

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

2008

Mobile genetic elements in *Coxiella burnetii*: Friends, foes or just indifferent?

rahul raghavan
The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

raghavan, rahul, "Mobile genetic elements in *Coxiella burnetii*: Friends, foes or just indifferent?" (2008).
Graduate Student Theses, Dissertations, & Professional Papers. 668.
<https://scholarworks.umt.edu/etd/668>

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

Mobile genetic elements in *Coxiella burnetii*: Friends, foes or just indifferent?

By

Rahul Raghavan

M.Sc (Med), Kasturba Medical College, Manipal, India 1998

Presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

The University of Montana

Summer 2008

Chairman, Board of Examiners

Dean, Graduate School

Date

Mobile genetic elements in *Coxiella burnetii*: Friends, foes or just indifferent?

Director: Dr. Michael F. Minnick

The genome of the obligate intracellular pathogen *Coxiella burnetii* contains a large number of mobile genetic elements including two group I introns and an intervening sequence (IVS) that interrupt the 23S rRNA gene; an intein within *dnaB* (encoding replicative DNA helicase) and a homing endonuclease. The introns are self-splicing ribozymes and able to inhibit ribosome function and retard bacterial growth through internal guide sequence (IGS)-dependent and –independent mechanisms. The introns were found to be highly conserved in all eight genomic groups of *C. burnetii*, suggesting a role in *C. burnetii*'s biology. It is not clear whether the introns are being positively selected because they promote bacterial persistence or whether they are slightly deleterious but were fixed in the population due to genetic drift.

The intein is able to self-splice, leaving the host protein intact and presumably functional. Similar inteins were found in two extremophilic bacteria (*Alkalilimnicola ehrlichei* and *Halorhodospira halophila*) that were found to be distantly related to *Coxiella*. Intein splicing appears to be a slow process, making it possible that before the intein is excised the DnaB precursor is non-functional, thereby reducing the pool of mature DnaB, thus creating a lag in replication and slowing growth.

The IVS is found inserted into helix 45 of *C. burnetii*'s 23S rRNA. Unlike introns and inteins, the flanking regions are not spliced back together after the IVS is excised, resulting in a fragmented 23S rRNA. The IVS encodes a hypothetical protein that is conserved between IVSs in a wide variety of bacteria. Phylogenetic analyses revealed that a similar ORF is present on an IVS inserted at the same locus in many *Leptospira* species, suggesting horizontal gene transfer as the mode of spread of this genetic element. It is possible that the dramatic drop in ribosome quantities that occurs when *Coxiella* transitions from LCV to SCV is aided by RNA fragmentation caused by the IVS. Another mobile genetic element found in *C. burnetii* is a homing endonuclease encoded within an intron. Here, we have characterized four “selfish” genetic elements in *C. burnetii* and show that they share an intimate relationship with their host.

ACKNOWLEDGMENTS

I sincerely thank my mentor, Mike Minnick for his guidance and friendship. I am grateful to him for letting me choose my own path yet not letting me fall off the edge. I want to thank my committee members, Scott Samuels, Steve Lodmell, Michele McGuirl and Keith Parker for their support and guidance through out the last four years. I am indebted to my comprehensive exam committee - Bill Granath, Scott Samuels, Frank Rosenzweig, Scott Wetzal and Jean Pfau - for helping me progress through that hurdle unscathed. I would like to thank Jim Battisti, the yin to Mike's yang, for keeping me grounded, and Sherry Coleman for her generosity and technical assistance. I am especially lucky to have worked with such talented technicians as Laura Smitherman, Kate Sappington and Linda Hicks. I thank them for making our lab a wonderful place to work in. I thank Nermi Parrow, my fellow traveler, for her thoughtfulness and for all the invigorating discussions. I dedicate this work to my grandma, parents, siblings and Kiran for their love and unwavering belief in my abilities, which makes everything possible.

TABLE OF CONTENTS

Chapter		Page
1	Introduction to <i>Coxiella burnetii</i> and Parasitic Genetic Elements	1
	I. <i>Coxiella burnetii</i>	1
	A. Introduction	1
	B. Historic Background	1
	C. Developmental Cycle	2
	D. Taxonomy and Genetic Diversity	3
	E. Epidemiology and Disease	5
	F. Diagnosis, Treatment and Prevention	6
	G. A Potential Bioweapon in Our Backyard	7
	H. Genome and Evolution	8
	II. Parasitic Genetic Elements	9
	A. Introduction	9
	B. Group I Introns	10
	C. Inteins	12
	D. Homing Endonucleases	15
	E. Intervening Sequences	18
	III. Research Significance and Goals	19
	A. Raison d'être	19
	B. Significance	20
	C. Research Goals	21
2	The Unusual 23S rRNA of <i>Coxiella burnetii</i>: Two Self-Splicing Group I Introns Flank a 34-Base-Pair Exon, and One Element Lacks the Canonical ΩG	26
	A. Abstract	26
	B. Introduction	27
	C. Materials and Method	29
	D. Results	32
	E. Discussion	42
	F. Acknowledgements	46
	G. References	47
3	Toxic Introns and Parasitic Intein in <i>Coxiella burnetii</i>: Legacies of a Promiscuous Past	50
	A. Abstract	50
	B. Introduction	51
	C. Materials and Methods	53
	D. Results	60
	E. Discussion	71
	F. Acknowledgements	83
	G. References	83

4.	Intron Toxicity is Caused by Both Sequence-Dependent and Sequence-Independent Mechanisms and Pentamidine Inhibits <i>Coxiella burnetii</i> Growth in Vero Cells	89
	A. Introduction	89
	B. Materials and Methods	90
	C. Results	94
	D. Discussion	103
5.	General Discussion	111
	References to Chapters One, Four and Five	117
	Appendix: Plasmid Vectors Used in This Study	127

LIST OF TABLES

Chapter	Table	Description	Page
2	2-1	Group I introns are conserved among eight genomic groups of <i>C. burnetii</i>	43
3	3-1	PCR primers and their targets	57
4	4-1	PCR primers and their targets	92
	4-2	Plasmid constructs used in this study	108

LIST OF FIGURES

Chapter	Figure	Description	Page
1	1-1	Intein modular structure	24
	1-2	Homing cycle	25
2	2-1	<i>C. burnetii</i> 23S rRNA gene linkage map	35
	2-2	In vivo analysis of introns and IVS	36
	2-3	In vitro analysis of introns and IVS	37
	2-4	Predicted secondary structures of introns	39
	2-5	Phylogenetic analyses of introns	40
3	3-1	Effect of <i>Coxiella</i> ribozymes on <i>E. coli</i> growth	61
	3-2	Ribozyme-ribosome association	63
	3-3	Ribozymes inhibit in vitro translation	66
	3-4	Inverse correlation between genome and ribozymes	67
	3-5	Intein splicing	70
	3-6	Phylogenetic analyses	72
	3-7	Potential ribozyme-ribosome interactions	76
4	4-1	Effect of introns lacking IGS on <i>E. coli</i> growth	96
	4-2	In vitro translation with IGS-less introns present	98
	4-3	Effect of IGS on <i>E. coli</i> growth	101
	4-4	Pentamidine inhibition of <i>Coxiella</i> growth	102
	4-5	Predicted secondary structure of IVS	104
	4-6	Phylogenetic analysis of CBU_2096	105

CHAPTER ONE

Introduction to *Coxiella burnetii* and parasitic genetic elements

I. *Coxiella burnetii*

A. INTRODUCTION

Coxiella burnetii is a gram-negative, obligate intracellular γ -*Proteobacterium*.

This category B select agent has a biphasic developmental cycle that alternates between a fragile, metabolically-active large-cell variant (LCV) and a durable, dormant small-cell variant (SCV). *C. burnetii* is distributed almost worldwide and has a broad range of susceptible hosts (86). Human infections are acquired through inhalation of aerosols of animal origin and can cause a wide array of conditions ranging from asymptomatic infection to an acute self-limiting febrile illness (Q fever) or severe, chronic diseases like endocarditis and hepatitis (26). Due to the lack of genetic systems, molecular details about the pathogenic mechanisms used by this highly infectious agent are not well understood. Consequently, safe and effective vaccines and therapeutic regimens are not widely available (99). Recently the entire genomes of several strains of *C. burnetii* were sequenced and the information gleaned from them has led to a deeper understanding of the pathogen's physiology and evolutionary history (117). *Coxiella*'s genome is thought to be undergoing reductive evolution but, interestingly, contains a large number of selfish genetic elements and 'young' pseudogenes, suggesting a recent shift from the extracellular environment to its current intracellular niche.

B. HISTORIC BACKGROUND

In 1937, Derrick proposed the name “Q fever” (Q for query) to describe a febrile illness of unknown etiology in Australian slaughterhouse workers (29). At about the same time, independent of Derrick’s work, Davis and Cox at the Rocky Mountain Laboratory in Hamilton, Montana identified a filterable “Nine Mile agent” from ticks, which caused a febrile illness in guinea pigs (28). In a remarkable mix of serendipity and science, the Q fever agent and the Nine Mile isolate were demonstrated to be identical by cross-immunity (128). Since the pathogen was isolated from ticks, was filterable and did not grow in axenic media, it was initially named *Rickettsia burnetii* (30). Although the name was later changed to *Coxiella burnetii*, the bacterium continued to be classified as a α -*Proteobacterium* in the *Rickettsiales* order along with genera *Rickettsia* and *Rochalimae* until recently, when it was moved to the gamma subdivision of *Proteobacteria* (131).

C. DEVELOPMENTAL CYCLE

C. burnetii replicates only within eukaryotic cells. However, unlike most obligate intracellular pathogens, *C. burnetii* can survive for months in the extracellular environment, where it is notoriously resistant to high temperature, ultraviolet light, various disinfectants and desiccation (87, 133). The durable qualities of *Coxiella* was mistakenly thought to be due an endospore by some earlier investigators (88). However, these traits are actually conferred by a biphasic developmental cycle that alternates between a durable and dormant SCV and a fragile, metabolically-active LCV. Although both forms can cause infections in vitro (132), most infections undoubtedly arise through contact with the SCV in nature. Once taken up by a host cell (e.g., an alveolar macrophage), the *Coxiella*-containing phagosome transits through the endocytic pathway

to finally fuse with the lysosomal compartment. SCV is activated by the acid pH of the developing vacuole (pH ~4.5), termed a parasitophorous vacuole (PV), which is indistinguishable from a secondary lysosome (127). Even though *Coxiella* prospers in the lysosomal environment, which is normally toxic, it causes autophagy-mediated delay of phagosome maturation. This phagosome stalling might allow the delivery of nutrients and other factors needed for SCV to develop into LCV over the course 1-2 days (22). The LCV is metabolically active, contains an extended chromosome, and replicates every 10-12 h to form additional LCVs. Following a 4-d exponential phase, LCVs subsequently convert to SCVs by condensing their chromosome and becoming metabolically quiescent (22). The PV content eventually converts to an SCV-dominated population and can enlarge to occupy nearly the entire cytoplasm pushing organelles to the cell's periphery, ultimately causing lysis. Once released, SCV's repeat the infection cycle. Bacterial factors that contribute to development, dormancy and slow growth rate are poorly characterized, in spite of the essential roles that these processes play in virulence, host chronicity and environmental persistence of *C. burnetii* (113).

D. TAXONOMY AND GENETIC DIVERSITY

It is of considerable importance to identify other bacteria closely related to *C. burnetii* due to its unique characteristics: it is the only bacterium known to thrive in the acidic phagolysosomal vacuoles of host cells; it has an unusual biphasic life cycle that involves a distinctive spore-like SCV; it is the only bacterium known to contain multiple introns within a gene and even though it is dependent on host cells for replication and growth, it can survive for a long period outside the cell. Earlier evolutionary studies

based on morphological characteristics grouped *C. burnetii* with *Rickettsia*, but recent phylogenetic analyses using 16S rRNA sequences have placed it within gamma *proteobacteria* (130). Currently, *C. burnetii* belongs to the order *Legionellales* with *Legionella* and *Rickettsiella* as its closest known relatives (111). *C. burnetii* is the only recognized species of its genus even though a number of bacteria with similar 16S sequences have been isolated from ticks and crustaceans (65). *C. burnetii* isolated from disparate geographic sources and various hosts show considerable genetic diversity. Using restriction fragment-length polymorphism and whole-genome microarray-based comparison, *C. burnetii* has been categorized into eight distinct genomic groups (I to VIII) (Table 1, Chapter 2), (10,56). Interestingly, the type of human disease caused by isolates in each group is similar; e.g. groups I, II and III contain only isolates from acute disease patients whereas group IV and group V isolates were all derived from chronic disease patients (62). Most isolates contain one of four large, low-copy plasmids called QpH1, QpRS, QpDG and QpDV. In some isolates (e.g. *C. burnetii* Scurry Q217, genomic group V) QpRS-like plasmid sequences are integrated into the chromosome (114). Even though plasmid functions have not yet been elucidated, the absolute conservation of plasmid sequences in all isolates suggests a critical role for the plasmid in some facet of *C. burnetii*'s biology. *C. burnetii* also displays genetic diversity in the production of lipopolysaccharide (LPS), the only defined virulence factor (50). *C. burnetii* isolates from natural sources with full length LPS are defined as "phase I". It is highly infectious and corresponds to the smooth colony variant observed in *Enterobacteriaceae*. Serial in vitro passage of phase I *Coxiella* in embryonated eggs and tissue culture results in phase II, avirulent bacteria with truncated LPS. Phase II LPS has

intact lipid A and some core sugars but O-antigen sugars are absent (51). The genetic basis for this phase variation is the lack of genes involved in biosynthesis of O-antigen sugars, including the rare sugar virenose (124). Recently, multispacer sequence typing has been used to place 173 isolates into 30 different genotypic groups (44).

E. EPIDEMIOLOGY AND DISEASE

Q fever is a zoonosis that has been reported in almost all countries except New Zealand (57). Reservoirs include many wild and domestic animals, birds and arthropods such as ticks (8). Animals are often chronically infected but asymptomatic, and shed *C. burnetii* into environment mainly during parturition. Human infection occurs mainly by breathing infectious aerosols or dust contaminated with birth fluids of domestic ruminants (8). Ingestion of raw milk is probably a minor route for the transmission of *C. burnetii*. Infrequently, sporadic cases of human Q fever have occurred after contact with an infected patient or by tick bite (34).

The infectious dose is estimated to be ten bacteria or fewer (123). Incubation periods can vary from a few days to several weeks, depending on the inoculating dose. *C. burnetii* infection remains asymptomatic in a majority of cases (107). Acute Q fever usually presents as an influenza-like illness with varying degrees of pneumonia, hepatitis, cough, nausea, vomiting, myalgia and arthralgia. Fatalities are rare with less than 1% resulting in death (99). Although the acute disease is self-limiting and recovery uneventful, *C. burnetii* persist in the body for a prolonged period of time and some patients (~9%) can develop a chronic debilitating disease (128). Patients with prior

coronary disease or immunocompromised state are more at risk for developing chronic Q fever. Endocarditis is the most severe manifestation of chronic Q fever and accounts for 60-70% of cases (99). Chronic endocarditis presents with intermittent fever, cardiac failure, hepatomegaly and splenomegaly. Even with timely antibiotic treatment, Q fever endocarditis has a very high mortality rate of up to 60% (107). Q fever fatigue syndrome is a non-fatal but debilitating complication seen in around 10% of people recovering from acute Q fever. It is characterized by fatigue and body and joint pain and is thought to be caused by high levels of cytokines including IL-6 and IL-10 (84). Q fever, if contracted during pregnancy, can be asymptomatic but may lead to low birth weight, premature birth, miscarriage or neonatal death (99).

F. DIAGNOSIS, TREATMENT AND PREVENTION

The non-specific, often flu-like symptoms make it difficult to suspect and diagnose Q fever. Moreover, the disease is endemic in much of the world thereby making serological titers difficult to interpret. Thus, it is very important to detect IgM antibodies and to demonstrate seroconversion using paired samples. The reference method for serological diagnosis is indirect immunofluorescence (IF) (40). In acute Q fever, IgM titers in excess of 1:50 against phase II antigen are diagnostic, whereas high titers against both phase I and phase II antigens are seen during chronic disease (40). Complement fixation (CF) and enzyme immunoassay (EIA) are also routinely used to diagnose *C. burnetii* infections (39).

Acute Q fever is usually self-limiting, but doxycycline (100 mg twice daily for 14 days) helps to reduce the duration and severity of symptoms (33). Chronic Q fever is

very difficult to treat and needs prolonged antibiotic therapy (86). Doxycycline, 100 mg twice daily and hydroxychloroquine (which raises the pH in the PVs), 200 mg, thrice daily for eighteen months, is the recommended therapeutic regimen (106).

Since Q fever is a zoonotic disease, control of the disease in animals will influence its prevalence in humans. Occupationally at-risk individuals like abattoir workers and researchers should ideally be vaccinated but no universally approved, safe vaccines are available. In Australia, a formalin-inactivated whole-cell vaccine (Q-Vax,) is available (70). Although highly immunogenic, this vaccine may induce adverse reactions in pre-sensitized individuals, so it is very important to perform pre-vaccination screening that includes history, skin test and serology. Other available vaccines include an acellular trichloroacetic acid extracted vaccine from the former Czechoslovakia (19), a chloroform-methanol residue vaccine from USA (129), and a live attenuated vaccine from the former USSR (42).

G. A POTENTIAL BIOWEAPON IN OUR BACKYARD

People all over the world, especially farming communities like Montana, are at risk for exposure to *C. burnetii*. Despite its apparent mild clinical symptoms, *C. burnetii* fulfills all requirements for a biological weapon: it is extremely infectious; it can be manufactured on a large scale and remains stable under production, storage and transportation conditions; it can be disseminated effectively as an aerosol and remains viable for a long time after dissemination (83). The Centers for Disease Control and Prevention (CDC) has classified *C. burnetii* as a category B select agent and it was listed among agents that can be used as weapons by the US government during its offensive

biological program during the 1940s (110). Even though *Coxiella* lacks the capacity to cause massive fatalities potentially caused by category A agents like small pox, anthrax, plague and viral hemorrhagic fevers, the degree of infectivity can rival that of anthrax and can lead to extensive acute and chronic morbidity (14).

H. GENOME AND EVOLUTION

The sequencing of the whole genome of *C. burnetii* (RSA493) in 2003 (117) was an important step in our understanding of this recalcitrant pathogen. The circular chromosome was found to be 1,995,275 bp in length with a G+C content of 42.6%. Of the 2095 open reading frames (ORFs) annotated, 693 (~33%) encode hypothetical proteins with no significant matches to other sequenced genes. 83 pseudogenes (genes disrupted by point mutations, frameshifts or truncations) and 29 insertion sequences (ISs) were also identified. The genomes of four other *C. burnetii* strains (Dugway, RSA331, RSA334 and MSU Goat) are being sequenced in the hope of identifying genetic markers characteristic of each strain (9, 118).

The availability of whole genome sequences has facilitated greater insights into the evolutionary history of *C. burnetii*. Many factors suggest that an ancestor of *C. burnetii* recently moved from the extracellular environment to the intracellular niche. Reductive evolution is a trait shared by all host-associated bacteria (93). The genome-reduction process seems to have just begun in *C. burnetii* as suggested by the high coding frequency (89%) when compared to that of other obligate intracellular bacteria, like *R. prowazekii* and *Mycobacterium leprae* (both 76%) (117). Genome size reduces in intracellular bacteria by the conversion of functional genes into pseudogenes

and later loss of these pseudogenes due to lack of selection (47). In *Coxiella*, most of the identified pseudogenes are ‘young’ (disrupted by a single frameshift), again suggesting that the process has just started. Another intriguing aspect of *Coxiella*’s biology uncovered by whole-genome sequencing is the large number of selfish/parasitic genetic elements found strewn across the genome. The high number of IS elements is a hallmark of bacteria that have recently shifted to a host-associated life style (93). In addition to the ISs, the sole 23S rRNA gene is interrupted by two group I introns and an intervening sequence (IVS), and the C-terminal region of the vital replicative DNA helicase gene (*dnaB*) contains an intein. Since these genetic elements gain access to a genome through horizontal gene transfer (HGT), it was surprising to find them in *C. burnetii*, whose intracellular location would likely preclude HGT. These selfish genes may have invaded *Coxiella*’s genome from co-infecting bacteria or more likely they may be relics from *C. burnetii*’s past as an environmental bacterium with greater opportunities for HGT (104).

II. PARASITIC GENETIC ELEMENTS

A. INTRODUCTION

Genomes of most organisms harbor DNA of alien origin that serve no known function; in fact most of the human genome consists of such ‘junk’ or ‘selfish’ DNA (77, 125). Even compact bacterial genomes that are packed with functional genes contain such selfish genetic elements underlining their universality. Since these elements do not contribute to host’s fitness but at the same time utilize host resources for their

perpetuation, it is appropriate to consider them as molecular parasites (7). A variety of parasitic genetic elements have been observed in all three domains of life (Eukarya, Bacteria and Archaea). Here I will focus on group I introns, inteins, intervening sequences (IVSs) and homing endonucleases (HEs) found in *C. burnetii*.

B. GROUP I INTRONS

Introns are non-coding, intragenic regions that are removed before the mature RNA is formed by splicing the exons (coding regions that flank introns) together. They are much more common in eukaryotes than in bacteria (60). Based on splicing mechanism, introns are classified into four classes: group I, group II/group III, spliceosomal and tRNA/archaeal introns (55). Spliceosomal introns are seen in eukaryotes and utilize spliceosomes (large protein-RNA complexes) for splicing (79), whereas tRNA introns splice with the help of specialized enzymes (82). Group I and group II introns are able to self-splice without the aid of any proteins (ribozymes) (121).

Group I introns are small RNAs (~ 250 to 500 nt in size) that have invaded protein, rRNA and tRNA coding genes in a variety of organisms, including algae, fungi, lichens, some lower eukaryotes and a few bacteria (55). In 1982, Thomas Cech and colleagues demonstrated the self-splicing ability of a group I intron from *Tetrahymena thermophila*, the first ribozyme to be described (74). Ribozymes are considered legacies of a primordial RNA world, where RNA possessed both information-encoding and catalytic properties, before the advent of DNA and protein-based life forms (43). All group I introns share a conserved secondary structure, which consists of paired elements (P) that assist in self-splicing by using a guanosine molecule as a cofactor (126). P4-P5-

P6 and P3-P7-P9 form two separately folding helices within the core (Figure 4, Chapter 2). Helix P3-P7-P9 contains the binding site for the guanosine molecule (G-binding site, GBS) and is the minimal catalytic domain required for splicing (61). P1 and P10 are complementary to 5' and 3' exons, respectively, and are together called the internal guide sequence (IGS) (136). In the first step of splicing, the 3'-OH group of the guanosine (or GMP or GTP) bound to GBS attacks the 5' splice site which is marked by a conserved G•U wobble pair formed when P1 binds to 5' exon. After the first step, this guanosine is covalently bound to the free 5' end of the intron and leaves the GBS. All known group I introns have a conserved terminal guanine called omega (Ω) G (See Chapter 2 for exception), which now occupies the GBS. In the second splicing step, the 3'-OH group of the free 5' exon attacks the 3' splice site (Ω G, now in GBS). This is a reaction that is chemically equivalent to the reverse of step 1, which results in the ligation of 5' and 3' and the release of the intron (120). The excised introns have been observed to circularize, but the significance of this property is not clearly understood (96). Also, reversal of the splicing process has been noted, which has led to the hypothesis that intron mobility can occur via reverse splicing (109).

The broad but sporadic distribution of group I introns suggests that they spread from one host to another through HGT. These parasitic elements move site-specifically to an intronless allele of the same gene by a process termed homing, using an encoded homing endonuclease (HE) (21) (see below for a discussion on HEs and homing). Another possible mechanism of intron mobility is by reverse splicing. In this model, intron RNA invades the target RNA by reverse splicing and then becomes inserted into the corresponding gene through reverse transcription coupled with

recombination (109). Since group I introns splice out before the protein is synthesized, they are thought to be evolutionarily neutral. Once they are fixed in a population, they tend to accrue mutations in non-functional regions and ultimately become lost by deletion (66).

In addition to the ability of group I introns to splice flanking exons together, it's been shown that they can mediate the splicing of exons contained on different RNAs (trans-splicing) (71). This property has potential as a gene therapy tool. In fact, group I introns have been used to convert sickle β -globin transcripts into RNAs encoding γ -globin in an effort to treat sickle cell disease (76). Another potential application of group I introns is as a genetic tool to knock-down expression of genes in bacteria like *C. burnetii*, in which disruption of genes has not yet been achieved. Here, group I introns that are tailored against a specific gene, could cleave and render the target mRNA nonfunctional (Raghavan and Battisti, unpublished results).

C. INTEINS

Similar to introns, another dogma-breaking discovery was that of genes that contain internal regions that were removed at the protein level, rather than as RNA. These elements are called inteins (internal proteins). They are transcribed and translated together with the host protein, but self-excise leaving flanking sequences (exteins) spliced together. The first intein was discovered in yeast vacuolar ATPase genes in 1990 (69). Since then hundreds of inteins have been discovered in Eukaryotes, Archaea and Bacteria. An intein database, InBase has been set up, which provides information on all known inteins (100). Inteins have been described in a variety of genes and the ratio of intein size to host protein size varies widely with some inteins four times as large as the

host protein and others only one-tenth the size of the host protein (100). Although the exact origin of inteins is not clear, the available evidence seems to suggest that inteins are extremely ancient and that all inteins originated from a single source (102). The protein-splicing domain of inteins is homologous to the C-terminal domain of the Hedgehog developmental proteins (Hog domain) found in all multicellular organisms; from nematodes to mammals (52).

Inteins have a modular organization consisting of three functional domains (Fig. 1.) (103). The N-terminal domain is made up of four motifs, A, B, N2 and N4. The C-terminal domain contains two motifs, F and G. A central endonuclease domain consists of four motifs, C, D, E and H. The N- and C-terminal motifs are involved in protein splicing and are found conserved in all known inteins, whereas motifs C, D, E and H are found only in some inteins that encode a HE. The first and the last amino acids of the intein and the first amino acid of the N-terminal extein are involved in the splicing reaction (46). The first amino acid (motif A) in all inteins is cysteine (C) or serine (S); the terminal intein amino acid is a conserved asparagine (N), and the first amino acid that follows the intein is C, S or threonine (T). Intein splicing involves four successive nucleophilic displacements (100). The first step is an N-O or N-S acyl shift where the OH or SH side chain of the N-intein amino acid (S or C) attacks the carbonyl carbon of the preceding amino acid to generate an ester/thioester intermediate linking the N-terminal extein to the side chain of the first intein amino acid, thereby breaking the peptide bond between N-extein and the intein. The second step is a transesterification, where the OH or SH side chain of the first C-extein amino acid (S, T or C) attacks the N-terminal ester/thioester bond formed in step 1. This results in the transfer of the N-extein to the

side chain of the first C-extein amino acid (S, T or C), forming a branched intermediate. In the third step, the peptide bond between the intein and the C-extein is broken by cyclization of the conserved C-terminal asparagine to form a succinimide, thereby resulting in intein excision. The N-extein is now attached via an ester bond to the side chain of the first C-extein amino acid. In the final step, the ester bond rapidly undergoes an acyl rearrangement to the thermodynamically more stable, normal peptide bond. In most inteins, the amino acid preceding the terminal asparagine is a histidine (H), which is thought to assist in asparagine cyclization. Part or all of the remaining residues are also important for proper intein folding to generate the active site (24, 25).

No intein function beneficial to a host has been identified, so they are thought to be under constant threat of elimination through negative selection. Hosts apparently “tolerate” inteins because they have negligible impact; moreover, it is not easy to get rid of inteins because there is no specific host mechanism for their excision. Since inteins are seen in highly conserved genes, their excision should be precise, otherwise, a deletion or insertion could occur in the crucial host-protein and would be selected against. Although precise deletions are very rare, eventually they will occur and the intein will be lost from the population. To prevent this, most inteins encode a HE (motifs C, D, E and F). Similar to group I introns, inteins can invade a new susceptible population using a process termed homing (see below for more on HEs and homing).

The protein splicing ability of inteins has been exploited as a biotechnology tool. Inteins can be used as tags to purify fusion proteins in place of traditional histidine tags. After purifying, the intein tag can be removed utilizing the self-cleaving property of the intein (135). This technique has been commercialized by New England Biolabs

(IMPACTTM). Other potential uses include the application of inteins to the semi-synthetic synthesis of cytotoxic proteins (35) and for introducing nuclear magnetic resonance (NMR) labels into part of a large protein (98).

D. HOMING ENDONUCLEASES (HEs)

The presence of a parasitic genetic element in a host represents a snapshot in the continuing struggle between the host's ability to purge the element and the element's ability to persist. One of the most successful parasitic genetic elements is the HE. They are simple and elegant parasitic elements that consist of only a single gene that encodes a single protein (16). HEs are small (< 40 kDa) proteins that recognize and cleave highly specific DNA target sequences without being overly toxic to the host organism. Based on conserved sequence motifs, HEs have been classified into three families (75). (1) The LAGLIDADG family has one or two conserved motifs called the dodecapeptide motifs with a consensus LAGLIDADG sequence (27). Most HEs found to date belong to this family. (2) The $\beta\beta\alpha$ -Me family contains two sub-families. The His-Cys box sub-family contains a ~30 amino acid region with two histidines and three cysteines (67), whereas the H-N-H sub-family contains a ~30 amino acid region with conserved histidine and asparagine residues (119). (3) The GIY-YIG family contains conserved 'GIY' and 'YIG' tripeptides flanking a ~10 amino acid segment (73).

In order to limit the negative impact on a host, and being negatively selected, HEs have associated themselves with other self-splicing elements like group I introns and inteins that are nearly neutral to selection (12). Together, HEs and introns/inteins have formed a successful, mutually beneficial partnership where the HE

provides mobility to the intron/intein and the intron/intein provides a “safe haven” for HE. The process by which the whole element moves from one site to another is called homing (68). The homing mechanism requires the protein to be translated. When an intron/intein-containing allele comes together with an intron/intein-minus allele, the HE protein binds to a homing site composed of the flanking exon/extein sequences and cleaves it. The host repairs this double-strand DNA break using homologous recombination between the alleles, which results in the insertion of the parasitic element into the new target. Now this site is “immune” to HE cleavage because the insertion element disrupts the target sequence. HEs have highly specific target sequences that span up to 40 bp of DNA that typically occur only once per host, thus minimizing potential negative impact on the host (92). They tolerate some variation in their recognition sequence, which helps them to move from one host to another (21). Another reason for their success is their adaptability to the new host, which conceivably resulted in the observed divergent DNA binding regions in similar HEs from different hosts (20, 81).

HEs go through a dynamic cycle that includes invasion, fixation, inactivation, elimination and eventual reinvasion (Fig.2) (45). Once an HE-containing parasitic element invades a new species, it spreads to all the individuals in that population until no HE-minus alleles are available and becomes fixed in that population. Once it is fixed in a population, the HE becomes non-functional (since there are no available target sequences) and starts to degenerate and eventually becomes lost (17). In fact, a large number of group I introns and inteins, including one intron and the intein in *C. burnetii*, have lost their respective HE genes. The element itself will maintain its sequence because any change will affect its splicing ability, thus negatively impacting the host. Eventually

the whole element is lost from the population by a precise deletion event. The parasitic genetic element reappears in the population only through a new HGT event. Thus, HGT is critical for the long term maintenance of a parasitic genetic element in a population. To prevent being purged from a population, some HEs utilize an intriguing strategy: they have evolved a maturase function (115); maturases are enzymes that promote intron splicing by stabilizing RNA folding. These HEs have evolved an RNA binding site in addition to their DNA binding site to function as maturase, showing their adaptability (80). Some HEs even confer beneficial functions upon their hosts. VDE (also known as PI-*SceI*), a HE found inserted in a self-splicing intein in the *VMA-1* gene of *Saccharomyces cerevisiae*, is one of the main regulators of the high-affinity glutathione transporter (90). HO (F-*SceII*), a freestanding HE in *S. cerevisiae* mediates mating-type switching (18). Some HEs of the $\beta\beta\alpha$ -Me family act as colicins, which help *Enterobacteriaceae* to eliminate competition (63). Similarly, intron-encoded HEs help the archaea *Sulfolobus acidocaldarius* (1) and the *Bacillus subtilis* phage SP82 (49) to eliminate competition.

The ability to introduce specific double-strand DNA breaks makes HEs a very useful genetic tool. Some HEs are commercially available like I-Ceu I, I-Sce I and PI-Psp I (New England Bio Labs). HEs along with other rare-cutting restriction enzymes have been used to map a variety of bacterial genomes, especially to analyze chromosomal organization (78). HEs have been used to study double-strand break repair mechanisms in phages, yeasts, plants and mammalian cells and to study chromosomal repair systems in *Drosophila* (64), (48). Recently, artificial HEs have been engineered with the aim of using them in human gene therapy (23).

E. INTERVENING SEQUENCES (IVS)

Ribosomal RNA, when isolated from bacteria like *Salmonella typhimurium*, was observed to be fragmented (134). In 1990, it was shown that a novel genetic element called IVS, which was found inserted in the 23S rRNA, was responsible for this fragmentation (15). Since then, a large number of IVSs have been identified in a variety of bacterial species (37). IVSs are parasitic genetic elements that form a stem-loop structure due to the presence of complementary sequences at their ends (Fig. 5, Chapter 4). RNase III, a double-strand specific endoribonuclease that processes the 30S rRNA precursor into pre16S and pre23S molecules, cleaves IVS post-transcriptionally (38). After IVS cleavage, the two flanking sequences are not religated, resulting in a fragmented rRNA. Since the RNase III processing site is found in the middle of the IVS stem-loop, the base of the helix holds the rRNA pieces together without affecting rRNA function (72).

IVSs are highly variable in length and sequence. The shortest one (55 nt) was reported in the 23S rRNA of *Rhizobium capsulatus* (38), whereas the longest known IVS (759 nt) is found in 23S rRNA of some *Leptospira* species (105). The *Leptospira* IVS contains an ORF that encodes a protein of unknown function. Intriguingly, an IVS with a similar ORF is present at the same position in the 23S rRNA of *C. burnetii* (2). Amino acid sequences are very well conserved between the ORFs in *Leptospira* and *C. burnetii* suggesting some important role, maybe in IVS mobility. We attempted to characterize this protein and found it to be lethal to *E. coli*. Only clones with the ORF out of frame or in opposite orientation to the promoter were obtained. Similar ORFs are found in a

number of IVSs from a variety of hosts. Figure 6 in Chapter 4 depicts a phylogenetic tree showing the evolutionary relationship between them. The closest relative to *Coxiella*'s ORF is found in *Leptospira*, a spirochete. This patchy distribution suggests HGT as the means of IVS spread, but the mechanism is not understood. The complementary ends of IVSs are similar to the DNA inverted repeats seen in mobile genetic elements, suggesting recombination as a means of IVS entry into host gene.

Since IVS is found in many bacteria and the resulting rRNA fragmentation does not affect ribosome function, several benefits of having IVS have been suggested. High abundance of IVS RNA found during exponential growth may sequester RNA binding enzymes or ribonucleases, thereby regulating their intracellular level (91). Fragmentation of rRNA may create more targets for degrading ribonucleases and enhance the efficiency of rRNA decay. It has been shown that rRNA fragmentation due to IVS influences the rate of rRNA degradation in *Salmonella typhimurium* during the transition from log phase to stationary phase (58). Thus, rRNA fragmentation might provide some selective advantage to the bacteria by aiding in the regulation of ribosome concentration, especially when exposed to growth-limiting conditions. In *C. burnetii* the dramatic transition from metabolically active LCV to dormant SCV (116) involves rapid decay of ribosomes, which might be aided by IVS-mediated 23S rRNA fragmentation.

III. RESEARCH SIGNIFANCE AND GOALS

A. RAISON D'ETRE

Publication of the whole genome sequence of *C. burnetii* (117) in 2003 was an import landmark because it provided a window into the biology of this intriguing

bacterium. Among the panoply of interesting characteristics that were revealed, the most remarkable (at least for me) was the presence of a large number of parasitic genetic elements in *C. burnetii*'s genome. This was fascinating because according to conventional wisdom, *C. burnetii*'s intracellular niche should have made it impervious to invasion by such parasitic genetic elements. Moreover, *C. burnetii*, an obligate intracellular bacterium, was found to be undergoing reductive evolution caused by the loss of apparently superfluous genes resulting in a compact ~2 MBP genome. *C. burnetii* is the only bacterium known to contain all three types of parasitic genetic elements, i.e. introns, inteins and intervening sequences. This rare abundance of parasitic genetic elements in an obligate intracellular bacterium hints at some unknown underlying biological phenomena that might have molded this pathogen's biology. Following in the foot steps of Theodosius Dobzhansky, who famously claimed that "nothing makes sense in biology except in the light of evolution" (141), it has been my goal to illuminate some of the evolutionary forces that have caused the accumulation of these molecular parasites in a host-associated bacterium and to understand the impact these elements have on *C. burnetii*'s biology/life style.

B. SIGNIFICANCE

Since we know very little about *C. burnetii*'s biology at the molecular level, even 70 years after its discovery, any information that adds to our understanding is very significant (99). Characterizing these parasitic genetic elements is important because they make good targets for antibiotic therapy since they are not found in human cells (89). Also, identification of factors that influence *C. burnetii*'s biology will help us combat this

potential bioterrorism agent more effectively. As there are no systems available for the genetic manipulation of *C. burnetii*, another benefit of studying these parasitic elements is their potential to be used as genetic tools. Studying the evolutionary history of these parasitic elements will undoubtedly shed light on the evolutionary history of the pathogen itself, which we know very little about. Finally, since parasitic genetic elements are not thought to affect the biology of the host, it will be significant if any of the elements are found to be influencing *C. burnetii*'s biology.

C. RESEARCH GOALS

Although group I introns are abundant in mitochondria and chloroplasts of lower eukaryotes (55), they are relatively rare in bacteria, and until recently were unknown in bacterial structural RNAs. Recently, self-splicing group I introns have been reported in the genus *Thermotoga* (95) and a non-splicing group I intron was reported in the 23S rRNA of *Simkania negevensis* (36). A putative group I intron was detected in the 23S rRNA gene of *C. burnetii* when the whole genome was sequenced (117). Preliminary analyses of this reported intron in our laboratory revealed that actually two group I introns (Cbu.L1917 and Cbu.L1951) interrupt the sole 23S rRNA gene of *C. burnetii*. The **first goal** was to investigate whether both Cbu.L1917 and Cbu.L1951 are able to splice out of *Coxiella*'s pre-23S rRNA in vivo. Most group I introns are ribozymes that are able to self-splice without the aid of proteins. The **second goal** was to analyze whether *C. burnetii*'s introns can self-splice in vitro. The **third goal** was to define the exact exon-intron boundaries and to infer the secondary structures of these introns so as to accurately identify the intron sub-types. Group I introns spread from one population to

another through horizontal gene transfer. The probable source of the introns found in *C. burnetii*'s 23S rDNA can be identified by phylogenetic analyses, which was the **fourth goal**. The 23S rRNA gene of all eight genotypes of *C. burnetii* was analyzed for sequence variation and presence or absence of these introns as the **fifth goal**. The introns should be conserved in all genotypes if they have a global effect on *C. burnetii*'s biology or they will be a patchy distribution with different strains having introns at various stages of degeneration as observed in red and green algae. In an earlier study, *E. coli* expressing *Tetrahymena* 26S rDNA group I self-splicing intron was found to grow slower than wild type (97). Our preliminary data also showed that when expressed in *E. coli* both Cbu.L1917 and Cbu.L1951 caused similar growth retardation. Since *Coxiella* also has a very slow growth rate, the **sixth goal** was to analyze the ribosomes of both *C. burnetii* and *E. coli* expressing the introns to check whether the introns can associate with ribosomes to potentially interfere with protein synthesis. The **seventh goal** was to analyze introns' impact on *E. coli* ribosomes by using an *E. coli* S30 in vitro transcription/translation system in conjunction with pBEST*luc* (encoding luciferase) in the presence of ribozymes or control RNA followed by a Luciferase Assay. If the introns affect the growth rate of *Coxiella*, then their RNA quantities should inversely correlate to the rate of growth. My **eighth goal** was to quantify *C. burnetii* genomes and ribozymes over time in co-cultures using qPCR and qRT-PCR, respectively. The IGS of Cbu.L1917 and Cbu.L1951 are complementary to helix 69 and helix 71 of the 23S rRNA, respectively. The **ninth goal** was to test whether introns use IGS to target ribosomes by transforming *E. coli* with introns that lack IGS and with IGSs that have a non-relevant RNA fused to it in place of the respective introns.

In an earlier study, Afseth and Mallavia (2) had suggested that the 23S rRNA of *C. burnetii* is fragmented due to the presence of an IVS. The **tenth goal** was to confirm that *C. burnetii*'s 23S rRNA fragmentation is caused by the IVS cleavage in vivo but not in vitro. The **eleventh goal** was to construct the secondary structure of IVS to understand where it is inserted in the 23S rRNA and to predict the possible RNase III processing sites. The IVS encodes a protein (CBU_2096) with unknown function. Understanding the evolutionary history of this protein may aid in understanding its function. To this end, the **twelfth goal** was to reconstruct the phylogenetic history of this protein.

When *C. burnetii*'s genome was sequenced, a putative intein was identified in the C-terminal region of the replicative DNA helicase (*dnaB*) gene. Since a non-functional intein could convert a functional gene into a pseudogene (47), and given the pseudogenization going on in *C. burnetii* (117), my **thirteenth goal** was to assess the intein's functionality. The intein was expressed in *E. coli* and protein splicing analyzed by SDS-PAGE. The **fourteenth goal** was to build phylogenetic trees to understand the evolutionary history of *C. burnetii*'s intein.

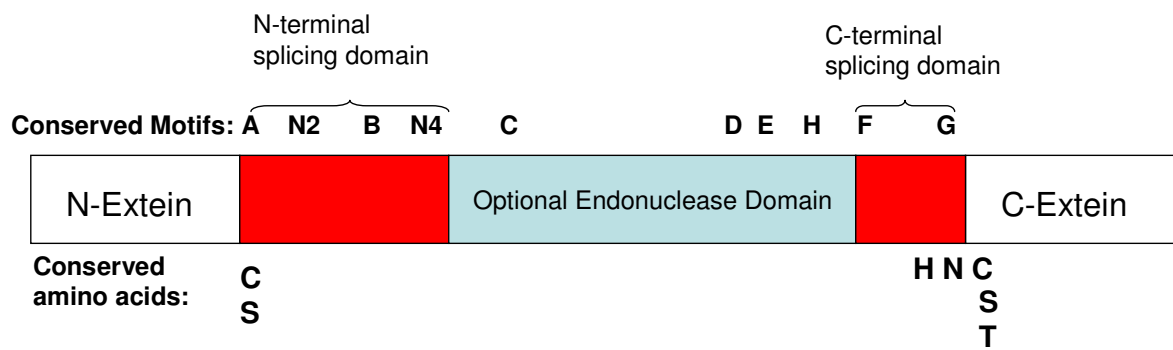


FIG. 1-1. Intein modular structure. An intein with flanking exteins is shown. The N- and C-terminal splicing domains (in red) and the optional endonuclease domain with their conserved motifs are shown. Conserved amino acids C or S at the 5' end of the intein, N at the 3' end of the intein and C, S or T at the first position on the 3' extein are also indicated.

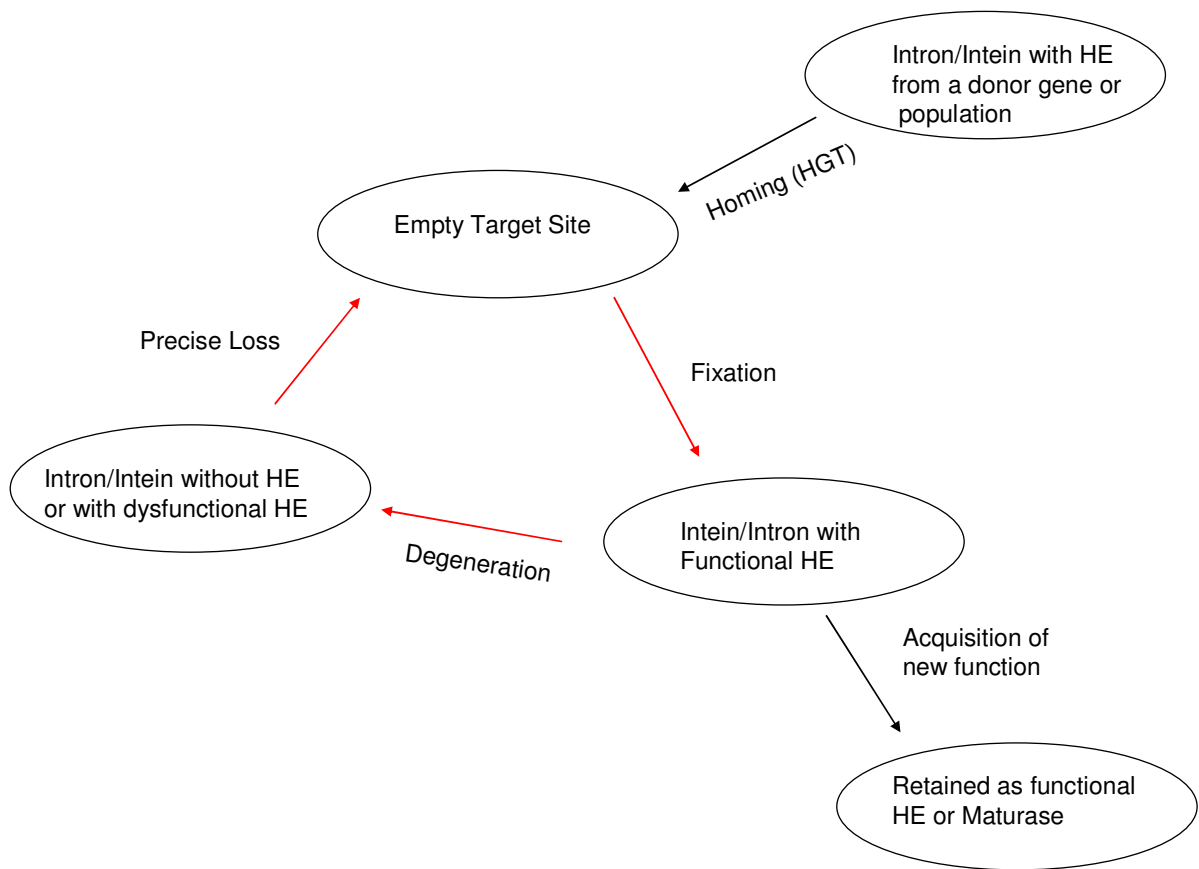


FIG. 1-2. Cyclic model of HEG gain, degeneration and loss or retention within host lineages. An empty site in a genome is invaded by an intron- or intein-associated HE via HGT. Once it is fixed in a population, due to the non availability of target sites, the HE starts to degenerate resulting in HE-less elements or elements with non-functional HEs. A precise deletion of the intron or intein leads to an empty site and the cycle continues through a new HGT event.

CHAPTER TWO

The unusual 23S rDNA of *Coxiella burnetii*: Two self-splicing group I introns flank a 34-bp exon and one element lacks the canonical omega G

Rahul Raghavan, Scott R. Miller, Linda D. Hicks and Michael F. Minnick. 2007.
J. Bacteriol. **189**: 6572-6579

A. ABSTRACT

We describe here the presence and characteristics of two self-splicing group I introns in the sole 23S rDNA of *Coxiella burnetii*. The two group I introns, Cbu.L1917 and Cbu.L1951 are inserted at sites 1917 and 1951 (*E. coli* numbering), respectively, in the 23S rDNA of *C. burnetii*. Both introns were found to be self-splicing in vivo and in vitro even though the terminal nucleotide of Cbu.L1917 is adenine and not the canonical conserved guanine, termed Ω G, as in Cbu.L1951 and all other group I introns described to date. Predicted secondary structures for both introns were constructed and revealed that Cbu.L1917 and Cbu.L1951 were group IB2 and group IA3 introns, respectively. We analyzed strains belonging to eight genomic groups of *C. burnetii* to determine sequence variation, presence or absence of the elements and found both introns to be highly conserved ($\geq 99\%$) amongst them. Although phylogenetic analysis didn't identify the specific identities of donors, it indicates that the introns were likely acquired independently; Cbu.L1917 from other bacteria like *Thermotoga subterranea* and Cbu.L1951 from lower eukaryotes like *Acanthamoeba castellanii*. We also confirmed the fragmented nature of mature 23S rRNA in *C. burnetii* due to the presence of an intervening sequence. The presence of three selfish elements in *C. burnetii*'s 23S rDNA

is very unusual for an obligate intracellular bacterium and suggests a recent shift to its current lifestyle from a previous niche with greater opportunities for lateral gene transfer.

B. INTRODUCTION

Group I introns are a distinct class of catalytic RNAs (ribozymes) that are considered a legacy of a primordial RNA world (40). They are able to self-splice by means of a two-step transesterification reaction using a guanosine molecule as a cofactor. The catalytic structure that is conserved among all group I introns consists of a specific arrangement of about 10 paired (P) elements that are capped by loops (L) and connected by junctions (J) (38). The core of each intron is comprised of two separately-folding helical domains made up of P4-P5-P6 and P3-P7-P9 (See Figures 4A and 4B). The P4-P6 domain structurally supports the P3-P9 domain, which contains the active site (17). Splice sites are determined by helix P1, which pairs with the 5' substrate strand (5' exon) and by helix P10, which pairs with the 3' substrate strand (3' exon). A conserved G•U wobble pair contributes to recognition of the 5' splice site, whereas the 3' splice site is recognized in part by a conserved terminal guanine, termed Ω G (23). Based on conserved secondary structure characteristics, Group I introns are further classified into 13 subgroups (23, 34). These selfish genetic elements are distributed widely in nature, albeit with a bias towards fungi, plants and red or green algae (14). Although group I introns are abundant in mitochondria and chloroplasts of lower eukaryotes, they are relatively rare in bacteria, and until recently were unknown in bacterial structural RNAs. Group I introns in 23S rDNA have been reported in hyperthermophilic bacteria of the genus *Thermotoga* (26) and in *Simkania negevensis*, a *Chlamydiales* member (7). A putative group I intron was detected in the sole 23S rDNA of *Coxiella burnetii* (Nine Mile phase I ; RSA493) when

the entire genome was sequenced (31); in addition, an intervening sequence (IVS) that produces a fragmented 23S rRNA has earlier been reported by Afseth et al. (1).

C. burnetii, the etiological agent of human Q fever, is a gram-negative obligate intracellular bacterium. This category B select agent is distributed almost worldwide and has a broad range of susceptible hosts including arthropods, fish, birds and wild and domestic mammals (6). Human infection is mainly acquired through inhalation of aerosols of animal origin and can cause a wide array of conditions ranging from asymptomatic infection to an acute self-limiting febrile illness or severe, chronic diseases like endocarditis and hepatitis (21).

The aim of this study was to characterize the splicing properties of the intron reported by Seshadri et al. (31) in the 23S rDNA of *C. burnetii* since both non-splicing and self-splicing introns are known to occur in bacterial 23S rRNA (7, 26). During our investigation, we discovered that contrary to the earlier report, two group I self-splicing introns flank a 34-bp exon and interrupt the 23S rDNA of *C. burnetii*. We also discovered that the terminal nucleotide of one of the introns was adenine and not the canonical conserved guanine, called Ω G. This is the first time that a group I intron without Ω G has been found in nature. This is significant because the Ω G is considered invariable and thought to play an essential role in RNA splicing (38, 40). This novel intron is also smaller than most group I introns as it does not encode a homing endonuclease, making it more difficult to identify in whole genome sequences than relatively larger introns that code for a LAGLIDADG homing endonuclease. Nevertheless, a nucleotide BLAST search against GenBank revealed a number of group I introns with conserved sequences. Secondary structures for both introns were also inferred by locating conserved sequences.

In addition to the Nine Mile strain, we sequenced the introns of eight other strains of *C. burnetii* representing seven additional genomic groups (3, 15) and found them to be highly conserved (Table 1), implying a possible role in *C. burnetii*'s biology.

C. MATERIALS AND METHODS

***C. burnetii* strains and cultivation.** The primary strain of *C. burnetii* used in this study was *C. burnetii* Nine Mile phase II (RSA 439; clone 4). Genomic DNA from strains belonging to seven other genomic groups used for determining the sequences of both the introns are listed in Table 1. *C. burnetii* Nine Mile phase II and genomic DNA of the other seven genotypes were generous gifts from Robert A. Heinzen and Paul A. Beare at Rocky Mountain Laboratories, Hamilton, Montana. *C. burnetii* Nine Mile phase II was propagated in African green monkey kidney (Vero) fibroblasts (CCL-81; American Type Culture Collection, Manassas, VA) grown in RPMI medium (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (HyClone, Logan, UT). Bacteria were purified from infected cells at 7 d post-infection by Renografin (Bracco Diagnostics, Princeton, NJ) gradient centrifugation as previously described (39).

Sequence analyses. All sequence data were obtained from plasmid DNA or PCR products with an automated DNA sequencer (ABI3130x1) and a BigDye Terminator Cycle Sequencing Ready Reaction kit (ABI, Foster City, CA). Sequence analysis was accomplished using the 'BLAST 2 sequences' tool at the NCBI website

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi> (36).

Intron secondary structure. Nucleotide BLAST searches (<http://www.ncbi.nlm.nih.gov/BLAST/>) were done to identify other introns with similar sequences (2). Secondary structures were inferred by manually locating conserved paired

helices (P1 – P10) using published intron secondary structures (20, 26) as a reference and were drawn using PowerPoint 2003 (Microsoft, Redmond, WA).

RNA and DNA preparations. RNA was isolated using a Ribopure Bacteria kit (Ambion, Austin, TX) and DNA was isolated using a High Pure PCR Template Preparation Kit (Roche Diagnostics, Basel, Switzerland) according to the manufacturers' recommendations.

RT-PCR, cloning and in vitro transcription. cDNA was made from RNA using an iScript cDNA Synthesis Kit (Bio-Rad) per instructions. PCR primers (Sigma-Aldrich, St. Louis, MO) for analyzing introns (intron F, GTGGCTGCGACTGTTTAC and intron R, ATTTCCGACCGTGCTGAG, and exon F, AACGGTCCTAAGGTAGCG, and exon R, TTCGCTACCTTAGGACCG) and IVS (IVS F, CGTGGTGGAAAGGGAAAC and IVS R, TGTCAGCATTCGCACTTC) were designed using Beacon Designer 6 software (BIO-RAD, Hercules, CA). Intron F and R primer sites flank the reported (31) intron sequence 266 bp upstream and 274 bp downstream, respectively, and IVS F and R primers flank the reported IVS sequence (1) 199 bp upstream and 78 bp downstream, respectively (Fig.1). Exon F and R primers are located in the 34-bp exon separating the two introns (see Results) and were used in combination with intron F and R primers to clone each intron independently into pCR2.i-TOPO vector as described below. PCR was done in a Mastercycler (Eppendorf AG, Hamburg, Germany). The reaction conditions were 94° C for 5 min, 30 cycles of 94 ° C for 1 min, 55 ° C for 1 min and 72 ° C for 1 min, and a single cycle of 72 ° C for 5 min. PCR products were cloned into pCR 2.1-TOPO using a TOPO TA cloning kit (Invitrogen) and selected for proper insertional orientation to utilize the T7 promoter on the plasmid. In vitro transcription was done

using a MAXIscript In vitro Transcription Kit (Ambion) and the resulting RNA was purified using NucAway Spin Columns (Ambion) per the manufacturer's instructions.

In vitro protein synthesis. The protein product from the gene coding for the homing endonuclease in Cbu.L1951 was produced in vitro from its endogenous promoter using an *E. coli* S30 Extract System for Circular DNA (Promega, Madison, WI) as per manufacturer's instructions and analyzed using 0.1% SDS-PAGE and autoradiography as previously described (24).

Phylogenetic analyses. Group IB introns (including Cbu.L1917) and Group IA3 introns (including Cbu.L1951) were preliminarily aligned with CLUSTALW (37). Other Group IB introns included in the alignment were *Thermotoga subterranea* (AJ556793), *Synechococcus* Sp. C9 (DQ421380), *Chlorosarcina brevispinosa* (L49150), *Chaetosphaeridium globosum* (AF494279), *Chlamydomonas zebra* (L43356), *Chlamydomonas humicola* (L42989), *Chlamydomonas monadina* (L49149), *Chlamydomonas frankii* (L43352), *Chlamydomonas komma* (L43502), *Mesostigma viride* (AF353999), *Oltmannsiellopsis viridis* (DQ291132), *Suillus luteus* (L47586, intron 2), *Chara vulgaris* (AY267353, introns 4, 5 and 6); the other Group IA3 introns were *Monomastix* sp. M722 (L44124), *Chlorella vulgaris* (AY008338), *Acanthamoeba castellanii* (U12380), *Chlorosarcina brevispinosa* (L49150), *Nephroselmis olivacea* (AF110138), *Stigeoclonium helveticum* (DQ630521). The alignments were next manually refined in BioEdit (11) to explicitly incorporate conserved secondary structures (P1-P9) and subsequently removed variable regions that could not be reliably aligned, as in Haugen and Bhattacharya (12). Alignments used for phylogeny reconstruction had 170 nucleotides (Group IB) and 237 nucleotides (Group IA3), respectively. Phylogenies were

reconstructed for both data sets by distance, maximum parsimony and maximum likelihood methods implemented in PAUP* version 4.0b10 (35). Phylogenies for Group IB introns were outgroup rooted with the distantly related *cox1* intron of *Catalpa fargesii* (AJ223411; (12)); Group IA3 phylogenies were unrooted. The models of sequence evolution used in the distance and likelihood methods for both data sets were selected by the Akaike Information Criterion, as implemented in Modeltest (28). For the Group IB data set, the model selected was the general time reversible model with among nucleotide site rate heterogeneity estimated by a discrete (4 rate categories) approximation of a gamma distribution (GTR + G model); for the Group IA3 data set, the model selected was the general time reversible model with among nucleotide site rate heterogeneity estimated by the proportion of invariant sites (GTR + I model). Heuristic searches (for 10 replicates of random sequence addition with branch swapping by the tree-bisection reconnection method) were performed for both the distance and likelihood analyses. Distance analyses were performed according to the least squares optimality criterion. Unweighted, unordered parsimony reconstructions were inferred by branch and bound, which is guaranteed to find the most parsimonious tree(s). Phylogenies were bootstrap-replicated with either 10,000 (distance and parsimony) or 100 (likelihood) replicates. To test whether alternative phylogenetic hypotheses explained the aligned sequence data equally well (i.e., whether their likelihood scores were statistically significantly different), we used the SH test of topological congruence (32), as implemented in PAUP*.

D. RESULTS

***C. burnetii* 23S rDNA contains two introns that splice in vivo.** Using the published genome sequence as a reference (31), PCR primers (intron F and R, Fig. 1) flanking the reported intron insertion site in 23S rDNA were synthesized. PCR was done using these primers and genomic DNA from *C. burnetii* Nine Mile phase II. PCR products were sequenced to confirm that the published 23S rDNA sequence of Nine Mile phase I was conserved in the phase II strain (data not shown). Since unspliced and self-splicing group I introns have been reported in 23S rRNA of bacteria (7, 26), splicing characteristics of the intron in the 23S rDNA of *C. burnetii* was analyzed using RT-PCR and PCR (intron F and R primers) on total RNA and genomic DNA, respectively, from 7-day-old cultures ($\sim 2.5 \times 10^{10}$ cells). While the resulting PCR product from cDNA was only 551 bp in size, the PCR product from genomic DNA using the same primer-pair was 1559 bp (Fig. 2, lanes 1 and 2), indicating that the intron was spliced out of the mature rRNA. When the PCR product from cDNA was sequenced, it revealed to our surprise that contrary to the previous report, two introns actually interrupted the 23S rDNA of *C. burnetii* as shown in Fig. 1. The two introns flank a 34-bp exon corresponding to bases 1918 to 1951 in domain IV of *E. coli* 23S rRNA (41). Based on their positions, the two introns were designated as Cbu.L1917 and Cbu.L1951 respectively, using standard nomenclature (18). Sequence data were deposited in GenBank and was assigned the accession number EF632073. Sequencing data also showed that the two introns are spliced in vivo to generate mature rRNA. Further, these data allowed us to determine the exact sequence and the 5' and 3' boundaries of both introns. We also analyzed the previously reported (1) intervening sequence (IVS) in the 23S rDNA of *C. burnetii* using RT-PCR and PCR (IVS F and R primers, Fig. 1) as described above. While genomic DNA gave the expected 722-

bp PCR product, no PCR product was observed with cDNA (Fig. 2, lanes 5 and 6), confirming that the mature 23S rRNA of *C. burnetii* is fragmented due to the excision of IVS without subsequent ligation, as previously reported (1).

Both introns are self-splicing in vitro. To determine whether the introns could self-splice in vitro, the PCR product generated from primers intron F and R (Fig. 1) and *C. burnetii* genomic DNA was cloned using a TOPO TA Cloning kit (Invitrogen). Utilizing the T7 promoter present in the pCR2.1-TOPO vector, RNA was transcribed and converted into cDNA to create a PCR template for intron F and R primers. As observed in vivo (Fig. 2), the PCR product from cDNA was 551 bp in size (Fig. 3, lane 2), indicating that the introns could self-splice in vitro in the transcription buffer. Subsequent sequencing of the 551-bp PCR product confirmed that both introns self-splice in vitro without the aid of any proteins (data not shown). Separate constructs containing Cbu.L1917 or Cbu.L1951 were also made (see Materials and Methods), and the introns were observed to be able to self-splice independently in vitro (data not shown). In contrast to introns, when PCR was done using IVS F and R primers (Fig. 1) both cDNA and genomic DNA gave similar sized bands (Fig. 3, lanes 5 and 6) showing that the IVS does not self-splice in vitro.

Both Cbu.L1917 and Cbu.L1951 are group I introns. Introns are classified into four major classes based on their splicing mechanisms as group I, groupII/III, spliceosomal introns and tRNA/archael introns (14). Hallmarks of group I introns include autocatalytic activity, conserved paired (P) helices, wobble pair G•U at the 5' splice site and a conserved terminal guanine, called Ω G. Based on our sequencing data and previous reports (12, 31) it was easy to determine that intron Cbu.L1951 is a group I intron.

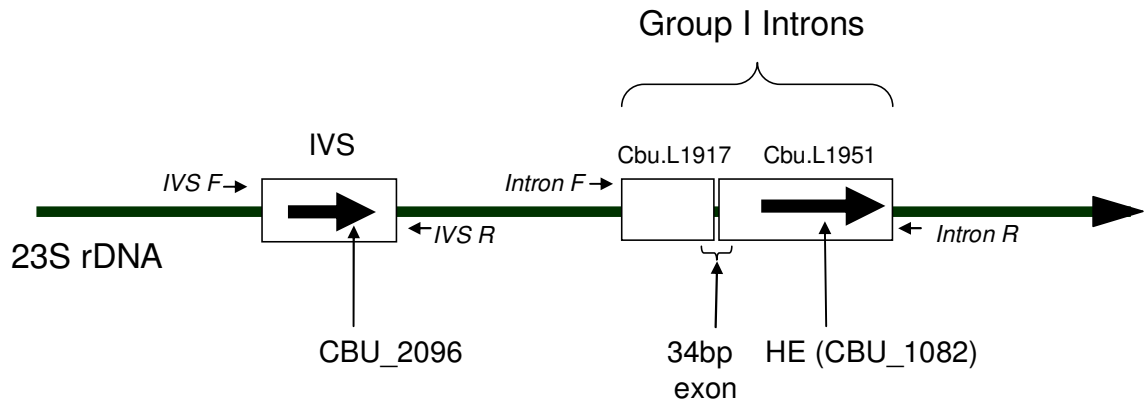


FIG. 2-1. *C. burnetii* 23S rDNA linkage map. Two group 1 introns, Cbu.L1917 and Cbu.L1951, flank an essential 34-bp 23S rDNA exon. Cbu.L1951 also encodes a putative homing endonuclease (HE; CBU_1082; arrow). Position of the intervening sequence (IVS) is also shown, with its nested ORF (CBU_2096; arrow). Primer sets used in the study are italicized and indicated by small arrows

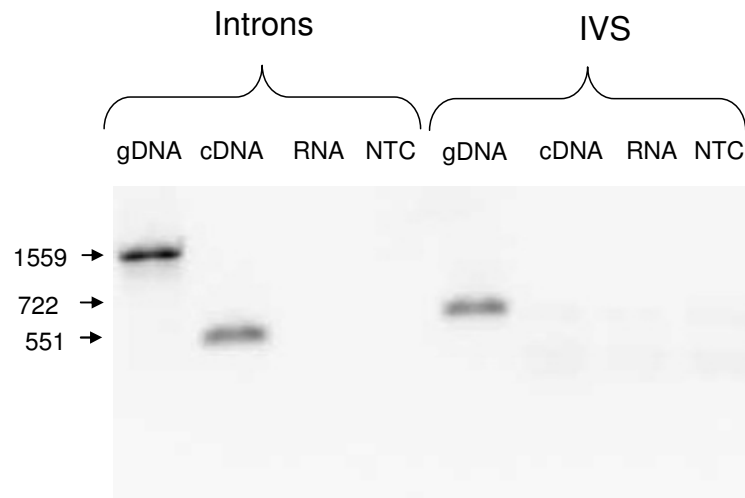


FIG. 2-2. In vivo analysis. PCR was done using Intron or IVS primer sets on genomic DNA (gDNA), cDNA or RNA Isolated from *C. burnetii* (Nine Mile Phase II) or a no-template control (NTC). An ethidium bromide-stained agarose gel (1% agarose; w/v) is shown. Sizes of the amplicons were determined from standards and are given to the left in base pairs.

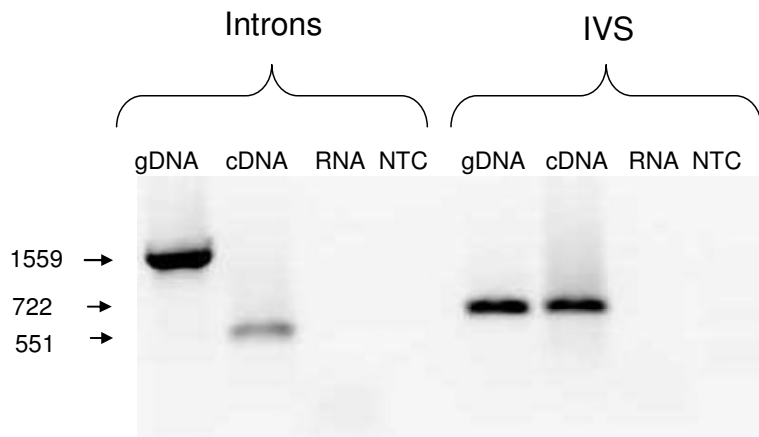


FIG. 2-3. In vitro analysis. PCR was done using Intron or IVS primer sets on genomic DNA (gDNA), cDNA (from RNA generated by in vitro transcription), in vitro transcribed RNA or a no-template control (NTC). An ethidium bromide-stained agarose gel (1% agarose; w/v) is shown. Sizes of the amplicons were determined from standards and given to the left in base pairs.

Cbu.L1951 has all the typical features of group I introns described above and also encodes a LAGLIDADG homing endonuclease (HE; Figs. 1 and 4B), which might play a role in its mobility (12). In fact, when expression of cloned HE from its endogenous promoter was checked using a prokaryotic in vitro transcription and translation kit (Promega), a ~17 KDa product was observed on autoradiographs that was not present on vector reactions (data not shown). The presence of the P2 helix and of helices P7.1 and P7.2 in Cbu.L1951 are characteristics of subgroup IA3 introns (10, 20). The internal guide sequence (IGS) that pairs with 5' and 3' exons and forms the substrate for the ribozyme was also recognizable (Helices P1 and P10 in Fig.4B) in Cbu.L1951. In contrast, it was difficult to classify intron Cbu.L1917 as it does not code for an HE and does not have the canonical Ω G. A BLAST search using the nucleotide sequence of Cbu.L1917 identified other group I introns in the same 23S rDNA position (1917) with conserved sequences. Using the published secondary structure of intron *Tsu.bL1917* as a reference (26), the secondary structure of Cbu.L1917 was predicted. From the secondary structure (Fig. 4A) it is clear that this intron belongs to group I, even though it lacks Ω G. The lack of P2 helix and an extensive P5 loop indicates that Cbu.L1917 belongs to subgroup IB2 (10). In addition, the IGS (P1 and P10 in Fig. 4 A) is much better defined in Cbu.L1917 than in Cbu.L1951.

Phylogenetic analyses. Phylogeny reconstructions of the Group IB intron data were generally similar, with the least squares distance phylogeny shown (Fig. 5A). The phylogenetic position of *C. burnetii* intron Cbu.L1917 was not definitively resolved. Cbu.L1917 formed a clade with the *Thermotoga subterranea* intron in the least squares and the three equally most parsimonious trees, but with only modest bootstrap support.

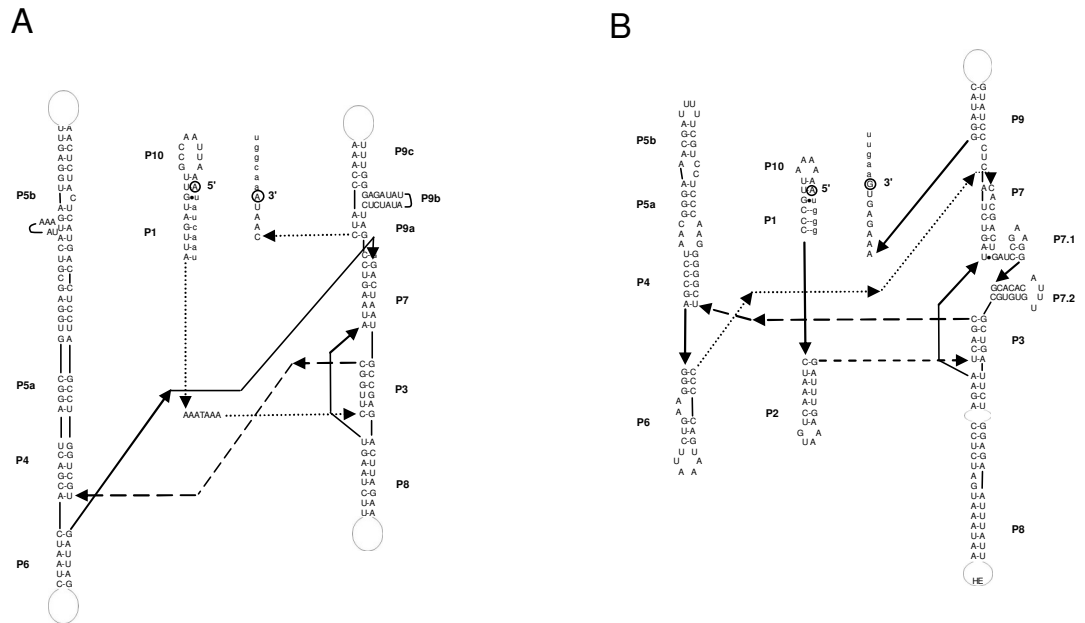


FIG. 2-4. Predicted secondary structures for (A) Cbu.L1917 and (B) Cbu.L1951.

Positions of conserved, paired helices common to group I introns are designated P1-P10.

The 5' and 3' terminal bases are encircled. The site of the homing endonuclease-encoding

ORF of Cbu.L1951 in P8 is indicated by "HE". Different line types are included to

facilitate easy visualization of intron secondary structure. Exons depicted in lower case

and introns in upper case alphabets.

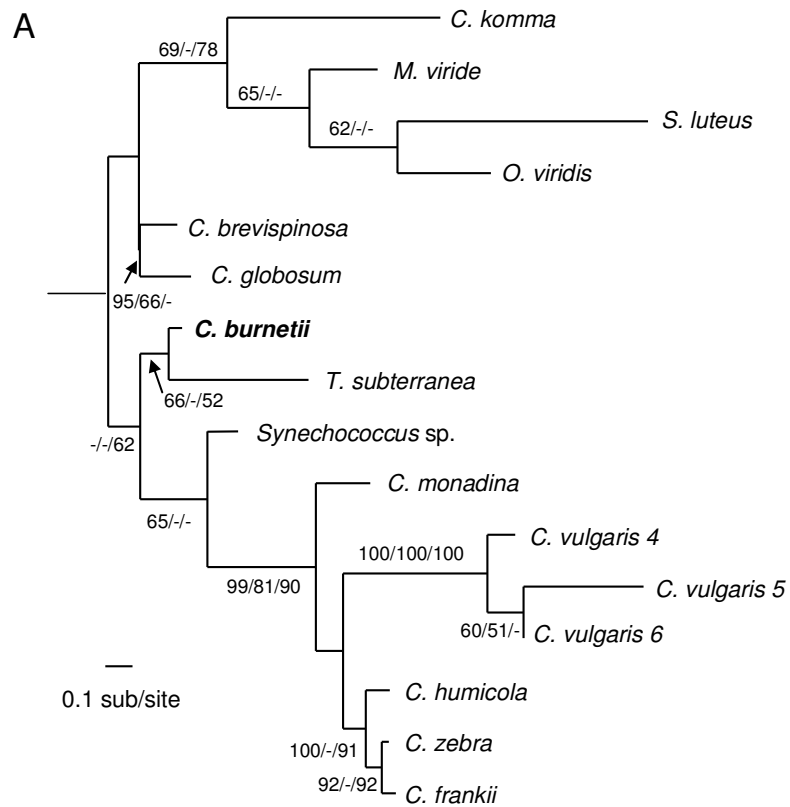
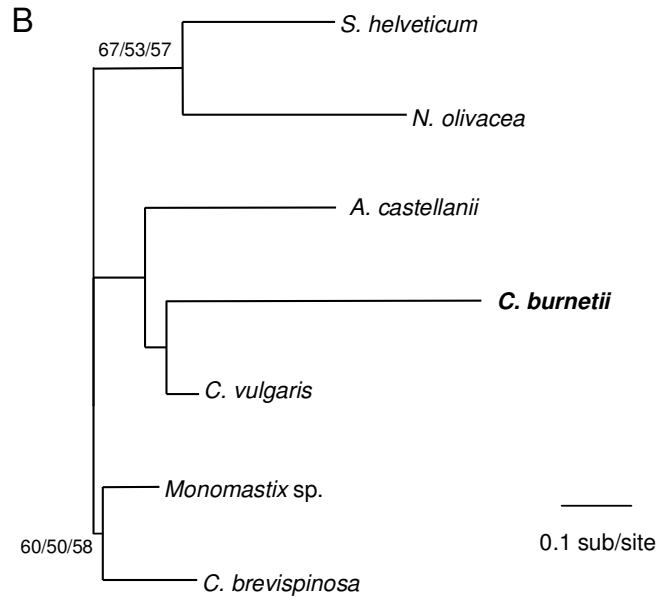


FIG. 2-5. Phylogenetic analyses of Cbu.L1917 and Cbu.L1951. (A) Least squares phylogeny for Group IB introns. The tree was rooted with the *cox1* intron of *Catalpa fargesii*. Bootstrap values for the least squares/maximum likelihood/maximum parsimony analyses are indicated at nodes for which the bootstrap value was >50% in at least one analysis. **(B)** Unrooted least squares phylogeny for Group IA3 introns. Bootstrap values as in (A).



In the likelihood phylogeny, Cbu.L1917 was basal to *Thermotoga*, *Synechococcus* and the *Chara/Chlamydomonas* clade. However, the likelihood score of an alternative tree constrained to the least squares topology (Fig. 5A) was not significantly different from the maximum likelihood tree by a SH test ($-\Delta L = 7.70463$, $P = 0.193$), indicating that both trees explain the data equally well.

Phylogeny reconstructions of the Group IA3 intron data were qualitatively similar (least squares distance phylogeny, Fig. 5B). L1951 introns (including Cbu.L1951) grouped together to the exclusion of introns with insertion sites at other locations (*Nephroselmis* L2593 and *Stigeoclonium* L730), as previously found (12). Although Cbu.L1951 formed a clade with the *Chlorella* and *Acanthamoeba* introns in all trees, its sister taxon was not definitively resolved (*Acanthamoeba* in the least squares and parsimony trees, *Chlorella* in the likelihood tree).

Intron sequences are conserved in other strains of *C. burnetii*. Genomic DNA from strains belonging to all eight genomic groups (3) of *C. burnetii* was analyzed for both introns to determine their presence, absence or sequence variation. The data indicate that both introns were present in a common ancestor of all eight genomic groups of *C. burnetii* and are highly conserved among them ($\geq 99\%$ nucleic acid sequence identity, with most variation occurring in the HE gene) (Table 1).

E. DISCUSSION

In the current study, we identified two self-splicing group I introns in the 23S rDNA of *C. burnetii*, one of which, Cbu.L1917, is the first of its kind. The lack of Ω G in Cbu.L1917 is intriguing since all other group I introns known to date, have guanine as the 3' terminal nucleotide (4). The conserved Ω G forms a Hoogsteen base triple with a G-C

TABLE 2-1. Introns are highly conserved amongst genomic groups of *C. burnetii*

Genomic group ^a	Strain	Source, Year	Cbu.L1917	Cbu.L1951
I	Nine Mile Phase I RSA493	Montana, tick, 1935	Reference Sequence	Reference Sequence
I	Nine Mile Phase II RSA493	Montana, tick, 1935	Ref ^b	Ref
II	Henzerling RSA331	Italy, human blood, 1945	A247 deleted	Ref
III	Idaho Goat Q195	Idaho, goat, 1981	Ref	G393A ^c and G402A ^c
IV	MSU Goat Q177 (Priscilla)	Montana, goat, 1980	Ref	Ref
IV	K Qi54	Oregon, human heart valve, 1976	Ref	Ref
V	G Q212	Nova Scotia, human heart valve, 1981	Ref	C403T ^c
VI	Dugway 7E9-12	Utah, rodents, 1958	Ref	Ref
VII	Q321	Russia, cow's milk	Ref	Ref
VIII	Le Bruges	France, unknown source	Ref	Ref

^aAs defined by (3, 15)

^bSame as Reference

^cWithin Homing Endonuclease ORF

pair in the P7 helix, flanked by two additional P7 base triples to form the guanosine-binding pocket (33, 40). Even though it is possible that the splicing efficiency of Cbu.L1917 is lower than that of Cbu.L1951, as shown by earlier studies substituting Ω G to adenine, cytosine or uracil greatly reduced the rates of 3' cleavage and exon ligation in vitro (22, 29), we could never detect intermediate forms containing spliced Cbu.L1951 and unspliced Cbu.L1917 in *C. burnetii* RNA preparations. Moreover, both in vitro and in *E. coli* expressing the introns, we could detect intermediate forms with either intron spliced out in comparable amounts when internal primer sets were used, indicating that both introns splice out independently and with similar efficiencies (data not shown). It would be interesting to learn the splicing mechanism used by Cbu.L1917 with adenine as its terminal base. Possibly the Ω A uses the same guanosine-binding pocket as other group I introns to facilitate 3' cleavage or it might use a novel mechanism to select the 3' splice site. It is also very likely that accessory proteins are involved in intron splicing in vivo in spite of the introns' ability to self-splice in vitro as previously described (19).

It is highly unusual that *C. burnetii*'s lone 23S rDNA contains two group I introns, especially considering the rarity of even single elements in bacterial rDNAs (26). In addition, the large ribosomal subunit RNA gene of *C. burnetii* also carries an intervening sequence (IVS) near its 5' end (Fig.1). In contrast to introns, the IVS is excised from the RNA without exon ligation, a cleavage process mediated by RNase III (8), resulting in a fragmented but functional rRNA. Moreover, unlike other obligate intracellular bacteria, which have few or no insertion sequences (IS) - presumably due to limited opportunities for lateral gene transfer- *C. burnetii*'s genome possesses a large number of IS elements (29 full-length and 3 degenerate). This unusually high number of selfish DNAs in the

genome of *C.burnetii*, especially the acquisition of three ‘selfish genes’ in its vital 23S rRNA gene seems antithetical to the observation that obligate intracellular bacteria undergo reductive evolution. However, this is likely a relic from a past niche with greater opportunities for lateral gene transfer with the pathogen recently shifting to its present obligate intracellular lifestyle; a hypothesis supported by the higher percentage of coding genes (89.1%) in its genome as opposed to other intracellular bacteria, like *Rickettsia prowazeki* and *Mycobacterium leprae* (both 76%). It is likely that given sufficient time *C. burnetii*’s genome will also undergo further reductive evolution, as suggested by Seshadri et al. (31).

Group I introns are genetic elements that can spread from an intron-positive strain to an intron-minus strain by two known mechanisms. The best studied and well understood is a process termed homing, where the intron moves to an intronless allele of the same gene using its encoded HE (5). The alternative to the HE-dependent mechanism is reverse splicing, where the intron recognizes and integrates into homologous or ectopic RNA sites coupled with reverse transcription and recombination (30). The Goddard-Burt model of intron evolution suggests that host populations go through cyclical intron-containing and intronless states with horizontal transmission being necessary for the long-term persistence of introns in any population (9). According to this model, a group I intron with a full length HE invades an intronless population by homing. The HE is lost once the intron becomes fixed in a population. Assuming no host benefit, the intron is also subsequently lost from the population. The intron reappears only when it regains access to the population via lateral gene transfer. This model seems appropriate for *C. burnetii*. Even though available sequence data were not sufficient to specifically identify the

donors of both introns, it is apparent that the introns have invaded the 23S rDNA of *C. burnetii* from organelles of lower eukaryotes or other bacteria (Fig. 5). With the introns fixed in the population (Table 1), they appear to be in the process of losing HEs, as suggested by the lack of HE in Cbu.L1917. However, in contrast to earlier studies that found introns in various stages of degeneration in multiple populations of the same organism (13, 25), screening eight genomic groups of *C. burnetii* revealed that both introns are highly conserved (>99%). This could either be due to insufficient evolutionary time for mutations to accumulate or due to selective pressure acting on the introns to maintain them. If the latter is the case, it would be highly informative to understand what function(s) the introns serve, as they are generally considered to be selfish genes that make no contribution towards the reproductive success of the host. Preliminary work in our laboratory suggests a role for the introns in modulating the growth rate of *C. burnetii*, akin to the intron-mediated ribosomal inhibition observed when *Tetrahymena* group I intron was expressed in *E. coli* (27). Currently, we are investigating this growth modulation to understand its potential role in *C. burnetii*'s lifecycle. Another exciting avenue for further investigation is the potential for using self-splicing introns as anti-microbial targets to treat Q fever. For example, pentamidine, a drug used for the treatment and prophylaxis of *Pneumocystis carinii* (*Pneumocystis jiroveci*) pneumonia (16) is thought to act by inhibiting rRNA group I intron self-splicing (42). Although pentamidine has toxic side effects, it might be a potential therapeutic agent for treating patients with chronic Q fever endocarditis.

F. ACKNOWLEDGMENTS

We are grateful to Patty McIntire (UM Murdock Sequencing Facility) for sequence analysis; Jim Battisti and Sherry Coleman (University of Montana), and Steven Zimmerly and Dawn Simon (University of Calgary, AB) for helpful discussions. This work was supported by the NIH Rocky Mountain Regional Center of Excellence for Biodefense and Emerging Infectious Disease grant, U54 AI065357-010023.

G. REFERENCES

1. **Afseth, G., Y. Y. Mo, and L. P. Mallavia.** 1995. Characterization of the 23S and 5S rRNA genes of *Coxiella burnetii* and identification of an intervening sequence within the 23S rRNA gene. *J. Bacteriol.* **177**:2946-9.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**:3389-402.
3. **Beare, P. A., J. E. Samuel, D. Howe, K. Virtaneva, S. F. Porcella, and R. A. Heinzen.** 2006. Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. *J. Bacteriol.* **188**:2309-24.
4. **Burke, J. M.** 1989. Selection of the 3'-splice site in group I introns. *FEBS Lett* **250**:129-33.
5. **Chevalier, B. S., and B. L. Stoddard.** 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucl. Acids Res.* **29**:3757-74.
6. **Cutler, S. J., M. Bouzid, and R. R. Cutler.** 2007. Q fever. *J Infect* **54**:313-8.
7. **Everett, K. D., S. Kahane, R. M. Bush, and M. G. Friedman.** 1999. An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J. Bacteriol.* **181**:4734-40.
8. **Evguenieva-Hackenberg, E.** 2005. Bacterial ribosomal RNA in pieces. *Mol. Microbiol.* **57**:318-25.
9. **Goddard, M. R., and A. Burt.** 1999. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. U. S. A.* **96**:13880-5.
10. **Gonzalez, P., G. Barroso, and J. Labarere.** 1999. Molecular gene organisation and secondary structure of the mitochondrial large subunit ribosomal RNA from the cultivated Basidiomycota *Agrocybe aegerita*: a 13 kb gene possessing six unusual nucleotide extensions and eight introns. *Nucl. Acids Res.* **27**:1754-61.
11. **Hall, T.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**:95-98.
12. **Haugen, P., and D. Bhattacharya.** 2004. The spread of LAGLIDADG homing endonuclease genes in rDNA. *Nucl. Acids Res.* **32**:2049-57.

13. **Haugen, P., V. A. Huss, H. Nielsen, and S. Johansen.** 1999. Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing-endonuclease pseudogenes in the *Bangiophyceae*. *Curr. Genet.* **36**:345-53.
14. **Haugen, P., D. M. Simon, and D. Bhattacharya.** 2005. The natural history of group I introns. *Trends Genet.* **21**:111-9.
15. **Hendrix, L. R., J. E. Samuel, and L. P. Mallavia.** 1991. Differentiation of *Coxiella burnetii* isolates by analysis of restriction-endonuclease-digested DNA separated by SDS-PAGE. *J. Gen. Microbiol.* **137**:269-76.
16. **Hughes, W. T.** 1991. Prevention and treatment of *Pneumocystis carinii* pneumonia. *Annu. Rev. Med.* **42**:287-95.
17. **Ikawa, Y., H. Shiraishi, and T. Inoue.** 2000. Minimal catalytic domain of a group I self-splicing intron RNA. *Nat. Struct. Biol.* **7**:1032-5.
18. **Johansen, S., and P. Haugen.** 2001. A new nomenclature of group I introns in ribosomal DNA. *RNA* **7**:935-6.
19. **Lambowitz, A. M., M. G. Capara, S. Zimmerly and P. S. Perlman.** 1999. Group I and Group II ribozymes as RNPs: clues to the past and guides to the future, p. 451-485. *In* R. F. Gesteland, T. R. Cech and J. F. Atkins (ed.), *The RNA world*, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, NY.
20. **Lonergan, K. M., and M. W. Gray.** 1994. The ribosomal RNA gene region in *Acanthamoeba castellanii* mitochondrial DNA. A case of evolutionary transfer of introns between mitochondria and plastids? *J. Mol. Biol.* **239**:476-99.
21. **Maurin, M., and D. Raoult.** 1999. Q fever. *Clin. Microbiol. Rev.* **12**:518-53.
22. **Michel, F., M. Hanna, R. Green, D. P. Bartel, and J. W. Szostak.** 1989. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342**:391-5.
23. **Michel, F., and E. Westhof.** 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**:585-610.
24. **Mitchell, S. J., and M. F. Minnick.** 1997. A carboxy-terminal processing protease gene is located immediately upstream of the invasion-associated locus from *Bartonella bacilliformis*. *Microbiology* **143** (Pt 4):1221-33.
25. **Muller, K. M., J. J. Cannone, R. R. Gutell, and R. G. Sheath.** 2001. A structural and phylogenetic analysis of the group IC1 introns in the order *Bangiales* (*Rhodophyta*). *Mol. Biol. Evol.* **18**:1654-67.
26. **Nesbo, C. L., and W. F. Doolittle.** 2003. Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc. Natl. Acad. Sci. U. S. A.* **100**:10806-11.
27. **Nikolcheva, T., and S. A. Woodson.** 1997. Association of a group I intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA* **3**:1016-27.
28. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817-8.
29. **Price, J. V., and T. R. Cech.** 1988. Determinants of the 3' splice site for self-splicing of the *Tetrahymena* pre-rRNA. *Genes Dev.* **2**:1439-47.
30. **Roman, J., and S. A. Woodson.** 1998. Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **95**:2134-9.

31. **Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg.** 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc. Natl. Acad. Sci. U. S. A. **100**:5455-60.
32. **Shimodaira, H., and M. Hasegawa.** 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16**:1114-1116.
33. **Stahley, M. R., and S. A. Strobel.** 2006. RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. Curr. Opin. Struct. Biol. **16**:319-26.
34. **Suh, S. O., K. G. Jones, and M. Blackwell.** 1999. A Group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of Group I introns. J. Mol. Evol. **48**:493-500.
35. **Swofford, D. L.** 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b4a. Sinauer Associates, Sunderland, Massachusetts.
36. **Tatusova, T. A., and T. L. Madden.** 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. **174**:247-50.
37. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. **22**:4673-80.
38. **Vicens, Q., and T. R. Cech.** 2006. Atomic level architecture of group I introns revealed. Trends Biochem. Sci. **31**:41-51.
39. **Williams, J. C., M. G. Peacock, and T. F. McCaul.** 1981. Immunological and biological characterization of *Coxiella burnetii*, phases I and II, separated from host components. Infect. Immun. **32**:840-51.
40. **Woodson, S. A.** 2005. Structure and assembly of group I introns. Curr. Opin. Struct. Biol. **15**:324-30.
41. **Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller.** 2001. Crystal structure of the ribosome at 5.5 Å resolution. Science **292**:883-96.
42. **Zhang, Y., A. Bell, P. S. Perlman, and M. J. Leibowitz.** 2000. Pentamidine inhibits mitochondrial intron splicing and translation in *Saccharomyces cerevisiae*. RNA **6**:937-51.

CHAPTER THREE

Toxic introns and parasitic intein in *Coxiella burnetii*: Legacies of a promiscuous past

Rahul Raghavan, Linda D. Hicks and Michael F. Minnick. 2008.

J. Bacteriol. **190** (In press)

A. ABSTRACT

The genome of the obligate intracellular pathogen *Coxiella burnetii* contains a large number of selfish genetic elements including two group I introns (Cbu.L1917 and Cbu.L1951) and an intervening sequence that interrupt the 23S rRNA gene, an intein (Cbu.DnaB) within *dnaB*, and 29 insertion sequences. Here, we describe the ability of the intron-encoded RNAs (ribozymes) to retard bacterial growth rate (toxicity) and examine the functionality and phylogenetic history of Cbu.DnaB. When expressed in *E. coli*, both introns repressed growth, with Cbu.L1917 being more inhibitory. Both ribozymes were found to associate with ribosomes of *Coxiella* and *E. coli*. In addition, ribozymes significantly reduced in vitro luciferase translation, again with Cbu.L1917 being more inhibitory. We analyzed the relative quantities of ribozymes and genomes throughout a 14-d growth cycle of *C. burnetii* and found that they are inversely correlated, suggesting that the ribozymes have a negative effect on *Coxiella*'s growth. We determined possible sites for ribozyme associations with 23S rRNA that could explain the observed toxicities. Further research is needed to determine whether the introns are being positively selected because they promote bacterial persistence or whether they were fixed in the population due to genetic drift. The intein, Cbu.DnaB is able to self-splice leaving the host protein

intact and presumably functional. Similar inteins were found in two extremophilic bacteria (*Alkalilimnicola ehrlichei* and *Halorhodospira halophila*) that are distantly related to *Coxiella* making it difficult to determine whether the intein was acquired by horizontal gene transfer or was vertically inherited from a common ancestor.

B. INTRODUCTION

Bacterial genomes are in a constant state of flux. Bacteria gain new DNA through horizontal gene transfer (HGT), whereas nucleotide deletion results in loss of DNA (54). While environmental bacteria often have large genomes (e.g., *Pseudomonas aeruginosa*, ~6 MB) that contain a large number of elements acquired through HGT, host-associated bacteria have relatively smaller genomes (e.g., *Rickettsia rickettsii*, ~1.2 MB; *Chlamydomyxa abortus*, ~1.1 MB) with little or no DNA of foreign origin (18,72). The advent of whole genome sequencing has facilitated the detailed comparative analyses of obligate intracellular bacterial genomes leading to identification of factors such as the prevalence of pseudogenes and insertion sequences (ISs) that can be used to distinguish early adapters from bacteria that have shifted recently from a free-living to a host-restricted lifestyle (48). A relevant case is the obligate intracellular pathogen *Coxiella burnetii* whose genome was found to be undergoing reductive evolution and to contain a large number of selfish genetic elements and ‘young’ pseudogenes, suggesting a recent shift to its current niche (66).

C. burnetii, a γ purple bacterium, is the causative agent of Q fever (12). Most human infections are acquired through inhalation of contaminated aerosols of animal origin and

can lead to an acute self-limiting febrile illness and severe, chronic cases of endocarditis or hepatitis (41). This category B select agent has a broad range of susceptible hosts including arthropods, fish, birds and wild and domestic mammals and is distributed almost worldwide (12). *C. burnetii* can survive for months in the extracellular environment, where it is notoriously resistant to heat, ultraviolet light, various disinfectants and desiccation (42, 75). These traits are conferred by a biphasic developmental cycle that alternates between a fragile, metabolically-active large-cell variant (LCV) and a durable, dormant small-cell variant (SCV) (8). Even though *C. burnetii* occupies an acidic parasitophorous vacuole where the opportunity for HGT is likely minimal, its genome contains 29 ISs strewn across the chromosome, two group I introns and an intervening sequence (IVS) within its sole 23S rRNA gene, and a putative intein in the C-terminal region of the replicative DNA helicase (*dnaB*) gene (59, 66).

Group I introns self-splice independent of proteins (ribozyme) and are considered a holdover from a primordial RNA world (77). These genetic elements spread efficiently into an intronless cognate site by a process called homing at the DNA level or by reverse splicing at the RNA level (60). The typical secondary structure of a group I intron consists of about 10 paired (P) elements (59,74). P1 and P10 are complementary to the 5' exon and 3' exon, respectively and are collectively called the internal guide sequence (IGS), which the intron uses to locate its 5' and 3' splice sites (31).

Inteins (internal proteins) are genetic elements similar to introns that have invaded the coding sequence of genes in Eukarya, Bacteria and Archaea. Inteins are found in proteins with diverse functions, however, enzymes involved in DNA replication and repair predominate (23, 56). Unlike introns, inteins are transcribed and translated together with

the host protein. They self-excise, leaving the flanking exons (external proteins) spliced together and the host protein intact (23). Four conserved motifs are involved in protein splicing and are related to the C-terminal domain of the Hedgehog protein of eukaryotes (13, 57). Most, but not all, inteins encode a homing endonuclease, which aids in mobility as observed in group I introns (15, 56). Chemical reactions involved in protein splicing have been extensively studied and are described elsewhere (9, 10, 40).

We recently showed that the 23S rRNA gene of *C. burnetii* contains two group I introns (Cbu.L1917 and Cbu.L1951) and elucidated their secondary structures, splicing properties and phylogenetic histories (59). Here, we describe the toxicity of both introns as it relates to bacterial growth and consider the possible biological consequences and evolutionary significance. We postulate that *Coxiella*'s past life as a free-living, environmental bacterium with extensive HGT combined with a recent shift to its current host-restricted lifestyle helps explain the occurrence of toxic introns that can potentially influence its biology. In addition, we analyzed an intein (Cbu.DnaB) found in the replicative DNA helicase (*dnaB*) gene and confirmed its functionality. Phylogenetic analyses revealed closely-related inteins at the same loci of *Alkalilimnicola ehrlichei*, an anaerobic, haloalkaliphilic, γ -proteobacterium and *Halorhodospira halophila*, an extremely halophilic γ -proteobacterium (29, 44). Further phylogenetic analyses using 16S and DnaB sequences point towards a possible close evolutionary relationship between *C. burnetii* and these extremophiles, making it difficult to delineate whether *C. burnetii* acquired the intein via HGT or whether the intein was vertically inherited from a common ancestor.

C. MATERIALS AND METHODS

***E. coli* growth assay.** Both introns were amplified using PCR primers (Table 1: L1917_flank and L1951_flank) and *C. burnetii* genomic DNA, as previously described (59). Amplicons were cloned into pCR2.1-TOPO (Invitrogen, Carlsbad, California). PCR and sequencing were performed to confirm that the inserts were in the proper orientation downstream of the *lac* promoter. *E. coli* (Top10F') transformed with each plasmid was grown to mid-logarithmic phase at 37°C in Luria-Bertani (LB) broth in the presence of 100 µg/ml ampicillin (Sigma-Aldrich, St. Louis, MO) and then used to inoculate 50 ml fresh LB with 100 µg/ml ampicillin and 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG, EMD Chemicals, Gibbstown, NJ) to an OD₆₀₀ of 0.03 at 0 h. Bacterial growth was assayed spectrophotometrically at 600 nm every 60 min for 5 h (37°C, shaking).

In vitro transcription to synthesize ribozymes and control RNA. Introns were amplified using *C. burnetii* genomic DNA and PCR primers designed to add a T7 promoter sequence at the 5' end of the amplicons (Table 1: L1917+T7P and L1951+T7P) to generate templates to synthesize the two ribozymes. As a control, the IVS was amplified using *C. burnetii* genomic DNA and a PCR primer set (Table 1: IVS+T7P) that adds a T7 promoter sequence at the 3' end of the amplicon such that an antisense RNA can be transcribed from it. In vitro transcription was performed using a MEGAscript Kit (Ambion, Austin, TX), and resulting RNA was purified using NucAway (Ambion) spin columns and TURBO DNase (Ambion) as per manufacturer's instructions.

In vitro transcription/translation and luciferase assay. Luciferase was synthesized in vitro using an *E. coli* S30 Extract System for Circular DNA and pBEST*luc* (Promega, Madison, WI) as per manufacturer's instructions in the presence of 25 µg of Cbu.L1917,

Cbu.L1951 or control anti-sense RNA. Luciferase activity was measured using a Luciferase Assay System (Promega) per manufacturer's instructions utilizing a luminometer (PerkinElmer, Waltham, MA).

Phylogenetic Analyses. Intein phylogeny was reconstructed using amino acid sequences. Inteins were identified either by BLAST or from the intein database, InBase,(55, 71). In addition to the intein from *C. burnetii*, (NP819887) other inteins used were from *H. halophila* SL1 (ABM61435), *A. ehrlichei* MLHE-1, (ZP00865879), *Rhodothermus marinus* (AAB66912.1), *Trichodesmium erythraeum* (ABG50752), *Lyngbya* sp. (EAW38924.1), *Microcystis aeruginosa* PCC 7806 (CA090722.1), *Herpetosiphon aurantiacus* ATCC 23779 (ABX06165.1), *Synechocystis* sp. PCC 6803 (NP442446.1), *Synechococcus* sp. JA-3-3Ab (ABC99714.1), *Guillardia theta* (O78411), and *Pseudomonas fluorescens* Pf-5 (AAY90835.1). Neighbor-joining trees that were bootstrap replicated (1000 replicates) were built using MEGA4 (64, 69). An optimal tree is shown with branch lengths depicting evolutionary distances computed by the Poisson correction method for the number of amino acid substitutions per site (80). All positions containing gaps were eliminated resulting in a total of 125 positions being used in the final dataset. The 16S nucleotide and DnaB amino acid sequences were obtained from GenBank, including *C. burnetii* (AE016828), *Rickettsiella grylli* (AAQJ00000000), *Legionella pneumophila* (AE017354), *A. ehrlichei* (CP000453), *H. halophila* (CP000544), *Francisella tularensis* (CP000803), *P. fluorescens* (CP000094), *P. syringae* (AE016853), *P. aeruginosa* (NC009656), *Yersinia pestis* (CP000668), *Salmonella typhimurium* (AE006468), *E. coli* K12 (CP000948), *Neisseria meningitidis* (AL157959), *N. gonorrhoeae* (AE004969), *Bartonella bacilliformis* (CP000524), *B. quintana*

(BX897700), *Rickettsia bellii* (CP000849), *R. prowazekii* (NC000963), *Chlamydophila abortus* (CR848038), and *C. pneumoniae* (BA000008). Neighbor-joining trees were built with 1000 bootstrap replicates using MEGA4 (69). Evolutionary distances in 16S trees were computed using the Maximum Composite Likelihood method in units of number of base substitutions per site, whereas evolutionary distances for the DnaB trees were computed similar to the intein tree (70). Positions containing gaps were removed with 1351 positions in the final 16S tree and 423 positions in the final DnaB tree. Maximum Parsimony trees were also built using the close-neighbor-interchange algorithm to confirm the evolutionary history inferred from the Neighbor-Joining trees (16, 50).

RNA and DNA preparations. DNA was isolated using a DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), and RNA was isolated using a RiboPure Bacteria Kit (Ambion) as per manufacturer's instructions.

***C. burnetii* culture.** *C. burnetii* Nine Mile phase II (RSA 439; clone 4) was used in this study and was propagated in African green monkey kidney (Vero) fibroblasts (CCL-81; American Type Culture Collection, Manassas, VA). Cultures were grown in RPMI medium (Invitrogen) supplemented with 10% fetal bovine serum (HyClone, Logan, UT). Bacteria were purified from infected cells by Renografin (Bracco Diagnostics, Princeton, NJ) gradient centrifugation as previously described (76). To generate SCVs, Vero cell monolayers were infected with *C. burnetii* and incubated for 4 weeks without replenishing the medium: the first week at 37°C in 5% CO₂ followed by 3 weeks at room temperature with the lids tightened, as previously described (8).

TABLE 3-1. PCR primers and their targets in *C. burnetii*

Target	Designation	Forward Sequence	Reverse Sequence
Cbu.L1917	L1917_flank	GTGGCTGCGACTGTTTAC	TTCGCTACCTTAGGACCG
	L1917_internal	GGACAATCAGCAGGAAAGAC	CGGACTCTATCATCACACTTA
	L1917+T7P	TAATACGACTCACTATAGGG- AAATTAACCGTTGTAGTTACG	TATTGACGTTATGTTAATCAT- GGG
Cbu.L1951	L1951_flank	AACGGTCCTAAGGTAGCG	ATTTCCGACCGTGCTGAG
	L1951_internal	TTTAGCAAAGGGCAATCC	TCCATAGTCACTTACTTCTTG
	L1951+T7P	TAATACGACTCACTATAGGG- AAAATTGCCCGCTAAACTG	CACTCTTTCCTATGTTTCCATA- GGG
IVS	IVS_flank	CGTGGTGGAAAGGGAAAC	TGTCAGCATTTCGCACTTC
	IVS+T7P	ATCAGAAGACAGATGACAG- ATGACAG	TAATACGACTCACTATAGGGA- TCAGAAGACAGATGGCAGAAG
23S rRNA	23S	GTGGCTGCGACTGTTTAC	ATTTCCGACCGTGCTGAG
Cbu.DnaB	Intein_flank	TTAAAGGACGACGATTGGC	CCCATTTCCGGTGTTTAGC
CBU_1927	ParB	AGAAGTTCCTGCGGTTATCC	CAACACTGGCAATCGGAAG

Sequence analyses. Sequence data were obtained with an automated DNA sequencer (ABI3130x1) and a BigDye Terminator Cycle Sequencing Ready Reaction kit (ABI, Foster City, CA). Sequence analysis was accomplished using the ‘BLAST 2 sequences’ tool at the NCBI website <http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi> (71).

Ribosome isolation and intron detection. *E. coli* containing both Cbu.L1917 and Cbu.L1951 cloned in pCR2.1TOPO (pUM1) was grown to OD₆₀₀ = 0.5 in the presence of 100 µg/ml ampicillin and 1 mM IPTG. The cells were centrifuged at 10000 x g for 10 min at 4°C and lysed in Buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NH₄Cl, 10 mM Mg Cl₂), containing 10 mg/ml lysozyme (Sigma) and 200 U/ml RNaseOUT (Invitrogen) by a freeze/thaw procedure (62). Cell debris was removed by centrifugation at 16000 x g for 10 min at 4°C. Lysates were loaded on a linear 5-20% sucrose gradient made in Buffer A containing 2 mM 2-mercaptoethanol and centrifuged for 15 h at 43000 x g (SW28 rotor, Beckman Coulter, Fullerton, CA) at 4°C. Fractions were collected from the gradient, assayed at 260 nm and ribosomes recovered by ethanol precipitation (67). *C. burnetii* ribosomes were isolated using a similar procedure except that bacteria were purified from Vero cells 72-h post infection (~2x10⁸ genomes) and lysed twice with a French pressure cell (10,000 psi) (67). RNA was isolated from ribosome fractions using a RiboPure-Bacteria Kit (Ambion) and cDNAs generated using an iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA). 23S rRNA was detected by PCR (Table 1: 23S). The presence of intron RNA was detected using primer sets specific for each intron (Table 1: L1917_internal and L1951_internal).

Quantitative RT-PCR (qRT-PCR). To prepare synchronized co-cultures, Vero cells were infected with SCVs prepared as above. RNA and DNA were isolated from the same

flask on days 0, 2, 4, 6, 8, 10, 12 and 14 post-infection using TRI-Reagent (Ambion) as per a protocol recommended by the manufacturer (<http://www.ambion.com/techlib/tn/123/13.html>). RNA was treated with TURBO DNase (Ambion) and quantified using spectrometry. One μg of RNA from each sample was converted to cDNA using an iScript cDNA Synthesis kit (Bio-Rad). qRT-PCR and qPCR were performed (Table 1: L1917_internal, L1951_internal) on cDNA and genomic DNA, respectively, using an iQ SYBR Green Supermix (Bio-Rad) on a MyiQ Single-Color Real-Time PCR detection System (Bio-Rad). Cycling parameters were 95°C for 5 min followed by 40 cycles of 95°C for 15 s and 60°C for 30 s. To control for DNA contamination of RNA samples, PCR was also performed on RNA that had not been reverse transcribed. Fold differences were calculated by comparing each value (2 -14 d) to that of day-0 (taken as 1.0). Amplified cDNA was normalized to genomic equivalents and plotted as fold difference.

Intein analyses. Cbu.DnaB, along with proximal flanking sequences, was amplified using the PCR primer set Intein_flank (Table 1) and *C. burnetii* genomic DNA. The amplicon was cloned in-frame into pQE-31 (Invitrogen) and cloned out-of-frame into pQE-30 (Invitrogen) utilizing BamH1 and Pst1 to produce pUM6 and pUM12, respectively. The whole *dnaB* was not cloned into the expression vector to avoid toxicity in *E. coli*, as reported previously (78). *E. coli* [M15 (pREP4)] transformed with pUM6, pUM12 or pQE-31 was grown to logarithmic phase (OD_{600} , 0.5) and 1 mM of IPTG was added to induce expression for 4 h at 37°C. Resulting cell pellets were solubilized in Laemmli sample buffer and used in gradient SDS-PAGE (10-15% acrylamide; w/v). Proteins were visualized by staining with Coomassie brilliant blue and appropriate bands

were excised from the gel and submitted to Alphalyse, Inc. (Palo Alto, CA), for matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) peptide mass fingerprinting and MALDI-TOF/TOF peptide sequencing, after trypsin digestion.

Statistical analyses and graphics. SigmaPlot 8.0 (Systat Software Inc., San jose, CA) was used for statistical analyses. PowerPoint 2003 and Excel 2003 (Microsoft, Redmond, WA) were used to generate graphs and figures.

D. RESULTS

Introns retard growth rate of *E. coli*. A previous article showed that *E. coli* expressing the 26S rRNA intron from *Tetrahymena thermophila* displayed a significantly decreased growth rate relative to controls, prompting us to check whether the same is true for the 23S rRNA introns of *C. burnetii* (51). Both Cbu.L1917 and Cbu.L1951 were cloned individually into pCR 2.1-TOPO (Invitrogen) to produce pUM2 and pUM3, respectively. As a control, the IVS located upstream of the introns in the 23S rRNA gene was similarly cloned but in opposite orientation to the vector's *lac* promoter to generate pUM4 and to transcribe a nonsense RNA of intermediate length (445 bases) relative to introns (Cbu.L1917, 288 bases; Cbu.L1951, 720 bases). Growth rates of *E. coli* strains transformed with pUM2, pUM3 or pUM4 were monitored spectrophotometrically for 5 h. As shown in Fig. 1, *E. coli* expressing either intron exhibited a significantly retarded growth rate when compared to the control, with Cbu.L1917 being more inhibitory than Cbu.L1951. Similar growth inhibition was observed when purified intron RNAs (25 µg) were electroporated directly into *E. coli* (not shown). Also, when Cbu.L1917 antisense

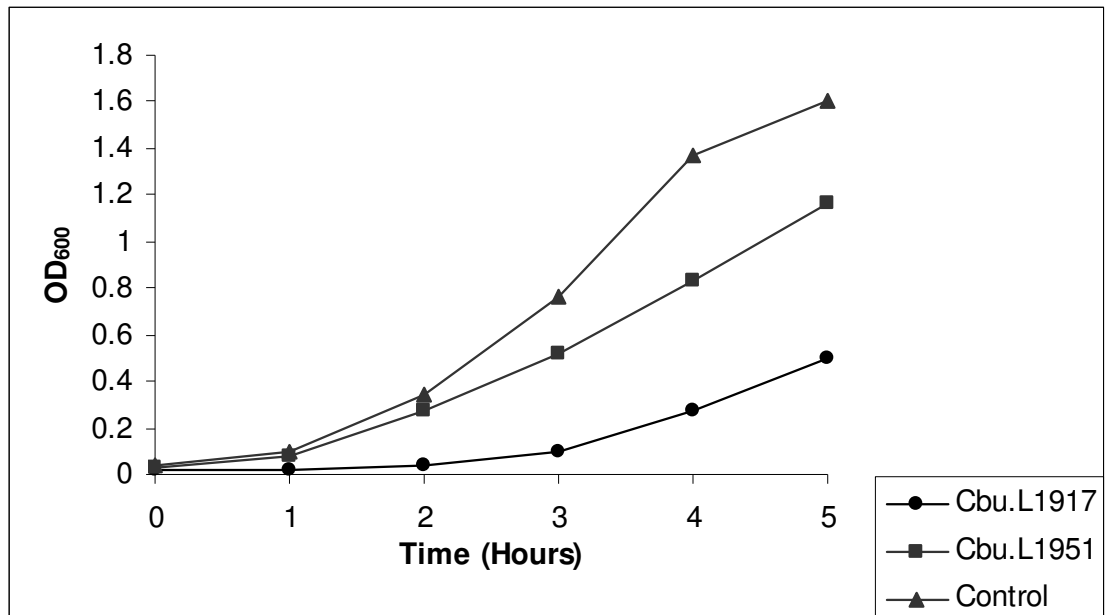


FIG.3-1. Effect of *Coxiella* ribozymes on *E. coli* growth. *E. coli* expressing cloned Cbu.L1917, Cbu.L1951 or an irrelevant control RNA (pUM2, pUM3 and pUM4, respectively) were induced with IPTG (0.1 mM) and assayed spectrophotometrically for growth at 37°C over 5 h at OD 600 nm. A representative growth curve is shown.

RNA was electroporated into *E. coli* transformed with pUM2, normal growth was restored (not shown).

Ribozymes associate with both *E. coli* and *C. burnetii* ribosomes. In the same report, *Tetrahymena* intron RNA (ribozyme) was found associated with *E. coli* ribosomes (51). We therefore analyzed ribosomes from both *C. burnetii* and from *E. coli* (pUM1), for ribosome-ribozyme complexes. Ribosomes were harvested from *C. burnetii* and *E. coli* (pUM1). RNA was isolated from ribosomal pellets, converted to cDNA and analyzed by PCR (Table 1: L1917_internal and L1951_internal) for the presence of ribozymes. Ribozymes were found associated with ribosomes of both *E. coli* (pUM1) and *C. burnetii* (Fig. 2A lanes 1 and 5, and Fig. 2B lanes 1, and 5, respectively). Further, PCR done with primers (Table 1: 23S) designed to give a large (1559 bp) amplicon if introns are unspliced and a small (551 bp) product if introns are spliced out, confirmed that the mature 23S rRNA does not contain unspliced introns (not shown). To verify these results, PCR analysis using a primer set (Table 1: ParB) for *Coxiella parB* (CBU_1927, encoding a chromosomal partitioning protein) was also performed. *parB* has been previously shown to be strongly expressed alongside rRNA genes by *Chlamydia trachomatis* (another obligate intracellular bacterium) during an entire 7d infection period in vitro (20, 30). Results showed that while *parB* mRNA is abundant in 0-8 d *Coxiella* co-cultures, it was not detectable in *Coxiella* ribosomal fractions, suggesting that intron-specific cDNA PCR products from ribosomes did not arise from contaminating RNA (not shown).

Introns inhibit in vitro transcription/translation. To directly analyze the hypothesized impact of intron RNA on protein synthesis, we used an *E. coli* S30 in vitro transcription/translation system in conjunction with pBEST*luc* (encoding luciferase) in

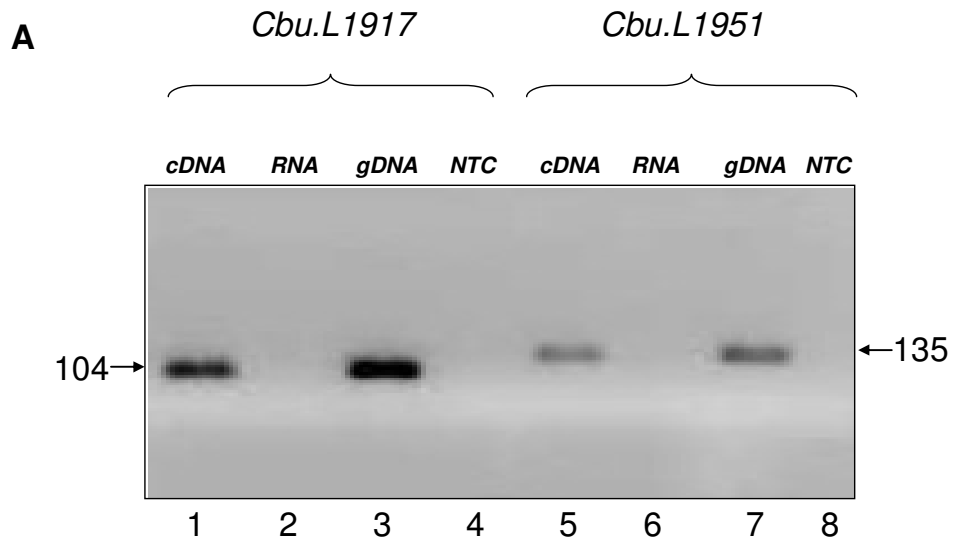
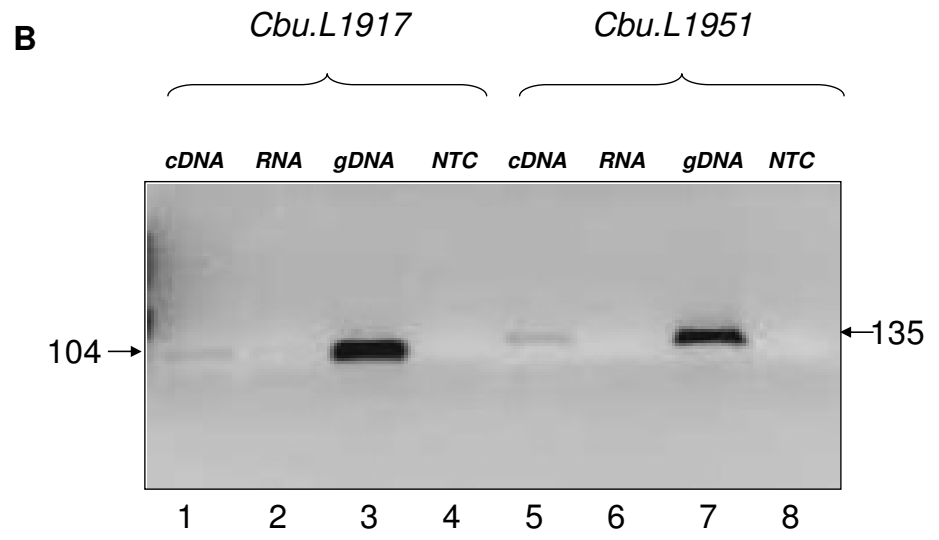


FIG. 3-2. Ribozyme-ribosome association in: (A) *E. coli*, (B) *C. burnetii*. PCR was done using L1917_internal or L1951_internal primer sets on RNA isolated from *E. coli* or *C. burnetii* ribosomes, cDNA made from RNA, *C. burnetii*-genomic DNA (gDNA), or a no-template control (NTC). An ethidium bromide-stained agarose gel (2% agarose; w/v) is shown. Amplicon sizes were determined from standards and are given on either side in base pairs



the presence of ribozymes or control RNA followed by a Luciferase Assay (Promega). Following a 60-min reaction, a luciferase substrate was added and luminescence immediately quantified by luminometry. Results show that both Cbu.L1917 and Cbu.L1951 RNAs significantly decrease (>30%, $p < 0.05$) luminescence as compared to the control RNA (Fig. 3).

Inverse correlation between *Coxiella* genome and intron RNA quantities. To build upon our observation that slow growth results from ribozyme-mediated inhibition of translation, we hypothesized that if the introns affect the growth rate of *Coxiella*, then their RNA quantities should inversely correlate to the rate of growth. Since *C. burnetii* is an obligate intracellular bacterium that does not grow in axenic medium, toxicity could only be measured indirectly. To this end, we compared *C. burnetii* genome content and ribozyme quantities measured over time in co-cultures. Total RNA and genomic DNA were isolated every 48 h from 0-14 d post-infection. *C. burnetii* genome and ribozyme quantities were estimated by qPCR and qRT-PCR, respectively, using primer sets specific for each intron (Table 1: L1917_internal, L1951_internal). Resulting data show a clear, inverse correlation between the quantities of *C. burnetii* genomes and either ribozyme, suggesting that the amount of intron RNA influences the rate of growth (Fig. 4). After an initial rapid decline when *C. burnetii* transitions from SCV to LCV (2 – 4 d post-infection) (8), intron RNA quantities stabilize at 6 -14 d at a nominal level. *Coxiella* genome amounts increased slowly from 0 – 2 d (lag phase), then rapidly from days 2 - 8 (log phase) and changed little between days 10 and 12 (stationary phase); however, it was observed to decrease between days 12 and 14 (our last time point).

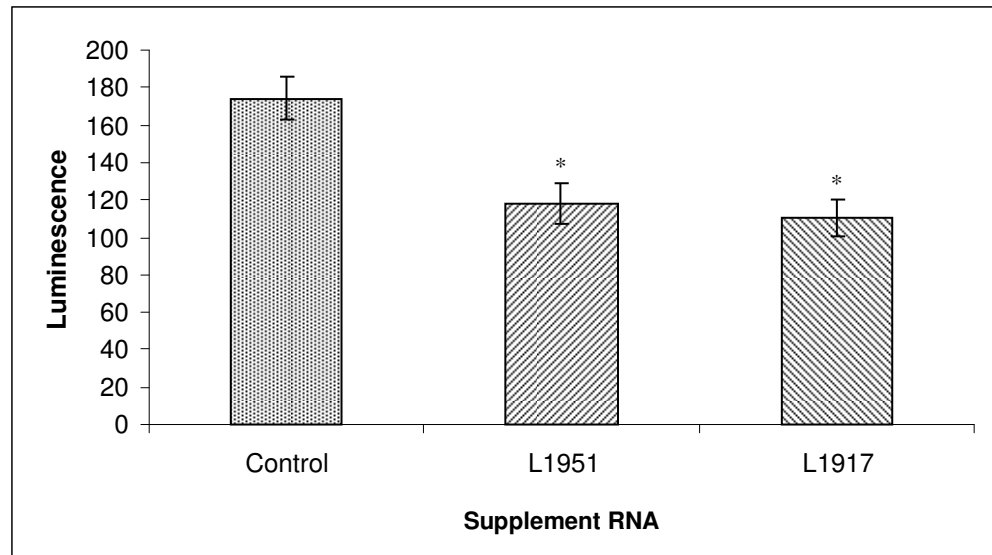


FIG. 3-3. Ribozymes inhibit in vitro translation of luciferase. In vitro translation of luciferase from pBEST luc in the presence of 25 μ g of Cbu.L1917, Cbu.L1951 or control RNA was done for 60 min and luminescence immediately measured thereafter. Data represent the means of three independent experiments \pm SD (* $p < 0.05$ by student's T test).

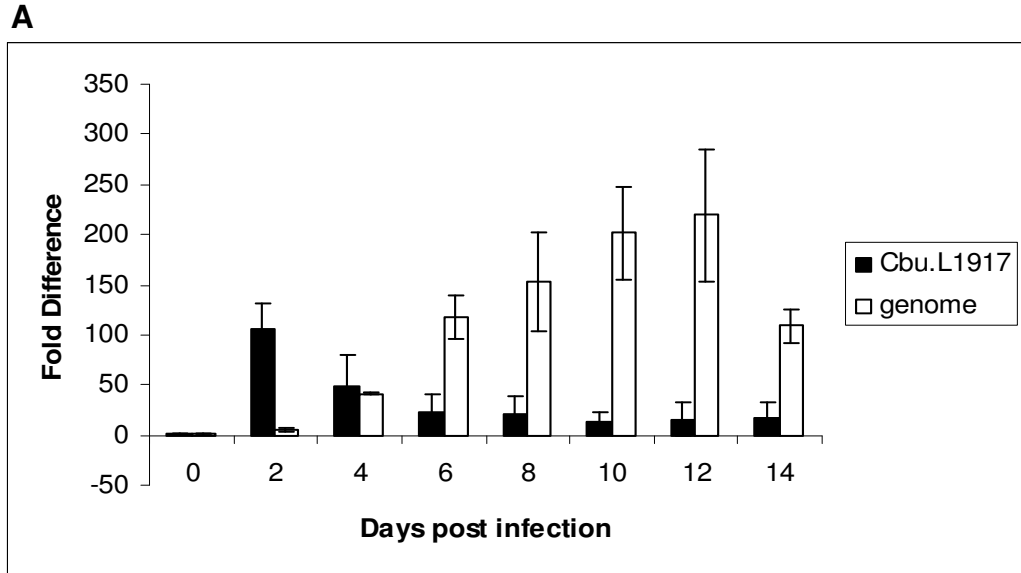
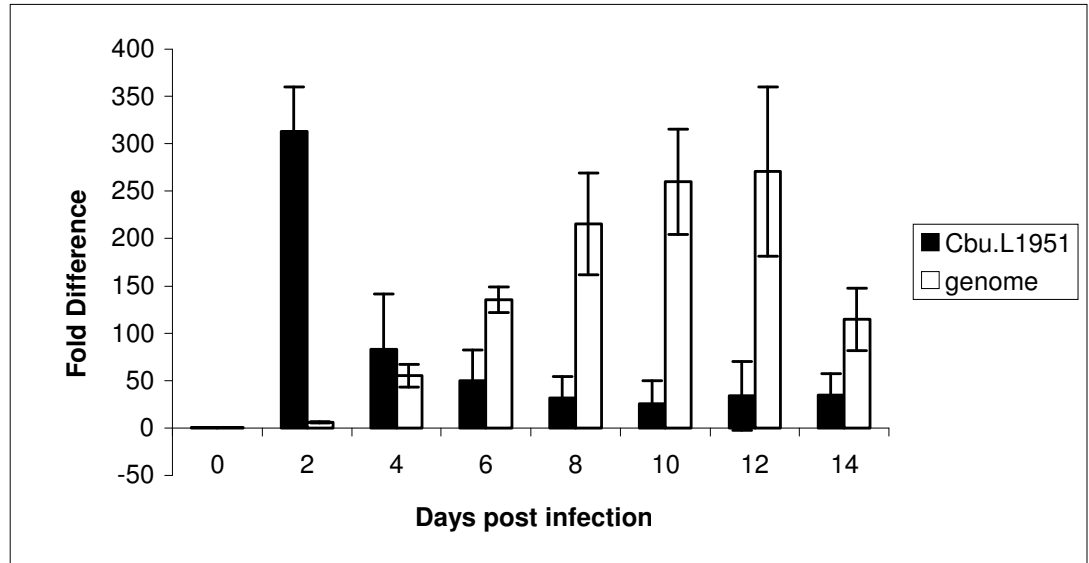


FIG. 3-4. Inverse correlation between quantities of *Coxiella* genome and ribozymes during growth. q-PCR and qRT-PCR data showing relative *C. burnetii* genome and Cbu.L1917 (**A**) and Cbu.L1951 (**B**) ribozyme quantities as a function of time. Data represent the means of three experiments \pm SD.

B



Intein splicing from DnaB. The putative intein (Cbu.DnaB) detected in the replicative DNA helicase gene (CBU_0868) was analyzed to determine its splicing properties (55, 66). PCR was performed on *C. burnetii* genomic DNA using the Intein_flank primer set (Table 1) to amplify the intein and proximal flanking sequences. The resulting amplicon was cloned into pQE31 (Invitrogen) to produce pUM6, which encodes a translational fusion protein. *E. coli* (pUM6) was used to study the intein's splicing activity at the protein level. As a control, the same insert was cloned out-of-frame into pQE30 (Invitrogen) to form pUM12. Expression was induced using IPTG and the protein products were analyzed using SDS-PAGE (Fig. 5). The intein (I, ~16.8 kDa) was observed to splice out of the precursor protein (P, ~35.6 kDa) leaving the two exteins (E, ~23.6 kDa) spliced together (Fig. 5). The molecular masses of I and P determined from the gel corresponds well with their in silico predicted values of 16.3 kDa, and 34.7 kDa, respectively. The "E" band seems to be running slightly slower than its predicted molecular mass of 18.3 kDa. The identity of P, E and I bands were confirmed using MALDI-TOF peptide fingerprinting and MALDI-TOF/TOF peptide sequencing (data not shown). The P, E or I bands were not visible in the vector-only or pUM12 lanes showing that the bands are insert-specific and are produced only when the insert is in-frame. The large precursor band suggests that intein splicing progresses slowly, unlike group I introns, which splice rapidly (59).

Phylogenetic analyses. Neighbor-joining trees were built using MEGA4 to reconstruct the plausible evolutionary history of the intein (69). Phylogeny reconstruction showed that Cbu.DnaB clustered with two very similar inteins inserted at the same position in *dnaB* genes (alleles) of *A. ehrlichii* (Aeh.DnaB2) and *H. halophila* (Hha.DnaB2), with the

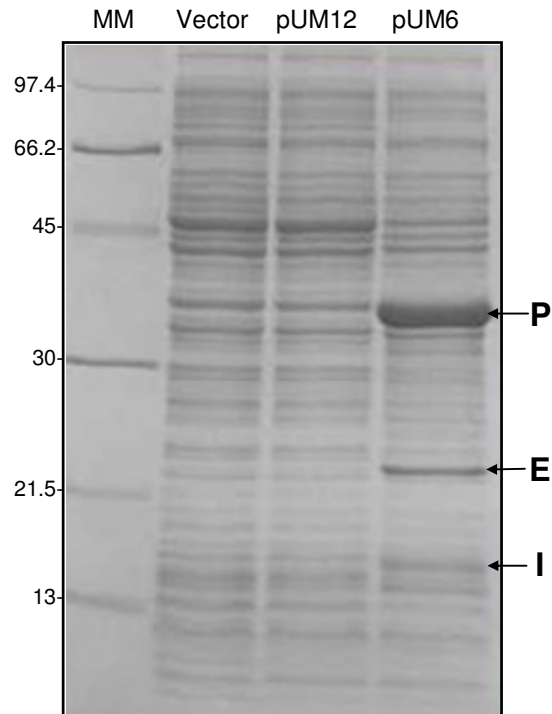


FIG. 3-5. Intein splicing. *E. coli* (pQE-31 “vector”, pUM6 or pUM12) were induced with IPTG for 4 h and protein profiles were analyzed using SDS-PAGE. **P**, precursor fusion protein (~35.6 kDa); **E**, spliced exteins (~23.6 kDa), **I**, intein (~16.8 kDa). Molecular masses (MM) were determined from standards and are given to the left in kDa.

exclusion of inteins inserted in other sites (non-alleles) (Fig. 6A). These extremophilic bacteria have not been previously shown to be related to *C. burnetii*, suggesting HGT was used by the intein to invade *C. burnetii*'s genome. To get a clearer picture, phylogenetic trees were constructed using 16S rRNA sequences and the amino acid sequences of the *dnaB* gene. DnaB was chosen to verify the 16S tree not only because it is the site of intein insertion but also due to the high degree of replicative DNA helicase conservation between bacterial species. Results show that although *L. pneumophila* and *R. gryllii* are the closest known relatives to *C. burnetii*, as previously reported (11, 63), *A. ehrlichei* and *H. halophila* are also related to *C. burnetii* albeit with weak bootstrap support (Figs. 6B and 6C). Similar results were obtained when phylogenetic trees were built using the Maximum Parsimony method (not shown) (16).

E. DISCUSSION

All genes can be considered selfish, but in most instances a gene increases its chance of maintenance by increasing the fitness of the host organism (14). However, many genetic elements like introns and inteins reside and replicate within the genome at the host's expense; hence it is perhaps more appropriate to describe them as parasitic (1, 23). Group I introns and inteins invade highly conserved host genes involved in vital functions utilizing an encoded HE. Curiously, HEs themselves are selfish/parasitic elements that have invaded introns and inteins. Based on the presence of conserved amino acid motifs, HEs are divided into four families: LAGLIDADG, GIY-YIG, HNH and His-cys box, with LAGLIDADG being the most common (25). While the route of entry for both group I introns and inteins might be the same, i.e., HE-mediated target-site cleavage and host-mediated double-strand DNA repair, their exit strategies are dramatically different;

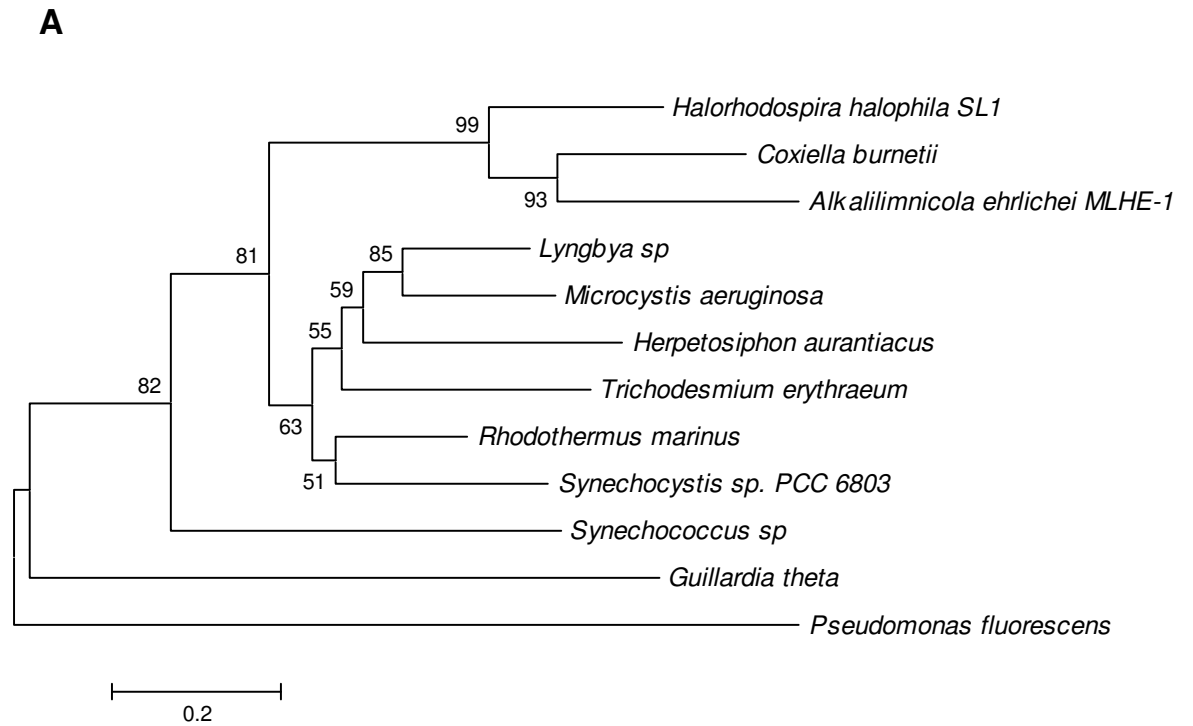
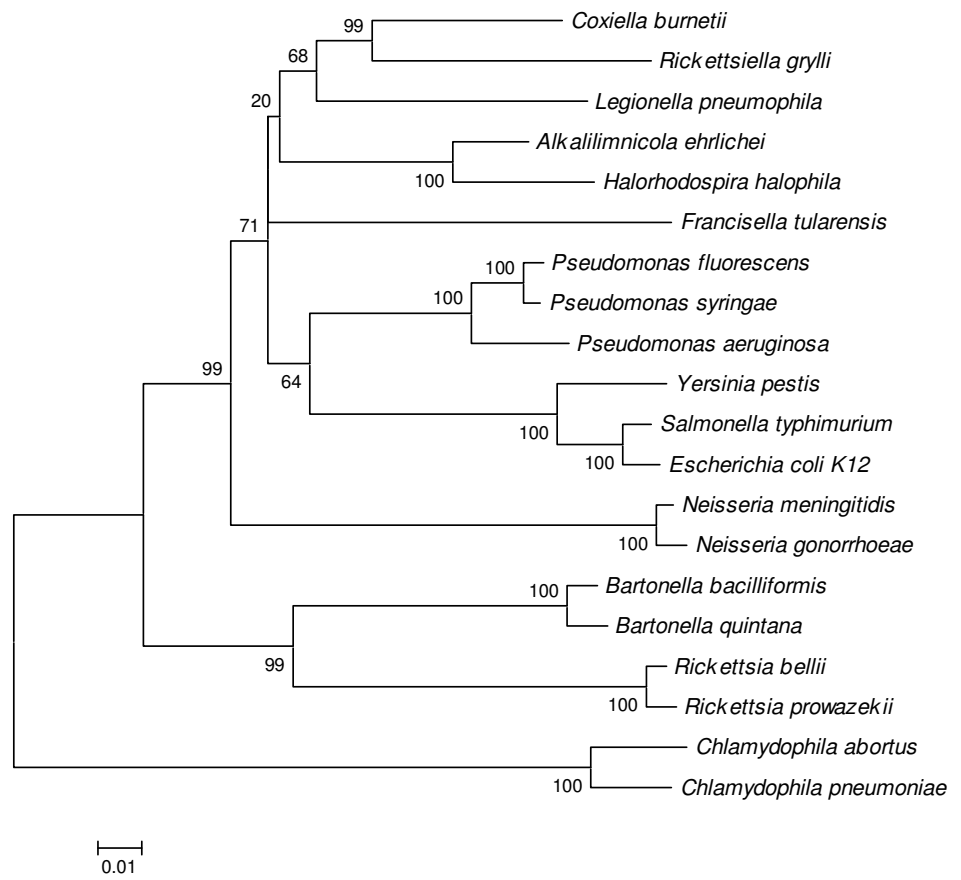
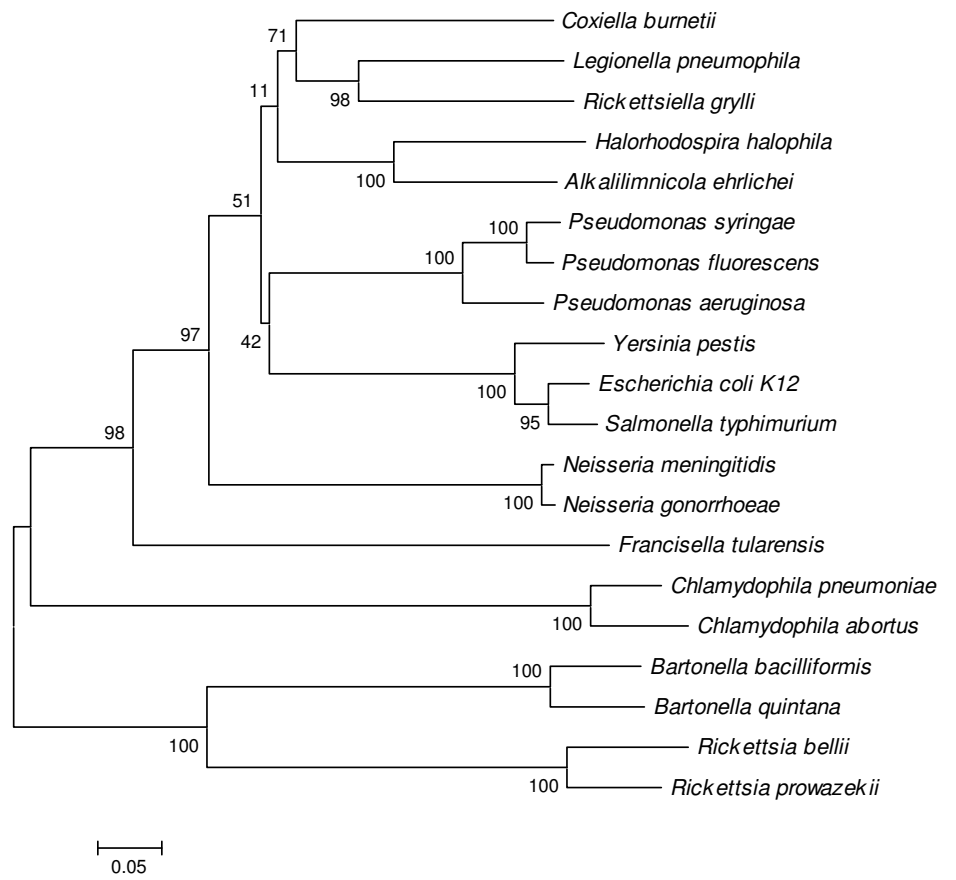


FIG. 3-6. Phylogenetic analyses. Neighbor-Joining trees for: (A) inteins, (B) 16S and (C) DnaB. Intein and DnaB phylogenies were computed using amino acid sequences (125 and 434 positions, respectively) and Poisson correction method. 16S tree was built using DNA sequence (1351 positions) and Maximum Composite Likelihood method. Bootstrap values (1000 replicates) are indicated at the nodes.

B



C



introns splice out as RNA whereas inteins splice out of proteins. In this study, we describe the “toxic” (i.e. growth-inhibitory) property exhibited by two self-splicing group I introns- Cbu.L1917 and Cbu.L1951 - that interrupt the 23S rRNA and the functionality of Cbu.DnaB, the intein found in the C-terminal portion of the replicative DNA helicase (DnaB) of *C. burnetii*.

Due to the lack of genetic systems for *C. burnetii*, we resorted to *E. coli* models for most of our experiments. However, both in vivo *E. coli* experiments and the in vitro S30 system are appropriate models for studying *Coxiella* introns because of the high degree of conservation between the 23S rRNA of *C. burnetii* and *E. coli*. In fact, exon sequences surrounding and separating the two introns in *C. burnetii* are 100% identical between the two bacteria. The observed association of ribozymes with the ribosomes of both bacteria supports the utility of the model. Since the IGS of both ribozymes are complementary to their respective splice junctions, it is very likely that intron RNAs associate with the 23S rRNA at these regions (Fig. 7). Binding of Cbu.L1917 RNA to helix 69 and Cbu.L1951 RNA to helix 71 of 23S rRNA would likely interfere with the formation of vital intersubunit (30S-50S) bridges B2a and B3, respectively (79). This interference could make the ribosome unstable, thereby affecting its ability to translate efficiently, leading to a retarded growth rate (37). Cbu.L1917 elicited greater inhibition of *E. coli* growth than Cbu.L1951, consistent with the in vitro observations where inhibition of luciferase translation by Cbu.L1917 RNA was slightly but consistently greater than by Cbu.L1951 RNA. Interestingly, predicted ribozyme-ribosome associations (Fig. 7) show more complementary interactions between Cbu.L1917's IGS and helix 69 than between Cbu.L1951's IGS and helix 71 (15 and 9 complementary bases, respectively). Moreover,

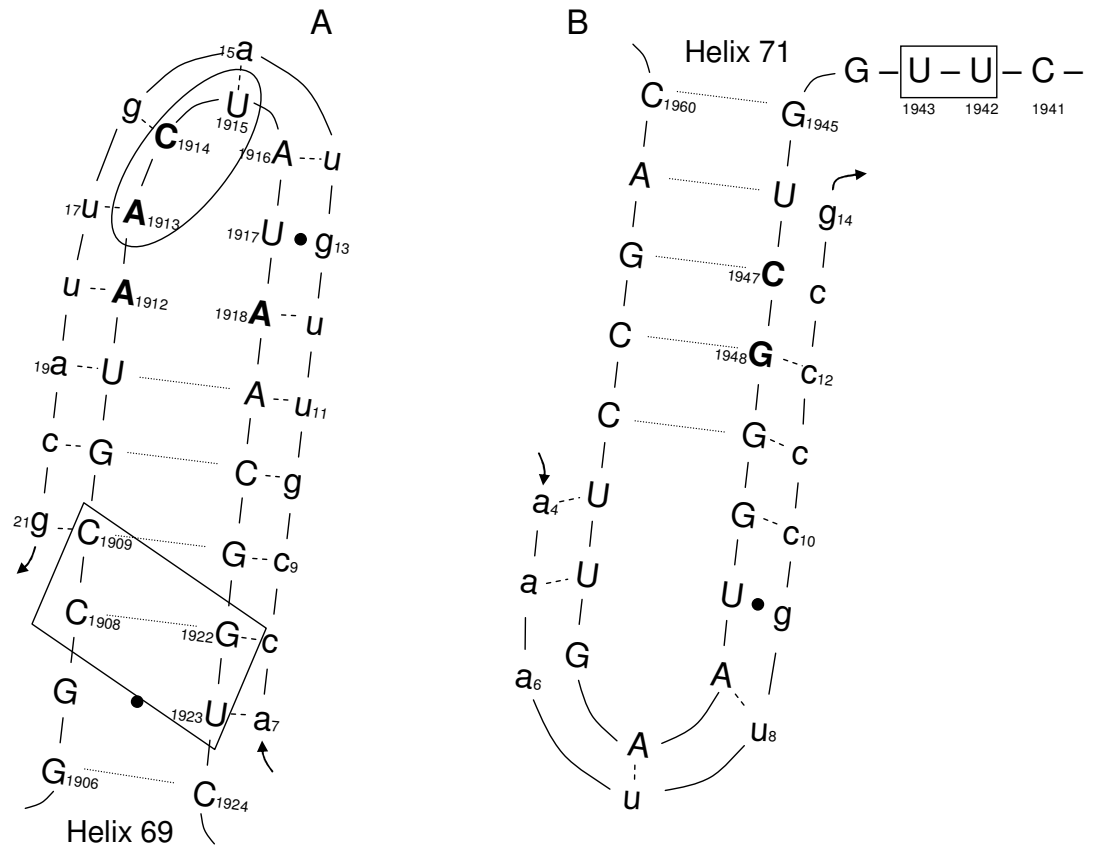


Figure 3-7. Potential ribozyme-ribosome interactions. (A) Masking of 23S rRNA helix 69 (upper case) by Cbu.L1917's IGS (lower case). Nucleotides at positions 1912, 1913, 1914 and 1918 (in bold) involved in forming the intersubunit bridge B2a. Positions 1913, 1914, and 1915 (within oval) interact with A-site tRNA. Positions 1908, 1909, 1922, and 1923 (within parallelogram) interact with P-site tRNA. Arrows, flanking portions of Cbu.L1917. **(B)** Binding of Cbu.L1951's IGS (lower case) with 23S rRNA helix 71 (upper case). Nucleotides 1947 and 1948 (in bold) involved in forming the intersubunit bridge B3. Positions 1942 and 1943 (within rectangle) interact with P-site tRNA. Arrows, flanking portions of Cbu.L1951.

if Cbu.L1917 RNA binds to helix 69 as predicted (Fig. 7A), it would mask contact sites with P-site tRNA (bases 1908, 1909, 1922, 1923) and with A-site tRNA (bases 1913, 1914 and 1915). It would also interfere with the formation of inter-subunit bridge B2a (bases 1912, 1913, 1914, and 1918). On the other hand, binding of Cbu.L1951 RNA to helix 71 would partially interrupt bridge B3 (base 1948) and may interfere with the contact sites with P-site tRNA (bases 1942 and 1943) (28, 36).

A potential disparity in target-affinity and structural hindrance could explain observed differences in growth inhibition caused by the two ribozymes. During the growth cycle of *C. burnetii*, Cbu.L1917 RNA is found consistently at lower amounts than Cbu.L1951, especially on day 2 of our experiment, when bacterial growth rate is poised to accelerate (Fig. 4), suggesting that RNA from Cbu.L1917 is more toxic than that of Cbu.L1951. Since equal amounts of both intron RNAs will be synthesized at any given time, the observed difference in concentration is likely due to differences in intron RNA half-lives, with Cbu.L1917 being more labile than Cbu.L1951, again supporting the possibility that Cbu.L1917 is more toxic than Cbu.L1951. In contrast to the relatively stable 16S rRNA levels observed during the growth cycle of *C. burnetii* (8), a rapid fall in the levels of both intron RNAs was observed when the bacteria transits from lag to log phase. This suggests a reduced half-life for the excised intron RNAs during log phase as compared to lag phase. This possibility fits well with earlier observations showing that the half-life of *Tetrahymena* intron RNA is much lower (5 sec) during rapid growth than in slow growth (30 sec) (5) and that the half-life of total *E. coli* mRNA is longer during lag and stationary phases when compared to log phase (33). Also, a study by Chan et al. has demonstrated that the half-life of an excised group I intron RNA is eight to 22 times

longer than that of a typical *E. coli* mRNA (6). Consequently, in *C. burnetii* it is possible that when the relatively fast-growing LCVs transition to metabolically quiescent SCVs, potentially toxic intron RNAs become more stable causing further growth retardation. However, this negative effect may not be significant since transcription and translation is minimal in SCVs.

Although group I introns are abundant in structural RNA genes of mitochondria and chloroplasts of lower eukaryotes, they are extremely rare in bacterial counterparts (19), (21). We recently described the unusual occurrence of two self-splicing group I introns in the 23S rRNA gene of *C. burnetii* (59). Given the predilection towards genome reduction in obligate intracellular bacteria, the presence of multiple introns in a single gene was intriguing by itself, but the additional observation that these introns can reduce bacterial growth rates raises some interesting questions: Why would a bacterium contain multiple introns in its vital and sole 23S rRNA gene when introns are so rare in other bacteria? Are the introns toxic but still tolerated by the bacterium? Is the growth-rate reduction caused by the introns beneficial to the bacterium or is it neutral? We believe that by examining the current life-style/biology of *C. burnetii* in light of its evolutionary history, we can elucidate this apparent paradox.

Bacterial genomes are constantly evolving under two opposing forces: gene loss and gene gain (3, 54). Mutational process, which is strongly biased towards deletion in bacteria, is responsible for most of the gene loss that results in condensed genomes with high densities of functional genes (45, 52). Bacteria acquire new genes via HGT; a well-established avenue for bacterial diversification and innovation, with acquisition of pathogenicity islands and antibiotic resistance being two of the most notorious examples

(49, 53). Usually these two opposing processes are more or less in equilibrium; however, when a bacterium shifts from a free-living to an intracellular life style, the balance between these two forces tips in favor of gene loss, resulting in shrunken genomes as observed in obligate intracellular pathogens and endosymbionts (4, 48). Evolutionary forces acting on an obligate intracellular bacterium like *C. burnetii* are quite different from those acting on an environmental bacterium. Constant availability of nutrients, low effective population size and bottlenecks during transmission render obligate intracellular bacteria susceptible to low purifying selection and genetic drift. This leads to the stochastic loss of several beneficial genes and accumulation of some slightly deleterious mutations (Muller's ratchet) (46, 47). In the obligate endosymbiotic bacterium *Buchnera aphidicola*, the aforementioned forces have not only reduced the genome but also have caused the loss of beneficial DNA repair genes and accumulation of slightly deleterious mutations in vital genes such as *groEL* and 16S rRNA (27, 34). A similar picture can be envisioned in *C. burnetii*. Available evidence suggests that an ancestor of *C. burnetii* recently moved from the environment into the intracellular niche with concomitant loss of some beneficial genes, like *recBCD* (66). The starter population would possibly have suffered severe bottlenecks before the extant bacterium, which is fully adapted to living in acidic parasitophorous vacuoles (PVs), arose. It has been shown that nutrient-laden vesicles of endocytic and/or autophagic pathway(s) fuse with *Coxiella*'s PVs providing a constant supply of nutrients (26, 61). Living in a privileged niche with adequate nutrients and low effective population size conceivably resulted in fixation of genes with low fitness, like the toxic introns. In bacteria like *E. coli*, living in unpredictable environments and with large effective populations, these low-fitness genes would be rapidly removed

by selection, hence the observed paucity of group I introns (17). Given the likelihood for reductive evolution and limited opportunities for HGT in *C. burnetii*, it is likely that these introns will be lost at some future time point unless the elements serve an adaptive role and are hence retained by positive selection. It is important to point out that while it is assumed that the intracellular niche occupied by *C. burnetii* precludes HGT, it is possible that some of the parasitic genetic elements originated from other co-infecting bacteria or from host cell organelles, like mitochondria and chloroplast [intracellular arena hypothesis, (4)].

The possibility that the introns are beneficial to *C. burnetii* is strengthened by the observation that intron sequences are highly conserved among all eight genomic groups of the pathogen, isolated from disparate hosts and geographic regions (59). It is conceivable that the growth rate reduction caused by the introns fosters chronic infection and persistence in the host. Slow growth rate is a trait shared by most bacteria that cause chronic infections (32, 38, 58, 68, 73). While acute Q fever is self-limiting, the immune system is often unable to clear all bacteria leading to persistence in the host. A study by Marmion et al. showed that over 88% of infected persons were PCR positive for *C. burnetii* twelve years after acute primary infection (39). By promoting slow growth, introns might in turn be promoting host survival and their own positive selection. Another possibility is that the introns are evolutionarily neutral, with minimal additional negative impact on the growth rate of *C. burnetii*, which already is extremely slow. In this scenario, the introns would eventually be lost from the population and would not reappear due to a lack of HGT, thus resulting in a more compact genome (22).

Like group I introns, inteins invade a new host via HE-mediated targeting (15). Once inteins are fixed in a population they tend to lose the HE gene (HEG) and accumulate mutations in regions not essential for splicing. Subsequently, the intein itself is lost from the population by accurate deletion and only through another HGT event can the intein regain access to the population (56). Cbu.DnaB has apparently lost its HEG, suggesting that it is in the second phase of its cycle and has been present for some time in *C. burnetii*. Interestingly, the G+C ratio and codon bias of the intein is similar to that of the *C. burnetii* genome (data not shown), again suggesting that intein-invasion occurred a long time ago, providing the bacterium with adequate time to ameliorate possible nucleotide differences (35). Given enough time, the intein will likely be lost, and due to the paucity of HGT may not be able to reinvade *C. burnetii*. Phylogenetic and BLAST analyses using the intein sequence identified inteins inserted at the same loci of *A. ehrlichei* and *H. halophila* as Cbu.DnaB's closest neighbors (71). To date, these are the only inteins found at this specific site in *dnaB* (alleles), and it has been previously shown that intein alleles are more closely related to each other than to non-alleles (inteins at different insertion sites) (55). The presence of homologous inteins in two extremophiles suggests that HGT is the most parsimonious explanation for their source. However, additional phylogenetic and BLAST analyses using 16S and DnaB sequences suggest that *C. burnetii*, *A. ehrlichei* (an haloalkaliphile found in soda lakes), and *H. halophila*, (one of the most halophilic bacteria known), share a closer evolutionary relationship than previously suspected. Thus, even though HGT is the most probable route of intein acquisition, it is possible that the element was inherited vertically from a common ancestor. Since a non-functional intein could convert a functional gene into a pseudogene

(23), and given the pseudogenization going on in *C. burnetii* (66), we were curious to assess the intein's functionality. Our analyses demonstrate that the intein is indeed functional (Fig. 5), hence the mature DnaB of *C. burnetii* is undoubtedly functional. In a recent analyses of the *C. burnetii* proteome, the spot identified as DnaB (No. 16) is clearly smaller (~ 50 kDa) than the molecular mass predicted from GenBank (accession no. NC002971; ~67 kDa), suggesting that the intein is removed to form the mature DnaB protein (65). These observations are not surprising since DnaB is essential (2).

A fascinating possibility is that the presence of the intein in an essential protein is beneficial to *C. burnetii*. Intein splicing appears to be a slow process (Fig. 5), and before the intein is excised, the DnaB precursor is probably non-functional, thereby reducing the pool of mature DnaB (24). Since the amount of DnaB in each bacterial cell is thought to be very low (43), intein-mediated scarcity of functional DNA helicase might create a lag in replication, thereby fostering slow growth. Again, similar to the argument made for introns, *Coxiella*'s slow growth rate may obviate the urgent need for mature DNA helicase proteins during replication, thus rendering the intein functionally neutral (molecular atavism). In any case, the availability of an obligate post-translational processing step provides opportunity for regulation. Although a regulatory role for inteins has not been demonstrated to date, it is possible that the intein excises in response to a signal, thus making available mature, functional proteins at the appropriate time. Indeed, inteins have been engineered with single amino acid substitutions to respond to pH and temperature (7). In conclusion, the parasitic genetic elements acquired by *C. burnetii* in the past, and retained in its present obligatory intracellular niche, undoubtedly influence

its biology. Further research is needed to delineate precise roles played by these intriguing elements in *C. burnetii*.

F. ACKNOWLEDGMENTS

We are grateful to Jim Battisti, Sherry Coleman, Nermi Parrow, Kate Sappington, William Knight, Dan Mummey, Sudhakar Agnihothram, John Gajewski and Meghan Lybecker (University of Montana) for helpful discussions and technical assistance. We thank Patty McIntire (UM Murdock Sequencing Facility) for sequence analyses. This work was supported by the NIH Rocky Mountain Regional Center of Excellence for Biodefense and Emerging Infectious Disease grant, U54 AI065357-030023.

G. REFERENCES

1. **Avise, J. C.** 2001. Evolving genomic metaphors: a new look at the language of DNA. *Science* **294**:86-7.
2. **Belle, J. J., A. Casey, C. T. Courcelle, and J. Courcelle.** 2007. Inactivation of the DnaB helicase leads to the collapse and degradation of the replication fork: a comparison to UV-induced arrest. *J Bacteriol* **189**:5452-62.
3. **Berg, O. G., and C. G. Kurland.** 2002. Evolution of microbial genomes: sequence acquisition and loss. *Mol Biol Evol* **19**:2265-76.
4. **Bordenstein, S. R., and W. S. Reznikoff.** 2005. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* **3**:688-99.
5. **Brehm, S. L., and T. R. Cech.** 1983. Fate of an intervening sequence ribonucleic acid: excision and cyclization of the *Tetrahymena* ribosomal ribonucleic acid intervening sequence in vivo. *Biochemistry* **22**:2390-7.
6. **Chan, W. K., G. Belfort, and M. Belfort.** 1988. Stability of group I intron RNA in *Escherichia coli* and its potential application in a novel expression vector. *Gene* **73**:295-304.
7. **Chong, S., K. S. Williams, C. Wotkowicz, and M. Q. Xu.** 1998. Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. *J Biol Chem* **273**:10567-77.
8. **Coleman, S. A., E. R. Fischer, D. Howe, D. J. Mead, and R. A. Heinzen.** 2004. Temporal analysis of *Coxiella burnetii* morphological differentiation. *J Bacteriol* **186**:7344-52.

9. **Cooper, A. A., Y. J. Chen, M. A. Lindorfer, and T. H. Stevens.** 1993. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J* **12**:2575-83.
10. **Cooper, A. A., and T. H. Stevens.** 1993. Protein splicing: excision of intervening sequences at the protein level. *Bioessays* **15**:667-74.
11. **Cordaux, R., M. Paces-Fessy, M. Raimond, A. Michel-Salzat, M. Zimmer, and D. Bouchon.** 2007. Molecular characterization and evolution of arthropod-pathogenic *Rickettsiella* bacteria. *Appl Environ Microbiol* **73**:5045-7.
12. **Cutler, S. J., M. Bouzid, and R. R. Cutler.** 2007. Q fever. *J Infect* **54**:313-8.
13. **Dassa, B., H. Haviv, G. Amitai, and S. Pietrokovski.** 2004. Protein splicing and auto-cleavage of bacterial intein-like domains lacking a C'-flanking nucleophilic residue. *J Biol Chem* **279**:32001-7.
14. **Dawkins, R.** 1976. *The Selfish gene*. Oxford University Press, Oxford, UK.
15. **Derbyshire, V., and M. Belfort.** 1998. Lightning strikes twice: intron-intein coincidence. *Proc Natl Acad Sci U S A* **95**:1356-7.
16. **Eck, R. V., Dayhoff, M.O.** 1966. *Atlas of protein sequence and structure*. National Biomedical Research Foundation, Silver Springs, MD.
17. **Edgell, D. R., M. Belfort, and D. A. Shub.** 2000. Barriers to intron promiscuity in bacteria. *J Bacteriol* **182**:5281-9.
18. **Ellison, D. W., T. R. Clark, D. E. Sturdevant, K. Virtaneva, S. F. Porcella, and T. Hackstadt.** 2008. Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa. *Infect Immun* **76**:542-50.
19. **Everett, K. D., S. Kahane, R. M. Bush, and M. G. Friedman.** 1999. An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* **181**:4734-40.
20. **Gerard, H. C., B. Krausse-Opatz, Z. Wang, D. Rudy, J. P. Rao, H. Zeidler, H. R. Schumacher, J. A. Whittum-Hudson, L. Kohler, and A. P. Hudson.** 2001. Expression of *Chlamydia trachomatis* genes encoding products required for DNA synthesis and cell division during active versus persistent infection. *Mol Microbiol* **41**:731-41.
21. **Giovannoni, S. J., S. Turner, G. J. Olsen, S. Barns, D. J. Lane, and N. R. Pace.** 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. *J Bacteriol* **170**:3584-92.
22. **Goddard, M. R., and A. Burt.** 1999. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A* **96**:13880-5.
23. **Gogarten, J. P., A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, and E. Hilario.** 2002. Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**:263-87.
24. **Hashimoto, H., M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, and Y. Kai.** 2001. Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J Mol Biol* **306**:469-77.
25. **Haugen, P., D. M. Simon, and D. Bhattacharya.** 2005. The natural history of group I introns. *Trends Genet* **21**:111-9.
26. **Heinzen, R. A., M. A. Scidmore, D. D. Rockey, and T. Hackstadt.** 1996. Differential interaction with endocytic and exocytic pathways distinguish

- parasitophorous vacuoles of *Coxiella burnetii* and *Chlamydia trachomatis*. Infect Immun **64**:796-809.
27. **Herbeck, J. T., D. J. Funk, P. H. Degnan, and J. J. Wernegreen.** 2003. A conservative test of genetic drift in the endosymbiotic bacterium *Buchnera*: slightly deleterious mutations in the chaperonin groEL. Genetics **165**:1651-60.
 28. **Hirabayashi, N., N. S. Sato, and T. Suzuki.** 2006. Conserved loop sequence of helix 69 in *Escherichia coli* 23 S rRNA is involved in A-site tRNA binding and translational fidelity. J Biol Chem **281**:17203-11.
 29. **Hoeft, S. E., J. S. Blum, J. F. Stolz, F. R. Tabita, B. Witte, G. M. King, J. M. Santini, and R. S. Oremland.** 2007. *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int J Syst Evol Microbiol **57**:504-12.
 30. **Hogan, R. J., S. A. Mathews, S. Mukhopadhyay, J. T. Summersgill, and P. Timms.** 2004. Chlamydial persistence: beyond the biphasic paradigm. Infect Immun **72**:1843-55.
 31. **Johnson, A. K., D. A. Baum, J. Tye, M. A. Bell, and S. M. Testa.** 2003. Molecular recognition properties of IGS-mediated reactions catalyzed by a *Pneumocystis carinii* group I intron. Nucleic Acids Res **31**:1921-34.
 32. **Kussell, E., R. Kishony, N. Q. Balaban, and S. Leibler.** 2005. Bacterial persistence: a model of survival in changing environments. Genetics **169**:1807-14.
 33. **Kuzj, A. E., P. S. Medberry, and J. L. Schottel.** 1998. Stationary phase, amino acid limitation and recovery from stationary phase modulate the stability and translation of chloramphenicol acetyltransferase mRNA and total mRNA in *Escherichia coli*. Microbiology **144 (Pt 3)**:739-50.
 34. **Lambert, J. D., and N. A. Moran.** 1998. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. Proc Natl Acad Sci U S A **95**:4458-62.
 35. **Lawrence, J. G., and H. Ochman.** 1997. Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol **44**:383-97.
 36. **Liiv, A., and M. O'Connor.** 2006. Mutations in the intersubunit bridge regions of 23 S rRNA. J Biol Chem **281**:29850-62.
 37. **Maivali, U., and J. Remme.** 2004. Definition of bases in 23S rRNA essential for ribosomal subunit association. RNA **10**:600-4.
 38. **Manabe, Y. C., and W. R. Bishai.** 2000. Latent *Mycobacterium tuberculosis*-persistence, patience, and winning by waiting. Nat Med **6**:1327-9.
 39. **Marmion, B. P., P. A. Storm, J. G. Ayres, L. Semendric, L. Mathews, W. Winslow, M. Turra, and R. J. Harris.** 2005. Long-term persistence of *Coxiella burnetii* after acute primary Q fever. QJM **98**:7-20.
 40. **Mathys, S., T. C. Evans, I. C. Chute, H. Wu, S. Chong, J. Benner, X. Q. Liu, and M. Q. Xu.** 1999. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene **231**:1-13.

41. **Maurin, M., A. M. Benoliel, P. Bongrand, and D. Raoult.** 1992. Phagolysosomes of *Coxiella burnetii*-infected cell lines maintain an acidic pH during persistent infection. *Infect Immun* **60**:5013-6.
42. **McCaul, T. F., T. Hackstadt, and J. C. Williams.** 1981. Ultrastructural and biological aspects of *Coxiella burnetii* under physical disruptions. Academic Press Inc., New York, NY.
43. **Messer, W., Weigel, C.** 1996. Initiation of chromosome replication. *In* C. F. Neidhardt (ed.), *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, 2 ed. ASM Press, Washington D.C.
44. **Meyer, T. E.** 1985. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim Biophys Acta* **806**:175-83.
45. **Mira, A., H. Ochman, and N. A. Moran.** 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**:589-96.
46. **Moran, N. A.** 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **93**:2873-8.
47. **Moran, N. A.** 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**:583-6.
48. **Moran, N. A., and G. R. Plague.** 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev* **14**:627-33.
49. **Narra, H. P., and H. Ochman.** 2006. Of what use is sex to bacteria? *Curr Biol* **16**:R705-10.
50. **Nei, M., Kumar, S.** 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.
51. **Nikolcheva, T., and S. A. Woodson.** 1997. Association of a group I intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA* **3**:1016-27.
52. **Ochman, H., and L. M. Davalos.** 2006. The nature and dynamics of bacterial genomes. *Science* **311**:1730-3.
53. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299-304.
54. **Ochman, H., and N. A. Moran.** 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**:1096-9.
55. **Perler, F. B.** 2002. InBase: the Intein Database. *Nucleic Acids Res* **30**:383-4.
56. **Pietrovski, S.** 2001. Intein spread and extinction in evolution. *Trends Genet* **17**:465-72.
57. **Pietrovski, S.** 1998. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* **7**:64-71.
58. **Proctor, R. A., C. von Eiff, B. C. Kahl, K. Becker, P. McNamara, M. Herrmann, and G. Peters.** 2006. Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Microbiol* **4**:295-305.
59. **Raghavan, R., S. R. Miller, L. D. Hicks, and M. F. Minnick.** 2007. The unusual 23S rRNA gene of *Coxiella burnetii*: two self-splicing group I introns flank a 34-base-pair exon, and one element lacks the canonical Ω G. *J Bacteriol* **189**:6572-9.

60. **Roman, J., and S. A. Woodson.** 1998. Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. *Proc Natl Acad Sci U S A* **95**:2134-9.
61. **Romano, P. S., M. G. Gutierrez, W. Beron, M. Rabinovitch, and M. I. Colombo.** 2007. The autophagic pathway is actively modulated by phase II *Coxiella burnetii* to efficiently replicate in the host cell. *Cell Microbiol* **9**:891-909.
62. **Ron, E. Z., R. E. Kohler, and B. D. Davis.** 1966. Polysomes extracted from *Escherichia coli* by freeze-thaw-lysozyme lysis. *Science* **153**:1119-20.
63. **Roux, V., M. Bergoin, N. Lamaze, and D. Raoult.** 1997. Reassessment of the taxonomic position of *Rickettsiella grylli*. *Int J Syst Bacteriol* **47**:1255-7.
64. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406-25.
65. **Samoilis, G., A. Psaroulaki, K. Vougas, Y. Tselentis, and G. Tsiotis.** 2007. Analysis of whole cell lysate from the intercellular bacterium *Coxiella burnetii* using two gel-based protein separation techniques. *J Proteome Res* **6**:3032-41.
66. **Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg.** 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci U S A* **100**:5455-60.
67. **Spedding, G.** 1990. Isolation and characterization of ribosomes from prokaryotes, eukaryotes and organelles, p. 1-27, *Ribosome and protein synthesis; a practical approach*. Oxford University Press, Oxford, England.
68. **Stewart, G. R., B. D. Robertson, and D. B. Young.** 2003. Tuberculosis: a problem with persistence. *Nat Rev Microbiol* **1**:97-105.
69. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**:1596-9.
70. **Tamura, K., M. Nei, and S. Kumar.** 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**:11030-5.
71. **Tatusova, T. A., and T. L. Madden.** 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**:247-50.
72. **Thomson, N. R., C. Yeats, K. Bell, M. T. Holden, S. D. Bentley, M. Livingstone, A. M. Cerdeno-Tarraga, B. Harris, J. Doggett, D. Ormond, K. Mungall, K. Clarke, T. Feltwell, Z. Hance, M. Sanders, M. A. Quail, C. Price, B. G. Barrell, J. Parkhill, and D. Longbottom.** 2005. The *Chlamydomophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. *Genome Res* **15**:629-40.
73. **Tierrez, A., and F. Garcia-del Portillo.** 2005. New concepts in *Salmonella* virulence: the importance of reducing the intracellular growth rate in the host. *Cell Microbiol* **7**:901-9.

74. **Vicens, Q., and T. R. Cech.** 2006. Atomic level architecture of group I introns revealed. *Trends Biochem Sci* **31**:41-51.
75. **Williams, J. C.** 1991. Infectivity, virulence and pathogenicity of *Coxiella burnetii* for various hosts. CRC Press, Boca Raton, FL.
76. **Williams, J. C., M. G. Peacock, and T. F. McCaul.** 1981. Immunological and biological characterization of *Coxiella burnetii*, phases I and II, separated from host components. *Infect Immun* **32**:840-51.
77. **Woodson, S. A.** 2005. Structure and assembly of group I introns. *Curr Opin Struct Biol* **15**:324-30.
78. **Wu, H., M. Q. Xu, and X. Q. Liu.** 1998. Protein trans-splicing and functional mini-inteins of a cyanobacterial DnaB intein. *Biochim Biophys Acta* **1387**:422-32.
79. **Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller.** 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**:883-96.
80. **Zuckermandl, E., Pauling L.** 1965. Evolutionary divergence and convergence in proteins. *In* V. Bryson (ed.), *Evolving genes and proteins* Academic Press, New York.

CHAPTER FOUR

Intron Toxicity is Caused by Both Sequence-Dependent and Sequence-Independent Mechanisms and Pentamidine Inhibits *Coxiella burnetii* growth in Vero Cells

A. INTRODUCTION

The two group I introns found in *C. burnetii* were observed to associate with bacterial ribosomes and interfere with their function (104). Since the IGSs of Cbu.L1917 and Cbu.L1951 are complementary to helix 69 and helix 71 of the 23S rRNA, respectively, we hypothesized that the introns associate with ribosomes using their IGS. This scenario seemed plausible since interfering with the function of helices 69 and 71 would affect ribosome function due to a failure to form inter-subunit bridges B2a and B3, respectively (138). To test this, we constructed introns without IGS and also chimeric RNAs that have the IGS but with non-relevant RNAs fused to it in place of the intron. Results show that introns without IGS could still inhibit *E. coli* growth, but IGS was found to be important for the inhibition of ribosome function in vitro. Since the introns are encoded in the sole 23S rRNA gene of *C. burnetii*, it is vital to remove them before mature 23S rRNA is formed. The dependence on intron splicing to form functional ribosomes makes any agent that can block this process a potentially attractive therapeutic agent against Q fever. We tested and determined that pentamidine, a known inhibitor of group I intron splicing in vitro, can inhibit *C. burnetii* growth in Vero cultures.

A predicted secondary structure shows that the IVS is inserted into helix 45 as a stem-loop structure with possible RNase III processing sites. The protein encoded by the ORF (CBU_2096) within the IVS was found to be highly toxic to *E. coli*, such that only

clones with the ORF out-of-frame or in opposite orientation to the promoter were recovered. The reconstructed phylogenetic history of this protein showed orthologous proteins encoded within IVSs of *Leptospira borgpetersenii* and *Carboxydotherrnus hydrogenoformans*.

B. MATERIALS AND METHODS

***E. coli* growth assay.** PCR was performed on *Coxiella* genomic DNA using primers (Table 1: L1917w/oIGS and L1951w/oIGS) that amplified the introns excluding their IGS regions, and the PCR products were cloned into pCR2.1-TOPO per manufacturer's instructions (Invitrogen) in the proper orientation downstream of the *lac* promoter. To construct chimeric RNAs, PCR primers L1917IGS+SVI and L1951IGS+SVI (Table 1) were used. These primers were designed by adding either intron's IGS sequence to the 5' end of each forward primer, such that the resulting amplicons have an IGS at the 5' end followed by the IVS sequence in reverse orientation in place of the intron. *E. coli* (Top10F') transformed with each plasmid was grown 2 h to mid-logarithmic phase at 37°C in LB broth in the presence of 100 µg/ml ampicillin (Sigma-Aldrich) and then used to inoculate 50 ml fresh LB with 100 µg/ml ampicillin and 1 mM IPTG to an OD₆₀₀ of 0.03 at 0 h. Bacterial growth was assayed spectrophotometrically at 600 nm at 60 min intervals (37°C, shaking).

In vitro transcription. To generate templates to synthesize RNAs, introns without IGS were amplified using *C. burnetii* genomic DNA and PCR primers designed to add a

T7 promoter sequence at the 5' end of the amplicon (Table 1: L1917w/oIGS+T7P, L1951w/oIGS+T7P). In vitro transcription was performed using a MEGAscript Kit (Ambion), and resulting RNA was purified using NucAway (Ambion) spin columns and TURBO DNase (Ambion) as per manufacturer's instructions.

In vitro transcription/ translation and luciferase assay. Utilizing pBEST*luc* as the template, luciferase was synthesized in vitro for 1 h using an *E. coli* S30 Extract System for Circular DNA (Promega), as per manufacturer's instructions, in the presence of 25 µg of either intron RNAs without their respective IGS segments or control RNA. Luciferase Assay Reagent (Promega) was added to the in vitro translation reactions and luminescence measured immediately using a luminometer (PerkinElmer).

Phylogenetic analyses. The phylogenetic history of CBU_2096 was reconstructed using its amino acid sequence. Other related proteins were identified by a BLAST search (3). Only sequences that were identified to be encoded in 23S rRNA were used. Also, when multiple strains of the same organism contained the same ORF, only one sequence was included in the analyses. In addition to CBU_2096 (YP_022738.1), the other ORFs used in this study were from *Leptospira borgpetersenii* serovar Hardjo-bovis L550 (YP_798587), *Carboxydotherrnus hydrogenoformans* Z-2901 (YP_359321), *Chlorobium phaeobacteroides* BS1 (ZP_00531740.1), *Flavobacteria bacterium* BAL38 (EAZ94741.1), *Anabaena variabilis* ATCC 29413 (YP_320317.1), *Prosthecochloris aestuarii* DSM 271 (ZP_00592577.1), *Acidobacteria bacterium* Ellin345 (YP_593439.1), *Crocospaera watsonii* WH 8501 (ZP_00516874.1), *Algoriphagus* sp. PR1

Table 4-1. Primers and their targets used in this study.

Target	Designation	Forward Primer	Reverse Primer
Cbu.L1917	L1917w/oIGS	TAAATAAACTTGGCTATATGC	TATTGACGTTATGTTAATCATG
	L1917IGS+SVI	ACCGTTGTAGTTACGCAGAAG ACAGATG	CCTGATAGTTATTGCGGAAGT CTACCC
	L1917w/oIGS +T7P	TAATACGACTCACTATAGGG- TAAATAAACTTGGCTATATGC	TATTGACGTTATGTTAATCATG
Cbu.L1951	L1951w/oIGS	GGTAGTGATAAAAGTCAGC	CACTCTTTCCTATGTTTCC
	L1951IGS+SVI	AAATTGCCCGCAGAAGAC AGATGACAG	CCTGATAGTTATTGCGGAAGT CTACCC
	L1951w/oIGS +T7P	TAATACGACTCACTATAGGG- TAGTGATAAAAGTCAGCTC	CACTCTTTCCTATGTTTCCATA- GGG

(ZP_01721093.1), *Geobacter bemidjiensis* (ZP_01777007.1), *Kordia algicida* OT-1 (ZP_02160952.1) and *Candidatus cloacamonas acidaminovorans* (CAO81281.1).

Neighbor-joining trees that were bootstrap replicated (1000 replicates) were built using MEGA4 (112), (122). An optimal tree is shown with branch lengths depicting evolutionary distances computed by the Poisson correction method for the number of amino acid substitutions per site (140). All positions containing gaps were eliminated, resulting in a total of 106 positions being used in the final dataset.

Pentamidine susceptibility testing. African green monkey kidney (Vero) fibroblasts (CCL-81; American Type Culture Collection) that were grown in 6-well tissue culture plates for 2 d were infected with *C. burnetii* Nine Mile phase II (RSA 439; clone 4) and allowed to grow in RPMI medium (Invitrogen) supplemented with 10% fetal bovine serum (HyClone) in the presence of 0, 10, 25, 50 or 100 µg/ml of pentamidine isethionate (Sigma). After 72 hrs, the cells were removed and total DNA extracted using a High Pure PCR template preparation kit (Roche) according to manufacturer's recommendations. Q-PCR was done using an iQ SYBR Green Supermix (Bio-Rad) on a MyiQ Single-Color Real-Time PCR detection System (Bio-Rad). Cycling parameters were 95°C for 5 min followed by 40 cycles of 95°C for 15 s and 60°C for 30 s. Fold differences were calculated by comparing each value on day-3 to that of day-0 (taken as 1.0).

Cell viability. Vero cell viability under pentamidine treatment was verified using a CytoTox 96 Assay (Promega) per manufacturer's recommendations. The supernatant from the 72-h Vero culture infected with *C. burnetii* and treated with 50 or 100 µg/ml of

pentamidine was collected and measured for the presence of lactate dehydrogenase (LDH) released from lysed cells. Supernatant from cultures without pentamidine and fresh RPMI supplemented with 10% fetal bovine serum were used as controls. Cell viability was also tested using the vital dye trypan blue. Both pentamidine-treated and untreated Vero cells were stained with trypan blue and viable cells counted using a hemocytometer and light microscopy.

Statistical analyses and graphics. PowerPoint 2003 and Excel 2003 (Microsoft) were used to generate graphs and figures. InStat (GraphPad, La Jolla, CA) was used to conduct paired t-tests and a P value of ≤ 0.05 was considered statistically significant.

C. RESULTS

Introns without IGS retard growth rate of *E. coli*. We previously showed that both Cbu.L1917 and Cbu.L1951 associate with bacterial ribosomes and inhibit their function causing growth rate retardation (Chapter 3). Based on the complementarities between the IGS of Cbu.L1917 and helix 69 and Cbu.L1951's IGS and helix 71 of 23S rRNA, we hypothesized that the introns were associating with the ribosomes using their IGS. To test this, we constructed and cloned Cbu.L1917 and Cbu.L1951 without their IGSs into pCR2.1-TOPO to produce pUM7 and pUM8, respectively. Growth rates of *E. coli* strains transformed with pUM7, pUM8 or pUM4 (Control, Table 3) were monitored spectrophotometrically for 5 h. Growth rates of *E. coli* transformed with pUM2 (intact Cbu.L1917) or pUM3 (intact Cbu.L1951) were also measured in parallel. As shown in

Figs. 1.A and 1.B, *E. coli* transformed with either pUM7 or pUM8 exhibited significantly retarded growth rates compared to the control, but they were not significantly different from that of pUM2 or pUM3. These data suggest that the IGS does not play a significant role in the *E. coli* growth retardation caused by the ribozymes.

In vitro transcription/translation in the presence of ribozymes lacking IGS. To test whether ribosomes were the target for introns that lack IGS, we used an *E. coli* S30 in vitro transcription/translation system in conjunction with pBEST*luc* (encoding luciferase). Following a 60-min reaction in the presence of intact introns, introns without IGS or control RNA, Luciferase Assay Reagent (Promega) was added and luminescence immediately quantified by luminometry. Results show that even though both intact and IGS-less Cbu.L1917 RNAs significantly ($P < 0.05$) decrease luminescence as compared to the control RNA, the inhibition caused by the intact ribozymes is significantly ($P < 0.05$) greater than that by the IGS-less species (Fig. 2.A). Intact Cbu.L1951 RNA significantly ($P < 0.05$) inhibits luminescence, whereas Cbu.L1951 RNA without IGS did not inhibit luminescence significantly, when compared to the control (Fig. 2.B). These results show that IGSs, which are complementary to exon junctions, play an important role in ribosome inhibition caused by both ribozymes. In addition, the data suggests that Cbu.L1917 also inhibits ribosomes through an antisense-independent mechanism.

IGS alone cannot retard *E. coli* growth rate. From the above experiments, it is clear that IGS is important for ribozyme-mediated inhibition of ribosome function in vitro, but not for retardation of *E. coli*'s growth rate. These data suggest that IGS-mediated

A.

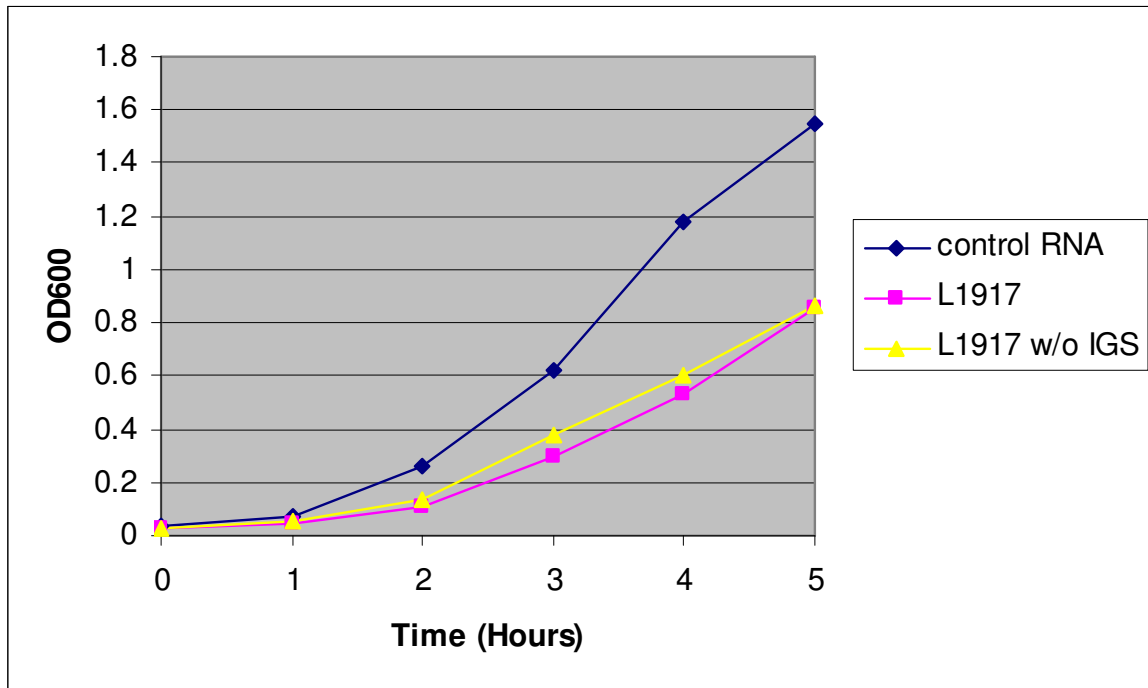
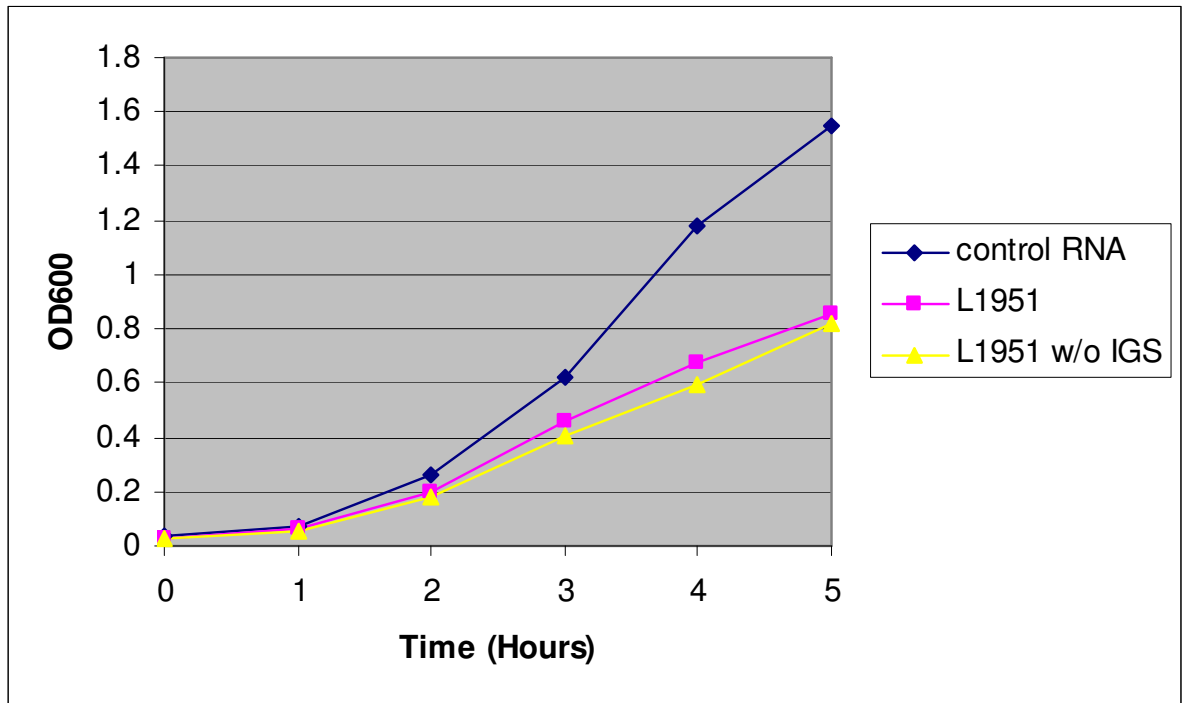


FIG.4-1. Effect of *Coxiella* introns lacking IGS on *E. coli* growth. (A) *E. coli* expressing cloned intact Cbu.L1917 (pUM2) or Cbu.L1917 without IGS (pUM7), and (B) *E. coli* expressing cloned intact Cbu.L1951 (pUM3) or Cbu.L1951 without IGS (pUM8) and an irrelevant control RNA (pUM4) were induced with IPTG (0.1 mM) and assayed spectrophotometrically at 600 nm for growth at 37°C over 5 h. A representative growth curve of three independent experiments is shown.

B.



A.

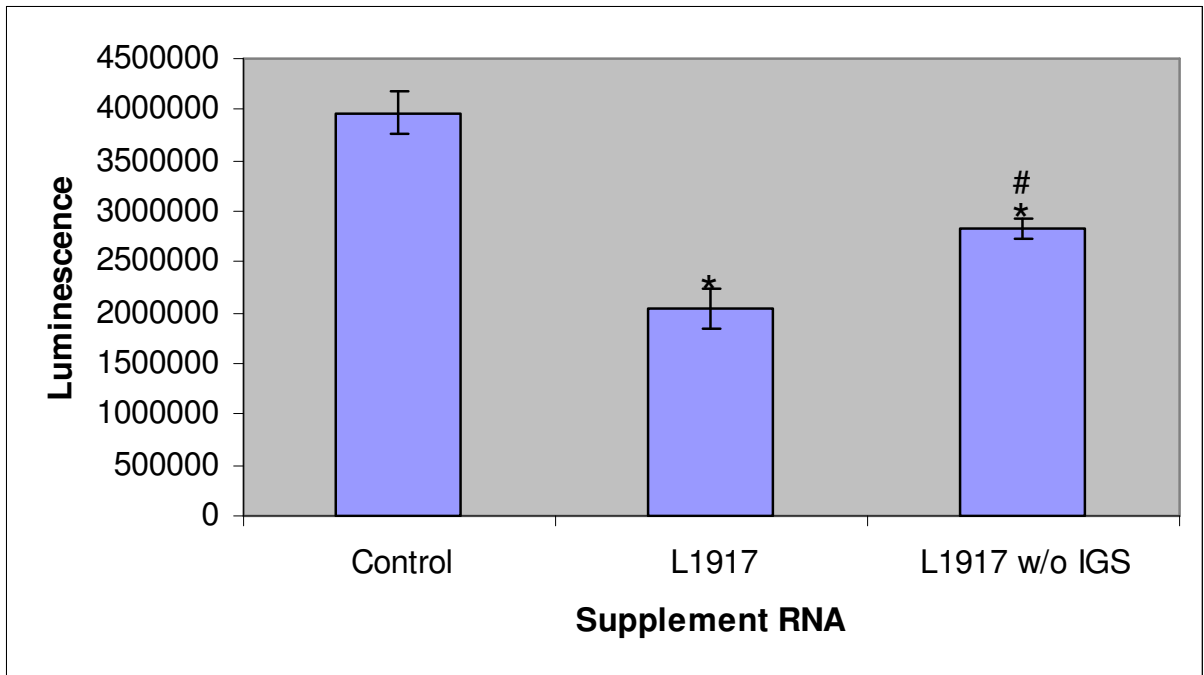
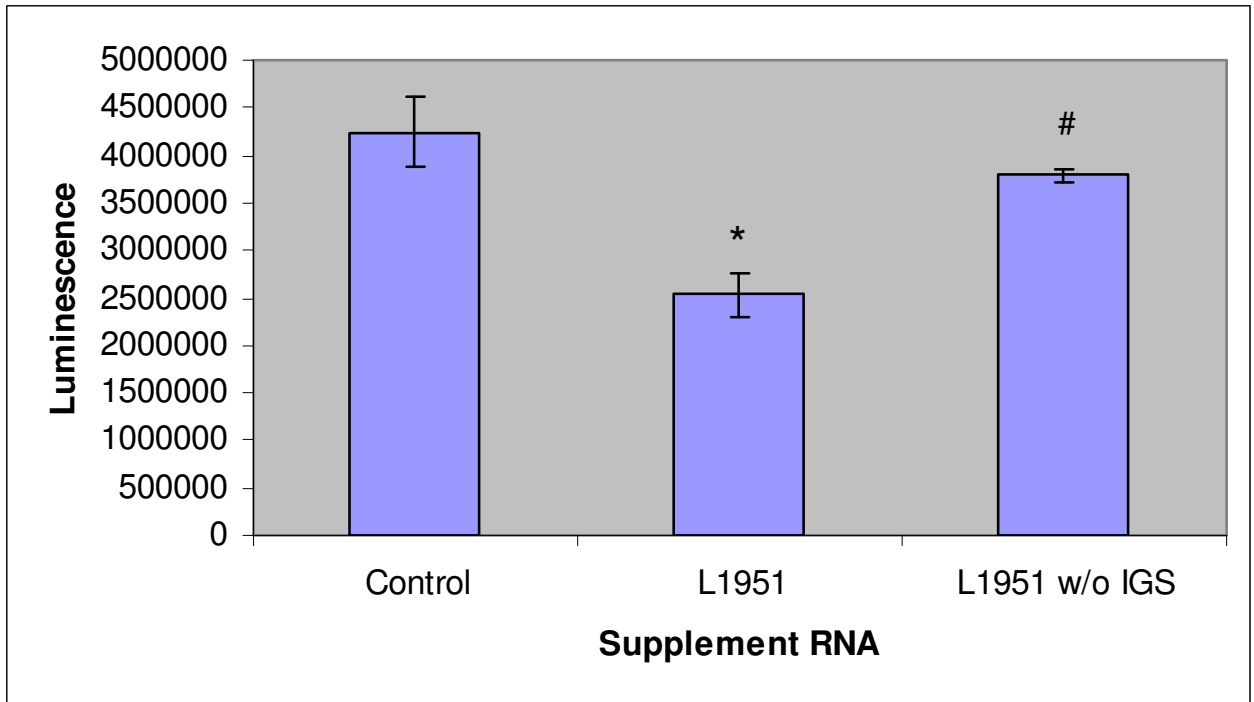


FIG.4-2. In vitro translation of luciferase in the presence of ribozymes lacking IGS.

In vitro translation of luciferase from pBEST luc in the presence of 25 μ g of intact Cbu.L1917 (L1917), Cbu.L1917 without IGS (L1917 w/o IGS) or control RNA (A) and intact Cbu.L1951 (L1951), Cbu.L1951 without IGS (L1951 w/o IGS) or control RNA (B) was done for 60 min and luminescence immediately measured thereafter. Data represent the means of three experiments \pm SD. *, ≤ 0.05 vs. control and #, ≤ 0.05 vs. intact ribozymes by paired t-test.

B.



ribosome inhibition is not a significant cause of the observed *E. coli* growth rate inhibition. To confirm this possibility, we constructed fusion RNAs that contain IGS at the 5' end but replaced the rest of the intron with a random stretch of RNA (see methods). We cloned these into pCR2.1-TOPO to produce pUM9 (Cbu.L1917's IGS) and pUM10 (Cbu.L1951's IGS). Growth rates of *E. coli* strains transformed with pUM9, pUM10 or pUM4 were monitored by OD₆₀₀ for 4 h. As shown in Fig. 3, all three strains have similar growth rates, showing that the IGS domains cannot inhibit bacterial growth by themselves.

Pentamidine inhibits *C. burnetii* growth in Vero cells. Since the two group I introns interrupt the sole 23S rRNA gene of *C. burnetii*, they have to be spliced out to form mature 23S rRNA. We tested the efficacy of pentamidine, a known inhibitor of group I intron splicing in vitro (89, 139), on *C. burnetii* grown in Vero cells. We found that pentamidine inhibits *C. burnetii* growth significantly (> 66%, $p < 0.05$) at all concentrations used (Fig. 4). To rule out pentamidine toxicity to Vero cells as a cause of the decreased growth, we also measured the level of lactate dehydrogenase (LDH) in the cocultures. LDH was not detected from the supernatant collected from Vero cultures treated with pentamidine (10, 50 or 100 µg/ml) demonstrating that pentamidine does not cause cell lysis (data not shown). We also performed Trypan blue staining and found no difference in viability between Vero cells treated with pentamidine (50 µg/ml and 100 µg/ml) and sham-treated cells (data not shown). Taken as a whole, these observations show that pentamidine is not toxic to Vero cells and has the potential to be used as an antimicrobial for *C. burnetii*.

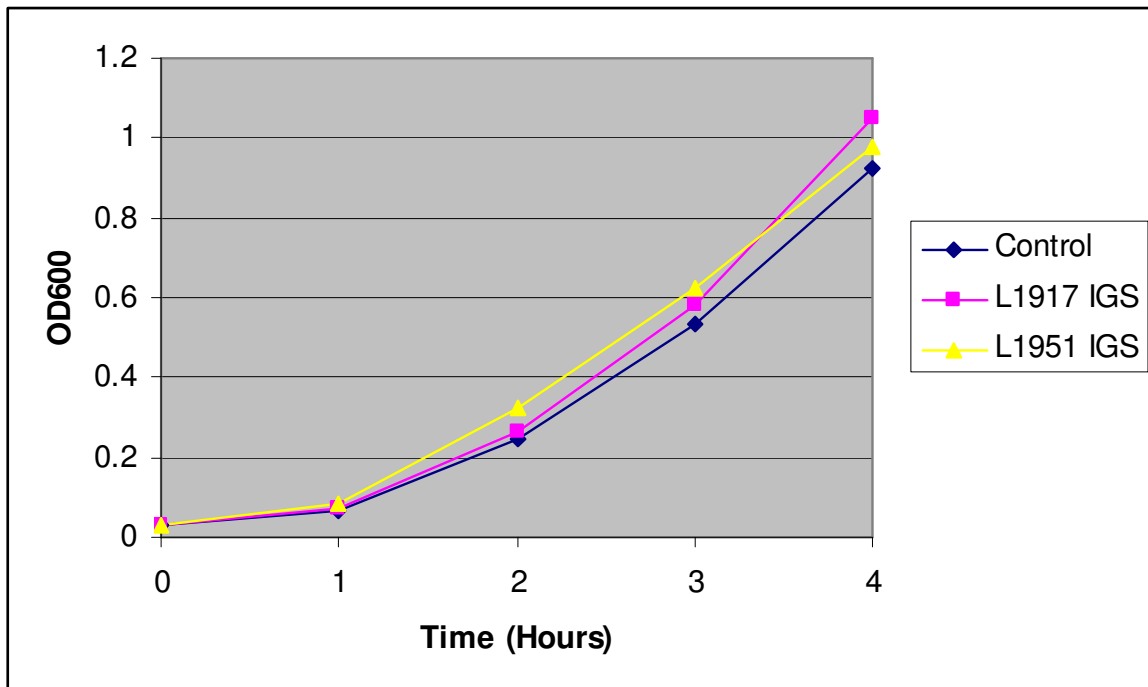


FIG. 4-3. Effect of IGS on *E. coli* growth. *E. coli* (pUM9 or pUM10) expressing fusion RNAs that have the IGS segments of Cbu.L1917 or Cbu.L1951, or an irrelevant control RNA (pUM4) were induced with IPTG (0.1 mM) and assayed spectrophotometrically at 600 nm for growth at 37°C over 4 h. A representative growth curve is shown.

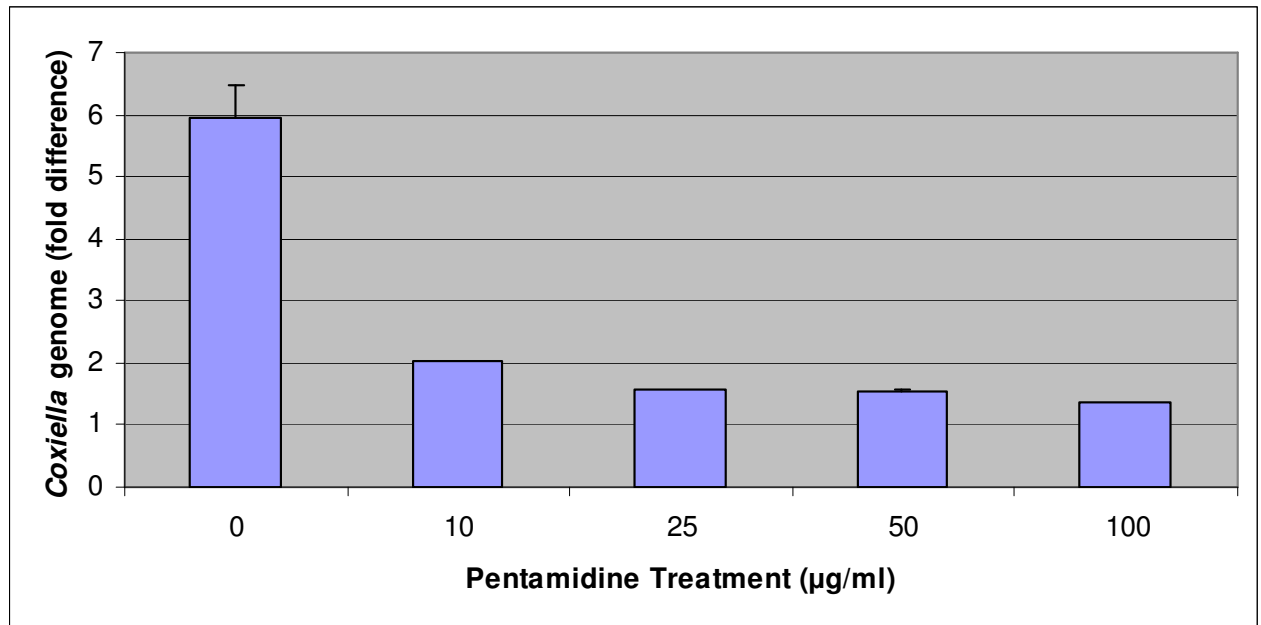


FIG. 4-4. Q-PCR data showing pentamidine inhibition of *Coxiella* growth. Vero cells were infected with *C. burnetii* and grown for 72 h in the presence of 0, 10, 25, 50 or 100 µg/ml of pentamidine. Data represent the means of three experiments \pm SD.

Secondary structure of IVS. We confirmed the published IVS sequence (117) and used it to predict a secondary structure. The IVS is found inserted into helix 45 of the 23S rRNA of *C. burnetii*. Since the 5' and 3' termini have long stretches of complementary bases, the IVS forms a stem-loop structure projecting out from helix 45 (Fig. 5). The stem loop has a repeating oligomer sequence which is characteristic of nuclease recognition-sites. Since RNase III cleaves near the middle, the base of the stem loop would hold the 23S rRNA fragments in place, thereby avoiding any deleterious effect from fragmentation.

Phylogenetic analysis of CBU_2096. We performed phylogenetic analyses to gain some insight into the unknown protein encoded within the IVS. Neighbor-joining trees were built using MEGA4 (112, 122). In the reconstructed phylogeny (Fig. 6), CBU_2096 clustered with a very similar ORF present in some *Leptospira* species on an IVS, inserted at the same site as in *C. burnetii*. Since *Leptospira* and *C. burnetii* are not closely related, the high degree of amino acid sequence conservation observed is most likely explained by HGT.

D. DISCUSSION

The two group I introns in *C. burnetii* are found inserted in the 23S rRNA gene and their IGS sequences are complementary to highly conserved helices in bacterial 23S

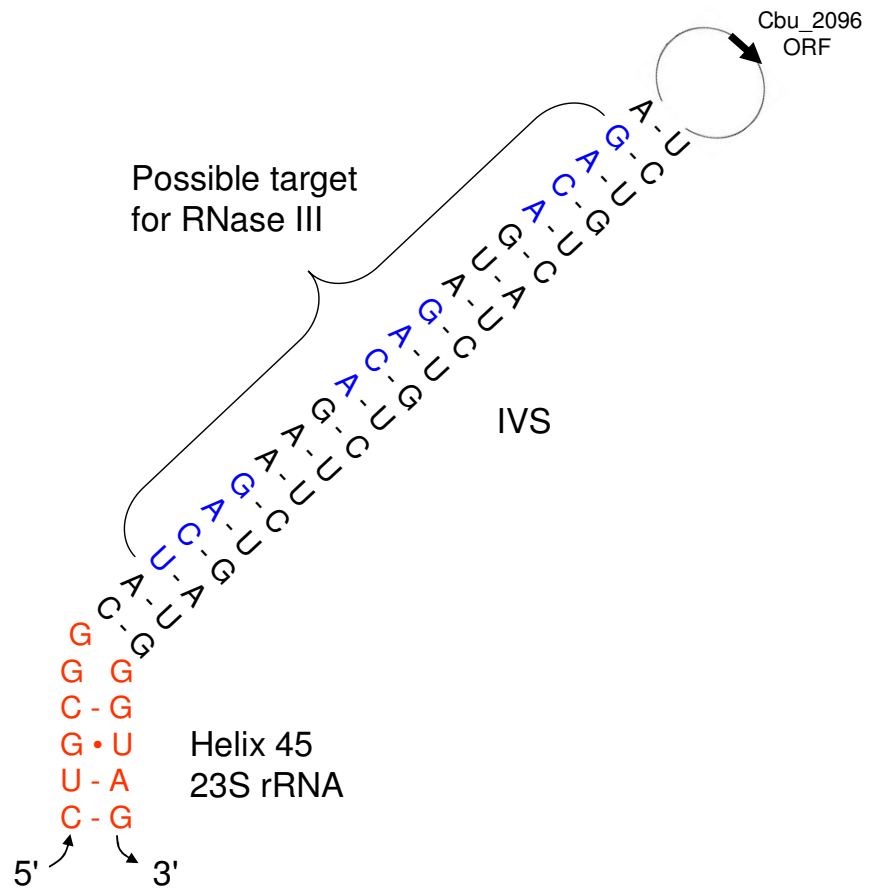


FIG. 4-5. Predicted secondary structure of IVS. IVS is found inserted into helix 45 (red) of *C. burnetii*'s 23S rRNA. Possible processing site for RNase III (blue) and the relative location of CBU_2096 are also depicted.

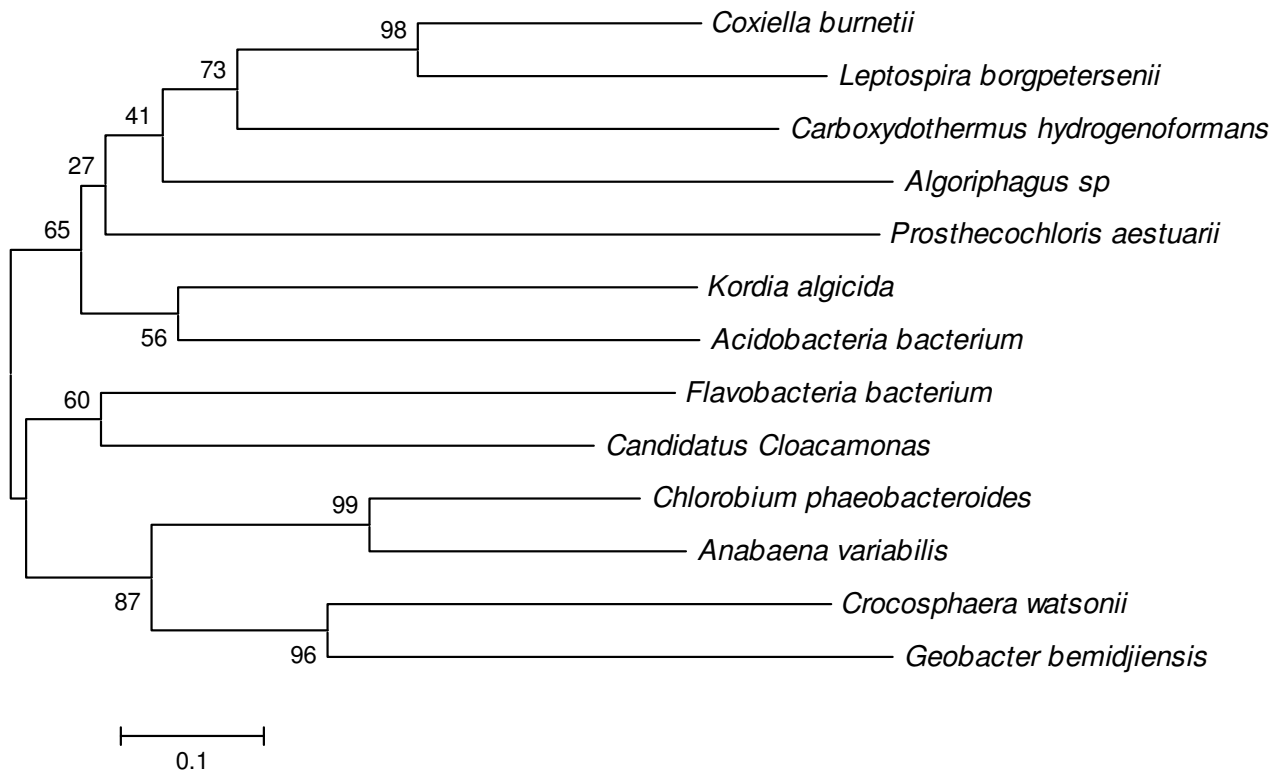


FIG. 4-6. Phylogenetic analysis of CBU_2096. Neighbor-Joining trees were constructed using amino acid sequences (106 positions) and Poisson correction method. Bootstrap values (1000 replicates) are indicated at the nodes.

rRNA. We previously observed that ribozymes can associate with both *C. burnetii* and *E. coli* ribosomes and hypothesized that the ribozyme-ribosome associations occurred via interactions between Cbu.L1917's IGS and helix 69 and Cbu.L1951's IGS and helix 71 (Fig. 7, Chapter 3). Indeed, a previous study had shown that a group I intron from *Tetrahymena*, when expressed in *E. coli*, associates with the ribosomes by means of its IGS and inhibits bacterial growth rate (97). Our data show that both introns inhibit ribosome function in vitro, via IGS. Binding of Cbu.L1917's IGS to helix 69 of 23S rRNA would likely obstruct the formation of the vital inter-subunit bridge B2a. Also, Cbu.L1917 would mask the contact sites between helix 69 and P- and A-site tRNAs. Cbu.L1951 bound to helix 71 would likely hinder formation of the inter-subunit bridge B3 and mask the contact site with P-site tRNA (138). These interferences could make the ribosomes unstable, thereby affecting their function.

To our surprise, our data also showed that both ribozymes retard *E. coli* growth rate even in the absence of IGSs. We also noted that, Cbu.L1917 but not Cbu.L1951 was able to inhibit ribosome function in vitro through an IGS-independent mechanism. Similar inhibitions of growth by oligonucleotides through both sequence-dependent and sequence-independent mechanisms have been observed in fungi, parasites and bacteria. For example, when DNA and RNA oligonucleotides that were mimics of, or complementary to, helix 69 were introduced into *Candida albicans*, it interfered with ribosome function and reduced cell growth. Interestingly, control oligos with no sequence similarity or complementarity to *Candida* rRNA also caused growth rate reduction through an unknown mechanism (31). In another study, antisense oligonucleotides that were complementary to immediate-early RNA of human cytomegalovirus were shown to

be able to inhibit virus replication. Again, control oligonucleotides without complementarity to any viral RNA were also shown to be able to cause growth inhibition (4). Oligonucleotides have also been shown to inhibit the growth of other viruses like HIV and Herpes simplex, bacteria like *M. tuberculosis* and *M. smegmatis*, and parasites like *Plasmodium falciparum* and *Toxoplasma gondii* through both sequence-dependent and independent manner (53). The underlying mechanisms that cause sequence-independent growth inhibition may vary. The intron RNAs may be inhibiting the function of proteins that play a role in ribosome assembly and function or may be interfering with tertiary interactions within the ribosomes (138). Further research is needed to better understand this phenomenon. Irrespective of their mode of action, small RNA and DNA sequences have the potential to be used as therapeutic agents to treat a number of human diseases. In fact, the U.S. Food and Drug Administration (FDA) has approved an oligonucleotide drug, Vitravene, for human use to treat cytomegalovirus retinitis (4).

By virtue of being inserted into the sole 23S rRNA gene, the two group I introns play an indirect but vital role in *C. burnetii* biology. In order to form functional ribosomes, it is imperative that the introns be removed from the pre-23S rRNA and the exons spliced together accurately. Any agent that blocks this process could potentially be used as a bactericidal agent to treat *C. burnetii* infections. Moreover, since group I introns are not found in humans, an agent that targets it should have less toxic side effects. One such agent is pentamidine, which acts by inhibiting rRNA group I intron self-splicing and is currently used in the treatment and prophylaxis of pneumonia caused by *Pneumocystis carinii* (59). Also, pentamidine has been shown to be able to inhibit the growth of *C. albicans* and *S. cereviceae* by inhibiting group I intron splicing (89, 139). We found that

Table 4-2. Plasmid constructs used in this study.

Designation	Plasmid Backbone	Insert
pUM1	pCR2.1-TOPO	Both Cbu.L1917 and Cbu.L1951
pUM2	pCR2.1-TOPO	Cbu.L1917
pUM3	pCR2.1-TOPO	Cbu.L1951
pUM4	pCR2.1-TOPO	IVS in reverse
pUM5	pCR2.1-TOPO	Cbu.DnaB+ Flanking sequences
pUM6	pQE-31	Cbu.DnaB+ Flanking sequences
pUM7	pCR2.1-TOPO	Cbu.L1917 without IGS
pUM8	pCR2.1-TOPO	Cbu.L1951 without IGS
pUM9	pCR2.1-TOPO	Cbu.L1917's IGS + IVS in reverse
pUM10	pCR2.1-TOPO	Cbu.L1951's IGS + IVS in reverse
pUM11	pQE-30	HE
pUM12	pQE-30	Cbu.DnaB+ Flanking sequences (out-of-frame)

pentamidine severely inhibits *C. burnetii* growth in Vero cell cultures showing its potential to be used to treat chronic Q fever, where treatment options are limited. Although more animal and human studies are needed before pentamidine can be used to treat human Q fever, the fact that it has been approved by FDA to treat *P. carinii* infections in humans would make it easier to employ the drug for this purpose.

The IVS is found inserted into helix 45 of *C. burnetii*'s 23S rRNA. Since the complementary ends of the IVS help form a helix, the element is thought to be removed by RNase III. These enzymes are a collection of endoribonucleases that cleaves double-stranded RNA (dsRNA) (32). The simplest members of this family are found in bacteria and play an important role in processing rRNA precursors (5). They also regulate translation by cleaving target mRNAs (108), (6). Typical substrates for these endoribonucleases are RNAs that have complementary segments that fold back and form dsRNA regions (137). Even though a consensus sequence has not yet been determined, recognition elements on the target dsRNA are thought to direct RNase III to the precise sites to be cleaved (101). The ORF found within the IVS codes for a hypothetical protein that is conserved between IVSs in a wide variety of bacteria. Our phylogenetic analyses revealed that a similar ORF is present on an IVS inserted at the same site in many *Leptospira* species, suggesting HGT was the mode of spread of this genetic element. Although the nucleotide sequences of the ORFs are not similar, the amino acid sequences of the encoded proteins are highly conserved between *Coxiella* and *Leptospira* suggesting an important function for the protein, possibly in IVS mobility. In our hands, the protein was found to be very toxic to *E. coli* such that we could only recover clones that did not have the insert in the proper orientation (we tested more than 50 clones from three

separate cloning experiments). If this protein aids in the horizontal transfer of IVS, it is conceivable that it cleaves the target site (helix 45) for IVS insertion akin to DNA cleavage caused by HEs found in introns and inteins. This ribosomal RNA cleavage might be the cause for the observed toxicity in *E. coli*, similar to the toxicity that has been reported by others and observed by us while trying to clone and express HE in *E. coli* (41, 85). Curiously, the IVS-encoded proteins analyzed in this study seem to fall into two broad groups (Fig. 6). Further study is needed to understand the functions of the IVS-encoded proteins.

CHAPTER FIVE

GENERAL DISCUSSION

Parasitic genetic elements are a fascinating phenomenon that transcends our conventional definitions of life-forms. They move from one organism to another horizontally, exploiting systems of genetic exchange and ensure their survival through vertical transmission to the next generation. Since they are thought not to contribute towards the reproductive success of the organism in which they reside, their evolutionary path is assumed to be independent of the host. The presence of two introns, an IVS and an intein in addition to 29 ISs in *C. burnetii*'s genome, which is undergoing reductive evolution (117), is intriguing. It seemed probable that the elements that reside in *C. burnetii*, with definite physiological and metabolic costs, might be imparting some influence on the host's biology. The overarching goal of this study was to understand the impact of parasitic genetic elements on *C. burnetii*'s biology, while characterizing in detail various properties of each element. In the course of this study, we found that the evolutionary history and biology of both the host and the genetic elements were intimately intertwined.

This study was inspired by the report by Seshadri et al. showing that the relatively small genome of *C. burnetii* - an obligate intracellular bacterium with no known avenues for HGT – contained a large number of parasitic genetic elements (117). We started by

analyzing the reported group I intron in the 23S rRNA gene. Using the published sequence as a reference, we amplified and cloned the intron and its proximal flanking regions. While studying the splicing properties of this intron, we discovered to our surprise and contrary to the original report, that actually two group I introns interrupt the 23S rRNA gene of *C. burnetii*. Significantly, this was the first time that a bacterial gene was shown to contain multiple introns (104). One intron was found to be inserted between bases 1917 and 1918 (*E. coli* numbering) and the other between bases 1951 and 1952 of the 23S rRNA gene. Based on conventional nomenclature guidelines, we designated each intron as Cbu.L1917 and Cbu.L1951, respectively. Cbu.L1917 is unique among all known group I introns due to the absence of the terminal guanosine (called Ω G). This is the first time a group I intron has been found in nature with terminal adenosine (A) in place of Ω G. This is significant because Ω G plays an important role in intron self-splicing, especially in determining the 3' splice junction. The mechanism of splicing of this novel intron is not yet known, but, in the laboratory, group I introns with Ω G mutated to A have been found to self-splice using A or G (present in high concentrations) as cofactors (11). Cbu.L1917 is also smaller than Cbu.L1951 due to the absence of HEG. Since both non-splicing and self-splicing introns have been reported in bacteria (36), (95), we analyzed the splicing properties of each intron. We found them to be able to self-splice in vitro without the aid of any proteins, and even though it is not known whether they utilize accessory proteins in vivo, they were also found to be able to splice out of the pre-23S rRNA in the cell. We determined the predicted secondary structures of both introns and found them to contain stem loops made by complementary sequences, which is a common feature of all group I introns. Based on their secondary

structures, we were able to classify each intron to its specific subgroup: Cbu.L1917 belongs to subgroup IB2 whereas Cbu.L1951 is a subgroup IA3 intron. Since introns are parasitic elements that have invaded the genome of *C. burnetii* or one of its ancestors, we next performed phylogenetic analyses to determine their evolutionary histories. Although sufficient data are not available to specifically identify the donors of each intron, we were able to determine that the introns moved horizontally from 23S rRNA genes of subcellular organelles like mitochondria or from other environmental bacteria like cyanobacteria. Since introns are thought to be evolutionarily neutral, we expected to find them in various stages of degeneration in disparate populations of *C. burnetii* as observed in various subpopulations of red and green algae (54, 94). However, our analysis of representative strains of *C. burnetii* belonging to all eight genomic groups and isolated from different hosts and geographic regions show that the introns are highly conserved, suggesting a possible role for the introns in *Coxiella*'s biology.

Since there are no genetic systems available to manipulate *Coxiella*, we used an *E. coli* model to deduce possible intron functions. We observed the intron RNAs are able to associate with *E. coli* ribosomes and interfere with protein synthesis resulting in retarded bacterial growth. We analyzed *C. burnetii*'s ribosomes and found the introns to be associated with them too, suggesting a role for the introns in *C. burnetii*'s slow growth rate. In addition, we noticed that the intron RNA amounts were inversely correlated with *C. burnetii* genome quantities, further implying a negative effect for the introns on bacterial growth. Further study is needed to understand the exact role played by the introns in the slow growth rate of *Coxiella* and whether this aids latency in hosts and promotes chronic diseases like endocarditis. Our study is one of the very few that have

shown that the introns have an impact on its host. Conversely, host's biology and lifestyle can affect the intron life cycle as well. Once fixed in a population, introns degenerate and are ultimately lost from that population, only to reappear through a HGT event. However, in an obligate intracellular bacteria like *Coxiella*, with few opportunities for HGT, introns would permanently be lost unless they are positively selected. Our data suggest that the introns are retained in *C. burnetii*, possibly because they foster slow growth and thereby persistence. In any case, the presence of introns within the sole 23S rRNA of *C. burnetii* makes them potential targets for antibiotic therapy. We tested and found pentamidine, a known inhibitor of group I intron self-splicing, to be able to inhibit *C. burnetii* growth in vitro.

The 23S rRNA gene of *C. burnetii* also contains an IVS upstream of the introns. This parasitic element also moves from one organism to another by HGT but the mechanism is not yet understood (37). The conserved ORF found within the IVS may be aiding in the horizontal transfer of this element. We found the protein encoded by this ORF to be toxic to *E. coli* presumably because it cleaves the 23S rDNA to insert the IVS in a fashion similar to the HEs found on introns and inteins (13). Our phylogenetic analyses revealed that orthologous proteins are encoded on similar IVSs inserted at the same site in a number of bacteria found in a wide variety of habitats. Unlike introns, the flanking regions are not spliced together when the IVS is cleaved, resulting in a mature 23S rRNA in two pieces. The negative consequence of having a fragmented large subunit rRNA is thought to be minimal, but some positive attributes have been suggested. In *Salmonella*, the fragmented 23S rRNA has been shown to improve the efficiency of ribosome turnover as the bacteria shifts from log phase to stationary phase (58). It is possible that

the dramatic drop in ribosome amounts that occur when *Coxiella* transitions from LCV to SCV is aided by the RNA fragmentation caused by the IVS. Again, due to the lack of a genetic system for *Coxiella*, we are unable to test this possibility at this time.

C. burnetii also contains a third kind of parasitic genetic element, an intein. This proteinaceous genetic parasite is found inserted in the C-terminal region of the replicative DNA helicase gene (*dnaB*) of *Coxiella* and is designated Cbu.DnaB. As in other host-associated bacteria, *C. burnetii*'s genome is undergoing reductive evolution by converting functional genes into pseudogenes and ultimately losing them due to lack of selection (47). Since a non-functional intein could convert *dnaB* into a pseudogene, we investigated the functionality of Cbu.DnaB. We found the intein to be able to cleave out, leaving the exteins spliced together and the host protein presumably functional, thus having no apparent negative impact on *C. burnetii*'s biology. However, it is possible that the time required to splice out the intein before functional DnaB can be formed, might cause a lag in *C. burnetii*'s replicative process. Also, the intein might be responding to some unknown environmental cues, like favorable pH or temperature, to make available functional DnaB so that bacterial replication can take place.

Incredibly, some of the genetic parasites are themselves parasitized by another mobile genetic element, HE. These are endoribonucleases that have invaded genetic elements like introns and inteins. They help in the mobility of the host genetic element through a process called homing. In *C. burnetii*, Cbu.L1951 contains an HE that belongs to the LAGLIDADG family. Both Cbu.L1917 and Cbu.DnaB do not encode a HE but their closest relatives found in other bacteria contain intact HEs suggesting that their HEs were lost subsequent to the invasion of these elements into *C. burnetii*. To prevent being lost

from a population, some HEs have evolved a maturase function whereby the intron needs the protein encoded by HEG to splice accurately (80). It is possible that the HE present in *C. burnetii* aids in the proper splicing of Cbu.L1951 and maybe even the splicing of Cbu.L1917 (altruistic parasite). To study the endonuclease and possibly the maturase action of HE (CBU_0182), we have cloned it into an expression vector (pQE30) to produce pUM11. During this experiment, we noticed that the HE was very toxic to *E. coli* possibly due to its ability to cleave 23S rRNA. Currently, we are in the process of purifying the HE protein to study whether its endonuclease function is intact and to check whether it has evolved a maturase function.

In conclusion, we have characterized four intriguing parasitic genetic elements in *C. burnetii* and shown that they share an intimate relationship with their host. Our study has resulted in many significant findings, which we hope will pave way for a better understanding of this enigmatic pathogen and promote better appreciation for the biological roles played by parasitic genetic elements.

REFERENCES FOR CHAPTERS ONE, FOUR AND FIVE

1. **Aagaard, C., J. Z. Dalgaard, and R. A. Garrett.** 1995. Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron-cells of *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci U S A* **92**:12285-9.
2. **Afseth, G., and L. P. Mallavia.** 1997. Copy number of the 16S rRNA gene in *Coxiella burnetii*. *Eur J Epidemiol* **13**:729-31.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-402.
4. **Anderson, K. P., M. C. Fox, V. Brown-Driver, M. J. Martin, and R. F. Azad.** 1996. Inhibition of human cytomegalovirus immediate-early gene expression by an antisense oligonucleotide complementary to immediate-early RNA. *Antimicrob Agents Chemother* **40**:2004-11.
5. **Apirion, D., and A. Miczak.** 1993. RNA processing in prokaryotic cells. *Bioessays* **15**:113-20.
6. **Aristarkhov, A., A. Mikulskis, J. G. Belasco, and E. C. Lin.** 1996. Translation of the adhE transcript to produce ethanol dehydrogenase requires RNase III cleavage in *Escherichia coli*. *J Bacteriol* **178**:4327-32.
7. **Avise, J. C.** 2001. Evolving genomic metaphors: a new look at the language of DNA. *Science* **294**:86-7.
8. **Babudieri, B.** 1959. Q fever: a zoonosis. *Adv. Vet. Sci.* **5**:81.
9. **Beare, P. A., S. F. Porcella, R. Seshadri, J. E. Samuel, and R. A. Heinzen.** 2005. Preliminary assessment of genome differences between the reference Nine Mile isolate and two human endocarditis isolates of *Coxiella burnetii*. *Ann N Y Acad Sci* **1063**:64-7.
10. **Beare, P. A., J. E. Samuel, D. Howe, K. Virtaneva, S. F. Porcella, and R. A. Heinzen.** 2006. Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. *J Bacteriol* **188**:2309-24.
11. **Been, M. D., and A. T. Perrotta.** 1991. Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site. *Science* **252**:434-7.
12. **Belfort, M.** 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev* **17**:2860-3.
13. **Belfort, M., and R. J. Roberts.** 1997. Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**:3379-88.
14. **Bellamy, R. J., and A. R. Freedman.** 2001. Bioterrorism. *QJM* **94**:227-34.
15. **Burgin, A. B., K. Parodos, D. J. Lane, and N. R. Pace.** 1990. The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* **60**:405-14.
16. **Burt, A., and Trivers, R.** 2005. *Genes in conflict: the biology of selfish genetic elements.* Harvard University Press, Cambridge, MA.
17. **Burt, A., and V. Koufopanou.** 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr Opin Genet Dev* **14**:609-15.

18. **Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, and K. H. Wolfe.** 2004. Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc Natl Acad Sci U S A* **101**:1632-7.
19. **Camacho, M. T., I. Outschoorn, E. Kovacova, and A. Tellez.** 2000. Distribution of immunoglobulin G (IgG) and A (IgA) subclasses following Q fever vaccination with soluble phase I *Coxiella burnetii* extract. *Vaccine* **18**:1773-7.
20. **Chevalier, B., M. Turmel, C. Lemieux, R. J. Monnat, Jr., and B. L. Stoddard.** 2003. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J Mol Biol* **329**:253-69.
21. **Chevalier, B. S., and B. L. Stoddard.** 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* **29**:3757-74.
22. **Coleman, S. A., E. R. Fischer, D. Howe, D. J. Mead, and R. A. Heinzen.** 2004. Temporal analysis of *Coxiella burnetii* morphological differentiation. *J Bacteriol* **186**:7344-52.
23. **Collins, C. H., Y. Yokobayashi, D. Umeno, and F. H. Arnold.** 2003. Engineering proteins that bind, move, make and break DNA. *Curr Opin Biotechnol* **14**:665.
24. **Cooper, A. A., Y. J. Chen, M. A. Lindorfer, and T. H. Stevens.** 1993. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J* **12**:2575-83.
25. **Cooper, A. A., and T. H. Stevens.** 1993. Protein splicing: excision of intervening sequences at the protein level. *Bioessays* **15**:667-74.
26. **Cutler, S. J., M. Bouzid, and R. R. Cutler.** 2007. Q fever. *J Infect* **54**:313-8.
27. **Dalgaard, J. Z., A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, and I. S. Mian.** 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res* **25**:4626-38.
28. **Davis, G., Cox, H.** 1938. A filter-passing infectious agent isolated from ticks. I. Isolation from *Dermacentor andersoni*, reactions in animals, and filtration experiments. *Public Health Rep* **53**:2259-67.
29. **Derrick, E. H.** 1937. Q fever, a new fever entity: clinical features, diagnosis and laboratory investigation. *Med J Aust* **2**:281.
30. **Derrick, E. H.** 1939. *Rickettsia burnetii*: the cause of "Q" fever. *Med J Aust* **1**:14.
31. **Disney, M. D., C. G. Haidaris, and D. H. Turner.** 2003. Uptake and antifungal activity of oligonucleotides in *Candida albicans*. *Proc Natl Acad Sci U S A* **100**:1530-4.
32. **Dridger, D., and C. Condon.** 2004. The continuing story of endoribonuclease III. *J Mol Microbiol Biotechnol* **8**:195-200.
33. **Dumler, S. J.** 2002. Q fever. *Curr Treat Options Infect Dis* **4**:437-445.
34. **Eklund, C. M., R.R. Parker, and D.B. Lackman.** 1947. A case of Q fever probably contracted by exposure to ticks in nature. *Public Health Rep* **62**:1413-1416.
35. **Evans, T. C., Jr., J. Benner, and M. Q. Xu.** 1998. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**:2256-64.

36. **Everett, K. D., S. Kahane, R. M. Bush, and M. G. Friedman.** 1999. An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* **181**:4734-40.
37. **Evguenieva-Hackenberg, E.** 2005. Bacterial ribosomal RNA in pieces. *Mol Microbiol* **57**:318-25.
38. **Evguenieva-Hackenberg, E., and G. Klug.** 2000. RNase III processing of intervening sequences found in helix 9 of 23S rRNA in the alpha subclass of *Proteobacteria*. *J Bacteriol* **182**:4719-29.
39. **Field, P. R., J. L. Mitchell, A. Santiago, D. J. Dickeson, S. W. Chan, D. W. Ho, A. M. Murphy, A. J. Cuzzubbo, and P. L. Devine.** 2000. Comparison of a commercial enzyme-linked immunosorbent assay with immunofluorescence and complement fixation tests for detection of *Coxiella burnetii* (Q fever) immunoglobulin M. *J Clin Microbiol* **38**:1645-7.
40. **Fournier, P. E., T. J. Marrie, and D. Raoult.** 1998. Diagnosis of Q fever. *J Clin Microbiol* **36**:1823-34.
41. **Gauthier, A., M. Turmel, and C. Lemieux.** 1991. A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron. *Curr Genet* **19**:43-7.
42. **Genig, V. A., E.N. Knyazeva, P.S. Chelnikov, and M.M. Miroshnichenko.** 1965. Experience on the mass immunization of humans with the M-44 vaccine against Q fever. I. Subcutaneous method of immunization. *Vopr. Virusol* **10**:319-323.
43. **Gilbert, W.** 1986. Origin of life: the RNA world. *Nature* **319**:618.
44. **Glazunova, O., V. Roux, O. Freylikman, Z. Sekeyova, G. Fournous, J. Tyczka, N. Tokarevich, E. Kovacava, T. J. Marrie, and D. Raoult.** 2005. *Coxiella burnetii* genotyping. *Emerg Infect Dis* **11**:1211-7.
45. **Goddard, M. R., and A. Burt.** 1999. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A* **96**:13880-5.
46. **Gogarten, J. P., A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, and E. Hilario.** 2002. Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**:263-87.
47. **Gomez-Valero, L., A. Latorre, and F. J. Silva.** 2004. The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. *Mol Biol Evol* **21**:2172-81.
48. **Gong, W. J., and K. G. Golic.** 2003. Ends-out, or replacement, gene targeting in *Drosophila*. *Proc Natl Acad Sci U S A* **100**:2556-61.
49. **Goodrich-Blair, H., and D. A. Shub.** 1996. Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* **84**:211-21.
50. **Hackstadt, T.** 1990. The role of lipopolysaccharides in the virulence of *Coxiella burnetii*. *Ann N Y Acad Sci* **590**:27-32.
51. **Hackstadt, T., M. G. Peacock, P. J. Hitchcock, and R. L. Cole.** 1985. Lipopolysaccharide variation in *Coxiella burnetii*: intrastain heterogeneity in structure and antigenicity. *Infect Immun* **48**:359-65.

52. **Hall, T. M., J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy, and D. J. Leahy.** 1997. Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* **91**:85-97.
53. **Harth, G., P. C. Zamecnik, J. Y. Tang, D. Tabatadze, and M. A. Horwitz.** 2000. Treatment of *Mycobacterium tuberculosis* with antisense oligonucleotides to glutamine synthetase mRNA inhibits glutamine synthetase activity, formation of the poly-L-glutamate/glutamine cell wall structure, and bacterial replication. *Proc Natl Acad Sci U S A* **97**:418-23.
54. **Haugen, P., V. A. Huss, H. Nielsen, and S. Johansen.** 1999. Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing-endonuclease pseudogenes in the *Bangiophyceae*. *Curr Genet* **36**:345-53.
55. **Haugen, P., D. M. Simon, and D. Bhattacharya.** 2005. The natural history of group I introns. *Trends Genet* **21**:111-9.
56. **Hendrix, L. R., J. E. Samuel, and L. P. Mallavia.** 1991. Differentiation of *Coxiella burnetii* isolates by analysis of restriction-endonuclease-digested DNA separated by SDS-PAGE. *J Gen Microbiol* **137**:269-76.
57. **Hilbink, F., M. Penrose, E. Kovacova, and J. Kazar.** 1993. Q fever is absent from New Zealand. *Int J Epidemiol* **22**:945-9.
58. **Hsu, D., L. M. Shih, and Y. C. Zee.** 1994. Degradation of rRNA in *Salmonella* strains: a novel mechanism to regulate the concentrations of rRNA and ribosomes. *J Bacteriol* **176**:4761-5.
59. **Hughes, W. T.** 1996. Recent advances in the prevention of *Pneumocystis carinii* pneumonia. *Adv Pediatr Infect Dis* **11**:163-86.
60. **Hurst, G. D., and J. H. Werren.** 2001. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* **2**:597-606.
61. **Ikawa, Y., H. Shiraishi, and T. Inoue.** 2000. Minimal catalytic domain of a group I self-splicing intron RNA. *Nat Struct Biol* **7**:1032-5.
62. **Jager, C., H. Willems, D. Thiele, and G. Baljer.** 1998. Molecular characterization of *Coxiella burnetii* isolates. *Epidemiol Infect* **120**:157-64.
63. **James, R., C. Kleanthous, and G. R. Moore.** 1996. The biology of E colicins: paradigms and paradoxes. *Microbiology* **142** (Pt 7):1569-80.
64. **Jasin, M.** 1996. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12**:224-8.
65. **Jasinskas, A., J. Zhong, and A. G. Barbour.** 2007. Highly prevalent *Coxiella* sp. bacterium in the tick vector *Amblyomma americanum*. *Appl Environ Microbiol* **73**:334-6.
66. **Jeffares, D. C., T. Mourier, and D. Penny.** 2006. The biology of intron gain and loss. *Trends Genet* **22**:16-22.
67. **Johansen, S., T. M. Embley, and N. P. Willassen.** 1993. A family of nuclear homing endonucleases. *Nucleic Acids Res* **21**:4405.
68. **Jurica, M. S., and B. L. Stoddard.** 1999. Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci* **55**:1304-26.
69. **Kane, P. M., C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebel, and T. H. Stevens.** 1990. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**:651-7.

70. **Kermode, M., K. Young, S. Hurley, B. Marmion** 2003. An economic evaluation of increased uptake in Q fever vaccination among meat and agricultural industry workers following implementation of the National Q Fever Management Program. *Aust N Z J Public Health* **27**:390-398.
71. **Kohler, U., B. G. Ayre, H. M. Goodman, and J. Haseloff.** 1999. Trans-splicing ribozymes for targeted gene delivery. *J Mol Biol* **285**:1935-50.
72. **Konkel, M. E., R. T. Marconi, D. J. Mead, and W. Cieplak, Jr.** 1994. Identification and characterization of an intervening sequence within the 23S ribosomal RNA genes of *Campylobacter jejuni*. *Mol Microbiol* **14**:235-41.
73. **Kowalski, J. C., M. Belfort, M. A. Stapleton, M. Holpert, J. T. Dansereau, S. Pietrokovski, S. M. Baxter, and V. Derbyshire.** 1999. Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res* **27**:2115-25.
74. **Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech.** 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**:147-57.
75. **Kuhlmann, U. C., G. R. Moore, R. James, C. Kleanthous, and A. M. Hemmings.** 1999. Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett* **463**:1-2.
76. **Lan, N., R. P. Howrey, S. W. Lee, C. A. Smith, and B. A. Sullenger.** 1998. Ribozyme-mediated repair of sickle beta-globin mRNAs in erythrocyte precursors. *Science* **280**:1593-6.
77. **Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chisoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al.** 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921.
78. **Liu, S. L., and K. E. Sanderson.** 1996. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc Natl Acad Sci U S A* **93**:10303-8.
79. **Logsdon, J. M., Jr.** 1998. The recent origins of spliceosomal introns revisited. *Curr Opin Genet Dev* **8**:637-48.

80. **Longo, A., C. W. Leonard, G. S. Bassi, D. Berndt, J. M. Krahn, T. M. Hall, and K. M. Weeks.** 2005. Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat Struct Mol Biol* **12**:779-87.
81. **Lucas, P., C. Otis, J. P. Mercier, M. Turmel, and C. Lemieux.** 2001. Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. *Nucleic Acids Res* **29**:960-9.
82. **Lykke-Andersen, J., C. Aagaard, M. Semionenkov, and R. A. Garrett.** 1997. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci* **22**:326-31.
83. **Madariaga, M. G., K. Rezai, G. M. Trenholme, and R. A. Weinstein.** 2003. Q fever: a biological weapon in your backyard. *Lancet Infect Dis* **3**:709-21.
84. **Marmion, B. P., M. Shannon, I. Maddocks, P. Storm, and I. Penttila.** 1996. Protracted debility and fatigue after acute Q fever. *Lancet* **347**:977-8.
85. **Marshall, P., and C. Lemieux.** 1991. Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast large subunit rRNA-encoding gene of *Chlamydomonas eugametos*. *Gene* **104**:241-5.
86. **Maurin, M., and D. Raoult.** 1999. Q fever. *Clin Microbiol Rev* **12**:518-53.
87. **McCaul, T. F., T. Hackstadt, and J. C. Williams.** 1981. Ultrastructural and biological aspects of *Coxiella burnetii* under physical disruptions. Academic Press Inc., New York, NY.
88. **McCaul, T. F., and J. C. Williams.** 1981. Developmental cycle of *Coxiella burnetii*: structure and morphogenesis of vegetative and sporogenic differentiations. *J Bacteriol* **147**:1063-76.
89. **Miletti, K. E., and M. J. Leibowitz.** 2000. Pentamidine inhibition of group I intron splicing in *Candida albicans* correlates with growth inhibition. *Antimicrob Agents Chemother* **44**:958-66.
90. **Miyake, T., H. Hiraishi, H. Sammoto, and B. Ono.** 2003. Involvement of the VDE homing endonuclease and rapamycin in regulation of the *Saccharomyces cerevisiae* GSH11 gene encoding the high affinity glutathione transporter. *J Biol Chem* **278**:39632-6.
91. **Mohanty, B. K., and S. R. Kushner.** 2002. Polyadenylation of *Escherichia coli* transcripts plays an integral role in regulating intracellular levels of polynucleotide phosphorylase and RNase E. *Mol Microbiol* **45**:1315-24.
92. **Moran, J. V., S. Zimmerly, R. Eskes, J. C. Kennell, A. M. Lambowitz, R. A. Butow, and P. S. Perlman.** 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol* **15**:2828-38.
93. **Moran, N. A., and G. R. Plague.** 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev* **14**:627-33.
94. **Muller, K. M., J. J. Cannone, R. R. Gutell, and R. G. Sheath.** 2001. A structural and phylogenetic analysis of the group IC1 introns in the order *Bangiales* (Rhodophyta). *Mol Biol Evol* **18**:1654-67.
95. **Nesbo, C. L., and W. F. Doolittle.** 2003. Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc Natl Acad Sci U S A* **100**:10806-11.

96. **Nielsen, H., T. Fiskaa, A. B. Birgisdottir, P. Haugen, C. Einvik, and S. Johansen.** 2003. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* **9**:1464-75.
97. **Nikolcheva, T., and S. A. Woodson.** 1997. Association of a group I intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA* **3**:1016-27.
98. **Otomo, T., K. Teruya, K. Uegaki, T. Yamazaki, and Y. Kyogoku.** 1999. Improved segmental isotope labeling of proteins and application to a larger protein. *J Biomol NMR* **14**:105-14.
99. **Parker, N. R., J. H. Barralet, and A. M. Bell.** 2006. Q fever. *Lancet* **367**:679-88.
100. **Perler, F. B.** 2002. InBase: the Intein Database. *Nucleic Acids Res* **30**:383-4.
101. **Pertzev, A. V., and A. W. Nicholson.** 2006. Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res* **34**:3708-21.
102. **Pietrokovski, S.** 2001. Intein spread and extinction in evolution. *Trends Genet* **17**:465-72.
103. **Pietrokovski, S.** 1998. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* **7**:64-71.
104. **Raghavan, R., S. R. Miller, L. D. Hicks, and M. F. Minnick.** 2007. The Unusual 23S rRNA Gene of *Coxiella burnetii*: Two Self-Splicing Group I Introns Flank a 34-Base-Pair Exon, and One Element Lacks the Canonical Ω G. *J Bacteriol* **189**:6572-9.
105. **Ralph, D., and M. McClelland.** 1993. Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. *Proc Natl Acad Sci U S A* **90**:6864-8.
106. **Raoult, D., P. Houpiqian, H. Tissot Dupont, J. M. Riss, J. Arditi-Djiane, and P. Brouqui.** 1999. Treatment of Q fever endocarditis: comparison of 2 regimens containing doxycycline and ofloxacin or hydroxychloroquine. *Arch Intern Med* **159**:167-73.
107. **Raoult, D., T. Marrie, and J. Mege.** 2005. Natural history and pathophysiology of Q fever. *Lancet Infect Dis* **5**:219-26.
108. **Regnier, P., and M. Grunberg-Manago.** 1989. Cleavage by RNase III in the transcripts of the met Y-nus-A-infB operon of *Escherichia coli* releases the tRNA and initiates the decay of the downstream mRNA. *J Mol Biol* **210**:293-302.
109. **Roman, J., and S. A. Woodson.** 1998. Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing *in vivo*. *Proc Natl Acad Sci U S A* **95**:2134-9.
110. **Rotz, L. D., A. S. Khan, S. R. Lillibridge, S. M. Ostroff, and J. M. Hughes.** 2002. Public health assessment of potential biological terrorism agents. *Emerg Infect Dis* **8**:225-30.
111. **Roux, V., M. Bergoin, N. Lamaze, and D. Raoult.** 1997. Reassessment of the taxonomic position of *Rickettsiella grylli*. *Int J Syst Bacteriol* **47**:1255-7.
112. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406-25.

113. **Samuel, J. E., K. Kiss, and S. Varghees.** 2003. Molecular pathogenesis of *Coxiella burnetii* in a genomics era. *Ann N Y Acad Sci* **990**:653-63.
114. **Savinelli, E. A., and L. P. Mallavia.** 1990. Comparison of *Coxiella burnetii* plasmids to homologous chromosomal sequences present in a plasmidless endocarditis-causing isolate. *Ann N Y Acad Sci* **590**:523-33.
115. **Schafer, B., B. Wilde, D. R. Massardo, F. Manna, L. Del Giudice, and K. Wolf.** 1994. A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Curr Genet* **25**:336-41.
116. **Seshadri, R., L. R. Hendrix, and J. E. Samuel.** 1999. Differential expression of translational elements by life cycle variants of *Coxiella burnetii*. *Infect Immun* **67**:6026-33.
117. **Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg.** 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci U S A* **100**:5455-60.
118. **Seshadri, R., and J. Samuel.** 2005. Genome analysis of *Coxiella burnetii* species: insights into pathogenesis and evolution and implications for biodefense. *Ann N Y Acad Sci* **1063**:442-50.
119. **Shub, D. A., H. Goodrich-Blair, and S. R. Eddy.** 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem Sci* **19**:402-4.
120. **Stahley, M. R., and S. A. Strobel.** 2006. RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol* **16**:319-26.
121. **Strobel, S. A., and J. C. Cochrane.** 2007. RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr Opin Chem Biol* **11**:636-43.
122. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**:1596-9.
123. **Tigertt, W. D., A. S. Benenson, and W. S. Gochenour.** 1961. Airborne Q fever. *Bacteriol Rev* **25**:285-93.
124. **Toman, R., and L. Skultety.** 1996. Structural study on a lipopolysaccharide from *Coxiella burnetii* strain Nine Mile in avirulent phase II. *Carbohydr Res* **283**:175-85.
125. **Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanagan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K.**

- Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al.** 2001. The sequence of the human genome. *Science* **291**:1304-51.
126. **Vicens, Q., and T. R. Cech.** 2006. Atomic level architecture of group I introns revealed. *Trends Biochem Sci* **31**:41-51.
127. **Voth, D. E., and R. A. Heinzen.** 2007. Lounging in a lysosome: the intracellular lifestyle of *Coxiella burnetii*. *Cell Microbiol* **9**:829-40.
128. **Waag, D. M.** 2007. *Coxiella burnetii*: host and bacterial responses to infection. *Vaccine* **25**:7288-95.
129. **Waag, D. M., M. J. England, R. F. Tammariello, W. R. Byrne, P. Gibbs, C. M. Banfield, and M. L. Pitt.** 2002. Comparative efficacy and immunogenicity of Q fever chloroform:methanol residue (CMR) and phase I cellular (Q-Vax) vaccines in *cynomolgus* monkeys challenged by aerosol. *Vaccine* **20**:2623-34.
130. **Weisburg, W. G., M. E. Dobson, J. E. Samuel, G. A. Dasch, L. P. Mallavia, O. Baca, L. Mandelco, J. E. Sechrest, E. Weiss, and C. R. Woese.** 1989. Phylogenetic diversity of the *Rickettsiae*. *J Bacteriol* **171**:4202-6.
131. **Weiss, E., and J.W. Moulder.** 1984. Genus III. *Coxiella*, p. 701-704. In N. R. Krieg, and J.G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 1. The Williams and Wilkins Co., Baltimore, MD.
132. **Wiebe, M. E., P. R. Burton, and D. M. Shankel.** 1972. Isolation and characterization of two cell types of *Coxiella burnetii* phase I. *J Bacteriol* **110**:368-77.
133. **Williams, J. C.** 1991. Infectivity, virulence, and pathogenicity of *Coxiella burnetii* for various hosts. In J. C. W. a. H. A. Thomson (ed.), *Q fever: the biology of Coxiella burnetii*. CRC Press, Boca Raton, FL.
134. **Winkler, M. E.** 1979. Ribosomal ribonucleic acid from *Salmonella typhimurium*: absence of the intact 23S species. *J Bacteriol* **139**:842-849.
135. **Wood, D. W., W. Wu, G. Belfort, V. Derbyshire, and M. Belfort.** 1999. A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* **17**:889-92.
136. **Woodson, S. A.** 2005. Structure and assembly of group I introns. *Curr Opin Struct Biol* **15**:324-30.
137. **Young, R. A., and J. A. Steitz.** 1978. Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proc Natl Acad Sci U S A* **75**:3593-7.
138. **Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller.** 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**:883-96.

139. **Zhang, Y., A. Bell, P. S. Perlman, and M. J. Leibowitz.** 2000. Pentamidine inhibits mitochondrial intron splicing and translation in *Saccharomyces cerevisiae*. *RNA* **6**:937-51.
140. **Zuckermandl, E., Pauling L.** 1965. Evolutionary divergence and convergence in proteins. *In* V. Bryson (ed.), *Evolving genes and proteins* Academic Press, New York.
141. **Dobzhansky, T.,** 1964. Biology, molecular and organismic. *Am. Zool.* **4**: 443-52.

APPENDIX: PLASMID VECTORS USED IN THIS STUDY

