

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Ingle, DJ;Walsh, CJ;Samuel, GR;Wick, RR;Davidovich, N;Fiocchi, E;Judd, LM;Elliman, J;Owens, L;Stinear, TP;Basso, A;Pretto, T;Newton, HJ

Title:

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

Date:

2025-04-01

Citation:

Ingle, D. J., Walsh, C. J., Samuel, G. R., Wick, R. R., Davidovich, N., Fiocchi, E., Judd, L. M., Elliman, J., Owens, L., Stinear, T. P., Basso, A., Pretto, T. & Newton, H. J. (2025). 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. *mSphere*, 10 (4), <https://doi.org/10.1128/msphere.01002-24>.

Persistent Link:

<https://hdl.handle.net/11343/360158>

License:

CC BY

# 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,<sup>1</sup> Calum J. Walsh,<sup>1,2</sup> Genevieve R. Samuel,<sup>3,4</sup> Ryan R. Wick,<sup>1,2</sup> Nadav Davidovich,<sup>5</sup> Eleonora Fiocchi,<sup>6</sup> Louise M. Judd,<sup>2</sup> Jennifer Elliman,<sup>7</sup> Leigh Owens,<sup>7</sup> Timothy P. Stinear,<sup>1,2</sup> Andrea Basso,<sup>6</sup> Tobia Pretto,<sup>6</sup> Hayley J. Newton<sup>3,4</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 13.

**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 species 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 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 homologs of several important *C. burnetii* effectors, including a homolog 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 insights 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.

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

Australian native freshwater redclaw crayfish, *Cherax quadricarinatus*, are bred for the commercial market. Aquaculture of these animals has led to the observation of infectious diseases that impact both wild and farmed *C. quadricarinatus* (reviewed in reference 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

**Editor** Sarah E. F. D'Orazio, University of Kentucky College of Medicine, Lexington, Kentucky, USA

Address correspondence to Hayley J. Newton, hayley.newton@monash.edu.

Danielle J. Ingle and Calum J. Walsh contributed equally to this article. Author order was determined based on alphabetical order of last name.

The authors declare no conflict of interest.

See the funding table on p. 14.

**Received** 27 November 2024

**Accepted** 5 February 2025

**Published** 10 March 2025

Copyright © 2025 Ingle et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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 do 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 reference 9). Most of these newly discovered Legionellales are likely nonpathogenic bacteria that have evolved mutualistic relationships with non-vertebrate hosts. However, recent data suggest 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 reference 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 reference 16). In contrast, internalized *C. burnetii* is trafficked through the endocytic pathway before the lysosomal environment triggers the 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 CL-Es, 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 is 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 Candidatus *C. cheraxi* represents a novel genus within Coxiellaceae. We propose to rename this crayfish pathogen Candidatus *Paracoxiella cheraxis*. 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 with known *C. burnetii* T4SS effectors. We have been able to demonstrate the translocation of the Candidatus *P. cheraxis* CpeB homolog 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.

## MATERIALS AND 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 run employed a PromethION. Both runs were carried out on R10.4.1 flow cells. Basecalling and quality filtering were performed by Dorado (version 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 1 kbp were discarded. Kraken2 (21) was used for the 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 them 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 using Minimap2 (26), and the mapped reads were retained for downstream analysis. The chromosomal and plasmid reads were then downsampled to 500× coverage using Filtlong ([github.com/rrwick/Filtlong](https://github.com/rrwick/Filtlong)). The shortest mapped read was 6,961 bp, and the mapped read  $N_{50}$  was 10,724 bp. Mapped reads were combined into a single high-quality assembly with Tricycler (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 version 1.9.1 (31). Reads that mapped to the assembled chromosome were also classified with Kraken2. Full details of the individual steps for quality controls, kraken assignment, and genome assembly are available in File S1.

### 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](https://gtdb.ecogenomic.org/tree?r=o__Coxiellales)) were used in comparative analyses to infer the relative location of Candidatus *P. cheraxi* 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 data set was first dereplicated by Assembly Dereplicator ([github.com/rrwick/Assembly-Dereplicator](https://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. GTDB-Tk (33), a marker gene-based tool for taxonomic classification and phylogenetic placement of prokaryotic genomes based on the Genome Taxonomy Database taxonomy system (34), was used to infer a phylogeny for this genome and the representative 60 genomes and to taxonomically classify our long-read assembly. Briefly, GTDB-Tk creates a protein multiple sequence alignment of bacterial marker genes in the genomes and uses

FastTree to infer a phylogeny with the WAG-GAMA model as default. The final protein alignment was 5,035 amino acids in length.

### 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 (version 1.0.14) 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 version 2 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.

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* strain 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 (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 similarity. For these three genomes, the results were filtered to the loci that encode the T4SS as identified by coordinates from SECRET4, and the hits with the highest bitscore for each protein were 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<sub>2</sub> and 2.5% O<sub>2</sub> as previously described (46). When required, chloramphenicol was added to *C. burnetii* cultures at 3 µg/mL. HeLa CCL2 cells (ATCC, VA, USA) were maintained in Dulbecco's modified Eagle's media supplemented with 10% heat-inactivated fetal bovine serum at 37°C in 5% CO<sub>2</sub>.

### Construction of *C. burnetii* translocation reporter strains

The open reading frame encoding for the Candidatus *P. cheracis* CpeB homolog (*PcCpeB*) was synthesized by IDT and cloned into pJB-CAT:Blam (47) using BamHI and XhoI restriction enzymes. The resulting pBlam-*PcCpeB* was introduced into *C. burnetii* strains using a standard electroporation protocol (11). Following the selection of transformants, the expression of Blam fusion proteins was confirmed using an anti-Blam (1:1,500, 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.

## RESULTS

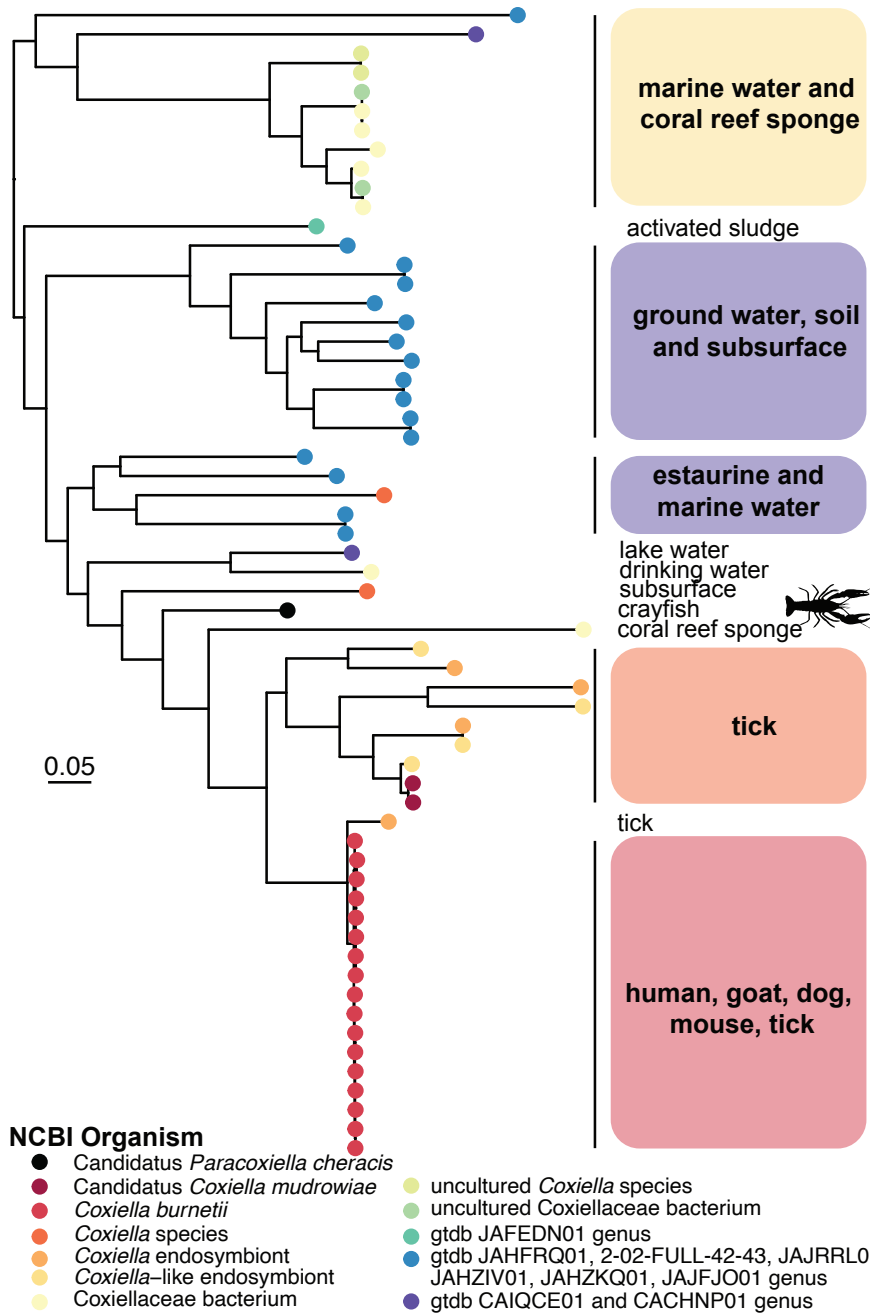
### A putative new genus in the Coxiellaceae family

Here, we describe the complete genome of a putative new genus within Coxiellaceae, Candidatus *C. cheraxi*, which 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 1.9–2.2 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 identity of 89%. This high level of similarity, even with the TO-98 data set produced using older minION chemistry (R9.4.1), provides confidence that this genome is representative of the same species as TO-98.

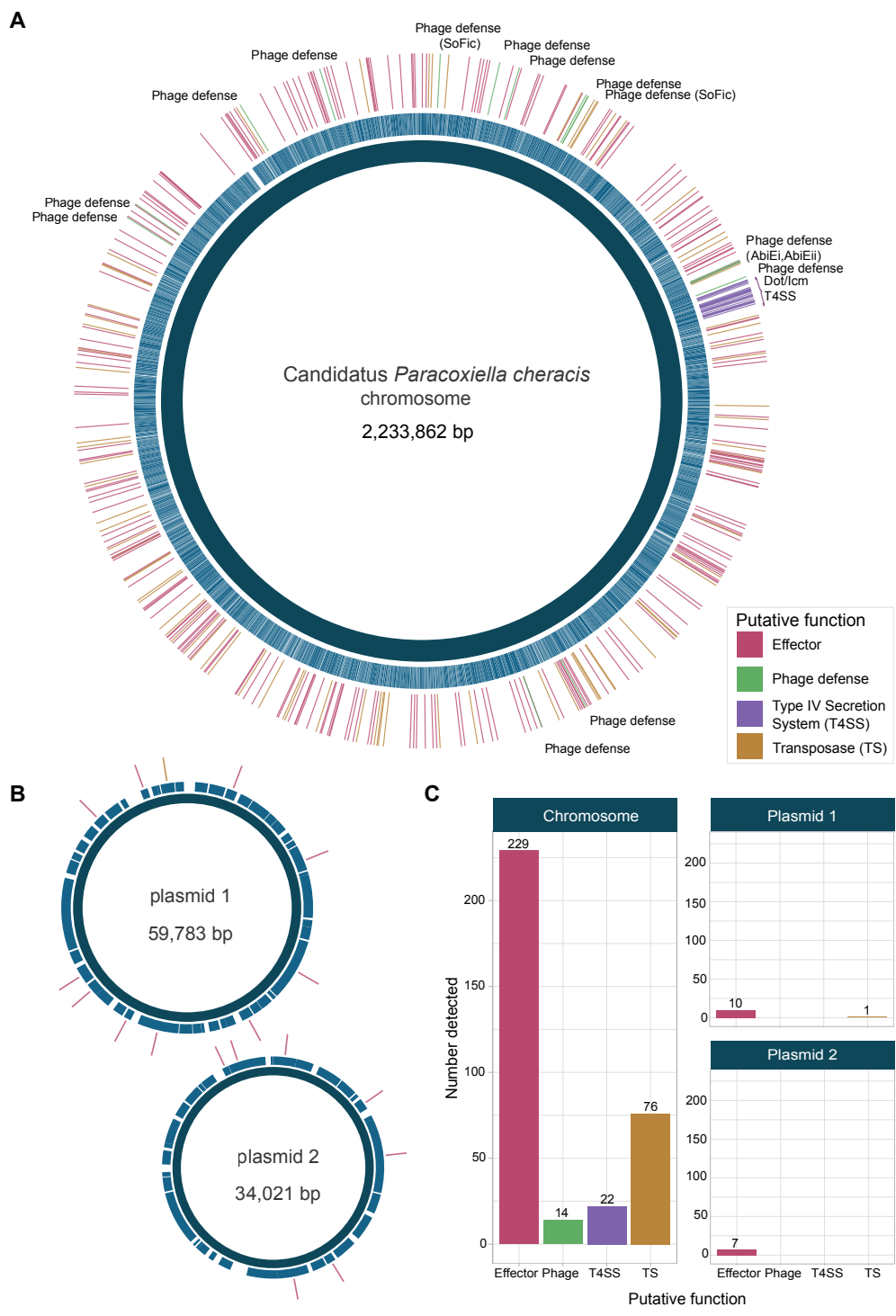
The Candidatus *C. cheraxi* genome was classified as Coxiellaceae by GTDB-Tk (32), meaning that the novel genome could be classified to the family level but no further, suggesting that this genome represents a novel genus, supported by a relative evolutionary divergence value of 0.79373 (32). Based on these 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 the genomes of other Coxiellaceae bacteria (Fig. 1) isolated from similar aquatic ecological niches and distant to *C. burnetii* (largely associated with mammalian hosts) and the C-LEs that are associated with ticks as hosts (Fig. 1). Interestingly, Kraken2 classification of the chromosomal reads using a GTDB-based database identified multiple bacterial families in the read set (Files S1 and S2). 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 (File S2), 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, the 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 (Fig. 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. Furthermore, these two plasmid sequences had no similarity 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, comprising AbiEi and AbiEii components, were identified in the genome. The remaining 10 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 (accession: [JRAA00000000](https://www.ncbi.nlm.nih.gov/nuccore/JRAA00000000)) (52).



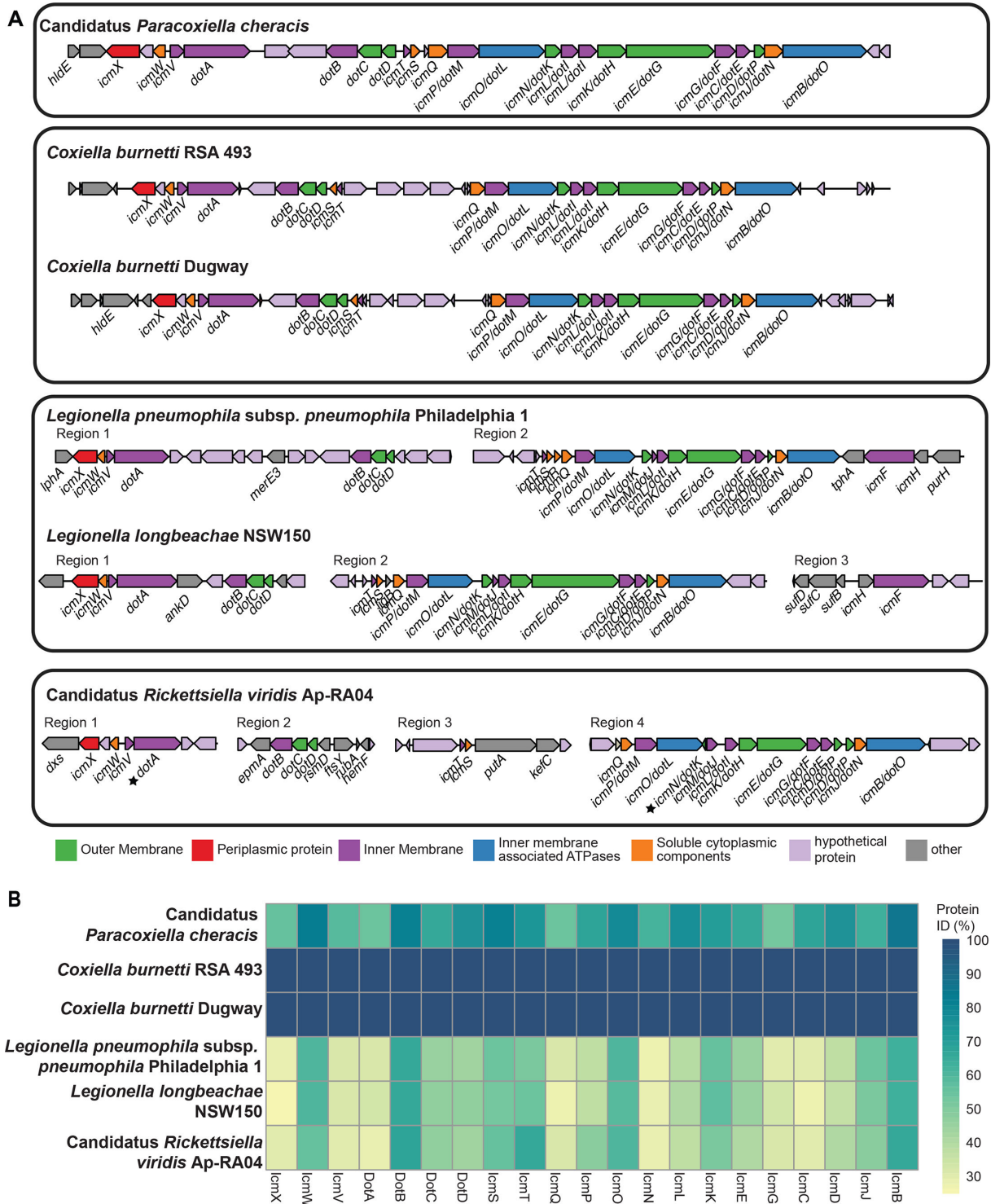
**FIG 1** Phylogeny of Coxiellales order demonstrating relationship to *Candidatus P. cheracis* genome. The inferred evolution of the *Candidatus P. cheracis* genome in representative isolates from Coxiellales order ([gtdb.ecogenomic.org/tree?r = o\\_Coxiellales](https://gtdb.ecogenomic.org/tree?r=o_Coxiellales)). 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 colored by the NCBI organism for the Coxiellales lineages and additional GTDB lineages at the genus level. The branch lengths represent the expected substitutions per site of the protein alignment of 5,035 amino acids. The associated hosts or sources from where the samples were collected are shown to the right of the tree. The *Candidatus P. cheracis* genome is indicated in black and by the crayfish icon (sourced from <https://www.phylopic.org/nodes/aa251b9d-1cb9-4327-929a-d56ef1bc8f82/crayfishes-silhouettes>).



**FIG 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 on the chromosome and plasmids. Genes of interest are shown in the outer ring with colors indicating broad function (T4SS, multicopy transposase (TS), phage defense, and putative effector). The total number of these four functional classes on each genetic element is shown on panel C.

### 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



**FIG 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.

arrangement were similar to the two isolates of *C. burnetii* examined here (Fig. 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 similarity between 21 genes of the Dot/Icm system differed within the six genomes by species and genus (Fig. 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 of the Dot/Icm proteins in Candidatus *P. cheracis* varied between 52% (IcmG) and 86% (IcmB) 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*) and 67% (DotB for Candidatus *R. viridis*). As previously reported, the Dot/Icm system was not detected in the three C-LE 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, a 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 (File S3). Bastion4 predicted 117 instances of IS elements as putative effectors. IS elements have not previously been reported as T4SS substrates; however, experimental evidence is required to determine whether these are false-positive results. Excluding IS elements, this approach identified a pool of 238 potential Dot/Icm effectors. 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, PcCpeB, can be translocated by the *C. burnetii* Dot/Icm system

Each of the predicted Dot/Icm effectors was 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  $\beta$ -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* homolog of *C. burnetii* effector CpeB (PcCpeB) was confirmed (Fig. 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, 65, 66) or BlaM-PcCpeB (Fig. 4C). HeLa cells were infected with these *C. burnetii* strains for 22 h before the addition of CCF2-AM and subsequent detection of fluorescence emission via both a fluorescence plate reader (Fig. 4A) and microscopy (Fig. 4B). BlaM-PcCpeB was translocated in a Dot/Icm-dependent manner. Interestingly, a comparison of the amino acid sequences of *C. burnetii*

TABLE 1 Candidatus *P. cheracis* proteins with similarity to *C. burnetii* Dot/Icm substrates

Candidatus <i>P. cheracis</i> gene	<i>C. burnetii</i> gene	Identity (%)	Coverage (%)	Role of <i>C. burnetii</i> effector
PBJCNB_00280	CBUD_0487	26	22	Experimentally shown to be a Dot/Icm substrate (54).
PBJCNB_01005	CBU_1863 (CvpE)	40	78	<i>Coxiella</i> vacuolar protein E localizes to the <i>Coxiella</i> -containing vacuole. Characterization of a deletion mutant demonstrated that CvpE contributes to intracellular replication of <i>C. burnetii</i> (47, 54).
PBJCNB_01155	CBU_0201 (AnkC)	84	95	Early studies using <i>L. pneumophila</i> as a surrogate system did not demonstrate translocation of AnkC (53, 55, 56). Similarly, no translocation was observed in <i>C. burnetii</i> using the CyaA reporter assay (57). However, using a BlaM reporter assay in <i>C. burnetii</i> , low level of translocation was observed at a later time point, 72 h post-infection (58).
PBJCNB_01790	CBUA_0013 (CpeB)	29	90	Plasmid-encoded <i>C. burnetii</i> effector that contributes to LC3-II accumulation in a Rab11-dependent manner (59).
PBJCNB_05220	CBU_0937 (CirC/MceB)	57	100	Identified as a DotF binding protein and demonstrated to be translocated by <i>L. pneumophila</i> (60). Two independent <i>cbu0937</i> transposon mutants were reported to have an intracellular replication defect (53, 61). Detected in mitochondria isolated from <i>C. burnetii</i> -infected cells (62). However, recent translocation reporter assays conducted in <i>C. burnetii</i> could not confirm translocation of this protein (57).
PBJCNB_05510	CBU_1079	76	97	5% translocation observed using the BlaM reporter in <i>L. pneumophila</i> (53). Not validated in <i>C. burnetii</i> using the CyaA reporter (57).
PBJCNB_08410	CBU_0513 (CinF)	83	99	Shown to be translocated as a BlaM-fusion protein by <i>L. pneumophila</i> with Dot/Icm-dependent translocation confirmed in <i>C. burnetii</i> (53, 63). Loss of CinF, through transposon disruption of silencing expression, inhibits intracellular replication of <i>C. burnetii</i> (53, 63). CinF is a protein phosphatase that dephosphorylates IκBα (63).
PBJCNB_09120	CBU_1425 (MceC)	74	100	Identified as a DotF binding protein and demonstrated to be translocated by <i>L. pneumophila</i> (60). Identified as being enriched in the mitochondria of <i>C. burnetii</i> -infected cells with the capacity to localize to the inner mitochondrial membrane and interact with components of the mitochondrial quality control machinery (62). Translocation by <i>C. burnetii</i> was not confirmed using the BlaM reporter assay (57).
PBJCNB_09430	CBU_0469	61	95	Low-level (1%) translocation of BlaM fusion protein in <i>L. pneumophila</i> (53). Translocation has not been confirmed in <i>C. burnetii</i> .
PBJCNB_10235	CBU_1566	79	99	BlaM fusion protein was shown to be translocated by <i>L. pneumophila</i> (53), but no translocation was observed for a Cya fusion in <i>C. burnetii</i> (57).
PBJCNB_10940	CBU_0175 (CoxK1)	82	100	BlaM fusion protein translocated by <i>C. burnetii</i> in a Dot/Icm-dependent manner (64). Shown to interact with the Rab7 GTPase-activating protein TBC1D5 (64).
PBJCNB_10315	CBU_1594 (MceD)	71	97	BlaM fusion protein was shown to be translocated by <i>L. pneumophila</i> (53), but no translocation was observed for a Cya fusion in <i>C. burnetii</i> (57). Identified as being enriched in the mitochondria of <i>C. burnetii</i> -infected cells (62).



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-LEs. 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 the adaptation of the effectors to the different host niches. The observed differences in Dot/Icm protein identity between these genomes are greater than the average values recently reported for 58 *Legionella* species (67). 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 the exposure of these proteins to hosts (67). In contrast, IcmB and DotB had the highest protein identity, which is consistent with previous work in *Legionella* species suggesting that these proteins are under functional selection (67).

Key challenges in the detection of obligate intracellular pathogens are 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 high sequence similarity 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 the diagnostic efforts of members of the *Bordetella* genus (68). *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 the 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 (69). 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 (69). Interestingly, only 8 Dot/Icm substrates are conserved among these 65 species of *Legionella* (69). 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, 59, 65). 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 (PcCpeB) 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* (70). Candidatus *P. cheracis* Mip shares 66% identity with *C. burnetii* Mip, and these proteins are conserved within the PPlase catalytic domain, which suggests that Mip inhibitors developed by Debowski et al. (70) may inhibit Candidatus *P. cheracis* Mip. 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 (71, 72) and for dealing with the significant number of IS elements throughout this genome (73). The detection of a T4SS and cohort of novel effector proteins that likely facilitate the 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 (74). If such an outbreak were to occur today, in addition to traditional histopathology, ONT metagenomics would enhance efforts to detect and characterize unknown emerging pathogens.

## ACKNOWLEDGMENTS

D.J.I. is supported by a National Health and Medical Research Council (NHMRC) Investigator Grant (GNT1195210). C.J.W. and R.R.W. are funded through the NHMRC (GNT1194325). G.R.S. is the recipient of an Australian Government Research Training Program Scholarship and Monash Graduate Excellence Scholarship. Research in H.J.N.'s laboratory was funded by the NHMRC (GNT2010841).

The funders had no role in study design, data collection and analysis, and the decision to publish or preparation of the manuscript.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia

<sup>2</sup>Centre for Pathogen Genomics, University of Melbourne, Parkville, Victoria, Australia

<sup>3</sup>Infection Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

<sup>4</sup>Department of Microbiology, Monash University, Clayton, Victoria, Australia

<sup>5</sup>Israeli Veterinary Services, Bet Dagan, Israel

<sup>6</sup>Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy

<sup>7</sup>College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia

## AUTHOR ORCID*s*

Danielle J. Ingle  <http://orcid.org/0000-0003-0707-6537>

Calum J. Walsh  <http://orcid.org/0000-0003-3113-9534>

Genevieve R. Samuel  <http://orcid.org/0009-0006-4402-1956>

Ryan R. Wick  <http://orcid.org/0000-0001-8349-0778>

Nadav Davidovich  <http://orcid.org/0000-0002-4467-6083>

Eleonora Fiocchi  <http://orcid.org/0000-0003-4601-9543>  
 Louise M. Judd  <http://orcid.org/0000-0003-3613-4839>  
 Jennifer Elliman  <http://orcid.org/0000-0001-6460-1604>  
 Timothy P. Stinear  <http://orcid.org/0000-0003-0150-123X>  
 Andrea Basso  <http://orcid.org/0000-0001-6784-7271>  
 Tobia Pretto  <http://orcid.org/0000-0001-7699-9728>  
 Hayley J. Newton  <http://orcid.org/0000-0002-9240-2001>

## FUNDING

Funder	Grant(s)	Author(s)
National Health and Medical Research Council	GNT1195210	Danielle J. Ingle
National Health and Medical Research Council	GNT1194325	Calum J. Walsh Ryan R. Wick
National Health and Medical Research Council	GNT2010841	Hayley J. Newton

## AUTHOR CONTRIBUTIONS

Danielle J. Ingle, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Calum J. Walsh, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Genevieve R. Samuel, Data curation, Investigation, Writing – review and editing | Ryan R. Wick, Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing | Nadav Davidovich, Investigation, Writing – review and editing | Eleonora Fiocchi, Investigation, Writing – review and editing | Louise M. Judd, Data curation, Investigation, Methodology, Writing – review and editing | Jennifer Elliman, Investigation, Writing – review and editing | Leigh Owens, Investigation, Writing – review and editing | Timothy P. Stinear, Funding acquisition, Writing – review and editing | Andrea Basso, Investigation, Writing – review and editing | Tobia Pretto, Investigation, Writing – review and editing | Hayley J. Newton, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing

## DATA AVAILABILITY

The complete genome and the long-read data are available at BioProject [PRJNA1189271](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1189271).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**File S1 (mSphere01002-24-S0001.txt).** Workflow for quality controls, kraken assignment, and genome assembly.

**File S2 (mSphere01002-24-S0002.txt).** The Kraken2 classification of the chromosomal reads using a GTDB-based database.

**File S3 (mSphere01002-24-S0003.xlsx).** Predicted Dot/Icm effector proteins based on Bastion4 and T4Sepp tools.

**Legends for supplemental material (mSphere01002-24-S0004.docx).** Legends for Files S1, S2, and S3.

## REFERENCES

1. Dragičević P, Bielen A, Petrić I, Hudina S. 2021. Microbial pathogens of freshwater crayfish: a critical review and systematization of the existing data with directions for future research. *J Fish Dis* 44:221–247. <https://doi.org/10.1111/jfd.13314>

2. Ketterer PJ, Taylor DJ, Prior HC. 1992. Systemic rickettsia-like infection in farmed freshwater crayfish, *Cherax quadricarinatus*, p 173–179. In Shariff M, Subasinghe RP (ed), Diseases in Asian aquaculture. Asian Fisheries Society.
3. Owens L, Muir P, Sutton D, Wingfield M. 1992. The pathology of microbial diseases in tropical Australian Crustacea, p 165–172. In Shariff M, Subasinghe RP (ed), Diseases in Asian aquaculture. Asian Fisheries Society.
4. Tan CK, Owens L. 2000. Infectivity, transmission and 16S rRNA sequencing of a rickettsia, *Coxiella cheraxi* sp. nov., from the freshwater crayfish *Cherax quadricarinatus*. Dis Aquat Organ 41:115–122. <https://doi.org/10.3354/dao041115>
5. Cooper A, Layton R, Owens L, Ketheesan N, Govan B. 2007. Evidence for the classification of a crayfish pathogen as a member of the genus *Coxiella*. Lett Appl Microbiol 45:558–563. <https://doi.org/10.1111/j.1472-765X.2007.02240.x>
6. Ruth Elliman J, Owens L. 2020. Confirmation that candidatus *Coxiella cheraxi* from redclaw crayfish (*Cherax quadricarinatus*) is a close relative of *Coxiella burnetii*, the agent of Q-fever. Lett Appl Microbiol 71:320–326. <https://doi.org/10.1111/lam.13336>
7. Lührmann A, Newton HJ, Bonazzi M. 2017. Beginning to understand the role of the type IV secretion system effector proteins in *Coxiella burnetii* pathogenesis. Curr Top Microbiol Immunol 413:243–268. [https://doi.org/10.1007/978-3-319-75241-9\\_10](https://doi.org/10.1007/978-3-319-75241-9_10)
8. Graells T, Ishak H, Larsson M, Guy L. 2018. The all-intracellular order legionellales is unexpectedly diverse, globally distributed and lowly abundant. FEMS Microbiol Ecol 94:fy185. <https://doi.org/10.1093/femsec/fiy185>
9. Duron O, Doublet P, Vavre F, Bouchon D. 2018. The importance of revisiting legionellales diversity. Trends Parasitol 34:1027–1037. <https://doi.org/10.1016/j.pt.2018.09.008>
10. Gardner BR, Bachmann NL, Polkinghorne A, Hufschmid J, Tadeipalli M, Marenda M, Graves S, Arnould JPY, Stenos J. 2023. A novel marine mammal *Coxiella burnetii*-genome sequencing identifies a new genotype with potential virulence. Pathogens 12:893. <https://doi.org/10.3390/pathogens12070893>
11. Carey KL, Newton HJ, Lührmann A, Roy CR. 2011. The *Coxiella burnetii* Dot/Icm system delivers a unique repertoire of type IV effectors into host cells and is required for intracellular replication. PLoS Pathog 7:e1002056. <https://doi.org/10.1371/journal.ppat.1002056>
12. Beare PA, Gilk SD, Larson CL, Hill J, Stead CM, Omsland A, Cockrell DC, Howe D, Voth DE, Heinzen RA. 2011. Dot/Icm type IVB secretion system requirements for *Coxiella burnetii* growth in human macrophages. mBio 2:e00175–11. <https://doi.org/10.1128/mBio.00175-11>
13. Vogel JP, Andrews HL, Wong SK, Isberg RR. 1998. Conjugative transfer by the virulence system of *Legionella pneumophila*. Science 279:873–876. <https://doi.org/10.1126/science.279.5352.873>
14. Segal G, Purcell M, Shuman HA. 1998. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. Proc Natl Acad Sci U S A 95:1669–1674. <https://doi.org/10.1073/pnas.95.4.1669>
15. Qiu J, Luo ZQ. 2017. *Legionella* and *Coxiella* effectors: strength in diversity and activity. Nat Rev Microbiol 15:591–605. <https://doi.org/10.1038/nrmicro.2017.67>
16. Hubber A, Roy CR. 2010. Modulation of host cell function by *Legionella pneumophila* type IV effectors. Annu Rev Cell Dev Biol 26:261–283. <https://doi.org/10.1146/annurev-cellbio-100109-104034>
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. <https://doi.org/10.1371/journal.pone.0054566>
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. <https://doi.org/10.1073/pnas.1921344117>
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 J 3:e41. <https://doi.org/10.24072/pcjournal.269>
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. <https://doi.org/10.1111/tbed.14375>
21. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
22. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
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. <https://doi.org/10.1038/s41587-023-01983-6>
24. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>
25. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with raven. Nat Comput Sci 1:332–336. <https://doi.org/10.1038/s43588-021-00073-4>
26. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
27. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Tricycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22:266. <https://doi.org/10.1186/s13059-021-02483-z>
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. <https://doi.org/10.1101/gr.215087.116>
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. <https://doi.org/10.1038/s41467-020-20236-7>
30. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, Ruan J, Xiao C-L, Wang D, Wu D-D, Wang S. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol 25:107. <https://doi.org/10.1186/s13059-024-03252-4>
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 7:000685. <https://doi.org/10.1099/mgen.0.000685>
32. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36:1925–1927. <https://doi.org/10.1093/bioinformatics/bty2848>
33. ChaumeilPA, MussigAJ, HugenholtzP, ParksDH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 38:5315–5316. <https://doi.org/10.1093/bioinformatics/bt ac672>
34. Parks DH, Chuvpochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. <https://doi.org/10.1093/nar/gkab776>
35. Sherry NL, Horan KA, Ballard SA, Gonçalves da Silva A, Gorrie CL, Schultz MB, Stevens K, Valcanis M, Sait ML, Stinear TP, Howden BP, Seemann T. 2023. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. Nat Commun 14:60. <https://doi.org/10.1038/s41467-022-35713-4>
36. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom 4:e000206. <https://doi.org/10.1099/mgen.0.000206>
37. Robertson J, Bessonov K, Schonfeld J, Nash JHE. 2020. Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. Microb Genom 6:mgen000435. <https://doi.org/10.1099/mgen.0.000435>
38. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, Nobrega FL, Jackson SA. 2021. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids Res 49:10868–10878. <https://doi.org/10.1093/nar/gkab883>
39. Wang J, Yang B, An Y, Marquez-Lago T, Leier A, Wilksch J, Hong Q, Zhang Y, Hayashida M, Akutsu T, Webb GI, Strugnell RA, Song J, Lithgow T. 2019. Systematic analysis and prediction of type IV secreted effector

- proteins by machine learning approaches. *Brief Bioinform* 20:931–951. <https://doi.org/10.1093/bib/bbx164>
40. Hu Y, Wang Y, Hu X, Chao H, Li S, Ni Q, Zhu Y, Hu Y, Zhao Z, Chen M. 2024. T4SEpp: a pipeline integrating protein language models to predict bacterial type IV secreted effectors. *Comput Struct Biotechnol J* 23:801–812. <https://doi.org/10.1016/j.csbj.2024.01.015>
  41. Bi D, Liu L, Tai C, Deng Z, Rajakumar K, Ou HY. 2013. SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res* 41:D660–D665. <https://doi.org/10.1093/nar/gks1248>
  42. Cazalet C, Gomez-Valero L, Rusniok C, Lomma M, Dervins-Ravault D, Newton HJ, Sansom FM, Jarraud S, Zidane N, Ma L, Bouchier C, Etienne J, Hartland EL, Buchrieser C. 2010. Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. *PLoS Genet* 6:e1000851. <https://doi.org/10.1371/journal.pgen.1000851>
  43. Brenner AE, Muñoz-Leal S, Sachan M, Labruna MB, Raghavan R. 2021. *Coxiella burnetii* and related tick endosymbionts evolved from pathogenic ancestors. *Genome Biol Evol* 13:evab108. <https://doi.org/10.1093/gbe/evab108>
  44. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290. <https://doi.org/10.1093/bioinformatics/btg412>
  45. Yu G, Smith DK, Zhu H, Guan Y, Lam T-Y. 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36. <https://doi.org/10.1111/2041-210X.12628>
  46. Omsland A, Beare PA, Hill J, Cockrell DC, Howe D, Hansen B, Samuel JE, Heinzen RA. 2011. Isolation from animal tissue and genetic transformation of *Coxiella burnetii* are facilitated by an improved axenic growth medium. *Appl Environ Microbiol* 77:3720–3725. <https://doi.org/10.1128/AEM.02826-10>
  47. Voth DE, Beare PA, Howe D, Sharma UM, Samoilis G, Cockrell DC, Omsland A, Heinzen RA. 2011. The *Coxiella burnetii* cryptic plasmid is enriched in genes encoding type IV secretion system substrates. *J Bacteriol* 193:1493–1503. <https://doi.org/10.1128/JB.01359-10>
  48. Newton HJ, Kohler LJ, McDonough JA, Temoche-Diaz M, Crabill E, Hartland EL, Roy CR. 2014. A screen of *Coxiella burnetii* mutants reveals important roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. *PLoS Pathog* 10:e1004286. <https://doi.org/10.1371/journal.ppat.1004286>
  49. Thomas DR, Garnish SE, Khoo CA, Padmanabhan B, Scott NE, Newton HJ. 2024. *Coxiella burnetii* protein CBU2016 supports CCV expansion. *Pathog Dis* 82:ftae018. <https://doi.org/10.1093/femspd/ftae018>
  50. Abou Abdallah R, Million M, Delerce J, Anani H, Diop A, Caputo A, Zgheib R, Rousset E, Sidi Boumedine K, Raoult D, Fournier PE. 2022. Pangenomic analysis of *Coxiella burnetii* unveils new traits in genome architecture. *Front Microbiol* 13:1022356. <https://doi.org/10.3389/fmicb.2022.1022356>
  51. Oren A, Garrity GM. 2022. CANDIDATUS LIST No. 3. Lists of names of prokaryotic *Candidatus* taxa. *Int J Syst Evol Microbiol* 72. <https://doi.org/10.1099/ijsem.0.005186>
  52. Dmytrenko O, Russell SL, Loo WT, Fontanez KM, Liao L, Roeselers G, Sharma R, Stewart FJ, Newton ILG, Woyke T, Wu D, Lang JM, Eisen JA, Cavanaugh CM. 2014. The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics* 15:924. <https://doi.org/10.1186/1471-2164-15-924>
  53. Weber MM, Chen C, Rowin K, Mertens K, Galvan G, Zhi H, Dealing CM, Roman VA, Banga S, Tan Y, Luo ZQ, Samuel JE. 2013. Identification of *Coxiella burnetii* type IV secretion substrates required for intracellular replication and *Coxiella*-containing vacuole formation. *J Bacteriol* 195:3914–3924. <https://doi.org/10.1128/JB.00071-13>
  54. Larson CL, Beare PA, Voth DE, Howe D, Cockrell DC, Bastidas RJ, Valdivia RH, Heinzen RA. 2015. *Coxiella burnetii* effector proteins that localize to the parasitophorous vacuole membrane promote intracellular replication. *Infect Immun* 83:661–670. <https://doi.org/10.1128/IAI.02763-14>
  55. Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320:1651–1654. <https://doi.org/10.1126/science.1158160>
  56. Voth DE, Howe D, Beare PA, Vogel JP, Unsworth N, Samuel JE, Heinzen RA. 2009. The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J Bacteriol* 191:4232–4242. <https://doi.org/10.1128/JB.01656-08>
  57. Larson CL, Pullman W, Beare PA, Heinzen RA. 2023. Identification of type 4B secretion system substrates that are conserved among *Coxiella burnetii* genomes and promote intracellular growth. *Microbiol Spectr* 11:e00696-23. <https://doi.org/10.1128/spectrum.00696-23>
  58. Burette M, Allombert J, Lambou K, Maarifi G, Nisole S, Di Russo Case E, Blanchet FP, Hassen-Khodja C, Cabantous S, Samuel J, Martinez E, Bonazzi M. 2020. Modulation of innate immune signaling by a *Coxiella burnetii* eukaryotic-like effector protein. *Proc Natl Acad Sci U S A* 117:13708–13718. <https://doi.org/10.1073/pnas.1914892117>
  59. Fu M, Zhang J, Zhao M, Zhang S, Dai L, Ouyang X, Yu Y, Wen B, Zhou D, Sun Y, Jiao J, Xiong X. 2022. *Coxiella burnetii* plasmid effector B promotes LC3-II accumulation and contributes to bacterial virulence in a SCID mouse model. *Infect Immun* 90:e00016-22. <https://doi.org/10.1128/iai.00016-22>
  60. Chen C, Banga S, Mertens K, Weber MM, Gorbaslieva I, Tan Y, Luo ZQ, Samuel JE. 2010. Large-scale identification and translocation of type IV secretion substrates by *Coxiella burnetii*. *Proc Natl Acad Sci U S A* 107:21755–21760. <https://doi.org/10.1073/pnas.1010485107>
  61. Crabill E, Schofield WB, Newton HJ, Goodman AL, Roy CR. 2018. Dot/Icm-translocated proteins important for biogenesis of the *Coxiella burnetii*-containing vacuole identified by screening of an effector mutant sublibrary. *Infect Immun* 86:e00758-17. <https://doi.org/10.1128/IAI.00758-17>
  62. Fielden LF, Scott NE, Palmer CS, Khoo CA, Newton HJ, Stojanovski D. 2021. Proteomic identification of *Coxiella burnetii* effector proteins targeted to the host cell mitochondria during infection. *Mol Cell Proteomics* 20:100005. <https://doi.org/10.1074/mcp.RA120.002370>
  63. Zhang Y, Fu J, Liu S, Wang L, Qiu J, van Schaik EJ, Samuel JE, Song L, Luo Z-Q. 2022. *Coxiella burnetii* inhibits host immunity by a protein phosphatase adapted from glycolysis. *Proc Natl Acad Sci U S A* 119:e2110877119. <https://doi.org/10.1073/pnas.2110877119>
  64. Martinez E, Huc-Brandt S, Brelle S, Allombert J, Cantet F, Gannoun-Zaki L, Burette M, Martin M, Letourneur F, Bonazzi M, Molle V. 2020. The secreted protein kinase CstK from *Coxiella burnetii* influences vacuole development and interacts with the GTPase-activating host protein TBC1D5. *J Biol Chem* 295:7391–7403. <https://doi.org/10.1074/jbc.RA119.010112>
  65. Martinez E, Allombert J, Cantet F, Lakhani A, Yandrapalli N, Neyret A, Norville IH, Favard C, Muriaux D, Bonazzi M. 2016. *Coxiella burnetii* effector CvpB modulates phosphoinositide metabolism for optimal vacuole development. *Proc Natl Acad Sci U S A* 113:E3260–E3269. <https://doi.org/10.1073/pnas.1522811113>
  66. Kohler LJ, Reed SCO, Sarraf SA, Artega DD, Newton HJ, Roy CR. 2016. Effector protein Cig2 decreases host tolerance of infection by directing constitutive fusion of autophagosomes with the *Coxiella*-containing vacuole. *mBio* 7:e01127-16. <https://doi.org/10.1128/mBio.01127-16>
  67. Gomez-Valero L, Chiner-Oms A, Comas I, Buchrieser C. 2019. Evolutionary dissection of the dot/Icm system based on comparative genomics of 58 *Legionella* species. *Genome Biol Evol* 11:2619–2632. <https://doi.org/10.1093/gbe/evz186>
  68. Martini H, Detemmerman L, Soetens O, Yusuf E, Piérard D. 2017. Improving specificity of *Bordetella pertussis* detection using a four target real-time PCR. *PLoS One* 12:e0175587. <https://doi.org/10.1371/journal.pone.0175587>
  69. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE, Rolando M, Pasricha S, Reuter S, Demirtas J, Crumbach J, Descorps-Declere S, Hartland EL, Jarraud S, Dougan G, Schroeder GN, Frankel G, Buchrieser C. 2019. More than 18,000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. *Proc Natl Acad Sci U S A* 116:2265–2273. <https://doi.org/10.1073/pnas.1808016116>
  70. Debowski AW, Bzdyl NM, Thomas DR, Scott NE, Jenkins CH, Iwasaki J, Kibble EA, Khoo CA, Scheuplein NJ, Seibel PM, Lohr T, Metters G, Bond CS, Norville IH, Stubbs KA, Harmer NJ, Holzgrabe U, Newton HJ, Sarkar-Tyson M. 2023. Macrophage infectivity potentiator protein, a peptidyl prolyl cis-trans isomerase, essential for *Coxiella burnetii* growth and pathogenesis. *PLoS Pathog* 19:e1011491. <https://doi.org/10.1371/journal.ppat.1011491>
  71. Luo S, Lu S, Fan H, Sun Z, Hu Y, Li R, An X, Uversky VN, Chen Z, Tong Y, Song L. 2021. The *Coxiella burnetii* QpH1 plasmid is a virulence factor for colonizing bone marrow-derived murine macrophages. *J Bacteriol* 203:e00588-20. <https://doi.org/10.1128/JB.00588-20>

72. Wachter S, Cockrell DC, Miller HE, Virtaneva K, Kanakabandi K, Darwitz B, Heinzen RA, Beare PA. 2022. The endogenous *Coxiella burnetii* plasmid encodes a functional toxin-antitoxin system. *Mol Microbiol* 118:744–764. <https://doi.org/10.1111/mmi.15001>
73. Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Brinkman FSL. 2020. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. *Microb Genom* 6:mgen000436. <https://doi.org/10.1099/mgen.0.000436>
74. Buytaers FE, Saltykova A, Denayer S, Verhaegen B, Vanneste K, Roosens NHC, Piérard D, Marchal K, De Keersmaecker SCJ. 2021. Towards real-time and affordable strain-level metagenomics-based foodborne outbreak investigations using oxford nanopore sequencing technologies. *Front Microbiol* 12:738284. <https://doi.org/10.3389/fmicb.2021.738284>