The complete genome sequence of the crayfish pathogen Candidatus *Paracoxiella cheracis* n.g. n.sp. provides insight into pathogenesis and the phylogeny of the Coxiellaceae family.

Danielle J. Ingle[1#], Calum J. Walsh[1,2#], Genevieve R. Samuel[3], Ryan R. Wick[1,2], Nadav Davidovich[4], Eleonora Fiocchi[5], Louise M. Judd[2], Jennifer Elliman[6], Leigh Owens[6], Timothy P. Stinear[1,2], Andrea Basso[6], Tobia Pretto[6], Hayley J. Newton[3]

[1] Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

[2] Centre for Pathogen Genomics, University of Melbourne, Parkville, Victoria Australia

[3] Infection Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, VIC 3800, Australia

[4] Israeli Veterinary Services, Bet Dagan, Israel

[5] Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy

[6] College of Public Health, Medical and Veterinary Sciences, James Cook University, 1 Solander Road, Townsville, QLD, 4811, Australia

[#] authors contributed equally

Corresponding author:

Hayley J. Newton

Ph: +619902 9146

Email: hayley.newton@monash.edu

Running Title: Analysis of the *Paracoxiella cheracis* genome

Keywords: Candidatus *Paracoxiella cheracis*, Nanopore long-read sequencing, Coxiellaceae, Dot/Icm type 4 secretion system, effector proteins

ABSTRACT

The Coxiellaceae bacterial family, within the order Legionellales, is defined by a collection of poorly characterized obligate intracellular bacteria. The zoonotic pathogen and causative agent of human Q fever, *Coxiella burnetii*, represents the best characterized member of this family. Coxiellaceae establish replicative niches within diverse host cells and rely on their host for survival, making them challenging to isolate and cultivate within a laboratory setting. Here we describe a new genus within the Coxiellaceae family that has been previously shown to infect economically significant freshwater crayfish. Using culture-independent long-read metagenomics, we reconstructed the complete genome of this novel organism and demonstrate that the previously referred to as Candidatus *Coxiella cheraxi* represents a novel genus within this family, herein denoted Candidatus *Paracoxiella cheracis*. Interestingly, we demonstrate that Candidatus *P. cheracis* encodes for a complete, putatively functional Dot/Icm type 4 secretion system that likely mediates the intracellular success of this pathogen. In silico analysis defined a unique repertoire of Dot/Icm effector proteins and highlighted homologues of several important *C. burnetii* effectors including a homologue of CpeB that was demonstrated to be a Dot/Icm substrate in *C. burnetii*.

IMPORTANCE

Using long-read sequencing technology we have uncovered the full genome sequence of Candidatus *Paracoxiella cheracis*, a pathogen of economic importance in aquaculture. Analysis of this sequence has revealed new insight into this novel member of the Coxiellaceae family, demonstrating that it represents a new genus within this poorly characterized family of intracellular organisms. Importantly, the genome sequence reveals invaluable information that will support diagnostics and potentially both preventative and treatment strategies within crayfish breeding facilities. Candidatus *P. cheracis* also represents a new member of Dot/Icm pathogens that rely on this system to establish an intracellular niche. Candidatus *P. cheracis* possesses a unique cohort of putative Dot/Icm substrates that constitute a collection of new eukaryotic cell biology manipulating effector proteins.

INTRODUCTION

Australian native freshwater redclaw crayfish, *Cherax quadricarinatus*, are bred for commercial market. Aquaculture of these animals has led to the observation of infectious diseases that impact both wild and farmed *C. quadricarinatus* (reviewed in (1)). In particular, a Rickettsial-like intracellular bacterium, designated TO-98, isolated from infected crayfish was recognized as the likely causative agent of mass mortality in farmed redclaw crayfish in the early 1990s (2, 3). Subsequent investigations were able to recapitulate disease in experimentally infected crayfish and demonstrate that TO-98 causes a lethal infection in redclaw crayfish (4). This study conducted 16s rRNA sequence analysis and revealed that TO-98 was closely related to the zoonotic human pathogen *Coxiella burnetii*, coining the name Candidatus *Coxiella cheraxi* (4).

The initial observation that this pathogen is a member of the Coxiellaceae family has been confirmed by examination of additional DNA sequences including comparison of specific gene sequences (5) and MinION sequencing data showing fragments of DNA with closest similarity to *C. burnetii* and *Coxiella*-like endosymbionts (C-LEs) (6). However, this fragmented genomic data does not allow complete and comprehensive understanding of the relationship between *C. burnetii* and Candidatus *C. cheraxi*.

C. burnetii, the causative agent of human Q fever, has a large reservoir in domesticated ruminants (7). *C. burnetii* is a Gammaproteobacteria within the order Legionellales and family Coxiellaceae. While the order Legionellales is best known for the two major human pathogens, *Legionella pneumophila* and *C. burnetii*, these bacteria are distantly related and the order is poorly characterized, including over 450 uncultured genera (8). The diversity of Legionellales has been revealed by increased sampling of different environments and ecological surveys employing metagenomic techniques that have demonstrated the presence of Legionellales within arthropods, amoeba and aquatic environments (reviewed in (9)). Most of these newly discovered Legionellales are likely nonpathogenic bacteria that have evolved mutualistic relationships with non-vertebrate hosts. However, recent data suggests that some poorly characterized Legionellales harbor the potential to be zoonotic

pathogens and/or pathogenic to important animal species. This includes a recently identified *Coxiella* species found in the placenta of fur seals (10).

Essential to the ability of *C. burnetii* and *L. pneumophila* to replicate intracellularly, and cause disease, is the Dot/Icm type IVB secretion system (T4SS) (11-14). This multiprotein apparatus is responsible for the delivery of a large cohort of effector proteins into the host cell where they modulate a range of cellular functions and pathways (reviewed in (15)). Dot/Icm-deficient *L. pneumophila* are rapidly destroyed via lysosomal degradation as many of the effectors act to remodel the phagosome and block endocytic maturation (reviewed in (16)). In contrast, internalized *C. burnetii* is trafficked through the endocytic pathway before the lysosomal environment triggers activation of the Dot/Icm system (17, 18). Dot/Icm-deficient *C. burnetii* are not destroyed by the lysosome but they are also incapable of intracellular replication (11, 12). Interestingly, comparative genomic studies show that the Dot/Icm system is lost or pseudogenized in C-LE, indicating its association with pathogenesis (19).

Despite *L. pneumophila* and *C. burnetii* relying on the same key virulence factor, the cohort of effector proteins translocated by this secretion system are pathogen specific. This genetic divergence is represented by phenotypic divergence through the establishment of distinct intracellular niches. The biochemical and functional characterization of these effector proteins is an active area of research as these novel proteins facilitate insight into both mechanisms of pathogenesis and novel strategies to manipulate eukaryotic cell biology.

Here we report the complete genome sequence of Candidatus *C. cheraxi* revealing a more distant relationship to *C. burnetii* than expected, notably that the Candidatus *C. cheraxi* represents a novel genus within Coxiellaceae. We propose to rename this crayfish pathogen Candidatus *Paracoxiella cheracis.* Interestingly, we report that this novel Coxiellaceae family member possesses a Dot/Icm T4SS most closely associated with *C. burnetii*. *In silico* analysis allowed us to identify a large cohort of putative effectors of this secretion system with only 12 out of 238 predicted effectors sharing identity to known *C. burnetii* T4SS effectors. We have been able to demonstrate translocation of the Candidatus *P. cheracis* CpeB homologue via the *C. burnetii* Dot/Icm system. This study provides additional insight into the poorly

characterized Coxiellaceae family and demonstrates that the Dot/Icm secretion system is a common tool used by these bacteria to communicate with their host. It also demonstrates the utility of long-read sequencing approaches for generating a complete metagenome-assembled genome for an obligate intracellular pathogen.

METHODS

Collection and preparation of genetic material.

Samples of hepatopancreas from infected redclaw crayfish were collected during an outbreak of crayfish rickettsiosis, which occurred after the crayfish were imported from Australia to Israel in 2019 (20). Briefly, 10,000 juvenile Redclaw crayfish (2–3 weeks old) were imported to Israel and quarantined in a facility. From April to July, juvenile mortality sharply increased despite the absence of clinical signs or gross lesions. Specimens collected for histological and molecular evaluations were sent to the Istituto Zooprofilattico Sperimentale delle Venezie in Legnaro, Italy, and revealed a severe infection by Candidatus *C. cheraxi.* DNA was extracted from these samples, as previously described (20), and sent to Monash University, Australia, for further analysis.

Long-read sequencing and genome assembly

Metagenomic sequencing of extracted DNA was performed on two separate sequencing runs. The first run was performed on an Oxford Nanopore Technologies (ONT) MinION while the second employed a PromethION. Both runs were carried out on R10.4.1 flow cells. Basecalling and quality filtering was performed by Dorado (v. 0.5.0) using models dna_r10.4.1_e8.2_400bps_sup@v4.1.0 and dna_r10.4.1_e8.2_400bps_sup@v4.3.0 respectively. Reads shorter than 1kbp were discarded. Kraken2 (21) was used for initial read-level classification of metagenomic data.

Initial assemblies were generated using Flye (22), MetaMDBG (23), Miniasm (24), and Raven (25), each of which produced a single, continuous sequence representing the bacterial chromosome, and most of which contained two medium-sized plasmids at a similar read depth to the chromosome. For each assembly, the length-filtered reads were mapped to the chromosome and plasmids Minimap2 (26) and mapped reads were retained for downstream analysis. The chromosomal and plasmid reads were then downsampled to 500× coverage using Filtlong (github.com/rrwick/Filtlong) and combined into a single high-quality assembly with Trycycler (27) incorporating 24 separate assemblies from six assemblers: Canu (28), Flye (22), Miniasm (24), NECAT (29), NextDenovo (30), and Raven (25). The final complete genome assembly was annotated with Bakta v1.9.1 (31). Full details of the individual steps

for quality controls, kraken assignment and genome assembly are available in **Supplementary File 1**.

Phylogeny of Coxiellaceae Family

The Genome Taxonomy Database Toolkit (GTDB-Tk) (32) classified this genome as Coxiellaceae. Representative isolates from the order Coxiellales (gtdb.ecogenomic.org/tree?r=o__Coxiellales) were used in comparative analyses to infer the relative location of Candidatus *P. cheracis* in the order. Given the large number of Coxiellales genomes in the GTDB database, the majority of which belong to the clinically important species *Coxiella burnetii*, the dataset was first dereplicated by Assembly Dereplicator (github.com/rrwick/Assembly-Dereplicator) to a Mash distance of 0.001, resulting in a representative set of 60 genomes capturing the available diversity of the Order. A phylogeny was inferred using GTDB-Tk (33), a marker gene-based tool for taxonomic classification and phylogenetic placement of prokaryotic genomes based on the Genome Taxonomy Database (GTDB) taxonomy system (34) was used to infer a phylogeny for the representative 60 genomes and to taxonomically classify our long-read assembly.

Investigation of the complete genome for features of interest.

The genome content of Candidatus *P. cheracis* was explored using publicly available tools. First, known antimicrobial resistance (AMR) determinants were screened with abritAMR (v1.0.14) with the AMRfinderPlus database (v2022-08-09.1) with no species flag and default parameters (35). Mob-typer, which is part of Mob-Suite, was used to identify plasmid replicons and mob genes with default parameters (36, 37). Phage defense mechanisms were identified with PADLOC v2 using the annotated gff file as input and default parameters (38).

Putative Dot/Icm effectors were screened in the complete genome with Bastion4 (39) and T4Sepp (40) tools. The predictions were compared, with hits for putative IS elements not included in further analysis due to the likelihood of being false positive results.

Components of the T4SS were identified with SECRET4 (41). T4SSs were also identified in selected, publicly available genomes that have been established to have functional T4SSs. These included two *C. burnetii* genomes as the type strains for the species (RSA493 accession: AE016828.3, and Dugway accession: CP000733.1), two genomes from the *Legionella* genus, *Legionella pneumophila* subsp. *pneumophila* str. Philadelphia 1 (accession: AE017354.1) and *Legionella longbeachae* NSW150 (accession:NC_013861.1)(42) and the Candidatus *Rickettsiella viridis* Ap-RA04 genome (accession: AP018005.1). We also screened representative isolates with complete genomes from *Coxiella* like-endosymbionts (C-LEs) (accessions: CP021379.1, CP011126.1, CP033868.1 and CP064834.1) with SECRET4 confirming the reported lack of a complete T4SS in these genomes (43).

Protein identity between the six genomes was undertaken with *blastp* using the Dot/Icm proteins from *C. burnetii* RSA493 (accession: AE016828.3) as a query sequence. For the three *Coxiella* genomes, the output was filtered to the top hit for each with a bitscore >100. A modified approach was undertaken for the two *Legionella* and Candidatus *Rickettsiella viridis* genomes due to the lower sequence homology. For these three genomes, the results were filtered to the loci that encode the T4SS and the hits with highest bitscore for each protein was included.

Visualization of the Candidatus *P. cheracis* genome with features of interest including phage defense, T4SS and putative Dot/Icm effectors was done in BRICK (github.com/esteinig/brick). The phylogenetic tree was visualized in R using ape (44) and ggtree (45). The R package gggenomes was used to plot the T4SS (github.com/thackl/gggenomes) and pheatmap was used to visualize the protein identity (github.com/raivokolde/pheatmap).

Cultivation of bacterial and tissue culture cells

C. burnetii Nine Mile phase II (NMII) strain RSA439 clone 4 was axenically cultivated in liquid ACCM-2 or ACCM-2 agarose at 37°C in 5% CO_2 and 2.5% O_2 as previously described (46). When required chloramphenicol was added to *C. burnetii* cultures at 3 µg/ml. HeLa CCL-2 cells (ATCC, VA) were maintained in Dulbecco's Modified

Eagle's Media (DMEM) supplemented with 10% heat inactivated fetal bovine serum at 37°C in 5% CO_2.

Construction of *C. burnetii* translocation reporter strains

The open reading frame encoding for the Candidatus *P. cheracis* CpeB homologue (*Pc*CpeB) was synthesized by IDT and cloned into pJB-CAT:BlaM (47) using BamHI and XhoI restriction enzymes. The resulting pBlaM-*Pc*CpeB was introduced into *C. burnetii* strains using a standard electroporation protocol (11). Following selection of transformants, expression of BlaM fusion proteins was confirmed using an anti-BlaM (1:1500, QED Biosciences). *C. burnetii* strains carrying either pJB-CAT:BlaM or pJB-CAT:BlaM:CBU0021 were used as negative and positive translocation controls (48).

BlaM reporter translocation assay

Translocation assays were conducted as previously reported (17, 49). HeLa CCL2 cells were seeded in 96-well black, clear bottom tissue culture plates at a density of 10^4 per well. Approximately 24 h later, cell monolayers were infected with *C. burnetii* strains at an MOI of 300. Following a 22 h incubation, wells were loaded with CCF2-AM substrate from the LiveBLAzer FRET loading kit (Thermo Fisher Scientific) and incubated at room temperature for 2 h. Wells were excited at 415 nm and the emission at 450 nm (blue) and 520 nm (green) was collected using a Clariostar fluorescence plate reader. Representative images were acquired using an Invitrogen EVOS™ FL Imaging System, images taken at 10× objective.

Data availability

The complete genome and the long-read data are available at BioProject PRJNA1189271.

RESULTS

A putative new genus in the Coxiellaceae family

Here we describe the complete genome of a putative new genus within Coxiellaceae, Candidatus *C. cheraxi,* that is the likely causative agent of mass mortality events in *C. quadricarinatus*. The Candidatus *C. cheraxi* genome consisted of one chromosome (2,233,862 bp) and two smaller plasmids (59,783 and 34,021 bp). The chromosome is similar in size to previously sequenced *C. burnetii* isolates ranging from approximately 1.9 Mb to 2.2 Mb (50). To confirm that this organism is related to the original Candidatus *C. cheraxi* TO-98 organism, the 21 available sequence reads from TO-98 (6) were mapped to this genome with an average identify of 89%. This is a high level of similarity, even with the TO-98 dataset produced using older minion chemistry (R9.4.1), provided confidence that this genome is representative of the same species as TO-98.

The Candidatus *C. cheraxi* genome was classified as Coxiellaceae by GTDB-Tk (32), meaning that the novel genome could be classified to the family level but no further, suggesting that this genome represents a novel genus, supported by a Relative Evolutionary Divergence (RED) value of 0.79373 (32). Based on this data we propose to consider this Coxiellaceae as a novel genus: Candidatus *Paracoxiella cheracis*, also adopting the adjusted genus nomenclature (51).

A phylogenetic tree inferred from bacteria-specific marker genes placed Candidatus *P. cheracis* near genomes of other Coxiellaceae bacteria (**Figure 1**) isolated from similar aquatic ecological niches and distant to *C. burnetii* (largely associated with mammalian hosts) and the *Coxiella*-like endosymbionts (C-LEs) that are associated with ticks as hosts (**Figure 1**). Interestingly, Kraken2 classification of the chromosomal reads using a GTDB-based database identified multiple bacterial families in the read set. These include 10.6% Alteromonadaceae, 4.1% Vibrionaceae, 3.7% Enterobacteriaceae, 2.9% Burkholderiaceae, 2.5% Legionellaceae, 2.3% Moraxellaceae and only 8.5% Coxiellaceae (**Supplementary File 2**), which are all members of the Gammaproteobacteria class. This demonstrates that Kraken2 results for an isolate in a novel taxon not represented in the database can be misleading and may falsely suggest a mixed sample. However,

assembly of the long-read data successfully recovered the bacterial chromosome and two plasmids.

The complete genome of Candidatus *P. cheracis* was explored for antimicrobial resistance, phage defense mechanisms, the Dot/Icm T4SS and putative effectors of this system (**Figure 2**). No known AMR mechanisms were identified, consistent with the AMR profiles of *C. burnetii* genomes. No replicon or mob genes were detected in the two plasmids. Further, these two plasmid sequences had no homology to publicly available data on NCBI. Several phage defense systems were identified on the chromosome in the novel genome. Two SoFic defense systems and the AbiE system, comprised of AbiEi and AbiEii components, were identified in the genome. The remaining ten defense systems identified were from the curated PADLOC database. Multiple putative effectors were identified (expanded upon in Predicted substrates of the *C. cheraxi* Dot/Icm secretion system), in addition to many copies and variants of a transposase that shares identity with a transposase from a symbiont of *Solemya velum*, the Atlantic awning clam [ref, accession: JRAA00000000] (52).

Detection and analysis of the Dot/Icm locus encoded by Candidatus *P. cheracis*

The Dot/Icm system was detected in the Candidatus *P. cheracis* genome, suggesting that the isolate has a functional T4SS. The site of integration into the chromosome and gene arrangement were similar to the two isolates of *C. burnetii* examined here (**Figure 3A**). The two *C. burnetii* genomes have additional genes between *icmT* and *icmQ,* some of which have been shown to be effector proteins (53), which are lacking in the Candidatus *C. cheraxi* genome. Additionally, *icmT* and *icmS* were encoded in the reverse orientation in the Candidatus *P. cheracis* genome compared to the two *C. burnetii* genomes. These two genes were in the same orientation as the two *Legionella* and Candidatus *R. viridis* genomes despite the Dot/Icm system being fragmented into multiple regions in these three genomes. Of note, the protein homology between 21 genes of the Dot/Icm system differed within the six genomes by species and genus (**Figure 3B**). The proteins from the *C. burnetii* Dugway isolate were homologous to the Dot/Icm proteins of *C. burnetii*

RSA493 (>99% protein identity for all). In contrast, the protein identity for the Dot/Icm proteins in Candidatus *P. cheracis* varied between 52% (IcmG) and 86% (IcmB) identity relative to the *C. burnetii* RSA493 proteins. This was higher than the protein similarity of the Dot/Icm proteins from the two *Legionella* and the Candidatus *R. viridis* genomes which ranged between 24% (IcmX in *L. longbeachae*) to 67% (DotB for Candidatus *R. viridis*). As previously reported, the Dot/Icm system was not detected in the three C-LEs genomes.

Predicted substrates of the Candidatus *P. cheracis* Dot/Icm secretion system

Given the conservation of genes encoding for the Dot/Icm apparatus we hypothesized that Candidatus *P. cheracis* possesses a unique repertoire of putative T4SS substrates. To identify these proteins, two effector prediction pipelines, Bastion4 and T4SEpp, were employed. Bastion4, bacterial secretion effector predictor for T4SS, is an ensemble effector predictor based on six distinct machine learning models (39). More recently, T4SEpp was developed using an integrated pipeline that incorporates homology-based predictions with machine learning models to generate a prediction score for the likelihood of a protein being a T4SS substrate (40). Combined, these tools predicted 355 putative Dot/Icm substrates within the Candidatus *P. cheracis* genome. Bastion4 predicted 296 effectors and T4SEpp predicted 114 with 55 putative effectors predicted by both approaches (**Supplementary File 3**). Bastion4 predicted 117 instances of IS elements as putative effectors which are likely false positives bringing the pool of potential Dot/Icm effectors to 238. Of particular interest, 66 putative effectors (25 predicted by both tools) have no significant similarity to protein sequences found in NCBI.

Candidatus *P. cheracis* CpeB, *Pc*CpeB, can be translocated by the *C. burnetii* Dot/Icm system.

Each of the predicted Dot/Icm effectors were examined for similarity to known Dot/Icm effectors of *L. pneumophila* and *C. burnetii*. No putative effectors showed similarity within the *L. pneumophila* effector cohort however 12 putative effectors have similarity to putative and experimentally characterized *C. burnetii* effectors (**Table 1**).

To test the feasibility of using *C. burnetii* as a surrogate bacterium to demonstrate Dot/Icm-dependent translocation of putative Candidatus *P. cheracis* effectors, the BlaM translocation reporter assay was employed. This assay is routinely used to demonstrate Dot/Icm substrate translocation into host cells and relies on bacterial delivery of a β-lactamase (BlaM) enzyme into the host cytosol where its activity can be measured by cleavage to the BlaM substrate CCF2-AM (11). Dot/Icm-dependent translocation of the Candidatus *P. cheracis* homologue of *C. burnetii* effector CpeB (*Pc*CpeB) was confirmed (**Figure 4A and B**). *C. burnetii* WT and Dot/Icm-deficient (*icmL::*Tn) strains were engineered to express BlaM alone (negative control), BlaM-CvpB, a well characterized *C. burnetii* Dot/Icm effector (positive control) (48, 54-56) or BlaM-*Pc*CpeB (**Fig 4C**). HeLa cells were infected with these *C. burnetii* strains for 22 h before addition of CCF2-AM and subsequent detection of fluorescence emission via both a fluorescence plate reader (**Fig 4A**) and microscopy (**Fig 4B**). BlaM-*Pc*CpeB was translocated in a Dot/Icm-dependent manner. Interestingly, comparison of the amino acid sequences of *C. burnetii* CpeB (*Cb*CpeB) and *Pc*CpeB showed only 29% identity however the region determined to mediate interaction between *Cb*CpeB and Rab11 shows greater conservation potentially indicating functional similarity (purple box, **Fig 4D**) (57).

DISCUSSION

In the absence of live samples of Candidatus *P. cheracis* n.g. n. sp., progressing our understanding of this novel crayfish pathogen is limited to the exploration of genetic material. This report of the complete genome sequence of this pathogen provides both an opportunity to expand our understanding of the evolutionary relationships between members of the family Coxiellaceae and formulate hypotheses regarding the host-pathogen interactions mediated by this organism.

The phylogenetic placement, ANI to reference genomes, and Relative Evolutionary Divergence of Candidatus *P. cheracis* provides evidence that this novel genome represents a new genus within the Coxiellaceae family and is more distantly related to both *C. burnetii* and C-LEs than previously appreciated. Of note, the Dot/Icm T4SS detected in Candidatus *P. cheracis* was found to have inserted into the chromosome near the same genes as in *C. burnetii* genomes. From this, we hypothesize that the Dot/Icm T4SS may have been acquired in the most recent common ancestor of both Candidatus *P. cheracis, C. burnetii* and C-LE. Moreover, the loss of the Dot/Icm T4SS in the C-LEs which are associated with invertebrate hosts suggests that the Dot/Icm T4SS did not confer a fitness advantage to warrant being maintained in C-LEs, which are maternally inherited in ticks (19).

The Dot/Icm T4SS has co-evolved in these different genetic backgrounds, resulting in the observed differences in protein identity between Candidatus *P. cheracis* and the two *C. burnetii* reference genomes, with adaptation of the effectors to the different host niches. The observed differences in Dot/Icm protein identity between these genomes is greater than the average values recently reported for 58 *Legionella* species (58). However, the Dot/Icm components of *Legionella* with the greatest variation in protein identity were the same in the Candidatus *P. cheracis* (relative to *C. burnetii*), specifically IcmX, DotA and IcmG/DotF, consistent with exposure of these protein to hosts (58). In contrast, IcmB and DotB had the highest protein identity, which is consistent with previous work in *Legionella* species suggestive that these proteins are under functional selection (58).

Key challenges in the detection of obligate intracellular pathogens is the difficulty in culturing the organisms, and the lack of diagnostic targets that can detect these

bacteria at low concentrations in often complex metagenomic samples. Here we identified that Candidatus *P. cheracis* possesses a high copy transposase that had homology to a symbiont of *Solemya velum*, the Atlantic awning clam. Future efforts to screen for the presence of Candidatus *P. cheracis* could be facilitated by PCR amplification of the transposase in aquatic samples. This would be akin to diagnostic efforts of members of the *Bordetella* genus (59). *Bordetella pertussis,* the causative agent of whooping cough, is commonly identified through the PCR of high copy number insertion sequences (IS) with the primary target IS481. This would provide a potentially non-invasive, sensitive screening strategy for ongoing surveillance of the crayfish farms that would facilitate the prevention and absence of this pathogen in aquaculture environments. Additionally, if proven as a sensitive strategy for detection of Candidatus *P. cheracis*, this PCR approach could also be employed to test different aquatic environments for the presence of this Coxiellaceae genus.

Comparative genomics of 65 *Legionella* species has reported an expansive repertoire of putative Dot/Icm effector proteins (60). While many species possess putative effectors with specific eukaryotic domains, such as GTPase, F-box and SET domains, these motifs are encoded within different proteins that are not orthologous (60). Interestingly, only 8 Dot/Icm substrates are conserved among these 65 species of *Legionella* (60). Using two distinct *in silico* tools, we have developed a list of 238 putative Dot/Icm effectors of Candidatus *P. cheracis*. Unlike *Legionella* species, very few putative Dot/Icm effectors encoded by Candidatus *P. cheracis* possess similarity to eukaryotic domains and a large proportion, ~28%, are novel. We observed 12 predicted effectors with similarity to known or putative *C. burnetii* Dot/Icm effectors, including CpeB and CvpE which have been shown to make important contributions to the intracellular success of *C. burnetii* (48, 54, 56, 57). Further expanding the collection of sequenced Coxiellaceae genomes will facilitate greater comparison of the effector repertoire in this family, allowing us to determine whether some of these 12 common effectors act as “core” Coxiellaceae effectors that facilitate their characteristic interaction with eukaryotic cells. We were able to demonstrate Dot/Icm-dependent translocation of Candidatus *P. cheracis* CpeB (*Pc*CpeB) using *C. burnetii* as a surrogate host. In the absence of methodology for the isolation and cultivation of Candidatus *P. cheracis,* the success of this approach provides an

alternative strategy for developing our understanding of the host-pathogen interactions mediated by Candidatus *P. cheracis*.

The full genome sequence of Candidatus *P. cheracis* has highlighted potential screening or diagnostic targets and unveiled likely pathogenesis strategies. In addition, the genome sequence shows that Candidatus *P. cheracis* encodes for Mip, a peptidyl-prolyl cis-trans isomerase that has been shown to be essential for intracellular replication and pathogenesis of *C. burnetii* (61). Candidatus *P. cheracis* Mip shares 66% identity with *C. burnetii* Mip and these proteins are conserved within the PPIase catalytic domain which suggests that Mip inhibitors developed by Debowski et al., may inhibit Candidatus *P. cheracis* Mip (61). Treatment with rationally designed small molecule inhibitors of Mip may represent a plausible intervention measure for future outbreaks of Candidatus *P. cheracis* in aquaculture facilities.

Here, we have further demonstrated the effective use of ONT sequencing to reconstruct the genome of a novel intracellular pathogen - a member of a family that is difficult to culture in standard laboratory conditions - directly from its host. The complete genome of a chromosome and two plasmids enabled the exploration of a previously unexplored pathogen that is important for Australian aquaculture. ONT was advantageous here for the capture of the two plasmids, elements that are often missed by short-read metagenomic binning approaches but can be critical for virulence (62, 63) and for dealing with the significant number of IS elements throughout this genome (64). The detection of a T4SS and cohort of novel effector proteins that likely facilitate intracellular success of the pathogen provides the first insights into the host pathogen interactions of this novel pathogen. ONT sequencing has shown promise in recent years for outbreak investigations due to its low-cost, portability, minimal sample preparation, and rapid, real-time data output (65). If such an outbreak was to occur today, in addition to traditional histopathology, ONT metagenomics would enhance efforts to detect and characterize unknown emerging pathogens.

ACKNOWLEDGEMENTS

DJI is supported by a National Health and Medical Research Council (NHMRC) Investigator Grant (GNT1195210). CJW and RRW are funded through the NHMRC (GNT1194325). GRS is the recipient of an Australian Government Research Training Program Scholarship and Monash Graduate Excellence Scholarship. Research in HJNs laboratory was funded by the NHMRC (GNT2010841). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

494 **REFERENCES**

495 1. Dragičević P, Bielen A, Petrić I, Hudina S. 2021. Microbial pathogens of
496 freshwater crayfish: A critical review and systematization of the existing data
497 with directions for future research. J Fish Dis 44:221-247.
498 2. Ketterer PJ, Taylor DJ, Prior HC. 1992. Systemic rickettsia-like infection in farmed
499 freshwater crayfish, *Cherax quadricarinatus*, p 173-179. *In* Shariff M, Subasinghe
500 RP, Arthur JR (ed), Diseases in Asian Aquaculture. Asian Fisheries Society.
501 3. Owens L, Muir P, Sutton D, Wingfield M. 1992. The pathology of microbial
502 diseases in tropical Australian Crustacea, p 165-172. *In* Shariff M, Subasinghe
503 RP, Arthur JR (ed), Diseases in Asian Aquaculture. Asian Fisheries Society.
504 4. Tan CK, Owens L. 2000. Infectivity, transmission and 16S rRNA sequencing of a
505 rickettsia, *Coxiella cheraxi* sp. nov., from the freshwater crayfish *Cherax*
506 *quadricarinatus*. Dis Aquat Organ 41:115-22.
507 5. Cooper A, Layton R, Owens L, Ketheesan N, Govan B. 2007. Evidence for the
508 classification of a crayfish pathogen as a member of the genus *Coxiella*. Lett
509 Appl Microbiol 45:558-63.
510 6. Ruth Elliman J, Owens L. 2020. Confirmation that candidatus *Coxiella cheraxi*
511 from redclaw crayfish (*Cherax quadricarinatus*) is a close relative of Coxiella
512 burnetii, the agent of Q-fever. Lett Appl Microbiol 71:320-326.
513 7. Luhrmann A, Newton HJ, Bonazzi M. 2017. Beginning to Understand the Role of
514 the Type IV Secretion System Effector Proteins in *Coxiella burnetii* Pathogenesis.
515 Curr Top Microbiol Immunol 413:243-268.
516 8. Graells T, Ishak H, Larsson M, Guy L. 2018. The all-intracellular order
517 Legionellales is unexpectedly diverse, globally distributed and lowly abundant.
518 FEMS Microbiol Ecol 94.
519 9. Duron O, Doublet P, Vavre F, Bouchon D. 2018. The Importance of Revisiting
520 Legionellales Diversity. Trends Parasitol 34:1027-1037.
521 10. Gardner BR, Bachmann NL, Polkinghorne A, Hufschmid J, Tadepalli M, Marenda
522 M, Graves S, Arnould JPY, Stenos J. 2023. A Novel Marine Mammal *Coxiella*
523 *burnetii*-Genome Sequencing Identifies a New Genotype with Potential
524 Virulence. Pathogens 12.
525 11. Carey KL, Newton HJ, Luhrmann A, Roy CR. 2011. The *Coxiella burnetii* Dot/Icm
526 system delivers a unique repertoire of type IV effectors into host cells and is
527 required for intracellular replication. PLoS Pathog 7:e1002056.
528 12. Beare PA, Gilk SD, Larson CL, Hill J, Stead CM, Omsland A, Cockrell DC, Howe
529 D, Voth DE, Heinzen RA. 2011. Dot/Icm type IVB secretion system requirements
530 for *Coxiella burnetii* growth in human macrophages. MBio 2:e00175-11.
531 13. Vogel JP, Andrews HL, Wong SK, Isberg RR. 1998. Conjugative transfer by the
532 virulence system of *Legionella pneumophila*. Science 279:873-6.
533 14. Segal G, Purcell M, Shuman HA. 1998. Host cell killing and bacterial conjugation
534 require overlapping sets of genes within a 22-kb region of the *Legionella*
535 *pneumophila* genome. Proc Natl Acad Sci U S A 95:1669-74.
536 15. Qiu J, Luo ZQ. 2017. *Legionella* and *Coxiella* effectors: strength in diversity and
537 activity. Nat Rev Microbiol 15:591-605.
538 16. Hubber A, Roy CR. 2010. Modulation of host cell function by *Legionella*
539 *pneumophila* type IV effectors. Annu Rev Cell Dev Biol 26:261-83.

17. Newton HJ, McDonough JA, Roy CR. 2013. Effector protein translocation by the *Coxiella burnetii* Dot/Icm type IV secretion system requires endocytic maturation of the pathogen-occupied vacuole. PLoS One 8:e54566.
18. Newton P, Thomas DR, Reed SCO, Lau N, Xu B, Ong SY, Pasricha S, Madhamshettiwar PB, Edgington-Mitchell LE, Simpson KJ, Roy CR, Newton HJ. 2020. Lysosomal degradation products induce *Coxiella burnetii* virulence. Proc Natl Acad Sci U S A 117:6801-6810.
19. Santos-Garcia D, Morel O, Henri H, El Filali A, Buysse M, Noël V, McCoy KD, Gottlieb Y, Klasson L, Zenner L, Duron O, Vavre F. 2023. Genomic changes during the evolution of the *Coxiella* genus along the parasitism-mutualism continuum. Peer Community Journal 3.
20. Davidovich N, Fiocchi E, Basso A, Budai J, Montesi F, Pretto T. 2022. An outbreak of crayfish rickettsiosis caused by *Coxiella cheraxi* in redclaw crayfish (*Cherax quadricarinatus*) imported to Israel from Australia. Transbound Emerg Dis 69:204-212.
21. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257.
22. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540-546.
23. Benoit G, Raguideau S, James R, Phillippy AM, Chikhi R, Quince C. 2024. High-quality metagenome assembly from long accurate reads with metaMDBG. Nat Biotechnol 42:1378-1383.
24. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103-10.
25. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1:332-336.
26. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.
27. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22:266.
28. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722-736.
29. Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, Wang YX, Xing JF, Huang ZJ, Wang DP, He LJ, Luo F, Wang JX, Liu YZ, Xiao CL. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat Commun 12:60.
30. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, Ruan J, Xiao CL, Wang D, Wu DD, Wang S. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol 25:107.
31. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom 7.
32. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925-7.

587 33. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory
588 friendly classification with the genome taxonomy database. Bioinformatics
589 38:5315-5316.
590 34. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2022.
591 GTDB: an ongoing census of bacterial and archaeal diversity through a
592 phylogenetically consistent, rank normalized and complete genome-based
593 taxonomy. Nucleic Acids Res 50:D785-d794.
594 35. Sherry NL, Horan KA, Ballard SA, Gonçalves da Silva A, Gorrie CL, Schultz MB,
595 Stevens K, Valcanis M, Sait ML, Stinear TP, Howden BP, Seemann T. 2023. An
596 ISO-certified genomics workflow for identification and surveillance of
597 antimicrobial resistance. Nat Commun 14:60.
598 36. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering,
599 reconstruction and typing of plasmids from draft assemblies. Microb Genom 4.
600 37. Robertson J, Bessonov K, Schonfeld J, Nash JHE. 2020. Universal whole-
601 sequence-based plasmid typing and its utility to prediction of host range and
602 epidemiological surveillance. Microb Genom 6.
603 38. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, Nobrega FL,
604 Jackson SA. 2021. Identification and classification of antiviral defence systems
605 in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids
606 Res 49:10868-10878.
607 39. Wang J, Yang B, An Y, Marquez-Lago T, Leier A, Wilksch J, Hong Q, Zhang Y,
608 Hayashida M, Akutsu T, Webb GI, Strugnell RA, Song J, Lithgow T. 2019.
609 Systematic analysis and prediction of type IV secreted effector proteins by
610 machine learning approaches. Brief Bioinform 20:931-951.
611 40. Hu Y, Wang Y, Hu X, Chao H, Li S, Ni Q, Zhu Y, Hu Y, Zhao Z, Chen M. 2024.
612 T4SEpp: A pipeline integrating protein language models to predict bacterial type
613 IV secreted effectors. Comput Struct Biotechnol J 23:801-812.
614 41. Bi D, Liu L, Tai C, Deng Z, Rajakumar K, Ou HY. 2013. SecReT4: a web-based
615 bacterial type IV secretion system resource. Nucleic Acids Res 41:D660-5.
616 42. Cazalet C, Gomez-Valero L, Rusniok C, Lomma M, Dervins-Ravault D, Newton
617 HJ, Sansom FM, Jarraud S, Zidane N, Ma L, Bouchier C, Etienne J, Hartland EL,
618 Buchrieser C. 2010. Analysis of the *Legionella longbeachae* genome and
619 transcriptome uncovers unique strategies to cause Legionnaires' disease. PLoS
620 Genet 6:e1000851.
621 43. Brenner AE, Muñoz-Leal S, Sachan M, Labruna MB, Raghavan R. 2021. *Coxiella*
622 *burnetii* and Related Tick Endosymbionts Evolved from Pathogenic Ancestors.
623 Genome Biol Evol 13.
624 44. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and
625 Evolution in R language. Bioinformatics 20:289-90.
626 45. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. GGTREE: an R package for
627 visualization and annotation of phylogenetic trees with their covariates and
628 other associated data. Methods in Ecology and Evolution 8:28-36.
629 46. Omsland A, Beare PA, Hill J, Cockrell DC, Howe D, Hansen B, Samuel JE,
630 Heinzen RA. 2011. Isolation from animal tissue and genetic transformation of
631 *Coxiella burnetii* are facilitated by an improved axenic growth medium. Appl
632 Environ Microbiol 77:3720-5.

633 47. Voth DE, Beare PA, Howe D, Sharma UM, Samoilis G, Cockrell DC, Omsland A,
634 Heinzen RA. 2011. The *Coxiella burnetii* cryptic plasmid is enriched in genes
635 encoding type IV secretion system substrates. J Bacteriol 193:1493-503.
636 48. Newton HJ, Kohler LJ, McDonough JA, Temoche-Diaz M, Crabill E, Hartland EL,
637 Roy CR. 2014. A screen of *Coxiella burnetii* mutants reveals important roles for
638 Dot/Icm effectors and host autophagy in vacuole biogenesis. PLoS Pathog
639 10:e1004286.
640 49. Thomas DR, Garnish SE, Khoo CA, Padmanabhan B, Scott NE, Newton HJ. 2024.
641 *Coxiella burnetii* protein CBU2016 supports CCV expansion. Pathog Dis 82.
642 50. Abou Abdallah R, Million M, Delerce J, Anani H, Diop A, Caputo A, Zgheib R,
643 Rousset E, Sidi Boumedine K, Raoult D, Fournier PE. 2022. Pangenomic analysis
644 of *Coxiella burnetii* unveils new traits in genome architecture. Front Microbiol
645 13:1022356.
646 51. Oren A, Garrity GM. 2022. CANDIDATUS LIST No. 3. Lists of names of prokaryotic
647 Candidatus taxa. Int J Syst Evol Microbiol 72.
648 52. Dmytrenko O, Russell SL, Loo WT, Fontanez KM, Liao L, Roeselers G, Sharma R,
649 Stewart FJ, Newton IL, Woyke T, Wu D, Lang JM, Eisen JA, Cavanaugh CM. 2014.
650 The genome of the intracellular bacterium of the coastal bivalve, *Solemya*
651 *velum*: a blueprint for thriving in and out of symbiosis. BMC Genomics 15:924.
652 53. Weber MM, Chen C, Rowin K, Mertens K, Galvan G, Zhi H, Dealing CM, Roman
653 VA, Banga S, Tan Y, Luo ZQ, Samuel JE. 2013. Identification of *Coxiella burnetii*
654 type IV secretion substrates required for intracellular replication and *Coxiella*-
655 containing vacuole formation. J Bacteriol 195:3914-24.
656 54. Martinez E, Allombert J, Cantet F, Lakhani A, Yandrapalli N, Neyret A, Norville IH,
657 Favard C, Muriaux D, Bonazzi M. 2016. *Coxiella burnetii* effector CvpB modulates
658 phosphoinositide metabolism for optimal vacuole development. Proc Natl Acad
659 Sci U S A 113:E3260-9.
660 55. Kohler LJ, Reed Sh C, Sarraf SA, Arteaga DD, Newton HJ, Roy CR. 2016. Effector
661 Protein Cig2 Decreases Host Tolerance of Infection by Directing Constitutive
662 Fusion of Autophagosomes with the *Coxiella*-Containing Vacuole. MBio 7.
663 56. Larson CL, Beare PA, Voth DE, Howe D, Cockrell DC, Bastidas RJ, Valdivia RH,
664 Heinzen RA. 2015. *Coxiella burnetii* effector proteins that localize to the
665 parasitophorous vacuole membrane promote intracellular replication. Infect
666 Immun 83:661-70.
667 57. Fu M, Zhang J, Zhao M, Zhang S, Dai L, Ouyang X, Yu Y, Wen B, Zhou D, Sun Y,
668 Jiao J, Xiong X. 2022. *Coxiella burnetii* Plasmid Effector B Promotes LC3-II
669 Accumulation and Contributes To Bacterial Virulence in a SCID Mouse Model.
670 Infect Immun 90:e0001622.
671 58. Gomez-Valero L, Chiner-Oms A, Comas I, Buchrieser C. 2019. Evolutionary
672 Dissection of the Dot/Icm System Based on Comparative Genomics of 58
673 *Legionella* Species. Genome Biol Evol 11:2619-2632.
674 59. Martini H, Detemmerman L, Soetens O, Yusuf E, Piérard D. 2017. Improving
675 specificity of *Bordetella pertussis* detection using a four target real-time PCR.
676 PLoS One 12:e0175587.
677 60. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE, Rolando M,
678 Pasricha S, Reuter S, Demirtas J, Crumbach J, Descorps-Declere S, Hartland EL,
679 Jarraud S, Dougan G, Schroeder GN, Frankel G, Buchrieser C. 2019. More than

680 18,000 effectors in the *Legionella* genus genome provide multiple, independent
681 combinations for replication in human cells. Proc Natl Acad Sci U S A 116:2265-
682 2273.
683 61. Debowski AW, Bzdyl NM, Thomas DR, Scott NE, Jenkins CH, Iwasaki J, Kibble EA,
684 Khoo CA, Scheuplein NJ, Seibel PM, Lohr T, Metters G, Bond CS, Norville IH,
685 Stubbs KA, Harmer NJ, Holzgrabe U, Newton HJ, Sarkar-Tyson M. 2023.
686 Macrophage infectivity potentiator protein, a peptidyl prolyl cis-trans isomerase,
687 essential for *Coxiella burnetii* growth and pathogenesis. PLoS Pathog
688 19:e1011491.
689 62. Luo S, Lu S, Fan H, Sun Z, Hu Y, Li R, An X, Uversky VN, Chen Z, Tong Y, Song L.
690 2021. The *Coxiella burnetii* QpH1 plasmid is a virulence factor for colonizing
691 bone marrow-derived murine macrophages. J Bacteriol 203.
692 63. Wachter S, Cockrell DC, Miller HE, Virtaneva K, Kanakabandi K, Darwitz B,
693 Heinzen RA, Beare PA. 2022. The endogenous *Coxiella burnetii* plasmid encodes
694 a functional toxin-antitoxin system. Mol Microbiol 118:744-764.
695 64. Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Brinkman FSL. 2020.
696 Metagenome-assembled genome binning methods with short reads
697 disproportionately fail for plasmids and genomic Islands. Microb Genom 6.
698 65. Buytaers FE, Saltykova A, Denayer S, Verhaegen B, Vanneste K, Roosens NHC,
699 Piérard D, Marchal K, De Keersmaecker SCJ. 2021. Towards Real-Time and
700 Affordable Strain-Level Metagenomics-Based Foodborne Outbreak
701 Investigations Using Oxford Nanopore Sequencing Technologies. Front Microbiol
702 12:738284.
703 66. Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. 2008. Ankyrin repeat
704 proteins comprise a diverse family of bacterial type IV effectors. Science
705 320:1651-4.
706 67. Voth DE, Howe D, Beare PA, Vogel JP, Unsworth N, Samuel JE, Heinzen RA. 2009.
707 The *Coxiella burnetii* ankyrin repeat domain-containing protein family is
708 heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated
709 secretion. J Bacteriol 191:4232-42.
710 68. Larson CL, Pullman W, Beare PA, Heinzen RA. 2023. Identification of Type 4B
711 Secretion System Substrates That Are Conserved among *Coxiella burnetii*
712 Genomes and Promote Intracellular Growth. Microbiol Spectr 11:e0069623.
713 69. Burette M, Allombert J, Lambou K, Maarifi G, Nisole S, Di Russo Case E, Blanchet
714 FP, Hassen-Khodja C, Cabantous S, Samuel J, Martinez E, Bonazzi M. 2020.
715 Modulation of innate immune signaling by a *Coxiella burnetii* eukaryotic-like
716 effector protein. Proc Natl Acad Sci U S A 117:13708-13718.
717 70. Chen C, Banga S, Mertens K, Weber MM, Gorbaslieva I, Tan Y, Luo ZQ, Samuel
718 JE. 2010. Large-scale identification and translocation of type IV secretion
719 substrates by *Coxiella burnetii*. Proc Natl Acad Sci U S A 107:21755-60.
720 71. Crabill E, Schofield WB, Newton HJ, Goodman AL, Roy CR. 2018. Dot/Icm-
721 translocated proteins important for biogenesis of the *Coxiella burnetii*-
722 containing vacuole identified by screening of an effector mutant sub-library.
723 Infect Immun doi:10.1128/iai.00758-17.
724 72. Fielden LF, Scott NE, Palmer CS, Khoo CA, Newton HJ, Stojanovski D. 2020.
725 Proteomic identification of *Coxiella burnetii* effector proteins targeted to the

726 host cell mitochondria during infection. Mol Cell Proteomics
727 doi:10.1074/mcp.RA120.002370.
728 73. Zhang Y, Fu J, Liu S, Wang L, Qiu J, van Schaik EJ, Samuel JE, Song L, Luo ZQ.
729 2022. *Coxiella burnetii* inhibits host immunity by a protein phosphatase adapted
730 from glycolysis. Proc Natl Acad Sci U S A 119.
731 74. Martinez E, Huc-Brandt S, Brelle S, Allombert J, Cantet F, Gannoun-Zaki L,
732 Burette M, Martin M, Letourneur F, Bonazzi M, Molle V. 2020. The secreted
733 protein kinase CstK from *Coxiella burnetii* influences vacuole development and
734 interacts with the GTPase-activating host protein TBC1D5. J Biol Chem
735 295:7391-7403.
736
737

738 **Table 1. Candidatus *P. cheracis* proteins with similarity to *C. burnetii* Dot/Icm**
739 **substrates**
740

Candidatus P. *cheracis* gene	***C. burnetii* gene**	**Identity (%)**	**Coverage (%)**	**Role of *C. burnetii* effector**
PBJCNB_00280	CBUD_0487	26	22	Experimentally shown to be a Dot/Icm substrate (56).
PBJCNB_01005	CBU_1863 (CvpE)	40	78	*Coxiella* vacuolar protein E localizes to the *Coxiella*-containing vacuole (CCV). Characterization of a deletion mutant demonstrated that CvpE contributes to intracellular replication of *C. burnetii* (47, 56).
PBJCNB_01155	CBU_0201 (AnkC)	84	95	Early studies using *L. pneumophila* as a surrogate system did not demonstrate translocation of AnkC (53, 66, 67). Similarly, no translocation was observed in *C. burnetii* using the CyaA reporter assay (68). However, using a BlaM reporter assay in *C. burnetii* low level of translocation were observed at a later time point, 72 h post-infection (69).
PBJCNB_01790	CBUA_0013 (CpeB)	29	90	Plasmid encoded *C. burnetii* effector that contributes to LC3-II accumulation in a Rab11-dependent manner (57).
PBJCNB_05220	CBU_0937 (CirC/MceB)	57	100	Identified as a DotF binding protein and demonstrated to be translocated by *L. pneumophila* (70). Two independent *cbu0937* transposon mutants

				were reported to have an intracellular replication defect (53, 71). Detected in mitochondria isolated from *C. burnetii* infected cells (72). However, recent translocation reporter assays conducted in *C. burnetii* could not confirm translocation of this protein (68).
PBJCNB_05510	CBU_1079	76	97	5% translocation observed using the BlaM reporter in *L. pneumophila* (53). Not validated in *C. burnetii* using the CyaA reporter (68).
PBJCNB_08410	CBU_0513 (CinF)	83	99	Shown to be translocated as a BlaM-fusion protein by *L. pneumophila* with Dot/Icm-dependent translocation confirmed in *C. burnetii* (53, 73).Loss of CinF, either through transposon disruption of silencing expression, inhibits intracellular replication of *C. burnetii* (53, 73). CinF is a protein phosphatase that dephosphorylates IκBα (73).
PBJCNB_09120	CBU_1425 (MceC)	74	100	Identified as a DotF binding protein and demonstrated to be translocated by *L. pneumophila* (70). Identified as being enriched in the mitochondria of *C. burnetii* infected cells with the capacity to

				localise to the inner mitochondrial membrane and interact with components of the mitochondrial quality control machinery (72). Translocation by C. burnetii not confirmed using the BlaM reporter assay (68).
PBJCNB_09430	CBU_0469	61	95	Low level (1%) translocation of BlaM fusion protein in L. pneumophila (53). Translocation has not been confirmed in *C. burnetii*.
PBJCNB_10235	CBU_1566	79	99	BlaM fusion protein was shown to be translocated by *L. pneumophila* (53) but no translocation was observed for a Cya fusion in *C. burnetii* (68).
PBJCNB_10940	CBU_0175 (CoxK1)	82	100	BlaM fusion protein translocated by *C. burnetii* in a Dot/Icm-dependent manner (74). Shown to interact with the Rab7 GTPase-activating protein TBC1D5 (74).
PBJCNB_10315	CBU_1594 (MceD)	71	97	BlaM fusion protein was shown to be translocated by *L. pneumophila* (53) but no translocation was observed for a Cya fusion in *C. burnetii* (68). Identified as being enriched in the mitochondria of *C. burnetii* infected cells (72).

741

FIGURE LEGENDS

Figure 1. Phylogeny of Coxiellales Order demonstrating relationship to Candidatus *P. cheracis* genome

The inferred evolution of the Candidatus *P. cheracis* genome in representative isolates from Coxiellales Order (gtdb.ecogenomic.org/tree?r=o__Coxiellales). The phylogeny was inferred using GTDB-Tk, a marker gene-based method of 120 domain-specific marker genes for bacteria (33). The tips of the tree are coloured by the NCBI organism for the Coxiellales lineages, and additional GTDB lineages at the genus level. The associated hosts or source from where the samples were collected are shown to the right of the tree. The Candidatus *P. cheracis* genome is indicated in green and by the crayfish icon (sourced from www.phylopic.org/).

Figure 2. Characterization of the Candidatus *P. cheracis* genome

The complete genome showing A) the chromosome and B) two plasmids. The middle ring shows the coding sequence (CDS) on the chromosome and plasmids. Genes of interest are shown in the outer ring with colours indicating broad function (T4SS, multicopy transposase (TS), phage defense and putative effector). The total number of these four functional classes on each genetic element are shown on panel C.

Figure 3. Structure and protein identity of the T4SS in the Candidatus *P. cheracis* genome

A) Characterization of the T4SS identified in Candidatus *P. cheracis, Coxiella burnetii*, *Legionella pneumophila, Legionella longbeachae* and Candidatus *Rickettsiella viridis.* Genes encoding the T4SS are colored by function within the system. B) Visualization of Dot/Icm protein identity relative to *Coxiella burnetii* RSA 493.

Figure 4. Candidatus *P. cheracis Pc*CpeB is a Dot/Icm substrate.

HeLa cells were infected for 24 h with either wildtype (NMII) or Dot/Icm-deficient (*icmL*::Tn) *C. burnetii* expressing the indicated BlaM fusion proteins. **(A)** Change in the fluorescence emission ratio 450nm::520nm was measured as a readout of BlaM cleaving the CCF2-AM substrate (y-axis). The graph depicts this ratio from 4

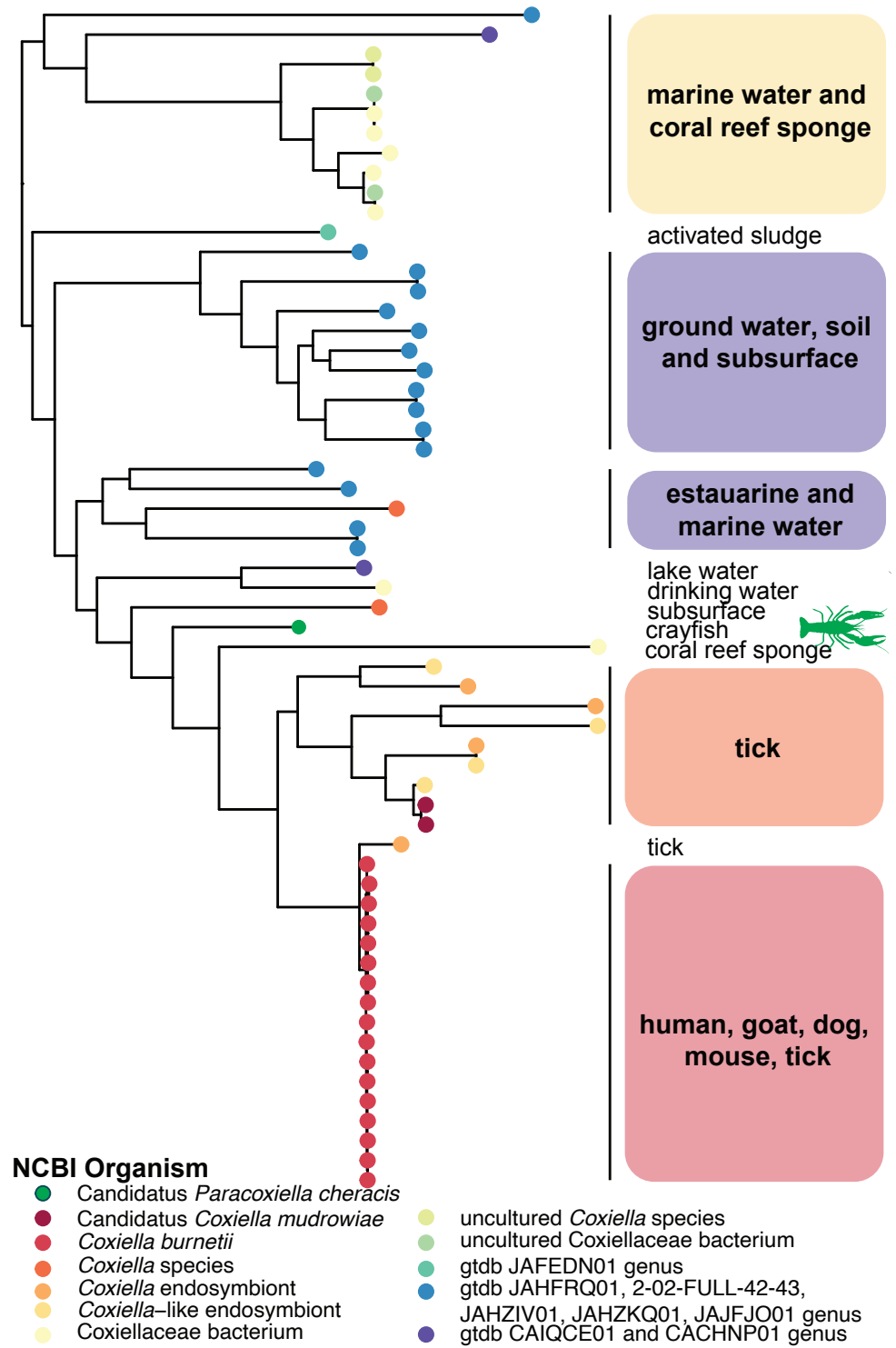
independent experiments (circles) with the bar representing the mean and SEM ratio. The threshold for translocation positive was set to 2.5× the BlaM alone negative control (dotted line) and Šidák's multiple comparisons test was used to determine P-values compared to the BlaM and between NMII and *icmL*::Tn strains expressing the same fusion proteins. **(B)** Representative fluorescent images of infected HeLa cells demonstrate the translocation of CvpB (*C. burnetii* effector positive control) and *Pc*CpeB by *C. burnetii* NMII but not *C. burnetii icmL*::Tn. Fluorescence intensity at 520 nm of uncleaved CCF2-AM is shown in green and cleaved CCF2-AM, at 450 nm is blue, scale bars represent 100 µm. **(C)** Immunoblot analysis of the indicated *C. burnetii* strains confirmed expression of BlaM (29 kDa), BlaM-CvpB (117.6 kDa) and BlaM-*Pc*CpeB (54.5 kDa). **(D)** ClustalOmega alignment showing the similarity between *C. burnetii* CpeB (*Cb*CpeB) and Candidatus *P. cheracis* CpeB (*Pc*CpeB). The two proteins share 26% identity (*) with additional conserved (:) and semi-conserved (.) substitutions. The purple boxed area represents amino acids 164 to 204 that were previously shown to mediate interaction of *Cb*CpeB with Rab11a (57).

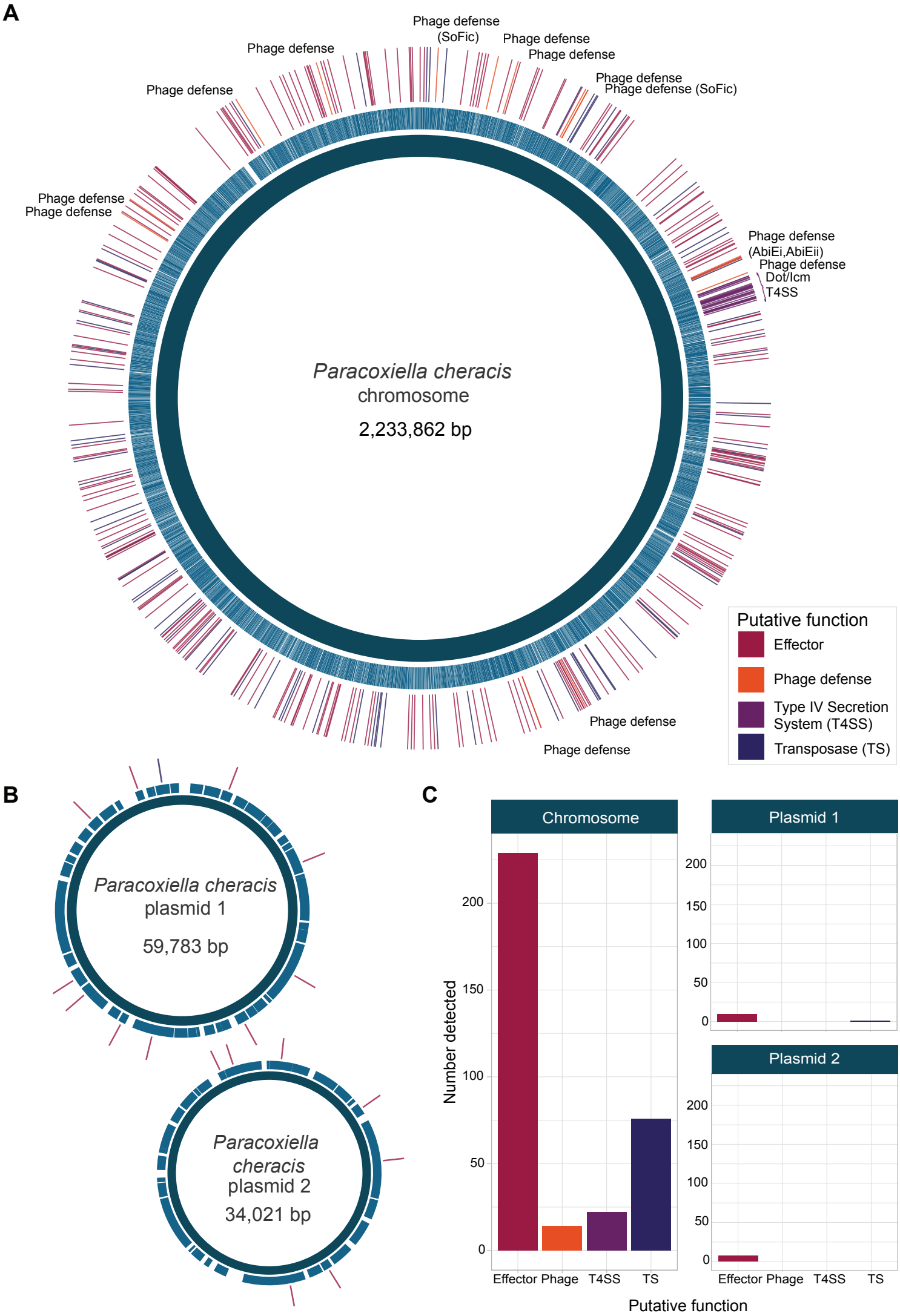
SUPPLEMENTARY FILES

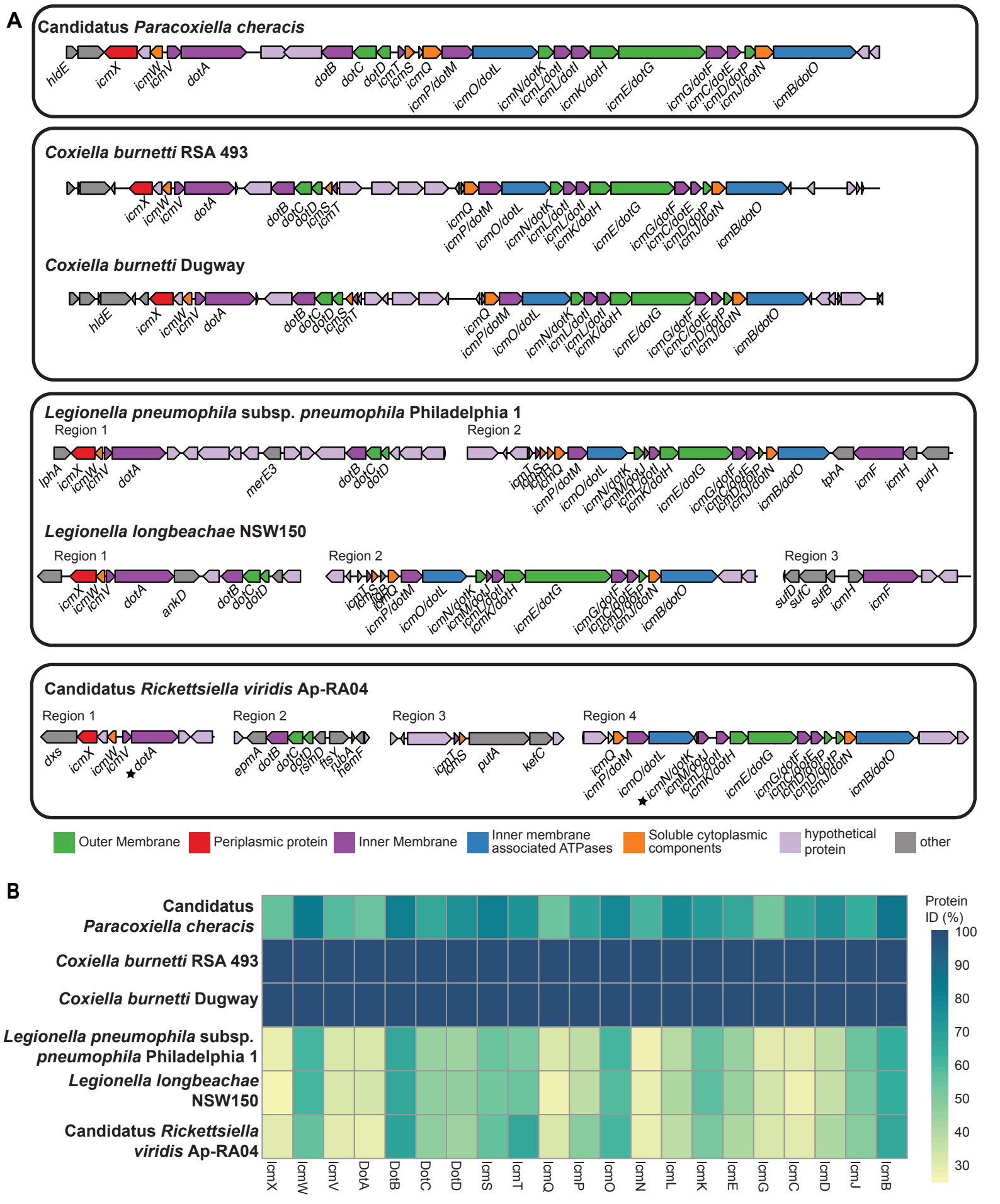
Supplementary File 1. Workflow for quality controls, kraken assignment and genome assembly.

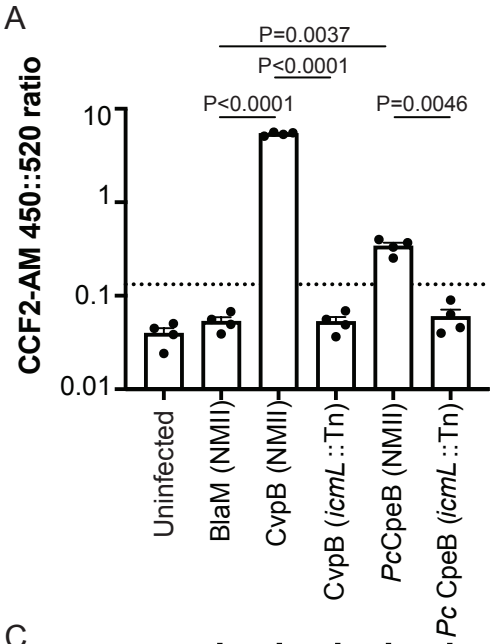
Supplementary File 2. The Kraken2 classification of the chromosomal reads using a GTDB-based database.

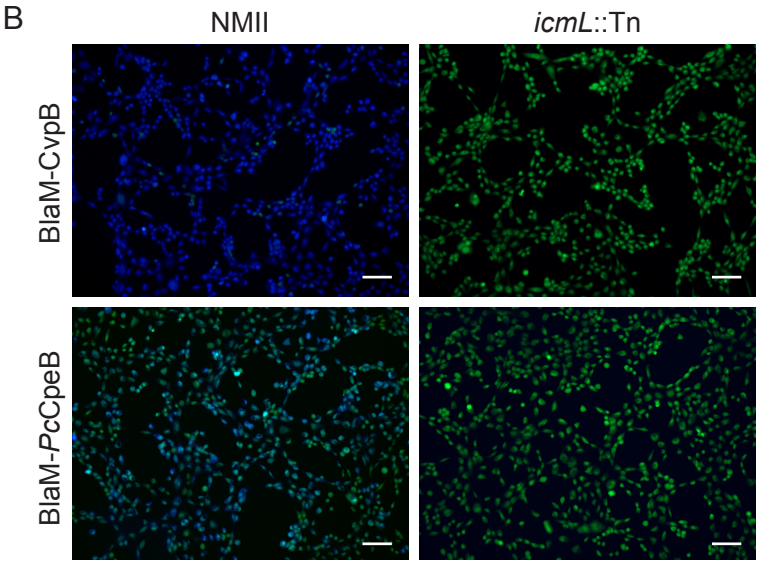
Supplementary File 3. Predicted Dot/Icm effector proteins based on Bastion4 and T4Sepp tools.

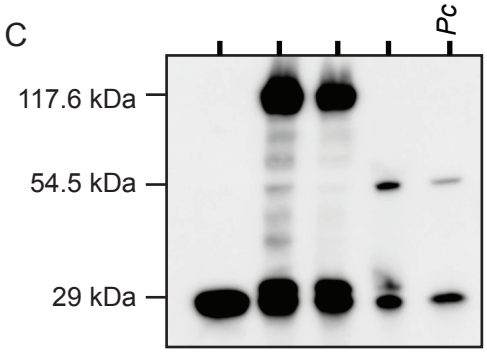

marine water and coral reef sponge
activated sludge
ground water, soil and subsurface
estauarine and marine water
lake water
drinking water
subsurface
crayfish
coral reef sponge
tick
tick
human, goat, dog, mouse, tick
NCBI Organism
Candidatus Paracoxiella cheracis
Candidatus Coxiella mudrowiae
Coxiella burnetii
Coxiella species
Coxiella endosymbiont
Coxiella–like endosymbiont
Coxiellaceae bacterium
uncultured Coxiella species
uncultured Coxiellaceae bacterium
gtdb JAFEDN01 genus
gtdb JAHFRQ01, 2-02-FULL-42-43, JAHZIV01, JAHZKQ01, JAJFJO01 genus
gtdb CAIQCE01 and CACHNP01 genus


A
Phage defense
(SoFic)
Phage defense
Phage defense
Phage defense
Phage defense
Phage defense (SoFic)
Phage defense
Phage defense
Phage defense
(AbiEi,AbiEii)
Phage defense
Dot/Icm
T4SS
Paracoxiella cheracis
chromosome
2,233,862 bp
Phage defense
Phage defense
Putative function
Effector
Phage defense
Type IV Secretion System (T4SS)
Transposase (TS)
B
Paracoxiella cheracis
plasmid 1
59,783 bp
Paracoxiella cheracis
plasmid 2
34,021 bp
C
Chromosome
Plasmid 1
Plasmid 2
Number detected
0
50
100
150
200
Effector
Phage
T4SS
TS
Putative function


A
Candidatus Paracoxiella cheracis
hldE icmX icmW icmV dotA dotB dotC dotD icmT icmS icmQ icmP/dotM icmO/dotL icmN/dotK icmL/dotI icmL/dotI icmK/dotH icmE/dotG icmG/dotF icmC/dotE icmD/dotP icmJ/dotN icmB/dotO
Coxiella burnetti RSA 493
icmX icmW icmV dotA dotB dotC dotD icmS icmT icmQ icmP/dotM icmO/dotL icmN/dotK icmL/dotI icmL/dotI icmK/dotH icmE/dotG icmG/dotF icmC/dotE icmD/dotP icmJ/dotN icmB/dotO
Coxiella burnetti Dugway
hldE icmX icmW icmV dotA dotB dotC dotD icmS icmT icmQ icmP/dotM icmO/dotL icmN/dotK icmL/dotI icmL/dotI icmK/dotH icmE/dotG icmG/dotF icmC/dotE icmD/dotP icmJ/dotN icmB/dotO
Legionella pneumophila subsp. pneumophila Philadelphia 1
Region 1
lphA icmX icmW icmV dotA merE3 dotB dotC dotD
Region 2
icmT icmS icmR icmQ icmP/dotM icmO/dotL icmN/dotK icmM/dotJ icmL/dotI icmK/dotH icmE/dotG icmG/dotF icmC/dotE icmD/dotP icmJ/dotN icmB/dotO tphA icmF icmH purH
Legionella longbeachae NSW150
Region 1
icmX icmW icmV dotA ankD dotB dotC dotD
Region 2
icmT icmS icmR icmQ icmP/dotM icmO/dotL icmN/dotK icmM/dotJ icmL/dotI icmK/dotH icmE/dotG icmG/dotF icmC/dotE icmD/dotP icmJ/dotN icmB/dotO
Region 3
sufD sufC sufB icmH icmF
Candidatus Rickettsiella viridis Ap-RA04
Region 1
dxs icmX icmW icmV ★dotA
Region 2
epmA dotB dotC dotD rsmD ftsY rubA hemF
Region 3
icmT icmS putA kefC
Region 4
icmQ icmP/dotM icmO/dotL ★icmN/dotK icmM/dotJ icmL/dotI icmK/dotH icmE/dotG icmG/dotF icmC/dotE icmD/dotP icmJ/dotN icmB/dotO
Outer Membrane
Periplasmic protein
Inner Membrane
Inner membrane associated ATPases
Soluble cytoplasmic components
hypothetical protein
other
B
Candidatus Paracoxiella cheracis
Coxiella burnetti RSA 493
Coxiella burnetti Dugway
Legionella pneumophila subsp. pneumophila Philadelphia 1
Legionella longbeachae NSW150
Candidatus Rickettsiella viridis Ap-RA04
IcmX IcmW IcmV DotA DotB DotC DotD IcmS IcmT IcmQ IcmP IcmO IcmN IcmL IcmK IcmE IcmG IcmC IcmD IcmJ IcmB
Protein ID (%)
100
90
80
70
60
50
40
30







D

```
CbCpeB MPYFFTLPKIDIKE--------KSASQTPEKVLREWLRGASNQEDASLFIGNSPIF----
PcCpeB MSAFFTVPKPQYGGESELLEREKKDDTSGKITLKDWVRNASNACEARVFVMSGTSPNTLT 60
       *  ***:** :           *. . : : .*::*:*.***  :* :*: ..

       ---------GEHQDTSGSSLPKVEIPEFINSIKAVEAVFKKEGLKDEEILNYILHFGGIN 99
       TAINESDLNAYMRDDDGE-----RIPLLSDADQYIKQLLEKHGLENHRIYKYIRMFANIN 115
                .  :* .*.     .** : :: : :: :::*.**::..* :**  *..**

       GYMTLGQVVFEEISMNKG-KSSRMENRRYSYRVKNKGEVEYIEQFDFYKKNTCDRNEIGD 158
       GFLPAGQIVMSQVKNSGEQSPIHTENPKYTFFVKDKTTVQYTEESDIVLNPTH-EYGKGE 174
       *::  **:*:.::. .   .  : ** :*:: **:*  *:* *: *:  : *  .   *:

       FIGQVKIVSIISKDGD-IQHACKKVEI-AVPGEKLNSFRKHFGDQLNFVNRMKLRLVELY 216
       TIAKLKVVSIISSNDQKIHHILDSASLTAYPGK-EPQCIKCFGDNRSVMVKLIHKLKEMF 233
        *.::*:*****.:.: *:*  ....: * **:    .  * ***: ..: ::  :* *::

       NSLVERFRAAINKADRTN-----NNCFPFCSR 243
       MAFVALFTKRSSCEQDYNGFNDVASNAASCRR 265
        ::*  *    .  :  *      .    * *
```