusogenic strains (91), and strains with tropism for different tissues (55, 62). The implications of these recombination events are only beginning to be elucidated.

The application of high throughput sequencing for *Chlamydia* is challenging because of the extensive culturing process required for generating sufficient quantities of genomic DNA for sequencing. Moreover, extended *in vitro* culturing may lead to changes in the chlamydial genome due to the absence of host immune pressures (92). To address these issues, a system for sequencing chlamydial genomes without the need for growth in cell culture was developed. Immunomagnetic cell separation (IMS), utilizing antibodies specific to chlamydial lipopolysaccharide, was used to isolate *Chlamydia* directly from clinical samples. Multiple displacement amplification (MDA) was then employed to amplify the isolated chlamydial genomic DNA and produce sufficient quantities of DNA for high throughput genomic sequencing. Utilizing this protocol, we sequenced and analyzed the chlamydial genomes collected from 10 clinical endo-cervical swab specimens isolated from the Seattle area collected between April 1993 and January 1998. The results revealed a geographically linked clade of similar chlamydial genomes

with variable *ompA* sequences, distinct recombination blocks, and evidence of in-patient mutation.

METHODS

C. trachomatis collection inclusion forming units (IFU) determination, and serotyping.

De-identified patient materials used for this study were selected from frozen specimens in the University of Washington Chlamydia Repository. This resource contains over 15,000 patient samples including isolates from culture-documented patients attending Seattle–King County Health Department sexually transmitted disease clinics from 1988–2006 (93). In the original diagnostic procedure, McCoy cells in 96-well trays were infected in a standardized dilution series with patient material, and incubated for chlamydial growth. Cells were fixed with methanol and inclusions were identified and quantified using genus-specific FITC labeled anti-chlamydial lipopolysaccharide monoclonal antibody (clone 2C1; Washington Research Foundation, Seattle, WA). Monolayers having 100 inclusions or less were individually counted to determine the inclusion number per swab. The average inclusion count per well was calculated by enumerating 3 fields on a Zeiss microscope, and multiplying the average count by predetermined conversion factors to calculate the number of inclusions per well (94). Serotypes of primary diagnostic cultures were determined using the methods of Suchland et al. (95). After culture manipulations for diagnostic analysis, primary sample material was returned to the repository and stored until used in this study.

Immunomagnetic cell separation of chlamydial elementary bodies.

Cervical swabs initially used for culture-based diagnostics were stored at -80C prior to processing. Specimens were selected for analysis on the basis of chlamydial IFU values (ranging from 4 ifu-15,000 IFU per swab), and serovar (Fig. 2.1). The identified specimens were removed from the freezer, thawed at 37ºC, and placed on ice. Three Pyrex beads (2mm diameter) were added to 1 mL of sample in PBS and vortexed (4ºC, 33 RCF for 3 minutes). Samples were then split into two aliquots, with one being stored at -80ºC and the other being used for the enrichment of chlamydial elementary bodies (EBs) through immunomagnetic separation (IMS). Initial enrichments used primary mouse antibodies directed at LPS (E6H1: (51)) or the major outer membrane protein (MOMP; (96)). In each case, antibody concentrations were standardized to a final concentration of 3.1 ng/µL. The primary antibody/cell suspension was incubated for 30 min at 4ºC on a rotating shaker (33 rpm), and 1.0 mL of MACS buffer (Miltenyi biotech; PBS (pH 7.2), EDTA (2 mM), FBS (0.55%)) was added. The sample was vortexed gently and centrifuged at 20,800 RCF for 10 minutes at room temperature. The supernatant was aspirated and pellet resuspended in 1.0 mL MACS buffer. IgG-specific secondary antibodies, conjugated to 50 nm paramagnetic particles (Miltenyi), were added and incubated using the same conditions as the primary antibody incubation. The labeled cells were loaded onto a disposable column and placed in a strong magnetic field (MACS Multi Stand Cat #003754). Bound material was washed twice with PBS, the columns were removed from the magnetic field, and target material collected with MACs buffer.

DNase treatment for removal of host DNA

Eluted material from the chlamydial cell separation was incubated (37ºC, 5% CO2, 60 minutes) with RQ1 DNAse (Promega). RQ1 stop buffer was added and the reaction was incubated at 65°C for 10 minutes, with gentle mixing. Total DNA was then extracted from the enriched EBs using the DNeasy Blood & Tissue kit (Qiagen Cat # 940014) as described previously (55).

Quantification of chlamydial genome copies

Real-time PCR analysis of this EB preparation was carried out using the Applied Biosystems Taqman Fast Universal PCR Master Mix with primers defining unique sequences of the 2nd allele (groEL_2; CT604) of the *C. trachomatis* chaperone gene Hsp60 (primers: CTHsp60F GATTCTCTCTTCCTCGCTGTCTTC, CTHsp60F GAGGGTTTTCCCTGTCTGTGC). A plasmid containing the groEL_2 ORF was created, quantified, and used as a standard curve in quantifying genome copy number from sample DNA.

Multiple displacement amplification (MDA).

Column-purified EBs were removed from -80°C storage, thawed quickly at 37°C, and placed on ice. Samples were centrifuged at 21,500 RCF for 10 minutes, the supernatant was aspirated, and the pellet re-suspended in PBS. MDA was performed as described by

the manufacturer (Qiagen Repli-g kit), using a 90 minute reaction time at 30°C. Amplified material was then stored at -20°C.

Genome sequencing.

MDA-amplified genomic DNA preparations from clinical swab samples were prepared for multiplex Illumina sequencing using the NEBNext DNA Library Prep Master Mix Set for Illumina (Cat# E6040S) kit and according to manufacturer-specified protocols (Illumina Inc., San Diego, CA). Sequencing was performed on the Illumina HiSeq 2000 platform at the Center for Genome Research and Biocomputing (CGRB) Core Lab facility at Oregon State University. Multiplexing of samples was conducted using a commercial kit (Illumina Multiplexing Sample Preparation Oligonucleotide Kit Cat# PE-400-1001). The clinical isolates were sequenced in two different groups; both were single-end, multiplexed runs with either 51 base pair read-lengths or 101 base read lengths, as indicated (Fig. 2.1). For two samples (J/31-98 and F/11-96), single IMS preparations were divided in half, with each half being processed independently through all subsequent steps. These parallel runs were used to assemble and crosscheck the sequence analysis. The initial 51-cycle run was completed with 2 multiplexed samples per lane, while the 101 cycle runs were performed with 3 and 4 multiplexed samples per lane.

Genome assembly and sequence analysis.

Genome sequence assemblies were done using the reference guided assembly software package Maq (97). Loci that could be not assembled using Maq were resolved by

assembling scaffolds from contiguous segments of sequence generated via the *de novo* assembly software package VCAKE (98) or through the use of an ad hoc Perl script pipeline. Any remaining ambiguous gaps in the assembled contiguous drafts were resolved through the use of traditional PCR and Sanger sequencing to walk across the ambiguous region. Macvector software (Macvector Inc.) was used for sequence manipulation and draft editing, and the MAFFT software package (99) was used for whole genome alignments and phylogenetic analysis. Specimens with final genomic assemblies containing ambiguous bases comprising over .05% of the genome were not included in whole genome comparisons (Fig. 2.1) All ORFs were annotated based on the D/UW-3/CX genome as published by Stephens and colleagues (8).

Regional recombination analyses.

A sliding window Perl script described by Jeffrey et al. (55), was initially used to compare compiled sequence information, for variation and recombination, against a database consisting of previously published chlamydial genome sequences (D/UW3 (NC_000117(8)), J/6276 (NCBI Taxonomic ID 564416 (51)), G/9768 (NC_017429 (55)), E/11023(NC_017431 (55)), and/or F/70 (NCBI Taxonomic ID _564418 (51)). Genomes were aligned using MAFFT and custom Perl scripts. To identify regions of recombination, the Recombination Identification program (RIP) available on the HIV sequence database website (http://www.hiv.lanl.gov/) was used. The software package PhiPack was used to assess overall probability of recombination in each region based on ascertaining the Pairwise Homoplasy Index (PHI) statistic for each region (100). The R statistical software

package was used to generate data plots for analysis (R_Development_Core_Team, 2008).

RESULTS

Comparison of different monoclonal antibodies for isolating C. trachomatis elementary bodies (EBs) from a clinical swab sample.

Optimization of the culture-independent sequencing technology began with a comparison of available surface-reactive monoclonal antibodies, to determine if antibody specificity affected EB harvesting efficiency. Monoclonal antibodies EVI-HI (specific to a genus common chlamydial LPS epitope), L2-I-V (specific to L2 MOMP), and HV-AV (Specific to A,C,H,I and J MOMP) were each tested as possible primary antibodies for IMS. HV-AV was used as a negative control as it is non-reactive to L2 MOMP. The use of the anti-LPS primary mAb, resulted in the elution of 95.5% of the genome copies in test samples, while the use of the surface-epitope-targeting anti-MOMP antibody L2-I-V led to the recovery of far less material in these assays (Fig. 2.2). Experiments in which either the primary or secondary antibodies were excluded demonstrated that the overall procedure was efficient and specific (Fig. 2.2B). Based on these results, we elected to use the anti-LPS mAb for all subsequent immunomagnetic separations.

Amplification of chlamydial genomes in samples with varying numbers of inclusion forming units (IFUs).

The minimal necessary number of chlamydial genome copies required for culture-independent Illumina sequencing was determined by creating a set of standards with

different amounts of genomic DNA and using these in independent MDA reactions. Amplification of samples containing less than 10 genome copies provided sufficient DNA for sequencing (Data not shown). Sufficient genomic DNA for sequencing and complete assembly was collected and amplified from samples containing maximal numbers of infectious chlamydiae (i.e. 15,000 IFU in specimen F/6-94), to as few as 4 IFUs per swab (i.e. specimen J/31-98 and J/27-98; Fig. 2.1).

Mitigating contamination from non-chlamydial DNA.

The clinical samples used for amplification and sequence analysis were expected to contain several nonchlamydial sources of DNA. For example, preliminary qPCR on post-IMS samples revealed a substantial amount of human DNA that was carried through the separation process. A DNase incubation step was included post-IMS to help minimize contaminating host DNA. Subsequent qPCR performed after the DNase treatment demonstrated that this was successful at reducing host DNA (not shown).

To assess the abundance of DNA from non-chlamydial sources in our final Illumina-generated sequence data, a total of 1000 randomly-selected reads from each of the clinical swabs were used in a BLAST-base search against the non-redundant nucleotide sequence database. The mentioned DNase reaction lead to minimal contribution of host DNA in sequence reads (Fig. 2.1). The percentage of non-chlamydial reads from all sources ranged from ~4% in the case of D/13-96, to ~75% and ~95% in J/31-98 and J/27-97 respectively. Levels of each of the individual contaminants were variable across all the sequenced samples (Fig. 2.1).

Comparative analysis of generated genome sequences.

Ten specimens were subjected to IMS purification, MDA whole genome amplification, and Illumina-based sequencing. For 8 of these specimens, unambiguous sequence data were collected for greater than 99.98% of the genome. The range of ambiguous bases per genome ranged from zero to 1,197. In two cases a parallel sequencing approach was used, leading to the generation of two independently assembled genome sequences that were compared against each other. This approach led to the lowest numbers of ambiguous bases, in specimens F/11-96 and J/31-98. For example, in F/11-93 and J/31-98, ambiguous bases were reduced from 100 (Run 1) and 168 (Run 2) to 0 and 3 respectively, when sequence data was combined.

The assembled unambiguous sequence data were compared to a set of chlamydial genomes published by our group and other investigators (Fig. 2.3)(8, 55, 62). In all specimens, the chlamydial plasmid was present and displayed over 99% identity with that of strain D/UW3. None of the changes present led to truncations in any of the plasmid coding sequences. In every specimen, CT135, a gene recently identified as necessary for *in vivo* pathogenicity, was intact and identical (52, 92). CT119, the ORF encoding IncA, was also intact in each of the genomes collected from these swabs.

Six assembled genomes carried a D, F, or J *ompA* genotype, but otherwise shared a highly similar genomic background (F/1-93, F/2-93, F/6-94, F11-96, D/13-96, J/27-97;Fig.

2.4). Each of these specimens were closely related to strain D(s)/2923, an IncA-negative isolate collected during the same time period and geographic area as the specimens analyzed in this study (55, 101)(Fig. 2.4). This group of specimens will be referred to as the D/13-like clade, and D(s)/2923 will be included in all subsequent comparisons.

While the D/13-like clade was common in this population of patients, this was not the only sequence group identified in this set of specimens. The four remaining specimens (D/14-96, Ia/20-97, J/31-98, and E/12-94) contained *C. trachomatis* with genome sequences that were significantly different than the D/13-like strains and aligned more closely with other published chlamydial genomes (Fig. 2.3).

Several different regions of apparent recombination were identified in the D/13-like clade. The most obvious example involved sequences surrounding *ompA* (CT681) the gene that codes for the major outer membrane protein (MOMP; Fig. 2.4). The *ompA* gene from the D/13-96 genome is identical to that found in D(s)/2923, while F/2-93, F/6-94, F/11-96 had similar flanking sequence but contained *ompA* sequence of a different serovar. Specimen J/27-97 was included in the *ompA* comparative analysis because sequences surrounding *ompA* were fully resolved, against the background of a poorly completed genome (Fig. 2.1). Analysis of the J/27-97 *ompA* region showed flanking sequences highly similar to the D/13-like clade, with a *ompA* genotype similar to the J serovars (Fig. 2.4). The location of the apparent integration sites for these *ompA* regions

were different in some sequences, indicating functional selection of that region by gene transfer that was not associated with sequence specific recombination hot spots.
Two other regions of apparent recombination within this group include open reading frames (ORFS) CT050-055 and CT193-218 (Fig. 2.4). These regions showed differences at their margins within the different specimens, indicating that they likely originated from distinct recombination events and were not the product of clonal expansion of a single recombinant lineage.

The genome sequence of F/6-94 is similar to D/13-96 in sharing several patterns of recombination. However, F/6-94 contains sequence representing a different integration event, leading to a large region of the chromosome (~7%, CT833-872) containing sequence with low similarity to the D13-like clade (Fig. 2.4). All of the examined regions (CT050-55, CT193-218, CT681, CT833-872) had significant support for recombination based on grouped patterns of homoplasic SNPs ($p < 0.05$).

The genome sequence carried in specimen J/31-98 was most similar to a previously published strain collected from a patient in the Seattle area, J/6276 (51) . This strain has a short region of the genome that is less related to J/6276, including ORFs CT660 through CT662. This ~12 kb region is most similar to the sequences within the D/13-like clade, and likely represents a short region of recombination that led to the observed differences between J/31-98 and J/6276.

Analysis of mutations within C. trachomatis from a single specimen.

For most of the specimens, a single sequencing run was completed following IMS enrichment of samples. This resulted in many instances of unsolvable bases, based on ambiguity among the collected reads at that position (Fig. 2.1). While in some cases these might represent a nonclonal population with a single patient, it was not possible to separate genuine polymorphisms of sequence from errors generated in the generation of sequence data. This problem was addressed in two specimens (F/11-96 and J/31-98) by independently sequencing the genomes twice and crosschecking the results with an ad hoc Perl script. Using this approach, sequences that represented genuine polymorphisms were represented similarly in both sequencing runs. For specimen F/11-96, this approach led to a complete genome sequence with no ambiguities, leading to the assessment that this patient was infected with a clonal *C. trachomatis* population. In contrast, examination of sequence reads for specimen J/31-98 identified a total of 3 variable bases, demonstrating a low level of clonal variation in the patient (Fig. 2.1). Each variable base is associated with amino acid changes in the respective coding sequences (CT042: glycogen hydrolase; CT308: ATP Synthase Subunit A; CT449: hypothetical.

The coding sequences for ORFs CT308 and CT449 are separated by approximately 150,000 base pairs. Because of the recently established paradigm that these organisms commonly recombine *in vitro* and *in vivo* (51, 55, 62, 63), we hypothesized that these two mutations would segregate independently in published *C. trachomatis* genomes.

However, comparison of these variable loci across all sequenced chlamydial genomes demonstrated that the different nucleotide polymorphisms are absolutely linked in published genomes from clinical strains. In this set of 47 published genomes collected from patients, 21 of the genomes carried an 'A' at position 2 in CT308, and a 'C' at position 320 in CT449. In contrast, the remaining 26 genomes had a 'G' at position 2 in CT308, and a 'T' at position 320 in CT449 . Our sequence data for J/31-98 contains consistent mixtures of these two polymorphic nucleotides in each of the two independent assembled sequences. Therefore, complete and published genome sequences from patients had specific combinations of nucleotides, each leading to a different amino acid in the encoded protein, while our culture-independent clinical specimen had variability at that position. In contrast, our group recently generated complete genome sequences from 10 recombinant clones generated in vitro, from parents that were members of the 47 strains discussed above (Brendan M. Jeffrey, 2013 in press). Examination of these polymorphic nucleotide positions in these recombinant progeny demonstrated that 3 of 10 independent *in vitro* generated recombinants had nucleotide combinations that were not consistent with that seen in the clinical strains. Therefore, the absolute linkage found at these physically distant nucleotide positions is only found in chlamydiae growing under undetermined selective pressures *in vivo*, and not in recombinant clones generated *in vitro*.

DISCUSSION

Culture and propagation of chlamydiae can be time and reagent intensive, and opens the door for attenuations *in vitro* that might alter the genome of chlamydiae found in the patient. To address these issues, we have developed a rapid, inexpensive method for sequencing chlamydial genomes using material collected directly from clinical swab samples, which can be used even in the presence of large amounts of non-chlamydial DNA contamination. While we have applied our method directly to uncultured patient specimens, a similar approach should also be useful for sequencing hard to grow specimens or low-passage *in vitro* generated mutant strains. For the latter approach, we consider the use of parallel independent amplifications and sequence analysis to be a relatively inexpensive and ultimately cost-saving approach to generating unambiguous DNA sequence. Contamination is a significant challenge faced when dealing with samples directly collected from patients. Enrichment of chlamydial EBs with the LPS-specific antibody, DNAse treatment of samples to reduce the level of nonchlamydial DNA, and the strength of high throughput sequencing technologies, collectively led to high numbers of sequence reads for each position in the target genome. This was accomplished against a background of variable levels of contamination from the host, other microbes collected in the specimen, and template-independent products that can occur artifactually in MDA procedures (Fig. 2.1). Regardless of these challenges, non-chlamydial reads from all sources did not reduce the overall success of sequencing, or assembly, of any of the samples processed.

A very recent report by Seth-Smith and colleagues discussed a similar technology for creating culture-independent *C. trachomatis* sequences from clinical specimens (102). Our approach is very similar to that described by these investigators, with the exception that we have included a RQ1 DNase treatment that leads to near total elimination of host DNA (Fig. 2.1). While this treatment adds another step to the process of generating a complete genome sequence, the strength of this approach lies in the much higher percentage of *Chlamydia*-specific reads within each Illumina sequencing run, allowing the processing of more multiplexed samples per run.

Sequence ambiguities still existed in several of our assembled genomes, as is true with many published genomes available in the databases. However, we greatly reduced the numbers of ambiguities through the use of a parallel sequencing strategy, in which a single IMS-harvested *Chlamydia* preparation is subjected to two independent MDA reactions and sequencing runs. This approach is recommended for two reasons- first it yields a much clearer overall dataset, and greater confidence can be attributed to remaining polymorphisms, because these polymorphisms will remain variable in independent sequencing runs. In our data, sequence ambiguities ranged from 21-1366 nucleotides in genomes sequenced a single time, but were reduced to zero and three in the two genomes that were sequenced twice. Variations at the three remaining positions in specimen J/31-98 were consistent both in the identities of the variable bases and in the percentage of a specific base at that position. Additionally, the success of a particular sequence analysis was not directly associated with the percentage of

chlamydia-specific reads, as samples with as little as 25% chlamydia-specific reads were successfully sequenced (strain J/31-98). The most significant factor in generating dependable and unambiguous sequence was the use of two parallel, independent amplifications and sequencing runs within a single clinical specimen.

Using this method we elucidated a geographically and temporally linked clade of *C. trachomatis*, termed the D/13-like clade, that display shared apparent recombination blocks. While each of these specimens were collected in the Seattle area, a similar genome sequence was identified in the genomic survey by Harris et al. (62); their strain, D/SotonD1, was collected from the United Kingdom. Chlamydiae in these specimens display a highly similar overall genomic content to that of the previously sequenced isolate D(s)/2923 (55). Each of these specimens were collected in the mid 1990's from patients in the Seattle area (Fig. 2.1). Sequence variability within the clade can be attributed to both recombination and mutation, with examples of each being evident in each specimen. For example, two highly related genomes, D/13-96 and Ds/2923, have 126 sequence differences, and no apparent blocks of recombination that distinguish the genomes. One polymorphism between these genomes led to inactivation of *incA* in Ds/2923, while the coding sequence of D/13-96 is intact and expresses a properly localized IncA polypeptide, based on fluorescence microscopy (not shown). These data are consistent with previous results from our group, in which highly related strains that were either IncA-positive or –negative can exist within single patients (101). In the present case, highly similar chlamydiae with differing *incA* genotypes were identified in

different patients, with each sample collected in the Seattle area at time points just over one year apart. Collectively these results demonstrate that variability at *incA* is common and is found in similar specimens both in single patients and in related specimens being circulated in a geographically linked population.

The use of multiple sequence runs was critical to the solving of complete genomes, and for the identification of apparently genuine nucleotide polymorphisms within a selected genome. This was evident during analysis of J/31-98, in which three nucleotide positions were variable at similar rates in each of the independent Illumina runs. These analyses identified polymorphic nucleotides that are uniquely matched in each sequenced clinical strain, but not in in vitro-generated recombinant strains. Such results lead to the hypothesis that there are functional linkages between [V-type ATP synthase subunit A (CT308) and the hypothetical inclusion protein (CT449)] that are important *in vivo*, but it is challenging to design experiments to test this hypothesis. Further, it is not likely that uncovering of such a linkage between these two nucleotide positions could have been accomplished using any technology besides a culture independent sequencing approach.

Recombination is also evident within clade members. Recombination-driven *ompA* variability is evident in specimens D/13-96, F/2-93, F/6-94, F/11-96, and J/27-98, where unique exchange events were identified leading to a phenotypic change of serovar (Fig. 2.4). Changes were also evident in other regions of the chromosome, in events that

appear to have occurred independent of one another (CT050-055, CT193-218, CT826-874) (Fig. 2.4). For example, the region containing CT050-055 (Fig. 2.4) is very similar within the clade, but the margins of apparent integration differ for each sequence. The data from each of the D/13-like clade members suggest that variation within a common genetic background, similar to that described by Harris et al., was supplemented with independent recombination events involving common regions of the genome. These results demonstrate that recombination and mutation shape the overall mosaicism that is evident in the published *C. trachomatis* genome sequences, and support a hypothesis that the described mosaicism is the rule, not the exception, in chlamydial genome composition. The multi locus sequence-typing (MLST) scheme designed by Dean and colleagues in 2009 (58), supported grouping these genomes by their similar mosaic structure. However, while serotyping and MLST analyses remain useful in understanding variation among chlamydial strains, the underlying mosaicism, against the background of overall genomic synteny and sequence identity, can only be assessed via more extensive whole genome sequence analysis.

We have developed a rapid system for sequencing chlamydial genomes directly from clinical swabs without the need for growth in cell culture. This method could theoretically be used to isolate and expand the genome of any microbe with a surface antigen that is targetable by antibodies, directly from a clinical or mixed sample. Through its use we have elucidated a geographically and temporally linked clade of *C.*

trachomatis that has undergone widespread recombination and is variable in the *ompA* sequence type carried by its members.

The use of the recently developed transformation system, next generation sequencing, and culture-independent methods allowing the sequencing of genomes directly from clinical samples, are a powerful combination of tools that will allow us to answer questions about *Chlamydia* that have never been possible to address before. Future efforts will continue in the investigation of the mechanisms and significance of recombination in chlamydial populations, and how recombination and genomic variation pertain to disease outcome.

Figure. 2.1. Clinical swab sample classification and sequencing output distribution.

Clinical Swab Strain	F/1-93	F/2-93	F/6-94	F/11-96	E/12-94	D/13-96	D/14-96	Ia/20-97	J/27-97	J/31-98
Sample Collection Date	Apr-93	Mar-93	Sep-94	Mar-96	Oct-94	Mar-96	Mar-96	Apr-97	Sep-97	Jan-98
Infectious Chlam. /swab	1500	15000	15000	10100	15000	6600	15000	650	4	4
ompA	F	F	F	F	E	D	D	Ia	J	J
Chlamydia Host Microbial No Sig. Hit										
# Chlamydia reads x 10^6	21.9	10.6	42.9	22.8	19.0	32.4	30.6	21.5	1.9	12.2
Read Length bps	101	101	51	101	51	101	101	101	101	101
Number of Runs	1	1	1	2	1	1	1	1	1	2
Unsolved Bases	94	180	41	0	1197	21	395	102	1366	3
Accession Number	256263	256277	256278	256365	256279	256281	256288	256289	256290	256295

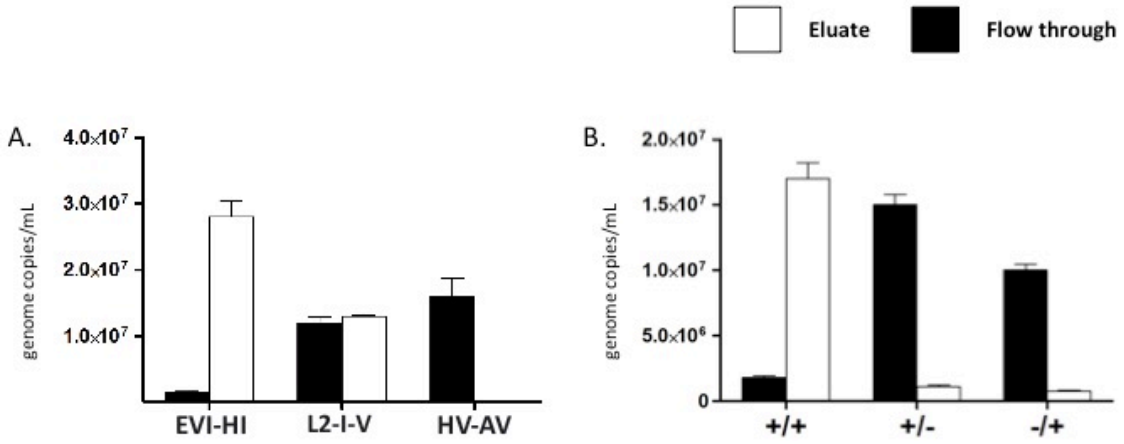

Figure. 2.1. Monoclonal antibody selection and verification.

Enrichment efficiency of C. trachomatis serovar L2, as a factor of surface-reactive mAbs. qPCR of CT604 was used to infer bacterial enrichment efficiency based on chlamydial genome abundance. The white bars indicate genome copy numbers in the eluate of an IMS reaction, while the black bars indicate genome copies in the flow-through. (a) A laboratory-generated swab sample was harvested using antibodies to chlamydial LPS (EVI-HI), serovar L2 MOMP (L2-I-V) or to a serologically different MOMP (HV-AV) and processed through the IMS protocol. (b) Successful IMS only occurred in the presence of both primary and secondary antibodies. +/+, Recovery using both primary and secondary antibodies; +/", no secondary antibody; –/+ indicates no primary antibody. This test was conducted using EVI-HI primary antibody.

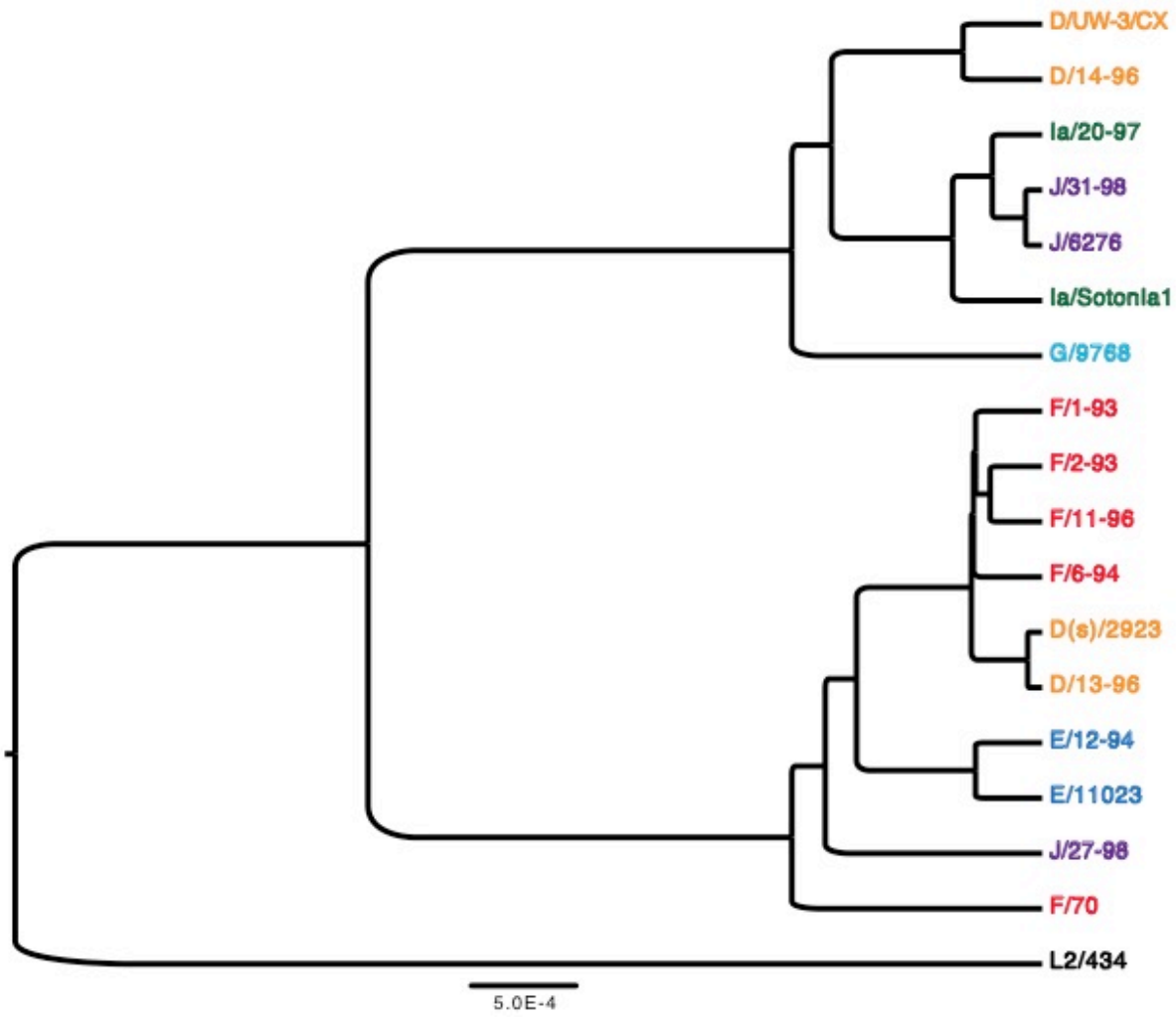

Figure. 2.3. MAFFT-generated whole-genome phylogeny aligning all 10 sequenced clinical swab samples with nine of the already published chlamydial genomes.

The duplicate rRNA sequences were removed to eliminate the repeat-associated error associated with highly homologous and non-informative sequence regions. The color and the first letter of each strain indicate ompA genotype. The branch lengths are proportional to the genetic distance between specimens, as indicated by the scale bar

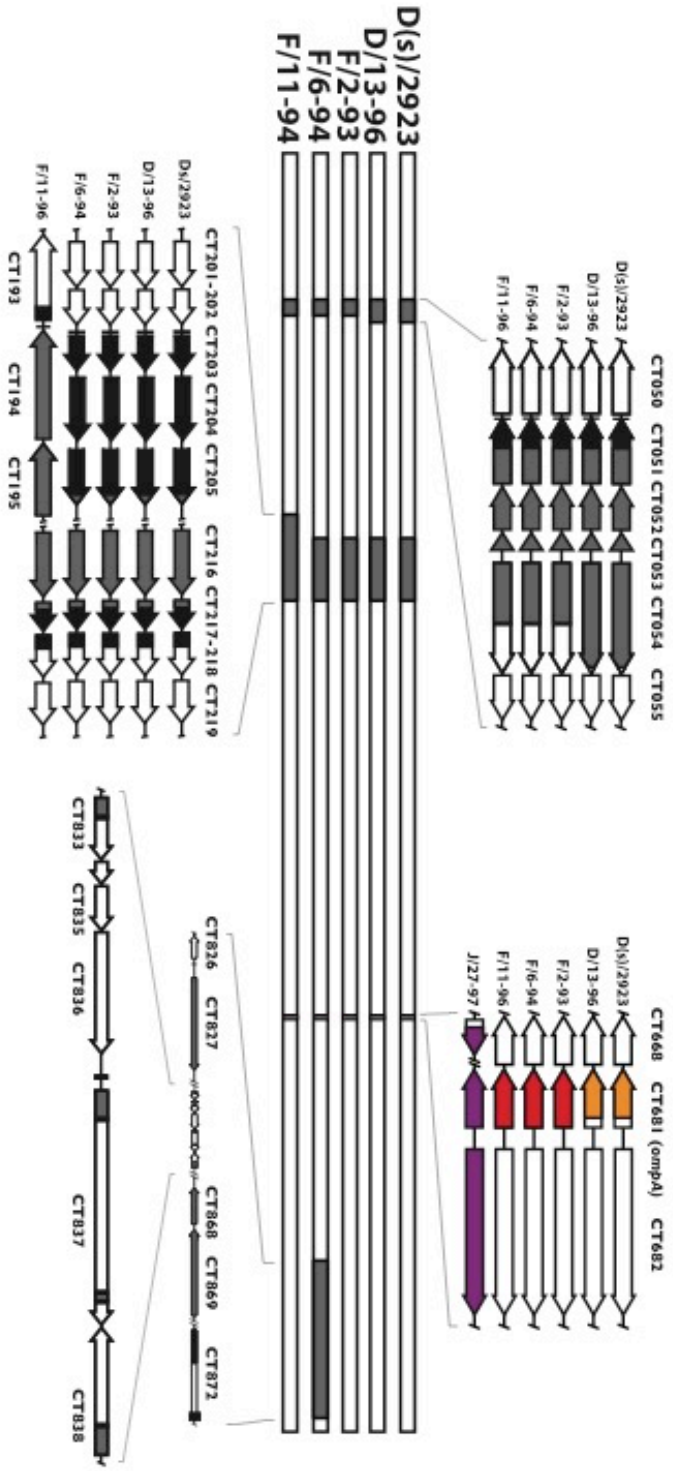

Figure. 2.4. Genome maps of clinical isolates examined in this study.

Each horizontal bar represents the complete genome sequence of the indicated strain or specimen, beginning with CT001. Sequence origin was identified by plotting informative sites using RIP. Regions colored grey indicate areas of apparent recombination that, in some strains, have margins that indicate independent recombination events. Black regions indicate conserved sequence between informative sites of sequences of different origin. The region showing the expanded sequences surrounding CT681 contains colour-coded sequence that identifies the ompA sequence origin (red, F; yellow, D; purple, J). J/27-97 is included in the ompA comparisons because the sequences shown were solved to a high degree of confidence, while the overall genome structure contained a relatively high level of ambiguous bases.

Culture-independent Sequencing of *Chlamydia abortus* Strains from the Western United States Identifies Intra-species Genomic Variation.

Timothy E. Putman, Kathy O'Reilly, Beth Mamer, Bernhard Kaltenboeck and Daniel D. Rockey

ABSTRACT

Ovine Enzootic Abortion (OEA) is a disease associated with infection of pregnant ewes by the bacterial pathogen *Chlamydia abortus.* The pathogen is present in sheep flocks in most countries, and can be a substantial burden both in terms of sheep health and economic productivity. In some areas, this disease is a major limitation to profitability. *Chlamydia*-induced abortion of sheep is relatively common in the Western United States. Using culture-independent technologies, recently developed in our laboratory, we have isolated and whole-genome sequenced 12 strains of *C. abortus* sourced from infected placenta tissues from farms, and acquired and sequenced 6 historical strains from a collaborating researcher's repository. Using comparative genomics we have obtained a broad view of the variation in genomic structure in populations of the pathogen across a large geographical region. This will help elucidate genetic differences that account for variation in the dynamics of natural infection, and infection in animals that have been vaccinated.

INTRODUCTION

Ovine Enzootic Abortion (OEA) is a disease associated with infection of pregnant ewes by the bacterial pathogen Chlamydia abortus. The pathogen is present in sheep flocks in most countries, and can be a substantial burden both in terms of sheep health and economic productivity (103-105). In some areas, this disease is a major limitation to profitability. *Chlamydia*-induced abortion of sheep is relatively common in the Western United States. Many states carry substantial burdens of this pathogen in sheep and, to a

limited extent, goatherds. In addition to economic loss by sheep farmers, *C. abortus* infection poses a risk of abortion to humans (106-108).

Control of this pathogen is challenging; neither infected male or female sheep are symptomatic, and pregnant ewes can appear completely normal prior to aborting their fetus. Often an "abortion storm" associated with C. abortus begins two weeks before expected delivery dates (105). Lambs birthed from infected ewes can range from relatively healthy to stillborn. Vaccines are available but are relatively unused, and, in the case of the live attenuated vaccine, their safety is questionable to sheep and to pregnant shepherds(108-112). Many questions remain as to the mechanism of immunity in sheep, and the means to discriminate between vaccinated and naturally infected animals (103, 104). While there are 2 publicly available genomes (S26/3 Thomson et. al 2005(113), LLG (109)), and multi-locus sequence typing studies have been performed (114), a comprehensive comparison of whole-genomes, isolated across a large geographical region, will offer greater insight into the nature of this misunderstood and important pathogen.

To this end, we have whole-genome sequenced 17 strains of *C. abortus* sourced from farms across the Western United States of America. Modifications to technologies previously developed in our laboratory, and others (102, 115), that target chlamydial elementary bodies (EBs) via immunomagnetic separation, were implemented to address genomic variation within this species.

METHODS

Specimen/sample collection and diagnosis

Clinical specimens were collected from sheep (n = 20) goats (n = 2) and cows (n = 2) (Table 3.1). State diagnostic laboratories in Oregon and Idaho collected abortion products from farms in various counties of Oregon, Idaho and Washington. Additional historical samples (ovine and bovine), were collected from farms in Idaho, Wyoming and Montana (Table 3.1). The specimens were grouped into 3 different categories, all of which contained gDNA that needed to be enriched and sequenced while minimizing host or cell culture gDNA contamination. These groups include clinical specimens of infected birth products collected from farm fields post abortion, repository samples containing chlamydial gDNA mixed in with gDNA from culture (chicken yolk-sack culture) and cultured chlamydial cells grown in cell monolayers. Preparation for sequencing each category of specimen required different protocols, all adapted from the culture-independent workflow developed in a previous study (115).

Intake and Processing:

Whole and partial placentas were submitted to diagnostic laboratory were selected for this study based on a positive Giminez stain indicating chlamydial infection (116). The tissue specimens were stored at -80°C. After thawing, approximately 5 gm of tissue from 3 to 5 sites was collected, and the tissue was macerated with a scalpel blade. From this, a ~5 gm tissue pool was transferred to a 2 mL screw cap polypropylene tube, ~1/4 full

with 2.5 mm zirconia/silica beads (BioSpec Products, Bartlesville, OK). One milliliter of Dulbecco's MEM (Gibco-Life Technologies, Grand Island, NY) was added to the tube and the tissues were gently homogenized for 2 min at a setting of 48 on a TissueLyser bead mill (Qiagen Inc., Valencia, CA). and the homogenized tissue was frozen at -80˚C. Tissue homogenates were removed from -80˚C, centrifuged at 106 RCF for 5 minutes and 0.5 mL of the supernatant was aspirated for qPCR quantification and further processing.

Historical repository samples

Samples were archived as either pure genomic DNA, EBs from yolk sack stock diluted in SPG, or yolk sack stock DNA. The gDNA samples were determined to have concentrations of chlamydial DNA below the required concentration for direct Illumina sequencing (.2nG/µL) and therefore, required further amplification via multiple displacement amplification (described below). Chlamydial DNA concentrations in the yolk sack preparations that were high enough for gDNA extraction (DNeasy) and Illumina prepping required no further amplification.

DNA purification

Two different DNA extraction techniques were used because the specimens came from different sources. DNA from some samples was extracted using the MagMAX Viral DNA Isolation Kit (Applied Biosystems- Life Technologies, Grand Island NY) on a BioSprint 96 Workstation (Qiagen Inc.) per the manufacturer's instructions. DNA within other

samples was extracted using the Qiagen DNeasy Blood and Tissue kit (cat# 69506). In both cases DNA was eluted in 100 µL of elution buffer.

Quantitative PCR

Quantitative PCR analysis for confirmation of chlamydial DNA presence and detection during preparation for sequencing was accomplished using the Applied Biosystems Taqman Fast Universal PCR Master Mix, on the Applied Biosystems Step One+ or Applied Biosystems 75000 quantitative PCR thermocyclers. Primers and probe, designed and verified by Pantchev et al. (114), were used, targeting sequence of the *C. abortus ompA* sequence, which encodes the serodeterminant major outer membrane protein (CpaOMPA-F GCAACTGACACTAAGTCGGCTACA, CpaOMPA-R ACAAGCATGTTCAATCGATAAGAGA, Probe: CpaOMPA-S FAM-TAAATACCACGAATGGCAAGTTGGTTTAGCG-TAMRA).

Immunomagnetic cell separation of chlamydial elementary bodies (IMS)

IMS enrichments from clinical samples were carried out as described (115) using primary mouse antibodies directed at the chlamydial genus-common LPS epitope (EVI-HI;(96)). Antibodies were used at 3.1 ng/µL in these purifications. The primary antibody/cell suspension was incubated for 30 min at 4C° on a rotating shaker (33 rpm). One milliliter of MACS buffer (Miltenyi biotech; PBS-pH 7.2, EDTA-2 mM, FBS-0.55%) was added and tubes were gently vortexed. Samples were then centrifuged at 20,800

RCF for 10 minutes at room temperature, the supernatant was aspirated, and pellets resuspended in 1.0 mL MACS buffer. IgG-specific secondary antibodies, conjugated to 50 nm paramagnetic particles (Miltenyi), were incubated with the resuspended material and the tubes centrifuged as described for the primary antibody incubation. The antibody bound homogenate was loaded onto a MACS LS column and placed in a MidiMACs magnetic field on a MACs MultiStand (Miltenyi) and washed three times with MACS Buffer. The columns were removed from the magnetic field and recovered chlamydiae were then collected into 1.5 mL MACs buffer.

Cultured samples

The use of IMS was unnecessary for samples that were purified and grown in cell culture. After 48 hours of growth in one well of a 6 well culture plate the wells were harvested by freeze/thaw (-80°C for 1 hr and 27°C for 30 minutes). The EB/cell lysate suspension was transferred to a clean 1.7 µL microcentrifuge tube and centrifuged at 20800 RCF for 10 min. The supernatant was discarded and the resultant pellet was resuspended in 500uL of water. This material was incubated at room temperature for 10 min while being pulse vortexed vigorously for 5 min. The suspension was again centrifuged at 20800 RCF for 10 min, and the resultant pellet was resuspended in 5uL of PBS and either used immediately or stored at -80°C for future use.

DNase treatment for removal of host DNA

EBs suspended in 5 µL PBS were incubated (37°, 30 min, interrupted at 15 minutes to gently pipet up and down 5 times) in a RQ1 DNase master mix (Promega) consisting of 0.1 µL RQ1 DNase (1:10 dilution of manufacturer's recommendation), 1uL RQ1 Reaction Buffer , and 4.9 µL H20. DNAse reactions were terminated by addition of 1 µL RQ1 stop buffer and a 10 min incubation at 65 °C). To reduce residual RQ1 DNase digestions, samples were centrifuged at 20,800 RCF for 10 minutes, the supernatant was aspirated, and the pellet re-suspended in 10 µL PBS.

Multiple displacement amplification (MDA)

EBs suspended in PBS were aliquoted and used as template for MDA amplification as described by the manufacturer (Qiagen Repli-g Mini kit). The reaction time was 14 hours at 30 °C, followed by a 3 minute incubation at 65°C to deactivate the polymerase. The two independent reactions for each sample were pooled and stored at -20 °C for genome sequencing.

Genome sequencing

Genomic DNA preparations from each sample were prepared for Illumina sequencing (Illumina Inc., San Diego, CA) with the Nextera XT DNA Sample Preparation Kit (Cat# FC-131-1024) using manufacturer-specified protocols for multiplexing. All sequencing was performed on the Illumina platforms at the Center for Genome Research and Biocomputing (CGRB) Core Lab facility at Oregon State University. (Table S1).

Genome assembly and sequence analysis

Genomic sequence assemblies were completed using the reference guided assembly software package Maq (97). Loci that could not be resolved by reference mapping were solved through the use of contiguous segments of sequence generated via the de novo assembly software package Velvet (109, 110, 117), or through the use of ad hoc Perl and Python script pipelines. Any remaining gaps in the assembled contiguous draft were resolved through using sequences derived by PCR and Sanger sequencing. Geneious software (V 7.1.5: Biomatters, Auckland, NZ), available from http://www.geneious.com/) was used for sequence manipulation and draft amendment. The MAFFT software package (99) was utilized for whole genome alignments and phylogenetic analysis. Alignment columns containing ambiguous bases were stripped prior to variant calling. All locus tags were numbered according to the S26/3 genome (113).

RESULTS

Quantifying source material

To quantify the number of copies of chlamydial genomes in the starting material, whole genomic DNA (gDNA) extractions from each tissue specimen were amplified via qPCR of *omp*A (Fig. 3.1). Placenta tissue from 12 specimens contained detectable amounts of chlamydial gDNA. Lung, fetus and umbilicus tissue specimens, and 1 placental specimen, were negative for chlamydial gDNA and not processed further. The placental specimen with the least initial chlamydial abundance, OP/3, was quantified at <50 chlamydial

genome copies/5 µL of gDNA extract (2.0E05 genome copies/gram tissue), while the specimen with the most, OP/2, was quantified at ~1,400,000 genome copies/5 µL of gDNA extract (5.6E09 genome copies/gram tissue).

Assessing Illumina Library Composition

To assess the proportion of reads in the Illumina libraries that were chlamydial specific, the subsample of reads from each Illumina library were identified and categorized. For each specimen, 1000 random reads were extracted from the sequencing data set and queried against the non redundant database via a nucleotide BLAST search. BLAST results were parsed to determine the number of reads in the subsample that had specific identity to *Chlamydi*a spp., non-chlamydial microbes, the eukaryotic host, or queries that returned no significant hit (Fig. 3.2). Percentage of Chlamydia-specific reads varied across the specimens. The Oregon placenta sample OP/3, which was quantified with the least amount of detectable chlamydial gDNA, resulted in no detectable *Chlamydia*-specific reads and subsequently no assembly was possible. Conversely, Oregon placental specimen OP/2 contained ~97% *Chlamydia* specific sequencing reads. The specimen with the least chlamydial gDNA that was successfully processed and sequenced was the placental specimen IP/D with ~6,600 genome copies and ~25% chlamydial specific reads. All Illumina libraries were then reference assembeled into drafts for analyses.

Genome Wide Variation

Whole genome phylogenetic analysis, of assembled draft genomes, led to the identification of two major clades and a major outlier (LLG;(109)) within the phylogeny *C. abortus* genomes. The two major clades consist of the two-strain clade of the closely related strains LW203 and LW508 (74 nucleotide differences), and a low diversity clade with the remaining 16 genomes examined (Fig. 3.3). The two genomes with the least nucleotide identity to each other were LW203 and B577, which exhibited 940 nucleotide variations (Fig. 3.2). Genome data collected from individual placentas at an Idaho farm (samples IP/C and IP/D) shared the most nucleotide identity, differing at only two nucleotide positions between each genome.

To understand how variation is distributed regionally across the collected genome sequences, the quantity of variant nucleotide positions per 10Kb was plotted across the genome sequence. The data are presented in a series of concentric bar graphs based on sample/specimen source groupings (Fig. 3.4). Sample and specimen groupings for analysis were established based on sample/specimen history and reference the S/26 genome sequenced by Thomson and colleagues in 2005. The groups consist the clinical specimens from Oregon (OR = OP/1, OP/2, OP/4, OP/5, OP/6, OP/7), the clinical specimens from Idaho (ID = IP/C, IP/D, IP/E, IP/F, IP/G), the killed bacterin vaccine strain (Vac), the historical repository strains (Rep = SV139, V2901, B577), and the outlier clade repository strains (LW = LW203 and LW508).

The regions of highest variability included those containing the genes that code for the polymorphic membrane proteins and sequences within the chlamydial plasticity zone. While recombination is widespread in *Chlamydia trachomatis* (61, 62, 88, 89, 118-120), there was no evidence of recombination found after extensive analyses using methods developed previously in our laboratory (51, 55, 91, 115).

Protein Diversity

Consistent with what has been observed by others (1, 121, 122), the coding sequence of *ompA* is highly conserved in all of the sequenced *C. abortus* specimens. Within all the tested strains, there are 4 nucleotide positions with variation in the collection of sequences, and 3 of these lead to amino acid substitutions in the MOMP amino acid sequence. The major genome outliers LW203 and LW508 share 3 SNPs at positions 115, 210 and 286. The mutation at position 210 is silent while those at position 115 and 286 code for amino acid substitutions (115: S<->G and 286: A<->T). The Oregon placenta isolates OP/6 and OP/7 share a nucleotide change at position 261 that leads to a substitution in the amino acid sequence (261: A<->T).

The *C. abortus* genome contains 18 genes that code for polymorphic membrane proteins (PMPs), a *Chlamydia*-specific family of Type V secretion pathway autotransported proteins (44, 45, 66, 69, 123-125), that are highly variable in number and sequence identity by species. Two of the 18 predicted PMPs are previously reported to contain deactivating frameshift mutations (pmp8G-CAB270 and pmp9G-CAB273) and

this is consistent in each genome examined in this study. The most variable gene within this group is pmp16G (CAB596), with 94 variant sites in the ~2500 bp coding sequence.

The transmembrane head (TMH) genes (CAB760-775) code for a group of Chlamydia-specific membrane spanning proteins. Consistent with previous findings, CAB762 and 768 are pseudogenes in all genomes. The 15 genes contain 37 nucleotide differences including 4 deletion mutations and 2 insertion mutations, resulting in in-frame changes to the amino acid sequence in CAB760, CAB766, CAB772, CAB773, and CAB775.

Phase Variation in PMPs

Expression patterns of the genes encoding for the PMPs varies in all chlamydial species (53, 69), and *C. abortus* is no exception (45). We used computational and laboratory experiments to examine the dynamics of surface antigen variability as a result of phase variation from replication slippage in poly guanine tracks in the coding sequences. A group of 4 PMPs in *C. abortus* containing homologous variable length homopolymeric tracks were investigated. Those include the gene duplicates: *pmp12G* and *pmp17G* (100% sequence identity) and the closely related: *pmp16G* and *pmp13G* (84.14% sequence identity). The gene *pmp13G* was chosen for further analysis, as the sequence immediately flanking the homopolymeric track was the most unique in this *pmp13G*. This allowed for clearer analysis in the read library. We searched all read libraries generated in this study for reads containing the sequence flanking the *pmp13G* homopolymeric track. Reads containing the flanking sequences were analyzed for

number of G's in the track, and binned into one of two categories depending on the translational result of that sequence (i.e. in frame vs. out of frame) (Fig. 3.5). Not all genome read libraries had substantial numbers of reads that met the search criteria. A final collection of 8 strains had sufficient reads that spanned the flanking sequence and the homopolymeric track (CSC, LW508, SV139, V2901, OP/1, OP/5, OP/6, OP7, OP/5c and OP/7c).

All of the historical repository samples displayed the majority of the reads with the correct amount Guanines for the coding sequence to be in-frame and produce a complete protein. Most of the uncultured clinical specimens had very few reads with in-frame polyguanine tracks.

Vaccine Strain

Colorado Serum Company (CSC) in Denver, CO supplied their killed bacterin vaccine strain for comparative analysis. The CSC vaccine strain fits in the clade with the largest number of high identity genomes (Fig. 3.2). The vaccine strain genome sequence differs from the most divergent genomes sequences in the collection, the historical samples LW203 and LW508, by 909 and 907 nucleotides respectively. The genome with the highest nucleotide identity to the CSC vaccine genome was the historical sample genome V2901, with 27 nucleotide differences, indicating a genome that is not an outlier from the majority of the C*. abortus* populations in terms of genome sequence identify and structure (Fig. 3.3, Fig. 3.4).

Major Outlying Strain LLG

The major variant strain genome, LLG, was published to GenBank in 2011 (109). Because of the high number of ambiguous bases in that draft genome (n = ~8,000), this strain was only included in limited analysis in this study. It was sourced from a caprine host in Europe (Greece) and varies from every other *C. abortus* genome by and average of ~6200 variant nonambiguous nucleotide positions. Considering its significant divergence in nucleotide sequence from all other known *C. abortus* genomes, analysis was done to verify its inclusion as a member of the *C. abortus* species.

Nucleotide comparison of two major species determining genes, *ompA* and the 16s ribosomal subunit, were done with strain LLG, all the strains included in this study, and 4 closely related chlamydial species (*C. psittaci, C. pecorum, C. pneumonia* and *C. caviae*). The 16s rRNA gene was identical in all *C. abortus* specimens with the exception of LLG (5 nucleotide differences). However, the LLG 16s sequence had higher nucleotide identity to the *C. abortus* 16s gene (99.68%), than any of the other chlamydial species (*C. psittaci*: 99.61%, *C. pecorum*: 99.11%, *C. pneumonia*: 99.55%, and *C. caviae*: 99.35%). Consistent with the 16s results, comparing the *ompA* sequences from the same set of *C. abortus* strains and closely related species, LLG has higher nucleotide identity to C. abortus than any of the other examined chlamydial *ompA* sequence (*C. psittaci*: 81.14%, *C. pecorum*: 68.43%, *C. pneumonia*: 72.19%, and *C. caviae:* 81.27%). There was no evidence that LLG acquired neither its 16s nor *ompA* genes from recombination (data not shown).

Deactivated open reading frames (ORFs) and Pseudogenes

A total of 38 coding sequences are disrupted by frameshift or nonsense mutation in at least 1 *C. abortus* genome. Of those, 22 are previously reported pseudogenes (126). Sequencing conducted in this study revealed in frame complete coding sequences for 3 of the previously reported pseudogenes in genomes LW203, LW508 or both (Fig. 3.6). A total of 18 genes, intact in at least 1 *C. abortus* genome, are deactivated in others by frame shift or nonsense mutation. These genes include 11 hypothetical proteins, 4 of which are in the PZ, a phosphatase, glycosal hydrolase, exported peptidase and predicted transmembrane protein (Fig. 3.6).

Plasticity Zone

The plasticity zone (PZ) is a region of the genome that contains the highest level of variation across genome sequence from members of the chlamydial species. In *C. abortus,* the ~11.75 Kb PZ consists of 11 genes (CAB539-551). As discussed by Thomson et al. 2005, the *C. abortus* PZ is considerably smaller than the PZ in its closest chlamydial relative *C. caviae.* The *C. caviae* PZ is ~35Kb in length and contains 22 genes (6). Similar to other sequenced chlamydial genomes, the *C. caviae* PZ encodes proteins involved in tryptophan biosynthesis and a major candidate toxin (113). In contrast, tryptophan biosynthesis genes, and a gene encoding the large chlamydial cytotoxin are absent in both the published *C. abortus* S/26 genome and in each of our sequenced genomes.

Analysis of the *C. abortus* PZ in the specimens and samples included in this study reveals intact reading frames of 2 PZ genes reported as pseudogenes in *C. abortus* S26/3. This includes an intact hypothetical protein (CAB543) in the historical samples LW203 and LW508, and an intact *guaB* (CAB551) in LW508 (Fig. 3.6). The remaining reported psuedogene in the PZ (CAB541), codes for a similar truncation in all samples and specimens examined in this study.

Genes CAB546-547 are combined to 1 ORF in LW203. This results from a 1 bp deletion near the end of the sequence of CAB546 that puts it in frame with the downstream overlapping reading frame of CAB547. Similarly, in genes CAB852 and CAB853, every genome analyzed in this study, with the exception of the highly similar S26 and SV139, has a frameshift mutation that merges the ORFs of CAB852 with the downstream ORF CAB853.

Discussion

The acute and widespread loss of unborn lambs from *C. abortus* infection is a potentially catastrophic risk that sheep farmers face each birthing season. While there are killed bacterin vaccine options available, refinement of such a product would improve farmers' chances of mitigating such risks. To address this issue, we have applied culture-independent methods, previously developed by our research group for studying the human chlamydial pathogen, *C. trachomatis*, to *C. abortus* infected tissue specimens taken from aborted placental tissue.

We obtained complete genome sequences and performed comparative analyses on 11 clinical specimens, 5 historical repository specimens, and the vaccine strain made available by Colorado Serum Company in Denver, CO. We included in the analyses the 2 *C. abortus* genomes available on NCBI's GenBank genome database (LLG NZ_CM001168.1, S26/3 NC_004552.2).

With the exception of the major outlier genome LLG (~6200 nucleotide differences from every other genome), the genome-wide variation in *C. abortus* is relatively low compared to other chlamydial species, with a maximum of 940 nucleotide differences between the most divergent genomes.

The initial genome sequenced of *C. abortus* S/26 (113) revealed homopolymeric tracks in polymorphic membrane proteins that could result in phase variation from replication slippage. Two studies by Wheelhouse and colleagues revealed that the PMPs are differentially expressed with *C. abortus* populations (44, 45). The investigation of homopolymeric tracks in our study of *pmp13G* (a near duplicate of the pmp16G characterized in the Wheelhouse study ~80 identity), revealed a mixed population in several strains, with an apparent pattern of phase, depending on the conditions that that population of genomes had been exposed to (i.e. *in vivo* vs. *in vitro* growth). All of the repository strains showed the majority of reads with the correct number of Gs in the polyguanine track to keep the gene in frame and express this surface protein. Most of the clinical isolates showed the majority of reads with that polyguanine track out of

frame. This indicates a possible selective immunological advantage by turning this gene off *in vivo*, yet the preference to turn it on *in vitro*, indicates a functional advantage of expressing the protein.

This ubiquitous pathogen remains cryptic in nature. Its low level of genomic variability, aggressive nature, and wide tissue tropism indicate a pathogen that has recently diverged from *C. psittaci,* and hasn't found its own niche in the world that it can exploit without causing a massive immune response. Much work remains as a reliable and effective vaccine would be the ultimate weapon for sheep farmers to use in the battle against *Chlamydia.*

Table 3.1. Clinical specimen and historical sample information.

Specimen	Category	Tissue	Host	Origin	County
OP/1	Clinical	Placenta	Ovine	Oregon	Curry
OP/2	Clinical	Placenta	Ovine	Oregon	Benton
OP/3	Clinical	Placenta	Ovine	Oregon	Curry
OP/4	Clinical	Placenta	Ovine	Oregon	Curry
OP/5	Clinical	Placenta	Ovine	Oregon	Curry
OP/6	Clinical	Placenta	Ovine	Oregon	Benton
OP/7	Clinical	Placenta	Ovine	Oregon	Benton
IP/A	Clinical	Placenta	Ovine	Idaho	Canyon
IP/B	Clinical	Fetus	Ovine	Idaho	Minidoka
IP/C	Clinical	Placenta	Ovine	Idaho	Owyhee
IP/D	Clinical	Placenta	Ovine	Idaho	Owyhee
IP/E	Clinical	Placenta	Ovine	Idaho	Twin Falls
IP/F	Clinical	Placenta	Ovine	Idaho	Twin Falls
IP/G	Clinical	Placenta	Ovine	Idaho	Twin Falls
IP/H	Clinical	Lung	Ovine	Idaho	Canyon
IP/I	Clinical	Lung	Ovine	Idaho	Franklin
IP/J	Clinical	Umbilicus	Caprine	Idaho	Minidoka
IP/K	Clinical	Umbilicus	Caprine	Idaho	Adams
B577	Historical		Ovine	California	
SV139	Historical		Bovine	Western US	
V2901	Historical		Ovine	Western US	
LW203	Historical		Ovine	Western US	
LW508	Historical		Bovine	Western US	
CSC/VS	Vaccine		Ovine	Idaho	

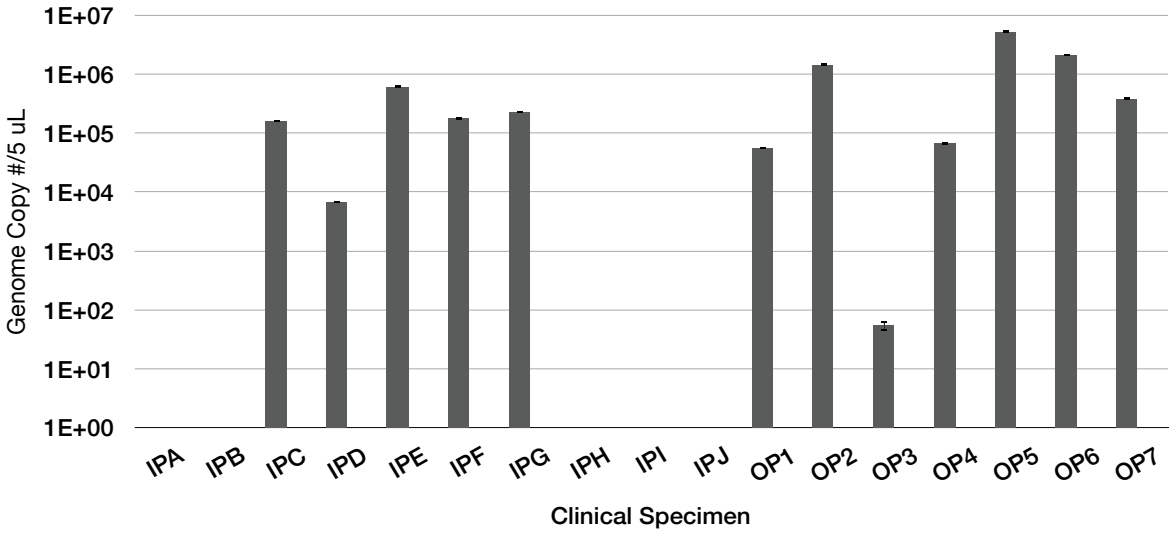

Figure 3.1. Quantification of source material from clinical tissue specimens.

Initial gDNA extracts (5uL) from all clinical tissue samples were used as template for qPCR targeting the gene that codes for the major outer membrane protein (*ompA*). Chlamydial DNA was not detected in 5 of the 10 Idaho specimens.

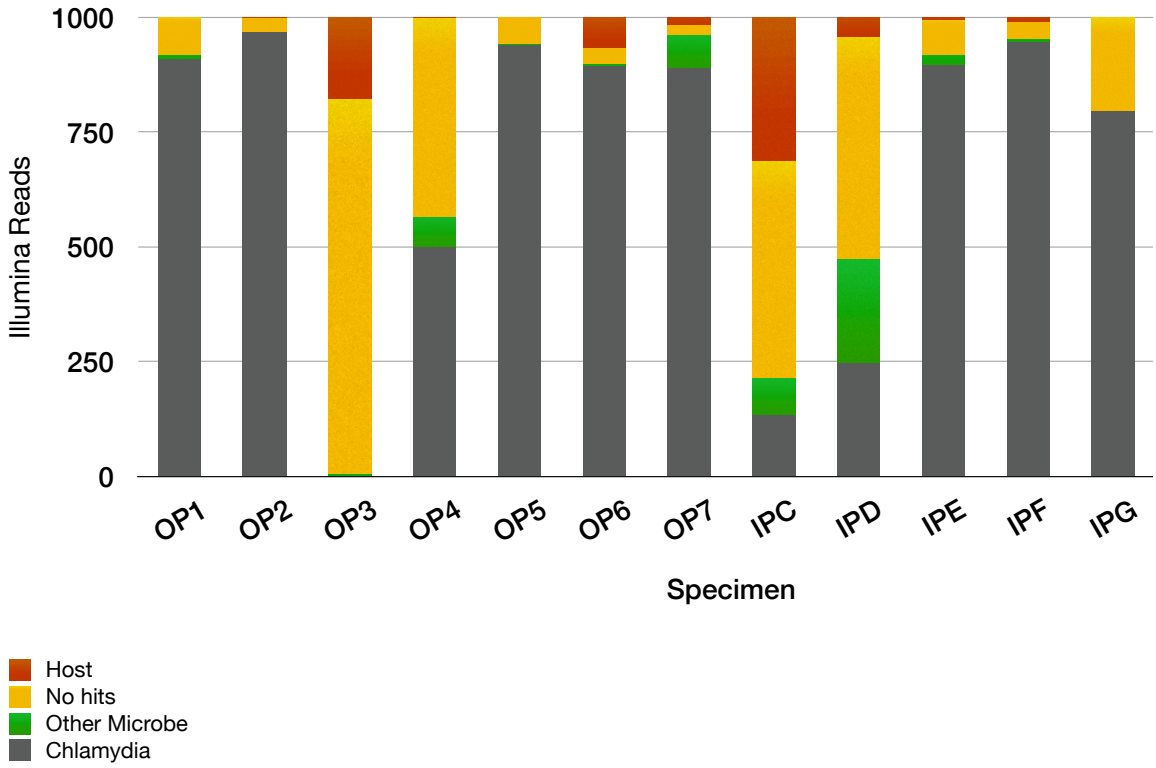

Figure 3.2. Illumina read library composition.

BLAST screen of 1000 random reads, extracted from Illumina libraries for each library prepped tissue specimen, to the NCBI non-redundant database. Categories represent the 4 sources of template identified in each specimen. Host (red);reads specific to the eukaryotic host, No hits (yellow); reads that return no significant hit, Other microbe (green); reads specific to non-chlamydial microbes present in tissue specimen, and Chlamydia (Grey); reads that only share identity with chlamydial species.

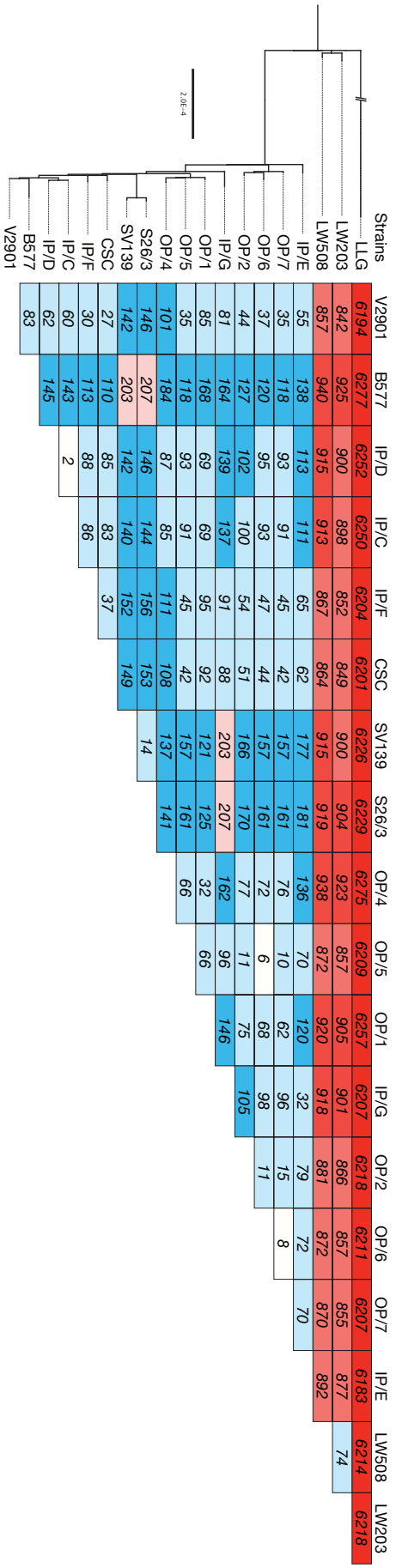

Figure 3.3. Whole genome phylogeny and distance matrix of all sequenced *C. abortus* specimens, samples, and existing genomes published to GenBank.

The phylogeny was constructed with the JK genetic distance model and neighbor joining tree building method. Scale bar indicates average number of substitutions per nucleotide site. Numbers in distance matrix indicate nucleotide differences between each strain. Color emphasizes degree of variation between each strain.

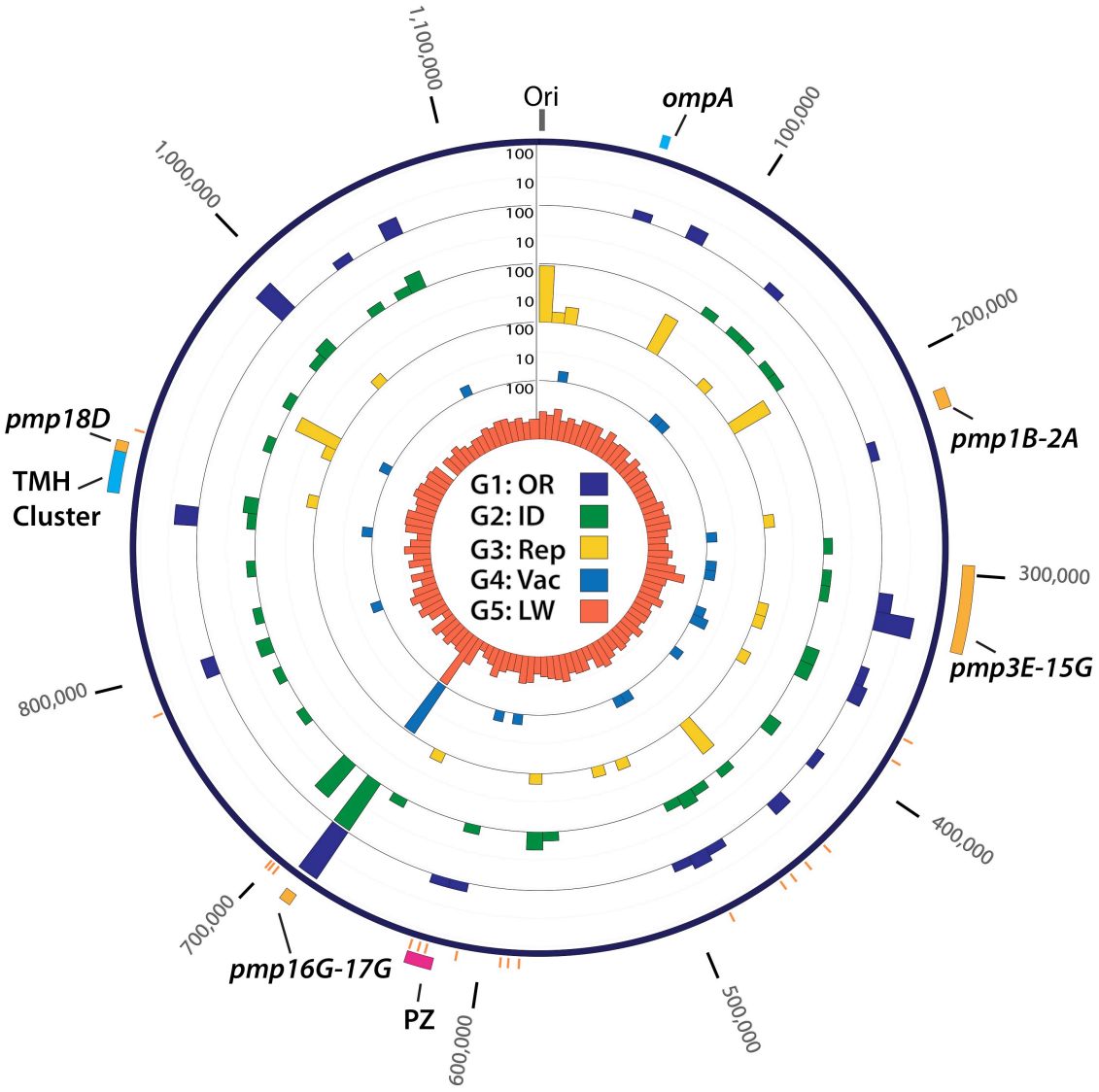

Figure 3.4. Circular representation of variant density across the *C. abortus* genome.

Outer ring indicates the C. abortus genome with tick marks identifying variant sites. Interior rings are histograms that indicate the number of SNPs per 10Kb region of the genome. The X-axis represents number of SNPs and is in log scale with a maximum of 100. The Y-axis includes a bar for each 10Kb region of the genome plotted in corresponding order with exterior ring genome map. Each ring represents a different comparative grouping based on sample/specimen origin. G1:ALL: Every genome involved in the study, G2:OR: All of the Oregon clinical tissue specimens, G3:ID: All of the Idaho clinical tissue specimens, G4:Rep: All repository/historical sample, G5:Vac: The killed bacterin Vaccine strain, G6:LW: The members of the 2nd major clade (LW203 and LW508). All variant sites are calculated from comparison to the published S26/3 genome (Thomson et al. 2005).

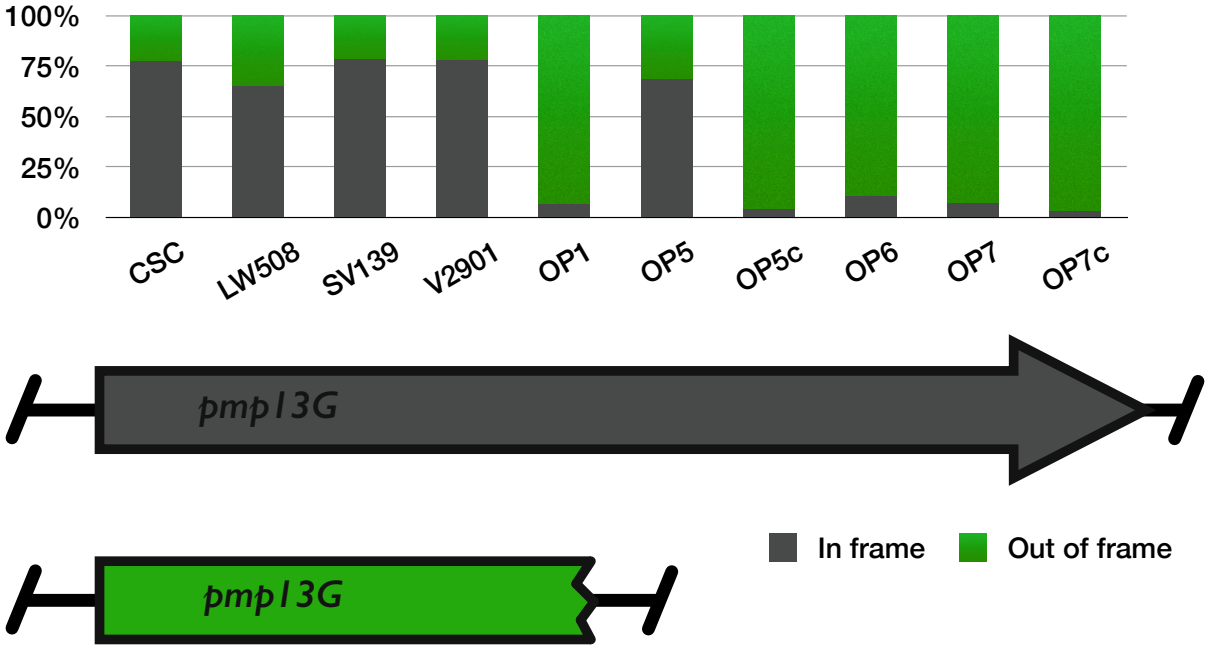

Figure 3.5. Phase variation in *pmp13G*.

Variable length poly G track in *pmp13G* (CAB281) produces mixed infection populations. For each sample/specimen included, percentage of reads were counted that contain homopolymeric tracks that produce an in tact reading frame (grey) or a truncation (green).

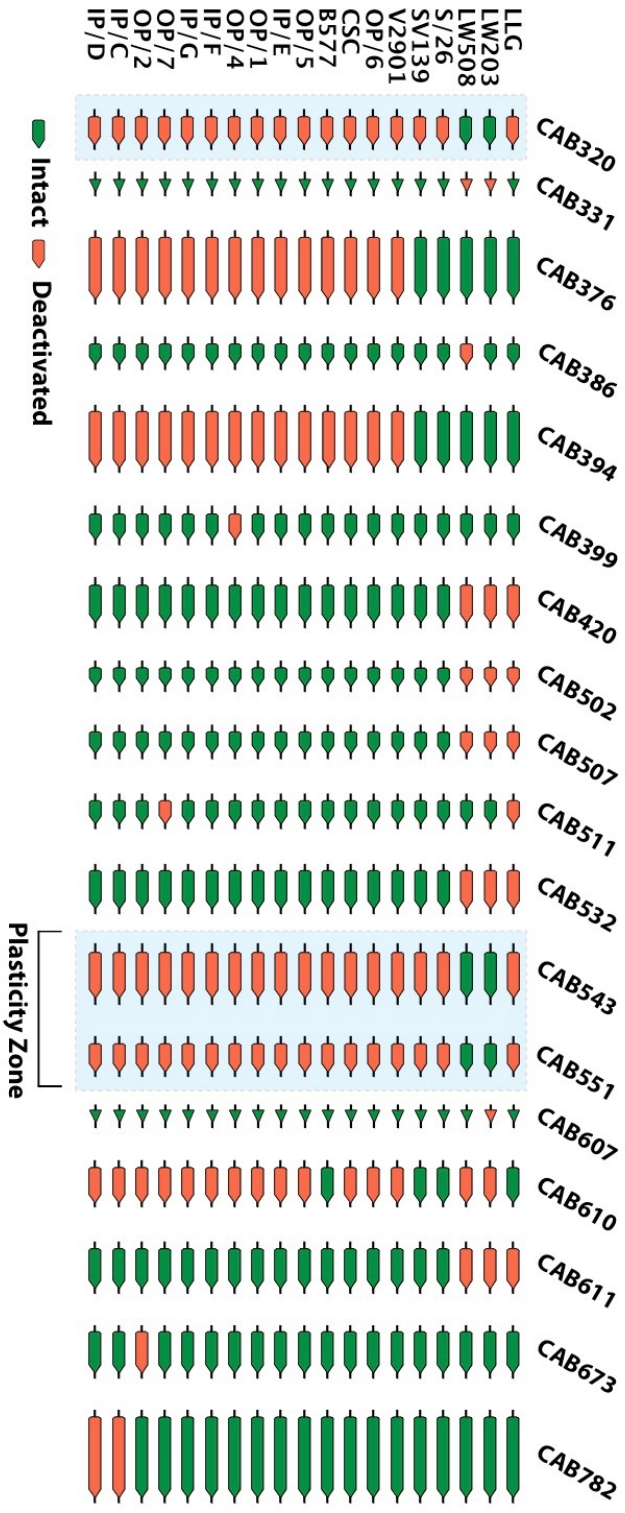

Figure 3.6. Deactivated Coding Sequences.

Comparisons of coding sequences that are deactivated in at least 1 *C. abortus* genome are indicated by the colored arrows. Green indicates in frame sequences while orange indicates disruption by frame shift or nonsense mutation. Yellow highlighting indicates coding sequences that have been previously reported as pseudogenes only.

Preparation of Chlamydial Genomes from Minimal Cell Culture Isolates Using DNase Treatment and Whole-genome Amplification.

Timothy E. Putman

INTRODUCTION

Prior to 2014, and the advent of culture-independent protocols previously described in this document (95, 115) , every published chlamydial genome, clinical or laboratory strain, was prepared for sequencing using a similar process. This process consists of 1) extensive growth in cell culture to propagate cells and genomic DNA, 2) sugar gradient centrifugation for separating chlamydial cells from host cells (127, 3) genomic DNA extraction (55, 95, 128, 129) and preparation for next-generation sequencing (NGS). This process uses extensive resources in the context of cell culture material and time. At best, these difficulties are cumbersome and expensive; at worst, serovars and species of chlamydia that are exceedingly difficult and slow to grow (Personal communication with Bob Suchland) require weeks of cultivation to acquire enough cells to proceed. Extensive cell culture can lead to changes in the genome (92, 130-132), causing concern for the *in vivo* relevance of any genome obtained from a strain that has been prepared for sequencing after in-cell propagation.

The advantages of the culture-independent approach are numerous and many applications are yet to be explored. While the original protocols were developed to obtain chlamydial cells directly from patients, and generate genomes sequences that accurately reflect that infection, we and other researchers in our field quickly identified further applications for culture free genome sequencing. The first such application that we have adapted our clinical protocols for is genome sequencing numerous strains in

parallel, from chemical mutagenesis studies that involve clonal isolation of chemical mutants with a phenotype of interest (133, 134).

Traditionally, clonal isolation of chlamydial cells that display a mutant phenotype of interest has been accomplished by limiting dilution of a population of interest and establishing that the phenotype is present in the final dilution (51, 95). The minimal number of cells remaining would then need to be propagated again to obtain sufficient gDNA for NGS. The development of reliable plaque isolation of clonal chlamydial populations from monolayers of mixed infection (54, 128, 135) has made limiting dilution unnecessary, but does not bypass the need for extensive propagation. Considering these inefficiencies, there has long been a need for the development of a system for amplifying clonal isolates with as little propagation as possible.

Using chlamydial material from a chemical mutagenesis study, supplied by researchers from the University of Indiana and University or Washington, we have developed a method for genome sequencing plaque isolates that mitigates the expense, difficulty, time and error associated with previous methods for prepping chlamydial genomic DNA for NGS.

The protocol makes use of key elements of the previously reported culture-independent protocols, and eliminates the elements that are unnecessary in an *in vitro* study. The protocol incorporates DNase digestion of host cell DNA, and whole-genome amplification to amplify enough genomic DNA for NGS, but does not require IMS for

enrichment due to the lack of other microbial contamination in the cell culture environment.

This streamlined and parsimonious approach allows one to go from an extremely low level clonal isolate, to enough gDNA for NGS in a substantially reduced amount of time, using a fraction of the materials previously required for the same outcome. This chapter describes the development and result of application of these protocols in detail.

METHODS

Organisms and Growth Conditions:

Chlamydia serovar L2 plaque isolates were provided by The Nelson Research Group at Univ. Indiana, and The Hybiske Research Group at University of Washington. The chlamydial cells were either direct plaque isolates following the plaque purification assay outlined in Binet et al 2005 (54, 128), or the same plaque isolates were used to inoculate a single well of a 96 well culture tray, and grown for 48 hours at 37 °C in 5% CO_2. Infected cells were harvested by freeze/thaw (-80°C > 24hrs/40°C for 10 min). The EB/cell lysate suspension was centrifuged at 20800 RCF for 10 min. The supernatant was discarded and the resultant pellet was re-suspended in 500 µL of water and the suspension was incubated for 30 min at 27ºC on a rotating shaker (33 RPM). The suspension was again centrifuged at 20800 RCF for 10 min, and the resultant pellet was re-suspended in 5 µL of PBS.

DNase treatment for removal of host DNA

EBs suspended in 5 µL PBS were incubated (37°, 30 min, interrupted at 15 minutes to gently pipet up and down 5 times) in a DNase master mix (Promega) consisting of 0.1 µL RQ1 DNase (1:10 dilution of manufacturer's recommendation), 1uL RQ1 Reaction Buffer , and 4.9 µL H_20. DNAse reactions were terminated by addition of 1 µL RQ1 stop buffer and a 10 min incubation at 65 °C). To reduce residual RQ1 DNase digestions, samples were centrifuged at 20,800 RCF for 10 minutes, the supernatant was aspirated, and the pellet re-suspended in 12 µL PBS.

Multiple displacement amplification (MDA).

Two independent MDA reactions were performed for each sample, as described by the manufacturer (Qiagen Repli-g Mini Kit cat # 150025), using a 16 hour reaction time at 30°C, followed by 3 min at 65°C to deactivate the enzyme and terminate the reaction. The reactions products for each sample were pooled together resulting in 100 µL of MDA product. This pooled product was then ethanol precipitated using the Qiagen supplemental protocol for purification of DNA amplified using Repli-g kits (https://www.qiagen.com/resources/download.aspx?id=80fb98ae-cec5-43a6-80af-f85090e7a572&lang=en). The purified MDA product was stored at -30˚C for future use.

Quantification of chlamydial DNA concentration and genome copies

To quantify total dsDNA we used a Qubit flourometer and 1 µL of 1:1 diluted post-MDA product. For detection of chlamydial specific DNA, Real-time PCR analysis of chlamydial

samples was carried out using the Applied Biosystems Taqman Fast Universal PCR Master Mix with primers defining unique sequences of the 2nd allele (groEL_2; CT604) of the *C. trachomatis* chaperone gene Hsp60 (primers: CTHsp60F GATTCTCTCTTCCTCGCTGTCTTC, CTHsp60F GAGGGTTTTCCCTGTCTGTGC). A plasmid containing the groEL_2 product from the primers CTHsp60 primer set was created, quantified, and used as a standard curve in quantifying genome copy number from sample DNA.

Genome sequencing

Genomic DNA preparations from each sample were prepared for Illumina sequencing (Illumina Inc., San Diego, CA) with the Nextera XT DNA Sample Preparation Kit (Cat# FC-131-1024) using manufacturer-specified protocols for multiplexing. All sequencing was performed on the Illumina HiSeq2000 and MiSeq platforms at the Center for Genome Research and Biocomputing (CGRB) Core Lab facility at Oregon State University.

Genome assembly and sequence analysis

Genomic sequence assemblies were completed through a custom Python pipeline that utilizes the reference guided assembly software package Maq (97) (Pipeline source code available at https://github.com/putmantime). Loci that were unresolved with sequence data after Maq reference guided mapping were resolved through the use of contiguous segments of sequence generated via the de novo assembly software package Velvet (55, 117) or through the use of ad hoc Python script pipelines that mine the Illumina read libraries for informative reads for ambiguous loci. Any remaining gaps in the assembled contiguous draft were resolved through Sanger sequencing specific PCR products

targeted to ambiguous loci. Geneious software (V 7.1.5: Biomatters, Auckland, NZ). Available from http://www.geneious.com/) was used for sequence manipulation and draft amendment. Genome annotation was carried using the published strain L2/434 (NC_010287) as reference. The MAFFT software package (99, 136) was utilized for whole genome alignments and phylogenetic analysis. Alignment columns containing ambiguous bases were stripped prior to variant calling. All locus tags are numbered according to the L2/434 genome (126).

RESULTS

Quantification of practice sample genome copy number

While designed to generate large quantities of DNA from relatively low levels of template, multiple displacement amplification performance is dependent on quality and quantity of starting template. To determine the number of chlamydial genome copies present in the starting template, qPCR was performed for each of 2 practice samples (L2-GFP, L2-PF) provided by our collaborators. The samples were provided as single, un-propagated plaque isolates, or a scale-up of a plaque from the same sample that had been propagated, following inoculation, for 48 hours in a single well of a 96 well cell culture tray. Performing qPCR on the un-propagated plaque isolate L2-PF and L2-GFP revealed a chlamydial genome copies/5µL of ~1.97×10^2 and ~4.97×10^2 respectively. The 48 hour, 96 well tray propagated scale-ups, L2-PF_SU and L2-GFP_SU, contained ~ 3.23×10^3 and ~2.23×10^3 chlamydial genome copies/5µL respectively (Fig. 4.1A) representing an order of magnitude increase from the plaque only template.

Quantification of practice sample genome copy number and total DNA after whole genome amplification

The Nextera XT DNA sample preparation kit requires 1 ng of genomic DNA for successful preparation. To ensure that chlamydial specific DNA was amplified, it is critical to have a means of quantifying the increase in chlamydial genome copies after the genome has been amplified. To determine the number of chlamydial genome copies present after MDA genome amplification in the practice samples, the same qPCR assay was performed. The number of chlamydial genome copies/5 µL in the MDA reaction using the un-propagated plaque isolate were quantified at ~1.60 x 10^4 and ~1.32 x 10^4 in L2-PF and L2-GFP respectively (Fig. 4.1A) representing an increase in genome copy number of ~2 orders of magnitude in the MDA reaction. When the 48 hr propagated isolates L2-PF_SU and L2_GFP_SU where used as template the number of chlamydial genome copies/5 µL was quantified at ~2.01 x 10^7 and ~6.21 x 10^7 respectively (Fig. 4.1A), representing a 3 order of magnitude increase in each case.

The Qubit Flourometer was used to determine the total amount of double stranded DNA present in the post MDA practice samples (Fig. 4.1B). In both L2-PF and L2-GFP plaque only preparations, the level of dsDNA after MDA amplification was too low to detect (<.02 ng/µL). Both the L2-PF-SU and L2-GFP-SU practice samples that had been propagated for 48 hours in a 96 well tray had concentrations of 47.3 and 22.6 ng/µL, which is more than sufficient for Illumina sequence preparation, requiring >/= 0.2ng/µL.

The lack of measurable DNA in the plaque template reactions resulted in the implementation of propagated template (i.e. at least 48 hours growth in a single well of a 96 well tray) in the preparation of future experimental isolates.

Proportion of chlamydial specific reads in practice Illumina libraries

After application of the template preparation protocol, the 12 genomes were submitted for Illumina Nextera XT genome sequence preparation and whole genome sequencing on the Illumina HiSeq 2000 at the Center for Genome Research and Biocomputing Core Laboratory at Oregon State University. I will first present the results of the practice sample data, then compare the results of two experimental groups that were prepared in two different ways.

To determine the amount of chlamydial specific short sequence reads in the Illumina libraries, a similar BLAST screen to those performed in previous chapters, using the NCBI non-redundant database, was performed on a random representative sample from each library (n = 1000) (Fig. 4.2A).

The proportion of *Chlamydia* specific reads existed as a wide range for the 12 total (including the 2 practice samples L2-PF and L2-GFP) clonal isolates submitted for sequencing. The two practice samples L2-PF and L2-GFP had the lowest percentage of chlamydial specific reads at ~13% and ~33% respectively.

Genome coverage analysis of practice libraries

Reference guided alignment against the reference genome, supported by *de novo* contig assembly, was used to create draft genomes sequences for each sample. Coverage and variant analysis were performed documenting all genomic differences from the reference strains.

Graphing the mean coverage/base across each genome revealed a wide range in mean coverage of ~42 and ~317 reads/base (Fig. 4.2B). The expected coverage, calculated based on the number of reads determined to be chlamydial specific (Fig. 4.2B), was statistically similar to the actual mean coverage/base calculated from the reference mapping process ($P > 0.05$) (Fig. 4.2B). The high variation in the mean coverage for each genome, and the possibility of amplification bias from whole genome sequencing, led to the investigation of coverage distribution across the genome sequence, and can be visualized as peaks and valleys in a histogram of the mean coverage/1000 bases (Fig. 4.2C). Because of the potential to overestimate the quantity of chlamydial specific DNA in the template sample, the target for qPCR, the highly conserved housekeeping gene *groEL_2,* is examined, relative to coverage, in these histograms. While GFP produced a complete unambiguous genome draft, the practice sample PF succumbed to too many regions of no coverage and failed to fully assemble.

Quantification of Group 1 and Group 2 Experimental Isolates

After initial development, 2 groups of experimental isolates (Group1 n = 10, Group 2 n = 10) were subjected to the protocol, and the input and output were compared for optimization.

Group 1 includes 10 clonal isolates from a chemical mutagenesis library. Prior to protocol application, the same qPCR assay was performed on each of the ten minimally propagated template samples. Starting templates were similar in quantity to the practice scale-up samples, ranging from 1.47 x 10^3 (43A3) to 9.4 x 10^5 (5G) (Fig. 4.3A).

After determining there was sufficient template in all isolates, the protocol for NGS prepping was performed. The genome copy number after MDA was quantified by qPCR, and the total double stranded DNA concentration was calculated by Qubit flourometer. (Fig. 4.3B). In all cases, MDA product genome copy number surpassed 10^7, ranging from 2.45 x 10^7 (36C2) to 1.09 x 10^{10} (43A3) copies and the concentrations of dsDNA ranged from 15.6 ng/µL (43A3) to 59 ng/µL (8G) (Fig. 4.3B).

A second group of 10 clonal isolates, generated in a similar manner for genome sequencing was examined as well. While using the same methods, after plaque isolation, this set of isolates was propagated in a single well of a 6 well plate instead of the 96 well plate used in Group 1. The initial templates in the Group 2 isolates displayed a range of genome copy numbers, from ~7.0 X 10^3 in α-MD to over ~5.6 x 10^7 in 7C11, with the majority in the ~10^6 and ~10^7 range (Fig. 4.3C). The post MDA product was quantified by Qubit flourometer alone, and revealed a range of total DNA

concentrations from no detectable DNA in α-MD and β-MD, to ~22 in KH-L2 and 9B3 (Fig. 4.3D). The loss of all genomic template in α-MD and β-MD is a notable outcome and speculation on the reason will be revisited in the discussion. All successfully prepared isolates were submitted for NGS.

Proportion of chlamydial specific reads in Group 1 and Group 2 Illumina libraries

After application of the template preparation protocol, the Group 1 clonal isolates were submitted for Illumina Nextera XT genome sequence preparation and whole genome sequencing on the Illumina HiSeq 2000 at the Center for Genome Research and Biocomputing Core Laboratory at Oregon State University.

To understand the composition of the Illumina libraries, in terms of the organisms that the DNA represented by each read originated from, a similar BLAST screen to those performed in previous chapters, using the NCBI non-redundant database, was performed on a random representative sample from each library (N= 1000) (Fig. 4.4).

The proportion of *Chlamydia* specific reads covered a wide range for the 10 Group 1 clonal isolates submitted for sequencing. The majority of the Group 1 samples had chlamydial specific read proportions in the 70-95% range with the minimum belonging to isolate 43A3 at ~41% (Fig 4.4A).

It is important to note that while Group 1 was sequenced on the Illumina HiSeq platform, Group 2 was sequenced on the Illumina MiSeq platform. The MiSeq has lower depth in sequencing generating ~25 million reads versus the ~180 million reads that 1 lane on the HiSeq 2000 generates. The comparison of Group1 to Group 2 is being made with respect to uniformity and consistency of coverage, which is not necessarily related to depth of coverage.

The same BLAST screen was performed on the representative samples of reads in the Group 2 isolate libraries (N = 1000), and revealed a similar range in chlamydial specific reads, with a library percentage as low as ~20% in 8F5 and several libraries over 90% chlamydial specific (Fig. 4.4B).

Genome coverage analysis of Group 1 and Group 2 libraries

Reference guided alignment against the reference genome, supported by *de novo* contig assembly, was used to create draft genome sequences for each sample. Complete and unambiguous drafts were produced for all but 2 genomes, 8G, 43A3 (as mentioned) ~23000 and ~1565 ambiguous bases remaining in their respective draft genomes (data not shown). To determine if any amplification bias resulted from the process of prepping and sequencing these genomes, the overall genome coverage was investigated in multiple ways.

Graphing the mean coverage/base across each genome revealed a wide range in all genomes analyzed in Group 1 (Fig. 4.5A) and Group 2 (Fig. 4.5B). In both groups, the expected coverage, calculated based on the number of reads determined to be chlamydial specific, was again similar to the actual mean coverage/base calculated from the reference mapping process. The Group 1 genomes displayed a range in mean coverage from the lowest covered genome 43A3 with ~278 mean coverage/base to the highest 34C2 with ~1106 mean coverage/base. The coverage distribution plots again visually revealed peaks and valleys associated with locus bias, and those that appeared to have the most bias correlated with those with the highest variance in there mean coverage (8G and 43A3) (Fig. 4.6A).

Similar to Group 1, the expected coverage in Group 2 was close to the actual mean coverage ranging from as low as ~100 in 8F5 to as high as ~425 in 9C2 (Fig. 4.5B). When plotting the coverage distribution, a much more uniform consistency of coverage was observed visually with all 8 sequenced genomes showing little sign of bias in distribution of coverage (Fig. 4.6B). To verify this beyond visual observation, the coefficient of variance (CV) for the mean coverage of each genome in each group was plotted and an unpaired, non-parametric t-test (Mann-Whitney) was applied to compare the collective variance of each group. As this was a pilot study, the genomes in Group 1 that had obvious locus bias (Fig. 4.5A) and did not produce complete unambiguous drafts (8G and 43A3), were not included in the CV comparison; the probability that these were failures from user error, rather than difference in the protocols was too great to include them.

However, even with out the outliers, Group 1 variation in mean coverage was significantly greater than the Group 2 variance (Fig. 4.5C, P= 0.0006), indicating a more even distribution of coverage, and less amplification bias, in the Group 2 genomes.

Variant analysis of practice and Group 1 and Group 2 samples.

To elucidate the variation in the 8 completely sequenced mutant isolates from Group 1, and 7 from Group 2 (L2-KH was the reference strain), variant analysis was performed. Nucleotide positions from Group 1, that contained any variation relative to the reference L2 strain used, were identified across the genome sequences numbering from 3 to 9 per isolate with coverage for the variants ranging from as low as ~44 to as high as ~1600 (Fig. 4.7A). Genomes that displayed less uniformity in coverage distribution across the genome, unsurprisingly displayed more variability in variant coverage (Fig. 4.7A). Similar to Group 1, Group 2 variants were located across the genome sequence ranging in number from as few as 2 in 9B3 to as many as 13 in 7H7. Two of the genomes displayed a wide range in coverage values, 7C11 and 7H7 (Fig. 4.7B). The same statistical test used in comparing the coefficient of variance for the mean genome coverage of each group was applied to the variant sites coverage. This determined that the variance in coverage values for Group 2 was significantly lower than the coverage values for Group 1 (P = 0.0140) (Data not shown) an expected result considering the mean coverage varied more in Group 1.

Coverage and library composition of samples prepped with traditional protocols.

To compare the newly developed protocol to traditional protocols for NGS genomic preparation, in context to coverage continuity and Illumina library composition, I have exhumed two Illumina libraries of cultured *C. trachomatis* isolates that were generated by our research group and prepared as previously described in Jeffrey et al 2010 (55). These were part of a study that has yet to be published. As these were not prepped in development of a protocol, no quantification data exists for the samples D/Trad and F/Trad.

To understand how our minimal culture compares in terms of Illumina library composition (specifically host DNA contamination) BLAST analyses was performed (Fig. 4.8A). The chlamydial specific read proportions were 86.3% and 66.5% in D/Trad and F/Trad. Mean coverage/base was ~366 and ~88 respectively (Fig. 4.8B) and visually the coverage uniform when plotted linearly (Fig. 4.8C).

DISCUSSION

Similar to preparation of clinical isolates, the process of isolation and generating enough quality genomic DNA for NGS in an obligate intracellular bacteria like *Chlamydia,* is not without difficulties. A single copy of a human host cell genome equates to over ~3000 copies of the chlamydial genome in terms of number of nucleotides involved; it would not take many intact cells to dominate a MDA reaction with host template. In addition to requiring the use of extensive cell culture propagation materials time, and energy,

with each passage of slow growing cells, there exists a higher chance that the genomic structure under investigation will be altered with *in vitro* specific artifact (92).

As mentioned previously, many strains have growth deficiencies in cell culture. One strain that highlights the necessity for protocols that bypass extended propagation is a tetracycline resistant clinical isolate of the swine pathogen *Chlamydia suis.* This strain, $R19^{tet}$ described in (35) grows very slowly for several passages in cell culture and then precipitously explodes with growth over the course of one passage (Personal communication with Bob Suchland and Art Anderson). This behavior implies a genomic change in cell culture that confers an advantage and leads to domination of that variant in the population of cells within that isolate. This is a problem when attempting to understand the pathogen, as it exists in its natural host. It is vitally important to develop a means of mitigating such artifact. That means must be robust and unbiased in its representation of the clinically relevant genome.

Other groups have tackled this problem by various methods, some more promising than others. One method is to target the chlamydial DNA itself and pull down with magnets; custom capture RNA bait sets are designed to enrich DNA specific to your organism of choice by targeting the 16S ribosomal DNA with a sequence specific probe and using whole genome amplification to obtain enough DNA for sequencing (127, 137, 138). This has its advantages and disadvantages. An advantage lies in that it circumvents the DNase digestion step allowing for the enrichment of extra cellular chlamydial DNA from a specimen, while our protocol would digest any chlamydial genomes not protected by an intact cell. A disadvantage is that this introduces a step that could introduce locus

bias by pulling down genomes targeted by one locus. If that genome is not fully intact, the rest of the DNA would be lost.

Another clever protocol targets the host DNA for physical removal by targeting invertebrate specific methyl-CpG islands. It uses an IgG primary antibody with a methyl-CpG binding domain fused to it (139), binding and magnetically removing host DNA from clinical microbiome specimens. This is a broad-spectrum approach, as it does not rely on DNase resistance or target sequences, and mitigates selection bias by targeting the contaminant, rather than the DNA of interest. It was designed for, and is very useful in shotgun analysis of microbiome samples, where a wide a variety of bacteria with different properties are the target. One drawback to this method, however, is that it does not target mitochondrial DNA as the mitochondria are hypothesized to have been free living prokaryotic organism before endosymbiosis resulted in mitochondria being endogenous to every eukaryotic cell (140). This results in host DNA that does not contain the methyl-CpG islands that allow this protocol to work.

Our protocol targets host DNA for digestion, but require that the genomic material of interest be protected from digestion. We take advantage of the spore like structure of the infectious chlamydial elementary body. The surface disulfide linkage creates a coat of armor that protects the cell from digestion by DNase (141-143).

While the protocol is not perfect, several factors have been examined in this chapter that validate the use of DNase and MDA for obtaining whole genome sequences from

minimally propagated chlamydial isolates. Comparison of outcomes to this proposed protocol, to those obtained from chlamydial cells that were prepared in the traditional manner, of extensive cell culture propagation, illustrates comparable quality in the product.

Using plaque isolates as template for MDA would be optimal, bypassing the need for any further propagation, but quantity of chlamydial genome copies was insufficient in the ~10^2 range. Our practice experiments demonstrated that chlamydial genome copies needed to be present in the ~10^3 range or the MDA reactions would fail. Our protocols use the Repli-G Mini Kit from Qiagen that was used in the design of our original protocol for clinical swab sample enrichment, and we are supplying less than the 10ng suggested template of gDNA with success in most cases. Since then, Qiagen has developed a Single-Cell Repli-G kit (Cat # 150343) that is much more sensitive (and more expensive) than the Mini kit only requiring 1-1000 cells as template. The protocol could be modified for future applications to use the more sensitive Single-Cell kit and plaques could then hypothetically be used as template.

When using a single well of a 96-well cell culture plate to propagate the plaque template (Group1), we had a great deal of success, but did observe instances of coverage bias and altogether failure. With this method limited to 1 trial run, it is difficult to say that it is not advised to use this amount of template for NGS. While uniformity of coverage was improved in Group 2, when a single well from a 6-well culture tray was used as

template, this was also the second time that this protocol had been performed; skill and experience likely play a roll in outcome.

To address the deficiencies observed in this study (locus bias and draft genome assembly failure), there are several factors to consider. In the case of the practice sample PF, I would attribute much of the failure to the fact that it was subject to experimentation in development of the protocol. While the Group 2 samples that failed, α-MD and β-MD, had lower initial chlamydial genome copy number than the rest of the Group 2 samples, there quantities were not lower than template used in the Group 1 preparations. Our collaborator informed us that a different person prepped these two isolates. Details were sparse on the difference, but one could speculate that the genome copies detected were extracellular or from damaged cells and did not survive the DNase digest. This could be attributed harvesting too early, and recovering mostly reticulate bodies that would not resist DNase, or harvesting to late and losing the majority of the viable elementary bodies from not having viable cells to infect.

The failure of 8G and 43A3 are hard to explain and may be attributed to pipetting error on my part. The nature of this protocol provides several opportunities to lose everything with an unsteady hand. These opportunities consist of several high-speed centrifugation steps, where a fragile and tiny pellet must be preserved. What is even more difficult is re-suspending the pellet in tiny volumes of reagent. During this process the chance of the pellet, sometimes too small to see, being thrown out in the liquid waste with the supernatant, is a real threat. In addition, not fully dissolving the pellet could lead to

incomplete digestion in the DNAse reaction resulting in host DNA overwhelming the sequencing. Experience and repetition have led to improved results with each iteration. For example, since the work described in this chapter was completed, a step has been added during the DNase incubation; the incubation is paused and the reaction is gently pipetted up and down a few times to further break up the pellet so enzyme can reach any extracellular DNA that was protected in the innards of the firm pellet, leading to more complete digestion in future experiments (Data not shown).

While traditional methods are useful and provide a high likelihood of uniform coverage across the genome, any protocol that is streamlined, easier and less expensive is a useful tool to have in the genomics too kit. Furthermore, the traditional methods are not foolproof, and the removal of host DNA when working with massively propagated isolates can fail.

While the two data sets representing traditional methods for sequence preparation, D/Trad and F/Trad, are not sufficient to represent all the attempts to prep chlamydial genomes for sequencing, their even distribution of coverage provide a benchmark that we can compare our methods to. There is nothing in the data to indicate that these protocols are deficient of that benchmark. While D/Trad has a reasonable mean coverage (~226.2) with a relatively low standard deviation (86.7), the coverage mean of F/Trad is relatively low (~135.5) with a high standard deviation (362.2), indicating that coverage distribution differs from genome to genome regardless of how it was prepped. Given the outcome of these pilot studies, the protocol has potential to be an extremely useful tool in the obligate intracellular genomics world. It was conceived of based on

successes with the clinical protocols, but the decision to exclude IMS was made with out knowing what the outcome would be. IMS is the most expensive and time-consuming process in the clinical protocol. Since its primary use is to remove other microbes that would not be present in cell culture, and the host contamination was being resolved by RQ1 DNase treatment, it was worth a try. Being able to prep so many genomes in parallel is a huge advantage and allows the generation of more accurate data than ever before.

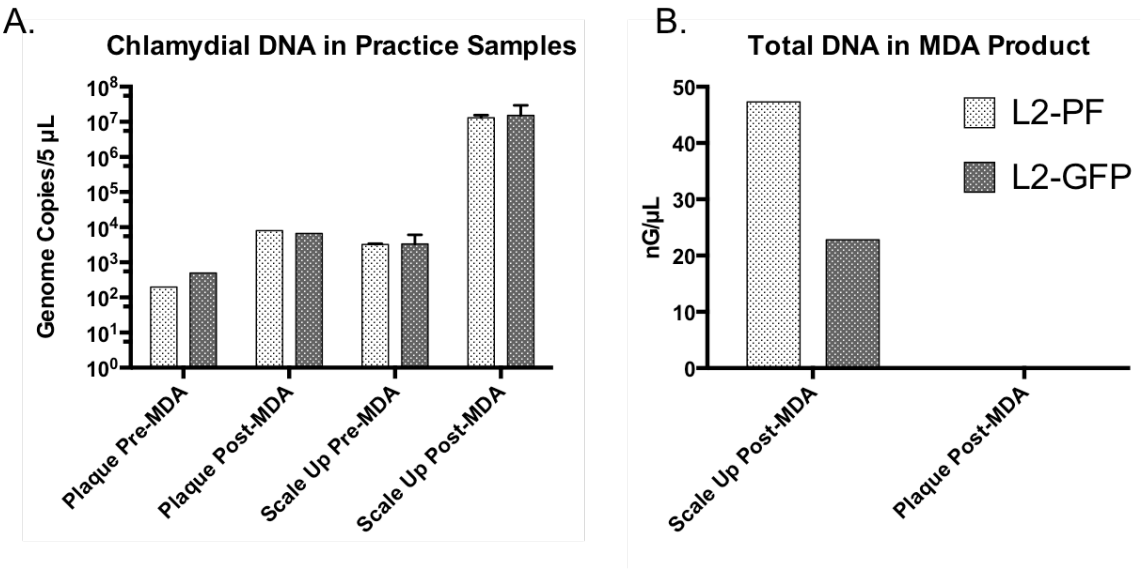

Figure 4.1. Practice sample quantification.

A. Genome copy number/ 5 µL of template (light grey) or post MDA product (dark grey) in the practice samples for the plaque only, and propagated (scale up) cells. **B.** Total concentration of DNA in the post MDA product. DNA was not detected when plaques were used as template.

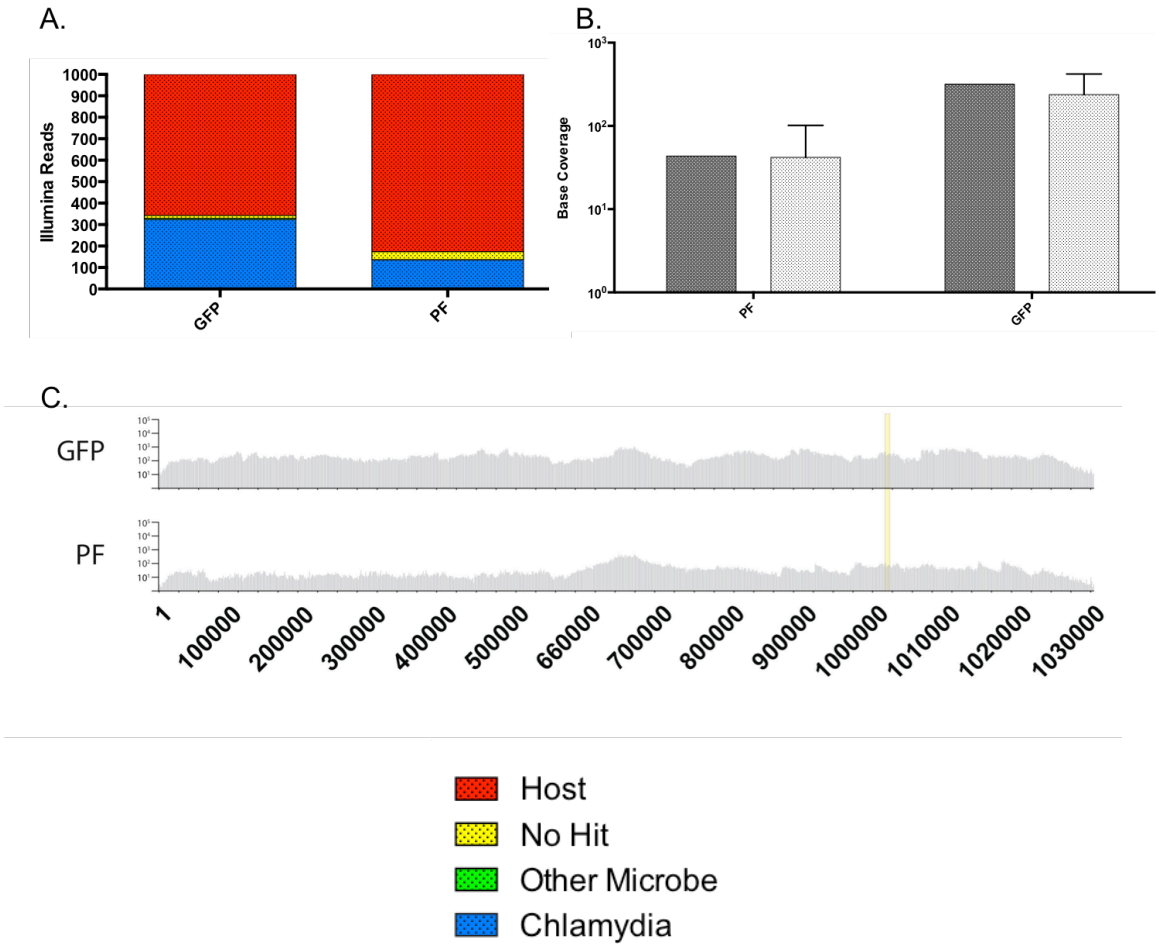

Figure 4.2. Practice sample Illumina library composition and coverage.

A. Illumina library composition of practice samples from a BLAST screen of 1000 random reads representing the possible sources of sequence (red = host, green = other microbe, yellow = no BLAST hit, blue = *Chlamydia*). **B.** Expected and mean coverage of reference mapped Draft assemblies. Expected coverage (dark grey) is calculated based on the number of *Chlamydia* specific reads in the library. Actual coverage (light grey) is calculated by mapping the reads to the L2/434 genome sequence. **C**. Coverage distribution for practice samples across reference genome. Histogram of average read depth/1000 bases in the genome. Yellow highlight indicates location of qPCR target (*groEL_2)*.

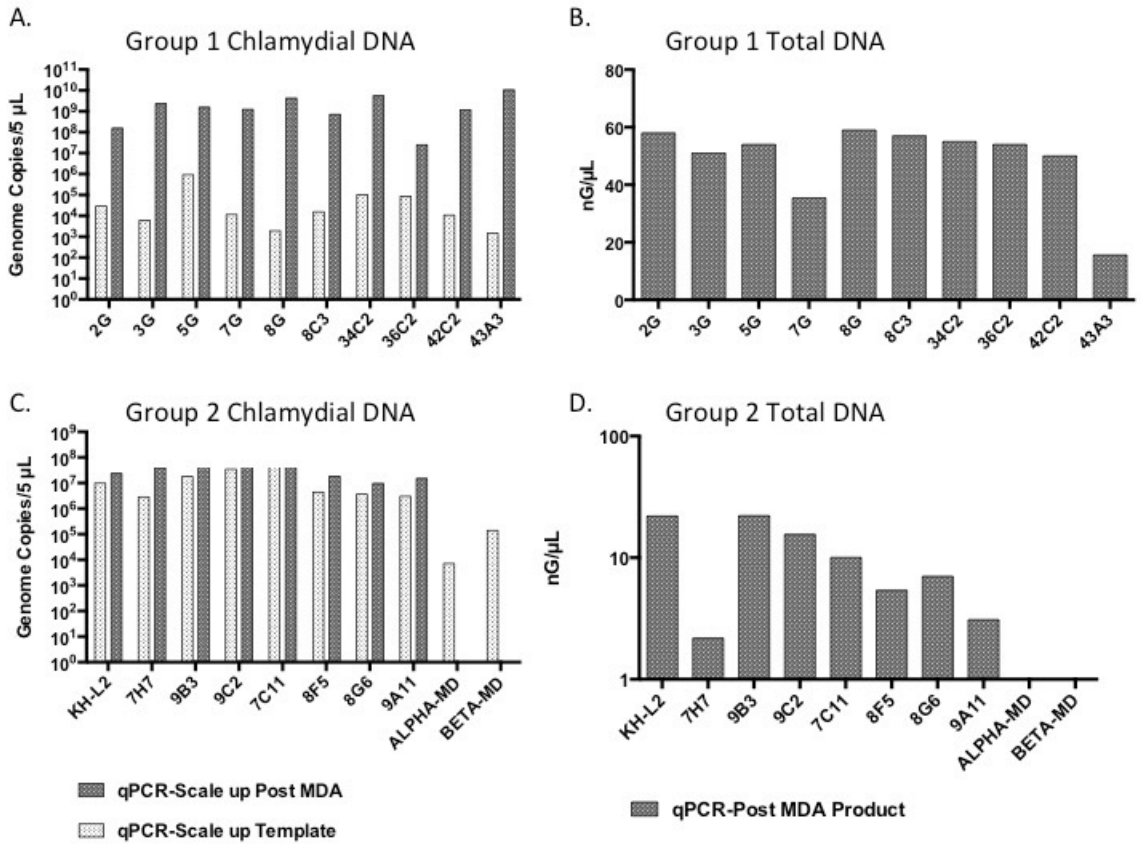

Figure 4.3. Group 1 and Group 2 genome copy number quantification.

Genome copy number/5 µL of template and post MDA product **A.** Group 1 **C**. Group 2. Total DNA after MDA reaction in **B**. Group 1 **C**. Group 2.

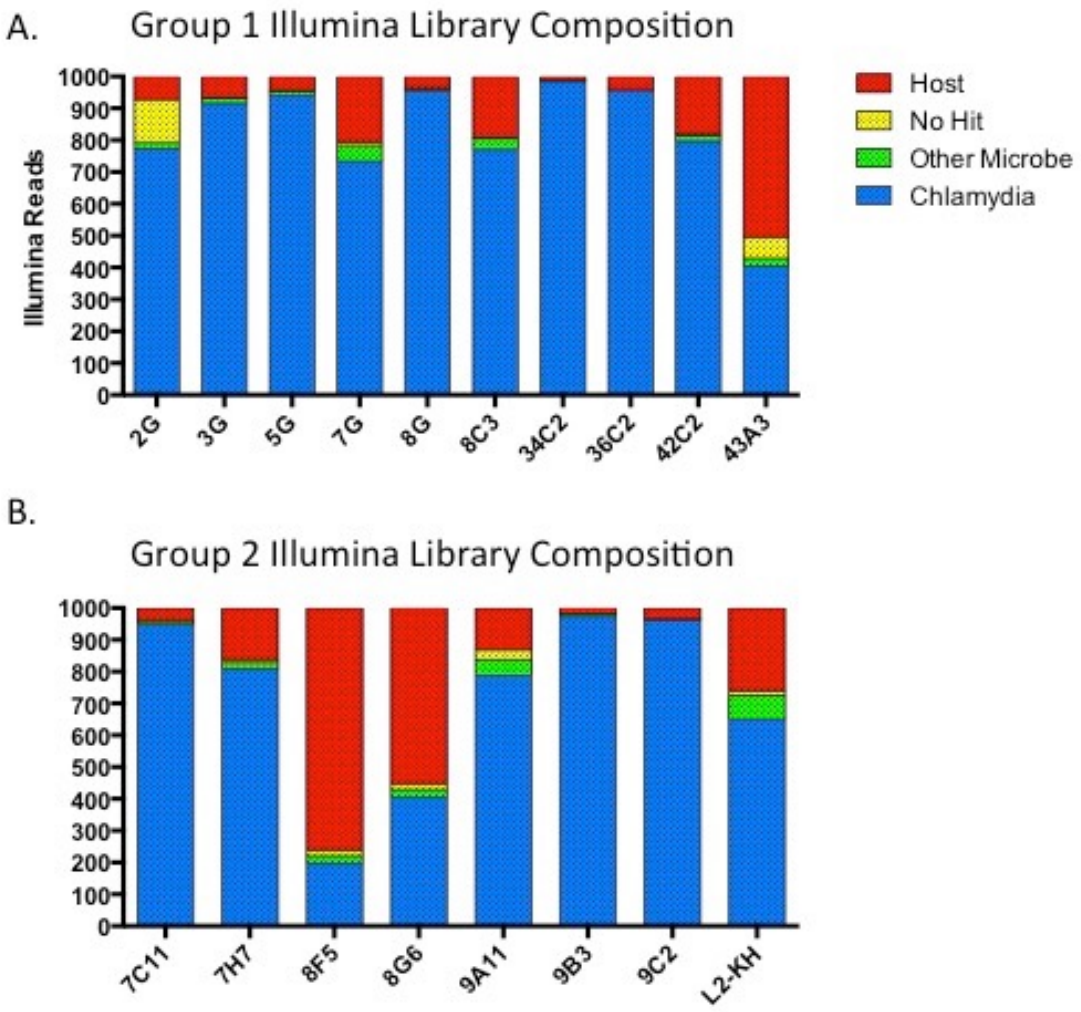

Figure 4.4. Group 1 and Group 2 Illumina library composition.

Illumina library composition from BLAST screen of 1000 random reads. For **A.** Group 1 **B.** Group 2.

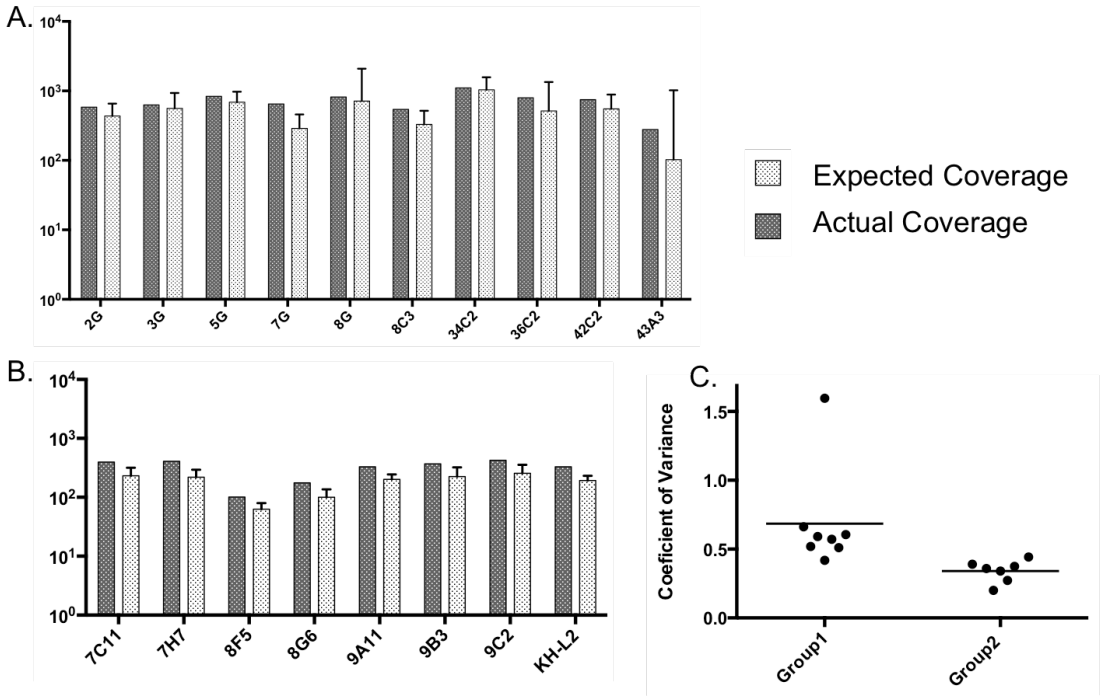

Figure 4.5. Group 1 and Group 2 coverage.

Expected and mean coverage for A. Group 1 and B. Group 2. C. Coefficients of variance for the two groups. Group 1 has significantly more variation in genome coverage than Group 2 indication locus bias (P = 0.0006).

A.

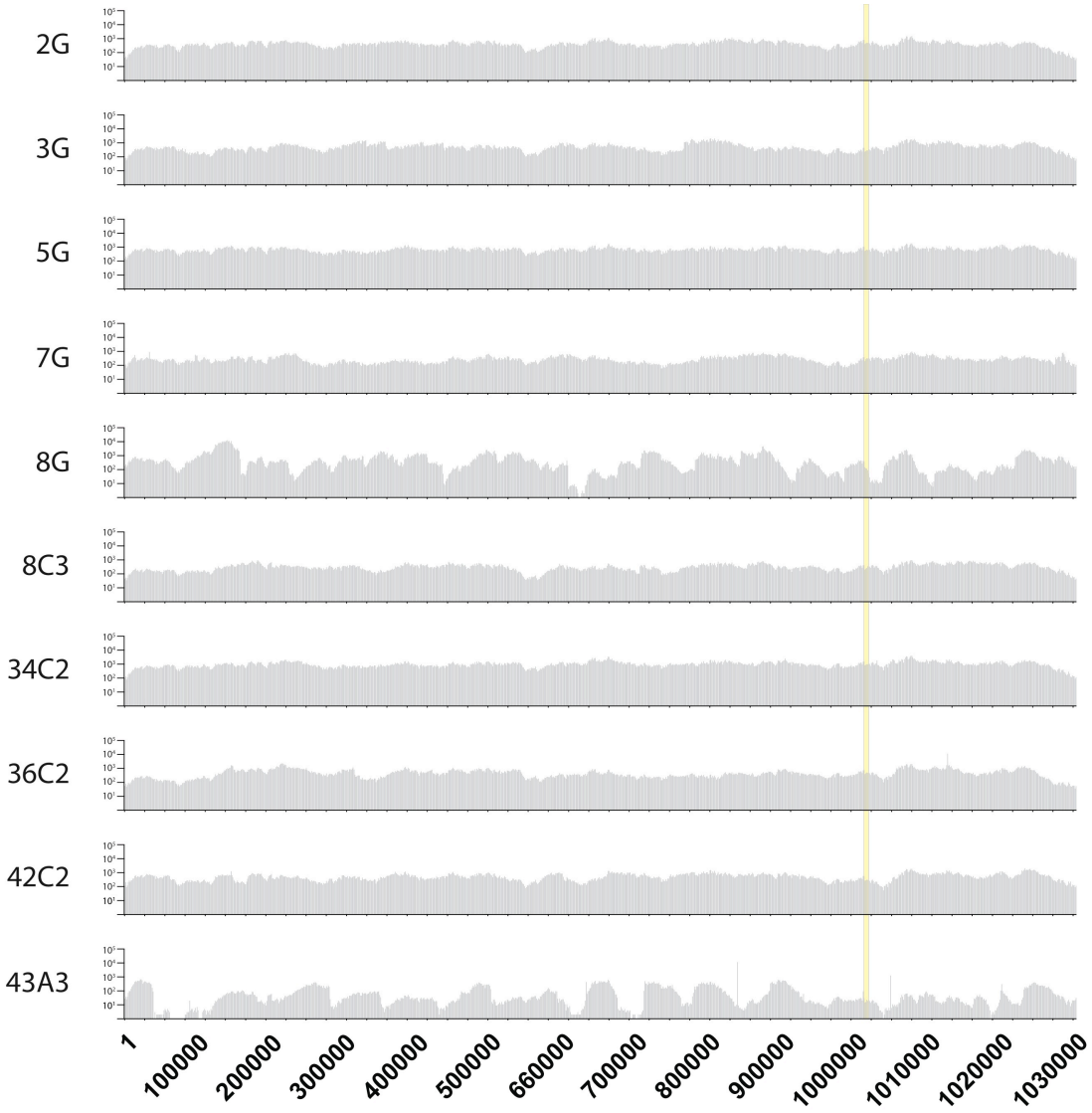

Figure 4.6. Coverage distribution.

B.

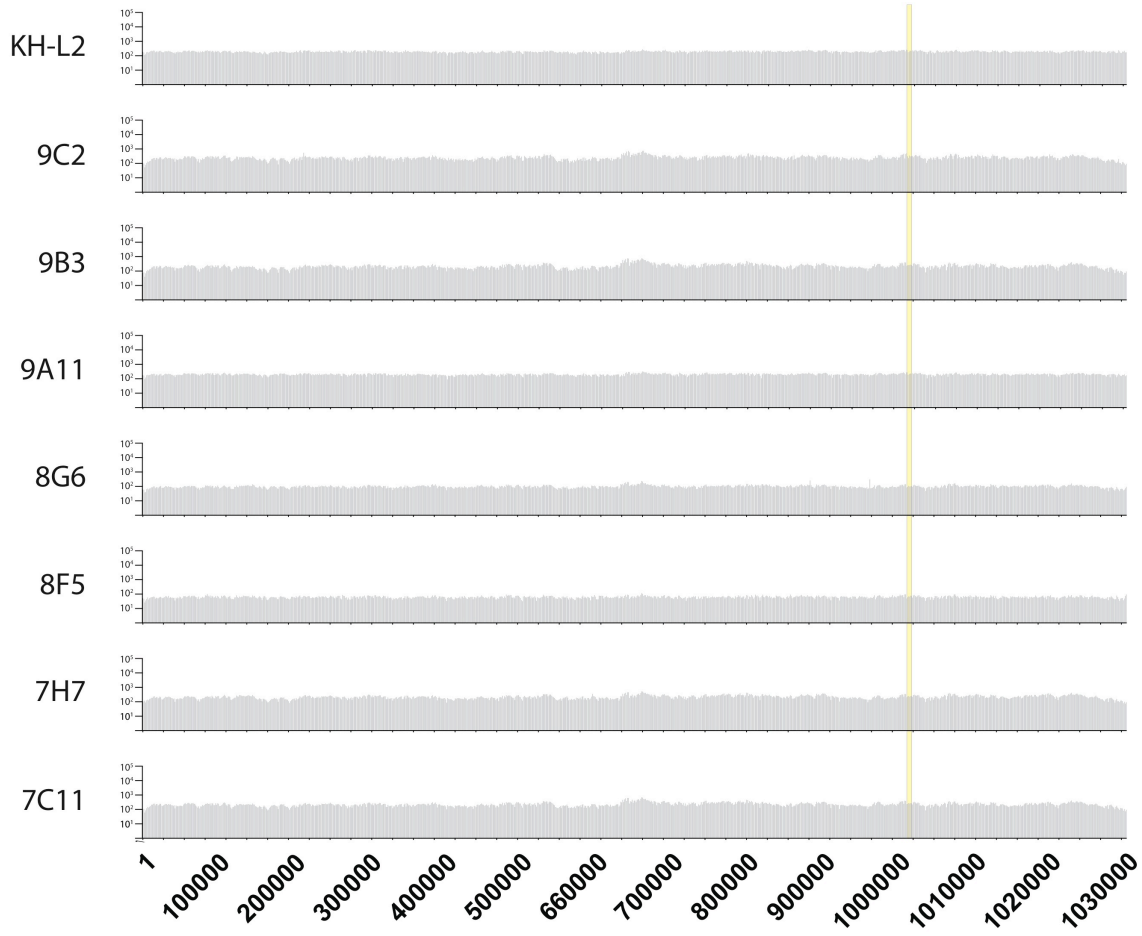

Figure 4.6. Coverage distribution (continued).

A. Average coverage per/1Kb in Group 1. **B**. Average coverage per/1Kb in Group 2. The yellow highlighted region indicates the target for quantitative PCR.

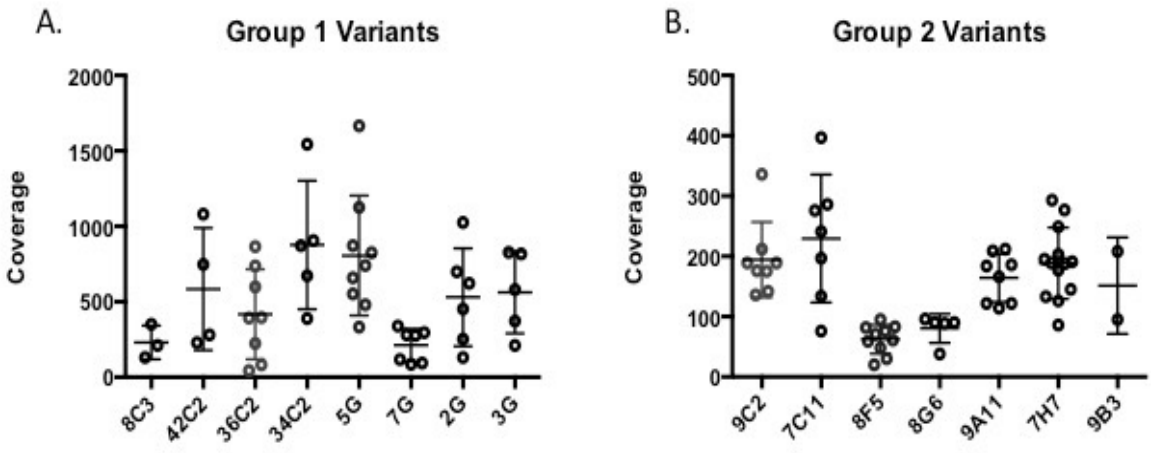

Figure 4.7. Group 1 and Group 2 variant analysis.

Number of single nucleotide polymorphisms and their coverage for A. Group 1 and B. Group 2.

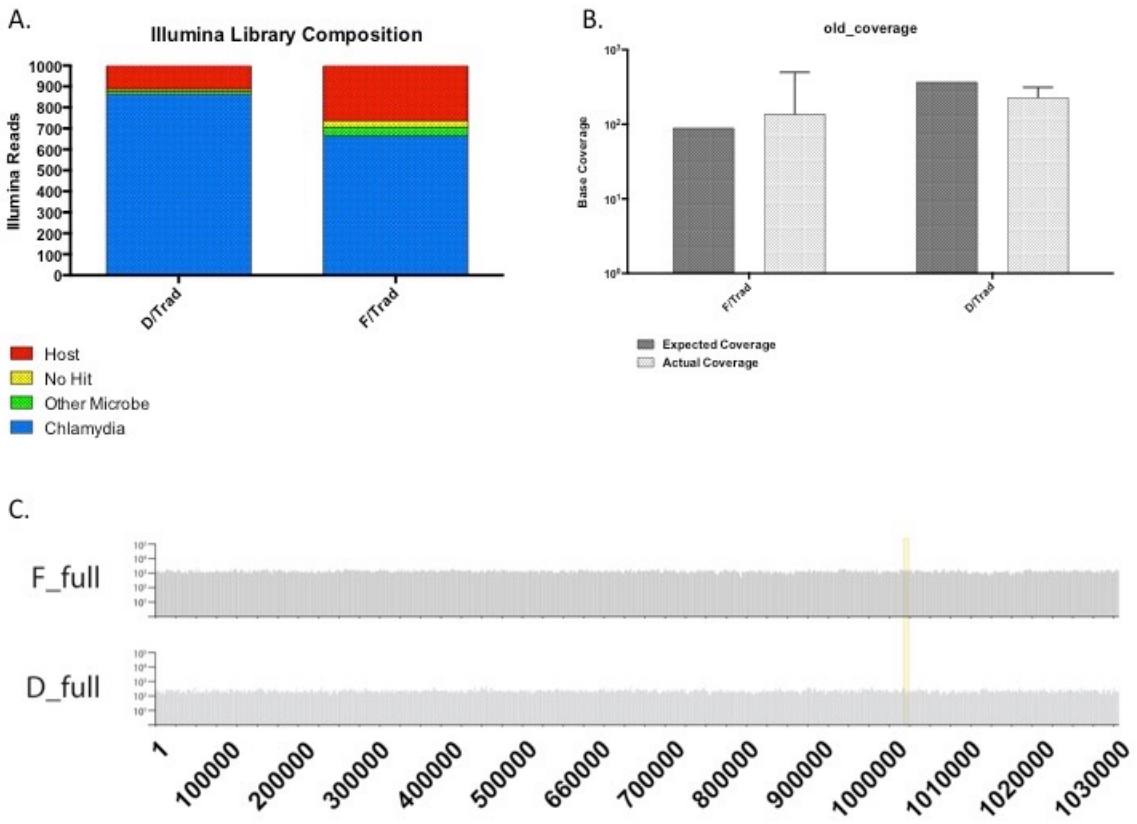

Figure 4.8. Traditional preparation Illumina library composition and coverage.

A. Illumina library composition of practice samples from BLAST screen of 1000 random reads. **B**. Expected and mean coverage of reference mapped Draft assemblies. **C.** Coverage distribution (average reads/1Kb) for traditionally prepared samples across reference genome. Yellow highlight indicates location of qPCR target.

GENERAL CONCLUSION

In 1998, 3 years after the first complete prokaryotic genome sequence was submitted to NCBI, *Haemophilus influenzae*, a publication in *Science* presented the first genome sequence of *Chlamydia trachomatis* (8). The contribution of knowledge and insight to the chlamydial research community could arguably rival decades of prior laboratory work. This genome sequence was among the first cohort of prokaryotic genome sequences published (~14th overall), and was an especially valuable genome sequence to its research community considering that, at the time, and until very recently, chlamydial species were genetically intractable. The inability to create site directed mutants, and genomic information being limited to sequences generated from PCR targeted regions, resulted in some very important questions about chlamydial biology going unanswered. The immense cost of this project (> $1,000,000) has come down significantly in the last 15 years due to technological advances primarily in the actual sequencing platforms.

However, obtaining sufficient high genomic template remained difficult and time consuming, representing a limiting factor to the amount of data that could be generated. The work presented in this document has streamlined the process in the preparation of chlamydial genomes for sequencing in terms of cost, time and difficulty, and improved clinical relevance of the sequences involved. This has allowed the sequencing of ~60 clinical and ~56 historical and laboratory strains, at less than $200 a

genome. The ability to include larger quantities of genomes in comparative studies has allowed much more robust comparative genome analysis.

An example of the utility of such methods is apparent in our study that expanded on the identification of widespread recombination in chlamydial genomes (55), previously by our research group. A sliding window method was applied to various clinical genomes to identify regions of recombination. While the exact location of crossing over is impossible to tell in most cases, due to conserved sequence masking such events, informative sites indicated extensive regions of recombination. In 2013, our group published findings that, when co-infection of cell culture were carried out, recombination is widespread, apparently random, and involves regions of double crossover greater than 50 Kb in length (91).

Our results, presented in Chapter 2 of this document, corroborate these findings in culture-independent clinical specimens, identifying multiple genomes that have genomic backgrounds that don't match the supposed origins of their *ompA* gene (115). Parallel sequencing of multiple clinical specimens at once identified a geographically and temporally linked clade of clinical strains that share regions of recombination, not resultant of clonal expansion of a single lineage. This is an important distinction, made by the fact that, while the internal sequence in these crossover regions are from similar origin, the sites of crossover appear different within this clade. These variable margins implicate recombination of specific regions, into different genomic backgrounds, as

analogous but independent events that are being selected for. The means by which *Chlamydia* complete crossing over of genetic information is not presently understood.

While the mechanisms of lateral gene transfer (LGT) in *Chlamydia* (144, 145) remain unknown, the size and scope of the regions of recombination, identified by us and previously by our group (55, 91, 115) are more prolific than would be expected from uptake of fragments of free floating DNA acquired from the environment. Possible mechanisms for this widespread crossover can only be speculated about without further experimentation, but a study investigating similar and interesting phenomenon in *Mycobacteria spp*. (146) presents an interesting mechanism that may be similar in nature to what we see in *Chlamydia.*

Multiple genome wide LGT events of large segments of the genomes were discovered in a comparative genome study of *M. smegmatis*. The conjugation-derived encounters between bacterial cells create a product that is meiotic like in nature, leading the authors to compare it to sexual reproduction in bacterial cells(147). This would create a means for prolific expansion of diversity in any capable organism. However, if *Chlamydia* undergoes its own version of the meiotic-like distributive conjugal transfer (DCT) of the chromosome, similarly to *M. smegmatis*, it must move the two parental genomes near enough to each other to undergo LGT. While chlamydial relatives such as *Protochlamydia amoebophila* encode the genes required for F-like conjugation(148), the reduced genome of *C. trachomatis* retains none of these genes, further there is no

evidence, or prior speculation, that *Chlamydia* take part in conjugation (144). This requires that mechanisms other than F-like conjugation need to be considered and investigated; an interesting mechanism to start with would be cellular fusion.

Chlamydia has a unique cell wall structure in multiple ways. As mentioned peptidoglycan is not present in chlamydial cells except to aid in septum formation during cell division (149-151). The emergence of peptidoglycan cell wall structures in bacteria is proposed to be an evolutionary contributor to bacterial genome stability (152). The lack of peptidoglycan in the cellular structure of C*hlamydia* is then curious because *Chlamydia,* as an obligate intracellular bacterium, has cultivated a stable, conserved and reduced genome; any changes to such a reduced and parsimonious genome are likely to be deleterious to the cell. Therefore, *Chlamydia* has evolved its own method for stabilizing its cellular structure.

While *Chlamydia* doesn't form a peptidoglycan sacculus to give stability to its cellular structure, or its genome, the *Chlamydia* specific proteins *omcA* and *omcB*, are a functional equivalent conferring rigidity to chlamydial EBs through sulfide amino acid crosslinking on the cell surface (153). Another relevant observation is that during this differentiation, the reticulate body is a spheroplast-like structure, the Gram-negative equivalent of a protoplast or single membrane. Bacteria (154). Spheroplasts are essentially Gram-negative cells without peptidoglycan and are capable of fusing with

other spheroplasts (155). This is implicated with the ability to undergo genetic recombination in. *Escherichia coli* (155) and *Mycobacterium aurum* (156).

Considering *Chlamydia* are Gram-negative with no structural peptidoglycan, undergo a spheroplast-like stage during the developmental cycle, and exhibit extensive recombination involving large tracks of the genome (55, 115), cellular fusion is an attractive hypothesis to pursue. Cellular fusion would support the idea of distributive transfer in *Chlamydia* by making available the entire genome, while maintaining genome stability, as the fusion would be *Chlamydia* specific. This type of distributive transfer makes sense in the context of a conserved genome with low variability. Rather than inserting fragments of free DNA, prone to disrupting synteny in the genome, genetic exchange would take place between two complete genomes maintaining similar structure and order.

To test a hypothesis that distributive, genome wide recombination occurs in *Chlamydia* by cellular fusion of chlamydial cells in spheroplast-like forms, a method for enrichment of fused chlamydial cells would first need to be optimized. Using similar immunomagnetic technologies to those presented in this document, a dual enrichment of co-infected chlamydial serovars would be one approach. This could include infecting a monolayer with 2 different serovars (i.e. D and E) and using primary antibodies to target the cell surface in a serovar specific way; differing from the previously described approach that targets LPS and would bind to any chlamydial cells in the sample.

During the first phase of the development cycle when chlamydial cells have been observed as spheroplast-like bodies, the cells would be subjected to IMS using anti-serovar monoclonal antibodies. An initial enrichment would target any chlamydial cell that has a D MOMP on its outer membrane. The eluate from this enrichment would then be incubated in H_20 to remove surface bound antibodies. A second round of IMS would then be performed, this time targeting E MOMP. This second enrichment would sequester any cell with an E MOMP that had been sequestered in the previous round because a D MOMP was associated with it. This dual enrichment would hypothetically result in enrichment for cells that are fused, as there would be physical attachments that linked the 2 MOMP proteins together, resulting in enrichment in both rounds of IMS.

This enriched population of cells could then be investigated in at least 3 ways. 1) Applying dual fluorescent antibody staining targeting each MOMP (E and D) with different wavelengths to look for a blended color under fluorescent microscopy, indicating cells labeled with both MOMP antibodies. 2) Transmission electron microscopy hypothesizing that the majority of cells in the eluate would be fused cells. 3) Genome sequencing this pool of fused cells to look for a higher rate of recombination in the dual enriched population than in the population prior to enrichment.

These are just suggestion of where to begin; there are many avenues to pursue including electro field fusion of spheroplasts as demonstrated in Ruthe et al. 1985 (157) where they successfully fused *E. coli* and *Salmonella typhimurium* spheroplasts and recovered viable cells with a combination of genetic markers, indicating recombination resulting from fusion. The widespread recombination uncovered by our work highlights the need to uncover the mechanism for recombination in *Chlamydia* and this would be an exciting place to start.

In conclusion it is an exciting time to be a chlamydial researcher; the ever improving technologies, such as those presented in this document, allow for robust genome sequencing studies that were not possible in the past. This leads to the ability to answer questions that were previously unmanageable, and learn more about *Chlamydia* than has ever been possible.

BIBLIOGRAPHY

1. **Collingro A**, **Tischler P**, **Weinmaier T**, **Penz T**, **Heinz E**, **Brunham RC**, **Read TD**, **Bavoil PM**, **Sachse K**, **Kahane S**, **Friedman MG**, **Rattei T**, **Myers GSA**, **Horn M**. 2011. Unity in Variety--The Pan-Genome of the Chlamydiae. Molecular Biology and Evolution **28**:3253–3270.
2. **Wright HR**, **Turner A**, **Taylor HR**. 2007. Trachoma and poverty: unnecessary blindness further disadvantages the poorest people in the poorest countries. Clin Exp Optom **90**:422–428.
3. **Geisler WM**, **Tang JM**, **Wang CB**, **Wilson CM**, **Kaslow RA**. 2004. Epidemiological and genetic correlates of incident Chlamydia trachomatis infection in North American adolescents. J Infect Dis **190**:1723–1729.
4. **Ma B**, **Forney LJ**, **Ravel J**. 2012. Vaginal Microbiome: Rethinking Health and Disease. Annu Rev Microbiol **66**:371–389.
5. **Clarke IN**. 2011. Evolution of Chlamydia trachomatis. Annals of the New York Academy of Sciences **1230**:E11–E18.
6. **Read TD**, **Myers G**, **Brunham RC**, **Nelson WC**, **Paulsen IT**, **Heidelberg J**, **Holtzapple E**, **Khouri H**, **Federova NB**, **Carty HA**, **Umayam LA**, **Haft DH**, **Peterson J**, **Beanan MJ**, **White O**, **Salzberg SL**, **Hsia RC**, **McClarty G**, **Rank RG**, **Bavoil PM**, **Fraser CM**. 2003. Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. Nucleic Acids Res **31**:2134–2147.
7. **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**, **Fraser CM**. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res **28**:1397–1406.
8. **Stephens RS**. 1998. Genome Sequence of an Obligate Intracellular Pathogen of Humans: Chlamydia trachomatis. Science **282**:754–759.
9. **Kalman S**, **Mitchell W**, **Marathe R**, **Lammel C**, **Fan J**, **Hyman RW**, **Olinger L**, **Grimwood J**, **Davis RW**, **Stephens RS**. 1999. Comparative genomes of Chlamydia pneumoniae and C. trachomatis. Nat Genet **21**:385–389.
10. **Voigt A**, **Schöfl G**, **Saluz HP**. 2012. The Chlamydia psittaci genome: a comparative analysis of intracellular pathogens. PLoS ONE **7**:e35097–e35097.
11. **Azuma Y**, **Hirakawa H**, **Yamashita A**, **Cai Y**, **Rahman MA**, **Suzuki H**, **Mitaku S**, **Toh H**, **Goto S**, **Murakami T**, **Sugi K**, **Hayashi H**, **Fukushi H**, **Hattori M**, **Kuhara S**, **Shirai M**. 2006. Genome sequence of the cat pathogen, Chlamydophila felis. DNA Res **13**:15–23.
12. **Blattner FR**, **Plunkett G**, **Bloch CA**, **Perna NT**, **Burland V**, **Riley M**, **ColladoVides**

J, **Glasner JD**, **Rode CK**, **Mayhew GF**, **Gregor J**, **Davis NW**, **Kirkpatrick HA**, **Goeden MA**, **Rose DJ**, **Mau B**, **Shao Y**. 1997. The complete genome sequence of Escherichia coli K-12. Science **277**:1453–&.

13. **Stover CK**, **Pham XQ**, **Erwin AL**, **Mizoguchi SD**, **Warrener P**, **Hickey MJ**, **Brinkman F**, **Hufnagle WO**, **Kowalik DJ**, **Lagrou M**, **Garber RL**, **Goltry L**, **Tolentino E**, **Westbrock-Wadman S**, **Yuan Y**, **Brody LL**, **Coulter SN**, **Folger KR**, **Kas A**, **Larbig K**, **Lim R**, **Smith K**, **Spencer D**, **Wong G**, **Wu Z**, **Paulsen IT**, **Reizer J**, **Saier MH**, **Hancock R**, **Lory S**, **Olson MV**. 2000. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature **406**:959–964.
14. **Pannekoek Y**, **van der Ende A**. 2006. Inclusion proteins of Chlamydiaceae. Drugs Today (Barc) **42**:65–73.
15. **Byrne GI**. 2010. Chlamydia trachomatisStrains and Virulence: Rethinking Links to Infection Prevalence and Disease Severity. J INFECT DIS **201**:126–133.
16. **Zhong G**. 2011. Chlamydia trachomatis secretion of proteases for manipulating host signaling pathways. Frontiers in Microbiology **2**:14–14.
17. **Clifton DR**, **Fields KA**, **Grieshaber SS**, **Dooley CA**, **Fischer ER**, **Mead DJ**, **Carabeo RA**, **Hackstadt T**. 2004. A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. Proceedings of the National Academy of Sciences **101**:10166–10171.
18. **Jiwani S**, **Ohr RJ**, **Fischer ER**, **Hackstadt T**, **Alvarado S**, **Romero A**, **Jewett TJ**. 2012. Chlamydia trachomatis Tarp cooperates with the Arp2/3 complex to increase the rate of actin polymerization. Biochem Biophys Res Commun **420**:816–821.
19. **Schwarzenbacher R**, **Stenner-Liewen F**, **Liewen H**, **Robinson H**, **Yuan H**, **Bossy-Wetzel E**, **Reed JC**, **Liddington RC**. 2004. Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. Journal of Biological Chemistry **279**:29320–29324.
20. **Fling SP**, **Sutherland RA**, **Steele LN**, **Hess B**, **D'Orazio S**, **Maisonneuve JF**, **Lampe MF**, **Probst P**, **Starnbach MN**. 2001. CD8(+) T cells recognize an inclusion membrane-associated protein from the vacuolar pathogen Chlamydia trachomatis. Proceedings of the National Academy of Sciences **98**:1160–1165.
21. **Misaghi S**, **Balsara ZR**, **Catic A**, **Spooner E**, **Ploegh HL**, **Starnbach MN**. 2006. Chlamydia trachomatis-derived deubiquitinating enzymes in mammalian cells during infection. Molecular Microbiology **61**:142–150.
22. **McLeod MP**, **Qin X**, **Karpathy SE**, **Gioia J**, **Highlander SK**, **Fox GE**, **McNeill TZ**, **Jiang H**, **Muzny D**, **Jacob LS**, **Hawes AC**, **Sodergren E**, **Gill R**, **Hume J**, **Morgan M**, **Fan G**, **Amin AG**, **Gibbs RA**, **Hong C**, **Yu X-J**, **Walker DH**, **Weinstock GM**. 2004. Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae. Journal of Bacteriology **186**:5842–5855.
23. **Sallstrom B**, **Andersson S**. 2005. Genome reduction in the alpha-proteobacteria. Curr Opin Microbiol **8**:579–585.
24. **Fraser CM**, **Norris SJ**, **Weinstock GM**, **White O**, **Sutton GG**, **Dodson R**, **Gwinn**

M, **Hickey EK**, **Clayton R**, **Ketchum KA**, **Sodergren E**, **Hardham JM**, **McLeod MP**, **Salzberg S**, **Peterson J**, **Khalak H**, **Richardson D**, **Howell JK**, **Chidambaram M**, **Utterback T**, **McDonald L**, **Artiach P**, **Bowman C**, **Cotton MD**, **Fujii C**, **Garland S**, **Hatch B**, **Horst K**, **Roberts K**, **Sandusky M**, **Weidman J**, **Smith HO**, **Venter JC**. 1998. Complete genome sequence of Treponema pallidum, the syphilis spirochete. Science **281**:375–388.

25. **Fadiel A**, **Eichenbaum KD**, **Semary El N**, **Epperson B**. 2007. Mycoplasma genomics: tailoring the genome for minimal life requirements through reductive evolution. Front Biosci (Landmark Ed) **12**:2020–2028.
26. **ROCKEY DD**. 2011. Unraveling the basic biology and clinical significance of the chlamydial plasmid. J Exp Med **208**:2159–2162.
27. **Carlson JH**, **Whitmire WM**, **Crane DD**, **Wicke L**, **Virtaneva K**, **Sturdevant DE**, **Kupko JJ**, **Porcella SF**, **Martinez-Orengo N**, **Heinzen RA**, **Kari L**, **Caldwell HD**. 2008. The Chlamydia trachomatis plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infection and Immunity **76**:2273–2283.
28. **Hacker J**, **Kaper JB**. 2000. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol **54**:641–679.
29. **Unemo M**, **Clarke IN**. 2011. The Swedish new variant of Chlamydia trachomatis. Current opinion in infectious diseases **24**:62.
30. **Kim JF**. 2001. Revisiting the chlamydial type III protein secretion system: clues to the origin of type III protein secretion. Trends Genet **17**:65–69.
31. **Dobrindt U**, **Chowdary MG**, **Krumbholz G**, **Hacker J**. 2010. Genome dynamics and its impact on evolution of Escherichia coli. Med Microbiol Immunol **199**:145–154.
32. **Dugan J**, **Rockey DD**, **Jones L**, **Andersen AA**. 2004. Tetracycline resistance in Chlamydia suis mediated by genomic islands inserted into the chlamydial inv-like gene. Antimicrobial Agents and Chemotherapy **48**:3989–3995.
33. **Dugan J**, **Andersen AA**, **ROCKEY DD**. 2007. Functional characterization of IScs605, an insertion element carried by tetracycline-resistant Chlamydia suis. Microbiology **153**:71–79.
34. **di Francesco A**, **Donati M**, **Rossi M**, **Pignanelli S**, **Shurdhi A**, **Baldelli R**, **Cevenini R**. 2008. Tetracycline-resistant Chlamydia suis isolates in Italy. Vet Rec **163**:251–252.
35. **Suchland R**, **Sandoz K**, **Jeffrey B**, **Stamm W**, **Rockey D**. 2009. Horizontal transfer of tetracycline resistance among Chlamydia spp. in vitro. Antimicrobial Agents and Chemotherapy **53**:4604–4611.
36. **Yuan Y**, **Zhang YX**, **Watkins NG**, **Caldwell HD**. 1989. Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 Chlamydia trachomatis serovars. Infection and Immunity **57**:1040–1049.
37. **Lampe MF**, **Suchland RJ**, **Stamm WE**. 1993. Nucleotide sequence of the variable domains within the major outer membrane protein gene from serovariants of Chlamydia trachomatis. Infection and Immunity **61**:213–219.

38. **BEATTY WL**, **BELANGER TA**, **DESAI AA**, **MORRISON RP**, **BYRNE GI**. 1994. Tryptophan Depletion as a Mechanism of Gamma-Interferon-Mediated Chlamydial Persistence. Infection and Immunity **62**:3705–3711.
39. **Roshick C**, **Wood H**, **Caldwell HD**, **McClarty G**. 2006. Comparison of gamma interferon-mediated antichlamydial defense mechanisms in human and mouse cells. Infection and Immunity **74**:225–238.
40. **Caldwell HD**, **Wood H**, **Crane D**, **Bailey R**, **Jones RB**, **Mabey D**, **Maclean I**, **Mohammed Z**, **Peeling R**, **Roshick C**, **Schachter J**, **Solomon AW**, **Stamm WE**, **Suchland RJ**, **Taylor L**, **West SK**, **Quinn TC**, **Belland RJ**, **Grant McClarty**. Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates
.
41. **Belland RJ**, **Scidmore MA**, **Crane DD**, **Hogan DM**, **Whitmire W**, **McClarty G**, **Caldwell HD**. 2001. Chlamydia trachomatis cytotoxicity associated with complete and partial cytotoxin genes. Proceedings of the National Academy of Sciences **98**:13984–13989.
42. **Carlson JH**, **Hughes S**, **Hogan D**, **Cieplak G**, **Sturdevant DE**, **McClarty G**, **Caldwell HD**, **Belland RJ**. 2004. Polymorphisms in the Chlamydia trachomatis cytotoxin locus associated with ocular and genital isolates. Infection and Immunity **72**:7063–7072.
43. **Crane DD**, **Dorward DW**, **Nelson DE**. 2006. Inhibition of Chlamydiae by Primary Alcohols Correlates with the Strain-Specific Complement of Plasticity Zone Phospholipase D Genes. Infection and ….
44. **Wheelhouse N**, **Aitchison K**, **Spalding L**, **Livingstone M**, **Longbottom D**. 2009. Transcriptional analysis of in vitro expression patterns of Chlamydophila abortus polymorphic outer membrane proteins during the chlamydial developmental cycle. Vet Res **40**:–47.
45. **Wheelhouse N**, **Sait M**, **Wilson K**, **Aitchison K**, **McLean K**, **Smith DGE**, **Longbottom D**. 2012. Expression patterns of five polymorphic membrane proteins during the Chlamydia abortus developmental cycle. Vet Microbiol **160**:525–529.
46. **Clifton DR**, **Dooley CA**, **Grieshaber SS**, **Carabeo RA**, **Fields KA**, **Hackstadt T**. 2005. Tyrosine phosphorylation of the chlamydial effector protein Tarp is species specific and not required for recruitment of actin. Infection and Immunity **73**:3860–3868.
47. **Carlson JH**, **Porcella SF**, **McClarty G**, **Caldwell HD**. Comparative Genomic Analysis of Chlamydia trachomatis Oculotropic and Genitotropic Strains†
.
48. **Dean D**, **Suchland RJ**, **Stamm WE**. 2000. Evidence for long-term cervical persistence of Chlamydia trachomatis by omp1 genotyping. J Infect Dis **182**:909–916.
49. **Rockey DD**, **Viratyosin W**, **Bannantine JP**, **Suchland RJ**, **Stamm WE**. 2002. Diversity within inc genes of clinical Chlamydia trachomatis variant isolates that occupy non-fusogenic inclusions. Microbiology **148**:2497–2505.

50. **Geisler WM**, **Suchland RJ**, **Rockey DD**, **Stamm WE**. 2001. Epidemiology and clinical manifestations of unique Chlamydia trachomatis isolates that occupy nonfusogenic inclusions. J Infect Dis **184**:879–884.
51. **Suchland RJ**, **Jeffrey BM**, **Xia M**, **Bhatia A**, **Chu HG**, **Rockey DD**, **Stamm WE**. 2008. Identification of Concomitant Infection with Chlamydia trachomatis IncA-Negative Mutant and Wild-Type Strains by Genomic, Transcriptional, and Biological Characterizations. Infection and Immunity **76**:5438–5446.
52. **Sturdevant GL**, **Kari L**, **Gardner DJ**, **Olivares-Zavaleta N**, **Randall LB**, **Whitmire WM**, **Carlson JH**, **Goheen MM**, **Selleck EM**, **Martens C**. 2010. Frameshift mutations in a single novel virulence factor alter the in vivo pathogenicity of Chlamydia trachomatis for the female murine genital tract. Infection and Immunity **78**:3660–3668.
53. **Tan C**, **Hsia R-C**, **Shou H**, **Carrasco JA**, **Rank RG**, **Bavoil PM**. 2010. Variable expression of surface-exposed polymorphic membrane proteins in in vitro-grown Chlamydia trachomatis. Cellular Microbiology **12**:174–187.
54. **Kari L**, **Whitmire WM**, **Carlson JH**, **Crane DD**, **Reveneau N**, **Nelson DE**, **Mabey DCW**, **Bailey RL**, **Holland MJ**, **McClarty G**, **Caldwell HD**. 2008. Pathogenic diversity among Chlamydia trachomatis ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis **197**:449–456.
55. **Jeffrey BM**, **Suchland RJ**, **Quinn KL**, **Davidson JR**, **Stamm WE**, **Rockey DD**. 2010. Genome Sequencing of Recent Clinical Chlamydia trachomatis Strains Identifies Loci Associated with Tissue Tropism and Regions of Apparent Recombination. Infection and Immunity **78**:2544–2553.
56. **Klint M**, **Fuxelius HH**, **Goldkuhl RR**, **Skarin H**, **Rutemark C**, **Andersson SGE**, **Persson K**, **Herrmann B**. 2007. High-Resolution Genotyping of Chlamydia trachomatis Strains by Multilocus Sequence Analysis. Journal of Clinical Microbiology **45**:1410–1414.
57. **Yibing Wang RJSLTCEAINCPM**. 2011. Evaluation of a High Resolution Genotyping Method for Chlamydia trachomatis Using Routine Clinical Samples. PLoS ONE **6**.
58. **Dean D**. 2009. Chlamydia trachomatis today: treatment, detection, immunogenetics and the need for a greater global understanding of chlamydial disease pathogenesis. Drugs of Today (Barcelona, Spain: 1998) **45**:25.
59. **Pedersen LN**, **Pødenphant L**, **Møller JK**. 2008. Highly discriminative genotyping of Chlamydia trachomatis using omp1 and a set of variable number tandem repeats. Clinical Microbiology and Infection **14**:644–652.
60. **Peuchant O**, **Le Roy C**, **Herrmann B**, **Clerc M**, **Bébéar C**, **de Barbeyrac B**. 2012. MLVA Subtyping of Genovar E Chlamydia trachomatis Individualizes the Swedish Variant and Anorectal Isolates from Men who Have Sex with Men. PLoS ONE **7**:–e31538.
61. **Naraporn Somboonna RWDMOMAPSJJACRHTDRDD**. 2011. Hypervirulent Chlamydia trachomatis Clinical Strain Is a Recombinant between Lymphogranuloma Venereum (L2) and D Lineages. mBio **2**.

62. **Harris SR**, **Clarke IN**, **Seth-Smith HMB**, **Solomon AW**, **Cutcliffe LT**, **Marsh P**, **Skilton RJ**, **Holland MJ**, **Mabey D**, **Peeling RW**, **Lewis DA**, **Spratt BG**, **Unemo M**, **Persson K**, **Bjartling C**, **Brunham R**, **de Vries HJC**, **Morré SA**, **Speksnijder A**, **Bébéar CM**, **Clerc M**, **de Barbeyrac B**, **Parkhill J**, **Thomson NR**. 2012. Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet **44**:413–419.

63. **Gomes JP**, **Bruno WJ**, **Nunes A**, **Santos N**, **Florindo C**, **Borrego MJ**, **Dean D**. 2007. Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots. Genome Research **17**:50–60.

64. **Spaargaren J**, **Fennema J**, **Morré SA**, **de Vries HJC**, **Coutinho RA**. 2005. New lymphogranuloma venereum Chlamydia trachomatis variant, Amsterdam. Emerg Infect Dis **11**:1090–1092.

65. **Albrecht M**, **Sharma CM**, **Reinhardt R**, **Vogel J**, **Rudel T**. 2010. Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. Nucleic Acids Res **38**:868–877.

66. **Jorgensen I**, **Valdivia RH**. 2008. Pmp-Like Proteins Pls1 and Pls2 Are Secreted into the Lumen of the Chlamydia trachomatis Inclusion. Infection and Immunity **76**:3940–3950.

67. **Joseph SJ**, **Didelot X**, **Gandhi K**, **Dean D**, **Read TD**. 2011. Interplay of recombination and selection in the genomes of Chlamydia trachomatis. Biology Direct **6**:28.

68. **Thalmann J**, **Janik K**, **May M**, **Sommer K**, **Ebeling J**, **Hofmann F**, **Genth H**, **Klos A**. 2010. Actin re-organization induced by Chlamydia trachomatis serovar D--evidence for a critical role of the effector protein CT166 targeting Rac. PLoS ONE **5**:e9887–e9887.

69. **João P Gomes ANWJBMJBCFDD**. 2006. Polymorphisms in the Nine Polymorphic Membrane Proteins of Chlamydia trachomatis across All Serovars: Evidence for Serovar Da Recombination and Correlation with Tissue Tropism. Journal of Bacteriology **188**:275.

70. **Bannantine J**, **Griffiths R**, **Viratyosin W**, **Brown W**, **Rockey D**. 2000. A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. Cellular Microbiology **2**:35–47.

71. **Dean D**, **Bruno WJ**, **Wan R**, **Gomes JP**, **Devignot S**, **Mehari T**, **de Vries HJC**, **Morré SA**, **Myers G**, **Read TD**. 2009. Predicting phenotype and emerging strains among Chlamydia trachomatis infections. Emerg. Infect. Dis. **15**:1385.

72. **Pannekoek Y**, **Dickx V**, **Beeckman DSA**, **Jolley KA**, **Keijzers WC**, **Vretou E**, **Maiden MCJ**, **Vanrompay D**, **van der Ende A**. 2010. Multi locus sequence typing of Chlamydia reveals an association between Chlamydia psittaci genotypes and host species. PLoS ONE **5**:e14179.

73. **Savy M**, **Hennig BJ**, **Doherty CP**, **Fulford AJ**, **Bailey R**, **Holland MJ**, **Sirugo G**, **Rockett KA**, **Kwiatkowski DP**, **Prentice AM**, **Cox SE**. 2010. Haptoglobin and Sickle Cell Polymorphisms and Risk of Active Trachoma in Gambian Children. PLoS ONE **5**:–e11075.

74. **Wang W**, **Stassen FR**, **Surcel HM**, **Ohman H**, **Tiitinen A**, **Paavonen J**, **de Vries HJC**, **Heijmans R**, **Pleijster J**, **Morré SA**, **Ouburg S**. 2009. ANALYSES OF POLYMORPHISMS IN THE INFLAMMASOME-ASSOCIATED NLRP3 AND miRNA-146A GENES IN THE SUSCEPTIBILITY TO AND TUBAL PATHOLOGY OF CHLAMYDIA TRACHOMATIS INFECTION. Drugs Today (Barc) **45**:95–103.
75. **Hartog den JE**, **Lyons JM**, **Ouburg S**, **Fennema JSA**, **de Vries HJC**, **Bruggeman CA**, **Ito JI**, **Peña AS**, **Land JA**, **Morré SA**, **Chlamydi IAS**, **Consortium E**. 2009. Tlr4 in Chlamydia Trachomatis Infections: Knockout Mice, Std Patients and Women with Tubal Factor Subfertility. Drugs Today (Barc) **45**:75–82.
76. **Rantala A**, **Lajunen T**, **Juvonen R**, **Bloigu A**, **Paldanius M**, **Silvennoinen-Kassinen S**, **Peitso A**, **Vainio O**, **Leinonen M**, **Saikku P**. 2011. Low mannose-binding lectin levels and MBL2 gene polymorphisms associate with Chlamydia pneumoniae antibodies. Innate Immun **17**:35–40.
77. **Hartog den JE**, **Ouburg S**, **Land JA**, **Lyons JM**, **Ito JI**, **Peña AS**, **Morré SA**. 2006. Do host genetic traits in the bacterial sensing system play a role in the development of Chlamydia trachomatis-associated tubal pathology in subfertile women? BMC Infect Dis **6**:122–122.
78. **Murillo LS**, **Land JA**, **Pleijster J**, **Bruggeman CA**, **Peña AS**, **Morré SA**. 2003. Interleukin-1B (IL-1B) and interleukin-1 receptor antagonist (IL-1RN) gene polymorphisms are not associated with tubal pathology and Chlamydia trachomatis-related tubal factor subfertility. Human Reproduction **18**:2309–2314.
79. **Natividad A**, **Hull J**, **Luoni G**, **Holland M**, **Rockett K**, **Joof H**, **Burton M**, **Mabey D**, **Kwiatkowski D**, **Bailey R**. 2009. Innate immunity in ocular Chlamydia trachomatis infection: contribution of IL8 and CSF2 gene variants to risk of trachomatous scarring in Gambians. BMC Medical Genetics **10**.
80. **Angels Natividad GCMJHMJBHMJKRDPKDCMRLB**. 2006. A coding polymorphism in matrix metalloproteinase 9 reduces risk of scarring sequelae of ocular Chlamydia trachomatis infection. BMC Medical Genetics **7**:40.
81. **Laisk T**, **Peters M**, **Saare M**, **Haller-Kikkatalo K**, **Karro H**, **Salumets A**. 2010. Association of CCR5, TLR2, TLR4 and MBL genetic variations with genital tract infections and tubal factor infertility. J Reprod Immunol **87**:74–81.
82. **Karimi O**, **Ouburg S**, **de Vries HJC**, **Peña AS**, **Pleijster J**, **Land JA**, **Morré SA**. 2009. Tlr2 Haplotypes in the Susceptibility to and Severity of Chlamydia Trachomatis Infections in Dutch Women. Drugs Today (Barc) **45**:67–74.
83. **Mei B**, **Luo Q**, **Du K**, **Huo Z**, **Wang F**, **Yu P**. 2009. Association of MICA gene polymorphisms with Chlamydia trachomatis infection and related tubal pathology in infertile women. Human Reproduction **24**:3090–3095.
84. 2011. Cytokine gene polymorphism and Chlamydia trachomatis-specific immune responses **72**:278–282.
85. **Ohman H**, **Tiitinen A**, **Halttunen M**, **Lehtinen M**, **Paavonen J**, **Surcel HM**. 2009. Cytokine polymorphisms and severity of tubal damage in women with Chlamydia-associated infertility. J Infect Dis **199**:1353–1359.
86. **Wang C**, **Tang J**, **Geisler WM**, **Crowley-Nowick PA**, **Wilson CM**, **Kaslow RA**.

2005. Human leukocyte antigen and cytokine gene variants as predictors of recurrent Chlamydia trachomatis infection in high-risk adolescents. J Infect Dis **191**:1084–1092.

87. **Wang Y**, **Kahane S**, **Cutcliffe LT**, **Skilton RJ**, **Lambden PR**, **Clarke IN**. 2011. Development of a Transformation System for Chlamydia trachomatis: Restoration of Glycogen Biosynthesis by Acquisition of a Plasmid Shuttle Vector. PLoS Pathog **7**:e1002258.
88. **Gomes JP**, **Bruno WJ**, **Nunes A**, **Santos N**, **Florindo C**, **Borrego MJ**, **Dean D**. 2007. Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots. Genome Research **17**:50–60.
89. **Jeffrey BM**, **Suchland RJ**, **Quinn KL**, **Davidson JR**, **Stamm WE**, **Rockey DD**. 2010. Genome Sequencing of Recent Clinical Chlamydia trachomatis Strains Identifies Loci Associated with Tissue Tropism and Regions of Apparent Recombination. Infection and Immunity **78**:2544–2553.
90. **Srinivasan S**, **Hoffman NG**, **Morgan MT**, **Matsen FA**, **Fiedler TL**, **Hall RW**, **Ross FJ**, **McCoy CO**, **Bumgarner R**, **Marrazzo JM**, **Fredricks DN**. 2012. Bacterial Communities in Women with Bacterial Vaginosis: High Resolution Phylogenetic Analyses Reveal Relationships of Microbiota to Clinical Criteria. PLoS ONE **7**:e37818.
91. **Jeffrey BM**, **Suchland RJ**, **Eriksen SG**, **Sandoz KM**, **ROCKEY DD**. 2013. Genomic and phenotypic characterization of in vitro-generated Chlamydia trachomatis recombinants. BMC Microbiol **13**:–142.
92. **Borges VT**, **Ferreira R**, **Nunes A**, **Sousa-Uva M**, **Abreu M**, **Borrego MJ**, **Gomes JP**. 2013. Infection, Genetics and Evolution. INFECTION, GENETICS AND EVOLUTION **17**:23–32.
93. **Suchland RJ**, **Eckert LO**, **Hawes SE**, **Stamm WE**. 2003. Longitudinal assessment of infecting serovars of Chlamydia trachomatis in Seattle public health clinics: 1988-1996. Sexually transmitted diseases **30**:357.
94. **Eckert LO**, **Suchland RJ**, **Hawes SE**, **Stamm WE**. 2000. Quantitative Chlamydia trachomatis cultures: Correlation of chlamydial inclusion-forming units with serovar, age, sex, and race. J Infect Dis **182**:540–544.
95. **Suchland RJ**, **Stamm WE**. 1991. Simplified Microtiter Cell-Culture Method for Rapid Immunotyping of Chlamydia-Trachomatis. Journal of Clinical Microbiology **29**:1333–1338.
96. **Zhang YX**, **Stewart SJ**, **Caldwell HD**. 1989. Protective monoclonal antibodies to Chlamydia trachomatis serovar-and serogroup-specific major outer membrane protein determinants. Infection and Immunity **57**:636–638.
97. **Li H**, **Ruan J**, **Durbin R**. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research **18**:1851–1858.
98. **Jeck WR**, **Reinhardt JA**, **Baltrus DA**, **Hickenbotham MT**, **Magrini V**, **Mardis ER**, **Dangl JL**, **Jones CD**. 2007. Extending assembly of short DNA sequences to handle error. Bioinformatics **23**:2942–2944.
99. **Katoh K**, **Misawa K**, **Kuma K**, **Miyata T**. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids

Res **30**:3059–3066.

100. **Bruen TC**, **Philippe H**, **Bryant D**. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172**:2665–2681.
101. **Robert J Suchland DDRJPBWES**. 2000. Isolates of Chlamydia trachomatis That Occupy Nonfusogenic Inclusions Lack IncA, a Protein Localized to the Inclusion Membrane. Infection and Immunity **68**:360.
102. **Seth-Smith HMB**, **Harris SR**, **Scott P**, **Parmar S**, **Marsh P**, **Unemo M**, **Clarke IN**, **Parkhill J**, **Thomson NR**. 2013. Generating whole bacterial genome sequences of low-abundance species from complex samples with IMS-MDA. Nat Protoc **8**:2404–2412.
103. **A S**, **J S**, **De Sa C**, **K L**, **A R**. 1994. Identification of subspecies- and serotype 1-specific epitopes on the 80- to 90-kilodalton protein region of Chlamydia psittaci that may be useful for diagnosis of chlamydial induced abortion. Am J Vet Res **55**:510–514.
104. **Longbottom D**, **Coulter LJ**. 2003. Animal chlamydioses and zoonotic implications. J Comp Pathol **128**:217–244.
105. **Sargison ND**, **Truyers IGR**, **Howie FE**, **Thomson JR**, **Cox AL**, **Livingstone M**, **Longbottom D**. 2015. Identification of the 1B vaccine strain of Chlamydia abortusin aborted placentas during the investigation of toxaemic and systemic disease in sheep. New Zealand Veterinary Journal 1–11.
106. **Thomson NR**, **Yeats C**, **Bell K**, **Holden MT**, **Bentley SD**, **Livingstone M**, **Cerdeño-Tárraga AM**, **Harris B**, **Doggett J**, **Ormond D**. 2005. The Chlamydophila abortus genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Research **15**:629–640.
107. **Jorgensen DM**. 1997. Gestational psittacosis in a Montana sheep rancher. Emerg Infect Dis **3**:191–194.
108. **Longbottom D**, **Livingstone M**. 2006. Vaccination against chlamydial infections of man and animals. The Veterinary Journal **171**:263–275.
109. **Sait M**, **Clark EM**, **Wheelhouse N**, **Livingstone M**, **Spalding L**, **Siarkou VI**, **Vretou E**, **Smith DGE**, **Lainson FA**, **Longbottom D**. 2011. Genome sequence of the Chlamydophila abortus variant strain LLG. Journal of Bacteriology **193**:4276–4277.
110. **Entrican G**, **Buxton D**, **Longbottom D**. 2001. Chlamydial infection in sheep: immune control versus fetal pathology. Journal of the Royal Society of Medicine **94**:273–277.
111. **Bouakane A**, **Benchaïeb I**, **Rodolakis A**. 2003. Abortive potency of Chlamydophila abortus in pregnant mice is not directly correlated with placental and fetal colonization levels. Infection and Immunity **71**:7219–7222.
112. **Biesenkamp-Uhe C**, **Li Y**, **Hehnen H-R**, **Sachse K**, **Kaltenboeck B**. 2007. Therapeutic Chlamydophila abortus and C. pecorum vaccination transiently reduces bovine mastitis associated with Chlamydophila infection. Infection and Immunity **75**:870–877.
113. **Thomson NR**. 2005. The Chlamydophila abortus genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome

Research **15**:629–640.

114. **Pantchev A**, **Sting R**, **Bauerfeind R**, **Tyczka J**, **Sachse K**. 2010. Detection of all Chlamydophila and Chlamydia spp. of veterinary interest using species-specific real-time PCR assays. Comp Immunol Microbiol Infect Dis **33**:473–484.
115. **Putman TE**, **Suchland RJ**, **Ivanovitch JD**, **ROCKEY DD**. 2013. Culture-independent sequence analysis of Chlamydia trachomatis in urogenital specimens identifies regions of recombination and in-patient sequence mutations. Microbiology **159**:2109–2117.
116. **GIMENEZ DF**. 1964. STAINING RICKETTSIAE IN YOLK-SAC CULTURES. Stain Technol **39**:135–140.
117. **Zerbino DR**, **Birney E**. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research **18**:821–829.
118. **Putman TE**, **Suchland RJ**, **Ivanovitch JD**, **ROCKEY DD**. 2013. Culture-independent sequence analysis of Chlamydia trachomatis in urogenital specimens identifies regions of recombination and in-patient sequence mutations. Microbiology **159**:2109–2117.
119. **Millman KL**, **Tavaré S**, **Dean D**. 2001. Recombination in the ompA Gene but Not the omcB Gene of Chlamydia Contributes to Serovar-Specific Differences in Tissue Tropism, Immune Surveillance, and Persistence of the Organism. Journal of Bacteriology **183**:5997–6008.
120. **Gomes JP**, **Bruno WJ**, **Borrego MJ**, **Dean D**. 2004. Recombination in the Genome of Chlamydia trachomatis Involving the Polymorphic Membrane Protein C Gene Relative to ompA and Evidence for Horizontal Gene Transfer. Journal of Bacteriology **186**:4295–4306.
121. **Van Loock M**, **Vanrompay D**, **Herrmann B**, **Vander Stappen J**, **Volckaert G**, **Goddeeris BM**, **Everett KDE**. 2003. Missing links in the divergence of Chlamydophila abortus from Chlamydophila psittaci. Int J Syst Evol Microbiol **53**:761–770.
122. **Psarrakos P**, **Papadogeorgakis E**, **Sachse K**, **Vretou E**. 2011. Chlamydia trachomatis ompA genotypes in male patients with urethritis in Greece - Conservation of the serovar distribution and evidence for mixed infections with Chlamydophila abortus. Mol Cell Probes **25**:168–173.
123. **Laroucau K**, **Souriau A**, **Rodolakis A**. 2001. Improved sensitivity of PCR for Chlamydophila using pmp genes. Vet Microbiol **82**:155–164.
124. **Mölleken K**, **Schmidt E**, **Hegemann JH**. 2010. Members of the Pmp protein family of Chlamydia pneumoniae mediate adhesion to human cells via short repetitive peptide motifs. Molecular Microbiology **78**:1004–1017.
125. **Kari L**, **Southern TR**, **Downey CJ**, **Watkins HS**, **Randall LB**, **Taylor LD**, **Sturdevant GL**, **Whitmire WM**, **Caldwell HD**. 2014. Chlamydia trachomatis polymorphic membrane protein D is a virulence factor involved in early host-cell interactions. Infection and Immunity **82**:2756–2762.
126. **Thomson NR**, **Holden MTG**, **Carder C**, **Lennard N**, **Lockey SJ**, **Marsh P**, **Skipp P**, **O'Connor CD**, **Goodhead I**, **Norbertzcak H**, **Harris B**, **Ormond D**, **Rance R**, **Quail MA**, **Parkhill J**, **Stephens RS**, **Clarke IN**. 2007. Chlamydia trachomatis:

Genome sequence analysis of lymphogranuloma venereum isolates. Genome Research **18**:161–171.

127. **Caldwell HD**, **Kromhout J**, **Schachter J**. 1981. Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. Infection and Immunity **31**:1161–1176.
128. **Binet R**, **Maurelli AT**. 2005. Frequency of spontaneous mutations that confer antibiotic resistance in Chlamydia spp. Antimicrobial Agents and Chemotherapy **49**:2865–2873.
129. **Suchland RJ**, **Rockey DD**, **Weeks SK**, **Alzhanov DT**, **Stamm WE**. 2005. Development of secondary inclusions in cells infected by Chlamydia trachomatis. Infection and Immunity **73**:3954–3962.
130. **Kari L**, **Whitmire WM**, **Olivares-Zavaleta N**, **Goheen MM**, **Taylor LD**, **Carlson JH**, **Sturdevant GL**, **Lu C**, **Bakios LE**, **Randall LB**. 2011. A live-attenuated chlamydial vaccine protects against trachoma in nonhuman primates. J Exp Med **208**:2217–2223.
131. **Burall LS**, **Rodolakis A**, **Rekiki A**, **Myers GSA**, **Bavoil PM**. 2009. Genomic Analysis of an Attenuated Chlamydia abortus Live Vaccine Strain Reveals Defects in Central Metabolism and Surface Proteins. Infection and Immunity **77**:4161–4167.
132. **Pan Q**, **Pais R**, **Ohandjo A**, **He C**, **He Q**, **Omosun Y**, **Igietseme JU**, **Eko FO**. 2015. Comparative evaluation of the protective efficacy of two formulations of a recombinant Chlamydia abortus subunit candidate vaccine in a mouse model. Vaccine **33**:1865–1872.
133. **Kelsi M Sandoz DDR**. 2010. Antibiotic resistance in Chlamydiae. Future microbiology **5**:1427.
134. **Rodolakis A**. 1983. In vitro and in vivo properties of chemically induced temperature-sensitive mutants of Chlamydia psittaci var. ovis: screening in a murine model. Infection and Immunity **42**:525–530.
135. **Banks J**, **Eddie B**, **Schachter J**, **Meyer KF**. 1970. Plaque formation by Chlamydia in L cells. Infection and Immunity **1**:259–262.
136. **Zhu H**, **He Z**, **Jia Y**. 2015. A Novel Approach to Multiple Sequence Alignment Using Multi-objective Evolutionary Algorithm Based on Decomposition. IEEE J Biomed Health Inform.
137. **Bachmann NL**, **Sullivan MJ**, **Jelocnik M**, **Myers GSA**, **Timms P**, **Polkinghorne A**. 2015. Culture-Independent Genome Sequencing of Clinical Samples Reveals an Unexpected Heterogeneity of Infections by Chlamydia pecorum. Journal of Clinical Microbiology **53**:1573–1581.
138. **Christiansen MT**, **Brown AC**, **Kundu S**, **Tutill HJ**, **Williams R**, **Brown JR**, **Holdstock J**, **Holland MJ**, **Stevenson S**, **Dave J**, **Tong CYW**, **Einer-Jensen K**, **Depledge DP**, **Breuer J**. 2014. Whole-genome enrichment and sequencing of Chlamydia trachomatis directly from clinical samples. BMC Infect Dis **14**.
139. **Gray TA**, **Krywy JA**, **Harold J**, **Palumbo MJ**, **Derbyshire KM**. 2013. Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. PLoS

Biol **11**:e1001602–e1001602.

140. **Sagan L**. 1967. On the origin of mitosing cells. Journal of Theoretical Biology **14**:225–IN6.
141. **Osaka I**, **Hefty PS**. 2014. Lipopolysaccharide-Binding Alkylpolyamine DS-96 Inhibits Chlamydia trachomatis Infection by Blocking Attachment and Entry. Antimicrobial Agents and Chemotherapy **58**:3245–3254.
142. **Grieshaber NA**, **Grieshaber SS**, **Fischer ER**, **Hackstadt T**. 2006. A small RNA inhibits translation of the histone-like protein Hc1 in Chlamydia trachomatis. Molecular Microbiology **59**:541–550.
143. **Grieshaber NA**, **Fischer ER**, **Mead DJ**, **Dooley CA**, **Hackstadt T**. 2004. Chlamydial histone–DNA interactions are disruptedby a metabolite in the methylerythritol phosphatepathway of isoprenoid biosynthesis 1–6.
144. **DeMars R**, **Weinfurter J**. 2008. Interstrain gene transfer in Chlamydia trachomatis in vitro: mechanism and significance. Journal of Bacteriology **190**:1605–1614.
145. **DeMars R**, **Weinfurter J**, **Guex E**, **Lin J**, **Potucek Y**. 2007. Lateral gene transfer in vitro in the intracellular pathogen Chlamydia trachomatis. Journal of Bacteriology **189**:991–1003.
146. **Gray TA, Krywy JA, Harold J, Palumbo MJ, and Derbyshire KM**. 2013. Distributive Conjugal Transfer in Mycobacteria Generates Progeny with Meiotic-Like Genome-Wide Mosaicism, Allowing Mapping of a Mating Identity Locus **11**:e1001602.
147. **Narra HP**, **Ochman H**. 2006. Of what use is sex to bacteria? Current Biology **16**:R705–R710.
148. **Greub G**, **Collyn F**, **Guy L**, **Roten C-A**. 2004. A genomic island present along the bacterial chromosome of the Parachlamydiaceae UWE25, an obligate amoebal endosymbiont, encodes a potentially functional F-like conjugative DNA transfer system. BMC Microbiol **4**:48–48.
149. **Jacquier N**, **Viollier PH**, **Greub G**. 2015. The role of peptidoglycan in chlamydial cell division: towards resolving the chlamydial anomaly. FEMS Microbiol Rev **39**:262–275.
150. **Liechti GW**, **Kuru E**, **Hall E**, **Kalinda A**, **Brun YV**, **VanNieuwenhze M**, **Maurelli AT**. 2014. A new metabolic cell-wall labelling method reveals peptidoglycan in Chlamydia trachomatis. Nature **506**:507–.
151. **Brown WJ**, **Rockey DD**. 2000. Identification of an antigen localized to an apparent septum within dividing chlamydiae. Infection and Immunity **68**:708–715.
152. **Errington J**. 2013. L-form bacteria, cell walls and the origins of life. Open Biology **3**:120143–120143.
153. **Liu X**, **Afrane M**, **Clemmer DE**, **Zhong G**, **Nelson DE**. 2010. Identification of Chlamydia trachomatis Outer Membrane Complex Proteins by Differential Proteomics. Journal of Bacteriology **192**:2852–2860.
154. **Chmel H**, **Bendinelli M**, **Friedman H**. 2013. Pulmonary Infections and Immunity.

155. **Reina S**, **Debbia EA**, **Schito GC**. 1993. Genetic recombination by spheroplast fusion in Escherichia coli K12. Cytobios **76**:91–95.
156. **Rastogi N**, **David HL**, **Rafidinarivo E**. 1983. Spheroplast fusion as a mode of genetic recombination in mycobacteria. J Gen Microbiol **129**:1227–1237.
157. **Ruthe HJ**, **Adler J**. 1985. Fusion of bacterial spheroplasts by electric fields. Biochim Biophys Acta **819**:105–113.