Pathogens and Disease, 2023, **81**, 1–18
DOI: 10.1093/femspd/ftad009
Advance access publication date: 16 May 2023
Research Article

Comparative transcriptomics and genomics from continuous axenic media growth identifies *Coxiella burnetii* intracellular survival strategies

Archana Yadav[1], Melissa N. Brewer[1,2], Mostafa S. Elshahed[1], Edward I. Shaw[1,3,*]

[1]Department of Microbiology and Molecular Genetics, Oklahoma State University,, 74078 Stillwater, OK, United States
[2]Biological Sciences, Southeastern Oklahoma State University, 74078 Durant, OK, United States
[3]Department of Biomedical Sciences, Philadelphia College of Osteopathic Medicine, 74078 Moultrie, GA, United States
*Corresponding author. Department of Bio-Medical Sciences, Philadelphia College of Osteopathic Medicine (PCOM), 31768, Moultrie, GA, United States. Tel: 229+-668-3188; E-mail: ed.shaw@pcom.edu
Editor: [Hayley Newton]

Abstract

Coxiella burnetii (Cb) is an obligate intracellular pathogen in nature and the causative agent of acute Q fever as well as chronic diseases. In an effort to identify genes and proteins crucial to their normal intracellular growth lifestyle, we applied a 'reverse evolution' approach where the avirulent Nine Mile Phase II strain of Cb was grown for 67 passages in chemically defined ACCM-D media and gene expression patterns and genome integrity from various passages was compared to passage number one following intracellular growth. Transcriptomic analysis identified a marked downregulation of the structural components of the type 4B secretion system (T4BSS), the general secretory (Sec) pathway, as well as 14 out of 118 previously identified genes encoding effector proteins. Additional downregulated pathogenicity determinants genes included several chaperones, LPS, and peptidoglycan biosynthesis. A general marked downregulation of central metabolic pathways was also observed, which was balanced by a marked upregulation of genes encoding transporters. This pattern reflected the richness of the media and diminishing anabolic, and ATP-generation needs. Finally, genomic sequencing and comparative genomic analysis demonstrated an extremely low level of mutation across passages, despite the observed Cb gene expression changes following acclimation to axenic media.

Keywords: *Coxiella burnetii*, reverse evolution, type 4B secretion system, effector proteins, intracellular survival

Introduction

Coxiella burnetii (Cb), the causative agent of acute and chronic Q fever (Maurin and Raoult 1999, McQuiston et al. 2002, Arricau-Bouvery and Rodolakis 2005, Raoult et al. 2005, Miller et al. 2006, van Schaik et al. 2013), is an obligate intracellular pathogen that infects macrophages, and successfully propagates in a parasitophorous vacuole termed the *Coxiella* containing vacuole (CCV) (Heinzen et al. 1999, Voth and Heinzen 2007, Rudolf Toman 2012 et al. 2012). Cb has evolved multiple strategies to tolerate and thrive in the CCV, in spite of the prevailing low pH (≈4.5), low O_2 content, oxygen radicals, and high level of degradative host factors such as acid hydrolases and defensins (Hackstadt and Williams 1981, Heinzen et al. 1999, Brennan et al. 2004, Omsland et al. 2009). Such remarkable ability has been the subject of a wide range of studies that employed a plethora of biochemical, genetic, imaging, and –omics-based approaches. Further, Cb employs a type 4B secretion system (T4BSS) to deliver effector proteins into the host throughout infection (Heinzen et al. 1999, Voth and Heinzen 2007, Voth and Heinzen 2009, Rudolf Toman 2012 et al. 2012, van Schaik et al. 2013). Cb effector proteins identified so far mediate a variety of biochemical activities and are known to target and modulate a broad array of host functions (Beare et al. 2011, Newton et al. 2013, van Schaik et al. 2013, Beare et al. 2014, Newton et al. 2014, Larson et al. 2015, Larson and Heinzen 2017, Crabill et al. 2018). Prior studies have employed bioinformatic tools (Chen et al. 2010), transposon mutagenesis (Beare et al. 2011, Carey et al. 2011, Weber et al. 2013, Martinez et al. 2014, Crabill et al. 2018), microscopic localization studies (Howe et al. 2003, Chen et al. 2010, Morgan et al. 2010, Voth et al. 2011), and cloning and infectivity testing to identify and characterize effector proteins. In addition, *Legionella pneumophila*, a close genetic neighbour of Cb with a very similar T4BSS (Segal and Shuman 1999, Sexton and Vogel 2002, Nagai and Kubori 2011), is known to use T4BSS-effector protein duality to infect its natural host cell, the amoeba. *Legionella pneumophila* has been extensively used as a proxy to identify putative effector proteins and propose molecular pathogenesis mechanisms in Cb (Zamboni et al. 2003, Zusman et al. 2003, Vogel 2004, Segal et al. 2005, Pan et al. 2008). Indeed, research on *L. pneumophila* has identified the structural features of the T4BSS, the nature of effector proteins secreted through the system, and possible function of some of these effectors.

Growth of Cb in an axenic media was first reported in 2009 using the undefined Acidified Citrate Cysteine Media (ACCM) (Omsland et al. 2009). Increased replication rates in the somewhat more defined ACCM-2 medium soon followed in 2011 (Omsland et al. 2011). Subsequently, a nutritionally fully defined media (ACCM-D) with an even greater replication rate and physiologic parallels to intracellular bacteria was developed (Sanchez et al. 2018). Growing Cb in axenic media is opening new venues for investigating mechanisms of Cb molecular pathogenesis (Beare et al. 2012,

Received 6 February 2023; **revised** 17 April 2023; **accepted** 15 May 2023

Rudolf Toman et al. 2012, Beare and Heinzen 2014, Martinez et al. 2015, Sandoz et al. 2016, Beare et al. 2018, Crabill et al. 2018). Theoretically, when grown in axenic media, the expression of genes required for intracellular survival and host cell manipulation is no longer required for Cb viability. Therefore, continuous maintenance and passaging the bacterium for extended periods of times under axenic conditions could potentially remove the powerful selective pressure exerted by the host cell, thus potentially minimize/silence expression in such genes. As such, we evaluated whether transcriptomic analysis of gene expression patterns as well as genomic identification of mutation and gene loss patterns in axenic grown versus Cb cultures derived from intracellular growth could be employed for identifying putative involvement of specific genes, as well as identification of novel genes necessary for Cb pathogenesis and survival in an intracellular environment. Similarly, continuous passaging could also lead to the propagation of mutations, DNA fragment losses, and rearrangements in genes/loci associated with intracellular survival, pathogenesis, and host cell manipulation. Such patterns could be regarded as 'reverse evolution', i.e. the opposite of the natural evolution trajectory of Cb from a free-living ancestor to an obligate intracellular pathogen. Specifically, we hypothesized that: (1) changes in gene expression within the first few passages upon transition from intracellular to axenic media growth would be observed, and such differences would be more pronounced in genes involved in subverting and coopting host metabolism, as well as genes enabling general adaptation to physiological conditions prevalent in its intracellular vacuolar environment, and (2) Cb could acquire and accumulate DNA mutations upon transition from intracellular to axenic media growth after repetitive passages since certain bacterial genes/proteins are no longer required for successful growth.

In this study, we transitioned Cb Nine Mile phase II from cell cultures into axenic defined media ACCM-D and subcultured it in long-term successive passages. We conducted transcriptomic and genomic sequencing on replicate samples at different time points (passages) to document temporal changes in gene expression patterns, as well as DNA mutations associated with adaptation to an axenic extracellular lifestyle.

Materials and methods

Microorganism and growth conditions

Coxiella burnetii (Cb) avirulent strain Nine Mile phase II (NMII), clone 4 (RSA439) was cultivated in rabbit epithelial RK13 cells (CCL-37; American Type Culture Collection) grown in Dulbecco's modified Eagle medium DMEM (ThermoFisher Scientific) supplemented with 5% fetal bovine serum (FBS) in T75 culture flasks. This method of collecting cells was adapted from Coleman et al. (2004). Briefly, the infected cell line was split into multiple nonvented and capped T150 culture flasks that were incubated at 37°C in 5% CO_2 for a week until confluent growth was observed. These flasks were then screwed tightly shut and left at room temperature for 2 weeks to induce cells to switch to the small cell variant (SCV) form. The cells were pelleted by ultracentrifugation (12 000 × *g*, 15 minutes) in 250 ml Nalgene round bottom tubes, scrapped off the round bottom tubes by using sterile 1× phosphate buffered saline (PBS), and then lysed by using Dounce homogenize. The lysed cells in PBS were then spun via centrifugation using Oakridge tubes in an ultracentrifuge at 12 000 × *g* for 15 minutes. The SCV pellets obtained were stored in SPG freezer media (0.7 M sucrose, 3.7 mM KH_2PO_4, 6.0 mM K_2HPO_4, 0.15 M KCl, 5.0 mM glutamic acid, pH 7.4) at −80°C.

Axenic growth in defined ACCM-D media

Cb cultures propagated intracellularly in rabbit epithelial RK13 cells were used to inoculate ACCM-D media (Sunrise Science Products, San Diego, CA). Approximately 10^6 genome equivalents per ml was used as an inoculum [determined using the RT-PCR procedure as previously described (Brennan and Samuel 2003)]. Cultures were grown in a T25 cell culture flasks at 5% O_2, 5% CO_2, and 37°C in a trigas incubator (Panasonic, MCO-170ML) for 7 days. Subsequent passages were achieved via a 1:1000 (6 μl into 6 ml) inoculum into freshly prepared ACCM-D media and incubation for 7 days. Axenically grown Cb cultures were routinely (every five passages) subjected to contamination check by inoculation into LB broth medium incubated under microaerophilic (5% O_2, 5% CO_2) conditions, LB broth medium incubated aerobically at 37°C, as well as ACCM-D medium incubated aerobically at 37°C.

Measuring growth and host cell infectivity

To determine the infectivity of axenic or intracellularly grown Cb; HeLa cells (CCL-2; American Type Culture Collection) were seeded onto 96-well culture plates at a density of 10^4 in Roswell Park Memorial Institute (RPMI) medium containing 2% FBS for 16 hours. Cb cultures grown in ACCM-D were pelleted at 12 000 × *g* at 4°C for 15 minutes. Serially passaged Cb were diluted in RPMI to normalize the number of genomes per volume, and 50 μl from various dilutions were inoculated onto the HeLa cell containing wells and centrifuged at 600 × *g* for 15 minutes at room temperature (Luedtke et al. 2017). Immediately following centrifugation, the inoculating media was replaced with 200 μl of fresh RPMI containing 2% FBS. The plates were incubated at 37°C and 5% CO_2 for 72 hours, fixed with ice-cold methanol for 10 minutes, then examined using indirect fluorescent antibody microscopy analysis as described previously (Luedtke et al. 2017). Briefly, Cb was stained using rabbit whole anti-Cb NMII antibody diluted 1:1000 in PBS containing 3% bovine serum albumin (BSA) as a blocking agent. Primary antibodies were detected using Alexa Fluor 488 labelled goat anti-rabbit IgG antibodies diluted 1:1000 in PBS containing 3% BSA (Invitrogen). Total DNA was stained using 4',6-diamidino-2-phenylindole (DAPI) diluted 1:10 000 in PBS containing 3% BSA (Molecular Probes) to illuminate host cell nuclei. The methanol fixed and stained cultures were visualized on a Nikon Eclipse TE2000-S and the number of maturing CCVs were counted and calculations performed to ascertain the number of fluorescence forming units, which indicates the infectivity of the Cb NMII in the ACCM-D samples.

Transcriptomics

RNA extraction

Cells from axenic media growth passages 1, 3, 5, 10, 12, 16, 21, 31, 42, 51, 61, and 67 were harvested for transcriptomic analysis. RNA was extracted using a combination of hot Trizol treatment (Moormeier et al. 2019) and the RNeasy Mini kit (Qiagen, Germany). Briefly, bacteria in 12 ml of ACCM-D culture (OD_{600} ∼ 0.3–0.4) were pelleted, resuspended in 700 μl of Trizol (ThermoFisher Scientific), boiled at 90°C for 10 minutes, and vortexed vigorously. Two hundred µl of Chloroform was then added, followed by centrifugation at 12 000 × *g* at 4°C for 10 minutes. After separation, 300 μl of 100% ethanol was added to the aqueous phase, which was then quickly transferred to the spin column provided in the RNeasy mini kit. On-column DNA digestion was conducted by adding 80 μl [10 μl 1 Unit/μl RNase free DNase I, (ThermoFisher Scientific) in 70 μl reaction buffer from the Master Pure Yeast RNA Purification kit, Epicenter] of DNase preparation. The

RNeasy mini kit's protocol was followed for washing and eluting RNA. RNA quality was assessed visually on a gel as well as using RNA screen tape (Agilent) and RNA integrity number (RIN) value measurements using Tapestation and Bioanalyzer systems (Agilent).

Transcriptome sequencing and assembly

RNA sequencing (RNA-Seq) was conducted on the Illumina platform, using Nextseq 500 sequencer at the Oklahoma State University Genomics and Proteomics core facility. After checking the quality of the reads using FastQC (Andrews 2010), Trimmomatic v0.38 (Bolger et al. 2014) was used to process raw reads and remove Illumina adapter sequences. HISAT2 v2.1.0 (Kim et al. 2019) was used to map the trimmed reads to the Chromosome (GenBank accession number: CP020616.1) and Plasmid (GenBank accession number: CP020617.1) of Cb NMII RSA 439. StringTie v2.1.4 (Kovaka et al. 2019) was used to assemble reads alignments into potential transcript and to generate a nonredundant set of transcripts. The Python script prepDE.py supplemented with StringTie tool was used to convert transcripts per kilobase million (TPM) and fragments per kilobase million (FKPM) to gene level raw count matrix. The raw count table was imported to DESeq2 package (Love et al. 2014) from Bioconductor in R programming language for further analysis. Comparison in expression patterns in replicates was assessed by using ' PCAtools' package in R.

Identification and analysis of differentially expressed genes (DEGs)

The overall strategy for comparative transcriptomics analysis is outlined in Fig. 1. DESeq2 was used to compute the fold change expression levels (reflected by logarithmic 2-fold expression change, i.e. L2fc) and its statistical significance (adjusted *P*-value, padj henceforth referred to as *P*-value) for every gene between passages when compared to passage one. DESeq2 tests the differential expression using negative binomial distribution and internally normalizes the counts by library size (Anders and Huber 2010). Genes with a *P*-value < .05 were labelled as significantly expressed. Only genes with TPM values > 10 in at least one passage were considered to minimize noise from minimally expressed genes. In most cases, DEGs in our temporal analysis were significantly expressed in more than one sampling point. In the few cases where differential expression was observed as a single spike in only one time point, a threshold of L2fc > 2 was considered as differentially expressed. Patterns of differential expression in DEGs was analysed and visualized by constructing plot count graphs using the function 'plotCounts' in DESeq2 package. The package 'EnhancedVolcano' was used to visualize gene expression patterns as volcano plots.

An empirical classification of differential expression patterns was implemented for grouping genes identified as exhibiting statistically significantly differential regulation (using the criteria described above). Expression levels of all DEGs were plotted as a function of time (see cartoon in Fig. 1) to identify temporal patterns. Verdicts are made based on observing the trendline in these expression graphs for individual genes. Using such patterns, DEGs were classified into (1) Early up/downregulated: Genes where upregulation/downregulation was observed early and the levels were sustained in subsequent late passages (Fig. 1). 'Early' is defined as occurring before Passage 31, while 'late' is defined as occurring after Passage 31. (2) Continuously up/downregulated: Genes where a constant/gradual increase/decrease in the magnitude of L2fc was observed throughout the sampling process (Fig. 1). (3) Late up/downregulated: Genes where differential expression was observed at or after Passage 31. (4) Variable: Genes where expression levels were significantly higher than passage one in some timepoints and significantly lower than passage one in other time points (Fig. 1).

Metabolic analysis and pathway mapping of DEGs

The subcellular protein localizations of proteins encoded by DEGs were predicted by using PSORTb (Yu et al. 2010). Transporter Classification Database (TCBD) was queried to find the putative transporter proteins. Pfam database (Mistry et al. 2020) was used to identify putative protein families for hypothetical proteins. BlastKOALA (Kanehisa et al. 2016) was used for functional annotation and assign KEGG Orthology (KO) numbers for the selected DEGs; and KEGG mapper (Kanehisa and Sato 2020) was then used to reconstruct metabolic pathway to visualize the DEGs in each pathway. The gene involvement in specific metabolic pathways were inferred from KEGG brite hierarchy file. Cluster of Orthologous genes (COGs) database (updated 2020) (Galperin et al. 2021) downloaded from NCBI, was used to classify the effector proteins into functional categories.

Genomics

DNA extraction and sequencing

Eight mls of Cb passaged cultures grown in ACCM-D for 7 days were pelleted by centrifugation at 12 000 × *g* and 4°C for 15 minutes. DNA extraction was conducted using Pure Link® Genomic DNA Kits (ThermoFisher Scientific) following the manufacturer's instructions. Sequencing was conducted at the Oklahoma State University Genomics and Proteomics core facility using Illumina's NextSeq® 500 System. DNA quality was assessed visually on a gel as well as using DNA Screentape and Bioanalyzer systems (Agilent).

Genome assembly and quality control

The KBase platform (Allen et al. 2017), which implements and integrates multiple bioinformatic tools, was used for DNA sequence data handling. Quality check was done using FastQC v0.11.5 (Andrews 2010). Trimmomatic v 0.36 (Bolger et al. 2014) was used to trim the Illumina adapter sequences. Assembly of Illumina reads to contigs was attempted using four different assemblers [Spades v3.13.0, Velvet v1.2.10, and IDBA-UD v1.1.3 and Unicycler (Davis et al. 2020)]. The quality of genome assemblies from these four assemblers were assessed using QUAST v1.4 (Gurevich et al. 2013) and the best assemblies were selected using metrices such as total length, largest N50, lesser number of contigs, and less Ns. CheckM (Parks et al. 2015) was used to assess quality and completion of genomes (Fig. 1).

Analysis of mutation frequencies

Breseq (Deatherage and Barrick 2014) was used to identify mutations/changes in the genome assemblies obtained, with Passage 1 used as a reference. The occurrence and frequency of both single nucleotide polymorphisms (SNPs) and deletion–insertion polymorphisms (DIPs) were examined (as outlined in Fig. 1). Breseq was run in polymorphism mode, which identifies the mutations occurring in a fraction of a population in addition to consensus mutations in the entire population in a sample. This allows for the visualization of the propagation of a particular mutation as a

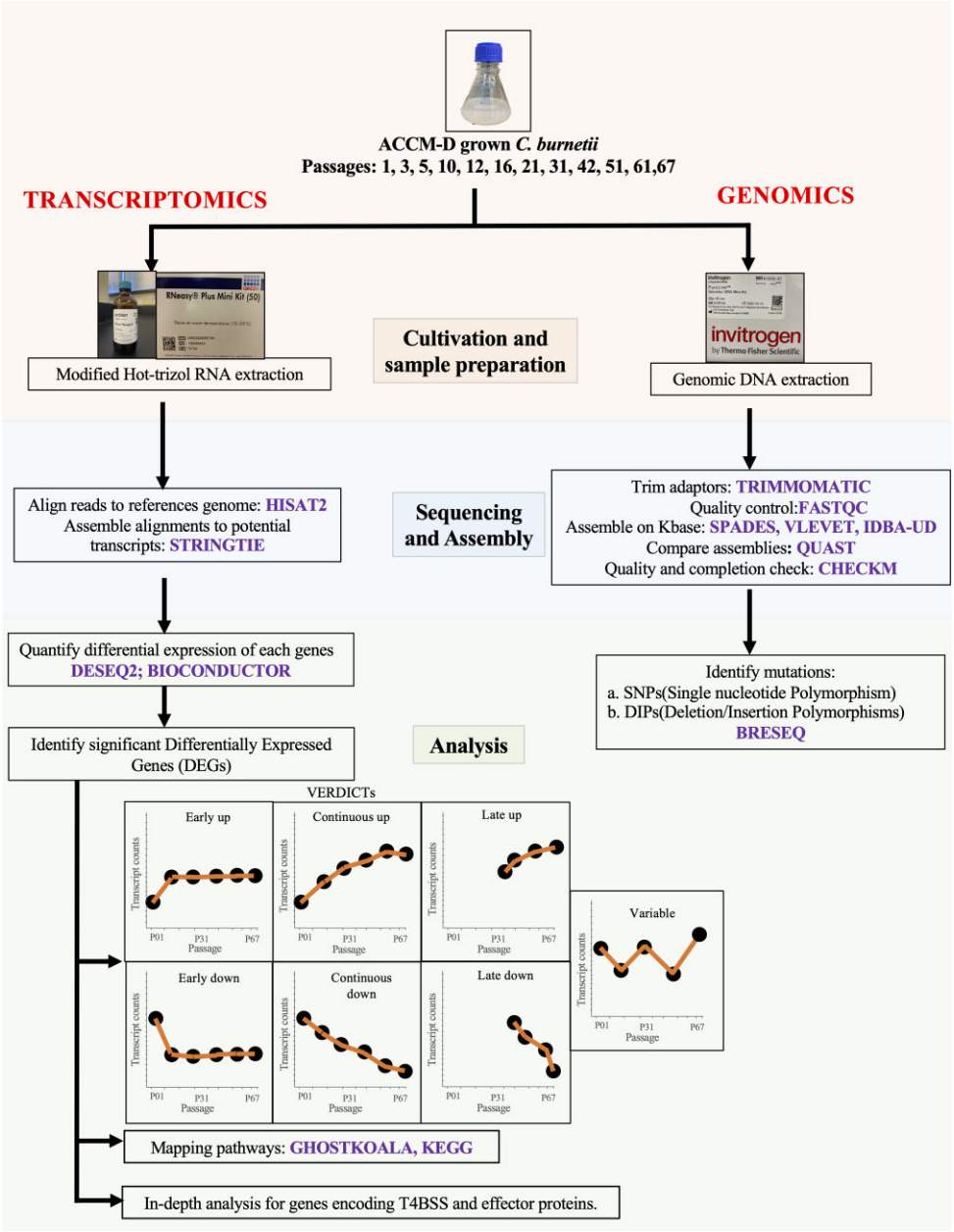


Figure 1. Flowchart representing the overall comparative transcriptomics and genomics strategy employed in this study. Cb NM II strain was grown and passaged in an axenic media for 67 times. DNA and RNA were extracted to study the changes in gene expression and genetic mutations for 12 of these passages. For transcriptomics, the reads were assembled and aligned to this strain of Cb. The expression levels of all other passages were compared to Passage 1 and differential gene expressions were designated early/late up, early/late down, continuous up/down, and variables based on their gene expression trends and using Passage 31 as midpoint (for 'early' versus 'late' differentiation) as shown in the figure. The gene expression changes in various pathways and metabolic systems were further analysed in detail. For genomics, reads from pooled replicates for 12 passages were assembled. To study and characterize the different types of genomic mutations occurring in later passages, their reads were aligned to the Passage 1 genome and the mutations (SNPs and DIPs) were identified using several bioinformatics tools.

frequency of evolved alleles and genetic diversity in the population.

Nucleotide sequences accession number

The whole-transcriptome and genome shotgun sequences were deposited in GenBank under the BioProject PRJNA796300 and BioSample accession numbers SAMN24840407–SAMN24840437 and SAMN24847762–SAMN24847773. The 31 transcriptomic assemblies were deposited in the SRA under project accession number SRX13723330–SRX13723360. Reads for 12 genomic assemblies can be found under SRA with accession SRX13726189–SRX13726200.

Results

Coxiella burnetii infectivity but not viability decreases with continuous passaging in axenic media

Following anecdotal observations, we sought to quantitatively assess whether serially passaged Cb infect cultured cells less

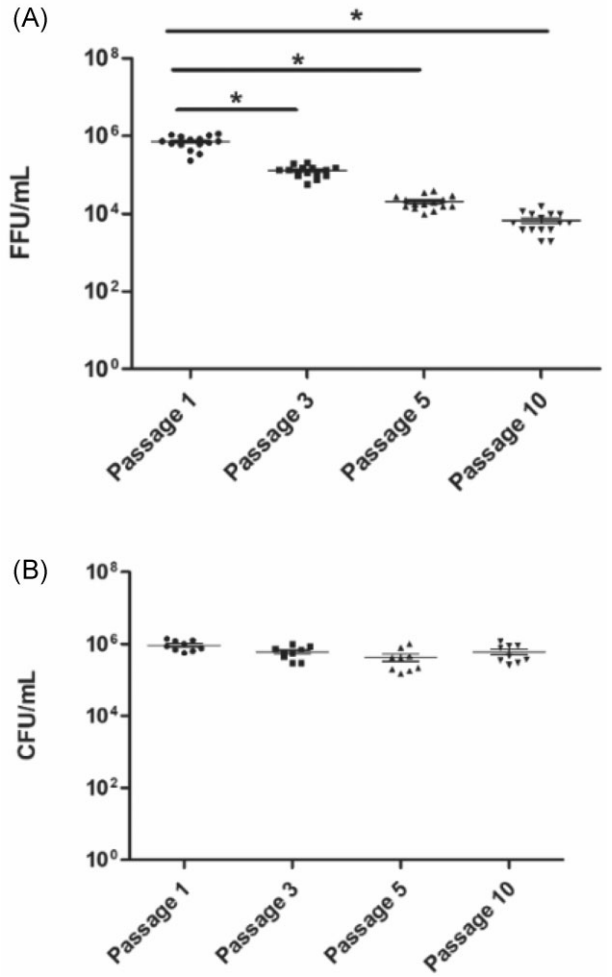


Figure 2. Infectivity and viability of Cb NMII in axenic media. Intracellular and axenic growth from three biological replicates of passaged ACCM-D growth (A) FFUs counts of infections of HeLa cells normalized by Cb genomes from passages 1, 3, 5, and 10. Significance between different passages are indicated by lines and $*P < .001$. (B) CFU enumeration of passages 1, 3, 5, and 10 Cb spread on ACCM-D plates normalized to genomes. No statistically significant difference was observed between groups.

readily than cell derived bacterial stocks. Using *C. burnetii* NMII serially passaged 1, 3, 5, and 10 times in ACCM-D, we initiated infections of Hela cells with bacterial dilutions normalized by the number of genomes in each sample. When the number of fluorescence forming units (FFU) per sample were calculated, they revealed a decrease in the number of *C. burnetii* filled vacuoles in tissue culture cells as the bacteria from subsequent passages were analyzed, respectively, resulting in a nearly two-log decrease between Passages 1 and 10 (Fig. 2A). This indicated that there were fewer bacteria per genome that were capable of initiating a typical infection following multiple passages in axenic media.

Next, we sought to determine if the decrease in infectivity of tissue culture cells was associated with a decrease in *in vitro* viability of the *C. burnetii* as measured by colony forming units on ACCM-D agar. To address this question, we plated dilutions of Passages 1, 3, 5, and 10 on ACCM-D agar plates and performed colony counts. Contrary to the decrease in infectious units (Fig. 2A), the colony counts indicated that there was no significant change in viable bacteria relative to genomes as the organism was serially passaged (Fig. 2B). This indicated that the number of live and replicative bacteria did not change during axenic growth, and therefore bacterial death was not responsible for the decrease in Cb infectivity of the cultured eukaryotic cells observed.

Transcriptional activity

To determine whether gene transcript expression levels changed in Cb during continuous axenic propagation, RNA sequencing was conducted on 12 different passages (1, 3, 5, 10, 12, 16, 21, 31, 42, 51, 61, and 67) with triplicates for all passages (except Passage 1, 3, and 5, which were conducted in duplicates). In general, a pattern where replicates from a single passage clustered together was observed, and principal component analysis indicate that replicate sampling were highly similar (Supplementary Fig. S1). A total of 162.2 Gb data were obtained, with 6.05–22.57 million reads per sample (Average = 10.14 million reads). Transcripts representing each of the 2217 genes in *C. burnetii* NMII strain (genome and plasmid) were identified in all samples, attesting to the depth of the sequencing effort conducted.

Expression level and overall pattern (early up, continuous up, late up, early down, continuous down, late down, variable) for every gene in the Cb genome is shown in Supplementary Table S1. A total of 845 genes were differentially expressed in at least one passage, with 464 upregulated and 371 downregulated (Fig. 3A) genes. The number of DEGs per passage ranged between 25 and 807 (Fig. 3B). A general pattern of an increasing number of DEGs per passage was observed through Passage 51, after which the number of DEGs dropped in Passage 61 and 67 (Fig. 3B). The ratio of upregulated to downregulated genes in each passage ranged between 0.14 (in Passage 5) and 1.26 (in Passage 3). Of the 371 downregulated genes, 249 expressed an early down pattern, 48 were continuous down, 43 were late down, and 31 were down in only one passage. Of the 464 upregulated genes, 288 were early up, 38 were continuous up, 85 were late up, and 53 were up in only one passage (Fig. 3C). Of the 845 DEGs, 81 were differentially regulated in 8–11 of the passages, 144 in 5–7 of the passages, 526 in 2–4 of the passages, and 84 in only 1 passage (Fig. 3D).

Visualization of DEGs patterns using volcano plots was used to provide an overview of the overall level of expression changes (Fig. 4). Transcript expression levels from each passage were compared to passage one. The labelled boxes within each plot analysis represents the 10 highest differentially expressed transcripts (i.e. smallest *P*-value). Visual inspections demonstrate that chaperons and T4BSS machinery proteins consistently represent an important component of highly downregulated genes in all passages. Below, we provide a more detailed assessment on differential expression patterns for various genes and pathways.

Secretory pathways are significantly downregulated in axenic growth media

The defective in organelle trafficking/intracellular multiplication (Dot/Icm) T4BSS in Cb has been shown to secrete protein effectors/pathogenic determinants into the host cell, a process required for Cb intracellular growth and pathogenesis (Voth and Heinzen 2009, Beare et al. 2011, Carey et al. 2011, Rudolf Toman et al. 2012, van Schaik et al. 2013). Interestingly, 19 out of the 24 components of the Cb T4BSS demonstrated significant differential expression (Fig. 5A, Supplementary Fig. S2a, and Supplementary Table S1). Out of these, 18 genes were downregulated and only 1 gene was upregulated (Fig. 5A and Supplementary Fig. S2a). Indeed, T4BSS encoded gene transcripts were some of the most significantly downregulated across the passages (Fig. 4).

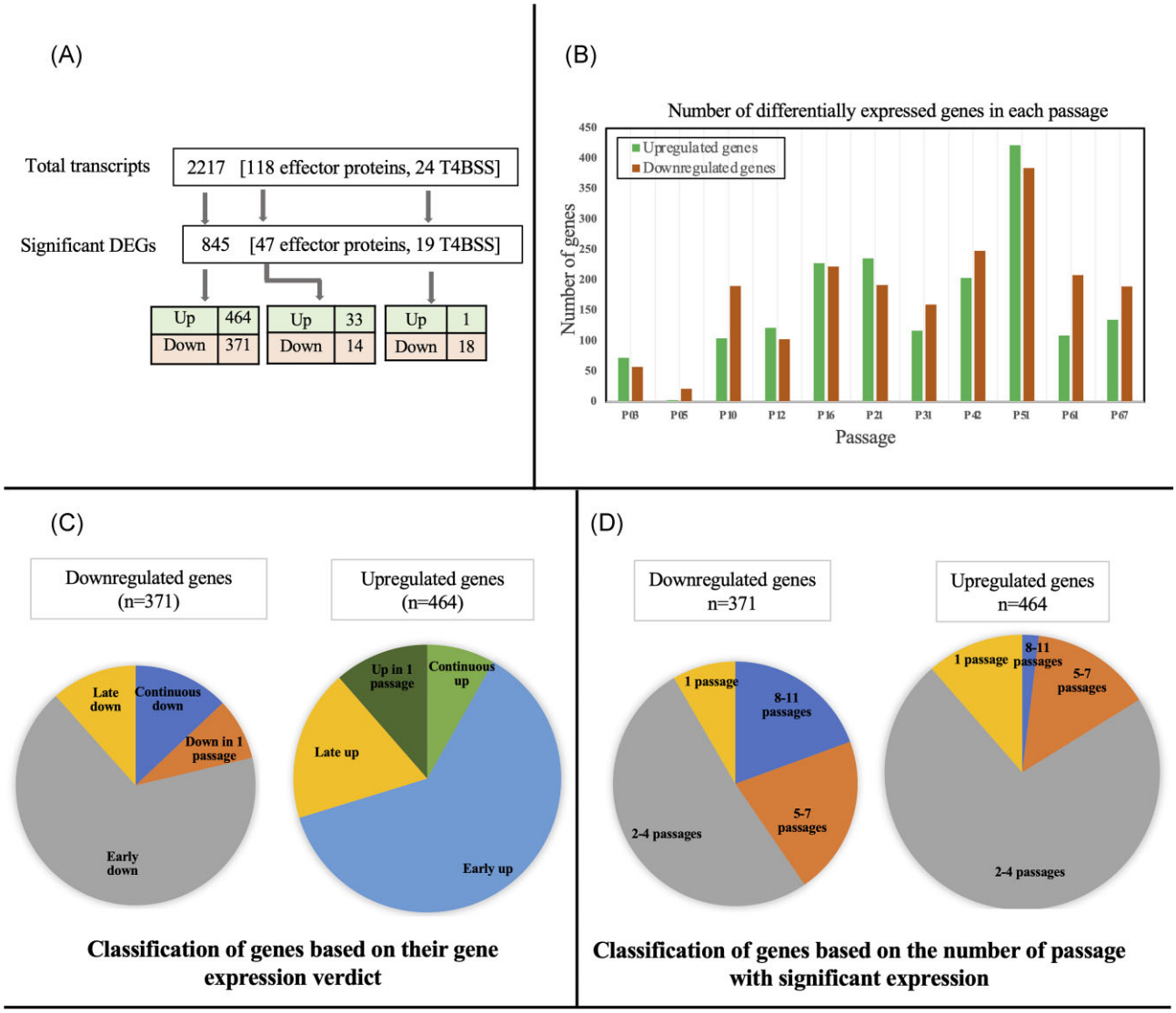


Figure 3. Overview for differential expression patterns in axenically grown Cb. (A) Summary for the number of total and DEGs, T4BSS, and effector proteins in this experiment. (B) Bar graph showing the number of upregulated (*P*-value < .05 and L2fc > 0) and downregulated (*P*-value < .05 and L2fc < 0) genes in each passage. (C) Classification of genes based on their gene expression verdicts, i.e. early up/down, late up/down, continuous up/down, up/down in one passage. (D) Classification of genes based on the number of passages showing significant expression change.

Within the T4BSS core transport complex, transcripts for genes *dotC*, *dotD*, *dotF*, and *dotG* were early downregulated, while only *dotH* indicated no significant change in gene expression in all passages. In addition, expression changes in genes encoding components of the T4BSS coupling protein complex (*dotL*, *dotM*, and *icmW*) demonstrated early down or continuous down (*dotN*), whereas *icmS* was the only component with no significant gene expression changes. Transcripts of the gene *dotB* was continuous downregulated whereas *dotA* and *icmX* were found to be early downregulated (Fig. 4 and Supplementary Fig. S2a). Besides the two main complexes, other components of the Cb T4BSS that were transcriptionally downregulated during continuous axenic media passaging includes genes *dotE*, *dotP*, *dotK*, *dotI*, *dotJ*, *icmT*, and *icmQ* (Fig. 5A). *icmF*, located in a separate locus than the majority of the T4BSS genes (Fig. 5B), was the only component that was transcriptionally upregulated in the system (Supplementary Fig. S2a).

Transcript expression of genes within additional secretory pathways in Cb were also analysed. Genes of the general secretary (Sec) pathway revealed a general trend of downregulation (Supplementary Fig. S2b and Supplementary Table S1). The Sec pathway provides a channel for polypeptide movement across the bacterial inner membrane (Green and Mecsas 2016). It is comprised of the proteins SecY, SecE, and SecG and an ATPase (SecA) that drives protein movement (Green and Mecsas 2016, Tsirigotaki et al. 2017). This pathway is known to secrete proteins from the cytosol through the cytoplasmic membrane (Mori and Ito 2001). We identified all of the Cb Sec pathway components, as shown in Fig. 5A. Transcripts for expression of the inner membrane proteins SecA, SecF, and SecE were early down and SecY and YajC were continuously down, whereas the targeting proteins SecB and SecG were down at later passages.

Expression patterns of T4BSS effector proteins previously implicated in Cb pathogenesis and intracellular survival

The differential expression in all 118 genes encoding T4BSS effector proteins previously identified in Cb through a variety of

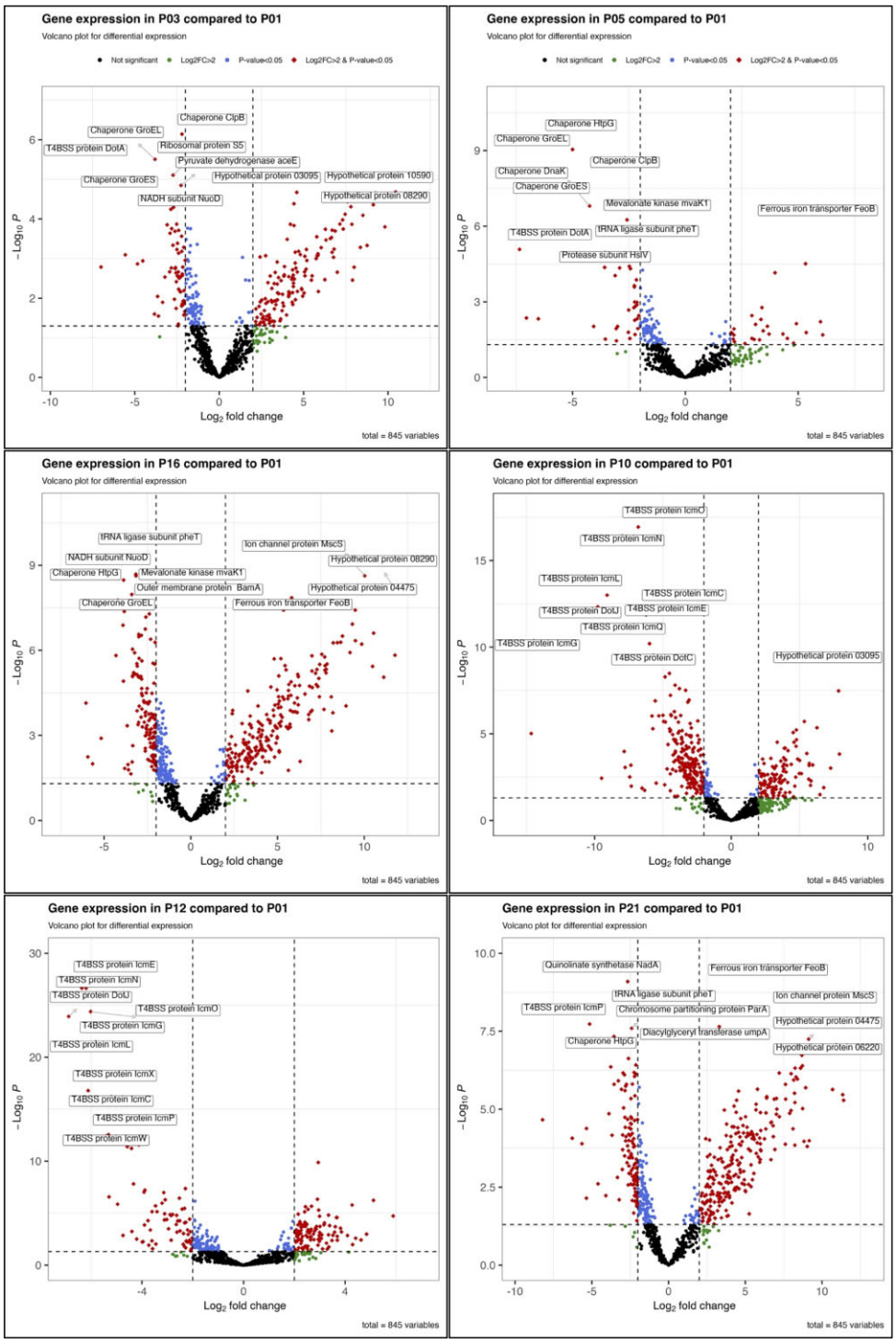


Figure 4. Volcano plots for expression patterns for 845 overall significant DEGs in different passages when compared to passage 01. The volcano plots were made using the package 'EnhancedVolcano' in R programming language. Here, for each passage, the nonsignificant DEGs would be the gene with *P*-value > .05. In these volcano plots, the black dots represent nonsignificant DEGs, green dots represent non-DEGs with L2fc > 2, blue dots represent significant DEGs (i.e. *P*-value < .05) with L2fc < 2 whereas the red diamonds represent significant DEGs with L2fc > 2. The 10 genes with the lowest *P*-values (and therefore the most significant expression changes) in all compared passages are labelled in boxes and these genes in most cases have the highest Log_2 fold change (L2fc) as shown in the figure.

effector screens (Chen et al. 2010, Carey et al. 2011, Voth et al. 2011, Weber et al. 2013, Larson et al. 2016) was examined. Forty-seven effector proteins were differentially expressed (column 'Effector proteins' in Supplementary Table S1, Fig. 3A). Interestingly, more genes were upregulated ($n = 33$) than downregulated ($n = 14$) (Fig. 3A).

The 14 genes encoding effector proteins that were transcriptionally downregulated indicated an up to 4-fold expression change, with the expression changes primarily beginning from passage 10 (Supplementary Fig. S3 and Table 1). These genes fell into COG functional groups of signal transduction mechanisms (*ankG*, *ankK*, and *ankD*), carbohydrate transport and metabolism

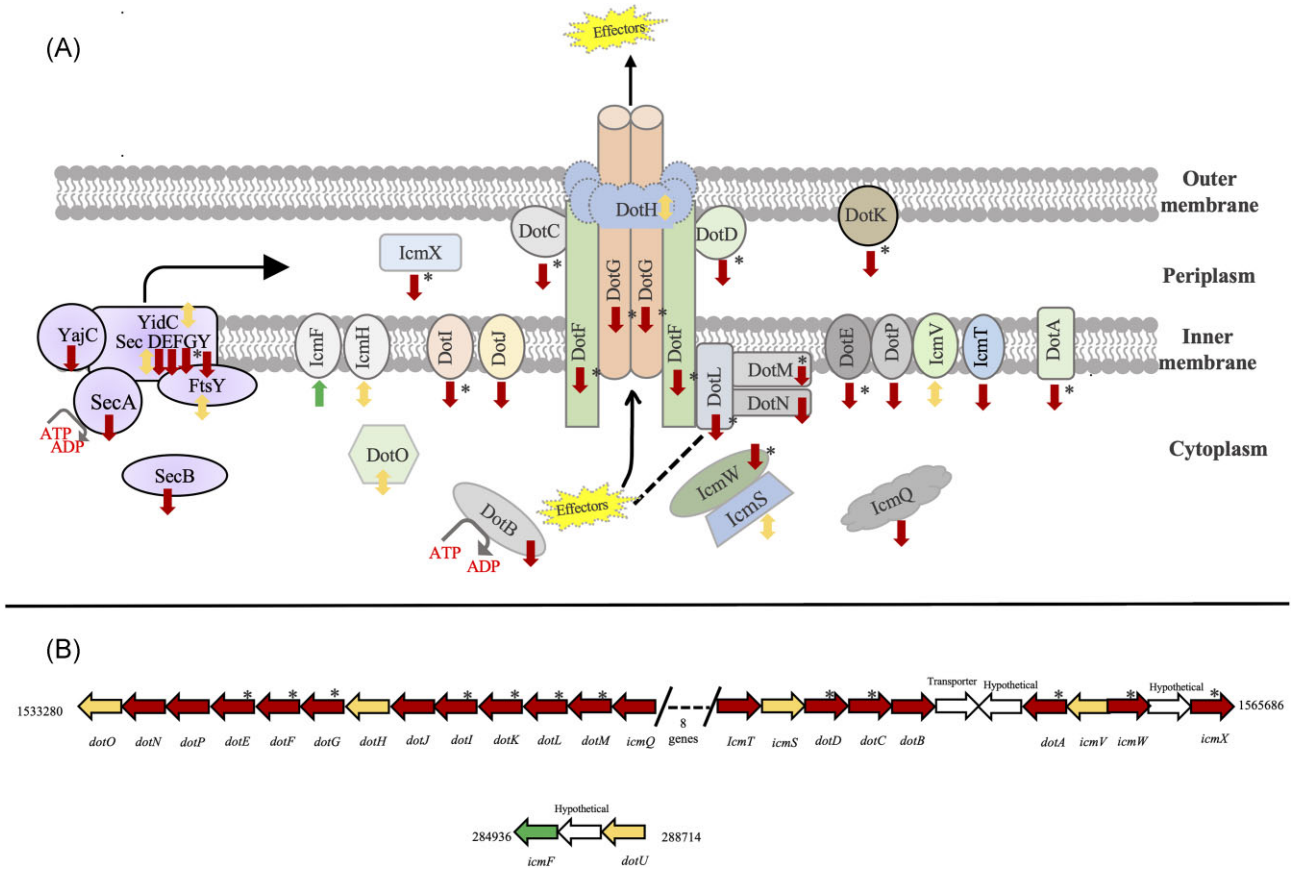


Figure 5. Cb T4BSS machinery and Sec expression changes during axenic passaging. (A) Membrane complex model with gene transcript expression profiles for components of the Cb T4BSS (right side) and Sec protein export pathway (left side). The T4BSS figure is adapted and modified from (van Schaik et al. 2013) and (Wang et al. 2018) whereas the cartoon for Sec pathway was adapted and modified from KEGG database (https://www.genome.jp/pathway/map03070). DEG patterns are denoted by arrows (downregulated in maroon, no significant expression changes in yellow, and upregulated in green). ∗ represents genes that were highly downregulated (L2fc ≤ −3). The T4BSS is comprised of a core transport complex (DotCDFGH), substrate recognition complex (DotLMN, IcmSW, LvgA), and other functional proteins. A total of 19 out of the 24 components comprising the Cb T4BSS demonstrated significant differential expression with 18 components being downregulated. The Sec pathway comprised of secYEG (constituting channel for polypeptide movement) and SecA (ATPase that drives protein movement) is known to secrete proteins from the cytosol through the cytoplasmic membrane (Mori and Ito 2001). (B) Gene locus map for all of the T4BSS the components in Cb NMII strain [CP020616.1]. The arrows are coloured filled according to DEGs patterns as (downregulated in maroon, no significant expression change in yellow, and upregulated in green). The arrows not filled are genes not related to the T4BSS pathway.

(B7L74_09 020), posttranslational modification, protein turnover and chaperones (*cpeH*), replication, recombination, and repair (*cig57*), lipid transport and metabolism (B7L74_03 275), and mobilome: prophages and transposons (B7L74_08 400) (Supplementary Table S2, Fig. 6, and Table 1). Genes B7L74_08 200 and *cpeF* were predicted as general function categories whereas there were no functional homologies for B7L74_07 850, *cig2*, *cirC*, and B7L74_03 065 in the COG database (Supplementary Table S2).

Expression patterns of additional pathogenic determinants in Cb

Expression patterns of additional pathogenic determinants unrelated to T4BSS effector proteins were examined. Of these, we noted three interesting patterns. First, the general downregulation of a wide range of chaperone proteins. Chaperone proteins primarily function as protein folding catalysts, but many are considered virulence factors for intracellular pathogens given that they encounter stress related to phagosome acidification and phagosome fusion with lysosomes (Neckers and Tatu 2008). Amongst the 16 genes annotated as chaperons in the Cb genome, 10 were transcriptionally downregulated during continuous *in vitro* passaging (Fig. 7A and Supplementary Table S1). Notable downregulated chaperones include glutaredoxins (*grxC* and *grxD*) that have been shown to be involved in CCV detoxification (Beare et al. 2009) and genes encoding heat shock proteins classes such as *dnaK*, *hptG*, *groEL*, and *dnaJ*. These proteins are known to help bacteria adapt to stressful conditions (Arnold et al. 2007, Genevaux et al. 2007; Supplementary Fig. S4). DnaK has been shown to be critical for survival of pathogenic bacteria inside the macrophage (Takaya et al. 2004) and is induced in Cb in high acid condition, similar to that found in the phagolysosome (Macellaro et al. 1998).

The second observation is the downregulation of several genes involved in lipopolysaccharide (LPS) biosynthesis. The LPS layer has long been known as a pathogenic determinant and important for host interaction in Cb (Williams and Waag 1991, Gajdosova et al. 1994, Hussein et al. 2001). Out of 35 genes related to LPS synthesis and O-antigen nucleotide sugar biosynthesis, 11 were downregulated (Fig. 7A) (Column 'LPS and O-antigen biosynthesis' in Supplementary Table S1). Out of the 11 downregulated genes, 4 are involved in the KDO2-lipid IVA Wbp pathway for LPS biosynthesis, whereas 8 genes are

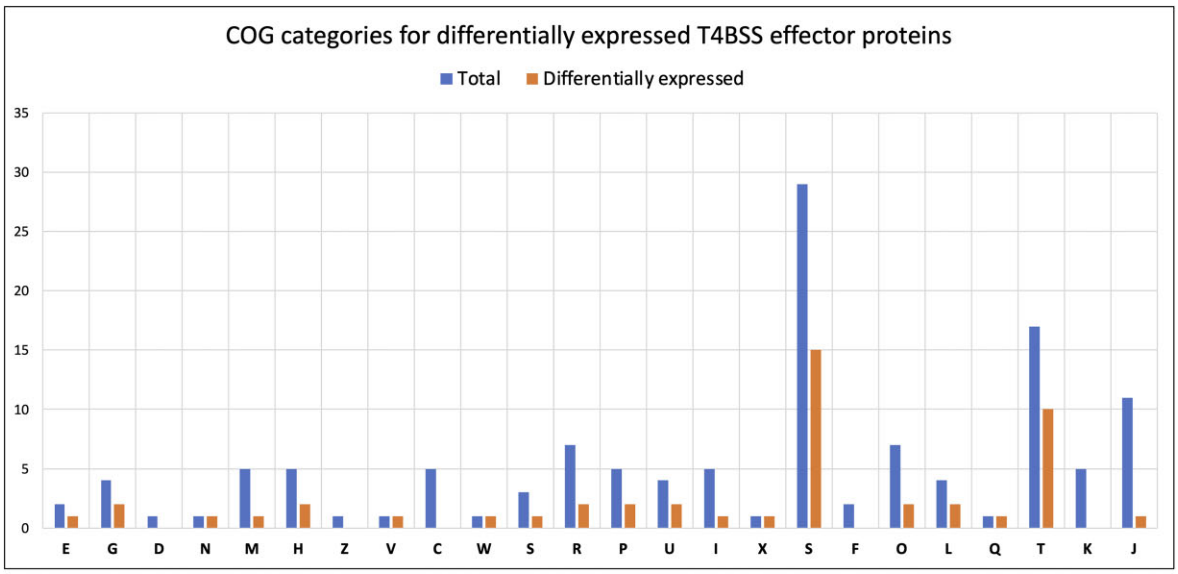
Table 1. List of 47 effector proteins differentially expressed in this study.*

Reference gene ID	Gene annotation	COG category**
B7L74_00 115	*Coxiella* vacuolar protein B (cvpB/cig2)	S
B7L74_01 850	Ankyrin repeat protein AnkD	T
B7L74_03 065	Hypothetical protein	S
B7L74_03 275	Hypothetical membrane spanning protein	I
B7L74_03 970	Ankyrin repeat protein AnkG	T
B7L74_04 780	*Coxiella* effector for intracellular replication cirC	S
B7L74_06 660	Ankyrin repeat protein AnkK	T
B7L74_07 850	Hypothetical protein	S
B7L74_08 200	Glutamyl-tRNA amidotransferase	R
B7L74_08 400	Hypothetical protein	X
B7L74_09 015	cig57 protein	L
B7L74_09 020	Hypothetical protein	G
B7L74_10 985	*Coxiella* plasmid effector protein C (cpeF)	R
B7L74_11 055	*Coxiella* plasmid effector protein H (CpeH)	O
B7L74_02 020	Hypothetical protein	S
B7L74_00 320	Molecular chaperone DnaJ	O
B7L74_00 640	Hypothetical membrane associated protein	S
B7L74_00 905	Serine/threonine protein kinase	T
B7L74_02 180	*Coxiella* effector for intracellular replication cirB	H
B7L74_02 295	Ankyrin repeat protein AnkF	T
B7L74_02 775	Hypothetical membrane associated protein	J
B7L74_03 285	Pyrroloquinoline quinone biosynthesis protein PqqC	H
B7L74_03 430	*Coxiella* vacuolar protein CvpA	L
B7L74_04 490	Hypothetical cytosolic protein	Q
B7L74_05 475	Hypothetical protein	S
B7L74_06 775	Hypothetical cytosolic protein	M
B7L74_06 960	Hypothetical cytosolic protein	S
B7L74_07 255	Hypothetical protein	P
B7L74_07 845	Hypothetical protein	S
B7L74_08 100	Hypothetical membrane spanning protein	W, U, N
B7L74_08 260	Hypothetical protein	S
B7L74_08 270	Ankyrin repeat protein	T
B7L74_08 290	Hypothetical protein	T
B7L74_08 875	Hypothetical protein	S
B7L74_09 030	Hypothetical protein	G
B7L74_10 395	Hypothetical protein	P
B7L74_10 785	Hypothetical exported protein	S
B7L74_10 940	Hypothetical protein	T
B7L74_10 945	Hypothetical protein	V
B7L74_10 950	QpH plasmid ptrotein CbhE	S
B7L74_00 375	Ankyrin repeat protein AnkA	T
B7L74_10 670	Hypothetical protein	S
B7L74_07 525	Hypothetical protein	U
B7L74_10 650	Hypothetical protein	E
B7L74_07 055	Hypothetical membrane associated protein	S
B7L74_09 355	Hypothetical membrane associated protein	T
B7L74_00 740	Ankyrin repeat protein AnkB	S

*Detailed quantitative expression values of each transporter at different passages are provided in Supplementary Table S1.
**COG categories: E: amino acid transport and metabolism, G: carbohydrate transport and metabolism, H: coenzyme transport and metabolism, I: lipid transport and metabolism, J: translation, ribosomal structure, and biogenesis, L: replication, recombination, and repair, M: cell wall/membrane/envelope biogenesis, N: cell motility, O: posttranslational modification, protein turnover, chaperones, P: inorganic ion transport and metabolism, Q: secondary metabolites biosynthesis, transport, and catabolism, R: general function prediction only, S: unknown function, T: signal transduction mechanisms, U: intracellular trafficking, secretion, and vesicular transport, V: defence mechanisms, W: extracellular structures, and X: mobilome: prophages, transposons.

involved in O-antigen nucleotide sugar biosynthesis (Supplementary Table S1). Genes involved in the first three steps, i.e. UDUDP-N-acetylglucosamine acyltransferase (*lpxA*), UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (*lpxC*), and P-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (*lpxD*), and the gene D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase (*gmhB*) in the LPS biosynthesis pathway are early down or continuously down (Supplementary Fig. S4 and Fig. 7C). In addition, three transporters related to LPS synthesis, i.e. a lipoprotein releasing system ATP-binding protein (*lolD*), a lipid flippase important in cell membrane formation (*pglK*), and a probable O-antigen/lipopolysaccharide transport ATP-binding protein (*rfbE*) were also early down (Fig. 7C). Eight downregulated genes including *wbpW*, *gmhB*, *galE*, *wbpD*, *galE*, *wbpI*, *cap1J*, and *glmU* are involved in O-antigen nucleotide sugar biosynthesis pathway, a 14 gene pathway, which is the first step in O-antigen biosynthesis where nucleotide sugars are assembled and activated by adding NTP (Samuel and Reeves 2003).

Finally, 3 of the 15 genes involved in peptidoglycan layer biosynthesis were downregulated (Supplementary Table S1). These



E=Amino acid transport and metabolism
G=Carbohydrate transport and metabolism
D=Cell cycle control, cell division, chromosome partitioning
N=Cell motility
M=Cell wall/membrane/envelope biogenesis
H=Coenzyme transport and metabolism
Z=Cytoskeleton
V=Defense mechanisms
C=Energy production and conversion
W=Extracellular structures
S=Function unknown
R=General function prediction only
P=Inorganic ion transport and metabolism
U=Intracellular trafficking, secretion, and vesicular transport
I=Lipid transport and metabolism
X=Mobilome: prophages, transposons
S=No hits on COG
F=Nucleotide transport and metabolism
O=Posttranslational modification, protein turnover, chaperones
L=Replication, recombination and repair
Q=Secondary metabolites biosynthesis, transport and catabolism
T=Signal transduction mechanisms
K=Transcription
J=Translation, ribosomal structure and biogenesis

Figure 6. COG classification graph for 118 identified effector proteins. The effector proteins were classified into COGs functional categories by using the COGs database downloaded from NCBI. The blue bars represent the total number of effector proteins in each COG functional group (x-axis) and orange bars represent the number of effectors that showed differential gene expression, thus showing the fraction of effectors that showed differential gene expression in each functional category.

genes are penicillin-binding protein PBP3/*ftsI*, penicillin binding protein PBP1A/*mrcA*, and undecaprenyl diphosphate synthase *uppS*. The peptidoglycan layer in Cb is an immunogenicity determinant and thickens substantially during LCV to SCV transition to help in environment resistance (Amano et al. 1984, Sandoz et al. 2016). *ftsI* and *mrcA* are the only two genes encoding penicillin binding proteins in this genome that are involved in peptide cross linking. This suggests that the peptidoglycan layer may be even thinner than is common intracellularly and could indicate a reduced requirement during axenic growth since the thick SCV peptidoglycan layer in Cb has been correlated to bacteria being more infectious (Sandoz et al. 2016).

Downregulation of multiple hypothetical proteins could suggest novel pathogenicity determinants

We posited that downregulated hypothetical proteins could represent previously unrecognized pathogenicity determinants, and that such identification could be a useful starting point for subsequent experimental validation. We identified 30 Cb genes encoding hypothetical proteins that were downregulated in more than four separate passages. Subcellular localization analysis software

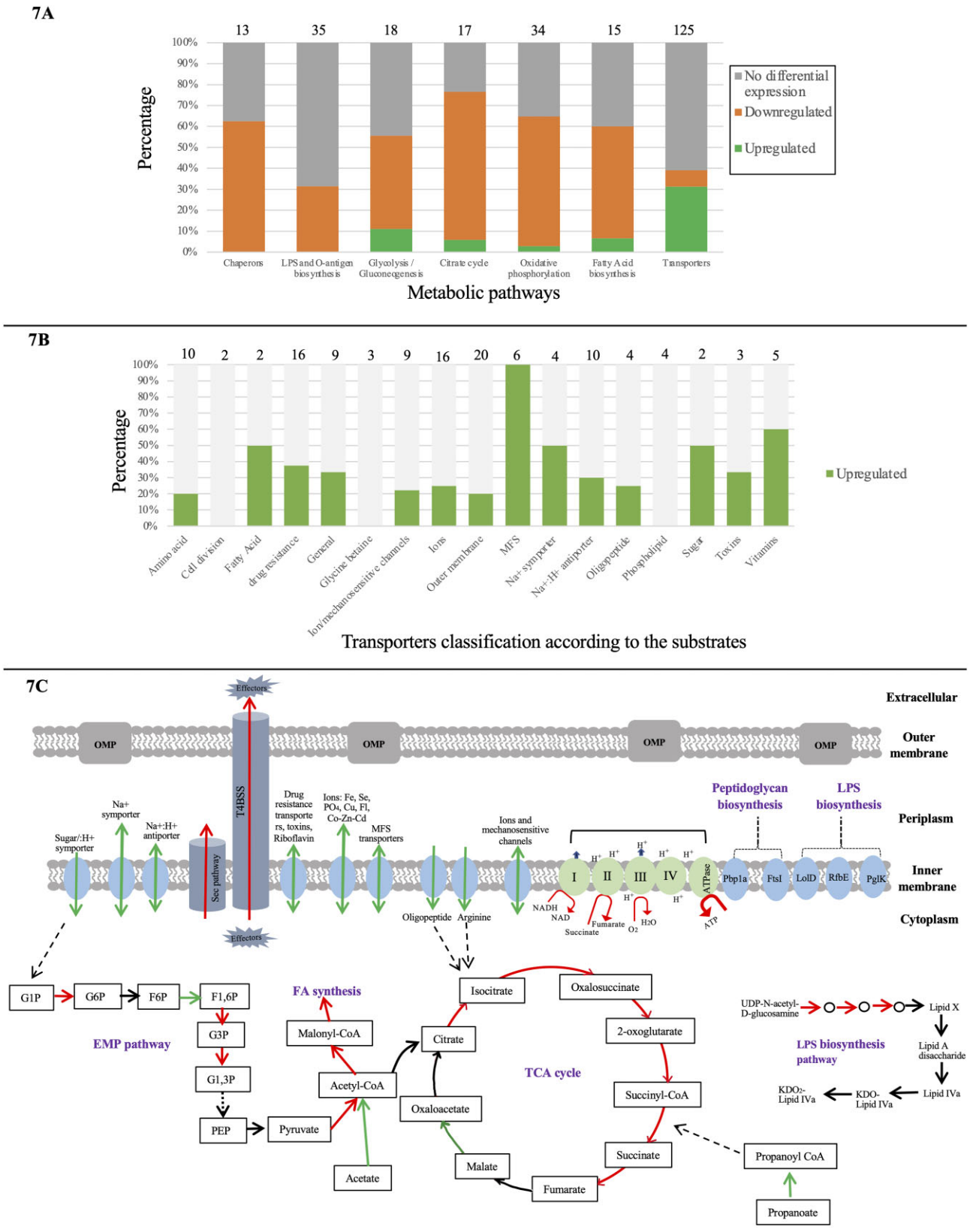


Figure 7. Gene expression changes in central metabolic pathways. (A) Bar graph showing the number of DEGs in each central metabolic pathway. The numbers on top of the bars represent the total number of genes classified into that metabolic category by KEGG. (B) Proportion of upregulated transporters (green) to the total number of transporters classified according to their substrates. The numbers on the top of each bar represents the total number of the genes that falls into this particular category of transporters. (C) Graphical representation of all metabolic pathways, i.e. glycolysis/EMP pathway, fatty acid synthesis, TCA cycle, LPS and peptidoglycan biosynthesis, T4BSS, Sec pathway, transporters, and electron transport chain (ETC) that are discussed in the manuscript text. The arrows in green represents the steps/pathways that are upregulated overall, red arrows represent the steps or the pathways that are overall downregulated, and black arrows represent the steps that does not show any significant expression change over several passages.

predicts that 13 are cytoplasmic, 8 inner membrane, 1 extracellular, 1 periplasmic, and 7 with unknown localization (Supplementary Table S4). Of the cytoplasmic proteins, Cig28 and an AMP binding protein (CBU_0787) possess a regulatory element recognized by PmrA, a sequence related to T4BSS expression and translocation, and thus a potential predictor of effector proteins (Beare et al. 2014), although they were subsequently shown not to be translocated by the Cb T4BSS (Zusman et al. 2007, Beare et al. 2014). Similarly, an uncharacterized protein, CBU_1234, has been shown to have a glutamate-rich C-terminal secretion signal (E-block), which is also a predictor of effector proteins (Weber 2014). Two Glycosyltransferase family 1 proteins (i.e. CBU_0839 and CBU_0841) have previously been linked to LPS mutations that lead to phase transitions (Beare et al. 2018). The eight proteins localized in the inner membrane included Cig3, an immunoreactive peptidase CBU_0215 (Weber 2014) previously shown to contain a regulatory element recognized by PmrA but not translocated by Dot/Icm system (Zusman et al. 2007, Beare et al. 2014), an immunoreactive protein CBU_1865 (Beare et al. 2008) and a DUF3971 domain-containing protein, CBU_1468, that has been shown to be important for intracellular replication (Newton et al. 2014; Supplementary Table S4).

The one differentially expressed hypothetical protein identified by the localization software as an extracellular protein was CBU_0962; a predicted short chain dehydrogenase with a yet unknown specific function (Bewley 2015). Lastly, proteins with unknown localization included an exported protein Cig40 (Weber 2014) with a regulatory element recognized by PmrA, a hypothetical surface antigen Com1 (Chen et al. 2010) and a hypothetical protein CBUA0012 located in an ORF containing other plasmid effectors, but shown to not be secreted by the T4BSS (Voth et al. 2011; Supplementary Table S4).

Transcriptional patterns of central metabolic pathways

Analysis of gene expression patterns of central metabolic pathways demonstrated a general trend of downregulation in genes encoding enzymes in central catabolic, amphibolic, and anabolic pathways, coupled with a broad upregulation in genes encoding transporters. An overall pattern of downregulation of glycolysis genes (8/18 genes) was observed (Fig. 7A), with several enzymes such as pyruvate dehydrogenases (*pdhC*, *pdhD*), fructose-bisphosphate aldolase (*fbaA*), glyceraldehyde 3-phosphate dehydrogenase (*gapA*), and phosphoenolpyruvate carboxykinase (*pckA*) downregulated early in the passaging (Supplementary Fig. S4 and Supplementary Table S1). Gene *pmm-pgm*, which encodes the enzyme phosphomannomutase/phosphoglucomutase and is involved in the first step of glycolysis, was down early as well (Supplementary Fig. S4). Similarly, analysis of genes encoding enzymes of the tricarboxylic acid (TCA) cycle, demonstrated an overall downregulation (12/17 genes), with a downregulation of ∼2-fold in isocitrate dehydrogenase (*IDH2*), the rate-limiting enzyme in the TCA cycle (Supplementary Fig. S4). As well, the downregulations in multiple carbohydrate dehydrogenases (e.g. pyruvate dehydrogenases *pdhC* and *pdhD*), succinate dehydrogenases (*sdhA*, *sdhB*, *sdhD*), as well as genes that are necessary for oxidation of glycolytic and TCA cycle sugar intermediates was observed. Finally, genes encoding components of the ETC were also downregulated. These include Complex I: NADH-quinone oxidoreductase (*nuoB*, *nuoD*, *nuoE*, *nuoF*, *nuoH*, *nuoI*, *nuoK*, *nuoL*, *nuoM*, *nuoN*), Complex II: Succinate dehydrogenase (*sdhA*, *sdhB*, *sdhD*), Complex III: Cytochrome oxidoreductase (*cyoC*, *cyoD*, *cydA*, *cydB*, *cydX*), and Complex V: F-type ATPase (*atpA*, *atpB*, *atpE*, *atpD*, *atpF*, *atpG*) (Fig. 7A and Supplementary Fig. S4). The downregulations in two complexes (*cyoC* and *cyoD*) of cytochrome o oxidase, which is known to be induced in oxygen rich growth conditions in bacteria, (Cotter et al. 1990) suggests a decreased affinity and/or competition for oxygen in the cell-free growth environment, as previously suggested (Kuley et al. 2015). Cytochrome d oxidase, which is shown to be expressed more in oxidative and nitrosative stress conditions (Cotter et al. 1990), also has expression of two of its components (i.e. *cydA* and *cydX*) early down and late down, respectively. Finally, an overall downregulation of fatty acids biosynthesis genes transcription (8/15) was also observed (Fig. 7A and Supplementary Fig. S4).

In contrast to the general trend of downregulation of the central metabolic machinery of Cb, a marked upregulation of genes encoding transporters was observed. Out of 125 general transporters, the transcription of 39 were upregulated and 10 were downregulated (Column 'Transporters' in Supplementary Table S1) (Fig. 7A). Transporters that were upregulated have double fold expression change (L2fc) ranging from 3 to 10 (Supplementary Fig. S4). Upregulated primary transporters included transporters for the amino acid arginine, oligopeptides, fatty acids, and vitamins such as riboflavin and thiamin (Fig. 7B and Supplementary Fig. S2) as well as a small number of transporters (4 out of 20 present) related to the synthesis and maintenance of the outer membrane (Supplementary Table S1). On the other hand, upregulated secondary transporters included MFS transporters, symporters, antiporters, and mechanosensitive ion channels. Of these, a notable observation was made where all six MFS transporters and two out of four Na+ symporters found in the Cb genome were found to be early upregulated (Supplementary Table S1). These MFS transporters transport various compounds such as monosaccharides, oligosaccharides, amino acids, peptides, vitamins, cofactors, drugs, nucleobases, nucleosides, and organic and inorganic anions and cations. Also, a large proportion of transporters related to drug resistance (6/16 present in the genome) and ion transporters mediating the uptake of ions such as copper, iron, fluoride, selenite, cobalt–cadmium–zinc, and phosphate were also upregulated (Supplementary Table S1 and Fig. 7B). In addition to transporters mediating substrate transport, transporters involved in pH homeostasis such as ions/mechanosensitive channels and Na+:H+ antiporter were also upregulated. (Fig. 7B and Supplementary Table S1). The Na+:H+ antiporter functions utilizes the proton motive force to efflux intracellular sodium ions for intracellular pH homeostasis (Ito et al. 2017) and these antiporters along with ion/mechanosensitive channels have been proposed to play an important role in pH homeostasis and survival within the acidic CCV (Seshadri et al. 2003). Lastly, 3 out of 9 transporters classified under general or unknown functions were upregulated as well (Fig. 7B).

Genomics reveals a stable Cb genome

For all 12 passages analysed, genomes with 100% completeness [assessed by identifying all 265 housekeeping marker genes specific for the Proteobacteria (Parks et al. 2015)] were obtained. N50 of genomic assemblies ranged between 49 903 and 75 629, N90 ranged between 15 966 and 20 406, and the number of contigs per genome ranged between 56 and 64 (Supplementary Table S5). Using Passage 1 as a reference, we identified 842 unique SNPs and 118 unique DIPs (Supplementary Table S5 and Fig. 8).

Of 842 unique SNPs, only 9 were identified in consensus mode (i.e. present in 100% of the population in one or more passages) while the remaining 833 SNPs were identified in population mode

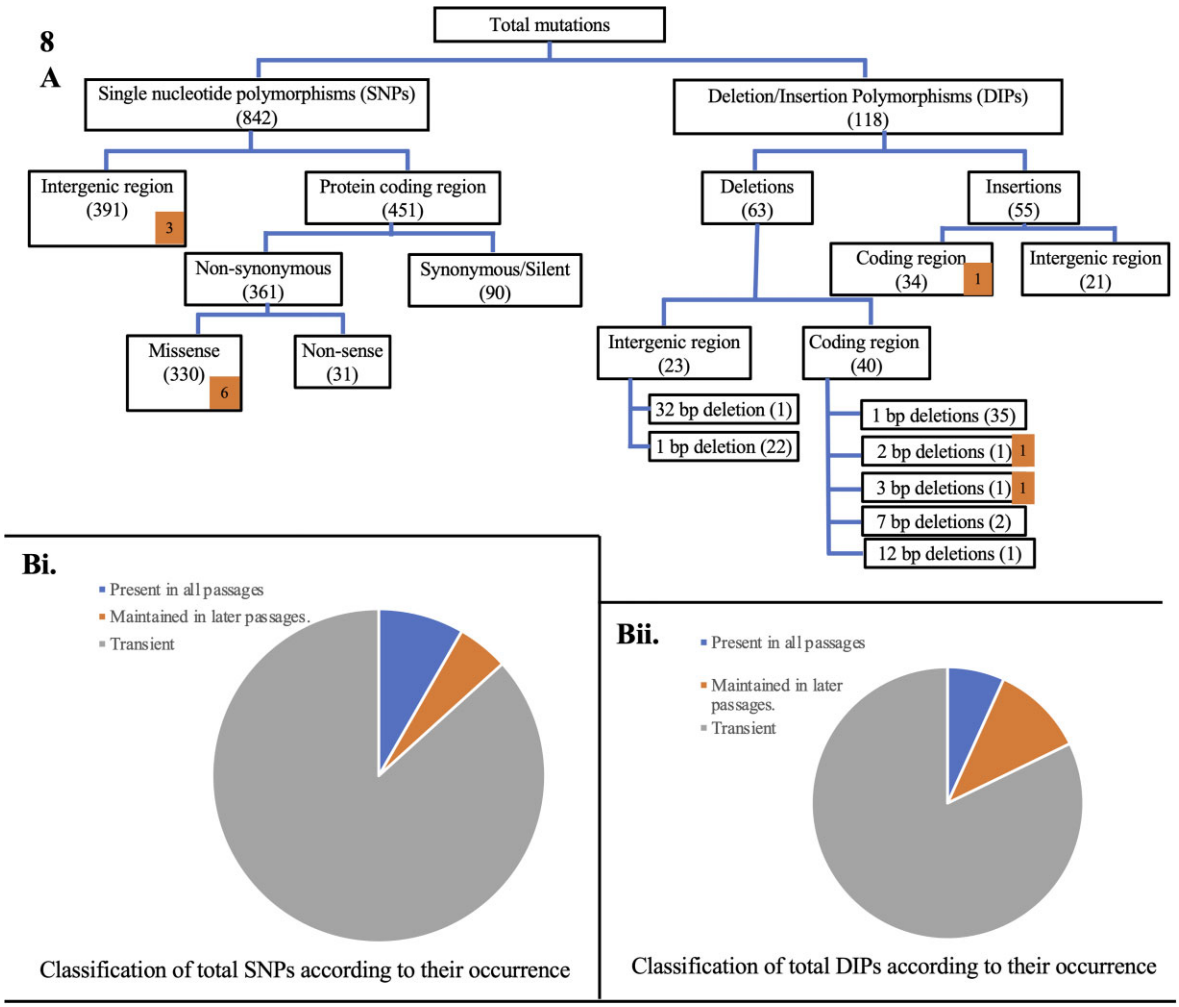


Figure 8. Breseq analysis of genomes from the 12 different passages. (A) Flowchart for classification of different types of SNPs and DIPs found in this experiment. The SNPs and DIPs are further classified into specific types as shown by the lower branches. Numbers in the orange blocks at the tip of some boxes represents the number of those mutations that occurred in 100% of the passages. (B) Classification of SNPs or DIPs by frequency of their occurrences, i.e. present in all passages (if a mutation is present even in a fraction of the population in all 12 passages), maintained in later passages (if the mutation is present in any fraction of the population in multiple later passages continuously), and transient (if the mutation is present in random multiple passages with no coherent expression in continuous passages). (Bi) Classification of SNPs according to its occurrence in number of passages. (Bii) Classification of DIPs according to the occurrence in number of passages.

(i.e. occurring in a fraction of the community), when sequenced (Fig. 8). Further, only 69 unique SNPs were identified to occur in all (i.e. Passage 3–61), and only 43 SNPs were maintained in later passages (Fig. 8B). More importantly, only one consensus mutation occurred in a gene that was downregulated in transcriptomic analysis. This gene GTP pyrophosphokinase SpoT (B7L74_01 590) had a one amino acid (aa) substitution (T to A) at position 262, which propagates to 100% of the population in the last nine passages analysed and is also noticeably early down in gene expression. SpoT is a signal transduction component and transcriptional regulator with a role in helping Cb cope with their normally low-nutrient and high stress conditions (Minnick and Raghavan 2012).

For the 118 unique DIPs, only 3 DIPs were identified in consensus mode and 115 in population mode (Fig. 8). Lengths of insertions and deletions were always very minor with 93% of DIPs representing an insertion or deletion of a single bp (Supplementary Table S3). The multibase pair deletions included deletions of 2, 3, 7, and 12 bp that occurred within coding regions, and the longest observed deletion, 32 base pairs, occurring in an intergenic region. However, none of these genes appeared significantly affected transcriptionally by the deletion as there were no significant transcriptomic changes. Of the genes that were downregulated, seven had DIPs mutations but all of those were only in a fraction of the Cb population within a passage (mostly 5%–10% of the population).

An interesting observation was the numerous mutations (SNPs and DIPs) over several passages in two genes, i.e. *lapA* and *lapB*. In LapA, the 97 aa long protein has a nonsense mutation at the 85th position in passage 13. LapB, an 389 aa long protein, on the other hand has two missense mutations in a large proportion of cells, with one mutation propagating to 100% of the population at passage 67. It also has numerous insertions in the coding region with a noticeable one being a 3 bp deletion in the coding region that propagates to later passages (Supplementary Table S3). Although these genes did not show any change in gene transcriptional expression modulated by mutations, it is possible that these genes are en route to simplifying the LPS and O-antigen layer in accordance with the absence of a stress environment, as seen in some bacteria (Maldonado et al. 2016).

Collectively, the low levels of consensus DNA mutations within the passage populations and possibly the lack of effects, suggest a very stable genome with minor levels of genomic mutations modulating transcriptional levels.

Discussion

Here, we attempted to identify genes and proteins crucial to Cb's intracellular growth lifestyle using a 'reverse evolution' approach paired with RNAseq and DNAseq comparative transcriptomics and genomics, respectively. We transitioned Cb Nine Mile phase II from cell cultures into the axenic defined media ACCM-D and subcultured it in a long-term successive passage model. Temporal changes in gene expression patterns, and DNA mutations associated with adaptation to an axenic extracellular lifestyle were identified. In general, we observe a significant number of differential expression (464 up, 371 down, 38% of overall Cb genes) through 67 passages. It is interesting to note that the majority (288 upregulated and 249 downregulated) of DEGs expressed an 'early up' or 'early down' expression pattern (Fig. 3 and Supplementary Table S1), suggestive of a relatively rapid adaptation during early passaging into this new axenic environment.

DEGs identified in this study could be grouped into different structural and functional categories (secretory apparatus, effector proteins, other pathogenicity determinants, hypothetical proteins, and central metabolic pathways). In general, a broad downregulation of genes (19/24) encoding T4BSS components showed significant expression change, with 18 genes showing a decrease in the expression whereas only one gene that was upregulated. T4BSS is the most crucial conduit for pathogenicity and effector proteins in Cb (Voth and Heinzen 2009, Carey et al. 2011, van Schaik et al. 2013). Components of the T4BSS span both membranes and the periplasm and are bridged by the core transport complex comprising proteins DotC, DotD, DotF, DotG, and DotH, which are predicted to provide a channel for export of effector substrates (Fig. 5A; Vincent et al. 2006). The coupling protein complex provides a link between substrates and transport complex and includes DotL, DotM, DotN, IcmS, and IcmW (Vincent et al. 2012). DotB is an essential cytoplasmic protein with an ATPase activity and unknown function, but its mutation has been linked to failure in secreting effector proteins during the infection of host cells (Beare et al. 2012). DotA and IcmX has been shown to be released from the bacteria (Luedtke et al. 2017). Besides these, other components of the T4BSS includes DotO localized in the cytoplasm, IcmX in periplasmic space, DotK in outer membrane, whereas IcmF, IcmH, DotI, DotJ, DotA, DotE, DotP, IcmV, and IcmT in the inner membrane. (Fig. 5A). The genes involved in the T4BSS in Cb are clustered in a single locus made up of two regions, with the exception of *icmF* and *dotU*, which are part of a separate operon (Fig. 5B). This is similar to the gene arrangement shown in the original Cb NMI sequence (Seshadri et al. 2003). Gene *icmF*, which has been shown to be involved in intramacrophage replication and inhibition of phagosome–lysosome fusion in *L. pneumophila* (VanRheenen et al. 2004, Zusman et al. 2004) and stabilization of the secretion complex (Sexton et al. 2004) was the only T4BSS component that showed transcriptional upregulation.

The observed downregulation of this experimentally verified central intracellular pathogenic determinant provides validation that gene downregulation under the experimental setting employed in this study could be regarded as a reasonable proxy of requirements for intracellular survival in cell-cultures. In addition to T4BSS, other secretory pathway such as the general secretory (sec) pathway and a component of the type I secretary pathway (i.e. TolC) also exhibited a general trend of overall downregulation (Fig. 5A, Supplementary Fig. S2a and Fig. S2b, and Supplementary Table S1). We interpret such overall lower expression of structural secretory apparatuses as a reflection of less need for these systems during interaction between Cb and the environment in a relatively rich axenic setting when compared to the organisms environmentally 'normal' intracellular setting.

Interestingly, while genes encoding the production of secretary machineries were downregulated, expression patterns of predicted T4BSS effector proteins were mixed, with 33 upregulated and 14 downregulated. Of the 14 downregulated genes (all of which were early downregulated), 9 have been experimentally verified based on experimental evidence of their translocation by the Dot/Icm system (Voth et al. 2009, Chen et al. 2010, Carey et al. 2011, Voth et al. 2011, Lifshitz et al. 2013, Maturana et al. 2013, Weber et al. 2013), three genes containing ankyrin repeat domains (*ankG*, *ankD*, and *ankK*) were considered effectors based on the presence of eukaryotic like domains and subsequently shown to be translocated by the Dot/Icm system (Voth et al. 2009) The Cb gene *cirC* (*Coxiella* effector for intracellular replication) was verified as an effector by transposon insertion mutation studies where its mutation was associated with a defect in CCV biogenesis (Weber et al. 2013). Lastly, a hypothetical protein B7L74_09 020 was verified to be an effector based on loss-of-function where its mutation was related to a smaller CCV phenotype (Crabill et al. 2018; Supplementary Table S2).

The eukaryotic type Ank domain in this protein family might have a role in host-cell attachment and allows the interaction of bacteria with a spectrum of host cell proteins and thus are considered particularly important in the pathogenic process (Batrukova et al. 2000, Voth et al. 2009, Pechstein et al. 2020, Cordsmeier et al. 2022). AnkD has both an Ank eukaryotic like domain and an F-box domain, but the function is not yet clear (Voth et al. 2009). AnkG in Cb NMII appears to associate with mitochondria and to be transported into the host cell nucleus (Eckart et al. 2014, Schäfer et al. 2017, 2020). AnkK has been shown to have an important role for bacterial growth inside of macrophages (Habyarimana et al. 2008), although it is not delivered to the host cell via T4BSS (Voth et al. 2009; Supplementary Table S2). *Coxiella* plasmid effector proteins (CpeF and CpeH) are in the plasmid T4BSS effector family of proteins, and are important for disrupting host cell mechanisms. Cig57 mutation has been linked to an intracellular replication defect for Cb, whereas a *Coxiella* vacuolar protein B (Cig2/CvpB) mutation causes both a growth defect and a CCV fusion defect (Newton et al. 2014). These two proteins are early downregulated in nine and seven passages, respectively. *Coxiella* effector for intracellular replication C/Mitochondrial *Coxiella* effector protein B (CirC/MceB) has been shown to be important for CCV biogenesis (Weber et al. 2013) and it is early downregulated in four passages. Of the remaining six downregulated effectors, four are hypothetical proteins with unknown functions, Cbu1752 has been shown to be important for vacuole biogenesis, and Cbu0635 appears important for host cell secretion (Supplementary Table S2).

Of the 33 upregulated effector proteins, the majority fall into COG categories of unclassified ($n = 11$), signal transduction mechanisms ($n = 7$), transportation and metabolism of coenzyme and inorganic ions ($n = 6$) (Fig. 6 and Table 1). These upregulated, and no expression change effector proteins ($n = 71$), could also be involved in mediating general survival functions or other cellular functions besides their involvement in directly association with the intracellular pathogenesis process. This could be one of the

explanations behind their upregulation, or no expression change, in this particular setting.

Such a pattern, where genes encoding the formation of the structural conduits (i.e. secretary pathways) are downregulated, but numerous genes encoding proteins thought to be secreted through these conduits (i.e. effector proteins) are upregulated is puzzling. We put forth the possibility that the expression of these effector proteins is controlled by Cb intracellular conditions, where high concentrations of intracellular metabolites (amino acids, inorganic salts, ATP/ADP ratio) regulate their expression. Under this scenario, high level of intracellular precursors in the Cb LCV is associated with growth inside the cell, and possibly in the 'rich' axenic media state of LCV growth. It remains to be seen whether translation of these effector transcripts to protein products and subsequent secretion occurs in Cb grown in axenic media. Limited reports suggest that Cb T4BSS effectors have not been observed during axenic media growth (Stead et al. 2013, Shaw, unpublished data)

Multiple additional pathogenic determinants were also downregulated in axenic media. Specifically, chaperons, LPS, and peptidoglycan synthesis. Amongst the 16 genes annotated as chaperons in the Cb genome, 10 were transcriptionally downregulated starting at early passages (Fig. 7A and Supplementary Table S1). The downregulation could be explained by the fact that chaperons play important roles for withstanding stress associated with intracellular survival (e.g. CCV detoxification) (Beare et al. 2009), survival inside the macrophage (Takaya et al. 2004), and low pH within the CCV (Macellaro et al. 1998). Downregulation in the genes involved in the synthesis of Lipid A and O-antigen as well as peptide cross linking during peptidoglycan layer synthesis suggests the possibility of further reduction in the need for these functions of Cb grown in axenic media, even within the Cb NMII strain used in this study. These changes in the bacterial cell outer surface could be tied to their role/importance in infection and intracellular host cell manipulation and bacterial survival and growth inside the host cell. These assumptions need to be verified with more experimental evidence. Here, we need to keep in mind that the strain we are using is the Cb avirulent strain NMII. These variants have a truncated LPS, due to a genomic deletion of approximately 25 Kbp of sequences that encodes the three sugars i.e. virenose, dihydrohydroxystreptose, and galactosaminuronyl-α(1,6)-glucosamine that comprises the LPS O-antigen biosynthesis (Amano et al. 1987, Denison et al. 2007). Thus, further downregulations in some of the remaining genes involved in LPS and O-antigen synthesis pathways could be due to their already altered functionality in the cell.

Analysis of central metabolic pathways showed a clear trend of downregulation of multiple catabolic (e.g. glycolysis and electron transport chain), amphibolic (e.g. citric acid cycle), and anabolic (e.g. FA synthesis) pathways, with a parallel upregulation of genes encoding transporters (Fig. 7C). Such a pattern could readily be explained by the nutrient rich growth environment (ACCM-D media) where Cb is grown in our continuous passage model. This media is a defined axenic medium for Cb growth and contains all 20 amino acids, salts (sodium phosphate and sodium bicarbonate), vitamins, minerals, and trace elements (Sandoz et al. 2016). As such, the need for expression of genes encoding enzymes that are components of these biosynthetic pathways decreases and subsequently, the overall need for ATP generation for biosynthetic purposes (hence a decrease in respiratory activity). Likewise, the upregulation in several structural genes encoding transporters could be explained by the bacteria increasing production of more channels/transporters to accommodate the increased presence of metabolites/substrates from the media as opposed to needing to produce them via metabolic pathways in the bacterial cytoplasm. Finally, 30 hypothetical proteins were downregulated, and their predicted localization, predicted role in pathogenesis and their general analysis provide some possible explanations for such pattern (see the Results section). Regardless, we suggest that these could be important, hitherto untested possible pathogenicity determinants. Future biochemical and genetic efforts to test such assumptions represent a ready avenue for future research directions.

We also hypothesized that continuous passaging could lead to the propagation of mutations, DNA fragment loses, and rearrangements in genes involved in intracellular survival, pathogenesis, and host cell manipulations could occur. While this is a well-known process, it does not appear that within the timeframe of the experiments here (67 passages) that this is the case. The careful study of each of these mutations, their corresponding gene expression changes, number of passages affected by the mutations and their transiency, and overall fraction of population affected by these mutations showed that the observed mutations do not appear to be significant at the gene or transcript expression level. Out of 960 unique mutations (SNPs and DIPs collectively) observed at different spots within the genomes of various passages, only 1 mutation (i.e. a SNP) in GTP pyrophosphokinase SpoT, a signal transduction component and transcriptional regular with a role in helping Cb cope with the low-nutrient and high stress conditions in the CCV, seems to have an effect on gene transcript expression (Minnick and Raghavan 2012). Early downregulation observed in this gene could be attributed to the fact that these cells are less stressed due to the growth conditions in rich ACCM-D media, and this mutation could have caused the decrease in transcript expression of this gene. Overall, these findings suggest that Cb gene expression changes significantly following acclimation to axenic media, although extensive genomic rearrangement does not occur. Genomics reveals a relatively stable Cb NMII genome over the 67 passages that were analysed here.

In conclusion, we present a detailed temporal analysis on how Cb transitions from intracellular growth (where a wide range of cellular processes are thought to be required to maintain survival and growth), to a defined, rich axenic media (where many of such processes are theoretically, no longer needed). As in any genome-wide transcriptomics survey, the approach is useful for uncovering patterns, confirming prior observations, and generating new insights and hypothesis. We stress that while transcriptional downregulation in axenic media compared to cell culture could broadly be associated with a given genes importance for survival in cell cultures, the precise nature of such correlation is yet unclear, and that differential expression patterns could further be modified on the translational and post translational levels. Experimental assessments and validation of many of the observed patterns may well open new avenues of Cb research. Nevertheless, our analysis is beneficial in providing information on how specific genes and pathways in Cb may be important for this unusual organisms intracellular survival, as well as to identify putatively novel pathogenicity determinants in this naturally intracellular pathogen.

Author contributions

Archana Yadav (Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft), Melissa N. Brewer (Data curation, Formal analysis, Investigation, Methodology - writing and editing), Mostafa S. Elshahed (Data curation,

Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing), and Edward I. Shaw (Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing)

Acknowledgements

The authors would like to thank Dr. Noha Youssef for insightful discussions into bioinformatics analysis.

Supplementary data

Supplementary data are available at *FEMSPD* online.

Conflict of interest. The authors declare no conflict of interest.

Funding

This work was supported by the National Institutes of Health, National Institutes of Allergy and Infectious Diseases Grant R03AI149144 to EIS.

References

Allen B, Drake M, Harris N *et al.* Using KBase to assemble and annotate prokaryotic genomes. *Curr Protoc Microbiol* 2017;**46**: 1E.13.1–18.

Amano K, Williams JC, McCaul TF *et al.* Biochemical and immunological properties of *Coxiella burnetii* cell wall and peptidoglycan-protein complex fractions. *J Bacteriol* 1984;**160**:982–88.

Amano K, Williams JC, Missler SR *et al.* Structure and biological relationships of *Coxiella burnetii* lipopolysaccharides. *J Biol Chem* 1987;**262**:4740–47.

Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;**11**:R106.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Last accessed date November 2021].

Arnold DL, Jackson RW, Waterfield NR *et al.* Evolution of microbial virulence: the benefits of stress. *Trends Genet* 2007;**23**:293–300.

Arricau-Bouvery N, Rodolakis A. Is Q fever an emerging or re-emerging zoonosis? *Vet Res* 2005;**36**:327–49.

Batrukova MA, Betin VL, Rubtsov AM *et al.* Ankyrin: structure, properties, and functions. *Biochemistry (Mosc)* 2000;**65**:395–408.

Beare PA, Chen C, Bouman T *et al.* Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a *Coxiella burnetii* protein microarray. *Clin Vaccine Immunol* 2008;**15**:1771–79.

Beare PA, Gilk SD, Larson CL *et al.* Dot/Icm type IVB secretion system requirements for *Coxiella burnetii* growth in human macrophages. *Mbio* 2011;**2**:e00175–11.

Beare PA, Heinzen RA. Gene inactivation in *Coxiella burnetii*. In: Vergunst A, O'Callaghan D (eds.), *Host–Bacteria Interactions*. New York, NY: Humana Press, 2014, 329–45.

Beare PA, Jeffrey BM, Long CM *et al.* Genetic mechanisms of *Coxiella burnetii* lipopolysaccharide phase variation. *PLoS Pathogens* 2018;**14**:e1006922.

Beare PA, Larson CL, Gilk SD *et al.* Two systems for targeted gene deletion in *Coxiella burnetii*. *Appl Environ Microbiol* 2012;**78**:4580–89.

Beare PA, Sandoz KM, Larson CL *et al.* Essential role for the response regulator PmrA in *Coxiella burnetii* type 4B secretion and colonization of mammalian host cells. *J Bacteriol* 2014;**196**:1925–40.

Beare PA, Unsworth N, Andoh M *et al.* Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus *Coxiella*. *Infect Immun* 2009;**77**:642–56.

Bewley KR. *The identification of immune-reactive proteins recognised in response to Coxiella burnetii infection. Ph. D. Thesis, School of Pharmacy and Biomedical Sciences*, University of Portsmouth, 2015, 230.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20.

Brennan RE, Russell K, Zhang G *et al.* Both inducible nitric oxide synthase and NADPH oxidase contribute to the control of virulent phase I *Coxiella burnetii* infections. *Infect Immun* 2004;**72**:6666–75.

Brennan RE, Samuel JE Evaluation of *Coxiella burnetii* antibiotic susceptibilities by real-time PCR assay. *J Clin Microbiol* 2003;**41**:1869–74.

Carey KL, Newton HJ, Luhrmann A *et al.* The *Coxiella burnetii* Dot/Icm system delivers a unique repertoire of type IV effectors into host cells and is required for intracellular replication. *PLoS Pathog* 2011;**7**:e1002056.

Chen C, Banga S, Mertens K *et al.* Large-scale identification and translocation of type IV secretion substrates by *Coxiella burnetii*. *Proc Natl Acad Sci USA* 2010;**107**:21755–60.

Coleman SA, Fischer ER, Howe D *et al.* Temporal analysis of *Coxiella burnetii* morphological differentiation. *J Bacteriol* 2004;**186**:7344–52.

Cordsmeier A, Rinkel S, Jeninga M *et al.* The *Coxiella burnetii* T4SS effector protein AnkG hijacks the 7SK small nuclear ribonucleoprotein complex for reprogramming host cell transcription. *PLOS Pathogens* 2022;**18**:e1010266.

Cotter PA, Chepuri V, Gennis RB *et al.* Cytochrome o (cyoABCDE) and d (cydAB) oxidase gene expression in *Escherichia coli* is regulated by oxygen, pH, and the fnr gene product. *J Bacteriol* 1990;**172**:6333–8.

Crabill E, Schofield WB, Newton HJ *et al.* Dot/Icm-translocated proteins important for biogenesis of the *Coxiella burnetii*-containing vacuole identified by screening of an effector mutant sublibrary. *Infect Immun* 2018;**86**:e00758–17.

Davis JJ, Wattam AR, Aziz RK *et al.* The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 2020;**48**:D606–D12.

Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* 2014;**1151**:165–88.

Denison AM, Massung RF, Thompson HA. Analysis of the O-antigen biosynthesis regions of phase II isolates of *Coxiella burnetii*. *FEMS Microbiol Lett* 2007;**267**:102–07.

Eckart RA, Bisle S, Schulze-Luehrmann J *et al.* Antiapoptotic activity of *Coxiella burnetii* effector protein AnkG is controlled by p32-dependent trafficking. *Infect Immun* 2014;**82**:2763–71.

Gajdosova E, Kovacova E, Toman R *et al.* Immunogenicity of *Coxiella burnetii* whole cells and their outer membrane components. *Acta Virologica* 1994;**38**:339–44.

Galperin MY, Wolf YI, Makarova KS *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 2021;**49**:D274–d81.

Genevaux P, Georgopoulos C, Kelley WL. The Hsp70 chaperone machines of *Escherichia coli*: a paradigm for the repartition of chaperone functions. *Mol Microbiol* 2007;**66**:840–57.

Green ER, Mecsas J. Bacterial secretion systems: an overview. *Microbiol Spectr* 2016;**4**.

Gurevich A, Saveliev V, Vyahhi N *et al.* QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**:1072–75.

Habyarimana F, Al-Khodor S, Kalia A *et al.* Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of

protozoan hosts and human macrophages. *Environ Microbiol* 2008;**10**:1460–74.
Hackstadt T, Williams JC. Biochemical stratagem for obligate parasitism of eukaryotic cells by *Coxiella burnetii*. *Proc Natl Acad Sci USA* 1981;**78**:3240–4.
Heinzen RA, Hackstadt T, Samuel JE. Developmental biology of *Coxiella burnettii*. *Trends Microbiol* 1999;**7**:149–54.
Howe D, Melnicakova J, Barak I *et al.* Fusogenicity of the coxiella burnetii parasitophorous vacuole. *Ann NY Acad Sci* 2003;**990**:556–62.
Hussein A, Kovacova E, Toman R. Isolation and evaluation of *Coxiella burnetii* O-polysaccharide antigen as an immunodiagnostic reagent. *Acta Virologica* 2001;**45**:173–80.
Ito M, Morino M, Krulwich TA. Mrp antiporters have important roles in diverse bacteria and archaea. *Front Microbiol* 2017;**8**:2325.
Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**:726–31.
Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* 2020;**29**:28–35.
Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;**37**:907–15.
Kovaka S, Zimin AV, Pertea GM *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* 2019;**20**:278.
Kuley R, Bossers-deVries R, Smith HE *et al.* Major differential gene regulation in *Coxiella burnetii* between in vivo and in vitro cultivation models. *BMC Genomics* 2015;**16**:953.
Larson CL, Beare PA, Voth DE *et al.* *Coxiella burnetii* effector proteins that localize to the parasitophorous vacuole membrane promote intracellular replication. *Infect Immun* 2015;**83**:661–70.
Larson CL, Heinzen RA. High-content imaging reveals expansion of the endosomal compartment during *Coxiella burnetii* parasitophorous vacuole maturation. *Front Cell Infect Microbiol* 2017;**7**:48.
Larson CL, Martinez E, Beare PA *et al.* Right on Q: genetics begin to unravel *Coxiella burnetii* host cell interactions. *Future Microbiol* 2016;**11**:919–39.
Lifshitz Z, Burstein D, Peeri M *et al.* Computational modeling and experimental validation of the *Legionella* and *Coxiella* virulence-related type-IVB secretion signal. *Proc Natl Acad Sci USA* 2013;**110**:E707–E15.
Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
Luedtke BE, Mahapatra S, Lutter EI *et al.* The *Coxiella Burnetii* type IVB secretion system (T4BSS) component DotA is released/secreted during infection of host cells and during in vitro growth in a T4BSS-dependent manner. *Pathog Dis* 2017;**75**:ftx047.
Macellaro A, Tujulin E, Hjalmarsson K *et al.* Identification of a 71-kilodalton surface-associated Hsp70 homologue in *Coxiella burnetii*. *Infect Immun* 1998;**66**:5882–88.
Maldonado RF, Sa-Correia I, Valvano MA. Lipopolysaccharide modification in gram-negative bacteria during chronic infection. *FEMS Microbiol Rev* 2016;**40**:480–93.
Martinez E, Cantet F, Bonazzi M. Generation and multi-phenotypic high-content screening of *Coxiella burnetii*&*nbsp*;transposon mutants. *J Vis Exp* 2015;**99**:e52851. https://doi.org/10.3791/52851
Martinez E, Cantet F, Fava L *et al.* Identification of OmpA, a *Coxiella burnetii* protein involved in host cell invasion, by multi-phenotypic high-content screening. *PLoS Pathog* 2014;**10**:e1004013.
Maturana P, Graham JG, Sharma UM *et al.* Refining the plasmid-encoded type IV secretion system substrate repertoire of *Coxiella burnetii*. *J Bacteriol* 2013;**195**:3269–76.
Maurin M, Raoult D. Q fever. *Clin Microbiol Rev* 1999;**12**:518–53.
McQuiston JH, Childs JE, Thompson HA. Q fever. *J Am Vet Med Assoc* 2002;**221**:796–99.
Miller JD, Shaw EI, Thompson HA *et al.* Fever and bioterrorism. In: Anderson B, Friedman H, Bendinelli M (eds.), *Microorganisms and Bioterrorism*. Boston, MA: Springer, 2006, 181–208.
Minnick MF, Raghavan R. Developmental biology of *Coxiella burnetii*. *Adv Exp Med Biol* 2012;**984**:231–48.
Mistry J, Chuguransky S, Williams L *et al.* Pfam: the protein families database in 2021. *Nucleic Acids Res* 2020;**49**:D412–D19.
Moormeier DE, Sandoz KM, Beare PA *et al.* *Coxiella burnetii* RpoS regulates genes involved in morphological differentiation and intracellular growth. *J Bacteriol* 2019;**201**:e00009–19.
Morgan JK, Luedtke BE, Shaw EI. Polar localization of the *Coxiella burnetii* type IVB secretion system. *FEMS Microbiol Lett* 2010;**305**:177–83.
Mori H, Ito K. The Sec protein-translocation pathway. *Trends Microbiol* 2001;**9**:494–500.
Nagai H, Kubori T. Type IVB secretion systems of *Legionella* and other gram-negative bacteria. *Front Microbiol* 2011;**2**:136.
Neckers L, Tatu U. Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell Host Microbe* 2008;**4**:519–27.
Newton HJ, Kohler LJ, McDonough JA *et al.* A screen of *Coxiella burnetii* mutants reveals important roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. *PLoS Pathog* 2014;**10**:e1004286.
Newton HJ, McDonough JA, Roy CR. Effector protein translocation by the *Coxiella burnetii* Dot/Icm type IV secretion system requires endocytic maturation of the pathogen-occupied vacuole. *PLoS One* 2013;**8**:e54566.
Omsland A, Beare PA, Hill J *et al.* Isolation from animal tissue and genetic transformation of *Coxiella burnetii* are facilitated by an improved axenic growth medium. *Appl Environ Microbiol* 2011;**77**:3720–5.
Omsland A, Cockrell DC, Howe D *et al.* Host cell-free growth of the Q fever bacterium *Coxiella burnetii*. *Proc Natl Acad Sci USA* 2009;**106**:4430–4.
Pan X, Lührmann A, Satoh A *et al.* Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 2008;**320**:1651–4.
Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55.
Pechstein J, Schulze-Luehrmann J, Bisle S *et al.* The *Coxiella burnetii* T4SS effector AnkF is important for intracellular replication. *Front Cell Infect Microbiol* 2020;**10**:559915.
Raoult D, Marrie T, Mege J. Natural history and pathophysiology of Q fever. *Lancet Infect Dis* 2005;**5**:219–26.
Rudolf Toman RAH, James ES, Jean-Louis M. *Coxiella Burnetii: Recent Advances and New Perspectives in Research of the Q Fever Bacterium. Advances in Experimental Medicine and Biology*. Netherlands: Springer, 2012.
Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 2003;**338**:2503–19.
Sanchez SE, Vallejo-Esquerra E, Omsland A. Use of axenic culture tools to study *Coxiella burnetii*. *Curr Protoc Microbiol* 2018;**50**:e52.
Sandoz KM, Beare PA, Cockrell DC *et al.* Complementation of arginine auxotrophy for genetic transformation of *Coxiella burnetii* by use of a defined axenic medium. *Appl Environ Microbiol* 2016;**82**:3042–51.

Sandoz KM, Popham DL, Beare PA *et al.* Transcriptional profiling of *Coxiella burnetii* reveals extensive cell wall remodeling in the small cell variant developmental form. *PLoS One* 2016;**11**:e0149957–e57.

Schäfer W, Eckart RA, Schmid B *et al.* Nuclear trafficking of the anti-apoptotic *Coxiella burnetii* effector protein AnkG requires binding to p32 and importin-α1. *Cell Microbiol* 2017;**19**.

Schäfer W, Schmidt T, Cordsmeier A *et al.* The anti-apoptotic *Coxiella burnetii* effector protein AnkG is a strain specific virulence factor. *Sci Rep* 2020;**10**:15396.

Segal G, Feldman M, Zusman T. The Icm/Dot type-IV secretion systems of *Legionella pneumophila* and *Coxiella burnetii*. *FEMS Microbiol Rev* 2005;**29**:65–81.

Segal G, Shuman HA. Possible origin of the *Legionella pneumophila* virulence genes and their relation to *Coxiella burnetii*. *Mol Microbiol* 1999;**33**:669–70.

Seshadri R, Paulsen IT, Eisen JA *et al.* Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci USA* 2003;**100**:5455–60.

Sexton JA, Miller JL, Yoneda A *et al. Legionella pneumophila* DotU and IcmF are required for stability of the Dot/Icm complex. *Infect Immun* 2004;**72**:5983–92.

Sexton JA, Vogel JP. Type IVB secretion by intracellular pathogens. *Traffic* 2002;**3**:178–85.

Stead CM, Omsland A, Beare PA, *et al.* Sec-mediated secretion by Coxiella burnetii. *BMC Microbiology* 2013;**13**:222. https://doi.org/10.1186/1471-2180-13-222

Takaya A, Tomoyasu T, Matsui H *et al.* The DnaK/DnaJ chaperone machinery of *Salmonella enterica* serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. *Infect Immun* 2004;**72**:1364–73.

Tsirigotaki A, De Geyter J, Šoštaric´ N *et al.* Protein export through the bacterial Sec pathway. *Nat Rev Microbiol* 2017;**15**:21–36.

van Schaik EJ, Chen C, Mertens K *et al.* Molecular pathogenesis of the obligate intracellular bacterium *Coxiella burnetii*. *Nat Rev Microbiol* 2013;**11**:561–73.

VanRheenen SM, Duménil G, Isberg RR. IcmF and DotU are required for optimal effector translocation and trafficking of the *Legionella pneumophila* vacuole. *Infect Immun* 2004;**72**:5972–82.

Vincent CD, Friedman JR, Jeong KC *et al.* Identification of the core transmembrane complex of the *Legionella* Dot/Icm type IV secretion system. *Mol Microbiol* 2006;**62**:1278–91.

Vincent CD, Friedman JR, Jeong KC *et al.* Identification of the DotL coupling protein subcomplex of the *Legionella* Dot/Icm type IV secretion system. *Mol Microbiol* 2012;**85**:378–91.

Vogel JP. Turning a tiger into a house cat: using *Legionella pneumophila* to study *Coxiella burnetii*. *Trends Microbiol* 2004;**12**:103–05.

Voth DE, Beare PA, Howe D *et al.* The *Coxiella burnetii* cryptic plasmid is enriched in genes encoding type IV secretion system substrates. *J Bacteriol* 2011;**193**:1493–503.

Voth DE, Heinzen RA. *Coxiella* type IV secretion and cellular microbiology. *Curr Opin Microbiol* 2009;**12**:74–80.

Voth DE, Heinzen RA. Lounging in a lysosome: the intracellular lifestyle of *Coxiella burnetii*. *Cell Microbiol* 2007;**9**:829–40.

Voth DE, Howe D, Beare PA *et al.* The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J Bacteriol* 2009;**191**:4232–42.

Wang S, Wang D, Du D *et al.* Advances in the assembly model of bacterial type IVB secretion systems. *Appl Sci* 2018;**8**:2368.

Weber MM, Chen C, Rowin K *et al.* Identification of *Coxiella burnetii* type IV secretion substrates required for intracellular replication and *Coxiella*-containing vacuole formation. *J Bacteriol* 2013;**195**:3914–24.

Weber MM. *Identification of C. burnetii type IV secretion substrates required for intracellular replication and Coxiella-containing vacuole formation. Ph. D. Thesis, Office of Graduate and Professional Studies*, Texas A&M University, 2014,126.

Williams JC, Waag DM. Antigens, virulence factors, and biological response modifiers of coxiella burnetii: strategies for vaccine development. *Q Fever: the Biology of Coxiella Burnetii*, Boca Raton : CRC Press. 1991, 175–222.

Yu NY, Wagner JR, Laird MR *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 2010;**26**:1608–15.

Zamboni DS, McGrath S, Rabinovitch M *et al. Coxiella burnetii* express type IV secretion system proteins that function similarly to components of the *Legionella pneumophila* Dot/Icm system. *Mol Microbiol* 2003;**49**:965–76.

Zusman T, Aloni G, Halperin E *et al.* The response regulator PmrA is a major regulator of the Icm/Dot type IV secretion system in *Legionella pneumophila* and *Coxiella burnetii*. *Mol Microbiol* 2007;**63**:1508–23.

Zusman T, Feldman M, Halperin E *et al.* Characterization of the icmH and icmF genes required for *Legionella pneumophila* intracellular growth, genes that are present in many bacteria associated with eukaryotic cells. *Infect Immun* 2004;**72**:3398–409.

Zusman T, Yerushalmi G, Segal G. Functional similarities between the Icm/Dot pathogenesis systems of *Coxiella burnetii* and *Legionella pneumophila*. *Infect Immun* 2003;**71**:3714–23.

Received 6 February 2023; **revised** 17 April 2023; **accepted** 15 May 2023