AN ABSTRACT OF THE DISSERTATION OF

Brendan Jeffrey for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on June 2, 2011.

Title: Understanding Chlamydial Biology Through Genome Sequencing and Comparative Genomics.

Abstract approved:

Daniel D. Rockey

The human pathogen *Chlamydia trachomatis* displays different phenotypes between serovars both *in vivo* and *in vitro*. The study of chlamydial biology has been hampered by the lack of a tractable genetic system. Here, chlamydial isolates with known phenotypes as well as strains generated though recombination were fully sequenced to address several aspects of chlamydial biology. The genomes of clonally isolated matched pairs of IncA-negative inclusion non-fusing strains and IncA-positive inclusion fusion strains from the same parents were very similar. In contrast, genome sequence of unmatched strains contained over 5,000 nucleotide polymorphisms. Additionally, transcriptional analysis, in vitro culture kinetics, and animal modeling show that the matched pairs are phenotypically more similar than the non-fusing strains isolated in absence of a serovar-matched wild-type strain. These results suggest that a change from an IncA positive strain to an IncA negative phenotype may involve multiple steps, including unidentified steps that lead to difference in growth rate and infectivity.

Serovar G is proportionally more common in rectal isolates from men having sex with men, whereas serovar E is less prevalent. Genome sequencing and comparative genomics of isolates from rectal or cervical sites of serovar G, E, and J revealed several open reading frames that may be associated with the rectal tropism. Further PCR-based and statistical analyses of several clinical

isolates show that ORFs CT144, CT154, and CT326 were highly associated with rectal tropism within the serovar G isolates.

Previous studies have shown that chlamydia can exchange DNA both in vitro and in vivo. We produced several chlamydial recombinant strains where the parental strains had known inclusion fusion, attachment to host cell, and secondary inclusion forming phenotypes. Genome sequencing of the parents and the recombinant strains and genome-wide association analyses were used to identify genetic loci that could be associated with observed phenotypes. Results show that the correct identification of *incA* as being associated with inclusion fusion supports the use of this approach to studying aspects of chlamydial biology. Genome-wide association studies of these recombinants also revealed several loci that may be associated with the high attachment efficiency seen in lymphogranuloma strains, such as the Pmp genes. Analysis also revealed several loci that may be associated with secondary inclusion formation.

Collectively, this work demonstrates that a combination of genome sequencing, comparative genomics, and the generation of recombinant chlamydial strains can function as a powerful tool for studying the biology of an organism that lacks classical genetic techniques.

Understanding Chlamydial Biology Through Genome Sequencing and Comparative Genomics

by

Brendan Jeffrey

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented June 2, 2011

Commencement June 2011

Doctor of Philosophy dissertation of Brendan Jeffrey presented on June 2, 2011

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.

Brendan Jeffrey, Author

ACKNOWLEDGEMENTS

I would like to sincerely thank the following individuals for various roles they have played in my graduate career: Dan Rockey for mentoring me throughout my research; Kelsi Sandoz for our many discussions on the subject of chlamydia, research or otherwise; Sara Weeks, Kelsey Quinn and Steven Eriksen for assistance with experiments; former Lab members Damir Alzhanov, Hency Chu, and Jae Dugan for their help in training me; and our collaborators Bob Suchland and Walt Stamm at the University of Washington. I would like to thank the Doctors Grieshabers for all their words of wisdom they have given me throughout the graduate school process, as well as my parents, Barbara and Berry, and my sister Simone for their support, emotionally and monetarily, throughout the last six years. And finally I would like to thank all my colleagues in the MCB program, and for Bombs Away Café for the sanity control. This work was supported by the following grants awarded to Dan Rockey: AI048769-01, AI088540-02, and AI086469-01.

CONTRIBUTION OF AUTHORS

Chapter 1: Brendan M. Jeffrey and Dan D. Rockey wrote the manuscript.

Chapter 2: Robert J. Suchland cloned the clinical isolates. Minsheng Xia and Ajay Bhatia performed the mouse experiments. Hencelyn G. Chu performed the transcriptional analysis. Brendan Jeffrey sequenced and assembled the genomes and performed the genomic analyses. Brendan M. Jeffrey, Robert J. Suchland, and Daniel D. Rockey wrote the manuscript, with Brendan M. Jeffrey and Robert J. Suchland contributing equally to this work.

Chapter 3: Robert J. Suchland cloned and expanded the clinical isolates. John R. Davidson wrote the Diffsort program. Kelsey L. Quinn and Brendan M. Jeffrey performed the PCR analyses. Brendan M. Jeffrey sequenced and assembled the genomes, and performed genomic and statistical analyses. Brendan M. Jeffrey and Daniel D. Rockey wrote the manuscript.

Chapter 4: Robert J. Suchland performed the recombination experiments, and performed the secondary inclusion phenotype experiments. Brendan M. Jeffrey and Steven G. Eriksen conducted the attachment assay experiments. Brendan M. Jeffrey sequenced and assembled the genomes, and performed all genomic and statistical analysis. Brendan M. Jeffrey, Robert J. Suchland, and Daniel D. Rockey wrote the manuscript, with Brendan M. Jeffrey and Robert J. Suchland contributing equally to this work.

TABLE OF CONTENTS

Page

Chlamydia Genetics: Progress and Challenges 2

INTRODUCTION 3

The chlamydial developmental cycle creates fundamental challenges 5

Technical limitations to transforming chlamydiae 7

Recent progress in genetic modification of chlamydiae 9

Genomics and our understanding of chlamydial recombination 10

Evidence for recombination in chlamydial 12

Recombination in clinical isolates 12

Recombinants created in the laboratory 15

CONCLUSIONS AND PERSPECTIVES 16

ACKNOWLEDGEMENTS 17

Identification of Concomitant Infection with *Chlamydia trachomatis* IncA-negative Mutant and Wild-type strains: Genomic, Transcriptional and Biological Characterization 21

ABSTRACT 22

INTRODUCTION 22

MATERIALS AND METHODS 24

Chlamydiae and chlamydial culture 24

Determination of infectious EB production 24

Antibodies and reagents for microscopy 25

Clonal isolation of independent IncA-positive and -negative strains in cell culture 25

Screening IncA-negative clinical isolates for the presence of minority IncA-positive sub-populations 25

PCR amplification and nucleotide sequence analysis 26

Mouse infections 27

Analysis of transcription in IncA-positive and -negative strains 28

Genomic DNA preparation and sequence analysis 29

RESULTS 30

Analysis of sequential same-serovar patient isolates 30

DNA sequence analysis of different matched pairs 31

TABLE OF CONTENTS (Continued)

Page

Genomic sequence analysis 32
Transcriptional analysis of IncA-positive and -negative strains 33
In vitro growth 34
Mouse infections 35
DISCUSSION 35
ACKNOWLEDGEMENTS 40
Genome Sequencing of Recent Clinical *Chlamydia trachomatis* Strains Identifies Loci Associated with Tissue Tropism and Regions of Apparent Recombination 50
ABSTRACT 51
INTRODUCTION 51
MATERIALS AND METHODS 53
Chlamydia strains and genomic DNA preparation 53
Genome Sequencing and Sequence Analysis 53
Genome-wide recombination analysis 54
Clinical isolate DNA preparation and PCR-based genotyping 55
Nucleotide sequence accession numbers 56
RESULTS 56
Comparative genome analysis of sequenced chlamydial isolates 56
Mapping of SNPs and larger changes in sequenced genomes of clinical isolates 57
Evidence for recombination in genomes of sequenced clinical isolates 58
Correlation of identified SNPs with tropism to the rectal or cervical site of infection 59
DISCUSSION 61
ACKNOWLEDGMENTS 65
Genome sequence analysis of recombinants between clinical *Chlamydia trachomatis* strains identifies genetic regions associated with known phenotypes 75
ABSTRACT 76
INTRODUCTION 77
MATERIALS AND METHODS 78

TABLE OF CONTENTS (Continued)

Page

Chlamydial strains and selection for resistance 78
Generation of recombinant chlamydial strains 79
Recombination rate determination 79
Chlamydia recombinant strain genomic DNA preparation 80
Genome sequencing and sequence analysis 80
Recombinant strain phenotype analysis 81
Genome-wide phenotype association analysis 82
Polymorphic membrane protein family charge analysis 82
Genome sequence accession numbers 83
RESULTS 83
Generation of recombinant strains 83
Genome sequence analysis of recombinant strains 84
A method to correlate genotype with phenotype in clinical C. trachomatis strains 87
Exploration of genotypes associated with attachment and secondary inclusion phenotypes 87
DISCUSSION 91
ACKNOWLEDGMENTS 96
General conclusion 114
BIBLIOGRAPHY 117

LIST OF FIGURES

Figure Page

1.1. Integration of the tetracycline resistance gene into chlamydiae. 18

2.1. Fluorescent microscopic analysis of McCoy cells infected with different IncA-positive and -negative matched pairs. 41

2.2. Orf analysis of regions showing variation found in the matched pair isolates. 42

2.3. Transcription of incA in prototype and clinical C. trachomatis strains. 43

2.4. Production of infectious chlamydia during growth in murine epithelial cells in vitro..... 44

2.5. Chlamydial shedding from mice infected vaginally with closely related and unrelated same-serovar strains of IncA-positive isolates compared to corresponding IncA- negative variants. 45

3.1. Phylogenetic analysis of sequenced chlamydial isolates based on whole genome alignment. 66

3.2. Number of substitutions per open reading frame between strains. 67

3.3. Location of insertions or deletions between strains. 68

3.4. Schematic diagram of open reading frames CT154 and CT326 in several sequenced C. trachomatis strains. 69

3.5. BLAST-based evidence for recombination in clinical C. trachomatis genome sequences. 70

3.6. PCR- and RFLP-based genotype analysis of variable regions in clinical isolates. 73

4.1. Recombinant geneology. 97

4.2. Genome maps of recombinant strains. 98

4.3. Schematic diagram of the CT740 to CT749 regions in selected recombinant sequences. 99

4.4. Genomic locations of crossover regions in the ten sequenced recombinants. 100

4.5. Attachment efficiency results of parental and recombinant strains. 101

LIST OF FIGURES (Continued)

Figure Page

4.6. Genome-wide association analysis of the non-fusing inclusion phenotype. 102

4.7. Genome-wide association analysis of the attachment phenotype. 103

4.8. Protein charge analysis of the polymorphic membrane proteins. 104

4.9. Fluorescent microscopic analysis of the secondary inclusion formation phenotype of recombinant strain RC-J/953. 105

4.10. Genome-wide association analysis of secondary inclusion formation phenotype. 106

LIST OF TABLES

Table Page

1.1. Antibiotics that have been used in generating stable or transient resistance phenotypes in *Chlamydia spp.* 19

1.2. Antibiotic resistance markers that cannot be used in selection for *Chlamydia* transformants. 20

2.1. Oligonucleotide primers used in these studies. 46

2.2. Culture points in single individuals persistently infected with *C. trachomatis* serovar Ia. 47

2.3. Summary of nucleotide differences among the sequenced chlamydial strains. 48

2.4. Characterization of genomic changes in the serovar J matched pair. 49

3.1. Pairwise analysis of the number of substitutions and insertion/deletion events identified in genome sequences from a collection of *C. trachomatis* isolates. 74

4.1. Phenotypes of parents and progeny in recombinant crosses. 107

4.2. Recombination rates. 108

4.3. Regions of shared crossovers between recombinant strains. 109

4.4. Polymorphic membrane protein charge analysis. 110

S4.1. Genes associated with attachment phenotype. 111

S4.2. Genes associated with secondary phenotype. 113

Understanding Chlamydial Biology Through Genome Sequencing and Comparative Genomics

Chlamydia Genetics: Progress and Challenges

Brendan M. Jeffrey, Anthony T. Maurelli, and Daniel D. Rockey

Chlamydia Biology and Pathogenesis

Edited by Patrick Bavoil and Ming Tan

ASM Press, in progress

INTRODUCTION

Chlamydiae are successful obligate intracellular pathogens that cause disease in animals from virtually every phylogenetic level. *Chlamydia trachomatis* and *Chlamydia pneumoniae* are the primary species causing disease in humans. *C. trachomatis* is the most common sexually transmitted bacterial species, and a causative agent in pelvic inflammatory disease, ectopic pregnancy and sterility in women (108). A specific serovariant group of *C. trachomatis* strains are responsible for causing trachoma, a disease that is the leading cause of preventable blindness worldwide (78). The major mechanisms involved in many aspects of chlamydial biology and pathogeneses are unclear at present. However, it is clear that chlamydial disease is a function of many processes, including attachment and invasion of host target cells (135), remodeling of the host cell endocytic process (102), the ability to replicate within the remodeled host environment (103), dissemination and persistence, all balanced against the hosts immunological response to infection. The ability to genetically manipulate various microorganisms has furthered our understanding of many aspects of prokaryote biological systems. This includes understanding of disease-causing factors involved in host microbe ecology and disease. The lack of a tractable genetic system in the Chlamydiales has greatly hindered our ability to study many aspects of the biology of this clinically important organism.

All chlamydiae encode the surface exposed protein OmpA (or MOMP), which is the major determinant in the serological classification within the genus. *C. trachomatis* contains three distinct serospecific groups, those that are associated with disseminating lymphogranuloma (L1 to L3), those associated with trachoma (A, B, and C), and those associated with non-invasive urogenital disease (D through K) (51). While new classification schemes have arisen in this field (69), serotypic characterization remains a useful method to discriminate amongst chlamydial species biotypes.

The chlamydiae undergo a biphasic developmental cycle that is unique to the prokaryotes. The metabolically inactive, infectious elementary body (EB)

attaches to a target host cell and is internalized into endocytic vesicles (30). These vesicles, termed inclusions, are modified by the chlamydia, leading to escape from the lysogenic pathway and creating a suitable environment for development and growth (51, 53). Within the inclusion, the infectious, metabolically inert EB form will differentiate into the non-infectious metabolically active reticulate body (RB). Replication of the *C. trachomatis* RBs occurs for approximately 18-24 hours, at which point the RB form de-differentiating into the infectious EB form can be observed. Depending on the chlamydial species, host cell type, and environmental conditions, the chlamydial developmental cycle is most often complete by 48 to 72 hours post infection, where the infectious EBs are then released from the infected host cell through a poorly understood mechanism (59).

Molecular genetic tools have become some of the most powerful means for analysis in any system, be it viral/prokaryotic/eukaryotic, big/small, or intracellular/extracellular. The chlamydiae are one of only a few bacterial pathogens for which there is no tractable genetic system (others include members of the genera *Lawsonia*, *Orientia*, *Tropheryma*, and the spirochete *Treponema pallidum*), and are arguably the most significant pathogens for which this technology remains unavailable. A consistent theme for these genetically intractable organisms is that they are often obligate intracellular organisms that cannot be propagated axenically, or are organisms that cannot be cultured outside of host organisms in any way. The absence of a useful genetic system has severely limited the ability for definitive study of gene and protein function in each of these genera or species.

The absence of a workable chlamydial genetic system has not been for lack of effort- a wide spectrum of individuals has tried to work on this problem. Experts in gram-negative plasmid biology expected that broad-host-range plasmids could be manipulated and transformed into recipient chlamydiae, followed by some form of selection that would allow isolation and propagation of plasmid-positive strains. This approach has been applied with success to genetic

modification of *Coxiella burnetii* (12, 129), an obligate intracellular species that has recently been cultured axenically (86, 87). This approach led to reports of transient maintenance of exogenous genes introduced into chlamydiae, using vectors with constitutive chlamydial promoters driving expression of chloramphenicol acetyltranserase, with a goal of generating chloramphenicol resistance in transformants (17, 123). Researchers also focused on the single chlamydial plasmid, a near ubiquitous, conserved element whose role in infection is presently being elucidated (27, 83, 85, 97). The presence of this plasmid formed the basis for several attempts to develop a chlamydial genetic system, as it logically might be a tool for manipulation and reintroduction into the organism. Chlamydiae also have bacteriophage (57, 64, 91, 96, 111), and experiments have been designed to exploit these phage as vectors for introduction of genes. Promiscuous transposons have been successfully used in many challenging bacterial systems (12, 110), and there is precedent for natural introduction of a transposon into a veterinary chlamydial pathogen (36, 37), described below). It is reasonable to expect that these elements should be useful in modification of the chlamydial genome. Each of the respected scientists involved in these projects has brought a different set of skills and approaches to the laboratory, yet the collective number of genes introduced into this system remains at zero.

So why is this the case? Why have all these successful research scientists been unable to introduce genes into this group of bacteria? The following sections will address a set of challenging aspects of the chlamydial system, and point out some avenues that can be exploited, perhaps opening the door to technologies for routine genetic modification of chlamydiae.

The chlamydial developmental cycle creates fundamental challenges

The obligatory intracellular lifestyle and unique biphasic developmental cycle of the chlamydiae contribute to the difficulty in the ability to genetically manipulate these organisms. It is likely that several aspects of this lifestyle have a role in these complications. The rigid cross-linked protein outer membrane of the metabolically inactive elementary bodies (EBs) may hinder the introduction of

exogenous DNA into this developmental form. Furthermore, the lack of metabolic activity at this stage may preclude interactions with the recipient genome. The challenges may be parallel to what is seen in the *Bacillus* and *Clostridium* spp. While the vegetative forms of these bacteria are readily made competent, the spore structures are refractory to acquiring or incorporating introduced DNA (Peter Setlow, personal communication). EBs are similarly resistant to physical stress (though not similarly resistant to the stresses of heat, desiccation, or time) and this environmental stability may lead to analogous resistance to DNA uptake. It is not likely this is a barrier however, as one of our laboratories was able to transform EBs in the only example of stable incorporation of an exogenously added resistance allele into the chromosome of *C. psittaci* (17). It is likely that other investigators also successfully electroporated DNA into EBs (84, 123), but there was no evidence for stable maintenance of any of the plasmid species that were introduced.

There might be several levels to this challenge of modifying a phenotype via introduction of DNA into EBs. The cell wall may be physically impermeable to DNA or the pores introduced by electroporation might be of a size that does not allow DNA molecules to efficiently enter. The recent report by Binet and Maurelli do not support this, as their electroporation experiments suggest that linear or circular DNA of at least 1,052 base pairs can be electroporated into EBs (123). Additionally, DNA in both EBs and in spores is bound by small protein molecules- in the chlamydiae this involves chlamydial histone-like proteins (10, 49, 50, 89) while in spore-forming bacteria, a set of Small Acid Soluble Proteins are tightly bound to genomic DNA (104). While these DNA-binding proteins likely physically wall the genome away from introduced donor DNA, they also place the DNA in a metabolically inert state, making it inaccessible for recombination machinery and homologous recombination. It is also possible that DNA not bound by histone-like proteins is not favorably recognized by replication or transcriptional machinery in a developing EB, perhaps blocking exogenous DNA from being functional within chlamydial developmental forms.

After infection and uptake, EBs differentiate into the metabolically active reticulate bodies (RB), an orchestrated process that turns the potentially hostile endosome into a chlamydia-friendly bacteria factory termed the inclusion (see chapters 6,7 and 9). Targeting exogenous DNA to the RB is an attractive possibility for gene introduction into this system. RBs have cell envelopes that lack the structural rigidity of EBs, but otherwise are generally similar to typical Gram-negative bacteria. RBs divide by binary fission, are metabolically active, and undergo classical DNA replication. There is also considerable recent evidence that RB genomes undergo recombination (see below). However, there are unique challenges when considering RBs as targets for DNA exchange. Functional RBs are present only within the inclusion and no technology exists that will allow isolated RBs to re-infect host cells. Therefore, a strategy of transforming purified RBs and then using them to infect cells- similarly to what has been accomplished with *Coxiella burnetii* (129)- is not currently an option. It is possible that DNA can be introduced to RBs within cells, but there are four lipid bilayers (the host plasma membrane, the inclusion membrane, the RB outer membrane, and the RB inner membrane) through which exogenous DNA will need to cross to enter the chlamydial cytoplasm. Perhaps a multiple-pulse electroporation strategy might be exploited to overcome this challenge. Intracellular RBs might be targeted through the use of a genetically tractable second intracellular DNA donor (such as *Shigella*, *Salmonella* or *Legionella*), but these candidate donor bacteria will likely be separated from chlamydiae by at least the inclusion membrane, and most intracellular bacteria do not cohabit the same vacuole within host cells (53, 100). Therefore, introduction of DNA into RBs will require the development of technologies to access the lumen of the inclusion, a concept that appears increasingly possible as we continue to understand vesicular trafficking within infected host cells (see chapter 8).

Technical limitations to transforming chlamydiae

The developmental cycle presents a unique set of challenges when thinking about transforming these organisms, but there are other considerations that also need to be addressed. The first of these is simply the challenge of manipulating

chlamydiae. The obligately intracellular nature of these bacteria, their capacity for persistence, and the stability of EBs can lead to significant misunderstandings when an individual thinks they have a novel resistance phenotype, or thinks that an antibiotic can be removed to see if their transformation worked (120). These issues can lead to false positives when examining a population of infected cells for acquisition of a genetic marker. Technical challenges also include the need to generate microbiological clones of a particular strain. Researchers have recently addressed the important concept that many chlamydial strains are polyclonal, and that isolation of microbiological clones is critical to evaluation of the biology of individual strains. Different laboratories use either limiting dilution (121) or plaque purification (43) of individual chlamydial inclusions (i.e. “colonies”) when they purify a phenotypically variant strain. It is conceivable that the inability to identify individual transformed chlamydial subpopulations might have led to false negatives in otherwise successful transformed chlamydial experiments. Researchers should consider the microbiological cloning of chlamydiae as a critical technique to develop prior to initiating any attempts at developing a chlamydial system.

An additional technical barrier may involve the interactions between antibiotic resistance mechanisms and the basic nature of the chlamydial developmental cycle. We have discussed the chlamydial histone-like proteins, and their ability to completely shut down transcription and translation during transition of RBs to EBs (10, 11). The majority of chlamydial genes are not transcribed during the first stages of successful transition from EB to RB and the remodeling of the endosomal vacuole (2, 14, 105). It is likely that all but the very earliest events in development require bacterial transcription and translation, leading to a stepwise orchestration of chlamydial gene and protein expression. Because only the metabolically dormant EBs are infectious, this genetic reactivation has to happen following the initial inoculation and in each subsequently infected cell. Chlamydiae are extraordinarily sensitive to antibiotics and exhibit minimum inhibitory concentrations that are much lower than those for many other bacteria (reviewed in (99)). It is likely that selection for antibiotic resistance to certain

antimicrobials has not been successful because the candidate antimicrobial resistance factor is not functional within an EB, and the antibiotic in the medium blocks activation of the resistance phenotype before it can function within the bacteria. The best examples of this are chloramphenicol and beta-lactam resistance, for which active enzymes will need to be present within the EB as it transitions to an RB. The phenotypic lag in each developmental cycle might be significantly limiting to antibiotic choice in the chlamydial system. It is suggested that the list of antibiotic resistance mechanisms known to be effective in chlamydiae be carefully evaluated before any vector design is initiated (Table 1.1). Additionally, there are antibiotics that, for a variety of reasons, should not be used in any vector design (Table 1.2). Such reasons include antibiotics used clinically in treatment of chlamydial infection, or antibiotics that do not enter mammalian cells, such as most aminoglycosides. This list of restricted antibiotics should be taken seriously- antibiotics that are clinically important *cannot be used* to develop vectors that encode resistance to them. The deliberate transfer of a drug resistance trait to microorganisms that are not known to acquire the trait naturally is prohibited by the NIH "Guidelines for Research Involving Recombinant DNA Molecules" (Available through http://oba.od.nih.gov/rdna/nih_guidelines_oba.html), if such acquisition could compromise the use of the drug to control these disease agents in humans or veterinary medicine.

Note that the *tet*(C) allele (described below) is an example of an exogenous resistance mechanism that is known to function in chlamydiae, but approval for use of this marker must be formally obtained from the NIH on a laboratory-by-laboratory basis. All other markers used in successful selection for resistance in chlamydiae have involved a mutated metabolic target gene for the antibiotic being used (17, 33, 34, 120). If antibiotic selection proves to remain challenging, it is possible that strategies using non-antibiotic-based selection strategies may be employable (95, 128).

Recent progress in genetic modification of chlamydiae

Even though there have been challenges to introduction of DNA into chlamydiae, there are recent examples of progress in modifying the chlamydial genome. Kari and colleagues have developed a mutagenesis and selection strategy that, while arduous, does allow the creation of otherwise isogenic mutants with changes at chlamydial loci of interest (74). This technique involves the use of heteroduplex identification in pools of chemically mutated EBs, followed by the use of an enzyme (Cel1) that targets and cleaves DNA at heteroduplexes. Specific cleavage products are identified using electrophoresis, and positive mutant strains are subsequently cloned via isolation of chlamydiae from plaque-based culture. The final product is then genome sequenced to verify that additional sites are not mutated. At the end of the procedure, a strain is identified that carries a mutation in a gene of interest, in a genomic background that is otherwise isogenic with the parent. The original manuscript provides genetic evidence that the *trp*AB operon is necessary for growth on media lacking tryptophan, a premise previously established by this research group using other techniques (23, 39). The Kari approach can theoretically be used for any non-essential chlamydial gene, and a strategy for the use of temperature-sensitive mutations for analysis of essential genes is proposed by the authors. This protocol will be a valuable tool for examining host microbe interactions following creation of relevant knockouts at loci of interest to researchers. While it has limits, this is truly the first example of a workable genetic system for the chlamydiae.

Genomics and our understanding of chlamydial recombination

In a system that currently lacks the ability for researchers to genetically manipulate chlamydia, genome sequencing efforts have proved highly effective in increasing our understanding of the biology of this organism. With the release of the first genome sequence of the urogenital *C. trachomatis* strain D/UW3-cx in 1998 (109), many questions about basic chlamydial biology could be addressed. The availability of this and many subsequent genome sequences allowed researchers to search for hints that chlamydiae might be amenable to DNA uptake. Genome sequences of contemporary chlamydial species are remarkably

syntenous (i.e. they share gene order throughout the chromosome), with very limited examples of recently acquired genomic islands. While ancestral chlamydial cousins encode type IV secretion assemblies on a genomic island (48, 55), there is no evidence for anything resembling conjugation machinery in the chlamydiae. This difference suggests that chlamydiae evolved into its genetically intractable niche by reductive evolution from genetically amenable ancestors (55). A single IS element is found in isolates of *C. suis* (36), but no human pathogenic chlamydiae encode transposons, nor are any recognizable recombination targets or Chi sites present in any chlamydial genome.

Examination of the first chlamydia genome identified a collection of genes that appear to share more identity with genes from plants than to other known bacterial genes (109), suggesting possible acquisition via exchange between a chlamydial ancestor and a host cell. However, subsequent analysis showed that these genes are of chloroplast origin, and that this association may be a reflection of the ancient relationship between the Chlamydiaceae and the chloroplast-cyanobacterium lineage rather than a more recent acquisition of plant-like genes by the Chlamydiceae (19). Recent work showing that the product of CT390, encoding a DAP aminotransferase, supports the model of a common ancestor for chlamydia and bacteria of the cyanobacterium-chloroplast lineage (79).

Chlamydiae encode a complete type III secretion (T3SS) machine, which has been acquired by many bacteria via horizontal acquisition of plasmids or other large genomic islands. In contrast, the chlamydial T3SS assembly is present on different regions of the chromosome, with no evidence of acquisition as a large cassette. There are also no G+C ratio differences between T3SS genes and the chlamydial chromosome, which is another common trait in T3SS genes and operons (58, 67, 114). These data suggest that chlamydiae did not acquire their T3SS genes via horizontal transfer from another organism. Collectively, genome sequence data paints a picture of a system that does not appear to sample DNA from other organisms.

Evidence for recombination in chlamydial

While the above discussion stresses that chlamydiae are generally not receptive to acquisition of foreign genes, there is strong evidence for active recombination systems in these organisms. The following sections will address experimental evidence that chlamydiae do have the tools to regularly integrate homologous DNA into their genome.

Recombination in clinical isolates

Studies of genetic diversity within *ompA* provided the first evidence of recombination and horizontal gene transfer in *C. trachomatis* (20, 73, 80). Recombination in *ompA* is associated with antigenic changes on the surface of the major outer membrane protein (MOMP), a serovariant and abundant surface protein on EBs and RBs. These studies were expanded to other highly variable regions of the genomes such as the plasticity zone (PZ) and the polymorphic protein (Pmp) family of proteins (45-47). Gomes *et al* provide compelling evidence that recombination occurs at various hot spots around the chlamydial genomes including the region around CT049, the PZ, the region surrounding *ompA,* and regions of genome containing the various *pmp* genes (45-47). The sequencing of clinical isolates by our group (62) revealed evidence for recombination throughout the genome, in addition to the sites described by others. An attractive model for these observations is that since patients can be infected with more than one strain of chlamydia (116), different strains may come in contact with one another (52, 117) and then undergo genome shuffling through a currently unknown mechanism. Hot spots identified in previous work (46) include genetic regions encoding many proteins that are suspected to reside in the bacterial outer membrane, and lead to the hypothesis that recombination observed may be influenced by selection of recombinants associated with immune evasion strategies.

Although it is clear that horizontal gene transfer and recombination occurs between chlamydial strains, until recently there was no evidence for chlamydia incorporating foreign DNA into its genome. The genetic characterization of

tetracycline resistance in a chlamydial pathogen of swine, *Chlamydia suis,* revealed a recent lateral gene transfer event from a non-chlamydial species (37). Resistant *C. suis* strains were initially identified in the upper Midwest of the USA (4), and similar resistant strains were recently identified in Italy (35). Tet-resistant strains carry related integrated exogenous genomic islands [Tet(C) islands] of between 5 and 12 kilobases. Each Tet(C) island encodes a tetracycline resistance gene [*tet*(C)] and a set of genes associated with plasmid maintenance and replication (Fig. 1.1) (37). Most, but not all, of the Tet(C) islands also carry a unique insertion sequence named IScs605. This bicistronic insertion sequence is structurally similar to the IS605 family of insertion elements commonly found in *Helicobacter* spp., (37). Functional analysis of IScs605 in an *E. coli* system revealed that one of the ORFs (the IS200 homolog) is sufficient to mediate transposition of the element, and that integration is targeted nearly universally to the pentameric sequence TTCAA (36). This pentamer is present at each probable integration site within the Tet(C) island and the target chlamydial chromosome.

The origins of the Tet(C) island remain unclear, but there are compelling data that recombination with some other organism was a significant player in the transfer. With the exception of the IS element, sequence analysis of the Tet(C) islands revealed a high degree of identity with plasmid pRAS3.2, carried by the fish pathogen *Aeromonas salmonicida* (37, 72). The very few differences between the shared sequence of pRAS3.2 and the Tet(C) islands include nucleotides in possible control elements for expression of *tet*(C), and within the apparent plasmid origin of replication (37). Insertion of these islands into the *C. suis* genome occurred at an identical position in the *inv*-like gene in all of the Tet resistant *C. suis* strains analyzed (37).

While the Tet(C) islands are largely identical to sequences from pRAS3.2, this plasmid does not carry the IScs605 insertion sequence. Recent sequence analysis from another fish pathogen, *Laribacter hongkongensis,* demonstrated that this organism carries a genomic island that shares very high identity with

IScs605 and *tet*R in the *C. suis* Tet(C) islands. This discovery suggested that IS elements found in *L. hongkongensis* contributed genes to the *A. salmonicida* pRAS3.2 plasmid and then facilitated a possible integration event into the *C. suis* genome into the inv-like gene, either directly or following passage in other unknown organisms (Fig. 1.1A).

The acquisition of these Tet(C) islands remains the only example of horizontal acquisition of foreign DNA by a contemporary chlamydial strain. How these *C. suis* strains acquired the various forms of the tetracycline resistance islands discussed by Dugan *et al* remains unclear, but a model can be proposed. The long-term prophylactic use of tetracycline in swine facilities would provide an environment where selective pressure for *C. suis* strains harboring the Tet islands was strong. Conversations with Art Andersen of the National Animal Disease Center in Ames, Iowa also indicated that pig farms in the upper Midwest were commonly fed waste from fish processing plants, perhaps providing a means for contact between the chlamydiae and organisms harboring the resistance plasmids. The resistant strains identified in Italy share *tet*(C) with the strains from the USA, but there are not enough data to assess whether these are similar or distinct integration events.

Therefore, chlamydiae are capable of acquiring DNA from the appropriate donor bacteria. This appears to be a rare event, likely driven by the long-term use of low levels of antibiotics as growth promoters. There is no evidence that a similar natural process has led to genomic island integration in chlamydiae pathogenic to humans. It is possible that the tools used by *C. suis* to integrate this island may be used by investigators to introduce genes of interest into chlamydiae, although efforts to use this have not been successful to date (DDR, unpublished).

Outside of the acquisition of tetracycline resistance by *C. suis*, the chlamydiae appear to be driving in the slow lane of evolutionary progress, working through evolution via single base changes or insertion/deletions, incorporating and exploiting several species-wide gene conversion events and

deleting a few genes –with significant consequences (23, 47, 82). If this interpretation is correct, it again does not bode well for studies in chlamydial genetics, because the organism has not retained an evolutionary strategy that involves routine acquisition and sampling of foreign DNA.

Recombinants created in the laboratory

Pioneering work by Demars and colleagues demonstrated the remarkable fact that *C. trachomatis* genomes do recombine and recombine regularly and extensively within the species. These investigators co-infected HeLa cells with different combinations of ofloxacin-, lincomycin-, rifampin-, and trimethoprim-resistant serovar L1 *C. trachomatis* strains, and selected for doubly resistant mutant strains. The choice of antibiotics was driven by the ability of bacteria to evolve antibiotic resistance through single amino acid changes in target proteins. These investigators demonstrated doubly resistant strains (i.e. progeny of recombination) occurring at frequencies that were significantly higher than the rate of mutation to resistance. A second manuscript by these investigators demonstrated that surprisingly large regions of the genome are transferred among strains.

More recently, one of our laboratories used the *C. suis* tetracycline resistant strains discussed above and successfully created a *C. trachomatis* L2 recombinant strain carrying the tetracycline resistance island (120). Sequence analysis of this recombinant strain revealed a probable homologous recombination event had occurred, where the donor DNA had been inserted into one of the *C. trachomatis* ribosomal operons, resulting in a recombinant strain with three ribosomal operons (120), showing the first example of a large segment of foreign DNA being inserted into a *C. trachomatis* genome in vitro. These studies provide much needed evidence that the chlamydiae can take up foreign DNA from its environment, through an unknown mechanism, and this DNA can be incorporated into the genome. Recombination has been documented both within *C. trachomatis*, and also between *C. trachomatis*, *C. suis* and *C. muridarum*, in each case via a mechanism involving homologous recombination.

There was evidence for recombination between strains that were both IncA-positive and IncA-negative, indicating that fusion of inclusions or cohabitation of a common inclusion is not required for recombination (120). It is noteworthy that we have no evidence for activity of the *C. suis* IScs605 transposase in any chlamydial recombinant- no target sequence has been identified in any of our over 100 in vitro-generated recombination sites (BMJ, DDR, unpublished).

These studies demonstrate a surprising concept: chlamydiae readily exchange DNA between strains and incorporate this DNA into its genome via homologous recombination. This was surprising because of the very limited evidence for acquisition of DNA from non-chlamydial sources. The knowledge that the chlamydiae contain genes involved in genetic recombination (54, 56) provides additional evidence that these organisms can undergo this process. DNA sequencing of several genes known to be variable between *C. trachomatis* strains, in addition to recent whole genome sequencing of clinical isolates have provided evidence that these organisms do undergo genetic exchange. These data suggest a similar integration strategy as described by Binet and Maurelli (17), in that exogenous DNA is recombined into homologous sequences within the genome.

CONCLUSIONS AND PERSPECTIVES

While each of these findings are interesting and have utility, they still do not functionally address the challenge of routine introduction of novel genes into the chlamydial system. The work of Binet and Maurelli (17), along with work of Wyrick and colleagues (123) demonstrate that chlamydial EBs can acquire DNA and this DNA can be integrated into the chromosome. However, the rates are quite low and there is still no evidence for exogenous maintenance of an introduced plasmid. There is evidence of natural insertion sequence-mediated delivery of DNA to a chlamydial genome, but exploiting this in a practical sense remains a challenge. There is also compelling and rather surprising evidence for abundant and frequent homologous recombination between strains, with no current understanding of how DNA might have actually been shared between

developmental forms. Genomics has been key to developing an understanding of the processes involved and will continue to have a significant role. But genomics cannot be used to wish a gene into a recipient chlamydiae, and it is likely that continued exploration of mechanisms used by strains to share DNA, as well as building on technologies in other systems, will someday allow industrious and clever researchers to develop a practical chlamydial genetic system.

ACKNOWLEDGEMENTS

The Rockey laboratory is supported by PHS awards AI069214 and AI086469. The Maurelli laboratory is supported by R01 AI044033-08 and U19 AI08044-02

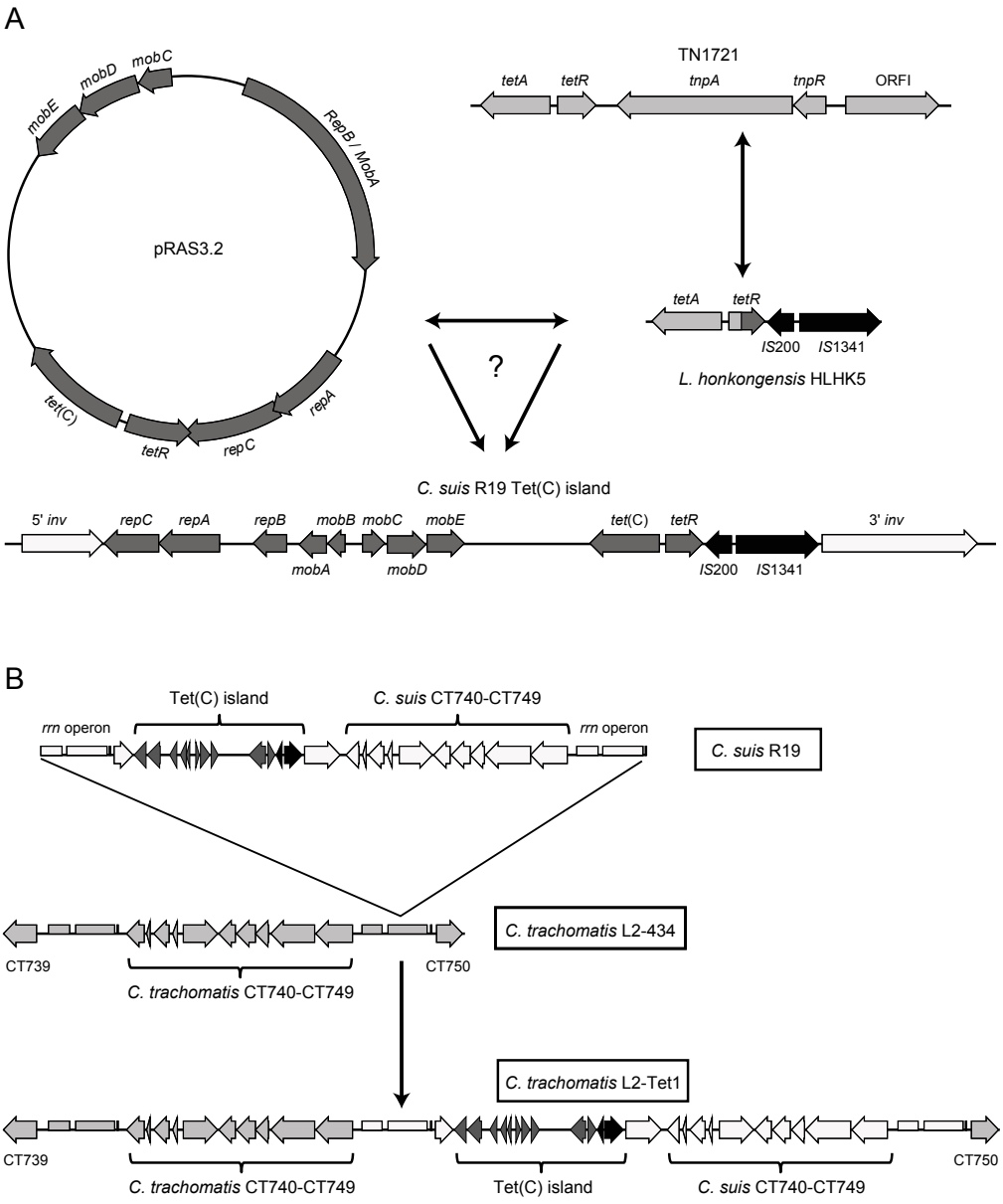


Figure 1.1. Integration of the tetracycline resistance gene into chlamydiae. Color coding in each panel represent identical sequences from each source. Panel A: The Tet(C) island found in a variety of *C. suis* strains is a mosaic of plasmid and insertion sequences from a variety of different possible sources. Note that the *IS*200/*S*1341 sequences in *L. hongkongensis* HLHK5 and in the *C. suis* Tet(C) island are identical and constitute IScs605 in *C. suis*. See text for details. Panel B. The presence of this naturally occurring island in *C. suis* was then exploited as a tool for laboratory-based recombination experiments similar to those conducted by Demars and colleagues (16). Ribosomal RNA operons in panel B are indicated by half-width ORF boxes.

Table 1.1. Antibiotics that have been used in generating stable or transient resistance phenotypes in *Chlamydia spp.*

Compound	Process targeted	Endogenous resistance gene	Reference for use in *Chlamydia spp.*
Chloramphenicol	Protein synthesis	none	(123)
Kasugamycin	Protein synthesis	16S rRNA	(17)
Nalidixic Acid	DNA supercoiling	*gyr*A	
Ofloxacin	DNA supercoiling	*gyr*A	(34, 120)
Rifampin	Transcription	*rpo*B	(34, 120)
Spectinomycin	Protein synthesis	16S rRNA	(17)
Tetracycline	Protein synthesis	*tet*(C)	(37, 75, 120)

Table 1.2. Antibiotic resistance markers that cannot be used in selection for *Chlamydia* transformants.

Antibiotics of clinical importance (1)	**Antibiotics that do not penetrate into mammalian cells**	**Antibiotics that are toxic to eukaryotic cells at concentrations needed to kill *Chlamydia spp.****
Amoxicillin	Kanamycin	Blasticidin S
Doxycycline	Gentamycin	Hygromycin B
Ofloxacin/levofloxacin	Streptomycin	Geneticin (G 418)
Azithromycin		Zeocin
Erythromycin		Phleomycin
Sulfisoxazole		Noursethricin

Identification of Concomitant Infection with *Chlamydia trachomatis* IncA-negative Mutant and Wild-type strains: Genomic, Transcriptional and Biological Characterization

Robert J. Suchland, Brendan M. Jeffrey, Minsheng Xia, Ajay Bhatia,

Hencelyn G. Chu, Daniel D. Rockey, and Walter E. Stamm

Infection and Immunity, American Society for Microbiology

1752 N Street N.W.

Washington DC, 20036

Volume 76, No. 12, pages 5438-5446

ABSTRACT

Clinical isolates of *Chlamydia trachomatis* that lack IncA on their inclusion membrane form nonfusogenic inclusions and have been associated with milder, subclinical infections in patients. The molecular events associated with the generation of IncA-negative strains, and their role in chlamydial sexually-transmitted infections are not clear. We explored the biology of the IncA-negative strains by analyzing their genomic structure, transcription, and growth characteristics in vitro and in vivo in comparison with IncA-positive *C. trachomatis* strains. Three clinical samples were identified that contained a mixture of IncA-positive and -negative same-serovar *C. trachomatis* populations, and two more such pairs were found in serial isolates from persistently infected individuals. Genomic sequence analysis of individual strains from each of two serovar-matched pairs showed that these pairs were very similar genetically. In contrast, the genome sequence of an unmatched IncA-negative strain contained over 5,000 nucleotide polymorphisms relative to the genome sequence of a serovar-matched but otherwise unlinked strain. Transcriptional analysis, in vitro culture kinetics and animal modeling demonstrated that IncA-negative strains isolated in the presence of a serovar-matched wild-type strain are phenotypically more similar to the wild-type than are IncA-negative strains isolated in the absence of a serovar-matched wild-type strain. These studies support a model suggesting that a change from an IncA-positive strain to the previously described IncA-negative phenotype may involve multiple steps, the first of which involves a translational inactivation of *incA*, associated with subsequent unidentified steps that lead to the observed decrease in transcript level, differences in growth rate, and differences in mouse infectivity.

INTRODUCTION

Chlamydia trachomatis is the most common bacterial STD in the United States and the leading cause of preventable blindness worldwide (78). Techniques for

its genetic analysis are severely limited, restricting understanding of the molecular biology of this widespread pathogen. We previously described *C. trachomatis* isolates that form multiple nonfusogenic inclusions within single cells infected with multiple elementary bodies (EB). Inclusions formed by these isolates uniformly lacked the protein IncA in the inclusion membrane (118). These IncA-negative variant strains were found in approximately 2% of the Chlamydial genital isolates in patients from Seattle area STD clinics. IncA-negative representatives of all routinely isolated serovars, except serovar G, were cultured from patients in these clinics (94). Nucleotide sequence analysis of the IncA-negative strains suggested that inactivating mutations occurred randomly across the gene. The genesis and persistence in vivo of these variant strains remain uncharacterized.

Because of the diversity and relative abundance of these strains in patient populations, we hypothesize that IncA-negative strains arise via mutation from wild-type strains, and that variant strains have a unique niche in the patient population. If these hypotheses are true, we would predict that individual patients can be identified that carry both variant and wild-type strains, and that in some cases the mixed populations should be nearly isogenic, with the exception of an inactivating mutation in *incA*. To test these hypotheses, we screened 55 low-passage IncA-negative isolates with anti-IncA monoclonal antibody (mAb) to identify possible minority IncA-positive populations. In 3 strains, serovar-matched minority populations were identified that expressed and localized IncA to the inclusion membrane. These “matched pairs” were separated from one another through limiting dilution cloning and compared with IncA-positive and -negative same-serovar strains isolated from independent patients. Analyses of genome sequence, *incA* transcription, and growth characteristics in vitro and in vivo demonstrated that the matched pair strains are genetically nearly isogenic and that the tested phenotypes of the IncA-positive and IncA-negative strains within a matched pair are more closely related than similar same-serovar pairs isolated from individual patients. The studies suggest that transition from an IncA-positive to an IncA-negative strain may represent an early stage in the

phenotypic variation of *C. trachomatis* within patients, and that other, as yet uncharacterized, changes may occur which further distinguish IncA-positive from IncA-negative strains.

MATERIALS AND METHODS

Chlamydiae and chlamydial culture

Prototype strains D/UW-3/Cx and J/UW-36/Cx were used for reference sequence and PCR analysis. Clinical strains J/6276, J(s)/6276, J(s)/893, J(s)/1980, J/p225, Ia/165, Ia/p202, Ia/9878, Ia/1008, Ia(s)/1010, Ia/1011, Ia(s)/1025, Ia/1026, D/p248, D(s)/2923, D(s)/5058, D(s)/8039, E/9383, E(s)/1968, F/70, F(s)/70, F/381, F(s)/4022, F(s)/8068, and D(s)/561 were propagated from frozen samples in the University of Washington Chlamydia Repository. The "(s)" designation indicates that the strain demonstrates the nonfusogenic phenotype in culture (118). Specimen collection, culture isolation techniques and serotyping were conducted as previously described (121). Infected HeLa or McCoy cells were incubated in Minimal Essential Medium with 10% fetal bovine serum (MEM-10; Sigma-Aldrich) with or without cycloheximide (1mg/ml) at 37 ^{o}C in 5% CO_2. Elementary bodies were partially purified by centrifugation of lysates of infected cells through a 30% renografin pad (21).

Determination of infectious EB production

McCoy cell monolayers in 12 mm shell vials were inoculated with chlamydiae suspended in SPG (0.25 M sucrose, 10 mM sodium phosphate, 5 mM L-glutamic acid). All infections were conducted at a multiplicity of infection (MOI) of 0.5. Inocula were centrifuged onto the monolayers for 1 h at 1,200 x *g*. The inocula were then removed and cells incubated in MEM-10. Infected monolayers were washed with PBS and lysed for EB collection by incubation in distilled water for 2 minutes at 0, 24 h and 48 h p.i. These lysates were plated in triplicate onto McCoy cell monolayers in 48-well plates using ten-fold dilutions for determination

of inclusion forming units (IFU). The average IFU output per well and the standard error of each average was determined for each of the triplicate cultures.

Antibodies and reagents for microscopy

Monoclonal antibody directed at *C. trachomatis* IncA (3H7) was produced as described (7). Monoclonal antibody E6-H1, specific for chlamydial lipopolysaccharide, was a gift from Harlan Caldwell. Anti-serovar-J MOMP (CC-1), anti- serovar-I/Ia MOMP PE-5 and anti-serovar-F MOMP (FC-2) were acquired through the Washington Research Foundation (Seattle, WA). All secondary antibodies were purchased from the Southern Biotechnology Associates, Inc (Birmingham, AL). The DNA-specific fluorescent label 4',6-diamidino-2-phenylindole dihydrochloride (DAPI; 2 ug/ml in mounting medium; Vectashield, Vector Laboratories) was used to label host cell and bacterial DNA. Antibody labeling was performed on chlamydiae grown in HeLa cell monolayers on 12 mm coverslips as previously described (7).

Clonal isolation of independent IncA-positive and -negative strains in cell culture

Cloned isolates were initially produced from patients persistently infected with *C. trachomatis* expressing serovar Ia. These patients were cultured several times over the course of two to five years and the recovered chlamydiae originally used for different studies (31). A limiting dilution approach was used to generate microbiological clones from these isolates (115). Immunofluorescence was used to determine the IncA- positive or IncA-negative status of cloned strains from the clinical samples.

Screening IncA-negative clinical isolates for the presence of minority IncA-positive sub-populations

Fifty-five clinical samples initially identified as containing IncA-negative *C. trachomatis* were examined for minority IncA-positive sub-populations. An aliquot of each tested isolate was inoculated onto McCoy cells at high multiplicity and cultured for 40 hours. Cells were then fixed with methanol and labeled with anti-IncA and anti-LPS mAbs. IncA-positive inclusions were identified in these

populations by scanning the entire coverslip using anti-IncA antibody at low magnification.

Previous EB production experiments demonstrated that IncA-positive isolates have a slight to moderate growth rate advantage over IncA-negative strains in vitro, depending on the particular strain (132). This growth rate difference was exploited for the isolation of minority IncA-positive clones against a majority population of IncA-negative strains in selected patient samples. For this purpose, the original clinical specimens were passed in shell vials until a titer of >1000 IFU/ml was attained. This culture was then lysed by sonic disruption and stored at -80 °C. A set of vials was then inoculated (MOI ~ 1) using this lysate as a source of EBs. Single vials of the subsequent culture were sonicated every hour between 12 and 20 h.p.i. Each sonicate was aliquoted and frozen at -80°C. After collecting and freezing the samples, one aliquot from each time point was thawed and inoculated onto McCoy cells in shell vials. These vials were cultured for 30 h.p.i. and labeled with anti-LPS mAb. This experiment determined the earliest time point that EB's could be harvested from the frozen aliquots of infected cell lysates. Because the IncA-positive strains tend toward faster growth than serovar-matched IncA-negative strains, collection of EB's from the earliest time point possible enriched for IncA-positive *C. trachomatis* from the mixed clinical isolates. Lysates from the earliest cultures showing viable EB production were serially passed in the same manner until it was determined that the IFU of the IncA-positive strain was greater than the IncA-negative strain. This process required several serial passages for each strain. Once this was accomplished, a two-fold limiting dilution method was used to microbiologically clone the IncA-positive chlamydial strain (115). Clonal isolation of the majority IncA-negative population from these isolates was accomplished directly via limiting dilution. Clones of these serovar matched IncA-positive and -negative pairs from each isolate were amplified and stored at -80 °C.

PCR amplification and nucleotide sequence analysis

All references to gene number are based on the published *C. trachomatis* D/UW3

genome (109). To assess the genetic similarity of these candidate mutant and wild-type strain pairs recovered from same patient specimen, we sequenced the entire open reading frames (ORFs) of 4 *C. trachomatis* polymorphic genes, namely *incA* (CT119)*, omp1* (CT681, encoding MOMP), *tarp* (CT456; reference (28) and *tox* (CT166; reference (33). To sequence the entire ORF, we used specific PCR primers that are in the flanking regions of both >100 base pair (bp) upstream of the 5' ends and >100 bp downstream of the 3' ends (Table 2.1). The complete ORF for each gene was amplified with PfuUltra (Stratagene), and both strands of the full-length gene were subjected to nucleotide sequence analysis. Consensus sequences were derived from assembled contigs with the Sequencher program (Gene Codes Corp., Ann Arbor, MI) and were compared with corresponding gene sequences in the reference D/UW-3 genome and in control strains of the same serovar.

Mouse infections

Seven- to eight-week-old female BALB/c mice were purchased from Charles River Laboratory (Wilmington, MA). Animals were injected with 1.25 mg of Depo-Provera (progesterone) on days 10 and 3 prior to infection. The mice were inoculated intravaginally directly to the cervix with 5.0 x 10^5 IFU/ mouse of appropriate *C. trachomatis* strains in a final volume of 20 ul using a positive displacement pipette. Swabs were collected on days 7 and 10. Swabbing was conducted with a Dacron swab that was inserted in the vagina and rotated 25 times. Swabs were placed in 1 ml of SPG buffer and immediately frozen at -80 °C. Recovered IFU were enumerated by inoculating serial 10-fold dilutions of swab sample onto wells of 48-well plates which were centrifuged as above and incubated for chlamydial growth for 48 h. Cells were then fixed with methanol and labeled with anti-LPS mAb. Fluorescent microscopy was used to count chlamydial inclusions from which the number of shed IFU/mouse was calculated. Five mice were inoculated per strain**.** Shedding of infectious chlamydiae were compared between IncA-positive and -negative strains using a Student's t-test and through Repeated Measures Effects ANOVA and a mixed model with random effects.

Analysis of transcription in IncA-positive and -negative strains

Total RNA was harvested by treating mock-infected or *C. trachomatis*-infected cells (MOI = 1) with the TriZol reagent (Invitrogen, Carlsbad, CA). Amplification grade RQ1 DNase (Promega, Madison WI) was added to remove residual genomic DNA. The RNA mini kit (Qiagen, Valencia, CA) protocol was used to remove residual DNAse and contaminants. RNA integrity was verified electrophoretically and by UV spectroscopy.

The NorthernMax-Gly kit (Ambion, Austin, TX) was used for Northern blot analysis of *incA* transcript in cells infected with IncA-positive and -negative strains. DNA probes (a 316-bp *incA* and a 320- bp *16S rRNA* PCR product) were labeled with digoxigenin (DIG) -labeled dNTPs (Roche Diagnostics Corporation, Indianapolis, IN) and incubated with blots at 52 °C for 12-16 h. Blots were washed and incubated with anti- DIG antibody prior to chemiluminescent detection.

Mapping of the transcriptional start for *incA* was conducted using a 5'-rapid amplification of cDNA ends procedure (5'-RACE; Invitrogen). cDNAs were generated via reverse transcription of 5 μg of total RNA using Superscript II polymerase (Invitrogen) and a gene-specific primer, RCE1 (Table 1.1). Sequencing of products was performed at the Center for Genome Research and Biocomputing at Oregon State University.

Quantitative RT-PCR was performed using the Superscript II™ First Strand Synthesis System (Invitrogen). cDNA from total RNA (5 ug) was produced by using a protocol supplied by the manufacturer. Separate PCR amplification reactions were used for target cDNAs from experimental and control samples. Control samples to normalize DNA content were cDNAs amplified with *16S rRNA* and/ or *groEL* primers. C_T values and corresponding copy numbers for the controls were obtained from a standard curve generated by PCR amplification of several dilutions of amplicons with known quantity were compared to copy numbers obtained from PCR reactions of the target experimental samples. All assays were performed in triplicate.

Genomic DNA preparation and sequence analysis

Elementary bodies of cloned chlamydial isolates D(s)/2923, F/70, F(s)/70, J/6276, and J(s)/6276 were purified as previously described (21). Purified EBs were incubated with RQ1 DNAse I (Promega) for 60 min and the DNAse inactivated with 2 mM EGTA. Chlamydial DNA was then extracted using a small scale genomic preparation kit (Qiagen) after suspending the EBs in 5 mM dithiothreitol. Isolate D(s)/2923 genomic DNA was sequenced using classical Sanger sequencing methods at the Joint Genome Institute (Walnut Creek, CA). DNA from isolates F/70, F(s)/70, J/6276, and J(s)/6276 was further processed for Solexa-based sequence analysis using commercial DNA preparation kits (Illumina Inc., San Diego, CA), following the manufacturer's instructions.

Solexa-derived draft genomes were first assembled using the reference-guided assembly program Maq (http://maq.sourceforge.net/). Regions in reference-guided assembled genome where Maq could not resolve sequence were then compared to contiguous sequences assembled through the use of de-novo assembly software (VCAKE; (61)) and a single contiguous draft sequence was produced. Single nucleotide polymorphisms (SNPs) between matched pairs were located using the Diffseq program from the Emboss software suite (92). Isolate genomes were also compared with the published *C. trachomatis* D/UW3 genome sequence (Genbank accession AE001273; reference (109)) using Diffseq to locate SNPs. The location and effect of individual polymorphisms were first determined using an in-house SNP parsing program (http://web.engr.oregonstate.edu/~davidsjo/). Any necessary manual gene variation analysis was performed using MacVector sequence analysis software (MacVector, Cary, NC). PCR analysis was used to verify 4 SNPs found between the serovar F matched pair, as well as 5 of the 51 SNPs found in the serovar J matched pair isolates.

Polymorphisms were then analyzed to determine which SNP in each serovar J pair was parental and which carried the product of the change. Each polymorphic nucleotide was compared to the corresponding nucleotide of two

published *C. trachomatis* strains (D/UW3 Genbank accession AE001273, and A/HAR-13 Genbank accession CP000051) and an additional four unpublished *C. trachomatis* genomes in our database (not shown). In cases where an individual SNP within one strain matched the corresponding nucleotide in all reference genomes, that strain was considered parental. The choice of parental strain for a particular SNP was recorded as unknown if there was not complete agreement between the sequence in one of the examined matched pair sequences and all of the reference strains used for comparison.

RESULTS

Analysis of sequential same-serovar patient isolates

Table 2.2 shows a timeline of culture results from two patients identified in a previous report as having multiple sequential same-serovar infections (31). Selected samples from these experiments were serotyped and screened for the nonfusing phenotype. Each tested isolate was confirmed as serovar Ia, but the nature of inclusion formation in cells infected at high MOI varied among isolates. Fluorescent microscopic analysis with anti-IncA antibody was used to determine the IncA phenotype in these isolates (Fig. 2.1 A-D). On one occasion each for two patients, the nonfusing (IncA-negative) phenotype was recovered in cloned isolates [isolate Ia(s)/1010 and Ia(s)/1025]. Isolates recovered and cloned at other times from the same patients were of serovar Ia and were IncA-positive (Fig. 2.1). These results supported the hypothesis that clonally related IncA-positive and IncA-negative *C. trachomatis* might be found in the same patients and vary in abundance within individuals. We then examined our collection of nonfusogenic clinical isolates with anti-IncA antibodies to determine the frequency of these IncA-negative isolates that might have a minority IncA-positive subpopulation. In these experiments, single IncA-positive inclusions were occasionally observed against a background of IncA-negative chlamydial inclusions (Fig. 2.1E and F). Five of 55 IncA-negative isolates contained a minority subpopulation of IncA-positive *C. trachomatis*, and three of these had an

identical serotype with the IncA-negative majority. The remaining isolates contained *C. trachomatis* that were not of the same serovar and most probably were collected from a patient that had acquired a mixed infection. The three matched pairs were of serovar F, serovar J and serovar D. The proportion of IncA-positive vs. IncA-negative strains in these populations ranged from approximately 0.001 (J/J(s); Fig. 2.1E) to 0.3 (D/Ds, not shown; F/F(s); Fig. 2.1F). Panel F (Fig. 2.1F) shows the serovar F matched pair inoculated onto cells at high MOI. Several cells in this image are infected with both strains, with the developmental forms localized to individual IncA-positive or-negative vacuoles. This demonstrates the nonfusogenicity of the IncA-negative inclusions (red) with each other and with the smaller number of IncA-positive inclusions (green). Note that several of the IncA-positive inclusions are linked by IncA-laden fibers. A diluted sample of this preparation yields single inclusions of both types in the infected cells, resulting in an image similar to that shown figure 1E.

DNA sequence analysis of different matched pairs

PCR-based and whole genome sequencing was used to address the question of the clonal relatedness of individual matched pairs. Dean et al. (31) demonstrated that the Ia MOMP sequences were identical or very nearly identical in the strains examined in their work. Our studies began with the nucleotide sequence analysis of four genes, *incA* (CT119), *omp1* (CT681) *tox* (CT166), and *tarp* (CT456) from 5 selected Ia strains isolated from a previous study of chlamydial persistence in patients. Within the complete *incA* ORF we found that the coding sequence for *incA* in the Ia(s) strains exhibited a single bp insertion at position 36, when compared to the *incA* sequence from the wild-type Ia strains (Fig. 2.2). Otherwise all sequences from the Ia strains collected from single patients were identical.

Preliminary sequence analysis was also conducted on the serovar J and F matched pairs. The nucleotide sequence analysis of *incA* from the IncA-negative strain J(s)/6276 had a single base pair deletion at position 18 that resulted in a reading frame shift and inactivation of the gene (3). This deletion was not

represented in the IncA-positive strain isolated from the same patient. *incA* was also sequenced from the strains F/70 and F(s)/70, a matched pair of serovar F. The coding sequence of *incA* was inactivated in strain F(s)/70 by a single deletion at position 515. Therefore, the sequences at *incA* were identical within each of the Ia, J, and F matched pairs, with the exception of the single insertion or deletion that interrupted the gene (Fig. 2.2).

Genomic sequence analysis

Five chlamydial genomes were sequenced to investigate the possibility that matched pairs were more closely isogenic than a serotype-matched, but otherwise unrelated, pair of IncA-positive and IncA-negative strains. The Illumina sequencing technology was used to generate genome sequence from cloned strains for each strain within the J and F matched pairs. The genome sequence of the F and F(s) matched pair differed within *incA* as described in the previous paragraph, and differed at four other loci within the genome (Table 2.3). The most significant difference was a 1,941 base pair deletion within the plasticity zone (91) of the F(s) isolate, leading to the deletion of ORFs CT160 and CT161, and the fusion of ORFs CT159 and 162 (Fig. 2.2). This deletion was not found in any other examined strain, including two additional serovar F(s) isolates (Fig. 2.2D). ORFs CT160 and CT161 encode hypothetical proteins with no known homologs in other organisms. There are three other single nucleotide polymorphisms (SNPs) that varied between these genomes, a substitution in CT135, a single nucleotide deletion that leads to a truncation in CT135 in the IncA-negative strain, and one in an intergenic region.

The two matched serovar J strains differed at 51 nucleotide positions within the genome, one of which was the change in *incA* as described above. Surprisingly, the only other ORF that is affected by a deletion in this matched pair is CT135, which, in this case, is truncated in the wild-type strain but intact in the IncA-negative strain (Fig. 2.2). Outside of the differences in *incA* and CT135, the SNPs in the serovar J matched pair include 13 within intergenic regions and 36 within coding regions (Table 2.3). Amino acids are changed in 24 of these

genes. In order to determine the probable parental sequence within an individual SNP, the nucleotide at each position was compared with genomic data from published (109) and unpublished genome sequences (Table 2.4). These comparisons showed that 24 of the SNPs most probably changed from parental sequences in the IncA-negative strain while 25 most probably changed from parental sequences in the IncA-positive strain. This suggested that both strains accumulated mutations at approximately the same rate. Similarly, there were no apparent differences in any category of genetic change between the fusogenic and nonfusogenic strain, including the number of transitions, transversions, deletions, or choice of mutated base. There was, however, a bias in the movement toward A + T in the cumulative mutations, with 32 of the total changes being (C or G) to (A or T). In contrast, 13 total changes went from (A or T) to (G or C).

Conventional shotgun sequencing was used to determine the genome sequence of strain D(s)/2923, an IncA-negative *C. trachomatis* isolate that was isolated in the absence of an IncA-positive match (118). Genome sequence analysis of strain D(s)/2923 identified a single nucleotide deletion in *incA* that resulted in a 99 amino acid C-terminal truncation. The complete genome differed from the prototype D/UW3 sequence by 5764 base pairs, including 47 cases where indels were differently present in genes of the two strains. The amino acid sequence of MOMP in D(s)/2923 was 99.0% identical to the D/UW3 MOMP sequence.

The above results demonstrate that the genome sequence of strains within a same-patient serovar-matched pair were very similar and, in one case, nearly isogenic. In contrast, an IncA-negative strain isolated in the absence of a serovar-matched IncA positive strain had many more SNPs relative to the genome sequence of a serovar-matched *C. trachomatis* strain.

Transcriptional analysis of IncA-positive and -negative strains

Northern blot analysis and quantitative RT-PCR were used to investigate transcript levels in cloned isolates of IncA-positive and -negative strains.

Examination of a set of unmatched IncA-positive and -negative strains demonstrated that *incA* transcript is found within both IncA-positive and -negative strains. However, transcript abundance was reduced approximately 10-fold in tested IncA-negative strains (Fig. 2.3). Two transcripts were also found in each strain, one that measured approximately full length for *incA* (~850 bp) and one slightly smaller species. Quantitative PCR was then used to examine cloned strains isolated from patients infected singly with either an IncA-positive or -negative strain, and each cloned isolate from the F and the J matched pairs. IncA-negative strains collected independently of a matched pair had levels of transcript that were consistent with the Northern blotting (Fig. 2.3C). However, both the IncA-positive and -negative members of each matched pair had *incA* transcript levels that were very similar. These data demonstrated that two groups of IncA-negative strains exist with respect to *incA* transcription: one of which had similar levels of transcript relative to a serotype-matched strain, and another type that had reduced levels of *incA* transcript.

The transcriptional start site of *incA*, as mapped with 5'-RACE analysis, was identical in all tested IncA -positive and -negative strains (Fig. 2.3B). Additionally, there was no evidence of changes in the sequences upstream of *incA* that might correlate to the reduced level of transcript.

In vitro growth

Each of the *C. trachomatis* isolates used in the mouse experiments, plus each of the strains from the Ia and F matched pairs, were used in EB output experiments following culture in McCoy cells (Fig. 2.4). Each strain was inoculated at an MOI of approximately 0.5. Between 1.3 and 3.1 x 10^4 EBs were harvested in cultures conducted immediately after inoculation and variation in this recovery was not a function of IncA positivity or negativity (not shown). Each strain infected and grew within the cells, but there were clear differences in the numbers of EB produced in the infections. Consistent with the work of Xia et al. (132), the IncA-negative strains isolated outside of a matched pair accumulated at least 10-fold fewer EBs at both 24 h (Fig. 2.4) and 48 h p.i. (not shown). In contrast, IncA-

positive and -negative members of each matched pair produced very similar numbers of infectious EB.

Mouse infections

We examined the ability of IncA-positive and-negative strains to colonize the murine genital tract. BALB/C mice were infected and then cultured on days 7 and 10 p.i. (Fig. 2.5). One matched pair (serovar J) was used in these studies and several other clinical isolates were tested, none of which were part of a matched pair. Wild-type strains of serovars J, D, and E were compared to serotype-matched IncA-negative strains.

The EB production by the J and J(s) strains in the matched pair were similar and not statistically different, while each IncA-negative strain outside of a matched pair was significantly different from each serotype-matched wild-type strain ($P< 0.001$, Student's t-test). ANOVA analyses also demonstrated that there was a significant difference ($P<0.001$) between chlamydial production in all IncA-positive vs. unmatched IncA-negative strains. These data support the hypothesis that IncA-negative strains that are part of a matched pair grow at a similar rate as IncA-positive strains in vivo. In contrast, IncA-negative strains that were not part of a matched pair had an in vivo colonization deficiency relative to both the wild-type strains and the single tested IncA-negative strain that was part of a matched pair.

DISCUSSION

We previously showed that *incA* is inactivated by frameshifts, deletions, and nonsense mutations, leading to strains of the organism that form nonfusogenic inclusions (94). Although there was some ambiguity in this issue in previous publications (94, 118), complete sequence analysis on cloned strains demonstrate that all truly IncA-negative nonfusogenic strains have an interrupted *incA* coding sequence. Retrospective epidemiological studies demonstrated that the nonfusing strains produced lower numbers of elementary bodies, fewer

inflammatory cells, less severe symptomology in patients (41) and may have growth defects in vitro (132). However, in our analysis of *C. trachomatis* isolated in the Seattle/King County area, the incidence of IncA-negative strains approaches 2% of all non-LGV genital serovars except serovar G (118). Therefore, we hypothesize that IncA-negative strains either are generated very commonly in vivo or manifest a unique phenotype allowing them to compete successfully with IncA-positive strains in patients. The first evidence that variation within *incA* might occur in a single patient was identified in a pair of patients infected with genetically similar strains that were either IncA-positive or -negative at different culture points during infection (Table 2.2). These strains were microbiologically cloned for a previous study (31), and the original culture is not available. But, the ability to clone, by limiting dilution, sequential IncA-positive and -negative strains suggested that the mixed population of IncA-positive and -negative *C. trachomatis* persists in patients *in vivo*. The identification of these serial cultures led us to examine the possibility that individual nonfusogenic isolates in our collection had a minority IncA-positive subpopulation. It was fully possible that we missed such mixed populations in the original screen, since the screen for variant isolates involved the formation of multiple inclusions in cells, a property that would be conserved in a mixed IncA-negative and -positive population (94). Several such strains were identified, including some that were of mixed serovars and three that were of the same serovar. The identification of these same serovar, IncA-negative and -positive “matched pairs” led to two hypotheses. First, deletions accumulate within the gene during culture in vivo, leading to a subpopulation that is IncA-negative, and, second, this variation led to diversity that allows IncA-negative strains to compete with wild-type strains in patients.

If the matched pairs represent very recent mutation events, possibly in the patient being cultured, there should be very little variation between the genomes in a matched pair. A rapid genome sequencing approach was used to demonstrate that this was the case. The serovar F matched pair, in which the IncA-positive strains represented approximately 20% of the population, had

differences at four positions in the genome. The serovar J matched pair, in which the IncA-positive subpopulation was represented at less than 0.1%, had differences at 51 positions. These data support the conclusion that the matched pairs had recently diverged from one another. In contrast, two serovar D strains that vary at *incA* but that were isolated independently, differed at over 5,000 base pairs. We hypothesize that, at least in the case of the F pair, this divergence may have occurred in the patient from which the sample was collected.

We hypothesize that the shift was from the IncA-positive phenotype to the IncA-negative phenotype for two reasons. First, the difference between IncA-positive and the IncA-negative matched pairs are either a deletion (in the F or J pairs), or an insertion in the sequence (the Ia pairs), which lead to different truncated protein products. There are no changes in the reading frame that suggests a suppressing mutation is present in the IncA-positive strains relative to the IncA-negative strains. Thus, in order to mutate from IncA-negative to positive, a progeny IncA-positive strain would need to have a correcting mutation at the exact same position in the gene. Secondly, one of the few differences between the F and F(s) matched pair is a 1,941-bp deletion in the F(s) strain. The presence of this deletion can more easily be explained by the IncA-positive strain being parental to the IncA-negative strain.

Genome sequencing identified four inactivated genes in the serovar F IncA-negative strain, relative to the matched IncA-positive strain. This included *incA*, CT135, and a large deletion that removed CT160 and 161 (Fig. 2.2). While there were 51 nucleotide differences in the serovar J matched pair, only 2 involved indels in coding sequences- the change inactivating *incA* and a change involving CT135. In this case, however, the IncA-positive, wild-type, strain had an inactivated CT135 ORF, while the IncA-negative variant had an intact ORF. This ORF was also shown to be one of approximately 20 variable genes among trachoma biovar strains (63). CT135 is an uncharacterized gene encoding a hypothetical protein that is conserved among *C. trachomatis*, *C. muridarum*, and *C. suis*. Distant homologs are present in the other chlamydiae, but they are not

in a syntenous region of the chromosome. Within *C. trachomatis*, the predicted CT135 protein contains a hydrophobicity motif suggestive of possible localization to the inclusion membrane (Fig. 2.2; (7)). There is no reason to believe that direct inactivation of CT135 is associated with any aspect of the IncA-negative phenotype, as the gene is intact in strain D(s)/2923 and in the IncA-negative member of the serovar J matched pair. However, the collected analysis of this gene suggests it is commonly variable in otherwise closely related clinical strains, perhaps varying the infectious process by differential localization of the encoded protein to the inclusion membrane.

The absence of IncA on the inclusion membrane is directly associated with lack of fusogenicity in clinical isolates, because IncA molecules on adjacent inclusions participate in homotypic vesicle fusion processes (32, 52, 118). However, we have previously demonstrated that there are growth rate differences, host gene expression differences, and patient phenotype differences between IncA-positive and IncA-negative strains. While these additional traits were associated with the absence of IncA, it may be that lacking IncA is a marker for these differences and not the actual cause of at least some of these phenotypes (132). It is possible that inactivation of IncA is a stage in a randomly selected process that leads to the clinical phenotypes observed in our previous work. The results in the present report provide four lines of evidence for this model. First, genome sequencing suggests that the strains isolated as matched pairs have very little sequence variation relative to strains isolated independently. Second, quantitative PCR data demonstrated no difference in *incA* transcript production in the serovar J and F matched pairs, while unmatched IncA-negative strains had a markedly reduced level of transcript. Third, the production of infectious chlamydiae was different between IncA-negative strains within a match vs. those isolated independently of a corresponding match. Each strain in the Ia, F, and J pairs, had a near equivalent number of EBs produced following infection, while the unmatched IncA-negative strains had at least 10-fold lower EB counts at both 24 and 48 h post-inoculation. Finally, the strains in the serovar J matched pair had similar infectivity in the mouse model, while all other IncA-

negative strains were significantly less infectious in mice. The in vivo and in vitro culture data suggest that these results can be explained by a reduction in growth rate by the unmatched IncA-negative strains. However, the reduction in *incA* transcription in the unmatched strains, under conditions where a constitutive gene (*groEL*) was standardized, emphasizes that a reduction in growth rate is likely associated with other changes in the genesis of distinct IncA-negative strains.

While we were unable to demonstrate a measurable growth rate difference in the in vitro culture of the members of the matched pairs, our cloning strategy for isolation of the IncA-positive minority members of a matched pair demonstrated that these strains could be educed from their IncA-negative partners by harvesting EBs produced at early time points following infection. This suggests that the IncA-positive strains are able to produce EBs more quickly in culture. It is possible that examining EB production earlier, both in the in vitro and the in vivo culture experiments, might identify a measurable growth rate difference between IncA-positive and –negative strains in a matched pair.

We have never identified a spurious IncA-negative strain being generated in any IncA-positive culture *in vitro*, and thus, it is not likely changes from IncA-positive to IncA-negative occur at an appreciable level in laboratory culture. IncA-negative strains are, however, found at a rate of approximately 2% in clinical isolates at Seattle area STD clinics. The changes are not a function of an expanding clone in the patient population, because the mutations inactivating *incA* are randomly distributed across the coding sequence and are found in strains representing several different serovars. These traits suggest that there exists an unidentified phenotype associated with the variants that facilitates their selection and maintenance *in vivo*. Additionally, the identification of matched pairs from clinical samples demonstrates that IncA-negative strains can grow to high levels with IncA-positive strains in the human host.

A role for the IncA-negative strains in the biology of chlamydial infection and disease remains to be elucidated. Antibody against IncA is present in sera from

infected patients (8), and a homolog of IncA is a CD8+ T cell antigen in *C. pneumoniae* (131). Thus, it is possible that IncA is a significant antigen involved in protection against infection. This might lead to selection of randomly generated IncA-negative strains that can avoid the host immune response. However, it is also possible that IncA-negative strains occupy an uncharacterized niche within the host, and their unique ability to occupy this niche allows for their survival within the larger population of IncA-positive strains. Studies on the biology and fitness of IncA-negative *C. trachomatis* continue in our laboratory.

ACKNOWLEDGEMENTS

Pacita Roberts provided important statistical analyses in this research. Sara Weeks is acknowledged for technical assistance. This research is supported in part by grants from the National Institutes of Health (AI48769, AI031448).

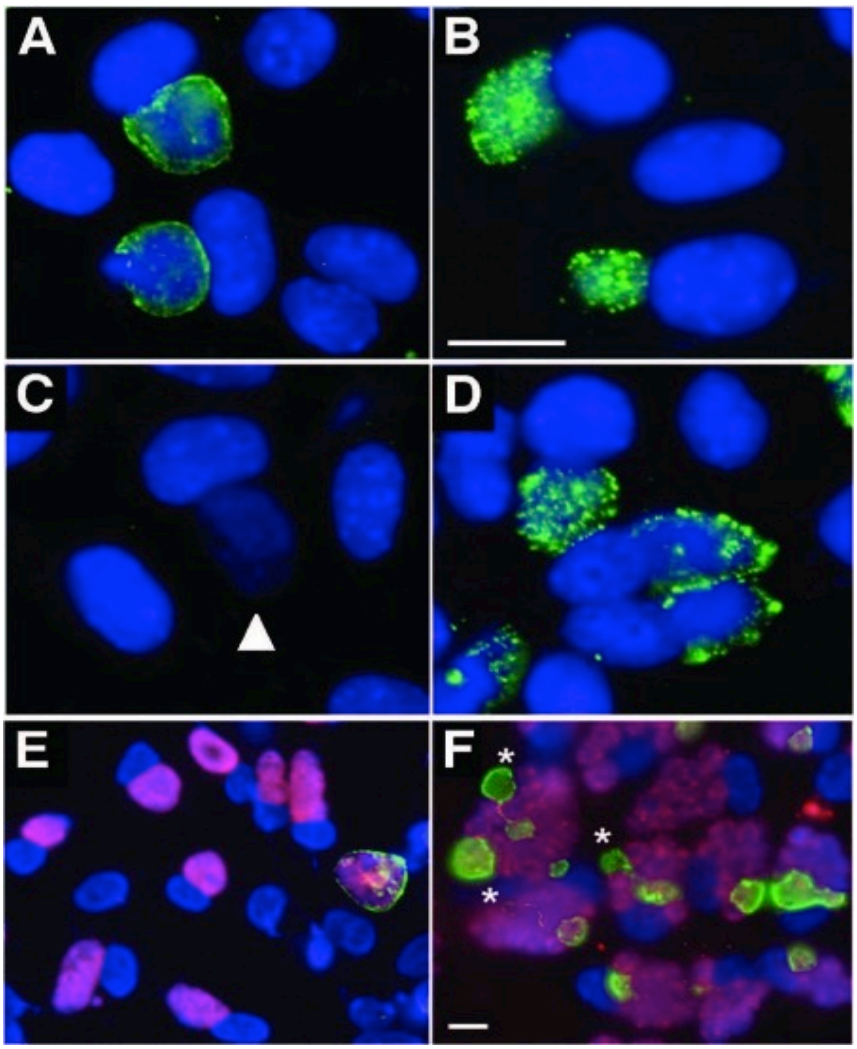


Figure 2.1. Fluorescent microscopic analysis of McCoy cells infected with different IncA-positive and -negative matched pairs.
In all panels, DNA is labeled blue with DAPI. Panels A-D: IncA labeling (green, panels A and C) and anti-serovar Ia MOMP labeling (green, panels B and D) of serial isolates Ia/1026 (panels A and B) and Ia(s)/1025 (panels C and D). Note the unlabeled, IncA-negative inclusion in panel C (arrowhead). The bar in B shows 10 micrometers for panels A-D. Panels E and F: primary isolation cultures for the serovar J and serovar F matched pairs (panels E and F, respectively). In both images, labeling of IncA is green and serotype-specific labeling of MOMP is red. Panel E: A rare (1/10,000) IncA-positive inclusion against background of a largely IncA-negative, serovar J *C. trachomatis* (Isolate 6276). Panel F: A high titer clinical sample that contains both IncA-positive and -negative *C. trachomatis* of serovar F (Isolate 70). The IncA-positive strain within this pair represents approximately 30% of the population. Note secondary inclusion formation (*) and intricate fiber production emanating from IncA-positive inclusions in some cells. The bar in F represents 10 micrometers for panels E and F.

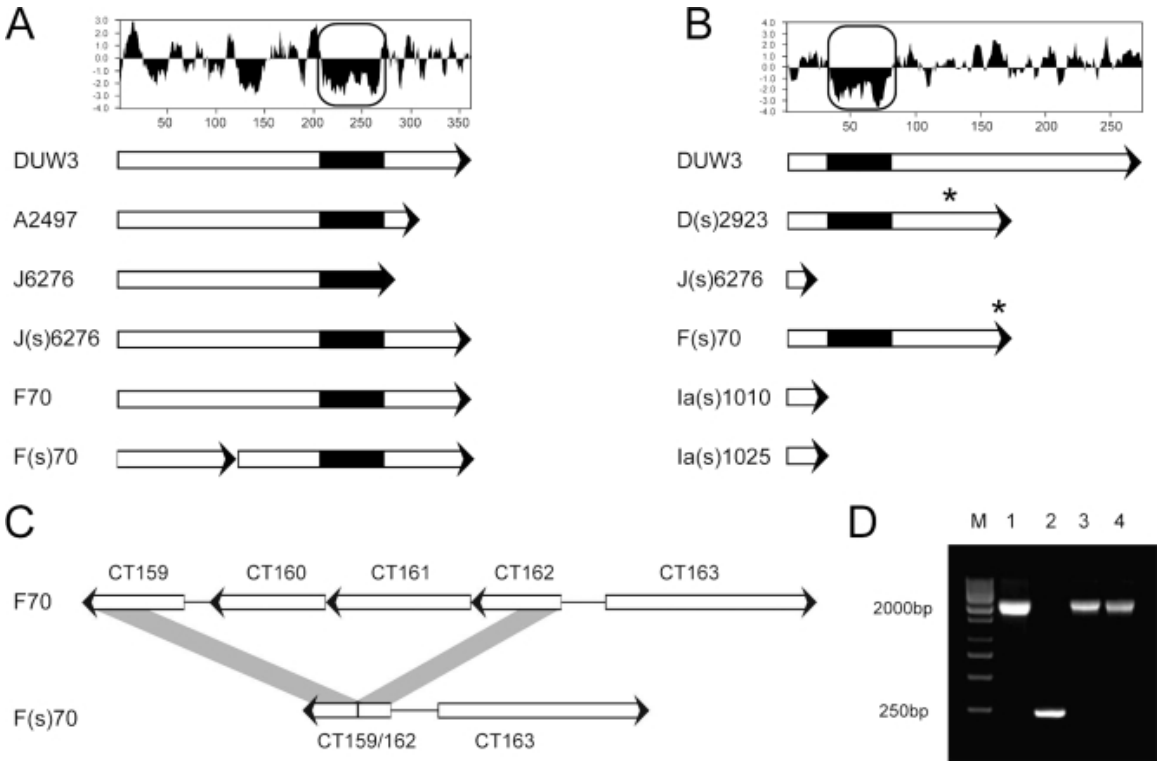


Figure 2.2. Orf analysis of regions showing variation found in the matched pair isolates.

Panel A. Kyte/Doolittle Hydrophilicity plot of gene CT135, and gene representations of lab strain D/UW3, ocular isolate A2496 (63) and matched pair isolates. Panel B. Kyte/Doolittle Hydrophilicity plot of the predicted CT119 (*incA*) protein sequence, and ORF representations of lab strain D/UW3 and clinical isolates showing the IncA-negative phenotype. The boxed regions in panels A and B highlight a putative inc motif. Asterisks indicate locations of frameshift mutations leading to downstream truncation. Panel C. A region of the plasticity zone that is deleted in isolate F(s)/70 relative to strain F/70, resulting in putative CT159/162 fusion. Panel D. PCR analysis confirming the deletion found in isolate F(s)/70, with the length of each product indicated to the left. Lanes: 1) Strain F/70; 2) Strain F(s)/70; 3) Strain F(s)/4022; 4; Strain F(s)/8068.

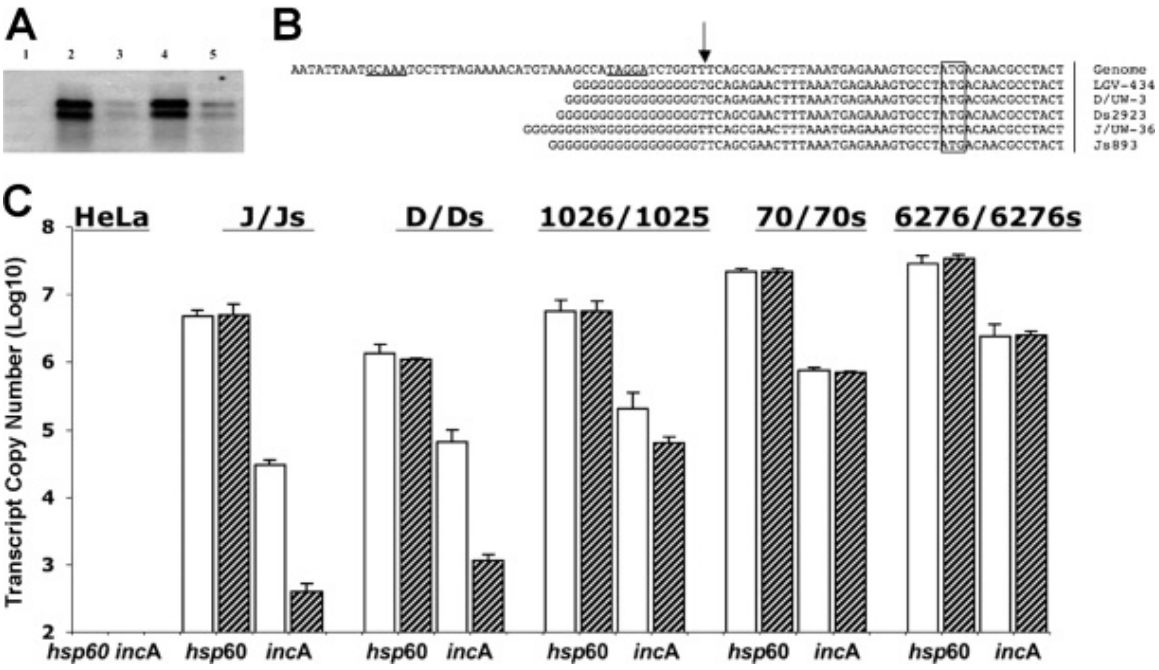


Figure 2.3. Transcription of incA in prototype and clinical C. trachomatis strains.
Panel A: Northern blot analysis of wild-type and IncA-negative *C. trachomatis* strains, probed with full length *inc*A. Lanes: 1: Uninfected HeLa cells. 2: J/UW36. 3: J(s)/893. 4: D/UW3. 5: D(s)/2923. Panel B: Results of 5'-RACE analysis of the *incA* transcriptional start site in wild-type and variant strains. The strains analyzed are indicated on the far right, with the D/UW3 genome sequence ion the top line. The arrow indicates the identified transcriptional start site. The boxed nucleotides shows the predicted translational start site. Underlined sequences on the top line indicate predicted –10 and –35 regions of the serovar D genome sequence. Panel C: Quantitative RT-PCR assays were performed for different *C. trachomatis* strains to determine the abundance of *incA* and *hsp60* transcripts. Cells were infected at an MOI of 3. In each case, the IncA-positive strain is indicated with the clear bar, while the IncA-negative strain is indicated with the hatched bar. To normalize transcript levels in each total RNA sample used for *incA* quantitation, the abundance of *groEL* transcripts was measured and used as control.

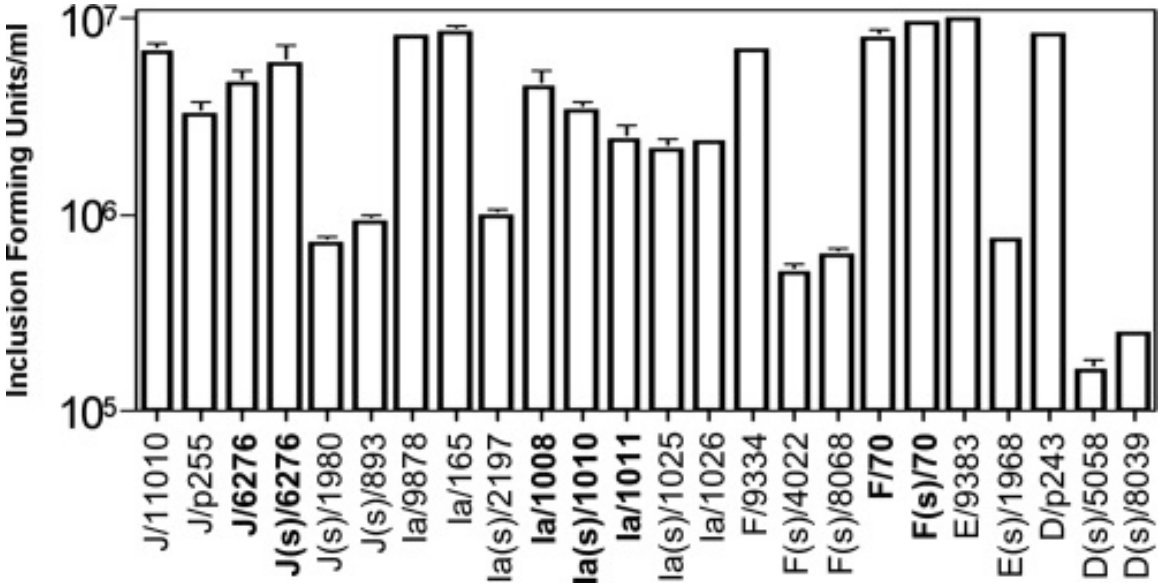


Figure 2.4. Production of infectious chlamydia during growth in murine epithelial cells in vitro.
Chlamydiae were collected for quantitative culture 24 h post infection of the primary monolayer. The vertical axis is logarithmic in scale. Each value represents an average of three wells and standard errors are indicated with error bars. An (s) designation indicates a strain that is IncA-negative. Adjacent strains shown in bold print represent members of a matched pair.

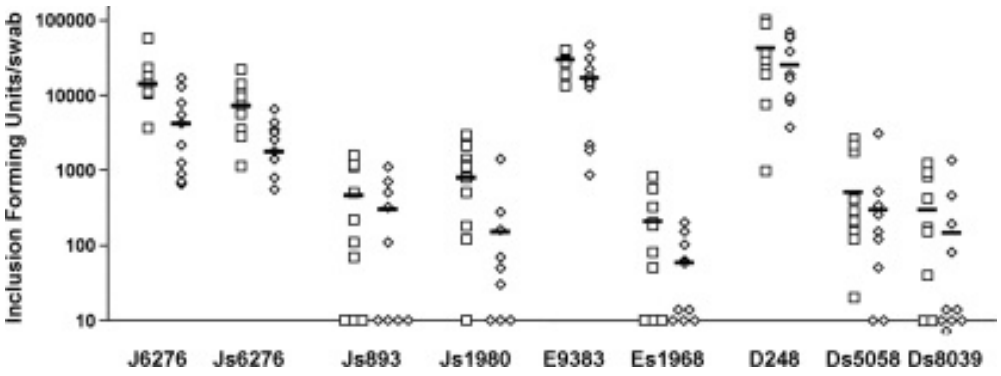


Figure 2.5. Chlamydial shedding from mice infected vaginally with closely related and unrelated same-serovar strains of IncA-positive isolates compared to corresponding IncA- negative variants.
The vertical axis is logarithmic and represents harvested inclusion forming units per mouse at the particular time point. The open squares show culture data from samples collected 7 d post-challenge. The diamonds show culture data from samples collected 10 d post-challenge. The bar in each data set shows the mean value of the isolates for a particular isolate on a particular day. Data points at the bottom of the graph represent culture values below our limits of detection.

Table 2.1. Oligonucleotide primers used in these studies.

Target	Primer or probe	Sequence (5′–3′)
incA	CT119F	ACATGTAAGCTGGTTTTTACATAG
	CT119R	TGGCACCCTGTGTACAGTTAG
omp1	681F	CTTTAACTAGGACGCAGTGCCGC
	681R	TGGACCCGACCGAAGCCGAGCC
	81F2	CTGAGATGTTTACAAATGCCGC
	U681R3	AGACCATTTAACTCCAATGTA
tarp	TPF1	TTCTCCAGATACTTCAGAAAGC
	TPF2	CAAAGACTCTGACGGAGCTGGTG
	TPR1	TGCACACATGTTTTCCAAGTC
	TPR2	CAGGGTAAACGACGTCTAGG
	TPR3	CCTCCAGGAATTCGTCCTCCG
CT166	CT166F	CTCCTCCAGAAGGAACGACAAC
	CT166FA	CTTCAGAGAATAAGGTCACTAC
	CT166F2	AAGAAAAACCTAAGACGACTCCG
	CT166F3	AACGCGTGCAGAGTATTTAGAG
	CT166R	CCCAGCATACATGTCTCCATC
incA RT-PCR	RT-PCR *incA* F	CACATTAGCAGGGAATGCTC
	RT-PCR *incA* R	AGAGCCTTTAAGATCTGC
incA	RCE1	CTTTTTGTAGAGGGTGAT
incA Q-PCR[a]	Q-PCR *incA* F	AACTTTAAATGAGAAAGTGCCTAT GACAA
	Q-PCR *incA* R	ATCAGAAAGCCAACAAGATGT
	incA probe[b]	TATTTTCTTAATTTGTCCATCAA AGAA
groEL Q-PCR	Q-PCR *groEL* F	TGTTACCGTTGCGAAAGAAGT
	Q-PCR *groEL* R	TCCAGCTTTGTCAGCAGTTT
	groEL probe[b]	ACATGAAAATATGGGCGCTC

[a] Q-PCR, quantitative PCR.

[b] Quantitative PCR probes are labeled with 6-carboxyfluorescein on the 5′ end and 6-carboxytetramethylrhodamine on the 3′ end.

Table 2.2. Culture points in single individuals persistently infected with *C. trachomatis* serovar Ia.
Patient 1 was culture/LCR-positive for a total period of three years. Patient 2 was culture/LCR-positive for a total period of six years.

Patient[a] and time point	Culture result	
	Strain	IncA status[b]
Patient 1		
$T = 0$	1025	−
$T = 6$ mo	1026	+
$T = 12$ mo	ND[c]	NT
Patient 2		
$T = 0$	1008	+
$T = 36$ mo	1010	−
$T = 48$ mo	1011	+

[a] Patient 1 was culture/LCR positive for a total period of 3 years. Patient 2 was culture/LCR positive for a total period of 6 years.
[b] −, negative; +, positive; NT, not tested (sample unavailable).
[c] ND, no designated strain.

Table 2.3. Summary of nucleotide differences among the sequenced chlamydial strains.

Strain comparison	Total no. of SNPs	No. of indels	No. of genes affected by indels	No. of amino acid substitutions in gene	No. of SNPs in intergenic region
F/70 vs F(s)/70	5	3[a]	6[b]	0	1
J/6276 vs J(s)/6276	51	2	2	36	13
D/UW3 vs D(s)/2923	5,764	168	54	2,376	1,018

[a] Two of these indels consist of a single base pair. The other indel is the 1,941-bp deletion found in strain F(s)70.

[b] The genes affected by these indels include two from genes affected by single base pair indels (CT119 and CT135) and four genes affected by the 1,941-bp deletion (CT159, CT160, CT161, and CT162).

Table 2.4. Characterization of genomic changes in the serovar J matched pair.

Change	Total no. of changes	No. of transitions	No. of transversions	No. of deletions	No. of the following changes:								
					A to G	C to T	G to A	T to C	C to A	T to A	C to G	C/G to A/T	A/T to C/G
From IncA positive to IncA negative	25	22	2	1	3	8	8	4	0	0	2	16	7
From IncA negative to IncA positive	24	19	4	1	6	10	3	0	3	1	0	16	6
Unknown	2												

Genome Sequencing of Recent Clinical *Chlamydia trachomatis* Strains Identifies Loci Associated with Tissue Tropism and Regions of Apparent Recombination

Brendan M. Jeffrey, Robert J. Suchland, Kelsey L. Quinn, John R. Davidson, Walter E. Stamm, and Daniel D. Rockey

Infection and Immunity, American Society for Microbiology

1752 N Street N.W.

Washington DC, 20036

Volume 78, No. 6, pages 2544-2553

ABSTRACT

The human pathogen *Chlamydia trachomatis* exists as multiple serovariants that have distinct organotropism for different tissue sites. Culture and epidemiologic data demonstrates that serovar G is more prevalent, while serovar E is less prevalent, in rectal isolates from men having sex with men (MSM). The relative prevalence of these serovars is the opposite in female cervical infections. In contrast, isolates of serovar J have approximately equal prevalence at each tissue site, and are the only C-class strains that are routinely cultured from MSM populations. These correlations led us to hypothesize that polymorphisms in open reading frame sequence will correlate with the different tissue tropisms in these serovars. To explore this, we sequenced and compared the genomes of clinical isolates representing both anorectal and cervical isolates of serovars E, G, and J, and compared these with each other, as well as with a set of previously sequenced genomes. We then used PCR- and restriction-digest-based genotyping assays on a large collection of recent clinical isolates to show that polymorphisms in ORFs CT144, CT154, and CT326 were highly associated with rectal tropism within the G serovar, and polymorphisms in CT869 and CT870 were associated with tissue tropism across all tested serovars. The collected genome sequences were also used to identify regions of likely recombination in recent clinical strains. This work demonstrates that whole genome sequencing and comparative genomics is an effective approach for discovering variable loci in *Chlamydia* spp. that are associated with clinical presentation.

INTRODUCTION

Chlamydia trachomatis is an obligate intracellular human pathogen that is the leading cause of preventable blindness worldwide and is the most common sexually transmitted bacterial infection in humans. The study of chlamydial biology is complicated by their obligate intracellular development and the lack of a routine system for directed mutagenesis. Chlamydial isolates are differentiated

into serovars based on sero-specificity to chlamydial major outer membrane protein (MOMP (22), encoded by *ompA* (109)). These serotypes fall into biological groupings associated with trachoma (serovars A-C), sexually transmitted non-invasive disease (serovars D - K) and invasive lymphogranuloma (L1 to L3) (101). Comparative genomic analysis of ocular and urogenital chlamydial species has proven to be an effective approach for discovering genetic loci that associate with observed tissue tropism (24, 25).

Studies conducted in Seattle, Washington and Birmingham, Alabama have shown a prevalence of serovar G rectal isolates in men having sex with men (MSM), whilst serovar E rectal isolates are less prevalent (9, 18, 42). This prevalence for serovar G and rectal tropism is in contrast to what is observed in studies of female cervical populations in the same geographical regions, where serovar E prevalence was significantly higher than serovar G (116). It is not clear whether the causes of these disparities are behavioral by means of network bottlenecks or if there are genuine biological differences between rectal-tropic and cervico-tropic strains.

The limited examples of horizontally acquired DNA in chlamydial species suggest that lateral gene transfer and recombination might be rare in this system. However, sequencing efforts have identified clear examples of recombination at a limited number of chlamydial loci, including *ompA* (20, 45-47, 73, 80). Recent studies have shown that chlamydiae do contain the necessary machinery for recombination (16, 109) and that lateral gene transfer can be selected for in cell culture following co-infection with strains carrying dissimilar drug markers both within and among chlamydial species (16, 33, 34, 120). Mechanisms of recombination and any role of recombination in chlamydial fitness in vivo remain to be investigated.

The distributions of serovar G strains across the heterosexual and MSM populations led us to hypothesize that strains with rectal tropism will have variable genes or loci compared to other urogenital isolates. To test this hypothesis, we sequenced eight chlamydial isolates representing serovars D, E,

F, J, and G, that were collected from the female cervix, male urethra, or male rectum. PCR and restriction fragment length polymorphism (RFLP) assays were then developed to determine if candidate ORFs identified in the genome sequence analysis were associated with observed tropism to the rectal site of infection. A set of candidate ORFs that were associated with rectal tropism in serovar G were discovered, and polymorphisms in the *pmp* genes were correlated with rectal tropism across all tested serovars. Analysis of these genomes also demonstrated that recombination appears common in clinical isolates, and occurs at locations across the chlamydial genome.

MATERIALS AND METHODS

Chlamydia strains and genomic DNA preparation

Chlamydia trachomatis clinical isolate strains Ds/2923, E/11023, E/150, F/70, G/9301, G/9768, G/11222, G/11074, and J/6276 were propagated from frozen samples stored at the University of Washington Chlamydia Repository (117). Isolate collection, clonal isolation techniques, serotyping and elementary body (EB) purification were conducted as previously described (119, 121). Purified EBs were incubated for 60 min with 4 units/mL RQ1 DNase (Promega) followed by treatment with 2 mM EGTA (RQ1 Stop solution, Promega) to inactivate the enzyme. DNA was then extracted from purified, DNAse-treated EBs using the Qiagen Genomic Tip kit (Qiagen, Valencia, CA) following the manufacturer's instructions. The initial suspension buffer supplemented for these purifications was supplemented with Dithiothreitol (5 mM) to facilitate EB lysis.

Genome Sequencing and Sequence Analysis

Isolates Ds/2923, E/11023, E/150, G/9301, G/9768, G/11222 were sequenced using classical Sanger sequencing methods at the Joint Genome Institute (Walnut Creek, CA). DNA from isolates J/6276, G/11074, and F/70 was processed for Illumina-based sequencing using commercial DNA preparation kits (Illumina Inc., San Diego, CA) following the manufacturer's instructions. Illumina-

derived genomes were first assembled using the reference-guided assembly program Maq (76). Regions in reference-guided assembled genomes where Maq could not resolve sequence were then compared to contiguous sequences assembled using de-novo assembly software VCAKE (61) and a single contiguous draft sequence was produced.

Whole genome phylogenetic analysis was performed using the alignment program MAFFT with default settings (65, 66). These comparisons included sequences generated in this study, as well as published genomes for serovars D, A, and L2 [D/UW3 genbank accession number AE001273 (109), A/HAR-13 genbank accession number CP000051 (25), L2-434/Bu genbank accession number AM884176 (125)]. Pairwise genome alignments produced using MAFFT (settings: iterative refinement = 2, default gap opening penalty = 1.53, default gap extension penalty = 0.123) were used to count the total number of substitutions and insertions/deletions (indels) between genome sequences. Regions of high variability between selected sequences were analyzed manually using MacVector sequence analysis software (MacVector, Cary, NC), and counts were adjusted accordingly. Isolate genome sequences were compared with the published *C. trachomatis* D/UW3 genome sequence (GenBank accession number AE001273) (109) using Diffseq from the Emboss Bioinformatics suite (92) and an in-house SNP parsing program (Diffsort: http://people.oregonstate.edu/~rockeyd/Diffsort) was used to determine the locations and translational effects of identified polymorphisms. Any gene variation not resolved by Diffsort was manually analyzed using MacVector sequence analysis software.

Genome-wide recombination analysis

DNA sequences from selected isolates were computationally extracted using sliding windows (1000 nucleotide windows, 800 nucleotide slides) and were used as a bioinformatic probe of a database of template genome sequences (D/UW3, J/6276, G/9768, E/11023, and/or F/70). A comparison of the BLAST raw score values of each window was plotted based on whether it was more similar to a

clade containing serovars J, G, or D, a clade containing serovars E or F, or if the probe sequence equally matched all template genomes. The following rules were used to assign a window to a particular genome sequence or clade: Queries that matched all serovars equally well were plotted along the 'All' line. Queries that did not match all genomes equally, but in which one matched template was either serovar E or F, were grouped with the "E or F" clade. Queries that matched any of serovars J, G, or D more closely than either serovar E or F were grouped with the "J, G, or D" clade. A single query did not match any of the template genomes and was categorized as a "No Hit" in this analysis. Whole genome results from this parsing were then graphed using the length of the complete D/UW3 genome as a reference, beginning with ORF CT001 (109).

Regions of apparent recombination in strain Ds/2923 were then characterized using the ClustalW program, and identified loci were aligned with corresponding sequences from strain D/UW3 and E/11023. These alignments were used to determine the number of informative sites shared by each strain. For these purposes, an informative site is any position in the examined sequences where there is a polymorphism between the template genomes being analyzed. Insertions and deletions of any size were counted as one informative site. An identical approach was used in the analysis of recombination in strain F/70, using strains D/UW3, J/6276, G/9768, E/11023 as templates.

Clinical isolate DNA preparation and PCR-based genotyping

Clinical isolate genome sequence variation data was used to design PCR and RFLP assays of genotypic variation in a population of clinical isolates stored in the repository. Isolates used for these studies included chlamydiae collected either in King County, WA (55 isolates), Lima, Peru (2 isolates, (98)), or Birmingham, AL (6 isolates). Genes to be analyzed were selected based on variation between either cervical and rectal isolates of serovar G, or variation found between serovars G, E, and D. This approach was also used to confirm that each *ompA* genotype was consistent with the MOMP phenotype identified by immunofluorescence (data not shown). McCoy cells in six well trays were

infected with cloned clinical isolates and chlamydiae were grown for 48 h. Genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue kit following the manufacturer's instructions, using an initial suspension buffer supplemented with 5 mM Dithiothreitol. Sequences in the polymorphic regions selected for SNP analysis were confirmed by traditional Sanger sequencing. The Fisher's exact test was used to determine any statistical association of identified SNPs with observed phenotypes. The alternate hypothesis in these statistical analyses stated that there was no correlation between genotype and phenotype at each of the tested loci. Statistical significance is reported as *P* values less than 0.01 or 0.001, and significance supports the hypothesis that differences at the tested locus correlate with tissue tropism to either the rectal or cervical site of infection.

Nucleotide sequence accession numbers

The *Chlamydia trachomatis* clinical isolate genome sequences sequenced at the Joint Genome Institute have been deposited at DDBJ/EMBL/GenBank under the given accession numbers [D(s)/2923: ACFJ00000000, E/11023: CP001890, E/150: CP001886, G/9301: CP001930, G/9768: CP001887, G/11222: CP001888].

The Whole Genome Shotgun project accession numbers for the strains sequenced using Illumina sequencing have been deposited at DDBJ/EMBL/GenBank under the given project accession numbers (J/6276: ABYD00000000, F/70: ABYF00000000, G/11074: CP001889).

RESULTS

Comparative genome analysis of sequenced chlamydial isolates

Pairwise alignment analysis using MAFFT (65, 66) of genomes from 12 recent clinical isolates demonstrated different levels of heterogeneity among strains (Table 3.1). The number of nucleotide substitutions between inter-serovar urogenital strains (not including L2 434/Bu) ranged from 6,494 (G/11222 versus

E/11023), to 1,638 substitutions (D/UW3 versus G/11222), and substitutions within serovars ranged from 1,287 (G/11074 versus G/11222) to three (G/9301 versus G/9768). The serovar G strains showing the highest level of similarity were cultured from the male urethra and male rectum of different patients, and their collection dates were separated by over 1 yr. A cervical non-fusogenic isolate, Ds/2923, was more similar to the sequenced serovars E and F isolates than to the published serovar D sequence (Table 3.1). A comparison of strains E/11023 and Ds/2923 identified 1,211 substitutions, while a comparison between D/UW3 and Ds/2923 identified 5,764 substitutions. The highest number of differences between genomes (8,811 nucleotides, 326 indels) was identified in a comparison of the publicly available sequenced ocular strain A/HAR-13 (25) and the publicly available sequenced lymphogranuloma strain L2-434/Bu. This represents a maximum level of variability of 0.87%.

The whole-genome phylogenetic tree shown in Fig. 3.1 indicated that there are at least two clades within our sequenced urogenital serovars, one group containing serovars D, G, and J, and a second group containing serovars E and F. These two groups were distinct from ocular strain A/HAR-13 and lymphogranuloma strain L2 434/BU. In this analysis, the genome of strain Ds/2923, which was originally serotyped by reactivity with serovar D-specific monoclonal antibodies, grouped in the clade with the serovar E and serovar F strains. These data are parallel to those shown in table 3.1 and confirm that the genome of Ds/2923 is more similar to serovars E or F than to the published serovar D genome.

Mapping of SNPs and larger changes in sequenced genomes of clinical isolates

Diffsort was used to determine the number of substitutions found per ORF between selected clinical isolates. These studies were undertaken to determine if variation is localized to specific regions within the chlamydial genome, and to identify possible loci containing variable regions between or within serovars (Fig. 3.2). These analyses identified specific regions throughout the genome that

contained higher levels of variation compared to the overall genome. These regions included possible recombination targets: ORFs CT049 to CT051 (46), plasticity zone ORFs CT144 to CT176, *ompA*, and the *pmp* genes (18, 24-26, 39, 91, 125, 126).

Similar pairwise analysis of clinical isolates identified a variety of insertions or deletions (indels) in the different strains (Fig. 3.3). The serovar G rectal strains encoded a 430 nucleotide insertion in the CT154 gene, which leads to an N-terminal truncation of CT154 and a putative new ORF (CT154.1). A second change found in the serovar G rectal strains was an in-frame 111 nucleotide insertion in gene CT326. These polymorphisms were limited to the sequenced serovar G rectal isolates- the cervical serovar G isolate (G/11222) contained neither of these insertions. Both rectal and cervical serovar E isolates lacked the insertion in CT154, but carried the 111 nucleotide insertion found in CT326. The structure of CT326 was complicated by a 25 nucleotide deletion in serovar E strains, relative to D/UW3. This deletion led to a truncated N-terminal CT326 and a C-terminal CT326 ORF (Fig. 3.4). Serovars J/6726, and previously sequenced strains L2-434/Bu and A/HAR-13 also contained the 111 nucleotide insertion in CT326, as well as similar but unique insertions in the CT154 region (Fig. 3.4). Consistent with work by Carlson *et al.* (25), insertions and deletions located in ORF CT456 (encoding Tarp) in several of the sequenced isolates.

Evidence for recombination in genomes of sequenced clinical isolates

The apparent genomic similarity of isolate Ds/2923 to genomes of serovar E and F strains led us to hypothesize that regions other than *ompA* might show evidence of recombination in this strain. A BLAST-based similarity approach using sliding windows of 1,000 nucleotides across the entire genome was used to uncover additional regions of recombination. These analyses confirmed that the majority of the Ds/2923 genome is similar to the genomes of serovars E and F isolates (Fig. 3.5, panel A), while *ompA* and nearby sequences are most similar to D/UW3. Fine mapping of these regions in Ds/2923 demonstrated that the apparent upstream crossover point adjacent to *ompA* is in the *rs2* gene (CT680),

at a position previously described as a hot spot for recombination in chlamydial genomes (46). The downstream crossover point for this recombination event is located within *ompA*, and results in a hybrid MOMP protein with variable domains from different serovars (133). These studies also uncovered several additional regions exhibiting higher similarity to the genome of D/UW3 than to E or F strains. The clearest examples of this are represented in ORFs CT171-CT183 and ORFs CT360-CT388 (bracketed, Fig. 3.5, panel A). Differences between genomes in these regions included SNPs, indels as large as 308 nucleotides (CT171), and ORF fusion events. These data support the conclusion that these regions were involved in recombination events between chromosomes, leading to the mosaic Ds/2923 genome.

To determine if genomes of other sequenced isolates exhibited a similar mosaic structure as with seen in Ds/2923, the BLAST similarity analysis was performed on isolate F/70 (Fig. 3.5, panel B). For this experiment, the genome of F/70 was removed as template in the E/F clade used for analysis of Ds/2923, and instead, used as probe of the remaining genomes. This analysis uncovered a different set of regions of apparent recombination than were observed in Ds/2923. One of these loci (ORFs CT153 - CT166) included the plasticity zone. This region also contains sequence homologous to TC0438 (*tox*) found in *C. muridarum* (117). Another possible site of exchange in F/70 included ORFs CT859 through CT868. This region is upstream of the *pmp* genes (CT869 - CT872, CT874), and has previously been hypothesized as a locus of lateral gene transfer in *C. trachomatis* (45, 47).

Correlation of identified SNPs with tropism to the rectal or cervical site of infection

To determine if polymorphisms found between these sequenced isolates were associated with tropism to the rectal site of infection, loci with varying degrees of differences were selected for PCR or RFLP analysis. This included genomic regions with nucleotide substitutions (i.e. CT144, CT869, CT870), some with indels that affected predicted protein length (CT154 and CT326, Fig. 3.4) and

some with single nucleotide polymorphisms that affect open reading frame length (CT158 and CT159). Analyses were conducted on a collection of clinical isolates representing serovars E, G, and J that were collected from the cervix, male rectum, or male urethra. While, in most cases, there was no apparent or statistical difference between a particular SNP and a tissue tropism, there were ORFs in which correlations were observed. Insertions in CT154 and CT326 ($P < 0.001$ or $P < 0.01$, respectively; Fisher's exact test) were associated with rectal tropism within serovar G (Fig. 3.4 and 3.6). Statistical analysis on these regions in serovars E and J indicated no association with genotype and tropism, but it is possible that some of these polymorphisms might reach significance (i.e. combinations of CT144, CT154, CT158, CT159, CT326 in serovar J; *pmp* genes in serovars E and J) if higher numbers of isolates were investigated. Because a high proportion of our samples were collected in the Seattle area, there was a possibility that we were examining a serovar G population that was restricted with regards to social network or geographic clustering. To address this concern, we included a limited number of rectal isolates of serovar G collected in Alabama and Peru. Each of these isolates (N = 4) contained the genotype associated with rectal infection for both ORFs CT154 and CT326. No significant association was found between polymorphisms in CT869 and CT870 and tropism for serovar G strains, but a comparison of rectal versus cervical sites of infection across all isolates identified a significant association between selected polymorphisms in these two ORFs (Fig. 3.6, panel C, $P < 0.001$).

While these analyses demonstrated that certain SNP patterns were associated with tropism to either the cervical or rectal site of infection, the overall structure of the chromosomes were mosaics of a variety of different SNP patterns. The mosaicism observed is most apparent in the cervical isolates of serovar G shown in Fig. 3.6, but is also evidenced in the serovar G rectal isolates (i.e. strains Gr-15P, Gr-16P, GR-30, GR-30AB). This mosaicism also supports the conclusion that the analyses in MSM populations did not simply identify clonal expansion in a geographically restricted area or a closed social network.

DISCUSSION

The study of chlamydiae has greatly benefitted from advances in genome sequencing technology. These bacteria contain small (~one megabase), largely syntenic AT-rich genomes, with very few repeat regions and almost no genomic islands. We (117) and others (25, 63) have explored the possibility that genome sequence analysis can be used to characterize functional roles of individual genes in this system, particularly when this system lacks practical genetic tools.

In the current study, we used genome sequencing and PCR-based analyses of polymorphisms to examine chlamydial recombination in clinical isolates, and to develop a technology for correlating chlamydial genotype with clinical phenotype. Our sequenced genomes were collected at Seattle/King County Sexual Health Clinics, and their isolation dates ranged from 10/1994 to 8/2005. Initial genome sequence data demonstrated a maximum sequence divergence of 0.87% between clinical strains, and a minimum difference of three substitutions and a single nucleotide indel. The two strains showing this high degree of similarity were serovar G isolates collected from the rectal (G/9768) and male urethral (G/9301) site of infection, with over a year separating the collection dates (11/2001 and 05/2000, respectively). The largest insertion in any sequenced clinical strain was 4,668 base pairs, found in our previously sequenced serovar F/70 and shown to be variable in serovar J strains (25) (117). The overall level of variation observed among *C. trachomatis* strains is similar to the variation seen when comparing members within the same genomic group of the obligate intracellular bacteria *Coxiella burnetii* (13), as well as comparisons between *Rickettsia rickettsii* isolates (38).

There are technical issues associated with generation of precise SNP counts between different strains. For example, we found 3,696 substitutions between published strains D/UW3 and A/HAR13, a number that is slightly higher than that determined by Carlson *et al* (25). These differences can be attributed to the

specifics in the programs used to count SNPs. One example of this is the settings used in the MAFFT to generate alignments for determining substitution counts and indel events. The default MAFFT settings, which were used in our study, result in a higher penalty for inserting a gap than extending a gap. Adjusting the settings for lower gap insertion penalty and higher extension penalty resulted in slightly fewer substitutions at the cost of increasing the number of indel events. Such differences in the analysis programs will lead to slight differences in count determinations, but the overall relationships among genomes will be conserved.

Phylogeny-based characterization of the genomes demonstrated two clades in our sequences; one clade containing serovars E and F, and the second clade containing serovars D, G, and J. These relationships are in agreement with previous work where phylogenetic analysis was performed, either by using a set of housekeeping genes for analysis (88), or with previous comparative studies of chlamydial phylogeny (25, 26, 91).

A BLAST-based analysis was used to examine if recombination is common in chlamydiae and could be reflected in the collected genomes. Strain Ds/2923 is a cervical isolate that was our original example of an IncA-negative, non-fusogenic strain (118), and its genome showed the best evidence for recombination in clinical strains. This strain, similar to all other characterized IncA-negative strains, contains a lesion in IncA that is correlated with the inactivated *incA* open reading frame, and inclusions formed by these strains do not fuse with inclusions formed by either IncA-negative or –positive *C. trachomatis* (118). Although serovar-specific monoclonal antibodies identified Ds/2923 as a serovar D strain, our analyses demonstrated that most of its chromosome is more similar to the genomes of serovar E or F strains. A set of ORFs that may have been recombination targets in this strain include *ompA*, the gene encoding MOMP, the major serovariant antigen in the chlamydiae. Possible recombination events involving *ompA* are not unique to this work. Early in the analysis of chlamydial gene sequences, hybrid *ompA* coding sequences

were identified by several investigators (20, 73, 80, 133) , and recent work by Gomes *et al.*, (46) demonstrated that a recombination hotspot is found just upstream of the *ompA* coding sequence. The Ds/2923 chromosome had an apparent recombination event exactly at the position identified by Gomes and colleagues. The downstream recombination event occurs between variable domains 3 and 4, also consistent with work of others who have identified *ompA* sequences encoding mosaic MOMPs with variable domains from different prototype serovars (73, 133).

Studies of possible recombination sites in Ds/2923 were expanded via a BLAST-based similarity analysis of sequenced genomes. These experiments identified additional regions in the Ds/2923 chromosome that were targets for recombination between strains. These candidate recombination loci involve several sites, including the plasticity zone (15, 24, 25, 39, 91, 125, 126) and ORFs CT360-CT389, a region encoding a set of metabolic pathways and hypothetical genes. Both of these genomic regions in Ds/2923 are more similar to the prototypic serovar D sequence than to our serovar E or F sequences. ORFs CT360-CT389 includes *aaxABC* (CT372 to CT374, respectively), which encode proteins that participate in an arginine-agmatine exchange system (106). Recent work by Giles *et al.* used an *E. coli* system to show that polymorphisms in *aaxB* led to inactivation of this gene in serovars D/UW3 and L2, and that function might be restored in D/UW3 by an $R^{115}G$ replacement (44). Genomic analysis demonstrated that isolates E/150, E/11023, contained this $R^{115}G$ replacement. Therefore, it is possible that there exists a phenotypic discrepancy at this locus, with serovar E and F strains being “wild type”, while D/UW3 and the mosaic Ds/2923 being deficient in this exchange pathway.

A subsequent analysis with the genome of strain F/70 identified similar examples of likely recombination events, but at different loci. These data are consistent with in vitro analyses by Demars *et al.* (33, 34) and our laboratory (120), and support the hypothesis that recombination is very common and involves many different locations across the chlamydial chromosome.

Our second hypothesis addressed the concept that genome sequencing of clinical strains could identify and help characterize genes and gene products important in the biology of the pathogen in vivo. Pioneering work in this area was conducted by Caldwell *et al*., who used sequence analysis of a limited number of ORFs to correlate the presence or absence of *trp* synthesis genes with ocular or genital tropism in *C. trachomatis* (23). We chose a different clinical phenotype for analysis- tropism to the rectal site of infection- as a trait for study. Our analyses identified four loci that were statistically associated with a particular tropism, only one of which – CT869-870 – was associated with rectal tropism across all serovars ($P < 0.001$). These ORFs encode two Pmp proteins, members of a family of chlamydial autotransporters important in different aspects of chlamydial biology (29, 68, 124, 130). The nucleotide variation found in these genes lead to amino acid changes that are not randomly distributed across the coding sequence, indicating the variable regions may be part of domains important in chlamydial biology. The variation and molecular evolution of these *pmp* genes is an active area of research (82), but any possible function of these different Pmp variants in attachment or intracellular development remains to be investigated.

Alterations in three other ORFs were statistically correlated with tropism in serovar G only. CT144 encodes a hypothetical 285 AA protein that varies at 14 AAs in the tested strains. Eleven of the 14 AA changes in CT144 were clustered to a 26 AA region of the gene product (data not shown). Further study of CT144 might determine if this variable domain may play a role in tissue tropism or pathogenesis. ORF CT154.1 is the result of a 430 nucleotide insertion that is also found in the genomes of serovars A and L2 (Fig. 3.4), and encodes a protein with no predicted function. Finally, the presence of a 111 nucleotide in-frame insertion in uncharacterized hypothetical gene CT326 was statistically correlated with tropism to the rectal culture site. Our association of these ORFs with tissue tropism is complicated by the fact that none of these ORFs were statistically associated with rectal tropism in strains of serovar E or serovar J. It is possible that production of these different proteins leads to phenotypic differences in the

context of serovar G, facilitating the apparent tissue tropism. Alternatively, it is possible that these SNP differences are within or linked to regions of the genome that encode unidentified proteins that collectively are important for this tropism. Possible functions for candidate proteins identified in these studies are currently being explored in our laboratory.

ACKNOWLEDGMENTS

Drs. Connie Celum and Will M. Geisler are gratefully acknowledged for supplying chlamydia collected from individuals in Lima, Peru and Birmingham, AL. Drs. Paul Richardson and Alla Lapidus are acknowledged for the genome sequencing and assembly conducted at the Joint Genome Institute. We thank Kelsi Sandoz for advice and consultation regarding PCR and RFLP analysis. Sara Weeks is acknowledged for technical assistance and editing of the manuscript. This research is supported by grants from National Institutes of Health (A148769, A1031448).

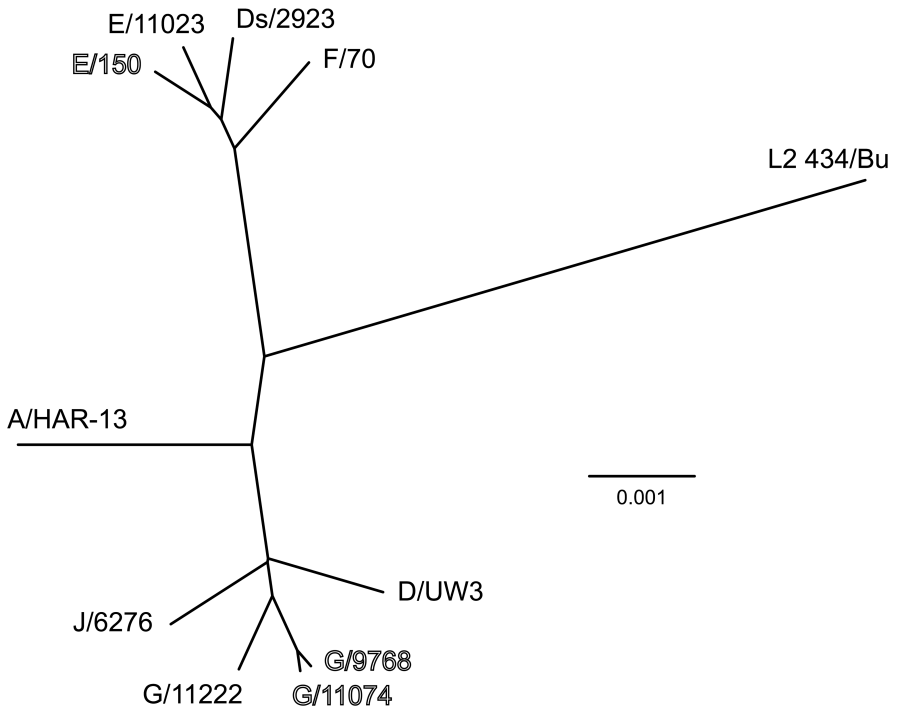


Figure 3.1. Phylogenetic analysis of sequenced chlamydial isolates based on whole genome alignment.
The tree was constructed using a modified UPGMA algorithm as part of the MAFFT alignment program. Strains collected from the male rectum are indicated with outlined text, and cervical isolates are standard text. The branch lengths are proportional to genetic distance between isolates.

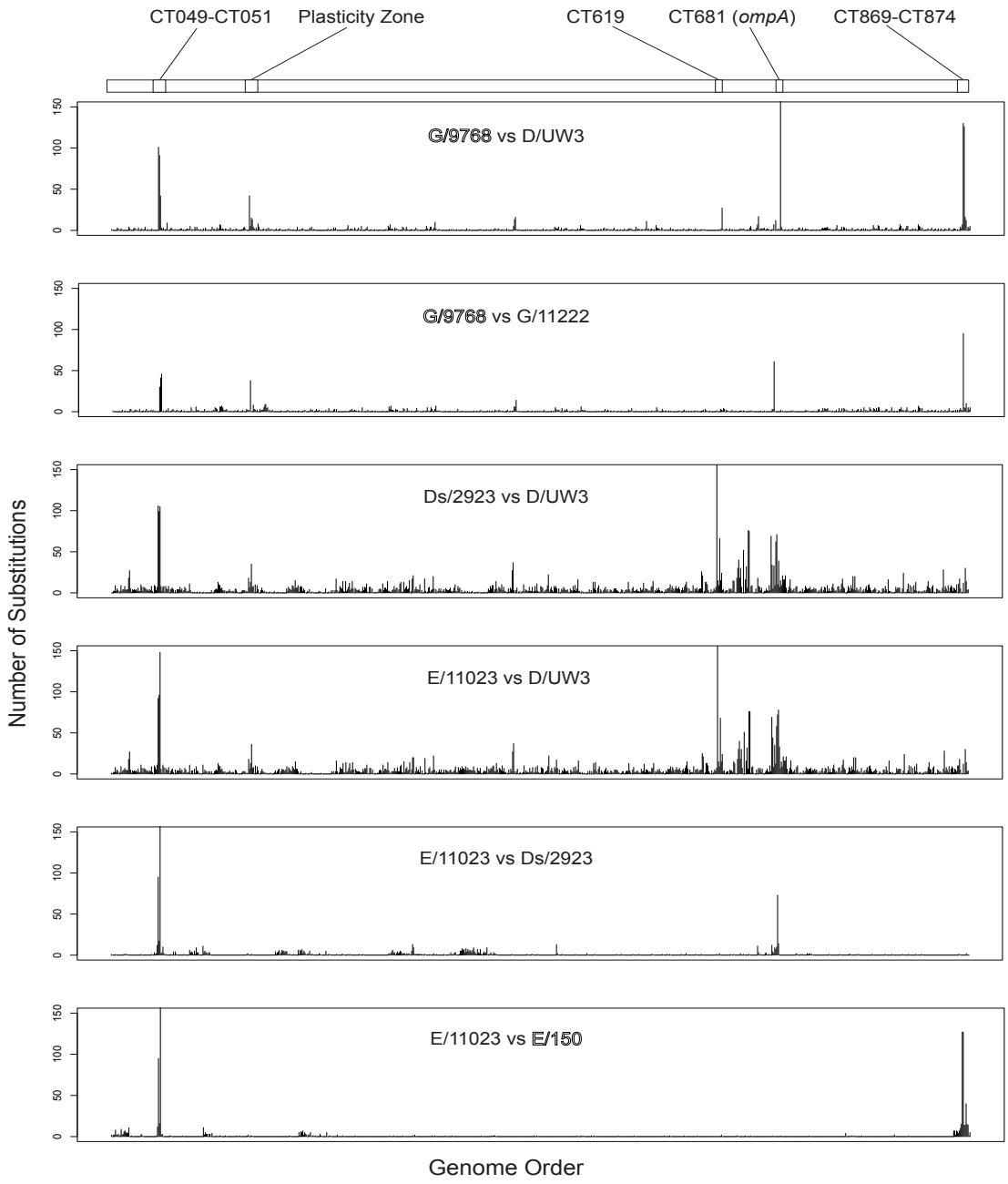


Figure 3.2. Number of substitutions per open reading frame between strains.

Each graph represents a Diffseq-generated whole genome alignment between the indicated strains. The genome order is represented on the X-axis, beginning with CT001, using the sequenced strain D/UW-3 gene designations as a reference. The numbers of substitutions found per ORF are represented on the Y-axis. Strains collected from the male rectum are indicated with outlined text, and cervical isolates are standard text.

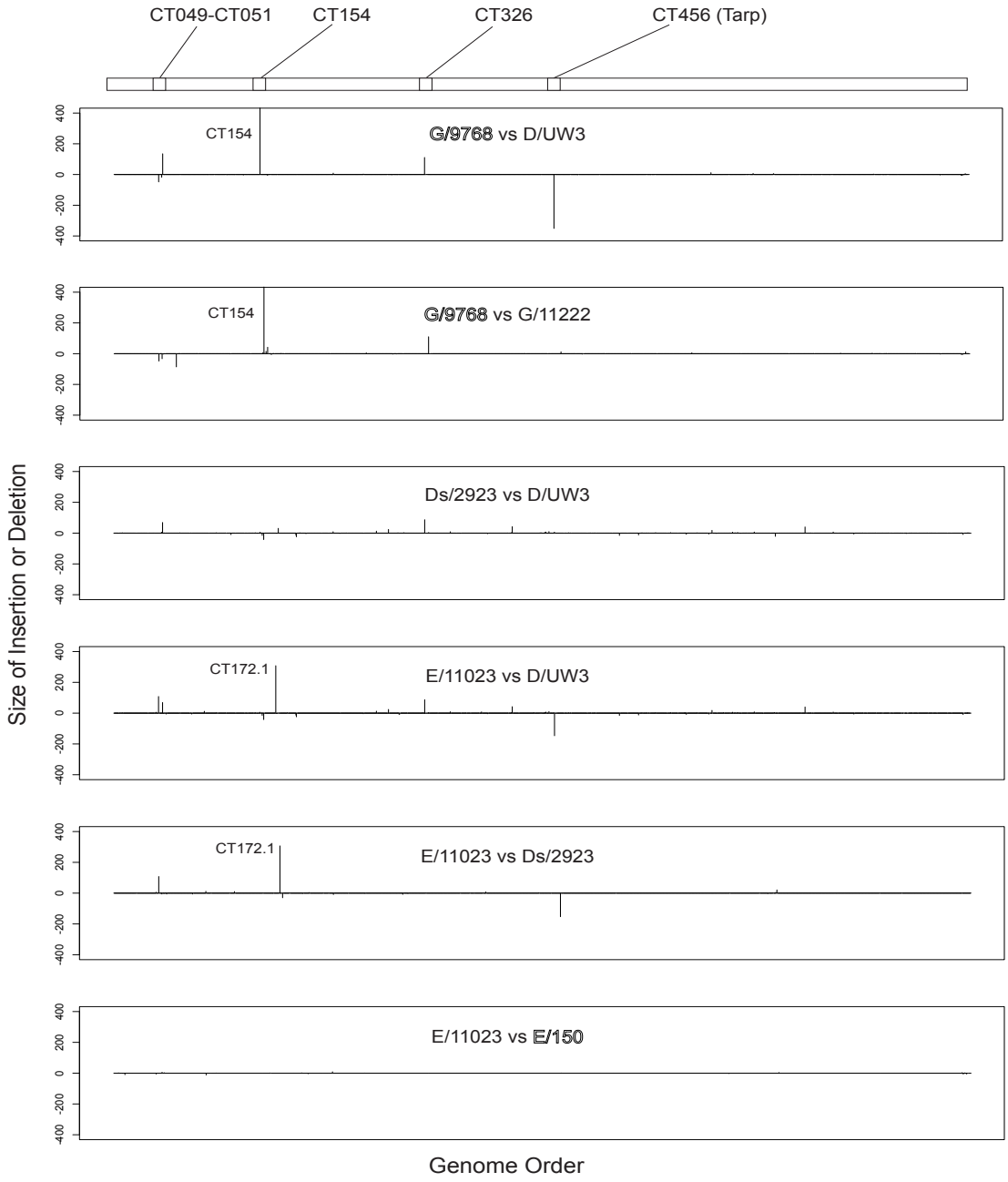


Figure 3.3. Location of insertions or deletions between strains.
The comparisons are identical and in the same order relative to those in Figure 2. The genome order is represented on the X-axis, beginning with CT001, using the sequenced strain D/UW-3 gene designations as a reference. The Y-axis is a measure of length for each insertion (above the X-axis) or deletion (below the X-axis). The strain listed first in graph title has an insertion or a deletion at a given site relative to the strain listed second. For example, G/9768 has a 437 base pair insertion in CT154 relative to D/UW3. Strains collected from the male rectum are indicated with outlined text, and cervical isolates are standard text.

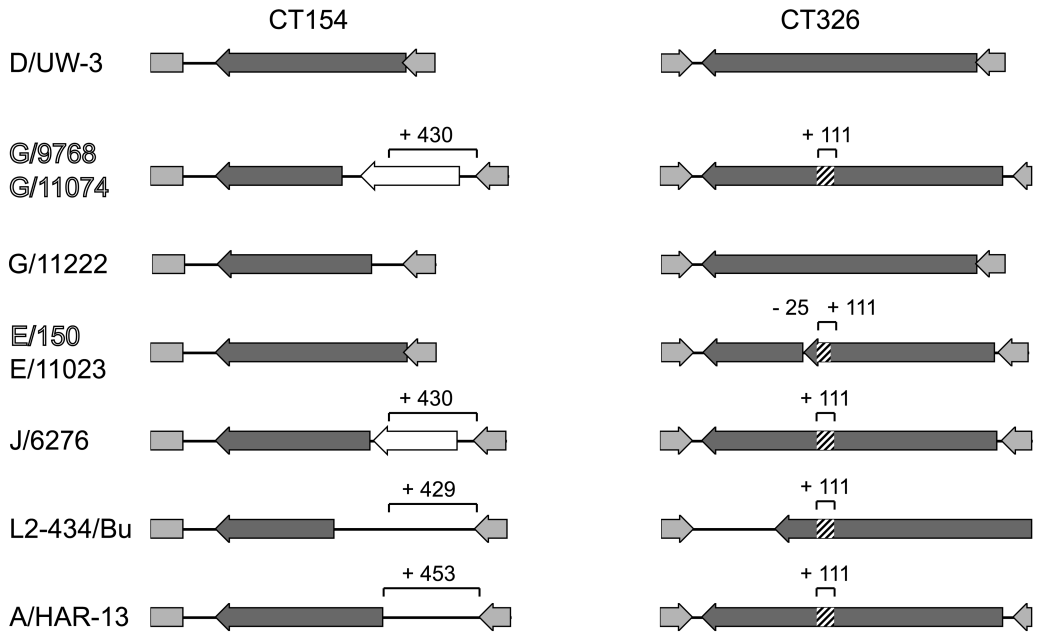


Figure 3.4. Schematic diagram of open reading frames CT154 and CT326 in several sequenced C. trachomatis strains.
The isolate labels are represented at left. Strains isolated from the male rectal site of infection are represented by outlined text, and the cervical isolates are shown in standard text. The indicated ORF is shown in dark grey, with lighter grey colors indicate flanking ORFs. Diagonal hash marks represent the location of insertions into the gene. Putative new ORFs resulting from an insertion are represented in white. The numbers above the brackets indicate length in base pairs of insertions relative to the sequenced D/UW-3 sequence.

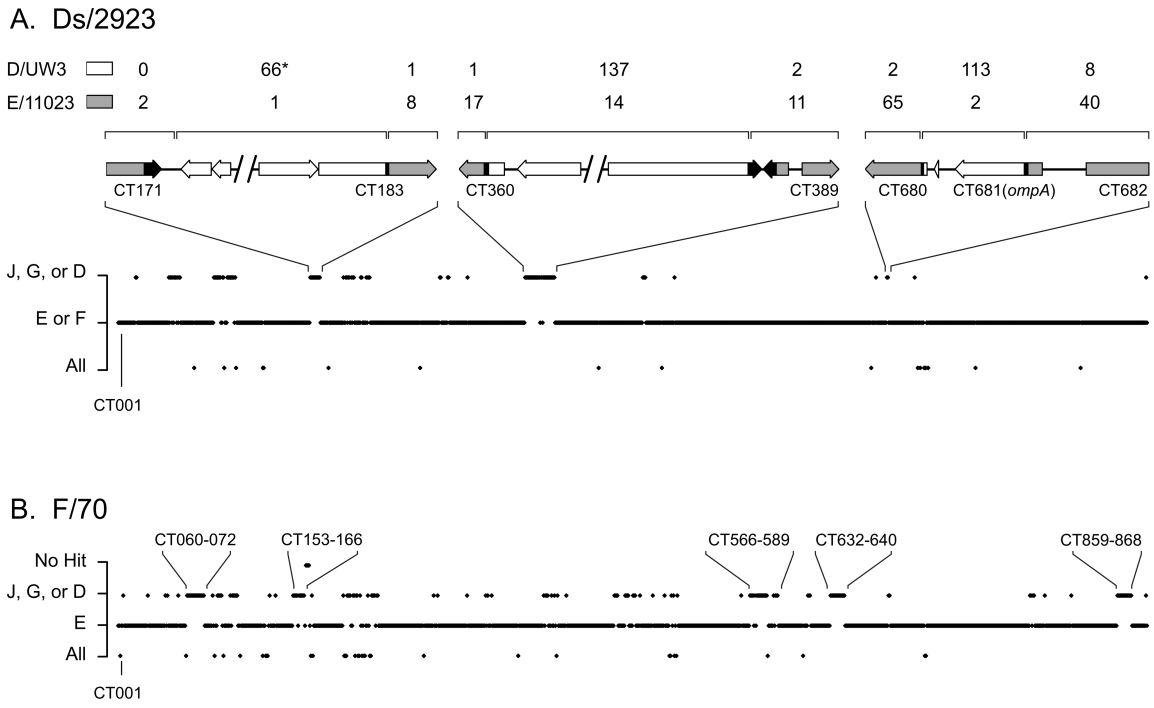
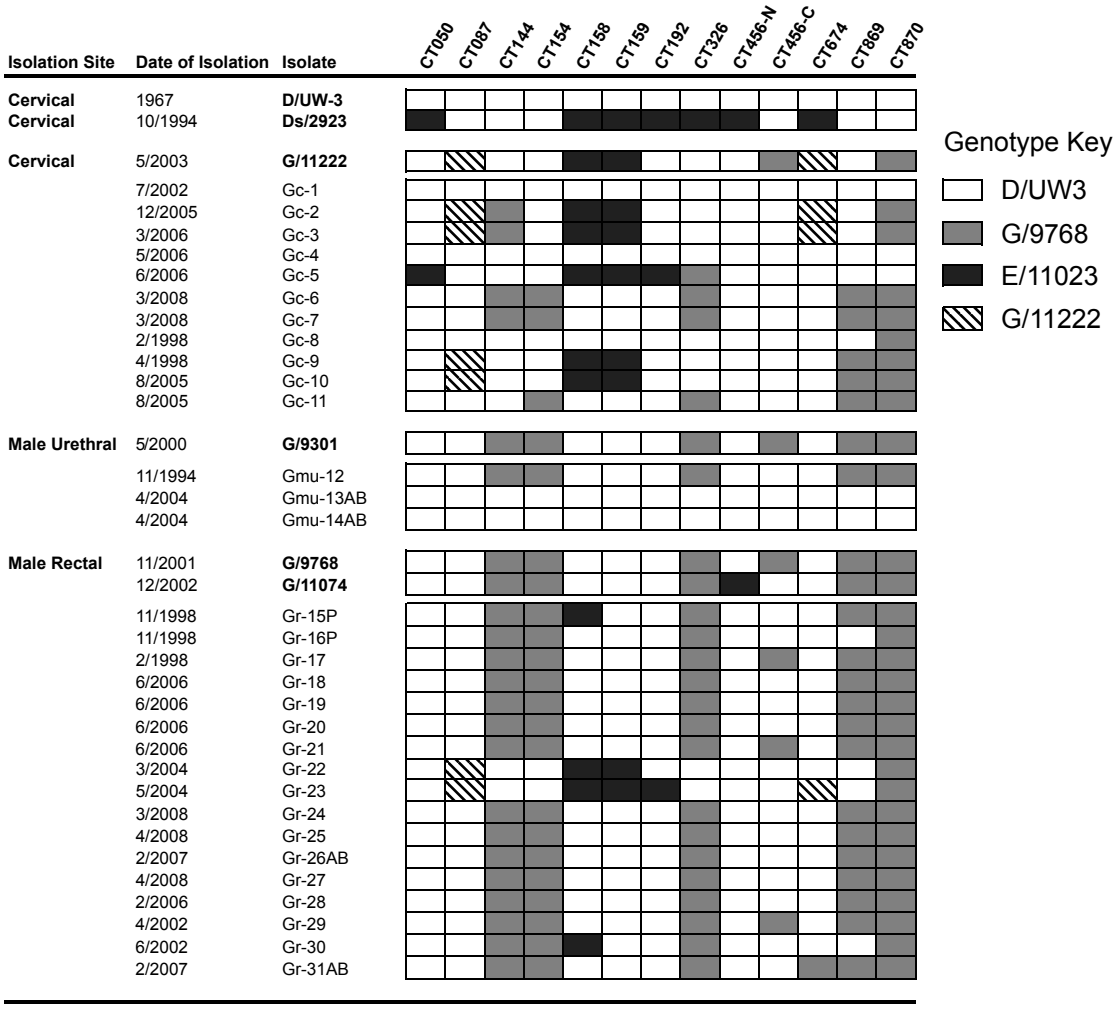
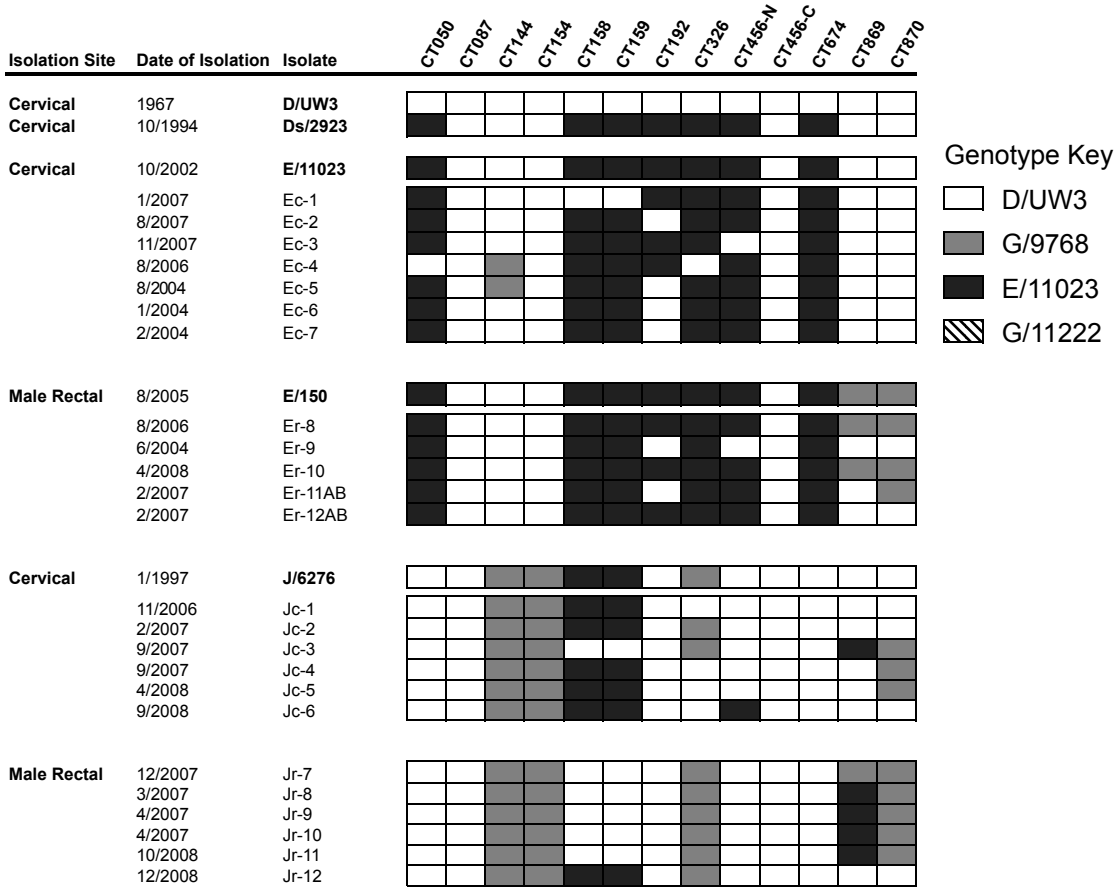


Figure 3.5. BLAST-based evidence for recombination in clinical C. trachomatis genome sequences.
Complete genome sequences for strain Ds/2923 (panel A) or strain F/70 (panel B) were used as probe against a genome database containing sequences of D/UW3, G/9768, J/6276, E/11023, and (in panel A) F/70. The complete genome sequence for each probe was evaluated using BLAST analysis of a 1 kbp sliding window, and each window was given a score based on whether it was more similar to, 1: a clade containing serovar J, G, or D, 2: a clade containing serovar E or F, or 3: if it was equally similar to all the genomes. The results for each window are plotted along the X axis, beginning with ORF CT001 as defined for the D/UW3 genome (109). All ORF designations in each panel are also based on the D/UW3 genome sequence. Panel A includes ORF maps showing expanded views of three regions showing apparent recombination. In the ORF maps, a gray box indicates that sequence is more similar to serovar E/11023, white boxes indicate the sequence is more similar to D/UW3, and black boxes indicate regions where apparent recombination events likely occurred. The brackets above each gene cartoon indicate the region from which informative sites were examined and counted. The numbers above the brackets reflect the number of shared informative sites between the Ds/2923 and D/UW3 sequences (top line) or the Ds/2923 and E/11023 sequences (bottom line). The asterisk indicates that one of the 66 shared informative sites is a 308 nucleotide deletion that both D/UW3 and Ds/2923 share relative to the E/11023 sequence. The bracketed areas in panel B show selected regions of the F/70 genome that are more similar to genomes in the J/G/or D clade than to the serovar E genome sequence.

A


CT050
CT087
CT144
CT154
CT158
CT159
CT192
CT326
CT456-N
CT456-C
CT674
CT869
CT870
Isolation Site
Date of Isolation
Isolate
Cervical 1967 D/UW-3
Cervical 10/1994 Ds/2923
Cervical 5/2003 G/11222
7/2002 Gc-1
12/2005 Gc-2
3/2006 Gc-3
5/2006 Gc-4
6/2006 Gc-5
3/2008 Gc-6
3/2008 Gc-7
2/1998 Gc-8
4/1998 Gc-9
8/2005 Gc-10
8/2005 Gc-11
Male Urethral 5/2000 G/9301
11/1994 Gmu-12
4/2004 Gmu-13AB
4/2004 Gmu-14AB
Male Rectal 11/2001 G/9768
12/2002 G/11074
11/1998 Gr-15P
11/1998 Gr-16P
2/1998 Gr-17
6/2006 Gr-18
6/2006 Gr-19
6/2006 Gr-20
6/2006 Gr-21
3/2004 Gr-22
5/2004 Gr-23
3/2008 Gr-24
4/2008 Gr-25
2/2007 Gr-26AB
4/2008 Gr-27
2/2006 Gr-28
4/2002 Gr-29
6/2002 Gr-30
2/2007 Gr-31AB
Genotype Key
D/UW3
G/9768
E/11023
G/11222

B



C

Statistical Comparison	CT050	CT087	CT144	CT154	CT158	CT159	CT192	CT326	CT456-N	CT456-C	CT674	CT869	CT870
Serovar G rectal versus cervical			*	**				*					
Serovar E rectal versus cervical													
Serovar J rectal versus cervical													
Global rectal versus cervical				*				**				**	**

Figure 3.6. PCR- and RFLP-based genotype analysis of variable regions in clinical isolates.
The strain, date of isolation, and the site of isolation are given at left of figure. Panel A includes data from serovar G isolates. Panel B includes data from serovars E and J. Panel C shows the results of statistical analyses with the Fisher's exact test, where one asterisk indicates $P < 0.01$, and two asterisks indicate $P < 0.001$. Isolates indicated in bold are those for which complete genome sequences were examined. Isolate labels ending in the letter P were collected from patients in Lima, Peru, and isolate labels ending in AB were collected from patients in Birmingham, Alabama. All other isolates were collected from patients in King County, WA area family planning clinics. CT designations at the top of the figure represent gene targets for the PCR and RFLP analysis. CT456-N and CT456-C refer to variable regions that are present at the amino and carboxy terminus of CT456. The color of the boxes within the figure refers to the detected genotype as follows. The color-coding for each SNP is as follows: White boxes represent the D/UW-3 genotype. Grey boxes represent the rectal serovar G genotype. Black boxes represent the serovar E genotype. Hashed boxes represent the cervical serovar G genotype.

Table 3.1. Pairwise analysis of the number of substitutions and insertion/deletion events identified in genome sequences from a collection of *C. trachomatis* isolates.

Substitutions are indicated on lower left half of matrix and indel events are indicated on upper right half of the matrix. Genome size in base pairs is indicated below the strain names in the column at left.

Isolate	Genome size (bp)	No. of substitutions or no. of indel events[a]										
		D/UW3	A/HAR13	L2-434/Bu	Ds/2923	E/11023	E/150	G/9301	G/9768	G/11222	G/11074	J/6276	F/70
D/UW3	1,042,519		157	297	187	210	210	93	93	78	91	102	244
A/HAR13	1,044,459	3,696		326	251	254	254	151	151	169	152	172	261
L2-434/Bu	1,038,842	8,163	8,811		319	309	314	310	310	309	308	312	318
Ds/2923	1,042,757	5,764	7,131	7,687		75	78	210	210	197	207	203	150
E/11023	1,043,025	6,135	7,143	7,737	1,211		54	231	231	220	228	235	118
E/150	1,042,996	6,138	7,027	7,766	1,325	1,130		205	205	197	203	225	135
G/9301	1,042,811	1,892	3,523	8,162	6,205	6,485	5,869		1	69	5	92	250
G/9768	1,042,810	1,893	3,524	8,163	6,206	6,486	5,776	3		69	6	92	250
G/11222	1,042,354	1,638	3,778	8,131	6,063	6,494	5,907	1,286	1,287		70	88	249
G/11074	1,042,875	1,893	3,529	8,153	6,207	6,490	5,874	41	42	1,287		90	247
J/6276	1,043,181	1,899	3,540	8,264	5,878	6,467	6,292	1,918	1,919	1,770	1,920		252
F/70	1,048,006	6,287	6,913	7,738	2,239	1,776	2,407	6,305	6,306	6,236	6,310	6,441	

[a] The numbers of substitutions are indicated on the lower left, and the numbers of indels are indicated on the upper right.

Genome sequence analysis of recombinants between clinical *Chlamydia trachomatis* strains identifies genetic regions associated with known phenotypes

Brendan M. Jeffrey, Robert J. Suchland, Steven G. Eriksen, Kelsi M. Sandoz, and Daniel D. Rockey

Submitted to *Molecular Microbiology*

Wiley-Blackwell publishing

ABSTRACT

Chlamydiae are refractory to introduction of foreign DNA, and techniques for generating isogenic deletion mutants are just being developed. Therefore, much of the understanding of chlamydial biology is still in its infancy. Pregenomic and postgenomic studies, however, demonstrate that chlamydiae are actively recombinogenic in vitro and in vitro. The molecular and cellular biology of this process is not well understood. In this study, we created and genome sequenced a set of 12 *C. trachomatis* recombinants, with a goal of exploring both the process of genetic exchange in chlamydia, and to correlate well-established *C. trachomatis* phenotypes with parental genotypes. Genome sequence analysis of the recombinants revealed no putative motifs that were present at target sites, and each example of gene integration was the product of a classical homologous recombination event. There was a single deletion in one recombinant progeny, and this was a product of homologous recombination deleting 17.1 kilobases between two rRNA operons. Recombination sites identified in this study did not correlate to previously identified hotspots in clinical strains, and we suggest that identification of such hotspots in clinical isolates represents a measure of selection in vivo as opposed to targeted recombination. Cell culture studies of different *C. trachomatis* serovars have revealed several phenotypes that differ between *C. trachomatis* strains. We next examined whether comparative genomics, could be used to identify genetic loci associated with observed phenotypes. These analyses identified several possible loci that were associated with differential attachment rates in the absence of centrifugation and the formation of secondary inclusions within infected cells. This work demonstrates that the generation of hybrid chlamydial genomes in combination with comparative genomics is an effective approach for detecting genetic loci associated with selected phenotypes, in an organism that is not amenable to routine genetic manipulation.

INTRODUCTION

Chlamydia trachomatis is a gram-negative obligate intracellular bacterium that is the leading cause of preventable blindness worldwide and the most common sexually transmitted infection (STI) in the United States (78). As is true for many chlamydial species, *Chlamydia trachomatis* isolates are differentiated into distinct serovars based on antibody specificity to the major outer membrane protein (MOMP), encoded by *ompA*. These serovars fall into three groups, those associated with trachoma (serovars A, B, and C), those associated with non-invasive sexually transmitted infections of the urogenital tract (serovars D though K), and those associated with invasive lymphogranuloma (LGV; serovars L1 to L3) (101). This historical classification system has recently been modified to a genotypic characterization of strains, both by sequencing of *ompA* and the inclusion of a variety of other markers in the analysis (69). Nevertheless, many of the biological differences among chlamydiae still can be grouped by the serovar-based classification scheme. Biological differences within the chlamydiae include host tropism, aspects of the disease process associated with different species of chlamydiae (i.e. human infections by *Chlamydia psittaci*, *C. pneumoniae*, and *C. trachomatis*) and disease processes caused by different serovariant strains within a species (i.e. trachoma, LGV, pelvic inflammatory disease, and cervicitis). Very little is known about how these different species or strains cause different diseases in the host.

Phenotypic variation among chlamydiae is also evident during in vitro culture. For some of these there is a known association between a particular genotype and the culture phenotype. An example of this is the requirement of tryptophan during growth, for which different strains have different complements of tryptophan utilization genes (24, 25). The genetic association with many other phenotypic traits is not clear. For example, different serovars have significantly different growth rates and increased efficiency of attachment and invasion (71) and the loci involved in these differences have not been identified. Our laboratories have shown that the serovar G and F isolates produce higher levels

of secondary inclusions and *incA* laden fibers in cell culture, compared to most isolates of serovars D, E, and J (119). Detailed analyses of these and other similar phenotypes are complicated by the relatively primitive molecular genetic techniques currently available for studying chlamydial biology.

Recent reports by our group and others showed that chlamydial isolates with differing drug resistance markers can be co-infected, and dual resistant recombinant progeny can be generated (34, 120). Additionally, evidence from pre-genomic and post-genomic analyses indicates that chlamydiae likely recombine in infected hosts (20, 45, 46, 62, 73, 80). *Chlamydia* spp. encode no homologs to bacterial gene transfer systems and, therefore, the mechanisms underlying recombination in this system remain very poorly characterized.

In the present study, a collection of genetically mosaic recombinant strains from parents with differing cell culture phenotypes were created, microbiogically cloned, and subjected to complete genome sequence analysis. These strains, the parentals used in the crosses, and selected clinical isolates were used to investigate the process of chlamydial genetic exchange, and to examine the association between phenotypic traits of clinical strains and strain genotype.

MATERIALS AND METHODS

Chlamydial strains and selection for resistance

Drug resistant *Chlamydia trachomatis* strains J/6276^{rif}, RC-J/6276$^{tet\text{-}rif}$, F(s)/70^{rif}, F(s)/70$^{tet\text{-}rif}$ and L2-434^{ofl} used for recombination experiments were generated as previously described (120). Briefly, strains were grown in McCoy cells at a multiplicity of infection (MOI) of approximately 1.0 in media containing sub-inhibitory concentrations (equivalent to half the minimum inhibitory concentration (MIC) of the appropriate drug. Serial passages of these strains were grown in the media containing desired antibiotics until resistant mutants emerged or until passage was completely negative. Some strains required several attempts until resistant mutants were isolated. Isolates were then cloned by a twofold dilution

method. The resulting cloned elementary bodies were grown to high titers and were partially purified by centrifugation of lysates of infected cells through a 30% MD-Gastroview® (Mallinckrodt Inc. St Louis) pad.

Generation of recombinant chlamydial strains

Recombinants isolated for genome analysis were generated from two sets of crosses. One involving two parental strains; L2-434^{ofl} and F(s)/70^{rif} and the other involving a three-way cross with the parental strains F(s)/70$^{tet\text{-}rif}$, J/6276^{rif} and L2-434^{ofl}. Recombination experiments were conducted as previously described (120). Briefly, crosses were performed in McCoy cells seeded in sets of individual shell vials (12 mm^2). The monolayers were then infected with different combinations of drug-resistant strains each at a multiplicity of infection of approximately 2.0, ensuring infections of cells with both strains. Cultures were incubated for 48 hours post-infection in the absence of antibiotics and were then harvested using the freeze-thaw method. Potential recombinants were selected by inoculating 50 µl of the freeze-thaw lysates from each shell vial onto a new shell vial monolayer and overlaying with a medium containing antibiotics at 4 times the MIC for each resistant parental strain. In the case of the three way cross (F(s) X J X L2), three different combinations of drug were applied to the infected monolayers. Combinations included ofloxacin/rifampicin, ofloxacin/tetracycline and ofloxacin/rifampicin/tetracycline.

Recombination rate determination

Vials were inoculated as described for the generation and selection of recombinants. For each cross, a primary recombination vial was generated and cultured in the absence of selective drug. Forty-eight hours post infection, infected cells were lysed via freeze-thaw and the lysate from each primary recombination vial was diluted and inoculated onto 100 identical shell vials (propagation vials) containing monolayers of McCoy cells. Selecting antibiotics were added to the media in these vials, and were always present on cells in any subsequent infection. Because the manifestation of the recombinant phenotype took variable lengths of time to be visible, each of these 100 propagation vials

was serially lysed and passed in antibiotic-containing media up to three times. After each of these culture periods, for each of the 100 vials, 1/3 of the lysate was inoculated onto individual McCoy cell monolayers, cultured in selecting drugs for 48 h, and examined via immunofluorescence microscopy for growing chlamydia. Culture-positivity in any of these microscopic analyses was considered indicative of a recombined chlamydial strain present in that lineage, and was quantitated as a measure of positivity in the original 100 propagation vials for that particular cross. Recombination rates were calculated by dividing the number positives from the original 100 vials with recombinants recovered by the total number of IFU generated from the primary recombination well.

Chlamydia recombinant strain genomic DNA preparation

Elementary body (EB) purification was conducted as previously described (119, 121). Purified EBs were incubated for 60 min with 4 units/mL RQ1 DNase (Promega) followed by treatment with 2 mM EGTA (RQ1 Stop solution, Promega) to inactivate the DNase. DNAse-treated EBs were suspended in Qiagen Genomic buffer B1 supplemented with Dithiothreitol (5 mM) and DNA was then extracted using the Qiagen Genomic Tip kit (Qiagen, Valencia, CA) following the manufacturer's instructions.

Genome sequencing and sequence analysis

Genomic DNA from recombinant strains was processed for Illumina-based paired-end sequencing using commercial DNA preparation kits (Illumina Inc., San Diego, CA) following the manufacturer's instructions. Each recombinant genome was first assembled using the reference-guided assembly program Maq (76) where both parental genomes were used as references. Regions in reference-guided assembled genomes where Maq could not resolve sequence were then compared to contiguous sequences assembled using de-novo assembly software Velvet (134) and a single contiguous draft sequence was produced.

Recombinant genome maps were produced by computationally parsing a whole genome alignment of the recombinant genome and the two parents used

to generate the recombinant. Whole genome alignments were produced using alignment program MAFFT with the default settings (65, 66). All crossover regions detected were manually analyzed using MacVector sequence analysis software (Cary, NC). Crossover regions were defined as the intervening homologous sequence between two informative sites (polymorphic base between the two parent genomes), where the recombinant base was the same as the first parent's base at a given informative site and the same as the second parent's base at the immediate upstream or downstream informative site.

To confirm clonality of the recombinant genomes, and to quality control our assembly process, two to four apparent crossover regions were amplified by PCR and sequenced using classical Sanger sequencing. In all cases the sequenced amplicon contained the appropriate informative sites from each parent involved in the cross (not shown).

The whole genome alignment produced by MAFFT of each recombinant and both parents was used to determine if any random mutations had occurred in the recombinant. A random mutation was defined as a nucleotide in the recombinant sequence that was different than the nucleotide of both parents at the same position in the alignment.

Recombinant strain phenotype analysis

To determine the attachment efficiency of parental and recombinant strains, McCoy cell monolayers were seeded in duplicate 24 well plates at 90% confluency, and triplicate wells of each plate were infected with identical inoculum with a target MOI of 1. Plates were either centrifuged at 2000 rpm for 1 hour or rocked for 1 hour, and wells were then washed 3 times with HBSS. DNA was extracted directly from each well using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA), where lysis buffer ATL supplemented with Dithiothreitol (5mM) was added to each well and sample was pipetted up and down 10 times to remove host cells and EBs. Genome copy number was determined for each treatment by qPCR, using probe specific for chlamydial Hsp60 (CT755). Attachment efficiency is defined as the genome copy number of

the rocked treatment divided by the genome copy number of the centrifuged treatment.

Genome-wide phenotype association analysis

Whole genome alignments were performed on all ten recombinants plus either the 3 parental genomes used (L2-434, Fs70, J6276) or with the 3 parental genomes and ten previously sequenced genomes (see accession numbers below). Whole genome alignments were performed using MAFFT with default settings (65, 66). Regions in alignment where MAFFT could not resolve alignment correctly were manually adjusted using alignment editor tools as part of MacVector sequence analysis software. Any position in the alignment where at least one genome had a variable base was further analyzed using the Fisher exact test as a metric to determine if the variable genotype could be associated with a given phenotype. The *p*-values calculated by the Fisher exact test were inverse Log-transformed and plotted using the statistical analysis program R (http://www.r-project.org/).

Polymorphic membrane protein family charge analysis

The annotated D/UW3 homologs of the Pmp proteins stored at Uniprot (http://www.uniprot.org/) were used to determine the auto-transporter domains of the L2-434, F(s)/70 and J/6276 Pmp proteins. Overall charge of the protein, not including the auto-transporter domain, at pH 7 was calculated with the Scripps ProteinCalculator (http://www.scripps.edu/~cdputnam/protcalc.html). The Kyte-Doolitle hydrophilicity plot tool as part of the MacVector sequence analysis software was used to determine which regions in the protein may be surface exposed. The charge program as part of the emboss bioinformatics suite (92), was used to calculate the difference in charge of individual amino acid changes in the Pmp proteins that were associated with the attachment phenotype. Pairwise clustalW (MacVector) alignments of the three parental Pmp sequences were used to count the number amino acid changes found between each parental sequence.

Genome sequence accession numbers

The genome sequences of the parental strains used to generate recombinant sequences and the previously sequence *C. trachomatis* strains used in the whole genome alignment studies are in the DDBJ/EMBL/GenBank database under the following accession numbers: D/UW3, AE001273; L2-434, AM884176; L2-UCH1, AM884177; D(s)/2923, ACFJ01000001; E/11023, CP001890; E/150, CP001886; G/9768, CP001887; G/11074, CP001889; G/11222, CP001888; F/70, ABYF01000001; F(s)/70, ABYG01000001, J/6276, ABYD01000001; J(s)/6276, ABYE01000001.

The *C. trachomatis* genome accession numbers of the recombinants used in this study have been deposited in the DDBJ/EMBL/GenBank database under the following accession numbers: RC-F/69, CP002671; RC-L2(s)/46, CP002672; RC-F(s)/852, CP002673; RC-J/943, CP002674; RC-J/953, CP002675; RC-L2(s)/3, CP002676; RC-F(s)/342, CP002677; RC-J(s)/122, CP002678; RC-J/966, CP002679; J/6276tet1, CP002680; RC-J/971, CP002681; RC-L2/55, CP002682.

RESULTS

Generation of recombinant strains

Our primary goal in this work was to use a set of strains with known phenotypic differences to characterize the genes associated with the phenotypes. To this end, a large collection of recombinant strains was generated, using parents of serovars J, F, and L2 (Table 4.1, Fig. 4.1). In some cases, crosses involved a pair of parents (i.e. crosses 1-6, 11,12), while in other cases, three way crosses were attempted (i.e. crosses 7-10). While progeny with each possible combination of two antibiotic resistance markers were routinely identified in the three-way crosses, no triply resistant strain was ever recovered from such a cross. Genome sequence analysis also did not identify any progeny strain of the three-way crosses that carried fragments of each of the three parents (Fig. 4.2).

Previous work by Demars and colleagues (34) demonstrated a recombination rate in their crosses of approximately 1.2 X 10^{-3} to 4 X 10^{-4}. This is slightly higher than the rate identified in our work for two biotype LGV parents (Table 4.2). We also used our current system to investigate a recombination rate in our crosses that involve both LGV and non-LGV parents. These assays are complicated by the observation that antibiotic resistance was differently manifested for many of the recombinants, leading to a need for multiple passages of some strains in the presence of selecting antibiotics (see experimental procedures). The recombination rate for each doubly resistant cross was approximately 10^{-6} (Table 4.2). Several of the crosses involved recombination between IncA-positive and IncA-negative parents (i.e. cross 2). There was no evidence for a difference in recombination rate in crosses that involved a single IncA-negative parent vs a cross that involved two IncA-positive (i.e. fusogenic) parent strains.

Genome sequence analysis of recombinant strains

The genomes of the twelve recombinant strains were sequenced using Illumina paired-end technology (Fig. 4.2). In all recombinant strains, the sequences surrounding the individual resistance markers were derived from the appropriate parental sequence, supporting the conclusion that these were recombinant strains and not strains that were mutated to antibiotic resistance in the selection process. Sequence comparison of each recombinant and its parents revealed a single random mutation in a recombinant strain [RC-L2(s)/3]. This mutation was a G (L2-434 sequence) to A [RC-L2(s)/3] transition substitution at position 293505, resulting in an alanine to valine amino acid change in the protein product of CT258. This same mutation was identified in the RC-J(s)/122 genome, a progeny of a cross in which RC-L2(s)/3 was a parent. There was no other evidence of random mutation in any other sequenced recombinant genome.

In every case except one, individual recombination events involved a classical double crossover homologous recombination event. The single event that varied from this pattern involved a crossover event that paired two rRNA

operons in one parent with a single rRNA operon in its recombination partner. This crossover was similar to that observed in the generation of the *C. trachomatis* L2Tet1 strain described in previous work (120). In the cross leading to L2Tet1, recombination between a *C. suis* strain carrying a tetracycline resistance locus involved the integration of adjacent ribosomal operons into a *single C. trachomatis* operon, yielding a progeny strain with three ribosomal loci, instead of two that are typically found in *C. trachomatis*. Strain L2Tet1 also is merodiploid at several loci between ribosomal operons, as it contains copies of *C. trachomatis* ORFS CT740-747 and homologs from *C. suis* (120). These poorly characterized ORFs are positioned between ribosomal operons in both species. In the present case, the opposite recombination event occurred. Recombinant strain RC-J(s)/122 (Table 4.1, cross 12), a progeny strain of recombinant RC-L2(s)/3 and RC-J/6276^{tet} exhibits a deletion of one of the three ribosomal operons commonly associated with our Tet^{r} *C. trachomatis* strains, returning the recombinant to the number of ribosomal operons found in *C. trachomatis* and other closely related species (Fig. 4.3). This also included the deletion of CT740-749, This progeny strain contains the *C. suis* homologs of CT740 to CT749, which apparently can complement any required function of the deleted *C. trachomatis* genes.

Genomic sequence analysis of the recombinant genomes showed that some of these isolates lacked the chlamydial plasmid. We originally hypothesized that this loss of plasmid was associated in some way with the recombination process. To explore this possibility, PCR analysis was performed on all isolates as well as the parents used in this study. These analyses demonstrated that both the J/6276^{rif} and the F(s)/70^{rif} parents were negative for the plasmid, whereas the L2-434^{ofl} parent was plasmid-positive (Table 4.1). Because plasmid was absent in parents used in the crosses, plasmid loss was likely a function of stress associated with antibiotic-based selection of strains prior to generating recombinant progeny.

The sequenced recombinant genomes allowed a comparative survey of the detected recombinant sites that were targeted in these genomes. The largest fragment that was laterally transferred during recombination was 412,907 base pairs, found in the RC-J(s)/122 recombinant strain. The smallest documented double crossover event was a 7 base pair fragment in the RC-L2(s)/3 strain. A total of 190 primary crossover regions were detected in the twelve recombinant strains. The distribution of these recombination sites was examined by mapping each crossover position from each of the 12 sequenced genomes to a single arbitrarily chosen parental genome [F(s)70; Fig. 4.4]. A higher concentration of crossovers surrounding the genetic markers used to select or screen for recombination was observed, particularly those surrounding the *rpoB* locus, and there were large regions of the chromosome that lacked evidence of recombination, such as the region surrounding CT001.

Comparative analysis of the 190 putative crossover regions revealed four sites where a crossover occurs at the same locus in unrelated recombinant strains (CT331/CT332, CT569/ CT570 CT585/CT586, CT778: Table 4.3). These crossover events occurred between common informative sites, but the precise position of the crossover cannot be more clearly established, because the sequences between informative sites are identical. In addition to these potentially shared crossover sites, we detected one region in recombinant strains RC-J/6276^{tet}, RC-L2(s)/3, RC-F(s)/852, and RC-F(s)/342 where crossover events occurred in the same gene (*hem*N_2; CT746), but were between different sets of informative sites. While these results indicate that recombination in different crosses can occur within the same locus or, perhaps, at the same position on the genome, there is no evidence that these positions represent recombination hotspots.

Chi sites have previously been described as sites for homologous recombination in bacteria (6, 107), and it is possible that a canonical chi site, or other sequence pattern, might be found at or near chlamydial recombination sites. The program MEME (5) was used to determine if any identified crossover

sites were linked to a common sequence motif. These and other analyses failed to identify any apparent genomic hotspot or target sequence in that might be associated with chlamydial recombination.

A method to correlate genotype with phenotype in clinical C. trachomatis strains

To determine if genomic analysis of the recombinant strains could be used to identify regions in the genome that may associate with observed phenotypes, an alignment of the twelve recombinant strains and the three parental sequences were aligned using the alignment program MAFFT (65). The Fisher's exact test was used as a metric to determine if the genotype at a given position in the genome was statistically associated with observed phenotypes. The known association of particular *incA* genotypes with a nonfusogenic phenotype (118) was used as a model to determine the utility of the procedure. A genome-wide association analysis using an alignment of the twelve recombinant strains and the three parental strains revealed a 23 kilobase genomic region surrounding *incA* (Fig. 4.7A) that contained polymorphic residues with the highest inverse-Log *p*-value possible by the Fisher's exact test (N = 15, 5 nonfusogenic strains, 10 fusogenic strains). A subsequent alignment was generated using the same 15 strains used in this study, plus an additional 10 currently available sequenced genomes where the inclusion fusion phenotype is known. The genome-wide association analysis of this larger data set revealed a single genomic site with the highest correlative value for this phenotype, and this position includes the lesion in strain F(s)/70 strain that interrupts the *incA* coding sequence.

Exploration of genotypes associated with attachment and secondary inclusion phenotypes

Differences in attachment efficiency in the presence or absence of centrifugation of inoculated cells is a clear phenotype differentiating *C. trachomatis* strains (71). In most cases, strains of serovar L2 are correlated with a highly efficient rate of attachment, while other serovars have a reduced ability to infect in the absence of centrifugation. To determine whether the recombinant strains displayed the

attachment efficiency similar to a high attachment efficiency parent (L2-434) or low attachment efficiency parent [F(s)/70 and J/6276], strains were assayed using qPCR to determine chlamydial genome copy numbers within infected McCoy cells using centrifugation versus static inoculation conditions. The results of these experiments revealed two groups of progeny strains that manifested significantly different attachment efficiencies (Fig. 4.5). One set behaved like the L2-434 parent, and a second group had a low attachment efficiency, similar to parents J/6276 and F(s)70.

The attachment phenotype of the parental and recombinant strains was used in the genome-wide association analysis to determine if regions in the chlamydial genome could be associated with the observed attachment phenotype. Five genomic regions were identified that had the highest possible inverse Log *p*-value (Fig. 4.8). These include regions containing ORFs CT001 – CT089, CT139 – CT146, CT203 - CT215, CT740 – CT746, and CT783 – CT875. The corresponding protein sequences in these regions from the 3 parents and 12 recombinants were aligned using the MUSCLE sequence alignment software, and were analyzed using the same approach as the genome-wide association method. A total of 124 proteins had at least one non-synonymous amino acid change that was associated with the attachment phenotype (Table S4.1). The chlamydial membrane proteins PmpE (14 non-synonymous amino acid changes), PmpF (110 AA changes), PmpG (28 AA changes), and PmpH (57 AA changes) were among the proteins with the highest number of non-synonymous amino acid changes that were associated with the attachment phenotype. Other relevant proteins that were associated with high attachment efficiency were ORFS CT0089, CT860-862, each of which encodes proteins involved in secretion of proteins via the Type III secretion machinery (40, 60).

Previous studies involving the attachment of *C. trachomatis* to host cells have implicated the involvement of several surface proteins including MOMP, HSP70 and OmcB (90, 112, 113, 122, 127, 135). Other outer membrane proteins are either neutralizing targets or function early in infection, such as Mip

and PorB (70, 77). None of these proteins was shown to be associated with the centrifugation-based efficiency of attachment as measured in this study. In addition to the possible roles OmcB and MOMP play in the binding of heparin and the subsequent attachment to host cells, Zhang and colleagues describe a mechanism where electrostatic forces on the bacterial surface might play a role in the attachment of *C. trachomatis* to host cells (135). To determine if the variation seen in the Pmp proteins described above between L2-434 and the two non-LGV parent strains might contribute to differing charges on the surface of the chlamydia, we calculated the overall charge at pH 7 of the each Pmp protein which included the set of Pmps associated with centrifugation-based attachment efficiency and PmpB, which is highly variable between strains but was not associated with high attachment in this study (Table 4.4, Fig. 4.8). Pmp proteins from three previously sequenced non-LGV urogenital strains (D/UW3, E/11023, and G/9768) were included in the analysis to determine if the charges of the urogenital strain Pmp proteins were similar to the two urogenital strains used as parents in this study. The results in table 4.4 show that the overall charge of the individual Pmp proteins are similar amongst each of the non-LGV urogenital strains. In the five Pmps that were associated with high attachment efficiency there was a common increase in net positive charge of the L2-434 homolog compared to non-LGV urogenital homolog (Table 4.4). The greatest increase in charge within Pmp proteins associated with attachment efficiency was seen in PmpG with an increase in overall charge from -22 in both J/6276 and F(s)/70 to -13 in L2-434. PmpD is highly conserved between L2-434 and the two parental urogenital strains (Fig. 4.9); thus, there was a minimal charge difference in the PmpD isoforms. The largest increase in charge seen in all Pmp proteins analyzed was in PmpB, which is not encoded by a locus associated with attachment efficiency in our assay. If the electrostatic charge of these membrane proteins does play a role in the efficiency of attachment, it is likely that the charge differences should be most evident in surface-exposed amino acids. The changes in charge identified in the different Pmps are predicted to be in regions of the protein that are hydrophilic and, possibly, surface exposed.

Formation of secondary inclusions in infected cells is another trait that varies among strains and serovars, with serovars G and F being generally higher and strains of serovar J and L2 being generally lower (119). We explored the genetic regions that might be associated with this phenotype in our collected recombinant strains. An added challenge to this phenotype is that analysis of secondary inclusions is not possible in strains that are IncA-negative, because our readout of secondary inclusions is dependent on antibodies to IncA. This had the effect of limiting the number of strains that were included in these comparisons. Of the eight strains tested, recombinants RC-J/953 and RC-J/971 showed extensive secondary inclusion and IncA-laden fiber production (Table 4.1; Fig. 4.9). This result is in contrast to both the parental strains J/6276 and L2-434 used to create RC-J/953 and RC-J/971, which are low secondary inclusion formers (119). The presence of recombinant progeny with high secondary inclusion phenotypes, in crosses where both parents exhibit low secondary inclusion formation, indicates a possible interaction between at least two chlamydial proteins in the manifestation of the secondary inclusion phenotype. To identify chromosomal loci in the two high secondary inclusion formers that might be associated with the formation of secondary inclusions, we aligned the four RC-J/900 series of recombinants, and searched for those regions that could possibly contribute to this phenotype. This analysis revealed eight possible chromosomal regions, containing 123 genes (Fig. 4.10, Table S4.2), where either gene products or intergenic regions from the J/6276 parent could interact with products contributed by the L2-434 parent.

Previous studies in our laboratory identified collections of clinical isolates that formed either high or low numbers of secondary inclusions (119). We sequenced the genomes of a subset of these strains (62, 117), and therefore, these strains were available to be added to our analysis of secondary inclusion formation. Sequenced isolates of serovar G (G/9768, G/11074, G/11222) are high secondary inclusion formers, while tested isolates of serovar E (E/11023 and E/150) were low secondary inclusion formers (not shown). Inclusion of genome sequence data from these strains led to a refining of the list of possible

candidates involved in secondary inclusion formation. We hypothesize that chlamydial inclusion membrane proteins (Inc proteins) are logical candidates for participating in secondary inclusion formation. The list in table S4.2 from the analysis of the four RC-J/900 series recombinants includes several such candidates, such as CT223 - 229, CT233, CT288, and CT565. Addition of the five clinical strains reduced the number of candidate ORFs to 55, and reduced the number of Inc proteins to only one, the hypothetical protein CT288.

DISCUSSION

One of the most significant challenges to the study of chlamydiae is the lack of a workable and routine genetic system. While there has been significant recent progress in this area (74), the correlation of genotype with phenotype remains an arduous task. One possible approach to this problem is the analysis of recombinant hybrid strains generated by mixing parental strains with known and differing phenotypic characteristics. Recombination among chlamydial strains in vitro was originally demonstrated by Demars and colleagues (33, 34), and we have recently explored this approach to demonstrate movement of tetracycline resistance genes among different chlamydial species. Although the methods for generating recombinants are becoming routine in this system, there is a general lack of understanding regarding the cellular and molecular mechanisms associated with this process.

The present study was initiated to address these challenges. We hypothesize that an investigation of both the process of genetic recombination in chlamydiae and the correlation of specific chlamydial genotypes with phenotypes can be addressed using a combination of contemporary genome sequencing technologies with our ability to create genetic recombinants among chlamydiae (16, 34, 120). Phenotypic distinctions varying among clinical strains represent the major focus of the analysis.

Evidence for recombination in chlamydiae was first provided in nucleotide sequence analysis of *ompA*, which encodes MOMP, the major serovariant antigen in chlamydiae (20, 73, 80). There is clear evidence that antigenic variation in some MOMP variants can be explained by recombination-mediated mosaic sequences in *ompA*. Evidence for recombination hotspots were also identified in CT049 and CT166 (46), leading to the premise that certain regions in the chlamydial genome are more frequently associated with recombination. The concept of recombination hotspots was addressed in whole genome sequence analysis as well. Work in our laboratory identified a serovar D strain whose genome contains primarily serovar E-like sequences (62). This genome has a hybrid D/E MOMP sequence, and apparent recombination sites within this strain are at or very near sites seen in other, independently isolated, clinical strains (46, 73). We investigated the possible clustering of recombination events in vitro through the genome sequence analysis of 12 recombinant genomes, leading to the identification of a total of 190 primary recombination sites. These analyses provided limited evidence for increased frequency of recombination at sites surrounding markers used for selection (Fig. 4.4), and a limited number of genome positions were identified as recombination targets in more than one independent progeny clone. None of the four positions that may have been used more than once in the generation of our recombinant clones is identified as a recombination hotspot in any other analysis, and at lest one protein identified as a recombination hostpot, the region surrounding *ompA*, was not targeted in any recombinant sequenced in this study. Furthermore, the four genomic regions that were involved in more than one recombination event in our study are so similar in sequence across all sequenced chlamydial genomes (not shown) that they would never have been considered hotspots in any analysis of clinical strains. Because of our inability to identify any hotspots surrounding previously identified recombination sites, we propose that most previously identified recombination hotspots were identified as such because: 1) there was significant, and probably very important, selection pressure for change at a locus (i.e. MOMP or Pmp antigenic variation), or: 2) the position being analyzed is identified

because there simply was more sequence heterogeneity in that region of the chromosome. The chlamydial genome is remarkably homogenous and syntenous, and we hypothesize that many recombination events have been missed because of the low level of genomic variability among strains.

Each recombination event identified in this study was a product of homologous recombination between highly related sequences. In a single case, two virtually identical rRNA sequences targeted a single rRNA operon in a recombination partner, leading to a deletion of several kilobases of DNA in the progeny [RC-J(s)/122, Fig. 4.3]. This was the only example of a deletion in any progeny strain. We also did not identify any evidence for duplication events in any of the recombinant progeny. This is consistent with the overall sequence homogeneity of sequenced chlamydial genomes, with some exceptions that stay consistent within biological groups, such as tryptophan utilization in urogenital versus ocular *C. trachomatis* strains (24, 25).

The three-way crosses attempted in this study were designed to examine the possibility that multiple parents could be involved in generation of a single recombinant progeny. We saw no evidence of a three-way cross in any of our selection experiments or in any genome sequence analysis, even though every combination of two-way crosses was recovered from those experiments. If the probability of a three-way event is a function of the probability of two independent recombination events, then this is not a surprising result. There is, however, one issue that is addressed by the absence of any evidence for contribution of three parents in a cross. In many of the recombinants identified by our group and in studies by Demars and colleagues (33, 34), multiple fragments from each parental genome are found in a recombinant progeny, often in regions of the chromosome that was not selected for with the tested antibiotics (Fig 4.2). It is possible that these differently recombined fragments involve sequential and independent recombination events occurring during the mixed infections used in this procedure. If involvement of multiple chlamydiae was a common occurrence in the generation of a cross, we hypothesized that some progeny from the three-

way crosses should carry fragments of each parent. Our results do not support this hypothesis. Therefore our current model is that the generation of recombinant progeny is the result of a single exchange event between two parents, and that these exchanges can involve very large fragments of the chromosomal DNA. Subsequent recombination events then lead to differential integration of fragments of the exchanged DNA, leading to the mosaicism seen in many of the recombinants.

The study of chlamydial pathogenesis is hampered by the absence of a routinely workable genetic system. One aspect of this is the challenges of directly associating identified phenotypes with particular genes or sets of genes. Three phenotypes were used in these analyses- the nonfusogenic phenotype, the efficiency of attachment in the absence of centrifugation, and the formation of secondary inclusions in cell culture. Each of these can be evaluated in our laboratory, and is differentially exhibited by clinical isolates of *C. trachomatis*. While the gene associated with the nonfusogenic phenotype is known [IncA; (118)], this served as a model phenotype to test our approach. Our system clearly identified *incA* as being associated with the nonfusogenic phenotype, and inclusion of both recombinant strains and clinical isolates with this phenotype augmented the strength of the analysis (Fig. 4.6).

We next analyzed phenotypes for which the genes leading to the phenotype are not known; attachment efficiency in the absence of centrifugation, and secondary inclusion formation within infected cells. Analysis of the attachment efficiency of the different recombinants revealed two groups having either high attachment efficiency characteristics, as exhibited by LGV strains, or low attachment efficiency characteristics, as exhibited by non-LGV urogenital strains (Fig. 4.5). A genome-wide association assay on these data revealed a number of loci that were quantitatively associated with the attachment efficiency phenotype seen in cell culture. While the list is extensive, we focused on surface proteins as the most logical candidates for affecting initial attachment events between chlamydiae and host cells. One collection of proteins that is

highly associated with our attachment efficiency phenotype were the Pmps, a group of chlamydia-specific autotransporters, some of which are known to be surface exposed and that are logical mediators of attachment to cells (81). Selected ORFS encoding proteins that function in Type III secretion were also associated with the attachment phenotype. We chose to focus our bioinformatic analysis on variability in Pmp structure. There were significant differences between the net charge of Pmp proteins between LGV and urogenital strains (Table 4.3, Fig. 4.8), with all examined Pmp proteins being more positively charged in the highly efficient attaching strains. This includes both the 4 Pmps that were associated with the phenotype (PmpEFGH) and PmpB, a highly variable Pmp in which sequence differences were not associated with attachment. Electrostatic forces on the surface of the chlamydia may play a role in attachment to host cells (135). It is possible that the net charge on key Pmp residues is associated with the centrifugation-based attachment differences evident between LGV and non-LGV strains. Chlamydial attachment is a complicated and multi-faceted process, and certainly proteins such as OmpA and OmcB are likely to participate as well. However, these proteins were not identified as critical to attachment efficiency as measured in this study.

The final phenotype we investigated in this analysis is the formation of secondary inclusions within infected cells, a phenotype that has differences not only between LGV and non-LGV but also differences within the non-LGV urogenital serovars (119). The most remarkable result in these assays was the identification of high secondary inclusion formers in crosses between parents exhibiting very low secondary inclusion formation phenotypes (Table 4.1, Fig. 4.9). This result can be explained one of two ways, both of which are based on the premise that more than one polymorphism is required for establishing the secondary inclusion phenotype. It is possible that the required polymorphisms are present within a single protein, and that the negative parents each contributed a fraction of the protein that led to the generation of a high secondary inclusion former. This is an unlikely possibility, because in no case was a recombination site shared between two progeny strains that were high secondary

inclusion formers. The second possibility is that the formation of secondary inclusions is the result of two or more functionally linked loci in the genome. We hypothesize that at least one of the proteins associated with secondary inclusion formation is a chlamydial Inc, as these proteins are localized to the surface of the inclusion within infected cells (93). In a genome-wide association analysis using the four RC-J/900 series recombinants, several Inc proteins were flagged as a possible candidate for the secondary inclusion phenotype, including CT223 to CT229, CT233, CT288, and CT565. However, when an additional five previously sequenced genomes with known secondary inclusion phenotypes were added to the analysis, only one putative Inc protein could be associated with this phenotype (CT288). At present it remains unclear as to any specific gene product that is associated with the formation of secondary inclusions within infected cells.

Continued examination of novel recombinants, including backcrosses to integrate more parental genome into recombinant strains, are proceeding in the laboratory, and we expect that these experiments will add clarity to the phenotypes we have discussed. We also continue to use the recombinants as tools to understand the basic processes associated with genetic exchange in the chlamydiae.

ACKNOWLEDGMENTS

We would like to thank Sara Weeks and Robert Heinzen for editing the manuscript. Chris Sullivan from the Center for Genome Research and Biocomputing at Oregon State University is acknowledged for his help with genome sequence analysis. This research was supported by grants AI088540-02 and AI086469-01 from the National Institutes of Health.

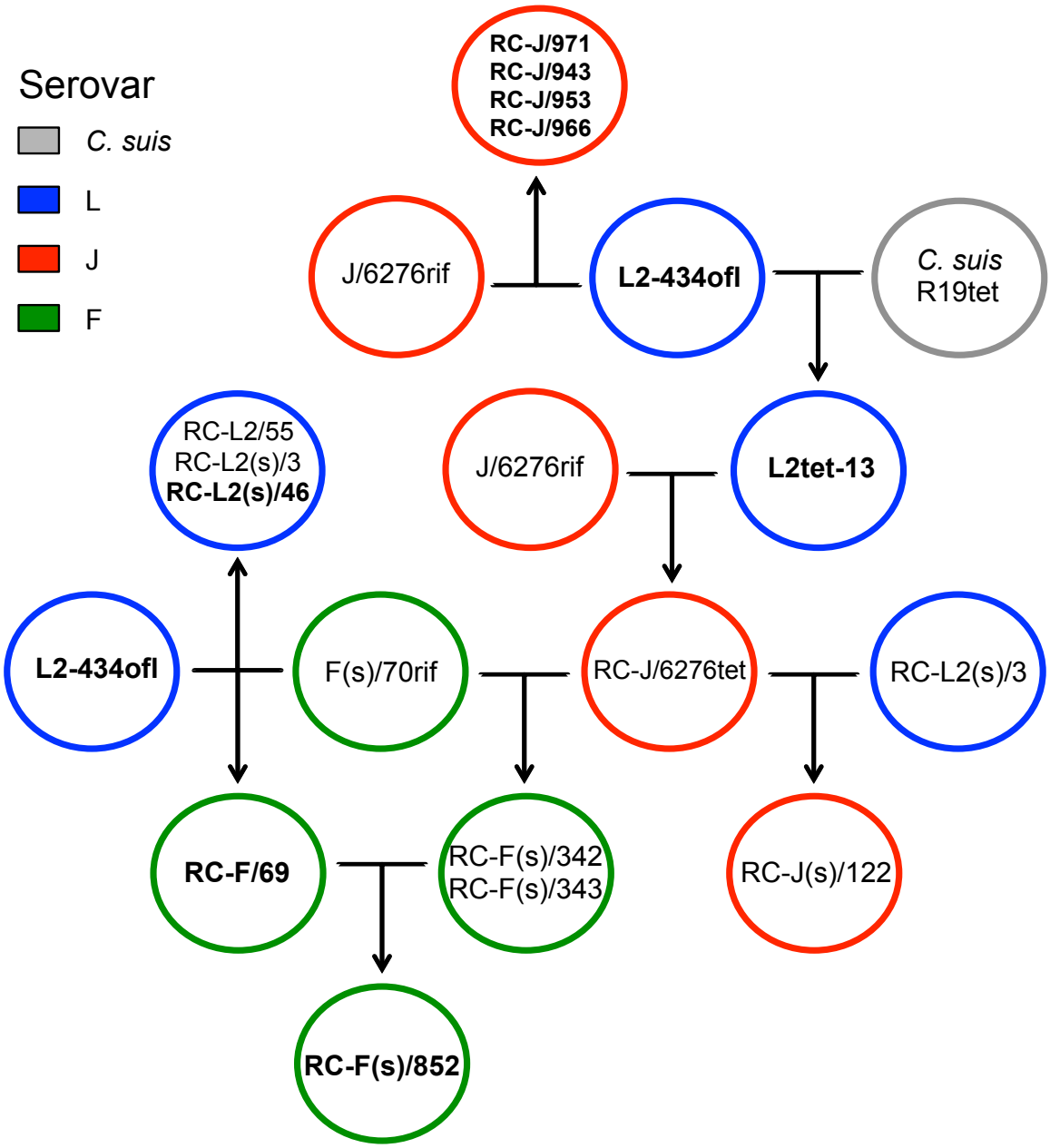


Figure 4.1. Recombinant geneology.
The figure shows the parental strains used to generate recombinant strains. The colors used indicate the *ompA* serotype of each strain. Strains containing the plasmid are shown in bold face.

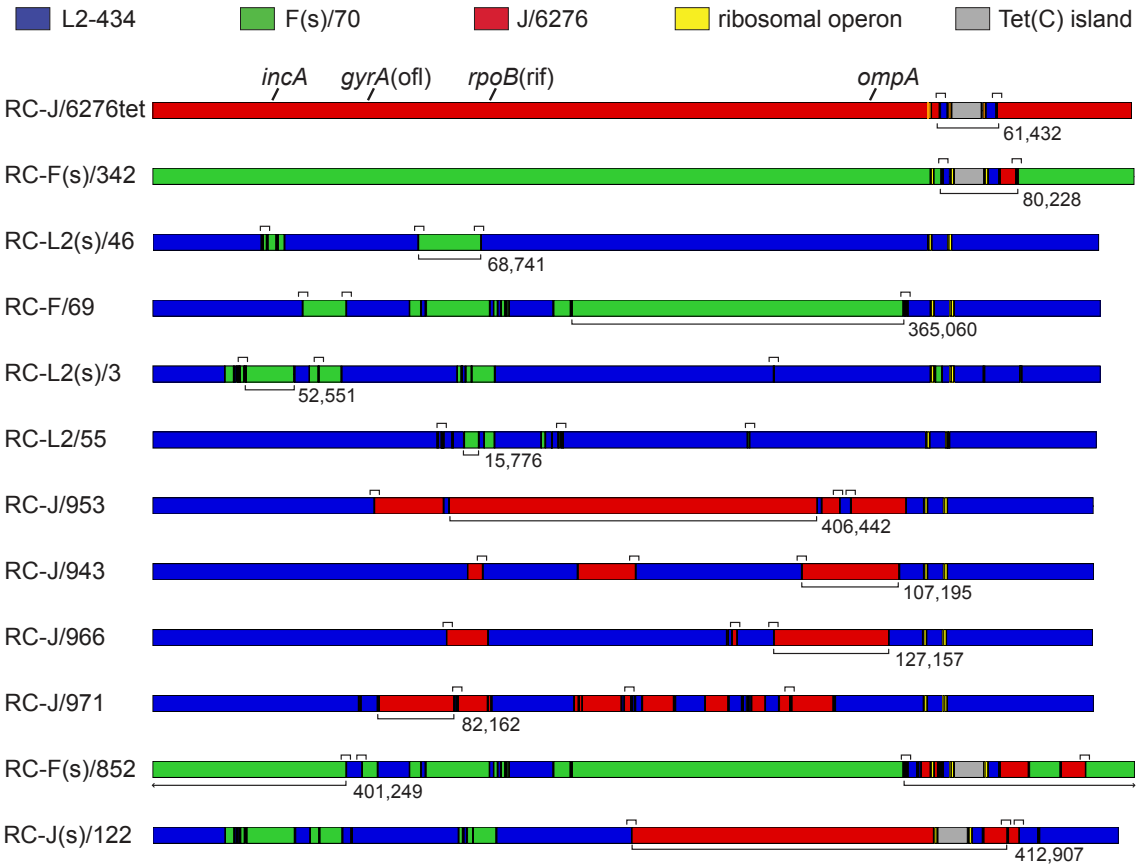


Figure 4.2. Genome maps of recombinant strains.
Colors used in recombinant maps indicate the parental genotype. The approximate location of the genetic markers used in the construction of the recombinant genomes is shown above the RC-L2(s)/46 genome map. The bracket and number below the genome map indicate the largest size of DNA received during recombination. The brackets above the genome map indicate crossover regions that were confirmed by PCR amplification and Sanger sequencing.

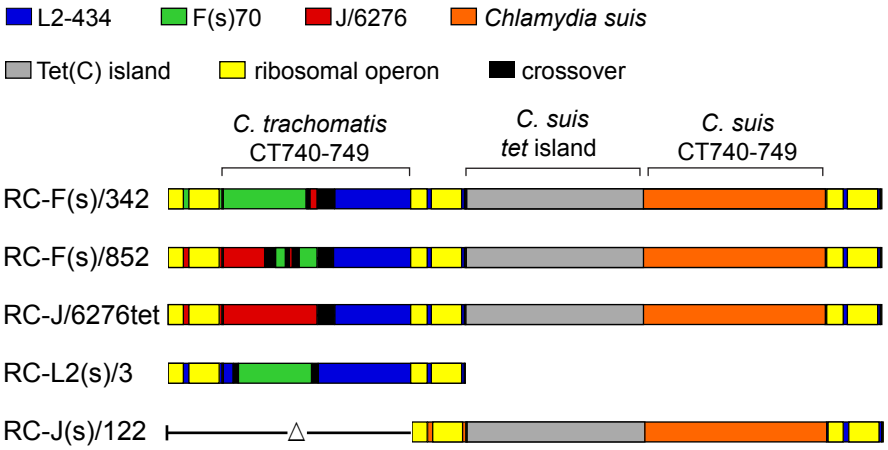


Figure 4.3. Schematic diagram of the CT740 to CT749 regions in selected recombinant sequences.
The colors used indicate the genotype of a given region. The ribosomal operons are shown in yellow, and crossover sites are shown in black. The deletion of the *C. trachomatis* homologous region of CT740 to CT749 in the RC-J(s)/122 sequence is indicated by the delta symbol.

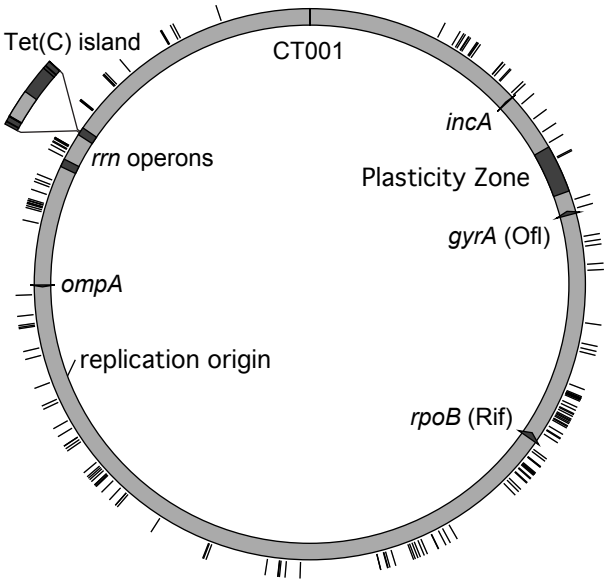


Figure 4.4. Genomic locations of crossover regions in the ten sequenced recombinants.
The sequenced strain D/UW3 gene designations were used as the reference, with the location of gene CT001 indicated at the top of representative genome. The black tick marks indicate location of a crossover region. The location of the plasticity zone is shown in gray.

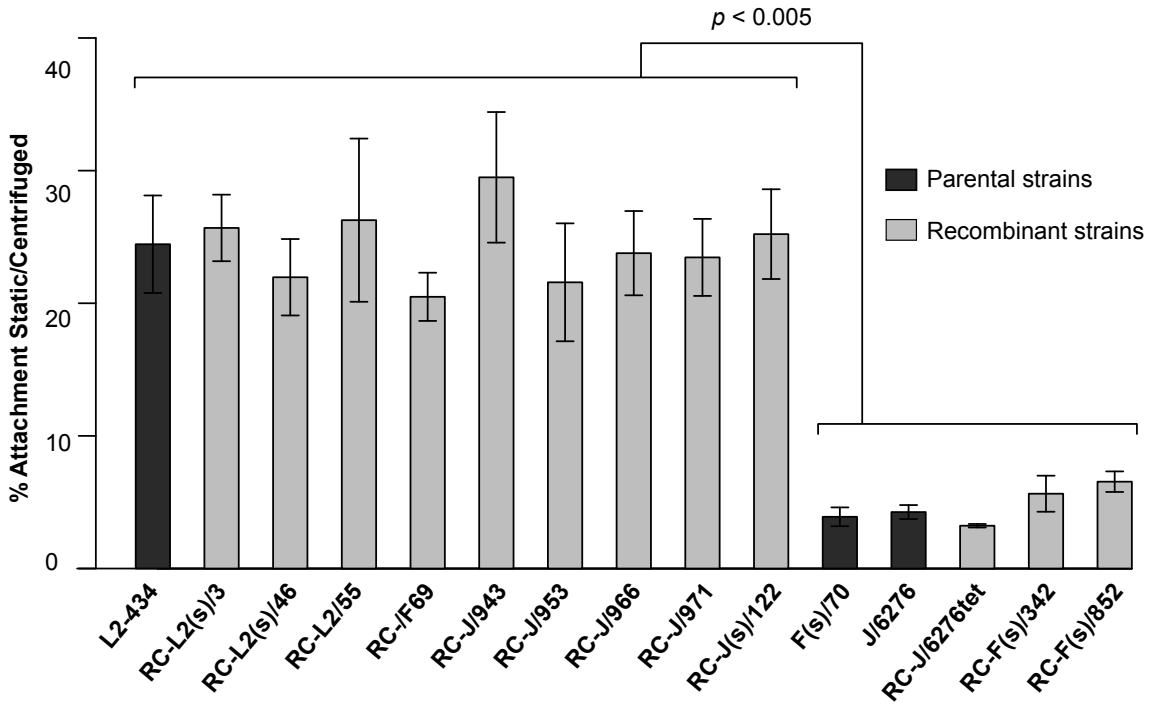


Figure 4.5. Attachment efficiency results of parental and recombinant strains.
Strains analyzed are represented on the x-axis, and percent attachment efficiency is represented on the y-axis. Dark gray bars represent parental strains, and light gray bars represent recombinant strains.

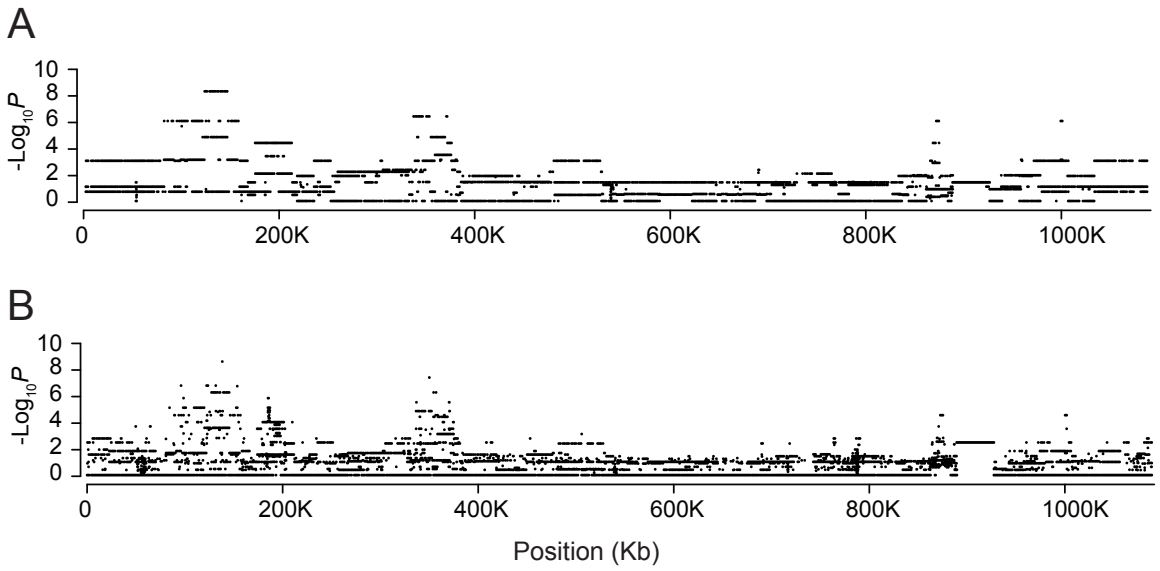


Figure 4.6. Genome-wide association analysis of the non-fusing inclusion phenotype.

Genome-wide *P* values from Fisher's exact test are given on the Y-axis. Genome position is indicated along X-axis, beginning with CT001 as defined for the DUW/3 genome (109). Black dots represent a statistic from a position in the alignment where there is variation in one or more of the genome sequences. Panel A shows results from Fisher's exact test using an alignment of the twelve recombinant genomes and the three parents used for creating the recombinants. Panel B shows results from Fisher's exact test using an alignment of the twelve recombinant genomes, the three parents used, and eleven published *C. trachomatis* genomes.

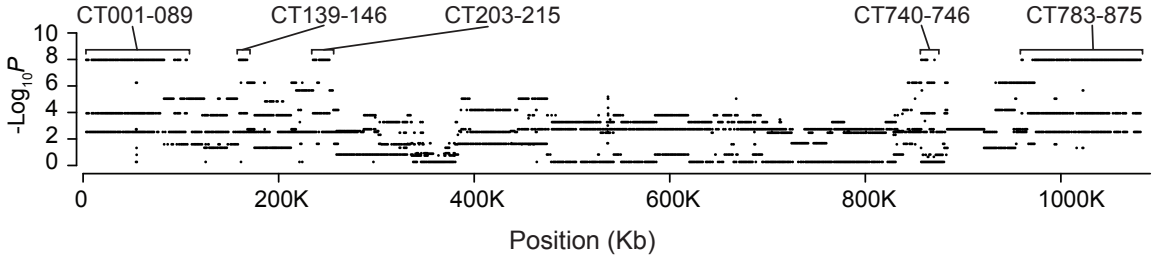


Figure 4.7. Genome-wide association analysis of the attachment phenotype.
Axes used are the same as in figure 6. The bracketed regions indicate the genes located in these regions. The results are from an alignment of the twelve recombinants and the three parents used for creating the recombinants.

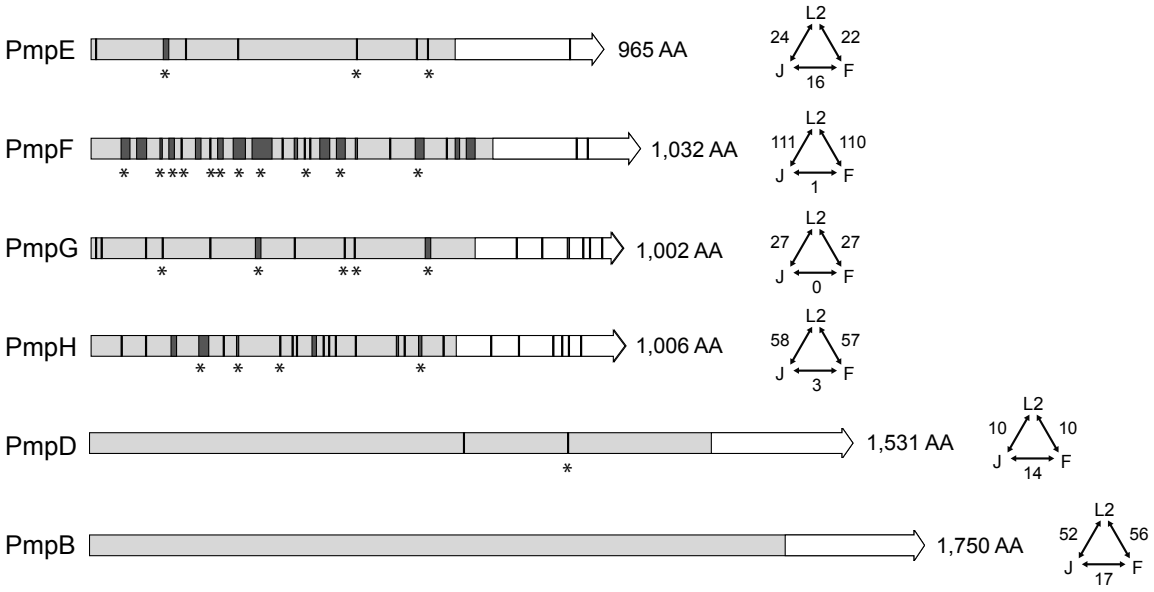


Figure 4.8. Protein charge analysis of the polymorphic membrane proteins. The length of the L2-434 protein is shown to the right of the cartoon. The white region at the C-terminus of the protein indicates the auto-transporter domain. The light gray region of protein include the N-terminal signal sequence and passenger domain of the proteins. The dark gray boxes are the amino acid regions in the L2-434 protein that were flagged as being associated with attachment. The asterisks below indicate those regions that are hydrophilic and possibly surface exposed as well as amino acid changes that result in a net positive charge change. The figure to the right of each protein are the results of a pairwise alignment analysis, where numbers represent the number of amino acid changes between two aligned strains.

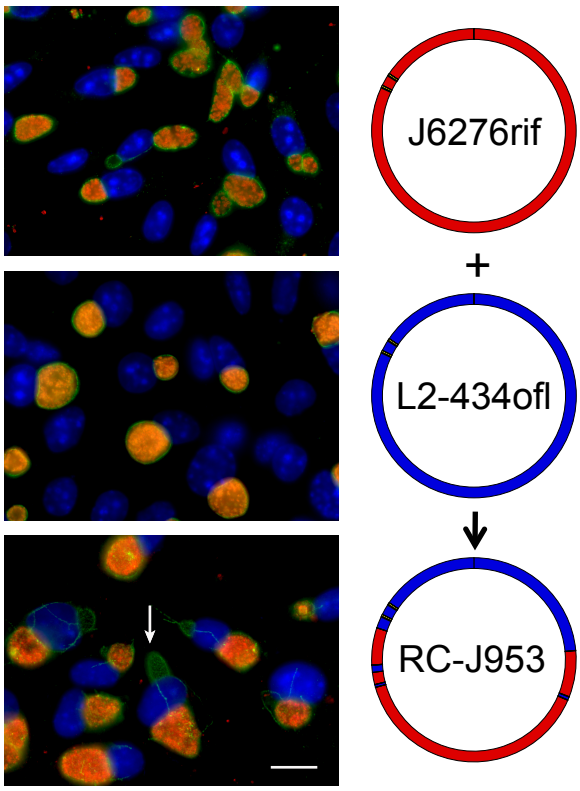


Figure 4.9. Fluorescent microscopic analysis of the secondary inclusion formation phenotype of recombinant strain RC-J/953.
McCoy cells were infected at an MOI of ~0.5, and images were taken 48 hours post-infection. All cells were labeled with anti-IncA (green), and anti-OmpA (red), and DNA is labeled with DAPI (blue). A representative secondary inclusion is indicated by the white arrow in the bottom panel. The strain being analyzed is shown at the right of each image. Scale bar, 10 μm.

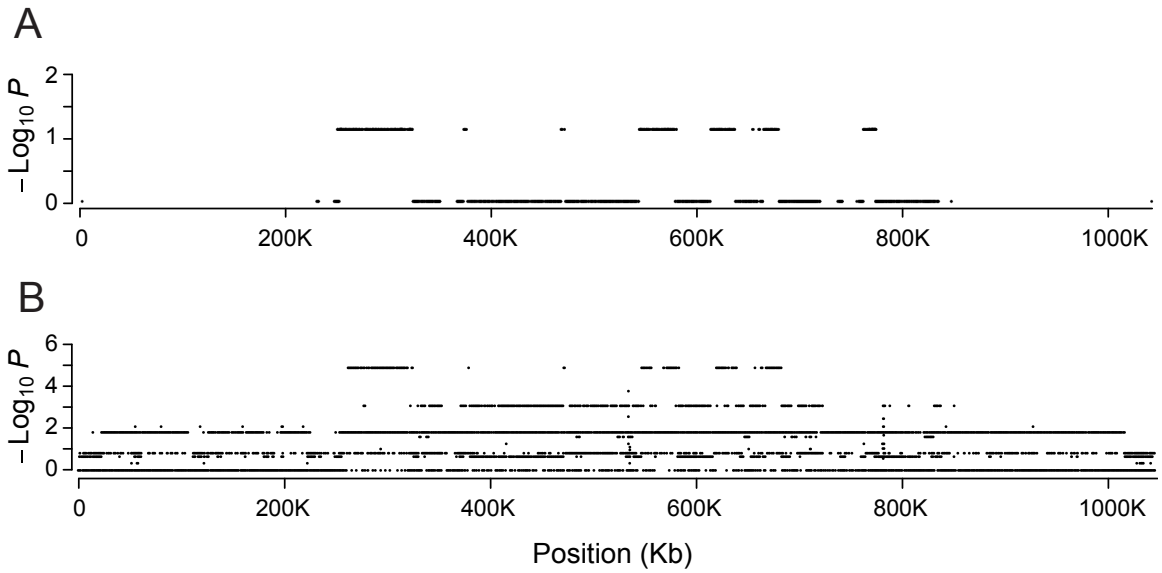


Figure 4.10. Genome-wide association analysis of secondary inclusion formation phenotype.

Axes used are the same as in figure 7. Panel A: Plotted results are from an alignment of the four RC-J/900 series of recombinants. Regions with highest inverse-Log *P* values are: CT223-289, CT332-333, CT410-412, CT470-501, CT545-566, CT581-600, and CT664-675. Analysis of alignment with the four RC-J/900 series and five previously sequenced isolates are shown in panel B. Regions with highest inverse-Log *P* values are: CT233-288, CT333, CT411, CT470-478, CT489-501, CT546-565, 581, CT586, and CT589-600.

Table 4.1. Phenotypes of parents and progeny in recombinant crosses.

				MIC (μg/ml)				Phenotype		
Cross	Recombinant	Parental strains	MOMP	Rifampicin	Ofloxacin	Tetracycline	Plasmid	Inclusion Fusion	Attachment	Secondary
1	**RC-J/6276tet**		J	8	0.5	8	-	+	-	2
		L2-434tet-13	L2	0.008	16	8	+	+	+	1
		J/6276rif	J	8	0.5	0.032	-	+	-	1
2	RC-F(s)/342		F	32	0.5	8	-	-	-	N/A
		RC-J/6276tet-rif	J	8	0.5	8	-	+	-	1
		F(s)/70rif	F	32	0.5	0.032	-	-	-	N/A
3	RC-L2(s)/46		L2	32	16	0.032	+	-	+	N/A
		L2-434ofl	L2	0.008	16	0.032	+	+	+	1
		F(s)70rif	F	32	0.5	0.032	-	-	-	N/A
4	**RC-F/69**		F	32	4	0.032	+	+	+	1
		L2-434ofl	L2	0.008	16	0.032	+	+	+	1
		F(s)70rif	F	32	0.5	0.032	-	-	-	N/A
5	**RC-L2(s)/3**		L2	32	4	0.032	-	-	+	N/A
		L2-434ofl	L2	0.008	16	0.032	+	+	+	1
		F(s)70rif	F	32	0.5	0.032	-	-	-	N/A
6	RC-L2/55		L2	32	4	0.032	-	+	+	N/A
		L2-434ofl	L2	0.008	16	0.032	+	+	+	1
		F(s)/70rif	F	32	0.5	0.032	-	-	-	N/A
7	RC-J/953		J	8	16	0.032	+	+	+	4
		L2-434ofl	L2	0.008	16	0.032	+	+	+	1
		RC-F(s)/343tet-rif	F	32	0.5	8	-	-	-	N/A
		J/6276rifR	J	8	0.5	0.032	-	+	-	1
8	RC-J/943		J	8	16	0.032	+	+	+	1
		L2-434/ofl	L2	0.008	16	0.032	+	+	+	1
		RC-F(s)/343tet-rif	F	32	0.5	8	-	-	-	N/A
		J/6276rif	J	8	0.05	0.032	-	+	-	1
9	RC-J/966		J	8	16	0.032	+	+	+	1
		L2-434/ofl	L2	0.008	16	0.032	+	+	+	1
		RC-F(s)/343tet-rif	F	32	0.5	8	-	-	-	N/A
		J/6276rif	J	8	0.5	0.032	-	+	-	1
10	RC-J/971		J	8	16	0.032	+	+	+	4
		L2-434/ofl	L2	0.008	16	0.032	+	+	+	1
		RC-F(s)/343tet-rif	F	32	0.5	8	-	-	-	N/A
		J/6276rif	J	8	0.5	0.032	-	+	-	1
11	RC-F(s)/852		F	32	4	8	+	-	-	N/A
		RC-F/69	F	32	4	0.032	+	+	+	1
		RC-F(s)/343tet-rif	F	32	0.5	8	-	-	-	N/A
12	RC-J(s)/122		J	32	4	8	-	-	+	N/A
		RC-L2(s)/3	L2	32	4	0.032	-	-	+	N/A
		RC-J/6276tet-rif	J	8	0.5	8	-	+	-	1

Table 4.2. Recombination rates.

Cross	Progeny from cross (IFU)	Selection antibiotics	Recombinant progeny (IFU)	Recombination rate
$L2^{oflR}$ X $L2^{rifR}$	2.8 X 10^7	rif/ofl	610	2.2 X 10^{-5}
$L2^{oflR}$ X $F(s)^{rifR}$	7.3 X 10^6	rif/ofl	26	3.7 X 10^{-6}
$L2^{oflR}$ X J^{rifR} X $F(s)^{tetR}$	1.2 X 10^7	rif/ofl/tet	0	0
$L2^{oflR}$ X J^{rifR} X $F(s)^{tetR}$	1.2 X 10^7	rif/ofl	32	2.6 X 10^{-6}
$L2^{oflR}$ X J^{rifR} X $F(s)^{tetR}$	1.2 X 10^7	tet/ofl	82	6.8 X 10^{-6}

Table 4.3. Regions of shared crossovers between recombinant strains.

	Strain 1	Strain 2
Recombinant	RC-L2(s)/3	RC-F(s)/342
Region of crossover	CT778 (*priA)*	CT778 (*priA)*
Coordinates	916870 : 917156	954495 : 955597
Size	286	1,102
Comments	F(s)70 - L2-434 hybrid CT778	F(s)70 - J/6726 hybrid CT778
Recombinant	RC-L2(s)/3	RC-J/966
Region of crossover	CT331 (*dxs*) and CT332 (*pykF)*	CT332 (*pykF)*
Coordinates	377279 : 377995	370626 : 37785
Size	716	159
Comments	F(s)70 CT331, L2-434 CT332	J/6276 - L2-434 hybrid CT332
Recombinant	RC-J/971	RC-J/966
Region of crossover	CT569 (*gspG*) and CT570 (*gspF*)	CT569 (*gspG*) and CT570 (*gspF*)
Coordinates	634854 : 636140	635246 : 636532
Size	1,287	1,287
Comments	J/6276 CT569, L2-434 CT570	J/6276 CT569, L2-434 CT570
Recombinant	RC-J/971	RC-L2/55
Region of crossover	CT585 (*trpS*) and CT586 (*uvrB*)	CT586 (*uvrB*)
Coordinates	655362 : 656561	656865 : 657292
Size	1,200	428
Comments	L2-434 CT585, J/6276 CT586	F(s)70 - L2-434 hybrid CT586

Table 4.4. Polymorphic membrane protein charge analysis.
The numbers below each of the Pmp is the charge of the protein at a pH of 7. The results of the three parental strains used in this study as well as three previously sequenced non-LGV urogenital strains are shown.

	Associated with attachment					Not associated with attachment			
Strain	**pmpD**	**pmpE**	**pmpF**	**pmpG**	**pmpH**	**pmpA**	**pmpB**	**pmpC**	**pmpI**
L2-434	-51.5	-4.5	4.6	-13	-8.6	10.9	5.8	-89.4	-11.6
J/6276	-54.5	-7.5	0.6	-22	-16.8	10.9	-20.9	-98.7	-12.4
F(s)/70	-53.8	-9.5	0.6	-22	-15.8	10.9	-24	-102.4	-11.6
D/UW3	-53.5	-7.5	0.6	-20.8	-16.8	10.9	-23.9	-100.4	-11.6
E/11023	-53.8	-7.5	0.6	-20.8	-16.8	10.9	-24	-102.4	-11.6
G/9768	-54.5	-7.2	-2.4	-20.8	-15.8	10.9	-20.9	-98.7	-11.6

Table S4.1. Genes associated with attachment phenotype.

D/UW3 locus	L2-434 locus	Gene product	NS AA changes	Membrane protein	indels	Elongation/truncation
CT036	CTL0291	hypothetical protein	20	putative membrane protein	yes	yes
CT049	CTL0305	hypothetical protein	30			
CT050	CTL0306	hypothetical protein	19	putative membrane protein	yes	
CT051	CTL0307	hypothetical protein	63	putative membrane protein	yes	
CT058	CTL0314	hypothetical protein	11	putative membrane protein	yes	
CT144	CTL0399	hypothetical protein	20	putative membrane protein	yes	
CT813	CTL0184	hypothetical protein	12	putative membrane protein		
CT833	CTL0205	*infC*	38		yes	yes
CT852	CTL0225	hypothetical protein	12	putative membrane protein	yes	yes
CT861	CTL0236	*copB2*	11	putative membrane protein		
CT868	CTL0247	hypothetical protein	36	putative membrane protein	yes	yes
CT869	CTL0248	*pmpE*	14	membrane protein	yes	
CT870	CTL0249	*pmpF*	110	membrane protein	yes	
CT871	CTL0250	*pmpG*	28	membrane protein		
CT872	CTL0251	*pmpH*	57	membrane protein	yes	
CT873	CTL0252	hypothetical protein	72		yes	yes
CT875	CTL0255	hypothetical protein	28			
CT001	CTL0256	hypothetical protein	2	putative membrane protein		
CT007	CTL0262	hypothetical protein	2	putative membrane protein		
CT011	CTL0266	hypothetical protein	2			
CT013	CTL0268	*cydA*	4	putative membrane protein		
CT014	CTL0269	*cydB*	3	putative membrane protein		
CT016	CTL0271	hypothetical protein	2			
CT017	CTL0272	hypothetical protein	3	putative membrane protein		
CT018	CTL0273	hypothetical protein	6	putative membrane protein		
CT019	CTL0274	*ileS*	5			
CT032	CTL0287	*metG*	2			
CT034	CTL0289	hypothetical protein	3	putative membrane protein		
CT042	CTL0298	hypothetical protein	3			
CT045	CTL0301	hypothetical protein	2			
CT046	CTL0302	*hct2*	5			
CT054	CTL0310	*sucA*	6			
CT057	CTL0313	*gcpE*	3			
CT059	CTL0315	*fer*	2			
CT060	CTL0316	*flhA*	4	putative membrane protein		
CT063	CTL0319	*gnd*	4			
CT082	CTL0338	hypothetical protein	3	putative membrane protein		
CT083	CTL0338A	hypothetical protein	2			
CT089	CTL0344	*copN*	5	putative membrane protein		
CT140	CTL0395	hypothetical protein	3	putative membrane protein		
CT142	CTL0397	hypothetical protein	8	putative membrane protein		
CT143	CTL0398	hypothetical protein	8			
CT146	CTL0401	*dnlJ*	2			
CT205	CTL0457	*pfkA*	2			
CT207	CTL0459	*pfkA_2*	2			
CT208	CTL0460	*gseA*	2	putative membrane protein		
CT209	CTL0461	*leuS*	5			
CT210	CTL0462	*hemL*	4			
CT212	CTL0464	hypothetical protein	2			
CT214	CTL0466	hypothetical protein	5	putative membrane protein		
CT215	CTL0467	*dhnA*	3			
CT783	CTL0152	hypothetical protein	2			
CT796	CTL0165	*glyQ*	2			
CT798	CTL0167	*glgA*	2			
CT806	CTL0175	*ptr*	2	putative membrane protein	yes	
CT809	CTL0178	hypothetical protein	7		yes	yes
CT812	CTL0183	*pmpD*	2	membrane protein	yes	
CT815	CTL0187	hypothetical protein	3			
CT816	CTL0188	*glmS*	4			
CT817	CTL0189	*tyrP*	2	putative membrane protein		
CT818	CTL0190	hypothetical protein	2	putative membrane protein		
CT821	CTL0193	*sucC*	4	putative membrane protein		

Table S4.1 (continued)

D/UW3 locus	L2-434 locus	Gene product	NS AA changes	Membrane protein	indels	Elongation/truncation
CT823	CTL0195	*htrA*	3			
CT824	CTL0196	hypothetical protein	4			
CT825	CTL0197	*rmuC*	4	putative membrane protein		
CT827	CTL0199	*nrdA*	7			
CT830	CTL0202	hypothetical protein	2			
CT837	CTL0209	hypothetical protein	6			
CT841	CTL0213	*ftsH*	3	putative membrane protein		
CT842	CTL0214	*pnp*	4			
CT847	CTL0219	hypothetical protein	2			
CT848	CTL0220	hypothetical protein	3			
CT850	CTL0223	hypothetical protein	2	putative membrane protein		
CT854	CTL0227	hypothetical protein	5	putative membrane protein		
CT856	CTL0231	hypothetical protein	3	putative membrane protein		
CT859	CTL0234	*ispH*	3			
CT860	CTL0235	*copD2*	6			
CT862	CTL0237	*lcrH*	2			
CT863	CTL0238	hypothetical protein	2			
CT865	CTL0244	hypothetical protein	2			
CT866	CTL0245	*glgB*	4			
CT867	CTL0246	hypothetical protein	9	putative membrane protein		
CT874	CTL0254	*pmpI*	3			
CT005	CTL0260	hypothetical protein	1	putative membrane protein		
CT010	CTL0265	*htrB*	1			
CT012	CTL0267	hypothetical protein	1	putative membrane protein		
CT015	CTL0270	hypothetical protein	1			
CT021	CTL0276	hypothetical protein	1	putative membrane protein		
CT022	CTL0277	*rpmE*	1			
CT025	CTL0280	*ffh*	1			
CT027	CTL0282	*trmD*	1			
CT035	CTL0290	hypothetical protein	1			
CT040	CTL0296	*ruvB*	1			
CT041	CTL0297	hypothetical protein	1			
CT047	CTL0303	hypothetical protein	1			
CT052	CTL0308	*hemN*	1			
CT056	CTL0312	hypothetical protein	1			
CT061	CTL0317	*fliA*	1			
CT062	CTL0318	*tyrS*	1			
CT065	CTL0321	hypothetical protein	1	putative membrane protein		
CT076	CTL0332	*smpB*	1			
CT079	CTL0335	hypothetical protein	1	putative membrane protein		
CT087	CTL0342	*malQ*	1			
CT206	CTL0458	hypothetical protein	1			
CT211	CTL0463	hypothetical protein	1			
CT784	CTL0153	*rnpA*	1			
CT795	CTL0164	hypothetical protein	1			
CT799	CTL0168	*rplY*	1			
CT804	CTL0173	*ispE*	1			
CT807	CTL0176	*plsB*	1		yes	
CT808	CTL0177	*cafE*	1			
CT820	CTL0192	*ftsY*	1			
CT826	CTL0198	*pssA*	1	putative membrane protein		
CT828	CTL0200	*nrdB*	1	putative membrane protein		
CT831	CTL0203	*murB*	1			
CT832	CTL0204	*nusB*	1			
CT836	CTL0208	*pheS*	1			
CT839	CTL0211	hypothetical protein	1	putative membrane protein		
CT840	CTL0212	*tilS*	1			
CT846	CTL0218	hypothetical protein	1	putative membrane protein		
CT853	CTL0226	hypothetical protein	1	putative membrane protein		
CT857	CTL0232	hypothetical protein	1	putative membrane protein		
CT858	CTL0233	*cpa*	1			
CT864	CTL0243	*xerD*	1			

Table S4.2. Genes associated with secondary phenotype.

D/UW3 locus	L2-434 locus	Gene product	NS AA changes
CT234	CTL0486	hypothetical protein	3
CT235	CTL0487	hypothetical protein	2
CT237	CTL0489	*fabG*	1
CT238	CTL0490	*fabD*	2
CT240	CTL0492	*recR*	1
CT241	CTL0493	hypothetical protein	1
CT244	CTL0496	hypothetical protein	3
CT245	CTL0497	*pdhA*	1
CT247	CTL0499	*pdhC*	3
CT250	CTL0501	*dnaA*	2
CT251	CTL0503	*oxaA*	1
CT252	CTL0504	*lgt*	1
CT254	CTL0506	hypothetical protein	1
CT256	CTL0508	hypothetical protein	1
CT257	CTL0509	hypothetical protein	1
CT259	CTL0511	hypothetical protein	1
CT262	CTL0514	hypothetical protein	2
CT264	CTL0516	hypothetical protein	2
CT266	CTL0518	hypothetical protein	2
CT269	CTL0521	*murE*	4
CT270	CTL0522	*pbp*	1
CT278	CTL0530	*nqrB*	1
CT281	CTL0533	*nqrE*	1
CT283	CTL0535	hypothetical protein	2
CT288	CTL0540	hypothetical protein	3
CT333	CTL0587	*uvrA*	1
CT411	CTL0668	*lpxB*	2
CT470	CTL0730	*recO*	2
CT472	CTL0733	hypothetical protein	1
CT474	CTL0735	hypothetical protein	2
CT475	CTL0736	*pheT*	2
CT477	CTL0738	hypothetical protein	3
CT478	CTL0739	*oppC2*	3
CT493	CTL0754	*polA*	4
CT494	CTL0755	*sohB*	2
CT495	CTL0756	hypothetical protein	5
CT497	CTL0759	*dnaB*	1
CT498	CTL0760	*gidA*	2
CT501	CTL0763	*ruvA*	1
CT551	CTL0813	*dacC*	1
CT552	CTL0814	hypothetical protein	1
CT554	CTL0817	*brnQ*	2
CT555	CTL0818	hypothetical protein	2
CT556	CTL0819	hypothetical protein	1
CT559	CTL0822	*sctJ*	2
CT560	CTL0823	hypothetical protein	1
CT561	CTL0824	*sctL*	1
CT586	CTL0849	*uvrB*	1
CT589	CTL0852	hypothetical protein	4
CT590	CTL0853	hypothetical protein	2
CT591	CTL0854	*sdhB*	3
CT594	CTL0858	hypothetical protein	1
CT595	CTL0859	*dsdD*	4
CT597	CTL0861	hypothetical protein	2
CT598	CTL0861A	hypothetical protein	2

General conclusion

The significant decrease in costs of genome sequencing along with the development of methods for assembling genomes and sequence analysis has made these approaches more applicable to studying biology. The central theme of this project involved the use of genome sequencing and comparative genome analysis to address aspects of chlamydial biology in a system where routine genetic manipulation of this organism is absent. This theme was applied to three specific areas: IncA-positive and IncA-negative populations in infected patients, tissue tropism among differing serovars, and using recombination and genome sequencing to study phenotypic differences in cell culture among serovars. Previous work had identified patients that were infected with mixed populations of IncA-positive and IncA-negative strains. We clonally isolated these same serovar-matched pairs and studied their behavior *in vivo* and *in vivo,* and genome sequenced the strains to determine what changes are occurring during infection within the host. Results of non-fusing inclusion and fusing inclusion isolates from the same patient show that these strains are genetically very similar, and that they have similar transcriptional, growth, and infectivity profiles, whereas non-fusing inclusion isolates from a non-matched pair background perform more poorly than the same serovar wild-type strains. It is likely that the IncA-negative strains acquire mutations over time, and these mutations are responsible for these strains to eventually fall out of the population. It is less clear what causes the IncA-negative strains to be selected for shortly after the *incA* deletion event. It is possible that there may be interplay between the bacteria and the host, where differing host dynamics such as immune status could contribute to the IncA-negative strains being selected for in these individuals. Homologous recombination has been proposed to be a method of generating genetic variety in bacterial species as well as genome maintenance, where the bacteria could repair deleterious lesions via recombination with DNA from a separate individual within the population (3). In the context of the non-fusogenic strains, it is possible that these strains become isolated from other chlamydia within an infected cell, where each inclusion contains a single

propagating clone. It is plausible that these physically isolated bacteria accumulate mutations over time, and in the absence of a heterozygous population of bacteria they cannot maintain or “fix” their genome through homologous recombination. The accumulation of these mutations could lead to a less fit organism, as represented by the slower growth and lower infection rates seen in these strains. A possible future avenue of research into these IncA-positive and IncA-negative populations would be to whole genome sequence on a larger set of matched pair isolates as well as IncA-negative isolates without an IncA-positive background. These data would provide insight into the extent of mutations in the IncA-negative strains, as well as identifying a set of genes that could be involved in the observed growth rates of the IncA-negative strains.

A second area of study where genome sequencing and comparative genomics was used was to find genetic loci that could be associated with observed tissue tropism differences among serovars. *C. trachomatis* serovar G strains exhibit a higher proportion of infection of the rectum in men having sex with men. Genome sequence of several chlamydia isolates including serovars G, E, and J isolated from the rectum or the cervix revealed several variable genes in the rectal serovar G compared to other serovars as well as serovar G strains isolated from the cervix. PCR analysis on additional clinical isolates revealed that ORFs CT144, CT154, and CT326 could be associated with the rectal tissue tropism phenotype. The functions of the products of these genes are currently not known, and further study is warranted to understand the roles these genes may play in the observed phenotype. One such study would be to generate antibodies to these proteins and determine their cellular location during infection.

In addition to tissue tropism differences observed between chlamydial serovars, differences in attachment to host cells in vitro between LGV strains and non-LGV urogenital strains, as well as differences in the development of secondary inclusions between serovars have been observed. To study these phenotypes, we developed assays to generate recombinant isolates from parents

with differing in vitro phenotypes, and using genome sequencing and genome-wide association assays to identify genetic loci that could be associated with the observed phenotypes. Several loci associated with higher attachment efficiency were detected. Among these loci was the family of polymorphic membrane proteins. Charge analysis of these proteins showed that strains with low attachment efficiency had a more negative charge compared to the recombinant strains with the high attachment efficiency. This adds strength to the hypothesis that the high attachment efficiency seen in the LGV strains could be attributed to the overall charge of the surface of the bacteria and its interaction with the host cell surface (135). To address the issue that in this study several loci including many many genes that were associated with attachment efficiency, a future approach would be to select for large pools of recombinants, where each pool would contain many recombinants with either high or low attachment efficiency. Deep sequencing on these pools and using the genome-wide association assays could reveal a smaller set of candidate genes that can be associated with the observed attachment efficiency phenotype. An analysis of the secondary inclusion phenotype differences observed between chlamydial serovars also revealed several loci that may be associated with this phenotype, including several putative inclusion membrane proteins. Many of the recombinants used in this study were IncA-negative strains and could not be assayed for secondary inclusion formation resulting in a small number of strains that could be used to study this phenotype. Serovar G strains are high secondary inclusion formers, and serovar E are low secondary formers (119). Generation of recombinants using strains from these serovars would be a good approach to finding genes that could be associated with this phenotype. Taken together, these studies show that the advent of next generation sequencing and the use of novel assays to generate hybrid chlamydial strains are powerful tools to study the biology of an organism that lacks classical genetic tools.

BIBLIOGRAPHY

1. 2010. Sexually Transmitted Diseases Treatment Guidelines. Centers for Disease Control and Prevention. MMWR **59(RR-12):**44-49.
2. **Albrecht, M., C. M. Sharma, R. Reinhardt, J. Vogel, and T. Rudel.** 2010. Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. Nucleic acids research **38:**868-877.
3. **Ambur, O. H., T. Davidsen, S. A. Frye, S. V. Balasingham, K. Lagesen, T. Rognes, and T. Tonjum.** 2009. Genome dynamics in major bacterial pathogens. FEMS Microbiol Rev **33:**453-470.
4. **Andersen A A, R. D. G.** 1998. Resistance to tetracycline and sulfadiazine in swine C. trachomatis isolates., p. 313-316. *In* S. R. S (ed.), Chlamydial Infections. Proceedings of the Ninth International Symposium on Human Chlamydial Infection.
5. **Bailey, T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:**28-36.
6. **Baldo, L., S. Bordenstein, J. J. Wernegreen, and J. H. Werren.** 2006. Widespread recombination throughout Wolbachia genomes. Molecular Biology and Evolution **23:**437-449.
7. **Bannantine, J. P., R. S. Griffiths, W. Viratyosin, W. J. Brown, and D. D. Rockey.** 2000. A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. Cell Microbiol **2:**35-47.
8. **Bannantine, J. P., W. E. Stamm, R. J. Suchland, and D. D. Rockey.** 1998. Chlamydia trachomatis IncA is localized to the inclusion membrane and is recognized by antisera from infected humans and primates. Infect Immun **66:**6017-6021.
9. **Barnes, R. C., A. M. Rompalo, and W. E. Stamm.** 1987. Comparison of Chlamydia trachomatis serovars causing rectal and cervical infections. J Infect Dis **156:**953-958.
10. **Barry, C. E., 3rd, T. J. Brickman, and T. Hackstadt.** 1993. Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. Molecular microbiology **9:**273-283.
11. **Barry, C. E., 3rd, S. F. Hayes, and T. Hackstadt.** 1992. Nucleoid condensation in Escherichia coli that express a chlamydial histone homolog. Science **256:**377-379.
12. **Beare, P. A., D. Howe, D. C. Cockrell, A. Omsland, B. Hansen, and R. A. Heinzen.** 2009. Characterization of a Coxiella burnetii ftsZ mutant generated by Himar1 transposon mutagenesis. Journal of Bacteriology **191:**1369-1381.
13. **Beare, P. A., J. E. Samuel, D. Howe, K. Virtaneva, S. F. Porcella, and R. A. Heinzen.** 2006. Genetic diversity of the Q fever agent, Coxiella burnetii, assessed by microarray-based whole-genome comparisons. J Bacteriol **188:**2309-2324.

14. **Belland, R. J., D. E. Nelson, D. Virok, D. D. Crane, D. Hogan, D. Sturdevant, W. L. Beatty, and H. D. Caldwell.** 2003. Transcriptome analysis of chlamydial growth during IFN-gamma-mediated persistence and reactivation. Proceedings of the National Academy of Sciences of the United States of America **100:**15971-15976.
15. **Belland, R. J., M. A. Scidmore, D. D. Crane, D. M. Hogan, W. Whitmire, G. McClarty, and H. D. Caldwell.** 2001. Chlamydia trachomatis cytotoxicity associated with complete and partial cytotoxin genes. Proc Natl Acad Sci USA **98:**13984-13989.
16. **Binet, R., and A. T. Maurelli.** 2009. Transformation and isolation of allelic exchange mutants of Chlamydia psittaci using recombinant DNA introduced by electroporation. Proc Natl Acad Sci U S A **106:**292-297.
17. **Binet, R., and A. T. Maurelli.** 2009. Transformation and isolation of allelic exchange mutants of Chlamydia psittaci using recombinant DNA introduced by electroporation. Proceedings of the National Academy of Sciences of the United States of America **106:**292-297.
18. **Boisvert, J.-F., L. A. Koutsky, R. J. Suchland, and W. E. Stamm.** 1999. Clinical Features of Chlamydia trachomatis Rectal Infection by Serovar Among Homosexually Active Men. Sexually transmitted diseases **26:**392-398.
19. **Brinkman, F. S., J. L. Blanchard, A. Cherkasov, Y. Av-Gay, R. C. Brunham, R. C. Fernandez, B. B. Finlay, S. P. Otto, B. F. Ouellette, P. J. Keeling, A. M. Rose, R. E. Hancock, S. J. Jones, and H. Greberg.** 2002. Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. Genome research **12:**1159-1167.
20. **Brunham, R., C. Yang, I. Maclean, J. Kimani, G. Maitha, and F. Plummer.** 1994. Chlamydia trachomatis from individuals in a sexually transmitted disease core group exhibit frequent sequence variation in the major outer membrane protein (omp1) gene. J Clin Invest **94:**458-463.
21. **Caldwell, H. D., J. Kromhout, and J. Schachter.** 1981. Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. Infect Immun **31:**1161-1176.
22. **Caldwell, H. D., and J. Schachter.** 1982. Antigenic analysis of the major outer membrane protein of Chlamydia spp. Infect Immun **35:**1024-1031.
23. **Caldwell, H. D., H. Wood, D. Crane, R. Bailey, R. B. Jones, D. Mabey, I. Maclean, Z. Mohammed, R. Peeling, C. Roshick, J. Schachter, A. W. Solomon, W. E. Stamm, R. J. Suchland, L. Taylor, S. K. West, T. C. Quinn, R. J. Belland, and G. McClarty.** 2003. Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Invest **111:**1757-1769.
24. **Carlson, J. H., S. Hughes, D. Hogan, G. Cieplak, D. E. Sturdevant, G. McClarty, H. D. Caldwell, and R. J. Belland.** 2004. Polymorphisms in the Chlamydia trachomatis cytotoxin locus associated with ocular and genital isolates. Infect Immun **72:**7063-7072.

25. **Carlson, J. H., S. F. Porcella, G. McClarty, and H. D. Caldwell.** 2005. Comparative genomic analysis of Chlamydia trachomatis oculotropic and genitotropic strains. Infect Immun **73:**6407-6418.
26. **Carlson, J. H., H. Wood, C. Roshick, H. D. Caldwell, and G. McClarty.** 2006. In vivo and in vitro studies of Chlamydia trachomatis TrpR:DNA interactions. Mol Microbiol **59:**1678-1691.
27. **Chen, D., L. Lei, C. Lu, A. Galaleldeen, P. J. Hart, and G. Zhong.** 2010. Characterization of Pgp3, a Chlamydia trachomatis plasmid-encoded immunodominant antigen. Journal of Bacteriology **192:**6017-6024.
28. **Clifton, D. R., K. A. Fields, S. S. Grieshaber, C. A. Dooley, E. R. Fischer, D. J. Mead, R. A. Carabeo, and T. Hackstadt.** 2004. A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. Proc Natl Acad Sci U S A **101:**10166-10171.
29. **Crane, D. D., J. H. Carlson, E. R. Fischer, P. Bavoil, R. C. Hsia, C. Tan, C. C. Kuo, and H. D. Caldwell.** 2006. Chlamydia trachomatis polymorphic membrane protein D is a species-common pan-neutralizing antigen. Proc Natl Acad Sci USA **103:**1894-1899.
30. **Dautry-Varsat, A., A. Subtil, and T. Hackstadt.** 2005. Recent insights into the mechanisms of Chlamydia entry. Cellular microbiology **7:**1714-1722.
31. **Dean, D., R. J. Suchland, and W. E. Stamm.** 2000. Evidence for long-term cervical persistence of Chlamydia trachomatis by omp1 genotyping. J Infect Dis **182:**909-916.
32. **Delevoye, C., M. Nilges, A. Dautry-Varsat, and A. Subtil.** 2004. Conservation of the biochemical properties of IncA from Chlamydia trachomatis and Chlamydia caviae: oligomerization of IncA mediates interaction between facing membranes. J Biol Chem **279:**46896-46906.
33. **DeMars, R., and J. Weinfurter.** 2008. Interstrain gene transfer in Chlamydia trachomatis in vitro: mechanism and significance. J Bacteriol **190:**1605-1614.
34. **Demars, R., J. Weinfurter, E. Guex, J. Lin, and Y. Potucek.** 2007. Lateral gene transfer in vitro in the intracellular pathogen Chlamydia trachomatis. J Bacteriol **189:**991-1003.
35. **Di Francesco, A., M. Donati, M. Rossi, S. Pignanelli, A. Shurdhi, R. Baldelli, and R. Cevenini.** 2008. Tetracycline-resistant Chlamydia suis isolates in Italy. Vet Rec **163:**251-252.
36. **Dugan, J., A. A. Andersen, and D. D. Rockey.** 2007. Functional characterization of IScs605, an insertion element carried by tetracycline-resistant Chlamydia suis. Microbiology (Reading, Engl) **153:**71-79.
37. **Dugan, J., D. D. Rockey, L. Jones, and A. A. Andersen.** 2004. Tetracycline resistance in Chlamydia suis mediated by genomic islands inserted into the chlamydial inv-like gene. Antimicrob Agents Chemother **48:**3989-3995.
38. **Ellison, D. W., T. R. Clark, D. E. Sturdevant, K. Virtaneva, S. F. Porcella, and T. Hackstadt.** 2008. Genomic comparison of virulent

Rickettsia rickettsii Sheila Smith and avirulent Rickettsia rickettsii Iowa. Infect Immun **76:**542-550.

39. **Fehlner-Gardiner, C., C. Roshick, J. H. Carlson, S. Hughes, R. J. Belland, H. D. Caldwell, and G. McClarty.** 2002. Molecular basis defining human Chlamydia trachomatis tissue tropism. A possible role for tryptophan synthase. J Biol Chem **277:**26893-26903.
40. **Fields, K. A., D. J. Mead, C. A. Dooley, and T. Hackstadt.** 2003. Chlamydia trachomatis type III secretion: evidence for a functional apparatus during early-cycle development. Molecular microbiology **48:**671-683.
41. **Geisler, W. M., R. J. Suchland, D. D. Rockey, and W. E. Stamm.** 2001. Epidemiology and clinical manifestations of unique Chlamydia trachomatis isolates that occupy nonfusogenic inclusions. J Infect Dis **184:**879-884.
42. **Geisler, W. M., W. L. Whittington, R. J. Suchland, and W. E. Stamm.** 2002. Epidemiology of anorectal chlamydial and gonococcal infections among men having sex with men in Seattle: utilizing serovar and auxotype strain typing. Sexually transmitted diseases **29:**189-195.
43. **Gieffers, J., R. J. Belland, W. Whitmire, S. Ouellette, D. Crane, M. Maass, G. I. Byrne, and H. D. Caldwell.** 2002. Isolation of Chlamydia pneumoniae clonal variants by a focus-forming assay. Infection and Immunity **70:**5827-5834.
44. **Giles, T. N., D. J. Fisher, and D. E. Graham.** 2009. Independent inactivation of arginine decarboxylase genes by nonsense and missense mutations led to pseudogene formation in Chlamydia trachomatis serovar L2 and D strains. BMC Evol Biol **9:**166.
45. **Gomes, J. P., W. J. Bruno, M. J. Borrego, and D. Dean.** 2004. Recombination in the genome of Chlamydia trachomatis involving the polymorphic membrane protein C gene relative to ompA and evidence for horizontal gene transfer. J Bacteriol **186:**4295-4306.
46. **Gomes, J. P., W. J. Bruno, A. Nunes, N. Santos, C. Florindo, M. J. Borrego, and D. Dean.** 2007. Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots. Genome Res **17:**50-60.
47. **Gomes, J. P., A. Nunes, W. J. Bruno, M. J. Borrego, C. Florindo, and D. Dean.** 2006. Polymorphisms in the nine polymorphic membrane proteins of Chlamydia trachomatis across all serovars: evidence for serovar Da recombination and correlation with tissue tropism. J Bacteriol **188:**275-286.
48. **Greub, G., F. Collyn, L. Guy, and C. A. Roten.** 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 microbiology **4:**48.
49. **Hackstadt, T., W. Baehr, and Y. Ying.** 1991. Chlamydia trachomatis developmentally regulated protein is homologous to eukaryotic histone

H1. Proceedings of the National Academy of Sciences of the United States of America **88:**3937-3941.

50. **Hackstadt, T., T. J. Brickman, C. E. Barry, 3rd, and J. Sager.** 1993. Diversity in the Chlamydia trachomatis histone homologue Hc2. Gene **132:**137-141.
51. **Hackstadt, T., E. R. Fischer, M. A. Scidmore, D. D. Rockey, and R. A. Heinzen.** 1997. Origins and functions of the chlamydial inclusion. Trends Microbiol **5:**288-293.
52. **Hackstadt, T., M. A. Scidmore-Carlson, E. I. Shaw, and E. R. Fischer.** 1999. The Chlamydia trachomatis IncA protein is required for homotypic vesicle fusion. Cell Microbiol **1:**119-130.
53. **Heinzen, R. A., M. A. Scidmore, D. D. Rockey, and T. Hackstadt.** 1996. Differential interaction with endocytic and exocytic pathways distinguish parasitophorous vacuoles of Coxiella burnetii and Chlamydia trachomatis. Infect Immun **64:**796-809.
54. **Hintz, N. J., D. G. Ennis, W. F. Liu, and S. H. Larsen.** 1995. The recA gene of Chlamydia trachomatis: cloning, sequence, and characterization in Escherichia coli. FEMS microbiology letters **127:**175-180.
55. **Horn, M., A. Collingro, S. Schmitz-Esser, C. L. Beier, U. Purkhold, B. Fartmann, P. Brandt, G. J. Nyakatura, M. Droege, D. Frishman, T. Rattei, H.-W. Mewes, and M. Wagner.** 2004. Illuminating the evolutionary history of chlamydiae. Science **304:**728-730.
56. **Hsia, R. C., and P. M. Bavoil.** 1996. Homologs of Escherichia coli recJ, gltX and of a putative 'early' gene of avian Chlamydia psittaci are located upstream of the 'late' omp2 locus of Chlamydia psittaci strain guinea pig inclusion conjunctivitis. Gene **176:**163-169.
57. **Hsia, R. C., L. M. Ting, and P. M. Bavoil.** 2000. Microvirus of chlamydia psittaci strain guinea pig inclusion conjunctivitis: isolation and molecular characterization. Microbiology **146 (Pt 7):**1651-1660.
58. **Hueck, C. J.** 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev **62:**379-433.
59. **Hybiske, K., and R. S. Stephens.** 2007. Mechanisms of host cell exit by the intracellular bacterium Chlamydia. Proceedings of the National Academy of Sciences of the United States of America **104:**11430-11435.
60. **Jamison, W. P., and T. Hackstadt.** 2008. Induction of type III secretion by cell-free Chlamydia trachomatis elementary bodies. Microb Pathog **45:**435-440.
61. **Jeck, W. R., J. A. Reinhardt, D. A. Baltrus, M. T. Hickenbotham, V. Magrini, E. R. Mardis, J. L. Dangl, and C. D. Jones.** 2007. Extending assembly of short DNA sequences to handle error. Bioinformatics **23:**2942-2944.
62. **Jeffrey, B. M., R. J. Suchland, K. L. Quinn, J. R. Davidson, W. E. Stamm, and D. D. Rockey.** 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.

63. **Kari, L., W. M. Whitmire, J. H. Carlson, D. D. Crane, N. Reveneau, D. E. Nelson, D. C. Mabey, R. L. Bailey, M. J. Holland, G. McClarty, and H. D. Caldwell.** 2008. Pathogenic diversity among Chlamydia trachomatis ocular strains in nonhuman primates is affected by subtle genomic variations. J Infect Dis **197:**449-456.
64. **Karunakaran, K. P., J. F. Blanchard, A. Raudonikiene, C. Shen, A. D. Murdin, and R. C. Brunham.** 2002. Molecular detection and seroepidemiology of the Chlamydia pneumoniae bacteriophage (PhiCpn1). Journal of clinical microbiology **40:**4010-4014.
65. **Katoh, K., G. Asimenos, and H. Toh.** 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol **537:**39-64.
66. **Katoh, K., K. Misawa, K. Kuma, and T. Miyata.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30:**3059.
67. **Kim, J. F.** 2001. Revisiting the chlamydial type III protein secretion system: clues to the origin of type III protein secretion. Trends in genetics : TIG **17:**65-69.
68. **Kiselev, A. O., W. E. Stamm, J. R. Yates, and M. F. Lampe.** 2007. Expression, processing, and localization of PmpD of Chlamydia trachomatis serovar L2 during the chlamydial developmental cycle. PLoS ONE **2:**e568.
69. **Klint, M., H.-H. Fuxelius, R. R. Goldkuhl, H. Skarin, C. Rutemark, S. G. E. Andersson, K. Persson, and B. Herrmann.** 2007. High-resolution genotyping of Chlamydia trachomatis strains by multilocus sequence analysis. J Clin Microbiol **45:**1410-1414.
70. **Kubo, A., and R. S. Stephens.** 2000. Characterization and functional analysis of PorB, a Chlamydia porin and neutralizing target. Molecular microbiology **38:**772-780.
71. **Kuo, C. C., and T. Grayston.** 1976. Interaction of Chlamydia trachomatis organisms and HeLa 229 cells. Infection and Immunity **13:**1103-1109.
72. **L'Abee-Lund, T. M., and H. Sorum.** 2002. A global non-conjugative Tet C plasmid, pRAS3, from Aeromonas salmonicida. Plasmid **47:**172-181.
73. **Lampe, M. F., R. J. Suchland, and W. E. Stamm.** 1993. Nucleotide sequence of the variable domains within the major outer membrane protein gene from serovariants of Chlamydia trachomatis. Infect Immun **61:**213-219.
74. **Laszlo Kari, M. M. G., Linnell B. Randall, Lacey D. Taylor, John H. Carlson, William M. Whitmire, Dezso Virok, Krithika Rajaram, Valeria Endresz, Grant McClarty, David E. Nelson and Harlan D. Caldwell.** 2011. Generation of targeted Chlamydia trachomatis null mutants. Proceedings of the National Academy of Sciences **in press:**in press.
75. **Lenart, J., A. A. Andersen, and D. D. Rockey.** 2001. Growth and development of tetracycline-resistant Chlamydia suis. Antimicrob Agents Chemother **45:**2198-2203.

76. **Li, H., J. Ruan, and R. Durbin.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res **18:**1851-1858.
77. **Lundemose, A. G., J. E. Kay, and J. H. Pearce.** 1993. Chlamydia trachomatis Mip-like protein has peptidyl-prolyl cis/trans isomerase activity that is inhibited by FK506 and rapamycin and is implicated in initiation of chlamydial infection. Molecular microbiology **7:**777-783.
78. **Mabey, D.** 2008. Trachoma: recent developments. Adv Exp Med Biol **609:**98-107.
79. **McCoy, A. J., N. E. Adams, A. O. Hudson, C. Gilvarg, T. Leustek, and A. T. Maurelli.** 2006. L,L-diaminopimelate aminotransferase, a trans-kingdom enzyme shared by Chlamydia and plants for synthesis of diaminopimelate/lysine. Proceedings of the National Academy of Sciences of the United States of America **103:**17909-17914.
80. **Millman, K. L., S. Tavaré, and D. Dean.** 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. J Bacteriol **183:**5997-6008.
81. **Molleken, K., E. Schmidt, and J. H. Hegemann.** 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.
82. **Nunes, A., P. Nogueira, M. Borrego, and J. Gomes.** 2008. Chlamydia trachomatis diversity viewed as a tissue-specific coevolutionary arms race. Genome Biology **9:**R153.
83. **O'Connell, C. M., and K. M. Nicks.** 2006. A plasmid-cured Chlamydia muridarum strain displays altered plaque morphology and reduced …. Microbiology.
84. **O'Connell, C. M. C., and A. T. Maurelli.** 1998. Introduction of foreign DNA into *Chlamydia* and stable expression of chloramphenicol resistance. Chlamydial Infections, Proceedings of the ninth international symposium on human chlamydial infection**:**519-522.
85. **Olivares-Zavaleta, N., W. Whitmire, D. Gardner, and H. D. Caldwell.** 2010. Immunization with the attenuated plasmidless Chlamydia trachomatis L2(25667R) strain provides partial protection in a murine model of female genitourinary tract infection. Vaccine **28:**1454-1462.
86. **Omsland, A., D. C. Cockrell, E. R. Fischer, and R. A. Heinzen.** 2008. Sustained axenic metabolic activity by the obligate intracellular bacterium Coxiella burnetii. J Bacteriol **190:**3203-3212.
87. **Omsland, A., D. C. Cockrell, D. Howe, E. R. Fischer, K. Virtaneva, D. E. Sturdevant, S. F. Porcella, and R. A. Heinzen.** 2009. Host cell-free growth of the Q fever bacterium Coxiella burnetii. Proc Natl Acad Sci USA **106:**4430-4434.
88. **Pannekoek, Y., G. Morelli, B. Kusecek, S. A. Morré, J. M. Ossewaarde, A. A. Langerak, and A. van der Ende.** 2008. Multi locus sequence typing

of Chlamydiales: clonal groupings within the obligate intracellular bacteria Chlamydia trachomatis. BMC Microbiol **8:**42.

89. **Perara, E., D. Ganem, and J. N. Engel.** 1992. A developmentally regulated chlamydial gene with apparent homology to eukaryotic histone H1. Proceedings of the National Academy of Sciences of the United States of America **89:**2125-2129.
90. **Raulston, J. E., C. H. Davis, T. R. Paul, J. D. Hobbs, and P. B. Wyrick.** 2002. Surface accessibility of the 70-kilodalton Chlamydia trachomatis heat shock protein following reduction of outer membrane protein disulfide bonds. Infection and Immunity **70:**535-543.
91. **Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser.** 2000. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res **28:**1397-1406.
92. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16:**276-277.
93. **Rockey, D. D., M. A. Scidmore, J. P. Bannantine, and W. J. Brown.** 2002. Proteins in the chlamydial inclusion membrane. Microbes Infect **4:**333-340.
94. **Rockey, D. D., W. Viratyosin, J. P. Bannantine, R. J. Suchland, and W. E. Stamm.** 2002. Diversity within inc genes of clinical Chlamydia trachomatis variant isolates that occupy non-fusogenic inclusions. Microbiology **148:**2497-2505.
95. **Roshick, C., H. Wood, H. D. Caldwell, and G. McClarty.** 2006. Comparison of gamma interferon-mediated antichlamydial defense mechanisms in human and mouse cells. Infection and Immunity **74:**225-238.
96. **Rupp, J., W. Solbach, and J. Gieffers.** 2007. Prevalence, genetic conservation and transmissibility of the Chlamydia pneumoniae bacteriophage (phiCpn1). FEMS microbiology letters **273:**45-49.
97. **Russell, M., T. Darville, K. Chandra-Kuntal, B. Smith, C. W. Andrews, Jr., and C. M. O'Connell.** 2011. Infectivity acts as in vivo selection for maintenance of the chlamydial cryptic plasmid. Infection and Immunity **79:**98-107.
98. **Sanchez, J., J. R. Lama, L. Kusunoki, H. Manrique, P. Goicochea, A. Lucchetti, M. Rouillon, M. Pun, L. Suarez, S. Montano, J. L. Sanchez, S. Tabet, J. P. Hughes, and C. Celum.** 2007. HIV-1, sexually transmitted infections, and sexual behavior trends among men who have sex with men in Lima, Peru. J Acquir Immune Defic Syndr **44:**578-585.
99. **Sandoz, K. M., and D. D. Rockey.** 2010. Antibiotic resistance in Chlamydiae. Future Microbiol **5:**1427-1442.
100. **Sauer, J. D., J. G. Shannon, D. Howe, S. F. Hayes, M. S. Swanson, and R. A. Heinzen.** 2005. Specificity of Legionella pneumophila and

Coxiella burnetii vacuoles and versatility of Legionella pneumophila revealed by coinfection. Infection and Immunity **73:**4494-4504.

101. **Schachter, J.** 1978. Chlamydial infections (first of three parts). N Engl J Med **298:**428-435.
102. **Scidmore, M. A., D. D. Rockey, E. R. Fischer, R. A. Heinzen, and T. Hackstadt.** 1996. Vesicular interactions of the Chlamydia trachomatis inclusion are determined by chlamydial early protein synthesis rather than route of entry. Infect Immun **64:**5366-5372.
103. **Scidmore-Carlson, M., and T. Hackstadt.** 2000. Chlamydia internalization and intracellular fate. Subcell Biochem **33:**459-478.
104. **Setlow, P.** 2007. I will survive: DNA protection in bacterial spores. Trends in Microbiology **15:**172-180.
105. **Shaw, A. C., K. Gevaert, H. Demol, B. Hoorelbeke, J. Vandekerckhove, M. R. Larsen, P. Roepstorff, A. Holm, G. Christiansen, and S. Birkelund.** 2002. Comparative proteome analysis of Chlamydia trachomatis serovar A, D and L2. Proteomics **2:**164-186.
106. **Smith, C. B., and D. E. Graham.** 2008. Outer and inner membrane proteins compose an arginine-agmatine exchange system in Chlamydophila pneumoniae. J Bacteriol **190:**7431-7440.
107. **Smith, G. R.** 2001. Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. Annu Rev Genet **35:**243-274.
108. **Stamm, W. E.** 1999. Chlamydia trachomatis infections: progress and problems. The Journal of infectious diseases **179 Suppl 2:**S380-383.
109. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis.** 1998. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science **282:**754-759.
110. **Stewart, P. E., J. Hoff, E. Fischer, J. G. Krum, and P. A. Rosa.** 2004. Genome-wide transposon mutagenesis of Borrelia burgdorferi for identification of phenotypic mutants. Applied and environmental microbiology **70:**5973-5979.
111. **Storey, C. C., M. Lusher, and S. J. Richmond.** 1989. Analysis of the complete nucleotide sequence of Chp1, a phage which infects avian Chlamydia psittaci. J Gen Virol **70 (Pt 12):**3381-3390.
112. **Su, H., N. G. Watkins, Y. X. Zhang, and H. D. Caldwell.** 1990. Chlamydia trachomatis-host cell interactions: role of the chlamydial major outer membrane protein as an adhesin. Infection and Immunity **58:**1017-1025.
113. **Su, H., Y. X. Zhang, O. Barrera, N. G. Watkins, and H. D. Caldwell.** 1988. Differential effect of trypsin on infectivity of Chlamydia trachomatis: loss of infectivity requires cleavage of major outer membrane protein variable domains II and IV. Infection and Immunity **56:**2094-2100.
114. **Subtil, A., A. Blocker, and A. Dautry-Varsat.** 2000. Type III secretion system in Chlamydia species: identified members and candidates. Microbes and infection / Institut Pasteur **2:**367-369.

115. **Suchland, R. J., A. Bourillon, E. Denamur, W. E. Stamm, and D. M. Rothstein.** 2005. Rifampin-resistant RNA polymerase mutants of Chlamydia trachomatis remain susceptible to the ansamycin rifalazil. Antimicrob Agents Chemother **49:**1120-1126.
116. **Suchland, R. J., L. O. Eckert, S. E. Hawes, and W. E. Stamm.** 2003. Longitudinal assessment of infecting serovars of Chlamydia trachomatis in Seattle public health clinics: 1988-1996. Sexually transmitted diseases **30:**357-361.
117. **Suchland, R. J., B. M. Jeffrey, M. Xia, A. Bhatia, H. G. Chu, D. D. Rockey, and W. E. Stamm.** 2008. Identification of concomitant infection with Chlamydia trachomatis IncA-negative mutant and wild-type strains by genomic, transcriptional, and biological characterizations. Infect Immun **76:**5438-5446.
118. **Suchland, R. J., D. D. Rockey, J. P. Bannantine, and W. E. Stamm.** 2000. Isolates of Chlamydia trachomatis that occupy nonfusogenic inclusions lack IncA, a protein localized to the inclusion membrane. Infect Immun **68:**360-367.
119. **Suchland, R. J., D. D. Rockey, S. K. Weeks, D. T. Alzhanov, and W. E. Stamm.** 2005. Development of secondary inclusions in cells infected by Chlamydia trachomatis. Infect Immun **73:**3954-3962.
120. **Suchland, R. J., K. M. Sandoz, B. M. Jeffrey, W. E. Stamm, and D. D. Rockey.** 2009. Horizontal transfer of tetracycline resistance among Chlamydia spp. in vitro. Antimicrob Agents Chemother **53:**4604-4611.
121. **Suchland, R. J., and W. E. Stamm.** 1991. Simplified microtiter cell culture method for rapid immunotyping of Chlamydia trachomatis. J Clin Microbiol **29:**1333-1338.
122. **Swanson, A. F., and C. C. Kuo.** 1994. Binding of the glycan of the major outer membrane protein of Chlamydia trachomatis to HeLa cells. Infection and Immunity **62:**24-28.
123. **Tam, J. E., C. H. Davis, and P. B. Wyrick.** 1994. Expression of recombinant DNA introduced into Chlamydia trachomatis by electroporation. Can J Microbiol **40:**583-591.
124. **Tan, C., R. C. Hsia, H. Shou, C. L. Haggerty, R. B. Ness, C. A. Gaydos, D. Dean, A. M. Scurlock, D. P. Wilson, and P. M. Bavoil.** 2009. Chlamydia trachomatis-infected patients display variable antibody profiles against the nine-member polymorphic membrane protein family. Infect Immun **77:**3218-3226.
125. **Thomson, N. R., M. T. Holden, C. Carder, N. Lennard, S. J. Lockey, P. Marsh, P. Skipp, C. D. O'Connor, I. Goodhead, H. Norbertzcak, B. Harris, D. Ormond, R. Rance, M. A. Quail, J. Parkhill, R. S. Stephens, and I. N. Clarke.** 2008. Chlamydia trachomatis: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res **18:**161-171.
126. **Thomson, N. R., C. Yeats, K. Bell, M. T. Holden, S. D. Bentley, M. Livingstone, A. M. Cerdeño-Tárraga, B. Harris, J. Doggett, D. Ormond, K. Mungall, K. Clarke, T. Feltwell, Z. Hance, M. Sanders, M.**

A. Quail, C. Price, B. G. Barrell, J. Parkhill, and D. Longbottom. 2005. The Chlamydophila abortus genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Res **15:**629-640.

127. **Ting, L. M., R. C. Hsia, C. G. Haidaris, and P. M. Bavoil.** 1995. Interaction of outer envelope proteins of Chlamydia psittaci GPIC with the HeLa cell surface. Infection and Immunity **63:**3600-3608.
128. **Vidal, L., J. Pinsach, G. Striedner, G. Caminal, and P. Ferrer.** 2008. Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in Escherichia coli. J Biotechnol **134:**127-136.
129. **Voth, D. E., P. A. Beare, D. Howe, U. M. Sharma, G. Samoilis, D. C. Cockrell, A. Omsland, and R. A. Heinzen.** 2011. The Coxiella burnetii Cryptic Plasmid is Enriched in Genes Encoding Type IV Secretion System Substrates. Journal of Bacteriology.
130. **Wehrl, W., V. Brinkmann, P. R. Jungblut, T. F. Meyer, and A. J. Szczepek.** 2004. From the inside out--processing of the Chlamydial autotransporter PmpD and its role in bacterial adhesion and activation of human host cells. Mol Microbiol **51:**319-334.
131. **Wizel, B., B. C. Starcher, B. Samten, Z. Chroneos, P. F. Barnes, J. Dzuris, Y. Higashimoto, E. Appella, and A. Sette.** 2002. Multiple Chlamydia pneumoniae antigens prime CD8+ Tc1 responses that inhibit intracellular growth of this vacuolar pathogen. J Immunol **169:**2524-2535.
132. **Xia, M., R. J. Suchland, R. E. Bumgarner, T. Peng, D. D. Rockey, and W. E. Stamm.** 2005. Chlamydia trachomatis variant with nonfusing inclusions: growth dynamic and host-cell transcriptional response. J Infect Dis **192:**1229-1236.
133. **Yuan, Y., Y. X. Zhang, N. G. Watkins, and H. D. Caldwell.** 1989. Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 Chlamydia trachomatis serovars. Infect Immun **57:**1040-1049.
134. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821-829.
135. **Zhang, J. P., and R. S. Stephens.** 1992. Mechanism of C. trachomatis attachment to eukaryotic host cells. Cell **69:**861-869.