2-6-2023

Comparative Transcriptomics and Genomics from Continuous Axenic Media Growth Identifies

Archana Yadav

Melissa N Brewer

Mostafa S Elshahed

Edward I. Shaw
Philadelphia College of Osteopathic Medicine, ed.shaw@pcom.edu

1   **Comparative Transcriptomics and Genomics from Continuous Axenic Media**
2   **Growth Identifies *Coxiella burnetii* Intracellular Survival Strategies**
3
4   Archana Yadav [a], Melissa N. Brewer [a,b], Mostafa S. Elshahed [a], and Edward I. Shaw [a,c] *
5
6
7   [a] Department of Microbiology and Molecular Genetics. Oklahoma State University. Stillwater, OK.USA.
8   [b] Biological Sciences. Southeastern Oklahoma State University. Durant, OK. USA.
9   [c] Department of Biomedical Sciences. Philadelphia College of Osteopathic Medicine. Moultrie, GA. USA
10
11
12
13   ***Corresponding author:**
14   Edward I. Shaw
15   Philadelphia College of Osteopathic Medicine (PCOM)
16   Department of Bio-Medical Sciences
17   Moultrie, GA. USA
18   Phone: 229-668-3188
19   Email: ed.shaw@pcom.edu
20
21
22
23
24   Keywords: *Coxiella burnetii*, Reverse evolution, Type 4B secretion system, Effector proteins,
25   Intracellular survival.
26
27
28

29   **Abstract**

30   *Coxiella burnetii* (Cb) is an obligate intracellular pathogen in nature and the causative agent of acute Q

31   fever as well as chronic diseases. In an effort to identify genes and proteins crucial to their normal

32   intracellular growth lifestyle, we applied a "Reverse evolution" approach where the avirulent Nine Mile

33   Phase II strain of Cb was grown for 67 passages in chemically defined ACCM-D media and gene

34   expression patterns and genome integrity from various passages was compared to passage number one

35   following intracellular growth. Transcriptomic analysis identified a marked downregulation of the

36   structural components of the type 4B secretion system (T4BSS), the general secretory (sec) pathway, as

37   well as 14 out of 118 previously identified genes encoding effector proteins. Additional downregulated

38   pathogenicity determinants genes included several chaperones, LPS, and peptidoglycan biosynthesis. A

39   general marked downregulation of central metabolic pathways was also observed, which was balanced by

40   a marked upregulation of genes encoding transporters. This pattern reflected the richness of the media and

41   diminishing anabolic and ATP-generation needs. Finally, genomic sequencing and comparative genomic

42   analysis demonstrated an extremely low level of mutation across passages, despite the observed Cb gene

43   expression changes following acclimation to axenic media.

44
45
46

Introduction

47

48   *Coxiella burnetii* (Cb), the causative agent of acute and chronic Q fever (Arricau-Bouvery and Rodolakis

49   2005; Maurin and Raoult 1999; McQuiston, Childs and Thompson 2002; Miller, Shaw and Thompson

50   2006; Raoult, Marrie and Mege 2005; van Schaik, Chen, Mertens*, et al.* 2013), is an obligate intracellular

51   pathogen that infects macrophages, and successfully propagates in a parasitophorous vacuole termed the

52   *Coxiella* containing vacuole (CCV) (Heinzen, Hackstadt and Samuel 1999; Rudolf Toman 2012; Voth

53   and Heinzen 2007). Cb has evolved multiple strategies to tolerate and thrive in the CCV, in spite of the

54   prevailing low pH ($\approx$ 4.5), low O_2 content, oxygen radicals, and high level of degradative host factors

55   such as acid hydrolases and defensins (Brennan, Russell, Zhang*, et al.* 2004; Hackstadt and Williams

56   1981; Heinzen, Hackstadt and Samuel 1999; Omsland, Cockrell, Howe*, et al.* 2009). Such remarkable

57   ability has been the subject of a wide range of studies that employed a plethora of biochemical, genetic,

58   imaging, and –omics-based approaches. Further, Cb employs a type 4B secretion system (T4BSS) to

59   deliver effector proteins into the host throughout infection (Heinzen, Hackstadt and Samuel 1999; Rudolf

60   Toman 2012; van Schaik, Chen, Mertens*, et al.* 2013; Voth and Heinzen 2007; Voth and Heinzen 2009).

61   Cb effector proteins identified so far mediate a variety of biochemical activities and are known to target

62   and modulate a broad array of host functions (Beare, Gilk, Larson*, et al.* 2011; Beare, Sandoz, Larson*, et*

63   *al.* 2014; Crabill, Schofield, Newton*, et al.* 2018; Larson, Beare, Voth*, et al.* 2015; Larson and Heinzen

64   2017; Newton, Kohler, McDonough*, et al.* 2014; Newton, McDonough and Roy 2013; van Schaik, Chen,

65   Mertens*, et al.* 2013). Prior studies have employed bioinformatic tools (Chen, Banga, Mertens*, et al.*

66   2010), transposon mutagenesis (Beare, Gilk, Larson*, et al.* 2011; Carey, Newton, Luhrmann*, et al.* 2011;

67   Crabill, Schofield, Newton*, et al.* 2018; Martinez, Cantet, Fava*, et al.* 2014; Weber, Chen, Rowin*, et al.*

68   2013), microscopic localization studies (Chen, Banga, Mertens*, et al.* 2010; Howe, Melnicakova, Barak*,*

69   *et al.* 2003; Morgan, Luedtke and Shaw 2010; Voth, Beare, Howe*, et al.* 2011) and cloning and infectivity

70   testing to identify and characterize effector proteins. In addition, *Legionella pneumophila*, a close genetic

71   neighbor of Cb with a very similar T4BSS (Nagai and Kubori 2011; Segal and Shuman 1999; Sexton and

72   Vogel 2002), is known to use T4BSS-effector protein duality to infect its natural host cell, the amoeba. *L.*

73   *pneumophila* has been extensively used as a proxy to identify putative effector proteins and propose

74   molecular pathogenesis mechanisms in Cb (Pan, Lührmann, Satoh*, et al.* 2008; Segal, Feldman and

75   Zusman 2005; Vogel 2004; Zamboni, McGrath, Rabinovitch*, et al.* 2003; Zusman, Yerushalmi and Segal

76   2003). Indeed, research on *L. pneumophila* has identified the structural features of the T4BSS, the nature

77   of effector proteins secreted through the system, and possible function of some of these effectors.

78   Growth of Cb in an axenic media was first reported in 2009 using the undefined Acidified Citrate

79   Cysteine Media (ACCM) media (Omsland, Cockrell, Howe*, et al.* 2009). Increased replication rates in the

80   somewhat more defined ACCM-2 medium soon followed in 2011 (Omsland, Beare, Hill*, et al.* 2011).

81   Subsequently, a nutritionally fully defined media (ACCM-D) with an even greater replication rate and

82   physiologic parallels to intracellular bacteria was developed (Sanchez, Vallejo-Esquerra and Omsland

83   2018). Growing Cb in axenic media is opening new venues for investigating mechanisms of Cb molecular

84   pathogenesis (Beare and Heinzen 2014; Beare, Jeffrey, Long*, et al.* 2018; Beare, Larson, Gilk*, et al.* 2012;

85   Crabill, Schofield, Newton*, et al.* 2018; Martinez, Cantet and Bonazzi 2015; Rudolf Toman 2012;

86   Sandoz, Beare, Cockrell*, et al.* 2016). Theoretically, when grown in axenic media, the expression of genes

87   required for intracellular survival and host cell manipulation is no longer required for Cb viability. As

88   such, continuous maintenance and passaging the bacterium for extended periods of times under axenic

89   conditions could potentially remove the powerful selective pressure exerted by the host cell, thus

90   potentially minimize/silence expression in such genes. As such, we posit that transcriptomic analysis of

91   gene expression patterns as well as genomic identification of mutation and gene loss patterns in axenic

92   grown versus Cb cultures derived from intracellular growth could be employed for identifying putative

93   involvement of specific genes, as well as identification of novel genes necessary for Cb pathogenesis and

94   survival in an intracellular environment. Similarly, continuous passaging could also lead to the

95   propagation of mutations, DNA fragment losses, and rearrangements in genes/loci associated with

96   intracellular survival, pathogenesis, and host cell manipulation. Such patterns could be regarded as

97   "reverse evolution" i.e., the opposite of the natural evolution trajectory of Cb from a free-living ancestor

98   to an obligate intracellular pathogen. Specifically, we hypothesized that: 1) changes in gene expression

99   within the first few passages upon transition from intracellular to axenic media growth would be

100   observed, and such differences would be more pronounced in genes involved in subverting and coopting

101   host metabolism, as well as genes enabling general adaptation to physiological conditions prevalent in its

102   intracellular vacuolar environment, and 2) Cb could acquire and accumulate DNA mutations upon

103   transition from intracellular to axenic media growth after repetitive passages since certain bacterial

104   genes/proteins are no longer required for successful growth.

105   In this study, we transitioned Cb Nine Mile phase II from cell cultures into axenic defined media

106   ACCM-D and subcultured it into a long-term successive passage. We conducted transcriptomic and

107   genomic sequencing on replicate samples at different time points (passages) to document temporal

108   changes in gene expression patterns, and DNA mutations associated with adaptation to an axenic

109   extracellular lifestyle.

110

Materials and Methods

Microorganism and growth conditions. *Coxiella burnetii* 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 in T75 culture flasks. This method of collecting cells was adapted from (Coleman, Fischer, Howe, *et al.* 2004). Briefly, the infected cell line was split into multiple non-vented 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 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 x g, 15 minutes) in 250 ml Nalgene round bottom tubes, scrapped off the round bottom tubes by using sterile 1X 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 x g for 15 minutes. The SCV pellets obtained were stored in SPG freezer media (0.7 M sucrose, 3.7 mM KH_2PO_4, 6.0mM 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 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.

6

135   **Measuring Growth and Host Cell Infectivity.** To determine the infectivity of axenic- or intracellularly

136   grown Cb; HeLa cells (CCL-2; American Type Culture Collection) were seeded onto 96 well culture

137   plates at a density of 10^4 in Roswell Park Memorial Institute (RPMI) medium containing 2% fetal bovine

138   serum (FBS) for 16 hours. Cb cultures grown in ACCM-D were pelleted at 12000 x g at 4°C for 15

139   minutes. Serially passaged Cb were diluted in RPMI to normalize the number of genomes per volume,

140   and 50 μl from various dilutions were inoculated onto the HeLa cell containing wells and centrifuged at

141   600 x g for 15 minutes at room temperature (Luedtke, Mahapatra, Lutter, *et al.* 2017). Immediately

142   following centrifugation, the inoculating media was replaced with 200 μl of fresh RPMI containing 2%

143   FBS. The plates were incubated at 37°C and 5% CO_2 for 72 hours, fixed with ice-cold methanol for 10

144   minutes, then examined using indirect fluorescent antibody microscopy analysis as described previously

145   (Luedtke, Mahapatra, Lutter, *et al.* 2017). Briefly, *C. burnetii* was stained using rabbit whole anti-*C.*

146   *burnetii* NMII antibody diluted 1:1000 in PBS containing 3% bovine serum albumin (BSA) as a blocking

147   agent. Primary antibodies were detected using Alexa Fluor 488 labeled goat anti-rabbit IgG antibodies

148   diluted 1:1000 in PBS containing 3% BSA (Invitrogen). Total DNA was stained using 4',6-diamidino-2-

149   phenylindole (DAPI) diluted 1:10000 in PBS containing 3% BSA (Molecular Probes) to illuminate host

150   cell nuclei. The methanol fixed and stained cultures were visualized on a Nikon Eclipse TE2000-S and

151   the number of maturing CCVs were counted and calculations performed to ascertain the number of

152   fluorescence forming units, which indicates the infectivity of the *C. burnetii* NMII in the ACCM-D

153   samples.

154   **Transcriptomics.**

155   **RNA extraction.** Cells from axenic media growth passages 1, 3, 5, 10, 12, 16, 21, 31, 42, 51, 61 and 67

156   were harvested for transcriptomic analysis. RNA was extracted using a combination of hot Trizol

157   treatment (Moormeier, Sandoz, Beare, *et al.* 2019) and the RNeasy Mini kit (Qiagen, Germany). Briefly,

158   bacteria in 12 ml of ACCM-D culture (OD_{600} ~ 0.3-0.4) were pelleted, resuspended in 700 μl of Trizol

159   (ThermoFisher Scientific), boiled at 90°C for 10 min, and vortexed vigorously. 200 μl Chloroform was

7

160   then added, followed by centrifugation at 12,000 x g at 4°C for 10 min. After separation, 300 μl of 100%

161   ethanol was added to the aqueous phase, which was then quickly transferred to the spin column provided

162   in the RNeasy mini kit. On-column DNA digestion was conducted by adding 80 μl (10 μl 1 Unit/μL

163   RNase free DNase I, (ThermoFisher Scientific) in 70 μl reaction buffer from the Master Pure Yeast RNA

164   Purification kit, Epicenter) of DNase preparation. The RNeasy mini kit's protocol was followed for

165   washing and eluting RNA. RNA quality was assessed visually on a gel as well as using RNA screen tape

166   (Agilent) and RNA integrity number (RIN) value measurements using Tapestation and Bioanalyzer

167   systems (Agilent).

168   **Transcriptome sequencing and assembly.** RNA sequencing (RNA-Seq) was conducted on the Illumina

169   platform, using Nextseq 500 sequencer at Oklahoma State University Genomics and Proteomics core

170   facility. Trimmomatic v0.38 (Bolger, Lohse and Usadel 2014) was used to process raw reads and remove

171   Illumina adapter sequences. HISAT2 v2.1.0 (Kim, Paggi, Park, *et al.* 2019) was used to map the trimmed

172   reads to the Chromosome (GenBank accession number: CP020616.1) and Plasmid (GenBank accession

173   number: CP020617.1) of Cb NMII RSA 439. StringTie v2.1.4 (Kovaka, Zimin, Pertea, *et al.* 2019) was

174   used to assemble reads alignments into potential transcript and to generate a non-redundant set of

175   transcripts. The Python script prepDE.py supplemented with StringTie tool was used to convert

176   transcripts per kilobase million (TPM) and fragments per kilobase million (FKPM) to gene level raw

177   count matrix. The raw count table was imported to DESeq2 package (Love, Huber and Anders 2014)

178   from Bioconductor in R programming language for further analysis.

179   **Identification and analysis of Differentially Expressed Genes (DEGs).** The overall strategy for

180   comparative transcriptomics analysis is outlined in Figure 1. DESeq2 was used to compute the fold

181   change expression levels (reflected by logarithmic two-fold expression change i.e., L2fc) and its statistical

182   significance (adjusted p-value, padj henceforth referred to as p-value) for every gene between passages

183   when compared to passage one. DEseq2 tests the differential expression using negative binomial

184   distribution and internally normalizes the counts by library size (Anders and Huber 2010). Genes with a

185   p-value < 0.05 were labeled as significantly expressed. Only genes with TPM values > 10 in at least one

8

186   passage were considered to minimize noise from minimally expressed genes. In most cases, differentially

187   expressed genes in our temporal analysis were significantly expressed in more than one sampling point. In

188   the few cases where differential expression was observed as a single spike in only one time point, a

189   threshold of L2fc > 2 was considered as differentially expressed. Patterns of differential expression in

190   DEGs was analyzed and visualized by constructing plot count graphs using the function "plotCounts" in

191   DESeq2 package. The package "EnhancedVolcano" was used to visualize gene expression patterns as

192   volcano plots. Differentially expressed patterns are classified into 1- Early up/downregulated, i.e.,

193   differential expression occurred in early (before passage 31) and the levels were sustained in subsequent

194   late passages (see Figure 1). 2- Continuously up/downregulated, i.e., a constant/gradual increase in the

195   magnitude of L2fc was observed throughout the sampling process (see Figure 1). 3- Late

196   up/downregulated, i.e., differential expression was observed at or after passage 31. 4- Variable, i.e.,

197   expression levels were significantly higher than passage one in some timepoints and significantly lower

198   than passage one in other time points (see Figure 1).

199   **Metabolic analysis and pathway mapping of DEGs.** The subcellular protein localizations of proteins

200   encoded by DEGs were predicted by using PSORTb (Yu, Wagner, Laird, *et al.* 2010). Transporter

201   Classification Database (TCBD) was queried to find the putative transporter proteins. Pfam database

202   (Mistry, Chuguransky, Williams, *et al.* 2020) was used to identify putative protein families for

203   hypothetical proteins. BlastKOALA (Kanehisa, Sato and Morishima 2016) was used for functional

204   annotation and assign KEGG Orthology (KO) numbers for the selected differentially expressed genes;

205   and KEGG mapper (Kanehisa and Sato 2020) was then used to reconstruct metabolic pathway to

206   visualize the differentially expressed genes in each pathway. The gene involvement in specific metabolic

207   pathways were inferred from KEGG brite hierarchy file. Cluster of Orthologous genes (COGs) database

208   (updated 2020) (Galperin, Wolf, Makarova, *et al.* 2021) downloaded from NCBI, was used to classify the

209   effector proteins into functional categories.

210   **Genomics.**

211   **DNA extraction and sequencing.** 8 ml of Cb cultures grown in ACCM-D for 7 days were pelleted by

212   centrifugation at 12,000 x g and 4°C for 15 minutes. DNA extraction was conducted using Pure Link®

213   Genomic DNA Kits (ThermoFisher Scientific) following the manufacturer's instructions. Sequencing was

214   conducted at Oklahoma State University Genomics and Proteomics core facility using Illumina's

215   NextSeq® 500 System. DNA quality was assessed visually on a gel as well as using DNA Screentape and

216   Bioanalyzer systems (Agilent).

217   **Genome assembly and quality control.** The KBase platform (Allen, Drake, Harris*, et al.* 2017), which

218   implements and integrates multiple bioinformatic tools, was used for DNA sequence data handling.

219   Trimmomatic v 0.36 (Bolger, Lohse and Usadel 2014) was used to trim the Illumina adapter sequences.

220   Quality check was done using FastQC v0.11.5 (Andrews 2010). Assembly of Illumina reads to contigs

221   was attempted using four different assemblers (Spades v3.13.0, Velvet v1.2.10 and IDBA-UD v1.1.3 and

222   Unicycler (Davis, Wattam, Aziz*, et al.* 2020)). The quality of genome assemblies from these four

223   assemblers were assessed using QUAST v1.4 (Gurevich, Saveliev, Vyahhi*, et al.* 2013) and the best

224   assemblies were selected using metrices such as total length, largest N50, lesser number of contigs and

225   less Ns. CheckM (Parks, Imelfort, Skennerton*, et al.* 2015) was used to assess quality and completion of

226   genomes (Figure 1).

227   **Analysis of mutation frequencies.** Breseq (Deatherage and Barrick 2014) was used to identify

228   mutations/changes in the genome assemblies obtained, with Passage one used as a reference. The

229   occurrence and frequency of both single nucleotide polymorphisms (SNPs) and deletion-insertion

230   polymorphisms (DIPs) were examined (as outlined in Figure 1). Breseq was run in polymorphism mode,

231   which identifies the mutations occurring in a fraction of a population in addition to consensus mutations

232   in the entire population in a sample. This allows for the visualization of the propagation of a particular

233   mutation as a frequency of evolved alleles and genetic diversity in the population.

234   **Nucleotide sequences accession number.** The whole-transcriptome and genome shotgun sequences were

235   deposited in GenBank under the BioProject PRJNA796300 and BioSample accession numbers

236   SAMN24840407-SAMN24840437 and SAMN24847762-SAMN24847773. The 31 transcriptomic

10

237   assemblies were deposited in the SRA under project accession number SRX13723330-SRX13723360.

238   Reads for 12 genomic assemblies can be found under SRA with accession SRX13726189-

239   SRX13726200.

240

241 **Results**

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

243 Following anecdotal observations, we sought to quantitatively assess whether serially passaged Cb infect

244 cultured cells less readily than cell derived bacterial stocks. Using *C. burnetii* NMII serially passaged 1, 3,

245 5, and 10 times in ACCM-D, we initiated infections of Hela cells with bacterial dilutions normalized by

246 the number of genomes in each sample. When the number of fluorescence forming units (FFU) per

247 sample were calculated, they revealed a decrease in the number of *C. burnetii* filled vacuoles in tissue

248 culture cells as the bacteria from subsequent passages were analyzed, respectively, resulting in a nearly

249 two-log decrease between Passages 1 and 10 (Figure 2A). This indicated that there were fewer bacteria

250 per genome that were capable of initiating a typical infection following multiple passages in axenic

251 media. Next, we sought to determine if the decrease in infectivity of tissue culture cells was associated

252 with a decrease in *in vitro* viability of the *C. burnetii* as measured by colony forming units on ACCM-D

253 agar. To address this question, we plated dilutions of passages 1, 3, 5, and 10 on ACCM-D agar plates

254 and performed colony counts. Contrary to the decrease in infectious units (Figure 2A), the colony counts

255 indicated that there was no significant change in viable bacteria relative to genomes as the organism was

256 serially passaged (Figure 2B). This indicated that the number of live and replicative bacteria did not

257 change during axenic growth, and therefore bacterial death was not responsible for the decrease in Cb

258 infectivity of the cultured eukaryotic cells observed.

259 **Transcriptional activity.** RNA sequencing was conducted on 12 different passages (1, 3, 5, 10, 12, 16,

260 21, 31, 42, 51, 61, and 67). A total of 162.2 Gb data were obtained, with 6.05 – 22.57 million reads per

261 sample (Average = 10.14 million reads). Transcripts representing each of the 2,217 genes in *C. burnetii*

262 NMII strain (genome and plasmid) were identified in all samples, attesting to the depth of the sequencing

263 effort conducted.

264 Expression level and overall pattern (Early up, Continuous up, Late up, Early down, Continuous

265 down, Late down, Variable) for every gene in the Cb genome is shown in Table S1. A total of 845 genes

266 were differentially expressed in at least one passage, with 464 upregulated and 371 downregulated

267   (Figure 3a) genes. The number of differentially expressed genes (DEGs) per passage ranged between 25

268   and 807 (Figure 3b). A general pattern of an increasing number of differentially expressed genes per

269   passage was observed through passage 51, after which the number of DEGs dropped in passage 61 and 67

270   (Figure 3B). The ratio of upregulated to downregulated genes in each passage ranged between 0.14 (in

271   passage 5) and 1.26 (in passage 3). Of the 371 downregulated genes, 249 expressed early down pattern,

272   48 were continuous down, 43 were late down, and 31 were down in only one passage. Of the 464

273   upregulated genes, 288 were early up, 38 were continuous up, 85 were late up and 53 were up in only one

274   passage (Figure 3C). Of the 845 DEGs, 81 were differentially regulated in 8-11 of the passages, 144 in 5-

275   7 of the passages, 526 in 2-4 of the passages and 84 in only one passage (Figure 3D).

276   Visualization of DEGs patterns using volcano plots was used to provide an overview of the

277   overall level of expression changes (see Figure 4). Transcript expression levels from each passage were

278   compared to passage one. The labeled boxes within each plot analysis represents the 10 highest

279   differentially expressed transcripts (i.e., smallest p-value). Visual inspections demonstrate that chaperons

280   and T4BSS machinery proteins consistently represent an important component of highly downregulated

281   genes in all passages. Below, we provide a more detailed assessment on differential expression patterns

282   for various genes and pathways.

283   **Secretory pathways are significantly downregulated in axenic growth media.** The defective in

284   organelle trafficking/intracellular multiplication (Dot/Icm) Type IVB secretion system (T4BSS) in Cb has

285   been shown to secrete the effectors and other pathogenic determinants into the host cell, a process

286   required for Cb intracellular growth and pathogenesis (Beare, Gilk, Larson*, et al.* 2011; Carey, Newton,

287   Luhrmann*, et al.* 2011; Rudolf Toman 2012; van Schaik, Chen, Mertens*, et al.* 2013; Voth and Heinzen

288   2009). Interestingly, 19 out of the 24 components of the Cb T4BSS demonstrated significant differential

289   expression (Figure 5A, Figure S1a, Table S1). Out of these, 18 genes were downregulated and only one

290   gene was upregulated (Figure 5A, Figure S1a). Indeed, T4BSS encoded gene transcripts were some of the

291   most significantly downregulated across the passages (see Figure 4).

13

292   Within the T4BSS core transport complex, transcripts for genes *dotC*, *dotD*, *dotF*, and *dotG* were

293   early downregulated, while only *dotH* indicated no significant change in gene expression in all passages.

294   In addition, expression changes in genes encoding components of the T4BSS coupling protein complex

295   (*dotL*, *dotM* and *icmW*) demonstrated early down or continuous down (*dotN*), whereas *icmS* was the only

296   component with no significant gene expression changes. Transcripts of the gene *dotB* was continuous

297   downregulated whereas *dotA* and *icmX* were found to be early downregulated (Figure 4 and Figure S1a).

298   Besides the two main complexes, other components of the Cb T4BSS that were transcriptionally

299   downregulated during continuous axenic media passaging includes genes *dotE, dotP, dotK, dotI, dotJ,*

300   *icmT* and *icmQ* (Figure 5a). *icmF*, located in a separate locus than the majority of the T4BSS genes

301   (Figure 5b), was the only component that was transcriptionally upregulated in the system (Figure S1a).

302   Transcript expression of genes within additional secretory pathways in Cb were also analyzed.

303   Genes of the general secretary (Sec) pathway revealed a general trend of downregulation (Figure S1b,

304   Table S1). The Sec pathway provides a channel for polypeptide movement across the bacterial inner

305   membrane (Green and Mecsas 2016). It is comprised of the proteins SecY, SecE and SecG and an

306   ATPase (SecA) that drives protein movement (Green and Mecsas 2016; Tsirigotaki, De Geyter, Šoštaric´,

307   *et al.* 2017). This pathway is known to secrete proteins from the cytosol through the cytoplasmic

308   membrane (Mori and Ito 2001). We identified all of the Cb Sec pathway components, as shown in Figure

309   5a. Transcripts for expression of the inner membrane proteins SecA, SecF and SecE were early down and

310   SecY and YajC were continuously down, whereas the targeting proteins SecB and SecG were down at

311   later passages.

312   **Expression patterns of T4BSS effector proteins previously implicated in Cb pathogenesis and**

313   **intracellular survival.** The differential expression in all 118 genes encoding T4BSS effector proteins

314   previously identified in Cb through a variety of effector screens (Carey, Newton, Luhrmann*, et al.* 2011;

315   Chen, Banga, Mertens*, et al.* 2010; Larson, Martinez, Beare*, et al.* 2016; Voth, Beare, Howe*, et al.* 2011;

316   Weber, Chen, Rowin*, et al.* 2013) was examined. Forty-seven effector proteins were differentially

14

317   expressed (column "Effector proteins" in Table S1, Figure 3a). Interestingly, more genes were

318   upregulated (n=33) than downregulated (n= 14) (Figure 3a).

319   The 14 genes encoding effector proteins that were transcriptionally downregulated indicated an

320   up to four-fold expression change, with the expression changes primarily beginning from passage 10

321   (Figure S2). These genes fell into COG functional groups of signal transduction mechanisms (*ankG*, *ankK*

322   and *ankD*), carbohydrate transport and metabolism (B7L74_09020), posttranslational modification,

323   protein turnover and chaperones (*cpeH*), replication, recombination, and repair (*cig57*), lipid transport and

324   metabolism (B7L74_03275) and mobilome: prophages and transposons (B7L74_08400) (Table S2,

325   Figure 6a, Figure 6b). Genes B7L74_08200 and *cpeF* were predicted as general function categories

326   whereas there were no functional homologies for B7L74_07850, *cig2*, *cirC* and B7L74_03065 in the

327   COG database (Table S2).

328   **Expression patterns of additional pathogenic determinants in Cb.** Expression patterns of additional

329   pathogenic determinants unrelated to effector proteins were examined. Of these, we noted three

330   interesting patterns. First, the general downregulation of a wide range of chaperone proteins. Chaperone

331   proteins primarily function as protein folding catalyst, but many are considered virulence factors for many

332   intracellular pathogens given that they encounter stress related to phagosome acidification and phagosome

333   fusion with lysosomes (Neckers and Tatu 2008). Amongst the 16 genes annotated as chaperons in the Cb

334   genome, 10 were transcriptionally downregulated during continuous *in vitro* passaging (Figure 7a) (Table

335   S1). Notable downregulated chaperones include glutaredoxins (*grxC* and *grxD)* that have been shown to

336   be involved in CCV detoxification (Beare, Unsworth, Andoh*, et al.* 2009) and genes encoding heat shock

337   proteins classes such as *dnaK, hptG, groEL* and *dnaJ*. These proteins are known to help bacteria adapt to

338   stressful conditions (Arnold, Jackson, Waterfield*, et al.* 2007; Genevaux, Georgopoulos and Kelley 2007)

339   (Figure S3). DnaK has been shown to be critical for survival of pathogenic bacteria inside the

340   macrophage (Takaya, Tomoyasu, Matsui*, et al.* 2004) and is induced in Cb in high acid condition, the

341   condition similar to the phagolysosome (Macellaro, Tujulin, Hjalmarsson*, et al.* 1998).

15

342   The second observation is the downregulation of several genes involved in lipopolysaccharide

343   biosynthesis. The Lipopolysaccharide (LPS) layer has long been known as a pathogenic determinant and

344   important for the host interaction in *C. burnetii* (Gajdosova, Kovacova, Toman*, et al.* 1994; Hussein,

345   Kovacova and Toman 2001; Williams and Waag 1991). Out of 35 genes related to LPS synthesis and O-

346   antigen nucleotide sugar biosynthesis, 11 were downregulated (Figure 7a) (Column "LPS and O-antigen

347   biosynthesis" in Table S1). Out of 11 downregulated genes, 4 genes are involved in KDO2-lipid IVA

348   Wbp pathway for LPS biosynthesis whereas 8 genes are involved in O-antigen nucleotide sugar

349   biosynthesis (Table S1). Genes involved in the first 3 steps i.e., UDUDP-N-acetylglucosamine

350   acyltransferase (*lpxA),* UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase *(lpxC)* and P-3-

351   O-[3-hydroxymyristoyl] glucosamine N-acyltransferase *(lpxD)* and the gene D-glycero-D-manno-heptose

352   1,7-bisphosphate phosphatase (*gmhB)* in LPS biosynthesis pathway are early down or continuously down

353   (Figure S3) (Figure 7C). In addition, 3 transporters related to LPS synthesis i.e., a lipoprotein releasing

354   system ATP-binding protein (*lolD),* a lipid flippase important in cell membrane formation (*pglK)* and a

355   probable O-antigen/lipopolysaccharide transport ATP-binding protein (*rfbE)* were also early down

356   (Figure 7C). 8 downregulated genes including *wbpW, gmhB, galE, wbpD, galE, wbpI, cap1J* and *glmU*

357   are involved in O-antigen nucleotide sugar biosynthesis pathway, a 14 genes pathway which is the first

358   step in O-antigen biosynthesis where nucleotide sugars are assembled and activated by adding NTP

359   (Samuel and Reeves 2003).

360   Finally, 3 of the 15 genes involved in peptidoglycan layer biosynthesis were downregulated

361   (Table S1). These genes are penicillin-binding protein PBP3/*ftsI*, penicillin binding protein PBP1A/*mrcA*

362   and undecaprenyl diphosphate synthase *uppS*. The peptidoglycan layer in Cb is an immunogenicity

363   determinant and thickens substantially during LCV to SCV transition to help in environment resistance

364   (Amano, Williams, McCaul*, et al.* 1984; Sandoz, Popham, Beare*, et al.* 2016). *ftsI* and *mrcA* are the only

365   two genes encoding penicillin binding proteins in this genome that are involved in peptide cross linking.

366   This suggests that the peptidoglycan layer may be even thinner than is in common intracellularly and

367   could indicate a reduced requirement during axenic growth since the thick SCV peptidoglycan layer in Cb

368   has been correlated to bacteria being more infectious (Sandoz, Popham, Beare, *et al.* 2016).

369   **Downregulation of multiple hypothetical proteins could suggest novel pathogenicity determinants.**

370   We posited that downregulated hypothetical proteins could represent previously unrecognized

371   pathogenicity determinants, and that such identification could be a useful starting point for subsequent

372   experimental validation. We identified 30 Cb genes encoding hypothetical proteins that were

373   downregulated in more than 4 separate passages. Subcellular localization analysis software predicts that

374   13 are cytoplasmic, 8 inner membrane, 1 extracellular, 1 periplasmic and 7 with unknown localization

375   (Table S4). Of the cytoplasmic proteins, Cig28 and an AMP binding protein (CBU_0787) possess a

376   regulatory element recognized by PmrA, a sequence related to T4BSS expression and translocation and

377   thus a potential predictor of effector proteins (Beare, Sandoz, Larson, *et al.* 2014), although they were

378   subsequently shown not to be translocated by the Cb T4BSS (Beare, Sandoz, Larson, *et al.* 2014;

379   Zusman, Aloni, Halperin, *et al.* 2007). Similarly, an uncharacterized protein, CBU_1234, has been shown

380   to have a glutamate-rich C-terminal secretion signal (E-block), which is also a predictor of effector

381   proteins (Weber 2014). Two Glycosyltransferase family 1 proteins (i.e. CBU_0839 and CBU_0841) have

382   previously been linked to LPS mutations that lead to phase transitions (Beare, Jeffrey, Long, *et al.* 2018).

383   The eight proteins localized in the inner membrane included Cig3, an immunoreactive peptidase

384   CBU_0215 (Weber 2014) previously shown to contain a regulatory element recognized by PmrA but not

385   translocated by Dot/Icm system (Beare, Sandoz, Larson, *et al.* 2014; Zusman, Aloni, Halperin, *et al.*

386   2007), an immunoreactive protein CBU_1865 (Beare, Chen, Bouman, *et al.* 2008) and a DUF3971

387   domain-containing protein CBU_1468 that has been shown to be important for intracellular replication

388   (Newton, Kohler, McDonough, *et al.* 2014) (Table S4).

389   The one differentially expressed hypothetical protein identified by the localization software as an

390   extracellular protein was CBU_0962; a predicted short chain dehydrogenase with a yet unknown specific

391   function (Bewley 2015). Lastly, proteins with unknown localization included an exported protein Cig40

392   (Weber 2014) with a regulatory element recognized by PmrA, a hypothetical surface antigen Com1(Chen,

17

393   Banga, Mertens, *et al.* 2010) and a hypothetical protein CBUA0012 located in an ORF containing other

394   plasmid effectors but has shown to be not secreted by Dot system (Voth, Beare, Howe, *et al.* 2011) (Table

395   S4).

396   **Transcriptional patterns of central metabolic pathways.** Analysis of gene expression patterns of

397   central metabolic pathways demonstrated a general trend of down-regulation in genes encoding enzymes

398   in central catabolic, amphibolic, and anabolic pathways, coupled with a broad upregulation in genes

399   encoding transporters. An overall pattern of downregulation of glycolysis genes (8/18 genes) was

400   observed (Figure 7A), with several enzymes such as pyruvate dehydrogenases (*pdhC, pdhD*), fructose-

401   bisphosphate aldolase (*fbaA*), glyceraldehyde 3-phosphate dehydrogenase (*gapA*), and

402   phosphoenolpyruvate carboxykinase (*pckA*) downregulated early in the passaging (Figure S3, Table S1).

403   Gene *pmm-pgm,* which encodes the enzyme phosphomannomutase/ phosphoglucomutase and is involved

404   in the first step of glycolysis, was down early as well (Figure S3). Similarly, analysis of genes encoding

405   enzymes of the tricarboxylic acid (TCA) cycle, demonstrated an overall downregulation (12/17 genes),

406   with a downregulation of ~2 fold in isocitrate dehydrogenase (*IDH2*), the rate-limiting enzyme in the

407   TCA cycle (Figure S3). As well, the downregulations in multiple carbohydrate dehydrogenases (e.g.,

408   pyruvate dehydrogenases *pdhC* and *pdhD*), succinate dehydrogenases (*sdhA, sdhB, sdhD*), as well as

409   genes that are necessary for oxidation of glycolytic and TCA cycle sugar intermediates was observed.

410   Finally, genes encoding components of the electron transport chain (ETC) were also downregulated.

411   These include Complex I: NADH-quinone oxidoreductase (*nuoB, nuoD, nuoE, nuoF, nuoH, nuoI, nuoK,*

412   *nuoL, nuoM, nuoN*), Complex II:Succinate dehydrogenase (*sdhA, sdhB, sdhD*), Complex III:Cytochrome

413   oxidoreductase (*cyoC, cyoD, cydA, cydB, cydX*), and Complex V:F-type ATPase (*atpA, atpB, atpE, atpD,*

414   *atpF, atpG*) (Figure 7A, Figure S3). The downregulations in two complexes (*cyoC* and *cyoD*) of

415   cytochrome o oxidase, which is known to be induced in oxygen rich growth conditions in bacteria,

416   (Cotter, Chepuri, Gennis, *et al.* 1990) suggests a decreased affinity and/or competition for oxygen in the

417   cell-free growth environment, as previously suggested (Kuley, Bossers-deVries, Smith, *et al.* 2015).

418   Cytochrome d oxidase, which is shown to be expressed more in oxidative and nitrosative stress conditions

18

419   (Cotter, Chepuri, Gennis*, et al.* 1990), also has expression of two of its components (i.e., *cydA* and *cydX*)

420   early down and late down, respectively. Finally, an overall downregulation of fatty acids biosynthesis

421   genes transcription (8/15) was also observed (Figure 7A, Figure S3).

422   In contrast to the general trend of downregulation of the central metabolic machinery of Cb*,* a

423   marked upregulation of genes encoding transporters was observed. Out of 125 general transporters, the

424   transcription of 39 were upregulated and 10 were downregulated (Column "Transporters" in Table S1)

425   (Figure 7a). Transporters that were upregulated have double fold expression change (L2fc) ranging from

426   3-10 (Figure S3). Upregulated primary transporters included transporters for the amino acid arginine,

427   oligopeptides, fatty acids, and vitamins such as riboflavin and thiamin (Figure S1) (Figure 7b) as well as a

428   small number of transporters (4 out of 20 present) related to synthesis and maintenance of outer

429   membrane (Table S1). On the other hand, upregulated secondary transporters included MFS transporters,

430   symporters, antiporters, and mechanosensitive ion channels. Of these, a notable observation was made

431   where all six MFS transporters and two out of four Na+ symporters found in Cb genome were found to be

432   early upregulated (Table S1). These MFS transporters transports various compounds such

433   as monosaccharides, oligosaccharides, amino acids, peptides, vitamins, cofactors, drugs, nucleobases,

434   nucleosides, and organic and inorganic anions and cations. Also, a large proportion of transporters related

435   to drug resistance (6 /16 present in the genome) and ion transporters mediating the uptake of ions such as

436   copper, iron, fluoride, selenite, cobalt-cadmium-zinc, and phosphate were also upregulated (Table S1)

437   (Figure 7b). In addition to transporters mediating substrate transport, transporters involved in pH

438   homeostasis such as ions/mechanosensitive channels and Na+:H+ antiporter were also upregulated.

439   (Figure 7b) (Table S1). The Na+:H+ antiporter functions to utilizes the proton motive force to efflux

440   intracellular sodium ions for intracellular pH homeostasis (Ito, Morino and Krulwich 2017) and these

441   antiporters along with ion/mechanosensitive channels have been proposed to play an important role in pH

442   homeostasis and survival within the acidic PL (Seshadri, Paulsen, Eisen*, et al.* 2003). Lastly, 3 out of 9

443   transporters classified under general or unknown functions were upregulated as well (Figure 7b).

444   **Genomics reveals a stable Cb genome.** For all 12 passages analyzed, genomes with 100% completeness

445   (assessed by identifying all 265 housekeeping marker genes specific for the Proteobacteria (Parks,

446   Imelfort, Skennerton, *et al.* 2015)) were obtained. N50 of genomic assemblies ranged between 49,903

447   and 75,629, N90 ranged between of 15,966 and 20,406, and the number of contigs per genome ranged

448   between 56 and 64 (Table S5). Using Passage one as a reference, we identified 842 unique single

449   nucleotide polymorphisms (SNPs) and 118 unique deletions/insertion polymorphisms (DIPs) (Table S5)

450   (Figure 8).

451   Of 842 unique SNPs, only 9 were identified in consensus mode (i.e., present in 100% of

452   population in one or more passages) while the remaining 833 SNPs were identified in population mode,

453   (i.e., occurring in a fraction of the community) when sequenced (Figure 8). Further, only 69 unique SNPs

454   were identified to occur in all (i.e., passage 3-61), and only 43 SNPs were maintained in later passages

455   (Figure 8b). More importantly, only one consensus mutation occurred in a gene that was downregulated

456   in transcriptomic analysis. This gene GTP pyrophosphokinase SpoT (B7L74_01590) had a one amino

457   acid (aa) substitution (T to A) at position 262, which propagates to 100% population in the last 9 passages

458   analyzed and is also noticeably early down in gene expression. SpoT is a signal transduction component

459   and transcriptional regulator with a role in helping *Coxiella* cope with the low-nutrient and high stress

460   condition (Minnick and Raghavan 2012).

461   For the 118 unique DIPs, only 3 DIPs were identified in consensus mode and 115 in population

462   mode (Figure 8). Lengths of insertions and deletions were always very minor with 93% of DIPs

463   representing an insertions or deletions of a single bp (Table S3). The multi base pair deletions included

464   deletions of 2, 3, 7 and 12 bp that occurred in coding region and the longest 32 base pair deletion

465   occurring in an intergenic region. However, none of these genes appeared significantly affected

466   transcriptionally by the deletion as there were no significant transcriptomic changes. Of the genes that

467   were downregulated, 7 had DIPs mutations but all of them being in only a fraction of Cb populations

468   within a passage (mostly 5-10% populations).

20

469   An interesting observation was the numerous mutations (SNPs and DIPs) over several passages in

470   two genes i.e., *lapA* and *lapB*. In LapA, a 97 aa long protein has a non-sense mutation at the 85th position

471   in passage 13. LapB, 389 aa long protein, on the other hand has two missense mutations in a large

472   proportion of cells, with one mutation propagating to 100% of the population at passage 67. It also has

473   numerous insertions in the coding region but the noticeable one is a 3 bp deletion in the coding region that

474   propagates to later passages (Table S3). Although these genes didn't show any change in gene

475   transcriptional expression modulated by mutations, it is possible that these genes are en route to

476   simplifying the LPS and O-antigen layer in accordance with the absence of a environment , as seen in

477   some bacteria (Maldonado, Sa-Correia and Valvano 2016).

478   Collectively, the low levels of DNA mutations within the passage populations and possibly the

479   lack of effects, suggest a very stable and minor level in genomic mutations in modulating transcriptional

480   levels.

481

482
483

484

<div style="text-align:center">Discussion</div>

485

486   Here, we attempted to identify genes and proteins crucial to Cb intracellular growth lifestyle using a

487   "reverse evolution" approach paired with RNAseq and DNAseq comparative transcriptomics and

488   genomics, respectively. We transitioned Cb Nine Mile phase II from cell cultures into the axenic defined

489   media ACCM-D and subcultured it in a long-term successive passage model. Temporal changes in gene

490   expression patterns, and DNA mutations associated with adaptation to an axenic extracellular lifestyle

491   were identified. In general, we observe a significant number of differential expression (464 up, 371 down,

492   38% of overall Cb genes) through 67 passages. It is interesting to note that the majority (288 upregulated

493   and 249 downregulated) of differentially expressed genes expressed an "early up" or "early down"

494   expression pattern (Figure 3, Table S1), suggestive of a relatively rapid adaptation (within 31 passages

495   out of 67 total passages) into this new axenic environment.

496   Differentially expressed genes identified in this study could be grouped into multiple structural

497   and functional categories (secretory apparatus, effector proteins, other pathogenicity determinants,

498   hypothetical proteins, and central metabolic pathways). In general, a broad (19/24 genes) encoding

499   T4BSS components showed significant expression change, with 18 genes showing a decrease in the

500   expression whereas only one gene that was upregulated. T4BSS is the most crucial conduit for

501   pathogenicity and effector proteins in Cb (Carey, Newton, Luhrmann, *et al.* 2011; van Schaik, Chen,

502   Mertens, *et al.* 2013; Voth and Heinzen 2009). Components of the T4BSS span both membranes and the

503   periplasm and are bridged by the core transport complex comprising proteins DotC, DotD, DotF, DotG

504   and DotH, which are predicted to provide a channel for export of effector substrates (Figure 5a) (Vincent,

505   Friedman, Jeong, *et al.* 2006). The coupling protein complex provides a link between substrates and

506   transport complex and includes DotL, DotM, DotN, IcmS and IcmW (Vincent, Friedman, Jeong, *et al.*

507   2012). DotB is an essential cytoplasmic protein with an ATPase activity and unknown function, but its

508   mutation has been linked to failure in secreting effector proteins during the infection of host cells (Beare,

509   Larson, Gilk, *et al.* 2012). DotA and IcmX has been shown to be released from the bacteria (Luedtke,

510   Mahapatra, Lutter, *et al.* 2017). Besides these, other components of the T4BSS includes DotO localized in

<div style="text-align:center">22</div>

511   the cytoplasm, IcmX in periplasmic space, DotK in outer membrane whereas IcmF, IcmH, DotI, DotJ,

512   DotA, DotE, DotP, IcmV and IcmT in inner membrane. (Figure 5a). The genes involved in the T4BSS in

513   Cb are clustered in a single locus made up of two regions, with the exceptionof *icmF* and *dotU,* which are

514   part of a separate operon (Figure 5b). This is similar to the gene rearrangement shown in the original

515   *Coxiella burnetii* sequence (Seshadri, Paulsen, Eisen*, et al.* 2003). Gene *icmF*, which has been shown to

516   be involved in intra-macrophage replication and inhibition of phagosome-lysosome fusion in *L.*

517   *pneumophila* (VanRheenen, Duménil and Isberg 2004; Zusman, Feldman, Halperin*, et al.* 2004) and

518   stabilization of the secretion complex (Sexton, Miller, Yoneda*, et al.* 2004) was the only T4BSS

519   component that showed transcriptional upregulation.

520   The observed downregulation of this experimentally verified central intracellular pathogenic

521   determinant makes biological sense and provides a general overall credence that gene downregulation

522   under the experimental setting employed in this study could be regarded as a reasonable proxy for

523   requirement for intracellular survival in cell-cultures. In addition to T4BSS, other secretory pathway such

524   as the general secretory (sec) pathway and a component of type I secretary pathway (i.e., TolC) also

525   exhibited a general trend of overall downregulation (Figure 5A, Figure S1a, Figure S1b, Table S1). In

526   general, we interpret such overall lower expression of structural secretory apparatuses as a reflection of

527   less need for these systems during interaction between Cb and the environment in a relatively rich axenic

528   setting when compared to the organisms environmentally "normal" intracellular setting.

529   Interestingly, while genes encoding the production of secretary pathways were downregulated,

530   expression patterns of predicted T4BSS effector proteins were mixed, with 33 upregulated and 14

531   downregulated. Of the 14 downregulated genes (all of which were early downregulated), nine have been

532   experimentally verified based on experimental evidence of their translocation by the Dot/Icm system

533   (Carey, Newton, Luhrmann*, et al.* 2011; Chen, Banga, Mertens*, et al.* 2010; Lifshitz, Burstein, Peeri*, et*

534   *al.* 2013; Maturana, Graham, Sharma*, et al.* 2013; Voth, Beare, Howe*, et al.* 2011; Voth, Howe, Beare*, et*

535   *al.* 2009; Weber, Chen, Rowin*, et al.* 2013), three genes containing ankyrin repeat domains (*ankG*, *ankD*

536   and *ankK*) were considered effectors based on the presence of eukaryotic like domains and subsequently

537   shown to be translocated by the Dot/Icm system(Voth, Howe, Beare*, et al.* 2009), gene *cirC (Coxiella*

538   effector for intracellular replication) was verified as effector by transposon insertion mutation studies

539   where its mutation was associated with a defect in Coxiella containing vacuole (CCV) biogenesis (Weber,

540   Chen, Rowin*, et al.* 2013), and lastly a hypothetical protein B7L74_09020 was verified to be an effector

541   based on loss-of-function mutation where its mutation was related to a smaller CCV phenotype (Crabill,

542   Schofield, Newton*, et al.* 2018)(Table S2).

543   AnkG, AnkD and AnkK are ankyrin repeat-containing effector proteins in Cb (Seshadri, Paulsen,

544   Eisen*, et al.* 2003). The eukaryotic type Ank domain in this protein family might have a role in host-cell

545   attachment and allows the interaction of bacteria with a spectrum of host cell proteins and thus are

546   particularly important in the pathogenic process (Batrukova, Betin, Rubtsov*, et al.* 2000; Cordsmeier,

547   Rinkel, Jeninga*, et al.* 2022; Pechstein, Schulze-Luehrmann, Bisle*, et al.* 2020; Voth, Howe, Beare*, et al.*

548   2009). AnkD has both eukaryotic like domain and F-box domain, but the function is not yet clear (Voth,

549   Howe, Beare*, et al.* 2009). AnkG has been shown to localize at the host microtubules and interferes with

550   the host apoptosis pathway by interacting with the host protein gC1qR (p32) (Luhrmann, Nogueira,

551   Carey*, et al.* 2010; Voth, Howe, Beare*, et al.* 2009). AnkK has been shown to have an important role for

552   bacterial growth inside of macrophages (Habyarimana, Al-Khodor, Kalia*, et al.* 2008), although it is not

553   delivered to the host cell via T4BSS (Voth, Howe, Beare*, et al.* 2009) (Table S2). *Coxiella* plasmid

554   effector proteins (CpeF and CpeH) are in the plasmid T4BSS effector family of proteins, and important

555   for disrupting host cell mechanisms. CpeF specifically localized in host cell during infection, and has

556   shown to cause a growth defect when mutated (Martinez, Cantet, Fava*, et al.* 2014; Voth, Beare, Howe*, et*

557   *al.* 2011) whereas CpeH localizes to the host cell's cytoplasm (Maturana, Graham, Sharma*, et al.* 2013).

558   Cig57 mutation has been linked to an intracellular replication defect for Cb, whereas a Cig2 mutation

559   causes both a growth defect and a CCV fusion defect (Newton, Kohler, McDonough*, et al.* 2014). These

560   two proteins are early downregulated in 9 and 7 passages, respectively. CirC has been shown to be

561   important for CCV biogenesis (Weber, Chen, Rowin*, et al.* 2013) and it is early downregulated in 4

562   passages. Of the remaining 6 downregulated effectors, 4 are hypothetical proteins with unknown

24

563   functions, Cbu1752 has been shown to be important for vacuole biogenesis, and Cbu0635 appears

564   important for host cell secretion (Table S2).

565   Of the 33 upregulated effector proteins, the majority fall into COG categories of unclassified

566   (n=11), signal transduction mechanisms (n=7), transportation and metabolism of coenzyme and inorganic

567   ions (n=6) (Figure 6, Table S1). These upregulated, and no expression change effector proteins (n=71),

568   could also be involved in mediating general survival functions or other cellular functions besides their

569   involvement in directly association with the intracellular pathogenesis process. This could be one of the

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

571   Such a pattern, where genes encoding the formation of the structural conduits (i.e., secretary

572   pathways) are downregulated, but numerous genes encoding proteins thought to be secreted through these

573   conduits (i.e., effector proteins) are upregulated is puzzling. We put forth the possibility that the

574   expression of these effector proteins is controlled by Cb intracellular conditions, where high

575   concentrations of intracellular metabolites (amino acids, inorganic salts, ATP/ADP ratio) regulate their

576   expression. Under this scenario, high level of intracellular precursors in LCV Cb is associated with

577   growth inside the cell, and possibly in the "rich" axenic media state of LCV growth. It remains to be seen

578   whether translation of these effector transcripts to protein products and subsequent secretion occurs in Cb

579   grown in axenic media. Limited reports suggest that Cb T4BSS effectors have not been observed during

580   axenic media growth (Stead, Omsland, Beare, *et al*. 2013; Shaw, Unpublished data)

581   Multiple additional pathogenic determinants were also downregulated in axenic media.

582   Specifically, chaperons, LPS, and peptidoglycan synthesis. Amongst the 16 genes annotated as chaperons

583   in the Cb genome, 10 were transcriptionally downregulated starting at early passages (Figure 7a) (Table

584   S1). The downregulation could be explained by the fact that chaperons play important roles for

585   withstanding stress associated with intracellular survival (e.g., CCV detoxification) (Beare, Unsworth,

586   Andoh*, et al.* 2009), survival inside the macrophage (Takaya, Tomoyasu, Matsui*, et al.* 2004) and low pH

587   within the CCV (Macellaro, Tujulin, Hjalmarsson*, et al.* 1998). . Downregulations in the genes involved

588   in synthesis of Lipid A and O-antigen biosynthesis as well as peptide cross linking during peptidoglycan

25

589   layer suggests the possible further reduction in the virulence or need for these functions of *Coxiella*

590   *burnetii* in axenic media, even within the CB NMII strain used in this study. These changes in the

591   bacterial cell coverings could be tied to role/importance of these genes in intracellular host cell

592   manipulation and bacterial survival and growth inside the host cell. These assumptions will certainly need

593   to be verified with more experimental evidence. Here, we need to keep in mind that the strain we are

594   using is the *Coxiella burnetii* avirulent strain Nine Mile phase II (NMII). These variants have a truncated

595   LPS, due to a genomic deletion of about 25 Kbp of sequences that encodes all the three sugars i.e.,

596   virenose, dihydrohydroxystreptose, and galactosaminuronyl-α(1,6)-glucosamine that comprises the LPS

597   O-antigen biosynthesis (Amano, Williams, Missler*, et al.* 1987; Denison, Massung and Thompson 2007).

598   Thus, further downregulations in some of the remaining genes involved in LPS and O-antigen synthesis

599   pathways could be due to their already altered functionality in the cell.

600   Analysis of central metabolic pathways showed a clear trend of downregulation of multiple

601   catabolic (e.g., glycolysis and electron transport chain), amphibolic (e.g., citric acid cycle) and anabolic

602   (e.g., FA synthesis) pathways, with a parallel upregulation of genes encoding transporters (Figure 7C).

603   Such pattern could readily be explained by the nutrient rich growth environment (ACCM-D media) where

604   Cb is grown in our continuous passage model. This media is a defined axenic medium for Cb growth and

605   contains all 20 amino acids, salts (sodium phosphate and sodium bicarbonate), vitamins, minerals, and

606   trace elements (Sandoz, Beare, Cockrell*, et al.* 2016). As such, the need for expression of genes encoding

607   enzymes that are components of these biosynthetic pathways decreases and subsequently, the overall need

608   for ATP generation for biosynthetic purposes (hence a decrease in respiratory activity). Likewise, the

609   upregulation in several structural genes encoding transporters could be explained by the bacteria

610   increasing production of more channels/transporters to accommodate the increased presence of

611   metabolites/substrates from the media as opposed to needing to produce them via metabolic pathways in

612   the bacterial cytoplasm. Finally, 30 hypothetical proteins were downregulated, and their predicted

613   localization, predicted role in pathogenesis and their general analysis provide some possible explanations

614   for such pattern (see results section). Regardless, we suggest that these could be important, hitherto

26

615 untested possible pathogenicity determinants. Future biochemical and genetic efforts to test such

616 assumptions represent a ready avenue for future research directions.

617 We also hypothesized that continuous passaging could also lead to the propagation of mutations,

618 DNA fragment loses, and rearrangements in genes involved in intracellular survival, pathogenesis, and

619 host cell manipulations could occur. While this is a well-known process, it doesn't appear that within the

620 timeframe of the experiments here (67 passages) that this is the case. The careful study of each of these

621 mutations, their corresponding gene expression changes, number of passages affected by the mutations

622 and their transiency, and overall fraction of population affected by these mutations showed that these

623 mutations do not appear to be significant at the gene or transcript expression level. Out of 960 unique

624 mutations (SNPs and DIPs collectively) observed at different spots within the genomes of various

625 passages, only one mutation (i.e., a SNP) in GTP pyrophosphokinase SpoT, a signal transduction

626 component and transcriptional regular with a role in helping *Coxiella* cope with the low-nutrient and high

627 stress conditions in the CCV, seems to have effect on gene transcript expression (Minnick and Raghavan

628 2012). Early downregulation observed in this gene could be attributed to the fact that these cells are less

629 stressed due to the growth condition in rich ACCM-D media, and this mutation could have caused the

630 decrease in transcript expression of this gene. Overall, these findings suggest that Cb gene expression

631 changes significantly following acclimation to axenic media, although extensive genomic rearrangement

632 does not occur. Genomics reveals a relatively stable Cb NMII genome over the 67 passages that were

633 analyzed here.

634 In conclusion, we present a detailed temporal analysis on how Cb transition from intracellular

635 growth, (where a wide range of cellular processes is thought to be required to maintain survival and

636 growth), to a defined, rich axenic media (where many of such processes are theoretically, no longer

637 needed). As any genome-wide transcriptomics survey, the approach is useful for uncovering patterns,

638 confirming prior observations, and generating new insights and hypothesis. We stress that while

639 transcriptional downregulation in axenic media compared to cell culture could broadly be associated with

640 a genes importance for survival in cell cultures, the precise nature of such correlation is yet unclear, and

641   that differential expression patterns could further be modified on the translational and post translational

642   levels. Experimental assessments and validation of many of the observed patterns may well open new

643   avenues of Cb research. Nevertheless, our analysis is beneficial in providinge information on how specific

644   genes and pathways in Cb may be important for this unusual organisms intracellular survival, as well as

645   to identify putatively novel pathogenicity determinants in this naturally intracellular pathogen.

646

647

648   Figure Legends

649   **Figure 1.** Flowchart representing the overall comparative transcriptomics and genomics strategy

650   employed in this study.

651   **Figure 2.** Intracellular vs Axenic Growth Following Serial Passage of *C. burnetii* NMII**.** Intracellular and

652   axenic growth from 3 biological replicates of passaged ACCM-D growth **A)** fluorescence forming units

653   (FFU) counts of infections of HeLa cells normalized by Cb genomes from passages 1, 3, 5, and 10.

654   Significance between different passages are indicated by lines and *p<0.001. **B)** CFU enumeration of

655   passages 1, 3, 5, and 10 Cb spread on ACCM-D plates normalized to genomes. No statistically significant

656   difference was observed between groups.

657   **Figure 3.** Overview for Differential Expression patterns in axenically-grown Cb. **A)** Summary for the

658   number of total and differentially expressed genes, T4BSS and effector proteins in this experiment. **B)**

659   Bar graph showing the number of upregulated (p-value<0.05 and L2fc > 0) and downregulated (p-

660   value<0.05 and L2fc < 0) genes in each passage. **C)** Classification of genes based on their gene

661   expression verdicts. **D**) Classification of genes based on the number of passages showing significant

662   expression change.

663   **Figure 4.** Volcano plots for 845 significant DEGs in different passages when compared to passage 01.

664   Black dots represent non-significant DEGs, green dots represent non-DEGs with L2fc >2, blue dots

665   represent significant DEGs with L2fc <2 whereas the red diamonds represent significant DEGs with L2fc

666   >2. The 10 genes with the lowest p-value in every passage comparison are labeled in boxes.

667   **Figure 5.** Cb T4BSS machinery and Sec expression changes during axenic passaging. **A)** Membrane

668   complex model with gene transcript expression profiles for components of the Cb T4BSS (right side) and

669   Sec protein export pathway (left side). DEG patterns are denoted by arrows (downregulated in maroon, no

670   significant expression changes in yellow, and upregulated in green). * Represents genes that were highly

671   downregulated (L2fc ≤ -3). **B)** Gene locus map for all the components in Cb NMII strain [CP020616.1].

672   The arrows are colored filled according to DEGs patterns as (downregulated in maroon, no significant

29

673   expression change in yellow, and upregulated in green). The arrows not filled are genes not related to the

674   T4BSS pathway.

675   **Figure 6.** Transcript expression changes for T4BSS effector proteins during continuous axenic passaging.

676   **A)** COG classification graph for 118 identified effector proteins. The blue bars represent total effector

677   proteins in each COG functional group (x-axis) and orange bars represents the effectors that showed

678   differential gene expression. **B**) Pie chart showing COG classification for 14 downregulated effector

679   proteins. **C**) Pie chart showing COG classification for 33 upregulated effector proteins.

680   **Figure 7.** Gene expression changes in central metabolic pathways. **A**) Bar graph showing numbers of

681   DEGs in each central metabolic pathway. The numbers on top of the bars represent the total number of

682   genes classified into that metabolic category by KEGG. **B**) Broad classification of differentially expressed

683   transporters according to their substrates. The number in parenthesis for each category represents the

684   proportion of upregulated genes **C**) Graphical representation of all metabolic pathways discussed in the

685   manuscript text.

686   **Figure 8.** DEseq2 Analysis of Genomes from Passages. **A**) Flowchart for classification of different types

687   of Single nucleotide polymorphisms (SNPs) and Deletion/Insertion polymorphisms (DIPs) found in this

688   experiment. Numbers in the orange blocks at the tip of some boxes represents the number of those

689   mutations that occurred in 100 % of the passages. **Bi**) Classification of SNPs according to its occurrence

690   in number of passages. **Bii**) Classification of DIPs according to the occurrence in number of passages.

691

30

692   **References**

693   Allen B, Drake M, Harris N et al. Using KBase to Assemble and Annotate Prokaryotic Genomes. *Curr*

694   *Protoc Microbiol* 2017;**46**: 1E 13 1-1E 13 18.

695   Amano K, Williams JC, McCaul TF et al. Biochemical and immunological properties of Coxiella burnetii

696   cell wall and peptidoglycan-protein complex fractions. *J bacteriol* 1984;**160**: 982-88.

697   Amano K, Williams JC, Missler SR et al. Structure and biological relationships of Coxiella burnetii

698   lipopolysaccharides. *J Biol Chem* 1987;**262**: 4740-47.

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

700   R106.

701   Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

702   Arnold DL, Jackson RW, Waterfield NR et al. Evolution of microbial virulence: the benefits of stress.

703   *Trends Genet* 2007;**23**: 293-300.

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

705   327-49.

706   Batrukova MA, Betin VL, Rubtsov AM et al. Ankyrin: structure, properties, and functions. *Biochemistry*

707   *(Mosc)* 2000;**65**: 395-408.

708   Beare PA, Chen C, Bouman T et al. Candidate antigens for Q fever serodiagnosis revealed by

709   immunoscreening of a Coxiella burnetii protein microarray. *Clin Vaccine Immunol* 2008;**15**:

710   1771-79.

711   Beare PA, Gilk SD, Larson CL et al. Dot/Icm Type IVB Secretion System Requirements for Coxiella

712   burnetii Growth in Human Macrophages. *mBio* 2011;**2**: e00175-11.

713   Beare PA, Heinzen RA. Gene inactivation in *Coxiella burnetii* host-bacteria interactions: Springer, 2014,

714   329-45.

715   Beare PA, Jeffrey BM, Long CM et al. Genetic mechanisms of *Coxiella burnetii* lipopolysaccharide

716   phase variation. *PLOS Pathog* 2018;**14**: e1006922.

717   Beare PA, Larson CL, Gilk SD et al. Two systems for targeted gene deletion in *Coxiella burnetii. Appl*

718   *Environ Microbiol* 2012;**78**: 4580-89.

719   Beare PA, Sandoz KM, Larson CL et al. Essential role for the response regulator PmrA in Coxiella

720   burnetii type 4B secretion and colonization of mammalian host cells. *J Bacteriol* 2014;**196**: 1925-

721   40.

722   Beare PA, Unsworth N, Andoh M et al. Comparative genomics reveal extensive transposon-mediated

723   genomic plasticity and diversity among potential effector proteins within the genus *Coxiella*.

724   *Infect Immun* 2009;**77**: 642-56.

725   Bewley KR. The identification of immune-reactive proteins recognised in response to *Coxiella burnetii*

726   infection Ph. D thesis.: University of Portsmouth, 2015, 230.

727   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.

728   *Bioinformatics* 2014;**30**: 2114-20.

729   Brennan RE, Russell K, Zhang G et al. Both inducible nitric oxide synthase and NADPH oxidase

730   contribute to the control of virulent phase I Coxiella burnetii infections. *Infect Immun* 2004;**72**:

731   6666-75.

732   Brennan RE, Samuel JE. Evaluation of *Coxiella burnetii* Antibiotic Susceptibilities by Real-Time PCR

733   Assay. *J Clin Microbiol* 2003;**41**: 1869-74.

734   Carey KL, Newton HJ, Luhrmann A et al. The Coxiella burnetii Dot/Icm system delivers a unique

735   repertoire of type IV effectors into host cells and is required for intracellular replication. *PLoS*

736   *Pathog* 2011;**7**: e1002056.

737   Chen C, Banga S, Mertens K et al. Large-scale identification and translocation of type IV secretion

738   substrates by Coxiella burnetii. *Proc Natl Acad Sci USA* 2010;**107**: 21755-60.

739   Coleman SA, Fischer ER, Howe D et al. Temporal analysis of Coxiella burnetii morphological

740   differentiation. *J Bacteriol* 2004;**186**: 7344-52.

741   Cordsmeier A, Rinkel S, Jeninga M et al. The Coxiella burnetii T4SS effector protein AnkG hijacks the

742   7SK small nuclear ribonucleoprotein complex for reprogramming host cell transcription. *PLOS*

743   *Pathog* 2022;**18**: e1010266.

744   Cotter PA, Chepuri V, Gennis RB et al. Cytochrome o (cyoABCDE) and d (cydAB) oxidase gene

745   expression in Escherichia coli is regulated by oxygen, pH, and the fnr gene product. *J Bacteriol*

746   1990;**172**: 6333-8.

747   Crabill E, Schofield WB, Newton HJ et al. Dot/Icm-Translocated Proteins Important for Biogenesis of the

748   Coxiella burnetii-Containing Vacuole Identified by Screening of an Effector Mutant Sublibrary.

749   *Infect Immun* 2018;**86**.

750   Davis JJ, Wattam AR, Aziz RK et al. The PATRIC Bioinformatics Resource Center: expanding data and

751   analysis capabilities. *Nucleic Acids Res* 2020;**48**: D606-D12.

752   Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-

753   generation sequencing data using breseq. *Methods Mol Biol* 2014;**1151**: 165-88.

754   Denison AM, Massung RF, Thompson HA. Analysis of the O-antigen biosynthesis regions of phase II

755   Isolates of Coxiella burnetii. *FEMS Microbiol Lett* 2007;**267**: 102-07.

756   Gajdosova E, Kovacova E, Toman R et al. Immunogenicity of Coxiella burnetii whole cells and their

757   outer membrane components. *Acta Virologica* 1994;**38**: 339-44.

758   Galperin MY, Wolf YI, Makarova KS et al. COG database update: focus on microbial diversity, model

759   organisms, and widespread pathogens. *Nucleic Acids Res* 2021;**49**: D274-d81.

760   Genevaux P, Georgopoulos C, Kelley WL. The Hsp70 chaperone machines of *Escherichia coli*: a

761   paradigm for the repartition of chaperone functions. *Mol Microbiol* 2007;**66**: 840-57.

762   Green ER, Mecsas J. Bacterial Secretion Systems: An Overview. *Microbiol Spectr* 2016;**4**.

763   Gurevich A, Saveliev V, Vyahhi N et al. QUAST: quality assessment tool for genome assemblies.

764   *Bioinformatics* 2013;**29**: 1072-75.

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

766   *pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ Microbiol*

767   2008;**10**: 1460-74.

768   Hackstadt T, Williams JC. Biochemical stratagem for obligate parasitism of eukaryotic cells by *Coxiella*

769   *burnetii. Proc Natl Acad Sci U S A* 1981;**78**: 3240-4.

770   Heinzen RA, Hackstadt T, Samuel JE. Developmental biology of *Coxiella burnettii. Trends Microbiol*

771   1999;**7**: 149-54.

772   Howe D, Melnicakova J, Barak I et al. Fusogenicity of the Coxiella burnetii parasitophorous vacuole.

773   *Annals of the New York Academy of Sciences* 2003;**990**: 556-62.

774   Hussein A, Kovacova E, Toman R. Isolation and evaluation of *Coxiella burnetii* O-polysaccharide

775   antigen as an immunodiagnostic reagent. *Acta virologica* 2001;**45**: 173-80.

776   Ito M, Morino M, Krulwich TA. Mrp antiporters have important roles in diverse Bacteria and Archaea.

777   *Front Microbiol* 2017;**8**: 2325.

778   Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci*

779   2020;**29**: 28-35.

780   Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional

781   characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**: 726-31.

782   Kim D, Paggi JM, Park C et al. Graph-based genome alignment and genotyping with HISAT2 and

783   HISAT-genotype. *Nat Biotechnolo* 2019;**37**: 907-15.

784   Kovaka S, Zimin AV, Pertea GM et al. Transcriptome assembly from long-read RNA-seq alignments

785   with StringTie2. *Genome Biology* 2019;**20**: 278.

786   Kuley R, Bossers-deVries R, Smith HE et al. Major differential gene regulation in Coxiella burnetii

787   between in vivo and in vitro cultivation models. *BMC Genomics* 2015;**16**: 953.

788   Larson CL, Beare PA, Voth DE et al. *Coxiella burnetii* effector proteins that localize to the

789   parasitophorous vacuole membrane promote intracellular replication. *Infect Immun* 2015;**83**: 661-

790   70.

791   Larson CL, Heinzen RA. high-content imaging reveals expansion of the endosomal compartment during

792   *Coxiella burnetii* parasitophorous vacuole maturation. *Front Cell Infect Microbiol* 2017;**7**: 48.

793   Larson CL, Martinez E, Beare PA et al. Right on Q: genetics begin to unravel *Coxiella burnetii* host cell

794   interactions. *Future Microbiol* 2016;**11**: 919-39.

795   Lifshitz Z, Burstein D, Peeri M et al. computational modeling and experimental validation of the

796   *Legionella* and *Coxiella* virulence-related type-IVB secretion signal. *Proc Nat Acad Sci USA*

797   2013;**110**: E707-E15.

798   Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data

799   with DESeq2. *Genome Biology* 2014;**15**: 550.

800   Luedtke BE, Mahapatra S, Lutter EI et al. The Coxiella Burnetii type IVB secretion system (T4BSS)

801   component DotA is released/secreted during infection of host cells and during in vitro growth in a

802   T4BSS-dependent manner. *Pathogens and Disease* 2017;**75**.

803   Luhrmann A, Nogueira CV, Carey KL et al. Inhibition of pathogen-induced apoptosis by a *Coxiella*

804   *burnetii* type IV effector protein. *Proc Natl Acad Sci U S A* 2010;**107**: 18997-9001.

805   Macellaro A, Tujulin E, Hjalmarsson K et al. Identification of a 71-kilodalton surface-associated Hsp70

806   homologue in *Coxiella burnetii. Infection and immunity* 1998;**66**: 5882-88.

807   Maldonado RF, Sa-Correia I, Valvano MA. Lipopolysaccharide modification in Gram-negative bacteria

808   during chronic infection. *FEMS Microbiol Rev* 2016;**40**: 480-93.

809   Martinez E, Cantet F, Bonazzi M. Generation and multi-phenotypic high-content screening of *Coxiella*

810   *burnetii* transposon mutants. *J Vis Exp* 2015, DOI 10.3791/52851: e52851.

811   Martinez E, Cantet F, Fava L et al. Identification of OmpA, a *Coxiella burnetii* protein involved in host

812   cell invasion, by multi-phenotypic high-content screening. *PLoS Pathog* 2014;**10**: e1004013.

813   Maturana P, Graham JG, Sharma UM et al. Refining the plasmid-encoded type IV secretion system

814   substrate repertoire of *Coxiella burnetii. J Bacteriol* 2013;**195**: 3269-76.

815   Maurin M, Raoult D. Q fever. *Clin Microbiol Rev* 1999;**12**: 518-53.

816   McQuiston JH, Childs JE, Thompson HA. Q fever. *J of the American Veterinary Medical Association*

817   2002;**221**: 796-99.

818   Miller JD, Shaw EI, Thompson HA. *Coxiella burnetii*, Q Fever, and Bioterrorism. In: Anderson B,

819   Friedman H, Bendinelli M (eds.) *Microorganisms and Bioterrorism*, DOI 10.1007/0-387-28159-

820   2_10. Boston, MA: Springer US, 2006, 181-208.

821   Minnick MF, Raghavan R. Developmental biology of *Coxiella burnetii*. *Adv Exp Med Biol* 2012;**984**:

822   231-48.

823   Mistry J, Chuguransky S, Williams L et al. Pfam: The protein families database in 2021. *Nucleic Acids*

824   *Research* 2020;**49**: D412-D19.

825   Moormeier DE, Sandoz KM, Beare PA et al. *Coxiella burnetii* RpoS regulates genes involved in

826   morphological differentiation and intracellular growth. *J Bacteriol* 2019;**201**.

827   Morgan JK, Luedtke BE, Shaw EI. Polar localization of the *Coxiella burnetii* type IVB secretion system.

828   *FEMS Microbiol Lett* 2010;**305**: 177-83.

829   Mori H, Ito K. The Sec protein-translocation pathway. *Trends Microbiol* 2001;**9**: 494-500.

830   Nagai H, Kubori T. Type IVB Secretion Systems of *Legionella* and other Gram-Negative Bacteria. *Front*

831   *Microbiol* 2011;**2**.

832   Neckers L, Tatu U. Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell*

833   *Host Microbe* 2008;**4**: 519-27.

834   Newton HJ, Kohler LJ, McDonough JA et al. A screen of *Coxiella burnetii* mutants reveals important

835   roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. *PLoS Pathog* 2014;**10**:

836   e1004286.

837   Newton HJ, McDonough JA, Roy CR. Effector Protein Translocation by the *Coxiella burnetii* Dot/Icm

838   Type IV Secretion System Requires Endocytic Maturation of the Pathogen-Occupied Vacuole.

839   *PLOS ONE* 2013;**8**: e54566.

840   Omsland A, Beare PA, Hill J et al. Isolation from animal tissue and genetic transformation of *Coxiella*

841   *burnetii* are facilitated by an improved axenic growth medium. *Appl Environ Microbiol* 2011;**77**:

842   3720-5.

843   Omsland A, Cockrell DC, Howe D et al. Host cell-free growth of the Q fever bacterium *Coxiella burnetii.*

844   *Proc Natl Acad Sci USA* 2009;**106**: 4430-4.

845   Pan X, Lührmann A, Satoh A et al. Ankyrin repeat proteins comprise a diverse family of bacterial type IV

846   effectors. *Science* 2008;**320**: 1651-4.

847   Parks DH, Imelfort M, Skennerton CT et al. CheckM: assessing the quality of microbial genomes

848   recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**: 1043-55.

849   Pechstein J, Schulze-Luehrmann J, Bisle S et al. The *Coxiella burnetii* T4SS Effector AnkF Is Important

850   for Intracellular Replication. *Front Cellular and Infection Microbiology* 2020;**10**.

851   Raoult D, Marrie T, Mege J. Natural history and pathophysiology of Q fever. *Lancet Infect Dis* 2005;**5**:

852   219-26.

853   Rudolf Toman RAH, James E. Samuel, Jean-Louis Mege. *Coxiella burnetii:* Recent advances and new

854   perspectives in research of the Q Fever bacterium. Advances in Experimental Medicine and

855   Biology, DOI 10.1007/978-94-007-4315-1: Springer Netherlands, 2012.

856   Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar

857   precursor synthesis and O-antigen assembly. *Carbohydrate Res* 2003;**338**: 2503-19.

858   Sanchez SE, Vallejo-Esquerra E, Omsland A. Use of Axenic Culture Tools to Study *Coxiella burnetii.*

859   *Curr Protoc Microbiol* 2018;**50**: e52.

860   Sandoz KM, Beare PA, Cockrell DC et al. Complementation of Arginine Auxotrophy for Genetic

861   Transformation of Coxiella burnetii by Use of a Defined Axenic Medium. *Appl Environ*

862   *Microbiol* 2016;**82**: 3042-51.

863   Sandoz KM, Popham DL, Beare PA et al. Transcriptional profiling of *Coxiella burnetii* reveals extensive

864   cell wall remodeling in the small cell variant developmental form. *PloS one* 2016;**11**: e0149957-

865   e57.

866   Segal G, Feldman M, Zusman T. The Icm/Dot type-IV secretion systems of *Legionella pneumophila* and

867   *Coxiella burnetii. FEMS Microbiol Rev* 2005;**29**: 65-81.

868   Segal G, Shuman HA. Possible origin of the *Legionella pneumophila* virulence genes and their relation to

869   Coxiella burnetii. *Mol Microbiol* 1999;**33**: 669-70.

870   Seshadri R, Paulsen IT, Eisen JA et al. Complete genome sequence of the Q-fever pathogen *Coxiella*

871   *burnetii. Proc Natl Acad Sci U S A* 2003;**100**: 5455-60.

872   Sexton JA, Miller JL, Yoneda A et al. *Legionella pneumophila* DotU and IcmF are required for stability

873   of the Dot/Icm complex. *Infect Immun* 2004;**72**: 5983-92.

874   Sexton JA, Vogel JP. Type IVB Secretion by intracellular pathogens. *Traffic* 2002;**3**: 178-85.

875   Stead CM, Omsland A, Beare PA et al. Sec-mediated secretion by *Coxiella burnetii*. BMC Microbiol.

876   2013; **13**:222.

877   Takaya A, Tomoyasu T, Matsui H et al. The DnaK/DnaJ chaperone machinery of *Salmonella enterica*

878   serovar Typhimurium is essential for invasion of epithelial cells and survival within

879   macrophages, leading to systemic infection. *Infect Immun* 2004;**72**: 1364-73.

880   Tsirigotaki A, De Geyter J, Šoštaric´ N et al. Protein export through the bacterial Sec pathway. *Nat Rev*

881   *Microbiol* 2017;**15**: 21-36.

882   van Schaik EJ, Chen C, Mertens K et al. Molecular pathogenesis of the obligate intracellular bacterium

883   *Coxiella burnetii. Nat Rev Microbiol* 2013;**11**: 561-73.

884   VanRheenen SM, Duménil G, Isberg RR. IcmF and DotU are required for optimal effector translocation

885   and trafficking of the Legionella pneumophila vacuole. *Infect Immunity* 2004;**72**: 5972-82.

886   Vincent CD, Friedman JR, Jeong KC et al. Identification of the core transmembrane complex of the

887   *Legionella* Dot/Icm type IV secretion system. *Mol Microbiol* 2006;**62**: 1278-91.

888   Vincent CD, Friedman JR, Jeong KC et al. Identification of the DotL coupling protein subcomplex of the

889   *Legionella* Dot/Icm type IV secretion system. *Mol Microbiol* 2012;**85**: 378-91.

890   Vogel JP. Turning a tiger into a house cat: using *Legionella pneumophila* to study *Coxiella burnetii.*

891   *Trends in Microbiol* 2004;**12**: 103-05.

892   Voth DE, Beare PA, Howe D et al. The Coxiella burnetii cryptic plasmid is enriched in genes encoding

893   type IV secretion system substrates. *J Bacteriol* 2011;**193**: 1493-503.

894   Voth DE, Heinzen RA. Lounging in a lysosome: the intracellular lifestyle of *Coxiella burnetii. Cell*

895   *Microbiol* 2007;**9**: 829-40.

896   Voth DE, Heinzen RA. Coxiella type IV secretion and cellular microbiology. *Curr Opin Microbiol*

897   2009;**12**: 74-80.

898   Voth DE, Howe D, Beare PA et al. The *Coxiella burnetii* ankyrin repeat domain-containing protein

899   family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J*

900   *Bacteriol* 2009;**191**: 4232-42.

901   Weber MM. Identification of *Coxiella burnetii* type IV secretion substrates required for intracellular

902   replication and *Coxiella*-containing vacuole formation volume Ph.D. thesis: Texas A&M

903   University 2014.

904   Weber MM, Chen C, Rowin K et al. Identification of *Coxiella burnetii* type IV secretion substrates

905   required for intracellular replication and Coxiella-containing vacuole formation. *J Bacteriol*

906   2013;**195**: 3914-24.

907   Williams JC, Waag DM. Antigens, virulence factors, and biological response modifiers of *Coxiella*

908   *burnetii:* strategies for vaccine development. Q fever: the biology of Coxiella burnetii 1991: 175-

909   222.

910   Yu NY, Wagner JR, Laird MR et al. PSORTb 3.0: improved protein subcellular localization prediction

911   with refined localization subcategories and predictive capabilities for all prokaryotes.

912   *Bioinformatics* 2010;**26**: 1608-15.

913   Zamboni DS, McGrath S, Rabinovitch M et al. *Coxiella burnetii* express type IV secretion system

914   proteins that function similarly to components of the Legionella pneumophila Dot/Icm system.

915   *Mol Microbiol* 2003;**49**: 965-76.

916 Zusman T, Aloni G, Halperin E et al. The response regulator PmrA is a major regulator of the icm/dot

917 type IV secretion system in *Legionella pneumophila* and *Coxiella burnetii*. *Mol Microbiol*

918 2007;**63**: 1508-23.

919 Zusman T, Feldman M, Halperin E et al. Characterization of the icmH and icmF genes required for

920 Legionella pneumophila intracellular growth, genes that are present in many bacteria associated

921 with eukaryotic cells. *Infect Immun* 2004;**72**: 3398-409.

922 Zusman T, Yerushalmi G, Segal G. Functional similarities between the icm/dot pathogenesis systems of

923 Coxiella burnetii and Legionella pneumophila. *Infect Immun* 2003;**71**: 3714-23.

924

925

ACCM-D grown *C. burnetii*
Passages: 1, 3, 5, 10, 12, 16, 21, 31, 42, 51, 61,67

TRANSCRIPTOMICS

GENOMICS

Modified Hot-trizol RNA extraction

Cultivation and sample preparation

Genomic DNA extraction

Align reads to references genome: **HISAT2**
Assemble alignments to potential transcripts: **STRINGTIE**

Sequencing and Assembly

Trim adaptors: **TRIMMOMATIC**
Quality control: **FASTQC**
Assemble on Kbase: **SPADES, VLEVET, IDBA-UD**
Compare assemblies: **QUAST**
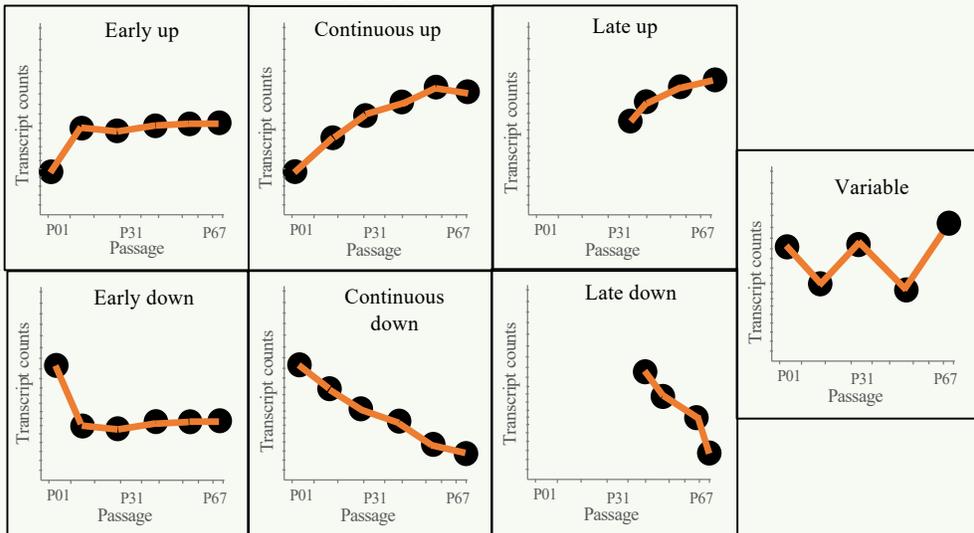Quality and completion check: **CHECKM**

Quantify differential expression of each genes
DESEQ2; BIOCONDUCTOR

Identify significant Differentially Expressed Genes (DEGs)

Analysis

Identify mutations:
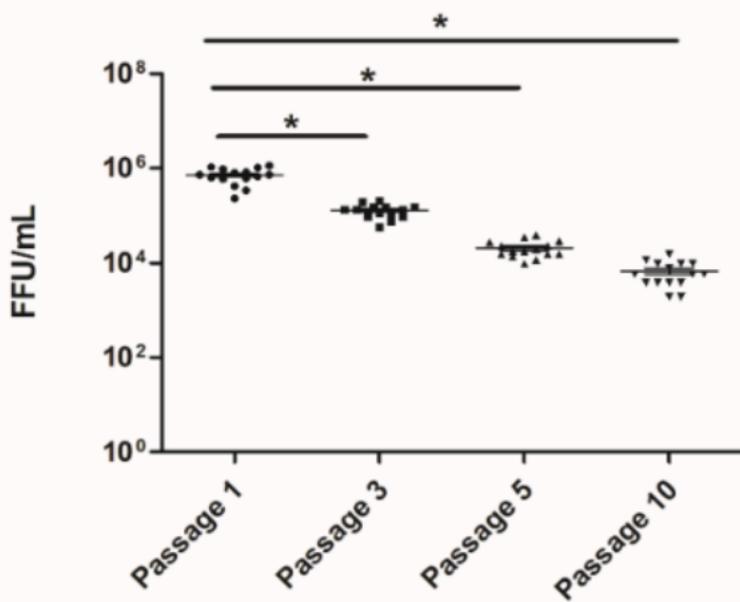a. SNPs(Single nucleotide Polymorphism)
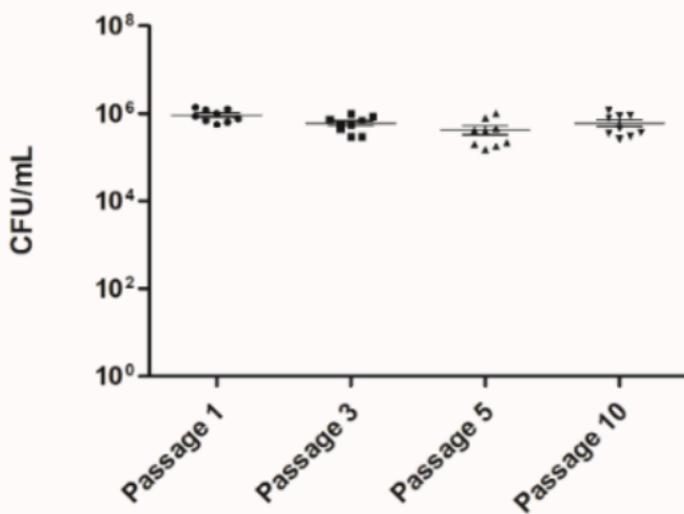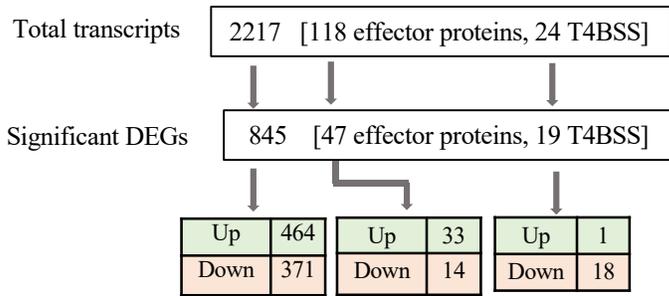b. DIPs(Deletion/Insertion Polymorphisms)
BRESEQ

VERDICTs

Mapping pathways: **GHOSTKOALA, KEGG**

In-depth analysis for genes encoding T4BSS and effector proteins.

A

FFU/mL

10^8
10^6
10^4
10^2
10^0

Passage 1   Passage 3   Passage 5   Passage 10

B

CFU/mL

10^8
10^6
10^4
10^2
10^0

Passage 1   Passage 3   Passage 5   Passage 10

A

Total transcripts

| 2217 | [118 effector proteins, 24 T4BSS] |

Significant DEGs

| 845 | [47 effector proteins, 19 T4BSS] |

| Up | 464 |
| Down | 371 |

| Up | 33 |
| Down | 14 |

| Up | 1 |
| Down | 18 |

B

Number of differentially expressed genes in each passage

C

Downregulated genes (n=371)

Upregulated genes (n=464)

Classification of genes based on their gene expression verdict

D

Downregulated genes n=371

Upregulated genes n=464

Classification of genes based on the number of passage with significant expression

A

Effectors

Outer
membrane

DotH

DotC

IcmX

DotG DotG

DotD

DotK

Periplasm

YajC

YidC
Sec DEFGY

IcmF IcmH

DotI DotJ

DotF

DotF

DotL

DotM

DotN

DotE DotP IcmV IcmT

DotA

Inner
membrane

SecA

FtsY

Cytoplasm

ATP
ADP

SecB

DotO

DotB

Effectors

IcmW

IcmS

IcmQ

ATP
ADP

B

1533280

dotO dotN dotP dotE dotF dotG dotH dotJ dotI dotK dotL dotM icmQ

8
genes

IcmT icmS dotD dotC dotB

Transporter Hypothetical

dotA icmV icmW

Hypothetical

icmX

1565686

Hypothetical

284936

icmF dotU

288714

6A.

Legend: ■ Total ■ Differentially expressed

Categories (x-axis): Amino acid transport and metabolism; Carbohydrate transport and metabolism; Cell cycle control, cell division...; Cell motility; Cell wall/membrane/envelope biogenesis; Coenzyme transport and metabolism; Cytoskeleton; Defense mechanisms; Energy production and conversion; Extracellular structures; Function unknown; General function prediction only; Inorganic ion transport and metabolism; Intracellular trafficking, secretion, and...; Lipid transport and metabolism; Mobilome: prophages, transposons; No hits on COG; Nucleotide transport and metabolism; Posttranslational modification, protein...; Replication, recombination and repair; Secondary metabolites biosynthesis,...; Signal transduction mechanisms; Transcription; Translation, ribosomal structure and...
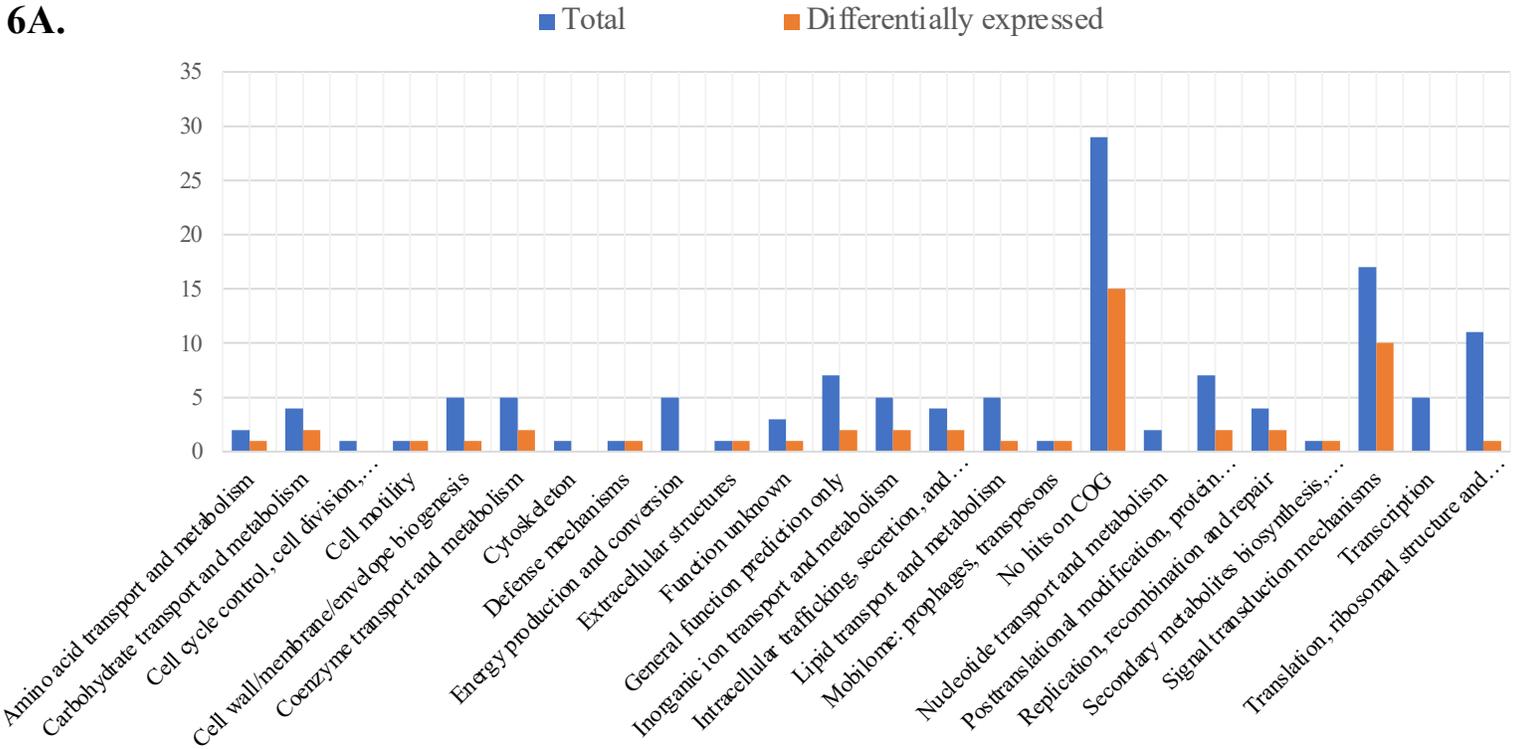
B.
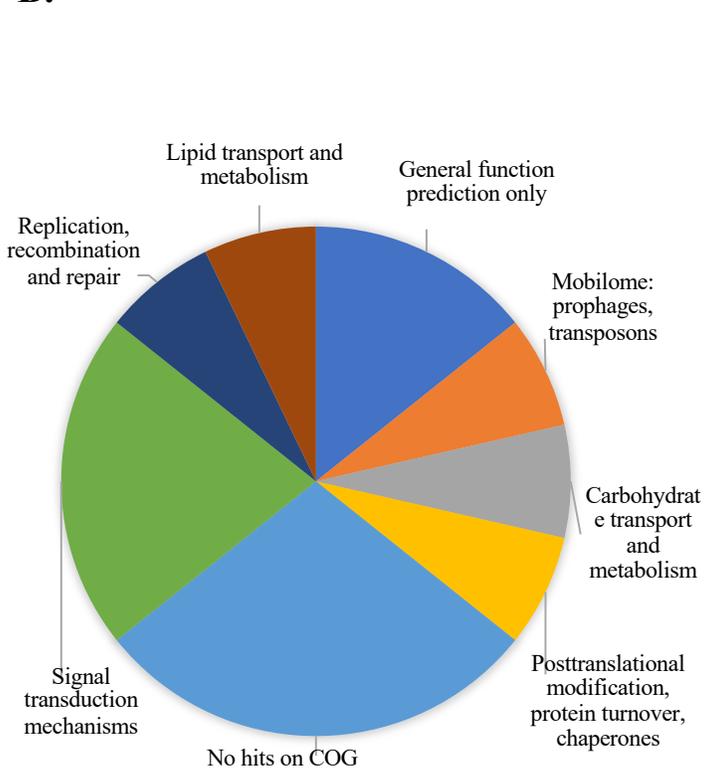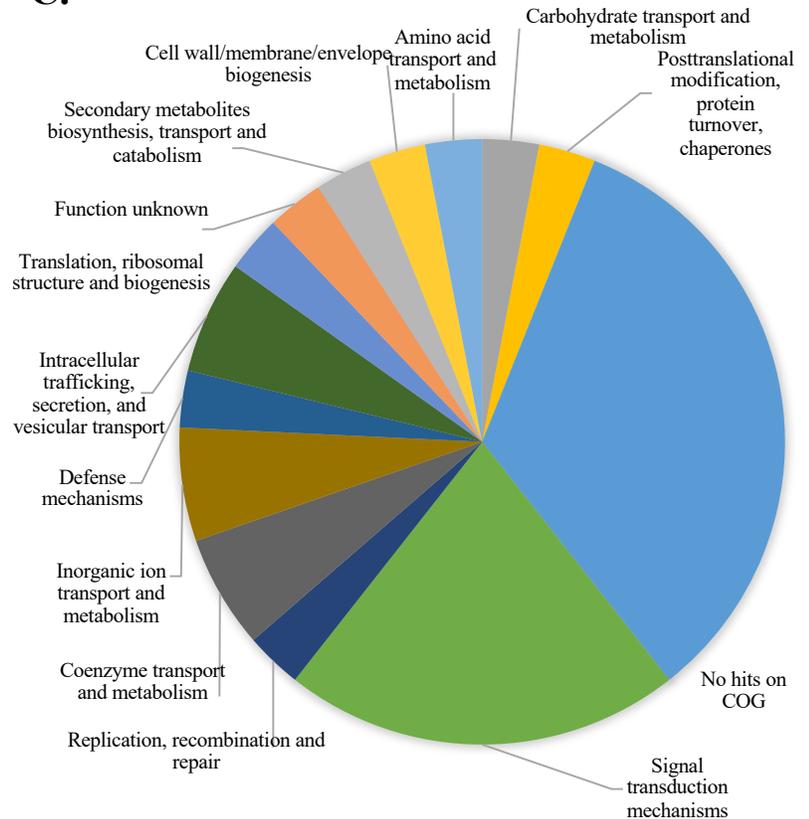
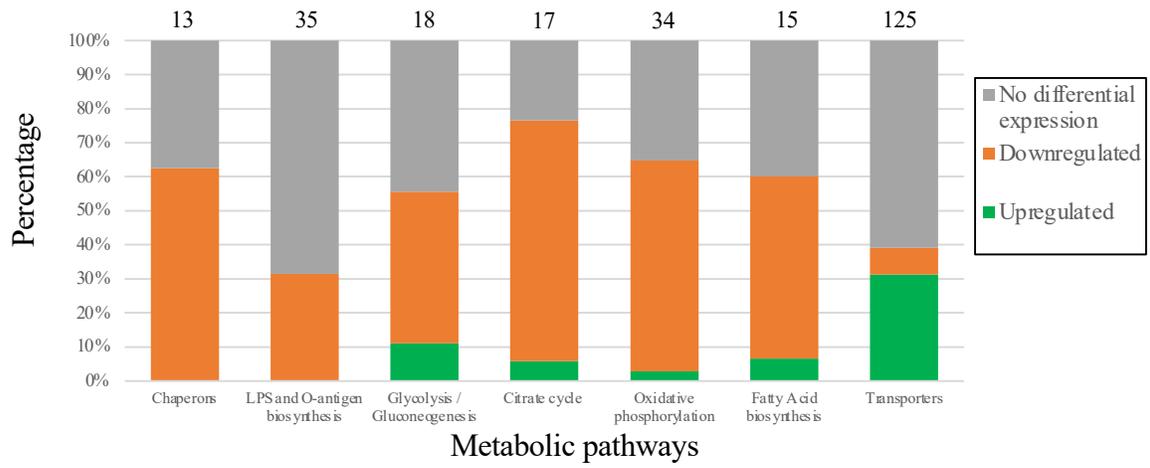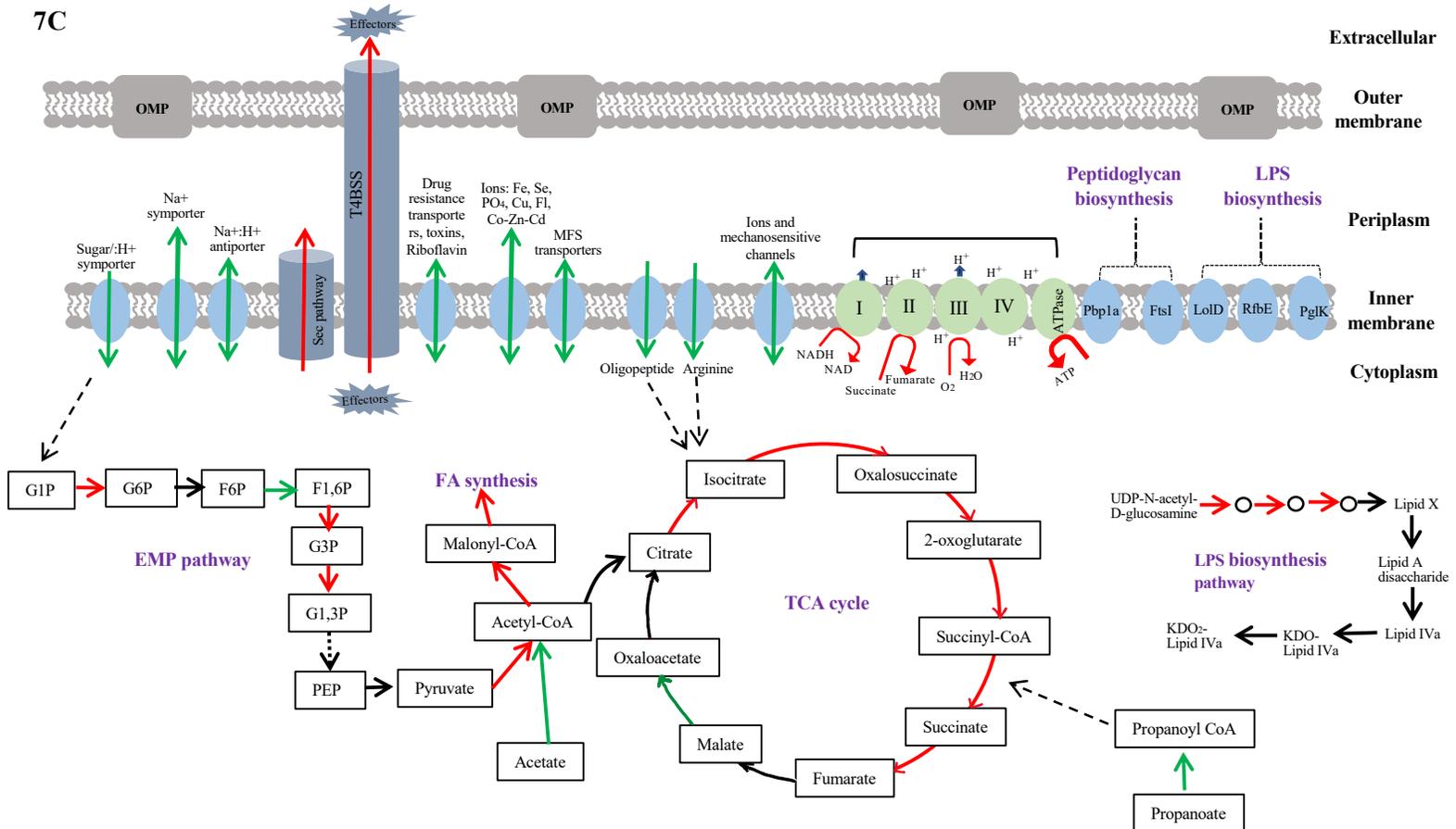Downregulated effectors (n=14)

C.

Upregulated effectors (n=33)

7A
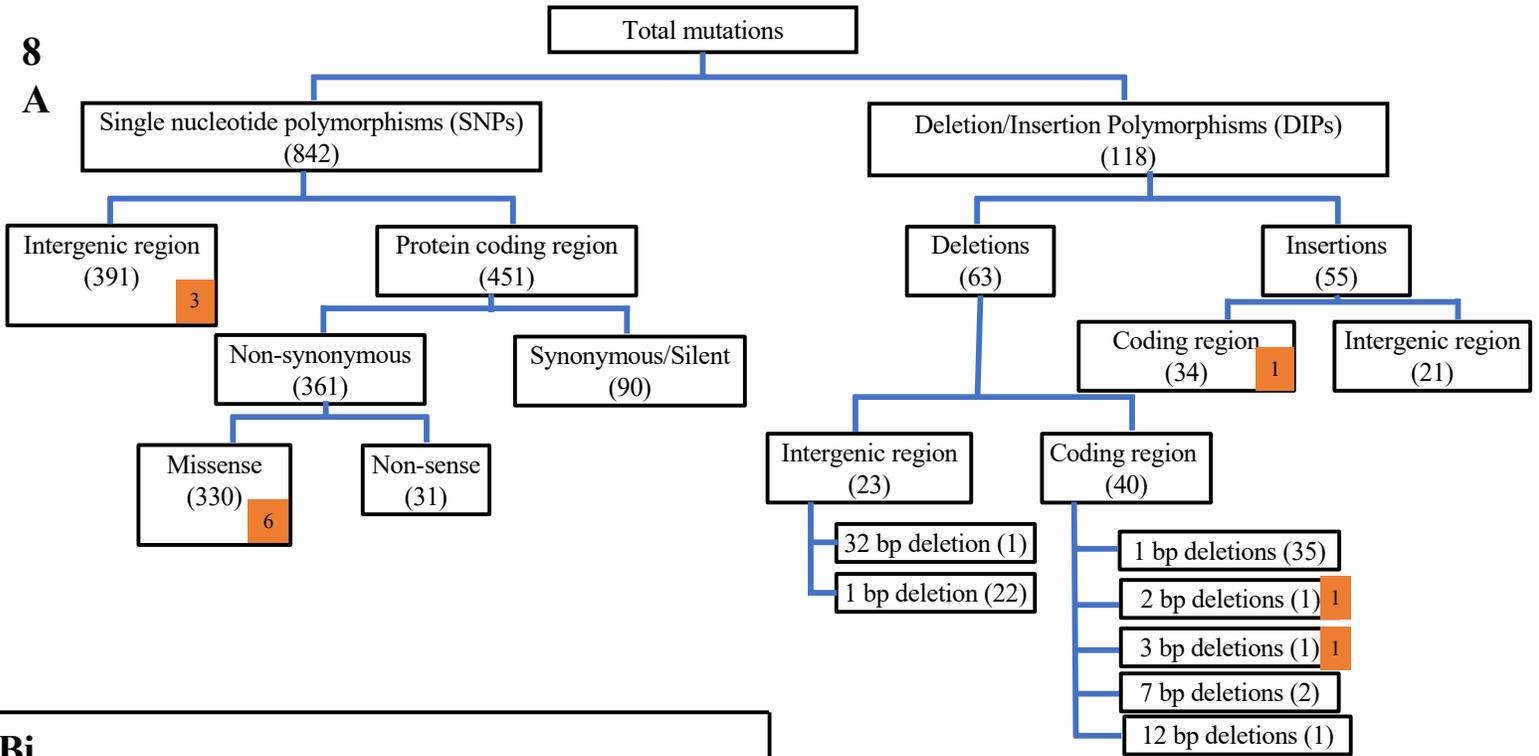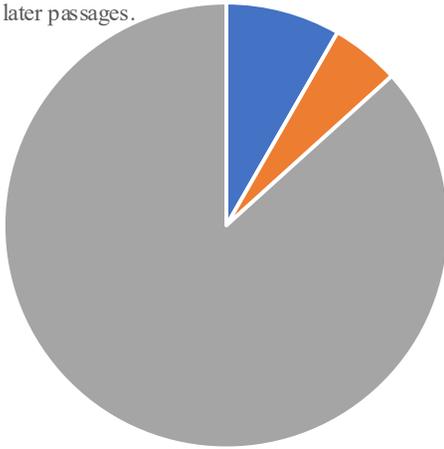
Percentage

Metabolic pathways

13 35 18 17 34 15 125

- No differential expression (grey)
- Downregulated (orange)
- Upregulated (green)

Categories: Chaperons | LPS and O-antigen biosynthesis | Glycolysis / Gluconeogenesis | Citrate cycle | Oxidative phosphorylation | Fatty Acid biosynthesis | Transporters

7B

Percentage

Transporters classification according to the substrates

10 2 2 16 9 3 9 16 20 6 4 10 4 4 2 3 5

- Not upregulated (grey)
- Upregulated (green)

Categories: Amino acid | Cell division | Fatty Acid | drug resistance | General | Glycine betaine | Ion/mechanosensitive channels | Ions | Outer membrane | MFS | Na+ symporter | Na+:H+ antiporter | Oligopeptide | Phospholipid | Sugar | Toxins | Vitamins

7C

Extracellular

Outer membrane

OMP   OMP   OMP   OMP

Periplasm

Inner membrane

Cytoplasm

Sugar/:H+ symporter | Na+ symporter | Na+:H+ antiporter | Sec pathway | T4BSS | Drug resistance transporters, toxins, Riboflavin | Ions: Fe, Se, PO4, Cu, Fl, Co-Zn-Cd | MFS transporters | Oligopeptide | Arginine | Ions and mechanosensitive channels

Effectors

Peptidoglycan biosynthesis | LPS biosynthesis

I | II | III | IV | ATPase | Pbp1a | FtsI | LolD | RfbE | PglK

H^+ H^+ H^+ H^+ H^+ H^+

NADH → NAD
Fumarate → Succinate
O_2 → H_2O
ATP

EMP pathway

G1P → G6P → F6P → F1,6P → G3P → G1,3P → PEP → Pyruvate

FA synthesis

Malonyl-CoA → Acetyl-CoA
Acetate → Acetyl-CoA
Acetyl-CoA → Citrate

TCA cycle

Isocitrate → Oxalosuccinate → 2-oxoglutarate → Succinyl-CoA → Succinate → Fumarate → Malate → Oxaloacetate → Citrate

Propanoyl CoA → Succinyl-CoA
Propanoate → Propanoyl CoA

LPS biosynthesis pathway

UDP-N-acetyl-D-glucosamine → ○ → ○ → ○ → Lipid X → Lipid A disaccharide → Lipid IVa → KDO-Lipid IVa → KDO2-Lipid IVa

8
A

Classification of total SNPs according to their occurrence

Classification of total DIPs according to their occurrence