Genome-wide Investigation into Chlamydial Factors Important in Host-Specific Pathogenesis

By

Zoe E. Dimond

Submitted to the graduate degree program in Molecular Biosciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree Doctor of Philosophy.

_____
Chairperson (P. Scott Hefty, Ph.D.)

_____ *
(Lynn E. Hancock, Ph.D.)

_____ *
(Susan M. Egan, Ph.D.)

_____ *
(David J. Davido, Ph.D.)

_____ *
(James Walters, Ph.D.)

*Committee Members
Date Defended: November 10th, 2020

The dissertation committee for Zoe E. Dimond certifies that this is the approved version of the following dissertation:

Genome-wide Investigation into Chlamydial Factors Important in Host-Specific Pathogenesis

_____

Chairperson (P. Scott Hefty, Ph.D.)

Date Approved: November 10th, 2020

Abstract

Chlamydia are unique obligate intracellular organisms responsible for disease in a wide variety of hosts. A defining feature of these bacteria is their biphasic developmental cycle, through which *Chlamydia* interact with the host cytoplasm using surface proteins and secreted effectors. As a major public health pathogen, *Chlamydia trachomatis* is the causative agent behind nearly two million sexually-transmitted infections in the United States every year. Symptoms of chlamydia can be permanent and life-threatening, including infertility, ectopic pregnancy and pelvic inflammatory disease, although for many, infections are asymptomatic. There remain many mysteries about how *Chlamydia* are able to cause disease in some patients while remaining unnoticeable in others. Prominently among these mysteries is how *Chlamydia* undergo ascension, or disseminate to the fallopian tubes to cause these complicated sequelae. Importantly, mouse models for *C. trachomatis* are limited, as *C. trachomatis* is an obligate human pathogen, unable to mimic human pathogenesis in the mouse. Closely related *C. muridarum*, an obligate mouse pathogen, shares 99% of genetic content with *C. trachomatis.* With this dissertation, I seek to answer how species with such genomic conservation can be host-specific.

Host-specificity is common amongst bacteria and is often mediated by factors such as surface receptors, immune evasion effectors and nutrient acquisition components. In *Chlamydia*, very few of these factors have been identified. Therefore, I used comparative genomics and chimeric recombinants to investigate the host-specificity of *Chlamydia.* Through the complete genome sequencing and annotation of *Chlamydia suis*, a close relative of both *C. trachomatis* and *C. muridarum*, I demonstrated that *C. suis* contains a subset of unique proteins, as well as shared putative host-interacting proteins with those found in *C. muridarum* and/or *C. trachomatis*.

C. muridarum and *C. trachomatis* genomes were crossed to generate a library of chimeras where *C. muridarum* genes were integrated into the *C. trachomatis* genome using homologous recombination. With these chimeras, I investigated the role of *C. muridarum* genes on pathogenesis phenotypes. First, I focused on the plasticity zone, a region of diversity in *Chlamydia*, thought to be important for mouse infections and host cell toxicity *in vitro*. Replacement of *C. trachomatis* plasticity zone genes with *C. muridarum* genes showed that the plasticity zone genes are partially involved in regulating the developmental cycle and inclusion morphology, as some plasticity zone chimeras had reduced progeny 36 hours post-infection and significantly smaller inclusion sizes than wildtype *C. trachomatis.* In the mouse model, *C. muridarum* plasticity zone genes did not enable the recombinants to ascend to upper reproductive tissues, nor were they more successful in avoiding clearance or establishing increased bacterial burdens in the lower reproductive tract than *C. trachomatis*. Three genes within the *C. muridarum* plasticity zone, referred to as large putative cytotoxins, which have long been considered responsible for an *in vitro* host cell toxicity phenotype of *C. muridarum,* were shown to be insufficient for this phenotype and additional genes may be involved. Additionally, an inclusion membrane protein from *C. muridarum* was unable to perform the function of the *C. trachomatis* ortholog, suggesting that this protein interacts with another chlamydial protein to stabilize the inclusion.

Second, chimeric recombinants across the genome were examined in a growth curve screen where two unique growth phenotypes were discovered; a poor growing mutant and a mutant which produces infectious progeny faster than *C. trachomatis.* Regions of recombination for these chimeras suggest candidate genes involved in these phenotypes. Mouse models for these

recombinants reveal that *C. muridarum* pathogenesis and ascension in the mouse is likely multifactorial as no recombinant was capable of ascension. Taken together, I have used comparative genomics and recombinant libraries to increase our knowledge of the plasticity zone, demonstrate the use of recombination in *Chlamydia* as a screening tool for host-specific phenotypes and suggest that *C. muridarum* and *C. trachomatis* orthologs are largely interchangeable, but that small coding changes can lead to biologically relevant phenotypes.

Acknowledgements

I would like to first and foremost thank my mentor Dr. Scott Hefty for all of his guidance. He has shown me what kind of scientist I strive to be and his tireless passion for scientific discovery is inspiring. Over the last four years, he has provided every opportunity for me to learn about this career that I have chosen and I cannot thank him enough for helping to prepare me for the next phase. I also want to thank my collaborators Dr. Daniel Rockey, Dr. Kevin Hybiske and Bob Suchland for their mentorship throughout my graduate career. I could not have asked for better role models. Additionally, I need to thank all of my committee members, Dr. Lynn Hancock, Dr. Susan Egan, Dr. David Davido, Dr. Jamie Walters and the late Dr. Steve Benedict. The support and advice from these amazing scientists will stay with me for years to come and help me become a better mentor one day.

Much of my dissertation is collaborative and I want to thank the members of the Hefty Lab who not only contributed scientifically to this work, but who made my time here more enjoyable. I want to particularly thank Scott LaBrie and Srishti Baid. Scott and Srishti have supported me through the challenges of graduate school and celebrated with me through the joys. Outside of the Hefty Lab, I want to thank Priyanka Goyal and Cindy Ly for their encouragement and friendship which has made all of this possible.

Finally, I want to thank my family, without whom I would never have been brave enough to start this journey in the first place. My parents, Allan and Carol Seigel, have always pushed me to be my best and have given me the love and support to follow my dream. And last, but perhaps most important, my husband Zack, for going on this crazy adventure with me and standing by me every step of the way. This would have been impossible without his steady support, constant encouragement, occasional pushing and unending love.

Table of Contents

Chapter 1.

Introduction

An evolutionary history of *Chlamydia*

Species contained within the phylum, *Chlamydiae*, are considered to be some of the most ancient groups of obligate intracellular bacteria, with a most recent common ancestor over 700 million years ago (1). Initial discovery of "Chlamydozoa," named for the Greek term for cloak, occurred when scientists, Halberstadter and von Prowazek traveled to Indonesia to investigate the causative agent behind syphilis, but discovered the ocular infection, trachoma, as an outcome (2, 3). Originally thought to be protozoa or even viral particles (4, 5), *Chlamydiae* serve as unique members in the kingdom of *Bacteria*. Most notably, every species within the phylum are obligate intracellular bacteria which require eukaryotic hosts for replication (2, 6, 7).
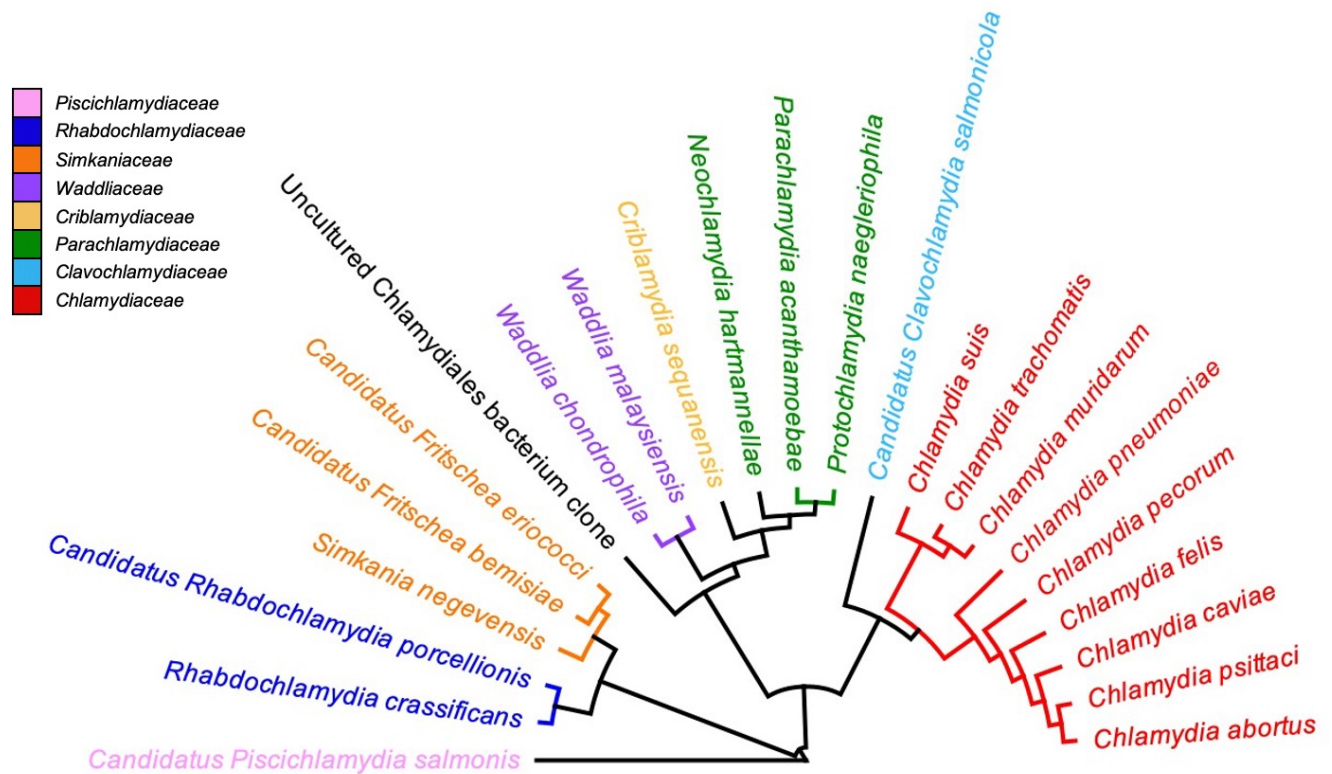
Figure 1-1. 16S rRNA phylogeny of *Chlamydiae*.

Many of the families within the phylum are classic endosymbionts. It is hypothesized that perhaps these bacteria have evolved over millennia from common amoebal commensals, of which many still exist, into symbionts of the more complex multicellular eukaryotes, and most recently developing into the mammalian pathogens within the most-studied genus, *Chlamydia* (5). This genus contains nine canonical chlamydial species (Figure 1-1, red). These species are known to infect a diverse group of hosts and cause a variety of diseases. Beyond this genus are several other families including *Waddliaceae, Parachlamydiaceae, Rhabdochlamydiaceae* and *Simkaniaceae,* which are largely detected in fish, invertebrates and protozoa (2, 7).

These non-*Chlamydia Chlamydiae*, often referred to as *Chlamydia*-like organisms, have been isolated from humans, and water sources as well, likely due to their presence in *Acanthamoeba*, a ubiquitous protozoon common in soil, marine and fresh water, air and even air conditioners (8). The complete diversity of *Chlamydiae* is still undiscovered, as novel species are added into the phylum each year with the advancement of metagenomic studies in regions like ocean sediment (2, 9). While these *Chlamydia*-like species have been detected in humans, the question of whether they contribute to human disease remains open. *Chlamydia*-like bacteria have been detected in human respiratory disease samples, conjunctivitis specimens, aborted ruminant placenta and even marsupial kidneys in cases of nephritis (10). Despite their presence in disease-state tissues, these bacteria have not been shown to meet the standard benchmark for pathogen determination, Koch's Postulates, which states that an agent must be present in all cases of disease and that the disease must be reproduced when the agent is inoculated into a new individual (2). However, many suggest that application of Koch's Postulates as a sole definition of pathogen requires flexible

interpretation as our understanding of the microbiome and bacterial pathogenesis improves (11-13).

Study of *Chlamydia*-like organisms has led to many advances in understanding the basic biology of all *Chlamydiae*, however. These advances include the discovery of peptidoglycan in chlamydial cell walls (14), mechanisms for cell division (15), and analysis of the metabolism involved in the developmental cycle of *Chlamydiae* (16). Importantly, these studies were enabled by the fact that 560 core genes have been identified as conserved between all *Chlamydiae* species with genomes available, despite the fact that the *Chlamydia*-like organism genomes are typically considered two to three times larger than the pathogenic *Chlamydiaceae* (17). These core genes are largely considered "housekeeping genes" and are involved with maintaining the conserved intracellular niche and unique developmental cycle of *Chlamydiae*, along with universal processes like transcription and translation (18). While the pathogenic potential of these long-categorized "environmental *Chlamydiae*" remains to be fully defined, the genus *Chlamydia* is made up of known pathogens which contribute to major public health and agricultural burdens (Figure 1-2).
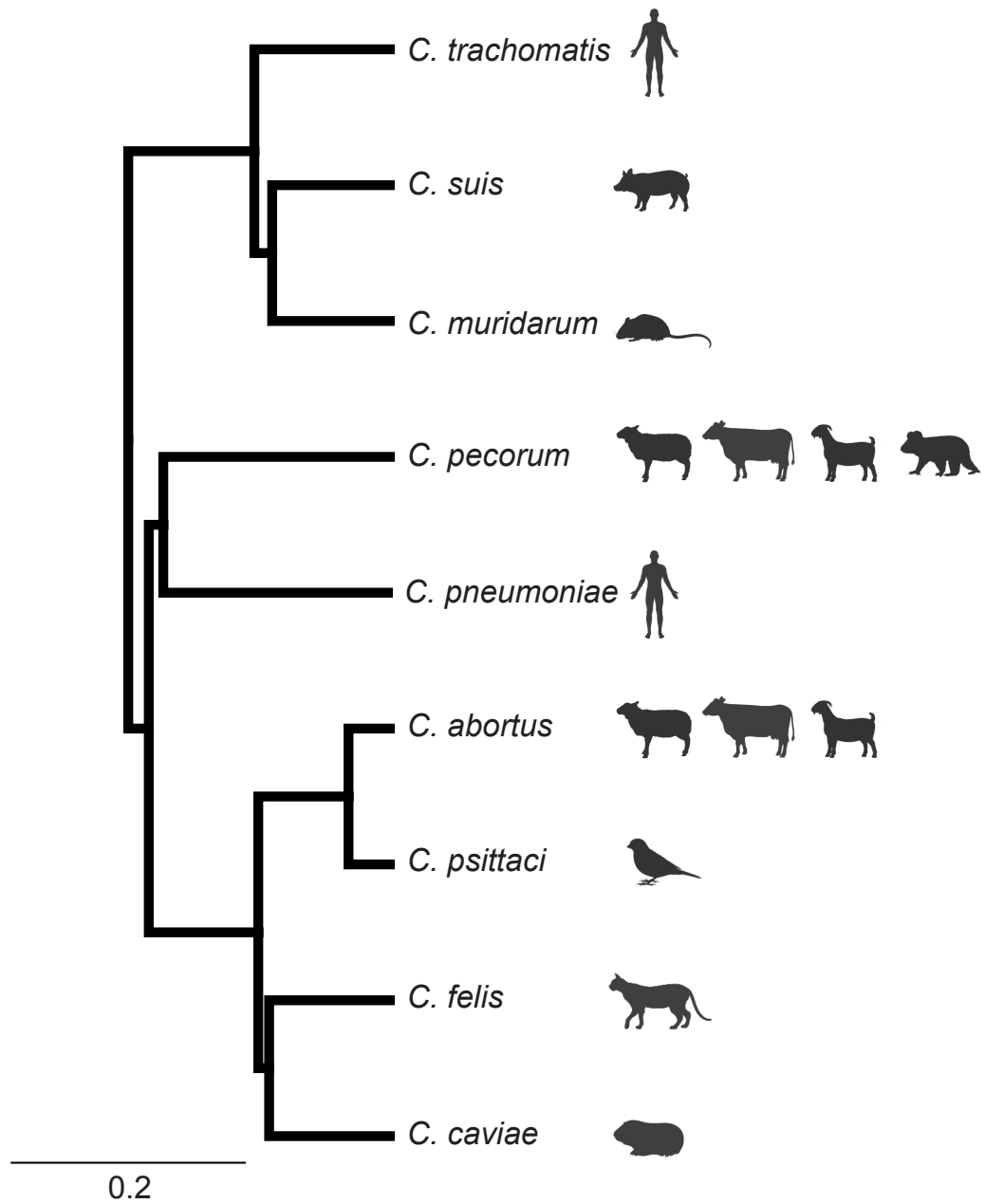
Figure 1-2. The nine canonical members of the genus, *Chlamydia* and the hosts they naturally infect. Scale bar represents substitutions per site.

The most significant human pathogen, *Chlamydia trachomatis,* is the causative agent for the most commonly reported bacterial sexually-transmitted disease worldwide as well as trachoma, the leading cause of non-congenital blindness (3, 19-21). The species is divided into distinct serovars based on the major outer membrane protein of *Chlamydia*, OmpA. These serovars correlate with tissue type and disease outcome; serovars A-C generally are known to cause ocular infections, leading to trachoma, while serovars D-K are associated with genital and reproductive infections. Additionally, serovar L1-3 is recognized as the cause of lymphogranuloma venereum (LGV), a disseminating infection affecting the rectum and local lymph nodes, frequently among men who have sex with men. Importantly, LGV strains are commonly used in the laboratory. While serovar is largely correlated with pathobiovar, *Chlamydia* are able to undergo horizontal gene transfer and serovar does not always correlate directly with genotype, as OmpA types can be exchanged (22).

The other major human pathogen, *C. pneumoniae*, a fairly ubiquitous bacteria, is considered a major cause of community-acquired pneumonia (23-25). Also in the genus are several species adapted to other, often mammalian, hosts. One such species is *C. suis* which has become a ubiquitous swine pathogen with many strains having acquired a tetracycline resistance cassette through horizontal gene transfer with another bacterial species (26-30). One of the most well-studied non-human *Chlamydia*, *C. muridarum,* infects mice in a variety of tissues including lungs, eyes, gastrointestinal tract and genital tract, and is often used as a model for human chlamydial infections (31-33). Another agriculturally important species is *C. abortus*, which infects a variety of livestock including cattle, sheep and goats (Figure 1-2) and has significant economic impact as it can lead to herd-wide spontaneous abortions (6, 34, 35). *C. pecorum*, a species that has evolved to infect multiple different hosts, is currently one of the leading causes in the decline of the koala

populations in Australia and is therefore also noteworthy (20, 36, 37). *C. caviae*, and *C. felis*, guinea pig and cat pathogens, respectively, are known to cause a variety of diseases, or even asymptomatic infections (32). The final canonical *Chlamydia*, *C. psittaci*, is an important pathogen causing the respiratory disease psittacosis, or parrot fever, largely infecting birds with known, but rare, zoonotic transmission to humans (32, 38, 39). Despite the wide range of host species infected by *Chlamydia,* each species in the phylum shares several distinctive properties including their developmental cycle.

The *Chlamydiae* developmental cycle

The unique biphasic developmental cycle of *Chlamydiae* is one of the defining characteristics of the phylum, differentiating it from other obligate intracellular bacteria like *Rickettsia* or *Coxiella* (40, 41). The term "biphasic" here refers to the two contrasting forms that *Chlamydiae* oscillate between during the developmental cycle; an infectious form called the Elementary Body (EB) and a replicative form called the Reticulate Body (RB) (Figure 1-3) (7, 41, 42). The EB is highly condensed, using two histone-like proteins to bind DNA and RNA forming a nucleoid (43, 44). Along with the DNA condensation, the EB outer membrane is cross-linked by two cysteine-rich outer membrane proteins which form disulfide bonds, strengthening the membrane against the harsh extracellular environment (45, 46). These processes contribute to a restriction of transcription, translation and metabolic activity within the EB (47). Elementary Bodies have long been considered metabolically inert, though recent advances appear to subvert that long-held assumption, as EBs are known to utilize the Type 3 Secretion System (T3SS) in invasion, are capable of metabolite uptake, and potentially use glucose-6-phosphate as an energy source (48-50). Conversely, the RB has traditionally been considered the metabolically-active form and

undergoes the bulk of transcription and translation, resulting in a replicative form (19, 51, 52). Because it is uncondensed, having both relaxed DNA and RNA and no cross-linking in the outer membrane, the RB is larger than the EB, approximately 1um in diameter where the EB is typically 3x smaller at 0.3um (53). However, due to the reduced strength of the outer membrane, the RB is osmotically sensitive and does not survive outside of the *chlamydia*-containing vacuole, called the inclusion (53). Because of this disparity in environmental stability, the developmental cycle of *Chlamydiae* relies on specific events to ensure bacterial survival.
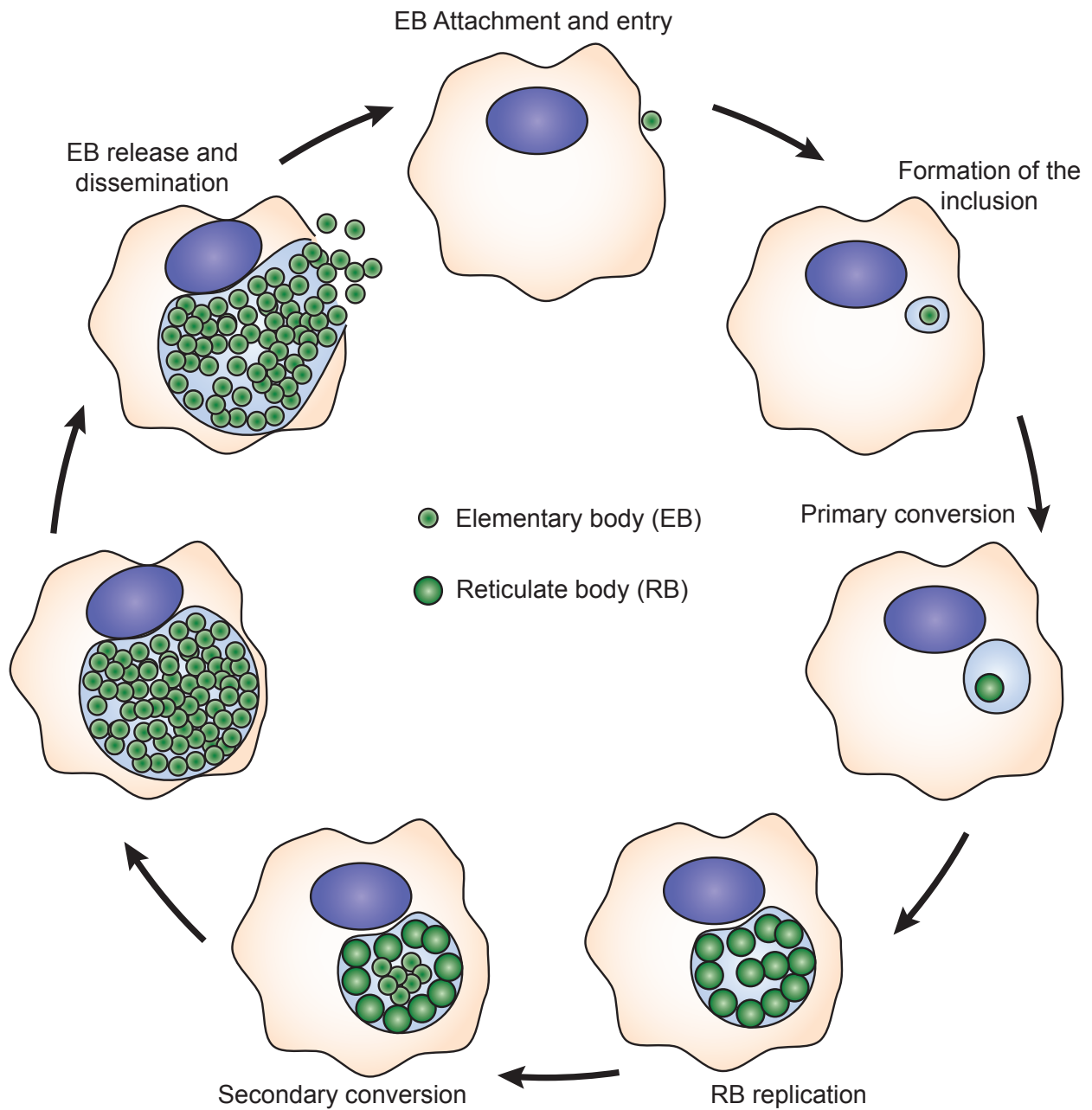
Figure 1-3. Developmental cycle for *Chlamydiae* demonstrating attachment and entry through primary conversion, replication, secondary conversion and release through host cell lysis (represented here) or through extrusion of the inclusion membrane.

Once an EB makes contact with a host cell membrane, it activates the trigger mechanism by injecting T3SS effector proteins, promoting cytoskeletal rearrangement resulting in internalization of the bacteria (54). Specific adherence mechanisms appear to vary by species and tissue type, suggesting that this is one mechanism for host-specificity within *Chlamydiae* (19, 55). Of the known T3SS effectors used in the invasion process, Translocated actin-recruiting phosphoprotein (TarP) is one of the most well-known and is conserved among *Chlamydia* (56-58). TarP, in conjunction with other effector proteins, recruits actin, allowing for internalization of the EB within an early endosomal vesicle, which is modified to form the inclusion (59, 60). All *Chlamydiae* secrete a family of proteins which localize to and insert into the inclusion membrane (59). These inclusion membrane proteins (Incs) differ between species and participate in a diverse array of functions such as membrane stabilization (61), centrosome and endoplasmic reticulum anchoring (62-64), vesicle fusion (65-67), and immune evasion (68, 69).

Inside the relative safety of the inclusion, the EB will undergo the primary conversion event, differentiating from EB to RB (19) (Figure 1-3). The RB then activates transcription and begins the process of asynchronous multiplication (70). There are contrasting data on the mechanism of RB division, with some studies supporting traditional binary fission (71-73) and newer reports suggesting a polarized budding mechanism (74, 75). In either case, the RB will undergo several rounds of division until the secondary conversion of the RB back into the EB. Again, this process is largely undefined in *Chlamydiae*, although current hypotheses include unequal division to reduce membrane size and promote condensation (71), and RB detachment from the inclusion membrane as a conversion catalyst (76). Once the inclusion is largely filled with infectious EBs,

bacteria can exit by two independent mechanisms: host cell lysis (pictured in Figure 1-3) or extrusion of the inclusion (77).

Throughout this developmental cycle, *Chlamydiae* consistently interact with the host cell in order to establish their intracellular niche, acquire nutrients, and subvert the host response. Once the bacterium has managed to enter the cell it has to establish its intracellular niche, which allows it to avoid intrinsic cell responses. One key aspect of this process is the formation and maintenance of the inclusion membrane, which acts as both a protective barrier, hiding the bacteria from the innate immune system of the host, and as the interface between the bacteria and the host cell cytoplasm (78). Nutrients obtained from the host, and effectors secreted into the host, must both pass through this membrane. Because of the host interactions required at every step of the developmental cycle, *Chlamydiae* pathogenesis is dependent on these specific host factors.

Chlamydia trachomatis as a major human pathogen

Chlamydia trachomatis is recognized as the most reported bacterial infection in the United States and is responsible for an estimated 1.8 million sexually-transmitted infections (STIs) each year, a rate of approximately 539.9 per 100,000 people (79). This rate has only increased over the last 20 years. As of 2016, worldwide cases were estimated to be around 127 million (80). In most cases, infections are asymptomatic which may result in these prevalence rates being far lower than in actuality. Symptoms do occur however, in an estimated 6-23% of all cases, which, according to one study, could indicate that around 95% of untreated chlamydia is due to a lack of symptoms (81, 82). These symptoms can range in severity from relatively mild; often genital discharge or dysuria, to severe. Severe symptoms include infertility or sterility and life threatening pelvic

inflammatory diseases (PID) and ectopic pregnancies (83, 84). Global infertility rates appear to be on the rise, and one major contributing factor is likely to be untreated sexually transmitted infections, of which chlamydia is the most common (85).

In the United States, chlamydia infections disproportionately affect female populations. Prevalence in women is nearly double that in men, indicating that infections in the female genital tract are more likely to be detected, either through routine testing or due to increased severity of symptoms (79, 84). Importantly, untreated female infections can result in the more severe symptoms listed above with often irreversible consequences (86). Treatment, which for chlamydia involves a single-dose of azithromycin or a seven-day course of doxycycline, is relatively simple and effective. However, this treatment relies on testing and detection, a major challenge for a largely asymptomatic infection, and depends on the use of broad-spectrum antibiotics, a problem contributing to the rise in antibiotic resistance (84). It is estimated that chlamydia infections are the costliest nonviral STIs with an annual financial burden around $500 million as of 2010 (87). The high prevalence rate and ability to go undetected is indicative of the "success" of this pathogen and of its adaptation to the human host. There is much we still do not understand about the pathogenesis of *C. trachomatis*, including how the bacteria cause damage to tissues leading to severe symptoms and why disease outcomes vary so widely between individuals (86). Answering these questions could be key to development of prophylactics, identifying risk groups to encourage routine screening and increasing the availability of treatment options.

What we do know about chlamydial pathogenesis in the genital tract is broad strokes of the infection. In a female, *C. trachomatis* EBs establish an initial infection in the epithelial tissues of

the vagina. In some, likely limited cases, the bacteria begin to disseminate towards the upper reproductive tract through a process known as ascension, where the infection spreads upward to the cervix, uterus and fallopian tubes (88). Within the fallopian tubes is where most of the complications associated with untreated *chlamydia* infections occur (Figure 1-4). Because these important sequelae occur in the upper reproductive tract, the bacteria's ability to ascend is essential to its ability to cause disease.
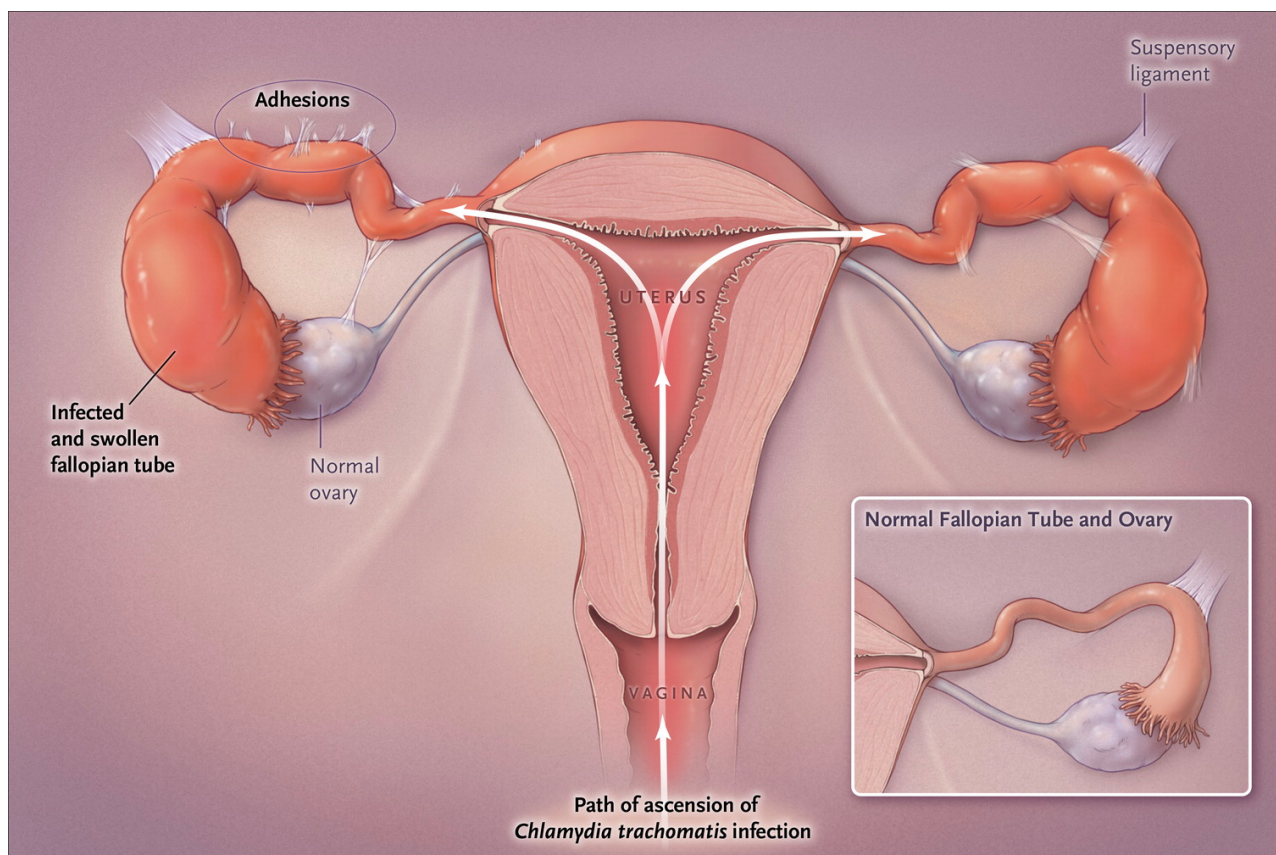
Figure 1-4. Standard model for *Chlamydia trachomatis* pathogenesis in the human female reproductive tract. *Chlamydia* initially infect the vaginal vault before ascending past the cervix, through the uterus and into the fallopian tubes where pathology occurs. Reproduced with permission from "Screening for *Chlamydia trachomatis* Infections in Women" by Wiesenfeld, HC, 2017 (88), Copyright Massachusetts Medical Society.

Despite the fact that *Chlamydia* cause active lysis of host cell tissue, this acute cell death does not appear to be the primary cause of tubal scarring and other large-scale tissue damage. *Chlamydia* are considered immunopathogenic, meaning that it is the immune response elicited by the infection that eventually causes the disease state and tissue damage, rather than the direct actions of the pathogen itself (89, 90). In what is referred to as the cellular pathogenesis paradigm, *Chlamydia* infect non-immune cells of the epithelium, stimulating the release of pro-inflammatory cytokines and chemokines (89, 91). This signal-rich environment leads to an influx of cellular immunity factors, particularly neutrophils, T cells and macrophages. Continued onslaught of the immune response damages epithelial cells, resulting in tissue damage. It is likely that persistent or repeat infections contribute to the severity of damage as well (92). B cells, the critical members of humoral immunity, will also migrate to the infected area and activate an adaptive response, although it has been shown that this response is not highly protective against reinfection. The weak adaptive response to initial *Chlamydia* infection has increased the challenges of vaccine development.

While much research has been done to determine the specific human immunological response, the limitation of these studies is largely that they cannot be performed with live, human subjects. Even with the ethical questions aside, exploring *chlamydia* pathogenesis over time in human subjects would be difficult. This is the primary reason for the development and use of animal models. Much of the current knowledge of *in vivo* chlamydial pathogenesis was discovered in the mouse model and conclusions are extrapolated to human infections. However, there are considerable limitations to this model and to the conclusions that can be drawn. First and foremost in these limitations is the host-specificity of *Chlamydia*.

Many serovars of *C. trachomatis* have been investigated in the mouse model and while the murine genital tract can be infected, these infections are mild, clear relatively quickly, and largely fail to ascend and cause major sequelae (93). In order to study the disease state in the mouse model, transcervical infections, where the upper reproductive tissues (uterine horns) are infected directly, are utilized. This allows for researchers to investigate formation of adhesions, swelling of the ovarian bursa called hydrosalpinx, and even infertility. However, because the bacteria cannot undergo ascension, factors involved directly in determining disease outcomes in a human infection become more challenging to investigate. In contrast, *C. muridarum*, the mouse pathogen, can cause a reproductive tract infection in mice that recapitulates many aspects of the human infection including ascension (93-95). One major challenge with using *C. muridarum* in the mouse model to directly correlate with human disease is that human and mouse anatomy is not structurally identical and utilize different immunological responses; mice do not have fallopian tubes, instead have two uterine horns which directly abut the ovarian bursa, for example. Despite these challenges, the *C. muridarum* mouse model continues to remain the most used model for *C. trachomatis* human infections.

Potential and known host-specificity factors

As discussed above, *Chlamydia* are known to have restricted host ranges, with *C. trachomatis* as an obligate human pathogen. This is not unique among bacteria, and many other organisms have similar host-specificity challenges. Examples of these include *Neisseria gonorrhoeae*, *Salmonella* Typhi*, Heliobacter pylori, Haemophilus influenzae, Vibrio cholerae, Streptococcus pneumoniae, Streptococcus pyogenes* and many more (96). The molecular basis for host-specificity in these bacterial pathogens is not fully understood, but researchers have found common mechanisms that

contribute to this host-specificity in different species, though common consensus suggests that many factors are involved in these restrictions. In general, host-specificity appears to be mediated by three aspects relating to pathogenesis; (I) specific binding of bacterial surface proteins to host surface proteins important in colonization, attachment and invasion, (II) adapted techniques for overcoming specific aspects of the host immune response, or (III) nutrient availability within the host and acquisition mechanisms of the bacteria (Figure 1-5) (96, 97).
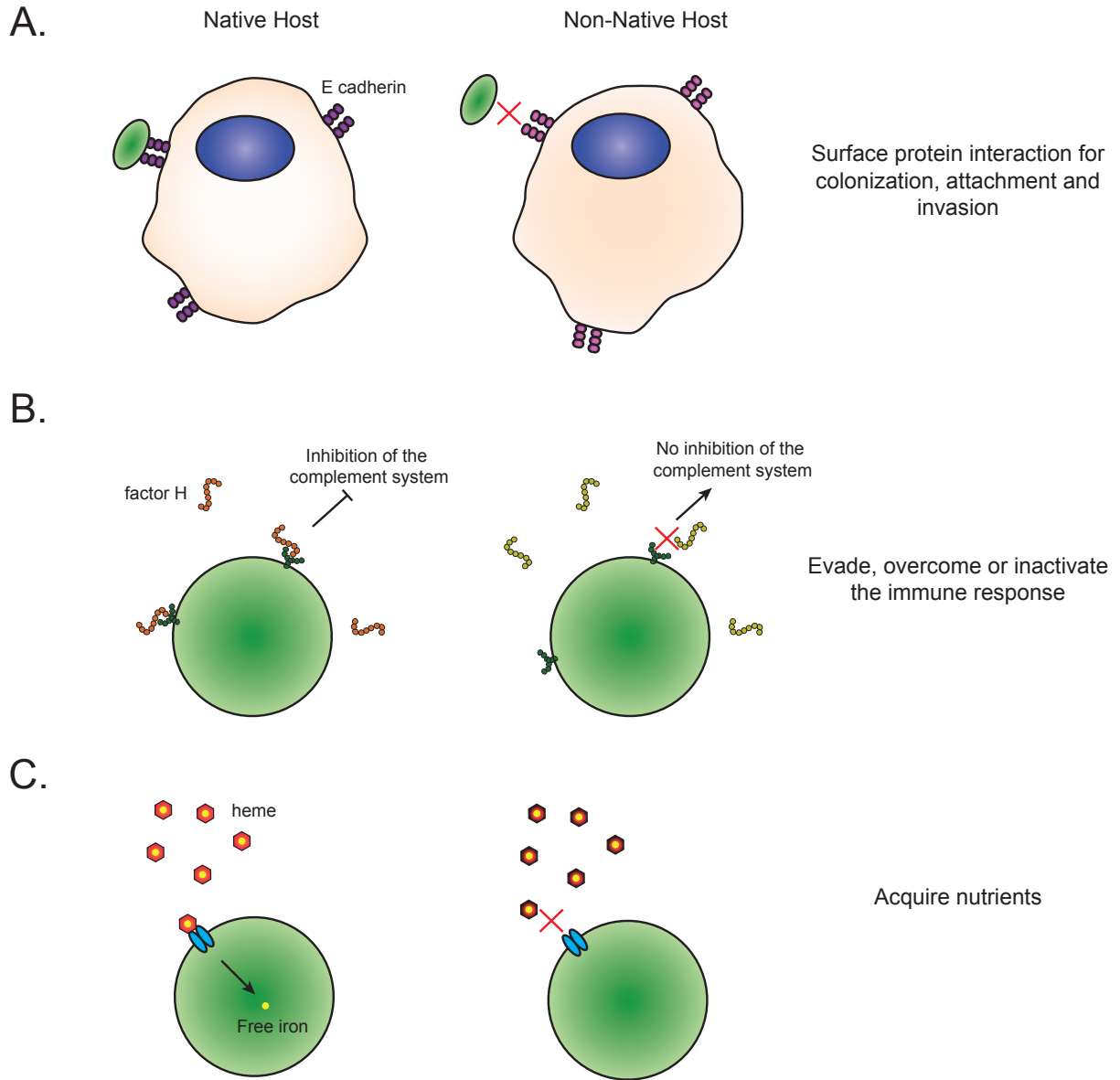
Figure 1-5. Examples of host specificity in bacteria. A) *Listeria monocytogenes* (green) can bind human E-cadherin to promote uptake but does not bind mouse E-cadherin due to slight differences in the proteins. B) *Neisseria* factor H binding protein (dark green) binds human but not mouse factor H to inhibit the complement pathway of the immune system. C) Hemoglobin receptors in *Staphylococcus aureus* (blue) recognize human hemoglobin more efficiently than mouse hemoglobin, leading to increased ability to acquire free iron in the human host.

One example of a surface protein aiding in invasion and known to be important for bacterial host-specificity is internalin A (InlA) of *Listeria monocytogenes*. InlA binds human E-cadherin, promoting uptake and invasion into epithelial cells (98, 99). However, InlA is unable to bind murine E-cadherin due to a single amino acid substitution disrupting the binding of InlA (100). In this case, a minor difference in host protein greatly affects the ability of the bacteria to infect a specific host.

Surface proteins can also be important in evading the specific immune response of the native host, as well. *Neisseria* factor H binding protein binds human, but not mouse, complement factor H which downregulates activation of complement, reducing bacterial killing (101, 102). Similar protein interactions have been seen in *S. pneumoniae* (103), *Borrelia burgdorferi* (104, 105) enteropathogenic *E. coli* (106) and *Staphylococcus aureus* (107). Secreted proteins can also play a role in key host-specific immune evasion. A well-studied example of this is a family of proteins known as immunoglobin A (IgA) proteases which cleave human IgA1, preventing antigen binding (108). IgA is the prominent antibody found in genital and nasal secretions, and many bacteria known to infect these sites have been shown to produce IgA proteases including; *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Haemophilus influenzae,* and *S. pneumoniae* (109) (96, 110).

Similarly, nutrient availability in the host can also play an important function in restricting pathogenesis. Iron acquisition mechanisms appear to be especially critical in host restriction. Bacteria require iron for growth and typically scavenge iron from the host. As a consequence, iron sequestration has become a successful strategy for preventing infection (111). In response, many bacteria have developed strategies to overcome the sequestration techniques in specific hosts.

Examples include the transferrin-iron import system in *Neisseria* (112) and a hemoglobin receptor in *S. aureus* (113, 114).

Each of these examples demonstrates that host restriction involves many facets of bacterial pathogenesis. Individually, each of these components are important virulence factors supporting the specialization of bacteria to a specific host. Working in concert, these factors enable specialization of a bacteria to a specific host. Discovery of these factors is important in understanding the basic biology of pathogenic bacteria and strategies in the "bacterial arms race", but perhaps more importantly, these discoveries can provide insight into human infection and be used in the design of therapeutics. *Neisseria* factor H binding protein, mentioned above, has been extensively studied as a vaccine candidate against bacterial meningitis (115). Discovery of species-specific virulence factors could similarly benefit *Chlamydia trachomatis* therapeutic and vaccine research.

Currently, only a few such species-specific factors are known in *Chlamydia*. The most well-demonstrated example involves immune evasion of the human, but not murine host for *C. trachomatis*. Genital tract-infecting *C. trachomatis* serovars contain a tryptophan biosynthesis (*trp*) operon used to generate tryptophan from indole for protein synthesis (116). This operon plays a critical role in human infection where interferon gamma (IFN-γ) stimulates an immune response in which indolamine 2,3-dioxygenase (IDO) converts free tryptophan to kynurenine, making it unavailable to pathogenic bacteria. In the genital microbiome, however, many commensal species produce indole, allowing *Chlamydia* to salvage tryptophan, and cause disease (116, 117). Mice have not been shown to utilize this IFN-γ-mediated IDO response pathway, eliminating the need

for the *trp* operon in *C. muridarum* (118). This single pathway may be responsible for restricting *C. muridarum* to the murine host. However, there are likely similar pathways in mice that restrict *C. trachomatis* infections that remain undiscovered.

There are a number of strategies to discover host-specific genes in bacterial species. One commonly used strategy, comparative genomics, is utilized to discover novel genes in one species, but not the other. Novel genes could be created by many reasons, from evolutionary gene loss, to the acquisition of entire pathogenicity islands through lateral gene transfer (97). Comparative genomic studies for *Chlamydia* species have revealed several unique features, particularly in a region termed the plasticity zone. However, specifically between *C. trachomatis* and *C. muridarum*, host restriction appears to be a paradox, as the genomes share over 90% of their gene content. As an additional challenge, *Chlamydia* contain many genus-specific genes, for which there is no known function, which comprise around 30% of the genome. Without the presence of multiple virulence plasmids, pathogenicity islands or a large subset of unique genes, discovery of species-specific factors within *Chlamydia* has been difficult.

While the *trp* operon is an example of a unique host-specific factor within *Chlamydia*, it is likely that conserved factors are significantly contributing to host specificity. One of the key hypotheses in determining host-specific genes important in *C. muridarum* pathogenesis of the mouse is that interactions between *C. muridarum* and mouse proteins will have direct *C. trachomatis* orthologous interactions in humans. As an example, *C. muridarum* receptors important in adherence to the mouse cell likely have direct homologs in *C. trachomatis*. However, while the *C. muridarum* receptor could be adapted to interact specifically with a mouse surface protein, *C.*

trachomatis would have adapted to the human homolog of the same, or similar surface protein. Those host-specific adaptations would lead to changes in protein coding sequence, so the genes encoding these receptors may appear divergent, but the function and cognate interacting factor, would be conserved. Therefore, discovery of conserved gene products important for *C. muridarum* pathogenesis in the mouse can directly correlate with host-specific factors for human infection by *C. trachomatis.*

Development of recombinant libraries in *Chlamydia*

One critical advancement allowing for functional studies in bacteria is the ability to perform genetic screens. In *Chlamydia*, genetic manipulation has been a major roadblock, until the last decade when transformation techniques were developed. The primary challenges stem from the obligate intracellular niche and biphasic developmental cycle, creating difficulties in that: (I) the DNA must make it past four membranes (host, inclusion and bacterial inner and outer); (II) there are differences in metabolic activity of the two forms; (III) there is a lack of available antibiotics which are host cell safe, membrane permeable and not used for clinical treatment; and (IV) the reduced genome has a high proportion of expected essential genes (119). Despite these difficulties several genetic tools have been developed for *Chlamydia* including TargeTron, allelic exchange and chemical mutagenesis (120). Recently, transposon mutagenesis has also been developed for *C. trachomatis* and *C. muridarum* (121, 122).

Transposon mutagenesis has been demonstrated in many bacterial species to be an important tool for discovering genetic correlates for specific phenotypes and determining essential genes. First discovered in corn, or *Zea mays*, transposons, also referred to as "jumping genes" have become

major technological boons over the past fifty years (123). These genetic elements can excise themselves from DNA to DNA, such as from a plasmid into a chromosome, through the use of their terminal inverted repeats and encoded transposase (124, 125). As a genetic tool, they allow for genome-wide random insertions in non-essential genes and often contain selectable markers as was the case in the *Chlamydia* transposon libraries. In *C. trachomatis*, 105 transposon mutants have been generated, 81 of which are within coding regions (Figure 1-6) (121). These transposon mutants have allowed for functional studies demonstrating that a putative *comEC* homolog plays a critical role in DNA uptake in *Chlamydia*. In *C. muridarum*, a smaller library of 33 transposon mutants was created as well (122). Importantly, each of these transposon mutants in *C. muridarum* contained a chloramphenicol resistance marker fused to a green fluorescent protein (GFP) tag (122).
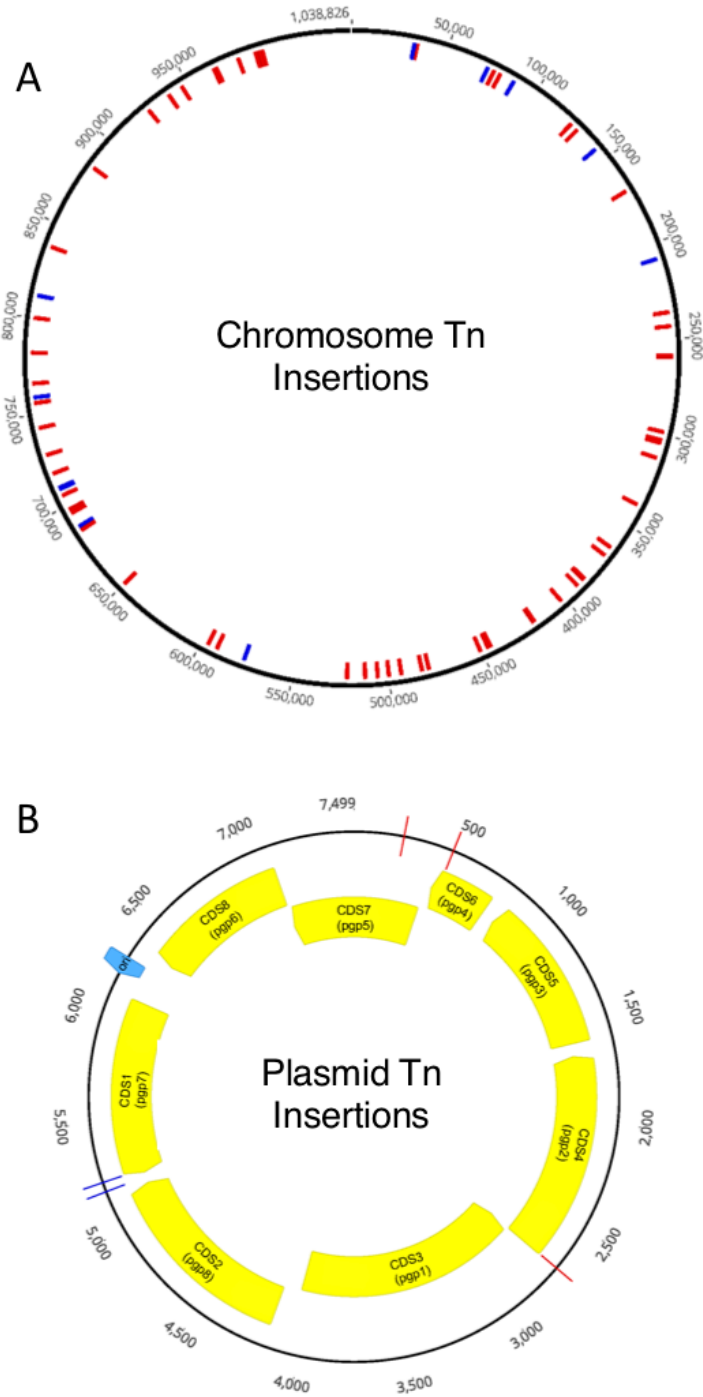
Figure 1-6. Genome maps of transposon insertions in *C. trachomatis* chromosome and plasmid. A) Transposon insertions with coding regions (red) and noncoding regions (blue) throughout the *C. trachomatis* L2 genome. B) Transposon insertions within the *C. trachomatis* L2 plasmid. Published in LaBrie et al, 2019 (121).

The generation of transposon libraries in *Chlamydia* led to an additional genetic tool: recombinant libraries. Homologous recombination is a universal process used for repair of double stranded breaks in DNA where donor DNA is used as the template, leading to the inclusion of new sequences or horizontal gene transfer (HGT). HGT was discovered in bacteria in 1928, before even the discovery of DNA, when Fredrick Griffith demonstrated that avirulent *S. pneumoniae* was able to uptake virulent genes from killed virulent strains (126). Recombination has historically been an invaluable tool in genetics for many organisms, from eukaryotes to bacteria. Classical genetics studies utilize *Saccharomyces cerevisiae*, or budding yeast, to generate genetic crosses which have been used to discover mechanisms of generating diversity in populations, and inheritance of a variety of traits such as specific sugar fermentation, resistance to copper, and production of byproducts like sulfide (127). Additionally, homologous recombination in yeast enabled a rise in reverse genetics studies with the ability to use plasmids to deliver DNA directly into the genome using HGT to generate gene disruptions and knockouts (128).

In bacteria, similar genetic crosses can be generated through bacterial transformation or conjugation. In transformation, as discovered by Griffith, bacteria take up DNA from their environment and use HGT to incorporate DNA into their genome. In conjugation, DNA is shared from one bacterium to another through the use of a conjugative pili. This sexual-like reproduction of DNA is determined by the presence of a fertility factor, F-plasmid. Studies generating bacterial crosses include generation of tetracycline-resistant *Staphylococcus* in a dual biofilm with resistant *Bacillus subtilus* (129) and adaptive evolution of *E. coli* to a new carbon source (7). Horizontal gene transfer has recently been demonstrated to occur in *Chlamydia in vitro* and *in vivo* (130-132). While this process has not been well-characterized in *Chlamydia,* it is expected to work by similar

mechanisms to other bacteria which have been long studied and well-defined. Instances of HGT in *Chlamydia in vivo* are few, particularly between species, but it has been shown to be possible in the laboratory. One prime example is the transfer of a tetracycline-resistance cassette acquired by *C. suis* into *C. trachomatis,* generating a tetracycline resistant *C. trachomatis* strain (133).

The tetracycline-resistant *C. trachomatis* demonstrated that interspecies recombination was possible, given strong selective pressure, like antibiotics. With creation the *C. muridarum* transposon library, selectable markers across the *C. muridarum* genome became available as well. Using these markers, a library of 142 recombinant *C. trachomatis* x *C. muridarum* crosses was generated (134) which helped define the constraints of intergenic recombination in *Chlamydia* (Figure 1-7). While determining the "rules of recombination" in *Chlamydia* has been interesting, this library also serves as a mechanism to investigate the function of unique *C. muridarum* genes, as well as incompatibilities between *C. trachomatis* and *C. muridarum* homologs.
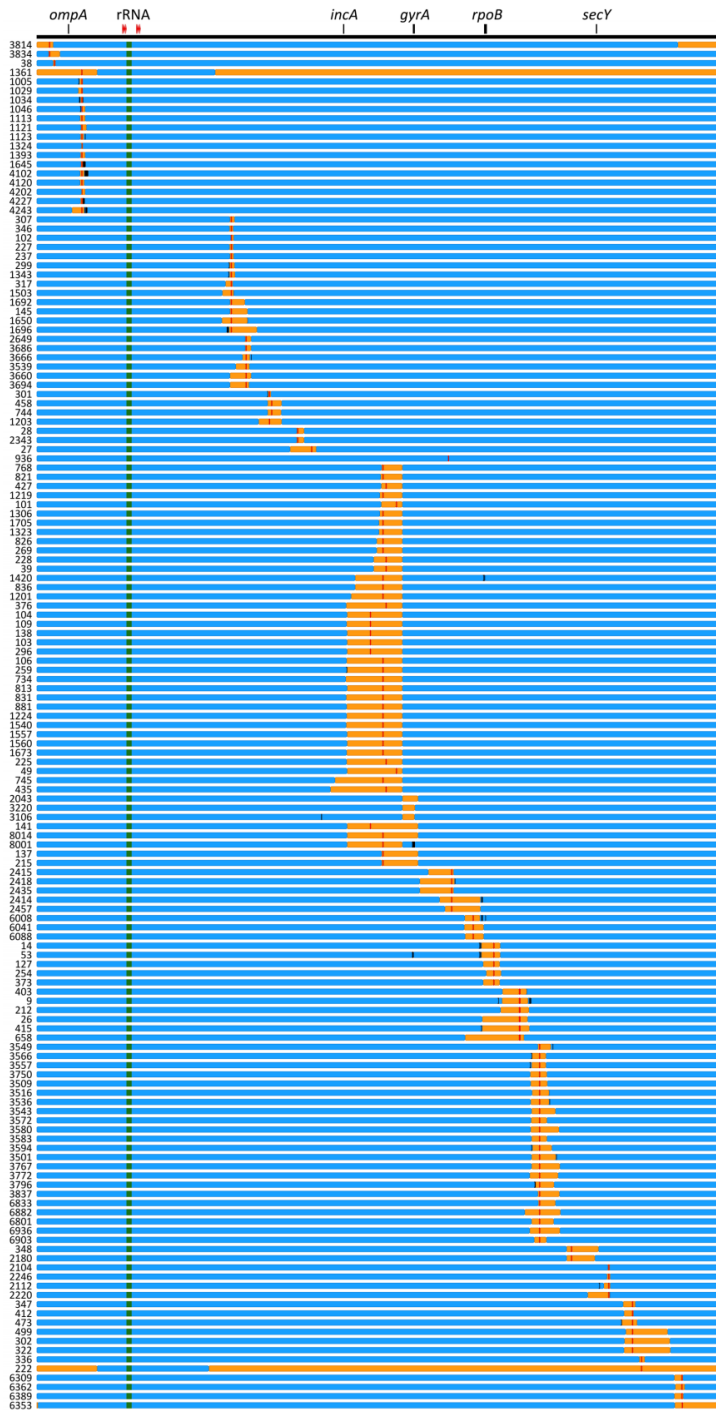
Figure 1-7. Graphic representation of the independent recombinant strains. Individual recombinants are identified by number at the left of each line. *C. trachomatis* sequence is indicated in blue, while *C. muridarum* sequence is indicated in tan. The vertical red bars indicate the location of the transposon in each strain, black bars indicate regions of secondary recombination in individual progeny strains and the location of the *tet*(C) is shown in green for each recombinant. The locations of selected chromosomal genes are indicated at the top of the figure.

This dissertation focuses on investigating host-specific pathogenesis factors across the chlamydial genome beginning with comparative genomics studies to identify unique effectors in *C. suis* as compared with *C. trachomatis* and *C. muridarum*, where the focus was on specific gene families, through investigating specific genomic regions using the *C. trachomatis* x *C. muridarum* libraries. In total, this research describes 1) the first independently annotated *C. suis* genome with direct comparisons of critical gene families in *Chlamydia,* 2) discovery that the *C. muridarum* plasticity zone, and the unique genes therein, is not sufficient to allow for *C. trachomatis* to overcome the mouse host-specificity barriers, 3) an incompatible, but orthologous gene that caused inclusion disruption and host apoptosis in the recombinant, 4) unique gene phenotypes associated with the recombination regions, and 5) investigation into the genome-wide chimeras which demonstrate that the majority of genetic material is interchangeable between *C. trachomatis* and *C. muridarum*. The next four chapters will describe these investigations in detail and provide discussion on the use of recombinant libraries as a tool for host-specific gene discovery, how we can integrate these findings into the broader chlamydial and microbiological body of knowledge and finally, future avenues for research based on these findings.

References

1. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, et al. Illuminating the evolutionary history of chlamydiae. Science. 2004;304(5671):728-30.
2. Horn M. Chlamydiae as symbionts in eukaryotes. Annu Rev Microbiol. 2008;62:113-31.
3. Budai I. Chlamydia trachomatis: milestones in clinical and microbiological diagnostics in the last hundred years: a review. Acta Microbiol Immunol Hung. 2007;54(1):5-22.
4. Moulder JW. The relation of the psittacosis group (Chlamydiae) to bacteria and viruses. Annu Rev Microbiol. 1966;20:107-30.
5. Subtil A, Collingro A, Horn M. Tracing the primordial Chlamydiae: extinct parasites of plants? Trends Plant Sci. 2014;19(1):36-43.
6. Sachse K, Bavoil PM, Kaltenboeck B, Stephens RS, Kuo CC, Rossello-Mora R, et al. Emendation of the family Chlamydiaceae: proposal of a single genus, Chlamydia, to include all currently recognized species. Syst Appl Microbiol. 2015;38(2):99-103.
7. Bayramova F, Jacquier N, Greub G. Insight in the biology of Chlamydia-related bacteria. Microbes Infect. 2018;20(7-8):432-40.
8. Collingro A, Kostlbacher S, Horn M. Chlamydiae in the Environment. Trends Microbiol. 2020;28(11):877-88.
9. Tsao HF, Scheikl U, Herbold C, Indra A, Walochnik J, Horn M. The cooling tower water microbiota: Seasonal dynamics and co-occurrence of bacterial and protist phylotypes. Water Res. 2019;159:464-79.
10. Corsaro D, Greub G. Pathogenic potential of novel Chlamydiae and diagnostic approaches to infections due to these obligate intracellular bacteria. Clin Microbiol Rev. 2006;19(2):283-97.
11. Vonaesch P, Anderson M, Sansonetti PJ. Pathogens, microbiome and the host: emergence of the ecological Koch's postulates. FEMS Microbiol Rev. 2018;42(3):273-92.
12. Falkow S. Molecular Koch's postulates applied to bacterial pathogenicity--a personal recollection 15 years later. Nat Rev Microbiol. 2004;2(1):67-72.
13. Inglis TJJ. Principia aetiologica: taking causality beyond Koch's postulates. J Med Microbiol. 2007;56(Pt 11):1419-22.
14. Klockner A, Buhl H, Viollier P, Henrichfreise B. Deconstructing the Chlamydial Cell Wall. Curr Top Microbiol Immunol. 2018;412:1-33.
15. Jacquier N, Viollier PH, Greub G. The role of peptidoglycan in chlamydial cell division: towards resolving the chlamydial anomaly. FEMS Microbiol Rev. 2015;39(2):262-75.
16. Konig L, Siegl A, Penz T, Haider S, Wentrup C, Polzin J, et al. Biphasic Metabolism and Host Interaction of a Chlamydial Symbiont. mSystems. 2017;2(3).
17. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, et al. Unity in variety--the pan-genome of the Chlamydiae. Mol Biol Evol. 2011;28(12):3253-70.
18. Psomopoulos FE, Siarkou VI, Papanikolaou N, Iliopoulos I, Tsaftaris AS, Promponas VJ, et al. The chlamydiales pangenome revisited: structural stability and functional coherence. Genes (Basel). 2012;3(2):291-319.
19. Elwell C, Mirrashidi K, Engel J. Chlamydia cell biology and pathogenesis. Nat Rev Microbiol. 2016;14(6):385-400.
20. Burnard D, Polkinghorne A. Chlamydial infections in wildlife-conservation threats and/or reservoirs of 'spill-over' infections? Vet Microbiol. 2016;196:78-84.

21.   Taylor HR, Burton MJ, Haddad D, West S, Wright H. Trachoma. Lancet. 2014;384(9960):2142-52.

22.   Suchland RJ, Dimond ZE, Putman TE, Rockey DD. Demonstration of Persistent Infections and Genome Stability by Whole-Genome Sequencing of Repeat-Positive, Same-Serovar Chlamydia trachomatis Collected From the Female Genital Tract. J Infect Dis. 2017;215(11):1657-65.

23.   Kauppinen M, Saikku P. Pneumonia due to Chlamydia pneumoniae: prevalence, clinical features, diagnosis, and treatment. Clin Infect Dis. 1995;21 Suppl 3:S244-52.

24.   Musher DM, Abers MS, Bartlett JG. Evolving Understanding of the Causes of Pneumonia in Adults, With Special Attention to the Role of Pneumococcus. Clin Infect Dis. 2017;65(10):1736-44.

25.   Di Pietro M, Filardo S, Romano S, Sessa R. Chlamydia trachomatis and Chlamydia pneumoniae Interaction with the Host: Latest Advances and Future Prospective. Microorganisms. 2019;7(5).

26.   Dimond ZE, Hefty PS. Comprehensive genome analysis and comparisons of the swine pathogen, Chlamydia suis reveals unique ORFs and candidate host-specificity factors. Pathog Dis. 2020.

27.   Chahota R, Ogawa H, Ohya K, Yamaguchi T, Everett KDE, Fukushi H. Involvement of multiple Chlamydia suis genotypes in porcine conjunctivitis. Transbound Emerg Dis. 2018;65(1):272-7.

28.   J. Lenart AAA, D.D Rockey. Growth and Development of Tetracycline-Resistant *Chlamydia suis*. Antimicrob Agents Chemother. 2001;45(8):2198-203.

29.   Manuela D H-CH, Humphrys M, Di Paolo M, Di Francesco A, Myers GSA. Genome Sequence of Chlamydia suis MD56, Isolated from the Conjunctiva of a Weaned Piglet. Genome Announc. 2014;2(3):e00425-14.

30.   Marti H, Kim H, Joseph SJ, Dojiri S, Read TD, Dean D. Tet(C) Gene Transfer between Chlamydia suis Strains Occurs by Homologous Recombination after Co-infection: Implications for Spread of Tetracycline-Resistance among Chlamydiaceae. Front Microbiol. 2017;8(156):156.

31.   Farris CM, Morrison RP. Vaccination against Chlamydia genital infection utilizing the murine C. muridarum model. Infect Immun. 2011;79(3):986-96.

32.   Cheong HC, Lee CYQ, Cheok YY, Tan GMY, Looi CY, Wong WF. Chlamydiaceae: Diseases in Primary Hosts and Zoonosis. Microorganisms. 2019;7(5).

33.   Rank R. Intracellular Pathogens 1: Chlamydiales. ASM Press. 2012;1:285-310.

34.   Baud D, Greub G. Intracellular bacteria and adverse pregnancy outcomes. Clin Microbiol Infect. 2011;17(9):1312-22.

35.   Entrican G, Wheelhouse N, Wattegedera SR, Longbottom D. New challenges for vaccination to prevent chlamydial abortion in sheep. Comp Immunol Microbiol Infect Dis. 2012;35(3):271-6.

36.   Jelocnik M BN, Kaltenboeck B, Waugh C, Woolford L, Speight KN, Gillett A, Higgins DP, Flanagan C, Myers GS, Timms P, Polkinghorne A. Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates Chlamydia pecorum strains from pigs, sheep, cattle, and koalas. BMC Genomics. 2015;16:893.

37.   Walker E, Lee EJ, Timms P, Polkinghorne A. Chlamydia pecorum infections in sheep and cattle: A common and under-recognised infectious disease with significant impact on animal health. Vet J. 2015;206(3):252-60.

38. Harkinezhad T, Geens T, Vanrompay D. Chlamydophila psittaci infections in birds: a review with emphasis on zoonotic consequences. Vet Microbiol. 2009;135(1-2):68-77.

39. Schofl G, Voigt A, Litsche K, Sachse K, Saluz HP. Complete genome sequences of four mammalian isolates of Chlamydophila psittaci. J Bacteriol. 2011;193(16):4258.

40. Walker DH, editor. Rickettsiae. 4th ed. Galveston, Texas: University of Texas Medical Branch at Galveston; 1996.

41. McClure EE, Chavez ASO, Shaw DK, Carlyon JA, Ganta RR, Noh SM, et al. Engineering of obligate intracellular bacteria: progress, challenges and paradigms. Nat Rev Microbiol. 2017;15(9):544-58.

42. Panzetta ME, Valdivia RH, Saka HA. Chlamydia Persistence: A Survival Strategy to Evade Antimicrobial Effects in-vitro and in-vivo. Front Microbiol. 2018;9:3101.

43. Goncalves OSL, Christiansen G, Holm A, Herrmann B, Klintstedt M, Petersen SB, et al. The repeated 36 amino acid motif of Chlamydia trachomatis Hc2 protein binds to the major groove of DNA. Res Microbiol. 2019;170(6-7):256-62.

44. Klint M, Thollesson M, Bongcam-Rudloff E, Birkelund S, Nilsson A, Herrmann B. Mosaic structure of intragenic repetitive elements in histone H1-like protein Hc2 varies within serovars of Chlamydia trachomatis. BMC Microbiol. 2010;10:81.

45. Allen JE, Stephens RS. Identification by sequence analysis of two-site posttranslational processing of the cysteine-rich outer membrane protein 2 of Chlamydia trachomatis serovar L2. J Bacteriol. 1989;171(1):285-91.

46. Christensen S, McMahon RM, Martin JL, Huston WM. Life inside and out: making and breaking protein disulfide bonds in Chlamydia. Crit Rev Microbiol. 2019;45(1):33-50.

47. Pedersen LB, Birkelund S, Christiansen G. Purification of recombinant Chlamydia trachomatis histone H1-like protein Hc2, and comparative functional analysis of Hc2 and Hc1. Mol Microbiol. 1996;20(2):295-311.

48. Grieshaber S, Grieshaber N, Yang H, Baxter B, Hackstadt T, Omsland A. Impact of Active Metabolism on Chlamydia trachomatis Elementary Body Transcript Profile and Infectivity. J Bacteriol. 2018;200(14).

49. Sixt BS, Siegl A, Muller C, Watzka M, Wultsch A, Tziotis D, et al. Metabolic features of Protochlamydia amoebophila elementary bodies--a link between activity and infectivity in Chlamydiae. PLoS Pathog. 2013;9(8):e1003553.

50. Omsland A, Sager J, Nair V, Sturdevant DE, Hackstadt T. Developmental stage-specific metabolic and transcriptional activity of Chlamydia trachomatis in an axenic medium. Proc Natl Acad Sci U S A. 2012;109(48):19781-5.

51. Belland RJ, Zhong G, Crane DD, Hogan D, Sturdevant D, Sharma J, et al. Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proc Natl Acad Sci U S A. 2003;100(14):8478-83.

52. Saka HA, Thompson JW, Chen YS, Kumar Y, Dubois LG, Moseley MA, et al. Quantitative proteomics reveals metabolic and pathogenic properties of Chlamydia trachomatis developmental forms. Mol Microbiol. 2011;82(5):1185-203.

53. Nunes A, Gomes JP. Evolution, phylogeny, and molecular epidemiology of Chlamydia. Infect Genet Evol. 2014;23:49-64.

54. Colonne PM, Winchell CG, Voth DE. Hijacking Host Cell Highways: Manipulation of the Host Actin Cytoskeleton by Obligate Intracellular Bacterial Pathogens. Front Cell Infect Microbiol. 2016;6:107.

55.   Caven L, Carabeo RA. Pathogenic Puppetry: Manipulation of the Host Actin Cytoskeleton by Chlamydia trachomatis. Int J Mol Sci. 2019;21(1).

56.   Ghosh S, Park J, Thomas M, Cruz E, Cardona O, Kang H, et al. Biophysical characterization of actin bundles generated by the Chlamydia trachomatis Tarp effector. Biochem Biophys Res Commun. 2018;500(2):423-8.

57.   Clifton DR, Fields KA, Grieshaber SS, Dooley CA, Fischer ER, Mead DJ, et al. 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. 2004;101(27):10166-71.

58.   Clifton DR, Dooley CA, Grieshaber SS, Carabeo RA, Fields KA, Hackstadt T. Tyrosine phosphorylation of the chlamydial effector protein Tarp is species specific and not required for recruitment of actin. Infect Immun. 2005;73(7):3860-8.

59.   Betts HJ, Wolf K, Fields KA. Effector protein modulation of host cells: examples in the Chlamydia spp. arsenal. Curr Opin Microbiol. 2009;12(1):81-7.

60.   Gitsels A, Sanders N, Vanrompay D. Chlamydial Infection From Outside to Inside. Front Microbiol. 2019;10:2329.

61.   Triboulet S, Subtil A. Make It a Sweet Home: Responses of Chlamydia trachomatis to the Challenges of an Intravacuolar Lifestyle. Microbiol Spectr. 2019;7(2).

62.   Mital J, Lutter EI, Barger AC, Dooley CA, Hackstadt T. Chlamydia trachomatis inclusion membrane protein CT850 interacts with the dynein light chain DYNLT1 (Tctex1). Biochem Biophys Res Commun. 2015;462(2):165-70.

63.   Bocker S, Heurich A, Franke C, Monajembashi S, Sachse K, Saluz HP, et al. Chlamydia psittaci inclusion membrane protein IncB associates with host protein Snapin. Int J Med Microbiol. 2014;304(5-6):542-53.

64.   Derre I, Swiss R, Agaisse H. The lipid transfer protein CERT interacts with the Chlamydia inclusion protein IncD and participates to ER-Chlamydia inclusion membrane contact sites. PLoS Pathog. 2011;7(6):e1002092.

65.   Pokrovskaya ID, Szwedo JW, Goodwin A, Lupashina TV, Nagarajan UM, Lupashin VV. Chlamydia trachomatis hijacks intra-Golgi COG complex-dependent vesicle trafficking pathway. Cell Microbiol. 2012;14(5):656-68.

66.   Lucas AL, Ouellette SP, Kabeiseman EJ, Cichos KH, Rucks EA. The trans-Golgi SNARE syntaxin 10 is required for optimal development of Chlamydia trachomatis. Front Cell Infect Microbiol. 2015;5:68.

67.   Moore ER, Mead DJ, Dooley CA, Sager J, Hackstadt T. The trans-Golgi SNARE syntaxin 6 is recruited to the chlamydial inclusion membrane. Microbiology. 2011;157(Pt 3):830-8.

68.   Flores R, Zhong G. The Chlamydia pneumoniae Inclusion Membrane Protein Cpn1027 Interacts with Host Cell Wnt Signaling Pathway Regulator Cytoplasmic Activation/Proliferation-Associated Protein 2 (Caprin2). PLoS One. 2015;10(5):e0127909.

69.   Wolf K, Plano GV, Fields KA. A protein secreted by the respiratory pathogen Chlamydia pneumoniae impairs IL-17 signalling via interaction with human Act1. Cell Microbiol. 2009;11(5):769-79.

70.   Abdelrahman YM, Belland RJ. The chlamydial developmental cycle. FEMS Microbiol Rev. 2005;29(5):949-59.

71.   Lee JK, Enciso GA, Boassa D, Chander CN, Lou TH, Pairawan SS, et al. Replication-dependent size reduction precedes differentiation in Chlamydia trachomatis. Nat Commun. 2018;9(1):45.

72. Bayramova F, Jacquier N, Greub G. Interactions Screenings Unearth Potential New Divisome Components in the Chlamydia-Related Bacterium, Waddlia chondrophila. Microorganisms. 2019;7(12).

73. Ranjit DK, Liechti GW, Maurelli AT. Chlamydial MreB Directs Cell Division and Peptidoglycan Synthesis in Escherichia coli in the Absence of FtsZ Activity. mBio. 2020;11(1).

74. Abdelrahman Y, Ouellette SP, Belland RJ, Cox JV. Polarized Cell Division of Chlamydia trachomatis. PLoS Pathog. 2016;12(8):e1005822.

75. Ouellette SP, Lee J, Cox JV. Division without Binary Fission: Cell Division in the FtsZ-Less Chlamydia. J Bacteriol. 2020.

76. Cortina ME, Ende RJ, Bishop RC, Bayne C, Derre I. Chlamydia trachomatis and Chlamydia muridarum spectinomycin resistant vectors and a transcriptional fluorescent reporter to monitor conversion from replicative to infectious bacteria. PLoS One. 2019;14(6):e0217753.

77. Hybiske K, Stephens RS. Mechanisms of host cell exit by the intracellular bacterium Chlamydia. Proc Natl Acad Sci U S A. 2007;104(27):11430-5.

78. Gauliard E OS, Rueden KJ, Ladant D. Characterization of interactions between inclusion membrane proteins from Chlamydia trachomatis. Front Cell Infect Microbiol. 2015;5(13).

79. Division of STD Prevention NCfHA, Viral Hepatitis, STD and TB Prevention. Sexually Transmitted Disease Surveillance. In: Prevention CfDCa, editor. 2018.

80. Rowley J VHS, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, et al. . Global and Regional Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2016. In: Organization WH, editor. 2019.

81. Farley TA, Cohen DA, Elkins W. Asymptomatic sexually transmitted diseases: the case for screening. Preventive Medicine. 2003;36(4):502-9.

82. Korenromp EL, Sudaryo MK, de Vlas SJ, Gray RH, Sewankambo NK, Serwadda D, et al. What proportion of episodes of gonorrhoea and chlamydia becomes symptomatic? Int J STD AIDS. 2002;13(2):91-101.

83. Cates W, Wasserheit JN. Genital chlamydial infections: Epidemiology and reproductive sequelae. American Journal of Obstetrics and Gynecology. 1991;164(6):1771-81.

84. Lane AB, Decker CF. Chlamydia trachomatis infections. Dis Mon. 2016;62(8):269-73.

85. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. PLoS Med. 2012;9(12):e1001356.

86. Haggerty CL, Gottlieb SL, Taylor BD, Low N, Xu F, Ness RB. Risk of sequelae after Chlamydia trachomatis genital infection in women. J Infect Dis. 2010;201 Suppl 2:S134-55.

87. Owusu-Edusei K, Jr., Chesson HW, Gift TL, Tao G, Mahajan R, Ocfemia MC, et al. The estimated direct medical cost of selected sexually transmitted infections in the United States, 2008. Sex Transm Dis. 2013;40(3):197-201.

88. HC W. Screening for Chlamydia trachomatis Infections in Women. N Engl J Med. 2017;376(8):765-73.

89. Stephens RS. The cellular paradigm of chlamydial pathogenesis. Trends in Microbiology. 2003;11(1):44-51.

90. Darville T HT. Pathogenesis of Genital Tract Disease due to Chlamydia trachomatis. J Infect Dis. 2010;15(201):S114-S25.

91. Rasmussen SJ, Eckmann L, Quayle AJ, Shen L, Zhang YX, Anderson DJ, et al. Secretion of proinflammatory cytokines by epithelial cells in response to Chlamydia infection suggests a central role for epithelial cells in chlamydial pathogenesis. J Clin Invest. 1997;99(1):77-87.

92. Burstein GR, Gaydos CA, Diener-West M, Howell MR, Zenilman JM, Quinn TC. Incident Chlamydia trachomatis infections among inner-city adolescent females. JAMA. 1998;280(6):521-6.

93. De Clercq E, Kalmar I, Vanrompay D. Animal models for studying female genital tract infection with Chlamydia trachomatis. Infect Immun. 2013;81(9):3060-7.

94. Lyons JM, Ito JI, Jr., Pena AS, Morre SA. Differences in growth characteristics and elementary body associated cytotoxicity between Chlamydia trachomatis oculogenital serovars D and H and Chlamydia muridarum. J Clin Pathol. 2005;58(4):397-401.

95. Morrison RP, Caldwell HD. Immunity to murine chlamydial genital infection. Infect Immun. 2002;70(6):2741-51.

96. Pan X, Yang Y, Zhang JR. Molecular basis of host specificity in human pathogenic bacteria. Emerg Microbes Infect. 2014;3(3):e23.

97. Baumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med. 2013;3(12):a010041.

98. Schubert W-D, Urbanke C, Ziehm T, Beier V, Machner MP, Domann E, et al. Structure of Internalin, a Major Invasion Protein of Listeria monocytogenes, in Complex with Its Human Receptor E-Cadherin. Cell. 2002;111(6):825-36.

99. Mengaud J, Ohayon H, Gounon P, Mège R-M, Cossart P. E-Cadherin Is the Receptor for Internalin, a Surface Protein Required for Entry of L. monocytogenes into Epithelial Cells. Cell. 1996;84(6):923-32.

100. Lecuit M, Dramsi S, Gottardi C, Fedor-Chaiken M, Gumbiner B, Cossart P. A single amino acid in E-cadherin responsible for host specificity towards the human pathogen Listeria monocytogenes. EMBO J. 1999;18(14):3956-63.

101. Granoff DM, Welsch JA, Ram S. Binding of complement factor H (fH) to Neisseria meningitidis is specific for human fH and inhibits complement activation by rat and rabbit sera. Infect Immun. 2009;77(2):764-9.

102. Ngampasutadol J, Ram S, Gulati S, Agarwal S, Li C, Visintin A, et al. Human factor H interacts selectively with Neisseria gonorrhoeae and results in species-specific complement evasion. J Immunol. 2008;180(5):3426-35.

103. Lu L, Ma Z, Jokiranta TS, Whitney AR, DeLeo FR, Zhang JR. Species-specific interaction of Streptococcus pneumoniae with human complement factor H. J Immunol. 2008;181(10):7138-46.

104. Hellwage J, Meri T, Heikkila T, Alitalo A, Panelius J, Lahdenne P, et al. The complement regulator factor H binds to the surface protein OspE of Borrelia burgdorferi. J Biol Chem. 2001;276(11):8427-35.

105. Kraiczy P, Hartmann K, Hellwage J, Skerka C, Kirschfink M, Brade V, et al. Immunological characterization of the complement regulator factor H-binding CRASP and Erp proteins of Borrelia burgdorferi. International Journal of Medical Microbiology Supplements. 2004;293:152-7.

106. Stein M, Kenny B, Stein MA, Finlay BB. Characterization of EspC, a 110-kilodalton protein secreted by enteropathogenic Escherichia coli which is homologous to members of

the immunoglobulin A protease-like family of secreted proteins. J Bacteriol. 1996;178(22):6546-54.

107. Haupt K, Reuter M, van den Elsen J, Burman J, Halbich S, Richter J, et al. The Staphylococcus aureus protein Sbi acts as a complement inhibitor and forms a tripartite complex with host complement Factor H and C3b. PLoS Pathog. 2008;4(12):e1000250.

108. Male CJ. Immunoglobulin A1 protease production by Haemophilus influenzae and Streptococcus pneumoniae. Infect Immun. 1979;26(1):254-61.

109. Reinholdt J, Kilian M. Comparative analysis of immunoglobulin A1 protease activity among bacteria representing different genera, species, and strains. Infect Immun. 1997;65(11):4452-9.

110. de Sousa-Pereira P, Woof JM. IgA: Structure, Function, and Developability. Antibodies (Basel). 2019;8(4).

111. Parrow NL, Fleming RE, Minnick MF. Sequestration and scavenging of iron in infection. Infect Immun. 2013;81(10):3503-14.

112. Noinaj N, Buchanan SK, Cornelissen CN. The transferrin-iron import system from pathogenic Neisseria species. Mol Microbiol. 2012;86(2):246-57.

113. Torres VJ, Pishchany G, Humayun M, Schneewind O, Skaar EP. Staphylococcus aureus IsdB is a hemoglobin receptor required for heme iron utilization. J Bacteriol. 2006;188(24):8421-9.

114. Pishchany G, McCoy AL, Torres VJ, Krause JC, Crowe JE, Jr., Fabry ME, et al. Specificity for human hemoglobin enhances Staphylococcus aureus infection. Cell Host Microbe. 2010;8(6):544-50.

115. McNeil LK, Zagursky RJ, Lin SL, Murphy E, Zlotnick GW, Hoiseth SK, et al. Role of factor H binding protein in Neisseria meningitidis virulence and its potential as a vaccine candidate to broadly protect against meningococcal disease. Microbiol Mol Biol Rev. 2013;77(2):234-52.

116. Aiyar A, Quayle AJ, Buckner LR, Sherchand SP, Chang TL, Zea AH, et al. Influence of the tryptophan-indole-IFNgamma axis on human genital Chlamydia trachomatis infection: role of vaginal co-infections. Front Cell Infect Microbiol. 2014;4:72.

117. Bonner CA, Byrne GI, Jensen RA. Chlamydia exploit the mammalian tryptophan-depletion defense strategy as a counter-defensive cue to trigger a survival state of persistence. Front Cell Infect Microbiol. 2014;4:17.

118. Roshick C, Wood H, Caldwell HD, McClarty G. Comparison of gamma interferon-mediated antichlamydial defense mechanisms in human and mouse cells. Infect Immun. 2006;74(1):225-38.

119. Hooppaw AJ, Fisher DJ. A Coming of Age Story: Chlamydia in the Post-Genetic Era. Infect Immun. 2015;84(3):612-21.

120. Bastidas RJ, Valdivia RH. Emancipating Chlamydia: Advances in the Genetic Manipulation of a Recalcitrant Intracellular Pathogen. Microbiol Mol Biol Rev. 2016;80(2):411-27.

121. LaBrie SD, Dimond ZE, Harrison KS, Baid S, Wickstrum J, Suchland RJ, et al. Transposon Mutagenesis in Chlamydia trachomatis Identifies CT339 as a ComEC Homolog Important for DNA Uptake and Lateral Gene Transfer. mBio. 2019;10(4).

122. Wang Y, LaBrie SD, Carrell SJ, Suchland RJ, Dimond ZE, Kwong F, et al. Development of Transposon Mutagenesis for Chlamydia muridarum. J Bacteriol. 2019;201(23).

123. Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, Hamer JE. Recent advances in large-scale transposon mutagenesis. Current Opinion in Chemical Biology. 2001;5(1):67-73.

124. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol. 2013;11(7):435-42.

125. Choi KH, Kim KJ. Applications of transposon-based gene delivery system in bacteria. J Microbiol Biotechnol. 2009;19(3):217-28.

126. Daubin V, Szollosi GJ. Horizontal Gene Transfer and the History of Life. Cold Spring Harb Perspect Biol. 2016;8(4):a018036.

127. Mortimer RK. Evolution and variation of the yeast (Saccharomyces) genome. Genome Res. 2000;10(4):403-9.

128. Duina AA, Miller ME, Keeney JB. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. Genetics. 2014;197(1):33-48.

129. Roberts AP, Pratten J, Wilson M, Mullany P. Transfer of a conjugative transposon, Tn5397 in a model oral biofilm. FEMS Microbiol Lett. 1999;177(1):63-6.

130. Demars R, Weinfurter J, Guex E, Lin J, Potucek Y. Lateral gene transfer in vitro in the intracellular pathogen Chlamydia trachomatis. J Bacteriol. 2007;189(3):991-1003.

131. Jeffrey BM, Suchland RJ, Quinn KL, Davidson JR, Stamm WE, Rockey DD. Genome sequencing of recent clinical Chlamydia trachomatis strains identifies loci associated with tissue tropism and regions of apparent recombination. Infect Immun. 2010;78(6):2544-53.

132. Gomes JP, Bruno WJ, Borrego MJ, Dean D. 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. 2004;186(13):4295-306.

133. Suchland RJ, Sandoz KM, Jeffrey BM, Stamm WE, Rockey DD. Horizontal transfer of tetracycline resistance among Chlamydia spp. in vitro. Antimicrob Agents Chemother. 2009;53(11):4604-11.

134. Suchland RJ, Carrell SJ, Wang Y, Hybiske K, Kim DB, Dimond ZE, et al. Chromosomal Recombination Targets in Chlamydia Interspecies Lateral Gene Transfer. J Bacteriol. 2019;201(23).

Collaborator's Contributions

This work was made possible by the efforts of many collaborators who performed a number of experiments in this work. In Chapter 1, Figures 1-6 and 1-7 were generated by me and each published previously as part of my contributing authorship in "Transposon Mutagenesis in *Chlamydia trachomatis* identifies CT339 as a ComEC Homolog Important for DNA Uptake and Lateral Gene Transfer" by Labrie et al, 2019, and "Chromosomal Recombination Targets in *Chlamydia* Interspecies Lateral Gene Transfer" by Suchland et al, 2019. As a contributing author in these works, I performed genomic analysis to determine transposon insertion sites, investigate recombination efficiencies, and recombination regions, as well as contributing to content and analysis. Along with these contributing authorships, similar analyses were performed in "Development of Transposon Mutagenesis for *Chlamydia muridarum*" by Wang et al, 2019, "*Chlamydia trachomatis*-containing vacuole serves as deubiquitination platform to stabilize MCL-1 and to interfere with host defense" by Fischer et al, 2017, "Structural and ligand binding analysis of the periplasmic sensor domain of RsbU in *Chlamydia trachomatis* support a role in TCA cycle regulation" by Soules et al, 2020 and finally a co-first authored work entitled "Demonstration of Persistent Infections and Genome Stability by Whole Genome Sequencing of Repeat-Positive, Same-Serovar *Chlamydia trachomatis* Collected from the Female Genital Tract" by Suchland and Dimond et al, 2017.

Chapter 2: Comprehensive Genome Analysis and Comparisons of the Swine Pathogen, *Chlamydia suis* Reveals Unique ORFs and Candidate Host-Specificity Factors" is a published manuscript in FEMS Pathogens and Disease with all analyses performed and written by me, with guidance and direction from my mentor, Dr. P. Scott Hefty.

Chapter 3 is a manuscript in review at Cellular Microbiology entitled "Inter-species Lateral Gene Transfer Focused on the *Chlamydia* Plasticity Zone Identifies Loci Associated with Immediate Cytotoxicity and Inclusion Stability" by Dimond et al. Within this work, I performed all animal model experiments, with assistance from our technician, Nancy Schwarting, and wrote an initial version of the manuscript. Robert Suchland, at the University of Washington, generated the recombinant clones and performed the cytotoxicity imaging and inclusion lysis imaging (Figure 3-6 and Figure 3-7). Srishti Baid performed the growth curve analysis and Scott LaBrie performed the confocal imaging (Figure 3-3). In addition, Scott LaBrie contributed the image and video analysis for the early lysis mutant (Figure 3-9). Katelyn Soules prepared the mRNA for RNAseq and performed RNAseq analyses (Figure 3-2 and Table 3-2). Jacob Stanley assisted in the verification of the RNAseq analysis. Steven Carrell, from Oregon State University, and I performed the genome sequencing of the recombinants used in this study. Forrest Kwong, from the University of Washington, performed the Annexin V and caspase analyses (Figure 3-9) and Yibing Wang, also from the University of Washington, generated the transposon mutants used to generate the chimeras.

Chapter 4, "Chlamydial interspecies recombination reveals candidate genes involved with species-specific growth and infection phenotypes", is a manuscript in preparation. Robert Suchland generated the recombinants used in this study and Steven Carrell and I performed genome sequencing on the recombinant strains. Scott LaBrie assisted with the confocal imaging used in (Figure XXX). I performed all other studies in this chapter.

Chapter 2.

Comprehensive Genome Analysis and Comparisons of the Swine Pathogen,

Chlamydia suis **Reveals Unique ORFs and Candidate Host-Specificity Factors**

Abstract

Chlamydia suis, a ubiquitous swine pathogen, has the potential for zoonotic transmission to humans and often encodes for resistance to the primary treatment antibiotic, tetracycline. Because of this emerging threat, comparative genomics for swine isolate R19 with inter- and intra-species genomes was performed. A 1.094Mb genome was determined through *de novo* assembly of Illumina high throughput sequencing reads. Annotation and subsystem analyses were conducted, revealing 986 putative genes (Chls_###) that are predominantly orthologs to other known *Chlamydia* genes. Subsequent comparative genomics revealed a high level of genomic synteny and overall sequence identity with other *Chlamydia* while 92 unique *C. suis* open reading frames were annotated. Direct comparison of *Chlamydia*-specific gene families that included the plasticity zone, inclusion membrane proteins, polymorphic membrane proteins, and the major outer membrane protein, demonstrated high gene content identity with *C. trachomatis* and *C. muridarum*. These comparisons also identified diverse components that potentially could contribute to host-specificity. This study constitutes the first genome-wide comparative analysis for *C. suis,* generating a fully annotated reference genome. These studies will enable focused efforts on factors that provide key species specificity and adaptation to cognate hosts that are attributed to chlamydial infections, including humans.

Introduction

Members of the genus, *Chlamydia,* are biologically distinctive, obligate intracellular bacterial pathogens that cause disease in both humans and agriculturally important livestock. *Chlamydia suis* is a near-ubiquitous swine pathogen causing reproductive disorders, conjunctivitis, enteritis, rhinitis and pneumonia (1-4). Prevalence of *C. suis* disease within regions throughout the world has been investigated and those reports each found high incidence in farmed pigs. Many *C. suis* strains have acquired a tetracycline-resistance cassette with clinical resistance confirmed in over seven countries (5-9). *In vitro* evidence has also demonstrated the ability for tetracycline resistance to be conveyed to *C. trachomatis* through horizontal gene transfer under conditions of co-infection (10, 11). Though traditionally restricted to its swine host, there is some evidence that *C. suis* may be zoonotically transmitted, although active human infection and clinical symptoms have not been observed (12-14).

The emerging potential threat of *C. suis* and ongoing concerns to the agricultural industry highlight a need for a comparison of the *Chlamydia suis* genome to prominent human pathogen, *Chlamydia trachomatis*. *C. trachomatis* is the most commonly reported bacterial infection worldwide and is the causative agent behind several important diseases including blinding trachoma and a set of different sexually transmitted conditions. Importantly, genome sequencing and comparative studies between *Chlamydia* species have enabled the identification and analysis of gene products that are diverse and may participate in host specific adaptations (15-18). While there have been many intraspecies analyses of *C. suis* with a focus on the tetracycline-resistance genes (19-22), interspecies comparisons with *C. trachomatis* and *C. muridarum*, the species most closely related to *C. suis*, have been limited. One key observation within *Chlamydia* is that despite the genetic

similarity between species, there appear to be many host-restricting factors responsible for the variation in disease between species. Studying the *C. suis* genome in comparison to other *Chlamydia* sp. is expected to identify candidate host-specificity factors.

The *C. suis* R19 strain is a tetracycline resistant strain isolated from a Nebraska farm (23) and has been widely studied for diagnostic and antibiotic resistance transfer capabilities (11, 13, 21). Previously, this strain was sequenced by Joseph et al with 14x coverage, was solved to the contig level and contains many ambiguous bases (20) Additionally, there have been recent advances in gene function annotation methods that are expected to enhance the understanding of genetic content within bacteria. The PathoSystems Resource Integration Center (PATRIC), an NIAID funded bioinformatic toolbox, allows for streamlined workflows for assembly, annotation and protein comparisons (24-25). While PATRIC contains a database of chlamydial genomes and proteins, this resource has not been utilized within the field, although, it has been demonstrated to be useful in the annotation of several other bacterial genomes (26-29). In this study, these public databases were used to annotate the *C. suis* genome, along with reannotation for *C. trachomatis* and *C. muridarum*, allowing for the identification of several new open reading frames and possible coded protein functions. These analyses have confirmed previous findings that the *C. suis* genome is highly similar in content and organization to both *C. trachomatis* and *C. muridarum*. These analyses also demonstrated that *C. suis* encodes for many genes similar to other *Chlamydia*. Finally, comparative analyses show specific genes which may play a role in the host-adaptation of *Chlamydia*.

Results

The *Chlamydia suis* genome architecture and composition.

The *C. suis* R19 strain genome was sequenced using Illumina HiSeq with over 1000X coverage, enabling an unbiased, *de novo* assembly of the genome (Figure 2-1). The completed genome includes a chromosome 1,094,719 bp in length and a plasmid of 7,496 bp. These data agree with previously performed studies on *C. suis* strains. Read density analysis of the plasmid relative to the chromosome supported approximately 2 copies of the plasmid per chromosome. The genome was annotated using both direct ORF prediction and BLAST homology-based annotation through Geneious and the PATRIC annotation platform. This analysis revealed 986 coding regions (***Chl**amydia **s**uis*; Chls_###), with 6 rRNAs and 37 tRNAs, and an overall 41.7% G+C content (Figure 2-2, Table 2-1). The origin of replication for this genome was assigned according to the reference genome *C. trachomatis* L2 434/Bu (NC_010287) (30). Of the 986 coding sequences identified in *C. suis*, 686 were assigned through comparisons to other *Chlamydia* and through inferred functional assignment of known motifs with a 75% identity threshold. The assigned genes were divided into 9 broad categories or defined as "other" using the subsystems approach and the RAST toolkit (31-32). Overall, the largest division, at 165 genes, were predicted to be involved with protein processing followed closely by metabolism-related genes (163 ORFs; Figure 1B). Additional subsystems include those involved with stress response, defense or virulence (50 genes), DNA processing (51 genes), energy (62 genes), RNA processing (31 genes), cellular processes (28 genes), cell envelope (28 genes) or membrane transport (3 genes). The remaining 300 genes had no homologues or identified motifs in Genbank and were annotated as hypothetical. The *C. suis* R19 tetracycline resistance cassette, likely integrated by the IS605 transposase and horizontal gene transfer, is thought to be spreading among *C. suis* strains (33-34). Many analyses

of the *tet* island demonstrate that the island is plastic, but that the key *tet* (C) gene was always

intact, as it is for R19, and that the cassette is transferable between *C. suis* strains by homologous

recombination without the need for the transposase (16-17, 21-22). As in the other species in the

genus, there is one plasmid with 8 genes encoding replicative machinery and putative virulence

factors (35).
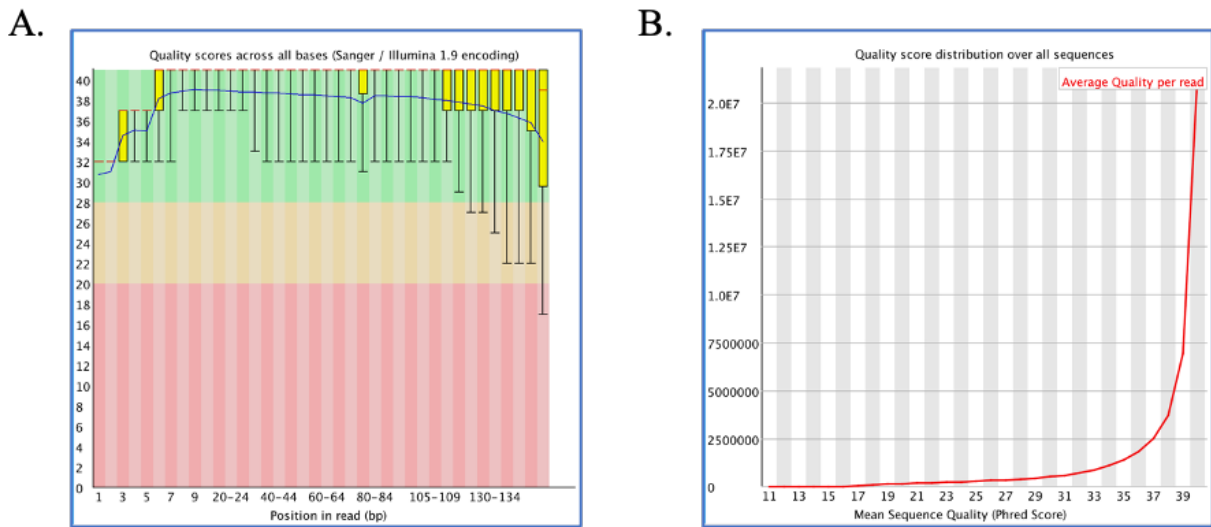
Figure 2-1. A) Quality scores for Illumina reads and B) quality score distribution demonstrate overall high quality for reads generated by FastQC.

Table 2-1. Comparative Characteristics of the *C. suis* R19 Genome

	Chlamydia suis R19	*Chlamydia trachomatis* 434/Bu	*Chlamydia muridarum* Nigg
Chromosome Length (bp)	1,094,719	1,038,842	1,072,950
Chromosomal genes (PATRIC)	986	962	983
NCBI reported chromosomal genes	N/A	937	952
Plasmid length (bp)	7,496	7,499	7,501
Plasmid genes	8	8	8
GC Content (%)	41.7	41.3	40.3
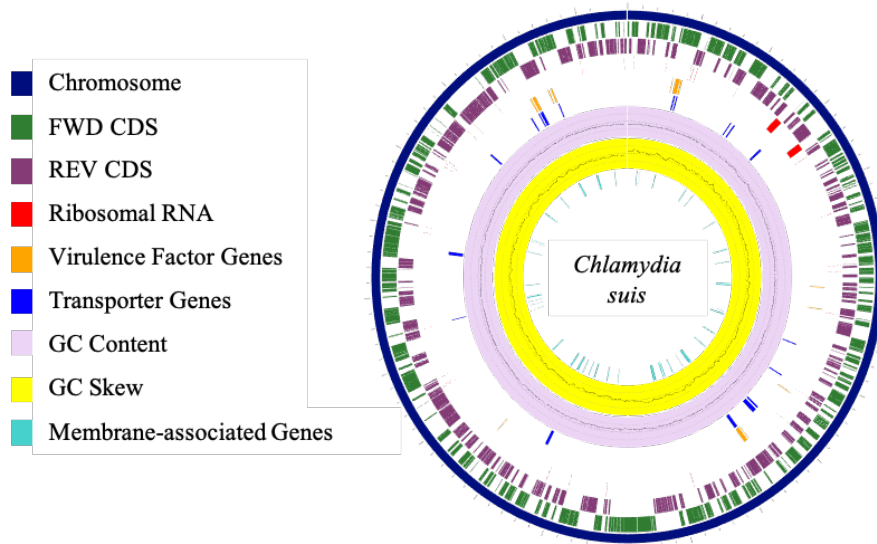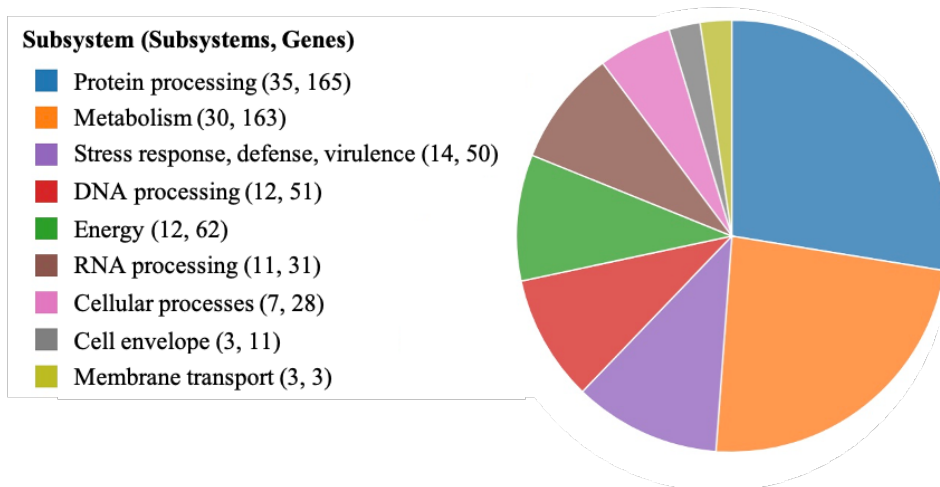
Figure 2-2. A) Annotated Genome for *Chlamydia suis* R19. Rings, from outside, show: Chromosome (navy), all annotated coding regions in the forward direction (green), annotated coding regions in the reverse direction (violet), non-coding annotations including ribosomal RNAs (red), known virulence factors including type 3 secretion components, as automatically annotated through PATRIC annotation platforms (gold), functionally annotated transporters (blue), relative GC content (lavender ring) and GC skew (yellow ring), annotated membrane-associated proteins assigned through structural prediction and relatedness to reference *Chlamydia* and inclusion membrane proteins determined through comparative annotation and canonical hydrophobicity plots (cyan). B) Subsystems of functionally assigned genes excluding the 282 hypothetical genes. 122 assigned genes were also not categorized into one of these subsystems and were left out of the chart. The remaining genes were divided into 9 subsystems including Protein processing, Metabolism, Stress response/defense/virulence, DNA processing, Energy, RNA processing, Cellular processes, Cell envelope and Membrane transport.

Comparative analysis of the *C. suis* coding regions to *C. trachomatis* and *C. muridarum* representative strains.

Chlamydia trachomatis L2/434/Bu contains a predicted 962 genes while the *C. muridarum* Nigg genome consists of 983 genes, as re-annotated under the PATRIC annotation platform (Table 2-1). There is approximately 90% synteny in the genome organization for these three genomes, as well as to other members of the phylum (Figure 2-3A). There are four key regions with a notable absence of conservation. The first of these (around 140-160kb) correlates with the ribosomal RNA operons. The second region (around 310kb) has diversity which can be attributed to a polymorphic membrane protein *(pmp)* operon where not all genes are conserved between *Chlamydia.* The third region (around 530kb), represents the plasticity zone which is where the greatest diversity is expected between species. The final region (around 850kb) is an additional *pmp* operon. Over 80% of proteins encoded by both *C. trachomatis* and *C. muridarum* have greater than 75% amino acid similarity with *C. suis* orthologs, further highlighting the evolutionary relationship of these species (Figure 2-4A). Both this genetic comparison, and genome-wide sequence phylogenies (Figure 2-3B) corroborate previous findings using 16S sequences that *C. suis* is in a phylogenetic clade with *C. trachomatis* and *C. muridarum* (36). Upon closer inspection of the genetic make-up of the *C. suis* genome, 20 proteins considered to be key for phylogenetic classification of *Chlamydia* are present, intact, and highly similar to the other *Chlamydiales* (37).

A.

List of tracks, from outside to inside:
1. *Chlamydia suis* R19
2. *Chlamydia trachomatis* D/UW-3/CX
3. *Chlamydia trachomatis* 434/Bu
4. *Chlamydia muridarum* Nigg
5. *Chlamydia pneumoniae* strain Wien 1
6. *Chlamydia pecorum* strain E58
7. *Chlamydia psittaci* GR9
8. *Chlamydia abortus* strain GIMC 2006:CabB577
9. *Chlamydia caviae* GPIC
10. *Chlamydia felis* Fe/C-56

Percent protein sequence identity

100	99	80	40
99.9	98	70	30
99.8	95	60	20
99.5	90	50	10

B.

Figure 2-3. A) Circular alignment of *C. suis* R19 coding regions (outer ring) with canonical *Chlamydia*. Protein sequence identity is indicated by colorimetric scale where purple/blue represents higher identity than orange/red. White or blank areas represent an absence of coding regions annotated at that site. This alignment was performed using the PATRIC proteome comparison tool. B) Phylogenetic analysis of *Chlamydia* using whole-genome sequences aligned through progressiveMauve demonstrates the positioning of representative *C. suis* genomes compared with selected *C. trachomatis* and *C. muridarum* reference genomes.

- 47 -
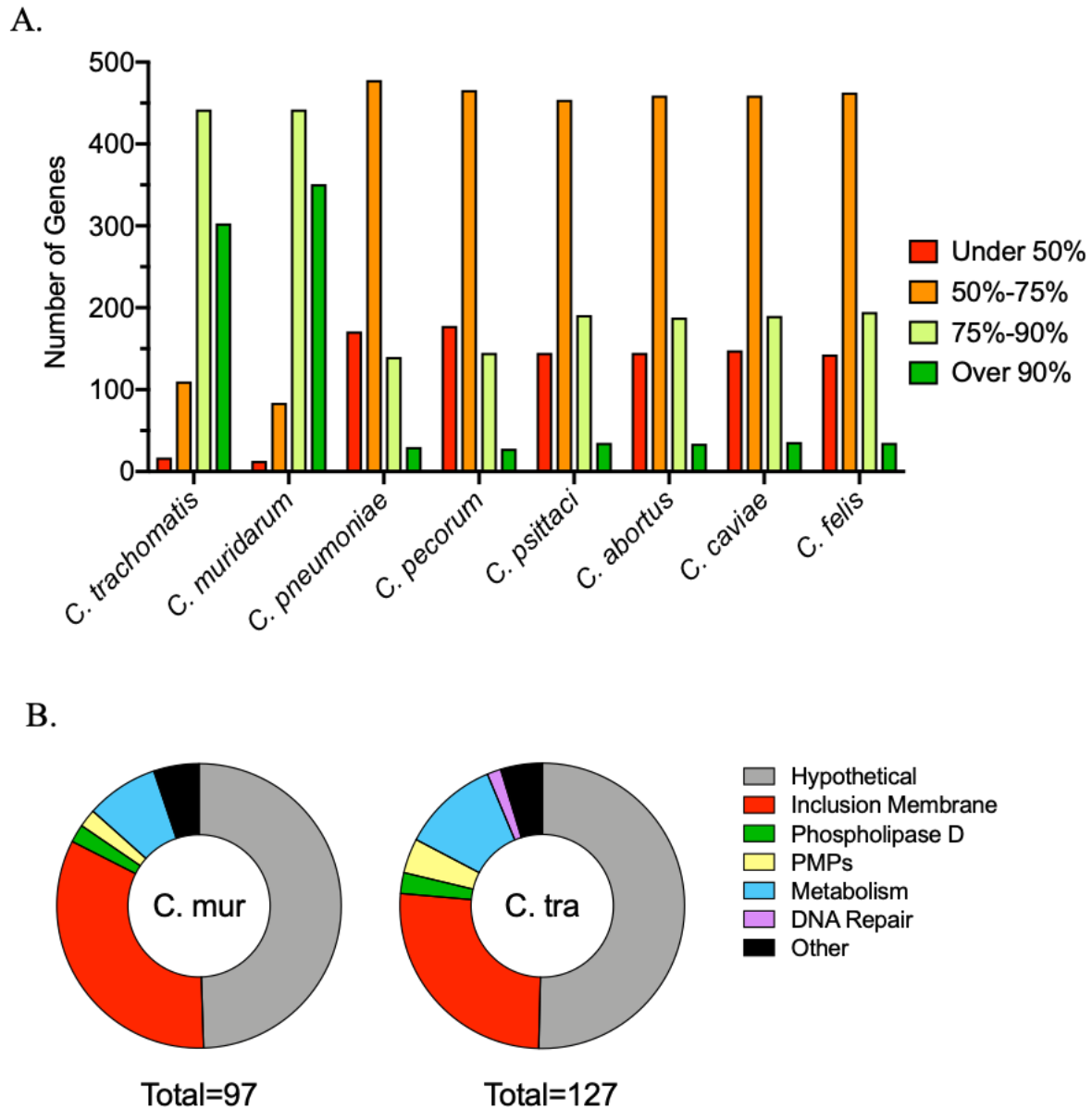
Figure 2-4. A) Amino acid similarity percentages for conserved protein coding sequences. *C. trachomatis* D/UW-3 was used in this comparison. B) Protein families for *C. muridarum* Nigg (C. mur) and *C. trachomatis* D/UW-3 (C. tra) that fall below 75% amino acid similarity with their *C. suis* orthologs. Colors correspond to protein families as defined by subsystems analysis and specific protein function.

C. suis encodes for 92 genes that are not present in either *C. trachomatis* or *C. muridarum* (Table 2-2). In each instance, *C. trachomatis* D/UW-3 and *C. muridarum* Nigg were investigated for syntenous orthologs and these appear to be unique to *C. suis*. Most of these are hypothetical proteins although several are known members of the tetracycline resistance island, an inclusion membrane protein, and genes within the plasticity zone. Along with these unique *C. suis* genes appear to be a subset of orthologous that are shared between either *C. suis* and *C. trachomatis* or *C. suis* and *C. muridarum* (Table 2-3).

Table 2-2. Unique *C. suis* Open Reading Frames

Gene	AA Length	Annotated Function	Gene	AA Length	Annotated function
Chls 011	90	hypothetical protein	Chls 331	51	hypothetical protein
Chls 021	40	hypothetical protein	Chls 334	41	hypothetical protein
Chls 028	54	hypothetical protein	Chls 339	41	hypothetical protein
Chls 035	43	hypothetical protein	Chls 363	62	hypothetical protein
Chls 037	71	hypothetical protein	Chls 378	42	hypothetical protein
Chls 056	41	hypothetical protein	Chls 396	55	hypothetical protein
Chls 058	45	hypothetical protein	Chls 404	285	hypothetical protein
Chls 061	54	hypothetical protein	Chls 405	105	hypothetical protein
Chls 072	41	hypothetical protein	Chls 406	91	hypothetical protein
Chls 080	49	hypothetical protein	Chls 407	44	hypothetical protein
Chls 100	62	hypothetical protein	Chls 421	166	hypothetical protein
Chls 105	37	hypothetical protein	Chls 438	45	hypothetical protein
Chls 113	44	hypothetical protein	Chls 452	38	hypothetical protein
Chls 123	459	Metal-dependent hydrolase	Chls 454	47	hypothetical protein
Chls 124	292	Replication protein RepA	Chls 460	53	hypothetical protein
Chls 125	108	hypothetical protein	Chls 461	40	hypothetical protein
Chls 126	107	Growth inhibitor, PemK family	Chls 462	67	hypothetical protein
Chls 127	417	Relaxase	Chls 464	54	hypothetical protein
Chls 128	167	Mobilization protein, MobB	Chls 468	92	hypothetical protein
Chls 129	118	Mobilization protein, MobC	Chls 474	38	hypothetical protein
Chls 130	227	Mobilization protein, MobD	Chls 475	261	Gamma-glutamylcyclotransferase
Chls 131	212	Mobilization protein, MobE	Chls 478	649	hypothetical protein
Chls 132	250	OfxX fusion product	Chls 480	37	hypothetical protein
Chls 133	101	Resolvase	Chls 503	29	hypothetical protein
Chls 134	396	Tetracycline resistance, MFS efflux pump, Tet(C)	Chls 536	51	hypothetical protein
Chls 135	211	Tetracycline resistance regulatory protein, TetR	Chls 540	50	hypothetical protein
Chls 136	151	Transposase	Chls 542	37	hypothetical protein
Chls 137	459	Mobile element protein	Chls 578	44	hypothetical protein
Chls 138	868	Hemagglutinin	Chls 602	38	hypothetical protein
Chls 143	44	hypothetical protein	Chls 634	52	hypothetical protein
Chls 150	71	hypothetical protein	Chls 642	41	hypothetical protein
Chls 151	37	hypothetical protein	Chls 647	94	hypothetical protein
Chls 175	52	hypothetical protein	Chls 648	196	hypothetical protein
Chls 188	51	hypothetical protein	Chls 649	46	hypothetical protein
Chls 213	44	hypothetical protein	Chls 679	38	hypothetical protein
Chls 218	39	hypothetical protein	Chls 682	39	hypothetical protein
Chls 234	47	hypothetical protein	Chls 692	40	hypothetical protein
Chls 235	41	hypothetical protein	Chls 727	44	hypothetical protein
Chls 236	58	hypothetical protein	Chls 729	37	hypothetical protein
Chls 237	48	hypothetical protein	Chls 753	47	hypothetical protein
Chls 248	30	hypothetical protein	Chls 785	66	hypothetical protein
Chls 269	50	hypothetical protein	Chls 788	78	hypothetical protein
Chls 285	41	hypothetical protein	Chls 790	51	hypothetical protein
Chls 289	342	Ribose-phosphate pyrophosphokinase	Chls 801	52	hypothetical protein
Chls 290	44	hypothetical protein	Chls 823	38	hypothetical protein
Chls 300	39	hypothetical protein	Chls 874	44	hypothetical protein

Light grey shading corresponds to the phage-origin tetracycline resistance island
Dark grey shading corresponds to the plasticity zone

Table 2-3. *C. suis* ORFs shared only with *C. trachomatis* or *C. muridarum*

Gene	AA Length	Annotated Function	Ortholog[a]	AA Similarity (%)	ORF Coverage (%)
Chls 472	370	hypothetical protein	CT161[b]	56.4	94.3
Chls 473	540	hypothetical protein	CT163	67.3	99.8
Chls 479	96	Transcriptional repressor protein TrpR	CT169	81.9	98.9
Chls 481	392	Tryptophan synthase beta chain (EC 4.2.1.20)	CT170	90.8	99.7
Chls 482	250	Tryptophan synthase alpha chain (EC 4.2.1.20)	CT171	68.8	99.6
Chls 484	580	hypothetical protein	CT161[b]	45.2	88.6
Chls 534	181	hypothetical protein	CT222	52.9	93.2
Chls 535	355	Inclusion membrane protein	CT223	41.1	65.6
Chls 538	102	Inclusion membrane protein	CT224	44.8	70.7
Chls 539	93	hypothetical protein	CT225	59.3	73.8
Chls 547	111	Inclusion membrane protein B	CT232	84.2	92.6
Chls 476	3362	hypothetical protein	TC0438	74.1	99.9
Chls 477	3224	hypothetical protein	TC0439	75.2	99.9
Chls 529	829	DNA helicase IV (EC 3.6.4.12)	TC0490	80.5	96.0
Chls 537	323	Inclusion membrane protein	TC0496	36.1	66.0
Chls 585	105	hypothetical protein	TC0541	59.8	99.1
Chls 617	184	Inclusion membrane protein	TC0574	56.9	57.1

[a]CT – *C. trachomatis* and TC – *C. muridarum*
[b]Chls 472 and Chls 484 show similarity to CT161 supporting a potential gene duplication in *C. suis* R19

C. suis protein-coding sequences with lower similarity (<75%) to orthologs in *C. trachomatis* and *C. muridarum* are largely proteins of unknown function and also include potential host-interacting components such as inclusion membrane proteins, polymorphic membrane proteins, and the phosopholipase D-like genes. Albeit a minority of those with low sequence similarity, proteins in other families such as those related to metabolism, DNA repair and other families, also are included (Figure 2-4B). In Figure 2-4B, genes within the "other" category include; the MAC/perforin, a putative eukaryotic transcriptional activator, competence protein ComEC (38), late transcription unit B and Tol-Pal cell envelope complex TolA protein.

Individual proteins encoded by *C. suis* were compared to orthologs in six other *Chlamydia* species. As expected, those proteins with the highest primary sequence similarity include components

critical for replication, transcription, translation and metabolism. Included in these highly conserved sequences are many hypothetical proteins that may contribute to those shared *Chlamydia*– and bacteria- specific processes. In contrast, for the other six chlamydial species represented here, this ratio flips and the majority of conserved proteins fall below the 75% similarity range (Figure 2-4A). These results, particularly the subset of lowly conserved orthologs, highlight possible host-specificity candidates within *Chlamydia suis*: the hypothetical proteins, plasticity zone components and membrane proteins.

The *C. suis* plasticity zone

The chlamydial plasticity zone (PZ) is a variable region present in all species of the genus that is hypothesized to contain genes responsible for host-specific virulence factors. The PZ of *C. muridarum* contains a series of phospholipases (PLDs), three genes often annotated as adherence factors that have cytotoxic domains similar to those in *Clostridium spp,* as well as a purine interconversion operon (39). In contrast, the significantly reduced PZ of *C. trachomatis* contains several of the phospholipases, but only contains a truncated version of one of the cytotoxin-resembling genes in specific serovars and has replaced the purine interconversion operon with a tryptophan biosynthesis pathway (Figure 2-5A) (40). Other *Chlamydia* contain permutations of these plasticity zones where some contain no phospholipases, variable cytotoxin numbers with two present in *C. pecorum* and one ortholog in *C. caviae*, *C. felis,* and *C. psittaci*. The presence or absence of the tryptophan operon and close investigation of the PZ has provided insight into the immunological interactions between host and bacteria (41-43).
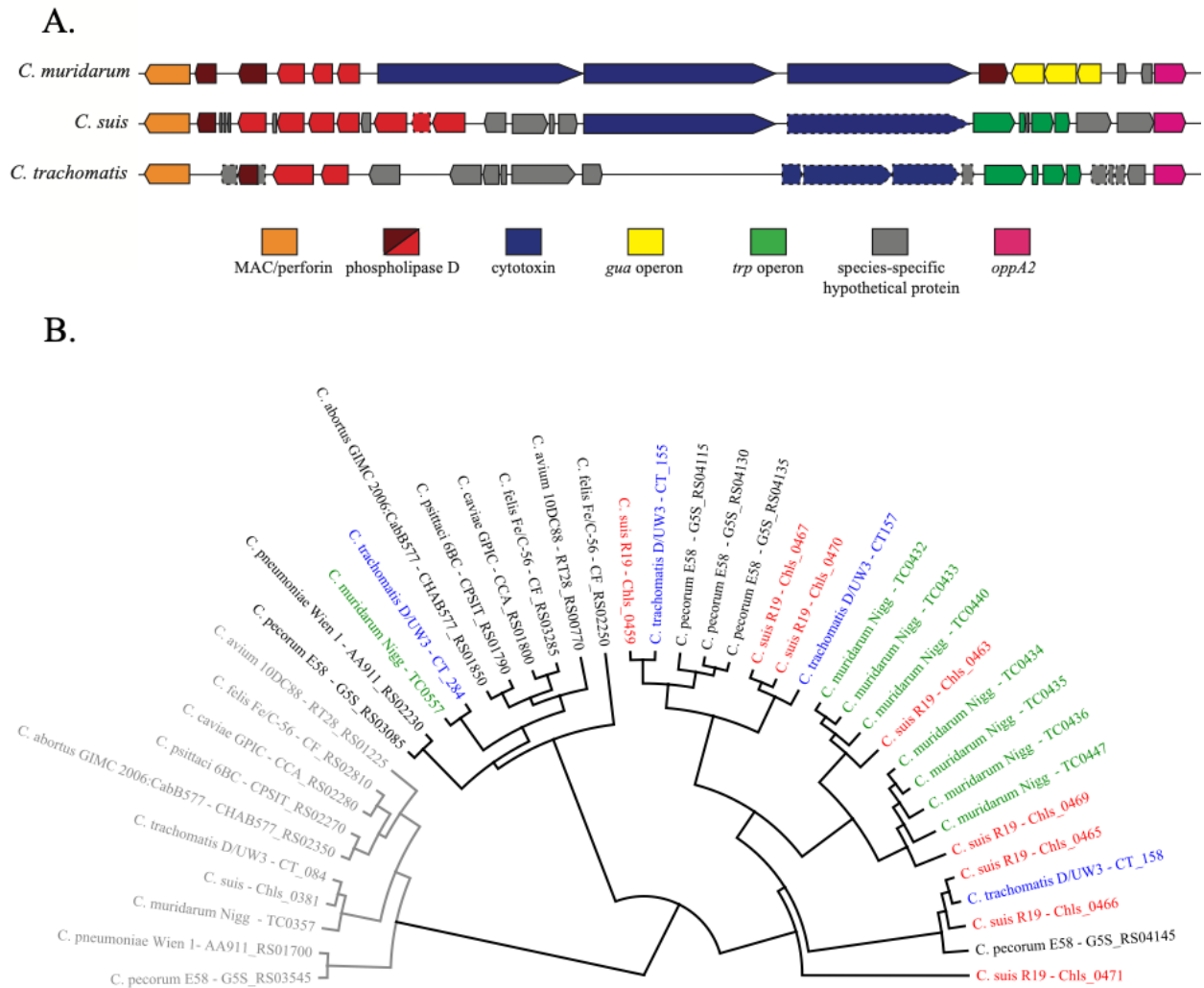
Figure 2-5. A) Gene organization for the plasticity zones of composite *C. suis*, *C. muridarum* and composite *C. trachomatis* strains. Mac/perforin annotated genes (orange) were used as the left margin for the PZ while the oppA2 genes (pink) was used as a right margin. Grey arrows represent hypothetical proteins which have no known function. Genes with dashed outlines are variable within the species. Key differences between these PZs include the presence and number of the cytotoxins or adherence factors (blue), the tryptophan biosynthesis operon (green) and phospholipase D-like genes (dark or light red depending on clade in B). B) Phylogenetic cladogram comparing amino acid sequences for phospholipase D-like genes from eleven species of *Chlamydia.* Total alignment was 1000aa in length. Marked in gray are the genes considered to be the ancestral chromosomal phospholipases. The PZ phospholipases from *C. suis* (red), *C. muridarum* (green) and *C. trachomatis* (blue) are indicated as well.

The *C. suis* PZ contains two putative cytotoxins, both of which are similar to two of the *C. muridarum* cytotoxin-like genes as also determined by Manuela et al for *C. suis* MD56 (Figure 2-5A) (44). To investigate the presence of these cytotoxins in other strains of *C. suis*, full or partial genomes were analyzed for 29 deposited SRA samples using the *C. suis* R19 genome as a reference (22). Interestingly, when these strains of *C. suis* were examined, seven were discovered to contain only one of the adherence factors, indicating that both copies are likely non-essential for the *in vivo* success of the organism, as also noted by Seth-Smith et al (22). It is unclear as of yet, whether the two cytotoxins have distinct biological functions *in vivo* or whether they play a role in host or biovar selectivity. Also found in the *C. suis* PZ is an intact tryptophan biosynthesis pathway, homologous to the one in *C. trachomatis* genital strains.

As is seen in *C. muridarum* and *C. trachomatis*, the PZ of *C. suis* contains a putative operon encoding phospholipases (PLDs). Importantly, these proteins have some of the most lowly conserved amino acid similarities. The operon in *C. suis,* however, seems to contain an increased number of PLDs relative to both *C. trachomatis* and *C. muridarum* (Fig. 2-5A). Phylogenetic analysis of the repertoire of PLD enzymes, as identified through their $HxKx_4Dx_6G$ domains, revealed a distinct clade of non-PZ PLD which mirror classic phylogenetic relationships for the *Chlamydia* (Figure 2-5B)*.* This indicates that this non-PZ PLD is likely truly ancestral and has evolved with the speciation of *Chlamydia*. The pzPLDs, however, are not so simply categorized. There is a great diversity in these sequences, with a subset of proteins more closely related the chromosomal PLD, perhaps representing duplication events. The clade more distantly related the chromosomal PLDs include the greatest number of pzPLDs, potentially pointing to a unique role

for this subset of PLDs. The division between clades could suggest differential selection for these PLDs which would support a distinct biological function.

C. suis membrane proteins

One key protein differentiating members of *Chlamydiaceae* is the major outer membrane protein (Momp, *ompA).* This porin-like structure is integrated into the outer membrane of all *Chlamydia* species and has been shown to make up approximately 60% of the chlamydial outer membrane (53). While structural analyses of this protein suggest it to be a general porin, its conserved and highly present nature indicates that this gene is crucial to bacterial survival in the host. In 2016, a study was conducted which looked at variation in *ompA* among strains of *C. muridarum* collected from wild mice and found a 99% identity between *ompA* genes (54). Similarly, several strains of *Chlamydia pecorum, Chlamydia abortus*, *Chlamydia pneumoniae*, and *Chlamydia psittaci* (55), have been deposited and close examination of these strains shows little *ompA* diversity. In contrast, *C. trachomatis ompA* has increased variability between strains when compared with *C. muridarum* and has allowed classification into unique serovars or clades based solely on *ompA*-typing.

While it has been shown that genotyping by *ompA* does not indicate genomic relatedness (56), these divisions do appear to correlate with specific pathobiovars. In *C. trachomatis*, *ompA* appears to be evolving at a faster rate than the rest of the genome, likely due to unique selective pressures, (57) and is characterized by four variable domains (VD) which are used to distinguish serovars. These variable domains are generally regions of lower hydrophobicity and potentially indicate outward facing domains, leading to the hypothesis for the biovar-specificity attributed to this gene.

The variability in *ompA* could confer an evolutionary advantage specific to the host organism and the specific tissues involved in infection.

A recent analysis by Chahota et al investigated the potential for serotyping by Momp in *C. suis* using PCR analysis of VD2 and VD3 (58). When the deposited contigs were investigated for completed genes for *ompA*, and compared to R19, similar results were found. *C. suis* Momp also contains four variable domains which directly correspond to those seen in *C. trachomatis*, though are divergent (Figure 2-6). Phylogenetic analysis of the *C. suis ompA* divides the strains into distinct clades, corroborating the observation by Chahota et al, that *C. suis* strains could be serotyped and a larger study could be performed to investigate the potential for distinctions between the clades.

There are several other important gene families that show some variability between the *Chlamydial*es, including a family of inclusion membrane proteins (Incs), and polymorphic membrane proteins (Pmps). Inclusion membrane proteins, in particular, may be targets for host specificity, as the conserved Incs share disproportionately lower amino acid sequence similarity between species. Both families have been shown to be correlated with tissue tropism and therefore, may play a role in speciation (59). Relatively little is known about the modifications made to this vacuole that allow for the bacteria's survival, but the Inc family of proteins are secreted into the inclusion membrane and are known to interact with the host cytosol. Inc proteins make up approximately 7-10% of the chlamydial proteome and have a high degree of diversity (60). Despite this diversity, most Incs are defined by a characteristic bilobed hydrophobicity motif. This domain is conserved among predicted Inc proteins and seems to be a *Chlamydia*-specific motif.
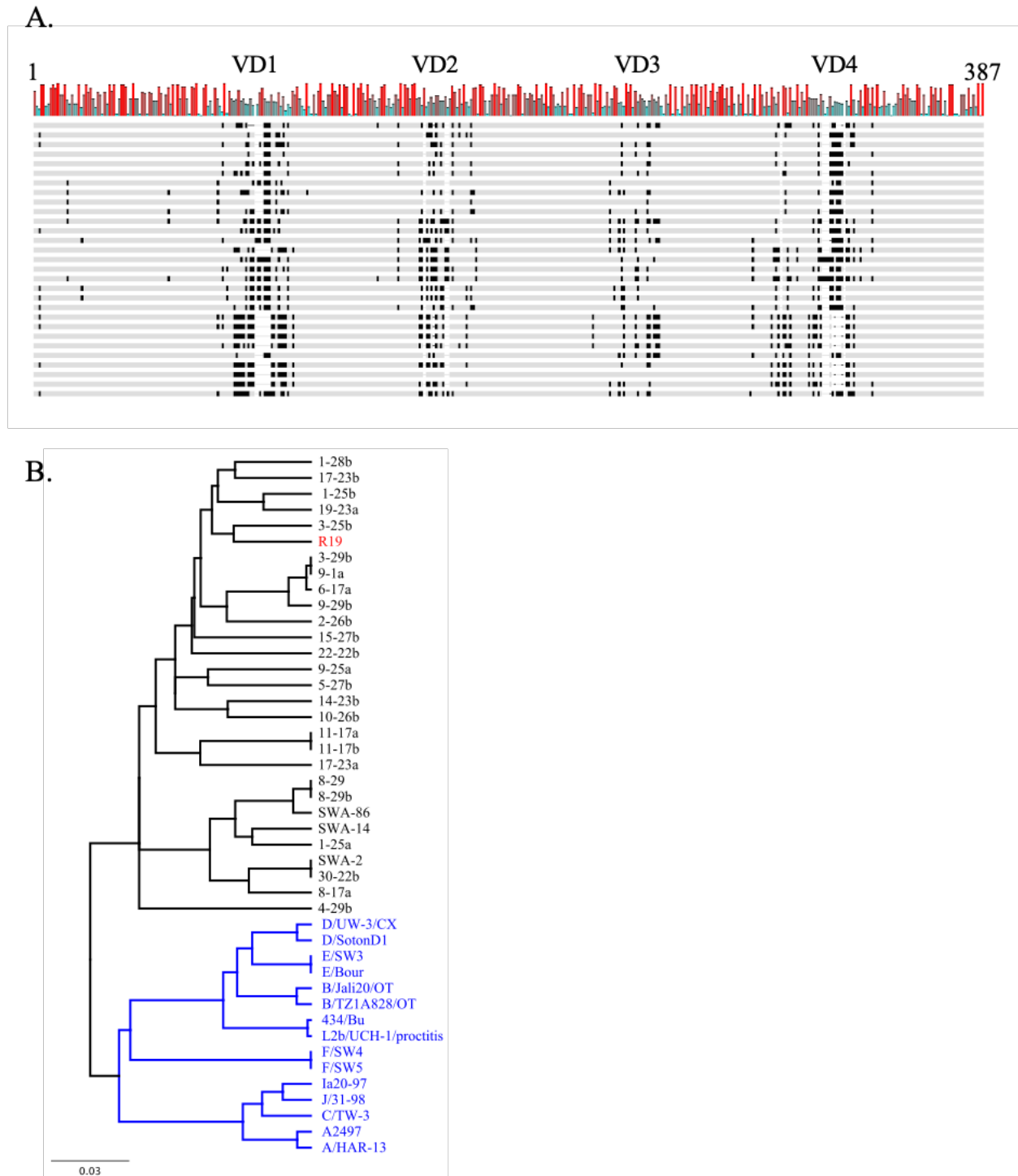
Figure 2-6. A) Alignment of *C. suis* Momp primary sequences with differences are marked in black. Regions of the greatest variation appear are labelled as variable domains (VD). Mean hydrophobicity is plotted in red and blue along the top. B) Phylogenetic tree from A), with the addition of select *C. trachomatis* Momp sequences (blue). R19 is indicated in red. Scale bar is in substitutions per site.

There are varying predictions for the numbers of Inc proteins across the genus with 23 conserved *incs* encoded in 5 examined species (61). The *C. suis* R19 genome encodes each of these conserved *inc*s as well as 34 non-conserved predicted Incs for a total of 57 predicted Incs. According to Lutter et al, *C. trachomatis* also encodes for 55 *incs* with 6 that differ from *C. muridarum,* which has 53 Incs (61). Analysis of the 57 putative inclusion membrane proteins in *C. suis* that were identified and aligned with *C. trachomatis* and *C. muridarum* revealed an additional 2 putative Incs for *C. muridarum*. A Venn diagram and phylogeny were constructed to display Incs shared between or specific to the three species and their evolutionary relationship (Figure 2-7). Briefly, 51 Incs are conserved within the three species of the clade. One Inc is unique to *C. suis* (Chls_474) and one to *C. muridarum* (TC0011) (Figure 2-7C, panels 1 and 8). Several Incs are shared between species. Three Incs are found in *C. trachomatis* and *C. suis* only; CT222 (Chls_534), CT224 (Chls_538) and CT225 (Chls_539). These are in an operon of inclusion membrane protein genes and possibly function together as a complex. While CT222 and Chls534 are syntenous, phylogenetic analysis suggests that these may not be direct orthologs (Figure 2-7C, panels 2 and 6). Additionally, two Incs are unique to *C. muridarum* and *C. suis*, TC0496 (Chls_537) and TC0573 (Chls_617). Interestingly, and unexpectedly, due to the placement of *C. suis* as an intermediate between *C. trachomatis* and *C. muridarum*, there appears to be one Inc that is absent in *C. suis* but found in the other two species (CT227/TC0498) (Figure 2-7C, panel 1). This Inc is closely related to the neighboring Inc CT226/TC0497 which is also present in *C. suis*. In general, the conservation of Incs between *C. suis*, *C. muridarum* and *C. trachomatis* suggests that the protein functions may be retained in all three species and that they may share cognate host interactions. Non-conserved Incs may indicate key differences in host interactions or the host environment in general. Functional

analysis of host-binding or -interacting partners for each of these Incs, could reveal their specific

roles in their host species.
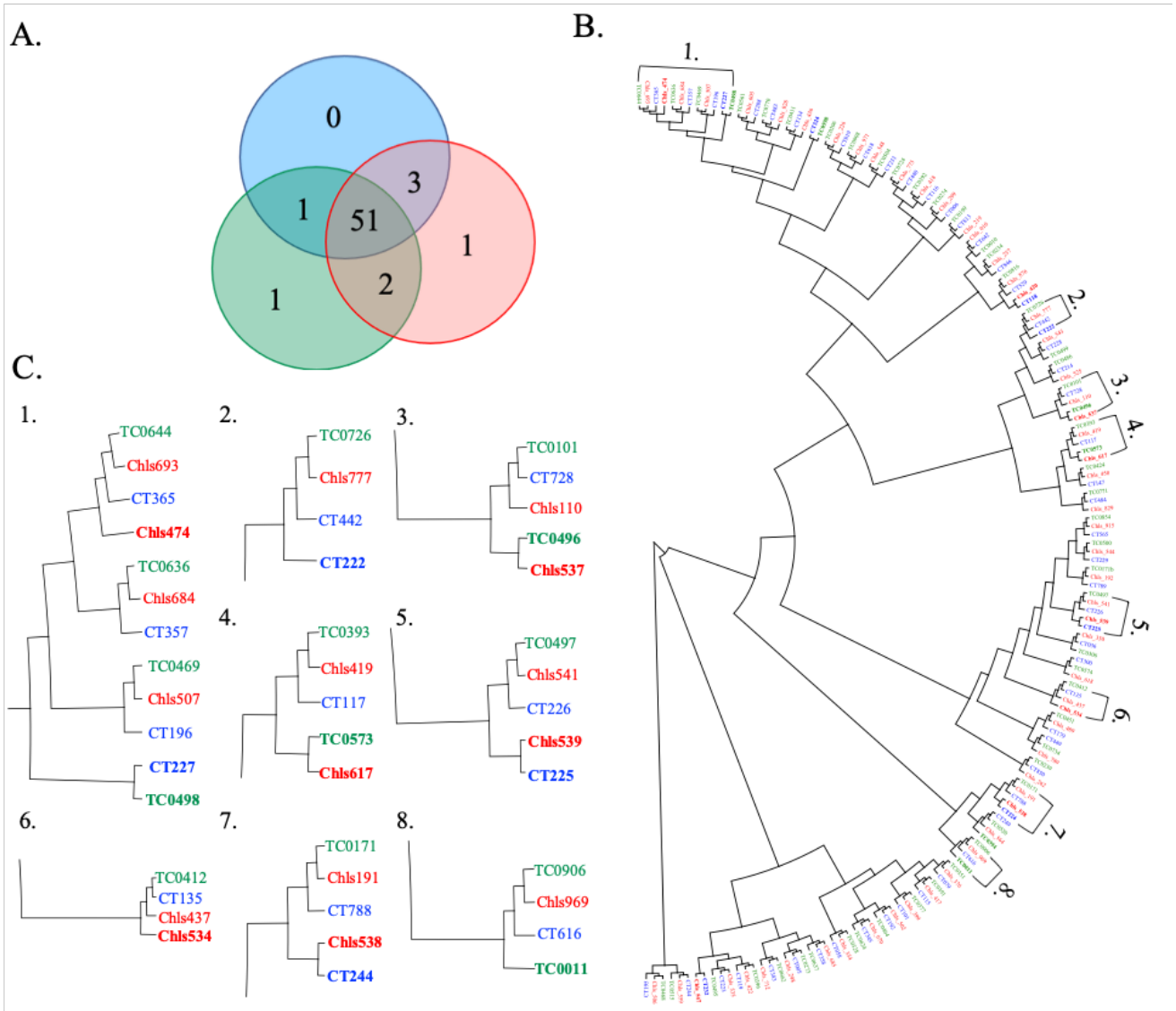
Figure 2-7. A) Venn diagram displaying the shared or specific inclusion membrane proteins (Incs) between *C. trachomatis* D/UW-3 (blue), *C. suis* R19 (red) and *C. muridarum* Nigg (green). B) Phylogenetic analysis of putative Inc protein sequences. C) Highlighted in boxes 1-8 are the Incs that are not conserved between *C. trachomatis* (blue), *C. muridarum* (green) and *C. suis* (red). Tree is transformed with ordered branching.

Like Incs, polymorphic membrane proteins (Pmps) share in varying numbers across species, though the specific roles for each Pmp are unknown. Pmps are outer membrane proteins and are immunogenic for humans. These genes make up approximately 1-2% of the chlamydial gene content and each contain a C-terminal phenylalanine and as well as multiple GGAI motifs which are associated with host cell adhesion (62). Nine putative *pmps* were annotated in *C. suis* R19 through a genome wide search for multiple GGAI motifs. Each of these correspond directly to a Pmp in *C. trachomatis* and *C. muridarum*. These Pmp primary sequences range from high identity between the three species to low identity: PmpA (79.8%), PmpI (78.8%), PmpH (77.0%), PmpB (74.9%), PmpG (74.0%), PmpE (73.9%), PmpD (72.9%), PmpF (65.2%) and PmpC (46.3%). Seven out of nine *C. suis* R19 Pmps are more similar to their *C. muridarum* orthologs while two (PmpH and PmpF) are more similar to *C. trachomatis.* Interestingly, Pmps have been shown to be involved in cellular tropism as six of the nine Pmps in *C. trachomatis* analyzed from different serovars were able to be phylogenetically clustered based on disease properties, suggesting a role for these membrane proteins in adhesion or differential biovar tropism (62).

Discussion

The *C. suis* R19 genome was assembled through whole genome sequencing and fully annotated using PATRIC. One key aspect of this study was the use of NIAID annotation service, PATRIC, for independent annotation. As indicated in Table 2-1, the PATRIC platform has annotated several genes not previously annotated by other methods in both *C. trachomatis* and *C. muridarum*. It is likely that these new annotations correspond with known open reading frames that were not originally included in the first depositions of these genomes. These could correspond with pseudogenes or untranscribed genes, as many are small <300bp open reading frames. However, in

recent years, it has been shown that small open reading frames (smORFs), previously discarded in many genome annotation pipelines, do produce transcribed and translated proteins in both bacteria and eukaryotes (63-64). These smORFs are predicted to have functional roles in the stress response and nutrient sensing (65). While it would be important to investigate the transcription of these genes in *Chlamydia suis*, these studies reveal their presence and potential importance, emphasizing the need for continual evaluation of genomic content and locus tag numbering system.

Interestingly, individual *Chlamydia* can display diverse but limited ranges of mammalian hosts. Livestock pathogens like *C. pecorum,* which cause chlamydiosis in a variety of animals particularly ruminants and swine, and is notably the leading cause of infectious disease in koalas and *C. abortus,* which also infects ruminants and has been associated with spontaneous abortions in swine and sheep, are among the few able to infect and cause disease in a multiple ruminants species, while the majority of *Chlamydia*, including *C. suis*, *C. trachomatis* and *C. muridarum* appear to be largely restricted to a single host (43, 66). Comparative genomic studies, like this report, will continue to identify molecular candidates potentially involved in restricting host range and enable direct genetic analyses to investigate the role of these host specific factors.

As with a full dN/dS analysis, comparing of amino acid similarity in primary sequences, allows for a direct analysis of evolutionary relationships between species. In this study, approximately 10-20% of the *C. suis* genome appears to have increased evolutionary pressure, likely stemming from the differences in host species and ability to interact with cognate host factors. Further study of this subset of genes could lead to insights in chlamydia-host interactions and chlamydial tissue

tropisms. Even in the more distantly related species, like *C. pecorum* which has the fewest highly conserved genes (over 90%) with *C. suis,* 88% of genes appear to be conserved to some degree.

Several distinct aspects of *C. suis* emerge, however, likely due to adaptations to the porcine host. These include the those in the plasticity zone: a tryptophan biosynthesis operon absent in *C. muridarum*, cytotoxins lost in *C. trachomatis* that show variable representation in *C. suis* and a unique subset of phospholipase D-like genes. Complete tryptophan biosynthesis operons are also found in *C. pecorum* and *C. felis.* Tryptophan levels are known to play a key role in inhibiting the pathogenesis of intracellular organisms, including *Chlamydia*. Human (45) and pig (46) innate immune responses are able to limit the available tryptophan through the indoleamine 2,3-dioxygenase (47) response pathway mediated by interferon gamma (IFN-γ). Briefly, IFN-γ, activated by chlamydial infection, will induce IDO to catalyze the breakdown of L-tryptophan into N-formylkynurenine effectively depleting available tryptophan available for the bacteria. Indole producers, like many members of the human vaginal microbiome, provide the necessary input for salvage by the *trp* operon. Importantly, ocular *C. trachomatis* do not encode the *trp* operon, potentially due to the absence of indole-producing microbes. Sherchand et al demonstrate that the *trp* operon can be deleterious to *Chlamydia* in the absence of indole-producers (48). It would follow that perhaps a strong negative selection on the *trp* operon due to a lack of indole-producers in *C. muridarum*-infected tissues would explain the absence of the *trp* operon in *C. muridarum*. There is no current evidence to support an IFN-γ-mediated IDO response in mice which further supports the negative selection of the *trp* operon in mouse-infecting *Chlamydia*. Overall, this pathway could provide a hypothesis as to why there are no incidences of *C. muridarum* infection

in humans, but there have been reports of *C. suis* present in human samples while detection of active human infection or symptomology has not been reported.

Eukaryotic phospholipase D has roles in lipid metabolism and vesicle regulation and this family of proteins has been exploited by pathogens and used to increase virulence (49). While the *in vivo* role of these genes in *Chlamydia* remains relatively unknown, a study done with *C. trachomatis* in HeLa cells provides evidence that the PZ phospholipases (pzPLDs) may be important for inclusion formation and lipid acquisition (50). Other bacterial phospholipase D genes have been characterized, however, and could provide insight into their role in *Chlamydia*. A study in *Neisseria gonorrhea* showed that a phospholipase D homolog acts to increase adherence and invasion to cervical cells by stimulating complement receptor type 3-mediated endocytosis (51). This effect was species specific, as other bacterial phospholipase D proteins were not able to rescue a knock-out mutant. In *Yersinia pestis*, a plasmid-encoded phospholipase D allows the pathogen to survive in its arthropod host further suggesting that these proteins may play a role in host-specificity (52). As noted, *C. suis* contains more PLD genes within its PZ than almost any other chlamydial species. In fact, several species including *C. caviae, C. pneumoniae* and many parachlamydiae are missing any PLD family genes in the plasticity zone but retain ancestral chromosomal PLD outside of the PZ (50). The variability in number and sequence of this operon, and the putative roles in virulence and host specificity from other bacterial species suggests that these enzymes could be providing an essential function in the manipulation of the unique host cell allowing the species to infect and survive within its specific host.

Outside of the plasticity zone, the presence of variable regions in *ompA* suggest a similar role to that in *C. trachomatis*, given that *C. suis* is known to inhabit several distinct tissues within pigs including eyes, gastrointestinal tract, respiratory tract and reproductive tracts. Other membrane proteins, *incs* and *pmps*, have a similar repertoire as found in *C. trachomatis* and *C. muridarum*. Additional studies on the roles of the non-conserved genes may provide insight into the biological role they may play as well as to provide insight into the host environment. One limitation of this study is in the use of few or single representative genomes from each species. To fully investigate the roles for these key families of proteins, a global analysis of *C. suis* strains, as well as other clinical *Chlamydia* isolates, collected from individual tissues or geographic locations, may be necessary. As with the small-scale studies performed by Chahota et al and Seth-Smith et al on *ompA* and the *tet* island, respectively, a larger-scale study could reveal the dynamics between host and gene content (22, 58).

Based on the evidence for *C. suis* evolution with *C. trachomatis* and *C. muridarum* and the presence of several genes present in *C. trachomatis* and absent in *C. muridarum*, *C. suis* may provide a better model for human chlamydial infection than *C. muridarum*. In order to evaluate this, several questions would need to be answered. Primarily, does *C. suis* infect, ascend and cause pathology in a mouse? One study has provided some evidence that *C. suis* may result in a more robust infection in mice than *C. trachomatis* (67), but the absence of direct infection analyses and comparison with *C. muridarum* leaves some uncertainty related to this hypothesis. Comparative analysis of the *C. suis* genome opens the door for further studies into the complexities of the chlamydial genetic repertoire, proteome and host specificity adaptations.

Methods

Whole Genome Sequencing and De Novo Assembly

Using the DNeasy Blood and Tissue Kit (Qiagen), DNA was isolated and purified from *C. suis* strain R19, provided by Dr. Daniel D Rockey (8). Library generation was performed at the University of Kansas Genome Sequencing Core using the NEBNext Ultra II DNA Library kit. Libraries were run on the Illumina Miseq PE100. Over 90 million 151 bp reads were obtained. Reads were trimmed and quality filtered using BBDuk. Subsequent de novo assembly was performed in the Geneious (version 9.1.8) software suite (https://www.geneious.com) using Velvet v1.2.10 with the Velvet optimizer to determine an optimal Kmer (68). When needed, to resolve larger gaps and to verify final assembly, PCR and Sanger sequencing was used. Predicted coverage to the assembled R19 genome was over 1000x. The origin of replication and the first nucleotide was assigned using *C. trachomatis* L2 434/Bu (NC_010287) as a reference and genomes were adjusted accordingly.

Annotation

Annotation was performed using the RAST tool kit (RASTtk) through PATRIC web resources (24-25) and using open reading frame prediction through Geneious. Functional assignments of Geneious-identified open reading frames were verified through BLAST analysis against protein sequences in the NCBI database. Subsystems were determined through PATRIC automatic annotation. Amino acid similarities were generated through PATRIC proteome comparisons with a minimum percent coverage of 30%, a minimum similarity of 10% and an E-value below 1e-5.

Inclusion membrane proteins were identified by manually searching translated open reading frames for two hydrophobic domains. Putative Inc proteins were analyzed for closest ortholog in *C. trachomatis* D/UW-3 and *C. muridarum* Nigg using the phylogenetic analyses methods for single gene phylogenies as described below. In each case, individual Incs were translated and compared against all known incs from all three species. Pmps were identified by manually searching the genome for the GGAI motif and C-terminal phenylalanine. All pmps were compared with those found in *C. trachomatis* D/UW-3 and *C. muridarum* Nigg using the same methods as for the incs (data not shown).

Phylogenetic Analyses

Comparison of genetic content was performed using the following reference genomes: *C. suis* SWA-2 (NZ_LT821323), *C. suis* SWA-14 (NZLT860207), *C. suis* SWA-86 (NZ_LT860209), *C. suis* 2-26b (NZ_LT999997), *C. suis* 3-25b (NZ_LT999998), *C. trachomatis* D/UW-3/CX, *C. trachomatis* L2 434/Bu, *C. trachomatis* A/HAR-13 (NC_007429.1), *C. trachomatis* B/TZ1A828/OT (NC_012687.1), *C. trachomatis* E/Bour (NC_020971.1), *C. trachomatis* F/SW4 (NC_017951.1), *C. trachomatis* G/SotonG1 (NC_020941.1), *C. trachomatis* Ia/SotonIA1 (NC_020970.1), *C. trachomatis* K/SotonK1 (NC_020965.1), *C. muridarum* Nigg (NC_002620), *C. muridarum* MopnTet14 (NZ_ACUJ01000001.3), *C. muridarum* Nigg3 (NZ_CP009760.1) *C. muridarum* Weiss (ACOW01000004.1), *C. pneumoniae* strain Wien 1 (NZ_LN846980), *C. pecorum* strain strain E58 (NC_015408), *C. psittaci* GR9 (NC_018620), *C. abortus* strain GIMC 2006:CabB577 (NZ_CP024084), *C. caviae* GPIC (NC_003361), *C. felis* Fe/C-56 (NC_007899), and *C. avium* 10DC88 (NZ_CP006571). Genomic alignments were performed in Geneious using MAFFT (v7.309) alignment tools and coding sequence alignments were performed with the

PATRIC proteome comparison web tool which uses bidirectional BLASTP with the following parameters: minimum coverage of 30%, minimum identity of 10%, BLAST E-value of 1e-5.

Phylogenetic trees comparing single genes were constructed using the Geneious Tree Builder for a global alignment with free end gaps and a cost matrix set to Blosum65. Distances were obtained from pairwise alignments of all sequence pairs. To assemble the trees, a Jukes-Cantor genetic distance model was selected and UPGMA tree-building was performed. Nucleotide-based alignments and analyses were also performed in all cases and resulting phylogenies showed similar relationships to protein sequences (Data not shown). Whole genome alignments and subsequent phylogenies to determine genomic evolutionary relationships were performed using the progressiveMauve algorithm with a match seed weight of 15 and a minimum LCB score of 30,000. Gaps were aligned using Muscle 3.6 and a phylogenetic tree was determined using the neighbor-joining method with bootstrapping.

References

1. Rogers DG AA, Hunsaker BD. Lung and nasal lesions caused by a swine chlamydial isolate in gnotobiotic pigs. . J Vet Diagn Invest. 1996;8(1):45-55.
2. Sachse K, Grossmann E, Berndt A, Schutt C, Henning K, Theegarten D, et al. Respiratory chlamydial infection based on experimental aerosol challenge of pigs with Chlamydia suis. Comp Immunol Microbiol Infect Dis. 2004;27(1):7-23.
3. Schiller I KR, Weilenmann R, Thoma R, Kaltenboeck B, Heitz P, Pospischil A. Mixed infections with porcine Chlamydia trachomatis/pecorum and infections with ruminant Chlamydia psittaci serovar 1 associated with abortions in swine. Vet Microbiol. 1997;58(2-4):251-60.
4. Rogers DG AA. Conjunctivitis caused by a swine Chlamydia trachomatis-like organism in gnotobiotic pigs. J Vet Diagn Invest. 1999;11(4):341-4.
5. Borel N RN, Di Francesco A, Donati M, Markov J, Masserey Y, Pospischil A. Selection for tetracycline-resistant Chlamydia suis in treated pigs. Vet Microbiol. 2012;156(1-2):143-6.
6. DiFrancesco A DM, Rossi M, Pignanelli S, Shurdhi A, Baldelli R, Cevenini R. Tetracycline-resistant Chlamydia suis isolates in Italy. Vet Rec. 2008;163(8):251-2.
7. Kaltenboeck B, Kousoulas KG, Storz J. Two-step polymerase chain reactions and restriction endonuclease analyses detect and differentiate ompA DNA of Chlamydia spp. J Clin Microbiol. 1992;30(5):1098-104.
8. J. Lenart AAA, D.D Rockey. Growth and Development of Tetracycline-Resistant *Chlamydia suis*. Antimicrob Agents Chemother. 2001;45(8):2198-203.
9. Schautteet K DCE, Miry C, Van Groenweghe F, Delava P, Kalmar I, Vanrompay D. Tetracycline-resistant Chlamydia suis in cases of reproductive failure on Belgian, Cypriote and Israeli pig production farms. J Med Microbiol. 2013;62(Pt 2):331-4.
10. Jeffrey BM, Suchland RJ, Eriksen SG, Sandoz KM, Rockey DD. Genomic and phenotypic characterization of in vitro-generated Chlamydia trachomatis recombinants. BMC Microbiol. 2013;13:142.
11. Suchland RJ, Sandoz KM, Jeffrey BM, Stamm WE, Rockey DD. Horizontal transfer of tetracycline resistance among Chlamydia spp. in vitro. Antimicrob Agents Chemother. 2009;53(11):4604-11.
12. Dean D RJ, Ruettger A, Kandel RP, Sachse K. Zoonotic Chlamydiaceae Species Associated with Trachoma, Nepal. Emerg Infect Dis. 2013;19(12):1948-55.
13. De Puysseleyr L DPK, Braeckman L, Morre SA, Cox E, Vanrompay D. Assessment of *Chlamydia suis* Infection in Pig Farmers. Transbound Emerg Dis. 2017;64(3):826-33.
14. De Puysseleyr K, De Puysseleyr L, Dhondt H, Geens T, Braeckman L, Morre SA, et al. Evaluation of the presence and zoonotic transmission of Chlamydia suis in a pig slaughterhouse. BMC Infect Dis. 2014;14(14):560.
15. Xie G, Bonner CA, Jensen RA. Dynamic diversity of the tryptophan pathway in chlamydiae: reductive evolution and a novel operon for tryptophan recapture. Genome Biol. 2002;3(9):research0051.
16. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res. 2000;28(6):1397-406.

17. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, et al. Comparative genomes of Chlamydia pneumoniae and C. trachomatis. Nat Genet. 1999;21(4):385-9.

18. Girjes AA, Carrick FN, Lavin MF. Remarkable sequence relatedness in the DNA encoding the major outer membrane protein of Chlamydia psittaci (koala type I) and Chlamydia pneumoniae. Gene. 1994;138(1-2):139-42.

19. Donati M H-CH, Humphrys M, Di Paolo M, Di Francesco A, Myers GS. Genome Sequence of Chlamydia suis MD56, Isolated from the Conjunctiva of a Weaned Piglet. Genome Announc. 2014.

20. Joseph SJ, Marti H, Didelot X, Read TD, Dean D. Tetracycline Selective Pressure and Homologous Recombination Shape the Evolution of Chlamydia suis: A Recently Identified Zoonotic Pathogen. Genome Biol Evol. 2016;8(8):2613-23.

21. Marti H, Kim H, Joseph SJ, Dojiri S, Read TD, Dean D. Tet(C) Gene Transfer between Chlamydia suis Strains Occurs by Homologous Recombination after Co-infection: Implications for Spread of Tetracycline-Resistance among Chlamydiaceae. Front Microbiol. 2017;8(156):156.

22. Seth-Smith HM WS, Bachmann N, Marti H, Qi W, Donati M, di Francesco A, Polkinghorne A, Borel N. The Chlamydia suis Genome Exhibits High Levels of Diversity, Plasticity, and Mobile Antibiotic Resistance: Comparative Genomics of a Recent Livestock Cohort Shows Influence of Treatment Regimes. Genome Biol Evol. 2017;9(3):750-60.

23. Andersen AA RD. Resistance to tetracycline and sulfadiazine in swine C. trachomatis isolates. Chlamydial Infections - Proceedings of the Ninth International Symposium on Human Chlamydial Infection. 1998 313-6.

24. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 2014;42(Database issue):D581-91.

25. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, et al. The PATRIC Bioinformatic Resource Center: expanding data and analysis capabilities. Nucleic Acids Res. 2020;48(D1):D606-12.

26. Viana MVC, Sahm A, Góes Neto A, Figueiredo HCP, Wattam AR, Azevedo V. Rapidly evolving changes and gene loss associated with host switching in Corynebacterium pseudotuberculosis. *PLoS One*. 2018;13(11):e0207304. Published 2018 Nov 12. doi:10.1371/journal.pone.0207304.

27. Banerjee K, Sekar P, Krishnan P, et al. Whole genome sequence analysis of NDM-1, CMY-4, and SHV-12 coproducing *Salmonella enterica* serovar Typhimurium isolated from a case of fatal burn wound infection. *Infect Drug Resist*. 2018;11:2491–2495. Published 2018 Nov 28. doi:10.2147/IDR.S167027.

28. Al Dahouk S, Köhler S, Occhialini A, et al. Brucella spp. of amphibians comprise genomically diverse motile strains competent for replication in macrophages and survival in mammalian hosts. *Sci Rep*. 2017;7:44420. Published 2017 Mar 16. doi:10.1038/srep44420.

29. Viana MV, Wattam AR, Govil Batra D, et al. Genome Sequences of Three *Brucella canis* Strains Isolated from Humans and a Dog. *Genome Announc*. 2017;5(8):e01688-16. Published 2017 Feb 23. doi:10.1128/genomeA.01688-16.

30. Thomson NR HM, Carder C, Lennard N, Lockey SJ, Marsh P, Skipp P, O'Connor CD, Goodhead I, Norbertzcak H, Harris B, Ormond D, Rance R, Quail MA, Parkhill J, Stephens

RS, Clarke IN. Chlamydia trachomatis: Genome sequence analysis of lymphogranuloma venereum isolates. Renome Res. 2008;18(1):161-71.

31.   Overbeek R BT, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic Acids Res. 2005;33(17):5691-702.

32.   Overbeek R OR, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014.

33.   Dugan J, Rockey DD, Jones L, Andersen AA. Tetracycline resistance in Chlamydia suis mediated by genomic islands inserted into the chlamydial inv-like gene. *Antimicrob Agents Chemother*. 2004;48(10):3989–3995. doi:10.1128/AAC.48.10.3989-3995.2004.

34.   Dugan J, Andersen AA, Rockey DD. Functional characterization of IScs605, an insertion element carried by tetracycline-resistance Chlamydia suis. Microbiology. 2007;153(Pt 1):71-9.

35.   DD R. Unraveling the basic biology and clinical significance of the chlamydial plasmid. J Exp Med. 2011;208(11):2159-62.

36.   Pudjiatmoko FH, Ochiai Y, Yamaguchi T, Hirai K. Phylogenetic Analysis of the Genus Chlamydia Based on 16s rRNA Gene Sequences. Int J Syst Bacteriol. 1997;47(2):425-31.

37.   Pillonel T, Bertelli C, Salamin N, Greub G. Taxogenomics of the order Chlamydiales. Int J Syst Evol Microbiol. 2015;65(Pt 4):1381-93.

38.   LaBrie SD, Dimond ZE, Harrison KS, et al. Transposon mutagenesis in Chlamydia trachomatis identifies CT339 and a ComEC homolog important for DNA uptake and lateral gene transfer. mBio. 2019;10(4). doi: 10.1128/mBio.01343-19.

39.   Read TD BR, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM Genome Sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res. 2000;28(6):1397-406.

40.   Stephens RS KS, Lammel C, Fan J, Marathe R, Aravind L, Mitchecll W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. Genome Sequence of an Obligate Intracellular Pathogen of Humans: Chlamydia trachomatis. Science. 1998;282(5389):754-9.

41.   Rajaram K, Giebel AM, Toh E, Hu S, Newman JH, Morrison SG, et al. Mutational Analysis of the *Chlamydia muridarum* Plasticity Zone. Infect Immun. 2015;83(7):2870-81.

42.   Sait M LM, Clark EM, Wheelhouse N, Spalding L, Markey B, Magnino S, Lainson FA, Myers GS, Longbottom D. Genome sequencing and comparative analysis of three Chlamydia pecorum strains associated with different pathogenic outcomes. BMC Genomics. 2014.

43.   Jelocnik M BN, Kaltenboeck B, Waugh C, Woolford L, Speight KN, Gillett A, Higgins DP, Flanagan C, Myers GS, Timms P, Polkinghorne A. Genetic diversity in the plasticity

zone and the presence of the chlamydial plasmid differentiates Chlamydia pecorum strains from pigs, sheep, cattle, and koalas. BMC Genomics. 2015;16:893.

44.   Manuela D H-CH, Humphrys M, Di Paolo M, Di Francesco A, Myers GSA. Genome Sequence of Chlamydia suis MD56, Isolated from the Conjunctiva of a Weaned Piglet. Genome Announc. 2014;2(3):e00425-14.

45.   Roshick C, Wood H, Caldwell HD, McClarty G. Comparison of gamma interferon-mediated antichlamydial defense mechanisms in human and mouse cells. Infect Immun. 2006;74(1):225-38.

46.   Meurens F, Summerfield A, Nauwynck H, Saif L, Gerdts V. The pig: a model for human infectious diseases. Trends Microbiol. 2012;20(1):50-7.

47.   Bailey M CZ, Lewis MC. The evolutionary basis for differences between the immune systems of man, mouse, pig and ruminants. Vet Immunol Immunopathol. 2013;152(1-2):13-9.

48.   Sherchand SP, Aiyar A. Ammonia generation by tryptophan synthase drives a key genetic difference between genital and ocular Chlamydia trachomatis isolates. Proc Natl Acad Sci U S A. 2019.

49.   Selvy PE LR, Lindsley CW, Brown HA. Phospholipase D - enzymology, functionality, and chemical modulation. Chem Rev. 2011;111(10):6064-119.

50.   Nelson DE CD, Taylor LD, Dorward DW, Goheen MM, Caldwell HD. Inhibition of Chlamydiae by Primary Alcohols Correlates with the Strain-Specific Complement of Plasticity Zone Phospholipase D Genes. Infect Immun. 2006;74(1):73-80.

51.   Edwards JL ED, Apicella MA. Gonococcal Phospholipase D Modulates the Expression and Function of Complement Receptor 3 in Primary Cervical Epithelial Cells. Infect Immun. 2003;71(11):6381-91.

52.   Hinnebusch BJ RA, Cherepanov P, Dixon JE, Schwan TG, Forsberg A. Role of Yersinia Murine Toxin in Survival of Yersinia pestis in the Midgut of the Flea Vector. Science. 2002;296(5568):733-5.

53.   Caldwell HD, Kromhout J, Schachter J. Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. Infect Immun. 1981;31(3):1161-76.

54.   Ramsey KH SI, Scripsema JH, Denman CJ, Bowlin AK, Myers GA, Rank RG. Strain and Virulence Diversity in the Mouse Pathogen Chlamydia muridarum. Infect Immun. 2009;77(8):3284-93.

55.   Schofl G VA, Litsche K, Sachse K, Saluz HP. Complete Genome Sequences of Four Mammalian Isolates of Chlamydophila psittaci. J Bacteriol. 2011;193(16):4258.

56.   Suchland RJ DZ, Putman TE, Rockey DD. Demonstration of Persistent Infections and Genome Stability by Whole-Genome Sequencing of Repeat-Positive, Same-Serovar Chlamydia trachomatis Collected From the Female Genital Tract. J Infect Dis. 2017;215(11):1657-65.

57.   Nunes A, Borrego MJ, Nunes B, Florindo C, Gomes JP. Evolutionary dynamics of ompA, the gene encoding the Chlamydia trachomatis key antigen. J Bacteriol. 2009;191(23):7182-92.

58.   Chahota R, Ogawa H, Ohya K, Yamaguchi T, Everett KD, Fukushi H. Involvement of multiple Chlamydia suis genotypes in porcine conjunctivitis. Transbound Emerg Dis. 2017.

59.   Almeida F BV, Ferreira R, Borreego MJ, Gomes JP, Mota LJ. Polymorphisms in Inc Proteins and Differential Expression of inc Genes among Chlamydia trachomatis Strains Correlate with Invasiveness and Tropism of Lymphogranuloma Venereum Isolates. J Bacteriol. 2012;194(23):6574-85.

60.   Dehoux P, Flores R, Dauga C, Zhong G, Subtil A. Multi-genome identification and characterization of chlamydiae-specific type III secretion substrates: the Inc proteins. BMC Genomics. 2011;12:109.

61.   Lutter EI, Martens C, Hackstadt T. Evolution and conservation of predicted inclusion membrane proteins in chlamydiae. Comp Funct Genomics. 2012;2012:362104.

62.   Gomes JP NA, Bruno WJ, Borrego MJ, Florindo C, Dean D. 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. 2006;118(1):275-86.

63.   Sharma M, Das M, Diana D, Wedderburn A, Anindya R. Identification of novel open reading frames in the intergenic regions of *Mycobacterium leprae* genome and detection of transcript by qRT-PCR. Microb Pathog. 2018;124:316-321. doi: 10.1016/j.micpath.2018.08.062.

64.   Couso, J., Patraquim, P. Classification and function of small open reading frames. *Nat Rev Mol Cell Biol* **18,** 575–589 (2017). https://doi.org/10.1038/nrm.2017.58.

65.   Khitun A, Ness TJ, Slavoff SA. Small open reading frames and cellular stress responses. *Mol Omics*. 2019;15(2):108–116. doi:10.1039/c8mo00283e.

66.   Borel N, Polkinghorne A, Pospichil A. A review on Chlamydial diseases in animals:Still a challenge for pathologists? Vet Pathol. 2018;55(3):374-390. doi: 10.1177/0300985817751218.

67.   Donati M, Di Paolo M, Favaroni A, Aldini R, Di Francesco A, Ostanello F, et al. A mouse model for Chlamydia suis genital infection. Pathog Dis. 2015;73(1):1-3.

68.   Kearse M MR, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9.

Chapter 3.

Inter-species lateral gene transfer focused on the *Chlamydia* plasticity zone

identifies loci associated with immediate cytotoxicity and inclusion stability

Abstract

Chlamydia muridarum actively grows in murine mucosae and is a representative model of human chlamydial genital tract disease. In contrast, *C. trachomatis* infections in mice are limited and rarely cause disease. The factors that contribute to these differences in host adaptation and specificity remain elusive. Overall genomic similarity leads to challenges in the understanding of these significant differences in tropism. A region of major genetic divergence termed the plasticity zone (PZ) is hypothesized to contribute to the host specificity. To evaluate this hypothesis, lateral gene transfer was used to generate multiple hetero-genomic strains that are predominately *C. trachomatis* but have replaced regions of the PZ with those from *C. muridarum*. *In vitro* analysis of these chimeras revealed *C. trachomatis*-like growth. Similarly, these chimeras infected mice poorly and were not detected in the upper genital tract. Growth-independent cytotoxicity has been ascribed to three large putative cytotoxins (LCT) encoded in the *C. muridarum* PZ. However, analysis of PZ chimeras supported that gene products other than the LCTs are responsible for this phenotype. Growth analysis of associated chimeras also lead to the discovery of an inclusion protein, CTL0402 (CT147) and TC0424 that is critical for the integrity of the inclusion and preventing apoptosis.

Introduction

Chlamydia are unique obligate intracellular pathogens responsible for diverse diseases in many animals including humans. Despite the wide range of hosts infected by *Chlamydia*, each species in the phylum shares several unique properties, including a biphasic developmental cycle (1). Throughout this developmental cycle, *Chlamydia* consistently interacts with the host cell in order to establish their intracellular niche, acquire nutrients, and subvert host responses. These diverse and complex molecular interactions function through chlamydial outer membrane proteins, inclusion membrane proteins, and secreted effectors (1, 2). Because the chlamydial developmental cycle is intrinsically tied to the host, many of these chlamydial proteins play key roles in pathogenesis. Chlamydial lineages have uniquely adapted to specific hosts or tissues and the unique molecular and cellular properties therein (2). While some species-specific components and interactions have been described (3-7), the vast majority of the chlamydial and cognate host factors contributing to host specificity remain elusive.

A well-established example of chlamydial host adaptation and specificity is provided by *C. trachomatis* and *C. muridarum* infection capabilities. In human genital tract infections, *C. trachomatis* colonizes the cervicovaginal epithelium of the lower genital tract and can ascend into the uterus and fallopian tubes of the upper genital tract (8). Once in the fallopian tubes, *Chlamydia* stimulates an immune response which leads to inflammation and eventual tissue damage. This can lead to a variety of symptomatic, chronic complications, including pelvic inflammatory disease, ectopic pregnancies, and infertility (9, 10). The mouse model of *C. muridarum* infection shares many key aspects of human *C. trachomatis* infection (11). In female mice, *C. muridarum* infects and replicates at relatively high levels in the lower genital tract tissues before ascending into the

uterine horns and ovarian ducts of mice, leading to hydrosalpinx and other pathologies (12-14). In contrast, when immunocompetent mice are infected with *C. trachomatis*, replication in the vaginal vault is limited and infections are quickly cleared. Furthermore, ascension of *C. trachomatis* past the cervix to the upper genital tract is rarely observed in mice (15-17). In contrast, there is no evidence of *C. muridarum* infecting humans, despite its common presence in *Chlamydia*-research laboratories (18, 19). These contrasting infection and pathologic capabilities of *C. trachomatis* and *C. muridarum* highlight the adaption to their specific mammalian hosts.

Despite these differences in host-specific infectivity, *C. trachomatis* and *C. muridarum* share much of their gene content and their genomes are highly syntenous (20, 21). These *Chlamydia* species have condensed genomes comprised of an approximately 1 Mb chromosome and a 7.5 kb plasmid (20, 22). *C. trachomatis* and *C. muridarum* have overall sequence identity of approximately 80% and the species share over 90% of their gene content (20, 23). Within these highly conserved genomes is a singular, highly distinctive region of genetic diversity known as the plasticity zone (PZ) (20, 21). In *C. trachomatis*, the PZ region is ~20 kb and encodes 22 proteins including homologs to MAC/perforin (*macP*), multiple phospholipase D-like genes (PLD), and many proteins of unknown function (hypothetical proteins). Uniquely encoded within the PZ of *C. trachomatis* oculogenital pathobiovars is the tryptophan biosynthesis operon (*trp* operon). These gene products are a key example of species-specific adaptation as they are known to be critical for enabling *C. trachomatis* to survive the human interferon gamma (IFN-γ) cellular response and maintain genital tract infections (21, 24-28). The *C. muridarum* PZ region is almost twice as large (~50 kb) and encodes a larger subset of PLD orthologs, three large cytotoxin homologs (LCT), a *gua* operon for purine interconversion, and a different subset of species-specific hypothetical

proteins (20, 21). The *C. muridarum* genome does not encode a tryptophan operon, reflecting the difference in mouse IFN-γ response (20, 21, 23). Three LCTs are encoded by *C. muridarum,* whereas *C. trachomatis* either does not encode (e.g. LGV strains) or has a single truncated LCT (e.g. serovar D). As such, host-specific phenotypes have been attributed to these genes including the growth-independent, immediate cytotoxicity as well as virulence in the mouse model, although direct functional studies have been challenging (8, 29).

As the PZ contains a substantially higher level of variability between otherwise conserved and syntenous genomes, and the PZ is known to contain putative host-specific genes, this region was hypothesized to be important for host-specific pathogenesis in *Chlamydia*. In order to investigate this hypothesis, *C. trachomatis* and *C. muridarum* chimeric recombinants (30) were generated which have the *C. trachomatis* PZ region replaced with full or partial regions of the *C. muridarum* PZ. These novel chimeras were used to evaluate relevant host-specific and gene-related phenotypes.

Results

Generation of PZ recombinant chimera strains

To evaluate the role of PZ genes in *Chlamydia*, horizontal gene transfer was performed between *C. trachomatis* and *C. muridarum*. The generation of these chimeric species was part of the creation of a recombinant library described previously (30). Briefly, dual-selection was used to cross a tetracycline-resistant *C. trachomatis* parent (31) with a chloramphenicol-resistant *C. muridarum* transposon mutant (*macP*::Tn *cat*) (32) containing an insertion near the PZ. Recombinant progeny resistant to both selectable markers were isolated and whole genome

sequenced. From these efforts, 36 primary recombinant clones were generated with recombination at the PZ. Of these, three clones contain a majority *C. trachomatis* genome with recombination regions largely restricted to the PZ, replacing the *C. trachomatis* PZ with full or partial *C. muridarum* PZ (Figure 3-1 and Table 3-1). One chimera RC768 (RC – <u>R</u>ecombinant <u>Chlamydia</u>) hereafter called the Left Partial PZ, contains *C. muridarum* genes *tc0431* through *tc0439*, replacing the *C. trachomatis* MAC/perforin gene (*ctl0408)*, the phospholipase-D genes (*ctl0409-411*) and species-specific hypotheticals, with the *C. muridarum* MAC/perforin (*tc0431*), phospholipase-D genes (*tc0432-436*) and the three LCTs (*tc0437-439*). However, the Left Partial PZ chimera retains the *C. trachomatis trp* operon (*ctl0420-423*), hypothetical proteins, and the oligopeptide permease 2 (*oppA2*). A second chimera, RC215, called the Full PZ, replaces the entire *C. trachomatis* PZ with the entire *C. muridarum* PZ. A third chimera generated using ofloxacin, rifampicin and tetracycline selection, RC2043, referred to as the Right Partial PZ, contains *C. muridarum* genes *tc0441* through *tc0471* (Figure 3-1 and Table 3-1). For the Full PZ and Right Partial PZ chimeras, the recombination extends approximately 25 kb to the right of the PZ, recombining an additional 23 genes after the *oppA2*. Therefore, the Full PZ mutant contains the entirety of *C. muridarum* genes exchanged into the Left Partial and Right Partial PZ chimeras. Genome sequencing confirms that there are no polymorphisms accumulated in either the *C. trachomatis* background or *C. muridarum* region of recombination in any chimera.
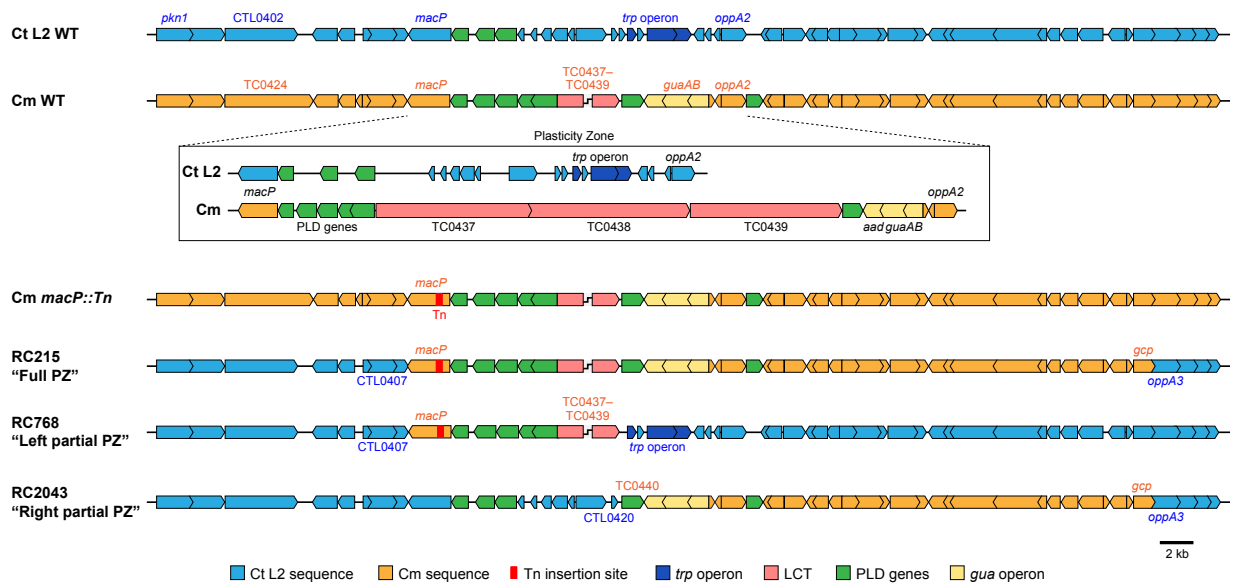
Figure 3-1. Schematic representations of *Chlamydia* and plasticity zone chimeras. Gene organization for the region surrounding the plasticity zone and detailed inset for the plasticity zone of the parental strains *C. trachomatis* (LGV/Bu/434) and *C. muridarum* (strain Nigg). Gene organization and location of the transposon insertion (red) for the *C. muridarum* mutant strain (*Cm macP::Tn)* used to generate *C. trachomatis* chimera RC215 (Full PZ), RC768 (Left Partial PZ), and RC2043 (Right Partial PZ). Genes in blue are *C. trachomatis* and orange genes represent those recombined from *C. muridarum*. Related plasticity zone genes share colors including: *trp* operon (dark blue), LCTs (pink), phospholipase D genes (green) and the *gua* operon (yellow).

Table 3-1. *C. muridarum* Genes in Full Plasticity Zone Chimera Strain

Partial Chimera	Gene	Name	Function	ID/Sim %[A]	E (e-)
L	*tc0431*	*macP*	MAC/perforin family protein	71/82	0
L	*tc0432*	PLD	phospholipase D family protein	32/51[B]	39
L	*tc0433*	PLD	phospholipase D family protein	33/50[B]	37
L	*tc0434*	PLD	phospholipase D family protein	36/53[B]	34
L	*tc0435*	PLD	phospholipase D family protein	29/52[B]	24
L	*tc0436*	PLD	phospholipase D family protein	34/51[B]	31
L	*tc0437*		large cytotoxin (LCT)	unique	
L	*tc0438*		large cytotoxin (LCT)	unique	
L	*tc0439*		large cytotoxin (LCT)	unique	
R	*tc0440*	PLD	phospholipase D family protein	36/54[B]	41
R	*tc0441*		hypothetical protein	unique	
R	*tc0442*	*guaA*	GMP synthase	unique	
R	*tc0443*	*guaB*	inosine-5`-monophosphate dehydrogenase, putative	unique	
R	*tc0444*		hypothetical protein	unique	
R	*tc0445*		hypothetical protein	unique	
R	*tc0446*	*oppA2*	peptide ABC transporter substrate-binding protein	73/86	0
R	*tc0447*	PLD	phospholipase D family protein	36/52[B]	29
R	*tc0448*	*dsbB*	disulfide formation protein	86/92	75
R	*tc0449*	*dsbG*	putative disulfide bond chaperone	89/94	163
R	*tc0450*		conserved hypothetical protein	82/91	0
R	*tc0451*		conserved hypothetical protein	66/77	67
R	*tc0452*		ABC transporter, ATP-binding protein	77/89	134
R	*tc0453*		conserved hypothetical protein	84/92	134
R	*tc0454*	*kdsB*	3-deoxy-manno-octulosonate cytidylyltransferase	87/93	169
R	*tc0455*	*pyrG*	CTP synthase	86/93	0
R	*tc0456*	*ruvX*	Holliday junction resolvase	84/94	92
R	*tc0457*	*zwf*	glucose-6-phosphate 1-dehydrogenase	92/97	0
R	*tc0458*	*devB*	6-phosphogluconolactonase	88/92	172
R	*tc0459*	*dnaX*	DNA polymerase III, tau subunit	82/92	0
R	*tc0460*	*tmk*	thymidylate kinase	81/92	0
R	*tc0461*	*gyrA*	DNA gyrase, subunit A	96/97	0
R	*tc0462*	*gyrB*	DNA gyrase, subunit B	96/97	0
R	*tc0463*		conserved hypothetical protein	80/98	78
R	*tc0464*	*inc*	Inclusion membrane protein	46/62	59
R	*tc0465*	*tgt*	queuine tRNA-ribosyltransferase	91/94	0
R	*tc0466*	*mgtE*	magnesium transporter	99/99	0
R	*tc0468*		conserved hypothetical protein	71/84	0
R	*tc0469*	*inc*	Inclusion membrane protein	57/73	32
R	*tc0470*	*tsaD*	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase	89/94	0

[A] Protein sequences compared to *C. trachomatis* LGV/434/Bu (GenBank AM884176.1)
[B] Protein identity and similarity to the closest PLD homolog in *C. trachomatis* LGV/434/Bu (GenBank AM884176.1)
Grey indicates plasticity zone genes
E is E-value calculated by NCBI Blast (33)

Transcriptional analysis supports expression of *C. muridarum* PZ genes in a *C. trachomatis* background

In order to determine the expression of *C. muridarum* PZ genes in a predominantly *C. trachomatis* background, RNA sequencing was performed at 24 hours post-infection (hpi). Sequencing reads aligning to the *C. muridarum* PZ were present for all genes in the Full PZ mutant and at levels equal to those of *C. muridarum*, with the exception of *tc0452*. Quantitative PCR was performed on *tc0452* and *tc0451* revealing that these genes are transcribed at very low levels, which could contribute to differences in the RNAseq. Using this method, the Full PZ transcript levels for *tc0452* and *tc0451* were not significantly different from the *C. muridarum* parental strain. These results demonstrate that transcription of *C. muridarum* PZ genes still occurred in *C. trachomatis* and at equal levels to parental *C. muridarum*. Analysis of transcript levels for the remainder of the genome revealed limited differential expression (Figure 3-2). Ten genes in the Full PZ exhibited significant differential expression compared to the parental *C. trachomatis* and the levels were two-fold or less for all but one of these genes (Table 3-2). The genes are dispersed throughout the genome with four overexpressed in the Full PZ mutant and six under expressed.
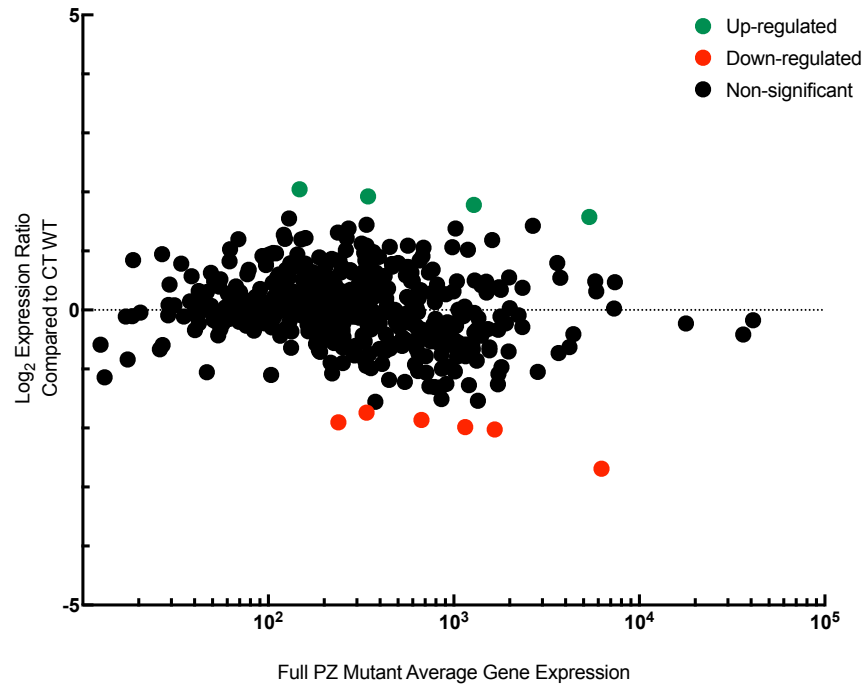
Figure 3-2. MA plot comparison of the Full PZ mutant gene expression to wild-type *C. trachomatis*. A) Only 10 genes were differentially regulated in the Full PZ mutant. These comparisons do not include the genes in the region of recombination with *C. muridarum.* Genes that are significantly up-regulated are highlighted in green; genes that are significantly down-regulated are highlighted in red. Significant regulation was determined by gene expression comparison ratios more than two standard deviations from the gene expression ratio average between two RNAseq runs.

Table 3-2. Genes differentially expressed[a] at 24 hpi in *C. trachomatis* Full PZ chimera

Locus Tag	Gene Name	Function	Differential Expression
CTL0631	*ltuA*	late transcription unit A	1.8
CTL0638	*hyp*	hypothetical	1.6
CTL0682	*hyp*	hypothetical	2.1
CTL0739	*oppC2*	oligonucleotide permease	1.9
CTL0037	*hyp*	hypothetical	-2.0
CTL0079	*pckA*	phosphoenolpyruvate carboxykinase	-1.9
CTL0145	*aas*	2-acylglycerophosphoethenolamine acyltransferase	-1.7
CTL0152	*hyp*	hypothetical	-2.1
CTL0702	*omcB*	membrane protein	-2.7
CTL0842	*copD*	T3SS protein	-1.9

[a]Transcript read counts greater than 2 SD from the average expression ratio

C. muridarum PZ genes are partially involved in *in vitro* growth and developmental cycle regulation

All *Chlamydia* undergo a well described developmental cycle (1, 34), although the timing of key events can vary by species. *C. muridarum* undergoes development within cells that is accelerated by about 6 hours relative to *C. trachomatis* (35). Initial detection and peak production of infectious progeny or elementary bodies (EBs) from *C. muridarum* occurs between 18-24 hpi. In contrast, peak infectious progeny in *C. trachomatis* infection is produced between 24-36 hpi. To investigate whether the PZ recombination affected the *in vitro* developmental cycle, temporal analysis of progeny production was evaluated. Analyzing the timing and levels of infectious progeny produced throughout the developmental cycle is expected to reveal potential defects in infectious and growth characteristics.

Consistent with prior observations, wildtype *C. muridarum* showed peak progeny production from 18-24 hours while *C. trachomatis* progeny peaked after 36 hpi (Figure 3-3A). The *C. muridarum* transposon parent (Cm *macP*::Tn) used to generate the chimeras was also evaluated to ensure the insertion was not disrupting *in vitro* growth. *C. muridarum macP*::Tn was not different from wildtype *C. muridarum* progeny production, suggesting that *macP (tc0431)* was not critical for *in vitro* growth. Each of the PZ chimeras displayed progeny production patterns that closely correlated to *C. trachomatis*. The Left Partial PZ chimeras mirrored wildtype *C. trachomatis* progeny production, while the Full PZ and Right Partial PZ chimeras matched the *C. trachomatis* parental production at 24 hpi but displayed about a log lower progeny production at 36 and 42 hpi.
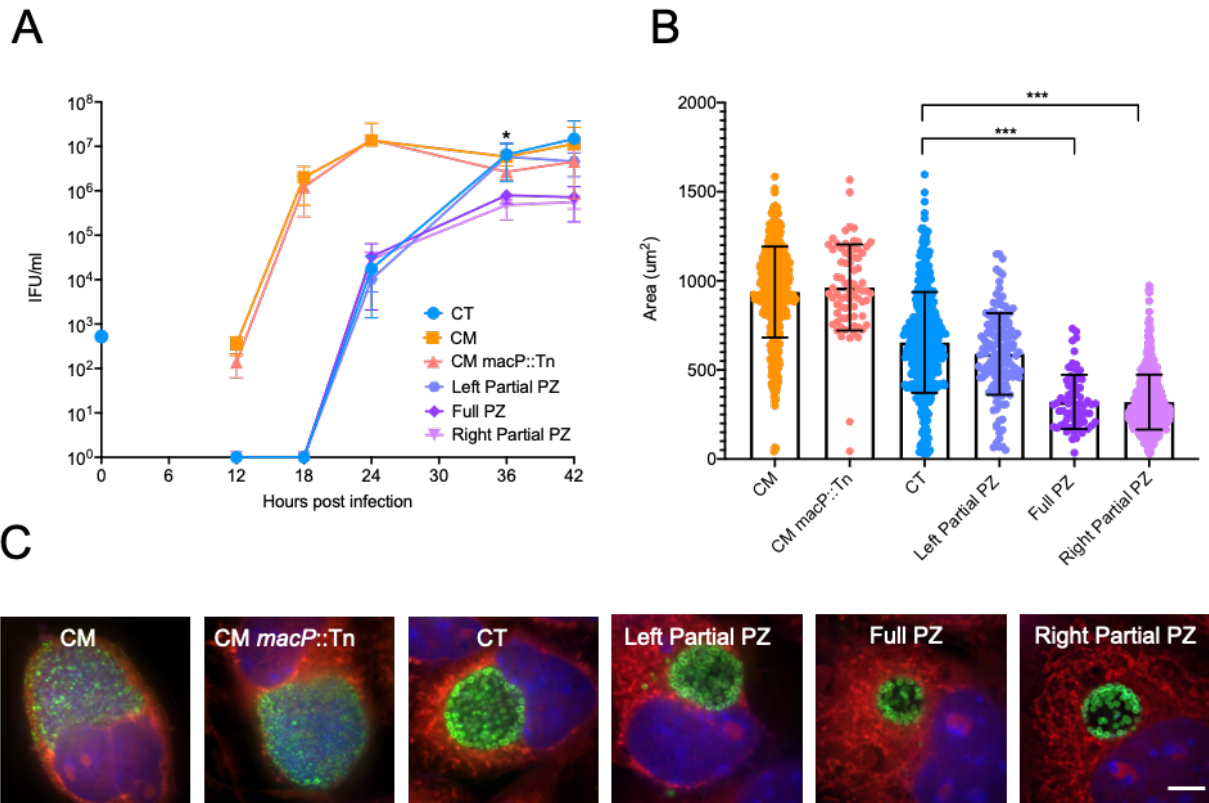
Figure 3-3. Growth and morphology of *Chlamydia* PZ chimeras. A) Progeny production was assessed from 0 to 42 hours post-infection. The blue dot on y-axis represents normalized initial infection concentration for all strains. Statistical analysis was performed by a two-tailed unpaired Student's t-test between samples at each time point. * indicates significant difference (*P*-value <0.05) for Full PZ and Right Partial PZ relative to *C. trachomatis*. B) Inclusion sizes were measured at 20X magnification at 24 hours post-infection (n≥70) and two-tailed unpaired students t-tests performed to compare all data sets (*** indicates *P*-value<0.0001). C) Immunofluorescent confocal microscopy of L929 mouse fibroblast cells infected with respective *Chlamydia* strains 24 hours post-infection. Green (*Chlamydia* MOMP), red (Evans Blue – host cytosol), blue (DAPI – DNA). Representative images of at least 5 fields of view. Scare bar = 5µm.

Inclusion development was visualized for each chimera at 24 hours post infection (Figure 3-3C). As expected at 24 hpi, *C. trachomatis* inclusions were localized next to host cell nuclei and contained both RBs around the inclusion membrane and interior EBs. Because of their faster growth rate, *C. muridarum* and *C. muridarum macP*::Tn inclusions were much larger and filled with EBs at a higher density than *C. trachomatis,* consistent with their faster growth rate. The Left Partial PZ chimera inclusion morphology was similar to *C. trachomatis* (Figure 3-3C), matching

the levels and pattern of progeny production (Figure 3-3A). Interestingly, the Full PZ and Right Partial PZ mutants had inclusions that were smaller and appeared to be less densely packed as *C. trachomatis* (Figure 3-3C). Inclusion area and progeny production are not always correlated (36), and inclusion size may indicate a defect in metabolism of lipid components or hijacking of host vesicles. At 24 hpi, inclusions were imaged and mean area was quantified (Figure 3-3B). *C. muridarum* and *C. muridarum macP*::Tn had significantly larger inclusions than *C. trachomatis*, consistent with their more rapid developmental cycle. The Left Partial PZ chimera had *C. trachomatis*-like inclusion sizes, also matching observations with progeny production (Figure 3-3A) and imaging (Figure 3-3C). In contrast, the Full PZ and the Right Partial PZ inclusions were significantly smaller when compared to both *C. trachomatis* and *C. muridarum*. These observations support that all of the chimeras exhibited growth phenotypes more similar to *C. trachomatis*, with the Left Partial PZ chimera matching those of *C. trachomatis*. These data also support that the 34 kb region of *C. muridarum* which crossed into the Full PZ and Right Partial PZ, but not the Left Partial PZ, may have removed *C. trachomatis* gene products that contribute to optimal growth, or have added *C. muridarum* PZ gene products that disrupt optimal development.

PZ Chimeras exhibit infection capabilities similar to *C. trachomatis* in the mouse model

Ascension from the vaginal vault to the upper genital tract is a critical pathogenic capability and distinguishing difference between *C. trachomatis* and *C. muridarum* in mice. As highlighted, *C. muridarum* reliably ascends in mice whereas *C. trachomatis* rarely is observed in the upper genital tract following vaginal infection. To evaluate the hypothesis that genes encoded in the *C. muridarum* PZ contribute to this capability and reflect species-specific adaptations, PZ chimeras and parental strains were intravaginally infected into mice. After seven days post-infection, both

vaginal vaults and uterine horns were harvested from the mice and bacterial burdens were assessed. *C. muridarum* and *C. muridarum macP*::Tn infected the vaginal vault and the upper genital tract. In contrast, *C. trachomatis* infected the lower genital tract but with a two-log lower bacterial burden than wildtype *C. muridarum* and was not detected in the uterine horns (Figure 3-4A). All PZ chimeras were unable to colonize the uterine horns and had reduced bacterial burdens in the lower genital tract when compared to *C. trachomatis*, although this difference was not statistically significant by an unpaired Student's t-test.
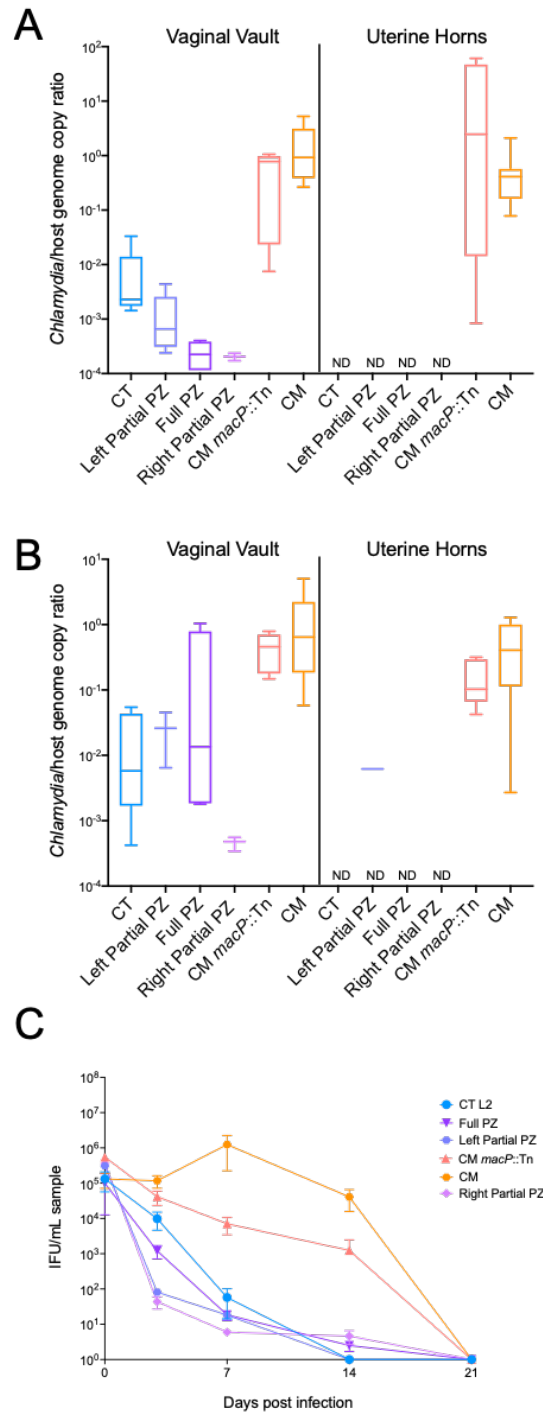
Figure 3-4. Mouse infections with PZ chimeras. C57Bl/6 mice were infected with 10^5 IFU either (A) intravaginally or (B) trans-cervically and lower genital tracts (vaginal vault) and upper genital tracts (uterine horns) were harvested seven days post infection. Bacterial burdens were assessed using ddPCR to compare genome copies of *Chlamydia* to host in each tissue. ND means not detectable when compared to mock infection cut-off levels. C) Bacterial shedding was measured by vaginal swab at 3, 7, 14, 21 and 28 days post- intravaginal infection.

The reduced infection levels may suggest that the absence or presence of PZ genes was a greater burden to the chimeric strains during murine infection. To further investigate this, vaginal shedding was examined (Figure 3-4C). *C. muridarum* shedding was detected until 21 days post-infection, with a peak at 7 days post-infection, while wildtype *C. trachomatis* was quickly cleared with organisms undetectable after 14 days post-infection. In parallel with the bacterial burdens observed in the vaginal vault (Figure 3-4A), all of the PZ chimeras were shed at levels lower than *C. trachomatis*. Again, this was a general trend, though not statistically significant by unpaired Student's t-test. Based on previous studies suggesting that the C57Bl/6 mouse model is more robust against chlamydial infection (37), the same experiment was conducted in the slightly more susceptible C3He/J mouse model (Figure 3-5A), and results mirrored the C57Bl/6 model with the exception of the Right Partial PZ, which was unable to be tested.
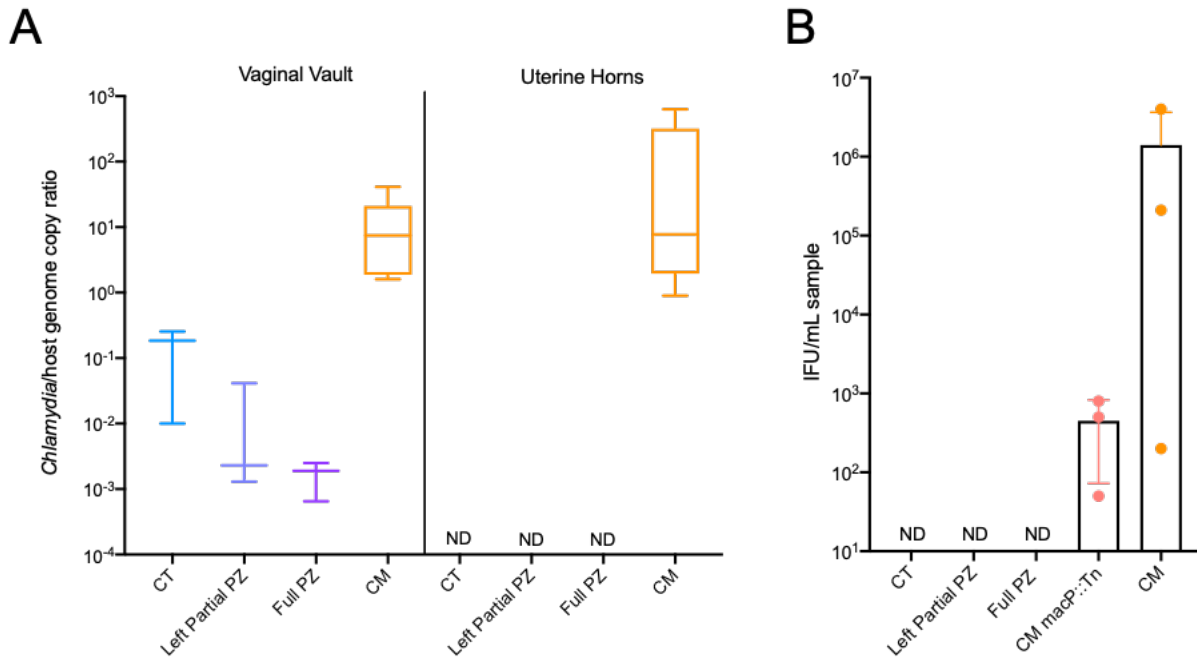
Figure 3-5. Additional mouse models to analyze the virulence and pathogenesis of the PZ chimeras. A) C3He/J mice were infected with 10^5 IFU intravaginally and organs were harvested at day 7. Bacterial burdens were assessed using ddPCR to compare genome copies of *Chlamydia* to host at each tissue (n>3). ND means not detectable when compared to mock infection false positive levels. B) C57Bl/6 mice were infected with 10^5 IFU intrarectally and rectal swabs were collected 7 days post-infection. The Right Partial PZ mutant was not able to be tested in these models

The inability to ascend in the mouse female reproductive tract could indicate an inability to replicate or establish a niche within the endometrial tissues. Therefore, transcervical infection was conducted by infecting the bacteria directly into the uterine horns using an implantation device to bypass the cervix (38). Once again, both vaginal vaults and uterine horns were harvested after seven days post-infection (Figure 3-4B). *C. trachomatis* was able to colonize the uterine horns with approximately two-log fold lower bacterial burdens than *C. muridarum*, similar to what is seen in the lower genital tract in an intravaginal infection (Figure 3-4A). Unexpectedly, both the Left and Full PZ chimeras had higher than *C trachomatis* bacterial burdens in the uterine horns, though this difference was not statistically significant. These data suggest that the cost associated

with the recombined PZ may be specific to colonization of the lower genital tract. Additionally, because *C. trachomatis* is able to colonize the upper reproductive tract, ascension is not merely a question of ability to infect endometrial tissue. Along with the upper reproductive tract, bacterial burdens in the lower reproductive tract were analyzed after transcervical infections. Interestingly, only the wildtype and transposon mutant *C. muridarum* strains were able to disseminate downward and colonize the lower reproductive tract while *C. trachomatis* and the PZ chimeras were not able to descend, which indicates that the bacterial dissemination restriction may occur in both directions and that the cervix may serve as a barrier to ascension.

Recent research by Morrison et al (29), demonstrated that *C. muridarum* mutants with nonsense mutation in LCTs (*tc0438* or *tc0439*) showed attenuation in the gastrointestinal model, but no attenuation in the genital infection model. The Full and Right Partial PZ chimeras contain all three LCTs, providing an opportunity to investigate the role of the LCTs in conferring a gastrointestinal advantage. Mice were infected intrarectally and swabs were collected at peak *C. muridarum* rectal shedding (Figure 3-5B). *C. muridarum* exhibited high bacterial burdens; however, the *C. muridarum macP*::Tn strain showed attenuation in the rectal model with approximately three-log lower detectable IFUs. This attenuation appears to be tissue-specific, as this mutant was not attenuated in the genital tract model (Figure 3-4A and B), suggesting a possible role for *macP* in gastrointestinal infections. Importantly, *C. trachomatis* was not detectable in the rectal swabs or tissue at either day 3 or day 7 post-infection (Figure 3-5B and data not shown). The Full PZ and Left Partial PZ mutants were also not detected in the rectal swabs or tissues. The Right Partial PZ was unable to be tested in this model, but the Full PZ data suggests that the genes included in the Right Partial recombination region are not sufficient for rectal infection. These data support that

the LCTs are not sufficient to confer the ability for *C. trachomatis* to infect rectal tissues, indicating that cognate molecular components present in *C. muridarum* are required for these infections.

Cytotoxin genes (LCT) from *C. muridarum* are insufficient for immediate host-cell toxicity. Another phenotype that is distinct between *C. trachomatis* L2 and *C. muridarum* is the growth-independent and immediate cellular toxicity upon exposure to high multiplicity of infection (MOI) with *C. muridarum* (8). This immediate cytotoxicity phenotype has been associated with the 3 LCTs encoded within the PZ and has been characterized by an early, approximately 4 hpi, breakdown of the host actin cytoskeleton and eventual host cell death (8). Belland et al investigated this phenotype both in *C. muridarum* and in *C. trachomatis* serovar L2, which did not have the cytotoxic phenotype. It was posited that the LCT genes, present in *C. muridarum* but absent in *C. trachomatis,* were responsible for producing this phenotype. The chimeras containing the three LCTs (Full and Left Partial) compared to the chimera without LCTs (Right Partial), along with parental *C. trachomatis* and *C. muridarum*, provided an opportunity to further investigate this phenotype and the contribution of the LCTs.

Figure 3-6. Cytotoxicity analysis of PZ Chimeras. A) Oregon phalloidin staining of host cells to determine cytotoxicity. Representative images showing host cells infected with *Chlamydia* at an MOI of 100 at four hours post-infection. Host nuclei stained with DAPI (blue) which host actin is stained with OregonGreen Phalloidin (green). B) Strains tested for cytotoxicity and associated graphic reflect genetic regions and content. *C. muridarum* (Cm) genes that are present in Full, Left Partial, and Right Partial chimera are indicated as well as the number of intact LCT cytotoxins. Genes in blue are *C. trachomatis* and orange genes represent those recombined from *C. muridarum*. Related plasticity zone genes share colors including: *trp* operon (dark blue), LCTs (pink), phospholipase D genes (green) and the *gua* operon (yellow). Cytotoxicity was rated semi-qualitatively where 3+ indicates that 100% of cells were affected.

Each PZ chimeric strain, along with the wildtype parent strains, were infected onto a cellular monolayer at a high MOI (~100) as was performed by Belland et al (8). After 4 hours of incubation, actin arrangement was analyzed by immunofluorescence microscopy (Figure 3-6A). Cells infected with *C. trachomatis* L2 had an elongated fibroblast shape with actin filaments visible within the cytosol and concentrated around the membrane, which were structurally not discernible from uninfected cells. In contrast, high MOI *C. muridarum* infected cells lost their fibroblastic shape and instead had a distinct cell rounding with puncta of concentrated actin in the center of the cell, indicating a breakdown of the actin filaments. When the PZ chimeras were analyzed for this phenotype, the Full PZ chimera exhibited cytotoxicity. Interestingly, the Left Partial PZ chimera was not cytotoxic despite containing all three LCTs. To verify the expression of the LCTs in the Left Partial PZ, ddPCR was used to measure transcript levels in the Left Partial PZ chimera. At 24 hpi, transcripts were detected for all three LCTs with levels similar to wildtype *C. muridarum* and the Full PZ chimera (data not shown). This suggested that the LCTs were not sufficient for this phenotype, and there may be genes to the right of the LCT within the Full PZ chimera that may play an important role. Therefore, the Right Partial PZ chimera, which contains *C. muridarum* genes to the right of the LCTs, but not including them (Figure 3-1), was investigated as well. Importantly, this chimera was also cytotoxic, supporting that gene products in the Right Partial PZ were critical to the cytotoxic phenotype and that the LCTs may not be associated with this phenotype.

As an additional measure, a chimera with a transposon insertion in the middle LCT gene (*tc0438*) was analyzed for the cytotoxicity phenotype (Figure 3-6B). This chimera was cytotoxic to the same level as wildtype *C. muridarum*, further indicating that disruption of this single LCT was not

sufficient to reduce cytotoxicity, as was consistent with prior observations (20). Overall, these results support that the LCT were not responsible for the cytotoxicity of *C. muridarum* and that one or more of the 30 genes within the right end of the PZ chimeras were contributing to this phenotype (Table 3-1). Based upon gene products that are more unique to *C. muridarum*, candidate genes within this region include two phospholipases (*tc0440* and *tc0447*), two inclusion membrane proteins (*tc0464* and *tc0469*), and three hypothetical proteins specific to *C. muridarum* (*tc0444*, *tc0445* and *tc0467*). While it is less likely that highly conserved bacterial proteins, particularly those involved in known pathways like RuvX (*tc0456*, Table 3-1), are responsible for this phenotype, this possibility cannot be excluded.

CT147 is an inclusion membrane protein that plays a role in preventing host cell death. During the generation of PZ chimeras, a strain termed RC826 was difficult to propagate. Fluorescent imaging of this strain at 26 hpi revealed that the integrity of the inclusion appeared to be compromised early in the development cycle and dispersal of *Chlamydia* within the host cell was noticeable due to early lysis (Figure 3-7B). In order to determine which gene(s) may be responsible for this phenotype, additional recombination was performed and certain sequenced chimeras were selected for investigation of this phenotype. Most importantly, RC1323 was generated, which has an identical genotype to RC826 except for a single gene difference: RC1323 retains the *C. trachomatis* gene *ctl0402* and RC826 has incorporated the *C. muridarum* ortholog *tc0424* (Figure 3-7A). RC1323 displayed inclusion morphology matching that of parental *C. trachomatis* (Figure 3-7B). Additional chimeras (Figure 3-8) were analyzed and recombinant strains that contained the native *ctl0402* appeared to have wildtype growth, while those with a recombined *tc0424,* exhibited a disrupted developmental cycle with early lysis (data not shown).

To confirm that the replacement of *ctl0402* with *tc0424* was responsible for the abnormal growth, a backcross was performed to replace *tc0424* in RC826 with the native *ctl0402*. The resulting chimera, RC3254, was able to complete the developmental cycle and did not cause early inclusion lysis and dispersal of EBs (Figure 3-7B).
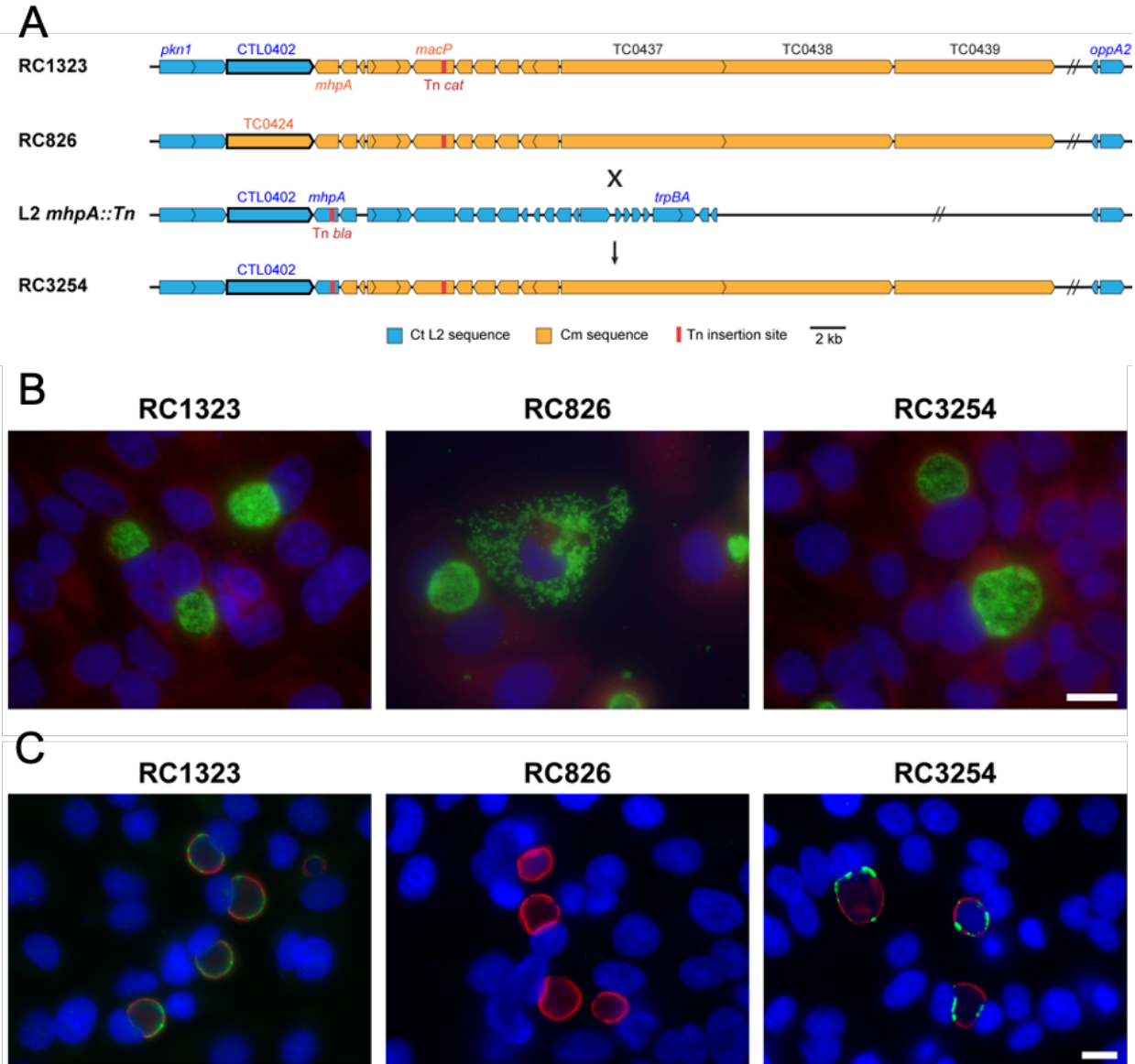
Figure 3-7. Microscopy of early inclusion lysis. A) Representative schematic showing the recombination that occurred to generate the backcross, RC3254, which rescues the early inclusion rupture phenotype. B) Microscopy images showing early lysis of the inclusion membrane and chlamydial dispersal (green) within the host cell at 26 hours post-infection. Host nuclei are stained with DAPI (blue) and host cytosol is stained with Evan's Blue (red). C) Infected samples fixed and stained at 18 hpi with CTL0402 antibody labelling (green) shows that the protein is recruited and inserted into the inclusion membrane, stained with IncG antibody (red). In the tc0424-containing recombinant, RC826, CTL0402 was not labelled, confirming the absence of this gene.
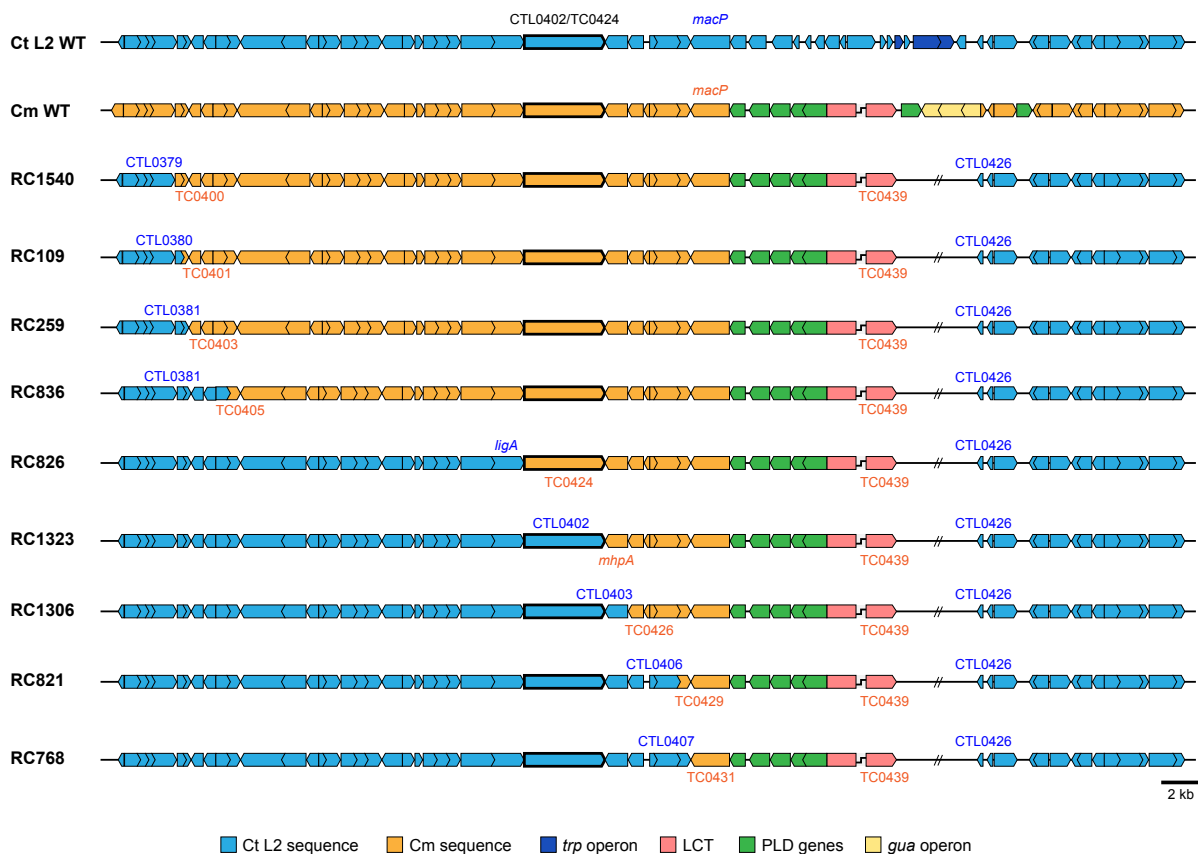
Figure 3-8. Schematic representations of the extended left end chimeras. Gene maps for the parental strains; wildtype *C. muridarum,* and wildtype *C. trachomatis* L2/tet, and the chimeras, RC#. Genes in blue are from *C. trachomatis* while orange genes represent those recombined from *C. muridarum*. Related plasticity zone genes share colors including: trp operon (dark blue), cytotoxins (pink), phospholipase D genes (green) and the gua operon (yellow). Genes with bold outlines correspond to the *CTL0402* or *TC0424* orthologs.

Both CTL0402 and TC0424 are putative inclusion membrane proteins, as they contain the characteristic bilobed hydrophobic domains (39). The presence of CTL0402 on the inclusion membrane in the chimeras was verified with fluorescent labeling (Figure 3-7). RC826 lacked labeling for CTL0402, while TC0424 was present around the inclusion membrane in RC1323 and RC3254, which was consistent with genome sequencing. The early lysis of the inclusion in strains with TC0424, rather than CTL0402, could be consistent with improper inclusion maintenance, a

possible function of these inclusion membrane proteins. To identify what was occurring mechanistically and temporally to cause the early lysis of the inclusions, live time-lapse microscopy was performed on host cells infected with recombinant strains as well as *C. muridarum* and *C. trachomatis* L2 controls. Inclusions appear to develop normally in *C. trachomatis*, *C. muridarum* and RC1323-infected cells. However, RC826-infected cells displayed a membrane-blebbing phenotype characteristic of apoptosis occurring as early as 16 hours post-infection, and the majority of infected cells exhibited this phenotype by 22 hours post-infection (Figure 3-9A). In wildtype *C. muridarum*-infected cells, occasional apoptosis is known, though this typically does not occur until around 20-22 hours.

Apoptosis can be assessed by examination of a variety of cellular markers, including the presence of phosphatidylserine in the outer leaflet of the plasma membrane and the detection of activated effector caspases-3/7 (40). To determine if the membrane-blebbing phenotype observed in the time lapse microscopy was apoptosis, infected host cells were labeled with Annexin V, to detect phosphatidylserine in the outer membrane and FLICA labeling of the activated caspases 3/7. In both experiments, apoptosis markers were detected in a significantly higher percentage of cells when infected with RC826 over RC1323 or wildtype *C. trachomatis* (Figure 3-9 B and C). Together, these data support that *C. muridarum* TC0424 expression in *C. trachomatis* as a replacement for CTL0402 is responsible for the early lysis phenotype which is also associated with the induction of cellular apoptosis.
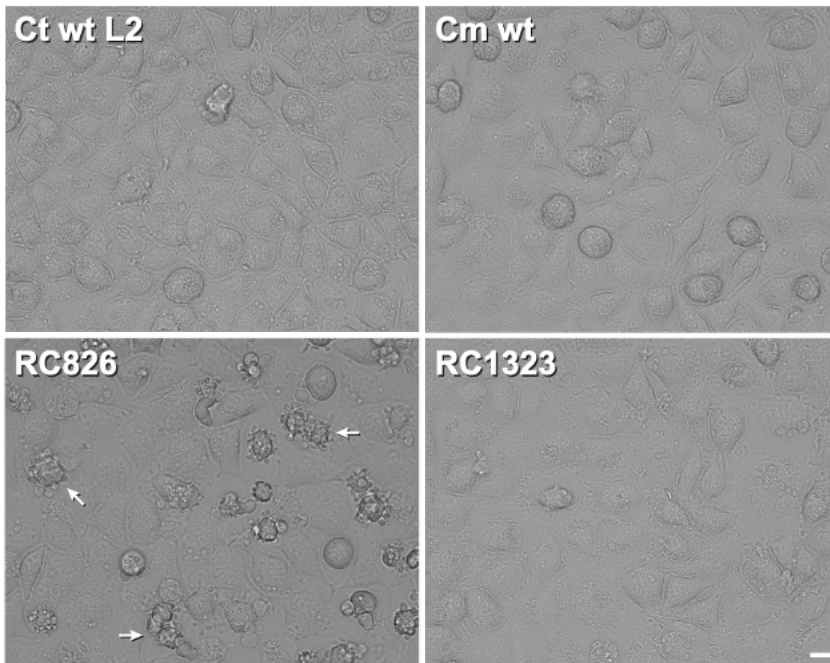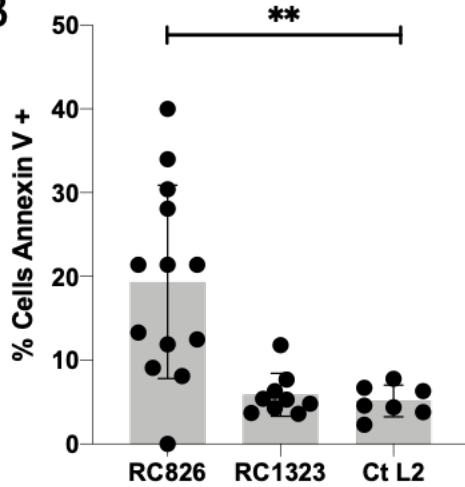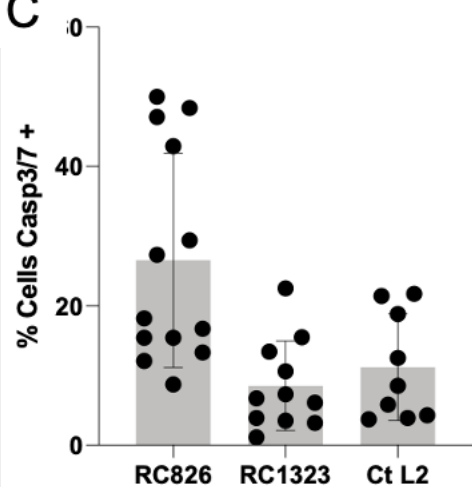
Figure 3-9. Qualitative and quantitative assays demonstrating host cell apoptosis of cells infected with RC826. A) Images taken from live infection videos (Supplementary Figure 6) at 16 hours post infection. Arrows point to cells undergoing apoptosis. B) Host cells infected with wildtype *C. trachomatis* or chimeric strains were stained with Annexin V AlexaFluor 594. Percent cells positive for external plasma membrane annexin V staining in each sample compared with an uninfected control was plotted. Significance was calculated by one-way ANOVA (adjusted p-value = 0.0006). C) Host cells infected as in B) and stained at 28 hours post-infection with Image-iT live caspase 3/7 kit and plotted as percent stained compared with uninfected controls.

Discussion

In this study, interspecies recombination was used to evaluate the role of the PZ in *C. trachomatis* and *C. muridarum*. Key recombinants around the plasticity zone were generated which resulted in the first *C. trachomatis* strain that contains the entire PZ from *C. muridarum*. Chimeras were also generated that contained Left and Right partial PZ regions. These chimeras allowed for the collective assessment of this multi-gene locus and contributions to various phenotypes, including those associated with mammalian infection. Among the primary observations was the limited ability of the PZ locus from *C. muridarum* to solely confer enhanced infection capabilities for *C. trachomatis* in mice. The absence of detectable enhancement in infection capabilities is not overly surprising as these processes are complex and, very likely, incorporate a suite of diverse molecular interactions. However, it would have been a powerful observation relative to the importance of PZ contents had enhanced infection phenotypes been observed. Another major observation was that of the various chimeras with inducing immediate cytotoxicity that support that gene products, not including the LCTs, are likely responsible for this phenotype. Additionally, an inclusion membrane protein (*ctl0402 - tc0424*) that displayed a species-specific incompatibility lead to loss of inclusion membrane integrity and cellular apoptosis was discovered. Overall, this interspecies LGT approach focused on a single locus has enabled multiple discoveries that are expected to serve as an example for future efforts as well as studies on specific gene contributions and mechanisms associated with these key phenotypes.

A central hypothesis for this study was that the *C. muridarum* plasticity zone contains gene products that contribute to host-specificity which could provide enhanced infection capabilities to *C. trachomatis* in lower and upper reproductive tissues of mice. Despite the presence of *C.*

muridarum PZ genes, these chimeras displayed very poor replication in the vaginal vault and no ascension into the upper genital tract similar to *C. muridarum* in this mouse model (Figure 3-4A). These observations should not be interpreted as PZ genes are not important for host specificity or infection processes, merely that this locus alone was unable to sufficiently complement *C. trachomatis* and enable a detectable 'gain of infection function' phenotype in murine vaginal infection. This may indicate that PZ gene products require contributions from specific *C. muridarum* gene products encoded elsewhere in the genome for vaginal infection in mice. Interestingly, the Full PZ chimera strain did trend toward higher bacterial burdens relative to *C. trachomatis* when upper reproductive tracts were infected directly (Figure 3-4B) suggesting that contents of the PZ may play a more critical role in uterine horn tissues or perhaps that the adverse effects of containing these *C. muridarum* genes are lessened in the upper reproductive tract. One explanation should include the difference in environment between these sites. Generally, the lower reproductive tract is lined with stratified, squamous epithelial cells which forms a more protective barrier than the columnar epithelium of the upper reproductive tissues (41). Additionally, there are differences in the immune cells of the upper and lower tracts; particularly, the vagina-facing ectocervix contain dendritic cells, which are not found in the upper reproductive tract. The endocervix, which is part of the upper reproductive tract, contains mucus-producing glands to prevent microbial pathogen access to the uterus (42).

Along these lines, prior studies have evaluated *C. muridarum* mutant strains that contain gene disruptions in certain PZ genes (21, 29), revealing no effect on vaginal infections and providing support for a less critical role for these gene products in the lower genital tract environment. Interestingly, disruptions to LCT genes have rendered strains deficient in gastrointestinal

infections providing further support that certain PZ genes may be critical in specific tissue environments (29). As multiple PZ gene products may contribute collectively to successful lower and upper genital tract infections, possibly in combination with other *C. muridarum* gene products, it will be important to evaluate *C. muridarum* strains with a disrupted PZ locus. Based upon the advances described herein, the application of LGT with selected *C. trachomatis* transposon clones (32, 36) are expected to enable the generation of reciprocal PZ chimeras and assessment of these gene products in the context of *C. muridarum*.

One of the more unexpected observations was the apparent discordance between immediate cytotoxicity and the associated PZ chimera genetic content. As highlighted, immediate cytotoxicity has been established with *C. muridarum*, although some strains of *C. trachomatis* (excluding LGV) have displayed this activity to a substantially reduced degree (8). This immediate cytotoxicity has been hypothesized to be correlated with three large LCT gene products with domains similar to *Clostridium difficile* A/B toxin. The combinatorial analysis of shared and unique loci revealed that the Left Partial PZ chimera, which encode three LCTs, did not exhibit immediate cytotoxicity (Fig. 3-6). The absence of cytotoxicity activity in the Left Partial PZ chimera is unlikely to be due to expression issues as equal LCT transcription was confirmed by the Left Partial and parental *C. muridarum*. Moreover, the Right Partial PZ chimera that does not encode the LCTs did induce cytotoxicity similar to that of parental *C. muridarum* and the Full PZ chimera.

 These observations call into question the genes that contribute to immediate cytotoxicity. There have been few direct analyses on LCTs and cytotoxicity to support causation, largely due to difficulty in cloning these genes and proteins. The most compelling observation in support of LCT

and cytotoxicity was from a recent study on *C. muridarum* strains that contain non-sense mutations in each of the LCTs (21). This study reported no loss of cell rounding for any of the LCT mutant strains, however two mutant strains did exhibit a decrease in lactate dehydrogenase (LDH) release. As these mutant strains had other mutations, contributions by other mutations to this loss of LDH release must be considered. Additional important studies on cytotoxicity include those by Thalmann et al. and Bothe et al. which focus on *C. trachomatis* serovar D protein *ct166,* which contains similarity to the *C. muridarum* cytotoxins (43) (44). *C. trachomatis* D is known to exhibit a version of cytotoxicity and these studies demonstrated that the cytotoxicity was attributed to actin rearrangement and was dependent on glucosylation of Rac1 (43). One critical aspect of this study was the use of CT166-expressing HeLa cells which demonstrated that CT166 was sufficient to cause reduced cell proliferation. Our study directly contradicts these findings, in that we show that the cytotoxins do not appear to be necessary for cytotoxicity in *C. muridarum*. A potential explanation for this difference is that *C. muridarum* cytotoxicity is mechanistically distinct from the cytotoxicity induced by CT166. Additional studies will be required to compare the mechanism of actin re-organization, determine the role of Rho GTPases in *C. muridarum* cytotoxicity and to define the genetic components involved in this process.

There are 30 genes encoded in the Right Partial PZ region that could contribute to the observed immediate cytotoxicity (Table 3-1). The vast majority of these genes are highly conserved with *C. trachomatis* and are less likely to contribute to this phenotype; however, there are some unique and less conserved genes to consider. There are three hypothetical proteins that are unique to *C. muridarum* (*tc0441*, *tc0444*, and *tc0445*) and two phospholipase D proteins (*tc0440* and *tc0447*) that share close to 50% similarity with their *C. trachomatis* orthologs. Phospholipase D proteins

typically hydrolyze phosphatidylcholine and phosphatidylethanolamine to generate phosphatidic acid, which can serve as a signaling molecule (45). They can also serve as secreted effector proteins, cause cytotoxicity, and are documented virulence factors in other bacteria (46). There are two inclusion membrane proteins also encoded in this region (*tc0464* and *tc0469*) which are also expected to be secreted effector proteins. Future studies focused on disruptions to these gene candidates, as well as obtaining and investigating Right Partial PZ chimera sub-clones that contain fewer genes, will be instrumental in determining more directly the genetic contributions to immediate cytotoxicity.

A powerful demonstration of the utility of chlamydial LGT studies, as well as significant biologic discovery, were the observations related to a single gene (*ctl0402*/*tc0424*) and early lysing of *Chlamydia* inclusions (Figure 3-7). The initial observation was of a chimera, RC826, that was difficult to propagate and genomic comparison to similar chimeras that did not exhibit this growth defect (Figure 3-8). The comparison to RC1323 enabled a single gene to be identified in association with the early lysis phenotype and genetic complementation was achieved using a Tn mutant strain (L2 *mhpA*::Tn) and backcrossing with RC826. This gene product is an inclusion membrane protein and these observations support that TC0424 in a *C. trachomatis* background leads to the integrity of the inclusion being compromised and lysing relatively early in the developmental cycle. This uncontrolled spilling of the inclusion contents is expected to induce apoptosis that was observed by time lapse microscopy and membrane markers (Fig 3-9).

These Inc orthologs share 59% identity and 76% similarity with almost complete sequence coverage and limited gaps (Table 3-1). As such, it is expected that the differences that lead to the

incompatibility are subtle molecular aspects that likely affect stable protein-protein interactions. These interactions could be with other inclusion membrane proteins likely in a complex, as is seen with many inclusion membrane proteins (47). This inability to form a stable heterocomplex could disrupt host cell interactions that lead to loss of inclusion integrity. Additionally, interacting partners could be host proteins, including those involved in apoptotic signaling pathways, as is the case for *Chlamydia pneumoniae* inclusion membrane protein Cpn1027 (48). This inclusion membrane protein (CT147) has been investigated in *C. trachomatis* serovar D and was annotated as early endosomal antigen-1 (49). It was transcribed very early (1 hpi) with protein detected by 8 hpi and eventually localized to the inclusion membrane. Protein partners have not been identified and future studies investigating this hypothesis will require these efforts. Additionally, it may be expected that investigations of other chimeras with greater genomic representation could reveal strains that phenocopy RC826 and lead to protein partner candidates. This direction seems promising given the success of these efforts to identify a single gene associated with inclusion maintenance process.

There have been other reported cases of inclusion membrane protein disruption leading to apoptosis of the host cell (50, 51). In these studies, random chemical (*C. muridarum)* or TargeTron directed (*C. trachomatis*) mutations in inclusion membrane proteins resulted in inclusion lysis and apoptosis. In the study performed in *C. muridarum*, inclusion lysis of a TC0574 mutant was mediated by interferon gamma (51). In contrast, the observations in this study were performed in cells untreated with interferon, suggesting that the mechanism of action for the TC0424/CTL0402 inclusion lysis is distinct from those observed in TC0574. In *C. trachomatis*, TargeTron mutations in *ct229*, *incC,* and *ct383* each resulted in inclusion membrane lysis, similar to what was observed

in this work (50). In those mutants, it was demonstrated that premature inclusion lysis led to host cell apoptosis. Similar correlation is expected here, as well although additional studies can be performed to determine the timing of these events, along with the mechanism for premature lysis. From the Weber et al. study, it appears that premature inclusion lysis was not caused by the same mechanism in all three mutants investigated (50), and it will become important to define these mechanisms in the TC0424/CTL0402 recombinant in subsequent studies.

Overall, this study demonstrates the utility of combining LGT and selected transposon insertion clones to investigate a multi-gene locus with capabilities to identify a single gene associated with important phenotypes. As this study focused on the PZ locus, the transposon libraries in *C. muridarum* (32) and *C. trachomatis* (36) with complementary selection markers provide for extensive possibilities to expand similar studies to other candidate loci. Similarly, these will enable larger scale screens to discover regions of interest associated with critical cellular processes and virulence-related phenotypes through mouse models. These should also enable the reciprocal exchange of the *C. trachomatis* PZ into a *C. muridarum* background and evaluation of these genetic modifications on vaginal and gastrointestinal infections in mice. Together, this and similar recent studies (30, 31, 52) provide a growing foundation for new genetic tools and approaches within *Chlamydia* to better decipher the basic biology and pathogenesis for this important intracellular pathogen.

Materials and Methods

Chlamydial Strains.

C. trachomatis serovar L2/tet (31), *C. muridarum* (ATCC VR-123), *C. muridarum macP*::Tn (UWCM007) (32), and chimeras were propagated in L929 mouse fibroblast cells (ATCC CCL-1) using RPMI 1640 medium (Invitrogen, Grand Island, NY) supplemented with 5% heat-inactivated fetal bovine serum (FBS) plus 10 µg/ml gentamycin (Fisher Scientific, Pittsburgh, PA) at 37°C and 5% CO_2. All EBs were stored in sucrose-phosphate-glutamic acid (SPG) media at −80°C. Chimeras used in this study (GeneBank accession numbers): RC768 (Left Partial PZ; CP042748), RC215 (Full PZ; CP042772), RC2043 (Right Partial PZ; CP042717), RC826 (CP042747), RC1323 (CP042791) and RC3254.

RNAseq Transcriptional Analysis.

PZ chimeric strains and their parents were grown in L929 cells at a MOI of 5. At 20 hpi, the infections were harvested for RNA using TRIzol (Invitrogen). RNA was purified by phenol/chloroform extraction followed by DNase treatment with TURBO DNase (Invitrogen). Residual DNA contamination was assessed by converting a portion of the RNA into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher), with control reactions containing no reverse transcriptase, and performing PCR on cDNA reactions. Ribosomal RNA was then depleted using NEBNext rRNA Depletion (Human/Mouse/Rat) (New England Biosciences) and RiboMinus Transcriptome Isolation bacterial (Invitrogen) kits. A final purification step was performed using the RNeasy Mini kit (Qiagen). rRNA depletion was assess by running samples on a 2% agarose gel. RNA quality was assessed using Qubit quantification and TapeStation gel analysis. Library prep for RNAseq was performed using NEBNext Standard mRNA library kit without the polyA selection step. Sequencing was performed on the NextSeq550 with single read 50bp read length. Geneious Prime (Version 2019.1.1) was used for the data

analysis. Reads were aligned to *Chlamydia muridarum* (NC002620) and/or *Chlamydia trachomatis* L2/tet9 (CP035484.1). An average of approximately 75 million reads were sequenced for each sample with approximately 3.3 million reads mapping to each *Chlamydia* genome. Transcript levels were normalized to reads per million. RNAseq analysis was done in duplicate with RNA harvested from two independent infections. Significant differential expression ratios were determined using two standard deviations from the average expression ratio for all RNAseq runs (Figure 3-2). To verify RNAseq, cDNA was generated and used in a PCR to confirm presence or absence of genes. Using the above RNA extraction protocol, RNA was extracted from the PZ chimeric strains and wildtype parents at 20 hpi. cDNA was prepared using the Applied Biosystems High Capacity cDNA reverse Transcription Kit using the manufacturer's instructions. PCR was performed to detect transcripts. To quantify transcript levels, ddPCR was performed with total cDNA using EvaGreen Supermix (Bio-rad) protocols (53). Positive droplets were normalized to *secY* levels using dual *C. trachomatis* and *C. muridarum secY* primers (Table 3-3).

Table 3-3. Primers used throughout this work

Gene		Forward		Reverse
rpsR		5'-TGTCCTTTTGTTTCCGCGGG-3'		5'-CTCCTACGAAAGGCAAAAGCCC-3'
Ct *macP*		5'-TGAGCGGCTGTGGAACATGTGACGG-3'		5'-GCATCGCAGCCTCTCATCCACATCC-3'
Cm *macP*		5'-GGAATCGTAAGTAGAGTAGCCAGATTCCG-3'		5'-CGGGGCTAATGCTGCAGAAAAACG-3'
Ct PLD (*CTL0411*)		5'-GCCCCATTTGATAGTGTACGGAGACAGGCGC-3'		5'-AACGAGCCCGATCTTCTTCGAGCCAAGTCG-3'
Cm PLD (*tc0436*)		5'-CGGACACACCTTGTAGATTTGATGGCAAGAGC-3'		5'- GCTCGCAGAACACTACACACCCAGCCC-3'
trpB		5'-GGTGCGGGACAACATGGAGTAGC-3'		5'-CGGGATAAGGTAAAGGCTAAGGCC-3'
guaA		5'-CGGTT-AAGAAATTCTAGATAAAGATGGCCAGC-3'		5'-CCGACACTTTCTGGCTGG-3'
secY		5'-CCACCCAAAAAGTAGCT-3'		5'-GGTAATATAGCTACAACAGC-3'
Cm LCT (*tc0437*)		5'-CCTTTACGGTTTCTCCAGC-3'		5'-GGTTCCGGCAAAGATTATCC-3'
Cm LCT (*tc0438*)		5'-CCTCTCCTGCTTACCATACC-3'		5'-CGCTTCACTTCTAAAGATACG-3'
Cm LCT (*tc0439*)		5'-GCTGGTTGCAAATCTATTCC-3'		5'-CCAAAGTAGTTTCTTCTCGC-3'
tc0451		5'-GAGAGTTTGTTATCTCTGGAAAGCATT-3'		5'-GCAGCTTGCTCAAGTTTTTGAAGAGG-3'
tc0452		5'-CTGTTGGTAGTAAAATATTTTTTCTAACCG-3'		5'-GGCAACAGAGCTTCTTTTTGTTGC-3'

Progeny production assay and inclusion size measurement.

Mouse fibroblast L929 cells were infected (550xg; 30min) with *C. trachomatis* L2 (LGV/434/Bu), *C. muridarum* (ATCC VR-123), the transposon mutant *C. muridarum* (*macP*::Tn *cat*) or one of the three PZ recombinants (MOI = 0.3-0.5). At 0, 12, 18, 24, 36 or 42-hours post infection, cells were washed with 1X Hanks' balanced salt solution (HBSS) chlamydial EBs were released through water lysis (50). In order to determine titers at each time point, samples were infected with serial dilution into 96-well plates and allowed to grow for 24 hours before methanol fixation and subsequent staining. Chlamydial inclusions were stained using the PathoDX Chlamydia Culture Confirmation Kit (ThermoFisher Scientific). In brief, the *Chlamydia* infected cellular monolayer was washed twice with 1X phenol-free HBSS and fixed in 100% methanol for 10 minutes before adding a 1:40 dilution of stain. Total DNA was stained using DAPI. Inclusions were visualized under 10X magnification using the EVOS FL Auto 2 microscope (Thermo Scientific, Waltham, MA) and were enumerated manually and/or on MIPAR software using similar parameters as mentioned below. Final progeny production was normalized to *C. trachomatis* wildtype MOI at 0 hours post-infection and normalized values were plotted as mean and standard deviation followed by two tailed unpaired t-test at 36 hrs to see if the chimeras were significantly different with a p-value<0.05. Inclusion sizes were measured at 24 hpi by capturing images of from the center of the field under 20X magnification using the EVOS FL Auto 2 microscope (Thermo Scientific, Waltham, MA). Quantification of inclusion area was done using MIPAR software (54) with a contrast range between 90 and 255 units, expected inclusion area of 30 μm^2 -1600 μm^2 and a roundness threshold of $\leq$ 0.7. Two tailed unpaired t-test was run between the chlamydial parents and chimeras to see if there was any significant difference in the inclusion sizes with a p-value< 0.0001.

Confocal Microscopy

L929 cells were seeded at 50% confluency 24 hours prior to infection in an 8-well ibiTreat µ-Slide (Ibidi, Martinsried, Germany) and were infected with respective *C. trachomatis, C. muridarum,* or recombinant mutant. At 24 hpi, infected cells were fixed with 100% methanol for 10 minutes at RT. Cells were washed once with HBSS and again with PBS then stained using 180 µl of the PathoDx *Chlamydia* Culture Confirmation Kit (Remel Europe Ltd., Dartford, UK), mouse anti-CT147 (kind gift from Dr. Guangming Zhong, UTHSCSA), or rabbit anti-IncG diluted in PBS 1 hour and 50 minutes RT in the dark or overnight at 4°C. 20µl of 1µM 4', 6-diamidino-2-phenylindole (DAPI) diluted 1:100 in PBS was then added to wells and allowed to stain for 10 minutes, RT in the dark. Stain was then removed, and the cells washed with PBS. A final overlay of Vectashield antifade mounting medium (Burlingame, CA) was added and slides were stored at 4°C in the dark until imaged. Cells were visualized on an Olympus IX81/3I spinning disk confocal inverted microscope at 150X magnification and captured on an Andor Zyla 4.2 sCMOS camera (Belfast, Northern Ireland). Microscope and camera were operated using SlideBook 6 software (Intelligent Imaging Innovations, Denver, USA). Exposure time remained consistent for all fields captured, with exposure for DAPI at 2 seconds, GFP (MOMP/LOS) 3 seconds, and Evans Blue (host cytoplasm) 3 seconds. 5-10 images were taken per strain. 5 Z stack images at 0.35µm apart were taken per field imaged. Images were processed in SlideBook 6 and a No Neighbors Deconvolution with a subtraction constant of 0.4 was applied to all images.

Mouse Infections

Female C57BL/6 mice or C3He/J mice (6 to 8 weeks old) were purchased from Jackson Laboratories and housed in accordance with the requirements specified by the University of

Kansas Institutional Care and Use Committee. Mice were treated subcutaneously with 2.5 mg medroxyprogesterone acetate (Depo-Provera, Pfizer, NY) upon arrival (day −7) to synchronize menstrual cycles. Infectious doses of parental or chimeric *Chlamydia* were diluted in sucrose-phosphate-glutamic acid (SPG) buffer along with an SPG-only mock control dose. Mice were inoculated intravaginally with 5 μl of infectious dose, for a final concentration around 1×10^5 IFU/mouse, by deposition of the dose into the vagina. Seven days post infection for C57BL/6 or five days post-infection for C3He/J, mice were humanely euthanized and the genital tracts including vaginal vault and uterine horns were collected in SPG. Organs were homogenized using a rotor/stator homogenizer (Biospec, Bartlesville, OK). DNA isolation was performed using a DNeasy Blood and Tissue Kit (Qiagen) to the manufacturer's instructions. Isolated DNA was then used to determine bacterial burden by droplet digital PCR (ddPCR). Primers and probes for *Chlamydial rpsR* and murine *rpp30* were used with ddPCR Supermix for Probes (Bio-Rad, Hercules, CA). ddPCRs were performed as previously reported (36) Bacterial burdens of each tissue were analyzed using QuantaSoft Software (Bio-Rad), and the results are reported as $\log_{10}$ ratios of *Chlamydia* DNA to host DNA (*rpsR*/*rpp30* copies). Box and whisker scatter plots were generated in GraphPad Prism 8. For the vaginal shedding, C57BL/6 mice were infected as above. At days 3, 7, 14, 21 and 28 post infection, vaginal swabs were collected by inserting a calgiswab (Puritan Medical Products, ME USA) 5 mm into the vagina and turning five times in each direction. Swabs were then placed in SPG and bead beating was used to release *Chlamydia* from the swab. Titers were measured identically to the progeny production assessment.

For the rectal models, female C57BL/6 were inoculated intrarectally with 5 μl of infectious dose, for a final concentration around 1×10^5 IFU/mouse, by deposition of the dose into the rectum.

Rectal swabs were taken at 3 and 7 days post-infection. After 7 days post-infection, mice were humanely euthanized and GI tracts including rectum, small intestine and spleen were collected in SPG. Rectal swabs were performed as described for vaginal swabs above.

Immediate Cytotoxicity

Methods used by Belland et al. were followed with minor alterations (8). Briefly, McCoy cells were grown to confluency (5.0 X 10^5 cells) on 12mm coverslips in shell vials and infected with purified EB's strains of *C. muridarum*, *C. trachomatis* L2, RC768, RC2043 and RC215 and *C. muridarum* transposon mutants diluted in 0.3 ml of sucrose-phosphate-glutamic acid (SPG) at a standardized MOI of 100 (5.0 X 10^7 infection forming units). Vials were rocked for 2h at 37°C and washed three times with PBS, MEM-10 was added, and the cells were incubated at 37°C for 4h rinsed with PBS then fixed in 3.7% paraformaldehyde in PBS for 10 min at room temperature. Fixed cells were incubated with Oregon Green phalloidin (Life Technologies) according to manufacturer's instructions. Cover slips were then washed with PBS and mounted with Vectashield plus DAPI. Cells were observed by fluorescent microscopy for immediate cytotoxicity. Cytotoxicity was scored semi-quantitatively (0-3+ scoring) by evaluation of cell rounding, detachment, actin rearrangement and lysis compared with uninfected control cells. A 3+ score indicated 100% of cells affected; 2+, 25-75% affected; 1+, <25% affected; 0, same as uninfected cell control.

Apoptotic assays

HeLa and McCoy cells were infected with wild type *C. trachomatis* L2 or chimera strains, and analyzed for phosphatidylserine externalization and caspase 3/7 activation using the Annexin V

Alexa Fluor 594 and Image-iT live caspase 3/7 assay kits (Molecular Probes), respectively, using established procedures (55). Infected cells fixed and stained according to the manufacturer's instructions, and cells were subsequently imaged on a Nikon Eclipse Ti-E inverted microscope equipped with a 60X objective and a cooled CCD camera. Images were acquired and processed using the Volocity software package (Perkin Elmer). Cells positive for external plasma membrane annexin V were scored using a software-based calculation of fluorescence intensity and distribution, along with positive and negative controls to train the software for accuracy.

Live *in vitro* imaging

L929 mouse fibroblast cells were infected in a 96-well plate at an MOI of ~5. At 16 hours post-infection, the infected plate was placed in an EVOS FL Auto 2 microscope with 5% CO2, 37°C, and humidity. One field of view for each strain was imaged by brighfield using a 40X objective every 20 minutes for 12 hours (16hpi-28hpi). Images for each strain were then compiled in ImageJ (NIH, Bethesda, MD) and processed into .mp4 format using Adobe Media Encoder 2019 (Adobe, San Jose CA).

References

1.   Abdelrahman YM, Belland RJ. The chlamydial developmental cycle. FEMS Microbiol Rev. 2005;29(5):949-59.
2.   Nunes A, Gomes JP. Evolution, phylogeny, and molecular epidemiology of Chlamydia. Infect Genet Evol. 2014;23:49-64.
3.   Clifton DR, Dooley CA, Grieshaber SS, Carabeo RA, Fields KA, Hackstadt T. Tyrosine phosphorylation of the chlamydial effector protein Tarp is species specific and not required for recruitment of actin. Infect Immun. 2005;73(7):3860-8.
4.   Gauliard E OS, Rueden KJ, Ladant D. Characterization of interactions between inclusion membrane proteins from Chlamydia trachomatis. Front Cell Infect Microbiol. 2015;5(13).
5.   Gomes JP NA, Bruno WJ, Borrego MJ, Florindo C, Dean D. 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. 2006;118(1):275-86.
6.   Dehoux P, Flores R, Dauga C, Zhong G, Subtil A. Multi-genome identification and characterization of chlamydiae-specific type III secretion substrates: the Inc proteins. BMC Genomics. 2011;12:109.
7.   Damiani MT, Gambarte Tudela J, Capmany A. Targeting eukaryotic Rab proteins: a smart strategy for chlamydial survival and replication. Cell Microbiol. 2014;16(9):1329-38.
8.   Belland RJ, Scidmore MA, Crane DD, Hogan DM, Whitmire W, McClarty G, et al. Chlamydia trachomatis cytotoxicity associated with complete and partial cytotoxin genes. Proc Natl Acad Sci U S A. 2001;98(24):13984-9.
9.   Wiesenfeld H. Screening for Chlamydia trachomatis Infections in Women. New England Journal of Medicine. 2017;376(22):2197-8.
10.   Darville T, Hiltke TJ. Pathogenesis of genital tract disease due to Chlamydia trachomatis. J Infect Dis. 2010;201 Suppl 2:S114-25.
11.   Phillips S, Quigley BL, Timms P. Seventy Years of Chlamydia Vaccine Research - Limitations of the Past and Directions for the Future. Front Microbiol. 2019;10:70.
12.   de la Maza LM, Pal S, Khamesipour A, Peterson EM. Intravaginal inoculation of mice with the Chlamydia trachomatis mouse pneumonitis biovar results in infertility. Infect Immun. 1994;62(5):2094-7.
13.   Shah AA, Schripsema JH, Imtiaz MT, Sigar IM, Kasimos J, Matos PG, et al. Histopathologic changes related to fibrotic oviduct occlusion after genital tract infection of mice with Chlamydia muridarum. Sex Transm Dis. 2005;32(1):49-56.
14.   Campbell J, Huang Y, Liu Y, Schenken R, Arulanandam B, Zhong G. Bioluminescence imaging of Chlamydia muridarum ascending infection in mice. PLoS One. 2014;9(7):e101634.
15.   O'Meara CP, Andrew DW, Beagley KW. The mouse model of Chlamydia genital tract infection: a review of infection, disease, immunity and vaccine development. Curr Mol Med. 2014;14(3):396-421.
16.   De Clercq E, Kalmar I, Vanrompay D. Animal models for studying female genital tract infection with Chlamydia trachomatis. Infect Immun. 2013;81(9):3060-7.
17.   Gondek DC, Olive AJ, Stary G, Starnbach MN. CD4+ T cells are necessary and sufficient to confer protection against Chlamydia trachomatis infection in the murine upper genital tract. J Immunol. 2012;189(5):2441-9.

18.   Pike RM. Laboratory-associated infections: summary and analysis of 3921 cases. Health Lab Sci. 1976;13(2):105-14.

19.   Miller CD, Songer JR, Sullivan JF. A twenty-five year review of laboratory-acquired human infections at the National Animal Disease Center. Am Ind Hyg Assoc J. 1987;48(3):271-5.

20.   Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res. 2000;28(6):1397-406.

21.   Rajaram K, Giebel AM, Toh E, Hu S, Newman JH, Morrison SG, et al. Mutational Analysis of the *Chlamydia muridarum* Plasticity Zone. Infect Immun. 2015;83(7):2870-81.

22.   Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, et al. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science. 1998;282(5389):754-9.

23.   Dimond ZE, Hefty PS. Comprehensive genome analysis and comparisons of the swine pathogen, Chlamydia suis reveals unique ORFs and candidate host-specificity factors. Pathog Dis. 2020.

24.   Fehlner-Gardiner C, Roshick C, Carlson JH, Hughes S, Belland RJ, Caldwell HD, et al. Molecular Basis Defining HumanChlamydia trachomatisTissue Tropism. Journal of Biological Chemistry. 2002;277(30):26893-903.

25.   Caldwell HD, Wood H, Crane D, Bailey R, Jones RB, Mabey D, et al. Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Invest. 2003;111(11):1757-69.

26.   Rottenberg ME, Gigliotti-Rothfuchs A, Wigzell H. The role of IFN-gamma in the outcome of chlamydial infection. Curr Opin Immunol. 2002;14(4):444-51.

27.   Roshick C, Wood H, Caldwell HD, McClarty G. Comparison of gamma interferon-mediated antichlamydial defense mechanisms in human and mouse cells. Infect Immun. 2006;74(1):225-38.

28.   Xie G, Bonner CA, Jensen RA. Dynamic diversity of the tryptophan pathway in chlamydiae: reductive evolution and a novel operon for tryptophan recapture. Genome Biol. 2002;3(9):research0051.

29.   Morrison SG, Giebel AM, Toh EC, Spencer HJ, Nelson DE, Morrison RP. Chlamydia muridarum Genital and Gastrointestinal Infection Tropism Is Mediated by Distinct Chromosomal Factors. Infection and Immunity. 2018;86(7).

30.   Suchland RJ, Carrell SJ, Wang Y, Hybiske K, Kim DB, Dimond ZE, et al. Chromosomal Recombination Targets in Chlamydia Interspecies Lateral Gene Transfer. J Bacteriol. 2019;201(23).

31.   Suchland RJ, Sandoz KM, Jeffrey BM, Stamm WE, Rockey DD. Horizontal Transfer of Tetracycline Resistance among Chlamydia spp. In Vitro. Antimicrob Agents Ch. 2009;53(11):4604-11.

32.   Wang Y, LaBrie SD, Carrell SJ, Suchland RJ, Dimond ZE, Kwong F, et al. Development of Transposon Mutagenesis for Chlamydia muridarum. J Bacteriol. 2019;201(23).

33.   Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, et al. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 2005;272(20):5101-9.

34. Gitsels A, Sanders N, Vanrompay D. Chlamydial Infection From Outside to Inside. Front Microbiol. 2019;10:2329.

35. Messinger JE, Nelton E, Feeney C, Gondek DC. Chlamydia Infection Across Host Species Boundaries Promotes Distinct Sets of Transcribed Anti-Apoptotic Factors. Front Cell Infect Microbiol. 2015;5:96.

36. LaBrie SD, Dimond ZE, Harrison KS, Baid S, Wickstrum J, Suchland RJ, et al. Transposon Mutagenesis in Chlamydia trachomatis Identifies CT339 as a ComEC Homolog Important for DNA Uptake and Lateral Gene Transfer. mBio. 2019;10(4).

37. Chen J, Zhang H, Zhou Z, Yang Z, Ding Y, Zhou Z, et al. Chlamydial induction of hydrosalpinx in 11 strains of mice reveals multiple host mechanisms for preventing upper genital tract pathology. PLoS One. 2014;9(4):e95076.

38. Helble JD, Reinhold-Larsson NV, Starnbach MN. Early Colonization of the Upper Genital Tract by Chlamydia muridarum Is Associated with Enhanced Inflammation Later in Infection. Infect Immun. 2019;87(9).

39. Bannantine JP GR, Viratyosin W, Brown WJ, Rockey DD. A secondary structure motif predictive of proteinlocalization to the chlamydial inclusion membrane. Cell Microbiol. 2000;2(1):35-47.

40. Segawa K, Nagata S. An Apoptotic 'Eat Me' Signal: Phosphatidylserine Exposure. Trends Cell Biol. 2015;25(11):639-50.

41. Lee SK, Kim CJ, Kim DJ, Kang JH. Immune cells in the female reproductive tract. Immune Netw. 2015;15(1):16-26.

42. Trifonova RT, Lieberman J, van Baarle D. Distribution of immune cells in the human cervix and implications for HIV transmission. Am J Reprod Immunol. 2014;71(3):252-64.

43. Thalmann J, Janik K, May M, Sommer K, Ebeling J, Hofmann F, et al. Actin re-organization induced by Chlamydia trachomatis serovar D--evidence for a critical role of the effector protein CT166 targeting Rac. PLoS One. 2010;5(3):e9887.

44. Bothe M, Dutow P, Pich A, Genth H, Klos A. DXD motif-dependent and -independent effects of the chlamydia trachomatis cytotoxin CT166. Toxins (Basel). 2015;7(2):621-37.

45. Selvy PE, Lavieri RR, Lindsley CW, Brown HA. Phospholipase D: enzymology, functionality, and chemical modulation. Chem Rev. 2011;111(10):6064-119.

46. Flores-Diaz M, Monturiol-Gross L, Naylor C, Alape-Giron A, Flieger A. Bacterial Sphingomyelinases and Phospholipases as Virulence Factors. Microbiol Mol Biol Rev. 2016;80(3):597-628.

47. Rockey DD, Scidmore MA, Bannantine JP, Brown WJ. Proteins in the chlamydial inclusion membrane. Microbes and Infection. 2002;4(3):333-40.

48. Flores R, Zhong G. The Chlamydia pneumoniae Inclusion Membrane Protein Cpn1027 Interacts with Host Cell Wnt Signaling Pathway Regulator Cytoplasmic Activation/Proliferation-Associated Protein 2 (Caprin2). PLoS One. 2015;10(5):e0127909.

49. Belland RJ, Zhong G, Crane DD, Hogan D, Sturdevant D, Sharma J, et al. Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proc Natl Acad Sci U S A. 2003;100(14):8478-83.

50. Weber MM, Lam JL, Dooley CA, Noriea NF, Hansen BT, Hoyt FH, et al. Absence of Specific Chlamydia trachomatis Inclusion Membrane Proteins Triggers Premature Inclusion Membrane Lysis and Host Cell Death. Cell Rep. 2017;19(7):1406-17.

51. Giebel AM, Hu S, Rajaram K, Finethy R, Toh E, Brothwell JA, et al. Genetic Screen in Chlamydia muridarum Reveals Role for an Interferon-Induced Host Cell Death Program in Antimicrobial Inclusion Rupture. mBio. 2019;10(2).

52. Jeffrey BM, Suchland RJ, Eriksen SG, Sandoz KM, Rockey DD. Genomic and phenotypic characterization of in vitro-generated Chlamydia trachomatis recombinants. BMC Microbiol. 2013;13:142.

53. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem. 2011;83(22):8604-10.

54. Sosa JM, Huber DE, Welk B, Fraser HL. Development and application of MIPAR™: a novel software package for two- and three-dimensional microstructural characterization. Integrating Materials and Manufacturing Innovation. 2014;3(1):123-40.

55. Sherrid AM, Hybiske K. Chlamydia trachomatis Cellular Exit Alters Interactions with Host Dendritic Cells. Infect Immun. 2017;85(5).

Chapter 4. Chlamydial Interspecies Recombination Reveals Candidate Genes Involved with Species-Specific Growth and Infection Phenotypes

Abstract

Chlamydia trachomatis is the causative agent for the most commonly reported bacterial sexually transmitted disease. *Chlamydia muridarum* murine infections are widely used as a model for human infection by CT. However, despite striking genome synteny and gene identity (~99%) between these species, CT replicates poorly in mice and rarely ascends to the upper genital tract and CM is not known to infect humans. Both genetic and regulatory components are hypothesized to contribute to the observed host specificity. To evaluate this hypothesis, lateral gene transfer was used to generate heterogenomic chimeric species of CT and CM. Sixteen recombinants were generated which encompass 79.3% of the CM genome transferred into the CT genome in regions from 17kb to 125kb in size. Analysis of *in vitro* characteristics revealed two recombinants with lower infectious progeny production and one recombinant with increased progeny at 21hpi compared to parental *C. trachomatis*. Examination of the regions of recombination and the use of subclones were able to narrow down the gene products responsible for these phenotypes to select candidates. Additionally, mouse infections reveal no ascension in any of the sixteen chimeras, however, there appear to be potential advantages to specific regions in the lower female reproductive tract. These studies reveal the utility for *Chlamydia* recombinants in investigating species-specific phenotypes and provide potential gene candidates for further investigations.

Introduction

Chlamydia trachomatis is an obligate intracellular pathogen and the causative agent behind the most commonly reported bacterial infection in the United States. The associated sexually transmitted infection affects over one million people every year (1). Severity of disease ranges from asymptomatic to permanent or dangerous sequelae such as infertility, ectopic pregnancy and pelvic inflammatory disease in women (2-4). These serious outcomes are attributed to the detection of the bacteria and the elicited prolonged inflammatory response (5, 6). While many studies have been performed to determine the factors involved in the immune response to *C. trachomatis* (5-7), one primary challenge is the host-specific nature of *Chlamydia.*

As an obligate human pathogen, recapitulating *C. trachomatis* infections within animal models has its challenges. Primarily, while the bacteria can be introduced into the vaginal epithelium of mice, the pathogenesis fails to reflect human female infections (8). *C. trachomatis* is cleared quickly, with no prolonged immune response. Additionally, bacteria are restricted to the lower reproductive tract and unable to disseminate to the murine equivalent endometrial tissues, a process termed ascension (3). Because of this key limitation, in many studies, the obligate mouse, and closely related pathogen, *Chlamydia muridarum* is used (9, 10). *C. muridarum* genital infections in mice are able to mimic the pathogenesis of *C. trachomatis* in humans, complete with the ascension process (8).

Restricted host range is not a unique concept among bacteria. For example, Gram negative bacteria *Salmonella enterica* subsp. Typhi, an obligate human pathogen, is closely related to the subspecies *S.* Typhimerium, which has a broad host range (11). Broader host range, in this case, is largely attributed to the acquisition of plasmids or pathogenicity islands through horizontal gene transfer

(HGT) (12). There are many more examples of bacteria with restricted host range including *Neisseria* spp., *Listeria, Escherichia, Streptococcus* and others (13). As an example, enterohemorrhagic *E. coli* (EHEC) and enteropathogenic *E. coli* (EPEC) infections, which do not cause the same disease in mice that they do in humans, are often modelled by related obligate mouse pathogen, *Citrobacter rodentium* (14, 15). EPEC and EHEC share approximately 68% of their genetic content with *C. rodentium*, including core type 3 secretion effectors, suggesting that the pathogenesis strategies of these related bacterium are conserved (16). Research into the host-specificity for these bacteria seem to suggest that restriction is largely determined by one or more of the following factors; specific receptor recognition, the ability to evade a specific immune mechanism or the availability/utilization of a specific nutrient in the host (13). As in *Salmonella,* the ability of a bacteria to overcome one of these barriers is often tied to genetic exchange and the presence of one or more specialized genes. These can be recognized through comparative genomic studies.

Many such comparative genomics studies have been performed within *Chlamydia* (17-20), and despite the differences in host range *C. trachomatis* and *C. muridarum* share 99% of their genetic content (17, 18). The majority of non-conserved genes fall into a region known as the plasticity zone, which was hypothesized to serve as a pathogenicity island and affect host-specificity. However, studies in Chapter 3 have demonstrated that the plasticity zone may not be playing as significant a role in this process. Because there are no other candidate pathogenicity islands, nor are there additional plasmids beyond the nearly identical native chlamydial plasmid, we must look at the rest of the genome for answers about this host restriction. Outside of the plasticity zone are many genes which could contribute to host-specificity. These include unique coding sequences, most of which have no orthologs outside of *Chlamydia,* along with conserved, but host-interacting

components, like type 3 secretion effectors, outer membrane proteins and inclusion membrane proteins (17, 21, 22). We hypothesized that these conserved genes are contributing to host-specificity and have evolved within *C. muridarum* and *C. trachomatis* independently to interact with mouse or human proteins, respectively.

All *Chlamydia* undergo a unique biphasic developmental cycle where the bacteria oscillate between an infectious, but metabolically inert form, the elementary body (EB) and a replicative form that is unable to survive outside of the host, the reticulate body (RB) (23, 24). Upon attachment by an EB, *Chlamydia* is taken up by the host cell and quickly establishes an intracellular niche within the chlamydia-containing vacuole, known as the inclusion (21, 24). Within the inclusion, the EB undergoes a primary conversion event, becoming an RB, which is able to begin replicating. During this period of replication, known as the eclipse period, the inclusion expands and serves as an interface between the bacteria and the host cell. Through a still uncharacterized process, the RBs begin a secondary conversion process, turning back into infectious progeny in preparation for host lysis or extrusion of EBs. While all *Chlamydia* undergo this general developmental cycle, timing of the conversion events can vary by species and *C. muridarum* has been documented to have a shorter eclipse period by approximately six hours (25) (Chapter 3, Figure 3-3). In a typical *C. trachomatis in vitro* infection, peak infectious progeny are produced between 24 and 36 hours post infection while this peak occurs between 18 to 24 hours for *C. muridarum*. Shorter conversion could be associated with increased progeny over time, resulting in more organisms and increasing the ability to disseminate further, overwhelm the immune system and outcompete microflora. Therefore, increased growth rate of *C. muridarum* may be a factor in its ability to infect mice.

One recent advance in the *Chlamydia* field allows for the exploration of the entire chlamydial genome for regions important in host-specific phenotypes; the generation of a *C. trachomatis* and *C. muridarum* recombinant library (26). This library includes a set of seventeen chimeric strains with large regions of *C. muridarum* exchanged into *C. trachomatis* by HGT which cover approximately 79% of *C. muridarum* genes (Figure 4-1). One of these seventeen recombinants was characterized in Chapter 3, and so was not included in this study. In order to identify mechanisms of host restriction in *Chlamydia,* each *C. muridarum* region was evaluated in a progeny production screen to evaluate *in vitro* fitness and in an *in vivo* screen for mouse-specific infection.

Results

Genome sequencing reveals limited polymorphisms restricted to *C. muridarum* regions

The recombinant library was generated through horizontal gene transfer, facilitated by the use of a tetracycline-resistant *C. trachomatis* and chloramphenicol-resistant *C. muridarum* transposon parents with insertions throughout the *C. muridarum* genome (26). Whole genome sequencing was performed for each of the sixteen chimeras used in this study (Figure 4-1) with an average read coverage of over 1000X. In all recombinants, only *C. trachomatis* plasmid was detected, indicating that *C. muridarum* genes were recombined into the *C. trachomatis* genome. Of the sixteen recombinants, five have acquired limited single nucleotide polymorphisms (SNPs) (Table 4-1). Interestingly, each detected SNP was in the *C. muridarum* region of recombination, rather than the *C. trachomatis* backbone. RC4243 has accumulated three silent mutations within *TC0082* (*mreB*), in the *C. muridarum* region of recombination. RC435 has acquired one missense mutation within

TC0408 (hypothetical protein, P280Q), reverting the amino acid to the *C. trachomatis* amino acid at the position, and a single base deletion in *TC0412* (hypothetical protein, early stop codon at 338/366 amino acids). RC2418 has three silent mutations; one in *TC0493* and two in *TC0536*. RC6882 contains a silent mutation in *TC0668* and *TC0727* (*omcB*), and a single base deletion in the non-coding region between *TC0707* and *TC0708*. This intergenic mutation is 72bp ahead of the translational start site for *TC0708*. Finally, RC322 contains two SNPS; a missense mutation in *TC0867* (type III secretion system membrane protein, S438P) and a silent mutation in *TC0879*.
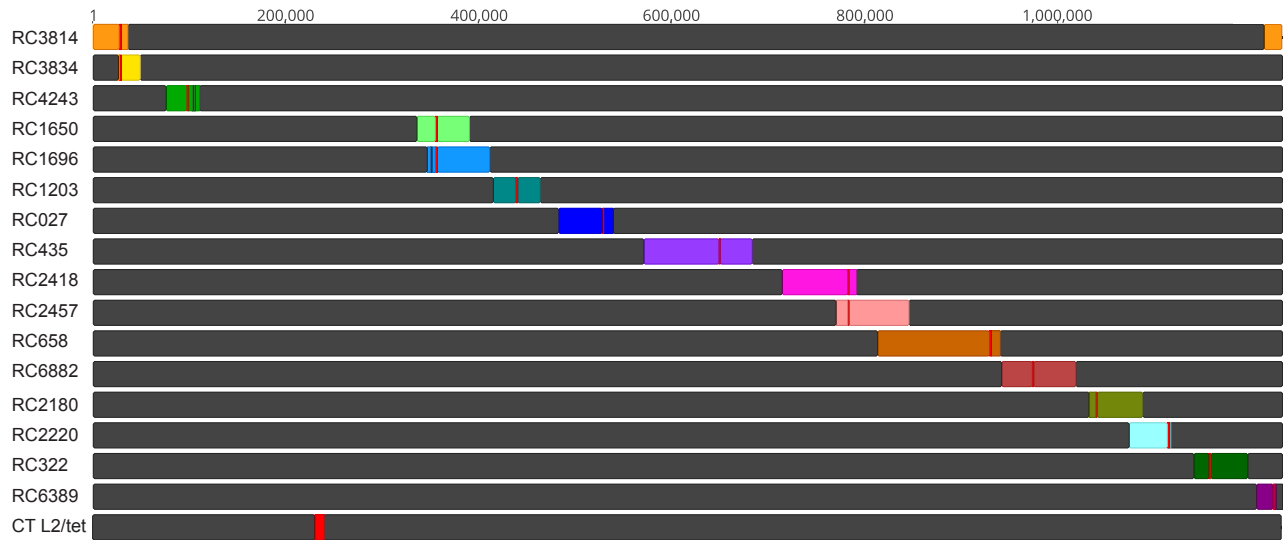
Figure 4-1. Schematic representation of the sixteen chimeras investigated in this study. Black bars represent *C. trachomatis* genomic regions. Colored bars represent specific *C. muridarum* region recombined by HGT into *C. trachomatis*. Red bars within colored regions indicate the location of the transposon insertion. Thick red bar in CT L2/tet signifies the tetracycline resistance cassette. Numbers along the left side denote recombinant (RC#).

Table 4-1. Polymorphisms detected in the recombinant genomes

Recombinant	Gene	Nucleotide change	Amino Acid change
RC4243	*tc0082*	T/A	V200V
RC4243	*tc0082*	A/G	E201E
RC4243	*tc0082*	T/G	S202S
RC435	*tc0408*	T/G	P280Q
RC435	*tc0412*	AGC/A-C	frameshift
RC2418	*tc0493*	C/A	P126P
RC2418	*tc0536*	A/C	S218S
RC2418	*tc0536*	A/G	H217H
RC6882	*tc0668*	A/C	R205R
RC6882	IGR[a] (*tc0707/708*)	G/-	NA
RC6882	*tc0727*	G/T	T36T
RC322	*tc0867*	T/C	S438P
RC322	*tc0879*	A/G	L184L

Because the recombination can occur within a coding region, chimeric genes may result in chimeric proteins which could lead to major changes to primary sequence and protein folding. Therefore, primary sequences were examined for all chimeric genes within the library. In all cases, recombinant genes did not appear to have frameshifts due to the recombination and primary sequences reveal intact coding sequences. Select clustal alignments of protein coding sequences reveal that primary sequence seamlessly transitions between *C. trachomatis* and *C. muridarum* (Table 4-2).

Table 4-2. Hybrid genes generated by the recombination in select strains

Recombinant	Genes	Amino Acid sequence

RC4243 · *CTL0060(CT691)/tc0063*

```
tc0063        MQVLASLFGRSPFSPLQAHVELVSASIEVLFPLFSAIKEGDYQRVEALAQLVSSKERQAD    60
CTL0060       MQVLASLFGQSPFAPLQAHLELVSSTINVLFPLFSALKEGDYERVGVLAQLVSSKERQAD    60
CTL0060/tc0   MQVLASLFGQSPFAPLQAHLELVSSTINVLFPLFSALKEGDYERVGVLAQLVSSKERQAD    60
              *********:***:*****:****::*:********:*****:**  .*************

tc0063        GVKNDIRSHLASGVFIPVSRLAMLEIISTQDSVADCAEDIAILLTVKELRFYPEFEEIFF   120
CTL0060       GMKNDVRRHLASGVFLPVSRAALLEIISIQDSLADCAEDIAILLTVKELQFYPEFEELFF   120
CTL0060/tc0   GMKNDVRRHLASGVFLPVSRAALLEIISIQDSLADCAEDIAILLTVKELQFYPEFEELFF   120
              *:***:* *******:**** *:***** ***:*****************:*******:**

tc0063        QFLQKTVQSFEVVAKAIREMDLLLESSFGGCRAEKTRVLVNEVSNLEHECDLLQRELMKI   180
CTL0060       EFLQKTVQSFEAVAKTIREMDRLLESSFGGNRADKTRVLVSEVSNLEHECDLLQRELMKV   180
CTL0060/tc0   EFLQKTVQSFEAVAKTIREMDRLLESSFGGNRADKTRVLVSEVSNLEHECDLLQRELMKV   180
              :**********.***:***** ******** **:******.**:******.**********:

tc0063        LFSENFSIETKDFVLWTQIIKRLSGISNNSEKLAYRVGMTLEEK*    224
CTL0060       FFSDDFAIGTKGFVLWMQIIKGISGISNNSEKLAYRVSMTLEEK*    224
CTL0060/tc0   FFSDDFAIGTKDFVLWTQIIKRLSGISNNSEKLAYRVGMTLEEK*    224
              :**::*:* **.**** **** :**************.*******
```

RC4243 · *tc0087/CTL0083(CT714)*

```
CTL0083          MKETIAYLGMGMWGFSLANLLANNGHRVVGWARNPALIEQLSVQRRHPAAPHISIPQNLS    60
tc0087/CTL0083   MKETIAYLGMGMWGFSLANLLANNGHRVVGWARNPALIEQLSVQRRHPAAPHISIPQNLS    60
tc0087           MKETIAYLGMGMWGFSLANLLANNGHRVVGWARNPSLIEQLSTQRQHPAAPHVTIPSNLS    60
                 ***********************************:******.**:******:.**.***

CTL0083          FTYHMEEALDGATMIVEGVTSAGMRPVLTQLKALTELRVPLVITSKGIEQNTGLLLSEIA   120
tc0087/CTL0083   FTYHMEEALDGATMIVEGVTSAGMRPVLTQLKALTELRVPLVITSKGIEQNTGLLLSEIA   120
tc0087           FTSSMEEALDGATMIVEGVTSAGMRPVLNQLKSITDLQIPLVITSKGIEQNTGLLLSEIA   120
                 **  ************************.***:*:*::.*.::**********************

CTL0083          LEIFGRPAAQHLGYLSGPSIASEVLRGCPCSVVISAYNPDTLKQIHRAFLTPTFRVYPNS   180
tc0087/CTL0083   LEIFGRPAAQHLGYLSGPSIASEVLRGCPCSVVISAYNPDTLKQIHRAFLTPTFRVYPNS   180
tc0087           LEIFGKPAAKYLGYLSGPSIASEVLRGCPCSVVISAYDPATLKQIHQAFLTPTFRVYPNS   180
                 *****:***::*******************************:* ******:**************

CTL0083          DLKGVALGGALKNVIAIACGISDGFRFGDNAKSGLVTRGLHEIRKFATIMGCRPDTLNGL   240
tc0087/CTL0083   DLKGVALGGALKNVIAIACGISDGFRFGDNAKSGLVTRGLHEIRKFATIMGCRPDTLNGL   240
tc0087           DLKGVALGGALKNVIAIACGISDGFRFGDNAKSGLVTRGLHEIRKFATIMGCRPDTLNGL   240
                 ************************************************************

CTL0083          AGLGDLCTTCFSAFSRNTLFGKLLAEGLTPEQAKTKIGMVVEGVYTALSAHQIATHHRID   300
tc0087/CTL0083   AGLGDLCTTCFSAFSRNTLFGKLLAEGLTPEQAKTKIGMVVEGVYTALSAHQIATHHRID   300
tc0087           AGLGDLCTTCFSAFSRNTLFGKMLAEGLTPEQAKTKIGMVVEGVYTALSAHQIATHHKID   300
                 **********************:*****************************************:**

CTL0083          MPITTSVYRVLYENLDIQEGIAQLLQRDTKEEYL*   334
tc0087/CTL0083   MPITTSVYRVLYENLDIQEGIAQLLQRDTKEEYL*   334
tc0087           MPITTGVYRVLYENLDIQKGIAQLLQRNTKEEYL*   334
                 *****.************:********:*******
```

RC1696   *CTL0165(CT796)*
                 /tc0178

```
tc0178           MSSQPLTLQDMMAAILRFWSEQGCIIHQGYDLEVGAGTFNPATFLQALGPEPFKTAYIEP      60
CTL0165          MSSQPLTLQAMMAAILNFWSEQGCIIHQGYDLEVGAGTFNPATFLQSLGPEPFRTAYIEP      60
CTL0165/tc0178   MSSQPLTLQDMMAAILRFWSEQGCIIHQGYDLEVGAGTFNPATFLQALGPEPFKTAYIEP      60
                 ********* ***** .*****************************:*****:******

tc0178           SRRPQDGRYGQHPNRLQKYHQLQVILKPVPENFLSLYLESLKVIGLNLVDHDIRFVHDDW     120
CTL0165          SRRPQDGRYGQHPNRLQKYHQLQVILKPVPENFLSLYLESLKVIGLNLVDHDIRFVHDDW     120
CTL0165/tc0178   SRRPQDGRYGQHPNRLQKYHQLQVILKPVPENFLSLYLESLKVIGLNLVDHDIRFVHDDW     120
                 ************************************************************

tc0178           ENPTIGAWGLGWEVWLNGMEITQLTYFQAVGSKPLEAISGEITYGVERIAMYLQKKNSVY     180
CTL0165          ENPTIGAWGLGWEVWLNGMEITQLTYFQAVGSKPLDAISGEITYGVERIAMYLQKKNSVY     180
CTL0165/tc0178   ENPTIGAWGLGWEVWLNGMEITQLTYFQAVGSKPLEAISGEITYGVERIAMYLQKKNSVY     180
                 ***********************************:************************

tc0178           DVMWNDSLTYGDITQHAEQAWSQYNFETANTSMWLKHFEDFAAEALATLDKGLPLPAYDF     240
CTL0165          DVMWNGSLTYGDITQYAEQAWSQYNFETANTTMWLKHFDDFSAEALATLDQGLPLPAYDF     240
CTL0165/tc0178   DVMWNDSLTYGDITQHAEQAWSQYNFETANTSMWLKHFEDFAAEALATLDKGLPLPAYDF     240
                 *****.*********:***************:******:**:*********:*********

tc0178           VIKASHAFNMLDSRGVISVTERTRYITKIRQLARAVADKYVIWRESLGFPLLKTIPSTPT     300
CTL0165          VIKASHAFNMLDSRGVISVTERTRYIAKIRQLARAAADKYVAWRESLGFPLLKTPPSTPT     300
CTL0165/tc0178   VIKASHAFNMLDSRGVISVTERTRYITKIRQLARAVADKYVIWRESLGFPLLKTIPSTPT     300
                 **************************:*********.*****.*********** *****

tc0178           VTAKQIPHICQDEDFLLEIGSEELPAAFVPTGIQQLESLAKKLLADHNISYNNLEVLGTP     360
CTL0165          VTPKKIPTICQPEDFLLEIGSEELPATFVPTGIQQLESLAKKLLADHGIAYKHLEVLGTP     360
CTL0165/tc0178   VTAKQIPHICQDEDFLLEIGSEELPAAFVPTGIQQLESLAKKLLADHNISYNNLEVLGTP     360
                 ** *:** *** **************:*******************.*:*:.*******

tc0178           RRLALRIQGLSHLTIRPEAEKKGPPLSLLFEEDGSVSSQGEQFFASHGLSISHRSALDQS     420
CTL0165          RRLALCIEGLSHVTIRPESEKKGPPLSLLFMTDGSVSPQGEQFFPSHGLSISHRSALDQP     420
CTL0165/tc0178   RRLALRIQGLSHLTIRPEAEKKGPPLSLLFEEDGSVSSQGEQFFASHGLSISHRSALDQS     420
                 ***** *:****:*****:*********** ***** ****** **************

tc0178           STICRVRSIKGTDYLFLVIPEERIETAAILVNELPLLIRSMRFPKKMTWDNGGVEYARPI     480
CTL0165          SAICRVRSINGTDYLFLVIPEERKETAAILVNELPQLIRSIRFPQKMTWDNGGVEYARPI     480
CTL0165/tc0178   STICRVRSIKGTDYLFLVIPEERIETAAILVNELPLLIRSMRFPKKMTWDNGGVEYA-PI     479
                 *:*******:************* ********** ****:***:************* **

tc0178           RWLVALYGDQVLPISLGFVSSGNISWGHRQLDNRQLTIPSSKEYIDILRDACVIVSQKER     540
CTL0165          RWLVALYGDQILPISLGFVSSGNTSWGHRQLDNRQLTIPSSNMYVDTLRSACVIVSQKER     540
CTL0165/tc0178   RWLVALYGDQILPISLGFVSSGNTSWGHRQLDNRQLTIPSSNMYVDTLRSACVIVSQKER     539
                 **********:************.*************** *:* **.**********

tc0178           RSIIEQGLQNLTGDQTVAIAPEHLIEETVFLTEHPFVICAQFNPDFCSLPKELLIAEMIN     600
CTL0165          RAIIKQGLQNLTGDQIVAIAPEHLVDETVFLTEHPFVISAQFDPAFCSLPKELLIAEMIQ     600
CTL0165/tc0178   RAIIKQGLQNLTGDQIVAIAPEHLVDETVFLTEHPFVISAQFDPAFCSLPKELLIAEMIQ     599
                 *:**:********** ********:.***********.***:* **************:

tc0178           HQRYFPTQNLQGEITNRFLIVCDNSPTDTIIEGNEKALAPRLTDGNFLFKQDLLTSLDSF     660
CTL0165          HQRYFPTQNMQGEITNRFLIVCDNSPTDSIVEGNEKALAPRLTDGNFLFKQDLLTPLSSF     660
CTL0165/tc0178   HQRYFPTQNMQGEITNRFLIVCDNSPTDSIVEGNEKALAPRLTDGNFLFKQDLLTPLSSF     659
                 *********:*****************:.*:*********************** *.**

tc0178           VEKLKSVTYFDALGSLADKTARLKLHLEETYPLLPLCPKEDIDTAVHYCKADLVSAVVNE     720
CTL0165          VEKLKSVTYFESLGSLADKTSRLKLHLEEAYALLPLCAKEDIDTAIHYCKADLVSSVVNE     720
CTL0165/tc0178   VEKLKSVTYFESLGSLADKTSRLKLHLEEAYALLPLCAKEDIDTAIHYCKADLVSSVVNE     719
                 **********::********.********:* ***** *******.*********:****

tc0178           FPELQGIMGRYYLQNAALSKAAAIAVGEHLQHITLGSSVSTTGALLSILDRVDNLLSCFI     780
CTL0165          FPELQGIMGRYYLQNASLSRAAALAIGEHLQHITLGSNISTTGALLSILDRIDNLLSCFI     780
CTL0165/tc0178   FPELQGIMGRYYLQNASLSRAAALAIGEHLQHITLGSNISTTGALLSILDRIDNLLSCFI     779
                 ****************:**:***.*:***********. .************:********

tc0178           LGLLPTSSHDPYALRRQSLELLTLLYTTQSSVDIEDLFSRLVRHFPTTIPNTVWSPEDVL     840
CTL0165          LGLLPTSSHDPYALRRQSLEILTLLYTTQSSVDIEDLFARLIRHFPSSIPNTVWSPEEVL     840
CTL0165/tc0178   LGLLPTSSHDPYALRRQSLEILTLLYTTQSSVDIEDLFARLIRHFPSSIPNTVWSPEEVL     839
                 ******************:***************:**:***.****::********:**

tc0178           NKLCSFVWGRLKTILSSLGFAKEVIAAVLTENCPKNPLTIINSARSIQELQNTQTLETIA     900
CTL0165          SKLNTFVWGRLRTILSSLGFDKEIIATVLTDNCPKNPLTIIQSAQSIQELKNTQILKTIA     900
CTL0165/tc0178   SKLNTFVWGRLRTILSSLGFDKEIIATVLTDNCPKNPLTIIQSAQSIQELKNTQILKTIA     899
                 .** :******:******** **:**:***:*********:**:*****:*****:*.***

tc0178           STHNRLKKILASLSFSVTEQIFSLITSEDMLFKQALERFKEATTSLPISSREYLLQLEDL     960
CTL0165          ATHNRLKKILASLSFSVTEQMFSLQSAEDLLFKQALDRFVEETTALPISSKDYLHLLKEL     960
CTL0165/tc0178   ATHNRLKKILASLSFSVTEQMFSLQSAEDLLFKQALDRFVEETTALPISSKDYLHLLKEL     959
                 :*******************:*** ::**:******:** * **:*****::**  *::*

tc0178           SQSTALFLDSVRIADDDENIRNQRIALLVATQKCFGFYAWDAL-                1003
CTL0165          AQSTELFLDSVRVASDDESTRNQRIALLIAAQKCFGFYAWDVL*                1003
CTL0165/tc0178   AQSTELFLDSVRVASDDESTRNQRIALLIAAQKCFGFYAWDVL*                1002
                 :*** *******:*.***. ********:*:**********.*
```

```
                                   ┌─┐                        ┌─┐
CTL0167         MKIIHTAIEFAPVIKAGGLGDALYGLAKALAANHTTEVVIPLYPKLFTLPKEQDLCSIQK      60
tc0181/CTL0167  MKIIHTAIEFAPVIKAGGLGDALYGLAKALAANHTTEVVIPLYPKLFTLPKEQDLCSIQK      60
tc0181          MKITHTAIEFAPVIKAGGLGDALYGLAKALAVNHTTEVVIPLYPKLFTSLHEQDLFATQK      60
                *** ***************************.****************  :**** : **

                └┐
CTL0167         LSYFFAGEQEATAFSYFYEGIKVTLFKLDTQPELFENAETIYTSDDAFRFCAFSAAAASY      120
tc0181/CTL0167  LSYFFAGEQEATAFSYFYEGIKVTLFKLDTQPELFENAETIYTSDDAFRFCAFSAAAASY      120
tc0181          IPYFFAGEQEATAFSYFYEGIKVTLLKLDSQPELFEEAETIYTNDDAFRFCAFSAAAAAY      120
                : **********************:***:******:****** .****************:*

                               ┌─┐         ┌─┐
CTL0167         IQKEGANIVHLHDWHTG--LVAGLLKQQPCSQLQKIVLTLHNFGYRGYTTREILEASSLN      178
tc0181/CTL0167  IQKEGANIVHLHDWHTGLGLVAGLLKQQPCPQLQKIVLTLHNFGYRGYTTREVLEASSLN      180
tc0181          IQKEGADIVHLHDWHVG--LVAGLLKQQPCPQLQKIVLTLHNFGYRGYTTREVLEASSLN      178
                ******:********.*  **********  ***********************:*******

CTL0167         EFYISQYQLFRDPQTCVLLKGALYCSDFVTTVSPTYAKEILEDYSDYEIHDAITARQHHL      238
tc0181/CTL0167  EFYLSHYQLFRDPQTCVLLKGALYCSDFVTTVSPTYAKEILQDYSDYEIHDAITARQHHL      240
tc0181          EFYLSHYQLFRDPQTCVLLKGALYCSDFVTTVSPTYAKEILQDYSDYEIHDAITARQHHL      238
                ***:*:****************************************:*************

                          ┌─┐                     ┌─┐
CTL0167         RGILNGIDTTIWGPETDPNLAKNYTKELFETPSIFFEAKAENKKALYERLGLSLEHSPCV      298
tc0181/CTL0167  KGILNGIDYTILDPETDPHLAKNYSKVLFEDPKAFFEAKAENKKALYETLGLSLDKSPCM      300
tc0181          KGILNGIDYTILDPETDPHLAKNYSKVLFEDPKAFFEAKAENKKALYETLGLSLDKSPCM      298
                :******* ** .*****:*****:* *** *. ************** *****::***:

CTL0167         CIISRIAEQKGPHFMKQAILHALENAYTLIIIGTCYGNQLHEEFANLQESLANSPNVRIL      358
tc0181/CTL0167  CIISRIAEQKGPEFMKQAILHALENAYTLIIIGTCYGGQIHKEFSNLQESLADSPNVRIL      360
tc0181          CIISRIAEQKGPEFMKQAILHALENAYTLIIIGTCYGGQIHKEFSNLQESLADSPNVRIL      358
                ************.************************.*:*:**:*******:*******

CTL0167         LTYSDVLARQIFAAADMICIPSMFEPCGLTQMIGMRYGTVPLVRATGGLADTVANGINGF      418
tc0181/CTL0167  LTYSDVLARQIFAAADMICIPSMFEPCGLTQMIGMRYGTVPVVRATGGLADTVTDGVNGF      420
tc0181          LTYSDVLARQIFAAADMICIPSMFEPCGLTQMIGMRYGTVPVVRATGGLADTVTDGVNGF      418
                *****************************************:************.:*:***

                ┌─┐           ┌─┐     ┌─┐
CTL0167         SFFNPHDFYEFRNMLLEAVTTYRTNHDKWQHIVRACLNFSSDLETAANKYLEIYKQ*      474
tc0181/CTL0167  SFSNPHDFHEFRNMLSKAIATYRDDQDKWQQIVRSCLEFSSDLETAANKYLEIYQQ*      476
tc0181          SFSNPHDFHEFRNMLSKAIATYRDDQDKWQQIVRSCLEFSSDLETAANKYLEIYQQ*      474
                ** *****:****** :*::*** ::****:***:**:*****************:**
```

RC1696 — CTL0168 (CT799)/tc0182

```
CTL0168         MELVVQSRETDKKSVIKKIRQQGGIPAVLYSGGKSLANIVVDARVFSKFLSTLESGALAS      60
tc0182/CTL0168  MELVVQSRETDKKSVIKKIRQQGGIPAVLYSGGKSLANIVVDARVFSKFLSTLESGALAS      60
tc0182          MELVVQSRETDKKSVIKKIRQQGGIPAVLYSGGKSLANVVIDARVFNKFLSTLESGALAS      60
                **************************************:*:*****:************

CTL0168         TVFTLSYEGREIKALVKDIQYHVTTYDVIHLDFEELVDGRDVRLNVPIRCINTVDCVGVK      120
tc0182/CTL0168  TVFTLSYEGREIKALVKDIQYHVTTYDVIHLDFEELVDGRDVRLNVPIRCINTVDCVGVK      120
tc0182          TVFTLSYEGREIKALVKDIQYHVTTYDVIHLDFEELVEDREVRLNIPIRCINTVDCVGVK      120
                *************************************:.*:****:*************

                                                     ┌─┐         ┌─┐
CTL0168         LGGSLRQVIRCIRVVCKPKDIVPFLELDVQSLGLSQTLKLSDICIPEGIRPVTSLKEVAV      180
tc0182/CTL0168  LGGSLRQVIRCIRVVCKPKDIVPFLELDVQSLGLSQTLKLSDICIPEGIRPVTSLKEVAV      180
tc0182          LGGSLRQVIRYMRVVCKPKDIVPFLELDVQSLGLSQTLKLSDIRIPEGIKPVTPLKEVAV      180
                ********** :************************** *****:*** ******

CTL0168         TVARR*      185
tc0182/CTL0168  TVARR*      185
tc0182          TVARR*      185
                ******
```

RC1696 — tc0232/CTL0216(CT844)

```
                                 ┌─┐              ┌─┐
tc0232          MCIEKDLFFMKKALDEARKAYELDEVPVGCIIVQGDAIIARGHNSVERLKDPTAHAEMIC      60
CTL0216         MCIEKDLFFMKKALDEARKAYEQDEVPVGCIIVHGDKIIARGHNSVEQLKDPTAHAEMIC      60
tc0232/CTL0216  MCIEKDLFFMKKALDEARKAYELDEVPVGCIIVQGDAIIARGHNSVERLKDPTAHAEMIC      60
                ********************* **********:** **********:.*************

                         ┌─┐
tc0232          ISAAAEYLQNWRLKDTTLYCTLEPCLMCAGAIQLARIPRIVWGAPDLRLGAGGSWLNVFL      120
CTL0216         ISAAAEYLENWRLKDTILYCTLEPCLMCAGAIQLARIPRIVWGAPDLRLGAGGSWLNVFL      120
tc0232/CTL0216  ISAAAEYLENWRLKDTILYCTLEPCLMCAGAIQLARIPRIVWGAPDLRLGAGGSWLNVFL      120
                ********:*******.*******************************************

                             ┌─┐      ┌─┐        ┌─┐
tc0232          EKHPFHQVECCPGICHQESEWLMKNFFWEKRKEKNEK*------              157
CTL0216         EKHPFHQVECCSGVCYQESEQLMKNFFLEKRKAKDEGRNSRAT*             163
tc0232/CTL0216  EKHPFHQVECCSGVCYQESEQLMKNFFLEKRKAKDEGRNSRAT*             163
                *********** *:*:**** ****** **** *:*
```

RC1650 — CTL0155(CT787)/tc0170

```
tc0170          MAKKSAVAREVKRRKLVEANFQKRAELRKLAKSLSVSEEERERAREALNKMKRDTSPSRL      60
CTL0155         MAKKSAVAREVKRRKLVEANFQKRAELRKLAKSLSVSEEERERAREALNKMRRDTSPSRL      60
CTL0155/tc0170  MAKKSAVAREVKRRKLVEANFQKRAELRKLAKSLSVSEEERERAREALNKMRRDTSPSRL      60
                ***************************************************:********

tc0170          HNRCLLTGRPRGYLRKFAISRICFRQMASMGDIPGVIKASW*      101
CTL0155         HNRCLLTGRPRGYLRKFAISRICFRQMASMGDIPGVVKASW*      101
CTL0155/tc0170  HNRCLLTGRPRGYLRKFAISRICFRQMASMGDIPGVIKASW*      101
                ************************************:*****
```

RC1696 — tc0181/CTL0167(CT798)

RC1696 — CTL0168 (CT799)/tc0182

RC1696 — tc0232/CTL0216(CT844)

RC1650 — CTL0155(CT787)/tc0170

RC1650   *tc0214/*
CTL0199(CT827)

```
tc0214           MVDLQEKQCTIVKRNGMFVPFDRNRIFQALEAAFRDTRRIDDHMPLPEDLENSIRSITHQ      60
CTL0199          MVDLQEKQCTIVKRNGMFVPFDRNRIFQALEAAFRDTRRIDDHMPLPEDLESSIRSITHQ      60
tc0214/CTL0199   MVDLQEKQCTIVKRNGMFVPFDRNRIFQALEAAFRDTRRIDDHMPLPEDLESSIRSITHQ      60
                 **********************************************.********

tc0214           VVKEVVQKITDGQVVTVERIQDMVESQLYINGLQDVARDYVVYRDDRKAHREKSWQSLSV     120
CTL0199          VVKEVVQKITDGQVVTVERIQDMVESQLYVNGLQDVARDYIVYRDDRKAHRKKSWQSLSV     120
tc0214/CTL0199   VVKEVVQKITDGQVVTVERIQDMVESQLYVNGLQDVARDYIVYRDDRKAHREKSWQSLSV     120
                 *****************************:.***********:**********:********

tc0214           IRRCGTTVHFNPMKISAALEKAFRATDRIEGMTPDFVREEVNALTQKIVAEIEERCSQQD     180
CTL0199          VRRCGTVVHFNPMKISAALEKAFRATDKTEGMTPSSVREEINALTQNIVAEIEECCSQQD     180
tc0214/CTL0199   IRRCGTTVHFNPMKISAALEKAFRATDKTEGMTPSSVREEINALTQNIVAEIEECCSQQD     180
                 :****.*******************: *****. ****:*****:******* *****

tc0214           SRIDIEQIQDIVEQQLMVVGHYATAKNYILYREARARVRDNRVEDQIVEEAPSEETFEVL     240
CTL0199          RRIDIEKIQDIVEQQLMVVGHYAIAKNYILYREARARVRDNREEDGSTEKTIAEEAFEVL     240
tc0214/CTL0199   RRIDIEKIQDIVEQQLMVVGHYAIAKNYILYREARARVRDNREEDGSTEKTIAEEAFEVL     240
                 *****:****************** ****************** **  .*:: :**:****

tc0214           SKDGSTYMITHSQLLARLARACSRFPETTDAALLTDMAFSNFYSGIKESEVVLACIMAAR     300
CTL0199          SKDGSTYTMTHSQLLAHLARACSRFPETTDAALLTDMAFANFYSGIKESEVVLACIMAAR     300
tc0214/CTL0199   SKDGSTYTMTHSQLLAHLARACSRFPETTDAALLTDMAFANFYSGIKESEVVLACIMAAR     300
                 ******* :*******:*************************:******************

tc0214           ANIEKEPDYAFVAAELLLDVVYKEALDRSRGDEDLEQVYRDHFKRYIMEGDSYRLNPELK     360
CTL0199          ANIEKEPDYALVAAELLLDVVYKEALGKSKYAEDLEQAHRDHFKRYIAEGDTYRLNAELK     360
tc0214/CTL0199   ANIEKEPDYALVAAELLLDVVYKEALGKSKYAEDLEQAHRDHFKRYIAEGDTYRLNAELK     360
                 **********:***************.:*:  *****.:******** ***:**** ***

tc0214           NLFDLDALANAMDLSRDLQFSYMGIQNLYDRYFNHDDGRRLETPQIFWMRVAMGLALKEQ     420
CTL0199          HLFDLDALADAMDLSRDLQFSYMGIQNLYDRYFNHHEGCRLETPQIFWMRVAMGLALNEQ     420
tc0214/CTL0199   HLFDLDALADAMDLSRDLQFSYMGIQNLYDRYFNHHEGCRLETPQIFWMRVAMGLALNEQ     420
                 :********:*******************************.:* ****************:**

tc0214           DKTYWAITFYNLLSTFRYTPATPTLFNSGMRHSQLSSCYLSTVQDDLVNIYKVISDNAML     480
CTL0199          DKTSWAITFYNLLSTFRYTPATPTLFNSGMRHSQLSSCYLSTVQDDLVNIYKVIADNAML     480
tc0214/CTL0199   DKTSWAITFYNLLSTFRYTPATPTLFNSGMRHSQLSSCYLSTVQDDLVNIYKVIADNAML     480
                 *** ****************************************************:*****

tc0214           SKWAGGIGNDWTAIRATGALIKGTNGKSQGVIPFIKVTNDTAVAVNQGGKRKGAVCVYLE     540
CTL0199          SKWAGGIGNDWTAIRATGALIKGTNGRSQGVIPFIKVTNDTAVAVNQGGKRKGAVCVYLE     540
tc0214/CTL0199   SKWAGGIGNDWTAIRATGALIKGTNGRSQGVIPFIKVTNDTAVAVNQGGKRKGAVCVYLE     540
                 **************************:**********************************

tc0214           VWHLDYEDFLELRKNTGDDRRRAHDVNTASWIPDLFFKRLQQKGSWTLFSPDDVPGLHDA     600
CTL0199          VWHLDYEDFLELRKNTGDERRRAHDVNTASWIPDLFFKRLQQKGTWTLFSPDDVPGLHDA     600
tc0214/CTL0199   VWHLDYEDFLELRKNTGDERRRAHDVNTASWIPDLFFKRLQQKGTWTLFSPDDVPGLHDA     600
                 *****************:******************************:*************

tc0214           YGEEFERLYEEYERKVDSGEIRLYKKVEAEDLWRKMLSMLFETGHPWMTFKDPSNIRSAQ     660
CTL0199          YGEEFERLYEEYERKVDTGEIRLFKKVEAEDLWRKMLSMVFETGHPWMTFKDPSNIRSAQ     660
tc0214/CTL0199   YGEEFERLYEEYERKVDTGEIRLFKKVEAEDLWRKMLSMVFETGHPWMTFKDPSNIRSAQ     660
                 *****************.:*****.***************:********************

tc0214           DHTGVVRCSNLCTEILLNCSETETAVCNLGSVNLVQHILDDGLDEEKLSETISIAVRMLD     720
CTL0199          DHKGVVRCSNLCTEILLNCSETETAVCNLGSINLVQHIVGDGLDEEKLSETISIAVRMLD     720
tc0214/CTL0199   DHKGVVRCSNLCTEILLNCSETETAVCNLGSINLVQHIVGDGLDEEKLSETISIAVRMLD     720
                 **.*****************************:******:.***********************

tc0214           NVIDINFYPTKEAKEANFAHRAIGLGVMGFQDALYKLDISYASQEAVEFADYSSELISYY     780
CTL0199          NVIDINFYPTKEAKEANFAHRAIGLGVMGFQDALYKLDISYASQEAVEFADYSSELISYY     780
tc0214/CTL0199   NVIDINFYPTKEAKEANFAHRAIGLGVMGFQDALYKLDISYASQEAVEFADYSSELISYY     780
                 ***********************************************************

tc0214           AIQASCLLAKERGTYSSYKGSKWDRGLLPIDTIQLLANYRGKDNLQMDTSVRKDWEPIRS     840
CTL0199          AIQASCLLAKERGTYSSYKGSKWDRGLLPIDTIQLLANYRGEANLQMDTSSRKDWEPIRS     840
tc0214/CTL0199   AIQASCLLAKERGTYSSYKGSKWDRGLLPIDTIQLLANYRGEANLQMDTSSRKDWEPIRS     840
                 *********************************************: ******* *********

tc0214           LIREHGMRNCQLMAIAPTATISNIIGVTQSIEPTYKHLFVKSNLSGEFTIPNVYLIEKLK     900
CTL0199          LVKEHGMRHCQLMAIAPTATISNIIGVTQSIEPTYKHLFVKSNLSGEFTIPNVYLIEKLK     900
tc0214/CTL0199   LVKEHGMRHCQLMAIAPTATISNIIGVTQSIEPTYKHLFVKSNLSGEFTIPNVYLIEKLK     900
                 *:.:*****:*************************************************

tc0214           KLGIWDADMLDDLKYFDGSLLEIERVPDHIKHIFLTAFEIEPEWILECASRRQKWIDMGQ     960
CTL0199          KLGIWDADMLDDLKYFDGSLLEIRIPDHLKHIFLTAFEIEPEWIIECASRRQKWIDMGQ     960
tc0214/CTL0199   KLGIWDADMLDDLKYFDGSLLEIRIPDHLKHIFLTAFEIEPEWIIECASRRQKWIDMGQ     960
                 **********************:.***:****************:***************

tc0214           SLNLYLAQPDGKKLSNMYLTAWKKGLKTTYYLRSSSATTVEKSFVDINKRGIQPRWMKNK    1020
CTL0199          SLNLYLAQPDGKKLSNMYLTAWKKGLKTTYYLRSSSATTVEKSFVDINKRGIQPRWMKNK    1020
tc0214/CTL0199   SLNLYLAQPDGKKLSNMYLTAWKKGLKTTYYLRSSSATTVEKSFVDINKRGIQPRWMKNK    1020
                 *****************************************************************

tc0214           SASAGIVVERASKTPVCSLEEGCEVCQ*        1047
CTL0199          SASAGIIVERAKKAPVCSLEEGCEACQ*        1047
tc0214/CTL0199   SASAGIIVERAKKAPVCSLEEGCEACQ*        1047
                 ******:****.*:**********.***
```

Three time-point progeny production screen reveals three chimeras with differences in progeny production relative to *C. trachomatis*

To evaluate growth rate of the recombinants, infectious progeny produced at three time points were analyzed to determine when secondary conversion occurred relative to *C. trachomatis* and *C. muridarum* wildtype (Figure 4-2). *C. muridarum* produced high levels of infectious progeny at 18 hours which remained stable through 24 hours, consistent with the early secondary conversion while *C. trachomatis* began converting to infectious progeny around 21 hours and reached its peak by 24 hours. The majority, thirteen out of sixteen, of the recombinants had a later conversion, similar to *C. trachomatis* (Figure 4-2A). Two recombinants, RC435 and RC4243, exhibited significantly poor growth (p-values <0.005), with reduced 21-hour progeny compared with *C. trachomatis* and one recombinant, RC1696, exhibited increased progeny at 21h (Figure 4-2B). For RC435, this growth was expected, as the chimera contains a *C. muridarum* inclusion membrane protein that has been previously demonstrated to be incompatible between species, resulting in early inclusion lysis and host cell apoptosis, *TC0424* (Chapter 3, Figure 3-7 and 3-9)
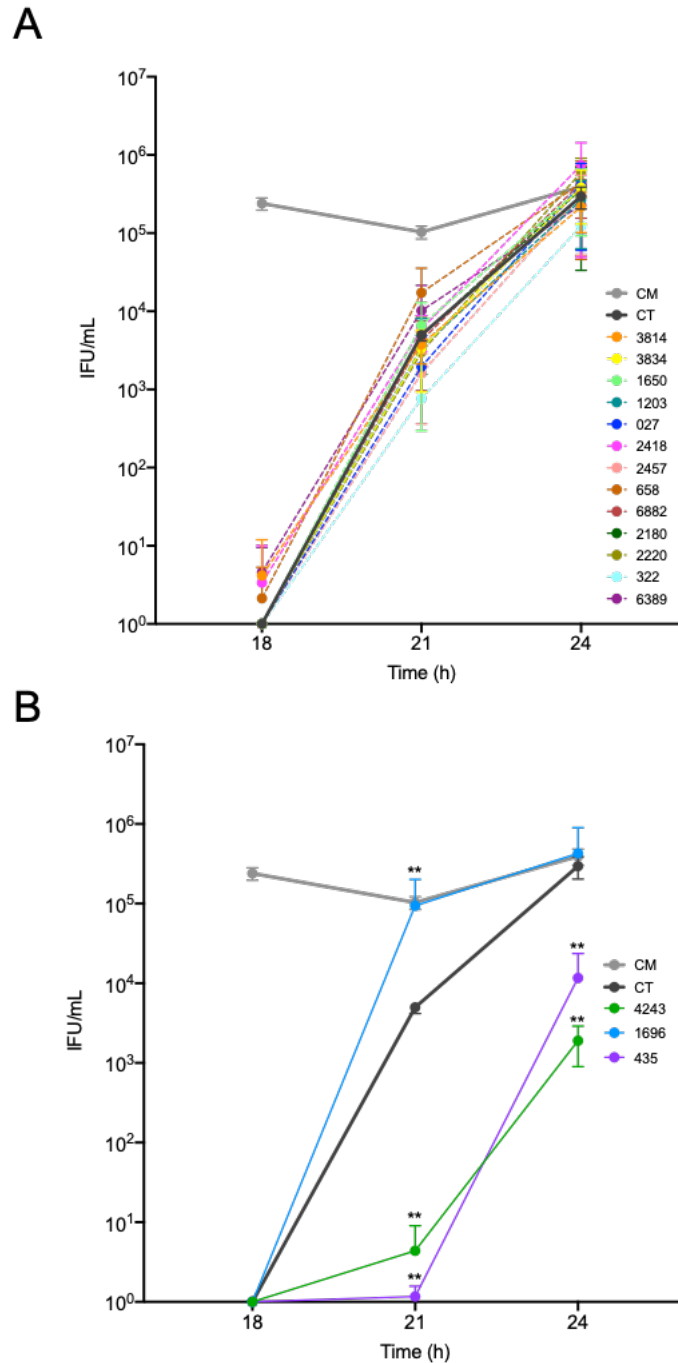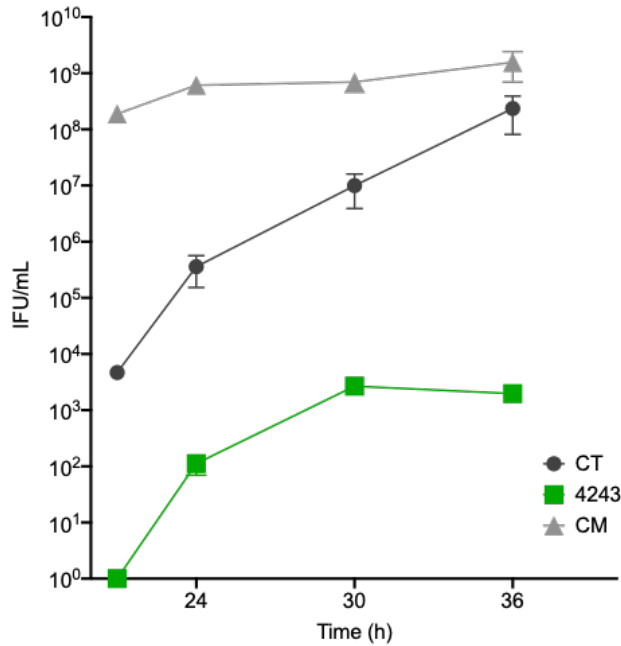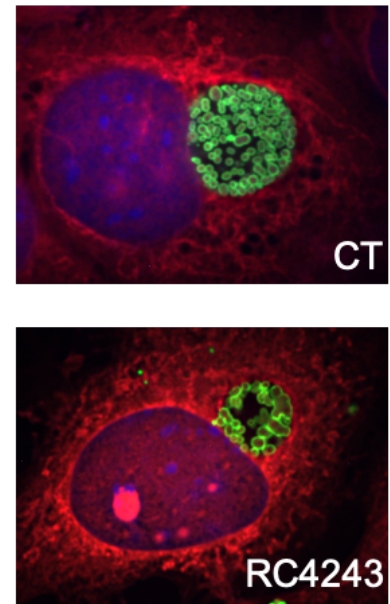
Figure 4-2. Three timepoint growth curves for recombinants (#) compared with wildtype *C. trachomatis* (CT) and *C. muridarum* (CM). Recombinants were infected at a MOI ~1 onto L929 mouse fibroblast cells. At 18, 21- or 24-hours post-infection, EBs were isolated and infectious progeny was titered to calculate inclusion forming units per mL. Growth curves were normalized to the initial infectious dose. A) Recombinant strains which have *C. trachomatis*-like growth and B) Recombinant strains with distinct growth phenotypes. ** (p-value<0.005) indicates significant difference from CT as determined by unpaired Student's t-test.

RC4243 *in vitro* attenuation does not appear to be a delay in the timing of the developmental cycle

From the initial growth screen, RC4243 grew poorly, peaking at 24 hours post-infection with nearly 2-log fold lower progeny than wildtype *C. trachomatis* (Figure 4-2B). Therefore, an extended growth curve was performed for RC4243 to determine if the attenuated growth at 21 and 24 hpi could be the result of less efficient cell division, a delay in secondary conversion or early lysis, as with the RC435 mutant. If RC4243 simply had slower cell division or secondary conversion rates, it would be expected that RC4243 would be able to reach a peak progeny level similar to *C. trachomatis,* but over an extended period of time. We investigated progeny production for RC4243 from 21-36hpi (Figure 4-3A). RC4243 was not able to successfully generate *C. trachomatis* progeny levels, indicating that the phenotype may be more complex than replication efficiency.

In order to further elucidate the reduced progeny production phenotype of RC4243, confocal imaging was used to visualize the inclusion at 24 hours post-infection (Figure 4-3B). When compared with *C. trachomatis*, RC4243 inclusions were qualitatively smaller and contained fewer progeny per inclusion. Additionally, individual bacteria appear larger in size, which, in *Chlamydia*, typically corresponds to an aberrant or persistence phenotype associated with antibiotic stress and nutrient deprivation (27, 28).

Figure 4-3. Extended growth curves for RC4243 compared with wildtype *C. trachomatis* (CT) and *C. muridarum* (CM). A) *Chlamydia* was infected at an MOI ~10 onto L929 mouse fibroblast cells. Every three hours between 21- and 36-hours post-infection, EBs were isolated by water lysis and infectious progeny was titered to calculate inclusion forming units per mL. B) Confocal imaging of parental *C. trachomatis and* RC4243 at 24hpi. Nuclei stained with Dapi for DNA, Host cytosol stained with Evan's Blue (red) and *Chlamydia* stained green.

The RC4243 region of recombination was then examined to determine potential genetic factors involved in the attenuated growth and morphology of this recombinant. RC4243 contains 24 *C. trachomatis* genes which have been replaced by *C. muridarum* orthologs along with one *C. muridarum*-specific hypothetical protein, *TC0071* (Table 4-3). One hypothesis for the poor growth of RC4243 stems from an incompatibility of *C. muridarum* genes to interact with *C. trachomatis* or host protein partners, as no *C. trachomatis*-specific genes are lost in this recombination. An analysis of the primary sequence similarity for the genes within this region could provide indicators that lowly conserved genes have evolved with specific host or chlamydial interacting

partners, and therefore, may not perform well in a recombinant strain. Of the 24 orthologs within RC4243, only one has an amino acid similarity below 75%, *TC0066*, a conserved hypothetical protein. While it is possible that this protein is responsible for the phenotype observed in RC4243, there are other possibilities as well. Primarily, through a transposon mutagenesis library, a mutant with an insertion in *CTL0065* (*CT696*) was shown to be highly attenuated *in vitro* (29). RC4243 does contain the *C. muridarum* ortholog for *CTL0065, TC0068*. These genes encode a conserved hypothetical protein and while they appear to be highly conserved between species, with amino acid similarity at 96.2%, it could be that the *C. muridarum* ortholog is unable to fully perform the function of *CTL0065* in *C. trachomatis*, resulting in a similar, albeit less severe, attenuation.

Similarly, among these orthologs is the chlamydial *mreB* (*CTL0078/TC0082*) which has been shown to be involved in cell division (30, 31). Known MreB interacting partner, RodZ (*CTL0264/TC0277*), is not included in the recombination region. One possible explanation for the fewer infectious progeny of RC4243 at 21hpi could be that replacing native *CTL0078* with *C. muridarum TC0082* reduces the efficiency of the MreB/RodZ interaction, decreasing cellular division. It has been suggested that asymmetrical division of RBs leading to reduced RB size is an essential factor in secondary conversion (32). Decreased cell division could result in a longer time to reach the size threshold needed to trigger conversion into EBs, resulting in fewer infectious progeny at earlier time points.

Table 4-3. RC4243 Region of Recombination

C. muridarum Gene	*C. trachomatis* homolog	Gene	Function	Direction [f]	% AA Identity	% AA Similarity [a]
tc0063[b]	*CTL0060* (*CT691*)		conserved hypothetical protein	+	83.1	92.4
tc0064	*CTL0061* (*CT692*)		Phosphate permease	+	88.5	94.6
tc0065	CTL0062 (*CT693*)	*pgk*	phosphoglycerate kinase	+	90.6	97.8
tc0066	*CTL0063* (*CT694*)		Conserved hypothetical protein	+	54.0	71.2
tc0067	*CTL0064* (*CT695*)		Conserved hypothetical protien	+	70.9	83.2
tc0068	*CTL0065* (*CT696*)		conserved hypothetical protein	+	90.6	96.2
tc0069	*CTL0066* (*CT697*)	*nth*	endonuclease III	-	80.4	86.0
tc0070	*CTL0067* (*CT698*)	*trmE/thdF*	tRNA modification GTPase / thiophene and furanoxidation protein	-	90.3	95.3
tc0071			hypothetical protein (40aa)	-	unique	unique[c]
tc0072	*CTL0068* (*CT699*)	*psdD*	phosphatidylserine decarboxylase	+	76.5	87.4
tc0073	*CTL0069* (*CT700*)		Conserved hypothetical protein	+	91.9	94.6
tc0074	*CTL0070* (*CT701*)	*secA*	preprotein translocase subunit	+	95.8	98.5
tc0075	*CTL0071* (*CT702*)		Conserved hypothetical protein	-	83.1	90.5
tc0076	*CTL0072* (*CT703*)	*engA*	GTP-binding protein, Era/ThdF family	-	90.6	96.5
tc0077	*CTL0073* (*CT704*)	*pcnB*	poly A polymerase	-	86.9	93.2
tc0078	*CTL0074* (*CT705*)	*clpX*	ATP-dependent Clp protease ATP-binding subunit	-	95.7	97.6
tc0079	*CTL0075* (*CT706*)	*clpP*	ATP-dependent Clp protease proteolytic subunit	-	98.0	98.5
tc0080[e]	*CTL0076* (*CT707*)	*tig*	trigger factor	-	91.0	95.2
tc0081[e]	*CTL0077* (*CT708*)		helicase, Snf2 family	+	96.2	98.2
tc0082	*CTL0078* (*CT709*)	*mreB*	rod shape-determining protein	+	98.9	99.2
tc0083	CTL0079 (*CT710*)	*pckA*	phosphoenolpyruvate carboxykinase	+	90.7	96.5
tc0084	*CTL0080* (*CT711*)		Conserved hypothetical protein	+	76.0	86.4
tc0085	*CTL0081* (*CT712*)		Conserved hypothetical type III secretion protein	+	82.9	91.8
tc0086	*CTL0082* (*CT713*)	*ompB*	Major outer membrane protein B	-	90.6	97.4
tc0087 [d]	*CTL0083* (*CT714*)	*gpsA*	glycerol-3-phosphate dehydrogenase	-	92.5	97.6

[a] As calculated by Blosum62 with a threshold of 1
[b] Margin falls within this open reading frame (723/837bp)
[c] Unique refers to genes specific to *C. muridarum* with no direct homolog in *C. trachomatis*
[d] Margin falls within this open reading frame (239/690 bp)
[e] Transposon insertion falls between these two genes

RC1696 exhibits an intermediate growth phenotype tied to the presence of six *C. muridarum* genes

One chimera, RC1696, exhibited an intermediate phenotype where secondary conversion appears to occur faster than CT, peaking around 21 hours, but not quite as early as CM (Figure 4-2). This difference was statistically significant (p-value <0.005). Of the 57 genes included within the region

of recombination for RC1696, five are unique to *C. muridarum* and may be contributing to the

accelerated growth rate (Table 4-4). Each of these are small hypothetical proteins, under 100 amino

acids, however, so exact functions are unknown. In each case, they appear to be *Chlamydia*

muridarum-specific (TC0179, TC0188, TC0191, TC0198), except for TC0220 which shows some

homology to *nusB* from *C. trachomatis. C. muridarum* has a nusB homolog (TC0219) directly

before this hypothetical protein. It is possible that this gene was acquired through some horizontal

gene transfer event, and only part of the gene was brought into CM. While this gene does contain

a complete open reading frame, it is unclear whether it is transcribed and translated.

Table 4-4. RC1696 region of recombination

C. muridarum Gene	*C. trachomatis* homolog	Gene	Function	Direction [i]	% AA Identity	% AA Similarity [a]
tc0178 [b]	CTL0165 (CT796)	*glyQ*	glycine--tRNA ligase	-		
tc0179			hypothetical protein	+	unique [c]	unique
tc0180	CTL0166 (CT797)	*pgsA*	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	+	91.6	94.1
tc0181 [d]	CTL0167 (CT798)	*glgA*	glycogen synthase	-		
tc0182 [e]	CTL0168 (CT799)	*rplY*	50S ribosomal protein L25/general stress proteinCtc	+		
tc0183	CTL0169 (CT800)	*pth*	peptidyl-tRNA hydrolase	+	80.6	88.9
tc0184	CTL0170 (CT801)	*rpsF*	30S ribosomal protein S6	+	99.1	99.1
tc0185	CTL0171 (CT802)	*rpsR*	ribosomal protein S18	+	98.9	97.6
tc0186	CTL0172 (CT803)	*rplI*	50S ribosomal protein L9	+	91.7	95.8
tc0187	CTL0173 (CT804)	*ispE*	4-diphosphocytidyl-2C-methyl-D-erythritol kinase	+	72.0	85.1
tc0188			hypothetical protein	+	unique	unique
tc0189 [f]	CTL0174 (CT805)		conserved hypothetical protein	-		
tc0190	CTL0175 (CT806)	*ptr*	insulinase/protease	-	81.8	91.9

tc0191			hypothetical protein	+	unique	unique
tc0192	CTL0176 (CT807)	*plsB*	glycerol-3-phosphate acyltransferase	-	89.5	94.3
tc0193	CTL0177 (CT808)	*cafA*	ribonuclease G	-	97.7	98.6
tc0194	CTL0178 (CT809)		conserved hypothetical protein	+	77.4	86.3
tc0195	CTL0181 (CT810)	*rpmF*	ribosomal protein L32	+	93.3	95.0
tc0196	CTL0182 (CT811)	*plsX*	fatty acid/phospholipid synthesis protein	+	85.7	92.9
tc0197	CTL0183 (CT812)	*pmpD*	polymorphic membrane protein D	+	71.9	83.3
tc0198			hypothetical protein	-	unique	unique
tc0199	CTL0184 (CT813)		inclusion membrane protein	+	38.0	55.3
tc0200	CTL0185 (CT814)		conserved hypothetical protein	-	92.0	97.0
tc0201	CTL0186 (CT814.1)		conserved hypothetical protein	-	78.5	89.3
tc0202	CTL0187 (CT815)	*glmM*	phosphoglucosamine mutase	+	93.9	97.0
tc0203	CTL0188 (CT816)	*glmS*	glucosamine--fructose-6-phosphate aminotransferase	+	85.7	92.3
tc0204	CTL0189 (CT817)	*tyrP*	tyrosine-specific transport protein	+	91.5	96.7
tc0205	CTL0190 (CT818)		tyrosine-specific transport protein	+	86.1	93.7
tc0206	CTL0191 (CT819)		conserved hypothetical protein	+	93.3	97.5
tc0207	CTL0192 (CT820)	*ftsY*	signal recognition particle receptor	-	88.1	94.0
tc0208	CTL0193 (CT821)	*sucC*	succinyl-CoA ligase subunit beta	+	88.1	94.8
tc0209	CTL0194 (CT822)	*sucD*	succinyl-CoA ligase subunit alpha	+	90.8	93.8
tc0210	CTL0195 (CT823)	*htrA*	serine protease	+	92.8	97.0
tc0211	CTL0196 (CT824)		Zn-dependent peptidase	+	91.0	95.9
tc0212	CTL0197 (CT825)	*rmuC*	DNA recombination protein	-	86.2	93.0
tc0213	CTL0198 (CT826)	*pssA*	CDP-diacylglycerol--serine O-phosphatidyltransferase	-	93.9	95.8
tc0214	CTL0199 (CT827)	*nrdA*	ribonucleoside-diphosphate reductase, alpha subunit	+	93.2	96.6
tc0215	CTL0200 (CT828)	*nrdB*	ribonucleoside-diphosphate reductase subunit beta	+	96.3	98.9
tc0216	CTL0201 (CT829)	*trmB*	tRNA (guanine-N(7)-)-methyltransferase	+	83.6	92.9

TC_t11	CTL_t36 (CT_t35)	*tRNA-Val*	tRNA-Val-1	-	100	100
tc0217	CTL0202 (CT830)		rRNA methylase	+	80.0	90.8
tc0218	CTL0203 (CT831)	*murB*	UDP-N-acetylenolpyruvoylglucosamine reductase	-	84.2	92.8
tc0219	CTL0204 (CT832)	*nusB*	transcription termination factor	-	81.7	85.2
tc0220			hypothetical protein	-	unique	unique
tc0221	CTL0205 (CT833)	*infC*	translation initiation factor IF-3	+	96.3	98.9
tc0222	CTL0206 (CT834)	*rpmI*	50S ribosomal protein L35	+	100	100
tc0223	CTL0207 (CT835)	*rplT*	ribosomal protein L20	+	93.5	98.4
tc0224	CTL0208 (CT836)	*pheS*	phenylalanine--tRNA ligase subunit alpha	+	91.0	96.8
TC_t12	CTL_t27 (CT_t36)	*tRNA-Ser*	tRNA-Ser-1	+	100	100
tc0225	CTL0209 (CT837)		conserved hypothetical protein	+	86.3	93.2
tc0226	CTL0210 (CT838)		conserved hypothetical protein	-	94.3	98.1
tc0227	CTL0211 (CT839)		conserved hypothetical protein	-	94.9	96.9
tc0228	CTL0212 (CT840)	*tilS*	tRNA(Ile)-lysidine synthase	+	79.2	87.6
tc0229	CTL0213 (CT841)	*ftsH*	ATP-dependent zinc metalloprotease, cell division protein	+	94.7	97.4
tc0230	CTL0214 (CT842)	*pnpA*	polyribonucleotide nucleotidyltransferase	-	91.2	96.8
tc0231	CTL0215 (CT843)	*rpsO*	ribosomal protein S15	-	100	100
tc0232 [g]	CTL0216 (CT844)		nucleoside deaminase	+		

[a] As calculated by Blosum62 with a threshold of 1

[b] Margin falls within this open reading frame (1143/3012 bp)

[c] Unique refers to genes specific to *C. muridarum* with no direct homolog in *C. trachomatis*

[d] Margin falls within this open reading frame (408/1425 bp)

[e] Margin falls within this open reading frame (490/558 bp)

[f] Transposon insertion within this open reading frame (1042/1353 bp)

[g] Margin falls within this open reading frame (197/492 bp)

[i] Dotted outlines signify Midlate expressed genes, dashed outlines indicate Late expressed genes, as reported by Nicholson et al, gray shading represent genes with TSSs identified by Albrecht et al (33)

In an effort to narrow down which genes may be responsible for this growth phenotype, RC1696 was directly compared with RC1650. RC1650, one of the fourteen chimeras which have *C. trachomatis*-like growth (Figure 4-2A), shares almost half (37 genes) of the region of recombination with RC1696 (Figure 4-4). Because this recombinant displays no growth phenotype in our previous studies, it can be concluded that genes within the region of recombination of RC1696 but not RC1650, are responsible for the intermediate growth phenotype of RC1696. This reduces candidate genes to 20 (Figure 4-4, Table 4-4). In order to reduce the pool of 20 genes further, additional recombinants from the initial library (26) were investigated. Four additional recombinants were discovered which overlap regions of recombination with RC1696 referred to as subclones (Figure 4-4).
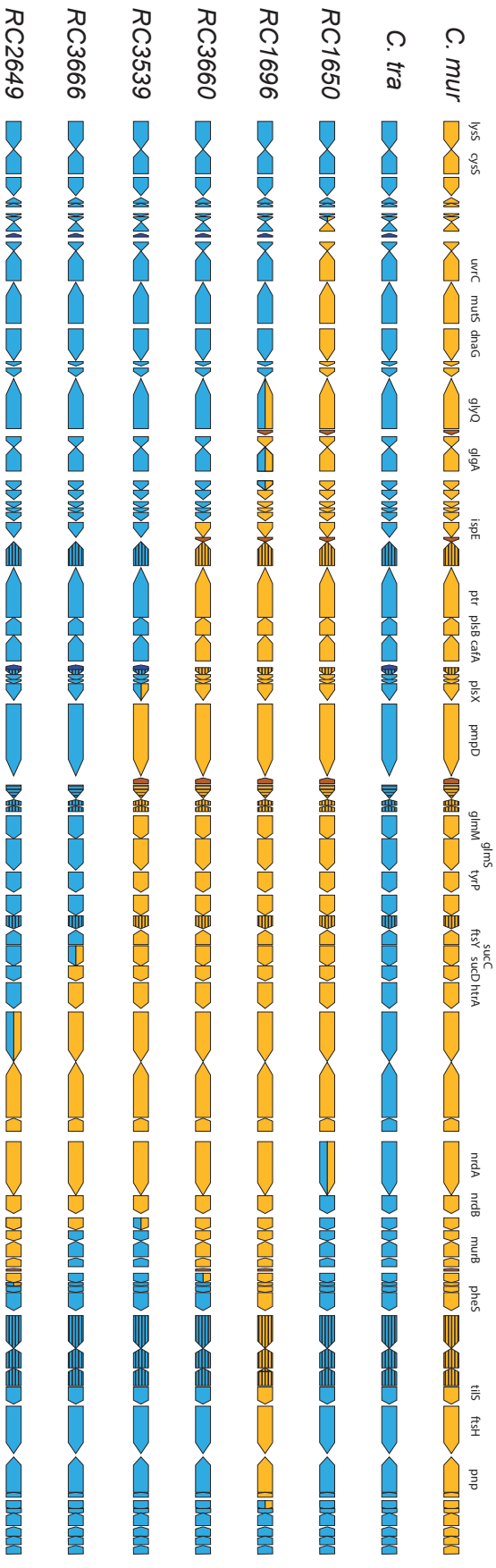
Figure 4-4. Gene map of the region of recombination for RC1650 and RC1696. RC3660, RC3539, RC3666 and RC2649 are subclone recombinants containing smaller regions of recombination at the same site, allowing for analysis of subsets of genes for growth phenotypes. Blue regions represent *C. trachomatis* genes while orange genes were recombined from *C. muridarum*. Select gene names are included across the top. Dark blue and dark orange genes are species-specific hypothetical genes. Horizontal striped genes represent conserved hypothetical proteins and vertical striped genes represent lowly conserved genes.

Two of these recombinants, RC3660 and RC3539, were also investigated for growth phenotypes using the three-time point growth curve. As in Figure 4-2, RC1650 grew similar to *C. trachomatis*, while RC1696 had increased 21h progeny (Figure 4-5A). Subclone RC3539 growth mirrored *C. trachomatis* as well, indicating that genes within RC3539 are likely not contributing to the phenotype seen in RC1696. In contrast, RC3660 growth was similar to RC1696, indicating that products within the region of recombination shared by both RC1696 and RC3660 are responsible for the phenotype. This reduced candidate genes to six; TC0216 to TC0221 (Table 4-4, Figure 4-4).

Through the growth curve experiments, RC1696 and RC1650 inclusions were noticeably smaller than those of either *C. muridarum* or *C. trachomatis* (Figure 4-5C). We were able to quantify these differences using MIPAR image analysis software to calculate inclusion area at 24 hours post infection (Figure 4-5B). *C. muridarum* inclusions at 24 hpi are significantly larger than *C. trachomatis*, which correlates with the accelerated growth rate. As expected, RC1696 and RC1650 inclusions were both statistically smaller, with the average inclusion size nearly half those of *C. trachomatis*. Interestingly, despite the accelerated growth rate at 21hpi of RC1696, there was not a difference between RC1650 and RC1696 in inclusion size. We investigated the subclones for the reduced inclusion size phenotype of RC1696 and RC1650 (Figure 4-5C). None of the subclones were significantly smaller than *C. trachomatis*.
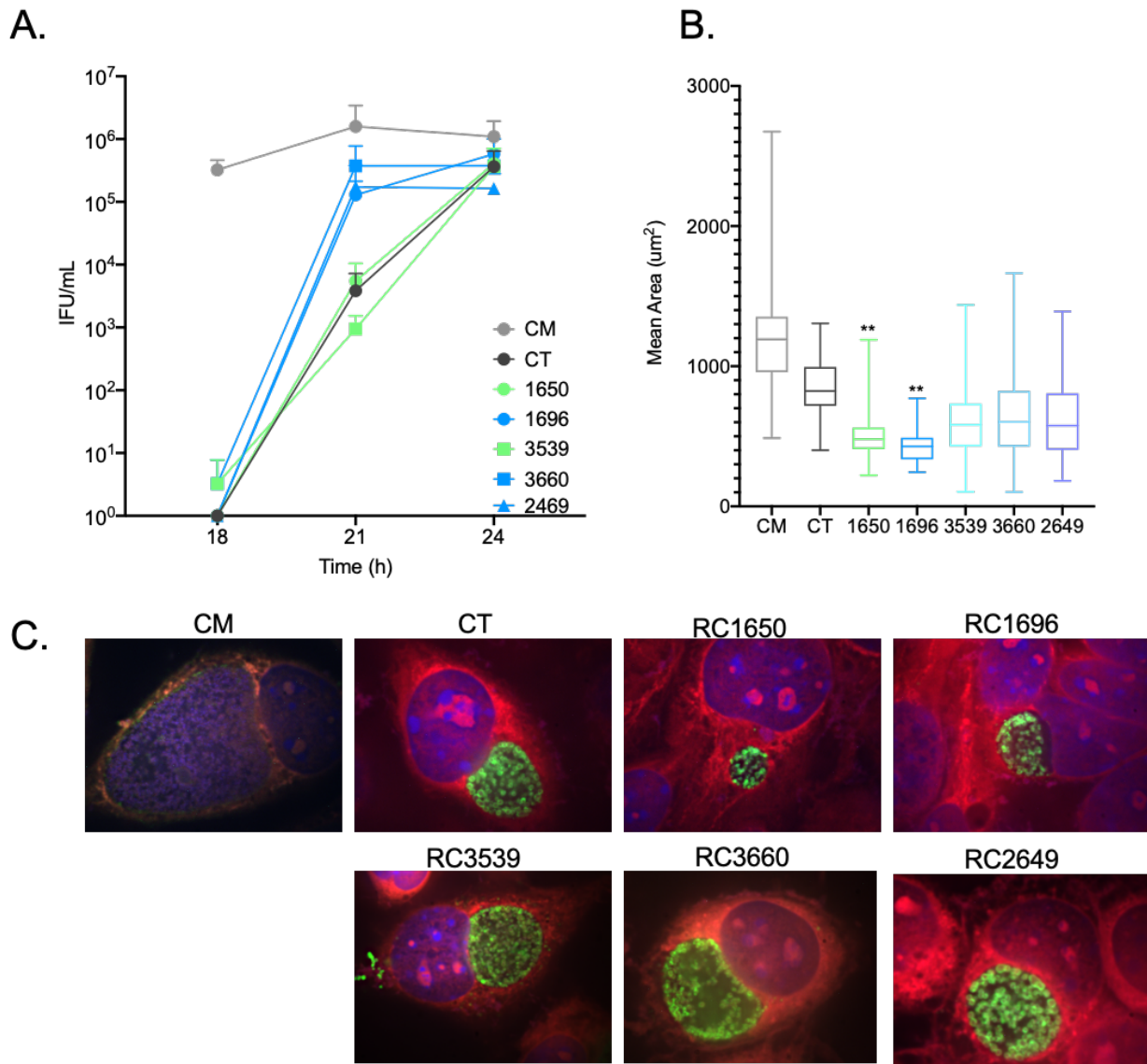
Figure 4-5. Growth characteristics for RC1696 subclones. A) Three timepoint growth curves for recombinants compared with wildtype *C. trachomatis* (CT) and *C. muridarum* (CM). Recombinants were infected at an MOI ~1 onto L929 mouse fibroblast cells. At 18, 21- or 24-hours post-infection, EBs were isolated and infectious progeny was titered to calculate inclusion forming units per mL. Data represents three individual experiments each in triplicate. B) Inclusion area was measured at 24 hpi at an approximately 50% infection in triplicate (n>50 inclusions per replicate). Statistical comparisons were done using an unpaired Student's t-test between samples. ** indicates significant difference to wildtype *C. trachomatis* (p-value <0.0005). C) Confocal imaging of RC1696, RC1650 and subclones along with the parent strains. Images were taken at 24 hours post-infection. Nuclei stained with Dapi for DNA, Host cytosol stained with Evan's Blue (red) and *Chlamydia* stained green.

***C. muridarum* genes were not sufficient in the regions of recombination to provide an advantage to *C. trachomatis* in the mouse model**

One of the major questions within the field is about the host-specific pathogenesis of *Chlamydia*. A major aspect of this pathogenesis is the ascension phenotype where *Chlamydia* disseminates from the lower reproductive tract to the organs of the upper reproductive tract such as the uterus and fallopian tubes. While fourteen of the recombinants showed no major defects *in vitro* in our initial screen, *C. muridarum* genes could be relevant for *in vivo* phenotypes and host specificity not seen in cell culture. Success in the *in vivo* environment requires many processes from acquiring the necessary nutrients, competing with microflora, evading the immune system and infecting a range of cell types. Therefore, each of the recombinants, including those with *in vitro* phenotypes (RC4243, RC435 and RC1696) were investigated in the mouse model using C57BL/6 mice.

Mice were infected intravaginally with either recombinant or parental strains with approximately 1-5 x 10^5 infectious forming units. Seven days post-infection bacterial burdens were analyzed for lower reproductive tracts (vaginal vault) and upper reproductive tracts (uterine horns). In the lower genital tract, *C. trachomatis* was able to establish a low burden infection, nearly 2.5 log lower burdens than seen in *C. muridarum*-infected vaginal vaults (Figure 4-6A). The majority recombinant-infected genital tracts had burdens at or below *C. trachomatis* levels, indicating that the *C. muridarum* genes within these regions of recombination were not sufficient to increase success in the lower genital tract of the murine host. For one chimera, RC027, half (n=3) of the mice infected had almost ten times higher bacterial burdens than *C. trachomatis*. While this was not statistically significant, due to half of the samples being below the threshold for detection, when we remove the lower subset, these would be significantly different from *C. trachomatis* by

unpaired Student's t-test (p-value = 0.0016). None of the recombinants were able to ascend to the upper reproductive tract and no chlamydia genome copies were detected in the uterine horns for *C. trachomatis* nor the recombinants (data not shown).

Because RC027 appeared to have two populations within the mice, a higher burden and a lower burden, we wanted to ensure that there was not a mixed population within this strain. First, the original genome sequencing was used to scan for evidence of a mixed population in the original cultures. No mixed sequencing was evident, suggesting that the original population was clonal. Second, PCR was performed for the recombination region in RC027 after isolation from the mouse to ensure the population remained clonal through the infection. PCRs confirmed a clonal population in all mice, as no *C. trachomatis* or *C. muridarum* –specific sequence was detected in the RC027-infected vaginal vaults. Based on these, it does not appear that the increased bacterial burden population is due to contamination with *C. muridarum*. However, it is possible that polymorphisms were acquired *in vivo* that allowed RC027 to be more successful in the mouse.
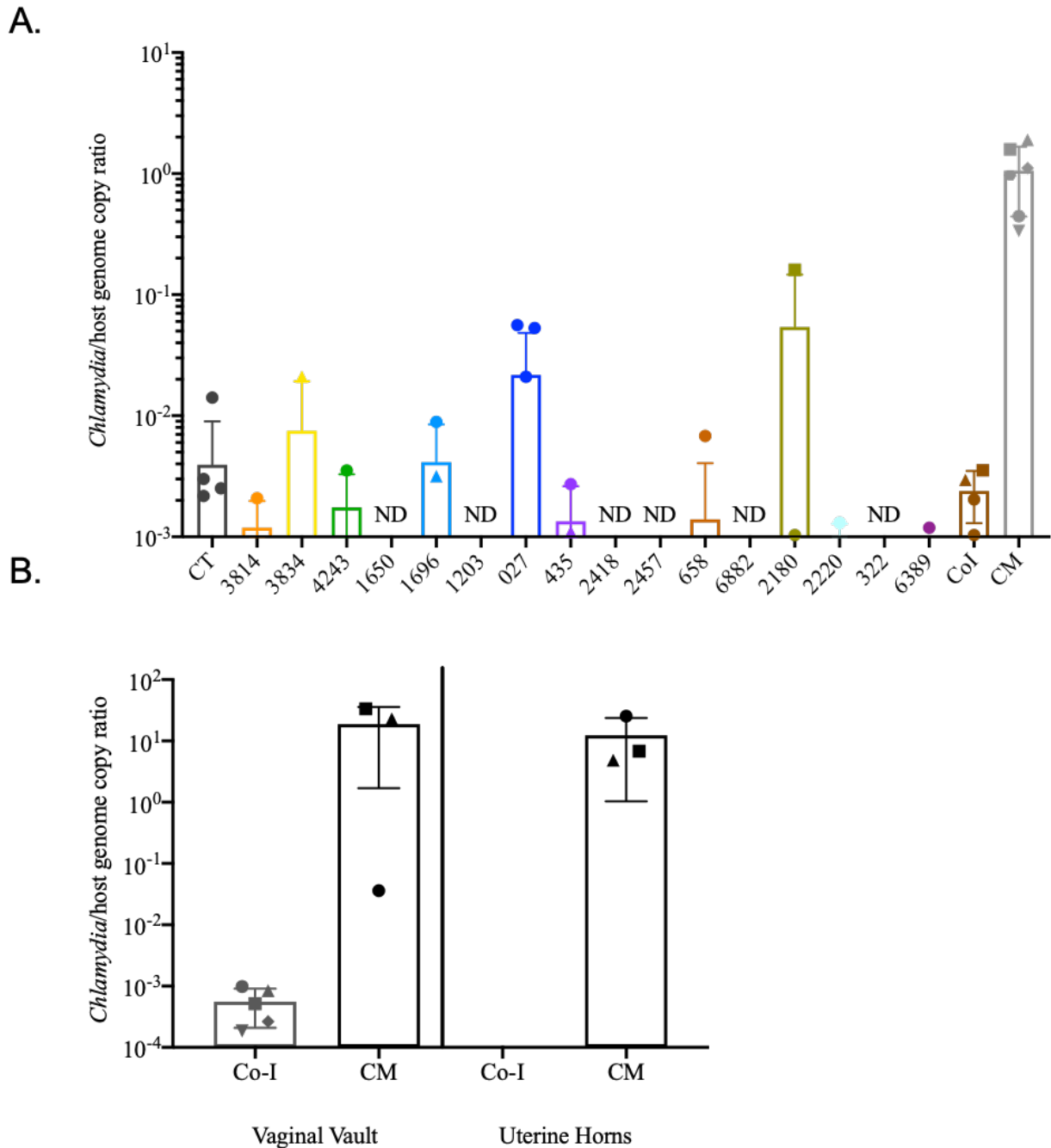
Figure 4-6. Mouse models for the recombinant strains. A) C57Bl/6 mice were infected with 105 IFU intravaginally. Seven days post-infection lower genital tracts were harvested and bacterial burdens were assessed using ddPCR to quantitate the ratio of chlamydial genome copies to host genome copies in the specific tissue. ND indicates not detected above mock-infection levels by ddPCR. CoI indicates the co-infection of each chimera at an infectious dose of 10^4 of each. Neither *C. trachomatis* nor any recombinant was detected in the upper reproductive tract. B) More susceptible C3He/J mice were infected intravaginally with either wildtype *C. muridarum* or a co-infection with all chimeras at the same dose as in A). Five days post-infection organs were harvested and analyzed by ddPCR.

None of the genes in any region of recombination were independently sufficient for *C. muridarum* bacterial burden levels in the lower reproductive tract nor ascension. It could be possible that these phenotypes are regulated by a variety of factors. Importantly, the chlamydial genome is relatively fractured compared with other bacterial species which are often organized into distinct operons of genes with related functions. In *Chlamydia*, related genes are not always organized into discrete operons and therefore, may not be near each other and brought into the same recombinant. Because of this, recombinant strains may be able to supplement *C. muridarum* genes missing from their region of recombination with another recombinant. To test this hypothesis, we infected mice with a co-infection of all recombinants at an infectious dose of $1x10^5$ IFU of each strain for a total dose of $1.7x10^6$ IFU (Figure 4-6A). No additive benefits were seen in the bacterial burdens in the lower genital tract of these mice and no ascension was detected. C57BL/6 mice are known to be less susceptible to infection than C3He/J mice (34), so we repeated these studies in the more susceptible mouse model. In C3He/J mice the results mirrored those in C57BL/6, however. The co-infection was unable to overcome the barriers to infection in the mouse and established only a low burden vaginal infection with no ascension (Figure 4-6B).

Discussion

In this study, we have used a chromosomal *C. trachomatis* x *C. muridarum* recombination library (26) in an initial screen to investigate the role of *C. muridarum* genes *in vitro* and *in vivo*. The initial three time-point growth screen analyzed the sixteen recombinants ability to undergo the developmental cycle including adhesion, invasion, establishment of the inclusion, and ability to convert between forms. Thirteen chimeras had growth phenotypes similar to *C. trachomatis*, while three recombinants, RC435, RC4243 and RC1696 each exhibited a unique growth phenotype.

The majority of recombinants did not appear to have *in vitro* growth phenotypes when compared with *C. trachomatis*. This could indicate that for the majority of the genome, *C. muridarum* genes are direct and complete functional homologs for their counterpart *C. trachomatis* genes. The possibility of this is not unlikely, due to the strict intracellular lifestyle, however, another possibility is that our cell culture screen was not able to capture the nuanced differences between *C. trachomatis* and *C. muridarum* gene functions. One major caveat of this screen was that these were performed only in immortalized mouse cells. A more expansive screen could include similar growth curves in primary cell lines, human and mouse vaginal epithelium. These studies could provide a way to detect species-specific phenotypes, along with more a more detailed look at the interactions between *Chlamydia* and host.

For RC4243, poor progeny production and aberrant-like morphology through 36 hpi suggests that there is an incompatibility within the region of recombination, and that *C. muridarum* genes are not able to perform the full function of the *C. trachomatis* ortholog. There are 25 genes within the recombination region for RC4243, including one *C. muridarum*-specific hypothetical protein (*TC0071,* Table 4-2). This unique hypothetical gene is only 40 amino acids in length, and it is possible that this gene is a pseudogene. We analyzed the upstream region for *TC0071* to look for putative ribosomal binding sites (RBS), as these can add support for the translation of mRNA made from this open reading frame. Ribosomal binding sites can be identified through analysis of the immediate 5' untranslated region (UTR) and identification of possible Shine-Dalgarno (S-D) sequences (35). The Shine-Dalgarno sequence is complementary to the 3' end of the 16S rRNA, which for prokaryotes, including *C. muridarum* has a conserved CCUCCU sequence (35).

Canonical S-D sequences, therefore are permutations of the complementary AGGAGG sequence. In *Chlamydia trachomatis*, these sequences have included AGGTG (36), AGGAG (37), and GAGGT (38). In the immediate 5' UTR for *TC0071,* we identified a possible RBS, although not the canonical sequence, which lends support to the correct annotation of TC0071 as a true coding sequence (Figure 4-7). It is possible that the addition of this gene into the *C. trachomatis* genome is creating the poor growth phenotype, however a more likely explanation is that the recombination has effectively deleted a *C. trachomatis* gene from the genome and replaced it with a *C. muridarum* gene unable to perform its function.

```
TC0071
TATACTTTTGGAGGAACTTTCTAAGCAAGAAAGGTAGATTAGGGAGAGTA 50
CGAGGTCTAGTTTGCTCCTTTGTTATTATAAATTTTTGAAATAGAATATT 100
TGGTAAATCGAGCGATGTTAGGCTGAAGGGATTTTTGAGAAAGGGGAACT 150
TCTTTATAGTTCCTTGTGTTTCGCATGTAGCAGACCTTCGCGAGAAGAGC 200


TC0179
AAGTTCCACAGGGATTAAGGAATATCTAGCATTCCTAGAGGAACATACTA 50
GTGGAAGCTATATTAATTGAACAAGGCGAGAAATATCCTGTTGATGAGGA 100
GAGCACAGTCTTATTAATGAGTTGTACTGTTGATAGGTTTTTTGAAAGTT 150
TTATGGGATCCTATTTATTCTGAATTCCTGACTGGAAAAACTAGCGATAA 200
ATTCTAGAAATGCTAAGAATGAGAGAACTAAGCAAGTGAGCGTTCTCGAG 250


TC0220
TATACCAAGAGTCATTTTTATGAAAGAGTAGACCTACTTATCTGTGATTG 50
ATTTGAAAATAAAAAATTTCTTGTCGAAGACTTCTCTTGAATAAGTTTAT 100
AGAATTCTTTTGAGGGAAAAAGGAGCGAAGTTTTCGGCGCAATACTTGAT 200
AATAATCCTTATCTATAGGTACCATGTAGCAGTCTGGAGCGGAGTCGTTT 250
```
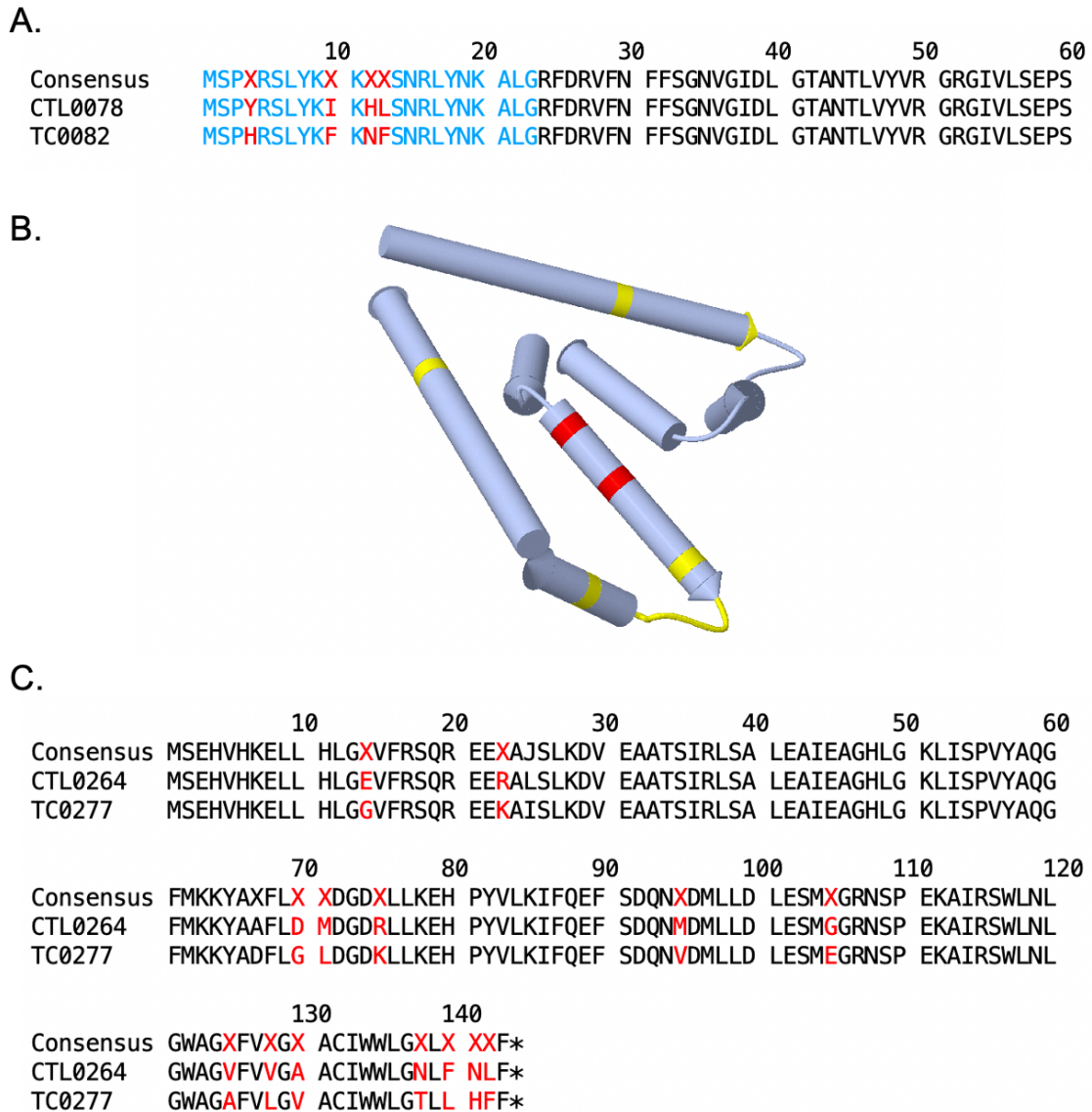
Figure 4-7. *C. muridarum* specific unique hypothetical proteins within the regions of recombination for RC4243 (*TC0071*) and RC1696 (*TC0179* and *TC0220*). Nucleotide sequence for each putative coding region is in red with the translational start site indicated (bold, underline). Putative ribosomal binding sites have been predicted by analyzing the sequence directly upstream of the start codon and are indicated in blue.

Candidate genes were suggested to be *TC0066*, a conserved hypothetical protein with amino acid similarity between species of 71.2%. This suggests that there are major differences between the *C. muridarum* and *C. trachomatis* orthologs and could indicate that these genes are involved in interacting with the host (human or mouse). Structural differences in protein sequence may suggest structural difference in cognate partners, which would be expected between human and mouse host proteins. Additionally, a likely target for the poor growth of RC4243 is the replacement of *CTL0065* with *TC0068*. CTL0065 (CT696) has been shown through transposon mutagenesis to be essential for proper growth and development of *C. trachomatis* (29). There is around 96.2% similarity between *C. muridarum* and *C. trachomatis*, but again, small changes could result in a similar, though less severe, *in vitro* attenuation.

Other candidate genes include *mreB*, which is important in chlamydial cell division. This protein is highly conserved between *C. muridarum* and *C. trachomatis*, however, with primary sequence similarity at 99.2% (Figure 4-8A). While it may seem less likely that such a highly conserved protein would not be interchangeable between species, it is important to remember that even single amino acid substitutions can be enough to render a protein ineffective. An example of this is with internalin A protein from *Listeria monocytogenes*, where a single amino acid substitution results in InlA binding human E-cadherin, but not mouse E-cadherin (39). Similarly, it is possible that small protein changes are having large effects on protein function for MreB. The region with the amino acid differences in MreB corresponds with a unique N-terminal extension found in chlamydial MreB homologs. This amphipathic extension region has been demonstrated to be essential for membrane localization (40). Intriguingly, this region was not important in homo-oligomerization of MreB, but was important for MreB interaction with both RodZ and FtsK (40).

RC4243 contains only the *C. muridarum mreB* homolog, retaining the *C. trachomatis* RodZ and FtsK proteins. Therefore, the amino acid differences in the *C. muridarum* MreB (Figure 4-8A) could be disrupting the normal MreB/RodZ or MreB/FtsK interaction. We examined the *C. trachomatis* RodZ structure (41) and compared the amino acid sequence to *C. muridarum* RodZ to see if there could be critical contact sites that are different between *C. trachomatis* and *C. muridarum*. There were 7 differences in the cytoplasmic domain for RodZ (Figure 4-8B), including one on the helix shown to be critical for MreB-interaction, although not at previously demonstrated contact sites (41). There were an additional 7 amino acid differences in the transmembrane C-terminal region of RodZ, as well (Figure 4-8C), although these are less likely to be important in MreB interaction. The aberrant-like large reticulate bodies in the confocal imaging of RC4243 could support that RC4243 is not able to efficiently undergo division leading to a reduced ability to convert to infectious EBs and a general reduction in infectious progeny at all time points.

A.

```
                              10         20         30         40         50         60
Consensus  MSPXRSLYKX KXXSNRLYNK ALGRFDRVFN FFSGNVGIDL GTANTLVYVR GRGIVLSEPS
CTL0078    MSPYRSLYKI KHLSNRLYNK ALGRFDRVFN FFSGNVGIDL GTANTLVYVR GRGIVLSEPS
TC0082     MSPHRSLYKF KNFSNRLYNK ALGRFDRVFN FFSGNVGIDL GTANTLVYVR GRGIVLSEPS
```

B.

C.

```
                              10         20         30         40         50         60
Consensus  MSEHVHKELL HLGXVFRSQR EEXAJSLKDV EAATSIRLSA LEAIEAGHLG KLISPVYAQG
CTL0264    MSEHVHKELL HLGEVFRSQR EERALSLKDV EAATSIRLSA LEAIEAGHLG KLISPVYAQG
TC0277     MSEHVHKELL HLGGVFRSQR EEKAISLKDV EAATSIRLSA LEAIEAGHLG KLISPVYAQG

                              70         80         90        100        110        120
Consensus  FMKKYAXFLX XDGDXLLKEH PYVLKIFQEF SDQNXDMLLD LESMXGRNSP EKAIRSWLNL
CTL0264    FMKKYAAFLD MDGDRLLKEH PYVLKIFQEF SDQNMDMLLD LESMGGRNSP EKAIRSWLNL
TC0277     FMKKYADFLG LDGDKLLKEH PYVLKIFQEF SDQNVDMLLD LESMEGRNSP EKAIRSWLNL

                             130        140
Consensus  GWAGXFVXGX ACIWWLGXLX XXF*
CTL0264    GWAGVFVVGA ACIWWLGNLF NLF*
TC0277     GWAGAFVLGV ACIWWLGTLL HFF*
```

Figure 4-8. A) Clustal omega alignment of *C. trachomatis* MreB (CTL0078) and *C. muridarum* MreB ortholog (TC0082), residues 1-60. These orthologs share 99.2% amino acid similarity, only differing at four residues (red) within the extended N-terminal region (blue). B) Predicted structure for known MreB-interacting partner, RodZ (cytoplasmic domain, residues 1-103). *C. trachomatis* L2/434 RodZ protein structure model with residues demonstrated to be important for MreB interaction (red). Yellow residues indicate those different in the *C. muridarum* RodZ ortholog. C) Clustal omega alignment of *C. trachomatis* RodZ (CTL0264) and *C. muridarum* RodZ ortholog (TC0277). These orthologs share 93.1% amino acid similarity with differences highlighted in red.

These candidate genes provide a basis for future studies, however, the possibility that any one of the other 21 genes may be involved in this phenotype. Future studies could be performed to determine the gene or genes involved in this phenotype. One such study is the investigation into subclones for RC4243, as was performed with RC1696. There are four subclones within the region of recombination for RC4243 from the library published by our group (26). These subclones are restricted to a small region of recombination however, only extending from *TC0076* to *TC0083*, spanning eight genes of the 25 in RC4243 (Table 4-3). The restricted recombination length of these subclones could suggest that the genes responsible for the poor growth phenotype are not within these subclones, as larger regions of recombination produce "sick" progeny like RC4243. Within these subclones is *mreB,* however, so examination of the subclones could provide evidence for the involvement, or lack thereof, of *mreB*.

An additional way to determine the gene(s) involved in the RC4243 growth phenotype could be to perform a backcross experiment, where we use a selectable marker near the region of recombination for RC4243, such as a transposon mutant from our library (29), and use homologous recombination to generate a cross within the RC4243 region of recombination to rescue RC4243 growth. Sequencing of the rescued mutant could provide information about which *C. muridarum* genes were retained in the recombinant, and therefore, which genes are not involved in the poor growth of RC4243. A possible experiment is outlined in Figure 4-9. Finally, directed mutagenesis of candidate genes could be used to pinpoint genes important for growth within this region. Determining incompatible *C. muridarum* genes could provide important information about how *Chlamydia* have evolved with their specific host and chlamydial basic biology. Importantly, there are many conserved hypothetical proteins within the region of recombination for RC4243, and it

is possible that these *Chlamydia*-specific proteins are involved in the developmental cycle of *Chlamydia* and studies such as these can help tease out the functional role of these unique proteins.
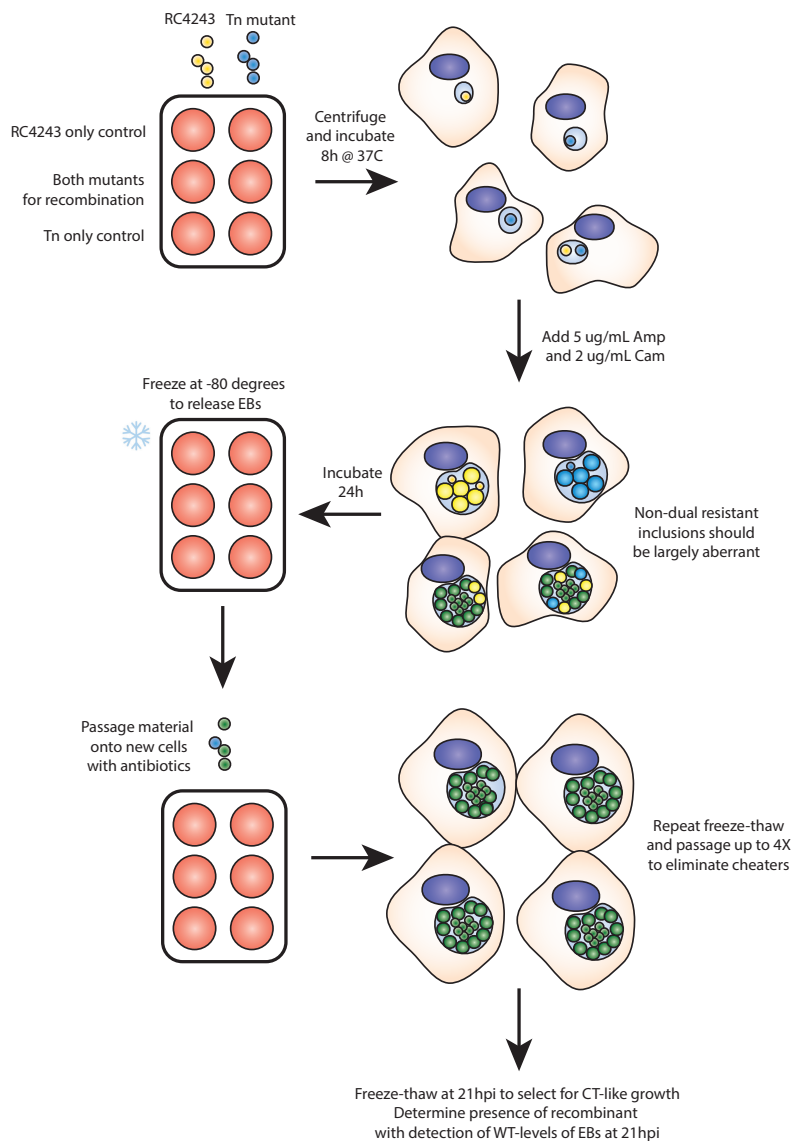
Figure 4-9. Schematic for the recombination experiments to restore wildtype CT-like growth for the RC4243 mutant. RC4243 and the TN mutant will be infected onto a 6-well plate at a MOI around 1, where two wells will get both mutants and the other wells will be infected with one or the other. Plate will be centrifuged to promote invasion and incubated for 1 hour at 37C to establish an infection. After 1 hour, 5 ug/mL of ampicillin and 2 ug/mL of chloramphenicol will be added to encourage recombination as RC4243 contains a CAT marker for chloramphenicol resistance and the TN mutant contains a bla marker for ampicillin resistance. After 24 hours of infection with antibiotics, non-recombinant *Chlamydia* should be aberrant, although cheaters will occur. The plate will be freeze-thawed at -80C to lyse host cells, releasing infectious EBs. Infectious material will be passaged to a new 6-well plate at a MOI of approximately 1 and treated with both antibiotics. At 21hpi, the plate will be freeze-thawed and passaged to a new plate. Concurrently, infectious material will also be infected onto a 96-well plate to titer infectious material. This process will continue for up to 4 passages until the recombinant wells have CT-like progeny at 21hpi, as determined by similarity to the TN control well.

Another interesting phenotype discovered in our initial screen is the intermediate growth of RC1696. Conversion of infectious progeny appears to happen faster in RC1696, as inclusion forming units reach peak levels three hours faster than in *C. trachomatis* (Figure 4-2B). Use of subclones and overlapping regions in wildtype-growing RC1650, allowed for narrowing down of the subset of candidate genes to six. Within the shared region of recombination, outside of the six candidate genes, is a tRNA for valine. Because this tRNA is completely conserved between *C. trachomatis* and *C. muridarum*, it is unlikely that this component would be involved in the phenotype.

However, that still leaves six possible genetic candidates. First, *TC0216* is a tRNA methyltransferase, *trmB*, a conserved family of transcription factors first discovered in archaea (42). Not much is known functionally about this regulator in bacteria, however studies suggest that it plays a role in the oxidative stress response in *P. aeruginosa* (43) and coordinate expression of metabolic pathways in *Halobacterium salinarum* (44). It is possible that structural differences between *C. muridarum* and *C. trachomatis* TrmB lead to differences in growth due to changes in expression of key genes. The second gene candidate is *TC0217*, a class I rRNA methyltransferase with similarity to conserved bacterial protein, RsmH (E-value 1.14e-03) which methylates 16S rRNA (45). In *Staphylococcus aureus* RsmH was demonstrated to be a putative virulence factor, as deletions of *rsmH* resulted in reduced virulence and increased sensitivity to oxidative stress (46). Again, like *trmB*, it is possible that *TC0217* is contributing to the increased progeny at 21 hpi of RC1696 through increased resistance to stress or another mechanism.

A third promising candidate gene, *TC0218*, *murB*, is a UDP-N-acetylmuramate dehydrogenase. In *Bacillus subtilis murB* is critical for peptidoglycan biosynthesis and considered essential. Interestingly, a conditional knockout of *murB* was shown to have almost three times the doubling time of wildtype *B. subtilus* when *murB* expression was reduced (47). *Chlamydia* do contain peptidoglycan, like *B. subtilis*, and again, it is possible that *C. muridarum* MurB is more efficient than *C. trachomatis murB*, resulting in faster growth rates, making *murB* a promising candidate for the faster growth of RC1696. The fourth gene of interest is *TC0219*, a *nusB* homolog, which is a transcription termination factor critical in most bacteria for Rho-dependent termination (48). The fifth gene, *TC0220*, is a small *C. muridarum*-specific hypothetical protein, 46 amino acids, with no similarity outside *Chlamydia* by protein or nucleotide BLAST. Within *Chlamydia* protein BLAST suggests that this protein may be related to *nusB* (*TC0219*), and although they are in the same direction, transcriptional start site data in *C. trachomatis* suggests that these may not be in an operon, as *nusB* in *C. trachomatis* has its own start site (33). As with *TC0071*, in RC4243, this unique hypothetical protein could be a pseudogene, perhaps an artifact of a gene related to *nusB*. In order to look at the possibility of *TC0220* translation, we again, identified a putative RBS for this open reading frame. Within the 5'UTR is a putative RBS, although it is a modified version of the canonical S-D sequence (Figure 4-7). However, it is possible that this sequence is acting as an RBS, perhaps at less efficiency than genes with a canonical. AGGAGG site. These permutations of the S-D sequence could act in altering protein expression as an alternative regulation mechanism. Finally, the sixth protein shared with the intermediate growing recombinants, *TC0221*, appears to be conserved bacterial translation initiation factor IF-3 (*infC*), one of three components required for bacterial protein synthesis.

It would not be surprising that the slightly accelerated growth of RC1696 could be mediated by transcriptional or translational machinery. Subtle changes to structure in these highly conserved proteins could increase efficiency of these processes, which would allow for faster replication, division or conversion. Further investigation into determining the exact gene or genes involved in this phenotype could begin to explain why the *C. muridarum* eclipse period is shorter than that of *C. trachomatis*. Additionally, it has been hypothesized that growth rate of *C muridarum* may provide an explanation for the ability to ascend. While the evidence for ascension of *C. trachomatis* in humans would suggest that this is not the case, there may still be advantages to the faster growth rate in the murine host which are not seen in the human host. Importantly, though, RC1696 did not have an advantage over *C. trachomatis*, nor ascend, in the mouse model, despite the growth phenotype *in vitro*.

It was expected that increased infectious progeny might be correlated with larger inclusion size, as *C. muridarum* inclusions are significantly larger than *C. trachomatis* throughout the developmental cycle (29). Interestingly, the increased 21-hour progeny production of RC1696 did not appear to correlate with inclusion size (Figure 4-5). Both RC1696 and RC1650 inclusions were significantly smaller than *C. trachomatis*, despite differences in growth, suggesting that the increased progeny phenotype and decreased inclusion size phenotype are independent. When the subclones were investigated for inclusion size, all three were *C. trachomatis*-like. There are nine genes shared in the region of recombination of RC1696 and RC1650, but not in the subclones (Table 4-4, Figure 4-4). Within these nine genes is one unique *C. muridarum* hypothetical protein (*tc0179*, 49aa), which has no similarity to a protein in any other species by BLAST homology. Once again, we investigated this unique hypothetical protein for putative S-D sequences to

determine the likelihood that this open reading frame produces translatable protein. A known chlamydial RBS, AGAGG, was identified approximately 10 bases upstream on the translational start codon (Figure 4-7). This RBS provides strong evidence that *TC0179* is not a pseudogene, despite its small size of only 49 amino acids. It is possible that TC0179 presence in *C. trachomatis* recombinants results in the reduced inclusion size, although, because *C. muridarum* inclusions are larger than *C. trachomatis* on average, this seems unlikely. Another possibility is that TC0179 interacts with another *C. muridarum* protein and that interaction, which is normally absent in *C. trachomatis* is interfering with normal *C. trachomatis* inclusion maintenance functions. Alternatively, one of the other nine genes may be responsible for this phenotype. The other eight genes include four highly conserved ribosomal proteins, which likely are not contributing to a reduced inclusion size, a tRNA-ligase, a phosphatidyl transferase, a tRNA hydrolase and a glycogen synthase. Of these, the glycogen synthase, *glgA*, may be the most likely candidate for the reduced inclusion phenotype. *GlgA* has been shown to be secreted into the inclusion lumen and host cytosol (49) and is thought to interact with host proteins (50). Further investigation into the reduced progeny phenotype could reveal a role for GlgA in the inclusion lumen.

One unexpected result from this initial examination of the sixteen recombinants was that none were able to ascend in the mouse model, despite containing *C. muridarum* genes (Figure 4-6A). This highlights the likelihood that host-specific phenotypes, like ascension are likely multifactorial. Even in a co-infection, in a more susceptible mouse, ascension was not observed (Figure 4-6B). It is possible that over a longer study, we may see increased progeny in the chimeras, particularly fast-growing RC1696. One future direction for these studies would be to investigate the fitness of the recombinants in the transcervical mouse model, where *Chlamydia* is

deposited past the cervix, directly into the uterine horns. It is likely that there may be phenotypes related to the *C. muridarum* genes in development of pathology, time to clearance and bacterial burden, as the uterine horns have less microbial competition.

Genome sequencing of the recombinants revealed that mutations only occurred within the *C. muridarum* region of recombination. This observation was particularly striking as it suggests that there may be differences in the rate of mutation between species. Several studies have been performed to look at the rate of mutation accumulation in *C. trachomatis*, and it has been reported that *in vitro*, *C. trachomatis* does not acquire many mutations after long-term laboratory passage (51-53). Estimated rates for *C. trachomatis* are between 1.10×10^{-9} and 3.84×10^{-10} per base pair per generation, depending on strain (51). Similar studies have not been performed in *C. muridarum*, but these results suggest that those rates could be higher. Increased mutation rate can correspond to an adaptive selective advantage in specific environments. Another potential explanation for the *C. muridarum*-only mutations could be that the gene acquire mutations to adapt to the *C. trachomatis* genomic background, and that these mutations are the direct result of the recombination. Studies examining the mutation rate for *C. muridarum* along with mutational analysis of the entire recombinant library could answer these questions.

Overall, this study provides an initial screen for the recombinant clones generated through horizontal gene transfer of *C. muridarum* into *C. trachomatis*. Our findings highlight the general interspecies compatibility, as most recombinants were able to successfully complete steps of the developmental cycle like wildtype *C. trachomatis*. More in depth analyses in the future should

reveal genes associated with *C. muridarum* growth rate, the developmental cycle and host-specificity in *Chlamydia*.

Materials and Methods

Genome Sequencing and Comparative Analyses

Genome sequencing was previously performed for each recombinant at Oregon State University on an Illumina Hiseq 3000 (29). Mutational analysis and plasmid assemblies were performed using the software platform Geneious (54). All alignments and assemblies were generated using the parental genomes of *C. muridarum* (NC_002620.2), *C. trachomatis* L2/tet9 (CP035484.1) and the *C. trachomatis* L2 434/Bu plasmid (NC_021049). Confirmatory PCRs and sanger sequencing were performed for many of the recombinants throughout the study to ensure accurate, clonal populations. A list of primers used within the study is available in Supplementary Table 4-5. Protein coding sequence analysis was performed through independent direct alignment of primary sequence for *C. muridarum* and *C. trachomatis* proteins. Identity was determined by calculating number of different amino acids over the total length. Similarity was determined using Blosum62 calculations with a threshold of 1.

Table 4-5. Primers used in this study

Name	Forward	Reverse
RC027 LM [a]	5'-GCCTCTTGGGGAACTCGG-3'	5'-GGTGAATTATGGTAGATCGCC-3'
RC027 RM [b]	5'-GCAAGAGAAGCTCTTTACGC-3'	5'-GCTGTTGATGCCTTTTCCTGC-3'
RC2180 LM	5'-GGTAACGTGATGTGTGCTCC-3'	5'-GGAACTGTCCTACAAGGGTC-3'
RC2180 RM	5'-GCTTCTCAGCAAATCCTCG-3'	5'-CCTCTTCTGAGGAAGGCG-3'
RC658 LM	5'-GGAGCAAGGGCGTTTAACAG-3'	5'-CGAAGGGGACTAAACGTAGG-3'
RC658 RM	5'-GCTTATCCGCTATAGAGTACC-3'	5'-GGGCTAGCATAGATTCAGC-3'
RC322 LM	5'-CGTAGTCGATCGTGTAGTC-3'	5'-GGACCGAGGGTCTTTCTC-3'
RC322 RM	5'-GGATACCCAAATCGGCTTGCAATT-3'	5'-CCGGGATTCTTCTATTAAATGAAGGATTC-3'
RC4243 LM	5'-GCAATCTTTCGAAGCTGTTGC-3'	5'-GCAAAGTCAACACGCCAGCC-3'
RC4243 LM	5'-GCTAAGAGAACAGTCAACAAG-3'	5'-GGTATGGTGGTCGAAGGGG-3'
RC1696 LM	5'-CGCGAGTTTTCAGTAAGTTTC-3'	5'-CGCAGGAAGTTTCTACCTTC-3'
RC1696 RM	5'-GGCTTGCTAATTATTGAAAGAAG-3'	5'-GCAAAAGGAGTCTTCTAGATC-3'
RC1696 sat [c] LM	5'-CGTCAACAAGGTGTTCAGG-3'	5'-CCTTTTGTTTGAGGAAGATGG-3'
RC1696 sat [c] RM	5'-GCTTTAGGATCCTCGAAAAGG-3'	5'-CCCAAAGAACAAGATCTTTGC-3'
RC1650 LM	5'-CGAAGAAGTCAGCAGTAGC-3'	5'-GGAAGATTGCAAAGATTCTCC-3'
RC1650 RM	5'-CGAGACATTCACAATATCTTGG-3'	5'-CGATATCAATGCGTCTATCC-3'
RC2649 LM	5'-CGAGGCATTGTAATCCCCGC-3'	5'-GGCAAACACTTAGGACAAGC-3'
RC2649 RM	5'-CGAGAGTTGGCTTATCCCG-3'	5'-CGGCCTTGCCTCTCTCG-3'
RC3666 LM	5'-CGATAGAAAACATCGTTGTCC-3'	5'-CCCTTATCTACATTCGTTCC-3'
RC3666 RM	5'-CCCTATTTTTGGGAAGATCG-3'	5'-CCCGCTTTGAAAGTTAGCG-3'
RC3660 LM	5'-GCTAGCGCTGGAACTCTCC-3'	5'-CCATACCGACTTTTTCTGC-3'
RC3660 RM	5'-GCGCAATACTTGATAATAATCC-3'	5'-GCTTTTCATCTTGGGCATGG-3'
RC3539 LM	5'-CCTATGATTCCGGCTGTTCC-3'	5'-GCATAAAGATCGGAACAATTCG-3'
RC3539 RM	5'-CCAATAAGATAAGGCTGGGG-3'	5'-CGAAACTTAGGCCAAGGATC-3'
Ct *macP*	5'-TGAGCGGCTGTGGAACATGTGACGG-3'	5'-GCATCGCAGCCTCTCATCCACATCC-3'
Cm *macP*	5'-GGAATCGTAAGTAGAGTAGCCAGATTCCG-3'	5'-CGGGGCTAATGCTGCAGAAAAACG-3'
CT *secY*	5'-CCACCCAAAAAGTAGCT-3'	5'-GGTAATATAGCTACAACAGC-3'
Murine r*pp30*	5'-CTCTTCCAGTGTGCAAGAAAGC-3'	5'-AGTGACTGATGAGCTACGAAGG-3'
Murine *rpp30* Probe	/5HEX/TGAGACGAG/ZEN/TCCTGAGTCTC/3IABkFQ/	
rpsR	5'-TGTCCTTTTGTTTCCGCGGG-3'	5'-CTCCTACGAAAGGCAAAAGCCC-3'
rpsR Probe	/56-FAM/TGGAAGACC/ZEN/ATCGATTACAAGGACGT/3IABkFQ/	

Chlamydial propagation and growth curves

Chlamydia strains were propagated in L929 mouse fibroblast cells (ATCC CCL-1) and grown in RMPI 1640 media (Invitrogen, Grand Island, NY) supplemented with 5% heat-inactivated fetal bovine serum (FBS) plus 10 μg/ml gentamycin (Fisher Scientific, Pittsburgh, PA). Cells were

incubated at 37°C with 5% CO_2. Accession numbers for all strains used within this study is in

Table 4-6.

Table 4-6. Genomes used in this study

Strain	Accession Number
C. trachomatis	CP035484.1
C. muridarum	NC_002620.2
RC3814	CP042677
RC3834	CP042676
RC4243	CP042670
RC1650	CP042719
RC1696	CP042781
RC1203	CP042795
RC027	CP042766
RC435	CP042745
RC2418	CP042771
RC2457	CP042769
RC658	CP042754
RC6882	CP042752
RC2180	CP042714
RC2220	CP042776
RC322	CP042700
RC6389	CP042659
RC3539	CP042690
RC3660	CP042683
RC2469	CP042708
RC3666	CP042682

Three time-point growth curves were performed by infecting L929 cells with indicated strains, centrifuged at 550xg for 30 minutes and incubated as above. At 18, 21- or 24-hours post-infection, cells were lysed by adding water onto the wells for 45 seconds. EBs were collected in 1X sucrose-phosphate-glutamic acid (SPG) buffer and frozen at -80 degrees. To determine titers at each time point, EBs were then freeze-thawed and infected as before in a serial dilution. After 24 hours of

incubation, cells were fixed in 100% methanol for 10 minutes, washed in 1X HBSS and stained according to the manufacturer's instructions with the Remel PathoDX Chlamydia Culture Confirmation Kit (ThermoFisher Scientific). Inclusions were visualized under 10X magnification using the EVOS FL Auto2 microscope and were enumerated using MIPAR software with a contrast range between 90 and 255 units, expected inclusion area of 20 μm^2 -1600 μm^2 and a roundness threshold of ≤ 0.7. Progeny was normalized to *C. trachomatis* infectious dose and normalized values were plotted. An unpaired Student's t-test was run between the *C. trachomatis* and each chimera at each time point to see if there was any significant difference in the progeny (p-values as reported). Extended growth curves were performed as above, but EBs were harvested at the indicated time points.

Confocal microscopy and inclusion size measurement

L929 cells were seeded at 30% confluency 24 hours prior to infection in an 8-well ibiTreat μ-Slide (Ibidi, Martinsried, Germany) and were infected with respective *C. trachomatis, C. muridarum,* or recombinant mutant. At 21 or 24 hpi, infected cells were fixed with 100% methanol for 10 minutes at room temperature. Cells were washed twice with HBSS then stained using 180 μl of the PathoDx *Chlamydia* Culture Confirmation Kit (Remel Europe Ltd., Dartford, UK) overnight at 4°C. 20μl of 1μM 4', 6-diamidino-2-phenylindole (DAPI) diluted 1:100 in PBS was then added to wells and allowed to stain for 10 minutes at room temperature. Stain was then removed, and the cells washed with PBS. A final overlay of Vectashield antifade mounting medium (Burlingame, CA) was added and slides were stored at 4°C in the dark until imaged. Cells were visualized on an Olympus IX81/3I spinning disk confocal inverted microscope at 150X magnification and captured on an Andor Zyla 4.2 sCMOS camera (Belfast, Northern Ireland). Microscope and camera were operated

using SlideBook 6 software (Intelligent Imaging Innovations, Denver, USA). Exposure time remained consistent for all fields captured, with exposure for DAPI at 2 seconds, GFP (MOMP/LOS) 3 seconds, and Evans Blue (host cytoplasm) 3 seconds. 5-10 images were taken per strain. 5 Z stack images at 0.35μm apart were taken per field imaged. Images were processed in SlideBook 6 and a No Neighbors Deconvolution with a subtraction constant of 0.4 was applied to all images.

Mouse Infections

Female C57BL/6 mice or C3He/J mice (6 to 8 weeks old) were purchased from Jackson Laboratories and housed in accordance with the requirements specified by the University of Kansas Institutional Care and Use Committee. Mice were treated subcutaneously with 2.5 mg medroxyprogesterone acetate (Depo-Provera, Pfizer, NY) upon arrival (day −7) to synchronize menstrual cycles. Infectious doses of parental or recombinant *Chlamydia* were diluted in sucrose-phosphate-glutamic acid (SPG) buffer along with an SPG-only mock control dose. Mice were inoculated intravaginally with 5 μl of infectious dose, for a final concentration around 1×10^5 IFU/mouse or 1.7×10^6 IFUs/mouse for the co-infection samples, by deposition of the dose into the vagina. Seven days post infection for C57BL/6 or five days post-infection for C3He/J, mice were humanely euthanized and the genital tracts including vaginal vault and uterine horns were collected in SPG. Organs were homogenized using a rotor/stator homogenizer (Biospec, Bartlesville, OK). DNA isolation was performed using a DNeasy Blood and Tissue Kit (Qiagen) to the manufacturer's instructions. Isolated DNA was then used to determine bacterial burden by droplet digital PCR (ddPCR). Primers and probes for *Chlamydial rpsR* and murine *rpp30* were used with ddPCR Supermix for Probes (Bio-Rad, Hercules, CA) (Supplementary Table X).

ddPCRs were performed as previously reported (29). Bacterial burdens of each tissue were analyzed using QuantaSoft Software (Bio-Rad), and the results are reported as $\log_{10}$ ratios of *Chlamydia* DNA to host DNA (*rpsR/rpp30* copies). Box and whisker scatter plots were generated in GraphPad Prism 8.

References

1. Division of STD Prevention NCfHA, Viral Hepatitis, STD and TB Prevention. Sexually Transmitted Disease Surveillance. In: Prevention CfDCa, editor. 2018.
2. Cates W, Wasserheit JN. Genital chlamydial infections: Epidemiology and reproductive sequelae. American Journal of Obstetrics and Gynecology. 1991;164(6):1771-81.
3. Lane AB, Decker CF. Chlamydia trachomatis infections. Dis Mon. 2016;62(8):269-73.
4. Korenromp EL, Sudaryo MK, de Vlas SJ, Gray RH, Sewankambo NK, Serwadda D, et al. What proportion of episodes of gonorrhoea and chlamydia becomes symptomatic? Int J STD AIDS. 2002;13(2):91-101.
5. Stephens RS. The cellular paradigm of chlamydial pathogenesis. Trends in Microbiology. 2003;11(1):44-51.
6. Darville T HT. Pathogenesis of Genital Tract Disease due to Chlamydia trachomatis. J Infect Dis. 2010;15(201):S114-S25.
7. Rasmussen SJ, Eckmann L, Quayle AJ, Shen L, Zhang YX, Anderson DJ, et al. Secretion of proinflammatory cytokines by epithelial cells in response to Chlamydia infection suggests a central role for epithelial cells in chlamydial pathogenesis. J Clin Invest. 1997;99(1):77-87.
8. De Clercq E, Kalmar I, Vanrompay D. Animal models for studying female genital tract infection with Chlamydia trachomatis. Infect Immun. 2013;81(9):3060-7.
9. Brunham RC, Rey-Ladino J. Immunology of Chlamydia infection: implications for a Chlamydia trachomatis vaccine. Nat Rev Immunol. 2005;5(2):149-61.
10. Phillips S, Quigley BL, Timms P. Seventy Years of Chlamydia Vaccine Research - Limitations of the Past and Directions for the Future. Front Microbiol. 2019;10:70.
11. Baumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med. 2013;3(12):a010041.
12. Foley SL, Johnson TJ, Ricke SC, Nayak R, Danzeisen J. Salmonella pathogenicity and host adaptation in chicken-associated serovars. Microbiol Mol Biol Rev. 2013;77(4):582-607.
13. Pan X, Yang Y, Zhang JR. Molecular basis of host specificity in human pathogenic bacteria. Emerg Microbes Infect. 2014;3(3):e23.
14. Collins JW, Keeney KM, Crepin VF, Rathinam VA, Fitzgerald KA, Finlay BB, et al. Citrobacter rodentium: infection, inflammation and the microbiota. Nat Rev Microbiol. 2014;12(9):612-23.
15. Savkovic SD, Villanueva J, Turner JR, Matkowskyj KA, Hecht G. Mouse model of enteropathogenic Escherichia coli infection. Infect Immun. 2005;73(2):1161-70.
16. Petty NK, Bulgin R, Crepin VF, Cerdeno-Tarraga AM, Schroeder GN, Quail MA, et al. The Citrobacter rodentium genome sequence reveals convergent evolution with human pathogenic Escherichia coli. J Bacteriol. 2010;192(2):525-38.
17. Dimond ZE, Hefty PS. Comprehensive genome analysis and comparisons of the swine pathogen, Chlamydia suis reveals unique ORFs and candidate host-specificity factors. Pathog Dis. 2020.
18. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res. 2000;28(6):1397-406.
19. Borges V, Gomes JP. Deep comparative genomics among Chlamydia trachomatis lymphogranuloma venereum isolates highlights genes potentially involved in pathoadaptation. Infect Genet Evol. 2015;32:74-88.

20. Bachmann NL, Polkinghorne A, Timms P. Chlamydia genomics: providing novel insights into chlamydial biology. Trends Microbiol. 2014;22(8):464-72.

21. Gitsels A, Sanders N, Vanrompay D. Chlamydial Infection From Outside to Inside. Front Microbiol. 2019;10:2329.

22. Nunes A, Gomes JP. Evolution, phylogeny, and molecular epidemiology of Chlamydia. Infect Genet Evol. 2014;23:49-64.

23. Abdelrahman YM, Belland RJ. The chlamydial developmental cycle. FEMS Microbiol Rev. 2005;29(5):949-59.

24. Triboulet S, Subtil A. Make It a Sweet Home: Responses of Chlamydia trachomatis to the Challenges of an Intravacuolar Lifestyle. Microbiol Spectr. 2019;7(2).

25. Messinger JE, Nelton E, Feeney C, Gondek DC. Chlamydia Infection Across Host Species Boundaries Promotes Distinct Sets of Transcribed Anti-Apoptotic Factors. Front Cell Infect Microbiol. 2015;5:96.

26. Suchland RJ, Carrell SJ, Wang Y, Hybiske K, Kim DB, Dimond ZE, et al. Chromosomal Recombination Targets in Chlamydia Interspecies Lateral Gene Transfer. J Bacteriol. 2019;201(23).

27. Panzetta ME, Valdivia RH, Saka HA. Chlamydia Persistence: A Survival Strategy to Evade Antimicrobial Effects in-vitro and in-vivo. Front Microbiol. 2018;9:3101.

28. Pokorzynski ND, Thompson CC, Carabeo RA. Ironing Out the Unconventional Mechanisms of Iron Acquisition and Gene Regulation in Chlamydia. Front Cell Infect Microbiol. 2017;7:394.

29. LaBrie SD, Dimond ZE, Harrison KS, Baid S, Wickstrum J, Suchland RJ, et al. Transposon Mutagenesis in Chlamydia trachomatis Identifies CT339 as a ComEC Homolog Important for DNA Uptake and Lateral Gene Transfer. mBio. 2019;10(4).

30. Ranjit DK, Liechti GW, Maurelli AT. Chlamydial MreB Directs Cell Division and Peptidoglycan Synthesis in Escherichia coli in the Absence of FtsZ Activity. mBio. 2020;11(1).

31. Liechti G, Kuru E, Packiam M, Hsu YP, Tekkam S, Hall E, et al. Pathogenic Chlamydia Lack a Classical Sacculus but Synthesize a Narrow, Mid-cell Peptidoglycan Ring, Regulated by MreB, for Cell Division. PLoS Pathog. 2016;12(5):e1005590.

32. Lee JK, Enciso GA, Boassa D, Chander CN, Lou TH, Pairawan SS, et al. Replication-dependent size reduction precedes differentiation in Chlamydia trachomatis. Nat Commun. 2018;9(1):45.

33. Albrecht M, Sharma CM, Reinhardt R, Vogel J, Rudel T. Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. Nucleic Acids Res. 2010;38(3):868-77.

34. Chen J, Zhang H, Zhou Z, Yang Z, Ding Y, Zhou Z, et al. Chlamydial induction of hydrosalpinx in 11 strains of mice reveals multiple host mechanisms for preventing upper genital tract pathology. PLoS One. 2014;9(4):e95076.

35. Shine J, Dalgarno L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci U S A. 1974;71(4):1342-6.

36. Danilition SL, Maclean IW, Peeling R, Winston S, Brunham RC. The 75-kilodalton protein of Chlamydia trachomatis: a member of the heat shock protein 70 family? Infect Immun. 1990;58(1):189-96.

37. Allen JE, Stephens RS. Identification by sequence analysis of two-site posttranslational processing of the cysteine-rich outer membrane protein 2 of Chlamydia trachomatis serovar L2. J Bacteriol. 1989;171(1):285-91.

38.   Stephens RS, Wagar EA, Edman U. Developmental regulation of tandem promoters for the major outer membrane protein gene of Chlamydia trachomatis. J Bacteriol. 1988;170(2):744-50.

39.   Lecuit M, Dramsi S, Gottardi C, Fedor-Chaiken M, Gumbiner B, Cossart P. A single amino acid in E-cadherin responsible for host specificity towards the human pathogen Listeria monocytogenes. EMBO J. 1999;18(14):3956-63.

40.   Lee J, Cox JV, Ouellette SP. Critical Role for the Extended N Terminus of Chlamydial MreB in Directing Its Membrane Association and Potential Interaction with Divisome Proteins. J Bacteriol. 2020;202(9).

41.   Kemege KE, Hickey JM, Barta ML, Wickstrum J, Balwalli N, Lovell S, et al. Chlamydia trachomatis protein CT009 is a structural and functional homolog to the key morphogenesis component RodZ and interacts with division septal plane localized MreB. Molecular Microbiology. 2015;95(3):365-82.

42.   Kim M, Park S, Lee SJ. Global transcriptional regulator TrmB family members in prokaryotes. J Microbiol. 2016;54(10):639-45.

43.   Thongdee N, Jaroensuk J, Atichartpongkul S, Chittrakanwong J, Chooyoung K, Srimahaeak T, et al. TrmB, a tRNA m7G46 methyltransferase, plays a role in hydrogen peroxide resistance and positively modulates the translation of katA and katB mRNAs in Pseudomonas aeruginosa. Nucleic Acids Res. 2019;47(17):9271-81.

44.   Schmid AK, Reiss DJ, Pan M, Koide T, Baliga NS. A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. Mol Syst Biol. 2009;5:282.

45.   Wei Y, Zhang H, Gao ZQ, Wang WJ, Shtykova EV, Xu JH, et al. Crystal and solution structures of methyltransferase RsmH provide basis for methylation of C1402 in 16S rRNA. J Struct Biol. 2012;179(1):29-40.

46.   Kyuma T, Kimura S, Hanada Y, Suzuki T, Sekimizu K, Kaito C. Ribosomal RNA methyltransferases contribute to Staphylococcus aureus virulence. FEBS J. 2015;282(13):2570-84.

47.   Real G, Henriques AO. Localization of the Bacillus subtilis murB gene within the dcw cluster is important for growth and sporulation. J Bacteriol. 2006;188(5):1721-32.

48.   Ciampi MS. Rho-dependent terminators and transcription termination. Microbiology (Reading). 2006;152(Pt 9):2515-28.

49.   Lu C, Lei L, Peng B, Tang L, Ding H, Gong S, et al. Chlamydia trachomatis GlgA is secreted into host cell cytoplasm. PLoS One. 2013;8(7):e68764.

50.   Sun Z, Sun Y, Li Y, Luan X, Chen H, Wu H, et al. Identification of HeLa cell proteins that interact with Chlamydia trachomatis glycogen synthase using yeast twohybrid assays. Mol Med Rep. 2020;21(3):1572-80.

51.   Borges V, Ferreira R, Nunes A, Sousa-Uva M, Abreu M, Borrego MJ, et al. Effect of long-term laboratory propagation on Chlamydia trachomatis genome dynamics. Infect Genet Evol. 2013;17:23-32.

52.   Borges V, Nunes A, Ferreira R, Borrego MJ, Gomes JP. Directional evolution of Chlamydia trachomatis towards niche-specific adaptation. J Bacteriol. 2012;194(22):6143-53.

53.   Ferreira R, Borges V, Nunes A, Nogueira PJ, Borrego MJ, Gomes JP. Impact of loci nature on estimating recombination and mutation rates in Chlamydia trachomatis. G3 (Bethesda). 2012;2(7):761-8.

54.   Kearse M MR, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. Geneious Basic: an

integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9.

Chapter V. Discussion

In the above chapters, I have described my work in determining genetic factors for host-specificity in *Chlamydia* using comparative genomics and cross-species recombinant libraries. Through these investigations we have bioinformatically determined potential host-specificity factors through analyzing unique and low similarity genes between three closely related species, determined that unique factors may not be acting alone as host-restriction factors, and utilized recombination as a new tool for investigating gene functions in *Chlamydia.* Within this chapter, I will discuss the overarching results from these individual studies, tying them into the current body of knowledge and present future directions that could be pursued given the presented works.

Genome-wide recombinants as a genetic tool for functional discovery

Recombinant libraries have been used as a tool for scientists in all fields of biology, from complex intercross libraries in *Drosophila* used to investigate inheritance of complex traits (1) to engineered viral libraries used to discover important disease correlates (2). We have generated a unique recombinant library in *Chlamydia* by using homologous recombination to generate interspecies crosses, bringing large regions from one species into the other. A subset of these clones was investigated for phenotypes relating to host-specificity, pathogenesis and basic biology.

Chlamydia trachomatis exhibits a restricted host range as an obligate human pathogen while closely related *Chlamydia muridarum* infects only mice. Because the vast majority of genes are conserved between these species, and because expression of genes in trans can be challenging in *Chlamydia*, recombinant libraries provided an opportunity to analyze not only the unique genes between species, but also those conserved genes for inter-compatibility and conserved host-pathogen interactions. First, we analyzed recombinants within the region known as the plasticity

zone (PZ), where unique genes have long been hypothesized, or demonstrated, to participate in host restrictive pathways. Second, we were able to look at regions of recombination throughout the genome. As a tool, the recombinant libraries proved useful in investigating specific phenotypes through screens like the three time-point growth curves from Chapter 4 (Figure 4-2). Using the overlapping libraries and "subclones" in both studies, large regions of recombination associated with a phenotype were able to be narrowed down to a small subset of genes, or even to an individual gene as in Chapter 3 with inclusion membrane protein TC0424 (Figure 3-7). These phenotypes generally speak to the large effect that small protein changes can have on the ability of a protein to perform its function with its cognate partners, falling back on the classic dogma that structure equals function.

One major benefit to recombinant libraries is that they allow for investigation of essential genes, which may otherwise be unable to be knocked out by other genetics approaches. Due to the obligate intracellular lifestyle of *Chlamydia* and the highly reduced genomes that go along with it, it is expected that a large proportion of genes are essential. With the recombinants, we can investigate the impact that small changes have on protein function for *C. trachomatis* by replacing these genes with an unidentical ortholog. As seen in Chapter 3 and 4, these small changes can have a large impact on *Chlamydia*. Another major advantage of this system is the ability to analyze a large subset of genes at the same time, which can be important for multi-gene processes. A third advantage to the recombinant libraries is the ability to perform backcrosses to "complement" a specific phenotype, as was performed in Chapter 3.

Along with the advantages to this system, there are a few distinct disadvantages, however. A key disadvantage to the recombinant libraries is that narrowing down a phenotype to a specific gene can be challenging and time consuming, as it requires potentially many recombination events to generate overlapping subclones. Correlating phenotype with gene can also require the use of forward genetics to confirm the interactions involved. These recombinant libraries, therefore, should be looked at as a screening tool, to provide a mechanism to scan the genome for regions of interest for additional studies, or as a way to analyze a large swath of genes at once, as with the PZ. Additional disadvantages include the need to use multiple selectable markers, limiting the choices for subsequent or follow-up studies. In our library, tetracycline and chloramphenicol were both used, limiting possible antibiotics for future studies like complementing in trans.

Finally, as was seen in this work, when the genomes are as highly conserved as they are between *C. trachomatis* and *C. muridarum*, the likelihood that genes are interchangeable can limit discernible phenotypes. In Chapter 4, we show that 13 of 16 chimeras had no gross growth defects in our initial screen (Figure 4-2A). This indicates that each were able to successfully adhere, invade, establish an inclusion, undergo the developmental cycle and escape the host cell. It is possible that there are minute phenotypes present in these recombinants, however, our screen demonstrates that these are not detrimental to *Chlamydia* in cell culture. What these data suggest is that *C. trachomatis* and *C. muridarum* genes are largely intercompatible and therefore likely play the same function. There were regions of the genome that were unable to be investigated in this system, as recombinants were unable to be generated at those sites (Chapter 1, Figure 1-7). These regions could speak to truly incompatible genes which result in "sick" *Chlamydia* and are unable to be propagated, or they could indicate constraints with recombination. Constraints could

include those that are technical, like not having a close enough antibiotic marker, or biological, like the DNA at these sites in inaccessible. Only continued generation of recombinants could help tease out these aspects.

The plasticity zone, what we know and what we don't

For almost 20 years, the plasticity zone has been a region of interest for researchers as a primary difference between chlamydial genomes. Back in 2000, with the first whole genome sequences of *C. muridarum,* then known as *C. trachomatis* strain Mouse Pneumonitis (MoPn) due to their similarity, a region near the terminus was discovered to have significant differences between *Chlamydia* genomes and was first referred to as the "plasticity zone" (3). Since then, many studies have been performed to try to figure out what the largely hypothetical genes within this region are doing. These studies have included comparative genomics studies to analyze LCT presence and mutations (4, 5), along with functional studies to determine the role of phospholipases (6, 7), the MAC/perforin (8) and cytotoxins (7, 9, 10).

In this study, we add to this growing body of work through additional comparative genomics using *C. suis* (Chapter 2) and the use of the PZ recombinants (Chapter 3). Comparisons of the PZ between *C. suis*, *C. muridarum* and *C. trachomatis* reveals the potential importance for phospholipase genes in *C. suis*, which contained a greater number of these genes than either *C. muridarum* or *C. trachomatis*. The exact function of these genes remains unclear, but the phylogenetic analyses point to the genes having distinct biological functions (Chapter 2, Figure 2-5B). What is known about these genes is that they do not appear to be unique among *Chlamydia*, and similar genes play roles in host-specific adherence and survival (11, 12). In our PZ

recombinants, we were not able to see any effects of the *C. muridarum* phospholipases in *C. trachomatis*, however, it is possible that two phospholipases within the region to the right of the cytotoxins could contribute to the cytotoxic phenotype described in Chapter 3, Figure 3-6. Interestingly, those two phospholipases, *tc0440* and *tc0447* do not appear to have direct *C. trachomatis* homologs (Chapter 2, Figure 2-5B) and appear in a clade with only *C. suis* homologs, *Chls0463* and *Chls0469.* As a future study, it may be worthwhile to investigate cytotoxicity in *C. suis* as a way to determine the role for these phospholipases.

Complicating the question of cytotoxicity is that *C. suis* contains one or two copies of the LCT genes. If *C. suis* is cytotoxic, it is possible that the LCT genes are playing some sort of role, although out PZ recombinant studies would suggest that they are either not contributing to this effect, as the Right Partial PZ mutant was cytotoxic without containing any LCTs, or that there are two distinct mechanisms of cytotoxicity, one mediated by the LCTs and another mediated by the genes to the right of the LCTs, potentially these phospholipases. If *C. suis* is not cytotoxic, it could indicate a different role for these cytotoxins, perhaps a role only necessary in a mouse or swine host. The variable presence of LCTs in *C. suis* could also suggest that they are only relevant in specific tissues. Both *C. muridarum* and *C. suis* are able to infect many tissues within the host from reproductive organs, which has largely been the focus of this work, to the lungs, eyes and gastrointestinal tract (13). It is possible that the cytotoxins are important for infection in the gastrointestinal tract, for example, which *C. trachomatis* does not appear to naturally infect, and therefore *C. trachomatis* would not require these genes. This idea that LCTs could be important in other tissues has also been discussion by Morrison et al, who demonstrated that two LCTs appeared to be important for virulence in the gastrointestinal mouse model but not the genital tract (9). The

biovar flexibility of *C. suis* adds support to this hypothesis. Mutational analysis of the *C. suis* LCTs in the pig infection model could provide complimentary results to Morrison's study.

There are still many remaining questions about the PZ, particularly in light of the data presented in Chapter 3. Because *C. muridarum* PZ genes were not sufficient for ascension in our model, we are still left wondering what these genes are for. Because *Chlamydia* genomes are so reduced, it is hypothesized that genes which are maintained in the genome are likely important. If this is true, then the genes of the PZ must serve a purpose, whether in maintenance of the developmental cycle and intracellular niche or in pathogenesis. It is still unclear what these functions could be, but the likely importance of this region will no doubt continue to make the PZ an area of interest for researchers.

Host-specificity is likely multifactorial

One of the pervading outcomes from the recombinant libraries has been that simply adding a *C. muridarum* gene to *C. trachomatis* does not appear to be sufficient to increase virulence within the mouse. We observed these results both with the PZ recombinants in Chapter 3 and the genome-wide recombinants in Chapter 4. Despite the fact that these recombinants contained genes unique to *C. muridarum*, no ascension was observed in the mouse, even when added in a co-infection. These results suggest that the ability to ascend, and likely other pathogenesis phenotypes, are the result of many genes working in tandem to increase the fitness of the pathogen within the host.

As described in Chapter 1, host-specificity genes largely fall into three categories; surface proteins, immune evasion and nutrient acquisition. It is important to note, though, that species must contain

genes in all of these categories to successfully infect a host, rather than any single gene in one category. The steps of pathogenesis for any pathogen require each of these steps as pathogens colonize and invade a host through the use of surface adhesions and receptors, evade the host immune system as they cause disease and disseminate to other cells or tissues, and compete for nutrients with the other microorganisms of the host as they continue to replicate. Without any one factor, a pathogen would be unable to be successful at infecting its host. At each step, a pathogen interacts with the host and every interaction is a possible point for the evolution of host-specificity.

As host species have evolved to increase efficiencies of their enzymes, to their unique environment or in response to pathogens, their protein sequences, and therefore structures, acquire changes. Pathogens depend on direct interactions with host proteins and over time, bacterial proteins must evolve with their host to maintain these interactions. Over time, and over a multitude of protein changes, this leads to specialization of bacteria to their specific host. Generalist pathogens, which do not have tightly restricted host ranges generally carry more genes than host-specific pathogens (14). These genes allow for the generalist to adapt to their environment which may vary widely between hosts, or even outside of a host, as is the case for many microorganisms like the saprophytic bacterium *L. monocytogenes* (15). Because specialists typically have lost extraneous genes allowing for survival in multiple environments, it is expected that remaining genes are each adapted for optimal interaction with the native host environment.

In *Chlamydia,* it is important to remember that these pathogens have evolved within hosts for millennia, likely undergoing many changes to perfect the interactions that allow for the success of the bacteria today. It is likely that every host-facing gene has evolved for interaction with the

specific host that it infects. Additionally, despite the fact that the recombinant libraries do allow for investigation of multiple factors in the same recombinant, the fractured nature of the *Chlamydia* genome may not allow complete pathways to be optimized for the non-native host. The *Chlamydia* genome is not organized with genes of similar biological roles in proximity of each other or in operons (16). Therefore, the recombination may bring one mouse-adapted *C. muridarum* gene into *C. trachomatis*, but not other critical components for the pathway, leading to an inability to successfully infect the mouse. Future experiments could help address the issue of multifactorality in determining host-specificity for *Chlamydia*.

Where do we go from here?

With the studies described here, we attempted to generate "gain-of-function" mutants by recombining *C. muridarum* genes into *C. trachomatis* in an effort to generate a strain of *C. trachomatis* able to overcome the host-specific barriers of the mouse. The initial idea behind these recombinants was that *C. muridarum* genes, whether in the PZ or elsewhere in the genome, would provide *C. trachomatis* with the ability to overcome the factors restricting its pathogenesis in the mouse. We were unable to generate recombinants which ascended in the mouse model, likely underlying how many different factors are involved in this process. It was always expected that producing a "gain-of-function" would be more challenging than to create a "loss-of-function" mutant by "breaking" an interaction. Because we used the *C. muridarum* transposon library to provide selectable markers and a static *C. trachomatis* selectable marker, our recombinants had largely *C. trachomatis* genomes.

We have attempted to reverse these libraries by utilizing the *C. trachomatis* transposon libraries (Chapter 1, Figure 1-6), however selection in these recombinants has been tricky. One primary reason for this is that the *C. trachomatis* transposon libraries are generated using ampicillin for selection, which *C. muridarum* has increased resistance to. In order to generate a library of *C. muridarum* genomes with regions of *C. trachomatis* genes recombined in, we need to overcome this barrier. Therefore, I analyzed increased concentrations of ampicillin that could be used to select for recombinants. Previously, a concentration of 5ug/mL was used in attempts to generate the reverse library. However, *C. muridarum* cheaters would overtake any possible recombinants. Using a transposon mutant from our *C. trachomatis* library, I demonstrated that a concentration between 50-200ug/mL could allow for maximal suppression of *C. muridarum* growth without reducing *C. trachomatis* viability (Figure 5-1). Using these concentrations, generation of a reverse recombinant library should be possible and could allow for the generation of loss-of-function" mutants.
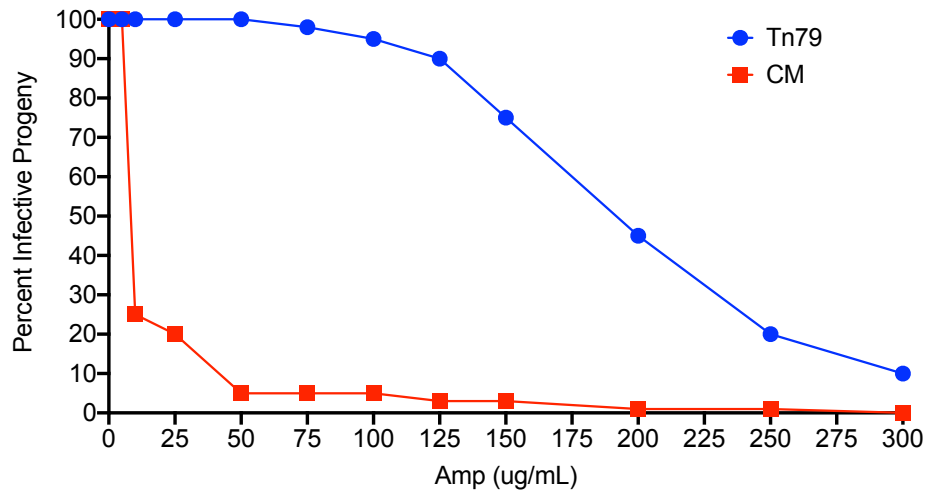
Figure 5-1. Ampicillin resistance for wildtype *C. muridarum* and *C. trachomatis*::tn/bla (Tn79). L929 mouse fibroblast cells were infected at a multiplicity of infection of 10 with *Chlamydia* and incubated in media containing ampicillin at the indicated dose. After 24 hours of incubation, infectious progeny were harvested using water lysis and infected in onto new L929 cells for an additional 24 hours without antibiotics. After incubation, cells were fixed with 100% methanol and stained with Dapi for host nuclei and the PathoDX *Chlamydia* Confirmation Kit for *Chlamydia*. Perfect infectivity was determined by calculating number of inclusions over number of host nuclei.

This reverse recombination library is expected to generate chimeras across the *C. trachomatis* genome. We anticipate that recombinants will have "sick" phenotypes, as with RC4243 in Chapter 4 (Figure 4-3) and RC826 in Chapter 3 (Figure 3-7), but also those which do not have *in vitro* phenotypes, as was seen in Chapter 3 (Figure 4-2). However, what is most intriguing is the possibility of recombinants which have a "loss-of-function" in the mouse model, indicating that they have lost an important factor in virulence of the mouse host by replacing it with its *C. trachomatis* homolog. Additionally, these recombinants could provide an opportunity to investigate the PZ from the reverse perspective, removing the *C. muridarum* unique genes, like the LCTs, and replacing them with the *C. trachomatis* genes. These future studies would continue the work in Chapter 3 to determine the role of these genes in *C. muridarum*.

Another important future direction of this work will be to determine the transcriptome of *C. muridarum*. Through these recombinants, we have shown that the addition of *C. muridarum* genes have been able to generate a faster-growing *C. trachomatis* (Chapter 4, Figure 4-5). Putative genes include several transcription factors which could help explain how *C. muridarum* growth rates are so much faster than *C. trachomatis.* One of the first steps to teasing out these differences will be a full transcriptome for *C. muridarum* across the developmental cycle. It is possible that the accelerated growth rate is due to earlier transcription of key conversion genes, increased efficiency of transcription factors leading to more copies of genes for nutrient acquisition and increased replication, or it could be that there are unique small RNAs which regulate this developmental cycle that have not been identified. A complete transcriptome for *C. muridarum*, with direct comparison to the transcriptome for *C. trachomatis* would be an important study and could provide insight into not only the developmental cycle, but which genes are not expressed *in vitro*. Genes not expressed *in vitro* may be important in pathogenesis and could indicate candidates important in the mouse infection.

Along with the *in vitro* transcriptome studies, an ideal study would be to compare *C. trachomatis* and *C. muridarum* transcriptomes within the mouse model to compare expression of genes *in vivo*. Similar experiments have been successful for identifying key genes involved in pathogenesis of the host, but not in culture in *S. aureus* (17), *Pseudomonas aeruginosa* (18)*,* and *Bordetella pertussis* (19). In the *S. aureus* study, *in vivo* and *in vitro* transcriptomes were directly compared and it was found that 240 genes were differentially expressed in the two conditions. Considering that the average *S. aureus* genome contains around 2600 genes (20), this would be around 9% of

the genome. Genes that were differentially expressed *in vivo* include surface-bound proteins, like staphylococcal cell-wall protein clumping factor B (*clfB*), which is an adhesion factor for squamous epithelial cells, along with those important for iron acquisition and those involved in immune evasion, like staphylococcal complement inhibitor (*scn*) (17). These genes fit into the paradigm that genes important for pathogenesis of the host fall into the three main categories described in Chapter 1 (Figure 1-5). Importantly, in this study it was also found that *S. aureus* transcriptomes between patients differed by nearly 45%, and not only differed between patients, but also within the same patient at different times. These data suggest that the host environment plays a major role in bacterial gene expression and could help explain why outcomes vary greatly between patients. In the *B. pertussis* study, 606 genes were differentially expressed *in vivo* when compared with the *in vitro* condition, around 16% of the genome (19) (21) while in the *P. aeruginosa* study over 50% of the genome was shown to be differentially expressed (18). Once again, in both studies, iron acquisition appeared to be a key component of *in vivo* pathogenesis. Iron acquisition in *Chlamydia* is still relatively poorly understood, although it is likely as important to its pathogenesis as it is with these other pathogens. In general, *Chlamydia* genomes lack homologs to the iron acquisition genes found in most other bacteria, likely having a unique or atypical method to overcome this host defense strategy (22, 23). These *in vivo* transcriptome studies could provide novel candidates in the chlamydial genome important for pathogenesis.

The challenge with this experiment will be to be able to isolate bacterial RNA from the mouse tissue, particularly since *C. trachomatis* infections in the mouse are weak. The examples provided above are not from intracellular bacteria, and the obligate intracellular pathogen will have additional difficulties being extracted from the host. Possible ways to overcome this barrier include

the use of immunomagnetic separation to separate *Chlamydia* from lysed mouse tissue to decrease the host RNA that would be present in the sample (24) or the use of single-cell RNA seq by isolating infected cells by fluorescence-activated cell sorting (FACS) as was performed in an Influenza A *in vivo* transcriptome study (25). One *in vivo* transcriptome study analyzed cervical samples from STI-positive women using Total RNA-seq, which involves sequencing of ribosomal-depleted RNAs from host, pathogens and the microbial community, without needing to extract the pathogen of interest (26). This pilot study was used as a proof of concept and it is unclear if there was enough sequencing depth to determine a pathogenic transcriptome from this technique, although it would certainly be a promising direction. The benefit of single-cell RNAseq or Total RNAseq methods over extracted EB methods would be to analyze how the host responds to chlamydial infection, which likely is also critical to answering questions about the pathogenicity of *Chlamydia.*

In summary, the works presented here continue a long history of investigation into the host-specificity of *Chlamydia*, adding novel evidence that virulence genes in *Chlamydia* may be conserved genes with evolved structural changes unique to the host-specific interaction, the plasticity zone may not be as critical to host-specificity as once predicted and that infection of a specific host is not determined by one singular gene. While this work alone cannot answer the key question of host-specificity in *Chlamydia*, it does continue to highlight the importance of these studies and provide evidence for how recombinant libraries could be used to tease out these answers. Future studies will build upon the data presented here, one day providing the means to increase our understanding of *Chlamydia* and to provide new ways to think about the treatment and prevention of this one-of-a-kind pathogen.

References

1. Long AD, Macdonald SJ, King EG. Dissecting complex traits using the Drosophila Synthetic Population Resource. Trends Genet. 2014;30(11):488-95.
2. Zhang W, Fu J, Liu J, Wang H, Schiwon M, Janz S, et al. An Engineered Virus Library as a Resource for the Spectrum-wide Exploration of Virus and Vector Diversity. Cell Rep. 2017;19(8):1698-709.
3. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, et al. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res. 2000;28(6):1397-406.
4. Carlson JH, Hughes S, Hogan D, Cieplak G, Sturdevant DE, McClarty G, et al. Polymorphisms in the Chlamydia trachomatis cytotoxin locus associated with ocular and genital isolates. Infect Immun. 2004;72(12):7063-72.
5. Sait M LM, Clark EM, Wheelhouse N, Spalding L, Markey B, Magnino S, Lainson FA, Myers GS, Longbottom D. Genome sequencing and comparative analysis of three Chlamydia pecorum strains associated with different pathogenic outcomes. BMC Genomics. 2014.
6. Nelson DE, Crane DD, Taylor LD, Dorward DW, Goheen MM, Caldwell HD. Inhibition of chlamydiae by primary alcohols correlates with the strain-specific complement of plasticity zone phospholipase D genes. Infect Immun. 2006;74(1):73-80.
7. Rajaram K, Giebel AM, Toh E, Hu S, Newman JH, Morrison SG, et al. Mutational Analysis of the *Chlamydia muridarum* Plasticity Zone. Infect Immun. 2015;83(7):2870-81.
8. Taylor LD, Nelson DE, Dorward DW, Whitmire WM, Caldwell HD. Biological characterization of Chlamydia trachomatis plasticity zone MACPF domain family protein CT153. Infect Immun. 2010;78(6):2691-9.
9. Morrison SG, Giebel AM, Toh EC, Spencer HJ, Nelson DE, Morrison RP. Chlamydia muridarum Genital and Gastrointestinal Infection Tropism Is Mediated by Distinct Chromosomal Factors. Infection and Immunity. 2018;86(7).
10. Belland RJ, Scidmore MA, Crane DD, Hogan DM, Whitmire W, McClarty G, et al. Chlamydia trachomatis cytotoxicity associated with complete and partial cytotoxin genes. Proc Natl Acad Sci U S A. 2001;98(24):13984-9.
11. Edwards JL ED, Apicella MA. Gonococcal Phospholipase D Modulates the Expression and Function of Complement Receptor 3 in Primary Cervical Epithelial Cells. Infect Immun. 2003;71(11):6381-91.
12. Hinnebusch BJ RA, Cherepanov P, Dixon JE, Schwan TG, Forsberg A. Role of Yersinia Murine Toxin in Survival of Yersinia pestis in the Midgut of the Flea Vector. Science. 2002;296(5568):733-5.
13. Cheong HC, Lee CYQ, Cheok YY, Tan GMY, Looi CY, Wong WF. Chlamydiaceae: Diseases in Primary Hosts and Zoonosis. Microorganisms. 2019;7(5).
14. Baumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med. 2013;3(12):a010041.
15. Orsi RH, den Bakker HC, Wiedmann M. Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol. 2011;301(2):79-96.

16. Thomson NR HM, Carder C, Lennard N, Lockey SJ, Marsh P, Skipp P, O'Connor CD, Goodhead I, Norbertzcak H, Harris B, Ormond D, Rance R, Quail MA, Parkhill J, Stephens RS, Clarke IN. Chlamydia trachomatis: Genome sequence analysis of lymphogranuloma venereum isolates. Renome Res. 2008;18(1):161-71.

17. Chaves-Moreno D, Wos-Oxley ML, Jauregui R, Medina E, Oxley AP, Pieper DH. Exploring the transcriptome of Staphylococcus aureus in its natural niche. Sci Rep. 2016;6:33174.

18. Damron FH, Oglesby-Sherrouse AG, Wilks A, Barbier M. Dual-seq transcriptomics reveals the battle for iron during Pseudomonas aeruginosa acute murine pneumonia. Sci Rep. 2016;6:39172.

19. Wong TY, Hall JM, Nowak ES, Boehm DT, Gonyar LA, Hewlett EL, et al. Analysis of the In Vivo Transcriptome of Bordetella pertussis during Infection of Mice. mSphere. 2019;4(2).

20. Lindsay JA, Holden MT. Staphylococcus aureus: superbug, super genome? Trends Microbiol. 2004;12(8):378-85.

21. Weigand MR, Peng Y, Loparev V, Johnson T, Juieng P, Gairola S, et al. Complete Genome Sequences of Four Bordetella pertussis Vaccine Reference Strains from Serum Institute of India. Genome Announc. 2016;4(6).

22. Pokorzynski ND, Thompson CC, Carabeo RA. Ironing Out the Unconventional Mechanisms of Iron Acquisition and Gene Regulation in Chlamydia. Front Cell Infect Microbiol. 2017;7:394.

23. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, et al. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science. 1998;282(5389):754-9.

24. Putman TE, Suchland RJ, Ivanovitch JD, Rockey DD. Culture-independent sequence analysis of Chlamydia trachomatis in urogenital specimens identifies regions of recombination and in-patient sequence mutations. Microbiology (Reading). 2013;159(Pt 10):2109-17.

25. Steuerman Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E, et al. Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. Cell Syst. 2018;6(6):679-91 e4.

26. O'Connell CM, Brochu H, Girardi J, Harrell E, Jones A, Darville T, et al. Simultaneous profiling of sexually transmitted bacterial pathogens, microbiome, and concordant host response in cervical samples using whole transcriptome sequencing analysis. Microb Cell. 2019;6(3):177-83.