Global Transcriptome Analysis of the Human Pathogens

Chlamydia trachomatis

and

Chlamydia pneumoniae

vorgelegt von
Diplom-Ingenieur
Marco Albrecht
aus Meerane

Von der Fakultät III - Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften

– Dr.-Ing. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Vera Meyer
Berichter: Prof. Dr. Roland Lauster
Berichter: Prof. Dr. Thomas Rudel
Berichter: Prof. Dr. Ulf Stahl

Tag der wissenschaftlichen Aussprache: 16.05.2011

Berlin 2011
D 83

CONTENT

1 Summary 3
1.1 Abstract 3
1.2 Zusammenfassung 5
2 Introduction 7
2.1 The Obligate Intracellular Bacterial Order Chlamydiales 7
2.1.1 The Chlamydial Life Cycle 8
2.1.2 Pathogenesis of *Chlamydia trachomatis* 10
2.1.3 Pathogenesis of *Chlamydia pneumoniae* 11
2.1.4 *Chlamydia* and Host Cell Apoptosis 11
2.1.5 The *Chlamydia* Genomes 12
2.1.6 Gene Regulation in *Chlamydia* 13
2.1.7 The Chlamydial Plasmids 15
2.2 Transcriptome Analysis by Differential RNA-Sequencing (dRNA-seq) 16
2.3 Bacterial Small Non-coding RNAs (sRNA) 17
2.3.1 Classes of Bacterial sRNAs 17
2.3.2 Mechanisms of RNA-Based Gene Regulation 18
2.3.3 Identification of Bacterial sRNAs 19
2.3.4 Target Identification of RNA-Based Gene Regulation 20
2.3.5 Small RNAs in *Chlamydia* 20
3 Aim of Study 22
4 Material 23
4.1 Bacterial Strains 23
4.2 Eukaryotic Cell Lines 23
4.3 Chemicals 23
4.4 Kits 24
4.5 Antibodies 24
4.6 Enzymes 24
4.7 Buffers and Solutions 24
4.8 Oligonucleotides 26
4.9 Lab Ware and Consumables 26
4.10 Equipment 26
4.11 Software 27
5 Methods 28
5.1 Cell Biology 28
5.1.1 General Handling of Human Cell Lines 28
5.2 Microbiology 28
5.2.1 Propagation and Purification of *Chlamydia trachomatis* 28
5.2.2 Propagation and Purification of *Chlamydia pneumoniae* 28
5.2.3 Infection and Estimation of Bacteria Stock Titre 29
5.2.4 Isolation of EB and RB of *Chlamydia trachomatis* 29
5.2.5 Isolation of EB and RB of *Chlamydia pneumoniae* 29
5.2.6 Electron Microscopy of Purified EB and RB 30
5.3 Biochemistry 30
5.3.1 Preparation of *C. pneumoniae* Lysates 30
5.3.2 Fractionation of Lysates on a Density Gradient 30

5.3.3 SDS-Polyacrylamide Gel Electrophoresis ... 31
5.3.4 Western Blot Analysis of Proteins ... 31
5.4 Molecular biology ... 32
5.4.1 Isolation of RNA from Purified Chlamydia ... 32
5.4.2 Polyacrylamide Gel Electrophoresis of RNA and Northern Blot ... 32
5.4.3 Next-Generation Sequencing of the *Chlamydia* Transcriptomes ... 33
5.5 Computational Biology ... 34
5.5.1 Processing of Sequencing Raw Data ... 34
5.5.2 Visualisation of Transcriptome Sequencing Data ... 34
5.5.3 Analysis of Transcriptional Start Sites and Read Numbers per Gene ... 34
5.5.4 Promoter Sequence Analysis ... 35
6 Results ... 36
6.1 Characterization of the *Chlamydia trachomatis* Transcriptome by dRNA-seq ... 36
6.1.1 Purification of *Chlamydia trachomatis* EB and RB ... 36
6.1.2 Deep Sequencing of Purified *Chlamydia trachomatis* EB and RB ... 36
6.1.3 Annotation of Transcriptional Start Sites (TSS) ... 39
6.1.4 Discovery of Putative Small Non-Coding RNAs of *Chlamydia trachomatis* ... 42
6.1.5 Differential Gene Expression in EB and RB ... 46
6.1.6 The Chlamydial Cryptic Plasmid Encodes Abundant sRNAs ... 48
6.2 Characterization of the *C. pneumoniae* Transcriptome by dRNA-seq ... 50
6.2.1 Purification of *C. pneumoniae* EB and RB ... 50
6.2.2 Deep Sequencing of Purified *C. pneumoniae* EB and RB ... 51
6.2.3 Annotation of Transcriptional Start Sites (TSS) ... 53
6.2.4 Discovery of Putative Small Non-Coding RNAs of *Chlamydia pneumoniae* ... 57
6.2.5 Analysis of Polycistronic Transcripts ... 63
6.2.6 Differentially Expressed Genes in *C. pneumoniae* EB and RB ... 65
6.2.7 Analysis of *C. pneumoniae* Promoters ... 67
7 Discussion ... 71
7.1 Transcriptome Analysis by dRNA-seq ... 71
7.2 Transcriptome Analysis of *Chlamydia trachomatis* ... 72
7.2.1 Comparison with Tiling Array Study ... 74
7.3 Transcriptome Analysis of *Chlamydia pneumoniae* ... 75
7.4 Comparison between *C. trachomatis* and *C. pneumoniae* Transcriptomes ... 77
7.5 Future Perspectives ... 80
8 References ... 81
9 Appendix ... 91
9.1 Detailed Lists of Transcriptional Start Sites (TSS) ... 91
9.1.1 TSS of *C. trachomatis* ... 91
9.1.2 TSS of *C. pneumoniae* ... 117
9.2 Abbreviations ... 151
9.3 List of Figures ... 153
9.4 List of Tables ... 154

1 SUMMARY

1.1 Abstract

Chlamydia are obligate intracellular pathogenic bacteria that share a unique biphasic life cycle. *Chlamydia trachomatis* is the most common bacterial cause of sexually transmitted infections and blinding trachoma. *Chlamydia pneumoniae* is responsible for a wide range of acute and chronic diseases, including respiratory infections and atherosclerosis. *Chlamydia* share a unique bi-phasic life cycle that switches from infectious but metabolically inactive elementary bodies (EB) to non-infectious but metabolically active reticular bodies (RB).

Since genetic tools to manipulate the genome and methods to culture the bacteria outside the host cell are lacking, genome sequence analysis has been the main approach to get insight into the biology of all *Chlamydiales*. Although the genomes of several strains have been sequenced yet, very little information is available on the gene structure of these bacteria. The genome sequences of *C. trachomatis* and *C. pneumoniae* are available since 1998 and 1999, respectively. Genome annotation and most information on transcript organisation are based on comparative computational approaches.

Massively parallel cDNA sequencing has revolutionized global transcriptome analysis in recent years. Here, a differential RNA-sequencing (dRNA-seq) approach is applied that uses enzymatic enrichment for primary transcripts to define the transcriptome of purified elementary bodies (EB) and reticulate bodies (RB) of the strains *C. trachomatis* L2b and *C. pneumoniae* CWL029, respectively. Using dRNA-seq technique, primary transcriptional start sites (TSS) of annotated genes and novel genes like small non-coding RNAs could be mapped in a thus far unprecedented resolution as a complement to the genome sequence.

By massive parallel sequencing a detailed map of the transcriptomes of two *Chlamydia* species was generated experimentally. For *C. trachomatis* 363 transcriptional start sites (TSS) of annotated genes could be determined. Semi-quantitative analysis of mapped cDNA reads revealed differences in the RNA levels of 84 genes isolated from EB and RB, respectively. Furthermore, 42 genome- and 1 plasmid-derived novel non-coding RNAs were discovered and in part validated by Northern blotting. The genome encoded non-coding RNA, ctrR0332, was found to be one of the most abundant and differentially expressed transcripts in EB and RB, implying an important role in the developmental cycle of *C. trachomatis*.

565 transcriptional start sites and 246 polycistronic transcripts could be determined for *C. pneumoniae*. Furthermore, numerous transcripts in intergenic regions and antisense to annotated ORFs which encode putative non-coding RNAs were identified and in part verified by Northern hybridisation. Most of them are *cis*- or *trans*-encoded non-coding RNAs. By experimental methods a distinct TSS or an affiliation to an operon could be identified for 861 out of 1074 genes (80%) of *C. pneumoniae*. Based on the transcriptome sequencing data the annotation of several genes could be

corrected. An analysis of differentially expressed genes identified more than 50 genes to be more abundant in EB, most of them with unknown function. Based on the experimentally verified TSS the promoter sequences of 531 genes were analysed for common motifs. This analysis revealed a standard sigma factor binding site whereas alternative sigma factor binding sites are not conserved among *Chlamydia*.

This detailed catalogue of TSS and operon structure will help to understand the organization, control and function of genes of these important pathogens. It is valuable information for further analysis of gene regulation. This is the first study that differentially analysed transcriptomes of purified EB and RB which will help to understand the regulation of the unique, yet unknown, conversion mechanisms between the two life cycle forms.

1.2 Zusammenfassung

Chlamydien sind obligat intrazelluläre humanpathogene Bakterien, welche einen einzigartigen zweiphasigen Lebenszyklus besitzen. *Chlamydia trachomatis* ist die häufigste Ursache für sexuell übertragbare Infektionen und Bindehautinfektionen die zur Erblindung führen. *Chlamydia pneumoniae* ist für ein breites Spektrum an akuten und chronischen Erkrankungen wie Atemwegsinfektionen und Atherosklerose verantwortlich. Chlamydien besitzen einen einzigartigen zweiphasigen Entwicklungszyklus, welcher zwischen infektiösen aber metabolisch inaktiven Elementarkörpern (EB) und nichtinfektiösen aber metabolisch aktiven Retikularkörpern (RB) alterniert.

Es gibt bis heute keine Möglichkeit zur genetischen Manipulation von Chlamydien und keine Methode um Chlamydien außerhalb der Wirtszelle zu kultivieren. Deshalb stellt die Genomanalyse die Hauptmethode dar, um Einblicke in die Biologie der *Chlamydiales* zu erhalten. Obwohl die Genome von mehreren Stämmen bereits sequenziert wurden, sind sehr wenige Informationen über die Genstruktur dieser Bakterien bekannt. Die Genomsequenzen von *C. trachomatis* und *C. pneumoniae* sind seit 1998 beziehungsweise 1999 bekannt. Die Annotation des Genoms und der Großteil der Informationen über die Organisation des Transkriptoms basiert auf vergleichenden computergestützten Analysen und Vorhersagen.

Durch die Einführung von Hochdurchsatz-Sequenziermethoden wurde die globale Transkriptomanalyse revolutioniert. In dieser Arbeit wurde eine Methode zur differentiellen RNA-Sequenzierung (dRNA-seq) angewendet, welche eine enzymatische Anreicherung von Primärtranskripten verwendet, um das primäre Transkriptom von aufgereinigten RB sowie EB der beiden Stämme *C. trachomatis* L2b und *C. pneumoniae* CWL029 zu bestimmen. Mit dieser dRNA-seq-Technik ist es möglich, primäre Transkriptionsstartstellen (TSS) von annotierten Genen zu bestimmen, sowie neue Gene, wie kleine nicht-codierende RNAs, zu entdecken. Diese Kartierung der TSS ist mit einer Auflösung bis zu einem Nukleotid möglich.

Durch Hochdurchsatzsequenzierung wurden experimentell detaillierte Transkriptomkarten der beiden *Chlamydien*-Spezies geschaffen. Für *C. trachomatis* konnten 363 TSS von annotierten Genen bestimmt werden. Durch semiquantitative Analyse der kartierten cDNA Sequenzen wurden 84 in EB und RB differentiell exprimierte Gene bestimmt. Weiterhin konnten 42 genom- und eine plasmid-kodierte kleine nicht-kodierende RNA entdeckt und teilweise experimentell durch Northern Blotting verifiziert werden. Die genomkodierte nichtkodierende RNA Ctr0332 war eine der am stärksten exprimierten und differentiell exprimierten Transkripte, was auf eine Rolle im Entwicklungszyklus von *C. trachomatis* hindeutet.

Bei der Analyse von *C. pneumoniae* wurden 565 TSS sowie 246 polycistronische Transkripte bestimmt. Außerdem wurde eine Vielzahl an Transkripten in intergenischen Regionen und antisense zu annotierten Genen entdeckt, die potentielle nichtkodierende RNAs darstellen. Diese wurden teilweise durch Northern Blotting verifiziert. Durch diesen experimentellen Ansatz der

Hochdurchsatzsequenzierung wurde für insgesamt 861 der 1074 annotierten Gene (entspricht 80%) von *C. pneumoniae* entweder eine primäre TSS oder eine Zugehörigkeit zu einem Operon identifiziert. Basierend auf diesem Transkriptom-Datensatz konnte die Annotation für eine Reihe von Genen korrigiert werden. Eine Analyse von differentiell regulierten Genen zeigte, dass über 50 Gene in EB angereichert sind, die meisten mit völlig unbekannter Funktion. Durch die experimentelle Bestimmung der TSS war es möglich, die Promotorsequenzen von 531 proteinkodierenden Genen zu analysieren. Diese Analyse zeigte, dass die Mehrheit der Gene durch den Standardsigmafaktor erkannt wird und es bei den untersuchten Chlamydien keine stark konservierten Promotormotive für alternative Sigmafaktoren gibt.

Dieser detaillierte Datensatz der primären TSS sowie der Operonstrukturen leistet einen fundamentalen Beitrag zum Verständnis der Genorganisation und -funktion dieser wichtigen Humanpathoge. Er bildet eine wertvolle Grundlage für zukünftige Untersuchungen der Genkontrolle. Dies ist die erste Arbeit, welche die differentiellen Transkriptome von aufgereinigten EB und RB untersucht hat. Die Ergebnisse können helfen den einzigartigen und bisher größtenteils unbekannten Konversionsmechanismus zwischen den beiden Entwicklungszyklusformen zu verstehen.

2 INTRODUCTION

2.1 The Obligate Intracellular Bacterial Order Chlamydiales

The bacterial order Chlamydiales comprises obligate intracellular Gram-negative pathogens that share a unique developmental cycle in which multiplication only occurs within a eukaryotic host cell (Moulder 1982). During the last decades numerous new organisms have been identified from different animal and environmental sources which led to the division of the family Chlamydiaceae into the two genera *Chlamydia* and *Chlamydophila* (Everett *et al.* 1999) shown in figure 1. This taxonomy is controversial and therefore this work refers to the use of a single genus, *Chlamydia*, as suggested by Schachter *et al.* (2001).

Bacteria of the family *Chlamydiaceae* are the etiological agents of numerous important human and animal diseases. The family *Chlamydiaceae* contains three human pathogenic species namely *Chlamydia trachomatis* (C. *trachomatis*), *Chlamydia pneumoniae (C. pneumoniae)*, and *Chlamydia psittaci*. *C. trachomatis* is the most prevalent cause of genital tract infection in the western world and the ocular serovars result in blinding trachoma in developmental countries. *C. pneumoniae* is the causing agent of community-acquired pneumonia, with millions of infections each year and chronic infections are associated with an enhanced risk of developing atherosclerosis (Kuo *et al.* 1993),

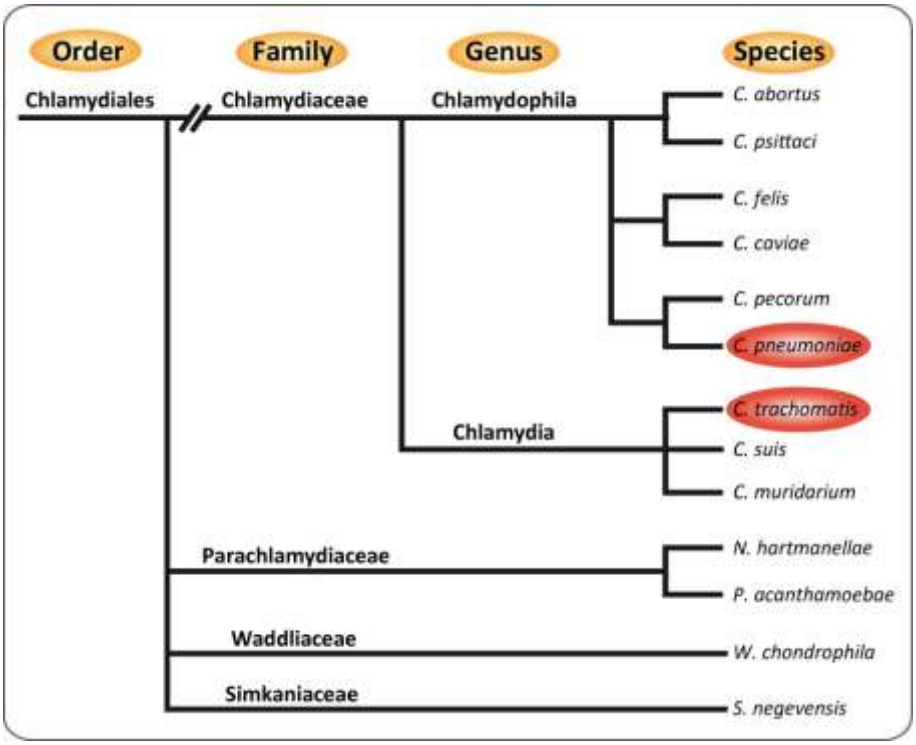

Figure 1: Taxonomy of the order Chlamydiales.
Phylogenetic tree of the order Chlamydiales based on ribosomal RNA homology. Species that have human as their major host are marked in red. The division of the family Chlamydiaceae into the two genera *Chlamydia* and *Chlamydophila* is controversially discussed in the community. This work refers to a single genus *Chlamydia.* Lengths of lines are not proportional to the real phylogenetical distances. Adapted from (Bush and Everett 2001).

cerebrovascular disease (Balin *et al.* 1998) and asthma (Hahn *et al.* 1991). *C. psittaci* can cause psittacosis in humans which is a zoonotic infectious disease. Transmission usually originates from contact to birds, especially poultry (Beeckman and Vanrompay 2009).

Although the genomes of several species are available for research for some years already, the mechanisms of pathogenicity are far from being understood (Subtil and Dautry-Varsat 2004). The major challenge when working with *Chlamydia* is the lack of an experimental genetic system and the unavailability of an *in vitro* culturing system outside a host cell. To date there is no transformation method available to maintain exogenous DNA inside *Chlamydia*. The consequence is the lack of any *in vivo* transcription assay.

2.1.1 The Chlamydial Life Cycle

The major feature of *Chlamydia* is their unique biphasic developmental cycle which alternates between two morphological forms, the elementary body (EB) and the reticulate body (RB) (Moulder 1991). EB are small, spore-like bacterial particles with a diameter of approximately 0.3 µm. They are metabolically inert, extracellular, infectious particles, able to attach to and invade susceptible cells. Their biological function is to survive extracellularly until a suitable host cell is found for infection. The bacterial nucleoid is highly condensed by bacterial histone-like proteins HctA and HctB (Brickman *et al.* 1993). EB contain significant quantities of mRNA and ribosomes, despite there is no transcriptional activity. EB have a unique outer membrane with little or no peptidoglycan which is cross-linked by inter- and intramolecular cysteine bonds and contains large amounts of outer membrane proteins such as OmpA, OmcA and OmcB. The cell wall of EB seems to contain components of a type three secretion system (Fields *et al.* 2003).

Chlamydia can infect numerous cell types and have developed several, possibly redundant, mechanisms for host cell attachment and entry. The exact interaction of bacterial adhesion and host cell receptor is still unclear. Several chlamydial membrane proteins such as OmpA, OmcB and the polymorphic outer membrane protein (Pmp) family have been suggested to play a role (Grimwood *et al.* 2001). The attachment of *Chlamydia* to the host cell seems to be a two-stage process. Initial attachment is reversible and occurs through electrostatic interactions between heparin sulphate containing glycosaminoglycan and OmpA (Stephens *et al.* 2001). The second step is irreversible and the exact mechanism is still unknown (Carabeo and Hackstadt 2001).

After adhesion (see figure 2, stage 1), a type three system secreted protein called Tarp (Translocated actin-recruiting phosphoprotein) is injected into the host cell. The protein is rapidly phosphorylated at tyrosine residues followed by actin recruitment (Clifton *et al.* 2004). The mechanism of internalization of the EB into the host is still unknown (Scidmore-Carlson and Hackstadt 2000). Inside the host cell the EB are located in a membrane bound vesicle called inclusion (Figure 2, stage 2). They become metabolically active, enlarge in size, and differentiate into RB that undergo binary fission (Figure 2, stages 3 and 5).

The reticulate bodies are larger than EB (1 µm in diameter) located inside the inclusion and bound by an inner and outer membrane like other Gram-negative bacteria. RB are non-infective and larger than EB due to their diffuse nucleic acid structure. RB divide by binary fission every 2-3 hours during the logarithmic growth phase and after a number of divisions the developmental process becomes asynchronous (McClarty 1994; Moulder 1991). Some RB start to differentiate back into EB (Figure 2, stage 5) followed by the release of the EB, which is accompanied by host cell death after 48 hours (*C. trachomatis*) to 96 hours (*C. pneumoniae*) depending on the species and growth conditions. Then a new infection cycle starts with EB attaching neighbouring cells or infecting another organism (Figure 2, stage 7).

The regulation of RB-to-EB conversion is largely unknown. Attempts to identify regulatory elements that switch gene regulation have failed so far. However, two histone-like proteins Hc1 and Hc2 have been identified that are expressed late in RB and are thought to initiate the conversion by

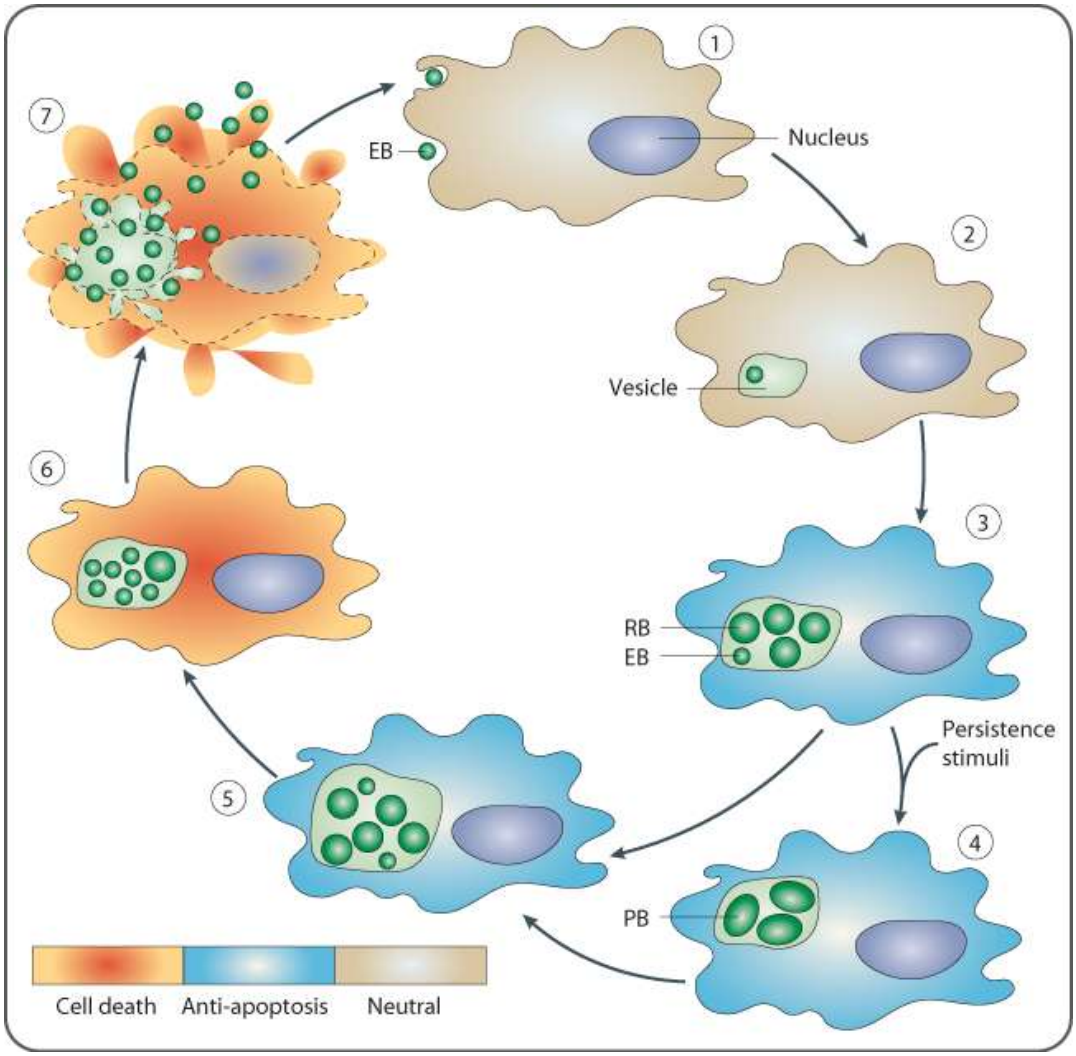

Figure 2: The life cycle of Chlamydiales.
(1) Elementary bodies (EB) attach to the host cell and invade the cell. (2) EB differentiate into reticulate bodies (RB) inside a vesicle called inclusion, followed by intracellular growth and division (3, 5). Persistence stimuli like cytokines or starvation of the host cell lead to the formation of persistence bodies (PB), which can enter the life cycle again after reactivation(5). Upon redifferentiation of the RB into EB, the lysis of the host cell is induced, and the EB escape from the host cell to start a new infection cycle. Figure adapted from Byrne and Ojcius (2004).

condensation of the DNA (Hackstadt *et al.* 1991; Perara *et al.* 1992). When expressed in *E. coli* Hc1 was shown to induce a compaction of chromatin into a spherical condensed nucleoid (Barry *et al.* 1992), whereas Hc2 expression leads to a more toroid nucleoid (Brickman *et al.* 1993). Hc2 has a higher affinity to RNA and represses translation more efficiently than Hc1 (Pedersen *et al.* 1996). Thus, these two histone-like proteins could act as global regulators by initiating transcriptional and translational silencing through chromatin condensation.

Chlamydia are associated with numerous chronic diseases that result from long lasting persistent infections which are defined by a long-term association between the bacteria and their host. In this stage the *Chlamydia* are in a viable but culture negative state. They are characterized by morphological aberrant enlarged RB (persistence bodies (PB), see figure 2, stage 4). *In vitro* persistence can be induced by several stimuli like antibiotics, amino acid starvation, depletion of iron, IFN-γ (reviewed in Hogan *et al.* (2004)) or co-infection with herpes viruses (Vanover *et al.* 2008).

2.1.2 Pathogenesis of *Chlamydia trachomatis*

C. trachomatis is the most common cause of sexually transmitted disease in humans. There are three biovariants of *C. trachomatis* that comprise 15 serovariants based on the major outer membrane protein OmpA. Serovars A-C cause trachoma, whereas serovars D-K and L1-L3 are associated with sexual transmitted diseases. Serovars D-K cause cervicitis in women and urogenital infections in men. Serovars L1, L2 and L3 cause lymphogranuloma venerum (LGV) and are more invasive than serovars A-K. They cause systemic infections, infect monocytes and disseminate to the local lymph nodes.

Most of the infected patients are asymptomatic and do not seek medical treatment. In women, this can cause pelvic inflammatory disease that can lead to infertility and ectopic pregnancies. Furthermore *C. trachomatis* causes trachoma which is the leading cause of preventable infectious blindness worldwide and is predominantly found in the poorest third world countries. In addition to these acute infections, *Chlamydia* are associated with a range of chronic diseases that are characterized by inflammation and scarring and lead to significant damage to the host.

The *C. trachomatis* strain used in this study is of serovar L2b. An outbreak of LGV among men who have sex with men (MSM) in the Netherlands in 2004 was caused by this serovariant which, for this reason, was also named the Amsterdam variant. All LGV cases reported in Amsterdam and France and a considerable percentage of infections in the UK and Germany were caused by this variant (Nieuwenhuis *et al.* 2004). It could be traced back to anal swabs isolated 1981 in San Francisco and was also found in Australia (Stark *et al.* 2007). LGV was considered a rare tropical disease in Germany and therefore not included in the national sentinel surveillance system of the Robert-Koch-Institute until the first outbreaks in Germany (v Krosigk *et al.* 2004). LGV is not a mandatorily notifiable disease in Germany since the new ‘Infektionsschutzgesetz’ (Protection Against Infection Act) introduced in 2001, and therefore no statistics of infection rates are available.

LGV is primarily an infection of lymphatics and lymph nodes caused by *C. trachomatis* serovars L1-L3 that gain entrance through skin breaks and abrasions or cross the epithelial cells of mucous membranes. The bacteria travel via the lymphatics and multiply within mononuclear phagocytes in regional lymph nodes. Pathophysiology and symptoms can be classified in three stages, whereas the first two stages can remain undetected. The primary stage is usually a genital ulcer, often unnoticed and transient. The secondary stage is characterized by inguinal adenitis, bubo formation and haemorrhagic proctitis, especially in MSM. The tertiary phase often occurs years after the initial infection. At this stage, an anogenitorectal syndrome may occur with resultant rectal strictures that are often irreversible. Furthermore, elephantiasis of the genitals is common (White 2009).

2.1.3 Pathogenesis of *Chlamydia pneumoniae*

The human pathogen *C. pneumoniae* is a major cause of pneumonia and is also associated with chronic diseases like atherosclerosis, asthma and Alzheimer's disease. Recent studies show that seropositivity of *C. pneumoniae* IgG antibodies correlates with increased lung cancer risk (Chaturvedi *et al.* 2010). *C. pneumoniae* was originally named TWAR agent (Taiwan Acute Respiratory Agent) from the clinical isolates TW-183 and AR-39 and is now considered a separate species of chlamydia (Everett *et al.* 1999). Both names, *Chlamydophila pneumoniae* and *Chlamydia pneumoniae*, are used in parallel in the literature since the classification is controversial (Schachter *et al.* 2001). This work uses the genus name *Chlamydia* to be consistent with the latest opinions of classification.

C. pneumoniae can cause a spectrum of infections that usually take a mild or sub-clinical course. It causes acute respiratory disease (Grayston *et al.* 1990) and accounts for 6-20% of community acquired pneumoniae cases in adults (Hammerschlag 2000). Almost all humans can expect to be infected with *C. pneumoniae* during their lifetime and infections can become chronic. Reinfections during the lifetime are common, leading to a seroprevalence of 80% in adults (Grayston 2000). After the primary infection the bacteria can persist in the host cells and represent a potential risk factor for chronic inflammatory lung disease (Laurila *et al.* 1997) and atherosclerosis (Campbell and Kuo 2004).

Currently, there is no vaccine available to prevent *C. pneumoniae* infection. However, acute infections can be treated with antibiotics like macrolides and doxycycline. Atypical persistent inclusions are resistant to antibiotics.

2.1.4 *Chlamydia* and Host Cell Apoptosis

Apoptosis is the major type of programmed cell death and, in contrast to necrosis, is not related to inflammatory processes. It plays an essential role in growth and maintenance of any multicellular organism including processes like embryogenesis, tissue homeostasis and immune regulation. It is a fundamental eukaryotic process that is highly evolutionary conserved. Two major pathways are involved in the controlled removal of cells. The extrinsic pathway is affected by ligands binding to cell surface receptors like CD95/Fas, which then initiate an intracellular cascade of

caspases. The activation and oligomerization of intracellular proteins induces the intrinsic pathway of apoptosis. Both pathways result in fragmentation of the cell. These fragments then can easily be phagocytosed by dendritic cells and macrophages, which results in a complete removal of the cell (Tait and Green 2010).

For most extracellular bacteria it has been shown that they are able to induce programmed cell death in order to invade host cell tissue and avoid immune defence. Induction of apoptosis is an energetically beneficial way to attack the host. This is, for example, practised by *Salmonella* (Knodler *et al.* 2005) and *Shigella spp.* (Zychlinsky *et al.* 1996), which efficiently kill macrophages by different mechanisms to escape the host's immune defence.

In contrast to this, several viral pathogens like herpes- and poxviridae have evolved or acquired genes that inhibit the host cell apoptosis machinery. This is advantageous for proliferation and persistence of the viral particles. Host cell caspases are a major point of attack as they are important mediators of apoptosis and consequently adequate targets for apoptosis inhibition. Viral FLIPs (FLICE [=caspase 8] like inhibitor proteins) are well studied proteins that inhibit apoptosis by preventing the formation of the death inducing signalling complex (DISC) (Thome *et al.* 1997). Homologs of vFlips have been identified in several other viruses and in mammalian cells (cFlips) (Irmler *et al.* 1997).

Chlamydia have a unique relationship to their host during productive infection and persistence. They are able to exploit host cell resources and manipulate the apoptosis regulation in different ways as a strategy for intracellular survival and propagation. There is a multitude of reports on *Chlamydia* and apoptosis which are partially conflicting. The emerging picture of this interaction is that anti-apoptotic activity is predominant at the early to mid-stages of the developmental cycle to ensure survival and productive growth, whereas in the later stages pro-apoptotic stimuli permit the release of elementary bodies (Byrne and Ojcius 2004; Miyairi and Byrne 2006).

Chlamydia infected HeLa cells are resistant to several apoptotic stimuli in the early to mid-phase of an acute infection (Fan *et al.* 1998). In microarray experiments it could be shown that anti-apoptotic genes are upregulated upon infection (Ren *et al.* 2003). Furthermore, pro-apoptotic proteins are degraded by *Chlamydia* by a yet unknown mechanism (Fischer *et al.* 2004).

2.1.5 The *Chlamydia* Genomes

Whole-genome sequencing of *Chlamydia* revolutionized the insights into the biology of this important pathogen. Due to the small genome size of only about 1 Mb *Chlamydia* was a good early candidate for genome sequencing. The *C. trachomatis* serovar D genome was released in 1998 (Stephens *et al.* 1998) as one of the first ten organisms with complete genomes available. Further sequenced chlamydial species and strains were *C. pneumoniae* CWL029 (Kalman *et al.* 1999), *C. muridarium* Nigg, *C. pneumoniae* AR039 (Read *et al.* 2000), *C. pneumoniae* J138 (Shirai *et al.* 2000),

and *C. trachomatis* L2b (Thomson *et al.* 2008). To date, more than 30 chlamydial genome sequences are available.

The genome of *C. trachomatis* L2b contains a single circular chromosome of 1,038,869 nt, has a GC-content of 41%, a coding capacity of 88% and is predicted to encode 934 genes, 874 of them for proteins. *C. pneumoniae* CWL029 harbours a circular chromosome of 1,230,230 nt (GC-content 40%, coding capacity 88%) that is predicted to contain 1122 genes, 1052 of them coding for proteins (Kalman *et al.* 1999). The *C. trachomatis* genome contains 70 genes without homologues in *C. pneumoniae*, 60 of them classified as encoding hypothetical proteins (Stephens *et al.* 1998). The ten remaining genes comprise two predicted thiol protease genes, the tryptophan operon (*trpA, B, R*), *trpC*, and four genes of the phospholipase-D superfamily. Orthologues were identified for 859 genes (80%) of the *C. pneumoniae* genes in the *C.trachomatis* genome. This demonstrates a high functional conservation between these two species. Similarity levels of orthologous genes lie in the range of 22-95% amino acid identity with an average of 62% (Stephens *et al.* 1998).

2.1.6 Gene Regulation in *Chlamydia*

The biphasic life cycle is unique to *Chlamydia* and needs a strong regulation since gene expression patterns vary enormously between the two distinct life cycle forms (Belland *et al.* 2003b; Maurer *et al.* 2007; Miura *et al.* 2008; Mukhopadhyay *et al.* 2006; Nicholson *et al.* 2003). Even though *Chlamydia* has a relatively small genome, remarkably little is known about gene regulation.

Regulation of gene expression is controlled at different levels in *Chlamydia*. DNA structure influences accessibility of genes, since in EB DNA is highly condensed. Promoter sequences and different sigma factors as well as transcription factors contribute to differential gene regulation. Post-transcriptional regulation is mainly influenced by mRNA stability and small non-coding RNA directing translation. Post translational modifications and protein stability comprise the last level of regulation.

2.1.6.1 *Chlamydia* RNA Polymerase

In eubacteria, transcription is initiated when the RNA polymerase holoenzyme ($\alpha_2\beta\beta'\sigma$) binds to its promoter upstream of the gene and forms an open complex. Subsequently the σ subunit is released and transcription starts. All known chlamydial RNA polymerases are highly conserved and consist of the core enzyme ($\alpha_2\ \beta\ \beta'$) encoded by *rpoA*, *rpoB* and *rpoC*, respectively. The holoenzyme includes one out of three sigma factors identified so far. These are the major sigma factor σ^{66} (homolog to the well-studied *E. coli* σ^{70}) and two alternative sigma factors σ^{28} and σ^{54} encoded by *rpoD*, *rpsD* and *rpoN*, respectively. Since sequence specificity of the RNA polymerase is controlled by the σ factor, it is likely that the different holoenzymes control different subsets of genes. The three σ factors are differentially expressed during the developmental cycle suggesting a role in stage-specific regulation (Mathews *et al.* 1999).

The RNA polymerase core enzyme genes and the major σ factor gene *rpoD* are expressed at relatively constant levels during the whole developmental cycle (Douglas and Hatch 2000). This is consistent with the expected function of regulating housekeeping genes. Promoter motifs have been identified biocomputationally based on homology to the σ^{70} family. Several targets like for example *ompA* and *omcB*, could be verified experimentally by antibody inhibition experiments (Mathews *et al.* 1993). The role of the two alternative σ factors is still unknown but some of the late genes expressed at the stage of RB-to-EB conversion seem to be directly regulated by σ^{28} (Fahr *et al.* 1995; Miura *et al.* 2008; Yu and Tan 2003).

The subregions that are known to bind the -35 and -10 promoter regions are strongly conserved between *E. coli* σ^{70} and the chlamydial σ^{66} factor (Lonetto *et al.* 1992). Based on the *E. coli* σ^{70} consensus sequence several chlamydial promoters could be identified and some experimentally verified (Fahr *et al.* 1992). Thereby conservation seems to be higher at the -10 region than in the -35 region. A mutational analysis of the rRNA P1 promoter defined the -10 box as TA(A/T)(A/T)AT, which is a close match to the *E. coli* σ^{70} consensus sequence TATAAT (Tan *et al.* 1998). Unexpectedly, point mutations in the -35 box did not have strong influences on promoter activity. Only a 5 bp substitution showed a 60% decrease of promoter activity (Tan *et al.* 1998). However, mutational analysis of the *dnaK* -35 box showed a 90% decrease in activity produced by a 5 bp substitution. A mutational analysis defined the optimal -35 promoter element as TTGACA, which is identical to the *E. coli* σ^{70} consensus sequence (Schaumburg and Tan 2003).

Only one promoter was identified that had a σ^{28} consensus sequence (Brickman *et al.* 1993). However, several promoter regions could be identified that did not resemble any known bacterial promoter consensus sequence. The basis for promoter identification is the identification of primary transcription start sites (TSS). This was done experimentally for only a few genes. Several genes, such as rRNA (Engel and Ganem 1987) and *ompA* (Stephens *et al.* 1988) were described to have more than one TSS, which raises the possibility of tandem promoters.

Concluding from the literature, there is only poor conservation between known chlamydial promoter sequences and also between RNA polymerase in *Chlamydia* and other bacteria. This was phrased “The RNA polymerase sigma subunit paradox” (Stephens 1990). A step to resolving this paradox is the identification of TSS to have a larger data set of promoter regions for computational analysis of conserved sequences. This is one major aim of this study.

Based on cross-species homology search to known regulators of σ factors several genes have been identified in *Chlamydia*. These genes were named *rsbU, rsbV1, rsbV2* and *rsbW*. Unlike in other species, in *Chlamydia* these genes are not adjacent to σ factors but scattered throughout the genome (Kalman *et al.* 1999; Stephens *et al.* 1998). The function of these genes is largely unknown.

2.1.6.2 Transcription Factors

The first chlamydial transcription regulators identified were the heat shock proteins. These comprise the molecular chaperones DnaK (70 kDa), GroEL1 (60 kDa), and GroES (10 kDa) (Wilson and Tan 2002). These genes are upregulated upon heat stress and are involved in protein folding and refolding of misfolded proteins. Besides temperature stress, these genes are induced upon other stressors like osmotic stress, pH change, ethanol, and antibiotics (Ramos *et al.* 2001). A homologue to the *E. coli* heat shock response σ^{32} factor has not yet been identified. But as in other bacteria the *dnaK* and *groE* promoters are negatively regulated by the repressor HrcA by binding to the so called CIRCE operator (Narberhaus 1999) which is an inverted repeat located upstream of both the *dnaK* and *groE* promoter (Schmiel and Wyrick 1994; Tan *et al.* 1996). GroEL1 has been shown to be a potential pathogenicity factor and is associated to the outer membrane where it mediates binding of EB to the host cell (Wuppermann *et al.* 2008). *Chlamydia* contain two additional *groE* genes, *groEL2* and *groEL3*, both of which are not regulated by heat shock (Karunakaran *et al.* 2003) and not associated to the surface (Wuppermann *et al.* 2008).

A class of chlamydial transcription factors senses and responds to amino acids. ArgR represses the transcription of genes for putative arginine transporters (Schaumburg and Tan 2006) and thereby seems to regulate arginine import from the host. TrpR is another transcriptional repressor that is regulated by tryptophan levels and controls expression of the tryptophan operon (Akers and Tan 2006). Supplementation of several amino acids has been shown to inhibit chlamydial growth. Some of them result in a small persistent population (Gussmann *et al.* 2008) suggesting that amino acid homeostasis plays an important role in chlamydial gene regulation.

Several other transcriptional regulators have been identified in *Chlamydia* recently. These comprise the metal-responsive DNA-binding protein DcrA (Wyllie and Raulston 2001) and two histone-like proteins, Hc1 and Hc2, with sequence similarities to eukaryotic histone-like proteins. They seem to be unique in the bacterial world and have been shown to repress transcription and translation by DNA condensation (Pedersen *et al.* 1994, 1996). These proteins are thought to play a key role in the conversion from RB to EB since they are expressed late in the developmental cycle.

The identification of transcriptional regulators in *Chlamydia* has relied on homology searches in databases and predictions from sequence information. Only a few transcription factors have been identified so far, compared to the over 300 known in *E. coli* (Perez-Rueda and Collado-Vides 2000). There seem to be more, yet unidentified, regulators present in *Chlamydia*. This work should give a basis for the identification of new transcription regulators and putative DNA-binding sites by offering a catalogue of transcription start sites and new insights in transcript structures.

2.1.7 The Chlamydial Plasmids

All *C. trachomatis* serovars identified so far contain a conserved cryptic plasmid of 7.5 kb which is not essential for growth. Only one *C. pneumoniae* strain (N16) contains a plasmid of 7.4 kb

in which the ORF1 is disrupted by a deletion and split into two smaller ORFs. Plasmid number estimates range from 4 to 10 copies per cell (Pickett *et al.* 2005). There is a strong selection on plasmid carrying bacteria since the plasmid confers a survival advantage of the bacteria and plasmid-deficient strains are rare. Three plasmid-free strains have been described so far (Farencena *et al.* 1997; Peterson *et al.* 1990; Stothard *et al.* 1998). Absence of the plasmid has been associated with several chlamydial phenotypes like a reduced ability to establish infection both *in vitro* (O'Connell and Nicks 2006) and *in vivo* (O'Connell *et al.* 2007). Plasmid free strains are unable to accumulate glycogen inside the inclusions (Matsumoto *et al.* 1998) and are unable to activate innate immune signalling of the host cell via toll like receptor 2 (Darville *et al.* 2003).

The plasmid encodes eight major open reading frames (ORF) of more than 100 bases and four 22 bp repeats between ORF1 and ORF8 that are indicative of an origin of replication. ORF1 to ORF7 are located on the same strand and ORF8 is located on the opposite strand. Antisense to ORF8 several convergent and overlapping transcripts were detected (Fahr *et al.* 1992). The biological functions of the plasmid encoded transcripts are still unknown.

Diagnostic systems for routine detection of *C. trachomatis* were mainly based on amplification of a certain plasmid region. In late 2006, a deletion mutant lacking a 377 bp region of the plasmid was first described in Sweden (Ripa and Nilsson 2006). Since detection methods were based on the amplification of this region, this led to several thousand false-negative results and a spread of this new variant.

2.2 Transcriptome Analysis by Differential RNA-Sequencing (dRNA-seq)

The transcriptome is the complete set of transcripts in an organism. This comprises all species of RNA molecules that are present under the specific physiological conditions and at a specific time point. These transcripts consist of mRNA, non-coding RNA (including housekeeping RNAs like tRNA, rRNA, etc.), and small RNAs (sRNA). The transcriptional structure and organization is also essential information of an organism's transcriptome. This includes gene structure, untranslated regions, polycistronic transcripts, and post-transcriptional modifications. Furthermore, the quantity of a certain transcript gives information about expression patterns during different conditions.

The development of massive parallel DNA sequencing technologies provides a new opportunity to experimentally gain access to the whole transcriptome of an organism. RNA-sequencing (RNA-seq) uses next-generation sequencing instruments to sequence cDNA libraries made of total RNA from cells or organisms (Wang *et al.* 2009) and gain millions of sequence reads from one biological sample. The use of a barcode which consists of a short unique nucleotide sequence that is introduced to each cDNA molecule during library preparation allows to pool different samples in one sequencing run. The sequences can be sorted by their barcode according to their affiliation to a biological sample after the sequencing run. This allows analysing multiple samples in one sequencing run and consequently save costs (Papenfort *et al.* 2010).

RNA is a dynamic molecule with short half-life and is physically labile. It is degraded rapidly by numerous nucleases and is also prone to hydrolysis. Isolated RNA contains many processed and partially degraded molecules that make it difficult to exactly map the native 5'end of an RNA molecule. Differential RNA-sequencing overcomes this drawback by specific enrichment of RNA molecules with a native 5'end. Bacterial primary transcripts possess a 5'triphosphate, whereas enzymatically processed RNAs including the abundant tRNAs and rRNAs have 5'monophosphate ends. The enzyme Terminator Exonuclease specifically degrades transcripts that have a 5'monophosphate end, but not those that have a 5'triphosphate, 5'capped, or 5'hydroxyl end. Thereby all RNA molecules that are not derived from primary transcription are removed. This technique allows to specifically sequence native 5'ends and determine TSS (Sharma *et al.* 2010).

2.3 Bacterial Small Non-coding RNAs (sRNA)

Bacteria are highly adaptive organisms that have compact genomes and stringent mechanisms of gene regulation. Traditionally, these mechanisms comprise alternative RNA polymerases (σ factors) and transcription factors that switch genes on and off as response to environmental signals. During infection, pathogenic bacteria must express their virulence factors and genes necessary for adaptation to their host. Recently, non-coding RNAs (sRNAs) were identified as new regulatory mechanism in all species they have been searched for. They are acting on all layers of gene regulation like transcription, mRNA stability and protein activity. Additionally, proteins have been identified that mediate the interaction of sRNAs with their targets. In bacteria, most sRNAs coordinate adaption processes in response to environmental signals (Mitschke *et al.* 2011).

2.3.1 Classes of Bacterial sRNAs

Most bacterial sRNAs discovered so far are acting by base-pairing (also termed antisense RNA) and can be divided into two classes, namely *cis*-encoded and *trans*-encoded sRNAs. *Cis*-encoded sRNAs are encoded on the same DNA locus. They can be part of the mRNA transcript such as thermosensors and riboswitches that comprise a structural part of the 5'end. *Cis*-RNAs that are encoded on the opposite DNA strand are acting by extensive base pairing due to their complementary sequence (Brantl 2007). sRNAs encoded antisense to the ribosome binding site (RBS) can prevent translation and trigger mRNA degradation. Most *cis*-encoded antisense sRNAs have been identified on mobile genetic elements like plasmids, transposons and bacteriophages and regulate copy number by a variety of mechanisms which have been reviewed by Brantl (2007). The physiological roles of *cis*-encoded antisense RNAs encoded on the bacterial chromosome are less well understood. One group of these sRNAs modulate gene expression of in operons. They are encoded antisense to intergenic regions of ORFs and by duplex formation with the polycistronic transcript transcription can be terminated (Windbichler *et al.* 2008) or the mRNA can be cleaved (Koo and Stephens 2003). Recent approaches of global transcriptome analysis have identified massive antisense transcription in several bacteria like *C. trachomatis* (Albrecht *et al.* 2010), *Neisseria gonorrhoea* (Albrecht *et al.*, unpublished

results), *Listeria monocytogenes* (Koo *et al.* 2006), *Helicobacter pylori* (Sharma *et al.* 2010), and others. This suggests that there are plenty of sRNAs of unknown function.

Trans-encoded sRNAs are likely to constitute the largest group of regulatory RNAs in bacteria. They are encoded at different loci than their target genes and act by base-pairing of short, partially imperfect sequences. These sRNAs regulate translation or stability of target mRNAs and are, in many aspects, functionally homologous to eukaryotic microRNAs (Grech *et al.* 2008). There is no correlation between the genomic location of *trans*-encoded sRNAs and their targets and it seems common that they target multiple mRNAs. Most known *trans*-encoded sRNAs negatively regulate protein expression by mRNA degradation or translational inhibition. The majority of characterized sRNAs binds to the 5'UTR of mRNAs and inhibits binding of the ribosome to the ribosome binding site by sterical hindrance. Besides, some sRNAs bind far upstream of the start codon of the repressed gene (Weissenmayer *et al.* 2011). Inhibition of ribosome binding is thought to be the major mechanism that leads to decreased protein concentrations whereas the cleavage of the sRNA-mRNA duplex by ribonucleases increases the regulation and makes it irreversible (Herbig and Nieselt 2011). Some *trans*-acting sRNAs have been shown to activate translation by disruption of inhibitory duplex structures of the mRNA that make the RBS inaccessible (Spencer *et al.* 2011).

Most known *trans*-acting sRNAs have been shown to be associated with the RNA chaperone Hfq. This protein forms a hexameric ring and is homologous to eukaryotic Sm-like proteins involved in mRNA splicing and decay. The current model comprises several aspects of Hfq-sRNA interaction. Hfq may actively remodel the RNAs to melt inhibitory secondary structures. Furthermore, it prevents sRNAs from being degraded by binding and protecting it from nucleases. Another aspect is the increase in local concentrations of sRNA and target mRNAs by binding and spatial colocalization. *In vitro*, Hfq is not necessary for sRNA:target interaction (Zagordi *et al.* 2010). Hfq could not be detected in any chlamydial species. Since all Hfq-proteins detected so far are highly conserved, it is unlikely that *Chlamydia* harbour an Hfq-like RNA chaperone. Furthermore, no other RNA-binding protein could be detected. However, the presence of *trans*-acting regulatory sRNAs in *Chlamydia* is very likely since such sRNAs have been discovered in all bacteria investigated so far, including others that lack Hfq like *Helicobacter pylori* (Sharma *et al.* 2010).

2.3.2 Mechanisms of RNA-Based Gene Regulation

Small RNAs have been shown to control various genes in response to environmental stimuli like iron sensing, quorum sensing and anaerobic stress. While transcription factors mainly control gene expression at the level of transcription, sRNAs interfere with posttranscriptional processes like mRNA accessibility, translation initiation and termination. Several sRNAs have been identified that play a role in the control of virulence genes of bacterial pathogens. One prominent candidate is RNAIII of *Staphylococcus aureus* that has a dual function. It encodes a 26 aa δ-hemolysin peptide and a functional sRNA that targets several virulence genes and a transcription factor (Pallen *et al.* 2010).

RNA thermosensors are *cis*-encoded hairpin structures that sense temperature (for example host body temperature) and regulate translational activity of the mRNA by conformational changes (Borges *et al.* 2010).

U6 RNA is a global regulator of gene expression, rather than targeting specific mRNAs by base-pairing. It has a secondary structure that mimics an open DNA promoter sequence and thereby binds the standard (*E. coli* σ^{70}) DNA-dependent RNA polymerase (RNAP) and competes with DNA promoter sequences. Thus, alternative RNAP containing other σ factors can access the DNA promoters and transcribe alternative genes (Wassarman and Storz 2000). During normal growth of *E. coli*, σ^{70} RNAP is mainly active and 6S RNA concentrations are low. When nutrients become limiting 6S RNA levels are increased and σ^{70} RNAP activity decreases. In parallel, RNAP containing alternative σ factors are becoming more active to switch gene expression from exponential to stationary phase (Wassarman and Saecker 2006).

Compared to protein based gene regulation mainly based on transcription factors, regulatory RNAs have several advantages. Due to the short length of sRNAs and the missing translation step they are cost efficient and provide faster reactions than proteins. Small RNAs are often degraded together with the mRNA they target and thereby a fast and efficient transition is achieved, which is often required during pathogenesis.

2.3.3 Identification of Bacterial sRNAs

Since the discovery of the first chromosomally encoded sRNA (Toledo-Arana and Solano 2010) and the first genome wide screens for sRNAs in *E. coli* (Tisoncik and Katze 2010), new innovative methods have been developed or adapted to discover sRNAs on a genome wide level. The first sRNAs were discovered as abundant bands in denaturing RNA gels (Vogel and Sharma 2005). With the availability of bacterial genomes different biocomputational algorithms have been developed to predict sRNAs based on their secondary structure and conservation between species (Siezen *et al.* 2010). Only a few of them could be experimentally verified, suggesting a high false positive rate of sRNA prediction by these algorithms. With the availability of microarrays covering intergenic and antisense regions global sRNA analysis was achieved (O'Connell *et al.* 2011). This method was even advanced by the use of high density tiling arrays that cover intergenic regions or whole genomes by partially overlapping probes (Koo *et al.* 2006).

The major break-through of genome wide sRNA discovery was achieved by sequencing of cDNA prepared from size fractionated or total RNA. By classical Sanger sequencing most abundant sRNAs could be identified (Abdelrahman *et al.* 2011). Due to the limited number of sequences and the presence of highly abundant RNAs (e.g. tRNA, rRNA) this approach was inapplicable to discover rare sRNAs. With the introduction of next-generation sequencing technology (also referred to as deep sequencing) this problem was overcome, since millions to billions of cDNA sequences can be obtained by this technology from a single biological sample. Furthermore, this technique can detect

sRNA transcripts with single nucleotide resolution at the 5' and 3'end. Using enzymatic treatment prior to cDNA synthesis classes of RNAs can be selected for sequencing. This allows, for example, to select for primary transcripts that are characterized by a 5'triphosphate (5'PPP) and thereby to exclude processed and degraded RNA species from the cDNA library (see Material and Methods for details). The major drawback of current deep sequencing technologies, that are bias introduction by adapter ligation and PCR amplification, will be overcome by the introduction of single nucleotide sequencing technologies within the next years (Filiatrault *et al.* 2010).

2.3.4 Target Identification of RNA-Based Gene Regulation

Several biocomputational approaches have been developed to identify targets of sRNAs on whole genome base (Fouquier d'Herouel *et al.* 2011; Tjaden *et al.* 2006). Since *trans*-acting sRNAs base pair only by short complementary sequences that may even contain mismatches, a reliable prediction of sRNA target interactions is challenging. Depending on the set parameters these algorithms predict a multitude of different targets and the true positive rate is questionable.

Experimental methods to identify sRNA targets are well established but more time consuming and cost intensive. Knockout and overexpression of sRNA genes can be easily performed in all bacteria that are accessible to genetic manipulations. Microarrays or proteomic approaches are then used to globally analyse deregulated genes or gene products, respectively. However, the constitutive removal or overexpression of sRNAs leads to a vast number of deregulated genes, most of them secondary or tertiary targets. This approach makes it difficult to elucidate primary sRNA targets that are deregulated by direct molecular interaction. Using an inducible overexpression system (pulse overexpression) followed by a rapid RNA isolation can efficiently be used to discover primary targets. Using this approach the sRNA is induced long enough to cause decay of the mRNA target and short enough to avoid alterations at the protein levels that can cause secondary effects (Hansen *et al.* 2010). The majority of sRNAs described so far, negatively regulates their target mRNAs. However, some sRNAs do not reduce the abundance of their target mRNA but rather inhibit translation. These targets can be missed by microarray screens.

Bacterial sRNAs are known to interact with proteins such as Hfq (Sittka *et al.* 2009; Tisoncik and Katze 2010), RNA-Polymerase (Wurm *et al.* 2010), CsrA (Liu *et al.* 1997) or RNase E (Morita *et al.* 2005). Precipitation of the endogenous protein can be applied to co-purify bound sRNAs as well as their targets. This approach can be used the other way around by tagging the sRNA (e. g. with an aptamer tag) and subsequent co-purification of the bound proteins (Windbichler *et al.* 2008).

2.3.5 Small RNAs in *Chlamydia*

When this work was started, only one sRNA was known in *Chlamydia*. Grieshaber *et al.* (2006a) performed a heterologous screen in *E. coli* to select for genes controlling histone-like protein Hc-1 expression by targeting its gene *hctA*. A small transcript called ihtA (inhibitor of *hctA* translation) was found to be expressed from an intergenic region and to negatively regulate expression

of Hc1. This sRNA has a length of 120 nt and consist of three stem loops including a rho-independent terminator. A single promoter could also be identified. This sRNA acts at the crucial regulatory step of RB-to-EB conversion since it effects global transcription by DNA condensation. During the initial conversion from EB to RB, ihtA levels are increased and the synthesis of Hc1 stops. Protein levels decrease probably by degradation and DNA condensation of chromatin is prevented. At the late stage of the infection cycle ihtA levels drop and Hc1 levels increase. This leads to a global inhibition of translation. However, ihtA does not affect translation of the second histone-like protein Hc2 encoded by *hctB* (Grieshaber *et al.* 2006b) which suggests that other sRNAs or other layers of regulation affect the life cycle conversion.

3 AIM OF STUDY

Chlamydia are important human pathogens that are causing severe infections that are major causes of blindness in developing countries (*C. pneumoniae*) and infertility in women in the western world (*C. trachomatis*). Since tools to genetically manipulate *Chlamydia* are still not available other approaches must be applied to gain insight into genetic organisation and gene regulation mechanisms. In this work a recently developed method of differential cDNA high-throughput sequencing (dRNA-seq) is applied to analyse the global transcriptomes of the important human pathogens *C. trachomatis* and *C. pneumoniae*. By enzymatic enrichment for native primary transcripts a catalogue of transcriptional start sites (TSS) of purified EB and RB is aspired. Furthermore, novel transcripts lacking in the annotation of the two genomes should be discovered, especially non-coding RNAs. Mechanisms of gene regulation and the use of alternative transcription factors for recognition of promoter sequences are far from being understood in *Chlamydia*. The identification of TSS down to single nucleotide resolution should help to identify promoter sequences and promote their analysis. This might give hints to the regulation of the unique bi-phasic life cycle of *Chlamydia*. Differential analysis of the EB and RB transcriptomes should unveil further details of EB-to-RB transition. The detailed map of TSS should complement the genomic annotation and confirm gene annotation that is mainly based on computational predictions and inter-species homology. Furthermore, errors in the annotation can be corrected by the availability of the global genome. In summary, this work should represent another milestone, after genome sequencing, to gain deeper insights into genome organisation and gene regulation of these two human pathogens.

4 MATERIAL

4.1 Bacterial Strains

Species	Strain	Origin
Chlamydia trachomatis	L2b/UCH-1/proctitis	MPI-IB Berlin
Chlamydia pneumoniae	CWL-029	ATCC VR-1310

4.2 Eukaryotic Cell Lines

Cell line	Origin	ATCC#
HeLa229	Cervix carcinoma	CCL-2.1
HEp-2	HeLa contaminant	CCL-23

4.3 Chemicals

Chemical	Supplier
γ^{32P}-ATP	Perkin Elmer
Acrylamide/bisacrylamide	Roth
Ammonium persulphate	Roth
Boric Acid	Sigma Aldrich
Bromophenol blue	Fluka
Chloroform	Roth
4',6-diamidino-2-phenylindole (DAPI)	Merck
Ethylenediaminetetraacetic acid (EDTA)	Sigma Aldrich
Formamide	Sigma Aldrich
L-Glutamate	Sigma
GlycoBlue	Ambion/Life Technologies
KH_2PO_4	Merck
K_2HPO_4	Merck
Magnesium chloride ($MgCl_2$)	Merck
Mowiol 4-88	Roth
Paraformaldehyde	Roth
Percoll Plus	GE Life Sciences
Phenol/Chloroform/Isopropanol (P:C:I)	Roth
Phenylmethylsulfonyl fluoride (PMSF)	Sigma Aldrich
Potassium chloride (KCl)	Merck
RapidHyb buffer	GE Life Sciences
RiboLock RNase Inhibitor	Fermentas
RiboRuler RNA Ladder	Fermentas

Sodium chloride	Roth
Sodium citrate	Roth
Sodium dodecyl sulphate (SDS)	Applichem
Sucrose	Roth
TEMED	Roth
Tris Base	Applichem
Triton X-100	Sigma Aldrich
Trizol Reagent	Invitrogen
Tween 20	Sigma Aldrich
Urea	Roth
Xylene Cyanol	Fluka

4.4 Kits

Kit	Supplier
Decade Marker kit	Ambion/Life Technologies
ECL Western Blotting substrate	Pierce/Thermo Scientific

4.5 Antibodies

Antibody	Origin
Chlamydial σ28 (rabbit polyclonal)	Ming Tan Lab, University of California, Irvine, USA
Bacterial HSP60 (mouse clone A57-B9)	Enzo Life Sciences, #ALX-804-072
Goat anti-rabbit HRP-conjugate	Santa Cruz Biotechnology, #sc-2004
Goat anti-mouse HRP-conjugate	Santa Cruz Biotechnology, #sc-2005

4.6 Enzymes

Enzyme name	Supplier
Calf intestine phosphatase (CIP)	Fermentas
Polynucleotide kinase (PNK)	Fermentas
Terminator Exonuclease (TEX)	Epicentre
Tobacco acid pyro phosphatase (TAP)	Epicentre
DNase I	Fermentas

4.7 Buffers and Solutions

SPG buffer:	250 mM	Sucrose
	3.8 mM	KH_2PO_4
	7.2 mM	K_2HPO_4
	0.5 mM	L-Glutamate
	pH 7.4	

RNA Gel loading buffer (1-2x):	95% (v/v)	Formamide
	18 mM	EDTA
	0.025% (w/v)	SDS
	Traces	Xylene Cyanol
	Traces	Bromophenol blue
10x TBE:	0.9 M	Tris base
	0.9 M	Boric acid
	20 mM	EDTA
20x SSC:	3 M	Sodium chloride
	300 mM	Sodium citrate
	pH 7.0	
4× SDS-Loading buffer:	70 mM	Tris-HCl, pH 6.8
	5% (v/v)	β-mercaptoethanol
	40% (v/v)	Glycerol
	3% (w/v)	SDS
	0.05% (w/v)	Bromophenol blue
SDS-PAGE buffer:	1.92 M	Glycine
	0.25 M	Tris
	1% (w/v)	SDS
SD buffer I:	25 mM	Tris, pH 10.4
	10% (v/v)	Methanol
SD buffer II:	300 mM	Tris, pH 10.4
	10% (v/v)	Methanol
SD buffer III:	25 mM	Tris, pH 9.4
	40 mM	Glycine
	10% (v/v)	Methanol
10× TBST buffer:	400 mM	Tris-HCl, pH 7.6
	2.75 M	NaCl

	1% (v/v)	Tween 20
Lysis buffer:	20 mM	Tris-HCl, pH7.5
	150 mM	KCl
	1 mM	$MgCl_2$
	1 mM DTT	1 mM PMSF
	200 U/ml	RNase inhibitor
Glutaraldehyde buffer:	2.5% (v/v)	Glutaraldehyde
	50 mM	Sodium cacodylate (pH 7.2)
	50 mM	KCl
	2.5 mM	$MgCl_2$
Cacodylate buffer:	2% (w/v)	OsO_4
	50 mM	Sodium cacodylate (pH 7.2)

4.8 Oligonucleotides

Sequences of oligonucleotides are listed in the respective results parts.

4.9 Lab Ware and Consumables

Product	Supplier
Glass beads	Roth
Lysing Matrix B tubes	MP Biomedicals
Hybond N+, nylon membrane	GE Life Sciences
Hybond P, PVDF membrane	GE Life Sciences
Hybaid hybridization flasks	Thermo Scientific
Microspin Sephadex G-25 columns	GE Life Sciences
Phase Lock tubes	Eppendorf

4.10 Equipment

Device	Supplier
Bioanalyzer 2100	Agilent Technologies
Cell incubator	Thermo Scientific
Chemiluminescence camera system	Intas
Class 2 work bench	Thermo Scientific
Electron microscope	Zeiss
Electrophoresis chamber, vertical	Amersham
FastPrep homogenizer	MP Biomedicals

Gradient maker	VWR
High speed centrifuge	Beckman Coulter
Hybridization oven	Peqlab
Microscope, UV-lamp	Leica
Nanodrop spectrophotometer	Thermo Scientific
Peristaltic pump	Amersham Pharmacia
Phospho storage plates	FujiFilm
Semi-dry blotter	Peqlab
Table top centrifuge (cooled)	Eppendorf
Typhoon Scanner	GE Life Sciences/Molecular Devices
Ultracentrifuge	Beckman Coulter
UV cross-linker	Stratagene
UV transilluminator, gel imager	Biostep
Vertical gel electrophoresis chamber	Hoefer
Wet blot chamber	Peqlab

4.11 Software

Device	Supplier
Endnote X4	Thomson Scientific
Illustrator CS4	Adobe
Integrated Genome Browser, version 4.6	Affymetrix
ImageMaster	GE Life Sciences
LabImage, ChemoStar	Intas
Photoshop CS3	Adobe
Word 2007	Microsoft
WU-BLAST 2.0	Washington University

5 METHODS

5.1 Cell Biology

5.1.1 General Handling of Human Cell Lines

HeLa and Hep-2 cells were grown in cell culture flasks or multi-well plates to a maximum confluency of 90 % in a humidified incubator at 37°C and 5% CO_2. For splitting, the cells were washed with PBS twice and incubated with trypsin/EDTA (Invitrogen) solution at 37°C. After entire detachment trypsin was inactivated by addition of growth medium (Dulbecco's modified Eagle medium (DMEM, Invitrogen) supplemented with 10% FBS (Biochrom) and sodium pyruvate (Invitrogen)). Cells were diluted and seeded again at an appropriate concentration.

5.2 Microbiology

5.2.1 Propagation and Purification of *Chlamydia trachomatis*

For infection with *C. trachomatis*, HeLa229 cells were seeded one day prior to infection in 150 cm² cell culture plates and grown to a confluency of 50-70%. The cell number was estimated and the cells were infected with a multiplicity of infection (MOI) of 5 in a volume of 10 ml infection medium (DMEM supplemented with 5% FBS) at 35°C and 5 % CO_2. The plates were swayed every 30 minutes and after 3 hours the medium was changed against culture medium (DMEM supplemented with 10% FBS). 72 hours post infection cells were scraped from the plates and collected in 50 ml tubes containing 5 ml of glass beads. The cells were disrupted by vortexing for 5 minutes at maximum speed with subsequent cooling on ice. In a first centrifugation step, host cell debris was pelleted at 1,500 rcf at 4°C for 10 minutes. In a second centrifugation step *Chlamydia* were pelleted at 25,000 rcf for 20 minutes at 4°C. The pellet was washed two times in SPG buffer (25,000 rcf, 20 min, 4°C), diluted in 5 ml cold SPG buffer and resuspended by passaging through a 26G needle. Aliquots of 50 µl were frozen at -80°C and the bacteria titre was estimated.

5.2.2 Propagation and Purification of *Chlamydia pneumoniae*

Hep-2 cells were seeded in 75 cm² cell culture flasks and grown to a confluency of approximately 90%. The infection medium consisted of DMEM supplemented with 5% FBS and cycloheximide at a final concentration of 1 µg/ml. The amount of bacteria corresponded to a MOI of 5. Infection with *C. pneumoniae* was carried out by centrifugation of the bacteria onto the cells for one hour at 4 °C and 1,500 rcf. Three hours post infection the cell culture medium was exchanged to remove excess bacteria from the suspension and thereby reduce cytotoxic effects. After two rounds of infection cycle (7-9 days) most cells were lysed by the bacteria. The remaining cells were scraped from the cell culture flasks and the cell-bacteria suspension was transferred to a 50 ml centrifuge tube containing 5 ml of glass beads. Cells were disrupted by vortexing of the suspension at maximum speed for 5 minutes. After transfer to a fresh centrifuge tube host cell debris was pelleted for 5 minutes at 4°C and 1,000 rcf. Bacteria were pelleted from the supernatant by

centrifugation for 20 minutes at 4°C and 25,000 rcf. The bacteria pellet was resuspended in SPG buffer and further purified by centrifugation through 30% sucrose in SPG solution in a swing out rotor for 1 hour at 4°C and 25,000 rcf. Finally the pellet was washed two times in SPG, resuspended in a small volume of SPG and 50 µl aliquots were stored at -80°C.

5.2.3 Infection and Estimation of Bacteria Stock Titre

Infection of human cell lines was carried out as described before, but in 24-well cell culture plates with glass cover slips in each well. For estimating the concentration of infectious bacterial particles in the stock a binary dilution series was performed. Cells were fixed with paraformaldehyde 20 or 40 hours after infection with *C. trachomatis* and *C. pneumoniae*, respectively. After staining the DNA with DAPI for 15 minutes at room temperature in the dark, the cover slips were mounted onto glass slides using Mowiol. Ten individual fields were photographed per dilution step using a UV microscope and 100x objective. From the ratio of chlamydial inclusions per nucleus the concentration of infectious bacteria in the stock solution was calculated.

5.2.4 Isolation of EB and RB of *Chlamydia trachomatis*

HeLa229 (ATCC CL-2.1) cells were cultured in DMEM containing 10% FBS and infected with *C. trachomatis* strain L2b with a MOI of 5 for 24 h. *C. trachomatis* containing cells were collected by scraping, pooled and disrupted with glass beads by vortexing for 5x 5 min with chilling on ice in between. *Chlamydia* were isolated by differential centrifugation followed by density gradient centrifugation as described before (Newhall *et al.* 1982) with several modifications to achieve better separation of EB and RB. Cell debris was pelleted at 1,500 rcf for 10 min at 4°C followed by centrifugation of the *Chlamydia* containing supernatant at 30,000 rcf for 30 min at 4°C. The bacterial pellet was washed twice in sucrose-phosphate-glutamate (SPG) buffer and resuspended in 2 ml SPG using a 26G syringe needle. The *C. trachomatis* suspension was layered on top of a Percoll Plus (GE Healthcare) solution containing 0.25 M sucrose. EB and RB were separated based on their different buoyant densities on an *in situ* forming continuous density gradient by centrifugation at 30,000 rcf for 30 min at 4°C in a fixed angle rotor (JA25.50; Beckman Coulter). Two distinct bands were visible containing EB (lower) and RB (upper). These were carefully collected using a syringe and washed twice in SPG buffer. Purity of EB and RB preparations was verified by electron microscopy of the bacterial pellets.

5.2.5 Isolation of EB and RB of *Chlamydia pneumoniae*

Hep-2 (ATCC CCL-23) cells were cultured in DMEM containing 10% FBS and infected with *C. pneumoniae* strain CDC/CWL29 (ATCC VR1310) with a MOI of 5 for 24 and 72 hours. *C. pneumoniae* containing cells were collected by scraping, pooled and disrupted with glass beads. *Chlamydia* were isolated by differential centrifugation followed by density gradient centrifugation in a discontinuous sucrose density step gradient. 2 ml of SPG buffer containing 45%, 40%, 35%, 30%, and 25% of sucrose, respectively, were layered on top of each other starting with the highest concentration in a 12 ml polycarbonate ultracentrifuge

tube. The gradients were precooled for 1h at 4°C. Cells were disrupted and crude *Chlamydia* pellets were obtained as described before for *C. trachomatis*. The bacterial pellet was resuspended in 1 ml of cold SPG buffer without using a syringe to avoid mechanical disruption of RB. Then *C. pneumoniae* suspension was layered on top of the sucrose step gradient followed by centrifugation for 60 minutes at 4°C and 30,000 rcf in a swing out rotor. After centrifugation EB were present as a distinct band at the interphase of the 40% and 35% sucrose. RB were present at the interphase of 35% and 30% sucrose. Cell debris was distributed as a diffuse band in the 25% sucrose fraction. EB and RB were carefully collected by capillary pipettes and washed in SPG buffer. Purity of the pellets was estimated by electron microscopy.

5.2.6 Electron Microscopy of Purified EB and RB

Purified EB and RB were pelleted by centrifugation for 20 min at 4°C and 25,000 rcf in a table top centrifuge. The Pellet was fixed for 45 minutes in glutaraldehyde buffer at room temperature. Then pellets were incubated in cacodylate buffer for 2 h at 4°C. Further preparation was performed by the department of electron microscopy (University of Würzburg, Biocenter, Prof. Dr. Georg Krohne). In brief, pellets were dehydrated, embedded, ultra-thin sectioned, and stained. Subsequent electron micrographs were digitalized by scanning.

5.3 Biochemistry

5.3.1 Preparation of *C. pneumoniae* Lysates

Four 75 cm² cell culture flasks with Hep-2 cells at a confluency of approximately 80% were infected with *C. pneumoniae* at a MOI of 5 for 40h and 80 hours, respectively. Cells were washed with ice cold PBS and collected by scraping on ice. Cells were pooled in a total volume of 20 ml cold PBS and disrupted by adding 5 ml of glass beads (1 mm diameter) and vortexing for 5 min at maximum speed with. Cell debris was removed by pelleting for 10 minutes at 1,000 rcf and 4°C. *Chlamydia* containing supernatant was centrifuged at 25,000 rcf for 20 minutes at 4°C to pellet bacteria and washed once with lysis buffer. Bacteria were resuspended in 2 ml lysis buffer containing Triton X-100 at a final concentration of 0.2% and portioned into two Lysis Matrix B tubes. Bacteria were disrupted by homogenization for 5x 20 seconds at 6.5 m/s with dry ice cooling in a FastPrep homogenizer. The lysates were cleared by centrifugation at 14,000 rcf at 4°C for 10 minutes.

5.3.2 Fractionation of Lysates on a Density Gradient

Molecular biology grade glycerol was used to prepare 1% and 40% stock solutions in lysis buffer containing 0.2% Triton X-100. The solutions were filtered through a 0.2 µm pore size filter. Continuous 1-40% glycerol gradients were formed in 12 ml ultracentrifuge tubes using a gradient maker device and a peristaltic pump. The cleared lysate was layered on top of the gradient and separated by centrifugation for 17h at 100,000 rcf and 4°C.The gradient was fractionated into 24 fractions of 500 µl each whereby the last fraction contained the resuspended pellet. 200 µl of each fraction were used for protein detection by Western blotting. The remaining 300 µl lysate from each fraction were transferred to PhaseLock tubes and RNA was

isolated by the phenol/chloroform method. One volume (300 µl) of P:C:I was added to the fractions followed by vigorous shaking for 30 seconds. Phases were separated by centrifugation at 14,000 rcf for 15 minutes at 4°C. The aqueous phases were transferred to fresh 1.5 ml tubes and 1.5 µl of GlycoBlue were added to each fraction as carrier to improve precipitation. RNA was precipitated by addition of 2.5 volumes ethanol containing 0.1 M sodium acetate, shaking, and incubation over night at -20°C. Subsequently RNA was pelleted by centrifugation for 20 minutes at 20,000 rcf and 4°C. The pellets were washed with 75% ethanol and air dried for 5 to 10 minutes. RNA was dissolved in 30 µl water and concentrations were estimated spectrophotometrically by a Nanodrop device.

5.3.3 SDS-Polyacrylamide Gel Electrophoresis

Proteins were separated on the basis of mass by electrophoresis in a polyacrylamide gel under denaturing conditions with sodium dodecyl sulphate (SDS-PAGE). Proteins in fractions from the glycerol gradient (see section 5.3.2.) were denatured in 4x SDS-loading buffer by heating at 95°C for 7 min. SDS is an anionic detergent that disrupts nearly all non-covalent interactions in native proteins. β-mercaptoethanol is also included in the sample buffer to reduce disulphide bonds. The SDS complexes with the denatured proteins were electrophoresed in a 12% polyacrylamide gel. Vertical gels are set in between two glass plates with 1.5 mm spacers between the plates. In this chamber, the acrylamide mix was poured and left to polymerize for at least 30 min at RT. The gels are composed of two layers: a 12% separating gel (pH 8.8) that separates the proteins according to size and a lower percentage (5%) stacking gel (pH 6.8) which assures that the proteins simultaneously enter into the separating gel. The proteins were separated at a voltage of 12 V/ cm electrode distance until the bromophenol blue dye reached the bottom of the gel.

5.3.4 Western Blot Analysis of Proteins

In order to make the proteins accessible to antibody detection, they were moved from within the gel onto a membrane made of PVDF using a semidry electro blotter. The gel was removed from its glass cassette and any stacking gel was trimmed away. The gel was immersed in SD buffer 3 for 15 min. The PVDF membrane was activated in methanol and subsequently equilibrated in SD buffer I for at least 5 min. Two pieces of filter paper soaked in SD buffer II were placed in the centre of the anode electrode plate. One piece of filter paper soaked in SD buffer I was placed on top of the first two sheets. The nitrocellulose membrane was placed on top of the filter papers. The gel was placed on top of the membrane. Three pieces of filter paper soaked in SD buffer 3 were placed on top of the membrane. To ensure an even transfer, air bubbles between the layers were removed by carefully rolling a pipette over the surface of each layer in the stack.

The protein transfer was performed at 2.5 mA/cm^2 for 2 h. For immunodetection, the membrane was incubated in TBST buffer containing 5% milk and 1% BSA for 1 h at RT on a shaker. Then the membrane was washed three times for 5 min in TBST buffer by vigorous shaking. The primary antibody was diluted in TBST buffer containing 1% BSA and added to the membrane. The membrane was incubated with the primary antibody overnight at 4°C on a shaker. After incubation, the membrane was washed again three

times for 5 min with TBST buffer by vigorous shaking. The appropriate peroxidase-conjugated secondary antibody was diluted in TBST buffer and added to the membrane. The membrane was incubated with the secondary antibody at RT for 1 h and washed three times with TBST buffer by vigorous shaking, each time for 5 min. This step was followed by the standard enhanced chemiluminescence reaction (ECL system).The membrane was incubated in a 1:1 mixture of ECL solution. This reaction is based on a peroxidase-catalysed oxidation of luminol, which leads to the emission of light photons that was detected by a chemiluminescence imager system.

5.4 Molecular biology

5.4.1 Isolation of RNA from Purified Chlamydia

Pelleted bacteria were resuspended in Trizol, mechanically disrupted in a homogenizer (FastPrep, MP Biomedicals) using 1.5 ml Lysing Matrix B tubes for 4 bursts of 25 sec each at maximum speed and dry ice cooling. Subsequently, RNA was isolated according to the manufacturer's protocol. Contaminating DNA was digested by DNase. Therefore the RNA was incubated with 1U DNase per µg of RNA in the supplied reaction buffer for 30 min at 37°C followed by isolation of RNA by phenol/chloroform/isoamyl alcohol and precipitation of RNA by 2.5 volumes of ethanol containing 0.1 M sodium acetate. The absence of DNA was controlled by PCR using primers to amplify genomic DNA of the *ompA* gene. RNA quality was determined on a Bioanalyzer 2100. Absence of 18S and 28S eukaryotic ribosomal RNA peaks supported the purity of the bacteria preparation.

5.4.2 Polyacrylamide Gel Electrophoresis of RNA and Northern Blot

50 µg of total RNA were diluted with an equal amount of 2x RNA gel loading buffer, denatured for 5 minutes at 95°C and quick chilled on ice. Then the RNA was separated on a denaturing 8% polyacrylamide gel (1x TBE buffer, 8 M Urea) in 1x TBE buffer at a constant current of 30 mA. For size detection the Ambion Decade marker (10 – 150 nt) and the RiboRuler low range marker (100 – 1000 nt) were prepared according to the manufacturer's protocols and 2 µl were loaded onto the gel, respectively. Electrophoresis was stopped when the bromophenol blue dye front had migrated to the bottom of the gel. Then the gel was soaked for 5 minutes in a 1 µg/ml solution of ethidium bromide and RNA was visualized on a UV gel imager and checked for degradation.

The transfer of RNA to a positively charged nylon membrane was carried out by wet blot transfer in 0.5x TBE buffer for 3 hours at 4°C and a constant voltage of 500 V. After blotting the RNA was UV-cross-linked to the membrane by exposure of 120 mJ.

Prehybridization of the membrane was carried out in 12 ml RapidHyb buffer at 42°C for at least 1 hour in a hybridization flask in a rotating hybridization oven. 10 pmol of the DNA probe were labelled with 3 µl of γ-^{32}P-ATP and T4-Polynucleotide kinase (PNK) in the supplied buffer for 1 hour at 37°C. After heat inactivation of the PNK for 3 minutes at 95°C, the labelled probe was purified by a Sephadex-25 gel

filtration column according to the manufacturer`s protocol. For hybridisation the radioactive labelled probes were directly added to the prehybridization buffer and incubated for 16 hours at 42°C.

Following incubation the membranes were washed three times for 7 minutes in prewarmed washing buffer at 42°C. Some probes needed an additional washing step for 10 min at 45°C to reduce background and improve specificity. Then the membranes were wrapped in plastic film and exposed to phospho-storage plates 12 to 72 hours according to the signal strength. The screens were read by the Typhoon scanner and results were visualized by LabImager image analysis software.

5.4.3 Next-Generation Sequencing of the *Chlamydia* Transcriptomes

5.4.3.1 Enrichment of Primary Transcription Start Sites

Primary transcripts of total RNA were enriched by selective degradation of RNAs containing 5' mono-phosphate (5'P) by treatment with a 5'phosphate dependent Terminator Exonuclease (TEX; Epicentre Biotechnology #TER51020). Primary bacterial transcripts (most mRNAs and sRNAs) are protected from exonucleolytic degradation by their tri-phosphate (5'PPP) RNA ends. Total RNA was freed of residual genomic DNA by treatment with 1U DNaseI per µg of RNA for 30 minutes at 37°C. For depletion of processed transcripts, equal amounts of *Chlamydia* RNA were incubated with TEX or in buffer alone for 60 min at 30°C. 1 unit TEX was used per 1 µg total RNA. Following organic extraction (25:24:1 v/v phenol/chloroform/isopropanol), RNA was precipitated overnight with 2.5 volumes of an ethanol/0.1M sodium acetate (pH 6.5) mixture, and treated with 1 unit TAP for 1 hour at 37°C to generate 5'mono-phosphates for linker ligation, and again purified by organic extraction and precipitation as above.

5.4.3.2 Preparation of cDNA Libraries

Libraries for 454 pyrosequencing of cDNA were constructed by *vertis* Biotechnology AG, Munich, Germany (http://www.vertisbiotech.com/) as described previously for eukaryotic microRNA (Berezikov *et al.* 2006) but omitting the RNA size fractionation step prior to cDNA synthesis. In brief, equal amounts of -/+ TEX treated RNA were poly(A)-tailed using poly(A) polymerase, followed by ligation of an RNA adapter to the 5´P RNA fragments. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and M-MLV-RNaseH- reverse transcriptase. Incubation temperatures were 42°C for 20 min, ramp to 55°C, followed by 55°C for 5 min. The cDNA was PCR-amplified to yield a concentration of 20-30 ng/µl using a high fidelity DNA polymerase. Each library contains a specific 4-mer barcode sequences, which is attached to the 5'end of the cDNA during PCR amplification.

cDNA cloning and pyrosequencing was performed as described before (Berezikov *et al.* 2006) but omitting size fractionation of RNA prior to cDNA synthesis. Equal amounts of total RNA were used for the generation of all cDNA libraries. For linker ligation RNA was treated with TAP to generate 5'mono-phosphates. After addition of specific 5'linkers with unique tags for each library and poly-A-tailing, the RNA was converted into a cDNA library. Four cDNA libraries were generated in total: total RNA and total RNA enriched for primary transcripts from EB and RB, respectively.

5.4.3.3 Pyrosequencing of cDNA

The libraries were sequenced on a Roche/454 GS-FLX system using the 'Amplicon A' and the 'SR70 sequencing kits' at the Max Planck Institute for Molecular Genetics in Berlin in the group of Richard Reinhardt. Sequencing raw data can be found and downloaded at the Gene Expression Omnibus (GEO) database under the accession number GSE24999 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24999) for *C. pneumoniae*. Sequencing raw data for *C. trachomatis* is available as supplemental data accompanying the paper by Albrecht *et al.* (2010) from the journal home page (http://nar.oxfordjournals.org/content/38/3/868/suppl/DC1).

5.5 Computational Biology

5.5.1 Processing of Sequencing Raw Data

From the multiplex sequencing runs the sequence reads were sorted by their specific four base barcode which were added during 5'linker ligation during cDNA synthesis. Clipping of 5'linker and poly-A-tails was performed by Cynthia M. Sharma (Department of Molecular Infection Biology, University of Würzburg). All cDNA sequences ≥ 18 nucleotides (nt) were considered for BLAST (Basic Local Alignment and Search Tool) search. The sequences were aligned to the *C. trachomatis* genome (NC_010280) and the *C. trachomatis* cryptic plasmid pL2 (NC_010285) or the *C. pneumoniae* CWL029 genome (NC_000922), respectively, using WU-BLAST 2.0 (http://blast.wustl.edu/) with the following parameters: -B=1 -V=1 -m=1 -n=-3 -Q=3 -R=3 -gspmax=1 -hspmax=1 -mformat=2 -e=0.0001.

5.5.2 Visualisation of Transcriptome Sequencing Data

For visualization of BLAST hit locations graph files were calculated and loaded into the Integrated Genome Browser (Affymetrix; http://www.affymetrix.com/partners_programs/programs/developer/tools/download_igb.affx; Version 4.56) as previously described (Tisoncik and Katze 2010). From the resulting BLAST data two graphs were calculated for every library, one for the sense and one for the antisense strand, respectively. Each graph represents the number of cDNA reads obtained from the sequencing for every nucleotide position. In all figures, graphs representing the untreated cDNA libraries are black and graphs representing the cDNA libraries enriched for TSS are coloured in red. The genome sequences and annotations were downloaded from the NCBI RefSeq FTP-server and also loaded into the IGB.

5.5.3 Analysis of Transcriptional Start Sites and Read Numbers per Gene

All annotations were made manually using the Integrated Genome Browser (IGB). By comparison of the specific pattern of the calculated sequence read distribution graphs between the untreated and enriched samples it is possible to identify the exact TSS down to single nucleotide resolution (see figures and description in results chapter). All available TSS were annotated by browsing through the genome and screening for unique TSS upstream of annotated genes. Thereby the TSS were added to the annotation table for each organism.

To quantify the gene expression and to allocate sequence reads to classes of RNA (e.g. tRNA, rRNA, mRNA) it is necessary to count sequence reads per gene. This was done by a Perl script developed by Cynthia Sharma. Using the new annotation which includes the 5'leader region (if available) for every gene, overlap numbers were counted for the sense and the antisense strand of every gene.

For quantification of gene expression transcripts with at least 20 sequences detected in total were considered for differential expression analysis. Individual sequence numbers were normalized to the total number of mRNA sequences per library. The threshold for a classification as differentially expressed was 2-fold.

5.5.4 Promoter Sequence Analysis

Sequences upstream of TSS were inspected for common promoter sequences. Genomic sequences from position -1 to -40 of every TSS were extracted as fasta file with the gene name (if available) or the gene number as identifier. Promoter motifs were searched using MEME suite (available for download at http://meme.nbcr.net/meme4_6_0/meme-download.html or with restricted range of functions as web server version at http://meme.nbcr.net/meme4_6_0/intro.html). Comparisons with published promoter sequences were done with standard text editors or BLAST search.

6 RESULTS

6.1 Characterization of the *Chlamydia trachomatis* Transcriptome by dRNA-seq

6.1.1 Purification of *Chlamydia trachomatis* EB and RB

Previous studies on transcriptomics and gene expression of *Chlamydia* have been performed with total RNA from infected host cell from different time points of infections. The infection starts to become asynchronous after several hours, depending on cell type, infection conditions and other parameters. Thus, later time points contain a mixture of late RB that are not re-differentiated and early EB that are derived from re-differentiated RB. To gain insights into the transcriptome of EB and RB these distinct life cycle forms were separated based on their different buoyant density by differential gradient centrifugation. Several different gradient systems have been tested and a self-forming Percoll gradient in SPG buffer gave the best results. Two distinct bands were visible in the gradient. These were collected, washed and pelleted. To proof the content of these two fractions, they were checked by electron micrographs. Figure 3 shows purified fractions from the gradient separation containing EB (Figure 3A) and RB (Figure 3B) marked by arrowheads. EB have highly condensed DNA, no metabolic activity and contain less water than RB. Therefore they have a higher density than most of the other cell components and organelles and are relatively pure. RB have similar density to cell organelles like mitochondria. Figure 3B clearly shows impurities of the RB fraction. To avoid longer processing times of the *Chlamydia* and consequently changes in the RNA content of the bacteria, especially RNA degradation, no additional purification step was added.

6.1.2 Deep Sequencing of Purified *Chlamydia trachomatis* EB and RB

In order to obtain a comprehensive image of the total transcriptome of *C. trachomatis*, cDNA libraries of total RNA were analysed by deep sequencing analysis. To allow detection and quantification of transcripts which differ between EB and RB, these developmentally distinct forms of *C. trachomatis* were purified, RNA was isolated and EB and RB cDNA libraries were generated. To identify primary transcriptional start sites, we sequenced two libraries for each growth form: one generated from the original, untreated total RNA and the other following enrichment of primary transcripts by selective enzymatic degradation of processed RNA species. The resulting four different libraries, derived from total RNA and RNA enriched for primary transcripts of EB and RB, were then subjected to deep sequencing to achieve a semi-quantitative comparison of gene expression. In total, 338,678 sequences were obtained from the four cDNA libraries. After removal of 5'-linkers and poly-A tail sequences, only reads with a length of at least 18 nt (309,695 reads; 91.44% of total reads) were considered for further analysis. Clipped read lengths up to 120 nt with 67.9% longer than 60 nt were obtained with similar read length distribution in all libraries. Most reads had a length of 60 to 90 nt (Figure 4A).

All reads were then aligned to the *C. trachomatis* genome (NC_010280) or the *C. trachomatis* cryptic plasmid (NC_010285) using the WU-BLAST 2.0 algorithm. In total, 263,949 sequences could be mapped to the genome (249,490 reads, 94.5%) and the plasmid (14,459 sequence reads, 5.5%). The RB libraries contained a large fraction of sequences which were neither found on the chlamydial genome nor on the plasmid. Further analyses unveiled, that these sequences were derived from contaminating human host cell mitochondria which apparently co-purified with RB during gradient purification. Table 1 gives exact read numbers obtained for every cDNA library, total BLAST hits and the number of sequence reads that could not be mapped to the genome and plasmid sequences.

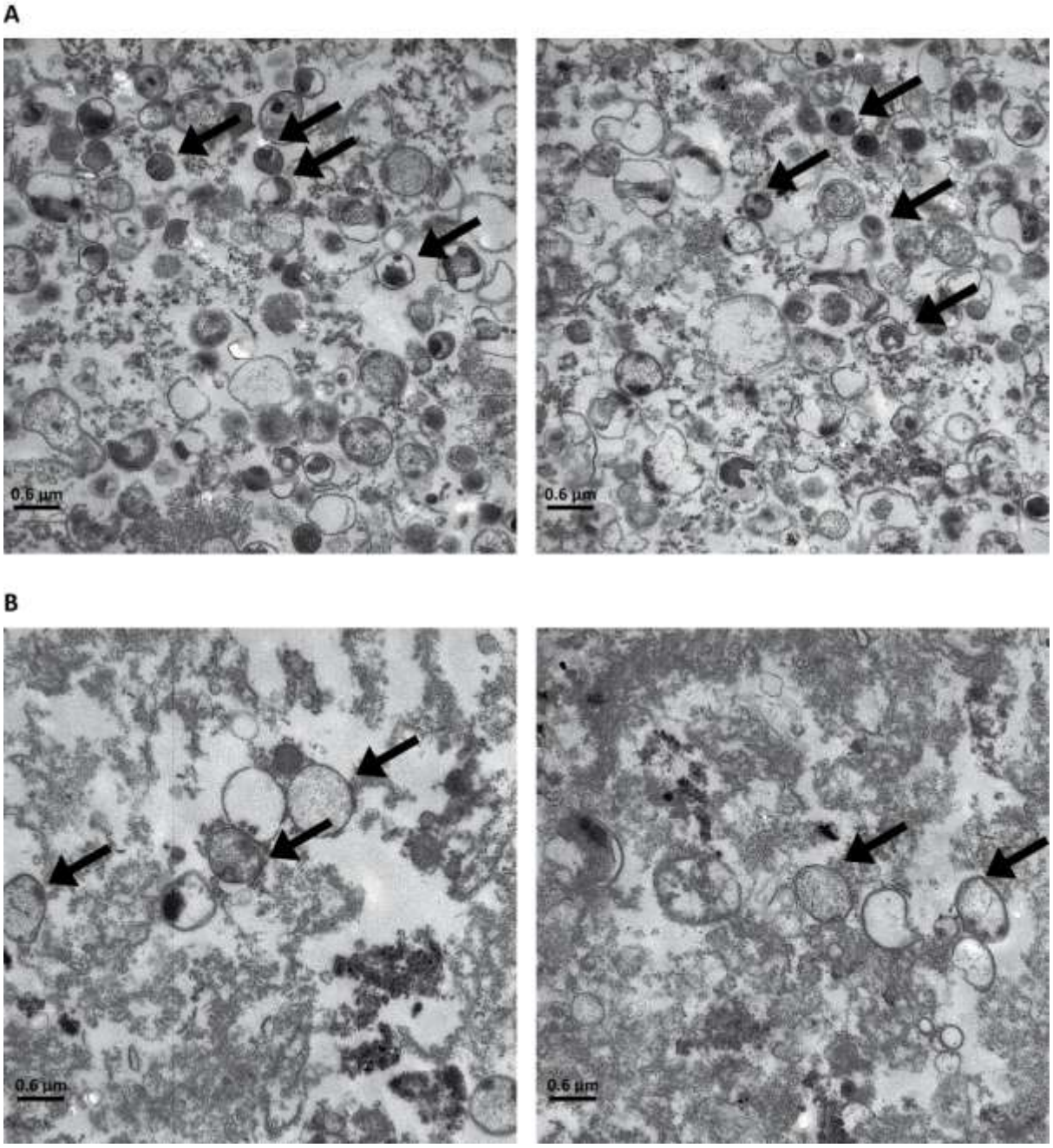

Figure 3: Electron micrographs of purified *C. trachomatis* EB and RB.
Bacterial pellets from gradient centrifugation were analyzed by electron microscopy to determine the purity of EB (A) and RB (B). The purification protocol was optimized to avoid contamination of the larger and less stable RB with the small condensed and mechanically stable EB. Arrowheads mark chlamydial particles.

Table 1: Numbers of sequence reads of the four *C. trachomatis* cDNA libraries.
All cDNA sequences ≥ 18 nt were aligned against the *C. trachomatis* genome and the plasmid. cDNAs that gave no BLAST hit are derived from contaminating human host RNA.

Library	Linker tag	Inserts	Inserts < 18 nt	Inserts ≥ 18nt	Total BLAST hits	Hits in genome	Hits in plasmid	Inserts ≥ 18nt but no BLAST hit
CTRB -	GCTC	73,784	9,065	64,719	47,546	42,517	5,029	22,202
CTRB +	GTAT	79,984	4,469	75,515	53,863	46,039	7,824	29,476
CTEB -	GTGA	86,157	8,498	77,659	74,538	74,113	425	3,546
CTEB +	GTTC	98,753	6,951	91,802	88,002	86,821	1,181	4,981
Total		338,678	28,983	309,695	263,949	249,490	14,459	60,205

Sequencing read numbers that could be mapped to either the genome or the plasmid ranged from 73,784 to 98,753 per library and were sorted by classes of RNA as shown in figure 4B. Plasmid reads represented 8.4% of the RB and 0.9% of the EB libraries, respectively. This suggests a stronger transcriptional activity of the cryptic plasmid in RB than in EB. Unexpectedly, the fraction of tRNA and rRNA reads is larger in the cDNA libraries enriched for primary transcripts. Since both, tRNA and rRNA are processed from a primary transcript by ribonucleases these molecules contain a 5'monophosphate and should be accessible to degradation by Terminator exonuclease. Their thermodynamically very stable secondary structure hinders the endonuclease from digesting the RNA. However, the fraction of mRNA reads was reduced significantly by the enrichment for primary transcripts since processing and degradation products were removed by the enzymatic enrichment procedure. Many mRNAs are several kb in length and contain multiple genes. They are processed by endonucleases and degraded by endo- and exonucleases. The total RNA libraries contain all these mRNA-derived fragments, whereas the enriched libraries contain only mRNA fragments with a native 5'end (see figure 4B). Sequence reads located in intergenic regions (IGR, figure 4B) can be part of 5' and 3'untranslated regions or derived from transcripts (e. G. Non-coding RNA) that are missing in the

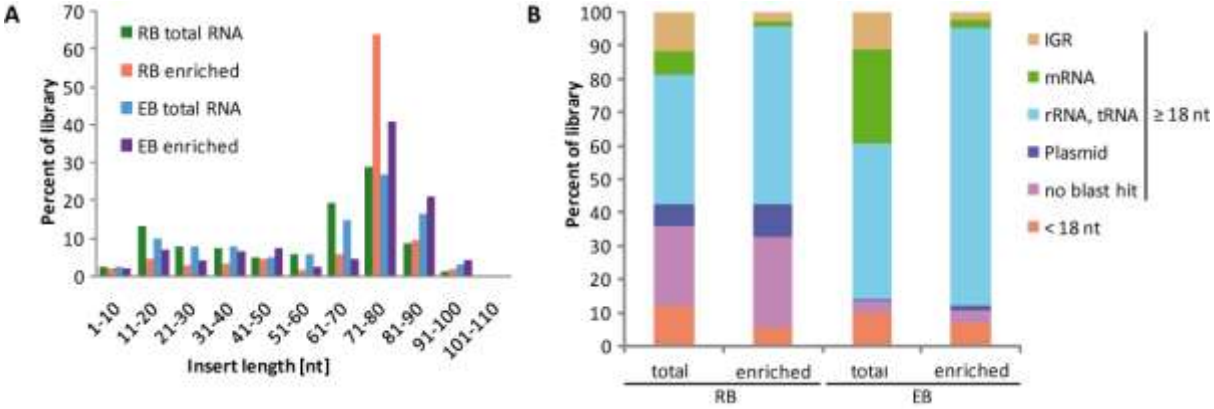

Figure 4: Characterization of the *C. trachomatis* cDNA libraries.
(A) Length distribution of reads after 5'end-linker and polyA-tail clipping of four sequenced *C. trachomatis* cDNA libraries generated from either total RNA or total RNA enriched for primary transcripts of reticulate bodies (RB) or elementary bodies (EB), respectively. Shown are relative numbers of groups of sequence length in relation to the total number of reads per library. (B) Sequence read distribution of the four cDNA libraries were grouped into different classes of RNAs. Transcripts antisense to ribosomal RNAs, transfer RNAs and messenger RNAs are not shown since they represent less than one percent of total sequences. Transcripts located in intergenic regions (IGR) can either be sRNA candidates, part of 5' or 3'UTRs of mRNAs or unannotated coding genes. The majority of the fraction of reads that could not be mapped to the chlamydial genome corresponds to contaminating host cell RNA, mainly mitochondrial RNAs.

annotation. Reads that are located antisense to annotated genes are not displayed in figure 4B since they account for less than 1% of the total read numbers of each library.

6.1.3 Annotation of Transcriptional Start Sites (TSS)

The above mentioned treatment of RNA enriches for sequence reads whose 5'end is a TSS. To globally map the TSS of the *C. trachomatis* transcriptome, sequence reads along the whole genome were visualized by calculation of graph files displayed by the Integrated Genome Browser (IGB). The graph files confer to bar diagrams of the whole genome, where the x-axis displays the nucleotide sequence and the y-axis confers to the number of sequence reads for the specific nucleotide position. All graphs of this work containing figures based on the IGB visualization display graphs of untreated libraries in black and those of enriched libraries in red. Figure 5 shows an example of a distinct TSS of the *inc* operon. The enriched libraries (RB+, EB+) contain sequence reads that all start at the same position, whereas the untreated libraries (RB-, EB-) contain reads distributed over the whole operon. This leads to the conclusion that the *inc* genes are co-transcribed from one distinct TSS indicated by the arrow in figure 5. It remains unclear, whether *incG* is also contained in the transcript or not transcribed at all.

TSS were manually annotated by inspection of sequenced regions upstream of open reading frames (ORFs) and also in intergenic regions and antisense to annotated genes. Sequencing reads starting at exactly the same nucleotide position defined the TSS. The distance between the newly annotated TSS and the translational start site defines the leader sequences (5'untranslated region).

This precise mapping of the TSS unveiled a broad variation in the length of the 5'untranslated regions (UTRs) with a peak at 21–40 nt and a number of 5'UTRs ranging over 100 nt in length (shown in figure 6). Three transcripts belong to the group of leaderless mRNAs since the TSS equals the translation start (CTLon_464, 600, 684) (Moll *et al.* 2002). Further investigation and probably re-

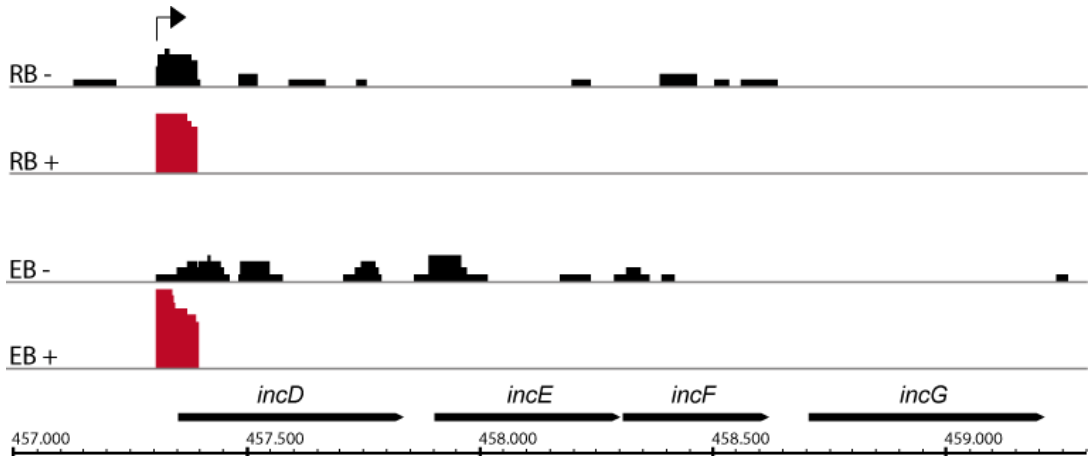

Figure 5: Visualization of pyrosequencing data of the *C. trachomatis inc*-operon.
Sequencing reads are shown as bar diagrams for all four C. trachomatis cDNA libraries. Red bars represent reads from TEX treated samples (RB + and EB +), whereas black bars represent untreated samples (RB - and EB -). In the untreated samples sequence reads were obtained that mapped to three of the four *inc* genes. In the treated samples all reads start from the same position (arrow) which corresponds to the TSS. This combined analysis of treated and untreated defines the TSS and suggests a polycistronic transcript that encodes *incD-G*. Note that the figure shows only the plus strand, since no reads were mapped against the minus strand and it does not contain annotated genes.

annotation is needed for five transcripts where the TSS is located downstream of the annotated translation start (CTLon_0311, 0537, 0742, 0755, 0757). These genes contain an alternative ORF downstream of the TSS which is supported by a putative ribosome binding site. For several genes two distinct TSS could be identified which indicates a differential regulation of gene expression by variation in leader length. These alternative TSS are listed in table 10 (Appendix) together with primary TSS.

In total, sequence reads could be detected for 548 out of 934 annotated genes (58.7%). Thereof TSS could be identified for 356 (65.0%) annotated genes located on the bacterial chromosome comprising 317 putative protein coding genes and 39 tRNA and rRNA genes. Table 10 (Appendix) gives a detailed list of all TSS identified. For the remaining expressed genes, TSS could not be specified since RNAs were not abundant enough to be identified or expressed as a polycistronic transcript. Furthermore, 16 distinct TSS could be identified in intergenic regions which lack an ORF and therefore indicate non-coding transcripts. In addition, 25 TSS representing small non-coding transcripts that are either partially or completely antisense to annotated ORFs (Table 2) could be identified.

Several transcripts are located partially or fully antisense to annotated genes. Figure 7 displays graphs of the sum of RB and EB libraries (untreated libraries in black overlaid with enriched libraries in red) and sequence reads on both strands including annotated genes. Most of them represent long 3' UTRs of annotated genes that overlap adjacent genes on the opposite strand as shown in figures 7A, B, E, H and I. Figure 7C displays long transcripts containing the tRNA:Leucine and an abundant 3' fragment of several hundred kb that overlaps the gene *xerD* on the opposite strand. A transcript antisense to gene CTLon_0243 coding for a hypothetical protein and the polymorphic outer membrane

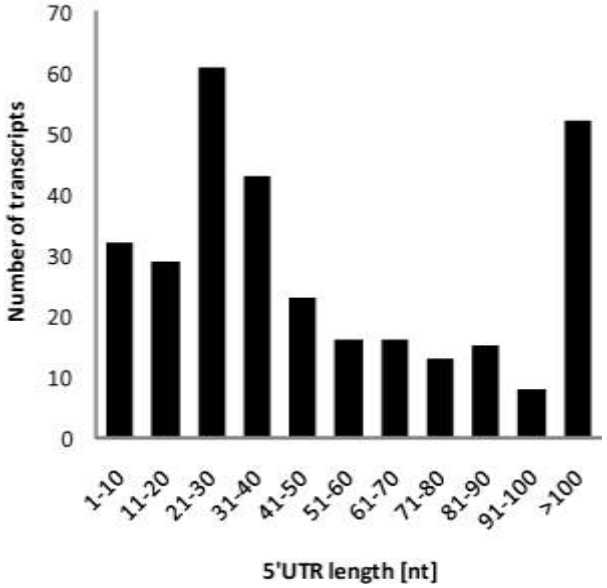

Figure 6: Length distribution of *C. trachomatis* 5'UTRs.
The distance between manually determined transcription start sites and the annotated translation starts was defined as 5'untranslated region and plotted over the number of transcripts. The length of the 5'UTRs ranges up to >100 nucleotides with a peak at 21-30 nucleotides.

protein gene *pmpE* is shown in figure 7D.

Two operons have been identified, that contain a gap where a gene on the opposite strand is annotated. The first operon encoding the genes CTLon_0428-0423 and CTLon_0425 is located on the opposite strand (Figure 7F).

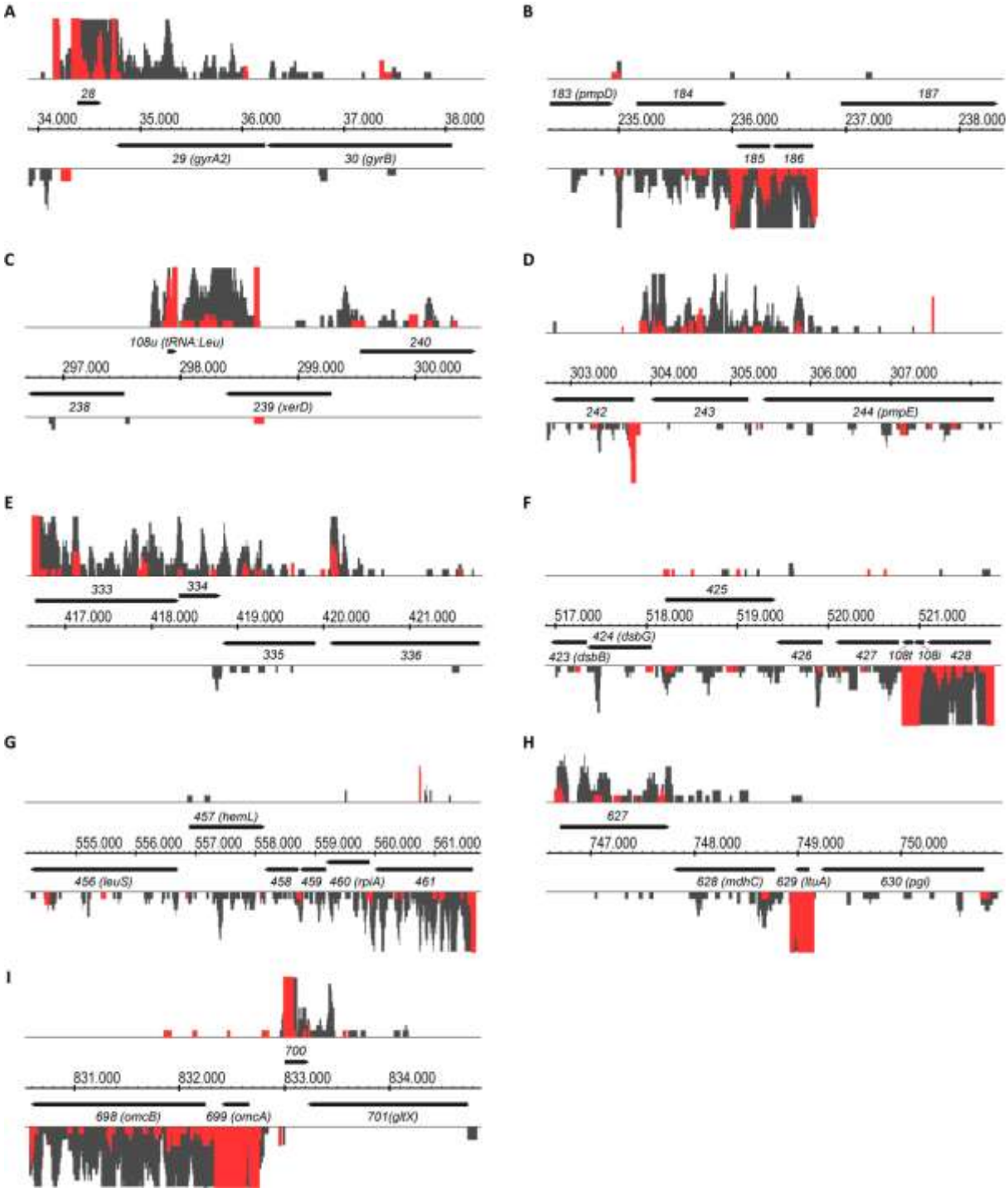

Figure 7: Antisense RNAs derived from long overlapping 3' UTRs in *C. trachomatis*.
Deep sequencing revealed nine long 3' UTRs which in part or completely overlap the gene located downstream on the opposite strand. Transcriptional activity is stronger on the antisense strand. These transcripts are long 3' UTRs overlapping the antisense downstream genes (A, B,C, E, H, I), operon transcripts containing an antisense ORF (F, G) or transcripts which are not annotated yet and located on the opposite strand of annotated ORFs (D).

The second operon containing the genes CTLon_0461-0456 and CTLon_0457 is located on the opposite strand. Since only very few sequence reads were obtained for the two genes annotated opposite to the two operons it remains unclear whether these genes are correctly annotated.

6.1.4 Discovery of Putative Small Non-Coding RNAs of *Chlamydia trachomatis*

Besides reads mapping to the annotated region of the genome a significant number of transcripts that could be mapped to intergenic regions was found. Furthermore, short abundant transcripts could be identified that are partially or completely antisense to annotated ORFs. Based on manual inspection for such transcripts with a defined primary transcription start site, 16 putative small RNAs located in intergenic regions as well as 25 small antisense RNAs could be detected from the sequencing data.

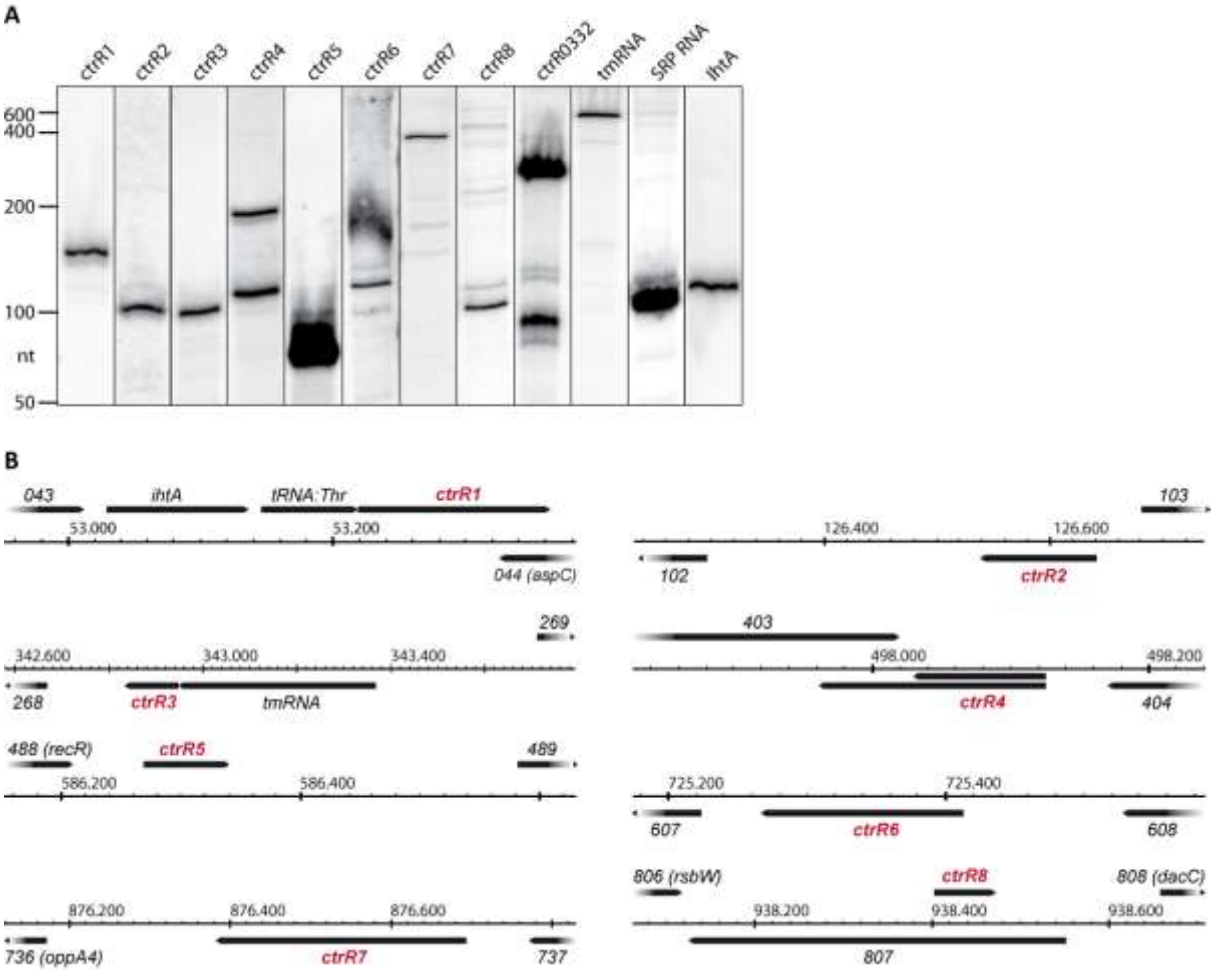

Figure 8: Validation of putative small RNAs of *C. trachomatis* by Northern Analysis.
(A) Expression of eight out of twelve new candidate sRNAs could be confirmed by Northern blotting whereas the length of the probed RNA corresponded to the calculated length from the sequencing data. The genomic location of the sRNAs is shown in (B). ctrR1 is not a primary transcript but presumably processed from a larger transcript containing tRNA:Thr. ctrR2, 3, 5-7 are located intergenic and primary transcripts. Two sRNAs are transcribed from ctrR4 and the longer transcript overlaps the gene 403 on the opposite strand. ctrR8 is located antisense to gene 807. ctrR0332 has previously been identified as ORF CTLon_0332, but represents two RNA species processed from a larger transcript lacking an ORF. Details on the genomic location and the sequence of ctrR0332 are given in supplementary figure S2. Two housekeeping RNAs (tmRNA & SRP RNA) and ihtA is the only so far reported chlamydial sRNA are shown in the right lanes. Genomic locations, sequence read numbers and hybridization probes are given in Table S2.

Table 2: Genomic Locations of putative non-coding RNAs and corresponding Northern probes.
Transcripts encoding putative and validated non-coding small RNAs are listed. Transcripts which were probed and detected by Northern hybridization are named CtrR1-CtrR8. Corresponding hybridization probes, hypothetical sizes calculated from the transcriptional start sites and ends as well as sizes detected in Northern hybridization are listed. The number of sequence reads from the cDNA libraries are given for RB and EB (n.d. = not detected).

	Start	End	Length	Strand	Gene	Northern signal	Northern probe	RB read number	EB read number
Processed transcripts									
ctrR1	53220	53363	144	+		140 nt	GAGGGGACGAAGGAATGGTTACTT	136	227
Primary transcripts located in intergenic regions									
	16349	16475	127	+				3	0
	16987	16891	97	-				3	0
ihtA	53029	53137	109	+		120 nt	GAGTCAGAAGCTATTCCATGGCGT	77	17
	64860	64946	87	+		n.d.	GATGTCTCTCCCTAGACCCAAGGA	30	12
ctrR2	126638	126536	103	-		100 nt	TGGGATGTTGAAGCGCCTTTATAA	1	2
	140853	140895	43	+				6	0
	214321	214403	83	+				2	0
	217101	217187	87	+		n.d.	GATGAACGGTCCTACGCGCTATAC	8	1
	220744	220679	66	-		n.d.	TGTGTGTTGCATATTCCAGCTCCT	11	2
ctrR3	342947	342838	110	-		100 nt	ACTCGTTATTTGGGAGGACACGAA	981	1761
tmRNA	343368	342948	421	-		500 nt	TCTAAGTCGAAACCTTTACACCCCC	310	823
	503748	503645	104	-		n.d.	TCGAACTATTTTAGTTTTGTTGTTCGC	2	1
ctrR5	586269	586341	73	+		80 nt	GGGTGCTGGTCACTAGCAAAAACT	8	3
SRP RNA	605142	605034	109	-		110 nt	GACCCGGTTCAGAAGTCTCTCATC	110	450
	681933	681813	121	-				2	2
ctrR6	725411	725268	144	-		120 nt	AAAAGCTATAGGGGAGGAGGGACC	75	19
	727681	727795	115	+				0	2
RNAseP RNA	775621	776032	412	+				55	4
	842374	842276	99	-				9	0
ctrR7	876694	876384	311	-		310 nt	GAAGCGGATCTAAGCTGTGTAAGC	71	15
Primary antisense/overlapping transcripts									
	34146	34257	112	+	CTLon_0027			22	0
	34316	34221	96	-	CTLon_0027			1	2
	72387	72249	139	-	CTLon_0059			36	5
	76386	76309	78	-	CTLon_0063			2	0
	97543	97475	69	-	CTLon_0077			2	0
	153722	153794	73	+	CTLon_0118			2	0
	360777	360842	66	+	CTLon_0283			5	0
	416601	416376	226	-	CTLon_0332			4	12
	429993	429917	77	-	CTLon_0343			7	2
ctrR4	498126	497963	164	-	CTLon_0403	110 nt + 180 nt	AGACACGGTCACCCCTAACTACCA	19	3
	500595	500676	82	+	CTLon_0404			4	1
	504613	504683	71	+	CTLon_0407			4	0
	560748	560770	23	+	CTLon_0461			3	8
	606237	606170	68	-	CTLon_0504			4	0
	620148	620389	242	+	CTLon_0518			7	0
	634134	634212	79	+	CTLon_0532			0	2
	648911	648987	77	+	CTLon_0544			4	0
	845728	845653	76	-	CTLon_0709			3	0
	910482	910558	77	+	CTLon_0777			3	1
ctrR8	938403	938509	107	+	CTLon_0807	110 nt	ATCTGATGTCTTAGCTCGTTCGGC	59	1
	967633	967559	75	-	CTLon_0833			1	2
	977322	977256	67	-	CTLon_0842			2	0
	987138	987170	33	+	CTLon_0847			3	2
	999992	999882	111	-	CTLon_0859			5	1
	1035912	1035954	43	+	CTLon_0888			7	2

In addition, one abundant processed RNA derived from a tRNA containing precursor could be identified (ctrR1, figure 8). The sRNA candidates included ihtA, the only previously identified chlamydial sRNA (Grieshaber *et al.* 2006a). The housekeeping RNAs trans-messenger RNA

Table 3: Putative targets of validated *C. trachomatis* sRNAs.
Based on binding probability of the newly identified small non-coding RNAs putative binding partners were identified using the TargetRNA algorithm (Tjaden *et al.* 2006). Hits were filtered for binding of the sRNA to the translation start codon or the ribosome binding site. None of the sRNAs is conserved in other bacteria than Chlamydiae. However some seed sequences are conserved among at least two other bacterial phyla. Targets putatively bound by these conserved sequence elements are underlined. Putative target mRNAs are listed in order of their binding score starting with the highest score.

sRNA	Start	End	Length	RB read number	EB read number	Putative target mRNAs
ihtA	53029	53137	109	77	17	hctA, 317, 097, mgtE
ctrR1	53220	53363	144	136	227	819, 370, 505
ctrR2	126638	126536	103	1	2	492, tufA, nlpD, 753, 443, dnaA, 035, 730, secG
ctrR3	342947	342838	110	981	1,761	248, 794, 190, pfrA, 218, 394, mip, rodA, 294, rbfA, def, ptsN_2, sctU
ctrR4	498126	497963	164	19	3	xerC, pknD, 626
ctrR5	586269	586341	73	8	3	dacC, gnd
ctrR6	725411	725268	144	75	19	824, rplL, 854, kdsA, 751, secG, ppa, 219
ctrR7	876694	876384	311	71	15	copB, 847, glnQ, 531, ssc2, 641, 623, 652, lpdA, ihfA, 597, tig, nusG, aroA
ctrR8	938403	938509	107	59	1	497, murE, eno, 080, 511, 013, 316, dnaQ
ctrR0332	416302	416641	340	1,039	16,960	333, atpK, 292, 586, gnd, 822, glgC
pL2-sRNA1	213	292	90	8,332	1,075	305, sucA, dnaG

(tmRNA), signal recognition particle RNA (SRP RNA), and RNaseP RNA were also present in the sequencing data. Table 2 lists all novel RNA candidates, their genomic coordinates and lengths.

To test whether these transcripts reflected small RNAs, total RNA from an EB-RB-mixture was analysed by Northern blot hybridization to check for the presence and the corresponding sizes of sRNA candidates. Oligonucleotide probes used for hybridization and the resulting sizes of the bands are listed in table 2. Expression of 9 out of 12 tested novel sRNAs was successfully validated by Northern blot analysis (Figure 8A), arguing for the presence of multiple sRNAs in *C. trachomatis*. Figure 8B shows the genomic location of these novel sRNAs. Six sRNAs (ctrR1–3, 5–7) are located in intergenic regions. ctrR4 encodes two sRNAs and the larger transcript overlaps the antisense located gene CTLon_0403 and ctrR8 is located antisense to gene CTLon_0807. Most of these sRNAs show a stable secondary structure when analysed by the RNAfold algorithm (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Table 2 lists the sRNA candidates that could be positively validated by Northern blot hybridization. The genomic coordinates and the length are given as well as the sequence read numbers in the EB and RB cDNA libraries. The abundances of the sRNAs vary in a wide range from only 3 reads (ctrR2) to several thousand reads (ctR3). The highly abundant RNA ctrR0332 and the plasmid encoded sRNA pL2-sRNA1 will be described in chapters 6.1.5 and 6.1.6, respectively.

In recent years, numerous small RNAs have been identified in bacteria and most of them act by direct interaction with their target mRNAs by base pairing to posttranscriptionally regulate gene expression (Waters and Storz 2009). To date, no method to functionally analyse the sRNA candidates *in vivo* in *Chlamydia* is available. An *in silico* search for putative targets regulated by the novel sRNAs was performed using the TargetRNA algorithm (Tjaden *et al.* 2006). This software calculates base-pairing probabilities of a given sRNA and all annotated mRNAs of a certain genome. The results were filtered for sRNA:target interactions incorporating the translation start codon or a putative ribosome binding site located upstream of the start codon. Candidate targets are listed in table 3 in the order of

Table 4: List of differentially expressed genes in *C trachomatis* EB and RB.
A semi-quantitaive analysis of differentially expressed genes returned 84 protein coding genes, 38 overrepresented in EB (A) and 46 more abundant in RB (B, next page). Sequence read numbers were counted for each ORF including the 5′-UTR if present and normalized by the total number of mRNA transcripts for each library. Genes with a total number of at least 20 sequence reads and a 2-fold regulation were considered as differentially expressed. A comparison to a microarray based gene expression study (Nicholson *et al.* 2003) shows a high correlation of genes abundant in the late infection cycle phase to EB and genes abundant in the midlate phases to RB.

A *Elementary Bodies*

Gene	*Description/Gene Product*	*Gene No.*	*Homo-logue*	*Micro-array*
	1-acyl-sn-glycerol-3-phosphate acyltransferase	CTLon_0144	CT775	late
omcB	60kD cysteine-rich outer membrane protein	CTLon_0698	CT443	late
	ABC transporter, ATP-binding protein	CTLon_0427	CT180	-
crpA	cysteine-rich membrane protein	CTLon_0697	CT442	late
omcA	cysteine-rich outer membrane protein	CTLon_0699	CT444	late
	hypothetical protein	CTLon_0186	CT814.1	late
	hypothetical protein	CTLon_0332	CT081	late
	hypothetical protein	CTLon_0250	CT875	late
	hypothetical protein	CTLon_0333	CT082	late
	hypothetical protein	CTLon_0880	CT622	late
	hypothetical protein	CTLon_0834	CT577	late
	hypothetical protein	CTLon_0242	CT867	-
	hypothetical protein	CTLon_0185	CT814	late
	hypothetical protein	CTLon_0428	CT181	late
	hypothetical protein	CTLon_0609	CT357	-
	hypothetical protein	CTLon_0255	CT005	late
	hypothetical protein	CTLon_0608	CT356	late
	hypothetical protein	CTLon_0461	CT214	late
	hypothetical protein	CTLon_0240	CT865	-
	hypothetical protein	CTLon_0334	CT083	late
	hypothetical protein	CTLon_0536	CT288	late
	hypothetical protein	CTLon_0477	CT229	-
	hypothetical protein	CTLon_0028	CT659	late
ltuA	late transcription unit A protein	CTLon_0629	CT377	midlate I
ltuB	late transcription unit B protein	CTLon_0331	CT080	late
aas	long chain fatty acid--[acyl-carrier-protein] ligase	CTLon_0145	CT776	late
mdhC	malate dehydrogenase	CTLon_0628	CT376	midlate I
	pseudogen	CTLon_0548	CT300	-
	pseudogen	CTLon_0610	CT358	-
	putative integral membrane protein	CTLon_0398	CT147	very late
	putative integral membrane protein	CTLon_0617	CT365	late
	putative lipoprotein	CTLon_0700	CT444.1	-
	putative oxidoreductase	CTLon_0627	CT375	late
	putative protein ligase	CTLon_0285	CT035	very late
copB	putative type III secretion system membrane prote	CTLon_0835	CT578	late
copD	putative type III secretion system protein	CTLon_0836	CT579	late
rpiA	ribose-5-phosphate isomerase A	CTLon_0460	CT213	late
scc2	type III secretion chaperone (low calcium resp. p. H	CTLon_0833	CT576	late

B ***Reticulate Bodies***

Gene	*Description/Gene Product*	*Gene No.*	*Homo-logue*	*Micro-array*
rplM	50S ribosomal protein L13	CTLon_0376	CT125	-
dppD	ABC transport protein, ATPase component	CTLon_0059	CT690	-
clpC	ATP-dependent Clp protease	CTLon_0534	CT286	midlate II
pgsA	CDP-diacylglycerol--glycerol-3-phosphate 3-phosph...	CTLon_0752	CT496	midlate I
ftsH	cell division protein	CTLon_0213	CT841	late
cydA	cytochrome d ubiquinol oxidase subunit I	CTLon_0263	CT013	midlate I
hct2	histone-like protein 2	CTLon_0297	CT046	late
ruvB	Holliday junction DNA helicase B	CTLon_0291	CT040	
	hypothetical protein	CTLon_0267	CT017	late
	hypothetical protein	CTLon_0859	CT602	-
	hypothetical protein	CTLon_0251	CT001	late
	hypothetical protein	CTLon_0624	CT372	midlate I
	hypothetical protein	CTLon_0152	CT783	midlate II
	hypothetical protein	CTLon_0537	CT289	midlate I
	hypothetical protein	CTLon_0876	CT618	midlate I
	hypothetical protein	CTLon_0572	CT324	-
	hypothetical protein	CTLon_0521	CT273	midlate I
euo	hypothetical protein	CTLon_0702	CT446	-
	hypothetical protein	CTLon_0607	CT355	-
	hypothetical protein	CTLon_0879	CT621	midlate I
	hypothetical protein	CTLon_0635	CT382.1	-
ihfA	integration host factor alpha-subunit	CTLon_0515	CT267	midlate I
copN	low calcium response protein E (TTSS effector protein)	CTLon_0340	CT089	midlate I
ompA	major outer membrane protein	CTLon_0050	CT681	midlate II
nqrA	Na(+)-translocating NADH-quinone reductase subunit A	CTLon_0002	CT634	midlate I
ndk	nucleoside diphosphate kinase	CTLon_0757	CT500	midlate I
pmpC	polymorphic outer membrane protein	CTLon_0667	CT414	midlate I
pmpF	polymorphic outer membrane protein	CTLon_0245	CT870	midlate I
pmpG	polymorphic outer membrane protein	CTLon_0246	CT871	midlate I
pmpH	polymorphic outer membrane protein	CTLon_0247	CT872	midlate I
pmpE	polymorphic outer membrane protein	CTLon_0244	CT869	late
pmpB	polymorphic outer membrane protein	CTLon_0666	CT413	midlate I
pmpI	polymorphic outer membrane protein	CTLon_0249	CT874	midlate I
	putative cation efflux protein	CTLon_0678	CT423	-
	putative DNA methyltransferase	CTLon_0733	CT477	-
	putative helicase	CTLon_0077	CT708	midlate I
	putative lipoprotein	CTLon_0501	CT253	-
pkn5	putative serine/threonine-protein kinase (TTSS effector	CTLon_0042	CT673	-
uhpC	putative sugar phosphate permease	CTLon_0801	CT544	midlate I
	putative type III secretion system chaperone	CTLon_0294	CT043	midlate I
nrdA	ribonucleotide-diphosphate reductase subunit alpha	CTLon_0199	CT827	midlate I
sfhB	ribosomal large subunit pseudogenuridine synthase D	CTLon_0027	CT658	-
tal	transaldolase B	CTLon_0561	CT313	midlate I
	type III secretion structural protein (outer membrane ri	CTLon_0043	CT674	midlate I
sctJ	type III secretion system protein, membrane componen	CTLon_0816	CT559	midlate I
	tyrosine-specific transport protein	CTLon_0190	CT818	-

their binding score, starting with the highest. Most targets have at least a 7 to 9 nt seed sequence. None of the newly identified sRNA is conserved within different species other than *Chlamydiae*. However, some motifs of ctrR2, -3, -6, -8 and the plasmid encoded sRNA pL2-sRNA1 are conserved among other species. Targets which are putatively bound by the newly identified sRNAs are underlined in table 3.

6.1.5 Differential Gene Expression in EB and RB

The *C. trachomatis* L2b/UCH-1/proctitis annotation contains 874 putative ORFs. A comparison of relative read numbers of a given RNA between EB and RB indicated a set of

differentially expressed genes in the two chlamydial developmental cycle stages (listed in table 3). The separate sequencing of EB and RB transcripts revealed 38 genes which are more abundant in EB and could be crucial for host cell invasion and infectivity. This subset includes 18 hypothetical proteins of unknown function. The 46 genes which were found to be upregulated in RB include outer membrane proteins and genes involved in bacterial division. Twelve genes are of unknown function.

To validate differential gene expression, the relative abundances of RNAs from the different chlamydial developmental stages were compared with published microarray expression data. Nicholson *et al.* (2003) classified clusters of temporal gene expression with 'early' (12 h p.i.), 'mid-late' (18 h p.i.), 'late' (24 h p.i.) and 'very late' (36 h p.i.). EB are metabolically inactive, and transcripts present in EB assure immediate translation in early infection events. Such transcripts are produced in the RB stage and stored in the EB, and therefore accumulate during the end of the infection cycle when the RB-to-EB transition is completed. These genes correlate with 'late' and 'very late' gene expression from microarray studies using RNA from time course experiments (table 4). Most of the transcripts overrepresented in EB are of unknown function with the interesting exception of the highly overrepresented RNA of an operon consisting of the four genes *CTLon_0833-0836*. This transcript encodes components of the T3SS such as the chaperone scc2, a protein of unknown function, and the ATPase copB and copD, all of which is consistent with the crucial role of the T3SS early in infection (Hefty and Stephens 2007). In the 'mid-late' stage, the RB is the predominant form of *Chlamydia*. Of the 48 genes that were found to be up-regulated in RB, 27 were classified as 'mid-

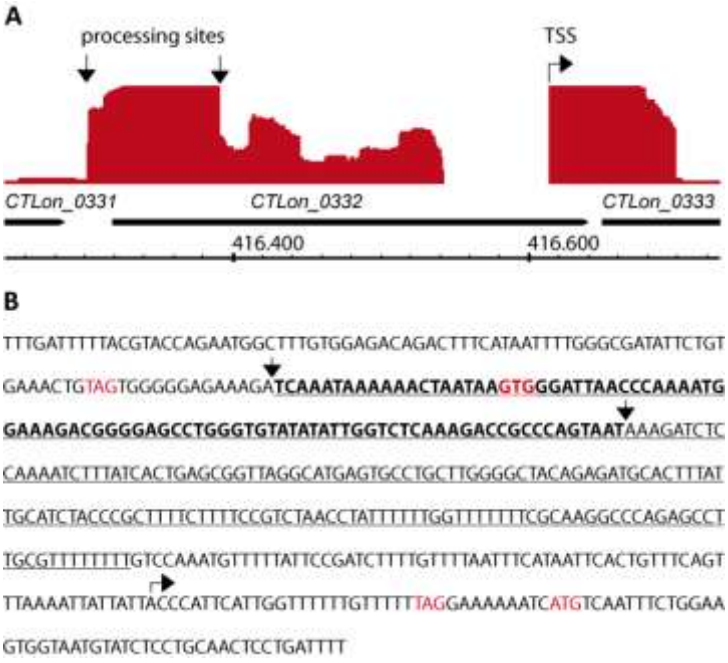

Figure 9: Sequencing read distribution of CTLon_0332.
(A) The region containing the annotated ORF CTLon_0332 including the 5'UTR that contains the highest number of sequencing reads (17,999) of all annotated ORFs. Two small RNAs of approx. 80 and 250 nucleotides in size were detected by Northern hybridization. These transcripts are shorter than the annotated ORF and contain no alternative ORF. Therefore this transcript is likely a non-coding RNA and the gene is renamed in ctrR0332. (B) Sequence of the genomic region shown in A). Processing sites of the abundant small RNA are marked by horizontal arrows. The sequence of the small RNA species is displayed in bold and underlined and the larger RNA is underlined. Red codons mark the start and the stop of the annotated ORFs CTLon_0331, 0332 and 0333, respectively.

late' and 5 as 'late' genes (Table 4). Arguably, the use of different platforms and the temporal spectrum of RB-to-EB conversion limit a direct comparison of expression data, i.e. obtained by microarrays versus direct transcriptome sequencing of purified bacteria. Nevertheless data from deep-sequencing correlated very well with the previous microarray analyses of infected cells.

One interesting finding was the most abundant transcript accounting for 20% and 78% of all putative mRNA sequence reads in the RB and EB libraries, respectively. Further analysis of this transcript revealed its localization inside a previously annotated ORF CTLon_0332 (homolog to CT081). The RNA was 97 nt shorter than predicted for the annotated ORF and started at the –19 position from the annotated translation start (Figure 9). The sequencing profile indicated processing from a larger transcript rather than a defined TSS. Northern blot analysis revealed two highly abundant RNA species of ~80 and 240 nt in size (Figure 8A, ctrR0332) which fits very well to the 89 and 242 nt calculated from the sequencing data. The longer RNA molecule does not contain an ORF. This data suggest that these transcripts encode regulatory RNAs rather than peptides.

Although the RNA sequencing data argued against the existence of ORFs in these RNAs, proteome database screens were performed with *C. trachomatis* for the presence of protein sequences possibly translated from the region covering these RNAs (Shaw *et al.* 2002; Skipp *et al.* 2005). No such proteins could be identified supporting the prediction of a non-coding RNA encoded in region 416,302–416,543.

Next, BLAST searches were performed to identify similar sequences in other bacteria. Sequences with significant homology were only present in all sequenced *C. trachomatis* strains (99–100%) and in *C. muridarium* Nigg (89%), a *Chlamydia* species isolated from mice with an average similarity of 90% to *C. trachomatis* orthologous genes (Read *et al.* 2000). These data suggest in summary, that CTLon_0332 represents a highly abundant non-coding RNA specific for human and mouse *C. trachomatis* isolates. We have therefore renamed ORF CTLon_0332 to ctrR0332 for *C. trachomatis* ncRNA0332.

6.1.6 The Chlamydial Cryptic Plasmid Encodes Abundant sRNAs

Virtually all isolates of *C. trachomatis* harbour a cryptic plasmid of 7.5 Kb (Palmer and Falkow 1986). A total of 14,459 cDNAs mapped to the chlamydial cryptic plasmid which has a size of 7.5 Kb and contains nine ORFs, eight of which on the minus strand. Figure 10A shows the distribution of cDNA reads mapped on the plasmid, including the annotated ORFs and regions of high transcriptional activity. A highly abundant (>10 000 reads) small transcript, named pL2-sRNA1, of ~ 80 nt length was identified. This sRNA is antisense to a predicted ORF coding for the putative partitioning protein PCTB_7 (pL2-07a, figure 10A). Moreover, high transcriptional activity can be detected antisense to the only predicted ORF on the plus strand, *pL2-02* (Ricci *et al.* 1993), encoding virulence plasmid integrase, pGP8-*D*. Northern blot experiments confirmed the expression of several small transcripts in 90–400 nt size range (Figure 10B, probes p4800, p4963, p5120) from this plasmid

region which lacks an ORF on the minus strand. The largest RNA species has a size of ~420 nt and is recognized by all probes. The second smallest RNA is ~220 nt long, and was detected by the probes p4963 and p5120. It has the same TSS as the longer RNA and could be derived by 3' processing of the latter end. The smaller RNA fragments, e.g. the 90 nt fragment detected by p4963 are most likely processing products of larger fragments, since no further TSS could be identified. TSS of the plasmid encoded transcripts are indicated by arrows in figure 10A including the TSS coordinates.

TSS have been identified for seven of the nine ORFs currently annotated on the plasmid. The remaining two annotated genes are expressed but lack a detectable TSS. The sequencing data suggests that pL2-04 and -05 as well as pL2-07a and -07 are transcribed from a common transcript, respectively. Interestingly, two distinct TSS could be identified for the pL2-02 transcript whereas the transcript starting at –240 nt is enriched in RB and the transcript starting at –58 nt is enriched in EB. This suggests a transcriptional regulation by alternative promoters depending on the developmental cycle phase.

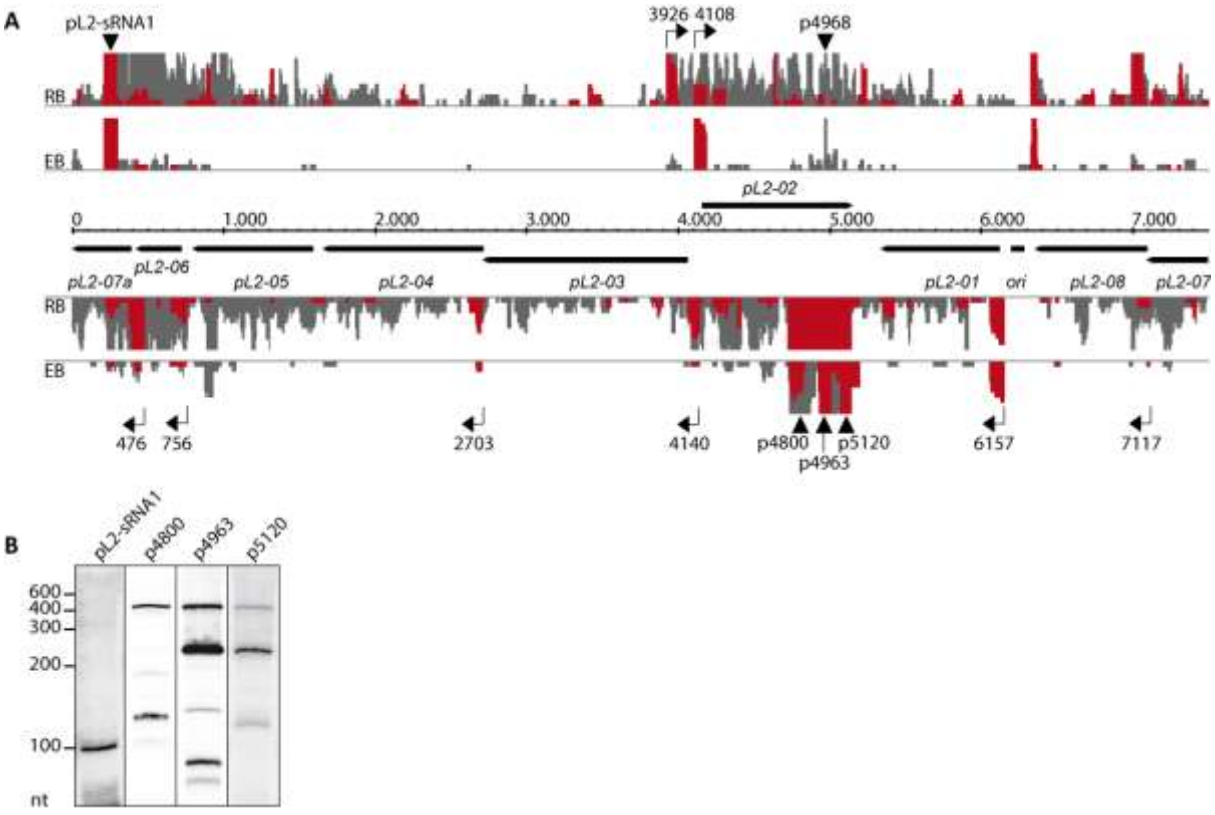

Figure 10: Transcriptome of the *C. trachomatis* cryptic plasmid pL2.
(A) Calculated sequence graphs for *C. trachomatis* cryptic plasmid. Graphs show the number of sequence reads for every nucleotide of the plasmid up to the cut-off value of ten. Total RNA library reads are shown in grey and are overlaid by red graphs for the TSS enriched libraries. Annotated ORFs are shown as black bars for the plus strand (upper) and minus strand (lower), respectively. Note that total numbers for single transcripts are much higher than ten but omitted for better visualization. TSS are indicated by horizontal arrows marked with the start positions. Hybridisation probe binding sites used for Northern detection are marked by vertical arrowheads and the probe name. (B) Northern analysis of C. trachomatis plasmid transcripts reveals a highly abundant small RNA pL2-sRNA1 of ~100 nucleotides in length antisense to pL2-07a. The region located antisense to the ORF pL2-02 encodes several transcripts in the range of ~80 to 450 nucleotides. Probe binding sites are marked by an arrow in 3A). The names of the transcripts correspond to the first nucleotide of the probe binding site.

6.2 Characterization of the *C. pneumoniae* Transcriptome by dRNA-seq

6.2.1 Purification of *C. pneumoniae* EB and RB

To gain insights into the transcriptome of EB and RB these distinct life cycle forms were separated based on their different buoyant density by differential gradient centrifugation. Several different gradient systems have been tested. The self-forming Percoll gradient in SPG buffer that was used for purification of *C. trachomatis* gave no satisfactory results for the isolation of *C. pneumoniae*. The band containing the RB resulted in a viscous suspension upon purification. Only a little amount of unbroken RB could be recovered. For isolation of viruses the use of potassium tartrate for equilibrium density-gradient centrifugation has been an established method since decades (McCrea *et al.* 1961). Tamura *et al*. (1967) successfully applied this method for the isolation of *C. pneumoniae* from mouse fibroblast cells grown in suspension. They reported that the number of broken RB observed by electron microscopy was increased after vigorous stirring and pipetting and that even complete disruption of RB can be achieved by treatment with ultrasound. These results indicate that *C. pneumoniae* RB are more fragile than *C. trachomatis* RB and that the recovery rate of intact RB is much lower. To overcome this problem, a new protocol was established based on a sucrose step gradient. This gradient consisted of SPG buffer containing 45%, 40%, 35%, 30% and 25% sucrose layered on top of each other in an ultracentrifuge tube. After centrifugation two distinct bands were visible in the gradient. The lower band was located at the 40/35% interphase and contained the denser EB. The lighter band containing RB was located at the 35/30% interphase. These were collected, washed and pelleted. To proof the content of these two fractions, they were checked by electron micrographs. Figure 11 shows purified fractions from the gradient separation containing RB (Figure 11A) and EB (Figure 11B) marked by arrowheads.

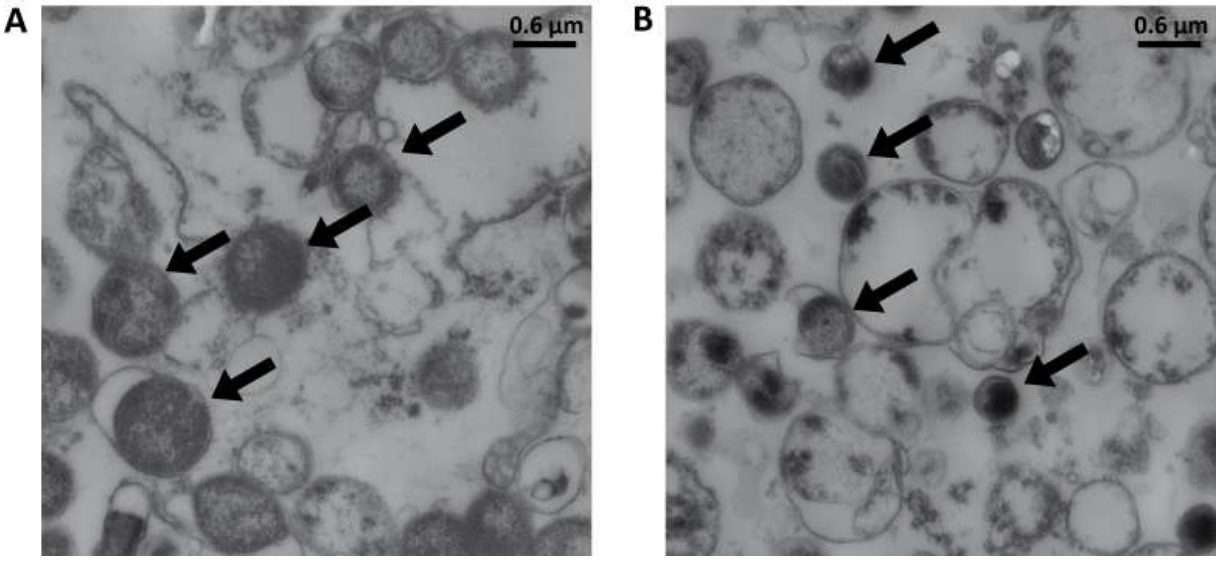

Figure 11: Electron micrographs of purified *C. pneumoniae* RB and EB.
RB (A) **and EB** (B) **were purified by differential centrifugation of disrupted infected host cells followed by density gradient centrifugation in a discontinuous sucrose gradient. EB (lower) and RB (upper) fractions were removed from the gradient and washed in SPG buffer. Purity of each fraction was estimated by electron microscopy. Arrows depict RB (A) and EB (B), respectively.**

EB have highly condensed DNA, no metabolic activity and contain less water than RB. Therefore they have a higher density than most of the other cell components and organelles and are relatively pure. RB have similar density to cell organelles like mitochondria. Figure 11B clearly shows impurities of the RB fraction. To avoid longer processing times of the *Chlamydia* and consequently changes in the RNA content of the bacteria, especially RNA degradation, no additional purification step was added.

6.2.2 Deep Sequencing of Purified *C. pneumoniae* EB and RB

The *C. pneumoniae* CWL029 chromosome (NC_000922) has a length of 1,230,230 nt and a coding density of 88%. RNA has been isolated from purified EB and RB for subsequent pyrosequencing of untreated and TSS-enriched RNA. The first sequencing run was performed with the Roche FLX chemistry. The achievable maximum sequence read length of this system is 100 bp. Read lengths derived from these libraries were much shorter than expected and a significant ratio of reads was shorter than 18 nt and could not be mapped to the genome due to the set criteria. A second set of cDNA libraries was made and sequenced using the Roche Titanium chemistry which obtains sequence reads of up to 400 bp in length. The resulting reads were significantly longer and the ratio of reads that could be mapped to the genome was greater. Sequence reads derived from both sequence runs were pooled for each library. The detailed numbers of reads obtained for each library and each sequence run, respectively, are given in figure 12 above the graphs and listed in table 5.

In total 1,437,231 sequence reads were obtained from the four cDNA libraries comprising more than 97 million nt. Of the resulting sequence reads 1,221,744 (85%) were at least 18 nt in length and were aligned against the *C. pneumoniae* genome. 854,242 sequence reads (70%) could be mapped to the *C. trachomatis* genome. Exact read numbers are given in figure 12 and listed in table 5 for each library. Concordant with literature a plasmid could not be detected in this strain. The remaining

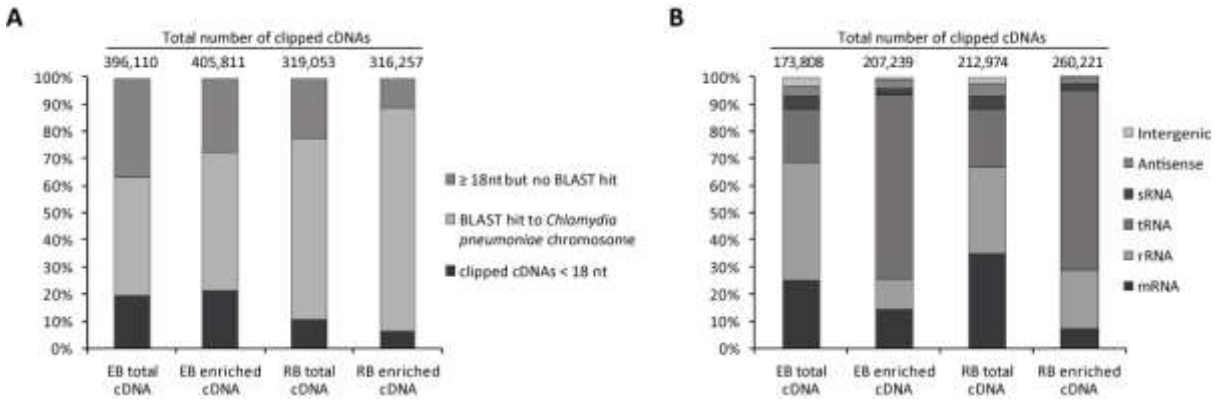

Figure 12: Classification of sequencing reads of *C. pneumoniae*.
Sequencing reads of at least 18 nucleotides in length were aligned against the *Chlamydia pneumonia* CWL029 genome. (A) **The fraction of sequences that could be mapped to the *Chlamydia* genome is given for each library. Sequence reads shorter than 17 nt were not considered for BLAST search.** (B) **The distribution of sequence reads among the different classes of RNA is shown for each library.**

Table 5: Numbers of sequence reads of the four *C. pneumoniae* cDNA libraries.
Sequence reads were pooled from two sequence runs with FLX and Titanium chemistry, respectively. All reads ≥ 18 nt were blasted against the *C. pneumoniae* genome. Sequence reads that gave no BLAST hit are derived from contaminating human host RNA.

Library	Tag FLX library	FLX cDNAs	Tag Titanium library	Titanium cDNAs	Pooled all	cDNAs < 18 nt	cDNAs ≥ 18nt	Total BLAST hits	cDNAs ≥ 18nt but no BLAST hit
CPEB-	GTAT	191,720	ACGTGC	204,390	396,110	76,018	320,092	173,808	146,284
CPEB+	GTGA	205,526	AGCGTA	200,285	405,811	86,244	319,567	207,239	112,328
CPRB-	GTTC	102,238	AGTCAG	216,815	319,053	33,697	285,356	212,974	72,382
CPRB+	TCAT	120,587	ATACTG	195,670	316,257	19,528	296,729	260,221	36,508
Total		620,071		817,160	1,437,231	215,487	1,221,744	854,242	367,502

sequences were of human origin or could not be mapped at all due to sequencing errors. Sequencing raw data files and processed data files which contain cDNA sequences in fasta format can be accessed at the Gene Expression Omnibus (GEO) of the NCBI under the accession number GSE24999 (http://www.ncbi.nlm.nih.gov/geo/).

For 982 of the 1122 (87.5%) genes from the primary annotation (Kalman *et al.* 1999) at least 10 sequence reads were obtained. The most abundant protein coding genes were *omcB*, *ompA*, *hctB* and *omcA* with more than 2,000 cDNA reads per locus. Of the remaining genes that were covered by less than 10 sequence reads per gene, 69% were genes of unknown function. They were either expressed at low levels under the conditions applied or are wrongly annotated. Sequence reads located in intergenic regions or antisense to annotated genes including candidates for non-protein-coding RNAs accounted for 8.5% of all sequence reads obtained.

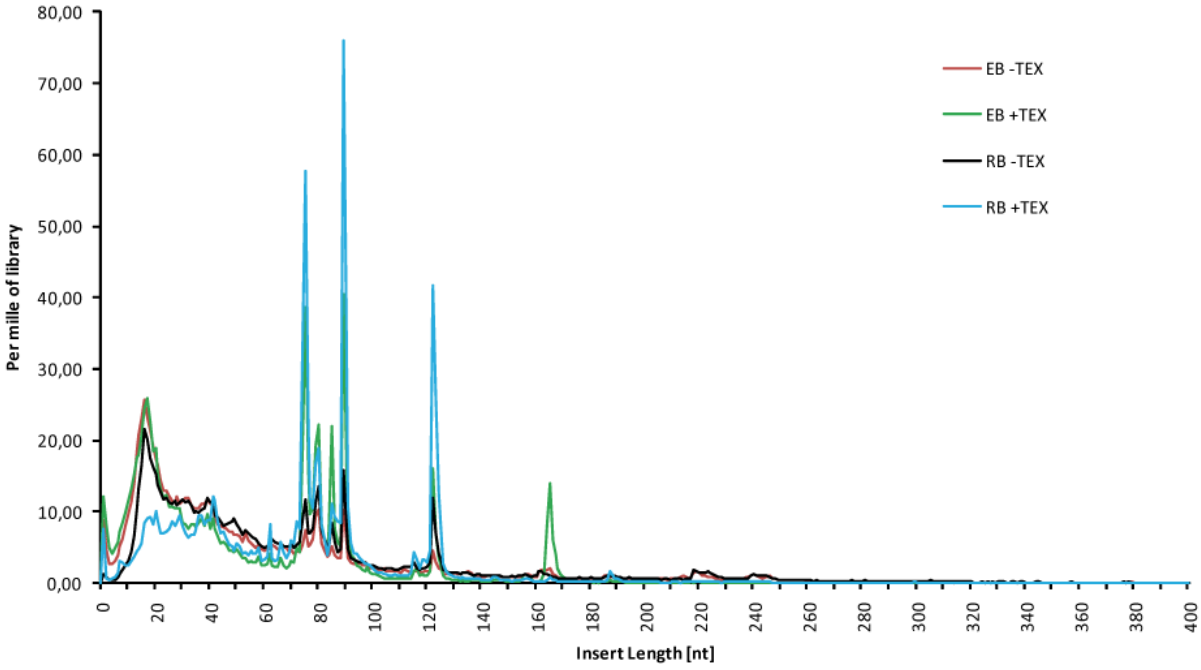

Figure 13: Sequence read length distribution of *C. pneumoniae*.
Graph showing the length distribution of sequence reads after linker and polyA removal of the four cDNA libraries. Peaks correspond to highly abundant RNA species like tRNAs (70–90 nt).

Figure 12A shows the ratios of classes of RNA among the four cDNA libraries. The ratio of RNA molecules shorter than 18 nt is larger in the two EB libraries. Also the fraction of cDNA reads that could not be mapped to the *C. pneumoniae* genome is significantly larger in the EB libraries. These sequences are derived from contaminating host cell RNA that was not depleted during *Chlamydia* isolation and purification. The fractions of reads that could be mapped to the genome are subdivided into the different classes of RNAs in figure 12B. Similar to the *C. trachomatis* libraries (Figure 4B), the fraction of mRNA reads is considerably decreased in the TEX treated libraries, whereas the fraction of rRNA and tRNA reads is increased (Figure 12B).

The average sequence read length of all cDNAs after linker and polyA clipping was 68.14 nt with read lengths up to 400 nt (shown in figure 13). Peaks in the length distribution are resulting from abundant RNAs like tRNAs (70 to 90 nt peaks) and 5S ribosomal RNA (123 nt peak). The peak at 165 nt is only present in the EB enriched library and refers to human U1 small nucleolar RNA that result from contaminating human RNA.

6.2.3 Annotation of Transcriptional Start Sites (TSS)

The primary annotation of the *C. pneumoniae* CWL029 genome contains 1,122 genes, thereof 1,074 protein coding and 43 structural RNAs. Treatment of the RNA with TEX prior to sequencing removes processed, fragmented, and degraded RNA molecules with a 5'monophosphate from the total RNA. By selective digestion of RNA with 5' monophosphates native 5'ends containing a triphosphate were enriched. This enables the exact determination of the TSS up to a single nucleotide resolution as demonstrated by previous studies (Albrecht *et al.* 2010; Sharma *et al.* 2010).

In total 531 primary TSS and 39 secondary TSS, located inside a primary transcript, could be identified by manual inspection of the sequencing data which are listed in table 11 (Appendix). This table contains the genome annotation of *C. pneumoniae* (NCBI RefSeq NC_000922) and all TSS that could be assigned to protein coding genes as well as non-coding RNAs and novel transcripts. The first column of table 11 defines the relative location of the TSS as intergenic (blank), antisense to annotated genes (AS), or sense to annotated genes (S). In the second column primary TSS are listed. In total, 39 secondary TSS were identified that are located downstream of a primary TSS and give rise to an alternative shorter transcript.

Based on the TSS map, the lengths of 5' leader sequences for the 437 mRNAs with assigned TSS were calculated as the distance between TSS and annotated translational starts. The length distribution of these leaders is given in figure 14. Leader sequences of the majority of mRNAs varied between 10 and 50 n in length. Leaders longer than 100 nt were found for 111 mRNAs and six mRNAs (CPn0036, *clpB*, *ung*, CPn0869, CPn0929 and *tyrP1*) have a leader of more than 400 nt. On the contrary, five genes (CPn0064, *yjjK*, *glgX*, CPn0600, and *yceA*) are transcribed as leaderless mRNAs whose TSS and translational start are identical. A comparison of the leader lengths between *C. pneumoniae* and *C. trachomatis* shows a very similar size distribution (see figure 14).

Two novel protein coding genes that were missing in the annotation have been identified. One of them is a homologue of *C. pneumoniae* strain AR39 gene CP0147 and was numbered CPn0600.1. Another previously unannotated expressed transcript is located antisense to CPn0955 and contains an ORF of 72 aa. This transcribed ORF was named CPn0655.1.

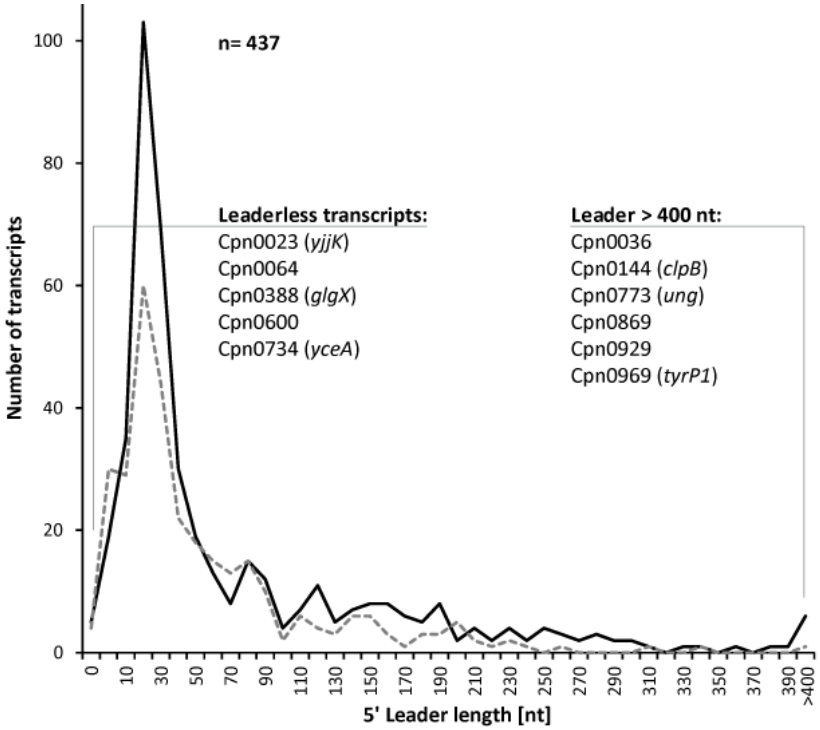

Figure 14: 5'leader length distribution and leaderless mRNAs of *C. pneumoniae*.
Graph showing the distribution of the distances between the TSS and the translational start of 437 annotated protein coding genes for which a distinct TSS could be assigned. Five Leaderless genes could be identified where the TSS corresponds to the annotated translational start. Six genes have very long leader sequences longer than 400 nt. For comparison, leader lengths of 320 *C. trachomatis* genes are included (dashed grey line).

Table 6: *C. pneumoniae* genes that have to be re-annotated.
The newly identified TSS have been compared to the translational start sites from the genome annotation. Ten genes have been identified where the TSS is located downstream of the translational start. These genes seem to be wrongly annotated and therefore should be considered for re-annotation.

	Annotation		Relative		
TSS	Start	End	location of TSS	Strand	Gene
151,170	151,164	151,781	+6	+	CPn0120 (*gmk*)
297,709	297,152	297,730	-21	-	CPn0264 (*ubiD*)
449,019	449,015	449,713	+4	+	CPn0403 (*yceC*)
461,462	460,218	461,498	-36	-	CPn0415
479,378	477,273	479,462	-84	-	CPn0434
575,352	575,143	575,364	-12	-	CPn0494
582,465	582,457	583,653	+8	+	CPn0501 (*hrcA*)
702,011	701,417	702,025	-14	-	CPn0611 (*coaE*)
856,966	856,957	857,697	+9	+	CPn0760
1,027,604	1,027,595	1,027,825	+9	+	CPn0896

The analysis of mRNA leader lengths unveiled 10 genes that have to be re-annoated because the transcriptional start is located downstream of annotated translational start. These genes are listed in table 6 including the relative location of the assigned TSS to the annotated translational start. Alternative shorter ORFs that are consistent with the TSS are present in all of these genes. The transcriptional regulator HrcA which is encoded as the first gene of the DnaK operon starts 8 bp downstream of the annotated CDS. There exists an in-frame start codon downstream of the annotated start and consequently the protein has a 12 amino acid shorter N-terminus than previously predicted.

Several genes have been described to have tandem promoters because two or more potential TSS have been mapped upstream of the gene. These are *C. trachomatis tuf* (Shen *et al.* 2000), the rRNA gene (Tan and Engel 1996) and *ompA* (Tan *et al.* 1996). In *C. pneumonia*, however, the *tuf* gene is co-transcribed as part of an operon and has no TSS upstream of the gene start. For the rRNA gene a single TSS could be identified and a processing site at position 1,000,490 which was previously reported to be a TSS in *C. muridarum* (Engel and Ganem 1987). Furthermore, the *ompA* gene was described to have tandem promoters in *C. trachomatis*. Tandem promoters with alternative TSS could be identified for 18 genes. The genomic coordinates of alternative TSS are listed in table 11. Interestingly, among these were genes with tandem promoters that were differentially used for transcription in EB and RB such as *rpsA*, CPn0365, *fabI*, CPn0408 and *infC*. Figure 15 displays the sequencing read distribution of the untreated and enriched cDNA libraries of EB and RB of these genes, respectively. Genomic coordinates of TSS are given above the arrows that depict TSS. For all five genes the major TSS used in EB is located downstream of the TSS used in RB resulting in a shorter leader sequence. Both, untreated and TEX treated libraries show a clear enrichment for the upstream TSS in RB and the downstream TSS in EB for the genes depicted in figure 15 A-D. The sequence read distribution in figure 15 E is not that distinct. In the untreated RB library the majority of reads starts from the downstream TSS at 1,139,785 that is present in EB. Thus, this site could be a processing site or both, a TSS and a processing site.

From the important group of polymorphic outer membrane proteins (Pmp) all 21 members were found to be expressed. An internal TSS was found to be located inside the annotated *pmp3.2* gene resulting in a transcript of 1.5 kb that contains an ORF of 454 aa in frame to the annotated protein of 746 aa. Furthermore, internal TSS were present in *pmp5.1*, *pmp10.1*, and *pmp17.1*.

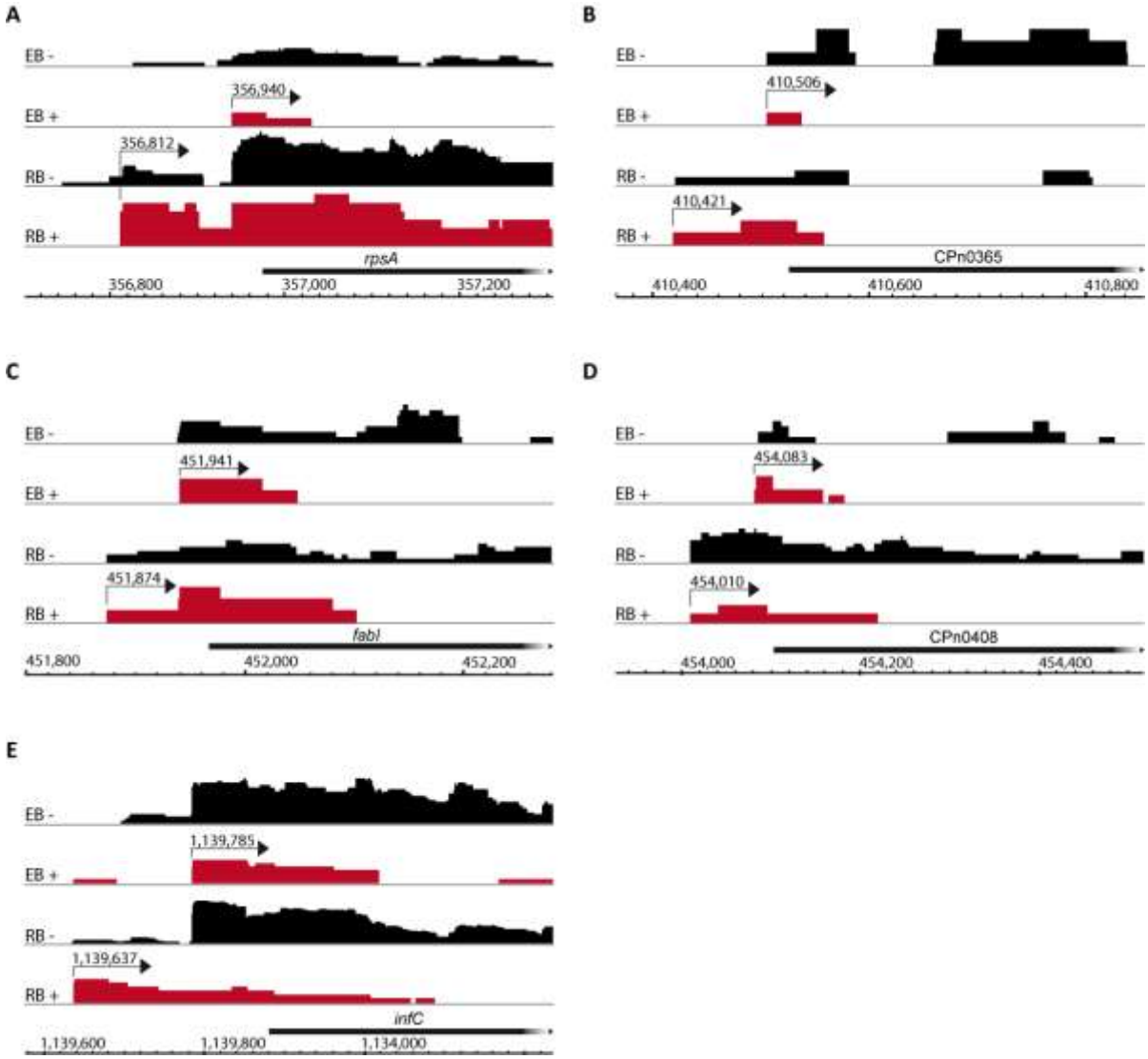

Figure 15: Genes with alternative transcription start sites in *C. pneumoniae* EB and RB.
Visualization of sequence reads for five genes that have alternative TSS in EB and RB. Black graphs represent reads from untreated libraries of EB (EB -) and RB (RB -), respectively, and red graphs represent reads obtained from TEX treated libraries of EB (EB +) and RB (RB +), respectively. Arrows depict TSS including the genomic coordinates of TSS. The transcriptional start is located downstream in EB which results in a shorter 5'leader sequence.

6.2.4 Discovery of Putative Small Non-Coding RNAs of *Chlamydia pneumoniae*

Numerous small transcripts lacking an ORF could be identified in intergenic regions, antisense or even sense to protein coding genes. In total, 75 TSS (listed in table 11) were identified that represent putative sRNAs. These comprise 20 putative *trans*-acting sRNAs encoded in intergenic regions, 47 putative *cis*-encoded antisense sRNAs and 8 sRNA candidates encoded sense to annotated ORFs. TSS and genomic coordinates of these novel transcripts are listed in table 11 including their relative location and nomenclature.

The 54 most promising candidates for novel sRNAs were tested by Northern hybridisation to check for presence of a distinct band of the corresponding size. Genomic coordinates, Northern probes used for hybridisation and the size of distinct bands obtained by hybridization are listed in table 7 in full detail. This table contains information on the location of the sRNA candidate (Column 1: antisense (AS), sense (S), intergenic region (IG)), the genomic coordinates of the sRNA, the DNA strand encoding the sRNA and the name of the putative sRNA. Novel validated sRNAs were named according to their relative location and numbered by the adjacent gene that is encoded upstream, antisense or sense to the sRNA, respectively. The last two columns of table 7 contain the theoretic size of the sRNA deduced from the deep sequencing information (if available) and the size of the detected sRNAs obtained from Northern hybridization. Based on the detection of a distinct band of the theoretical size thirteen of these sRNA candidates were positively validated as shown in figure 16B. The genomic location of these sRNAs is depicted in figure 16A. Nine novel *trans*-acting sRNAs are transcribed from intergenic regions, two *cis*-acting antisense sRNAs are transcribed from annotated protein coding genes, and two sRNAs are encoded inside the coding regions. Arrows in figure 16A depict TSS of sRNAs and surrounding protein coding genes, as determined from deep sequencing data. All predicted house-keeping RNAs were found, these include tRNAs, 5S, 16S and 23S rRNAs, signal recognition particle RNA (SRP RNA, 4.5S RNA, figure 16B), trans messenger RNA (tmRNA) and RNaseP RNA (M1 RNA). Furthermore, the homologue of the previously described sRNA ihtA in *C. trachomatis* (Grieshaber *et al.* 2006a) could be detected (CPIG701, figure 16B). 26 of the tested sRNA candidates gave no signal in Northern hybridisation, probably due to weak expression or insufficient probe binding. Overall, 14 candidates gave a signal that did not correspond with the theoretical size obtained from the sequencing data.

Two novel sRNAs CPIG0953 and CPIG 0954 are located sense and antisense to the annotated genes *ychB* and CPn0955, respectively. Sequencing read distribution among these genes and the presence of the novel sRNAs suggest that these genes are wrongly annotated. Furthermore, homologous genes in other *Chlamydia* species are lacking. Therefore, the two sRNAs are considered as intergenically encoded.

Table 7: Northern blot results of putative sRNAs of *C. pneumoniae*.
54 out of 75 sRNA candidates have been tested for expression by Northern blotting. 13 of these could be positively verified since the detected band corresponded to the theoretical size of the sRNA candidate calculated from the sequencing data. These verified sRNAs were named according to the adjacent gene located upstream, antisense or sense, respectively. A highly abundant sRNA was found to be encoded in ORF CPn0332. A sRNA candidate that turned out to encode a protein coding gene missing in the annotation was named CPn0600.1. The signal recognition particle RNA (SRP RNA) was used as positive control. Columns are labeled as follows: Position: AS = antisense to annotated ORF, IG = intergenic region, S = sense to annotated ORF; Start/End, genomic coordinates of sRNA; Strand: sense (+) or antisense (-) strand encoding sRNA; Putative Candidate: preliminary name; Validated Candidate: final name of validated sRNA; Probe Name: name of oligonucleotide used for Northern hybridization, contains coordinate of first nucleotide and strand; Probe Sequence: nucleotide sequence of the probe in 5' to 3' direction; Signal: indicates whether a signal was detected in Northern hybridization; Theoretic Size: length in nt calculated from sequencing data; Northern Size: length in nt detected by Northern hybridization.

Position	Start	End	Strand	Putative candidate	Validated Candidate	Probe Name	Probe Sequence	Signal	Theoretic Size [nt]	Northern Size [nt]
AS	28,995	28,435	-	pCpn01		Cpn_sRNA_028.927-	AGGGAGTGACCTTGATCACACAAG	-		
AS	49,662	50,571	+	pCpn02		Cpn_sRNA_049.751+	GGATTCCTCTCACATCCTTGGCT	+	531	400
AS	59,034	58,846	-	pCpn04		Cpn_sRNA_058.998-	TCAAGAGAGCCGGTATCTAGAGGG	-		
AS	78,886	77,712	-	pCpn06		Cpn_sRNA_078.499-	GCTTGGGTGGTTCTCGCTCTACT	+		90,150
AS	94,794	94,609	-	pCpn07		Cpn_sRNA_094.755-	TTATCGTGGAAGTTGGCTCGAAG	-		
AS	102,542	102,308	-	pCpn08		Cpn_sRNA_102.496-	TGTAGTTTGGGGAATCCGACAGA	-		
IG	104,806	104,677	-	pCpn09		Cpn_sRNA_104.773-	TGCAACCCGTTTCTTATTTGTGG	+	130	23
AS	132,469	132,027	-	pCpn10		Cpb_sRNA_132.426-	TGCTTATAGGCCTTCCTTTCATCA	+		200
S	132,303	132,814	+	pCpn11		Cpn_sRNA_132.332+	TAGGGAACGTTAGGGTAGCTGCC	-		
IG	175,970	176,208	+	pCpn12	CPIG0142	Cpn_sRNA_176.193+	GGAACCTCTATAGATCCAACCAGCTC	+	80	80
AS	184,478	183,992	+	pCpn13		Cpn_sRNA_184.426-	TTCACACTTCAGAAGGAAAACAAGG	-		
AS	195,128	195,218	+	pCpn14	CPAS0152	Cpn_sRNA_195.152+	CAAGACAGAAGTATGTCCGAAGCA	+	90	80
AS	237,544	237,186	-	pCpn15		Cpn_sRNA_237.483-	TCGACGAACTTCAGACTTTCGTTT	+	260	220
IG	248,672	248,719	+	pCpn16	CPIG0207	Cpn_sRNA_248.715+	GAAGGCTAAATCACAGGTGGTCC	+	45	~40
AS	251,272	250,976	-	pCpn17		Cpn_sRNA_251.144-	GCAATAACCCCTGCAGGTTTAGC	-		
AS	270,299	270,431	-	pCpn19		Cpn_sRNA_270.341+	TGAGACTCCCGATTTGAAAATCG	-		
AS	278,732	278,895	+	pCpn21		Cpn_sRNA_278.807+	CTGCCGCCTCACAATCTACAATC	-		
AS	282,641	282,703	+	pCpn24		Cpn_sRNA_282.701+	ATTAGGCCTTGCACGAGCTCTCT	+	63	
IG	298,666	298,738	+	pCpn25		Cpn_sRNA_298.729+	CCCCCTAACATAGGGGGTTCTAAA	-		
AS	317,557	317,607	+	pCpn26		Cpn_sRNA_317.601+	AGAAGCTTTCCAGGCTGCCATAC	+	50	50
AS	325,089	325,380	+	pCpn29		Cpn_sRNA_325.142+	CGATGCCTTTTGTAGCTTTGCTTA	-		
AS	330,522	330,471	-	pCpn31		Cpn_sRNA_330.493-	TTTTATTTACAGCAGCCCTGGTCC	-		
AS	333,613	333,433	-	pCpn32	CPAS0294	Cpn_sRNA_333.440-	TGCCCATCCGTAGCTTTATCTTTT	+	110	95
IG	333,613	333,433	-	pCpn32	CPIG0294	Cpn_sRNA_333.553-	GCACGCCGTAATCACTATTGAAAA	+	65	65
AS	333,759	333,823	+	pCpn33		Cpn_sRNA_333.824+	CAGTTAGGAGTGGATCCAAAAGAA	-		
AS	349,722	350,069	+	pCpn35		Cpn_sRNA_349.746+	ATCATTCGACGGTAATCTCATCCA	+	235	120
S	378,551	378,788	-		CPn0332	Cpn_sRNA_378.724-	TAATTACCCAGGCTTCCTGTCGTT	+	240	80-250
AS	389,562	389,909	+	pCpn36		Cpn_sRNA_389.598+	TGGAAGTCTTTTCACGGGTTCTC	-		
AS	410,234	410,104	-	pCpn37		Cpn_sRNA_410.349+	CTCGCGAGTAAATCAAGCCTTTC	-		
IG	410,321	410,359	+	pCpn38		Cpn_sRNA_410.350+	GCTCGCGAGTAAATCAAGCCTTT	-	40	-
AS	428,595	418,822	+	pCpn39		Cpn_sRNA_428.757+	TAGAGTTTCAGGACTTGCGCAGG	-		
IG	445,308	445,452	+	pCpn41	CPIG0397	Cpn_sRNA_445.394+	GGACGCAACCACTAAGAGCTAACA	+	144	140
IG	447,422	447,312	-	SRP_RNA		Cpn_sRNA_447.387-	TCGCAGGTTCCTTCTCTTAAAAGG	+	110	95,110
IG	503,607	503,533	+	pCpn46		Cpn_sRNA_503.534-	AATGGGCAGAGTACGTGATCTCAC	-		
IG	513,197	512,905	-	pCpn47		Cpn_sRNA_513.011+	TTCGTTTGGGAGAATTTCTTGAGA	-		
IG	523,061	522,987	-	pCpn48	CPS0657	Cpn_sRNA_523.009-	GATTGTGAAAGAGGCTTTTGCCC	+	73	70
IG	548,000	548,167	+	pCpn50		Cpn_sRNA_548.042-	CGCAAGGGAAACTTGAAGAGTTTT	-		
AS	578,934	579,257	+	pCpn51		Cpn_sRNA_578.979+	TTCTGTTGGGTGTTGTCCTTGGT	-		
AS	589,097	588,471	-	pCpn52		Cpn_sRNA_589.023-	ATCGGCATCGATTATGCTCAAGA	-		
IG	654,455	654,421	-	pCpn53	CPIG0564	Cpn_sRNA_654.431-	TCCACCTTCAAAGAGAGCTCTTTA	+	35	35,40
AS	660,667	660,351	-	pCpn54		Cpn_sRNA_660.507-	GTCATTAGAGGGACCTCCGGAGA	-		
AS	671,049	671,467	+	pCpn55		Cpn_sRNA_671.075+	AGATGCCACGCACACACTACGTA	-		
IG	692,664	692,320	-		CPn0600.1	Cpn_sRNA_692.520-	GCAACAGGAAGCACAAGAGCTAAA	+	344	420
S	738,367	738,277	-	pCpn57	CPS0657	Cpn_sRNA_738.296-	CTTCCGCGATAGTATCTCCGAGAT	+	88	90
IG	767,056	767,090	+	pCpn58		Cpn_sRNA_767.081+	TTCCATGGATATGGAATGGCTTTT	+	34	130
IG	775,135	775,010	-	pCpn59	CPIG0692	Cpn_sRNA_775.105-	CAGCATGGAATACACTAAGGCGC	+	122	120
IG	784,727	784,788	+	pCpn60		Cpn_sRNA_784.787+	GAGCAATCTTTTCCCTCTCTATGA	+	61	110
IG	786,909	786,802	-	pCpn61	CPIG0701	Cpn_sRNA_786.828-	TACGGCCGACTCTACATCTTCTTG	+	106	100
S	818,360	818,173	-	pCpn62		Cpn_sRNA_818.395-	GCCTCCAATCCTTGAGCAGAAAG	-		
IG	844,006	844,038	+	pCpn64		Cpn_sRNA_844.029+	TGAAGATAAAATCTTCATCCCCGA	+	31	30
IG	858,486	858,441	-	pCpn66		Cpn_sRNA_858.452-	CCCTCCTAAAGTATTAGAAGGGGG	+	45	450
IG	1,088,328	1,088,465	+	pCpn68	CPIG0953	Cpn_sRNA_1.088.374+	TAAAGGGTTGTGAGCTCGTGAAAA	+	137	130
IG	1,088,976	1,088,925	-	pCpn69	CPIG0954	Cpn_sRNA_1.088.955-	CTTTCGAAAGCCATGTGCTTATGT	+	50	52
AS	1,136,800	1,136,420	-	pCpn72		Cpn_sRNA_1.136.753-	TTTATGGAAAGAACGCCGTCCTT	-		
AS	1,206,657	1,206,332	-	pCpn74		Cpn_sRNA_1.206.564-	GCGATGACTTTCCAAGTTAGCGA	+	320	65,150,240

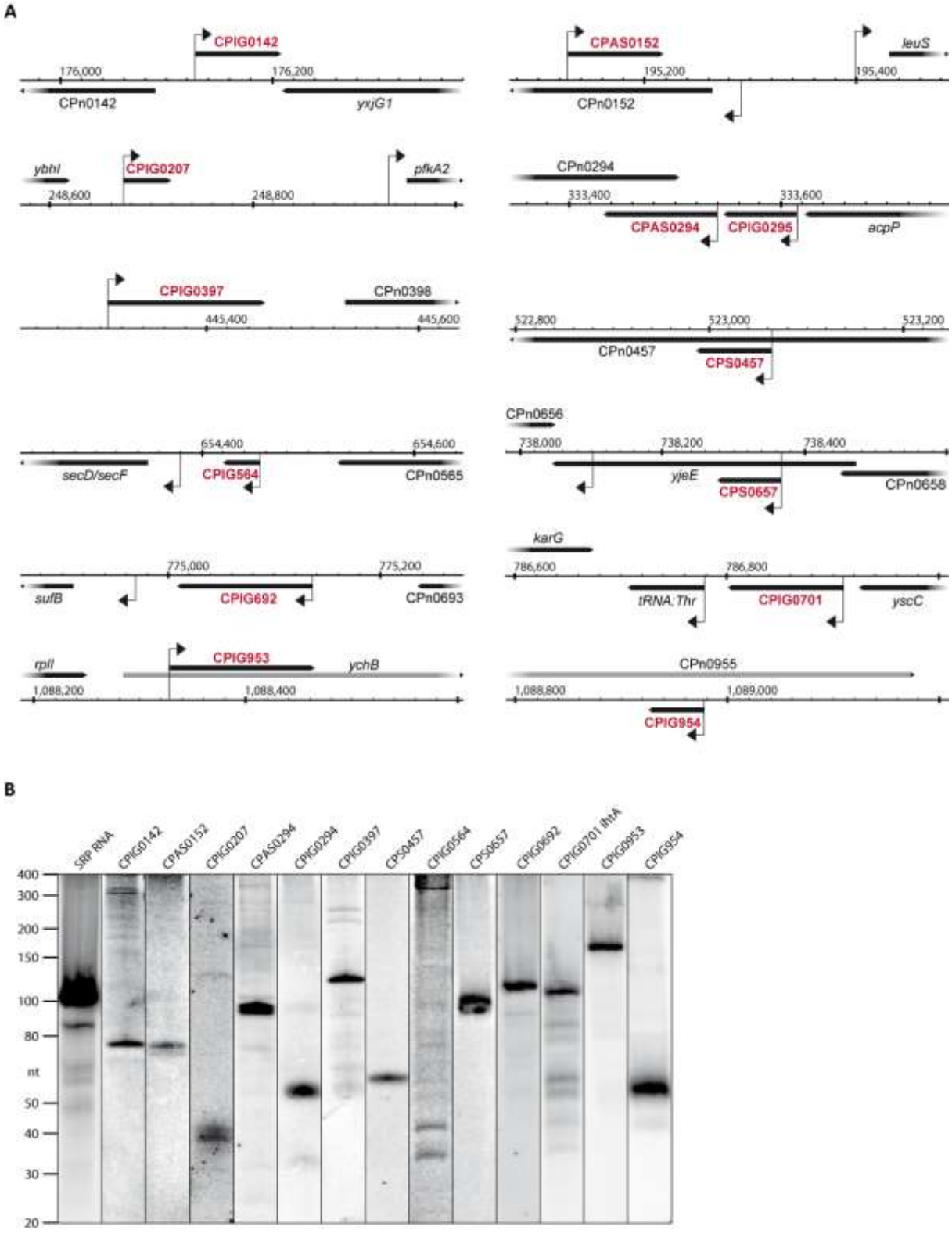

Figure 16: Novel non-coding sRNAs of *C. pneumoniae*.
(A) Short novel transcripts lacking an ORF were identified in intergenic regions, antisense or sense to annotated protein coding genes.The x-axis represents genomic coordinates with annotated genes on the sense (above) and antisense strand (below). Newly assigned names of sRNAs are coloured in red. Arrows represent TSS of sRNAs and surrounding genes as determined from deep sequencing data. (B) Northern Blot analysis of the small transcripts verifies the theoretical sizes calculated from the deep sequencing data. The housekeeping RNA SRP-RNA was used as positive control for size verification. CPIG0701 is a homologue of the only previously identified sRNA in *C. trachomatis.* The two genes depicted by grey bars (*ychB* and CPn0955) seem to be wrongly annotated since transcription is very unlikely according to the sequencing data set.

CPn0332 is one of the most abundant transcripts with a total of 40,170 sequence reads in the four cDNA libraries. The transcript is located downstream of *ltuB* (late transcription unit B), lacks an own TSS and is co-transcribed with *ltuB* (Figure 17A). It was described for *C. trachomatis* as an accumulating fragment of the *ltuB* transcript before (Fahr *et al.* 1995). The transcript is 18 nt shorter than the annotated gene CPn0332 and no alternative ORF is present. Northern Blot analysis reveals a full length RNA species of approximately 250 nt length which fits well to the theoretical size of 238 nt. Several smaller fragments could be detected by the probe which range from 70 to 110 nt in length (Figure 17B). The full length sequence shares homology among all available *Chlamydia* genomes and a very similar transcript could be identified in *C. trachomatis*. The alignment of the genomic sequences of this locus is shown in figure 17C. It contains several highly conserved regions and a conserved intrinsic terminator stem-loop followed by poly-T. The start codon of the annotated ORF is

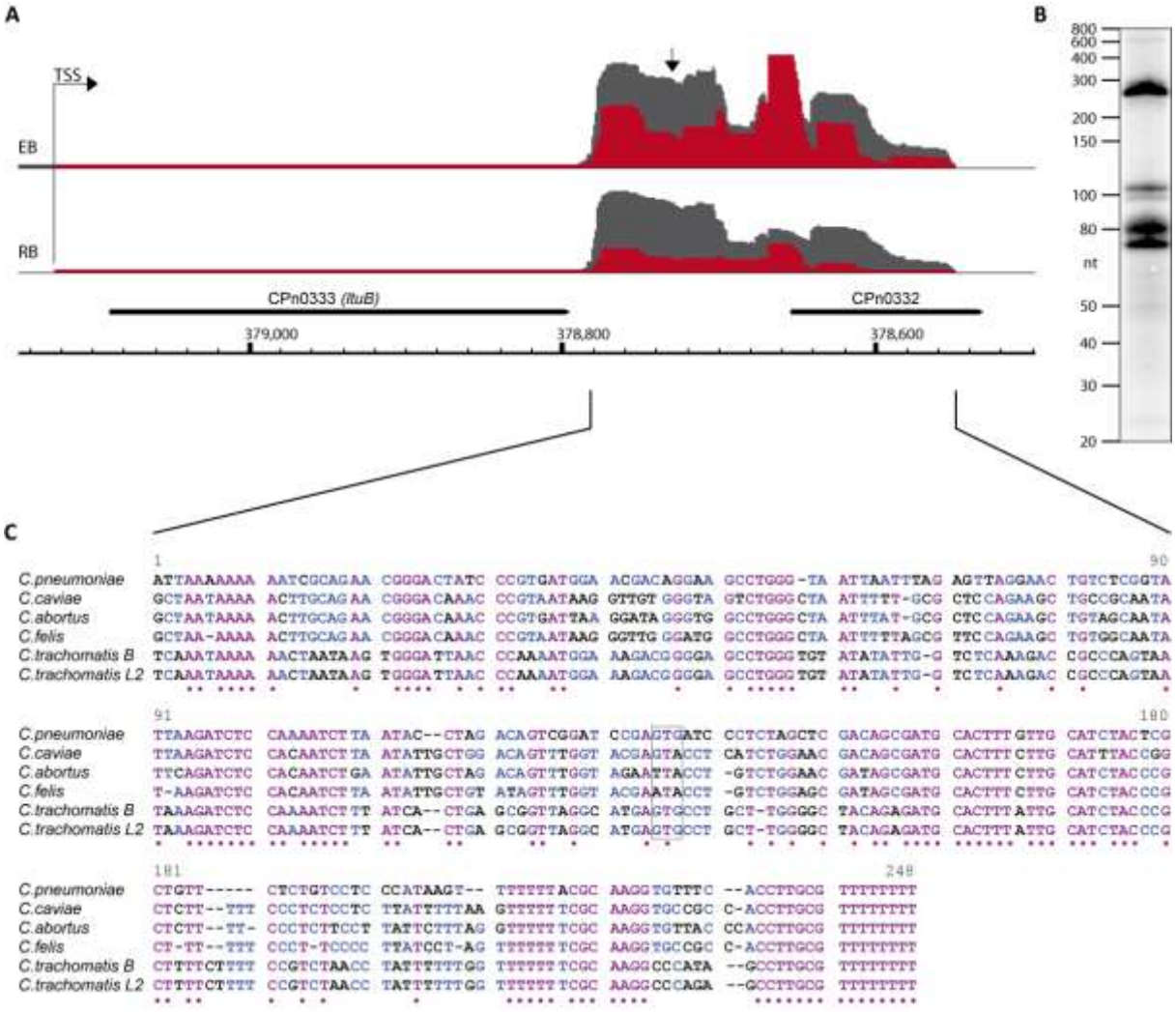

Figure 17: An abundant putative small RNA encoded in Cpn0332.
(A) Visualization of sequencing reads of the CPn0332 locus. Grey bars represent untreated total RNA libraries and red bars represent libraries of RNA treated with TEX and thus are enriched for primary transcripts. Upper line shows sequence reads obtained from EB and lower from RB, respectively. Black bars show annotated genes and coordinates of the genome. A highly abundant transcript is co-transcribed with *ltuB*. It ends 18 nt upstream of the annotated stop codon of CPn0332 and lacks an alternative ORF. (B) Northern Blot analysis of the abundant transcript reveals several bands ranging from approx. 70 to 250 nt. Probe binding site is marked by vertical arrow in (B). (C) Alignment of the locus downstream of *ltuB* shows homology between all available *Chlamydia* genomes and the two *C. trachomatis* strains B and L2. The Rectangle marks the start codon of the annotated gene CPn0332 which is not conserved.

not conserved among all *Chlamydia* which supports the finding of a non-coding RNA encoding locus.

Miura *et al.* (2008) searched for transcripts that are expressed at the late stage of the infection cycle. Thereby they identified putative σ^{28} promoters upstream of *ltub* and the annotated ORF CPn0332. According to the identified TSS the putative σ^{28} promoter postulated upstream of ORF CPn0332 is located inside the transcript. In many proteobacteria 6S RNA was identified to be an abundant non-coding RNA that globally regulates transcription during growth phases by inhibition of standard sigma factor RNA polymerase (Wassarman and Saecker 2006) and thereby enhancing alternative sigma factor activity. 6S RNA mimics an open promoter complex and a part of this RNA resembles a DNA promoter sequence (Toledo-Arana and Solano 2010). We tested whether the σ^{28} binding site could be functional which would result in binding of σ^{28} RNA polymerase to this RNA. We therefore tested by gradient fractionation whether CPn0332 RNA co-sediments with σ^{28} RNA polymerase. A total bacterial lysate of purified *C. pneumoniae* was separated on a 1-40% continuous

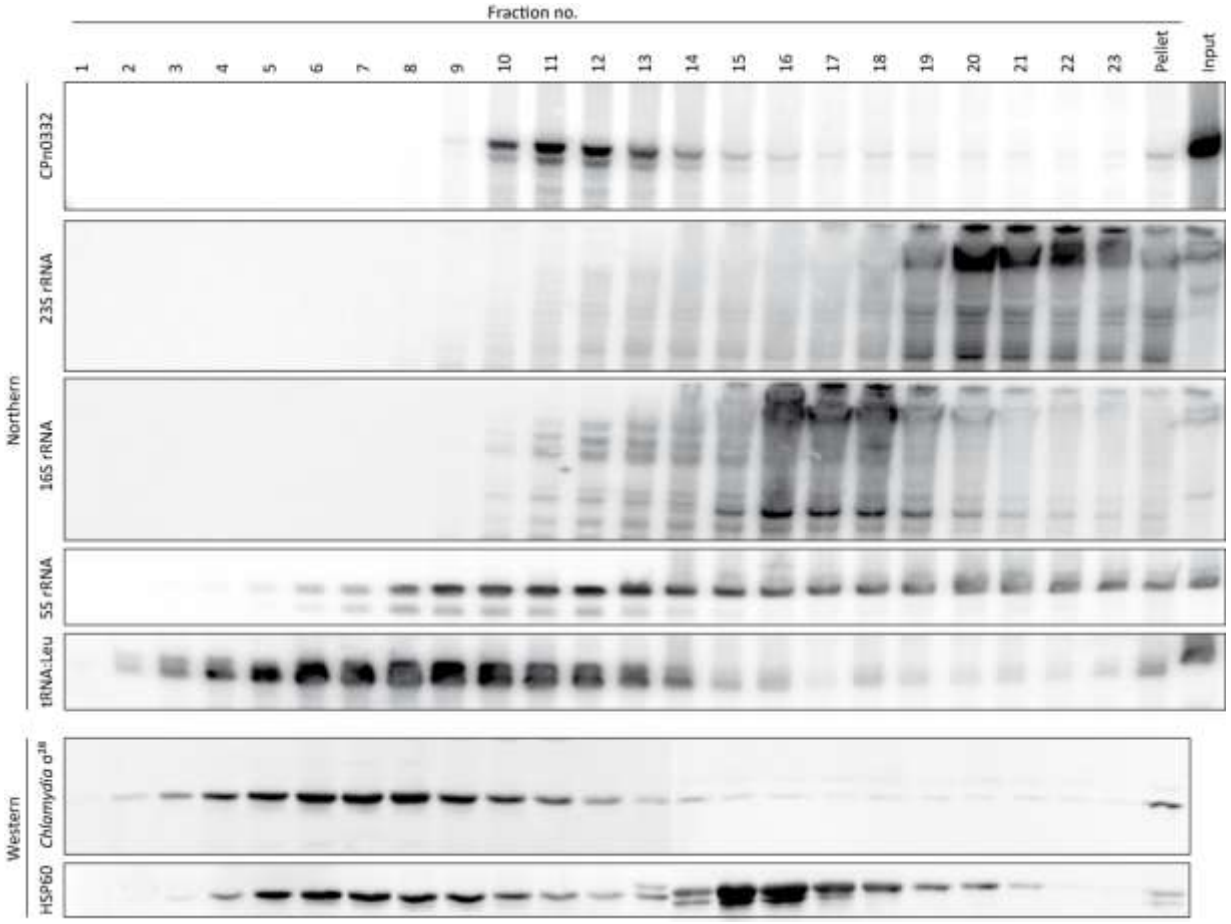

Figure 18: Density gradient fractionation of *C. pneumoniae* lysates.
***Cpn* whole cell lysate was separated on a 1-40% continuous glycerol gradient. From each of the 24 fractions total RNA was isolated and a Northern blot was performed using probe directed against the abundant RNA encoded by CPn0332, as well as ribosomal RNAs. From the same fractions a Western blot was performed using an antibody directed against chlamydial σ28 factor. An association of RNA CPn0332 with σ^{28} RNA polymerase can be excluded since they are located in different fractions of the density gradient. The ubiquitously expressed heat shock protein 60 (encoded by *groEL*) was used a control. Note that, due to the large number of samples, the Western blot picture is assembled from two blots that were treated the same way and in parallel. Fraction number 1 is the lightest on top of the gradient and fraction 24 is the heaviest and contains the insoluble resuspended pellet.**

glycerol gradient and 24 fractions were obtained from the gradient. From each fraction total RNA was isolated and was subjected to Northern hybridization for CPn0332. Additionally, Western blotting was performed from the same 24 samples and σ^{28} was detected by a polyclonal rabbit antibody (Figure 18). However, an association of RNA CPn0332 with σ^{28} could not be confirmed since the RNA CPn0332 accumulates in fractions 10 to 13, whereas σ^{28} is predominantly present in fractions 5 to 9 (see figure 18). Furthermore an association of CPn0332 with ribosomes can be excluded, since the RNA does not co-sediment with ribosomal RNAs (Figure 18). Although an association of CPn0332 RNA with σ^{28} RNA polymerase and ribosomes could be excluded, RNA polymerase itself and other σ factors could be tested as soon as antibodies for these proteins are available. Also, an identification of binding partners by aptamer-tagging technology could shed light on the biological role of this sRNA (Windbichler *et al.* 2008).

6.2.5 Analysis of Polycistronic Transcripts

The combined analysis of cDNA libraries derived from total RNA and RNA enriched for TSS allowed the analysis of the operon structure of *C. pneumoniae*. Figure 19 shows an example of two operons that encode genes of the type three secretion system (T3SS) (Hefty and Stephens 2007). Sequence reads of untreated libraries (Figure 19, black graphs) show that sequence reads were present for the entire operon. In contrast, sequence reads of TEX treated libraries (Figure 19, red graphs) define distinct TSS sites. The first operon shown in figure 19A encodes five genes of the T3SS. This operon has a primary TSS upstream of the *yscU* gene and an internal TSS that is located upstream of *lcrE* and inside the CDS of *lcrD*. This results in a long transcript comprising all five genes and a shorter transcript derived from the internal promoter that encodes the three genes *lcrE*, *sycE* and *MalQ*. The operon shown in figure 18B encodes six genes and has only one distinct TSS. The presence of a multitude of sequence reads among the entire operon leads to the conclusion that a long polycistronic transcript is produced from the single TSS which comprises all genes.

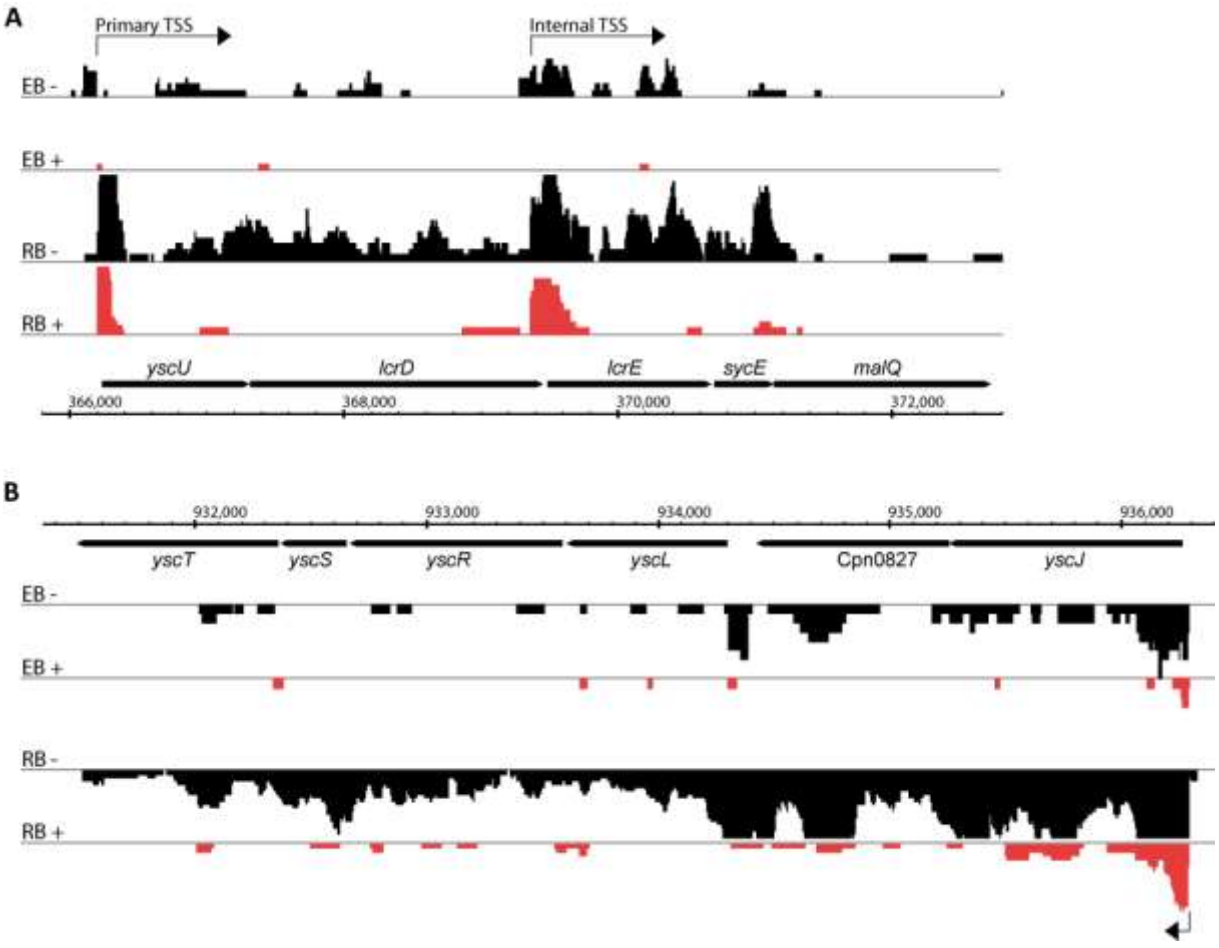

Figure 19: Two operons encoding genes of the type three secretion system (T3SS) in *C. pneumoniae*.
Polycistronic transcripts can be identified by comparing untreated and TEX-treated cDNA libraries. In the untreated libraries (black bars) sequence reads are distributed among adjacent genes and also cover gene junctions. The TEX-treated libraries (red bars) clearly show the TSS of a polycistronic transcript. (A) Sequencing data and Hefty and Stephens (31) suggest that one long transcript contains five genes (*yscU-malQ*). Furthermore, an internal additional TSS belongs to a shorter transcript containing only three genes (*lcrE-malQ*). Black arrows depict TSS. The operon shown in (B) contains six genes and is transcribed from a single TSS.

Table 8: Operon numbers and lengths of *C. pneumoniae*.
246 Operons that encode a total of 752 annotated genes were identified by analysis of deep sequencing data. The operon size listed corresponds to the number of encoded genes. The number of operons corresponds to the number of operons identified in the respective size and the number of genes is the product of operon size and number of operons of that respective size.

Operon size [genes]	Number of operons	Number of genes
2	129	258
3	61	183
4	30	120
5	11	55
6	6	36
7	1	7
8	3	24
10	2	20
12	2	24
25	1	25
Sum	246	752

The genome of *C. pneumoniae* contains 799 adjacent gene pairs. This study elucidated whether each adjacent gene pair is co-transcribed from the experimental sequencing data. 246 polycistronic transcripts could be identified containing between 2 and 25 ORFs each (Table 8). These encode a total of 752 genes. Taken together, a distinct TSS or an affiliation to a polycistronic transcriptional unit with a distinct TSS could be assigned to 861 out of 1074 protein coding genes (80%). Table 11 gives a precise map of TSS and the ORF structure of the *C. pneumoniae* transcriptome. Adjacent genes belonging to the same transcriptional unit are indicated by the same colour of the line. If adjacent genes of one transcriptional unit are interrupted in the table by genes encoded on the opposite strand, these are coloured in violet.

Several algorithms for operon prediction have been published in recent years. The present data set of operons of *C. pneumoniae* was compared with published operon predictions available at MicrobesOnline (Dehal *et al.* 2010) and DOOR database (Mao *et al.* 2009), respectively. Of the 799 pairs of adjacent genes 721 pairs (90.2 %) could be classified as either co-transcribed or individually transcribed. The remaining 78 pairs could not be classified since sequence read numbers were too low and thus, discrimination between co-transcription and individual transcription was not possible. The comparison with theoretical operon prediction algorithms reveals that 78.6 % (DOOR) and 81.1 % (MicrobesOnline) of the predictions coincide with the experimental data, respectively. Consequently the consistency is of the same magnitude as found for other bacteria like *E. coli* and *B. subtilis* (Dam *et al.* 2007; Mao *et al.* 2009; Price *et al.* 2005).

6.2.6 Differentially Expressed Genes in *C. pneumoniae* EB and RB

Previous studies on gene expression during the course of the *C. pneumoniae* developmental cycle were based on RNA isolation from host cell bacteria mixtures. Since the cycle becomes increasingly asynchronous with time this results in a mixture of EB, RB, and intermediate forms at the late time points of infection. Here, EB and RB were isolated and purified by differential gradient centrifugation to obtain total RNA from the two distinct life cycle forms.

786 protein coding genes were considered for differential expression analysis (for details see methods section). The threshold for differential expression was set as twofold more abundant in either EB or RB. In total, 53 genes were more abundant in EB (listed in table 9) and 404 in RB. Interestingly, 77% and 26% were found to be hypothetical proteins of unknown function in EB and RB, respectively. Gene families enriched in RB comprise genes involved in DNA and RNA synthesis, cell division, energy metabolism as well as the polymorphic outer membrane proteins. Among the few known transcripts more abundant in EB is the *ltuA* (late transcription unit A) gene. The *ltuB* gene is only 1.2-fold more abundant in EB than in RB. Since this gene is transcribed late during the developmental cycle and is very abundant, it seems to be accumulated in EB. Homologous genes were searched in *C. trachomatis* and protein identity based on amino acid sequence homology was estimated (Table 9, right section). Eleven genes from *C. pneumoniae* have no homologue in *C. trachomatis*. Furthermore, five *C. pneumoniae* genes are homologous to the same gene in *C. trachomatis* (CTLon_0309). Several of the remaining genes are only poorly conserved in *C. trachomatis* with protein identity levels of only 20-30%. This suggests that the genes enriched in *C. pneumoniae* are highly species specific and less conserved.

The chlamydial infection starts with EB entering the cell and conversion to RB after several hours. Late in the infection cycle, after more than 60 hours, EB start to reconvert to RB. Recently, a microarray based study of gene expression during an infection time course was carried out by Mäurer *et al.* (2007). Comparison of the deep sequencing based semi-quantitative approach with microarray analysis shows that 87% of the genes that were found to be more abundant in EB have their expression maximum at 6 or 72 hours post infection. In contrast 73% of the genes enriched in RB have their expression maximum at the intermediate time points 12 to 60 hours post infection. This clearly shows a good correlation between these two approaches.

Table 9: List of protein coding genes more abundant in *C. pneumoniae* EB.
Ratio of gene expression was calculated for every gene from the ratio of sequence read numbers in EB and RB. The ratio was normalized by the total number of sequence reads for protein coding transcripts for EB and RB, respectively. All genes with at least 20 sequence reads were considered. 53 genes were at least 2-fold more abundant in EB. For comparison to the *C. trachomatis* data set homologous genes in *C. trachomatis* strain L2b are given, including the identity on amino acid level. The last column lists the ratio of expression in EB vs. RB from the *C. trachomatis* dRNA-seq data; n.d., not detected.

				Chlamydia trachomatis		
Gene	*Product*	*Synonym*	*Ratio EB/RB*	*Homologue*	*Identity*	*Ratio EB/RB*
rpsI	30S ribosomal protein S9	CPn0246	2.08	CTLon_0377	0.783	n.d.
gcpE	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	CPn0373	5.18	CTLon_0308	0.604	n.d.
rpmA	50S ribosomal protein L27	CPn0545	2.15	CTLon_0672	0.854	n.d.
hemN_1	coproporphyrinogen III oxidase	CPn0380	3.48	CTLon_0303	0.644	n.d.
gltX	glutamyl-tRNA synthetase	CPn0560	3.17	CTLon_0701	0.797	n.d.
hrcA	heat-inducible transcription repressor	CPn0501	2.32	CTLon_0650	0.630	n.d.
	hypothetical protein	CPn0069	2.00	-	-	
	hypothetical protein	CPn0945	2.00	CTLon_0165	0.316	n.d.
	hypothetical protein	CPn1006	2.00	CTLon_0221	0.833	n.d.
	hypothetical protein	CPn0308	2.03	-	-	
	hypothetical protein	CPn0108	2.05	CTLon_0268	0.333	n.d.
	hypothetical protein	CPn0807	2.06	CTLon_0837	0.609	n.d.
	hypothetical protein	CPn0225	2.07	CTLon_0184	0.237	n.d.
	hypothetical protein	CPn0066	2.11	-	-	
	hypothetical protein	CPn1027	2.13	CTLon_0473	0.375	n.d.
	hypothetical protein	CPn0065	2.16	CTLon_0536	0.235	2.33
	hypothetical protein	CPn0285	2.18	CTLon_0309	0.231	n.d.
	hypothetical protein	CPn0370	2.21	CTLon_0309	0.256	n.d.
	hypothetical protein	CPn0585	2.22	CTLon_0137	0.215	n.d.
	hypothetical protein	CPn0383	2.22	CTLon_0298	0.446	n.d.
	hypothetical protein	CPn0369	2.25	CTLon_0309	0.591	n.d.
	hypothetical protein	CPn0497	2.26	CTLon_0641	0.583	n.d.
	hypothetical protein	CPn0367	2.37	CTLon_0309	0.246	n.d.
	hypothetical protein	CPn0381	2.38	CTLon_0301	0.243	n.d.
	hypothetical protein	CPn0332	2.44	CTLon_0332	-	11.12
	hypothetical protein	CPn0234	2.50	CTLon_0428	0.605	5.12
	hypothetical protein	CPn0169	2.53	-	-	
	hypothetical protein	CPn0241	2.61	CTLon_0386	0.219	n.d.
	hypothetical protein	CPn0163	2.67	-	-	
	hypothetical protein	CPn0648	2.68	CTLon_0786	0.360	n.d.
	hypothetical protein	CPn0474	2.76	CTLon_0617	0.402	n.d.
	hypothetical protein	CPn0365	2.85	CTLon_0309	0.211	n.d.
	hypothetical protein	CPn0461	2.87	-	-	
	hypothetical protein	CPn0814	2.87	CTLon_0830	0.479	n.d.
	hypothetical protein	CPn0565	3.17	CTLon_0705	0.301	n.d.
	hypothetical protein	CPn0152	3.29	CTLon_0400	0.461	n.d.
	hypothetical protein	CPn0166	3.55	-	-	
	hypothetical protein	CPn0240	3.87	CTLon_0386	0.240	n.d.
euo	hypothetical protein	CPn0561	4.09	CTLon_0702	0.633	n.d.
	hypothetical protein	CPn0439	4.19	CTLon_0060	0.241	n.d.
	hypothetical protein	CPn0516	4.30	CTLon_0301	0.276	n.d.
	hypothetical protein	CPn0220	5.34	-	-	
	hypothetical protein	CPn0458	5.59	-	-	
	hypothetical protein	CPn0150	5.77	CTLon_0398	0.219	5.52
	hypothetical protein	CPn0243	5.85	CTLon_0385	0.283	n.d.
	hypothetical protein	CPn0167	6.33	-	-	
	hypothetical protein	CPn0943	6.65	CTLon_0164	0.356	n.d.
	late transcription unit A protein	CPn1026	2.33	CTLon_0629	0.556	n.d.
oppA_3	oligopeptide binding protein	CPn0197	2.85	CTLon_0390	0.439	n.d.
	predicted OMP	CPn0343	5.49	-	-	
tgt	queuine tRNA-ribosyltransferase	CPn0219	4.22	CTLon_0440	0.755	n.d.
yabC_1	SAM-dependent methytransferase	CPn0382	2.31	CTLon_0299	0.630	n.d.
dsbD	Thio:disulfide interchange protein	CPn0786	2.51	CTLon_0852	0.541	n.d.

6.2.7 Analysis of *C. pneumoniae* Promoters

Bacterial RNA polymerase contains several σ factors that bind to conserved nucleotide sequences upstream of the TSS. Based on manual determination of single TSS and biocomputational analysis of promoter sequences, a few sequence motifs have been identified so far. Here, the sequences 40 bp upstream of the 531 determined primary TSS have been extracted and analysed for common motifs. Using the sequence motif search algorithm MEME (Bailey and Elkan 1994), a common motif could be found that resembles the *E. coli* σ^{70} consensus sequence. For this motif search a shift ±1 nt was allowed for the promoter sequences. The consensus sequence of 450 of the 531 promoter regions is shown in figure 20A. The -35 box consensus motif is TTGA which is shorter than the *E. coli* consensus sequence (TTGACA) and the -10 box resembles the *E. coli* sequence (TATAAT) whereas only TANNNT is highly conserved (see figure 20A). In between the -10 and the -35 box there are two A/T rich stretches around positions -17 and -26. Taken together, the sequence in figure 20A resembles the putative consensus promoter sequence of *C. pneumoniae* σ^{66} RNA polymerase. A further common promoter motif was detected for 24 genes, whereof 10 genes belong to the polymorphic outer membrane protein family (Pmp) as shown in figure 20B. These promoters share the motif CTTG at the -35 region and GTAT at the -10 box with long T-rich regions in between. A further search for shorter common promoter motifs was carried out using the -35 and -10 regions relative to the TSS. The most abundant motifs are given in figure 21A for the -35 box and figure 21B for the -10 box. The most prominent motif resembles the σ^{66} motif shown in figure 20.

Several predicted and validated promoter sequences have been reported for *Chlamydia.* The most conserved sequence has been reported for bacterial σ^{54} promoters which is TGGCAC-N_5-TTGC (Barrios *et al.* 1999). Based on sequence homology search Studholme and Buck (2000) identified a putative σ^{54} promoter sequence upstream of *C. pneumoniae* gene AAD18864 (CPn0725) located at positions 810,800 to 810,815. This site is located completely inside the transcript of CPn0725 and overlaps with the CDS. Therefore, this seems to be a false positive σ^{54} promoter. Mathews and Timms (2000) used a similar computational approach to search for putative σ^{54} promoter sequences in the *Chlamydia* genomes based on the eubacterial consensus sequence and identified a further putative σ^{54} binding site upstream of CPn0693. The sequencing data and also the Microbes Online operon prediction (Price *et al.* 2005) suggest a co-transcription of CPn0693 and CPn0694 from a TSS upstream of CPn0694, arguing against a potential σ^{54} promoter upstream of CPn0693. In total, the authors identified nine putative σ^{54} promoters in *C. trachomatis*. A homology search was carried out for these genes in *C. pneumoniae*. None of the nine promoters could be confirmed in *C. pneumoniae* since either the homologous gene is lacking or the homologous gene has no putative σ^{54} promoter in the region upstream of the TSS.

To further elucidate the presence of σ^{54} promoters, the 531 extracted promoter sequences (positions -1 to -40 relative to TSS) were checked for the least conserved σ^{54} core sequence GG-$N_{9\text{-}11}$-GC of the -24 and -12 box, respectively. In this data set of 531 promoter sequences no putative σ^{54}

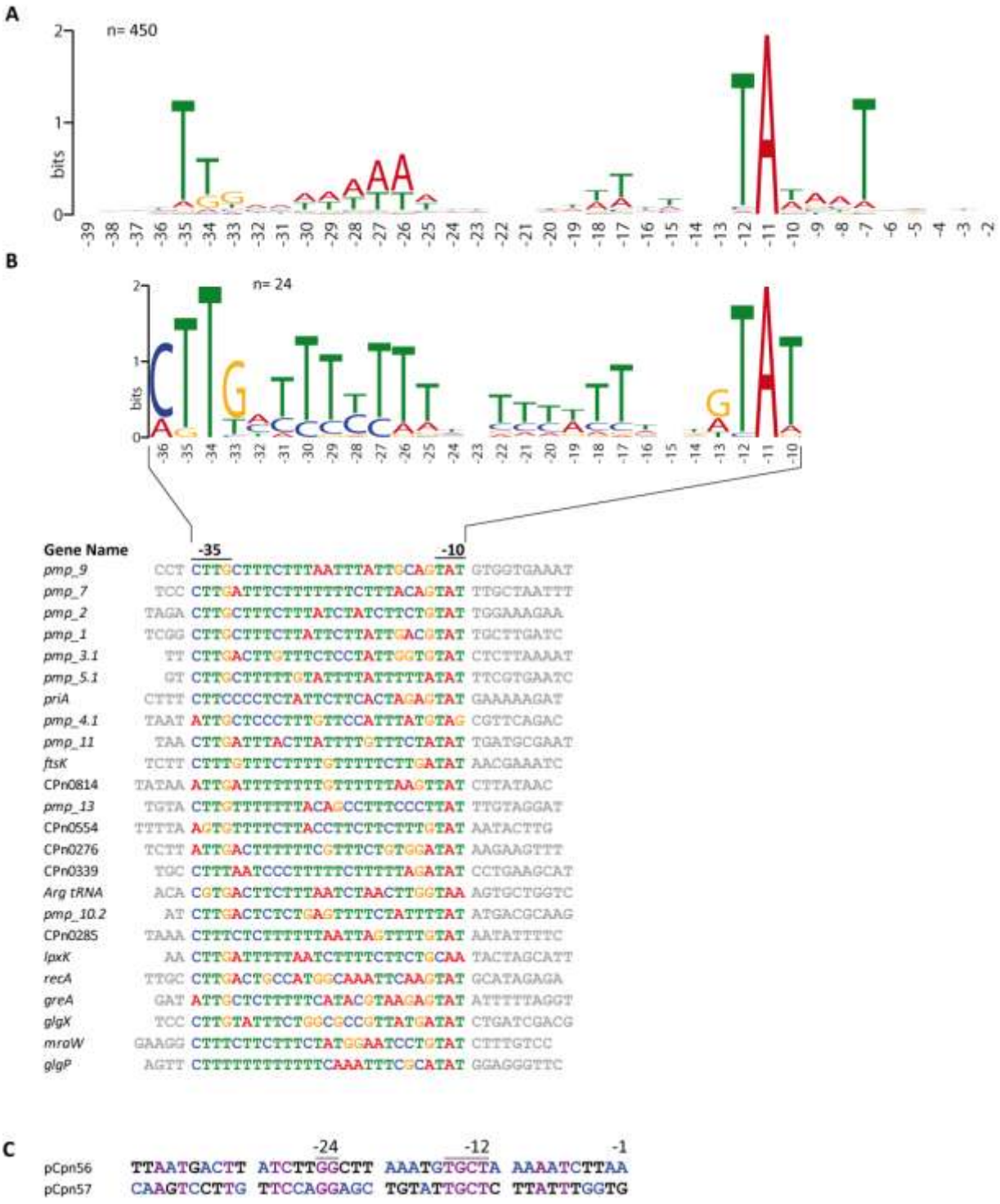

Figure 20: Promoter motifs of *C. pneumoniae*.
Sequences upstream of 531 TSS have been extracted (positions -1 to -40) and common sequence motifs were searched using MEME software. (A) The most prominent sequence alignment results from 450 out of 531 promoter sequences that were aligned. The motif resembles previously described σ^{66} promoter sequences and has the typical *E. coli* σ^{70} consensus sequences TATAAT at the -10 box. The -35 box is less conserved but also similar to the *E. coli* consensus sequence TTGACA. (B) The second most prominent motif is derived from an alignment of 24 genes comprising 10 genes of the polymorphic outer membrane protein family (Pmp). The sequence contains well conserved -35 and -10 boxes and a T-rich spacer in between. (C) A search for the minimum conserved σ^{54} consensus sequence GG-N_{9-11}-GC returned only two putative σ^{54} promoters that belong to sRNA candidates. No such sequence motif could be found in promoter regions of protein coding genes.

promoter sequence could be identified for annotated protein coding genes. However, two putative σ^{54} promoters were identified upstream of novel sRNA candidates pCPn56 and pCPn57 (Figure 20C) which share sequence homology only at the -12 and -24 promoter boxes, respectively.

The third sigma factor identified in *Chlamydia* so far is σ^{28}. Together with σ^{66} it belongs to the group of σ^{70} factors. σ^{28} RNA polymerase was shown to be expressed at the late stage of infection. Using an *in silico* prediction algorithm Yu *et al*. (2006b) identified putative σ^{28}-regulated genes of *C. trachomatis*. They could verify 5 genes to be controlled by σ^{28}. These genes comprise *tlyC1, bioY, dnaK, tsp* and *pgk.* Two of these genes are expressed in *C. pneumoniae* from an own promoter that resembles the *C. trachomatis* promoter (*tsp* and *pgk*). The genes *tlyC1* and *dnaK* are co-transcribed as part of a polycistronic transcript and *bioY* (probable biotin synthase) has no homologous gene in *C. pneumoniae.*

Several studies have characterized temporal gene expression during the developmental cycle of *Chlamydia* using microarrays (Belland *et al.* 2003a; Maurer *et al.* 2007; Miura *et al.* 2008; Nicholson *et al.* 2003). These studies identified cluster of genes that are expressed at the late stage of infection which corresponds to the stage prior to conversion of RB to EB. This set of genes includes *hctB* that was shown to be recognized by σ^{28} (Yu and Tan 2003). The *C. pneumoniae hctB* promoter resembles contains the extended σ^{28} consensus sequence TNAAG-N_{14}-GCCGATA derived from several γ-proteobacteria sequences (Song *et al.* 2007) with a spacer of 15 nt. In the set of 531 promoter sequences no further sequence was found that resembles the described σ^{28} promoter sequence of *hctB*

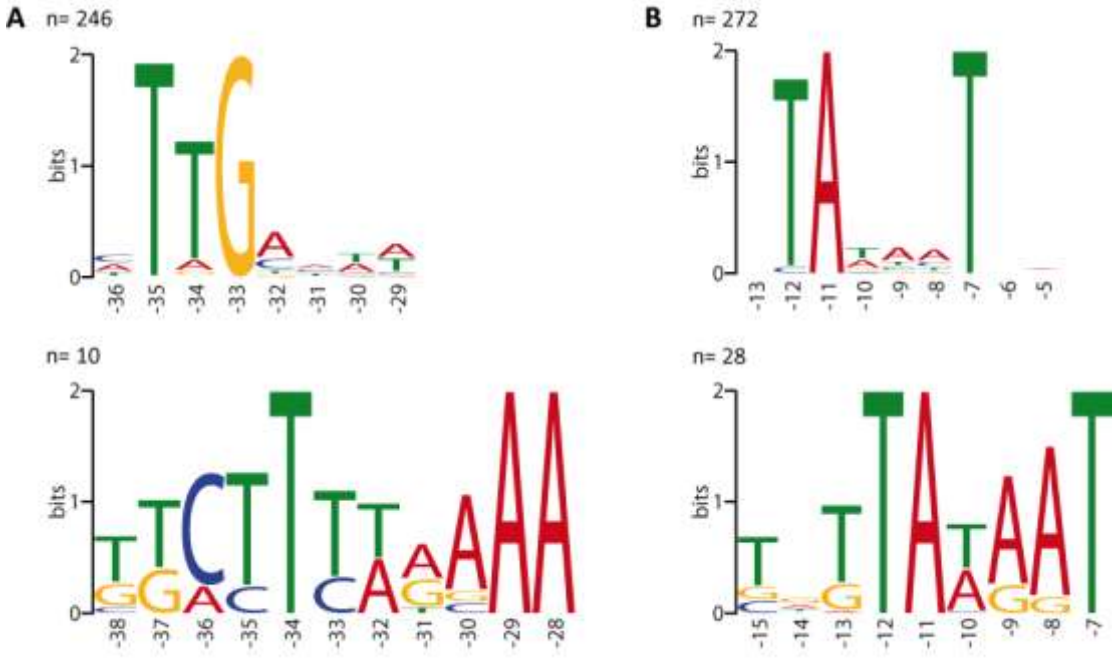

Figure 21: Most prominent sequence motifs of the -35 and -10 boxes of *C. pneumoniae.*
Promoter sequences were extracted from 531 primary TSS from -28 to -38 (A) and from -5 to -15 (B) and searched for common sequence motifs using MEME, respectively. The two most prominent motifs are shown in (A) for the -35 box and in (B) for the -10 box. The number of promoter sequences are given for each motif.

TNAAG-N_{15}-GCC. A search for the same motif but using a variable spacer length of 12 to 16 nt in length returned only *tsp* (tail specific protease; CPn0555) that exactly matches the consensus sequence with a spacer of 14 nt.

A search for the σ^{28} consensus sequence of the -35 box (TNAAG) returned 22 more sequences. None of these sequences contains the least conserved -10 box sequence GCC or CGA which was shown to be the preferred -10 sequence in a mutational analysis of the *hctB* promoter by Yu *et al.* (2006a). Three of the genes that are expressed late in the developmental cycle have been predicted to have σ^{28} promoter sequences by Miura *et al.* (2008) based on homology to the known *hctB* promoter -35 box AAAGTTT. The TSS data set argues against the existence of these promoter sites. For example, the predicted promoter of *adk* is located inside the transcript, CPn0332 is co-transcribed with CPn0333 and does not have an own promoter and the predicted *ltuB* -35 box is located at position -26 upstream of the TSS. Furthermore, these authors showed a homologous region upstream of the genes CPn0331, *omcA*, CPn0678 and *hctA*. Since these region starts at different distances relative to the corresponding TSS of these genes (CPn0331: -86, *omcA*: -33, CPn0678: -80, *hctA*: -65), it is unlikely that these sequences are part of a common promoter.

7 DISCUSSION

7.1 Transcriptome Analysis by dRNA-seq

Differential RNA-seq is a powerful method to determine the transcriptome of *Chlamydia*. It is suitable to catalogue all species of RNA, correct annotations, determine the transcriptional structure (i.e. TSS, 3'termini, operons) and allows quantifying gene expression in the full dynamic range of cellular RNA expression. Contrary to microarray approaches, RNA-seq is able to analyze all transcripts without *a priori* knowledge of the nucleotide sequences. In the meantime this method has been successfully applied to analyse transcriptomes of several organisms like *Helicobacter pylori* (Sharma *et al.* 2010), cyanobacteria *Synechocystis* (Weissenmayer *et al.* 2011), archea *Methanosarcina mazei* (Thomason and Storz 2010) and to identify novel regulatory RNAs in *Bacillus subtilis* (Herbig and Nieselt 2011), *Rhodobacter sphaeroides* (Pallen *et al.* 2010) as well as *Staphylococcus aureus* (Leroy *et al.* 2010).

Sequencing depth is an important parameter that has to be considered when analysing and interpreting results. Low abundant RNA molecules can be missed even if millions of sequencing reads are generated. Sequencing depth for *C. trachomatis* was not sufficient to determine operon structures whereas for *C. pneumoniae* a TSS or an affiliation to an operon could be determined for more than 80% of the annotated genes with only 3-fold sequencing depth compared to *C. trachomatis*.

One major drawback of the second generation sequencing platforms is the need for enzymatic modification (i. e. adapter ligation, polyA-tailing) and amplification of the cDNA. Enzymes used for cDNA library construction introduce a sequence dependent bias and therefore the total amount of certain RNA species does not necessarily correlate with the endogenous concentration. Comparison of sequencing results is only possible between cDNA samples that were treated exactly in the same way and were sequenced on the same platform (Linsen *et al.* 2009). The number of sequence reads is not reflecting the true biological abundance of the RNA molecule. Comparison of RNA-Seq data with microarray data has shown that microarrays give better results in terms of absolute quantification but sequencing is suitable better for relative quantification (Willenbrock *et al.* 2009).

The high abundance of sequence reads derived from Rho-independent promoters in both, the total RNA und TEX treated RNA samples, suggests that these RNA fragments are widely inaccessible for degrading endogenous nucleases as well as TEX. These fragments accumulated in the cell and were inaccessible for TEX and thereby could not be degraded, despite they do not have a 5'triphosphate. Consequently abundant transcripts enriched in the TEX treated libraries could also be derived from stable secondary structures. Hence, it is difficult to distinguish between real internal TSS and accumulated RNA fragments with secondary structure. For questionable internal TSS these have to be confirmed experimentally by a different method such as 5'RACE specific for native 5'ends.

7.2 Transcriptome Analysis of *Chlamydia trachomatis*

The availability of genome sequences for many different bacterial species and their successive annotation has initiated transcriptome analysis as a means to measure gene activity. Using microarray technology, annotated RNAs have been quantified very successfully for various biological conditions. A shortcoming of these probe-dependent approaches is, however, the indirect nature of the RNA quantification which allows only the detection and quantification of previously predicted RNAs with only limited alterations. Moreover, the precise starts of the 5'ends of most bacterial RNAs cannot be determined by conventional transcriptome analyses but require investigation of individual genes by 5'RACE or RNA run-off experiments. These limitations are especially relevant for bacteria that cannot be genetically manipulated and grown under laboratory conditions like obligate intracellular bacteria. The gene structure and function of this large group of bacteria including many important human pathogens remains therefore poorly understood.

Differential RNA-Seq has been applied to define the TSS of chlamydial genes at single nucleotide level. Due to the high resolution of the technology, a bona fide non-coding RNA ctrR0332 has been identified as a species-specific and most abundant RNA located in an annotated ORF (CTLon_0332) in *C. trachomatis.* In addition, numerous plasmid- and chromosomally encoded new sRNA species could be identified by sequencing and subsequently verified by Northern blot analysis.

The definition of 356 TSS in a single sequencing run, most of them to the single nucleotide level, demonstrates the power of the method applied. The enzymatic depletion of RNAs with 5'monophosphates was an essential step for the resolution of the TSS. Noise caused by RNA break-down products in many cases presumably would have complicated the unequivocal identification of TSS in the sequences obtained from untreated RNA. Such a precise mapping of TSS can now be used to characterize the respective promoters and other gene control elements by computational approaches.

Most of the discovered TSS clearly correlated with annotated ORFs. Others are located antisense and overlap unexpressed annotated ORFs like e.g. CTLon_0175, CTLon_0184 and CTLon_0239. It has been shown previously that these transcripts are expressed and contain ORFs (Jensen *et al.* 2006). However, numerous annotated transcripts could not be detected, some of them probably due to the limited depth of sequencing or the low expression. On the other hand, not all annotated 934 genes were expected to have a defined TSS, since *C. trachomatis* genes are frequently organized in operons such as the ten operons coding for a type three secretion system (Hefty and Stephens 2007). The operon structure can be analysed by comparing total RNA and enriched cDNA libraries. In enriched libraries a distinct TSS is present only for the first gene of an operon, whereas sequence reads are obtained for genes located downstream in the non-enriched libraries. Because of the polycistronic organisation of transcripts, the number of TSS is expected to be lower than the total ORF number. Collectively, the number of 356 identified TSS is likely to approach the real number of native transcript ends in *C. trachomatis.*

Differential transcriptomics of EB and RB, the developmentally different stages of the chlamydial developmental cycle, has been performed before from chromosomal genes but not from plasmid-encoded genes (Belland *et al.* 2003b; Nicholson *et al.* 2003). These studies used infected cells at different stages of the cycle as a source to prepare RNA. Since the infection course is asynchronous, later time points result in a mixture of EB and RB. To overcome this challenge and to avoid massive background of host RNA, libraries from purified EB and RB were generated that provide proof of principle for successful transcriptome analysis from purified chlamydial particles.

The relative abundances of RNAs from the different chlamydial developmental stages were compared with the published microarray expression data. Nicholson *et al.* (2003) classified clusters of temporal gene expression with 'early' (12 h p.i.), 'mid-late' (18 h p.i.), 'late' (24 h p.i.) and 'very late' (36 h p.i.). EB are metabolically inactive, and transcripts present in EB assure immediate translation in early infection events. Such transcripts are produced in the RB stage and stored in the EB, and therefore accumulate during the end of the infection cycle when the RB to EB transition is completed. These genes correlate with 'late' and 'very late' gene expression from microarray studies using RNA from time course experiments (table 4). Most of the transcripts overrepresented in EB are of unknown function with the interesting exception of the highly overrepresented RNA of an operon consisting of the four genes *CTLon_0833-0836*. This transcript encodes components of the T3SS such as the chaperone Scc2, a protein of unknown function, and the ATPase CopB and CopD, all of which is consistent with the crucial role of the T3SS early in infection (Hefty and Stephens 2007). In the 'mid-late' stage, the RB is the predominant form of *Chlamydia*. Of the 48 genes we found to be up-regulated in RB, 27 were classified as 'mid-late' and 5 as 'late' genes. Arguably, the use of different platforms and the temporal spectrum of RB-to-EB conversion limit a direct comparison of expression data, i.e. obtained by microarrays versus direct transcriptome sequencing of purified bacteria. Nevertheless, the semi-quantitative analysis of relative RNA abundances encoded by chromosomal genes from purified bacterial stages by deep sequencing correlated very well with the previous microarray analyses of infected cells. This indicates that the method is also suitable for relative quantitative measurements of RNA abundance.

Of all annotated ORFs, the most abundant transcript was that of ORF CTLon_0332. Its transcript quantity is comparable to the class of rRNAs or tRNAs, although it was previously annotated as mRNA. Precise mapping of the TSS and RNA length by Northern blot revealed two highly abundant non-coding RNAs. They are shorter than the annotated transcript, do not contain an ORF and are present only in *C. trachomatis* and the closely related mouse strain *C. muridarium*. It is noteworthy that CTLon_0332 is about tenfold overrepresented in EB which may indicate a role in the EB-to-RB transition.

The experimental prediction of 16 potential non-coding RNAs in chromosomal IGRs and 25 located antisense within annotated ORFs was another remarkable finding. A total of 9 sRNAs was validated on Northern blots, which firmly demonstrates that *C. trachomatis* expresses multiple sRNAs.

Since chlamydial genomes are small and highly optimized, intergenic regions are rare. The presence of these sRNAs may indicate a strong evolutionary selection for their expression in *C. trachomatis* due to their putative regulatory functions. The only previously known chlamydial sRNA also identified here was IhtA involved in translational regulation of histone-like protein Hc-1 (Grieshaber *et al.* 2006a), supporting an important role of sRNAs in regulating gene expression *C. trachomatis*. Most bacteria express the conserved RNA-binding protein Hfq, a pleiotropic regulator involved in the sRNA-mediated control of mRNA stability or translation (Valentin-Hansen *et al.* 2004). Chlamydia lack an Hfq homologue but the translational control of Hc-1 expression by the *trans*-encoded sRNA ihtA constitutes a riboregulatory event that is typically facilitated by Hfq in *E. coli*, the model bacterium of sRNA research. Whether a protein functionally similar to Hfq controls sRNA function in *C. trachomatis* remains to be shown. Clearly, the sRNAs identified here provide a departure point to discover a general sRNA-binding protein in Chlamydia by biochemical methods, e.g., aptamer-facilitated affinity purification (Said *et al.* 2009).

The results demonstrate that the chlamydial cryptic plasmid also expresses several sRNAs. Some were found in antisense orientation to ORFs implying a role in the regulation of plasmid-coded genes. One of these, pL2-sRNA1 was among the most abundant RNAs identified in *C. trachomatis* with more than 10,000 reads. pL2-sRNA1 codes for an RNA antisense to the gene for the putative partitioning protein PCTB_7, a ParA homolog. ParA proteins are membrane bound ATPases involved in the partitioning of chromosomes and plasmids during replication. It is therefore likely that pL2-sRNA1 is involved in the control of plasmid partitioning. The role of the other plasmid-encoded sRNAs is less clear. They may also be involved in the control of plasmid replication as has been shown for many different replicons (Weaver 2007). Yet another possibility is the regulation of chromosomal gene expression. Plasmid-free strains of *C. trachomatis* down-regulate the activity of chromosomal *glgA* gene encoding glycogen synthase leading to accumulation of glycogen in the chlamydial inclusion and an attenuated phenotype in mouse genital infection model (Carlson *et al.* 2008).

Whereas the general and precise identification of TSS will help to better understand the genome organization, the previously unknown highly abundant sRNAs described here raise several questions on their role in the physiology of *C. trachomatis*. Transcriptome analyses by deep-sequencing of other chlamydial species are on the way and will ultimately help to clarify whether species-specific abundant sRNAs are common among other members of the *Chlamydiales*.

7.2.1 Comparison with Tiling Array Study

During the course of this study a similar approach was published that used tiling arrays to analyse the transcription of sRNAs from intergenic regions during *C. trachomatis* growth and persistence (Abdelrahman *et al.* 2011). The authors identified numerous sRNAs in *C. trachomatis* and analysed expression of these sRNAs during normal growth compared to interferon-γ and carbenicillin

induced persistency. Based on the annotation of *C. trachomatis* D, all intergenic regions longer than 50 bp were represented on a microarray as 25 nt probes covering the whole intergenic region on both strands. With these tiling arrays the authors identified 33 novel sRNAs encoded on the bacterial chromosome. Overall, 14 of them could be verified by Northern blotting. This study identified 41 putative sRNAs from the sequencing data and 9 out of 12 sRNAs could be validated by Northern blotting. Since the major focus of this work was not the exhaustive description of sRNAs, not all candidates were tested by Northern blotting. Only nine out of the 33 sRNA candidates identified by Abdelrahman *et al.* (2011) were identified in this study.

The comparison clearly shows that the total number of sRNAs in *C. trachomatis* is in the same range of probably less than 50. The use of different RNA isolation methods and different methods to analyse RNA expression introduces a bias and leads to divergent results. As a consequence, in *C. pneumoniae* several sRNAs could be missed even though the sequencing depth was greater than for *C. trachomatis.* The total number of sRNAs identified in *C. pneumoniae* is smaller than the expected number compared to other bacterial species.

7.3 Transcriptome Analysis of *Chlamydia pneumoniae*

Analysis of the transcriptome of this obligate intracellular human pathogen gives important information for understanding gene expression and regulation. The data contributes to correct errors in the gene annotation and gives more detailed insight into the genome organization. Several findings show that *Chlamydia* share a high degree of transcriptome complexity. More specifically, this are the identification of novel non-coding transcripts in intergenic regions, sense and antisense to known genes, long 5'leader sequences of up to several hundred nucleotides and a high number of polycistronic transcripts with internal TSS. The discovery of alternative TSS gives insights into possible mechanisms of developmental stage specific gene regulation. The developmental use of alternative promoters has been demonstrated previously for the *C. trachomatis* cryptic plasmid gene pL2-02 (Albrecht *et al.* 2010; Ricci *et al.* 1993). Alternative promoters could be detected by stage specific transcription factors resulting in different lengths of mRNA leader sequences and the presence or absence of regulatory elements. Furthermore a longer leader sequence could form a stable secondary structure that masks the Shine-Dalgarno sequence and thereby prevents translation.

Genes more abundant in EB are mostly of unknown function and the mechanism of EB-to-RB conversion is poorly understood. RNA CPn0332 could play a role in this mechanism. It is 2.4-fold upregulated in EB compared to RB. Homology search and the fact that the abundant transcript, derived from the 3' part of the *ltuB* transcript, lacks an ORF suggest that the *ltuB* transcript encodes a protein and a highly abundant non-coding RNA of unknown function. In many proteobacteria 6S RNA was identified to be an abundant non-coding RNA that globally regulates transcription during growth phases by inhibition of standard sigma factor RNA polymerase (Wassarman and Saecker 2006) and thereby enhances alternative sigma factor activity. To test whether CPn0332 encodes a regulatory

RNA homologous to 6S RNA, short abundant antisense transcripts were searched that encode a pRNA like molecule as described for 6S RNA. However, we did not find a distinct short RNA like it was reported for other bacteria (Sharma *et al.* 2010; Wassarman and Saecker 2006). 6S RNA was reported to interact with RNA-Polymerase by mimicking an open promoter complex. In this work, an association with σ^{28} RNA polymerase could be excluded. Therefore, other RNA polymerases such as σ^{54} and σ^{66} remain to be tested as soon as antibodies are available for these proteins. However, the function of this highly abundant sRNA remains unclear. An identification of binding partners by aptamer-tagging technology could shed light on the biological role of this sRNA. Therefore, *in vitro* transcribed sRNA containing an aptamer tag at the 5' or 3' end is incubated with native lysates of *Chlamydia*. Subsequently, the tagged sRNA is captured by an aptamer-binding protein immobilized on a column and bound protein or RNA can be isolated, purified and characterized (Papenfort and Vogel 2010).

Comparative analysis of the sequencing data sets confirmed differentially expressed genes in EB and RB as has been shown previously by microarray experiments of infection time courses (Maurer *et al.* 2007; Miura *et al.* 2008). The majority of the 404 genes that were more than twofold upregulated in RB were housekeeping genes of central metabolism. The 53 genes more abundant in EB comprise 41 hypothetical proteins of unknown function. To get further inside into this group of proteins homologues of these genes and gene products were searched. Eleven of these 41 genes have no homologue in *C. trachomatis* and the majority of the remaining hypothetical proteins is only weakly conserved on amino acid sequence level. A search for these proteins in the Pfam protein family database (Finn *et al.* 2010) unveiled that 13 genes enriched in EB contain the IncA domain (Pfam PF04156). This domain is derived from the IncA protein which is secreted into the host cell cytoplasm via the T3SS and then integrates into the inclusion membrane. IncA is essential for fusion of inclusion vesicles (Fouquier d'Herouel *et al.* 2011). This set of IncA domain containing proteins includes CPn0585 which has been demonstrated to be localized in the inclusion membrane and interacts with host cell Rab-GTPases (Papenfort *et al.* 2010), as well as CPn1027 (Koo *et al.* 2006) and CPn0308 (Windbichler *et al.* 2008) which have also been shown to be localized in the inclusion membrane. A comparison of this semi-quantitative approach with microarray analysis of gene expression (Maurer *et al.* 2007) shows that the 13 genes enriched in EB which contain an IncA domain were found to have their maximum expression at 6h or 72h during the infection course which corresponds to the developmental stages of EB. Since EB lack transcriptional activity the mRNAs accumulated in EB are stored for immediate expression upon conversion into RB. Thus, the IncA domain containing proteins could be among the first effector proteins secreted into the host cell to be incorporated into the inclusion membrane.

The experimental identification of 531 primary TSS for the first time allowed to globally analyse promoter sequences. This analysis of promoters shows that most genes are controlled by the standard σ^{66} promoter that has a common motif which is less conserved than in other bacteria. Since no

common promoter motif could be identified for genes overrepresented in EB and RB, respectively, it is likely that differential expression of these subsets of genes is not accomplished by the use of alternative σ factors. Other sequence motifs may be present that act as *cis*-regulatory elements to control gene expression. However, since the *Cpn* genome is densely packed and intergenic regions are short, gene regulation could be effected by other mechanism such as sRNAs which have been identified in this work.

7.4 Comparison between *C. trachomatis* and *C. pneumoniae* Transcriptomes

In both species that were subjected to deep sequencing analysis a set of novel non-coding RNAs was identified. Only three of these novel sRNAs are conserved between *C. trachomatis* and *C. pneumoniae.* These are CPIG0564, CPIG0692, and CPIG0701 (the previously described sRNA ihtA). All of them are encoded in intergenic regions between highly conserved genes. Some of the sRNA genes of *C. pneumoniae* are encoded adjacent to genes that have no homologues in *C. trachomatis* (CPIG0142, CPS0457, CPIG0953, CPIG0954). For some of them a homologous gene is present but for the homologous putative sRNA locus no sequence reads were obtained (CPAS0152, CPAS0294, CPIG0294, CPIG0397, CPS0657). The sRNA CPIG0207 is encoded in an intergenic region which is absent in *C. trachomatis* since the intergenic space of the two adjacent homologous genes has a length of only 18 nt. The low conservation of non-coding RNAs among *C. trachomatis* and *C. pneumoniae* suggests a lower conservation of non-protein-coding regions than protein coding ORFs. An *in silico* search for homologous sequences of conserved sRNAs in other species unveiled that ihtA is the only sRNA that is highly conserved among all *Chlamydiales*. Figure 22 shows an alignment of the *ihtA* gene sequences from six different species. Target sequences of the mRNA of HctA that were predicted using TargetRNA algorithm (Grech *et al.* 2008) are given on top (*C. pneumoniae*) and below (*C. trachomatis*) the alignment, respectively, including the base pairing. The predicted target binding site incorporates the start codon and several bases upstream. Thereby, binding of the ribosome and translation of the HctA protein could be prevented.

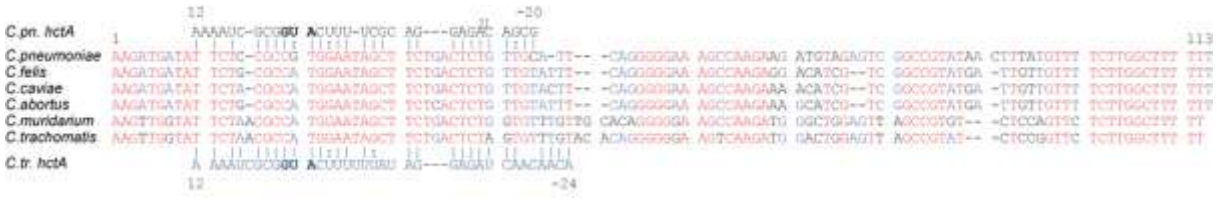

Figure 22: The conserved sRNA *ihtA*.
The *ihtA* gene is one out of three sRNA genes that have been found to be conserved in *C. trachomatis* and *C. pneumoniae.* Furthermore, this sRNA is conserved among all *Chlamydiae*. The alignment of this gene from six different species shows a high sequence homology. The section of the experimentally validated target gene *hctA* that is supposed to contain the putative target site is shown on top (for *C. pneumoniae)* and at the bottom (for *C. trachomatis*), respectively. The start codons are depicted in bold letters.

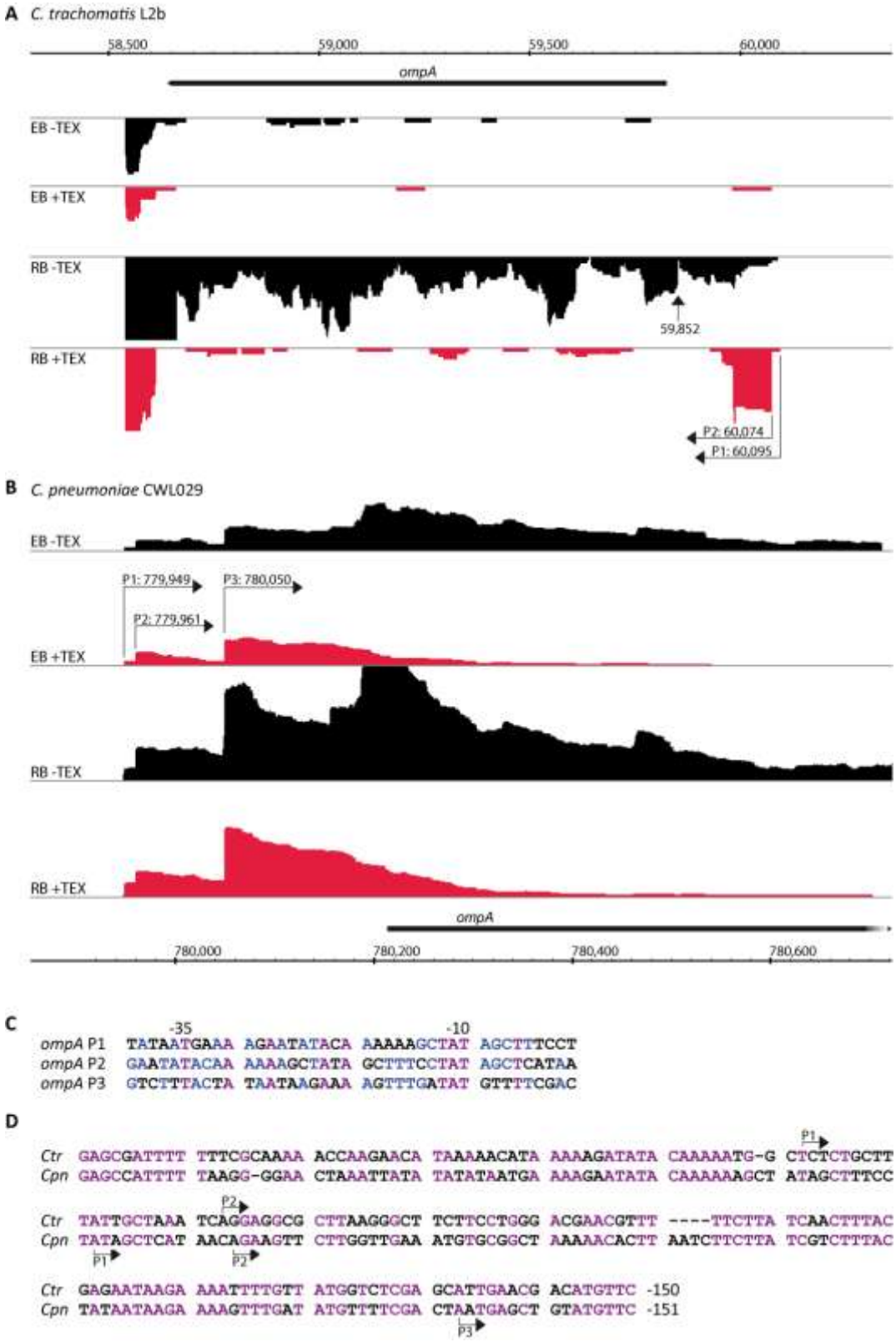

Figure 23: TSS of the major outer membrane protein gene *ompA*.
Sequence read distribution is shown for the ompA gene of *C. trachomatis* (minus strand) and *C. pneumoniae* (plus strand) for EB and RB and untreated libraries (black) as well as TEX treated libraries (red). (A) The ompA gene of *C. trachomatis* has a single major TSS at position 60,074 (P2, -248 to translational start), a minor TSS represented by only 1 cDNA read (P1 -269 to TSS), and a processing site at position 59,852 (-25) as earlier described (see text for details). (B) Sequencing data suggests three TSS for the *C. pneumoniae ompA* gene at the indicated positions P1 to P3 which correspond to positions -266, -254 and -165 relative to the translational start site, respectively, as indicated by arrows. (C) Alignment of promoter sequences of three *C. pneumoniae ompA* TSS P1 to P3 show poor conservation of -10 and -35 box. (D) Alignment of promoter sequences of *C. trachomatis* and *C. pneumoniae* with TSS depicted as arrows for *C. pneumoniae* (on top of alignment) and *C. trachomatis* (below alignment).

Promoters of conserved genes among *C. trachomatis* and *C. pneumoniae* often share the same TSS. However for some important genes the transcriptional start differs between the species. The *ompA* gene is coding for the major outer membrane protein of *Chlamydia* which constitutes more than 60% of the outer membrane total protein content (Caldwell *et al.* 1981). With a total of 3,749 sequence reads *ompA* was the second most abundant protein coding gene after *omcB* (9,009 reads) in terms of read numbers per gene. Figure 23 shows the sequencing read distribution of *ompA* of *C. trachomatis* and *C. pneumoniae*, respectively, including the TSS depicted by arrows. Note that the Rho-independent terminator downstream of the *ompA* coding sequence is highly enriched in the sequencing reads due to a thermodynamic highly stable secondary structure The *C. trachomatis ompA* gene was first described to have two tandem promoters which give rise to two transcripts that are differentially expressed during the life cycle (Stephens *et al.* 1988). Douglas and Hatch (1995) could show that *in vitro* transcription occurs only from the longer TSS (Figure 23A, position 60,074). They suggest that the shorter fragment is processed from the longer primary transcript. The sequencing read distribution confirms this assumption. Only one major primary TSS was found upstream of the *ompA* gene at position 60,074 (P2, figure 23A). A minor TSS represented by only one cDNA sequence read is located 26 bp upstream (P1, figure 23A). The -25 position (at 59,852) seems to be a processing site since there are no sequencing reads starting at this position in the TEX treated libraries. In contrast, a number of transcripts starting at this position exist in the untreated library. Interestingly, the *ompA* gene seems to have three distinct TSS in *C. pneumoniae* (see figure 23B), all of them harbouring a σ^{66} promoter that is partially conserved (Figure 23C). By inspecting the region upstream of the coding sequence in the libraries enriched for primary transcripts, three start sites are present. There are two minor TSS located at -266 and -254 (positions 779,949 and 779,961, respectively) and a major TSS located at -165 (position 780,050) identifiable. Interestingly, only P2 is conserved among the two species, even though -10 and -35 promoter regions are partially conserved among P1 and P3.For both species there are more sequence reads in the RB libraries than in the EB libraries which is in agreement with earlier studies (Stephens *et al.* 1988).

7.5 Future Perspectives

The data sets derived from this work can help to further examine the mechanisms of gene regulation in *Chlamydia* even though a genetic system is still not available. The first chlamydial sRNA has been identified by a heterologous screen in *E. coli.* This demonstrates that expression as well as interaction of sRNA and putative targets of *Chlamydia* is possible in E. coli. Other sRNAs and putative targets (i.e. derived from *in silico* predictions) can be tested in such systems. Fluorescence based reporter systems are available to easily get a quantitative measure of gene expression by heterologous expression (Urban and Vogel 2009).

For biochemical identification of targets bound by RNA, aptamer tagging technology has become a well established method (Said *et al.* 2009). Furthermore, co-sedimentation in density gradients can be applied to get a first idea of whether a sRNA is interacting with proteins.

The major drawbacks of RNA-seq are the enzymatic treatment and amplification steps during cDNA preparation that are thought to introduce a bias (Hansen *et al.* 2010; Linsen *et al.* 2009). Within the next years the so called 'third generation sequencing' techniques are entering the market. By using real single nucleotide sequencing these methods are supposed to circumvent any amplification step that could introduce a bias. Theoretically, these machines will be able to sequence all molecules of a single cell (host including bacteria) which will lead to a better understanding of pathogen-host-interaction. Additionally, protocols are being developed for existing sequencers (i.e. Illumina Genome Analyser) that avoid amplification steps (Mamanova *et al.* 2010). Furthermore, the sequencing throughput is steadily increasing while the costs per base are tremendously decreasing. In parallel computational biology is developing faster algorithms and software packages to efficiently analyse the large data sets derived from sequencing runs. Furthermore computing speed and storage capacity increases exponentially over time and thus, massively more sequences can be obtained in future. This facilitates the identification of rare transcripts, rarely used alternative TSS, and better coverage of polycistronic transcripts.

Besides the acute infection cycle, *Chlamydia* can become persistent which results in a long lasting relationship of bacteria and host cell. Transcriptomes of persistent infection models will be analysed and compared to the acute infection presented in this work. Since the mechanism of persistence is largely unknown this will help to identify the genes that are involved in control of persistent growth. Moreover the method could be used to determine transcriptomes of other species of the phylum *Chlamydiae* to get insights into evolutionary transcriptomics.

8 REFERENCES

Abdelrahman, Y. M., Rose, L. A. and Belland, R. J. (2011). "Developmental expression of non-coding RNAs in Chlamydia trachomatis during normal and persistent growth." *Nucleic Acids Res* **39**(5): 1843-1854.

Akers, J. C. and Tan, M. (2006). "Molecular mechanism of tryptophan-dependent transcriptional regulation in Chlamydia trachomatis." *J Bacteriol* **188**(12): 4236-4243.

Albrecht, M., Sharma, C. M., Reinhardt, R., Vogel, J. and Rudel, T. (2010). "Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome." *Nucleic Acids Res* **38**(3): 868-877.

Bailey, T. L. and Elkan, C. (1994). "Fitting a mixture model by expectation maximization to discover motifs in biopolymers." *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

Balin, B. J., Gerard, H. C., Arking, E. J., Appelt, D. M., Branigan, P. J., Abrams, J. T., Whittum-Hudson, J. A. and Hudson, A. P. (1998). "Identification and localization of Chlamydia pneumoniae in the Alzheimer's brain." *Med Microbiol Immunol* **187**(1): 23-42.

Barrios, H., Valderrama, B. and Morett, E. (1999). "Compilation and analysis of sigma(54)-dependent promoter sequences." *Nucleic Acids Res* **27**(22): 4305-4313.

Barry, C. E., 3rd, Hayes, S. F. and Hackstadt, T. (1992). "Nucleoid condensation in Escherichia coli that express a chlamydial histone homolog." *Science* **256**(5055): 377-379.

Beeckman, D. S. and Vanrompay, D. C. (2009). "Zoonotic Chlamydophila psittaci infections from a clinical perspective." *Clin Microbiol Infect* **15**(1): 11-17.

Belland, R. J., Nelson, D. E., Virok, D., Crane, D. D., Hogan, D., Sturdevant, D., Beatty, W. L. and Caldwell, H. D. (2003a). "Transcriptome analysis of chlamydial growth during IFN-gamma-mediated persistence and reactivation." *Proc Natl Acad Sci U S A* **100**(26): 15971-15976.

Belland, R. J., Zhong, G., Crane, D. D., Hogan, D., Sturdevant, D., Sharma, J., Beatty, W. L. and Caldwell, H. D. (2003b). "Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis." *Proc Natl Acad Sci U S A* **100**(14): 8478-8483.

Berezikov, E., Thuemmler, F., van Laake, L. W., Kondova, I., Bontrop, R., Cuppen, E. and Plasterk, R. H. (2006). "Diversity of microRNAs in human and chimpanzee brain." *Nat Genet* **38**(12): 1375-1377.

Borges, V., Ferreira, R., Nunes, A., Nogueira, P., Borrego, M. J. and Gomes, J. P. (2010). "Normalization strategies for real-time expression data in Chlamydia trachomatis." *Journal of Microbiological Methods* **82**(3): 256-264.

Brantl, S. (2007). "Regulatory mechanisms employed by cis-encoded antisense RNAs." *Curr Opin Microbiol* **10**(2): 102-109.

Brickman, T. J., Barry, C. E., 3rd and Hackstadt, T. (1993). "Molecular cloning and expression of hctB encoding a strain-variant chlamydial histone-like protein with DNA-binding activity." *J Bacteriol* **175**(14): 4274-4281.

Bush, R. M. and Everett, K. D. (2001). "Molecular evolution of the Chlamydiaceae." *Int J Syst Evol Microbiol* **51**(Pt 1): 203-220.

Byrne, G. I. and Ojcius, D. M. (2004). "Chlamydia and apoptosis: life and death decisions of an intracellular pathogen." *Nat Rev Microbiol* **2**(10): 802-808.

Caldwell, H. D., Kromhout, J. and Schachter, J. (1981). "Purification and partial characterization of the

major outer membrane protein of Chlamydia trachomatis." *Infect Immun* **31**(3): 1161-1176.

Campbell, L. A. and Kuo, C. C. (2004). "Chlamydia pneumoniae--an infectious risk factor for atherosclerosis?" *Nat Rev Microbiol* **2**(1): 23-32.

Carabeo, R. A. and Hackstadt, T. (2001). "Isolation and characterization of a mutant Chinese hamster ovary cell line that is resistant to Chlamydia trachomatis infection at a novel step in the attachment process." *Infect Immun* **69**(9): 5899-5904.

Carlson, J. H., Whitmire, W. M., Crane, D. D., Wicke, L., Virtaneva, K., Sturdevant, D. E., Kupko, J. J., 3rd, Porcella, S. F., Martinez-Orengo, N., Heinzen, R. A., Kari, L. and Caldwell, H. D. (2008). "The Chlamydia trachomatis plasmid is a transcriptional regulator of chromosomal genes and a virulence factor." *Infect Immun* **76**(6): 2273-2283.

Chaturvedi, A. K., Gaydos, C. A., Agreda, P., Holden, J. P., Chatterjee, N., Goedert, J. J., Caporaso, N. E. and Engels, E. A. (2010). "Chlamydia pneumoniae infection and risk for lung cancer." *Cancer Epidemiol Biomarkers Prev* **19**(6): 1498-1505.

Clifton, D. R., Fields, K. A., Grieshaber, S. S., Dooley, C. A., Fischer, E. R., Mead, D. J., Carabeo, R. A. and Hackstadt, T. (2004). "A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin." *Proc Natl Acad Sci U S A* **101**(27): 10166-10171.

Dam, P., Olman, V., Harris, K., Su, Z. and Xu, Y. (2007). "Operon prediction using both genome-specific and general genomic information." *Nucleic Acids Res* **35**(1): 288-298.

Darville, T., O'Neill, J. M., Andrews, C. W., Jr., Nagarajan, U. M., Stahl, L. and Ojcius, D. M. (2003). "Toll-like receptor-2, but not Toll-like receptor-4, is essential for development of oviduct pathology in chlamydial genital tract infection." *J Immunol* **171**(11): 6187-6197.

Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., Keller, K., Novichkov, P. S., Dubchak, I. L., Alm, E. J. and Arkin, A. P. (2010). "MicrobesOnline: an integrated portal for comparative and functional genomics." *Nucleic Acids Res* **38**(Database issue): D396-400.

Douglas, A. L. and Hatch, T. P. (1995). "Functional analysis of the major outer membrane protein gene promoters of Chlamydia trachomatis." *J Bacteriol* **177**(21): 6286-6289.

Douglas, A. L. and Hatch, T. P. (2000). "Expression of the transcripts of the sigma factors and putative sigma factor regulators of Chlamydia trachomatis L2." *Gene* **247**(1-2): 209-214.

Engel, J. N. and Ganem, D. (1987). "Chlamydial rRNA operons: gene organization and identification of putative tandem promoters." *J Bacteriol* **169**(12): 5678-5685.

Everett, K. D., Bush, R. M. and Andersen, A. A. (1999). "Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms." *Int J Syst Bacteriol* **49 Pt 2**: 415-440.

Fahr, M. J., Douglas, A. L., Xia, W. and Hatch, T. P. (1995). "Characterization of late gene promoters of Chlamydia trachomatis." *J Bacteriol* **177**(15): 4252-4260.

Fahr, M. J., Sriprakash, K. S. and Hatch, T. P. (1992). "Convergent and overlapping transcripts of the Chlamydia trachomatis 7.5-kb plasmid." *Plasmid* **28**(3): 247-257.

Fan, T., Lu, H., Hu, H., Shi, L., McClarty, G. A., Nance, D. M., Greenberg, A. H. and Zhong, G. (1998). "Inhibition of apoptosis in chlamydia-infected cells: blockade of mitochondrial cytochrome c release and caspase activation." *J Exp Med* **187**(4): 487-496.

Farencena, A., Comanducci, M., Donati, M., Ratti, G. and Cevenini, R. (1997). "Characterization of a new

isolate of Chlamydia trachomatis which lacks the common plasmid and has properties of biovar trachoma." *Infect Immun* **65**(7): 2965-2969.

Fields, K. A., Mead, D. J., Dooley, C. A. and Hackstadt, T. (2003). "Chlamydia trachomatis type III secretion: evidence for a functional apparatus during early-cycle development." *Mol Microbiol* **48**(3): 671-683.

Filiatrault, M. J., Stodghill, P. V., Bronstein, P. A., Moll, S., Lindeberg, M., Grills, G., Schweitzer, P., Wang, W., Schroth, G. P., Luo, S. J., Khrebtukova, I., Yang, Y., Thannhauser, T., Butcher, B. G., Cartinhour, S. and Schneider, D. J. (2010). "Transcriptome Analysis of Pseudomonas syringae Identifies New Genes, Noncoding RNAs, and Antisense Activity." *Journal of Bacteriology* **192**(9): 2359-2372.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. and Bateman, A. (2010). "The Pfam protein families database." *Nucleic Acids Res* **38**(Database issue): D211-222.

Fischer, S. F., Vier, J., Kirschnek, S., Klos, A., Hess, S., Ying, S. and Hacker, G. (2004). "Chlamydia inhibit host cell apoptosis by degradation of proapoptotic BH3-only proteins." *J Exp Med* **200**(7): 905-916.

Fouquier d'Herouel, A., Wessner, F., Halpern, D., Ly-Vu, J., Kennedy, S. P., Serror, P., Aurell, E. and Repoila, F. (2011). "A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen Enterococcus faecalis." *Nucleic Acids Res*.

Grayston, J. T. (2000). "Background and current knowledge of Chlamydia pneumoniae and atherosclerosis." *J Infect Dis* **181 Suppl 3**: S402-410.

Grayston, J. T., Campbell, L. A., Kuo, C. C., Mordhorst, C. H., Saikku, P., Thom, D. H. and Wang, S. P. (1990). "A new respiratory tract pathogen: Chlamydia pneumoniae strain TWAR." *J Infect Dis* **161**(4): 618-625.

Grech, B., Mathews, S. and Timms, P. (2008). "Phylogenetic comparison of the known Chlamydia trachomatis sigma(66) promoters across to Chlamydia pneumoniae and Chlamydia caviae identifies seven poorly conserved promoters." *Res Microbiol* **159**(7-8): 550-556.

Grieshaber, N. A., Grieshaber, S. S., Fischer, E. R. and Hackstadt, T. (2006a). "A small RNA inhibits translation of the histone-like protein Hc1 in Chlamydia trachomatis." *Mol Microbiol* **59**(2): 541-550.

Grieshaber, N. A., Sager, J. B., Dooley, C. A., Hayes, S. F. and Hackstadt, T. (2006b). "Regulation of the Chlamydia trachomatis histone H1-like protein Hc2 is IspE dependent and IhtA independent." *J Bacteriol* **188**(14): 5289-5292.

Grimwood, J., Olinger, L. and Stephens, R. S. (2001). "Expression of Chlamydia pneumoniae polymorphic membrane protein family genes." *Infect Immun* **69**(4): 2383-2389.

Gussmann, J., Al-Younes, H. M., Braun, P. R., Brinkmann, V. and Meyer, T. F. (2008). "Long-term effects of natural amino acids on infection with Chlamydia trachomatis." *Microb Pathog* **44**(5): 438-447.

Hackstadt, T., Baehr, W. and Ying, Y. (1991). "Chlamydia trachomatis developmentally regulated protein is homologous to eukaryotic histone H1." *Proc Natl Acad Sci U S A* **88**(9): 3937-3941.

Hahn, D. L., Dodge, R. W. and Golubjatnikov, R. (1991). "Association of Chlamydia pneumoniae (strain TWAR) infection with wheezing, asthmatic bronchitis, and adult-onset asthma." *JAMA* **266**(2): 225-230.

Hammerschlag, M. R. (2000). "Chlamydia pneumoniae and the lung." *Eur Respir J* **16**(5): 1001-1007.

Hansen, K. D., Brenner, S. E. and Dudoit, S. (2010). "Biases in Illumina transcriptome sequencing caused by random hexamer priming." *Nucleic Acids Res* **38**(12): e131.

Hefty, P. S. and Stephens, R. S. (2007). "Chlamydial type III secretion system is encoded on ten operons

preceded by sigma 70-like promoter elements." *J Bacteriol* **189**(1): 198-206.

Herbig, A. and Nieselt, K. (2011). "nocoRNAc: Characterization of non-coding RNAs in prokaryotes." *Bmc Bioinformatics* **12**.

Hogan, R. J., Mathews, S. A., Mukhopadhyay, S., Summersgill, J. T. and Timms, P. (2004). "Chlamydial persistence: beyond the biphasic paradigm." *Infect Immun* **72**(4): 1843-1855.

Irmler, M., Thome, M., Hahne, M., Schneider, P., Hofmann, K., Steiner, V., Bodmer, J. L., Schroter, M., Burns, K., Mattmann, C., Rimoldi, D., French, L. E. and Tschopp, J. (1997). "Inhibition of death receptor signals by cellular FLIP." *Nature* **388**(6638): 190-195.

Jensen, K. T., Petersen, L., Falk, S., Iversen, P., Andersen, P., Theisen, M. and Krogh, A. (2006). "Novel overlapping coding sequences in Chlamydia trachomatis." *FEMS Microbiol Lett* **265**(1): 106-117.

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S. (1999). "Comparative genomes of Chlamydia pneumoniae and C. trachomatis." *Nat Genet* **21**(4): 385-389.

Karunakaran, K. P., Noguchi, Y., Read, T. D., Cherkasov, A., Kwee, J., Shen, C., Nelson, C. C. and Brunham, R. C. (2003). "Molecular analysis of the multiple GroEL proteins of Chlamydiae." *J Bacteriol* **185**(6): 1958-1966.

Knodler, L. A., Finlay, B. B. and Steele-Mortimer, O. (2005). "The Salmonella effector protein SopB protects epithelial cells from apoptosis by sustained activation of Akt." *J Biol Chem* **280**(10): 9058-9064.

Koo, I. C. and Stephens, R. S. (2003). "A developmentally regulated two-component signal transduction system in Chlamydia." *J Biol Chem* **278**(19): 17314-17319.

Koo, I. C., Walthers, D., Hefty, P. S., Kenney, L. J. and Stephens, R. S. (2006). "ChxR is a transcriptional activator in Chlamydia." *Proc Natl Acad Sci U S A* **103**(3): 750-755.

Kuo, C. C., Shor, A., Campbell, L. A., Fukushi, H., Patton, D. L. and Grayston, J. T. (1993). "Demonstration of Chlamydia pneumoniae in atherosclerotic lesions of coronary arteries." *J Infect Dis* **167**(4): 841-849.

Laurila, A. L., Von Hertzen, L. and Saikku, P. (1997). "Chlamydia pneumoniae and chronic lung diseases." *Scand J Infect Dis Suppl* **104**: 34-36.

Leroy, Q., Lebrigand, K., Armougom, F., Barbry, P., Thiery, R. and Raoult, D. (2010). "Coxiella burnetii Transcriptional Analysis Reveals Serendipity Clusters of Regulation in Intracellular Bacteria." *PLoS ONE* **5**(12).

Linsen, S. E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., Fritz, B., Wyman, S. K., de Bruijn, E., Voest, E. E., Kuersten, S., Tewari, M. and Cuppen, E. (2009). "Limitations and possibilities of small RNA digital gene expression profiling." *Nat Methods* **6**(7): 474-476.

Liu, M. Y., Gui, G., Wei, B., Preston, J. F., 3rd, Oakford, L., Yuksel, U., Giedroc, D. P. and Romeo, T. (1997). "The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli." *J Biol Chem* **272**(28): 17502-17510.

Lonetto, M., Gribskov, M. and Gross, C. A. (1992). "The sigma 70 family: sequence conservation and evolutionary relationships." *J Bacteriol* **174**(12): 3843-3849.

Mamanova, L., Andrews, R. M., James, K. D., Sheridan, E. M., Ellis, P. D., Langford, C. F., Ost, T. W., Collins, J. E. and Turner, D. J. (2010). "FRT-seq: amplification-free, strand-specific transcriptome sequencing." *Nat Methods* **7**(2): 130-132.

Mao, F., Dam, P., Chou, J., Olman, V. and Xu, Y. (2009). "DOOR: a database for prokaryotic operons."

Nucleic Acids Res **37**(Database issue): D459-463.

Mathews, S. A., Douglas, A., Sriprakash, K. S. and Hatch, T. P. (1993). "In vitro transcription in Chlamydia psittaci and Chlamydia trachomatis." *Mol Microbiol* **7**(6): 937-946.

Mathews, S. A. and Timms, P. (2000). "Identification and mapping of sigma-54 promoters in Chlamydia trachomatis." *J Bacteriol* **182**(21): 6239-6242.

Mathews, S. A., Volp, K. M. and Timms, P. (1999). "Development of a quantitative gene expression assay for Chlamydia trachomatis identified temporal expression of sigma factors." *FEBS Lett* **458**(3): 354-358.

Matsumoto, A., Izutsu, H., Miyashita, N. and Ohuchi, M. (1998). "Plaque formation by and plaque cloning of Chlamydia trachomatis biovar trachoma." *J Clin Microbiol* **36**(10): 3013-3019.

Maurer, A. P., Mehlitz, A., Mollenkopf, H. J. and Meyer, T. F. (2007). "Gene expression profiles of Chlamydophila pneumoniae during the developmental cycle and iron depletion-mediated persistence." *PLoS Pathog* **3**(6): e83.

McClarty, G. (1994). "Chlamydiae and the biochemistry of intracellular parasitism." *Trends Microbiol* **2**(5): 157-164.

McCrea, J. F., Epstein, R. S. and Barry, W. H. (1961). "Use of potassium tartrate for equilibrium density-gradient centrifugation of animal viruses." *Nature* **189**: 220-221.

Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J. and Hess, W. R. (2011). "An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803." *Proc Natl Acad Sci U S A* **108**(5): 2124-2129.

Miura, K., Toh, H., Hirakawa, H., Sugii, M., Murata, M., Nakai, K., Tashiro, K., Kuhara, S., Azuma, Y. and Shirai, M. (2008). "Genome-wide analysis of Chlamydophila pneumoniae gene expression at the late stage of infection." *DNA Res* **15**(2): 83-91.

Miyairi, I. and Byrne, G. I. (2006). "Chlamydia and programmed cell death." *Curr Opin Microbiol* **9**(1): 102-108.

Moll, I., Grill, S., Gualerzi, C. O. and Blasi, U. (2002). "Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control." *Mol Microbiol* **43**(1): 239-246.

Morita, T., Maki, K. and Aiba, H. (2005). "RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs." *Genes Dev* **19**(18): 2176-2186.

Moulder, J. W. (1982). "The relation of basic biology to pathogenic potential in the genus Chlamydia." *Infection* **10 Suppl 1**: S10-18.

Moulder, J. W. (1991). "Interaction of chlamydiae and host cells in vitro." *Microbiol Rev* **55**(1): 143-190.

Mukhopadhyay, S., Good, D., Miller, R. D., Graham, J. E., Mathews, S. A., Timms, P. and Summersgill, J. T. (2006). "Identification of Chlamydia pneumoniae proteins in the transition from reticulate to elementary body formation." *Mol Cell Proteomics* **5**(12): 2311-2318.

Narberhaus, F. (1999). "Negative regulation of bacterial heat shock genes." *Mol Microbiol* **31**(1): 1-8.

Newhall, W. J., Batteiger, B. and Jones, R. B. (1982). "Analysis of the human serological response to proteins of Chlamydia trachomatis." *Infect Immun* **38**(3): 1181-1189.

Nicholson, T. L., Olinger, L., Chong, K., Schoolnik, G. and Stephens, R. S. (2003). "Global stage-specific gene regulation during the developmental cycle of Chlamydia trachomatis." *J Bacteriol* **185**(10): 3179-3189.

Nieuwenhuis, R. F., Ossewaarde, J. M., Gotz, H. M., Dees, J., Thio, H. B., Thomeer, M. G., den Hollander, J. C., Neumann, M. H. and van der Meijden, W. I. (2004). "Resurgence of lymphogranuloma venereum in Western Europe: an outbreak of Chlamydia trachomatis serovar l2 proctitis in The Netherlands among men who have sex with men." *Clin Infect Dis* **39**(7): 996-1003.

O'Connell, C. M., AbdelRahman, Y. M., Green, E., Darville, H. K., Saira, K., Smith, B., Darville, T., Scurlock, A. M., Meyer, C. R. and Belland, R. J. (2011). "Toll-like receptor 2 activation by Chlamydia trachomatis is plasmid dependent, and plasmid-responsive chromosomal loci are coordinately regulated in response to glucose limitation by C. trachomatis but not by C. muridarum." *Infect Immun* **79**(3): 1044-1056.

O'Connell, C. M., Ingalls, R. R., Andrews, C. W., Jr., Scurlock, A. M. and Darville, T. (2007). "Plasmid-deficient Chlamydia muridarum fail to induce immune pathology and protect against oviduct disease." *J Immunol* **179**(6): 4027-4034.

O'Connell, C. M. and Nicks, K. M. (2006). "A plasmid-cured Chlamydia muridarum strain displays altered plaque morphology and reduced infectivity in cell culture." *Microbiology* **152**(Pt 6): 1601-1607.

Pallen, M. J., Loman, N. J. and Penn, C. W. (2010). "High-throughput sequencing and clinical microbiology: progress, opportunities and challenges." *Current Opinion in Microbiology* **13**(5): 625-631.

Palmer, L. and Falkow, S. (1986). "A common plasmid of Chlamydia trachomatis." *Plasmid* **16**(1): 52-62.

Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M. and Vogel, J. (2010). "Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA." *Proc Natl Acad Sci U S A* **107**(47): 20435-20440.

Papenfort, K. and Vogel, J. (2010). "Regulatory RNA in bacterial pathogens." *Cell Host Microbe* **8**(1): 116-127.

Pedersen, L. B., Birkelund, S. and Christiansen, G. (1994). "Interaction of the Chlamydia trachomatis histone H1-like protein (Hc1) with DNA and RNA causes repression of transcription and translation in vitro." *Mol Microbiol* **11**(6): 1085-1098.

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

Perara, E., Ganem, D. and Engel, J. N. (1992). "A developmentally regulated chlamydial gene with apparent homology to eukaryotic histone H1." *Proc Natl Acad Sci U S A* **89**(6): 2125-2129.

Perez-Rueda, E. and Collado-Vides, J. (2000). "The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12." *Nucleic Acids Res* **28**(8): 1838-1847.

Peterson, E. M., Markoff, B. A., Schachter, J. and de la Maza, L. M. (1990). "The 7.5-kb plasmid present in Chlamydia trachomatis is not essential for the growth of this microorganism." *Plasmid* **23**(2): 144-148.

Pickett, M. A., Everson, J. S., Pead, P. J. and Clarke, I. N. (2005). "The plasmids of Chlamydia trachomatis and Chlamydophila pneumoniae (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents." *Microbiology* **151**(Pt 3): 893-903.

Price, M. N., Huang, K. H., Alm, E. J. and Arkin, A. P. (2005). "A novel method for accurate operon predictions in all sequenced prokaryotes." *Nucleic Acids Res* **33**(3): 880-892.

Ramos, J. L., Gallegos, M. T., Marques, S., Ramos-Gonzalez, M. I., Espinosa-Urgel, M. and Segura, A. (2001). "Responses of Gram-negative bacteria to certain environmental stressors." *Curr Opin Microbiol* **4**(2): 166-171.

Read, T. D., Brunham, R. C., Shen, C., Gill, S. R., Heidelberg, J. F., White, O., Hickey, E. K., Peterson, J.,

Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S. L., Eisen, J. and Fraser, C. M. (2000). "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39." *Nucleic Acids Res* **28**(6): 1397-1406.

Ren, Q., Robertson, S. J., Howe, D., Barrows, L. F. and Heinzen, R. A. (2003). "Comparative DNA microarray analysis of host cell transcriptional responses to infection by Coxiella burnetii or Chlamydia trachomatis." *Ann N Y Acad Sci* **990**: 701-713.

Ricci, S., Cevenini, R., Cosco, E., Comanducci, M., Ratti, G. and Scarlato, V. (1993). "Transcriptional analysis of the Chlamydia trachomatis plasmid pCT identifies temporally regulated transcripts, anti-sense RNA and sigma 70-selected promoters." *Mol Gen Genet* **237**(3): 318-326.

Ripa, T. and Nilsson, P. (2006). "A variant of Chlamydia trachomatis with deletion in cryptic plasmid: implications for use of PCR diagnostic tests." *Euro Surveill* **11**(11): E061109 061102.

Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H. and Vogel, J. (2009). "In vivo expression and purification of aptamer-tagged small RNA regulators." *Nucleic Acids Res* **37**(20): e133.

Schachter, J., Stephens, R. S., Timms, P., Kuo, C., Bavoil, P. M., Birkelund, S., Boman, J., Caldwell, H., Campbell, L. A., Chernesky, M., Christiansen, G., Clarke, I. N., Gaydos, C., Grayston, J. T., Hackstadt, T., Hsia, R., Kaltenboeck, B., Leinonnen, M., Ojcius, D., McClarty, G., Orfila, J., Peeling, R., Puolakkainen, M., Quinn, T. C., Rank, R. G., Raulston, J., Ridgeway, G. L., Saikku, P., Stamm, W. E., Taylor-Robinson, D. T., Wang, S. P. and Wyrick, P. B. (2001). "Radical changes to chlamydial taxonomy are not necessary just yet." *Int J Syst Evol Microbiol* **51**(Pt 1): 249; author reply 251-243.

Schaumburg, C. S. and Tan, M. (2003). "Mutational analysis of the Chlamydia trachomatis dnaK promoter defines the optimal -35 promoter element." *Nucleic Acids Res* **31**(2): 551-555.

Schaumburg, C. S. and Tan, M. (2006). "Arginine-dependent gene regulation via the ArgR repressor is species specific in chlamydia." *J Bacteriol* **188**(3): 919-927.

Schmiel, D. H. and Wyrick, P. B. (1994). "Another putative heat-shock gene and aminoacyl-tRNA synthetase gene are located upstream from the grpE-like and dnaK-like genes in Chlamydia trachomatis." *Gene* **145**(1): 57-63.

Scidmore-Carlson, M. and Hackstadt, T. (2000). "Chlamydia internalization and intracellular fate." *Subcell Biochem* **33**: 459-478.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R., Stadler, P. F. and Vogel, J. (2010). "The primary transcriptome of the major human pathogen Helicobacter pylori." *Nature* **464**(7286): 250-255.

Shaw, A. C., Gevaert, K., Demol, H., Hoorelbeke, B., Vandekerckhove, J., Larsen, M. R., Roepstorff, P., Holm, A., Christiansen, G. and Birkelund, S. (2002). "Comparative proteome analysis of Chlamydia trachomatis serovar A, D and L2." *Proteomics* **2**(2): 164-186.

Shen, L., Shi, Y., Douglas, A. L., Hatch, T. P., O'Connell, C. M., Chen, J. M. and Zhang, Y. X. (2000). "Identification and characterization of promoters regulating tuf expression in Chlamydia trachomatis serovar F." *Arch Biochem Biophys* **379**(1): 46-56.

Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T. (2000). "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA." *Nucleic Acids Res* **28**(12): 2311-2314.

Siezen, R. J., Wilson, G. and Todt, T. (2010). "Prokaryotic whole-transcriptome analysis: deep sequencing and tiling arrays." *Microb Biotechnol* **3**(2): 125-130.

Sittka, A., Sharma, C. M., Rolle, K. and Vogel, J. (2009). "Deep sequencing of Salmonella RNA associated

with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes." *RNA Biol* **6**(3).

Skipp, P., Robinson, J., O'Connor, C. D. and Clarke, I. N. (2005). "Shotgun proteomic analysis of Chlamydia trachomatis." *Proteomics* **5**(6): 1558-1573.

Song, W., Maiste, P. J., Naiman, D. Q. and Ward, M. J. (2007). "Sigma 28 promoter prediction in members of the Gammaproteobacteria." *FEMS Microbiol Lett* **271**(2): 222-229.

Spencer, W. C., Zeller, G., Watson, J. D., Henz, S. R., Watkins, K. L., McWhirter, R. D., Petersen, S., Sreedharan, V. T., Widmer, C., Jo, J. Y., Reinke, V., Petrella, L., Strome, S., Von Stetina, S. E., Katz, M., Shaham, S., Ratsch, G. and Miller, D. M. (2011). "A spatial and temporal map of C. elegans gene expression." *Genome Research* **21**(2): 325-341.

Stark, D., van Hal, S., Hillman, R., Harkness, J. and Marriott, D. (2007). "Lymphogranuloma venereum in Australia: anorectal Chlamydia trachomatis serovar L2b in men who have sex with men." *J Clin Microbiol* **45**(3): 1029-1031.

Stephens, R. S., Ed. (1990). *Molecular Genetics of Chlamydia*. Chlamydia Infection. Cambridge, Cambridge University Press.

Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V. and Davis, R. W. (1998). "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis." *Science* **282**(5389): 754-759.

Stephens, R. S., Koshiyama, K., Lewis, E. and Kubo, A. (2001). "Heparin-binding outer membrane protein of chlamydiae." *Mol Microbiol* **40**(3): 691-699.

Stephens, R. S., Wagar, E. A. and Edman, U. (1988). "Developmental regulation of tandem promoters for the major outer membrane protein gene of Chlamydia trachomatis." *J Bacteriol* **170**(2): 744-750.

Stothard, D. R., Williams, J. A., Van Der Pol, B. and Jones, R. B. (1998). "Identification of a Chlamydia trachomatis serovar E urogenital isolate which lacks the cryptic plasmid." *Infect Immun* **66**(12): 6010-6013.

Studholme, D. J. and Buck, M. (2000). "Novel roles of sigmaN in small genomes." *Microbiology* **146 (Pt 1)**: 4-5.

Subtil, A. and Dautry-Varsat, A. (2004). "Chlamydia: five years A.G. (after genome)." *Curr Opin Microbiol* **7**(1): 85-92.

Tait, S. W. and Green, D. R. (2010). "Mitochondria and cell death: outer membrane permeabilization and beyond." *Nat Rev Mol Cell Biol* **11**(9): 621-632.

Tamura, A., Matsumoto, A. and Higashi, N. (1967). "Purification and chemical composition of reticulate bodies of the meningopneumonitis organisms." *J Bacteriol* **93**(6): 2003-2008.

Tan, M. and Engel, J. N. (1996). "Identification of sequences necessary for transcription in vitro from the Chlamydia trachomatis rRNA P1 promoter." *J Bacteriol* **178**(23): 6975-6982.

Tan, M., Gaal, T., Gourse, R. L. and Engel, J. N. (1998). "Mutational analysis of the Chlamydia trachomatis rRNA P1 promoter defines four regions important for transcription in vitro." *J Bacteriol* **180**(9): 2359-2366.

Tan, M., Wong, B. and Engel, J. N. (1996). "Transcriptional organization and regulation of the dnaK and groE operons of Chlamydia trachomatis." *J Bacteriol* **178**(23): 6983-6990.

Thomason, M. K. and Storz, G. (2010). Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annual Review of Genetics, Vol 44*. **44:** 167-188.

Thome, M., Schneider, P., Hofmann, K., Fickenscher, H., Meinl, E., Neipel, F., Mattmann, C., Burns, K., Bodmer, J. L., Schroter, M., Scaffidi, C., Krammer, P. H., Peter, M. E. and Tschopp, J. (1997). "Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors." *Nature* **386**(6624): 517-521.

Thomson, N. R., Holden, M. T., Carder, C., Lennard, N., Lockey, S. J., Marsh, P., Skipp, P., O'Connor, C. D., Goodhead, I., Norbertzcak, H., Harris, B., Ormond, D., Rance, R., Quail, M. A., Parkhill, J., Stephens, R. S. and Clarke, I. N. (2008). "Chlamydia trachomatis: genome sequence analysis of lymphogranuloma venereum isolates." *Genome Res* **18**(1): 161-171.

Tisoncik, J. R. and Katze, M. G. (2010). "What is systems biology?" *Future Microbiol* **5**(2): 139-141.

Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S. and Storz, G. (2006). "Target prediction for small, noncoding RNAs in bacteria." *Nucleic Acids Res* **34**(9): 2791-2802.

Toledo-Arana, A. and Solano, C. (2010). "Deciphering the physiological blueprint of a bacterial cell Revelations of unanticipated complexity in transcriptome and proteome." *Bioessays* **32**(6): 461-467.

Urban, J. H. and Vogel, J. (2009). "A green fluorescent protein (GFP)-based plasmid system to study post-transcriptional control of gene expression in vivo." *Methods Mol Biol* **540**: 301-319.

v Krosigk, A., Meyer, T., Jordan, S., Graefe, K., Plettenberg, A. and Stoehr, A. (2004). "[Dramatic increase in lymphogranuloma venereum among homosexual men in Hamburg]." *J Dtsch Dermatol Ges* **2**(8): 676-680.

Valentin-Hansen, P., Eriksen, M. and Udesen, C. (2004). "The bacterial Sm-like protein Hfq: a key player in RNA transactions." *Mol Microbiol* **51**(6): 1525-1533.

Vanover, J., Sun, J., Deka, S., Kintner, J., Duffourc, M. M. and Schoborg, R. V. (2008). "Herpes simplex virus co-infection-induced Chlamydia trachomatis persistence is not mediated by any known persistence inducer or anti-chlamydial pathway." *Microbiology* **154**(Pt 3): 971-978.

Vogel, J. and Sharma, C. M. (2005). "How to find small non-coding RNAs in bacteria." *Biol Chem* **386**(12): 1219-1238.

Wang, Z., Gerstein, M. and Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." *Nat Rev Genet* **10**(1): 57-63.

Wassarman, K. M. and Saecker, R. M. (2006). "Synthesis-mediated release of a small RNA inhibitor of RNA polymerase." *Science* **314**(5805): 1601-1603.

Wassarman, K. M. and Storz, G. (2000). "6S RNA regulates E. coli RNA polymerase activity." *Cell* **101**(6): 613-623.

Waters, L. S. and Storz, G. (2009). "Regulatory RNAs in bacteria." *Cell* **136**(4): 615-628.

Weaver, K. E. (2007). "Emerging plasmid-encoded antisense RNA regulated systems." *Curr Opin Microbiol* **10**(2): 110-116.

Weissenmayer, B. A., Prendergast, J. G. D., Lohan, A. J. and Loftus, B. J. (2011). "Sequencing Illustrates the Transcriptional Response of Legionella pneumophila during Infection and Identifies Seventy Novel Small Non-Coding RNAs." *PLoS ONE* **6**(3).

White, J. A. (2009). "Manifestations and management of lymphogranuloma venereum." *Curr Opin Infect Dis* **22**(1): 57-66.

Willenbrock, H., Salomon, J., Sokilde, R., Barken, K. B., Hansen, T. N., Nielsen, F. C., Moller, S. and Litman, T. (2009). "Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing." *RNA* **15**(11): 2028-2034.

Wilson, A. C. and Tan, M. (2002). "Functional analysis of the heat shock regulator HrcA of Chlamydia trachomatis." *J Bacteriol* **184**(23): 6566-6571.

Windbichler, N., von Pelchrzim, F., Mayer, O., Csaszar, E. and Schroeder, R. (2008). "Isolation of small RNA-binding proteins from E. coli: evidence for frequent interaction of RNAs with RNA polymerase." *RNA Biol* **5**(1): 30-40.

Wuppermann, F. N., Molleken, K., Julien, M., Jantos, C. A. and Hegemann, J. H. (2008). "Chlamydia pneumoniae GroEL1 protein is cell surface associated and required for infection of HEp-2 cells." *J Bacteriol* **190**(10): 3757-3767.

Wurm, R., Neusser, T. and Wagner, R. (2010). "6S RNA-dependent inhibition of RNA polymerase is released by RNA-dependent synthesis of small de novo products." *Biol Chem* **391**(2-3): 187-196.

Wyllie, S. and Raulston, J. E. (2001). "Identifying regulators of transcription in an obligate intracellular pathogen: a metal-dependent repressor in Chlamydia trachomatis." *Mol Microbiol* **40**(4): 1027-1036.

Yu, H. H., Di Russo, E. G., Rounds, M. A. and Tan, M. (2006a). "Mutational analysis of the promoter recognized by Chlamydia and Escherichia coli sigma(28) RNA polymerase." *J Bacteriol* **188**(15): 5524-5531.

Yu, H. H., Kibler, D. and Tan, M. (2006b). "In silico prediction and functional validation of sigma28-regulated genes in Chlamydia and Escherichia coli." *J Bacteriol* **188**(23): 8206-8212.

Yu, H. H. and Tan, M. (2003). "Sigma28 RNA polymerase regulates hctB, a late developmental gene in Chlamydia." *Mol Microbiol* **50**(2): 577-584.

Zagordi, O., Klein, R., Daumer, M. and Beerenwinkel, N. (2010). "Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies." *Nucleic Acids Research* **38**(21): 7400-7409.

Zychlinsky, A., Thirumalai, K., Arondel, J., Cantey, J. R., Aliprantis, A. O. and Sansonetti, P. J. (1996). "In vivo apoptosis in Shigella flexneri infections." *Infect Immun* **64**(12): 5357-5365.

9 APPENDIX

9.1 Detailed Lists of Transcriptional Start Sites (TSS)

9.1.1 TSS of *C. trachomatis*

Table 10: Transcription start sites and sequencing read numbers of *C. trachomatis* genes.
The exact transcription start sites of the annotated genes including novel non-coding RNAs are listed. For some genes there are two transcription starts listed which can be the consequence of alternative transcription start sites. Numbers in TSS and location refer to the first nucleotide of the transcript or the annotated gene coordinates, respectively. Genes without TSS are either too low expressed or transcribed as part of a polycistronic transcript. The table gives the following information: TSS: transcription start site; Location: start and end coordinates of gene; Length: length of the corresponding protein in amino acids; Strand: sense (+) or antisense (-) strand; Gene: gene name or gene number if no name is available; Synonym: gene number according to NCBI Reference Sequence NC_010280.1; Product: name of the gene product; EB-, EB+, RB-, RB+: sequence read numbers of the four cDNA libraries derived from either EB or RB total RNA (-) or total RNA enriched for primary transcripts (+).

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	1	1,017	338	+	hemB	CTLon_0001	delta-aminolevulinic acid dehydratase	0	0	1	0
2,454	1,043	2,440	465	-	nqrA	CTLon_0002	Na(+)-translocating NADH-quinone reductase subunit A	15	3	18	18
2,440	1,043	2,440	465	-	nqrA	CTLon_0002	Na(+)-translocating NADH-quinone reductase subunit A	15	3	18	18
2,916	2,462	2,896	144	-	CTLon_0003	CTLon_0003	hypothetical protein	6	1	5	0
2,993	3,010	5,157	715	+	greA	CTLon_0004	transcript cleavage factor/unknown domain fusion protein	1	0	5	1
5,246	5,166	5,238		-	tRNA-Ala	CTLon_0108h	tRNA	139	274	72	135
	5,345	6,547	400	+	aspC	CTLon_0005	aromatic amino acid aminotransferase	0	0	2	1
6,511	6,522	7,532	336	+	CTLon_0006	CTLon_0006	exported protein	1	0	2	0
	7,548	10,628	1026	-	recB	CTLon_0007	exodeoxyribonuclease V beta chain	0	0	2	0
	10,630	13,650	1006	-	recC	CTLon_0008	exodeoxyribonuclease V gamma chain	0	0	2	2
13,649	13,663	15,342	559	+	CTLon_0009	CTLon_0009	putative membrane efflux protein	0	0	2	3
16,237	15,362	16,177	271	-	CTLon_0010	CTLon_0010	hypothetical protein	2	0	5	4
16,999	17,070	19,643	857	+	topA	CTLon_0011	DNA topoisomerase I/SWI domain fusion protein	22	5	21	1
	19,673	20,677	334	-	CTLon_0012	CTLon_0012	putative oxidoreductase	0	0	0	0
21,111	20,656	20,952	98	-	CTLon_0013	CTLon_0013	integral membrane protein	0	0	0	2
	21,057	22,436	459	+	CTLon_0014	CTLon_0014	hypothetical protein	4	0	1	0
	22,436	23,014	192	+	CTLon_0015	CTLon_0015	hypothetical protein	0	0	0	0
	23,005	24,279	424	+	CTLon_0016	CTLon_0016	hypothetical protein	0	0	1	1
	24,282	24,818	178	+	CTLon_0017	CTLon_0017	hypothetical protein	0	0	0	0
	25,078	26,136	352	+	recA	CTLon_0018	recombinase A	0	0	1	0
	26,422	28,248	608	+	CTLon_0019	CTLon_0019	putative exported lipoprotein	0	0	3	1
28,218	28,245	29,735	496	+	recD	CTLon_0020	exodeoxyribonuclease V alpha chain	0	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
29,771	29,801	29,980	59	+	CTLon_0021	CTLon_0021	hypothetical protein	0	0	0	1
	30,081	30,800	239	-	CTLon_0022	CTLon_0022	ABC transporter ATP-binding protein	0	0	0	0
	30,809	31,297	162	-	CTLon_0023	CTLon_0023	hypothetical protein	2	0	2	0
32,176	31,294	32,103	269	-	kdsA	CTLon_0024	2-dehydro-3-deoxyphosphooctonate aldolase	0	0	4	1
32,387	32,410	32,482		+	tRNA-Arg	CTLon_0108j	tRNA	69	3868	31	1829
32,559	32,589	32,882	97	+	CTLon_0025	CTLon_0025	hypothetical protein	2	0	12	0
	32,882	33,199	105	+	CTLon_0026	CTLon_0026	hypothetical protein	1	0	1	0
33,248	33,281	34,288	335	+	sfhB	CTLon_0027	ribosomal large subunit pseudouridine synthase D	4	2	16	12
34,324	34,388	34,624	78	+	CTLon_0028	CTLon_0028	hypothetical protein	146	55	50	15
	34,763	36,235	490	-	gyrA2	CTLon_0029	DNA topoisomerase IV subunit A	0	0	0	0
	36,250	38,067	605	-	gyrB	CTLon_0030	DNA topoisomerase IV subunit B	0	0	3	0
	38,447	39,454	335	+	hemA	CTLon_0031	glutamyl-tRNA reductase	1	0	0	0
40,095	40,173	40,574	133	+	sycE	CTLon_0032	Type III secretion system chaperone	0	0	3	1
	40,578	43,067	829	+	CTLon_0033	CTLon_0033	phosphopeptide binding protein (to be a TTSS protein)	3	0	14	2
	43,111	43,362	83	+	CTLon_0034	CTLon_0034	hypothetical protein	0	0	7	0
	43,389	43,640	83	+	CTLon_0035	CTLon_0035	hypothetical protein	0	0	4	0
	43,659	44,108	149	+	CTLon_0036	CTLon_0036	hypothetical protein	2	0	5	0
	44,128	44,799	223	+	CTLon_0037	CTLon_0037	hypothetical protein	1	0	10	0
	44,801	46,129	442	+	sctN	CTLon_0038	type III secretion system ATPase	1	0	11	0
	46,152	46,658	168	+	CTLon_0039	CTLon_0039	hypothetical protein	0	0	4	0
	46,662	47,513	283	+	CTLon_0040	CTLon_0040	hypothetical protein	0	0	9	0
	47,523	48,644	373	+	sctQ	CTLon_0041	type III secretion system protein	3	0	5	1
	48,779	50,251	490	+	pkn5	CTLon_0042	putative serine/threonine-protein kinase (TTSS effector protein)	1	0	18	5
49,966	50,248	53,013	921	+	CTLon_0043	CTLon_0043	Type III secretion structural protein (outer membrane ring)	1	0	41	4
53,147	53,147	53,219		+	tRNA-Thr	CTLon_0108an	tRNA	10	100	5	39
	53,327	54,397	356	-	aspC	CTLon_0044	ATP:guanido phosphotransferase	0	0	3	0
54,937	54,387	54,908	173	-	CTLon_0045	CTLon_0045	hypothetical protein	2	0	1	0
55,085	55,013	55,085		-	tRNA-Lys	CTLon_0108z	tRNA	151	615	45	194
55,176	55,102	55,176		-	tRNA-Glu	CTLon_0108q	tRNA	134	675	75	298
	55,280	55,819	179	-	rrf	CTLon_0046	ribosome recycling factor	0	0	7	0
	55,816	56,553	245	-	pyrH	CTLon_0047	uridylate kinase	0	0	2	0
	56,566	57,414	282	-	tsf	CTLon_0048	elongation factor Ts	0	0	4	0
	57,411	58,259	282	-	rpsB	CTLon_0049	30S ribosomal protein S2	0	0	1	0
58,422	58,342	58,412		-	tRNA-Gly	CTLon_0108r	tRNA	94	669	38	127
60,074	58,643	59,827	394	-	ompA	CTLon_0050	major outer membrane protein	12	3	301	47
60,420	60,435	63,677	1080	+	pbpB	CTLon_0051	penicillin-binding protein	0	0	4	0
63,719	63,741	64,748	335	+	CTLon_0052	CTLon_0052	TPR repeat-containing protein	0	0	4	0
	65,090	66,541	483	+	CTLon_0053	CTLon_0053	cysteine desulfurase activator complex subunit SufB	1	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	66,544	67,311	255	+	CTLon_0054	CTLon_0054	ABC transporter ATP-binding protein	0	0	0	0
	67,315	68,502	395	+	CTLon_0055	CTLon_0055	hypothetical protein	0	0	0	0
	68,495	69,700	401	+	sufS	CTLon_0056	cysteine desulfurase	0	0	0	0
69,773	69,836	70,681	281	+	parB	CTLon_0057	putative chromosome partitioning protein	0	0	0	1
	70,673	71,503	276	-	dppF	CTLon_0058	ABC transport protein, ATPase component	0	0	1	1
72,550	71,496	72,461	321	-	dppD	CTLon_0059	ABC transport protein, ATPase component	5	1	18	24
72,588	72,622	73,296	224	+	CTLon_0060	CTLon_0060	hypothetical protein	1	3	2	4
	73,299	74,579	426	+	CTLon_0061	CTLon_0061	low-affinity inorganic phosphate transporter	1	0	4	0
74,686	74,707	75,918	403	+	pgk	CTLon_0062	phosphoglycerate kinase	0	0	13	3
76,149	76,178	77,146	322	+	CTLon_0063	CTLon_0063	hypothetical protein	10	0	2	0
	77,197	78,393	398	+	CTLon_0064	CTLon_0064	hypothetical protein	16	1	2	0
78,435	78,479	79,657	392	+	CTLon_0065	CTLon_0065	hypothetical protein	5	0	2	0
	79,654	80,289	211	-	nth	CTLon_0066	endonuclease III	0	0	0	0
	80,296	81,630	444	-	trmE	CTLon_0067	tRNA modification GTPase TrmE	0	0	1	0
81,848	81,898	82,803	301	+	psdD	CTLon_0068	phosphatidylserine decarboxylase	2	1	1	0
82,845	82,869	84,194	441	+	CTLon_0069	CTLon_0069	hypothetical protein	0	0	6	1
	84,453	87,362	969	+	secA	CTLon_0070	preprotein translocase subunit SecA	1	0	14	0
88,023	87,456	87,983	175	-	CTLon_0071	CTLon_0071	hypothetical protein	0	0	4	0
	88,060	89,532	490	-	engA	CTLon_0072	GTP-binding protein EngA	0	0	0	1
90,916	89,648	90,880	410	-	pcnB	CTLon_0073	polyA polymerase	0	0	4	1
	90,895	92,154	419	-	clpX	CTLon_0074	ATP-dependent protease ATP-binding subunit	0	0	1	0
	92,164	92,775	203	-	clpP	CTLon_0075	ATP-dependent Clp protease proteolytic subunit	0	0	1	0
	92,942	94,270	442	-	tig	CTLon_0076	trigger factor	0	0	5	0
94,379	94,307	94,378		-	tRNA-Gly	CTLon_0108s	tRNA	148	1683	42	401
94,598	94,629	98,120	1163	+	CTLon_0077	CTLon_0077	putative helicase	1	0	12	7
98,073	98,125	99,225	366	+	mreB	CTLon_0078	cell shape determining protein	1	0	3	2
	99,222	101,021	599	+	pckA	CTLon_0079	phosphoenolpyruvate carboxykinase	0	0	5	2
101,109	101,133	103,436	767	+	CTLon_0080	CTLon_0080	hypothetical protein	4	0	13	0
	103,463	104,635	390	+	CTLon_0081	CTLon_0081	hypothetical protein	0	0	8	1
105,758	104,701	105,723	340	-	ompB, porB	CTLon_0082	outer membrane protein B	1	0	12	5
105,758	105,646	105,723	25	-	ompB	CTLon_0082a	hypothetical protein	0	0	2	3
	105,857	106,861	334	-	gpdA	CTLon_0083	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	0	0	3	0
	106,858	108,225	455	-	CTLon_0084	CTLon_0084	UDP-N-acetylglucosamine pyrophosphorylase	0	0	0	0
108,630	108,237	108,602	121	-	CTLon_0085	CTLon_0085	hypothetical protein	0	0	0	2
	108,595	109,899	434	-	fliI	CTLon_0086	type III secretion system ATPase	0	0	1	0
	109,973	110,497	174	-	CTLon_0087	CTLon_0087	hypothetical protein	2	0	1	0
111,694	110,502	111,506	334	-	fliF	CTLon_0088	type III secretion system protein	1	3	8	1
	111,780	112,562	260	-	CTLon_0089	CTLon_0089	hypothetical protein	0	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	112,559	113,713	384	-	CTLon_0090	CTLon_0090	cysteine desulfurase	0	0	1	0
114,378	113,668	114,348	226	-	pgmA	CTLon_0091	phosphoglyceromutase	0	2	0	2
114,519	114,643	115,368	241	+	CTLon_0092	CTLon_0092	ribosomal large subunit pseudouridine synthase B	5	4	4	3
115,513	115,487	116,011	174	+	CTLon_0093	CTLon_0093	hypothetical protein	0	0	0	6
	116,038	116,592	184	+	birA	CTLon_0094	biotin--protein ligase	1	0	1	0
	116,599	117,738	379	-	CTLon_0095	CTLon_0095	cell cycle protein	0	0	0	0
	117,830	119,809	659	-	CTLon_0096	CTLon_0096	cation transporting ATPase	0	0	4	1
120,600	119,833	120,579	248	-	CTLon_0097	CTLon_0097	putative integral membrane protein	0	0	6	0
121,931	120,616	121,902	428	-	serS	CTLon_0098	seryl-tRNA synthetase	1	0	1	1
	121,944	123,071	375	+	ribD	CTLon_0099	riboflavin biosynthesis protein (diaminohydroxyphosphoribosylaminopyrimidine deaminase)	0	0	1	0
123,109	123,181	124,455	424	+	ribA	CTLon_0100	bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein	4	1	3	2
	124,424	124,897	157	+	ribH	CTLon_0101	riboflavin synthase subunit beta	0	0	2	0
126,486	124,948	126,294	448	-	CTLon_0102	CTLon_0102	hypothetical protein	7	0	8	1
126,652	126,679	127,344	221	+	CTLon_0103	CTLon_0103	putative lipoprotein	16	3	13	9
126,828	126,679	127,344	221	+	CTLon_0103	CTLon_0103	putative lipoprotein	16	3	13	9
	127,461	128,828	455	+	dagA	CTLon_0104	Na(+)-linked D-alanine glycine permease	5	0	7	1
	128,886	129,338	150	+	CTLon_0105	CTLon_0105	hypothetical protein	2	1	1	0
	129,347	130,006	219	-	CTLon_0106	CTLon_0106	SET domain-containing protein	0	0	1	0
	130,003	130,791	262	-	CTLon_0107	CTLon_0107	Zn-dependent hydrolase	0	0	1	0
133,311	130,795	133,194	799	-	ftsK	CTLon_0108	cell division protein	2	0	15	1
133,499	133,374	133,447		-	tRNA-His	CTLon_0108a	tRNA	21	148	18	54
133,792	133,979	135,528		+	16S_rRNA	CTLon_0108b	16S ribosomal RNA	8178	1124	6622	477
	135,844	138,711		+	23S_rRNA	CTLon_0108c	23S ribosomal RNA	12641	1896	14550	1376
138,827	138,832	138,948		+	5S_rRNA	CTLon_0108d	5S ribosomal RNA	14991	8156	10819	17269
140,795	139,465	140,760	431	-	dmpP	CTLon_0109	Na(+)-translocating NADH-quinone reductase subunit F	0	1	2	2
141,274	140,903	141,247	114	-	CTLon_0110	CTLon_0110	preprotein translocase subunit YajC	0	0	3	4
	141,352	142,542	396	-	CTLon_0111	CTLon_0111	rRNA methyltransferase	0	0	0	0
143,154	142,730	143,107	125	-	hctA	CTLon_0112	histone H1--like developmental protein	0	0	9	1
143,447	143,503	145,971	822	+	CTLon_0113	CTLon_0113	hypothetical protein	0	0	1	2
	145,963	147,237	424	-	hemG	CTLon_0114	protoporphyrinogen oxidase	0	0	0	0
	147,234	148,607	457	-	hemN	CTLon_0115	coproporphyrinogen III oxidase	0	0	0	0
	148,580	149,590	336	-	hemE	CTLon_0116	uroporphyrinogen decarboxylase	0	0	0	0
	149,603	152,842	1079	-	mfd	CTLon_0117	transcription-repair coupling factor	2	0	0	1
	152,818	155,445	875	-	alaS	CTLon_0118	alanyl-tRNA synthetase	1	0	0	0
	156,023	157,572		+	16S_rRNA	CTLon_0108e	16S ribosomal RNA	0	0	0	0
	157,888	160,755		+	23S_rRNA	CTLon_0108f	23S ribosomal RNA	0	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	160,876	160,992		+	5S_rRNA	CTLon_0108g	5S ribosomal RNA	0	0	0	0
	161,270	163,270	666	+	tktB	CTLon_0119	transketolase	0	0	2	0
	163,267	164,136	289	-	amn	CTLon_0120	AMP nucleosidase	0	0	1	0
	164,392	164,928	178	+	efp	CTLon_0121	elongation factor P	1	0	0	0
	164,944	165,168	74	-	CTLon_0122	CTLon_0122	hypothetical protein	0	0	0	0
166,135	165,215	166,087	290	-	CTLon_0123	CTLon_0123	hypothetical protein	0	1	1	0
167,735	166,168	167,706	512	-	groEL	CTLon_0124	60 kDa chaperonin GroEL	0	0	5	2
167,861	167,784	167,854		-	tRNA-Cys	CTLon_0108o	tRNA	39	470	28	77
	168,154	169,506	450	+	murF	CTLon_0125	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	1	0	0	0
169,613	169,652	170,662	336	+	mraY	CTLon_0126	phospho-N-acetylmuramoyl-pentapeptide-transferase	0	0	0	1
	170,674	171,924	416	+	murD	CTLon_0127	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	0	0	3	7
171,412	171,921	172,658	245	+	nlpD	CTLon_0128	muramidase	0	0	6	0
	172,674	173,831	385	+	ftsW	CTLon_0129	cell division protein	0	0	2	1
	173,740	174,798	352	+	murG	CTLon_0130	N-acetylglucosaminyl transferase	0	0	1	0
	174,803	177,214	803	+	murC	CTLon_0131	UDP-N-acetylmuramate--alanine ligase	5	0	2	0
177,727	177,251	177,670	139	-	CTLon_0132	CTLon_0132	hypothetical protein	2	1	1	0
178,709	177,828	178,634	268	-	CTLon_0133	CTLon_0133	hypothetical protein	0	0	1	0
	178,788	179,120	110	+	rbsV	CTLon_0134	anti-sigma F factor antagonist	0	1	0	0
	179,124	180,143	339	+	miaA	CTLon_0135	tRNA delta(2)-isopentenylpyrophosphate transferase	0	0	0	0
	180,148	181,200	350	-	CTLon_0136	CTLon_0136	hypothetical protein	0	0	0	0
181,243	181,290	182,972	560	+	CTLon_0137	CTLon_0137	hypothetical protein	0	0	7	6
183,412	183,426	183,785	119	+	CTLon_0138	CTLon_0138	hypothetical protein	2	1	1	1
	183,796	185,052	418	+	fabF	CTLon_0139	3-oxoacyl-(acyl carrier protein) synthase II	0	0	0	0
	185,049	185,501	150	+	CTLon_0140	CTLon_0140	hypothetical protein	1	1	3	0
	185,582	186,211	209	-	ppa	CTLon_0141	inorganic pyrophosphatase	0	0	0	0
	186,459	187,499	346	+	ldh	CTLon_0142	leucine dehydrogenase	0	1	3	0
	187,465	188,493	342	-	cysQ	CTLon_0143	3'(2'),5'-bisphosphate nucleotidase	0	0	0	0
188,738	188,791	189,552	253	+	CTLon_0144	CTLon_0144	1-acyl-sn-glycerol-3-phosphate acyltransferase	22	7	3	2
	189,560	191,173	537	+	aas	CTLon_0145	long chain fatty acid--[acyl-carrier-protein] ligase	16	5	6	0
	191,216	192,349	377	+	bioF	CTLon_0146	8-amino-7-oxononanoate synthase	12	0	4	0
	192,268	194,529	753	-	priA	CTLon_0147	primosome assembly protein PriA	0	0	0	0
195,283	194,484	195,173	229	-	CTLon_0148	CTLon_0148	hypothetical protein	0	0	3	1
	195,329	195,823	164	+	CTLon_0149	CTLon_0149	protein disulfide isomerase	0	0	1	0
	196,055	197,635	526	+	lysS	CTLon_0150	lysyl-tRNA synthetase	0	0	2	0
	197,632	199,125	497	-	cysS	CTLon_0151	cysteinyl-tRNA synthetase	0	0	0	1
199,293	199,378	200,427	349	+	CTLon_0152	CTLon_0152	hypothetical protein	2	0	11	17
	200,505	200,867	120	-	rnpA	CTLon_0153	ribonuclease P	3	0	1	0
201,071	200,880	201,017	45	-	rnpA	CTLon_0154	50S ribosomal protein L34	2	1	6	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
201,220	201,494	201,631	45	+	rpmJ	CTLon_0155	50S ribosomal protein L36	1	0	1	0
	201,655	201,960	101	+	rpsN	CTLon_0156	30S ribosomal protein S14	1	0	5	0
	201,994	202,494	166	-	CTLon_0157	CTLon_0157	hypothetical protein	0	0	1	4
	202,616	202,867	83	-	CTLon_0158	CTLon_0158	putative integral membrane protein	0	0	0	0
203,104	203,147	203,641	164	+	CTLon_0159	CTLon_0159	hypothetical protein	0	0	5	2
	203,688	205,427	579	-	uvrC	CTLon_0160	excinuclease ABC subunit C	0	0	1	0
207,971	205,509	207,971	820	-	mutS	CTLon_0161	DNA mismatch repair protein	0	0	7	0
	208,034	208,302	88.6667	+	CTLon_0162	CTLon_0162	pseudo	0	0	0	0
	208,304	210,091	595	+	dnaG	CTLon_0163	DNA primase	0	0	1	0
210,158	210,176	210,463	95	+	CTLon_0164	CTLon_0164	hypothetical protein	8	6	0	2
	210,590	211,081	163	+	CTLon_0165	CTLon_0165	hypothetical protein	0	0	0	0
	211,164	214,175	1003	-	glyQ	CTLon_0166	glycyl-tRNA synthetase	2	0	1	0
214,413	214,629	215,237	202	+	pgsA	CTLon_0167	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	5	2	4	4
216,701	215,244	216,668	474	-	glgA	CTLon_0168	glycogen synthase	2	3	1	1
216,897	216,897	216,968		+	tRNA-Gln	CTLon_0108p	tRNA	158	2247	101	637
217,101	217,245	217,802	185	+	rplY	CTLon_0169	50S ribosomal protein L25/general stress protein Ctc	2	1	5	5
	217,810	218,349	179	+	pth	CTLon_0170	peptidyl-tRNA hydrolase	3	0	5	0
	218,487	218,825	112	+	rpsF	CTLon_0171	30S ribosomal protein S6	1	0	1	0
	218,842	219,087	81	+	rpsR	CTLon_0172	30S ribosomal protein S18	1	0	1	0
	219,108	219,611	167	+	rplI	CTLon_0173	50S ribosomal protein L9	0	0	0	0
	219,675	220,541	288	+	ispE	CTLon_0174	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	0	0	2	0
222,280	220,844	222,196	450	-	CTLon_0175	CTLon_0175	putative integral membrane protein	0	0	2	1
225,263	222,375	225,245	956	-	ptr	CTLon_0176	exported insulinase/protease	0	2	6	0
	225,297	226,298	333	-	plsB	CTLon_0177	glycerol-3-phosphate acyltransferase	1	0	1	1
	226,311	227,849	512	-	cafE	CTLon_0178	ribonuclease E	0	0	2	5
	228,413	228,733	106	+	CTLon_0179	CTLon_0179	hypothetical protein	0	0	0	0
	228,673	228,999	108	+	CTLon_0180	CTLon_0180	hypothetical protein	1	0	1	0
228,965	228,983	229,162	59	+	rpmF	CTLon_0181	50S ribosomal protein L32	0	0	2	0
	229,184	230,149	321	+	plsX	CTLon_0182	putative glycerol-3-phosphate acyltransferase PlsX	0	1	0	0
	230,369	234,961	1530	+	pmpD	CTLon_0183	polymorphic outer membrane protein	0	1	4	0
	235,162	235,956	264	+	CTLon_0184	CTLon_0184	hypothetical protein	0	0	0	0
	236,040	236,339	99	-	CTLon_0185	CTLon_0185	hypothetical protein	157	16	18	1
236,752	236,360	236,722	120	-	CTLon_0186	CTLon_0186	hypothetical protein	113	20	4	3
	236,967	238,343	458	+	CTLon_0187	CTLon_0187	phosphoglucosamine mutase	1	0	0	0
	238,354	240,174	606	+	glmS	CTLon_0188	D-fructose-6-phosphate amidotransferase	1	0	2	0
	240,277	241,473	398	+	tyrP	CTLon_0189	tyrosine-specific transport protein	2	0	3	0
241,603	241,656	242,849	397	+	CTLon_0190	CTLon_0190	tyrosine-specific transport protein	2	0	8	18

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
242,848	242,891	243,607	238	+	CTLon_0191	CTLon_0191	putative membrane transport protein	4	1	13	1
	243,704	244,558	284	-	ftsY	CTLon_0192	cell division protein	1	0	3	0
244,609	244,628	245,788	386	+	sucC	CTLon_0193	succinyl-CoA synthetase subunit beta	1	0	13	1
	245,803	246,678	291	+	sucD	CTLon_0194	succinyl-CoA synthetase subunit alpha	0	0	5	0
246,748	246,818	248,311	497	+	htrA	CTLon_0195	serine protease	0	0	17	1
248,444	248,513	251,437	974	+	CTLon_0196	CTLon_0196	metalloprotease-insulinase	2	0	8	1
	251,451	252,734	427	-	rmuC	CTLon_0197	DNA recombination protein	0	0	0	1
	252,738	253,529	263	-	pssA	CTLon_0198	CDP-diacylglycerol--serine O-phosphatidyltransferase	0	0	1	0
253,980	254,164	257,307	1047	+	nrdA	CTLon_0199	ribonucleotide-diphosphate reductase subunit alpha	0	0	12	9
	257,345	258,385	346	+	nrdB	CTLon_0200	ribonucleotide-diphosphate reductase subunit beta	2	0	3	0
	258,645	259,319	224	+	trmB	CTLon_0201	tRNA (guanine-N(7))-methyltransferase	0	0	0	0
259,392	259,310	259,383		-	tRNA-Val	CTLon_0108as	tRNA	94	578	37	71
	259,435	260,019	194	+	CTLon_0202	CTLon_0202	putative SAM-dependent methyltransferase	0	0	0	0
260,971	260,016	260,891	291	-	murB	CTLon_0203	UDP-N-acetylenolpyruvoylglucosamine reductase	0	0	0	3
261,622	261,024	261,530	168	-	nusB	CTLon_0204	transcription antitermination protein NusB	0	1	0	1
261,819	261,962	262,522	186	+	infC	CTLon_0205	translation initiation factor IF-3	1	0	5	2
	262,509	262,694	61	+	rpmI	CTLon_0206	50S ribosomal protein L35	0	0	1	0
	262,713	263,084	123	+	rplT	CTLon_0207	50S ribosomal protein L20	0	0	1	0
	263,091	264,119	342	+	pheS	CTLon_0208	phenylalanyl-tRNA synthetase subunit alpha	0	0	5	0
264,120	264,120	264,204		+	tRNA-Ser	CTLon_0108aj	tRNA	49	405	41	157
	264,430	266,406	658	+	CTLon_0209	CTLon_0209	hypothetical protein	1	0	2	0
	266,403	267,503	366	-	CTLon_0210	CTLon_0210	putative membrane transport protein	0	2	0	0
	267,506	268,570	354	-	CTLon_0211	CTLon_0211	putative membrane transport protein	3	0	0	0
	268,668	269,633	321	+	tilS	CTLon_0212	tRNA(Ile)-lysidine synthase	0	0	3	0
269,757	269,794	272,535	913	+	ftsH	CTLon_0213	cell division protein	2	0	33	20
	272,750	274,837	695	-	pnp	CTLon_0214	polynucleotide phosphorylase/polyadenylase	1	0	8	1
275,170	274,866	275,135	89	-	rpsO	CTLon_0215	30S ribosomal protein S15	3	0	3	2
	275,361	275,852	163	+	CTLon_0216	CTLon_0216	cytosine deaminase	0	0	0	0
	275,884	276,102	72	+	CTLon_0217	CTLon_0217	hypothetical protein	0	0	0	0
	276,074	276,778	234	-	CTLon_0218	CTLon_0218	putative integral membrane protein	0	0	0	0
	276,913	277,431	172	-	CTLon_0219	CTLon_0219	hypothetical protein	5	0	0	0
	277,444	277,950	168	-	CTLon_0220	CTLon_0220	hypothetical protein	1	1	2	0
278,487	277,978	278,457	159	-	CTLon_0221	CTLon_0221	hypothetical protein	7	0	3	1
278,547	278,774	278,962	62	+	CTLon_0222	CTLon_0222	hypothetical protein	0	1	2	7
278,837	278,981	280,198	405	+	CTLon_0223	CTLon_0223	putative integral membrane protein	0	0	2	0
	280,155	281,030	291	-	map	CTLon_0224	methionine aminopeptidase	0	0	1	0
	281,063	281,704	213	-	CTLon_0225	CTLon_0225	putative integral membrane protein	0	0	0	0
282,405	281,721	282,320	199	-	CTLon_0226	CTLon_0226	putative integral membrane protein	2	0	2	1

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
284,308	282,515	284,284	589	-	CTLon_0227	CTLon_0227	ABC transporter permease	0	0	1	0
	284,337	285,729	463.333	+	fumC	CTLon_0228	pseudo	0	0	0	0
	285,811	287,514	567	-	CTLon_0229	CTLon_0229	sulfate transporter	0	0	1	0
	287,745	287,906	53	-	CTLon_0230	CTLon_0230	hypothetical protein	0	0	0	0
	287,946	288,155	69	-	CTLon_0231	CTLon_0231	hypothetical protein	0	0	0	0
	288,269	289,534	421	+	CTLon_0232	CTLon_0232	putative membrane transport protein	0	0	2	1
	289,644	291,449	601	+	cpa	CTLon_0233	putative exported protease	0	0	4	0
291,514	291,544	292,467	307	+	ispH	CTLon_0234	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	2	0	0	0
	292,519	294,000	493	-	copD2	CTLon_0235	hypothetical protein	0	0	0	0
	294,012	295,532	506	-	copB2	CTLon_0236	putative type III secretion translocator	0	0	1	0
	295,560	296,156	198	-	lcrH	CTLon_0237	type III secretion chaperone, scc3	0	0	1	0
297,562	296,078	297,526	482	-	CTLon_0238	CTLon_0238	hypothetical protein	0	0	2	0
297,878	297,895	297,976		+	tRNA-Leu	CTLon_0108u	tRNA	40	207	15	21
	298,387	299,289	300	-	xerD	CTLon_0239	site-specific tyrosine recombinase XerD	0	1	0	0
	299,537	300,526	329	+	CTLon_0240	CTLon_0240	hypothetical protein	15	4	2	2
300,512	300,541	302,757	738	+	glgB	CTLon_0241	glycogen branching enzyme	15	1	15	2
303,816	302,768	303,787	339	-	CTLon_0242	CTLon_0242	hypothetical protein	19	18	3	1
	304,014	305,225	403	-	CTLon_0243	CTLon_0243	hypothetical protein	2	0	1	0
	305,402	308,299	965	-	pmpE	CTLon_0244	polymorphic outer membrane protein	2	1	19	3
	308,302	311,400	1032	-	pmpF	CTLon_0245	polymorphic outer membrane protein	2	0	29	1
311,536	311,565	314,606	1013	+	pmpG	CTLon_0246	polymorphic outer membrane protein	7	0	50	24
	314,637	317,666	1009	+	pmpH	CTLon_0247	polymorphic outer membrane protein	7	0	61	3
321,430	318,772	321,408	878	-	pmpI	CTLon_0249	polymorphic outer membrane protein	0	0	1	0
319,013	318,772	321,408	878	-	pmpI	CTLon_0249	polymorphic outer membrane protein	12	1	12	21
321,651	321,683	323,455	590	+	CTLon_0250	CTLon_0250	hypothetical protein	678	55	50	7
323,900	323,600	323,872	90	-	CTLon_0251	CTLon_0251	hypothetical protein	1	1	36	47
324,043	324,073	324,375	100	+	gatC	CTLon_0252	aspartyl/glutamyl-tRNA amidotransferase subunit C	0	0	4	6
	324,387	325,862	491	+	gatA	CTLon_0253	aspartyl/glutamyl-tRNA amidotransferase subunit A	2	1	3	0
	325,864	327,330	488	+	gatB	CTLon_0254	aspartyl/glutamyl-tRNA amidotransferase subunit B	0	0	1	0
328,556	327,429	328,520	363	-	CTLon_0255	CTLon_0255	hypothetical protein	136	6	21	1
	328,648	329,217	189	-	CTLon_0256	CTLon_0256	hypothetical protein	1	0	0	0
329,456	329,531	330,481	316	+	CTLon_0257	CTLon_0257	hypothetical protein	1	1	1	1
	330,497	331,399	300	-	rnhC	CTLon_0258	ribonuclease HIII	0	0	0	0
	331,655	332,086	143	+	CTLon_0259	CTLon_0259	hypothetical protein	0	0	0	0
	332,073	333,440	455	-	htrB	CTLon_0260	lipid A biosynthesis lauroyl acyltransferase	0	0	3	0
	333,572	334,828	418	-	CTLon_0261	CTLon_0261	hypothetical protein	0	0	3	0
	334,825	335,619	264	-	CTLon_0262	CTLon_0262	putative integral membrane protein	0	0	1	0
335,862	335,894	337,234	446	+	cydA	CTLon_0263	cytochrome d ubiquinol oxidase subunit I	0	0	23	1

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	337,246	338,307	353	+	cydB	CTLon_0264	cytochrome d ubiquinol oxidase subunit II	0	0	4	0
339,749	338,405	339,709	434	-	CTLon_0265	CTLon_0265	hypothetical protein	0	0	7	4
339,889	339,918	340,646	242	+	CTLon_0266	CTLon_0266	hypothetical protein	1	0	7	8
340,734	340,832	342,133	433	+	CTLon_0267	CTLon_0267	hypothetical protein	0	0	19	3
342,799	342,195	342,668	157	-	CTLon_0268	CTLon_0268	hypothetical protein	10	2	5	0
343,368	343,368	342,948		-	tmRNA		transfer messenger RNA	677	146	223	87
	343,714	346,824	1036	+	ileS	CTLon_0269	isoleucyl-tRNA synthetase	0	0	1	0
348,784	346,883	348,769	628	-	lepB	CTLon_0270	signal peptidase I	0	0	3	0
349,737	348,955	349,698	247	-	CTLon_0271	CTLon_0271	hypothetical protein	2	0	0	1
349,710	349,774	350,100	108	+	rpmE	CTLon_0272	50S ribosomal protein L31 type B	1	0	4	1
	350,288	351,367	359	+	prfA	CTLon_0273	peptide chain release factor 1	0	1	5	0
	351,351	352,223	290	+	CTLon_0274	CTLon_0274	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	0	0	3	1
352,198	352,220	353,566	448	+	ffh	CTLon_0275	signal recognition particle, subunit FFH/SRP54	0	0	3	0
	353,557	353,907	116	+	rpsP	CTLon_0276	30S ribosomal protein S16	0	0	3	0
	353,923	354,981	352	+	trmD	CTLon_0277	tRNA (guanine-N(1)-)-methyltransferase/unknown domain fusion protein	0	0	0	1
	355,007	355,372	121	+	rplS	CTLon_0278	50S ribosomal protein L19	0	0	0	0
	355,436	356,089	217	+	rnhB	CTLon_0279	ribonuclease HII	0	0	2	1
	356,080	356,697	205	+	gmk	CTLon_0280	guanylate kinase	0	0	0	0
	356,690	356,992	100	+	CTLon_0281	CTLon_0281	hypothetical protein	0	0	0	2
	356,992	358,644	550	+	metG	CTLon_0282	methionyl-tRNA synthetase	0	0	1	0
	358,783	361,023	746	-	recD	CTLon_0283	exodeoxyribonuclease V alpha chain	0	0	0	0
362,456	361,271	362,251	326	-	CTLon_0284	CTLon_0284	putative membrane transport/efflux protein	16	5	8	5
362,608	362,633	363,406	257	+	CTLon_0285	CTLon_0285	putative protein ligase	68	13	12	1
	363,371	364,528	385	-	CTLon_0286	CTLon_0286	hypothetical protein	0	0	0	0
	364,589	364,875	94.6667	-	CTLon_0287	CTLon_0287	pseudo	0	0	0	0
365,017	364,939	365,010		-	tRNA-Asn	CTLon_0108m	tRNA	394	5837	175	1570
	365,074	365,418	114	-	CTLon_0288	CTLon_0288	hypothetical protein	0	0	2	0
	365,440	366,012	190	-	dcd	CTLon_0289	deoxycytidine triphosphate deaminase	0	0	2	0
	366,099	366,251	50	+	CTLon_0290	CTLon_0290	hypothetical protein	1	0	0	0
367,049	366,293	367,297	334	+	ruvB	CTLon_0291	Holliday junction DNA helicase B	1	0	15	7
	367,299	368,111	270	+	CTLon_0292	CTLon_0292	hypothetical protein	1	0	8	0
	368,173	370,173	666	-	CTLon_0293	CTLon_0293	putative glycosyl hydrolase	0	0	2	2
370,821	370,290	370,793	167	-	CTLon_0294	CTLon_0294	putative type III secretion system chaperone	0	0	21	4
	371,259	371,732	157	+	ssb	CTLon_0295	single-strand DNA-binding protein	2	0	1	0
	372,075	373,574	499	+	CTLon_0296	CTLon_0296	leucyl aminopeptidase	0	0	4	0
373,683	373,712	374,383	223	+	hct2	CTLon_0297	histone-like protein 2	5	0	11	4

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
374,494	374,538	375,482	314	+	CTLon_0298	CTLon_0298	hypothetical protein	0	0	0	2
375,545	375,577	376,290	237	+	CTLon_0299	CTLon_0299	hypothetical protein	0	0	0	0
	376,381	377,853	490	+	CTLon_0300	CTLon_0300	hypothetical protein	0	0	11	3
	377,912	379,579	555	-	CTLon_0301	CTLon_0301	hypothetical protein	1	0	3	0
	379,615	381,312	565	-	CTLon_0302	CTLon_0302	hypothetical protein	0	0	4	1
	381,374	382,504	376	-	hemN	CTLon_0303	coproporphyrinogen III oxidase	0	0	0	1
383,175	382,494	382,940	148	-	CTLon_0304	CTLon_0304	hypothetical protein	0	0	2	0
383,237	383,272	385,983	903	+	sucA	CTLon_0305	alpha-ketoglutarate decarboxylase	0	0	5	0
	385,987	387,084	365	+	sucB	CTLon_0306	dihydrolipoamide acetyltransferase	0	0	5	0
	387,113	387,844	243	-	CTLon_0307	CTLon_0307	hypothetical protein	0	0	0	0
	387,853	389,661	602	-	gcpE, ispG	CTLon_0308	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	2	0	3	0
	389,813	390,904	363	-	CTLon_0309	CTLon_0309	hypothetical protein	0	1	3	0
	390,984	391,259	91	-	fer	CTLon_0310	ferredoxin	0	0	1	0
391,371	391,297	391,371		-	tRNA-Pro	CTLon_0108ah	tRNA	54	361	35	143
391,503	391,441	393,258	605	+	flhA	CTLon_0311	type III secretion system protein	2	0	1	4
393,277	393,366	394,127	253	+	fliA	CTLon_0312	RNA polymerase sigma factor sigma-28	0	0	3	2
	394,286	395,524	412	+	tyrS	CTLon_0313	tyrosyl-tRNA synthetase	0	0	5	0
	395,534	396,976	480	+	gnd	CTLon_0314	6-phosphogluconate dehydrogenase	0	0	3	0
	397,037	398,845	602	-	lepA	CTLon_0315	GTP-binding protein LepA	0	0	2	0
400,831	399,031	400,617	528	-	CTLon_0316	CTLon_0316	ADP,ATP carrier protein	16	4	11	4
401,461	400,970	401,446	158	-	CTLon_0317	CTLon_0317	hypothetical protein	2	0	3	1
401,905	402,107	403,087	326	+	CTLon_0318	CTLon_0318	ABC transport protein, solute binding component	2	1	4	7
	403,080	403,859	259	+	CTLon_0319	CTLon_0319	ABC transport protein, ATP-binding component	1	0	1	0
	403,860	405,215	451	+	CTLon_0320	CTLon_0320	ABC transport protein, membrane permease	0	0	1	0
	405,205	406,161	318	+	CTLon_0321	CTLon_0321	ABC transport protein, membrane permease	1	0	3	0
	406,188	407,327	379	+	dxr	CTLon_0322	1-deoxy-D-xylulose 5-phosphate reductoisomerase	0	0	7	0
407,402	407,523	409,382	619	+	CTLon_0323	CTLon_0323	putative protease	0	0	5	1
410,321	409,329	410,306	325	-	CTLon_0324	CTLon_0324	exported protein	6	2	3	3
	410,464	411,561	365	-	recF	CTLon_0325	recombination protein F	0	0	2	0
	411,561	412,811	416	-	dnaN	CTLon_0326	DNA polymerase III subunit beta	1	0	7	0
	412,941	413,396	151	+	smpB	CTLon_0327	SsrA-binding protein	0	0	1	0
	413,374	414,324	316	-	apbE	CTLon_0328	thiamine biosynthesis lipoprotein	1	0	0	0
415,220	414,294	415,157	287	-	folD	CTLon_0329	methylenetetrahydrofolate dehydrogenase (NADP+)	0	1	1	0
	415,275	415,670	131	-	CTLon_0330	CTLon_0330	putative integral membrane protein	1	0	1	0
415,962	415,995	416,288	97	+	ltuB	CTLon_0331	late transcription unit B protein	110	13	13	3
	416,321	416,641	106	+	CTLon_0332	CTLon_0332	hypothetical protein	16487	473	991	48
416,614	416,651	418,333	560	+	CTLon_0333	CTLon_0333	hypothetical protein	153	13	10	3
418,307	418,330	418,812	160	+	CTLon_0334	CTLon_0334	hypothetical protein	20	1	5	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	418,826	419,911	361	-	CTLon_0335	CTLon_0335	phosphatidylcholine-hydrolyzing phospholipase D (PLD) protein	0	0	5	0
	420,083	421,822	579	-	CTLon_0336	CTLon_0336	putative decarboxylase	1	0	0	0
	421,942	422,211	89	-	rpmB	CTLon_0337	50S ribosomal protein L28	3	0	1	0
	422,360	423,943	527	-	malQ	CTLon_0338	4-alpha-glucanotransferase	0	0	2	0
	423,947	424,387	146	-	scc1, sycE	CTLon_0339	type III secretion chaperone	0	0	0	0
425,805	424,425	425,690	421	-	copN, lcrE	CTLon_0340	low calcium response protein E (TTSS effector protein)	3	0	25	7
	425,708	427,834	708	-	lcrD	CTLon_0341	low calcium response protein D (to be part of the TTSS apparatus)	0	0	13	0
428,948	427,834	428,916	360	-	sctU	CTLon_0342	type III secretion system protein	0	0	4	11
	429,173	430,273	366	+	CTLon_0343	CTLon_0343	translation-associated GTPase	0	0	0	0
	430,282	431,187	301	-	ribF	CTLon_0344	bifunctional riboflavin kinase/FMN adenylyltransferase	0	0	0	0
	431,144	431,869	241	-	truB	CTLon_0345	tRNA pseudouridine synthase B	0	0	3	0
	431,907	432,278	123	-	rbfA	CTLon_0346	ribosome-binding factor A	0	0	0	0
	432,285	434,975	896	-	infA	CTLon_0347	translation initiation factor IF-2	2	0	2	2
	434,932	436,236	434	-	nusA	CTLon_0348	transcription elongation factor NusA	0	0	2	0
438,099	436,354	438,063	569	-	rpsA	CTLon_0349	30S ribosomal protein S1	1	0	5	3
438,246	438,308	439,363	351	+	trxB	CTLon_0350	thioredoxin reductase	0	0	1	1
	439,369	439,728	119	+	acpS	CTLon_0351	4'-phosphopantetheinyl transferase	0	0	0	0
	439,729	440,190	153	-	CTLon_0352	CTLon_0352	hypothetical protein	0	0	0	0
440,288	440,326	440,787	153	+	CTLon_0353	CTLon_0353	hypothetical protein	0	0	1	2
	440,822	441,718	298	+	CTLon_0354	CTLon_0354	hypothetical protein	0	0	1	0
	441,708	442,604	298	-	fabI	CTLon_0355	enoyl-(acyl carrier protein) reductase	0	0	0	0
442,831	442,916	444,886	656	+	CTLon_0356	CTLon_0356	hypothetical protein	5	1	2	0
445,933	445,086	445,910	274	-	CTLon_0357	CTLon_0357	tRNA pseudouridine synthase A	2	1	1	0
	445,957	447,063	368	+	CTLon_0358	CTLon_0358	putative DNA glycosylase	0	0	0	0
	447,117	447,872	251	+	CTLon_0359	CTLon_0359	hypothetical protein	0	0	0	0
448,776	447,885	448,673	262	-	CTLon_0360	CTLon_0360	hypothetical protein	3	0	6	2
	448,802	450,436	544	-	hsp60_1	CTLon_0361	chaperonin GroEL	1	0	16	1
450,878	450,475	450,783	102	-	groES	CTLon_0362	co-chaperonin GroES	1	1	10	0
	450,955	452,781	608	-	pepF	CTLon_0363	oligoendopeptidase F	0	1	3	1
453,007	453,083	455,686	867	+	clpB	CTLon_0364	chaperone-protease ClpB	1	0	12	3
	455,712	457,172	486	+	CTLon_0365	CTLon_0365	hypothetical protein	0	0	5	0
457,308	457,358	457,840	160	+	incD	CTLon_0366	inclusion membrane protein D	12	12	10	9
	457,907	458,305	132	+	incE	CTLon_0367	inclusion membrane protein E	7	0	1	0
	458,310	458,624	104	+	incF	CTLon_0368	inclusion membrane protein F	2	0	4	0
	458,711	459,214	167	+	incG	CTLon_0369	inclusion membrane protein G	0	0	0	0
	459,380	460,120	246	-	incA	CTLon_0370	inclusion membrane protein A	0	0	0	0
	460,247	460,522	91	-	CTLon_0371	CTLon_0371	hypothetical protein	0	0	1	0
	460,621	461,307	228	+	araD	CTLon_0372	ribulose-phosphate 3-epimerase	2	2	2	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
461,094	461,294	461,851	185	+	efp	CTLon_0373	elongation factor P	0	0	3	0
	461,864	462,358	164	+	accB	CTLon_0374	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	0	0	1	0
	462,362	463,735	457	+	accC	CTLon_0375	acetyl-CoA carboxylase biotin carboxylase subunit	1	0	5	0
463,936	463,958	464,410	150	+	rplM	CTLon_0376	50S ribosomal protein L13	2	0	15	3
	464,436	464,825	129	+	rpsI	CTLon_0377	30S ribosomal protein S9	0	0	0	3
	464,873	465,724	283	-	CTLon_0378	CTLon_0378	chitinase	0	0	0	0
466,604	465,820	466,557	245	-	adk	CTLon_0379	adenylate kinase	0	0	2	2
466,726	466,764	467,408	214	+	CTLon_0380	CTLon_0380	putative ABC-membrane transport protein, inner membrane component	0	0	1	2
	467,405	468,106	233	+	CTLon_0381	CTLon_0381	ABC transporter, ATP-binding component	0	0	4	1
	468,109	471,525	1138	-	CTLon_0382	CTLon_0382	hypothetical protein	0	0	1	0
	471,522	472,799	425	-	CTLon_0383	CTLon_0383	hypothetical protein	0	1	1	0
	472,877	473,680	267	-	CTLon_0384	CTLon_0384	putative methyltransferase	0	0	0	0
473,889	474,137	474,550	137	+	CTLon_0385	CTLon_0385	hypothetical protein	2	0	0	0
	474,608	475,690	360	+	CTLon_0386	CTLon_0386	hypothetical protein	3	0	1	0
	475,780	476,499	239	+	CTLon_0387	CTLon_0387	putative phospholipase-carboxylesterase family protein	0	0	2	0
	476,518	477,363	281	+	CTLon_0388	CTLon_0388	hypothetical protein	0	0	1	1
	477,341	478,288	315	+	CTLon_0389	CTLon_0389	microsomal dipeptidase	0	0	1	1
479,650	478,285	479,565	426	-	oppA	CTLon_0390	oligopeptide transport system binding protein	1	0	3	1
479,646	479,741	480,427	228	+	CTLon_0391	CTLon_0391	exported protein	8	0	2	1
480,589	480,611	481,057	148	+	CTLon_0392	CTLon_0392	protein translocase	1	0	1	1
481,121	481,126	481,198		+	tRNA-Thr	CTLon_0108ao	tRNA	174	1293	78	252
481,209	481,209	481,291		+	tRNA-Tyr	CTLon_0108ar	tRNA	30	304	24	41
	481,457	482,314	285	+	CTLon_0393	CTLon_0393	hypothetical protein	1	0	13	0
484,146	482,316	483,158	280	+	CTLon_0394	CTLon_0394	hypothetical protein	0	0	17	2
	483,158	484,024	288	+	CTLon_0395	CTLon_0395	hypothetical protein	0	0	2	2
	484,210	486,054	614	+	pkn1	CTLon_0396	serine-threonine-protein kinase	1	0	9	1
	486,065	488,056	663	+	dnlJ	CTLon_0397	NAD-dependent DNA ligase LigA	0	0	4	0
488,129	488,154	492,503	1449	+	CTLon_0398	CTLon_0398	putative integral membrane protein	142	26	18	2
494,174	492,566	494,089	507	-	mhpA	CTLon_0399	FAD-dependent monooxygenase	6	1	12	0
495,317	494,289	495,236	315	-	CTLon_0400	CTLon_0400	hypothetical protein	0	0	0	1
495,602	495,635	495,793	52	+	rpmG	CTLon_0401	50S ribosomal protein L33	0	4	2	3
	495,829	497,340	503	+	CTLon_0402	CTLon_0402	lipoprotein releasing systen, inner membrane component	0	0	6	0
	497,345	498,022	225	+	CTLon_0403	CTLon_0403	lipoprotein release ATP-binding component	1	0	0	0
500,636	498,173	500,605	810	-	CTLon_0404	CTLon_0404	MAC/perforin family protein	1	0	2	1
	500,791	502,249	485.333	-	CTLon_0405	CTLon_0405	pseudo	1	0	4	0
503,748	502,342	503,280	312	-	CTLon_0406	CTLon_0406	phosphatidylcholine-hydrolyzing phospholipase D (PLD) protein	1	1	1	3
	503,794	504,996	400	-	CTLon_0407	CTLon_0407	phosphatidylcholine-hydrolyzing phospholipase D (PLD) protein	1	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	505,122	506,240	372	-	CTLon_0408	CTLon_0408	pseudo	0	0	0	0
	506,850	507,352	166.667	-	CTLon_0409	CTLon_0409	pseudo	0	0	0	1
	507,345	508,091	248	-	CTLon_0410	CTLon_0410	hypothetical protein	0	0	0	1
	508,231	508,472	79.6667	-	CTLon_0411	CTLon_0411	pseudo	0	0	0	0
	508,703	510,232	509	+	CTLon_0412	CTLon_0412	hypothetical protein	1	0	2	1
	510,285	510,545	86	+	CTLon_0413	CTLon_0413	hypothetical protein	0	0	2	0
	510,827	511,021	64	+	CTLon_0414	CTLon_0414	pseudo	0	0	0	0
	511,090	511,392	100	+	CTLon_0415	CTLon_0415	pseudo	0	0	0	0
511,430	511,448	511,732	94	+	trpR	CTLon_0416	Trp operon repressor	5	1	1	1
	512,081	513,259	392	+	trpB	CTLon_0417	tryptophan synthase subunit beta	0	0	0	0
	513,252	514,013	253	+	trpA	CTLon_0418	tryptophan synthase subunit alpha	1	0	0	1
	514,204	514,725	173	-	CTLon_0419	CTLon_0419	hypothetical protein	1	0	1	0
	514,765	514,956	63	-	CTLon_0420	CTLon_0420	hypothetical protein	0	0	0	0
	515,002	515,138	44.6667	-	CTLon_0421	CTLon_0421	pseudo	0	0	0	0
	515,333	516,922	529	+	oppA2	CTLon_0422	oligopeptide transport system binding protein	0	0	0	0
	516,949	517,356	135	-	dsbB	CTLon_0423	putative disulfide oxidoreductase	2	1	1	0
518,078	517,353	518,069	238	-	dsbG	CTLon_0424	disulfide bond chaperone	13	0	1	1
518,194	518,215	519,429	404	+	CTLon_0425	CTLon_0425	putative integral membrane protein	0	0	5	4
	519,431	519,943	170	-	CTLon_0426	CTLon_0426	hypothetical protein	12	1	0	0
	520,087	520,779	230	-	CTLon_0427	CTLon_0427	ABC transporter, ATP-binding protein	26	2	6	0
520,890	520,817	520,890		-	tRNA-Ile	CTLon_0108t	tRNA	751	1010	157	147
520,968	520,896	520,968		-	tRNA-Ala	CTLon_0108i	tRNA	1088	2159	293	484
521,827	521,087	521,797	236	-	CTLon_0428	CTLon_0428	hypothetical protein	220	53	22	13
	522,165	522,929	254	+	kdsB	CTLon_0429	3-deoxy-manno-octulosonate cytidylyltransferase	1	0	0	0
	522,905	524,524	539	+	pyrG	CTLon_0430	CTP synthetase	0	0	0	0
	524,511	524,957	148	+	CTLon_0431	CTLon_0431	Holliday junction resolvase-like protein	0	1	0	1
525,069	525,074	526,597	507	+	zwf	CTLon_0432	glucose-6-phosphate 1-dehydrogenase	0	0	4	2
	526,622	527,392	256	+	devB	CTLon_0433	6-phosphogluconolactonase	0	1	2	1
	527,389	528,261	290	-	dnaX	CTLon_0434	DNA polymerase III subunit delta'	2	0	0	0
	528,275	528,886	203	-	tdk	CTLon_0435	thymidylate kinase	1	0	1	0
	528,888	531,398	836	-	gyrA	CTLon_0436	DNA gyrase subunit A	1	0	6	1
	531,413	533,827	804	-	gyrB2	CTLon_0437	DNA gyrase subunit B	9	0	6	1
	533,830	534,180	116	-	CTLon_0438	CTLon_0438	hypothetical protein	0	0	1	0
	534,327	535,022	231	-	CTLon_0439	CTLon_0439	hypothetical protein	0	0	0	0
	535,126	536,244	372	-	tgt	CTLon_0440	queuine tRNA-ribosyltransferase	1	0	0	0
	536,371	537,783	470	-	mgtE	CTLon_0441	magnesium transport protein	0	0	1	0
	538,223	539,314	363	-	CTLon_0442	CTLon_0442	putative integral membrane protein	0	0	2	0
	539,540	539,860	106	+	CTLon_0443	CTLon_0443	hypothetical protein	0	0	5	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	539,936	540,952	338	+	gcp	CTLon_0444	O-sialoglycoprotein endopeptidase	0	0	3	0
	540,916	542,472	518	+	oppA3	CTLon_0445	oligopeptide transport system binding protein	0	0	2	0
	542,631	543,572	313	+	oppB	CTLon_0446	oligopeptide transport system membrane permease	0	0	0	0
	543,604	544,449	281	+	oppC	CTLon_0447	oligopeptide transport system membrane permease	0	0	0	0
	544,442	545,275	277	+	oppD	CTLon_0448	oligopeptide transport system ATP-binding protein	0	0	2	0
	545,290	546,033	247	+	oppF	CTLon_0449	oligopeptide transport system ATP-binding protein	0	0	0	0
546,300	546,345	547,094	249	+	CTLon_0450	CTLon_0450	hypothetical protein	0	0	9	4
	547,124	548,539	471	+	CTLon_0451	CTLon_0451	sodium:sulfate symporter	0	0	3	0
	548,559	550,220	553	+	pfkA	CTLon_0452	diphosphate--fructose-6-phosphate 1-phosphotransferase	0	0	4	0
	550,229	551,077	282	+	CTLon_0453	CTLon_0453	hypothetical protein	0	0	4	0
	551,059	552,705	548	+	pfkA_2	CTLon_0454	diphosphate--fructose-6-phosphate 1-phosphotransferase	0	0	4	0
552,834	552,760	552,833		-	tRNA-Met	CTLon_0108ab	tRNA	67	527	40	146
552,931	552,852	552,924		-	tRNA-Met	CTLon_0108ac	tRNA	2259	19740	561	8337
	552,951	554,246	431	-	gseA, kdtA	CTLon_0455	3-deoxy-D-manno-octulosonic-acid transferase	5	0	2	0
	554,243	556,702	819	-	leuS	CTLon_0456	leucyl-tRNA synthetase	24	4	8	2
556,871	556,900	558,168	422	+	hemL	CTLon_0457	glutamate-1-semialdehyde aminotransferase	1	0	1	0
	558,160	558,729	189	-	CTLon_0458	CTLon_0458	hypothetical protein	9	1	3	0
	558,749	559,195	148	-	CTLon_0459	CTLon_0459	hypothetical protein	16	1	2	0
	559,185	559,913	242	-	rpiA	CTLon_0460	ribose-5-phosphate isomerase A	20	1	7	0
561,699	560,008	561,651	547	-	CTLon_0461	CTLon_0461	hypothetical protein	276	60	48	14
561,864	561,955	563,001	348	+	dhnA	CTLon_0462	fructose-bisphosphate aldolase	1	0	5	2
	563,015	564,415	466	+	xasA	CTLon_0463	glutamate/gamma-aminobutyrate antiporter	1	0	0	0
564,509	564,509	565,273	254	+	CTLon_0464	CTLon_0464	hypothetical protein	7	3	2	1
565,352	565,404	566,255	283	+	surE	CTLon_0465	stationary phase survival protein SurE	4	0	5	4
	566,367	567,275	302	+	ubiA	CTLon_0466	prenyltransferase	0	0	0	0
	567,272	567,850	192	+	CTLon_0467	CTLon_0467	aromatic acid decarboxylase	0	0	0	0
	567,847	568,743	298	-	CTLon_0468	CTLon_0468	hypothetical protein	1	0	3	1
568,840	568,767	568,840		-	tRNA-Asp	CTLon_0108n	tRNA	164	567	73	488
568,920	568,848	568,920		-	tRNA-Val	CTLon_0108at	tRNA	138	703	51	193
	569,033	569,173	46	-	CTLon_0469	CTLon_0469	hypothetical protein	0	0	0	0
	569,200	569,586	128	-	CTLon_0470	CTLon_0470	hypothetical protein	0	0	0	0
	569,728	570,534	268	-	CTLon_0471	CTLon_0471	hypothetical protein	0	0	1	0
	570,926	571,369	147	-	CTLon_0472	CTLon_0472	hypothetical protein	0	0	0	0
	571,477	571,845	122	-	CTLon_0473	CTLon_0473	hypothetical protein	0	0	1	0
	571,944	572,459	171	-	CTLon_0474	CTLon_0474	hypothetical protein	0	0	0	0
573,010	572,588	572,989	133	-	CTLon_0475	CTLon_0475	hypothetical protein	0	0	0	1
	573,270	573,860	196	-	CTLon_0476	CTLon_0476	hypothetical protein	1	0	1	0
574,694	574,010	574,657	215	-	CTLon_0477	CTLon_0477	hypothetical protein	17	6	7	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
574,858	574,883	576,130	415	+	CTLon_0478	CTLon_0478	putative sodium:dicarboxylate symport protein	4	0	2	1
	576,314	577,786	490	+	CTLon_0479	CTLon_0479	sodium-dependent amino acid transporter	0	0	0	0
577,837	577,891	578,238	115	+	incB	CTLon_0480	inclusion membrane protein B	0	0	4	0
	578,315	578,851	178	+	incC	CTLon_0481	inclusion membrane protein C	1	0	0	0
	578,909	581,695	928	+	CTLon_0482	CTLon_0482	putative membrane transport protein	2	0	8	2
581,673	581,724	582,137	137	+	CTLon_0483	CTLon_0483	Crp family transcriptional regulator	0	0	0	0
	582,198	582,431	77	-	acpP	CTLon_0484	acyl carrier protein	0	0	2	0
	582,800	583,546	248	-	fabG	CTLon_0485	3-ketoacyl-(acyl-carrier-protein) reductase	1	0	9	0
	583,543	584,469	308	-	fabD	CTLon_0486	malonyl-CoA-[acyl-carrier-protein] transacylase	0	0	3	0
585,523	584,486	585,469	327	-	fabH	CTLon_0487	3-oxoacyl-(acyl carrier protein) synthase III	0	0	3	1
	585,607	586,209	200	+	recR	CTLon_0488	recombination protein RecR	0	0	1	0
586,440	586,583	588,961	792	+	CTLon_0489	CTLon_0489	outer membrane protein (variable surface antigen)	0	0	8	5
	589,028	589,549	173	+	ompH	CTLon_0490	Outer membrane protein	1	1	1	0
589,559	589,577	590,641	354	+	lpxD	CTLon_0491	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	1	0	14	1
	590,638	591,834	398	-	CTLon_0492	CTLon_0492	hypothetical protein	0	0	0	0
	592,056	593,078	340	+	pdhA	CTLon_0493	pyruvate dehydrogenase E1 component alpha subunit	0	0	1	1
	593,071	594,057	328	+	pdhB	CTLon_0494	pyruvate dehydrogenase E1 component beta subunit	0	0	1	0
	594,062	595,351	429	+	pdhC	CTLon_0495	branched-chain alpha-keto acid dehydrogenase subunit E2	0	0	0	1
597,907	595,378	597,822	814	-	glgP	CTLon_0496	glycogen phosphorylase	1	2	3	8
597,956	597,978	598,328	116	+	CTLon_0497	CTLon_0497	hypothetical protein	8	9	2	0
	598,382	599,752	456	-	dnaA	CTLon_0498	chromosomal replication initiation protein	0	0	4	1
602,370	599,829	602,192	787	-	oxaA	CTLon_0499	putative inner membrane protein translocase component YidC	0	0	6	0
603,349	602,502	603,320	272	-	lgt	CTLon_0500	prolipoprotein diacylglyceryl transferase	0	0	4	1
603,479	603,571	604,218	215	+	CTLon_0501	CTLon_0501	putative lipoprotein	1	0	13	7
604,055	604,222	604,992	256	+	CTLon_0502	CTLon_0502	inner membrane protein	0	0	4	0
605,142	605,142	605,034		-	SRP RNA		7SL-RNA, signal recognition particle RNA	64	276	18	92
	605,200	605,583	127	-	CTLon_0503	CTLon_0503	hypothetical protein	0	0	0	0
605,749	605,772	607,016	414	+	CTLon_0504	CTLon_0504	putative membrane transport protein	0	0	0	1
	607,006	608,220	404	+	CTLon_0505	CTLon_0505	hypothetical protein	0	0	0	0
	608,224	609,348	374	-	CTLon_0506	CTLon_0506	putative cysteine desulfurase	1	0	4	0
610,158	609,354	610,100	248	-	CTLon_0507	CTLon_0507	protein phosphatase 2C	0	0	4	0
	610,419	610,910	163	+	CTLon_0508	CTLon_0508	hypothetical protein	0	0	2	0
	610,919	611,617	232	+	dnaQ	CTLon_0509	DNA polymerase III subunit epsilon	0	0	5	0
	611,614	612,384	256	+	CTLon_0510	CTLon_0510	hypothetical protein	0	0	0	0
	612,377	612,967	196	+	CTLon_0511	CTLon_0511	hypothetical protein	0	0	0	0
	612,988	614,928	646	-	CTLon_0512	CTLon_0512	ABC transporter, ATP-binding and membrane components	1	0	3	1
615,895	614,894	615,868	324	-	accA	CTLon_0513	acetyl-CoA carboxylase carboxyltransferase subunit alpha	8	0	5	1
	616,001	617,182	393	-	CTLon_0514	CTLon_0514	hypothetical protein	1	0	10	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
617,661	617,300	617,602	100	-	ihfA	CTLon_0515	integration host factor alpha-subunit	3	0	28	30
	617,822	618,601	259	-	amiA	CTLon_0516	N-acetylmuramoyl-L-alanine amidase	0	0	1	0
620,009	618,516	619,967	483	-	murE	CTLon_0517	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	0	0	1	5
	620,290	622,233	647	-	pbp	CTLon_0518	penicillin-binding protein	0	0	0	0
	622,220	622,507	95	-	CTLon_0519	CTLon_0519	hypothetical protein	0	0	2	0
	622,507	623,409	300	-	mraW	CTLon_0520	S-adenosyl-methyltransferase MraW	0	0	0	0
623,665	623,687	624,253	188	+	CTLon_0521	CTLon_0521	hypothetical protein	3	4	15	26
	624,260	624,679	139	+	CTLon_0522	CTLon_0522	hypothetical protein	1	0	6	0
624,859	624,921	626,288	455	+	dnaA2	CTLon_0523	chromosomal replication initiation protein	1	0	7	1
	626,340	626,924	194	+	CTLon_0524	CTLon_0524	hypothetical protein	0	0	3	0
	626,896	627,555	219	-	CTLon_0525	CTLon_0525	hypothetical protein	0	0	0	1
627,525	627,557	629,068	503	+	nqrB	CTLon_0526	Na(+)-translocating NADH-quinone reductase subunit B	0	0	6	0
	629,072	630,022	316	+	nqrC	CTLon_0527	Na(+)-translocating NADH-quinone reductase subunit C	0	0	2	0
	630,012	630,653	213	+	nqrD	CTLon_0528	Na(+)-translocating NADH-quinone reductase subunit D	0	0	1	0
633,960	630,659	631,393	244	+	nqrE	CTLon_0529	Na(+)-translocating NADH-quinone reductase subunit E	0	0	0	0
	631,417	631,770	117	-	gcsH	CTLon_0530	glycine cleavage system protein H	2	0	1	1
	631,790	633,862	690	-	CTLon_0531	CTLon_0531	hypothetical protein	1	1	2	0
	634,057	635,481	474	-	CTLon_0532	CTLon_0532	phospholipase D	1	0	0	1
	635,487	636,206	239	-	lplA	CTLon_0533	lipoate-protein ligase A	1	0	0	0
636,444	636,471	639,035	854	+	clpC	CTLon_0534	ATP-dependent Clp protease	1	1	19	4
640,245	639,016	640,092	358	-	trmU	CTLon_0535	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	4	2	1	0
640,300	640,323	642,017	564	+	CTLon_0536	CTLon_0536	hypothetical protein	191	139	46	47
643,239	642,104	643,249	381	-	CTLon_0537	CTLon_0537	hypothetical protein	4	0	41	2
	643,306	643,983	225	-	CTLon_0538	CTLon_0538	PTS family membrane transport protein IIA subunit	0	0	0	0
	643,986	644,462	158	-	ptsN_2	CTLon_0539	PTS family membrane transport protein IIA subunit	0	1	0	0
644,923	644,464	644,901	145	-	dut	CTLon_0540	deoxyuridine 5'-triphosphate nucleotidohydrolase	0	1	1	0
645,892	644,938	645,864	308	-	accD	CTLon_0541	acetyl-CoA carboxylase subunit beta	1	1	2	1
	645,935	646,555	206	-	sodM	CTLon_0542	superoxide dismutase	2	0	0	0
648,538	646,687	648,468	593	-	mrsA	CTLon_0543	phosphoglucomutase	0	0	3	0
	648,620	649,087	155	-	CTLon_0544	CTLon_0544	hypothetical protein	0	0	0	0
	649,196	649,891	231	+	rnc	CTLon_0545	ribonuclease III	0	0	0	0
	649,875	651,239	454	+	radA	CTLon_0546	DNA repair protein RadA	0	0	1	1
651,213	651,283	651,942	219	+	hemC	CTLon_0547	porphobilinogen deaminase	0	0	1	0
651,986	652,336	652,684	115.333	+	CTLon_0548	CTLon_0548	pseudo	25	5	0	0
	652,738	655,542	934	-	pknD	CTLon_0549	serine/threonine-protein kinase	0	0	0	0
	655,557	658,376	939	-	valS	CTLon_0550	valyl-tRNA synthetase	0	1	3	3
659,112	658,503	659,018	171	-	CTLon_0551	CTLon_0551	hypothetical protein	0	0	3	2

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	658,939	659,364	141	-	atpK	CTLon_0552	V-type ATP synthase subunit K	0	0	0	0
	659,426	661,375	649	-	atpI	CTLon_0553	V-type ATP synthase subunit I	0	0	2	0
	661,381	661,992	203	-	atpD	CTLon_0554	V-type ATP synthase subunit D	0	0	1	0
	661,977	663,293	438	-	atpB	CTLon_0555	V-type ATP synthase subunit B	1	0	2	0
	663,296	665,071	591	-	atpA	CTLon_0556	V-type ATP synthase subunit A	0	0	0	0
	665,065	665,865	266	-	CTLon_0557	CTLon_0557	hypothetical protein	0	0	1	0
666,719	666,032	666,658	208	-	atpE	CTLon_0558	V-type ATP synthase subunit E	4	0	2	2
666,717	666,767	667,477	236	+	CTLon_0559	CTLon_0559	hypothetical protein	2	1	8	1
	667,485	667,856	123	-	CTLon_0560	CTLon_0560	hypothetical protein	1	0	8	0
668,909	667,901	668,884	327	-	tal	CTLon_0561	transaldolase B	1	1	38	43
	668,996	673,186	1396	-	rpoC	CTLon_0562	DNA-directed RNA polymerase subunit beta'	0	0	5	0
	673,211	676,969	1252	-	rpoB	CTLon_0563	DNA-directed RNA polymerase subunit beta	0	2	3	2
	677,330	677,722	130	-	rplL	CTLon_0564	50S ribosomal protein L7/L12	1	0	14	1
	677,754	678,272	172	-	rplJ	CTLon_0565	50S ribosomal protein L10	0	0	6	0
	678,294	678,992	232	-	rplA	CTLon_0566	50S ribosomal protein L1	1	0	2	0
679,496	679,015	679,440	141	-	rplK	CTLon_0567	50S ribosomal protein L11	1	0	5	0
	679,546	680,094	182	-	nusG	CTLon_0568	transcription antitermination protein NusG	2	0	2	0
	680,098	680,346	82	-	secE	CTLon_0569	preprotein translocase subunit SecE	0	0	3	0
680,448	680,376	680,448		-	tRNA-Trp	CTLon_0108aq	tRNA	40	289	19	89
	680,489	681,673	394	-	tufA	CTLon_0570	elongation factor Tu	2	0	5	0
681,790	681,719	681,790		-	tRNA-Thr	CTLon_0108ap	tRNA	37	408	17	155
682,352	682,021	682,242	73	-	infA2	CTLon_0571	translation initiation factor IF-1	0	0	0	0
682,644	682,653	683,564	303	+	CTLon_0572	CTLon_0572	hypothetical protein	3	2	18	18
	683,561	684,007	148	+	CTLon_0573	CTLon_0573	hypothetical protein	0	0	0	0
	684,059	686,030	656.333	-	CTLon_0574	CTLon_0574	pseudo	0	0	0	0
	686,220	686,402	60	-	CTLon_0576	CTLon_0576	hypothetical protein	0	0	0	0
687,001	687,012	687,084		+	tRNA-Met	CTLon_0108aa	tRNA	12	61	3	11
687,809	687,125	687,751	208	-	trpC	CTLon_0577	N-(5'-phosphoribosyl)anthranilate isomerase	2	1	4	0
	687,947	688,771	274	+	tpiS	CTLon_0578	triosephosphate isomerase	0	0	1	0
	688,785	690,335	516	+	xseA	CTLon_0579	exodeoxyribonuclease VII large subunit	0	0	1	1
	690,319	690,537	72	+	xseB	CTLon_0580	exodeoxyribonuclease VII small subunit	0	0	0	0
	690,544	690,816	90	+	CTLon_0581	CTLon_0581	hypothetical protein	0	0	0	0
	690,813	692,735	640	+	dxs	CTLon_0582	1-deoxy-D-xylulose-5-phosphate synthase	0	0	0	0
692,801	692,832	694,289	485	+	pykF	CTLon_0583	pyruvate kinase	0	1	2	3
	694,312	699,672	1786	+	uvrA	CTLon_0584	excinuclease ABC subunit A	0	0	3	0
699,824	699,890	701,290	466	+	dnaX	CTLon_0585	DNA polymerase III subunits gamma and tau	1	0	7	2
	701,283	701,573	96	+	CTLon_0586	CTLon_0586	hypothetical protein	0	0	0	0
	701,583	703,298	571	-	ptsI	CTLon_0587	phosphoenolpyruvate-protein phosphotransferase	0	0	2	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
703,654	703,298	703,627	109	-	ptsH	CTLon_0588	phosphocarrier protein HPr	0	0	0	1
	703,765	704,226	153	-	CTLon_0589	CTLon_0589	hypothetical protein	0	0	0	1
	704,283	704,465	60	-	CTLon_0590	CTLon_0590	hypothetical protein	0	0	0	0
704,333	704,510	706,039	509	+	CTLon_0591	CTLon_0591	hypothetical protein	0	0	0	0
708,173	706,111	708,147	678	-	CTLon_0592	CTLon_0592	2-oxoisovalerate dehydrogenase alpha subunit	4	0	1	3
	708,182	709,360	392	-	dnaJ	CTLon_0593	heat shock chaperone protein	0	0	6	0
709,629	709,389	709,565	58	-	rpsU	CTLon_0594	30S ribosomal protein S21	1	1	2	2
710,420	709,763	710,395	210	-	CTLon_0595	CTLon_0595	O-sialoglycoprotein endopeptidase	6	0	8	0
710,628	710,705	713,164	819	+	lon	CTLon_0596	ATP-dependent protease La	2	0	9	0
	713,266	713,631	121	+	CTLon_0597	CTLon_0597	hypothetical protein	0	0	0	0
713,809	713,974	714,888	304	+	CTLon_0598	CTLon_0598	ribonuclease Z	4	0	1	1
	714,950	715,897	315	+	xerC	CTLon_0599	site-specific tyrosine recombinase XerC	0	0	0	2
715,951	715,951	717,540	529	+	CTLon_0600	CTLon_0600	ABC transporter, ATP-binding component	3	0	3	4
	717,576	718,166	196	+	CTLon_0601	CTLon_0601	Maf-like protein	0	0	0	0
	718,148	719,848	566	+	CTLon_0602	CTLon_0602	putative lipoprotein	0	0	2	0
	719,855	721,957	700	+	CTLon_0603	CTLon_0603	hypothetical protein	0	0	3	0
722,418	722,031	722,339	102	-	secG	CTLon_0604	preprotein translocase subunit SecG	0	1	2	5
	722,495	723,040	181	-	def	CTLon_0605	peptide deformylase	0	0	1	0
	723,315	724,148	277	-	ksgA	CTLon_0606	dimethyladenosine transferase	1	0	1	0
725,411	724,164	725,225	353	-	CTLon_0607	CTLon_0607	hypothetical protein	12	9	15	61
727,695	725,530	727,644	704	-	CTLon_0608	CTLon_0608	hypothetical protein	33	3	6	0
727,859	727,859	727,945		+	tRNA-Ser	CTLon_0108ak	tRNA	1193	18380	272	5562
728,024	728,101	728,433	110	+	CTLon_0609	CTLon_0609	hypothetical protein	22	6	4	0
	728,459	728,993	177.333	+	CTLon_0610	CTLon_0610	pseudo	32	4	9	0
729,837	729,053	729,643	196	-	CTLon_0611	CTLon_0611	putative inner membrane protein	0	0	3	3
	729,895	730,521	208	-	CTLon_0612	CTLon_0612	hypothetical protein	7	2	4	1
	730,683	731,543	286	-	dapA	CTLon_0613	dihydrodipicolinate synthase	8	0	0	0
	731,553	732,848	431	-	lysC	CTLon_0614	aspartate kinase	4	0	5	0
	732,841	733,845	334	-	asd	CTLon_0615	aspartate-semialdehyde dehydrogenase	3	1	2	0
734,669	733,855	734,616	253	-	dapB	CTLon_0616	dihydrodipicolinate reductase	4	1	0	1
734,683	733,855	734,616	253	-	dapB	CTLon_0616	dihydrodipicolinate reductase	4	1	0	1
736,545	734,793	736,520	575	-	CTLon_0617	CTLon_0617	putative integral membrane protein	99	36	19	5
	736,690	738,012	440	+	aroA	CTLon_0618	3-phosphoshikimate 1-carboxyvinyltransferase	1	2	4	0
737,686	737,954	738,508	184	+	aroL	CTLon_0619	shikimate kinase	0	0	0	0
	738,501	739,574	357	+	aroC	CTLon_0620	chorismate synthase	0	0	0	1
	739,571	740,692	373	+	aroB	CTLon_0621	3-dehydroquinate synthase	0	0	1	0
	740,673	742,115	480	+	aroE	CTLon_0622	bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase protein	1	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	742,157	742,942	261	+	CTLon_0623	CTLon_0623	hypothetical protein	0	0	0	0
743,004	743,083	744,411	442	+	CTLon_0624	CTLon_0624	hypothetical protein	1	1	26	4
	744,480	745,067	195	+	CTLon_0625	CTLon_0625	pseudo	0	0	10	1
	745,083	746,534	483	+	CTLon_0626	CTLon_0626	arginine/ornithine antiporter	0	0	0	0
746,653	746,706	747,764	352	+	CTLon_0627	CTLon_0627	putative oxidoreductase	53	7	10	2
748,855	747,806	748,786	326	-	mdhC	CTLon_0628	malate dehydrogenase	28	1	7	0
749,153	748,976	749,116	46	-	ltuA	CTLon_0629	late transcription unit A protein	148	99	24	21
750,849	749,231	750,808	525	-	pgi	CTLon_0630	glucose-6-phosphate isomerase	2	0	12	1
	750,921	752,264	447	-	CTLon_0631	CTLon_0631	putative nucleotide-binding protein	0	0	3	0
	752,277	753,077	266	-	phnP	CTLon_0632	metal-dependent hydrolase	0	0	1	0
753,904	753,106	753,879	257	-	artJ	CTLon_0633	arginine transport substrate-binding protein	1	0	1	4
	753,948	754,784	278	-	aroG	CTLon_0634	3-deoxy-7-phosphoheptulonate synthase	8	0	2	0
755,132	754,921	755,112	63	-	CTLon_0635	CTLon_0635	hypothetical protein	11	5	9	14
755,258	755,295	756,026	243	+	CTLon_0636	CTLon_0636	hypothetical protein	1	2	0	0
	756,037	757,656	539	-	CTLon_0637	CTLon_0637	hypothetical protein	0	0	6	0
	757,671	758,006	111	-	CTLon_0638	CTLon_0638	hypothetical protein	0	0	5	1
	758,003	758,872	289	-	CTLon_0639	CTLon_0639	hypothetical protein	1	0	2	1
	759,159	761,234	691	+	CTLon_0640	CTLon_0640	hypothetical protein	0	0	3	0
	761,248	761,595	115	-	CTLon_0641	CTLon_0641	hypothetical protein	0	0	1	0
	761,777	763,003	408	+	CTLon_0642	CTLon_0642	hypothetical protein	0	0	7	0
763,069	763,111	764,295	394	+	dapL	CTLon_0643	L,L-diaminopimelate aminotransferase	1	0	0	2
	764,303	765,310	335	+	CTLon_0644	CTLon_0644	ABC transporter substrate-binding component	0	0	1	0
766,482	765,319	766,449	376	-	CTLon_0645	CTLon_0645	hypothetical protein	11	0	3	0
766,617	766,650	768,395	581	+	proS	CTLon_0646	prolyl-tRNA synthetase	0	0	5	0
768,459	768,464	769,642	392	+	CTLon_0647	CTLon_0647	heat-inducible transcription repressor	0	0	7	0
	769,639	770,211	190	+	grpE	CTLon_0648	HSP-70 cofactor	0	0	2	0
	770,237	772,219	660	+	dnaK	CTLon_0649	molecular chaperone DnaK	0	0	12	1
772,351	772,512	774,596	694	+	vacB	CTLon_0650	exoribonuclease II	0	1	8	1
774,713	774,852	775,616	254	+	CTLon_0651	CTLon_0651	hypothetical protein	2	0	9	2
775,625	775,625	776,032		+	RNaseP RNA		RNaseP RNA	3	1	53	2
	776,039	777,025	328	-	CTLon_0652	CTLon_0652	carbohydrate isomerase	0	0	2	0
778,371	777,057	778,223	388	-	sucB	CTLon_0653	branched-chain alpha-keto acid dehydrogenase subunit E2	2	1	1	1
	778,469	779,707	412	-	gltT	CTLon_0654	Sodium:dicarboxylate symport protein	4	0	1	1
780,929	779,704	780,813	369	-	CTLon_0655	CTLon_0655	tetraacyldisaccharide 4'-kinase	1	8	5	4
	780,979	781,788	269	-	CTLon_0656	CTLon_0656	23S rRNA methyltransferase	0	0	0	0
	781,776	782,603	275	-	CTLon_0657	CTLon_0657	N6-adenine-specific DNA methylase	0	0	0	0
	782,600	783,199	199	-	ribC	CTLon_0658	riboflavin synthase subunit alpha	0	0	0	0
783,438	783,592	784,056	154	+	CTLon_0659	CTLon_0659	transcriptional regulator NrdR	0	0	1	3

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	784,070	784,444	124	+	dksA	CTLon_0660	dnaK suppressor protein	0	0	1	0
784,737	784,450	784,953	167	+	lspA	CTLon_0661	lipoprotein signal peptidase	0	0	2	2
785,019	785,055	786,416	453	+	CTLon_0662	CTLon_0662	Sodium/alanine symporter protein	3	4	7	1
	786,632	787,909	425	+	pcnB	CTLon_0663	polyA polymerase	0	0	7	0
	787,970	789,793	607	+	lpxB	CTLon_0664	lipid-A-disaccharide synthase	2	0	1	1
	789,900	792,827	975	+	pmpA	CTLon_0665	polymorphic outer membrane protein	0	0	0	0
792,931	792,966	798,215	1749	+	pmpB	CTLon_0666	polymorphic outer membrane protein	5	0	22	10
798,356	798,391	803,715	1774	+	pmpC	CTLon_0667	polymorphic outer membrane protein	0	0	13	7
803,959	803,873	803,959		-	tRNA-Ser	CTLon_0108al	tRNA	84	383	72	180
804,246	804,268	805,098	276	+	CTLon_0668	CTLon_0668	metal transporter, metal-binding component	3	1	1	0
	805,095	805,805	236	+	CTLon_0669	CTLon_0669	metal transporter, ATP-binding component	0	0	1	0
	805,796	806,677	293	+	CTLon_0670	CTLon_0670	metal transporter, membrane permease component	1	0	0	0
	806,593	807,600	335	-	CTLon_0671	CTLon_0671	GTPase ObgE	0	0	0	0
	807,690	807,941	83	-	rpmA	CTLon_0672	50S ribosomal protein L27	0	1	0	0
	807,972	808,295	107	-	rplU	CTLon_0673	50S ribosomal protein L21	1	0	0	0
808,690	808,612	808,684		-	tRNA-Phe	CTLon_0108ad	tRNA	15	135	2	34
808,851	808,845	809,546	233	+	CTLon_0674ae	CTLon_0674ae	putative inner membrane protein	0	0	2	0
809,647	809,731	809,892	53	+	CTLon_0675af	CTLon_0675af	hypothetical protein	0	0	5	1
	809,909	810,070	53	+	CTLon_0676	CTLon_0676	hypothetical protein	0	0	0	1
	810,084	810,569	161	+	CTLon_0677	CTLon_0677	hypothetical protein	1	0	0	0
810,632	810,702	811,811	369	+	CTLon_0678	CTLon_0678	putative cation efflux protein	0	0	11	24
811,953	812,000	812,350	116	+	rsbV	CTLon_0679	anti-sigma F factor antagonist	0	0	7	1
812,437	812,528	814,393	621	+	CTLon_0680	CTLon_0680	hypothetical protein	2	0	6	1
	814,590	815,699	369	+	CTLon_0681	CTLon_0681	hypothetical protein	0	0	2	0
	815,672	816,493	273	+	CTLon_0682	CTLon_0682	hypothetical protein	0	0	0	0
	816,429	817,118	229	+	ubiE	CTLon_0683	ubiquinone/menaquinone biosynthesis methyltransferase	0	0	0	1
818,133	817,144	818,133	329	-	CTLon_0684	CTLon_0684	hypothetical protein	7	1	5	4
	818,194	819,021	275	-	dapF	CTLon_0685	diaminopimelate epimerase	0	0	4	0
	818,990	819,568	192	-	clpP	CTLon_0686	ATP-dependent Clp protease proteolytic subunit	0	0	3	0
821,311	819,580	821,073	497	-	glyA	CTLon_0687	serine hydroxymethyltransferase	0	0	4	6
	821,437	822,108	223	+	CTLon_0688	CTLon_0688	uroporphyrinogen-III synthase	0	0	0	0
	822,211	822,747	178	+	ispF	CTLon_0689	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	0	0	0	0
	822,744	823,796	350	-	CTLon_0690	CTLon_0690	putative oxidoreductase	1	0	1	0
	823,813	824,130	105	-	rpsJ	CTLon_0691	30S ribosomal protein S10	0	0	2	0
	824,138	826,222	694	-	fusA	CTLon_0692	elongation factor G	0	0	10	0
	826,264	826,737	157	-	rpsG	CTLon_0693	30S ribosomal protein S7	0	0	1	0
827,189	826,787	827,176	129	-	rpsL	CTLon_0694	30S ribosomal protein S12	2	1	8	0
	827,418	827,756	112	+	CTLon_0695	CTLon_0695	hypothetical protein	0	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	827,914	829,863	649	+	tsp	CTLon_0696	carboxy-terminal processing protease	1	0	2	0
830,525	829,970	830,422	150	-	crpA	CTLon_0697	cysteine-rich membrane protein	31	3	3	1
	830,600	832,261	553	-	omcB	CTLon_0698	60kD cysteine-rich outer membrane protein	487	28	81	6
832,765	832,408	832,674	88	-	omcA	CTLon_0699	cysteine-rich outer membrane protein	2824	576	195	89
832,994	833,011	833,235	74	+	CTLon_0700	CTLon_0700	putative lipoprotein	80	684	8	87
	833,232	834,752	506	-	gltX	CTLon_0701	glutamyl-tRNA synthetase	0	0	0	0
835,780	835,023	835,574	183	-	euo	CTLon_0702	hypothetical protein	4	0	20	1
	835,978	837,732	584	-	recJ	CTLon_0703	single-stranded-DNA-specific exonuclease	1	0	2	0
	837,850	842,052	1400	-	secF, secG	CTLon_0704a	bifunctional preprotein translocase subunit SecD/SecF	0	0	8	0
	841,984	842,052	22	-	secF	CTLon_0704	hypothetical protein	0	0	0	0
842,954	842,472	842,804	110	-	CTLon_0705	CTLon_0705	hypothetical protein	2	1	3	1
	843,266	844,027	253	+	uppS	CTLon_0706	undecaprenyl pyrophosphate synthetase	0	0	1	1
	844,033	844,950	305	+	cdsA	CTLon_0707	phosphatidate cytidylyltransferase	0	0	0	0
	844,947	845,597	216	+	cmk	CTLon_0708	cytidylate kinase	0	0	0	0
	845,594	846,244	216	+	plsC	CTLon_0709	1-acyl-sn-glycerol-3-phosphate acyltransferase	0	0	1	0
	846,259	847,950	563	+	argS	CTLon_0710	arginyl-tRNA synthetase	1	0	0	0
	847,990	849,324	444	-	murA	CTLon_0711	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	1	0	2	0
	849,536	852,541	1001	+	tarp	CTLon_0712	translocated actin-recruiting protein	12	2	17	0
853,448	852,593	853,309	238	-	CTLon_0713	CTLon_0713	hypothetical protein	1	0	3	3
	853,492	853,995	167	-	CTLon_0714	CTLon_0714	ribosomal-protein-alanine acetyltransferase	0	2	2	0
	853,992	855,099	368	-	prfB	CTLon_0715	peptide chain release factor 2	0	1	3	1
855,365	855,419	855,679	86	+	CTLon_0716	CTLon_0716	hypothetical protein	0	0	1	4
	855,737	856,726	329	+	CTLon_0717	CTLon_0717	putative metallo-phosphoesterase	1	0	8	0
	856,716	857,375	219	+	ispD	CTLon_0718	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	0	0	0	0
	857,384	858,187	267	+	truA	CTLon_0719	tRNA pseudouridine synthase A	0	0	0	0
	858,139	858,813	224	-	CTLon_0720	CTLon_0720	HAD superfamily hydrolase	0	0	0	0
	858,906	859,547	213	+	CTLon_0721	CTLon_0721	hypothetical protein	0	0	0	0
859,610	859,677	859,759		+	tRNA-Leu	CTLon_0108v	tRNA	79	261	93	124
	859,841	860,170	109	+	CTLon_0722	CTLon_0722	hypothetical protein	0	0	2	1
	860,148	861,206	352	+	atoS	CTLon_0723	two component regulator, histidine kinase	0	0	2	0
	861,249	862,409	386	+	atoC	CTLon_0724	two component system response regulator	0	0	2	0
862,467	862,476	862,549		+	tRNA-Arg	CTLon_0108k	tRNA	12	43	12	25
	862,636	863,172	178	+	CTLon_0725	CTLon_0725	hypothetical protein	0	0	0	0
	863,142	863,873	243	+	recO	CTLon_0726	DNA repair protein RecO	0	0	0	0
	863,870	864,472	200	-	CTLon_0727	CTLon_0727	hypothetical protein	0	0	0	1
864,684	864,603	864,684		-	tRNA-Leu	CTLon_0108x	tRNA	21	218	23	53
	864,950	865,744	264	+	CTLon_0728	CTLon_0728	hypothetical protein	1	0	0	0
	865,750	866,064	104	+	CTLon_0729	CTLon_0729	hypothetical protein	0	0	1	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	866,003	866,932	309	-	CTLon_0730	CTLon_0730	hypothetical protein	0	0	0	0
	867,020	869,392	790	+	pheT	CTLon_0731	phenylalanyl-tRNA synthetase subunit beta	0	0	2	0
	869,389	870,354	321	+	CTLon_0732	CTLon_0732	hypothetical protein	0	0	4	1
870,356	870,373	870,885	170	+	CTLon_0733	CTLon_0733	putative DNA methyltransferase	5	1	17	2
	870,882	872,618	578	-	oppC2	CTLon_0734	oligonucleotide transport system permease	0	0	2	0
	872,620	874,035	471	-	oppB2	CTLon_0735	oligopeptide transport system membrane permease	0	0	1	0
876,262	874,080	876,170	696	-	oppA4	CTLon_0736	oligopeptide transport system, binding protein	0	0	3	0
	876,776	877,507	243	-	CTLon_0737	CTLon_0737	hypothetical protein	0	0	1	0
878,474	877,669	878,322	217	-	CTLon_0738	CTLon_0738	hypothetical protein	0	1	0	3
878,752	878,790	879,155	121	+	CTLon_0739	CTLon_0739	hypothetical protein	7	2	0	0
	879,134	880,132	332	-	CTLon_0740	CTLon_0740	hypothetical protein	0	0	1	0
	880,289	881,233	314	-	hemZ	CTLon_0741	ferrochelatase	0	0	1	0
882,029	881,255	882,040	261	-	fliY	CTLon_0742	glutamine-binding protein	0	0	2	3
	882,086	882,658	190	-	CTLon_0743	CTLon_0743	methyltransferase	0	0	0	0
	882,655	883,389	244	-	CTLon_0744	CTLon_0744	putative phosphohydrolase	3	0	2	0
884,819	883,478	884,803	441	-	glgC	CTLon_0745	glucose-1-phosphate adenylyltransferase	5	6	9	4
884,971	884,994	885,245	83	+	CTLon_0746	CTLon_0746	hypothetical protein	0	0	1	1
	885,255	886,649	464	-	rho	CTLon_0747	transcription termination factor Rho	3	0	3	0
	886,646	887,254	202	-	CTLon_0748	CTLon_0748	dephospho-CoA kinase	2	0	2	1
	887,248	889,848	866	-	polA	CTLon_0749	DNA polymerase I	0	0	10	0
890,899	889,865	890,860	331	-	sohB	CTLon_0750	exported protease IV	0	0	5	4
	891,000	892,622	540	-	CTLon_0751	CTLon_0751	putative nucleotide transport protein	7	1	3	0
893,455	892,834	893,340	168	-	pgsA	CTLon_0752	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	9	1	10	14
893,389	892,834	893,340	168	-	pgsA	CTLon_0752	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	9	1	10	14
893,615	893,522	893,610		-	tRNA-Ser	CTLon_0108am	tRNA	480	4753	183	1125
	893,736	893,885	49	-	CTLon_0753	CTLon_0753	hypothetical protein	1	0	0	0
893,815	893,854	895,272	472	+	dnaB	CTLon_0754	replicative DNA helicase	0	0	0	4
895,702	895,567	897,399	610	+	gidA	CTLon_0755	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	0	0	5	2
	897,389	898,090	233	+	lplA	CTLon_0756	lipoate-protein ligase A	0	0	0	0
898,672	898,147	898,572	141	-	ndk	CTLon_0757	nucleoside diphosphate kinase	5	3	7	16
898,375	898,147	898,572	141	-	ndk	CTLon_0757	nucleoside diphosphate kinase	5	3	7	16
	898,732	899,334	200	-	ruvA	CTLon_0758	Holliday junction DNA helicase motor protein	0	0	3	0
899,904	899,354	899,866	170	-	ruvC	CTLon_0759	Holliday junction resolvase	0	0	2	0
900,556	899,974	900,423	149	-	CTLon_0760	CTLon_0760	hypothetical protein	0	0	10	6
900,769	900,613	900,695		-	tRNA-Leu	CTLon_0108y	tRNA	98	776	37	163

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	900,815	901,681	288	-	CTLon_0761	CTLon_0761	hypothetical protein	1	0	5	0
	901,692	902,696	334	-	gapA	CTLon_0762	glyceraldehyde 3-phosphate dehydrogenase	0	0	1	1
	902,740	903,165	141	-	rplQ	CTLon_0763	50S ribosomal protein L17	0	0	1	0
	903,174	904,307	377	-	rpoA	CTLon_0764	DNA-directed RNA polymerase subunit alpha	1	0	6	0
	904,328	904,726	132	-	rpsK	CTLon_0765	30S ribosomal protein S11	0	0	4	0
905,140	904,748	905,116	122	-	rpsM	CTLon_0766	30S ribosomal protein S13	4	3	3	1
	905,172	906,545	457	-	secY	CTLon_0767	preprotein translocase subunit SecY	0	0	1	0
	906,568	907,002	144	-	rplO	CTLon_0768	50S ribosomal protein L15	1	0	1	0
	906,995	907,492	165	-	rpsE	CTLon_0769	30S ribosomal protein S5	0	0	0	0
	907,507	907,878	123	-	rplR	CTLon_0770	50S ribosomal protein L18	0	0	0	0
	907,900	908,451	183	-	rplF	CTLon_0771	50S ribosomal protein L6	0	0	2	0
	908,479	908,880	133	-	rpsH	CTLon_0772	30S ribosomal protein S8	1	0	2	0
	908,898	909,440	180	-	rplE	CTLon_0773	50S ribosomal protein L5	0	0	0	0
	909,442	909,777	111	-	rplX	CTLon_0774	50S ribosomal protein L24	1	0	2	0
	909,790	910,158	122	-	rplN	CTLon_0775	50S ribosomal protein L14	0	0	2	0
910,449	910,175	910,426	83	-	rpsQ	CTLon_0776	30S ribosomal protein S17	0	1	2	0
	910,419	910,637	72	-	rpmC	CTLon_0777	50S ribosomal protein L29	0	0	2	0
	910,639	911,055	138	-	rplP	CTLon_0778	50S ribosomal protein L16	0	0	1	0
	911,088	911,762	224	-	rpsC	CTLon_0779	30S ribosomal protein S3	0	0	3	0
	911,772	912,107	111	-	rplV	CTLon_0780	50S ribosomal protein L22	1	0	0	0
	912,126	912,392	88	-	rpsS	CTLon_0781	30S ribosomal protein S19	0	0	2	0
	912,398	913,252	284	-	rplB	CTLon_0782	50S ribosomal protein L2	0	0	0	0
	913,276	913,611	111	-	rplW	CTLon_0783	50S ribosomal protein L23	0	0	2	0
	913,627	914,295	222	-	rplD	CTLon_0784	50S ribosomal protein L4	0	1	3	0
915,286	914,304	914,969	221	-	rplC	CTLon_0785	50S ribosomal protein L3	0	0	7	1
916,343	915,404	916,300	298	-	CTLon_0786	CTLon_0786	hypothetical protein	4	0	6	2
	916,466	917,416	316	-	fmt	CTLon_0787	methionyl-tRNA formyltransferase	0	0	5	0
	917,406	918,248	280	-	lpxA	CTLon_0788	UDP-N-acetylglucosamine acyltransferase	0	0	4	0
	918,260	918,721	153	-	fabZ	CTLon_0789	(3R)-hydroxymyristoyl-ACP dehydratase	2	0	1	0
919,597	918,718	919,578	286	-	lpxC	CTLon_0790	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	0	4	4	4
	919,677	921,305	542	-	cutE	CTLon_0791	apolipoprotein N-acyltransferase	2	0	1	2
	921,369	921,851	160	-	CTLon_0792	CTLon_0792	acyl-CoA hydrolase	0	0	5	0
	921,987	922,739	250	-	dnaQ2	CTLon_0793	DNA polymerase III subunit epsilon	0	0	4	0
	922,743	923,216	157	-	CTLon_0794	CTLon_0794	putative nucleotide-binding protein	2	0	1	0
924,104	923,195	923,911	238	-	CTLon_0795	CTLon_0795	hypothetical protein	0	1	3	6
924,376	924,376	924,459		+	tRNA-Leu	CTLon_0108w	tRNA	25	160	12	33
	924,633	924,941	102	+	trxA	CTLon_0796	thioredoxin	0	0	0	0
925,473	924,991	925,446	151	-	CTLon_0797	CTLon_0797	SpoU family rRNA methylase	0	1	0	2

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	925,462	926,193	243	-	mip	CTLon_0798	peptidyl-prolyl cis-trans isomerase	0	0	7	1
	926,330	928,078	582	-	aspS	CTLon_0799	aspartyl-tRNA synthetase	0	0	1	0
929,426	928,059	929,345	428	-	hisS	CTLon_0800	histidyl-tRNA synthetase	2	2	2	1
929,618	929,781	931,151	456	+	uhpC	CTLon_0801	putative sugar phosphate permease	4	0	17	23
	931,165	934,878	1237	+	dnaE	CTLon_0802	DNA polymerase III subunit alpha	1	0	11	2
935,990	935,107	935,976	289	-	CTLon_0803	CTLon_0803	hypothetical protein	7	0	4	4
936,100	936,128	937,084	318	+	CTLon_0804	CTLon_0804	hypothetical protein	0	0	8	0
	937,109	937,693	194	+	CTLon_0805	CTLon_0805	hypothetical protein	0	0	0	2
937,677	937,680	938,120	146	+	rsbW	CTLon_0806	sigma regulatory factor-histidine kinase	0	0	0	0
938,565	938,127	938,552	141	-	CTLon_0807	CTLon_0807	hypothetical protein	4	2	4	5
	938,660	939,973	437	+	dacC	CTLon_0808	D-alanyl-D-alanine carboxypeptidase	0	0	3	0
	940,106	940,513	135	+	CTLon_0809	CTLon_0809	hypothetical protein	0	0	0	0
	940,510	941,616	368	+	CTLon_0810	CTLon_0810	putative methyltransferase	0	0	0	0
	941,766	943,001	411	+	brnQ	CTLon_0811	branched-chain amino acid transport system carrier protein	2	0	0	1
	943,175	946,774	1199	+	CTLon_0812	CTLon_0812	helicase	0	0	3	2
946,853	946,953	947,432	159	+	CTLon_0813	CTLon_0813	hypothetical protein	0	0	8	5
947,513	947,641	949,038	465	+	lpdA	CTLon_0814	dihydrolipoamide dehydrogenase	0	0	3	1
	949,035	949,970	311	+	lipA	CTLon_0815	lipoyl synthase	0	0	1	0
950,056	950,074	951,051	325	+	sctJ	CTLon_0816	type III secretion system protein, membrane component	1	0	36	8
	951,052	951,888	278	+	CTLon_0817	CTLon_0817	hypothetical protein	0	0	5	0
	952,164	952,835	223	+	sctL	CTLon_0818	type III secretion system protein	0	0	8	0
	952,848	953,768	306	+	sctR	CTLon_0819	type III secretion system protein	0	0	5	0
	953,780	954,064	94	+	sctS	CTLon_0820	type III secretion system, membrane protein	0	0	0	0
	954,071	954,940	289	+	sctT	CTLon_0821	type III secretion system, membrane protein	1	0	5	1
955,503	955,019	955,462	147	-	CTLon_0822	CTLon_0822	hypothetical protein	55	10	13	10
	955,552	956,544	330	-	CTLon_0823	CTLon_0823	hypothetical protein	1	0	0	0
	956,550	957,074	174	-	CTLon_0824	CTLon_0824	hypothetical protein	0	0	0	0
	957,059	957,514	151	-	CTLon_0825	CTLon_0825	hypothetical protein	0	0	0	0
	957,498	957,827	109	-	gspG	CTLon_0826	general secretion pathway protein G	0	0	0	0
	957,890	959,065	391	-	gspF	CTLon_0827	general secretion pathway protein F	0	0	1	0
	959,077	960,582	501	-	gspE	CTLon_0828	general secretion pathway protein E	0	0	3	1
	960,566	962,848	760	-	gspD	CTLon_0829	general secretion pathway protein D	2	1	3	1
964,101	962,849	964,078	409	-	CTLon_0830	CTLon_0830	hypothetical protein	0	0	0	0
	964,207	965,277	356	-	pepP	CTLon_0831	proline dipeptidase	0	0	2	0
	965,288	967,018	576	-	mutL	CTLon_0832	DNA mismatch repair protein	4	0	3	0
967,258	967,298	967,996	232	+	scc2	CTLon_0833	type III secretion chaperone (low calcium response protein H)	106	11	8	0
	968,013	968,372	119	+	CTLon_0834	CTLon_0834	hypothetical protein	55	3	6	0
	968,409	969,872	487	+	copB	CTLon_0835	putative type III secretion system membrane protein	88	27	15	1

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	969,903	971,222	439	+	copD	CTLon_0836	putative type III secretion system protein	120	10	10	0
	971,300	972,283	327	-	CTLon_0837	CTLon_0837	putative integral membrane protein	0	0	1	1
	972,588	974,495	635	+	thrS	CTLon_0838	threonyl-tRNA synthetase	11	0	2	1
	974,947	975,714	255	+	minD	CTLon_0839	ParA family chromosome partitioning ATPase	0	0	0	0
	975,719	976,510	263	+	gp6D	CTLon_0840	virulence plasmid protein pGP6-D-related protein	0	0	6	1
976,449	976,476	977,027	183	+	CTLon_0841	CTLon_0841	hypothetical protein	0	0	3	0
	977,206	978,246	346	+	trpS	CTLon_0842	tryptophanyl-tRNA synthetase	0	0	2	0
978,253	978,271	980,277	668	+	uvrB	CTLon_0843	excinuclease ABC subunit B	0	0	3	0
980,320	980,439	981,713	424	+	eno	CTLon_0844	phosphopyruvate hydratase	0	0	1	1
981,811	981,840	983,792	650	+	rbsU	CTLon_0845	sigma regulatory family protein-PP2C phosphatase	0	0	3	0
	983,917	985,725	602	+	CTLon_0846	CTLon_0846	hypothetical protein	0	0	0	0
	985,740	988,604	954	-	CTLon_0847	CTLon_0847	hypothetical protein	0	0	11	0
	988,701	989,399	232	-	sdhB	CTLon_0848	succinate dehydrogenase iron-sulfur subunit	0	0	0	0
	989,478	991,358	626	-	sdhA	CTLon_0849	succinate dehydrogenase flavoprotein subunit	0	0	1	0
	991,355	992,216	286.333	-	CTLon_0850	CTLon_0850	pseudo	0	0	0	1
	992,363	993,154	263	-	CTLon_0851	CTLon_0851	putative hydrolase	0	0	1	0
	993,239	995,317	692	-	dsdD	CTLon_0852	thiol:disulfide interchange protein	1	0	3	2
995,442	995,470	996,168	232	+	exbB	CTLon_0853	macromolecule transporter	0	0	1	2
	996,165	996,572	135	+	CTLon_0854	CTLon_0854	macromolecule transport protein	0	0	0	0
	996,575	997,282	235	+	CTLon_0855	CTLon_0855	histone H1-I	1	0	0	0
	997,282	998,577	431	+	tolB	CTLon_0856	translocation protein TolB	0	0	4	2
998,517	998,574	999,140	188	+	pal	CTLon_0857	peptidoglycan-associated lipoprotein	0	0	5	0
	999,130	999,732	200	+	CTLon_0858	CTLon_0858	putative soluble transglycosylase	0	0	6	0
999,781	999,803	1,000,195	130	+	CTLon_0859	CTLon_0859	hypothetical protein	0	0	13	10
	1,000,192	1,000,779	195	-	ahpC	CTLon_0860	thioredoxin peroxidase	0	0	0	0
1,000,890	1,000,804	1,000,876		-	tRNA-Arg	CTLon_0108l	tRNA	59	343	47	106
	1,000,987	1,002,588	533	-	groEL2	CTLon_0861	60 kDa chaperonin GroEL2	0	0	7	0
	1,002,668	1,003,894	408	-	CTLon_0862	CTLon_0862	hypothetical protein	0	0	1	1
1,004,039	1,004,091	1,004,720	209	+	CTLon_0863	CTLon_0863	putative deoxyribonucleotide triphosphate pyrophosphatase	0	0	0	1
	1,004,694	1,004,933	79	-	CTLon_0864	CTLon_0864	hypothetical protein	0	0	0	0
	1,004,918	1,005,607	229	-	ung	CTLon_0865	uracil-DNA glycosylase	0	0	0	0
1,005,599	1,005,688	1,007,592	634	+	uvrD	CTLon_0866	DNA helicase	0	0	6	10
1,007,467	1,007,596	1,008,906	436	+	rpoN	CTLon_0867	RNA polymerase factor sigma-54	0	0	4	0
	1,009,014	1,009,709	231	-	CTLon_0868	CTLon_0868	hypothetical protein	0	0	4	0
	1,009,706	1,010,437	243	-	CTLon_0869	CTLon_0869	hypothetical protein	0	0	5	1
	1,010,409	1,010,888	159	-	folA	CTLon_0870	dihydrofolate reductase	0	0	0	0
	1,010,885	1,012,237	450	-	folP	CTLon_0871	2-amino-4-hydroxy-6-hydroxymethyldihydropteridi ne pyrophosphokinase	0	0	0	0

TSS	Location		Length	Strand	Gene	Synonym	Product	EB -	EB +	RB -	RB+
	1,012,234	1,012,608	124	-	folX	CTLon_0872	dihydroneopterin aldolase	0	0	1	1
	1,012,627	1,014,342	571	-	rpoD	CTLon_0873	RNA polymerase sigma factor	2	1	1	0
	1,014,491	1,015,780	429	-	CTLon_0874	CTLon_0874	putative integral membrane protein	4	0	2	0
1,015,996	1,016,229	1,016,525	98	+	rpsT	CTLon_0875	30S ribosomal protein S20	0	1	1	1
1,016,671	1,016,758	1,017,558	266	+	CTLon_0876	CTLon_0876	hypothetical protein	5	0	33	6
1,020,266	1,017,615	1,020,242	875	-	CTLon_0877	CTLon_0877	hypothetical protein	8	3	8	0
	1,020,357	1,022,873	838	+	CTLon_0878	CTLon_0878	hypothetical protein	0	0	10	0
1,025,466	1,022,937	1,025,435	832	-	CTLon_0879	CTLon_0879	hypothetical protein	10	0	24	5
1,027,643	1,025,663	1,027,618	651	-	CTLon_0880	CTLon_0880	hypothetical protein	134	11	12	2
1,029,182	1,027,725	1,029,026	433	-	CTLon_0881	CTLon_0881	hypothetical protein	2	0	2	3
	1,029,270	1,030,853	527	-	mviN	CTLon_0882	putative virulence protein	0	0	0	0
1,031,027	1,031,053	1,031,919	288	+	nfo	CTLon_0883	endonuclease IV	0	0	1	2
	1,032,015	1,032,644	209	-	rpsD	CTLon_0884	30S ribosomal protein S4	1	0	4	2
1,033,014	1,033,028	1,034,011	327	+	CTLon_0885	CTLon_0885	hypothetical protein	0	0	1	0
	1,034,084	1,034,959	291	-	ispA	CTLon_0886	dimethylallyltransferase	0	0	1	1
	1,035,015	1,035,632	205	-	glmU	CTLon_0887	glucosamine-1-phosphate acetyltransferase	0	0	4	0
1,036,391	1,035,685	1,036,368	227	-	CTLon_0888	CTLon_0888	transcriptional regulatory protein	0	0	11	2
1,036,490	1,036,548	1,036,802	84	+	CTLon_0889	CTLon_0889	hypothetical protein	2	1	0	1
1,036,902	1,036,902	1,036,975		+	tRNA-Pro	CTLon_0108ag	tRNA	28	149	22	43
1,038,598	1,036,983	1,038,572	529	-	CTLon_0890	CTLon_0890	hypothetical protein	2	0	6	1

9.1.2 TSS of *C. pneumoniae*

Table 11: Transcription start sites and sequencing read numbers of *C. pneumoniae* genes.
The exact transcription start sites of the annotated genes including novel non-coding RNAs are listed. For some genes a secondary transcription starts listed which can be the consequence of alternative transcription start sites. Numbers in TSS and location refer to the first nucleotide of the transcript or the annotated gene coordinates, respectively. Genes without TSS are either too low expressed or transcribed as part of a polycistronic transcript. The table gives the following information: First column: indicates the relative location of the TSS to annotated genes (S= sense, AS= antisense, IG= intergenic region, LL= leaderless transcript); TSS: transcription start site; 2nd TSS: secondary TSS located downstream of primary TSS or inside an operon or ORF; Length: length of the corresponding protein in amino acids; Location: start and end coordinates of gene; Strand: sense (+) or antisense (-) strand; Gene: gene name or gene number if no name is available; Synonym: gene number according to NCBI Reference Sequence NC_010280.1; Product: name of the gene product; EB-, EB+, RB-, RB+: sequence read numbers of the four cDNA libraries derived from either EB or RB total RNA (-) or total RNA enriched for primary transcripts (+). Novel transcripts are coloured in red. Background colour depicts genes that are co-transcribed as operon. Adjacent genes that are transcribed as polycistronic transcript are coloured in green or orange. Genes that are adjacent and co-transcribed but interrupted in the table by genes on the antisense strand are coloured in violet.

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	322	309	1,230,195	1	282	-	93	CPn0001	CPn0001	hypothetical protein	60	19	534	123
				573	878	+	101	gatC	CPn0002	aspartyl/glutamyl-tRNA amidotransferase subunit C	0	0	4	0
				895	2,373	+	492	gatA	CPn0003	aspartyl/glutamyl-tRNA amidotransferase subunit A	14	2	34	1
			3,906	2,370	3,836	+	488	gatB	CPn0004	aspartyl/glutamyl-tRNA amidotransferase subunit B	11	1	29	2
	4,099		7,060	4,127	6,895	+	922	pmp_1	CPn0005	polymorphic Outer membrane protein G family	37	7	136	24
				7,138	7,293	-	51	CPn0006	CPn0006	hypothetical protein	0	0	0	0
		9,166		7,605	10,499	+	964	CPn0007	CPn0007	hypothetical protein	13	4	15	4
				10,975	11,688	+	237	CPn0008	CPn0008	hypothetical protein	2	3	0	1
				11,815	13,122	+	435	CPn0009	CPn0009	hypothetical protein	4	0	3	0
	13,281	13,446		13,435	14,328	+	297	CPn0010	CPn0010	hypothetical protein	8	3	9	3
			15,764	14,379	15,749	+	456	CPn0010.1	CPn0010.1	hypothetical protein	3	1	9	0
	15,855			15,892	16,617	+	241	CPn0011	CPn0011	hypothetical protein	5	2	18	1
			18,302	16,644	18,215	+	523	CPn0012	CPn0012	hypothetical protein	15	4	43	3
	18,557		21,340	18,584	21,109	+	841	pmp_2	CPn0013	polymorphic Outer membrane protein G family	29	6	147	30
	21,361			21,392	21,924	+		pmp_3.1	CPn0014	polymorphic Outer membrane protein	59	7	315	73
S		22,791	24,277	21,835	24,177	+		pmp_3.2	CPn0015	polymorphic Outer membrane protein	59	7	235	33
	24,392			24,416	26,191	+		pmp_4.1	CPn0016	polymorphic Outer membrane protein	24	20	82	44
			27,246	26,094	27,173	+		pmp_4.2	CPn0017	polymorphic Outer membrane protein	17	1	12	1
S	27,289	28,592		27,522	29,006	+		pmp_5.1	CPn0018	polymorphic Outer membrane protein	7	4	49	10
AS	28,995		28,435			-		pCpn01	pCpn01	putative sRNA, antisense to pmp_5.1	24	5	36	15
				29,007	30,359	+		pmp_5.2	CPn0019	polymorphic Outer membrane protein	2	0	5	0
			30,569	30,600	32,687	-	695	pmp	CPn0020	putative outer membrane protein	2	0	24	1
				32,704	34,410	-	568	pmp	CPn0021	putative outer membrane protein	19	0	36	4
				34,392	34,982	-	196	maf	CPn0022	Maf-like protein	11	0	40	5

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
LL	36,603			35,011	36,603	-	530	yjjK	CPn0023	ABC transporter ATPase	34	6	86	26
				36,658	37,596	-	312	xerC	CPn0024	site-specific tyrosine recombinase XerC	2	0	9	1
				37,681	38,604	-	307	atsA	CPn0025	ribonuclease Z	3	0	12	0
				38,759	39,625	-	288	CPn0026	CPn0026	hypothetical protein	5	0	6	3
	42,397			39,775	42,234	-	819	lon	CPn0027	Lon ATP-dependent protease	26	6	122	9
	43,345			42,540	43,325	-	261	CPn0028	CPn0028	hypothetical protein	6	1	4	1
				43,387	43,755	-	122	CPn0029	CPn0029	hypothetical protein	0	0	2	0
	43,869			43,891	44,532	+	213	gcp_1	CPn0030	O-sialoglycoprotein endopeptidase	8	1	5	0
	44,623			44,711	44,887	+	58	rpsU	CPn0031	30S ribosomal protein S21	11	8	55	6
				44,923	46,101	+	392	dnaJ	CPn0032	chaperone protein DnaJ	15	0	23	2
			48,264	46,138	48,174	+	678	pdhA/pdhB	CPn0033	(pyruvate) oxoisovalerate dehydrogenase Alpha & Beta fusion	16	2	52	7
	49,482			48,207	49,457	-	416	CPn0034	CPn0034	hypothetical protein	2	0	4	2
	51,209			49,566	51,029	-	487	CPn0035	CPn0035	hypothetical protein	12	3	36	2
AS	49,662		50,571			+		pCpn02	pCpn02	putative sRNA, antisense to CPn0035	21	6	38	5
AS	50,572			51,002	51,799	+	265	CPn0036	CPn0036	hypothetical protein	8	2	8	1
				51,792	52,118	+	108	ptsH	CPn0037	PTS phosphocarrier protein Hpr	6	1	4	2
			53,981	52,119	53,834	+	571	ptsI	CPn0038	PTS PEP phosphotransferase	8	4	21	1
			53,866	53,960	54,250	-	96	ybaB	CPn0039	hypothetical protein	4	0	3	0
	55,670			54,315	55,643	-	442	dnaX_1	CPn0040	DNA polymerase III subunits gamma and tau	8	2	55	3
AS	54,775		55,003			+		pCpn03	pCpn03	putative sRNA, antisense to CPn0040	4	1	7	3
				55,996	57,345	+	449	CPn0041	CPn0041	hypothetical protein	4	0	2	0
				57,403	58,185	+	260	CPn0042	CPn0042	hypothetical protein	4	0	6	0
	58,205			58,447	60,375	+	642	CPn0043	CPn0043	hypothetical protein	2	0	2	0
AS	59,034		58,846			-		pCpn04	pCpn04	putative sRNA, antisense to CPn0043	3	8	13	16
				60,419	60,781	+	120	CPn0044	CPn0044	hypothetical protein	2	1	11	2
				61,069	62,793	+	574	CPn0045	CPn0045	hypothetical protein	0	2	11	0
				62,790	63,266	+	158	CPn0046	CPn0046	hypothetical protein	1	0	3	0
				63,455	63,655	+	66	CPn0047	CPn0047	hypothetical protein	0	0	14	0
				63,687	65,804	+	705	yqfF	CPn0048	inner membrane protein	15	2	16	1
				65,814	66,296	-	160	CPn0049	CPn0049	hypothetical protein	8	2	12	2
	67,031			66,496	66,813	-	105	CPn0050	CPn0050	hypothetical protein	52	17	101	27
				66,833	67,114	+	93	CPn0051	CPn0051	hypothetical protein	0	0	1	0
				67,301	68,005	-	234	hemC	CPn0052	porphobilinogen deaminase	1	0	2	0
				67,983	69,344	-	453	sms	CPn0053	DNA repair protein RadA	5	0	10	0
				69,310	70,023	-	237	rnc	CPn0054	ribonuclease III	0	0	2	0
	70,107			70,129	70,593	+	154	CPn0055	CPn0055	hypothetical protein	8	1	13	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
AS	70,195		70,138			-		pCpn05	pCpn05	putative sRNA, antisense to CPn0055	2	0	4	1
				70,953	72,749	+	598	mrsA	CPn0056	phosphomannomutase	6	2	36	1
	72,799			72,934	73,557	+	207	sodM	CPn0057	superoxide dismutase (Mn)	7	0	65	12
	73,613			73,639	74,565	+	308	accD	CPn0058	acetyl-CoA carboxylase subunit beta	38	7	71	4
				74,616	75,053	+	145	dut	CPn0059	deoxyuridine 5'-triphosphate nucleotidohydrolase	15	1	55	1
				75,055	75,531	+	158	ptsN_1	CPn0060	PTS IIA protein	4	0	18	1
			76,269	75,534	76,211	+	225	ptsN_2	CPn0061	PTS IIA protein + HTH DNA-binding domain	7	0	15	2
	76,282		77,783	76,308	77,693	+	461	CPn0062	CPn0062	hypothetical protein	49	24	347	102
AS	78,886		77,712			-		pCpn06	pCpn06	putative sRNA, antisense to CPn0058	49	27	23	8
				78,112	78,270	+	52	CPn0063	CPn0063	hypothetical protein	0	0	0	0
LL	78,295	78,346		78,346	78,579	+	77	CPn0064	CPn0064	hypothetical protein	0	0	0	0
	78,900		80,739	78,924	80,654	+	576	CPn0065	CPn0065	hypothetical protein	411	158	269	65
	80,891			80,925	82,658	+	577	CPn0066	CPn0066	hypothetical protein	65	35	36	24
				82,953	84,056	+	367	CPn0067	CPn0067	hypothetical protein	19	6	22	4
				84,328	84,903	-	191	CPn0068	CPn0068	hypothetical protein	4	0	5	1
	84,960		87,101	85,236	87,089	+	617	CPn0069	CPn0069	hypothetical protein	28	13	20	6
			87,167	87,205	87,378	-	57	CPn0070	CPn0070	hypothetical protein	4	1	6	0
	87,532		87,449	87,450	87,522	-	73	Met tRNA	CPnt01	Met tRNA	56	122	103	383
				87,596	88,045	-	149	CPn0071	CPn0071	hypothetical protein	3	0	11	6
				88,054	89,061	-	335	CPn0072	CPn0072	hypothetical protein	4	2	18	3
	89,239			89,356	89,577	+	73	infA	CPn0073	translation initiation factor IF-1	12	11	30	43
	89,625		89,728	89,657	89,728	+	72	Thr tRNA	CPnt02	Thr tRNA	179	473	224	692
				89,774	90,958	+	394	tuf	CPn0074	elongation factor Tu	85	6	91	11
	90,958		91,070	90,998	91,070	+	73	Trp tRNA	CPnt03	Trp tRNA	420	586	493	948
				91,102	91,353	+	83	secE	CPn0075	preprotein translocase subunit SecE	12	0	43	0
				91,358	91,906	+	182	nusG	CPn0076	transcription antitermination protein NusG	9	2	49	2
	91,956			92,013	92,438	+	141	rplK	CPn0077	50S ribosomal protein L11	17	1	38	4
				92,465	93,163	+	232	rplA	CPn0078	50S ribosomal protein L1	43	2	40	2
				93,179	93,691	+	170	rplJ	CPn0079	50S ribosomal protein L10	18	2	30	1
			94,164	93,735	94,124	+	129	rplL	CPn0080	50S ribosomal protein L7/L12	59	0	75	2
				94,261	98,019	+	1252	rpoB	CPn0081	DNA-directed RNA polymerase subunit beta	30	0	103	18
				98,043	102,224	+	1393	rpoC	CPn0082	DNA-directed RNA polymerase subunit beta'	49	4	127	20
AS	94,794		94,609			-		pCpn07	pCpn07	putative sRNA, antisense to CPn0081	4	3	12	3
	102,305			102,332	103,315	+	327	tal	CPn0083	transaldolase B	34	10	183	34
			103,894	103,362	103,754	+	130	CPn0084	CPn0084	ferredoxin	13	3	83	7
AS	102,542		102,308			-		pCpn08	pCpn08	putative sRNA, antisense to CPn0083	9	3	16	1
	104,571		103,701	103,763	104,506	-	247	CPn0085	CPn0085	hypothetical protein	1	1	16	9

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
IG	104,806		104,677			-		pCpn09	pCpn09	putative sRNA, intergenic between CPn0085 and CPn0086	0	0	9	1
	104,836			104,904	105,530	+	208	atpE	CPn0086	V-type ATP synthase subunit E	28	0	92	3
				105,579	106,379	+	266	CPn0087	CPn0087	hypothetical protein	7	1	21	1
				106,373	108,148	+	591	atpA	CPn0088	V-type ATP synthase subunit A	10	0	14	2
				108,153	109,469	+	438	atpB	CPn0089	V-type ATP synthase subunit B	2	1	7	0
				109,454	110,083	+	209	atpD	CPn0090	V-type ATP synthase subunit D	6	1	11	1
				110,074	112,056	+	660	atpI	CPn0091	V-type ATP synthase subunit I	3	1	29	4
				112,151	112,576	+	141	atpK	CPn0092	V-type ATP synthase subunit K	4	1	7	2
		112,585		112,509	113,018	+	169	CPn0093	CPn0093	hypothetical protein	3	2	11	8
				113,152	115,974	+	940	valS	CPn0094	valyl-tRNA synthetase	14	1	22	5
				115,995	118,793	+	932	pknD	CPn0095	serine/threonine-protein kinase	5	2	26	2
				118,834	124,314	-	1826	uvrA	CPn0096	excinuclease ABC subunit A	17	6	29	5
	124,514			124,555	126,009	+	484	pyk	CPn0097	pyruvate kinase	9	5	40	15
				126,088	127,491	-	467	htrB	CPn0098	lipid A biosynthesis lauroyl acyltransferase	3	0	16	1
				127,593	127,868	+	91	CPn0099	CPn0099	hypothetical protein	0	0	0	0
				127,879	129,141	-	420	CPn0100	CPn0100	hypothetical protein	8	0	37	2
	130,012			129,138	129,932	-	264	ybbP	CPn0101	hypothetical protein	11	1	33	11
	129,929			130,123	131,469	+	448	cydA	CPn0102	cytochrome oxidase subunit I	10	0	52	3
				131,480	132,514	+	344	cydB	CPn0103	cytochrome oxidase subunit II	12	2	30	20
AS	132,469		132,027			-		pCpn10	pCpn10	putative sRNA, antisense to CPn0083	8	0	13	5
S	132,303		132,814			+		pCpn11	pCpn11	putative novel transcript, overlapping with CPn0103	5	0	12	0
	133,953		132,500	132,673	133,875	-	400	CPn0104	CPn0104	hypothetical protein	40	2	214	6
	134,889			134,026	134,847	-	273	CPn0105	CPn0105	hypothetical protein	36	2	116	14
				135,091	136,377	+	428	phoH	CPn0106	ATPase	8	0	16	1
				136,389	137,162	-	257	CPn0107	CPn0107	hypothetical protein	9	2	6	2
	137,887			137,300	137,857	-	185	CPn0108	CPn0108	hypothetical protein	40	2	22	4
	138,511		138,071	138,073	138,493	-		tmRNA	Cpns01	tmRNA	1655	733	973	899
	138,622		141,874	138,655	141,786	+	1043	ileS	CPn0109	isoleucyl-tRNA synthetase	21	8	34	8
	143,867		141,671	141,824	143,734	-	636	lepB	CPn0110	Signal peptidase I	21	3	83	10
	144,728		143,924	143,931	144,686	-	251	CPn0111	CPn0111	hypothetical protein	2	7	30	9
	144,711		145,163	144,767	145,096	+	109	rpmE2	CPn0112	50S ribosomal protein L31 type B	38	5	25	5
				145,335	146,408	+	357	prfA	CPn0113	peptide chain release factor 1	10	3	23	2
				146,398	147,264	+	288	hemK	CPn0114	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	3	1	9	2
	147,255			147,279	148,625	+	448	ffh	CPn0115	Signal recognition particle GTPase	15	1	34	3
				148,616	148,975	+	119	rpsP	CPn0116	30S ribosomal protein S16	5	0	15	3

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				148,989	150,074	+	361	trmD	CPn0117	tRNA (guanine-N(1)-)-methyltransferase/unknown domain fusion protein	9	1	23	10
				150,102	150,467	+	121	rplS	CPn0118	50S ribosomal protein L19	3	1	22	11
				150,523	151,167	+	214	rnhB	CPn0119	ribonuclease HII	4	1	10	1
S	151,170			151,164	151,781	+	205	gmk	CPn0120	guanylate kinase	4	1	29	7
				151,778	152,071	+	97	CPn0121	CPn0121	hypothetical protein	3	0	9	1
				152,071	153,726	+	551	metG	CPn0122	methionyl-tRNA synthetase	15	7	44	5
	156,258		153,630	153,771	155,969	-	732	recD_1	CPn0123	exodeoxyribonuclease V (Alpha subunit)	16	4	56	8
S	157,334			156,614	158,071	+	485	CPn0124	CPn0124	hypothetical protein	2	2	2	8
				158,096	158,608	+	170	CPn0125	CPn0125	hypothetical protein	1	1	13	0
				158,809	161,088	+	759	CPn0126	CPn0126	hypothetical protein	7	1	25	1
	162,175		161,040	161,127	162,143	-	338	ytfF	CPn0127	cationic amino acid transporter	21	18	36	8
				162,277	163,056	+	259	CPn0128	CPn0128	biotin protein ligase	1	0	1	0
				163,061	163,717	-	218	CPn0129	CPn0129	hypothetical protein	10	0	11	2
				163,748	164,245	-	165	CPn0130	CPn0130	hypothetical protein	9	0	6	2
	164,402		164,318	164,318	164,389	-	72	Asn tRNA	CPnt04	Asn tRNA	1632	6724	1800	9102
				164,549	165,583	+	344	CPn0131	CPn0131	hypothetical protein	9	1	15	1
				165,587	166,564	+	325	CPn0132	CPn0132	hypothetical protein	7	0	14	1
	167,359			166,561	167,334	-	257	CPn0133	CPn0133	hypothetical protein	1	0	8	3
			167,397	167,464	169,098	-	544	groEL	CPn0134	chaperonin GroEL	320	18	345	15
				169,140	169,448	-	102	groES	CPn0135	co-chaperonin GroES	68	11	87	4
	171,442			169,566	171,401	-	611	pepF	CPn0136	oligopeptidase	25	17	169	6
				171,499	172,254	-	251	ybgI	CPn0137	Acr family transporter	5	1	17	0
				172,697	174,019	-	440	hemL	CPn0138	glutamate-1-semialdehyde aminotransferase	6	0	12	3
				174,090	174,656	-	188	yqgE	CPn0139	hypothetical protein	3	0	11	0
				174,670	175,110	-	146	yqdE	CPn0140	hypothetical protein	5	0	22	0
				175,107	175,802	-	231	rpiA	CPn0141	ribose-5-phosphate isomerase A	1	0	6	1
				175,813	176,091	-	92	CPn0142	CPn0142	hypothetical protein	0	0	0	3
IG	175,970		176,208			+		pCpn12	pCpn12	putative sRNA, intergenic and antisense to terminator of CPn0083	2	2	16	4
			176,165	176,211	177,335	-	374	yxjG_1	CPn0143	hypothetical protein	3	0	9	1
AS	177,206			177,963	180,563	+	866	clpB	CPn0144	Clp protease ATPase	21	2	52	6
				180,777	182,372	+	531	CPn0145	CPn0145	hypothetical protein	11	3	30	1
	182,429			182,613	183,098	+	161	CPn0146	CPn0146	hypothetical protein	5	7	14	1
				183,225	183,674	+	149	CPn0147	CPn0147	hypothetical protein	21	0	41	3
	183,810			183,846	185,705	+	619	CPn0148	CPn0148	S/T protein kinase	42	8	96	20
AS	184,478		183,992			+		pCpn13	pCpn13	putative sRNA, antisense to CPn0148	24	8	82	15
				185,715	187,703	+	662	ligA	CPn0149	NAD-dependent DNA ligase LigA	8	2	14	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				187,834	192,447	+	1537	CPn0150	CPn0150	hypothetical protein	85	29	23	2
	194,171		192,485	192,622	194,142	-	506	mhpA	CPn0151	monooxygenase	27	5	40	2
	195,291			194,315	195,265	-	316	CPn0152	CPn0152	hypothetical protein	23	3	8	2
AS	195,128		195,218			+		pCpn14	pCpn14	putative sRNA, antisense to CPn0152	9	8	19	11
				195,433	197,895	+	820	leuS	CPn0153	leucyl-tRNA synthetase	17	2	16	2
				197,892	199,205	+	437	gseA	CPn0154	3-deoxy-D-manno-octulosonic-acid transferase	5	1	13	2
	199,225			199,229	199,301	+	73	Met tRNA	CPnt05	Met tRNA	4762	30570	6208	32521
			199,390	199,317	199,390	+	74	Met tRNA	CPnt06	Met tRNA	606	1588	590	1534
				199,485	199,691	-	68	CPn0155	CPn0155	hypothetical protein	7	1	4	0
				199,767	200,117	-	116	CPn0156	CPn0156	hypothetical protein	0	0	1	0
				200,295	200,723	-	142	CPn0157	CPn0157	hypothetical protein	6	0	4	0
				200,891	201,430	-	179	CPn0158	CPn0158	hypothetical protein	1	0	1	0
				201,464	201,772	-	102	CPn0159	CPn0159	hypothetical protein	1	1	2	0
				202,124	203,791	-	555	pfkA_1	CPn0160	diphosphate--fructose-6-phosphate 1-phosphotransferase	11	2	36	2
	204,653			203,795	204,622	-	275	CPn0161	CPn0161	acetyltransferase	3	1	28	10
				204,800	205,828	-	342	CPn0162	CPn0162	hypothetical protein	2	0	9	1
	205,864	206,244		206,026	206,397	+	123	CPn0163	CPn0163	hypothetical protein	14	5	6	3
	206,478			206,498	207,001	+	167	CPn0164	CPn0164	hypothetical protein	64	24	257	47
	206,990			206,998	207,585	+	195	CPn0165	CPn0165	hypothetical protein	199	24	132	11
				207,630	207,965	+	111	CPn0166	CPn0166	hypothetical protein	26	2	9	1
				207,974	208,306	-	110	CPn0167	CPn0167	hypothetical protein	19	1	2	2
				208,414	208,641	-	75	CPn0168	CPn0168	hypothetical protein	10	1	1	0
				208,707	209,501	-	264	CPn0169	CPn0169	hypothetical protein	16	0	7	1
				210,022	211,026	-	334	CPn0170	CPn0170	hypothetical protein	11	2	10	1
				211,146	212,435	-	429	guaA	CPn0171	GMP synthase	16	0	1	2
				212,437	213,177	-	246	guaB_1	CPn0172	inosine 5'-monophosphase dehydrogenase	7	0	4	1
	213,523			213,238	213,501	-	87	guaB_2	CPn0172a	inosine 5'-monophosphase dehydrogenase	2	2	16	3
				213,712	213,987	-	91	CPn0173	CPn0173	hypothetical protein	2	0	3	0
				214,257	214,727	+	156	CPn0174	CPn0174	hypothetical protein	6	0	13	0
				214,898	215,278	+	126	CPn0175	CPn0175	hypothetical protein	5	1	5	1
				215,286	216,521	+	411	CPn0176	CPn0176	hypothetical protein	8	1	9	2
	217,609		216,526	216,605	217,459	-	284	CPn0177	CPn0177	hypothetical protein	55	6	84	8
				217,786	218,052	-	88	CPn0178	CPn0178	hypothetical protein	2	1	1	0
				218,053	218,403	-	116	CPn0179	CPn0179	hypothetical protein	4	1	1	0
				218,352	218,906	-	184	CPn0180	CPn0180	hypothetical protein	5	2	2	0
				218,774	219,175	-	133	CPn0181	CPn0181	hypothetical protein	6	0	5	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				219,331	220,695	-	454	accC	CPn0182	acetyl-CoA carboxylase biotin carboxylase subunit	31	4	91	6
				220,692	221,195	-	167	accB	CPn0183	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	4	0	16	4
				221,218	221,775	-	185	efp_1	CPn0184	elongation factor P	12	3	33	7
	222,621			221,762	222,451	-	229	rpe	CPn0185	ribulose-phosphate 3-epimerase	21	0	38	11
	222,871			222,899	224,071	+	390	IncA	CPn0186	IncA	16	2	23	4
	224,206			224,248	225,048	+	266	CPn0187	CPn0187	methylase	8	2	32	0
				225,111	226,403	+	430	CPn0188	CPn0188	hypothetical protein	17	3	29	6
				226,400	229,828	+	1142	CPn0189	CPn0189	hypothetical protein	12	4	50	1
				229,919	231,277	+	452	CPn0190	CPn0190	hypothetical protein	5	1	3	1
				231,311	231,991	-	226	glnQ	CPn0191	amino acid ABC transporter ATPase	4	1	0	0
				231,981	232,634	-	217	glnP	CPn0192	amino acid ABC transporter ATPase	0	1	4	0
	233,359			232,683	233,126	-	147	argR	CPn0193	arginine repressor	3	6	10	1
	233,186			233,210	234,244	+	344	gcp_2	CPn0194	putative DNA-binding/iron metalloprotein/AP endonuclease	11	6	18	7
S	234,004			234,190	235,788	+	532	oppA_1	CPn0195	oligopeptide binding protein	16	0	76	7
				235,939	237,522	+	527	oppA_2	CPn0196	oligopeptide binding protein	5	1	7	0
AS	237,544		237,186			-		pCpn15	pCpn15	putative sRNA, antisense to CPn0196	8	5	11	2
				237,578	238,885	+	435	oppA_3	CPn0197	oligopeptide binding protein	17	1	6	2
	238,989			239,169	240,749	+	526	oppA_4	CPn0198	oligopeptide binding protein	7	3	44	4
	240,984			241,042	241,986	+	314	oppB_1	CPn0199	oligopeptide permease	17	10	62	3
				242,017	242,871	+	284	oppC_1	CPn0200	oligopeptide permease	1	1	11	0
				242,864	243,718	+	284	oppD	CPn0201	oligopeptide transport ATPase	1	0	18	1
			244,846	243,715	244,503	+	262	oppF	CPn0202	oligopeptide transport ATPase	1	0	1	2
				245,008	245,805	+	265	CPn0203	CPn0203	hypothetical protein	0	0	0	0
				245,817	246,005	+	62	CPn0204	CPn0204	hypothetical protein	0	0	0	0
				246,133	246,330	+	65	CPn0205	CPn0205	hypothetical protein	3	2	22	4
S	246,181			246,409	247,164	+	251	CPn0206	CPn0206	hypothetical protein	10	0	29	3
				247,208	248,620	+	470	ybhI	CPn0207	dicarboxylase translocator	10	4	23	3
IG	248,672		248,719			+		pCpn16	pCpn16	putative sRNA, intergenic between CPn0207 and CPn0208	21	96	34	53
				248,953	250,605	+	550	pfkA_2	CPn0208	diphosphate--fructose-6-phosphate 1-phosphotransferase	12	0	35	3
				251,036	251,275	+	79	CPn0209	CPn0209	hypothetical protein	0	0	0	0
AS	251,272		250,976			-		pCpn17	pCpn17	putative sRNA, antisense to CPn0209	2	1	10	9
				251,437	252,384	-	315	CPn0210	CPn0210	hypothetical protein	3	0	4	1
				252,460	252,756	-	98	CPn0211	CPn0211	hypothetical protein	4	1	2	0
				252,885	254,066	-	393	CPn0212	CPn0212	hypothetical protein	4	0	6	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				254,187	254,342	-	51	CPn0213	CPn0213	hypothetical protein	3	0	0	0
				254,443	255,657	-	404	CPn0214	CPn0214	hypothetical protein	1	0	7	3
				255,756	257,015	-	419	CPn0215	CPn0215	hypothetical protein	10	2	15	3
	257,637			257,171	257,608	-	145	CPn0216	CPn0216	hypothetical protein	6	2	11	4
				257,896	258,582	+	228	ypdP	CPn0217	hypothetical protein	4	0	2	0
				258,579	259,058	-	159	CPn0218	CPn0218	hypothetical protein	11	3	5	0
				259,357	260,475	+	372	tgt	CPn0219	queuine tRNA-ribosyltransferase	15	5	6	0
AS	260,651		260,059			-		pCpn18	pCpn18	putative sRNA, intergenic overlaps with CPn0083	5	5	11	3
	260,663		261,390	260,696	261,241	+	181	CPn0220	CPn0220	hypothetical protein	84	34	25	3
	261,616	261,631		261,657	262,067	+	136	CPn0221	CPn0221	hypothetical protein	31	14	94	21
				262,504	262,845	+	113	CPn0222	CPn0222	hypothetical protein	9	0	1	0
	262,938			262,956	263,336	+	126	CPn0223	CPn0223	hypothetical protein	9	2	8	2
				263,435	263,677	+	80	CPn0224	CPn0224	hypothetical protein	9	0	0	2
				263,873	264,544	+	223	CPn0225	CPn0225	hypothetical protein	10	8	11	0
		264,614		264,566	264,970	+	134	CPn0226	CPn0226	hypothetical protein	0	0	0	0
			264,492	265,006	265,416	-	136	dsbB	CPn0227	putative disulfide oxidoreductase	3	1	29	2
	266,135			265,409	266,110	-	233	dsbG	CPn0228	disulfide bond chaperone	3	2	12	0
	266,307		267,752	266,328	267,563	+	411	CPn0229	CPn0229	hypothetical protein	11	3	37	2
				267,573	268,253	-	226	CPn0230	CPn0230	hypothetical protein	5	1	6	2
				268,250	268,957	-	235	tauB	CPn0231	nitrate/iron ABC transporter ATPase	29	5	25	9
				268,992	269,065	-	74	Ile tRNA	CPnt07	Ile tRNA	3696	12290	4475	10432
	269,203			269,070	269,142	-	73	Ala tRNA	CPnt08	Ala tRNA	1583	2815	1726	4065
	270,152			269,229	270,122	-	297	CPn0232	CPn0232	putative 5'-methylthioadenosine nucleosidase	3	2	23	1
AS	270,299		270,431			-		pCpn19	pCpn19	putative sRNA, antisense to CPn0233	1	1	6	5
				270,245	270,424	-	59	CPn0233	CPn0233	hypothetical protein	0	4	2	3
	271,269		270,487	270,545	271,240	-	231	CPn0234	CPn0234	hypothetical protein	228	92	135	27
S	270,939		270,890			-		pCpn20	pCpn20	putative sRNA, sense to CPn0234	30	45	10	15
	271,379			271,416	272,180	+	254	kdsB	CPn0235	3-deoxy-manno-octulosonate cytidylyltransferase	17	13	20	4
				272,156	273,769	+	537	pyrG	CPn0236	CTP synthetase	15	5	28	1
				273,762	274,217	+	151	yggF	CPn0237	Holliday junction resolvase-like protein	4	0	3	0
				274,303	275,841	+	512	zwf	CPn0238	glucose-6-phosphate 1-dehydrogenase	8	0	23	3
				275,899	276,675	+	258	devB	CPn0239	6-phosphogluconolactonase	4	0	10	0
	277,941			276,695	277,861	-	388	CPn0240	CPn0240	hypothetical protein	27	28	17	1
				278,200	279,354	-	384	CPn0241	CPn0241	hypothetical protein	59	11	29	5
AS	278,732		278,895			+		pCpn21	pCpn21	putative sRNA, antisense to CPn0241	23	4	5	6
AS	279,422		279,724			+		pCpn22	pCpn22	putative sRNA, antisense to CPn0242	0	5	0	4
	279,955			279,484	279,918	-	144	CPn0242	CPn0242	hypothetical protein	23	6	22	2

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	280,591		280,084	280,130	280,555	-	141	CPn0243	CPn0243	hypothetical protein	152	65	41	6
	280,733			280,918	281,559	+	213	adk	CPn0244	adenylate kinase	33	2	36	3
			282,704	281,645	282,502	+	285	ydhO	CPn0245	polysaccharide hydrolase-invasin repeat family	13	10	10	12
AS	281,677		281,267			-		pCpn23	pCpn23	putative sRNA, antisense to CPn0244 and CPn0245	0	0	7	5
			282,485	282,548	282,952	-	134	rpsI	CPn0246	30S ribosomal protein S9	38	3	19	6
	283,439			282,966	283,415	-	149	rplM	CPn0247	50S ribosomal protein L13	102	29	373	22
AS	282,641		282,703			+		pCpn24	pCpn24	putative sRNA, antisense to CPn0246	3	12	6	13
			283,615	283,647	284,327	-	226	ycfV	CPn0248	ABC transporter ATPase	2	0	11	0
				284,330	285,841	-	503	CPn0249	CPn0249	hypothetical protein	44	12	56	24
AS	286,087			285,899	286,057	-	52	rpmG	CPn0250	50S ribosomal protein L33	8	9	16	12
				286,060	287,562	+	500	CPn0251	CPn0251	hypothetical protein	18	3	38	6
				288,456	289,262	-	268	CPn0254	CPn0254	hypothetical protein	0	0	2	0
				289,326	290,165	-	279	CPn0255	CPn0255	hypothetical protein	5	0	3	0
				290,395	291,264	-	289	CPn0256	CPn0256	hypothetical protein	1	0	4	0
				291,264	292,127	-	287	CPn0257	CPn0257	hypothetical protein	0	0	0	0
			293,316	293,317	293,399	-	83	Tyr tRNA	CPnt09	Tyr tRNA	330	373	252	477
	293,497			293,405	293,477	-	73	Thr tRNA	CPnt10	Thr tRNA	705	1680	909	2688
				293,545	294,045	-	166	secA_1	CPn0260	protein translocase subunit	4	2	12	0
	294,276			294,302	295,036	+	244	ydaO	CPn0261	PP-loop superfamily ATPase	2	1	12	0
	295,081			295,091	295,936	+	281	surE	CPn0262	stationary phase survival protein SurE	8	0	29	0
	296,065			296,075	296,147	+	73	Val tRNA	CPnt11	Val tRNA	676	1236	815	1681
			296,226	296,151	296,224	+	74	Asp tRNA	CPnt12	Asp tRNA	721	2465	1536	2395
				296,249	297,139	+	296	yqfU	CPn0263	hypothetical protein	24	0	23	2
S	297,709			297,152	297,730	-	192	ubiD	CPn0264	aromatic acid decarboxylase	6	1	2	4
				297,727	298,620	-	297	ubiA	CPn0265	prenyltransferase	1	0	7	0
IG	298,666		298,738			+		pCpn25	pCpn25	putative sRNA, intergenic between CPn0265 and CPn 266	838	26	454	18
	299,159			299,184	299,879	+	231	CPn0266	CPn0266	hypothetical protein	5	3	0	0
				300,122	300,913	+	263	CPn0267	CPn0267	hypothetical protein	3	0	1	0
				300,935	301,321	+	128	CPn0268	CPn0268	hypothetical protein	0	0	0	0
				301,473	302,450	-	325	CPn0269	CPn0269	dipeptidase	5	0	17	2
				302,465	303,325	-	286	ywlC	CPn0270	SuA5 superfamily-related protein	3	0	11	0
	303,609			303,634	304,365	+	243	CPn0271	CPn0271	lysophospholipase esterase	7	1	26	4
				304,337	305,233	-	298	dnaX_2	CPn0272	DNA polymerase III subunit delta'	1	2	5	1
				305,224	305,844	-	206	tmk	CPn0273	thymidylate kinase	9	0	13	1
				305,849	308,353	-	834	gyrA_1	CPn0274	DNA gyrase subunit A	44	6	105	11
				308,369	310,786	-	805	gyrB	CPn0275	DNA gyrase subunit B	37	6	93	6

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	311,277			310,790	311,137	-	115	CPn0276	CPn0276	hypothetical protein	21	3	89	8
				311,401	311,910	-	169	CPn0277	CPn0277	hypothetical protein	2	0	12	0
S		312,788		312,057	312,875	-	272	CPn0278	CPn0278	outer membrane lipoprotein protein	7	2	16	4
				312,872	313,537	-	221	CPn0279	CPn0279	ABC transporter permease	6	1	24	1
				313,547	314,572	-	341	dppF_1	CPn0280	peptide ABC transporter ATPase	3	0	2	1
	315,030			315,057	316,106	+	349	dhnA	CPn0281	fructose-bisphosphate aldolase	9	6	79	9
				316,126	317,532	+	468	xasA	CPn0282	amino acid transporter	13	1	44	0
				317,529	318,497	-	322	CPn0283	CPn0283	hypothetical protein	7	2	12	3
AS	317,557		317,607			+		pCpn26	pCpn26	putative sRNA, antisense to CPn0284	11	14	21	23
S	317,606		317,562			-		pCpn27	pCpn27	putative sRNA, sense to 3'end of CPn0284	1	2	10	3
AS	317,678		317,840			+		pCpn28	pCpn28	putative sRNA, antisense to CPn0284	5	2	7	2
				318,548	319,045	-	165	CPn0284	CPn0284	hypothetical protein	3	1	7	3
	320,629			319,048	320,595	-	515	CPn0285	CPn0285	hypothetical protein	30	13	22	3
S	322,355			320,647	322,059	-	470	mgtE	CPn0286	Mg2+ transporter	15	1	63	9
				322,086	324,221	-	711	CPn0287	CPn0287	hypothetical protein	9	3	36	5
	325,824			324,568	325,716	-	382	CPn0288	CPn0288	hypothetical protein	3	0	11	2
AS	325,089		325,380			+		pCpn29	pCpn29	putative sRNA, antisense to CPn0288	9	0	18	5
	325,791			325,812	326,999	+	395	aaaT	CPn0289	neutral amino acid (glutamate) transporter	6	0	16	0
				327,042	328,526	+	494	CPn0290	CPn0290	Na-dependent transporter	6	0	14	0
	328,638			328,667	329,197	+	176	incB	CPn0291	inclusion membrane protein B	15	12	20	5
				329,228	329,839	+	203	incC	CPn0292	inclusion membrane protein C	9	1	17	1
				329,949	332,726	+	925	CPn0293	CPn0293	hypothetical protein	26	26	65	11
S	330,088		330,142			+		pCpn30	pCpn30	putative sRNA, sense to CPn0293	7	0	14	4
AS	330,522		330,471			-		pCpn31	pCpn31	putative sRNA, antisense to CPn0293	8	19	14	15
				333,092	333,505	+	137	CPn0294	CPn0294	cAMP-dependent protein kinase regulatory subunit	4	0	3	2
IG	333,613	333,540	333,433			-		pCpn32	pCpn32	putative sRNA, intergneic between CPn0294 and CPn0295	281	153	382	219
				333,624	333,863	-	79	acpP	CPn0295	acyl carrier protein	10	0	75	5
AS	333,759		333,823			+		pCpn33	pCpn33	putative sRNA, antisense to CPn0295	0	0	17	11
				334,019	334,765	-	248	fabG	CPn0296	3-ketoacyl-(acyl-carrier-protein) reductase	16	0	40	7
S	336,015			334,771	335,697	-	308	fabD	CPn0297	acyl-carrier-protein S-malonyltransferase	8	1	23	0
	336,733			335,714	336,721	-	335	fabH	CPn0298	3-oxoacyl-(acyl carrier protein) synthase III	19	2	59	15
AS	336,484		336,546			+		pCpn34	pCpn34	putative sRNA, antisense to CPn0298	4	1	10	2
	336,771			336,816	337,418	+	200	recR	CPn0299	recombination protein RecR	7	2	2	4
	337,734			337,783	340,155	+	790	yaeT	CPn0300	OMP85 family membrane protein	27	3	91	7
	340,226			340,250	340,765	+	171	CPn0301	CPn0301	OmpH-like Outer membrane protein	19	3	84	31
		340,775	342,104	340,787	341,869	+	360	lpxD	CPn0302	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	68	6	128	15

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
			341,899	341,918	342,958	-	346	CPn0303	CPn0303	hypothetical protein	1	2	12	1
	343,095			343,133	344,161	+	342	pdhA	CPn0304	pyruvate dehydrogenase Alpha	24	5	71	9
				344,154	345,140	+	328	pdhB	CPn0305	pyruvate dehydrogenase Beta	17	1	33	0
			346,492	345,145	346,434	+	429	pdhC	CPn0306	branched-chain alpha-keto acid dehydrogenase subunit E2	6	2	28	6
	349,012	347,404	346,447	346,512	348,986	-	824	glgP	CPn0307	glycogen phosphorylase	20	1	83	17
	349,211		349,695	349,234	349,599	+	121	CPn0308	CPn0308	hypothetical protein	58	32	43	13
	351,008			349,592	350,974	-	460	dnaA	CPn0309	chromosomal replication initiation protein	7	0	30	9
AS	349,722		350,069			+		pCpn35	pCpn35	putative sRNA, antisense to CPn0309	14	6	18	12
				351,046	353,433	-	795	60IM	CPn0310	putative inner membrane protein translocase component YidC	18	11	65	6
				353,572	354,438	-	288	lgt	CPn0311	prolipoprotein diacylglyceryl transferase	4	0	19	2
	354,500			354,524	354,979	+	151	CPn0312	CPn0312	hypothetical protein	12	1	56	2
				354,990	355,358	+	122	acpS	CPn0313	4'-phosphopantetheinyl transferase	0	0	2	0
	356,442			355,350	356,285	-	311	trxB	CPn0314	thioredoxin reductase	15	0	48	3
	356,812	356,940		356,977	358,719	+	580	rpsA	CPn0315	30S ribosomal protein S1	40	2	95	20
				358,820	360,124	+	434	nusA	CPn0316	transcription elongation factor NusA	23	1	46	0
				360,081	362,753	+	890	infB	CPn0317	translation initiation factor IF-2	18	1	54	9
				362,767	363,129	+	120	rbfA	CPn0318	ribosome-binding factor A	1	1	18	2
				363,175	363,882	+	235	truB	CPn0319	tRNA pseudouridine synthase B	9	0	16	1
				363,860	364,786	+	308	ribF	CPn0320	bifunctional riboflavin kinase/FMN adenylyltransferase	0	0	4	1
	365,990			364,764	365,858	-	364	ychF	CPn0321	GTP-dependent nucleic acid-binding protein EngD	16	5	20	4
	366,208			366,249	367,331	+	360	yscU	CPn0322	type III secretion system protein	8	1	42	24
				367,331	369,463	+	710	lcrD	CPn0323	low calcium response D	15	1	56	10
S	369,367			369,492	370,691	+	399	lcrE	CPn0324	low calcium response E	20	1	73	2
				370,708	371,151	+	147	sycE	CPn0325	secretion chaperone	3	0	24	2
				371,148	372,728	+	526	malQ	CPn0326	4-alpha-glucanotransferase	1	0	4	1
	372,893			372,945	373,214	+	89	rpmB	CPn0327	50S ribosomal protein L28	33	9	83	7
				373,241	374,995	+	584	CPn0328	CPn0328	hypothetical protein	16	4	19	5
			376,325	375,088	376,149	+	353	CPn0329	CPn0329	phopholipase D superfamily protein	8	6	15	2
				376,199	376,675	-	158	CPn0330	CPn0330	hypothetical protein	36	15	49	12
	378,474			376,698	378,437	-	579	CPn0331	CPn0331	hypothetical protein	530	314	761	168
S			378,551	378,533	378,655	-	40	CPn0332	CPn0332	hypothetical protein	12252	14203	9173	4542
	379,124			378,797	379,090	-	97	CPn0333	CPn0333	late transcription unit B protein	182	83	231	53
	379,267			379,311	379,826	+	171	CPn0334	CPn0334	hypothetical protein	6	1	9	1
				379,817	380,677	+	286	folD	CPn0335	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	11	2	6	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				380,650	381,594	+	314	yojL	CPn0336	hypothetical protein	8	5	15	9
	382,052			381,572	382,027	-	151	smpB	CPn0337	SsrA-binding protein	5	3	14	1
	382,209			382,278	383,378	+	366	dnaN	CPn0338	DNA polymerase III subunit beta	12	1	92	15
S	383,265			383,419	384,037	+		CPn0339	CPn0339	-	12	0	7	0
				383,841	384,159	+		CPn0340	CPn0340	-	0	0	0	0
				384,159	384,498	+		CPn0341	CPn0341	-	0	0	0	0
	384,611			384,621	385,065	+		CPn0342	CPn0342	-	26	10	25	7
				384,998	385,598	+		CPn0343	CPn0343	-	17	9	6	0
				385,555	387,420	-	621	yaeL	CPn0344	metalloprotease	19	1	34	5
	388,599			387,433	388,572	-	379	yaeM	CPn0345	1-deoxy-D-xylulose 5-phosphate reductoisomerase	19	3	94	3
AS	389,562		389,909			+		pCpn36	pCpn36	putative sRNA, antisense to CPn0346 and CPn0347	2	1	2	2
				388,701	389,675	-	324	ytgD	CPn0346	integral membrane protein	1	1	7	1
				389,675	391,021	-	448	ytgC	CPn0347	integral membrane protein	4	0	18	2
				391,024	391,803	-	259	ytgB_1	CPn0348	rRNA methylase	8	3	24	3
	392,804			391,787	392,770	-	327	ytgA	CPn0349	solute protein binding family	17	3	39	5
	393,126			393,181	393,687	+	168	CPn0350	CPn0350	hypothetical protein	41	10	94	41
S		395,393	395,491	393,888	395,435	+	515	adt_1	CPn0351	ADP/ATP translocase	165	22	133	4
	395,555			395,574	396,833	+	419	CPn0352	CPn0352	hypothetical protein	7	3	14	2
				396,893	397,138	+	81	CPn0353	CPn0353	hypothetical protein	2	1	0	1
S	397,135			397,167	398,510	+	447	CPn0354	CPn0354	hypothetical protein	7	2	23	7
	399,916			398,588	399,889	-	433	CPn0355	CPn0355	hypothetical protein	13	6	55	14
				400,106	400,459	-	117	CPn0356	CPn0356	hypothetical protein	0	0	1	0
				400,466	401,317	-	283	CPn0357	CPn0357	hypothetical protein	4	1	20	1
				401,575	401,751	-	58	CPn0358	CPn0358	hypothetical protein	1	0	6	0
	401,862		403,914	402,012	403,820	+	602	lepA	CPn0359	GTP-binding protein LepA	14	4	30	7
S	405,483		403,877	403,919	405,358	-	479	gnd	CPn0360	6-phosphogluconate dehydrogenase	50	20	72	17
				405,379	406,647	-	422	tyrS	CPn0361	tyrosyl-tRNA synthetase	25	13	42	5
	407,848	407,845		407,052	407,825	-	257	rpsD	CPn0362	RNA polymerase sigma factor sigma-28	28	12	107	5
	410,234	409,801		407,940	409,688	-	582	flhA	CPn0363	type III secretion system protein	9	0	30	4
	409,824		409,924	409,848	409,922	+	75	Pro tRNA	CPnt13	Pro tRNA	162	258	91	394
				409,966	410,241	+	91	fer4	CPn0364	ferredoxin IV	12	0	24	1
AS	410,234		410,104			-		pCpn37	pCpn37	putative sRNA, antisense to CPn0364	1	1	2	1
IG	410,321		410,359			+		pCpn38	pCpn38	putative sRNA, intergenic between CPn0264 and CPn0365	29	28	40	26
	410,421	410,506		410,528	411,547	+	339	CPn0365	CPn0365	hypothetical protein	15	3	5	3
				411,976	412,443	+	155	CPn0366	CPn0366	hypothetical protein	7	1	0	0
	413,086			413,102	413,839	+	245	CPn0367	CPn0367	hypothetical protein	62	13	39	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				413,790	414,110	+	106	CPn0368	CPn0368	hypothetical protein	6	0	4	0
	414,327			414,351	415,565	+	404	CPn0369	CPn0369	hypothetical protein	148	79	100	28
				415,800	416,915	+	371	CPn0370	CPn0370	hypothetical protein	40	14	28	3
	417,079			417,147	417,506	+	119	CPn0371	CPn0371	hypothetical protein	10	2	7	2
	417,661			417,687	418,004	+	105	CPn0372	CPn0372	hypothetical protein	7	1	10	6
	418,354			418,380	420,221	+	613	gcpE	CPn0373	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	41	4	9	2
				420,218	420,964	+	248	CPn0374	CPn0374	hypothetical protein	2	2	15	1
				421,121	421,618	+	165	CPn0375	CPn0375	hypothetical protein	4	2	11	1
				421,854	422,297	+	147	CPn0376	CPn0376	hypothetical protein	1	1	2	0
				422,344	423,438	-	364	sucB_1	CPn0377	dihydrolipoamide succinyltransferase	12	0	40	9
	426,201			423,442	426,168	-	908	sucA	CPn0378	2-oxoglutarate dehydrogenase E1 component	20	7	128	13
	426,180			426,322	426,768	+	148	CPn0379	CPn0379	hypothetical protein	10	2	7	1
				426,758	427,879	+	373	hemN_1	CPn0380	coproporphyrinogen III oxidase	21	1	6	2
	429,828			428,034	429,809	-	591	CPn0381	CPn0381	hypothetical protein	36	26	29	4
AS	428,595		418,822			+		pCpn39	pCpn39	putative sRNA, antisense to CPn0381	3	0	4	0
			429,951	430,033	430,749	-	238	yabC_1	CPn0382	SAM-dependent methytransferase	67	57	54	14
S	430,107		429,959			-		pCpn40	pCpn40	putative sRNA, sense to 3' end of CPn0282	0	4	16	12
	431,759	431,566		430,746	431,693	-	315	CPn0383	CPn0383	hypothetical protein	267	176	190	63
	432,409		431,793	431,859	432,377	-	172	hctB	CPn0384	histone-like protein 2	978	616	871	267
			432,451	432,519	434,018	-	499	pepA	CPn0385	leucyl aminopeptidase	44	2	85	7
				434,043	434,525	-	160	ssb	CPn0386	single-stranded DNA-binding protein	9	0	43	5
	435,228	435,225		434,696	435,196	-	166	CPn0387	CPn0387	hypothetical protein	50	24	175	89
LL	435,329			435,329	437,323	+	664	glgX	CPn0388	glycogen hydrolase (debranching)	12	1	36	2
S	438,387		437,099	437,316	438,134	-	272	CPn0389	CPn0389	hypothetical protein	23	5	112	21
	439,225			438,131	439,144	-	337	ruvB	CPn0390	Holliday junction DNA helicase RuvB	2	3	44	32
				439,507	439,692	-	61	CPn0391	CPn0391	hypothetical protein	0	0	0	0
AS	439,638			439,814	440,386	+	190	dcd	CPn0392	deoxycytidine triphosphate deaminase	17	2	62	2
				440,379	440,726	+	115	CPn0393	CPn0393	hypothetical protein	9	0	12	1
				440,736	441,971	+	411	tlyC_1	CPn0394	CBS domain protein (hemolysin homolog)	8	0	23	4
			443,279	441,964	443,178	+	404	CPn0395	CPn0395	hypothetical protein	6	0	18	0
S	444,625			443,238	444,353	-	371	yhfO	CPn0396	NifS-related aminotransferase	13	5	23	8
	445,182			444,378	445,115	-	245	CPn0397	CPn0397	PP2C phosphatase family protein	2	0	9	3
IG	445,308		445,452			+		pCpn41	pCpn41	putative sRNA, intergenic upstream of CPn0398	253	348	757	907
				445,533	445,703	+	56	CPn0398	CPn0398	hypothetical protein	18	5	75	5
				445,879	446,526	+	215	CPn0399	CPn0399	hypothetical protein	8	0	39	2
				446,536	447,309	+	257	CPn0400	CPn0400	hypothetical protein	3	0	19	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
IG	447,422		447,312			-		SRPRNA	CPns03	SRP_RNA	2355	2465	1914	1402
S	448,178			447,492	447,884	-	130	CPn0401	CPn0401	hypothetical protein	4	7	49	26
				447,885	448,994	-	369	mutY	CPn0402	adenine glycosylase	4	0	2	0
S	449,019			449,015	449,713	+	232	yceC	CPn0403	pseudouridine synthetase family protein	5	0	2	2
				449,868	450,887	-	339	CPn0404	CPn0404	hypothetical protein	3	0	0	0
				450,963	451,739	-	258	CPn0405	CPn0405	hypothetical protein	2	0	2	0
	451,874	451,941	452,950	451,969	452,868	+	299	fabI	CPn0406	enoyl-(acyl carrier protein) reductase	36	5	49	9
				452,855	453,742	-	295	CPn0407	CPn0407	HAD family phosphatase/phosphatase	2	0	8	1
	454,010	454,083		454,105	454,584	+	159	CPn0408	CPn0408	hypothetical protein	10	6	22	2
	454,588			454,645	455,130	+	161	CPn0409	CPn0409	hypothetical protein	20	6	21	15
S	455,041			455,123	455,836	+	237	dnaQ_1	CPn0410	DNA polymerase III subunit epsilon	27	5	31	5
				455,833	456,612	+	259	CPn0411	CPn0411	hypothetical protein	3	2	10	2
				456,590	457,249	+	219	CPn0412	CPn0412	hypothetical protein	0	0	3	0
				457,224	459,203	-	659	msbA	CPn0413	transport ATP binding protein	8	1	28	3
				459,169	460,143	-	324	accA	CPn0414	acetyl-CoA carboxylase carboxyltransferase subunit alpha	17	2	68	3
S	461,462			460,218	461,498	-	426	CPn0415	CPn0415	hypothetical protein	34	8	137	50
AS	461,510		461,668			+		pCpn42	pCpn42	putative sRNA, antisense to CPn0416	5	0	9	1
			461,486	461,554	461,856	-	100	himD	CPn0416	integration host factor Alpha	109	23	332	40
S	463,078			462,241	463,035	-	264	amiA	CPn0417	N-acetylmuramoyl alanine amidase	12	4	48	19
	462,136		462,215	462,141	462,214	+	74	Arg tRNA	CPnt14	Arg tRNA	133	695	162	1018
	464,527			462,950	464,401	-	483	murE	CPn0418	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	16	4	20	8
AS	463,808		463,986			+		pCpn43	pCpn43	putative sRNA, antisense to CPn0418	4	0	4	1
				464,873	466,834	-	653	pbp3	CPn0419	transglycolase/transpeptidase	4	0	12	0
				466,821	467,108	-	95	CPn0420	CPn0420	hypothetical protein	1	1	5	0
	468,029			467,105	467,998	-	297	mraW	CPn0421	S-adenosyl-methyltransferase MraW	5	0	13	0
AS	467,414		467,532			+		pCpn44	pCpn44	putative sRNA, antisense to CPn0421	0	0	5	4
	468,217			468,242	468,787	+	181	CPn0422	CPn0422	hypothetical protein	23	1	96	16
			469,277	468,791	469,219	+	142	CPn0423	CPn0423	hypothetical protein	8	0	25	1
	469,398			469,612	470,964	+	450	dnaA	CPn0424	chromosomal replication initiation protein	44	24	226	158
				470,980	471,567	+	195	CPn0425	CPn0425	hypothetical protein	1	0	13	2
				471,533	472,111	-	192	CPn0426	CPn0426	hypothetical protein	0	0	0	0
				472,207	473,718	+	503	nqr2	CPn0427	Na(+)-translocating NADH-quinone reductase subunit B	5	1	45	2
				473,722	474,684	+	320	nqr3	CPn0428	Na(+)-translocating NADH-quinone reductase subunit C	4	0	17	1
AS	473,904		473,535			-		pCpn45	pCpn45	putative sRNA, antisense to CPn0428	0	0	8	3

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				474,681	475,322	+	213	nqr4	CPn0429	Na(+)-translocating NADH-quinone reductase subunit D	2	0	11	3
	475,287			475,326	476,096	+	256	nqr5	CPn0430	Na(+)-translocating NADH-quinone reductase subunit E	2	0	18	2
				476,148	476,483	-	111	CPn0431	CPn0431	hypothetical protein	2	1	0	0
				476,511	476,816	-	101	CPn0432	CPn0432	hypothetical protein	7	0	5	0
				476,926	477,273	-	115	gcsH	CPn0433	glycine cleavage system protein H	2	0	10	1
S	479,378			477,273	479,462	-	729	CPn0434	CPn0434	hypothetical protein	34	2	82	14
			479,406	479,472	480,902	-	476	CPn0435	CPn0435	phospholipase D superfamily protein	22	33	32	6
	481,711			480,899	481,618	-	239	lplA_1	CPn0436	lipoate protein ligase-like protein	12	6	8	4
	481,787		484,440	481,816	484,353	+	845	clpC	CPn0437	ClpC protease	35	13	223	10
	485,497		484,231	484,331	485,416	-	361	mnmA	CPn0438	tRNA-specific 2-thiouridylase MnmA	6	8	26	6
	485,532			485,553	486,080	+	175	CPn0439	CPn0439	hypothetical protein	39	4	11	2
			486,763	486,105	486,743	+	212	CPn0440	CPn0440	hypothetical protein	4	2	18	0
	486,858			486,891	487,841	+	316	CPn0441	CPn0441	hypothetical protein	18	5	24	12
				488,013	488,531	+	172	CPn0442	CPn0442	hypothetical protein	11	1	3	1
	488,670			488,729	489,982	+	417	CPn0443	CPn0443	hypothetical protein	329	227	273	103
			494,713	490,287	494,510	+	1407	pmp_6	CPn0444	polymorphic Outer membrane protein G/I family	125	48	316	44
	494,742			494,772	497,582	+	936	pmp_7	CPn0445	polymorphic Outer membrane protein G family	74	12	263	16
			500,455	497,626	500,418	+	930	pmp_8	CPn0446	polymorphic Outer membrane protein G family	62	10	258	22
	500,539		503,697	500,568	503,354	+	928	pmp_9	CPn0447	polymorphic Outer membrane protein G/I family	38	8	168	6
IG	503,607		503,533			+		pCpn46	pCpn46	putative sRNA, intergenic between CPn0447 and CPn0448	5	5	8	5
				503,695	504,810	-	371	yxjG_2	CPn0448	hypothetical protein	2	0	7	0
S		506,746	505,175	505,326	507,231	-		pmp_10.1	CPn0449	-	118	16	322	46
	508,137			507,176	508,112	-		pmp_10.2	CPn0450	-	52	11	334	36
	508,248			508,275	511,061	+	928	pmp_11	CPn0451	polymorphic Outer membrane protein G family	31	5	168	28
			512,944	511,319	512,863	+	514	pmp_12	CPn0452	polymorphic Outer membrane protein (truncated) A/I family	19	7	32	9
IG	513,197		512,905			-		pCpn47	pCpn47	putative sRNA, intergenic between CPn0452 and CPn 0453	2	0	2	15
	513,209			513,234	516,155	+	973	pmp_13	CPn0453	polymorphic Outer membrane protein G family	43	19	391	53
			519,308	516,182	519,118	+	978	pmp_14	CPn0454	polymorphic Outer membrane protein H family	192	94	731	137
				519,455	520,348	-	297	CPn0455	CPn0455	hypothetical protein	3	0	2	2
	521,566			520,324	521,532	-	402	CPn0456	CPn0456	hypothetical protein	4	0	7	0
				522,117	523,865	-	582	CPn0457	CPn0457	hypothetical protein	44	26	46	49
S	523,061		522,987			-		pCpn48	pCpn48	putative sRNA, sense to CPn0457	35	23	38	46
				524,233	526,320	-	695	CPn0458	CPn0458	hypothetical protein	38	15	10	2
				526,616	527,005	-	129	CPn0459	CPn0459	hypothetical protein	1	0	0	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				526,989	527,840	-	283	CPn0460	CPn0460	hypothetical protein	1	0	0	0
	528,667			527,841	528,638	-	265	CPn0461	CPn0461	hypothetical protein	12	22	12	3
	531,147			529,034	531,052	-	672	CPn0462	CPn0462	hypothetical protein	5	4	4	3
AS	531,107		531,468			+		pCpn49	pCpn49	putative sRNA, antisense to CPn0463	1	2	21	0
				531,188	532,357	-	389	CPn0463	CPn0463	hypothetical protein	1	1	0	1
				532,363	532,842	-	159	CPn0464	CPn0464	hypothetical protein	3	1	0	1
	533,440			532,868	533,212	-	114	CPn0465	CPn0465	hypothetical protein	3	4	0	0
	533,701			533,724	536,540	+	938	pmp_15	CPn0466	polymorphic Outer membrane protein E family	93	15	302	31
				536,633	539,437	+	934	pmp_16	CPn0467	polymorphic Outer membrane protein E family	22	6	237	22
S	539,655			539,631	540,435	+		pmp_17.1	CPn0468	-	30	1	97	10
				540,398	541,463	+		pmp_17.2	CPn0469	-	3	0	13	2
				541,357	542,535	+	392	pmp_17.3	CPn0470	polymorphic Outer membrane protein (Frame-shift with CPn0469)	0	0	4	0
				542,564	545,404	+	946	pmp_18	CPn0471	polymorphic Outer membrane protein E/F family	2	1	6	1
				545,578	547,905	-	775	CPn0472	CPn0472	hypothetical protein	295	299	431	192
IG	548,000		548,167			+		pCpn50	pCpn50	putative sRNA, intergenic between CPn0472 and CPn0473	18	64	17	24
				548,067	549,593	-	508	CPn0473	CPn0473	hypothetical protein	391	287	357	111
	551,592			549,804	551,573	-	589	CPn0474	CPn0474	hypothetical protein	123	69	79	9
S	553,868			551,682	553,844	-	720	glgB	CPn0475	glycogen branching enzyme	62	10	57	27
				553,855	554,844	-	329	CPn0476	CPn0476	hypothetical protein	9	2	6	0
	556,141			554,841	556,106	-	421	yqeV	CPn0477	hypothetical protein	14	7	22	15
				556,207	557,625	-	472	hflX	CPn0478	GTP binding protein	20	8	26	6
	558,521			557,613	558,425	-	270	phnP	CPn0479	metal dependent hydrolase	229	69	359	80
				558,647	559,303	-	218	CPn0480	CPn0480	hypothetical protein	11	4	11	7
				559,336	560,946	-	536	CPn0481	CPn0481	hypothetical protein	22	11	43	5
	561,771	561,754		560,958	561,737	-	259	artJ	CPn0482	arginine periplasmic binding protein	40	13	57	32
				561,836	564,967	+	1043	CPn0483	CPn0483	hypothetical protein	131	73	412	60
				564,970	565,827	+	285	aroG	CPn0484	3-deoxy-7-phosphoheptulonate synthase	6	6	10	13
	565,998		566,281	566,038	566,232	+	64	CPn0485	CPn0485	hypothetical protein	24	25	47	22
	567,946		566,337	566,402	567,784	-	460	CPn0486	CPn0486	putative proline permease	14	2	36	10
				568,109	569,740	-	543	CPn0487	CPn0487	hypothetical protein	12	0	14	3
				569,764	570,096	-	110	ycfF	CPn0488	HIT family hydrolase	11	2	22	6
	570,987			570,093	570,965	-	290	CPn0489	CPn0489	hypothetical protein	36	20	63	19
	571,191		573,520	571,279	573,336	+	685	CPn0490	CPn0490	hypothetical protein	3	1	39	1
	574,568		573,214	573,333	574,352	-	339	CPn0491	CPn0491	hypothetical protein	24	14	123	36
				574,652	574,807	+	51	CPn0492	CPn0492	hypothetical protein	3	0	0	0
				574,852	575,004	-	50	CPn0493	CPn0493	hypothetical protein	5	6	8	7

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
S	575,352			575,143	575,364	-	73	CPn0494	CPn0494	hypothetical protein	29	20	22	13
				575,603	576,796	+	397	aspC	CPn0495	L,L-diaminopimelate aminotransferase	11	1	22	8
				576,793	577,815	+	340	CPn0496	CPn0496	hypothetical protein	3	1	16	3
				577,817	578,089	-	90	CPn0497	CPn0497	hypothetical protein	36	57	35	17
	579,076			578,082	579,035	-	317	CPn0498	CPn0498	hypothetical protein	116	69	328	59
AS	578,934		579,257			+		pCpn51	pCpn51	putative sRNA, antisense to CPn0498 and CPn0499	18	6	22	4
	580,398			579,202	580,359	-	385	CPn0499	CPn0499	hypothetical protein	232	170	224	76
	580,542			580,659	582,365	+	568	proS	CPn0500	prolyl-tRNA synthetase	39	15	64	6
S	582,465			582,457	583,653	+	398	hrcA	CPn0501	heat-inducible transcription repressor	27	17	17	7
				583,650	584,204	+	184	grpE	CPn0502	HSP-70 cofactor	7	1	6	2
			586,272	584,234	586,216	+	660	dnaK	CPn0503	molecular chaperone DnaK	174	20	141	13
	586,356			586,487	588,517	+	676	vacB	CPn0504	ribonuclease family protein	21	8	49	12
				588,519	589,109	+	196	CPn0505	CPn0505	3-methyladenine DNA glycosylase	0	0	0	0
AS	589,097		588,471			-		pCpn52	pCpn52	putative sRNA, antisense to CPn0505	2	1	19	2
	589,136		589,856	589,172	589,843	+	223	CPn0506	CPn0506	hypothetical protein	11	2	34	5
	589,893			589,961	590,125	+	54	CPn0507	CPn0507	hypothetical protein	73	14	178	12
				590,142	590,303	+	53	CPn0508	CPn0508	hypothetical protein	19	4	20	0
				590,335	590,811	+	158	CPn0509	CPn0509	metalloenzyme	26	5	30	3
				590,813	591,976	+	387	tlyC_2	CPn0510	CBS Domains (hemolysin homolog)	9	2	11	1
	592,107			592,141	592,491	+	116	rsbV_1	CPn0511	sigma regulatory factor	3	0	19	2
	592,523			592,553	594,415	+	620	CPn0512	CPn0512	hypothetical protein	26	11	127	19
				594,647	595,756	+	369	CPn0513	CPn0513	hypothetical protein	12	3	33	2
				595,729	596,523	+	264	CPn0514	CPn0514	hypothetical protein	1	0	5	0
				596,471	597,184	+	237	ubiE	CPn0515	ubiquinone/menaquinone biosynthesis methyltransferase	0	0	4	1
				597,252	598,814	-	520	CPn0516	CPn0516	hypothetical protein	5	12	3	2
				598,792	599,631	-	279	CPn0517	CPn0517	hypothetical protein	11	1	7	0
	600,805			599,829	600,803	-	324	CPn0518	CPn0518	hypothetical protein	83	2	248	16
				600,901	601,674	-	257	dapF	CPn0519	diaminopimelate epimerase	15	3	24	2
				601,643	602,218	-	191	clpP_1	CPn0520	ATP-dependent Clp protease proteolytic subunit	12	3	28	6
				602,238	603,797	-	519	glyA	CPn0521	serine hydroxymethyltransferase	30	9	49	10
				603,987	604,658	+	223	hemD	CPn0522	uroporphyrinogen-III synthase	0	0	0	0
	604,691			604,723	605,055	+	110	CPn0523	CPn0523	hypothetical protein	11	2	3	2
	605,079			605,103	606,182	+	359	CPn0524	CPn0524	hypothetical protein	173	121	177	54
				606,522	607,286	+	254	CPn0525	CPn0525	hypothetical protein	85	25	137	14
			607,694	607,342	607,649	+	308	M1RNA	CPns02	ribonuclease P RNA	154	100	125	43
				607,707	608,696	-	329	kpsF	CPn0526	GutQ/KpsF family sugar-P isomerase	5	2	15	2

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				608,723	609,904	-	393	sucB_2	CPn0527	branched-chain alpha-keto acid dehydrogenase subunit E2	14	0	12	0
				609,918	611,162	-	414	gltT	CPn0528	glutamate symport	7	1	13	3
	612,339			611,162	612,259	-	365	lpxK	CPn0529	tetraacyldisaccharide 4'-kinase	16	5	14	12
				612,457	613,254	-	265	spoU_1	CPn0530	rRNA methylase	3	1	9	1
				613,242	614,069	-	275	CPn0531	CPn0531	SAM dependent methyltransferase	0	0	5	0
				614,072	614,674	-	200	ribC	CPn0532	riboflavin synthase subunit alpha	4	0	5	2
	614,882			614,930	615,388	+	152	nrdR	CPn0533	transcriptional regulator NrdR	22	1	97	3
				615,413	615,787	+	124	dksA	CPn0534	DnaK suppressor	9	0	37	2
				615,793	616,299	+	168	lspA	CPn0535	lipoprotein signal peptidase	15	1	42	2
	616,307			616,345	617,694	+	449	dagA_1	CPn0536	D-Ala/Gly permease	41	4	111	5
	617,802			617,833	618,192	+	119	CPn0537	CPn0537	hypothetical protein	66	75	135	46
				618,212	618,514	+	100	CPn0538	CPn0538	hypothetical protein	94	46	114	14
				618,705	621,548	+	947	pmp_19	CPn0539	polymorphic membrane protein A family	232	215	321	106
	621,658			621,694	626,865	+	1723	pmp_20	CPn0540	polymorphic membrane protein B family	69	32	360	50
	627,018		626,894	626,901	626,987	-	87	Ser tRNA	CPnt15	Ser tRNA	475	1291	468	1273
				627,170	628,006	+	278	CPn0541	CPn0541	solute binding protein	6	0	13	4
				628,003	628,740	+	245	CPn0542	CPn0542	ABC transporter ATPase	4	1	5	0
				628,725	629,606	+	293	CPn0543	CPn0543	metal transport protein	4	0	18	0
				629,522	630,529	-	335	obgE	CPn0544	GTPase ObgE	8	2	0	0
			630,553	630,630	630,884	-	84	rpmA	CPn0545	50S ribosomal protein L27	37	19	24	9
				630,909	631,229	-	106	rplU	CPn0546	50S ribosomal protein L21	15	0	27	1
	631,453			631,373	631,445	-	73	Phe tRNA	CPnt16	Phe tRNA	187	294	114	406
			632,325	631,661	632,191	+	176	ispF	CPn0547	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	11	0	11	5
				632,188	633,231	-	347	cysJ	CPn0548	sulfite reductase	13	0	22	0
				633,252	633,569	-	105	rpsJ	CPn0549	30S ribosomal protein S10	4	0	29	0
				633,577	635,661	-	694	fusA	CPn0550	elongation factor G	39	1	77	10
				635,695	636,168	-	157	rs7	CPn0551	30S ribosomal protein S7	25	1	30	1
	636,619			636,216	636,587	-	123	rpsL	CPn0552	30S ribosomal protein S12	30	60	68	30
AS	638,116			636,809	637,747	-	312	CPn0553	CPn0553	hypothetical protein	9	6	51	8
	637,840			637,854	638,144	+	96	CPn0554	CPn0554	hypothetical protein	0	0	0	0
	638,279		640,423	638,298	640,244	+	648	tsp	CPn0555	tail-specific protease	6	3	31	6
			640,280	640,322	640,912	-	196	crpA	CPn0556	hypothetical protein	11	15	27	14
			641,090	641,191	642,861	-	556	omcB	CPn0557	60 kDa cysteine-rich OMP	2216	2857	2331	1605
	643,396			643,028	643,300	-	90	omcA	CPn0558	9 kDa-cysteine-rich lipoprotein	653	385	1183	261
	643,725			643,742	643,930	+	62	CPn0559	CPn0559	hypothetical protein	20	22	22	10
				644,095	645,612	-	505	gltX	CPn0560	glutamyl-tRNA synthetase	41	4	17	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	646,490			645,868	646,404	-	178	euo	CPn0561	hypothetical protein	91	22	32	3
				646,915	648,036	-	373	CPn0562	CPn0562	CHLPS 43 kDa protein homolog_1	9	4	3	0
				648,290	650,056	-	588	recJ	CPn0563	ssDNA exonuclease	13	0	32	5
	654,381			650,142	654,350	-	1402	secD/secF	CPn0564	bifunctional preprotein translocase subunit SecD/SecF	72	9	162	11
IG	654,455		654,421			-		pCpn53	pCpn53	putative sRNA, intergenic between CPn0564 and CPn0565	4	3	1	5
				654,530	655,630	-	366	CPn0565	CPn0565	hypothetical protein	12	3	5	1
				656,141	656,893	+	250	yaeS	CPn0566	undecaprenyl pyrophosphate synthase	7	0	34	0
				656,894	657,820	+	308	cdsA	CPn0567	phosphatidate cytidylytransferase	6	1	34	4
				657,817	658,467	+	216	cmk	CPn0568	cytidylate kinase	1	2	13	5
S	658,443			658,464	659,102	+	212	plsC	CPn0569	glycerol-3-P acyltransferase	6	1	30	1
				659,107	660,792	+	561	argS	CPn0570	arginyl-tRNA synthetase	2	0	18	4
AS	660,667		660,351			-		pCpn54	pCpn54	putative sRNA, antisense to CPn0570	4	0	14	13
	662,154			660,746	662,122	-	458	murA	CPn0571	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	14	11	32	4
	662,315			662,352	664,619	+	755	CPn0572	CPn0572	hypothetical protein	243	139	363	72
				664,688	665,404	-	238	yebC	CPn0573	hypothetical protein	3	0	13	8
				665,391	665,945	-	184	CPn0574	CPn0574	hypothetical protein	6	1	10	1
				665,979	666,494	-	171	yhhY	CPn0575	amino group Acetyl transferase	2	0	10	0
	667,621			666,491	667,598	-	369	prfB	CPn0575a	peptide chain release factor 2	17	8	26	6
	667,849		668,252	667,895	668,158	+	87	CPn0577	CPn0577	SWIB (YM74) complex protein	24	33	63	90
				668,406	669,368	+	320	yaeI	CPn0578	phosphohydrolase	4	2	16	1
				669,361	669,996	+	211	ispD	CPn0579	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	1	1	7	1
				669,993	670,796	+	267	truA	CPn0580	tRNA pseudouridine synthase A	2	0	5	0
	671,464			670,742	671,434	-	230	CPn0581	CPn0581	phosphoglycolate phosphatase	6	1	8	8
AS	671,049		671,467			+		pCpn55	pCpn55	putative sRNA, antisense to CPn0581	8	0	16	7
				671,503	672,180	+	225	CPn0582	CPn0582	hypothetical protein	3	0	0	0
	672,226		672,324	672,236	672,318	+	83	Leu tRNA	CPnt17	Leu tRNA	2320	17928	4355	2186
	672,373			672,400	672,720	+	106	CPn0583	CPn0583	hypothetical protein	1	0	10	0
				672,707	673,801	+	364	atoS	CPn0584	two component sensor protein	1	0	32	2
		674,430	673,761	673,862	675,817	-	651	CPn0585	CPn0585	hypothetical protein	11	3	7	1
	676,002			676,026	677,186	+	386	atoC	CPn0586	two component regulator	3	1	6	2
	677,242		677,337	677,264	677,337	+	74	Arg tRNA	CPnt18	Arg tRNA	76	84	75	163
	677,382			677,441	678,127	+	228	yvyD	CPn0587	hypothetical protein	108	93	127	62
				678,084	678,629	+	181	CPn0588	CPn0588	hypothetical protein	89	62	77	31
				678,640	679,398	+	252	recO	CPn0589	DNA repair protein RecO	20	22	29	7
				679,513	680,112	-	199	CPn0590	CPn0590	hypothetical protein	8	2	12	1
	680,265		680,177	680,178	680,259	-	82	Leu tRNA	CPnt19	Leu tRNA	300	844	417	1576

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	680,347			680,373	681,023	+	216	yagE	CPn0591	YagE family protein	15	19	54	22
				681,153	681,464	+	103	yidD	CPn0592	hypothetical protein	0	6	3	2
AS	682,602			681,388	682,476	-	362	CPn0593	CPn0593	hypothetical protein	10	6	21	11
	682,558			682,583	684,961	+	792	pheT	CPn0594	phenylalanyl-tRNA synthetase subunit beta	23	9	79	15
S	684,803			684,958	685,929	+	323	CPn0595	CPn0595	hypothetical protein	19	2	73	7
				685,939	686,460	+	173	ada	CPn0596	methyltransferase	1	0	8	0
				686,476	688,215	-	579	oppC_2	CPn0597	oligopeptide permease	7	1	30	6
				688,216	689,697	-	493	oppB_2	CPn0598	oligopeptide permease	8	0	44	2
				689,679	691,802	-	707	oppA_5	CPn0599	oligopeptide binding lipoprotein	24	3	59	4
LL	692,147			691,824	692,147	-	107	CPn0600	CPn0600	hypothetical protein	20	14	166	72
IG	692,664		692,320	692,373	692,570	-	65	CPn0600.1	CPn0600.1	contains ORF, homologue to CP0147	251	48	531	137
				692,733	693,053	-	106	CPn0601	CPn0601	hypothetical protein	1	1	11	1
				693,101	694,105	-	334	CPn0602	CPn0602	hypothetical protein	8	0	8	0
				694,205	695,188	+	327	hemH	CPn0603	ferrochelatase	7	0	14	0
	695,974	695,574	695,056	695,193	695,945	-	250	fliY	CPn0604	glutamine binding protein	18	5	162	78
				696,147	696,707	-	186	yhhF	CPn0605	methylase	7	0	11	0
				696,704	697,444	-	246	CPn0606	CPn0606	hypothetical protein	5	2	9	1
	698,912			697,570	698,895	-	441	glgC	CPn0607	glucose-1-P adenyltransferase	25	6	44	23
S	700,033			699,013	699,645	-	210	CPn0608	CPn0608	orotate phosphoribosyltransferase	61	16	214	60
				699,705	699,989	+	94	CPn0609	CPn0609	hypothetical protein, unlikely, antisense transcription	0	0	0	0
				700,026	701,420	-	464	rho	CPn0610	transcription termination factor Rho	39	11	84	10
S	702,011			701,417	702,025	-	202	coaE	CPn0611	dephospho-CoA kinase	66	9	174	17
AS	702,670		702,941			+		pCpn56	pCpn56	putative sRNA, antisense to CPn0612	1	0	11	5
S		702,655		702,019	704,631	-	870	polA	CPn0612	DNA polymerase I	19	7	74	14
				704,655	705,656	-	333	sohB	CPn0613	protease	9	4	30	1
	707,543			705,780	707,402	-	540	adt_2	CPn0614	ADP/ATP translocase	39	14	214	30
	708,181			707,631	708,137	-	168	pgsA_1	CPn0615	glycerol-3-P phosphatidyltransferase	26	8	66	19
	708,491		708,352	708,354	708,441	-	88	Ser tRNA	CPnt20	Ser tRNA	899	3470	1207	5369
	708,693			708,791	710,140	+	449	dnaB	CPn0616	replicative DNA helicase	9	1	41	2
	710,222			710,484	712,319	+	611	gidA	CPn0617	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	35	9	68	40
				712,306	713,013	+	235	lplA_2	CPn0618	lipoate-protein ligase A	2	0	5	0
				713,010	713,444	-	144	ndk	CPn0619	nucleoside diphosphate kinase	1	1	7	1
				713,516	714,139	-	207	ruvA	CPn0620	Holliday junction DNA helicase RuvA	14	1	23	2
				714,141	714,647	-	168	ruvC	CPn0621	Holliday junction resolvase	11	1	19	1
	715,777			714,790	715,752	-	320	CPn0622	CPn0622	hypothetical protein	25	2	87	10
	716,055		715,888	715,889	715,971	-	83	Leu tRNA	CPnt21	Leu tRNA	211	1354	399	6669
			716,106	716,160	716,993	-	277	CPn0623	CPn0623	hypothetical protein	63	7	223	8

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	718,072			717,008	718,015	-	335	gapA	CPn0624	glyceraldehyde-3-phosphate dehydrogenase	53	3	170	10
				718,057	718,485	-	142	rplQ	CPn0625	50S ribosomal protein L17	2	0	31	2
				718,492	719,616	-	374	rpoA	CPn0626	DNA-directed RNA polymerase subunit alpha	13	3	41	5
				719,637	720,038	-	133	rs11	CPn0627	30S ribosomal protein S11	7	0	28	1
	720,452			720,060	720,428	-	122	rpsM	CPn0628	30S ribosomal protein S13	9	1	36	7
				720,484	721,857	-	457	secY	CPn0629	preprotein translocase subunit SecY	19	0	58	4
				721,882	722,316	-	144	rplO	CPn0630	50S ribosomal protein L15	9	1	30	0
				722,309	722,806	-	165	rpsE	CPn0631	30S ribosomal protein S5	12	0	21	0
				722,824	723,195	-	123	rplR	CPn0632	50S ribosomal protein L18	4	2	26	1
				723,206	723,757	-	183	rplF	CPn0633	50S ribosomal protein L6	9	0	43	1
				723,784	724,185	-	133	rpsH	CPn0634	30S ribosomal protein S8	14	1	35	1
				724,203	724,745	-	180	rplE	CPn0635	50S ribosomal protein L5	10	1	23	6
				724,747	725,082	-	111	rplX	CPn0636	50S ribosomal protein L24	4	1	16	1
				725,096	725,464	-	122	rplN	CPn0637	50S ribosomal protein L14	13	0	14	3
				725,487	725,747	-	86	rpsQ	CPn0638	30S ribosomal protein S17	7	0	9	1
				725,740	725,958	-	72	rpmC	CPn0639	50S ribosomal protein L29	5	1	19	5
				725,961	726,377	-	138	rplP	CPn0640	50S ribosomal protein L16	10	3	22	3
				726,406	727,077	-	223	rpsC	CPn0641	30S ribosomal protein S3	26	1	48	1
				727,093	727,428	-	111	rplV	CPn0642	50S ribosomal protein L22	15	0	22	2
				727,447	727,713	-	88	rpsS	CPn0643	30S ribosomal protein S19	1	1	4	1
				727,719	728,573	-	284	rplB	CPn0644	50S ribosomal protein L2	22	4	30	5
				728,595	728,930	-	111	rplW	CPn0645	50S ribosomal protein L23	12	0	19	2
				728,947	729,621	-	224	rplD	CPn0646	50S ribosomal protein L4	26	3	22	1
	730,436			729,654	730,313	-	219	rplC	CPn0647	50S ribosomal protein L3	67	9	90	9
			730,540	730,602	731,603	-	333	CPn0648	CPn0648	hypothetical protein	46	7	21	4
				731,707	732,672	-	321	fmt	CPn0649	methionyl-tRNA formyltransferase	16	0	25	4
				732,662	733,501	-	279	lpxA	CPn0650	UDP-N-acetylglucosamine acyltransferase	13	0	40	0
				733,514	733,975	-	153	fabZ	CPn0651	(3R)-hydroxymyristoyl-ACP dehydratase	17	0	41	1
				733,987	734,835	-	282	lpxC	CPn0652	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	25	4	78	12
				734,865	736,490	-	541	lnt	CPn0653	apolipoprotein N-acyltransferase	9	1	20	1
				736,500	736,967	-	155	yciA	CPn0654	acyl-CoA thioesterase	2	2	13	4
S	738,102			737,098	737,847	-	249	dnaQ_2	CPn0655	DNA polymerase III subunit epsilon	31	4	136	4
				737,872	738,051	+	59	CPn0656	CPn0656	hypothetical protein	0	0	0	0
				738,048	738,473	-	141	yjeE	CPn0657	hypothetical protein	43	12	110	20
S	738,367		738,277			-		pCpn57	pCpn57	putative sRNA, sense to CPn0657	37	12	99	18
	739,262			738,452	739,168	-	238	CPn0658	CPn0658	hypothetical protein	27	4	88	17
	739,381		739,487	739,403	739,486	+	84	Leu tRNA	CPnt22	Leu tRNA	47	107	107	290

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
			739,893	739,533	739,841	+	102	trxA	CPn0659	thioredoxin	42	37	82	18
				739,857	740,327	-	156	spoU_2	CPn0660	rRNA methylase	228	231	309	213
			740,220	740,324	741,100	-	258	mip	CPn0661	FKBP-type peptidyl-prolyl cis-trans isomerase	21	2	66	10
				741,169	742,923	-	584	aspS	CPn0662	aspartyl-tRNA synthetase	11	0	25	0
	744,357			742,898	744,190	-	430	hisS	CPn0663	histidyl-tRNA synthetase	24	4	33	3
				744,554	744,757	-	67	CPn0664	CPn0664	hypothetical protein	1	1	8	2
	744,948			745,001	746,368	+	455	uhpC	CPn0665	hexosphosphate transport	30	1	131	15
				746,388	750,110	+	1240	dnaE	CPn0666	DNA polymerase III subunit alpha	3	0	33	0
	751,072			750,174	751,058	-	294	CPn0667	CPn0667	hypothetical protein	27	19	63	10
	751,181			751,209	752,165	+	318	CPn0668	CPn0668	hypothetical protein	12	5	45	4
				752,179	752,778	+	199	CPn0669	CPn0669	hypothetical protein	9	3	59	8
S	752,485			752,765	753,199	+	144	rsbW	CPn0670	sigma regulatory factor-histidine kinase	4	0	23	0
	753,644		752,884	753,202	753,630	-	142	CPn0671	CPn0671	hypothetical protein	42	10	113	28
AS	753,478		755,258	753,741	755,051	+	436	dacF	CPn0672	D-Ala-D-Ala carboxypeptidase	11	2	15	0
				755,287	755,466	+	59	CPn0673	CPn0673	hypothetical protein	0	0	0	0
				755,574	756,668	-	364	fmu	CPn0674	RNA methyltransferase	46	35	44	15
				756,765	757,919	-	384	CPn0675	CPn0675	hypothetical protein	122	63	175	34
				758,048	759,217	-	389	CPn0676	CPn0676	hypothetical protein	510	572	674	208
				759,253	760,401	-	382	CPn0677	CPn0677	hypothetical protein	440	276	721	140
	761,349			760,679	761,320	-	213	CPn0678	CPn0678	hypothetical protein	335	258	396	124
	762,960		761,551	761,722	762,930	-	402	pgk	CPn0679	phosphoglycerate kinase	11	0	80	8
				762,968	764,248	-	426	ygo4	CPn0680	phosphate permease	4	0	33	2
	764,944			764,255	764,929	-	224	CPn0681	CPn0681	hypothetical protein	12	1	70	11
	764,961			764,984	765,958	+	324	dppD	CPn0682	peptide ABC transporter ATPase	5	1	9	0
				765,948	766,922	+	324	dppF_2	CPn0683	peptide ABC transporter ATPase	2	0	8	2
IG	767,056		767,090			+		pCpn58	pCpn58	putative sRNA, intergenic between CPn0683 and CPn0684	9	41	18	41
	768,161	767,646		767,178	768,038	-	286	parB	CPn0684	chromosome partitioning protein	9	2	38	13
				768,068	768,220	+	50	CPn0685	CPn0685	hypothetical protein	0	0	0	0
				768,173	768,361	-	62	CPn0686	CPn0686	hypothetical protein	0	0	1	0
	768,435			768,564	769,217	+	217	CPn0687	CPn0687	hypothetical protein	12	0	62	5
				769,382	770,140	+	252	CPn0688	CPn0688	hypothetical protein	7	0	8	0
				770,184	771,404	-	406	yfhO_1	CPn0689	NifS-related aminotransferase	4	0	14	1
				771,433	772,680	-	415	CPn0690	CPn0690	ABC transporter membrane protein	9	2	25	3
				772,682	773,452	-	256	abcX	CPn0691	ABC transporter ATPase	4	0	17	6
	774,970			773,458	774,912	-	484	CPn0692	CPn0692	cysteine desulfurase activator complex subunit SufB	17	9	52	17
IG	775,135		775,010			-		pCpn59	pCpn59	putative sRNA, intergenic between CPn0692 and CPn0693	1330	504	2761	1048

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				775,237	776,256	-	339	CPn0693	CPn0693	hypothetical protein	7	1	5	4
	779,681			776,327	779,599	-	1090	pbp2	CPn0694	PBP2-transglycolase/transpeptidase	6	4	36	17
	779,949	780,050	781,476	780,216	781,385	+	389	ompA	CPn0695	major Outer membrane protein	793	142	2402	412
	781,600		781,680	781,610	781,680	+	71	Gly tRNA	CPnt23	Gly tRNA	416	959	573	1592
				781,769	782,602	+	277	rpsB	CPn0696	30S ribosomal protein S2	19	0	27	3
				782,602	783,450	+	282	tsf	CPn0697	elongation factor Ts	7	2	31	1
				783,458	784,204	+	248	pyrH	CPn0698	uridylate kinase	7	0	21	0
				784,182	784,724	+	180	frr	CPn0699	ribosome recycling factor	5	0	16	2
IG	784,727		784,788			+		ihtA	pCpn60	putative sRNA, intergenic between CPn0699 and CPnt24, ihtA	32	20	34	21
	784,817		784,898	784,822	784,896	+	75	Glu tRNA	CPnt24	Glu tRNA	518	1573	635	2479
	784,915		784,994	784,922	784,994	+	73	Lys tRNA	CPnt25	Lys tRNA	492	1382	715	3003
				785,097	785,612	+	171	CPn0700	CPn0700	hypothetical protein	5	0	29	2
				785,599	786,675	+	358	karG	CPn0701	ATP:guanido phosphotransferase	4	0	25	0
	786,786		786,708	786,708	786,780	-	73	Thr tRNA	CPnt26	Thr tRNA	281	716	298	1115
IG	786,909		786,802			-		pCpn61	pCpn61	putative sRNA, intergenic between CPnt26 and CPn0702	149	252	456	365
S	789,768			786,926	789,685	-	919	yscC	CPn0702	Yop proteins translocation protein C/general secretion pathway protein	40	6	358	39
				789,682	791,190	-	502	CPn0703	CPn0703	S/T protein kinase	6	2	83	9
				791,206	792,321	-	371	fliN	CPn0704	type III secretion system protein	19	4	128	24
				792,331	793,173	-	280	CPn0705	CPn0705	hypothetical protein	17	1	67	4
				793,177	793,683	-	168	CPn0706	CPn0706	hypothetical protein	3	0	51	0
				793,701	795,029	-	442	yscN	CPn0707	type III secretion system ATPase	43	19	159	40
				795,031	795,705	-	224	CPn0708	CPn0708	hypothetical protein	21	3	41	9
				795,739	796,188	-	149	CPn0709	CPn0709	hypothetical protein	13	1	62	5
				796,207	796,461	-	84	CPn0710	CPn0710	hypothetical protein	6	2	74	0
				796,483	796,731	-	82	CPn0711	CPn0711	hypothetical protein	15	1	52	7
				796,778	799,315	-	845	CPn0712	CPn0712	FHA domain-containing protein	48	20	295	18
	800,030			799,329	799,721	-	130	CPn0713	CPn0713	hypothetical protein	43	7	220	18
				800,088	801,107	-	339	hemA	CPn0714	glutamyl-tRNA reductase	9	0	23	2
	801,418			801,657	803,465	+	602	gyrB_2	CPn0715	DNA topoisomerase IV subunit B	8	1	53	4
			804,954	803,469	804,905	+	478	gyrA_2	CPn0716	DNA topoisomerase IV subunit A	2	0	12	0
	804,983			805,010	805,309	+	99	CPn0717	CPn0717	hypothetical protein	7	5	65	24
				805,309	805,629	+	106	CPn0718	CPn0718	hypothetical protein	0	1	3	0
	805,894			805,916	806,893	+	325	sfhB	CPn0719	pseudouridine synthase	20	6	29	4
	806,940		807,418	807,003	807,239	+	78	CPn0720	CPn0720	hypothetical protein	103	56	177	23
	807,417		807,341	807,341	807,413	-	73	Arg tRNA	CPnt27	Arg tRNA	0	0	0	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	807,643			807,683	808,492	+	269	kdsA	CPn0721	2-dehydro-3-deoxyphosphooctonate aldolase	37	16	38	6
				808,489	808,977	+	162	CPn0722	CPn0722	hypothetical protein	10	4	9	0
				808,984	809,706	+	240	yhbG	CPn0723	ABC transporter ATPase	15	4	12	2
	810,554		809,572	809,703	810,527	-	274	CPn0724	CPn0724	hypothetical protein	8	8	35	7
	810,824			810,584	810,811	-	75	CPn0725	CPn0725	hypothetical protein	23	10	30	12
	813,406			810,877	813,372	-	831	CPn0726	CPn0726	hypothetical protein	249	123	319	83
			816,307	813,577	816,195	+	872	CPn0727	CPn0727	hypothetical protein	99	132	150	47
	818,506		816,447	816,522	818,477	-	651	CPn0728	CPn0728	hypothetical protein	428	363	437	208
S	818,360		818,173			-		pCpn62	pCpn62	putative sRNA, sense to CPn0728	52	45	33	11
S	818,441		818,386			-		pCpn63	pCpn63	putative sRNA, sense to CPn0728	50	88	50	92
			818,529	818,589	819,857	-	422	CPn0729	CPn0729	hypothetical protein	4	3	14	4
	821,805			819,960	821,603	-	547	mviN	CPn0730	integral membrane protein	4	4	21	14
				821,587	821,763	+	58	CPn0731	CPn0731	hypothetical protein	0	0	0	0
	821,903		823,072	822,098	822,979	+	293	nfo	CPn0732	endonuclease IV	6	4	32	2
	823,872		823,035	823,098	823,727	-	209	rpsD	CPn0733	30S ribosomal protein S4	43	7	114	15
LL	823,944		825,408	823,944	824,918	+	324	yceA	CPn0734	hypothetical protein	27	3	61	8
	825,859		824,924	825,000	825,668	-	222	CPn0735	CPn0735	uridine kinase	9	4	28	5
AS	827,880			825,989	827,686	-	565	ygeD	CPn0736	efflux protein	17	5	35	12
AS	827,493			827,685	830,759	+	1024	recC	CPn0737	exodeoxyribonuclease V, gamma	4	0	8	0
				830,746	833,898	+	1050	recB	CPn0738	exodeoxyribonuclease V, Beta	9	10	21	9
				833,858	834,871	-	337	CPn0739	CPn0739	rod shape-determining protein MreC	2	1	16	2
				834,861	836,048	-	395	tyrB	CPn0740	aromatic amino acid aminotransferase	11	4	25	1
	836,112		836,191	836,119	836,191	+	73	Ala tRNA	CPnt28	Ala tRNA	791	1920	1199	956
	838,367			836,182	838,350	-	722	greA	CPn0741	transcript cleavage factor/unknown domain fusion protein	35	6	67	6
	838,445			838,463	838,891	+	142	CPn0742	CPn0742	hypothetical protein	71	24	152	23
				838,962	840,365	+	467	nqrA	CPn0743	Na(+)-translocating NADH-quinone reductase subunit A	85	20	158	10
	841,559		840,298	840,386	841,384	-	332	hemB	CPn0744	delta-aminolevulinic acid dehydratase	5	1	18	0
				841,739	841,903	-	54	CPn0745	CPn0745	hypothetical protein	0	0	0	0
AS	841,835			841,975	843,570	+	531	CPn0746	CPn0746	hypothetical protein	46	6	137	9
	843,652			843,674	843,743	+		CPn0747	CPn0747	-	4	1	25	3
				843,724	843,913	+		CPn0747.1	CPn0747.1	-	4	0	3	0
	843,923		843,999	843,926	843,999	+	74	Pro tRNA	CPnt29	Pro tRNA	177	531	186	986
IG	844,006		844,038			+		pCpn64	pCpn64	putative sRNA, intergenic, probably processed from CPnt029	531	20	551	21
			843,950	844,118	844,987	-	289	ispA	CPn0748	geranyl transtransferase	6	1	23	2
				845,003	845,629	-	208	glmU	CPn0749	UDP-GlcNAc pyrophosphorylase	18	1	30	3

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	846,433			845,704	846,411	-	235	tctD	CPn0750	HTH transcriptional regulatory protein + receiver Doman	9	0	74	7
	846,551		848,451	846,608	848,437	+	609	CPn0751	CPn0751	hypothetical protein	3	3	34	3
	848,576			848,604	850,085	+	493	recD_2	CPn0752	exodeoxyribonuclease V, Alpha	3	2	19	5
			850,101	850,158	851,006	-	282	CPn0753	CPn0753	hypothetical protein	38	1	48	4
	851,432			851,037	851,336	-	99	rpsT	CPn0754	30S ribosomal protein S20	23	4	49	8
AS	851,106		851,419			+		pCpn65	pCpn65	putative sRNA, antisense to CPn0754	1	1	5	1
	851,499		852,866	851,597	852,802	+	401	CPn0755	CPn0755	hypothetical protein	18	2	64	8
	852,920			852,961	854,679	+	572	rpoD	CPn0756	RNA polymerase sigma factor	37	6	132	11
				854,733	855,137	+	134	folX	CPn0757	dihydroneopterin aldolase	5	4	40	3
				855,110	856,462	+	450	folP	CPn0758	dihydropteroate synthase	5	1	18	2
				856,488	857,000	+	170	folA	CPn0759	dihydrofolate reductase	1	2	21	12
S	856,966			856,957	857,697	+	246	CPn0760	CPn0760	hypothetical protein	10	6	33	18
S	857,670		858,485	857,704	858,378	+	224	CPn0761	CPn0761	hypothetical protein	27	0	30	2
IG	858,486		858,441			-		pCpn66	pCpn66	putative sRNA, intergenic between CPn0761 and CPn0762	47	15	34	14
	859,746			858,536	859,597	-	353	recA	CPn0762	recombinase A	27	5	49	11
				859,969	860,511	-	180	ygfA	CPn0763	formyltetrahydrofolate cycloligase	4	0	4	0
				860,521	861,807	-	428	CPn0764	CPn0764	hypothetical protein	5	1	12	1
				861,798	862,382	-	194	CPn0765	CPn0765	hypothetical protein	9	0	8	0
	863,808			862,391	863,782	-	463	CPn0766	CPn0766	hypothetical protein	36	5	43	1
AS	863,729			863,884	864,180	+	98	CPn0767	CPn0767	hypothetical protein	20	2	22	0
			865,266	864,159	865,166	+	335	yohI	CPn0768	oxidoreductase	8	3	18	5
			864,790	865,118	867,733	-	871	topA	CPn0769	DNA topoisomerase I/SWI domain fusion protein	19	0	44	3
	868,280			868,340	869,134	+	264	CPn0770	CPn0770	hypothetical protein	4	0	18	2
				869,141	870,463	-	440	rpoN	CPn0771	RNA polymerase factor sigma-54	4	0	30	3
				870,466	872,385	-	639	uvrD	CPn0772	DNA helicase	7	0	32	6
	871,886	872,467		872,488	873,198	+	236	ung	CPn0773	uracil-DNA glycosylase	5	0	14	1
				873,195	873,428	+	77	CPn0774	CPn0774	hypothetical protein	0	0	3	0
	874,120		873,365	873,411	874,031	-	206	yggV	CPn0775	putative deoxyribonucleotide triphosphate pyrophosphatase	3	0	15	3
AS	874,005			874,246	875,490	+	414	CPn0776	CPn0776	hypothetical protein	7	0	15	2
	875,552			875,601	877,181	+	526	groEL	CPn0777	chaperonin GroEL	11	1	134	6
	877,389		877,473	877,400	877,473	+	74	Arg tRNA	CPnt30	Arg tRNA	613	2160	860	2156
			878,456	877,505	878,095	+	196	ahpC	CPn0778	Thio-specific antioxidant (TSA) peroxidase	45	3	75	7
	878,504		877,812	878,092	878,481	-	129	CPn0779	CPn0779	hypothetical protein	0	0	11	1
				878,588	879,205	-	205	amiB	CPn0780	N-acetylmuramoyl-L-Ala amidase	13	0	19	0
				879,195	879,773	-	192	pal	CPn0781	peptidoglycan-associated lipoprotein	5	0	33	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
S	881,108			879,770	881,065	-	431	tolB	CPn0782	translocation protein TolB	8	2	62	7
				881,097	881,885	-	262	CPn0783	CPn0783	hypothetical protein	1	0	5	2
				881,889	882,296	-	135	exbD	CPn0784	biopolymer transport protein	0	1	6	0
	883,029			882,293	882,991	-	232	exbB	CPn0785	proton channel protein TolQ	4	1	30	2
AS	882,877		885,388	883,185	885,296	+	703	dsbD	CPn0786	Thio:disulfide interchange protein	85	20	44	9
				885,619	886,404	+	261	yabD	CPn0787	PHP superfamily (urease/pyrimidinase) hydrolase	6	2	2	1
	886,514			886,542	887,435	+	297	sdhC	CPn0788	succinate dehydrogenase	6	0	50	6
				887,439	889,319	+	626	sdhA	CPn0789	succinate dehydrogenase flavoprotein subunit	15	4	49	9
			890,425	889,330	890,106	+	258	sdhB	CPn0790	succinate dehydrogenase iron-sulfur subunit	4	1	31	5
S	893,026		889,978	890,108	893,050	-	980	CPn0791	CPn0791	hypothetical protein	11	0	64	12
				893,105	894,919	-	604	CPn0792	CPn0792	hypothetical protein	4	0	34	2
	896,896			894,916	896,823	-	635	rbsU	CPn0793	sigma regulatory family protein - PP2C phosphatase (RsbW antagonist)	6	1	62	10
	897,128			897,174	898,007	+	277	CPn0794	CPn0794	hypothetical protein	4	8	32	8
				898,128	899,198	+	356	CPn0795	CPn0795	hypothetical protein	8	0	25	1
	899,255		901,421	899,301	901,343	+	680	CPn0796	CPn0796	hypothetical protein	129	109	234	198
	901,541			901,600	902,697	+	365	CPn0797	CPn0797	hypothetical protein	13	7	23	8
				902,846	903,859	+	337	CPn0798	CPn0798	hypothetical protein	6	0	13	4
			903,684	903,937	904,986	-	349	CPn0799	CPn0799	hypothetical protein	14	0	24	2
	906,654			905,246	906,532	-	428	eno	CPn0800	phosphopyruvate hydratase	21	5	54	6
				906,724	908,697	-	657	uvrB	CPn0801	excinuclease ABC subunit B	22	15	35	1
				908,706	909,740	-	344	trpS	CPn0802	tryptophanyl-tRNA synthetase	25	10	59	8
				909,749	910,303	-	184	CPn0803	CPn0803	hypothetical protein	7	8	45	4
				910,307	911,059	-	250	gp6D	CPn0804	hypothetical protein	18	15	25	5
S	911,867	911,844		911,064	911,831	-	255	minD	CPn0805	chromosome partitioning ATPase	59	60	78	44
				911,864	913,771	-	635	thrS	CPn0806	threonyl-tRNA synthetase	79	49	100	9
	913,943		914,930	913,971	914,882	+	303	CPn0807	CPn0807	hypothetical protein	22	4	16	0
			914,901	914,953	916,287	-	444	CPn0808	CPn0808	hypothetical protein	396	321	607	140
				916,304	917,785	-	493	CPn0809	CPn0809	hypothetical protein	224	224	409	163
				917,822	918,184	-	120	CPn0810	CPn0810	hypothetical protein	99	59	192	53
	918,937			918,205	918,900	-	231	lcrH_1	CPn0811	low calcium response protein H	134	112	218	72
				919,123	920,865	+	580	mutL	CPn0812	DNA mismatch repair protein	7	1	17	3
				920,870	921,937	+	355	pepP	CPn0813	aminopeptidase P	13	2	24	3
	922,026			922,107	923,360	+	417	CPn0814	CPn0814	hypothetical protein	57	11	25	5
				923,361	925,625	+	754	gspD	CPn0815	general secretion protein D	18	6	42	6
				925,615	927,105	+	496	gspE	CPn0816	genral secretion protein E	13	2	11	3
				927,115	928,290	+	391	gspF	CPn0817	general secretion protein F	3	1	12	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				928,314	928,685	+	123	CPn0818	CPn0818	putative outer membrane protein	9	5	5	0
				928,689	929,135	+	148	CPn0819	CPn0819	hypothetical protein	4	2	2	3
				929,120	929,662	+	180	CPn0820	CPn0820	hypothetical protein	3	3	8	2
				929,667	930,671	+	334	CPn0821	CPn0821	hypothetical protein	8	3	9	0
	930,719		931,296	930,756	931,232	+	158	CPn0822	CPn0822	hypothetical protein	201	25	707	61
				931,498	932,367	-	289	yscT	CPn0823	Yop proteins translocation protein T	4	1	32	2
				932,375	932,662	-	95	yscS	CPn0824	Yop proteins translocation protein S	0	0	26	1
				932,674	933,594	-	306	yscR	CPn0825	type III secretion system protein	3	0	55	6
				933,609	934,310	-	233	yscL	CPn0826	type III secretion system protein	5	3	52	2
				934,431	935,264	-	277	CPn0827	CPn0827	hypothetical protein	7	0	121	6
	936,290			935,264	936,271	-	335	yscJ	CPn0828	Yop proteins translocation lipoprotein J	23	5	266	31
				936,744	937,301	+	185	CPn0829	CPn0829	hypothetical protein	1	1	1	0
				937,444	937,962	+	172	CPn0830	CPn0830	hypothetical protein	2	0	2	0
				938,267	938,437	+	56	CPn0831	CPn0831	hypothetical protein	0	0	0	0
				938,824	939,747	-	307	lipA	CPn0832	lipoyl synthase	6	1	48	1
	941,207			939,744	941,129	-	461	lpdA	CPn0833	dihydrolipoamide dehydrogenase	13	1	82	18
	941,358		942,240	941,553	942,017	+	154	CPn0834	CPn0834	hypothetical protein	36	13	331	90
				942,042	945,689	-	1215	CPn0835	CPn0835	SWI/SNF family helicase_1	9	1	46	0
	946,991			945,719	946,879	-	386	brnQ	CPn0836	branched chain amino acid ABC transporter	12	5	33	3
				947,142	947,771	-	209	nth	CPn0837	endonuclease III	0	0	3	9
				947,778	949,106	-	442	trmE	CPn0838	tRNA modification GTPase TrmE	5	1	8	0
	949,237			949,257	950,162	+	301	psdD	CPn0839	phosphatidylserine decarboxylase	6	1	23	0
	950,197			950,222	951,547	+	441	CPn0840	CPn0840	hypothetical protein	48	4	158	8
			954,678	951,731	954,643	+	970	secA	CPn0841	preprotein translocase subunit SecA	58	15	196	38
				954,706	954,883	-	177	CPn0842	CPn0842	-	0	0	0	0
S	955,064			954,990	955,191	-	201	Cpn0843	Cpn0843	-	14	6	25	2
				955,267	956,730	-	487	engA	CPn0844	GTP-binding protein EngA	15	1	21	3
S	958,289			956,847	958,079	-	410	pcnB_1	CPn0845	Poly A polymerase	39	3	104	23
				958,109	959,374	-	421	clpX	CPn0846	ATP-dependent protease ATP-binding subunit ClpX	23	2	43	4
	960,117			959,384	959,995	-	203	clpP	CPn0847	ATP-dependent Clp protease proteolytic subunit	18	9	105	17
				960,174	961,502	-	442	tig	CPn0848	trigger factor	18	1	15	0
	961,615		961,536	961,536	961,607	-	72	Gly tRNA	CPnt31	Gly tRNA	1436	2342	1205	2721
	961,738	961,755		961,788	965,288	+	1166	CPn0849	CPn0849	SWI/SNF family helicase_2	47	10	312	55
				965,293	966,393	+	366	mreB	CPn0850	rod shape-determining protein MreB	21	1	141	1
				966,396	968,198	+	600	pckA	CPn0851	phosphoenolpyruvate carboxykinase	13	0	36	5
				968,316	970,616	+	766	CPn0852	CPn0852	hypothetical protein	29	7	101	8
				970,637	971,806	+	389	CPn0853	CPn0853	hypothetical protein	7	4	27	6

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	972,868		971,542	971,803	972,837	-	344	porB	CPn0854	Outer membrane protein B	35	3	148	17
				972,991	973,995	-	334	gpsA	CPn0855	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	13	1	16	0
				973,992	975,377	-	461	CPn0856	CPn0856	UDP-glucose pyrophosphorylase	7	1	52	1
				975,389	975,757	-	122	CPn0857	CPn0857	hypothetical protein	1	0	12	1
				975,754	977,055	-	433	fliI	CPn0858	type III secretion system ATPase	9	0	32	1
				977,052	977,588	-	178	CPn0859	CPn0859	hypothetical protein	4	1	24	0
	978,677			977,605	978,630	-	341	fliF	CPn0860	type III secretion system protein	7	0	32	9
				978,922	979,722	-	266	CPn0861	CPn0861	NifU-related protein	5	0	17	0
				979,719	980,873	-	384	yfhO_2	CPn0862	NifS-related aminotransferase	9	1	44	3
	981,540			980,828	981,514	-	228	pgmA	CPn0863	phosphoglyceromutase	11	1	33	10
	981,654			981,670	982,377	+	235	yjbC	CPn0864	pseudouridine synthase	15	2	45	5
AS			983,199	982,418	982,945	+	175	CPn0865	CPn0865	hypothetical protein	2	4	25	6
	983,521		982,878	982,913	983,491	-	192	birA	CPn0866	biotin--protein ligase	2	1	9	2
				983,423	984,670	+	415	rodA	CPn0867	rod shape protein	1	0	12	1
				984,667	986,643	-	658	zntA	CPn0868	metal transport P-type ATPase	3	0	13	1
S	987,830			986,655	987,401	-	248	CPn0869	CPn0869	hypothetical protein	8	2	25	3
	988,761			987,445	988,728	-	427	serS	CPn0870	seryl-tRNA synthetase	5	0	14	1
	988,738			988,772	989,902	+	376	ribD	CPn0871	riboflavin deaminase	5	0	1	3
	989,935			989,963	991,219	+	418	ribA/ribB	CPn0872	bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II protein	8	2	26	6
			991,740	991,233	991,697	+	154	ribH	CPn0873	6,7-dimethyl-8-ribityllumazine synthase	0	0	0	0
	993,186		991,167	991,746	993,107	-	453	CPn0874	CPn0874	hypothetical protein	34	12	80	18
	993,343			993,372	994,025	+	217	CPn0875	CPn0875	hypothetical protein	95	37	211	26
				994,144	995,520	+	458	dagA_2	CPn0876	D-alanine/glycine permease	198	124	173	66
			996,078	995,533	995,985	+	150	ybcL	CPn0877	hypothetical protein	105	88	133	57
				995,989	996,654	-	221	CPn0878	CPn0878	SET domain protein	3	0	7	1
				996,642	997,439	-	265	yycJ	CPn0879	metal dependent hydrolase	11	1	32	3
	999,901			997,441	999,861	-	806	ftsK	CPn0880	cell division protein FtsK	36	7	190	20
	1,000,039		999,948	999,949	1,000,022	-	74	His tRNA	CPnt32	His tRNA	132	593	152	1029
	1,000,372			1,000,564	1,002,115	+	1552	16SRNA	CPnr01	16S ribosomal RNA	25447	2266	27402	4393
			1,005,361	1,002,415	1,005,278	+	2864	23SRNA	CPnr02	23S ribosomal RNA	40601	4966	28707	10253
	1,005,389		1,005,578	1,005,393	1,005,509	+	117	5SRNA	CPnr03	5S ribosomal RNA	9676	15997	12325	42094
				1,005,667	1,006,212	+	181	CPn0881	CPn0881	hypothetical protein	0	0	1	0
				1,006,268	1,007,407	+	379	CPn0882	CPn0882	hypothetical protein	1	0	1	0
	1,008,900		1,007,149	1,007,570	1,008,865	-	431	nqr6	CPn0883	Na(+)-translocating NADH-quinone reductase subunit F	17	2	62	10
	1,009,385		1,008,958	1,009,006	1,009,359	-	117	yajC	CPn0884	preprotein translocase subunit YajC	31	15	70	27
	1,010,763			1,009,430	1,010,635	-	401	ygcA	CPn0885	rRNA methyltransferse	13	1	26	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	1,011,322		1,010,848	1,010,905	1,011,276	-	123	hctA	CPn0886	histone-like developmental protein	117	36	357	43
	1,011,680			1,011,692	1,014,160	+	822	CPn0887	CPn0887	CHLTR phosphoprotein	1	2	27	2
				1,014,116	1,015,423	-	435	hemG	CPn0888	protoporphyrinogen oxidase	0	0	8	0
				1,015,459	1,016,835	-	458	hemN_2	CPn0889	coproporphyrinogen III oxidase	1	1	13	0
				1,016,816	1,017,805	-	329	hemE	CPn0890	uroporphyrinogen decarboxylase	0	0	5	0
				1,017,816	1,021,073	-	1085	mfd	CPn0891	transcription-repair coupling	3	0	30	7
				1,021,043	1,023,661	-	872	alaS	CPn0892	alanyl-tRNA synthetase	16	0	24	1
	1,023,863		1,025,991	1,023,894	1,025,891	+	665	tktB_1	CPn0893	transketolase	8	1	51	2
	1,026,791		1,025,691	1,025,885	1,026,766	-	293	amn	CPn0894	AMP nucleosidase	1	2	25	5
	1,026,961			1,026,988	1,027,560	+	190	efp_2	CPn0895	elongation factor P	5	3	34	1
S	1,027,604			1,027,595	1,027,825	+	76	CPn0896	CPn0896	hypothetical protein	24	13	206	134
				1,027,850	1,028,737	-	295	CPn0897	CPn0897	phosphohydrolase	1	0	2	1
				1,028,901	1,030,460	-	519	groEL_3	CPn0898	heat shock protein-60	5	0	26	2
	1,030,619		1,030,533	1,030,533	1,030,603	-	71	Cys tRNA	CPnt33	Cys tRNA	547	616	756	1012
	1,030,761	1,030,805		1,030,875	1,032,218	+	447	murF	CPn0899	Muramoyl-DAP ligase	16	0	98	19
				1,032,235	1,033,284	+	349	mraY	CPn0900	phospho-N-acetylmuramoyl-pentapeptide-transferase	5	0	28	0
S	1,034,026			1,033,287	1,034,540	+	417	murD	CPn0901	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	17	8	125	41
AS	1,034,404		1,034,262			-		pCpn67	pCpn67	putative sRNA, antisense to CPn0901	4	3	13	1
				1,034,543	1,035,244	+	233	nlpD	CPn0902	muramidase	9	0	49	2
				1,035,263	1,036,420	+	385	ftsW	CPn0903	cell division protein FtsW	5	2	23	2
				1,036,326	1,037,399	+	357	murG	CPn0904	undecaprenyldiphospho-muramoylpentapeptide beta-N- acetylglucosaminyltransferase	4	0	11	0
			1,040,089	1,037,409	1,039,838	+	809	murC/ddlA	CPn0905	bifunctional D-alanyl-alanine synthetase A/UDP-N-acetylmuramate--L-alanine ligase	14	0	69	1
	1,040,403		1,039,840	1,039,912	1,040,340	-	142	CPn0906	CPn0906	hypothetical protein	3	0	16	2
S	1,040,944			1,040,442	1,040,780	-	112	cutA	CPn0907	periplasmic divalent cation tolerance protein	5	0	11	4
				1,040,777	1,041,589	-	270	CPn0908	CPn0908	hypothetical protein	0	0	15	1
	1,041,606			1,041,637	1,041,969	+	110	rsbV_2	CPn0909	sigma regulatory factor	5	3	23	5
			1,043,107	1,041,979	1,043,007	+	342	miaA	CPn0910	tRNA delta(2)-isopentenylpyrophosphate transferase	23	2	40	1
	1,044,069		1,042,767	1,042,982	1,044,043	-	353	CPn0911	CPn0911	hypothetical protein	7	5	17	2
	1,044,085			1,044,129	1,045,763	+	544	CPn0912	CPn0912	hypothetical protein	37	2	162	13
	1,045,741			1,045,760	1,045,948	+	62	CPn0913	CPn0913	hypothetical protein	0	0	9	1
				1,045,999	1,046,400	+	133	CPn0914	CPn0914	hypothetical protein	3	0	10	0
	1,046,455			1,046,461	1,046,820	+	119	ybeB	CPn0915	iojap superfamily protein	9	2	41	9
				1,046,837	1,048,087	+	416	fabF	CPn0916	3-oxoacyl-(acyl carrier protein) synthase II	13	0	43	6
			1,048,599	1,048,090	1,048,542	+	150	CPn0917	CPn0917	hydrolase/phosphatase-like protein	9	0	25	2
	1,049,266		1,048,547	1,048,576	1,049,223	-	215	ppa	CPn0918	inorganic pyrophosphatase	7	3	4	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	1,049,353		1,050,504	1,049,378	1,050,433	+	351	ldh	CPn0919	leucine dehydrogenase	26	11	61	13
	1,051,427		1,050,343	1,050,428	1,051,405	-	325	cysQ	CPn0920	sulfite synthesis/biphosphate phosphatase	18	5	40	2
	1,051,471	1,052,006		1,051,535	1,052,296	+	253	CPn0921	CPn0921	snglycerol-3-P acyltransferase	16	1	65	20
				1,052,314	1,053,930	+	538	aas	CPn0922	long chain fatty acid--[acyl-carrier-protein] ligase	7	6	46	10
			1,055,227	1,053,984	1,055,096	+	370	bioF_1	CPn0923	8-amino-7-oxononanoate synthase	7	3	29	2
S	1,057,294		1,054,869	1,055,025	1,057,274	-	749	priA	CPn0924	primosome assembly protein PriA	18	3	46	7
	1,057,978			1,057,223	1,057,900	-	225	CPn0925	CPn0925	hypothetical protein	14	4	37	4
	1,058,013			1,058,060	1,058,560	+	166	CPn0926	CPn0926	thioredoxin disulfide isomerase	17	1	62	12
	1,059,831			1,058,667	1,059,809	-	380	CPn0927	CPn0927	CHLPS 43 kDa protein homolog_2	4	3	5	2
				1,059,881	1,061,008	-	375	CPn0928	CPn0928	CHLPS 43 kDa protein homolog_3	3	2	0	3
	1,062,743			1,061,183	1,062,292	-	369	CPn0929	CPn0929	CHLPS 43 kDa protein homolog_4	36	7	38	7
	1,062,786		1,063,564	1,062,857	1,063,333	+	158	CPn0930	CPn0930	hypothetical protein	24	45	76	15
	1,063,903		1,065,900	1,064,138	1,065,721	+	527	lysS	CPn0931	lysyl-tRNA synthetase	47	23	118	63
				1,065,718	1,067,142	-	474	cysS	CPn0932	cysteinyl-tRNA synthetase	13	1	13	1
	1,067,436	1,067,447	1,068,776	1,067,535	1,068,581	+	348	CPn0933	CPn0933	disulfide bond isomerase	85	15	240	79
				1,068,523	1,068,942	-	139	rnpA	CPn0934	ribonuclease P	23	20	44	25
	1,069,148			1,068,954	1,069,091	-	45	rpmH	CPn0935	50S ribosomal protein L34	23	9	52	7
	1,069,283			1,069,336	1,069,473	+	45	rpmJ	CPn0936	50S ribosomal protein L36	12	4	60	1
			1,069,881	1,069,496	1,069,801	+	101	rpsN	CPn0937	30S ribosomal protein S14	15	2	32	5
				1,069,846	1,070,322	-	158	CPn0938	CPn0938	hypothetical protein	4	0	14	0
	1,070,661		1,071,300	1,070,728	1,071,198	+	156	CPn0939	CPn0939	hypothetical protein	63	8	77	17
				1,071,201	1,073,012	-	603	uvrC	CPn0940	excinuclease ABC subunit C	10	3	13	10
	1,075,602			1,073,015	1,075,501	-	828	mutS	CPn0941	DNA mismatch repair protein MutS	16	5	36	10
				1,075,985	1,077,757	+	590	dnaG	CPn0942	DNA primase	7	2	6	1
	1,077,949			1,077,978	1,078,241	+	87	CPn0943	CPn0943	hypothetical protein	39	3	8	0
				1,078,512	1,079,000	+	162	CPn0944	CPn0944	hypothetical protein	11	0	2	0
			1,079,724	1,079,070	1,079,663	+	197	CPn0945	CPn0945	hypothetical protein	34	7	17	9
				1,079,742	1,082,786	-	1014	glyQ	CPn0946	glycyl tRNA synthetase	8	0	19	1
	1,083,101		1,084,197	1,083,442	1,084,062	+	206	pgsA_2	CPn0947	glycerol-3-P phosphatydyltransferase	12	0	63	7
	1,085,493		1,083,965	1,084,044	1,085,474	-	476	glgA	CPn0948	glycogen synthase	1	2	16	2
	1,085,598		1,085,677	1,085,605	1,085,676	+	72	Gln tRNA	CPnt34	Gln tRNA	469	1346	366	2304
	1,085,802			1,085,929	1,086,486	+	185	ctc	CPn0949	50S ribosomal protein L25/general stress protein Ctc	39	11	93	14
				1,086,488	1,087,030	+	180	pth	CPn0950	peptidyl-tRNA hydrolase	9	0	3	3
				1,087,122	1,087,460	+	112	rpsF	CPn0951	30S ribosomal protein S6	15	1	21	4
				1,087,478	1,087,726	+	82	rpsR	CPn0952	30S ribosomal protein S18	4	1	9	1
			1,089,306	1,087,742	1,088,251	+	169	rplI	CPn0953	50S ribosomal protein L9	721	116	579	212
				1,088,285	1,088,711	+		ychB	CPn0954	-	0	0	0	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
S	1,088,328		1,088,465			+		pCpn68	pCpn68	putative sRNA, sense to CPn0954	721	116	579	212
				1,088,611	1,089,178	+		CPn0955	CPn0955	-	0	0	0	0
AS	1,088,976		1,088,927			-		pCpn69	pCpn69	putative sRNA, antisense to CPn0955	213	267	738	458
AS	1,089,368		1,088,973	1089132	1,089,350	-	72	CPn0955.1	CPn0955.1	ORF 72 aa, no homologies	10	7	53	21
				1,089,560	1,090,912	+	450	CPn0956	CPn0956	hypothetical protein	4	5	36	2
				1,090,960	1,093,788	-	942	ide	CPn0957	insulinase family/protease III	5	0	34	4
				1,093,790	1,094,785	-	331	plsB	CPn0958	glycerol-3-P acyltransferase	2	0	10	1
	1,096,453			1,094,796	1,096,343	-	515	cafE	CPn0959	axial filament protein/ribonuclease E	19	9	60	112
	1,096,600			1,096,764	1,097,105	+	113	CPn0960	CPn0960	hypothetical protein	13	1	48	2
				1,097,118	1,097,300	+	60	rpmF	CPn0961	50S ribosomal protein L32	1	0	1	2
				1,097,316	1,098,278	+	320	plsX	CPn0962	putative glycerol-3-phosphate acyltransferase PlsX	11	4	21	8
	1,098,360		1,103,283	1,098,398	1,103,227	+	1609	pmp_21	CPn0963	putative outer membrane protein D	108	9	279	16
				1,103,298	1,104,758	-	486	CPn0964	CPn0964	hypothetical protein	1	1	22	4
				1,104,922	1,106,736	-	604	lpxB	CPn0965	lipid-A-disaccharide synthase	12	0	24	1
	1,108,110			1,106,745	1,108,037	-	430	pcnB_2	CPn0966	Poly A polymerase	12	1	67	2
	1,108,335			1,108,512	1,109,888	+	458	glmM	CPn0967	phosphoglucosamine mutase	9	1	11	9
				1,109,895	1,111,724	+	609	glmS	CPn0968	glucosamine--fructose-6-phosphate aminotransferase	26	3	62	11
S	1,111,398			1,111,812	1,113,002	+	396	tyrP_1	CPn0969	tyrosine transporter	46	7	206	34
S	1,113,002		1,113,441	1,113,053	1,113,373	+	106	yccA_1	CPn0969.1	transport permease (truncated)	93	21	363	38
AS	1,113,175		1,113,137			-		pCpn70	pCpn70	putative sRNA, antisense to CPn0969.1	4	5	3	3
				1,113,461	1,114,651	+	396	tyrP_2	CPn0970	tyrosine transporter	11	2	65	11
S	1,114,451	1,114,649	1,115,482	1,114,702	1,115,418	+	238	yccA_2	CPn0971	transport permease	131	102	338	141
	1,116,325		1,115,300	1,115,427	1,116,299	-	290	ftsY	CPn0972	cell division protein FtsY	7	4	61	3
	1,116,351			1,116,370	1,117,530	+	386	sucC	CPn0973	succinyl-CoA synthetase subunit beta	7	3	58	10
			1,118,733	1,117,544	1,118,425	+	293	sucD	CPn0974	succinyl-CoA synthetase subunit alpha	7	1	30	3
				1,119,104	1,119,640	+	178	CPn0975	CPn0975	hypothetical protein	1	0	0	1
				1,120,082	1,121,188	+	368	CPn0976	CPn0976	hypothetical protein	3	3	2	0
AS	1,121,326		1,120,974			-		pCpn71	pCpn71	putative sRNA, antisense to CPn0976	1	0	5	1
				1,121,371	1,122,405	+	344	CPn0977	CPn0977	hypothetical protein	10	3	14	4
				1,122,665	1,123,696	+	343	CPn0978	CPn0978	hypothetical protein	7	1	11	1
	1,123,891		1,125,731	1,123,980	1,125,446	+	488	htrA	CPn0979	DO Serine protease	37	11	93	13
				1,125,501	1,126,982	-	493	CPn0980	CPn0980	hypothetical protein	3	0	17	0
AS	1,126,927			1,127,031	1,129,955	+	974	CPn0981	CPn0981	zinc metalloprotease	19	6	70	4
				1,129,959	1,131,194	-	411	yigN	CPn0982	hypothetical protein	4	3	28	8
	1,132,023			1,131,203	1,132,000	-	265	pssA	CPn0983	glycerol-Serine phosphatidyltransferase	19	2	82	7
	1,132,183			1,132,379	1,135,513	+	1044	nrdA	CPn0984	ribonucleotide-diphosphate reductase subunit alpha	43	4	142	28
				1,135,534	1,136,574	+	346	nrdB	CPn0985	ribonucleotide-diphosphate reductase subunit beta	22	2	58	4

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
	1,136,694			1,136,724	1,137,398	+	224	trmB	CPn0986	tRNA (guanine-N(7)-)-methyltransferase	8	2	7	0
AS	1,136,800		1,136,420			-		pCpn72	pCpn72	putative sRNA, antisense to CPn0985 and CPn0986	0	1	10	4
	1,137,468		1,137,382	1,137,389	1,137,462	-	74	Val tRNA	CPnt35	Val tRNA	178	332	208	526
				1,137,516	1,138,118	+	200	ytgB_2	CPn0987	YtgB-like rRNA methylase	0	0	7	0
			1,137,912	1,138,072	1,138,986	-	304	murB	CPn0988	UDP-N-acetylenolpyruvoylglucosamine reductase	1	4	13	0
	1,139,570			1,139,013	1,139,495	-	160	nusB	CPn0989	transcription antitermination protein NusB	12	4	17	1
	1,139,637	1,139,785		1,139,883	1,140,443	+	186	infC	CPn0990	translation initiation factor IF-3	68	11	131	8
				1,140,421	1,140,615	+	64	rpmI	CPn0991	50S ribosomal protein L35	7	0	19	1
				1,140,634	1,140,999	+	121	rplT	CPn0992	50S ribosomal protein L20	18	0	26	0
				1,141,014	1,142,033	+	339	pheS	CPn0993	phenylalanyl-tRNA synthetase subunit alpha	21	7	76	48
S	1,141,938		1,142,118	1,142,034	1,142,117	+	84	Ser tRNA	CPnt36	Ser tRNA	508	935	824	1348
				1,142,398	1,144,443	+	681	CPn0994	CPn0994	hypothetical protein	2	0	9	2
				1,144,412	1,145,512	-	366	CPn0995	CPn0995	hypothetical protein	3	0	23	0
AS	1,146,729			1,145,516	1,146,589	-	357	CPn0996	CPn0996	hypothetical protein	16	1	29	6
				1,146,708	1,147,667	+	319	mesJ	CPn0997	PP-loop superfamily ATPase	3	2	6	2
	1,147,818		1,150,653	1,147,855	1,150,587	+	910	ftsH	CPn0998	ATP-dependent zinc protease	82	19	171	39
			1,150,692	1,150,763	1,152,847	-	694	pnp	CPn0999	polynucleotide phosphorylase/polyadenylase	45	16	80	9
	1,153,188			1,152,888	1,153,157	-	89	rpsO	CPn1000	30S ribosomal protein S15	17	13	122	13
				1,153,405	1,153,872	+	155	yfhC	CPn1001	cytosine deaminase	1	0	2	0
				1,153,862	1,154,092	+	76	CPn1002	CPn1002	hypothetical protein	3	0	3	4
				1,154,089	1,154,796	-	235	CPn1003	CPn1003	hypothetical protein	45	105	44	54
				1,154,876	1,155,397	-	173	CPn1004	CPn1004	hypothetical protein	26	31	49	8
				1,155,412	1,155,933	-	173	CPn1005	CPn1005	hypothetical protein	21	8	19	2
				1,155,987	1,156,472	-	161	CPn1006	CPn1006	hypothetical protein	10	9	11	1
AS	1,156,295	1,156,597		1,156,689	1,156,910	+	73	CPn1007	CPn1007	hypothetical protein	24	2	26	1
				1,156,928	1,158,226	+	432	CPn1008	CPn1008	hypothetical protein	22	3	24	2
S	1,159,103			1,158,183	1,159,058	-	291	map	CPn1009	methionine aminopeptidase	3	0	17	4
				1,159,064	1,159,672	-	202	CPn1010	CPn1010	hypothetical protein	0	1	4	0
				1,159,899	1,160,306	-	135	CPn1011	CPn1011	hypothetical protein	3	3	2	1
				1,160,418	1,162,193	-	591	yzeB	CPn1012	ABC transporter permease	4	2	8	0
AS	1,161,960		1,163,809	1,162,245	1,163,627	+	460	fumC	CPn1013	fumarate hydratase	11	2	44	5
AS	1,163,407		1,163,214			-		pCpn73	pCpn73	putative sRNA, antisense to CPn1013	4	0	14	6
	1,165,454		1,163,475	1,163,729	1,165,426	-	565	ychM	CPn1014	Sulfate transporter	17	9	84	9
	1,165,613			1,165,634	1,166,896	+	420	CPn1015	CPn1015	hypothetical protein	6	2	27	7
	1,167,005			1,167,042	1,168,901	+	619	CPn1016	CPn1016	hypothetical protein	71	11	168	19
				1,169,006	1,169,938	+	310	ispH	CPn1017	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	13	0	30	7
				1,169,898	1,170,632	+	244	CPn1018	CPn1018	hypothetical protein	7	0	13	1

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
				1,170,635	1,172,128	-	497	CPn1019	CPn1019	hypothetical protein	2	1	15	3
				1,172,147	1,173,679	-	510	CPn1020	CPn1020	hypothetical protein	1	1	9	3
				1,173,695	1,174,213	-	172	lcrH_2	CPn1021	low calcium response protein H	1	2	5	0
	1,175,700			1,174,213	1,175,673	-	486	CPn1022	CPn1022	hypothetical protein	8	1	33	3
	1,175,832		1,175,944	1,175,863	1,175,944	+	82	Leu tRNA	CPnt37	Leu tRNA	130	286	112	489
	1,176,005			1,176,035	1,176,334	+	99	CPn1023	CPn1023	hypothetical protein	61	20	122	16
	1,177,252			1,176,331	1,177,236	-	301	xerD	CPn1024	site-specific tyrosine recombinase XerD	10	2	4	1
				1,177,302	1,178,882	+	526	pgi	CPn1025	glucose-6-phosphate isomerase	140	89	220	43
				1,178,997	1,179,140	+	47	CPn1026	CPn1026	late transcription unit A protein	44	13	27	4
				1,179,175	1,180,758	+	527	CPn1027	CPn1027	hypothetical protein	92	46	71	11
				1,181,016	1,182,002	+	328	mdhC	CPn1028	malate dehydrogenase	40	14	77	16
				1,182,008	1,182,847	+	279	CPn1029	CPn1029	hypothetical protein	29	17	89	9
	1,183,924			1,182,840	1,183,886	-	348	CPn1030	CPn1030	D-amino acid dehyrogenase	7	4	35	7
				1,184,095	1,185,552	-	485	arcD	CPn1031	arginine/ornithine antiporter	15	6	133	30
				1,185,563	1,186,150	-	195	CPn1032	CPn1032	hypothetical protein	15	2	84	6
	1,187,533			1,186,184	1,187,500	-	438	CPn1033	CPn1033	hypothetical protein	17	2	102	10
				1,187,729	1,188,517	-	262	CPn1034	CPn1034	hypothetical protein	1	0	7	1
				1,188,567	1,190,000	-	477	aroDE	CPn1035	bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase protein	1	0	4	0
				1,189,981	1,191,135	-	384	aroB	CPn1036	3-dehydroquinate synthase	4	2	13	3
		1,191,347		1,191,120	1,192,199	-	359	aroC	CPn1037	chorismate synthase	1	0	7	2
				1,192,196	1,192,726	-	176	aroL	CPn1038	shikimate kinase	1	0	9	0
				1,192,662	1,193,999	-	445	aroA	CPn1039	3-phosphoshikimate 1-carboxyvinyltransferase	7	2	10	0
				1,194,070	1,194,741	-	223	CPn1040	CPn1040	hypothetical protein	2	6	11	7
				1,194,723	1,195,994	-	423	bioA	CPn1041	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	1	0	10	2
				1,195,931	1,196,590	-	219	bioD	CPn1042	dethiobiotin synthetase	7	0	7	1
				1,196,569	1,197,717	-	382	bioF_2	CPn1043	oxononanoate synthase	16	1	14	6
				1,197,696	1,198,691	-	331	bioB	CPn1044	biotin synthase	9	0	11	4
				1,198,898	1,199,590	-	230	CPn1045	CPn1045	hypothetical protein	2	0	5	0
				1,199,587	1,200,675	-	362	phhA	CPn1046	aromatic amino acid hydroxylase	4	0	13	0
AS	1,200,541			1,200,552	1,201,346	+	264	dapB	CPn1047	dihydrodipicolinate reductase	0	0	21	2
	1,201,356			1,201,606	1,202,607	+	333	asd	CPn1048	aspartate-semialdehyde dehydrogenase	11	6	101	23
				1,202,595	1,203,917	+	440	lysC	CPn1049	aspartate kinase	4	1	22	2
				1,203,926	1,204,801	+	291	dapA	CPn1050	dihydrodipicolinate synthase	2	1	12	2
				1,204,962	1,205,273	+	103	CPn1051	CPn1051	hypothetical protein	1	0	1	1
				1,205,417	1,206,172	+	251	CPn1052	CPn1052	hypothetical protein	2	0	0	0
				1,206,153	1,206,704	+	183	CPn1053	CPn1053	hypothetical protein	2	0	1	0

	TSS	2nd TSS	TES	Location		Strand	Length	Gene	Synonym	Product	EB -	EB +	RB -	RB +
AS	1,206,657		1,206,332			-		pCpn74	pCpn74	putative sRNA, antisense to CPn1053	0	1	12	2
				1,207,034	1,209,469	+	811	CPn1054	CPn1054	hypothetical protein	0	0	0	0
	1,209,540			1,209,694	1,210,524	+	276	CPn1055	CPn1055	hypothetical protein	1	1	3	1
				1,210,527	1,211,231	+	234	CPn1056	CPn1056	hypothetical protein	0	0	1	0
	1,211,470			1,211,497	1,213,599	+	700	CPn1057	CPn1057	hypothetical protein	21	18	64	15
	1,213,715			1,213,748	1,214,839	+	363	CPn1058	CPn1058	hypothetical protein	28	4	84	7
				1,214,848	1,215,681	+	277	ksgA	CPn1059	dimethyladenosine transferase	9	1	12	3
				1,215,724	1,217,658	-	644	tktB_2	CPn1060	1-deoxy-D-xylulose-5-phosphate synthase	11	0	34	4
S	1,218,258			1,217,663	1,217,920	-	85	CPn1061	CPn1061	hypothetical protein	27	3	117	22
	1,219,849			1,218,156	1,219,820	-	554	xseA	CPn1062	exodeoxyribonuclease VII large subunit	3	3	25	16
	1,219,910	1,220,437		1,219,951	1,220,715	+	254	tpiA	CPn1063	triosephosphate isomerase	6	2	14	3
				1,220,719	1,220,898	+	59	CPn1064	CPn1064	hypothetical protein	1	0	9	0
				1,220,925	1,221,095	-	56	CPn1065	CPn1065	hypothetical protein	2	0	1	0
	1,221,016			1,221,135	1,221,491	+	118	secG	CPn1066	preprotein translocase subunit SecG	10	0	9	1
	1,221,562			1,221,735	1,222,295	+	186	def	CPn1067	peptide deformylase	12	0	24	2
				1,222,362	1,223,258	-	298	rnhB_2	CPn1068	ribonuclease HIII	0	4	10	0
	1,223,425		1,224,326	1,223,513	1,223,944	+	143	yfgA	CPn1069	HTH transcriptional regulator	2	4	51	6
	1,225,683			1,224,141	1,225,511	-	456	CPn1070	CPn1070	hypothetical protein	17	2	20	2
				1,225,882	1,227,324	-	480	CPn1071	CPn1071	hypothetical protein	0	1	0	0
	1,227,947			1,227,969	1,228,838	+	289	CPn1072	CPn1072	hypothetical protein	6	1	24	1
			1,229,875	1,229,011	1,229,835	+	274	CPn1073	CPn1073	putative outer membrane protein	5	0	25	0
	1,230,230		1,229,941	1,229,945	1,230,028	-	84	Ser tRNA	CPnt38	Ser tRNA	6746	37742	11226	64682

9.2 Abbreviations

ATP ..... adenosine triphosphate
BSA ..... bovine serum albumin
C. pneumoniae ..... *Chlamydia (Chlamydophila) pneumoniae*
C. trachomatis ..... *Chlamydia trachomatis*
cDNA ..... complementary DNA
CDS ..... coding sequence
cIAP ..... cellular inhibitor of apoptosis protein
dRNA-seq ..... differential cDNA deep sequencing
DNA ..... deoxyribonucleic acid
dNTP ..... desoxynucleosid triphosphate
DTT ..... dithiothreitol
EB ..... elementary body
EDTA ..... ethylene-diamine-tetraacetic acid
HEPES ..... 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IGR ..... intergenic region
kDa ..... kilo-Dalton
LGV ..... Lymphogranuloma venerum
MOI ..... multiplicity of infection
mRNA ..... messenger RNA
MSM ..... men who have sex with men
ncRNA ..... non-coding RNA
nt ..... nucleotide(s)
ORF ..... open reading frame
PAGE ..... polyacrylamide gel electrophoresis
PB ..... persistence body
PBS ..... phosphate buffered saline
PCR ..... polymerase chain reaction
PNK ..... polynucleotide kinase
PVDF ..... polyvinylidene fluoride
RB ..... reticulate body
rcf ..... relative centrifugal force
RNA ..... ribonucleic acid
RNAi ..... RNA interference
RNAP ..... DNA-dependent RNA polymerase
rpm ..... revolutions per minute
SDS ..... sodium dodecyl sulphate

siRNA	small interfering RNA
sRNA	small regulatory RNA
T3SS	type three secretion system
TBST	tris-buffered saline with Tween
TEMED	tetramethylethylenediamine
TEX	Terminator Exonuclease
TSS	transcriptional start site
U	enzyme Unit
UTR	untranslated region
UV	ultra violet
v/v	volume per volume
w/v	weight per volume

9.3 List of Figures

Figure 1: Taxonomy of the order Chlamydiales. 7
Figure 2: The life cycle of Chlamydiales. 9
Figure 3: Electron micrographs of purified *C. trachomatis* EB and RB. 37
Figure 4: Characterization of the *C. trachomatis* cDNA libraries. 38
Figure 5: Visualization of pyrosequencing data of the *C. trachomatis inc*-operon. 39
Figure 6: Length distribution of *C. trachomatis* 5'UTRs. 40
Figure 7: Antisense RNAs derived from long overlapping 3' UTRs in *C. trachomatis*. 41
Figure 8: Validation of putative small RNAs of *C. trachomatis* by Northern Analysis. 42
Figure 9: Sequencing read distribution of CTLon_0332. 47
Figure 10: Transcriptome of the *C. trachomatis* cryptic plasmid pL2. 49
Figure 11: Electron micrographs of purified *C. pneumoniae* RB and EB. 50
Figure 12: Classification of sequencing reads of *C. pneumoniae*. 51
Figure 13: Sequence read length distribution of *C. pneumoniae*. 52
Figure 14: 5'leader length distribution and leaderless mRNAs of *C. pneumoniae*. 54
Figure 15: Genes with alternative transcription start sites in *C. pneumoniae* EB and RB. 56
Figure 16: Novel non-coding sRNAs of *C. pneumoniae*. 59
Figure 17: An abundant putative small RNA encoded in Cpn0332. 60
Figure 18: Density gradient fractionation of *C. pneumoniae* lysates. 61
Figure 19: Two operons encoding genes of the type three secretion system (T3SS) in *C. pneumoniae*. 63
Figure 20: Promoter motifs of *C. pneumoniae*. 68
Figure 21: Most prominent sequence motifs of the -35 and -10 boxes of *C. pneumoniae*. 69
Figure 22: The conserved sRNA *ihtA*. 77
Figure 23: TSS of the major outer membrane protein gene *ompA*. 78

9.4 List of Tables

Table 1: Numbers of sequence reads of the four *C. trachomatis* cDNA libraries. 38
Table 2: Genomic Locations of putative non-coding RNAs and corresponding Northern probes..... 43
Table 3: Putative targets of validated *C. trachomatis* sRNAs. 44
Table 4: List of differentially expressed genes in *C trachomatis* EB and RB. 45
Table 5: Numbers of sequence reads of the four *C. pneumoniae* cDNA libraries. 52
Table 6: *C. pneumoniae* genes that have to be re-annotated. 55
Table 7: Northern blot results of putative sRNAs of *C. pneumoniae*..... 58
Table 8: Operon numbers and lengths of *C. pneumoniae.* 64
Table 9: List of protein coding genes more abundant in *C. pneumoniae* EB. 66
Table 10: Transcription start sites and sequencing read numbers of *C. trachomatis* genes. 91
Table 11: Transcription start sites and sequencing read numbers of *C. pneumoniae* genes. 117