Metagenomic Analysis of Fish-Associated *Ca.* Parilichlamydiaceae Reveals Striking Metabolic Similarities to the Terrestrial *Chlamydiaceae*

Alyce Taylor-Brown[1], Trestan Pillonel[2], Gilbert Greub[2], Lloyd Vaughan[3,4], Barbara Nowak[5], and Adam Polkinghorne[1,*]

[1]USC Animal Research Centre, Faculty of Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland, Australia

[2]Institute of Microbiology, University of Lausanne and University Hospital Center, Switzerland

[3]Institute of Veterinary Pathology, University of Zurich, Switzerland

[4]Pathovet AG, Tagelswangen, Switzerland

[5]Institute of Marine and Antarctic Studies, University of Tasmania, Newnham, Tasmania, Australia

*Corresponding author: E-mail: apolking@usc.edu.au.

Accepted: September 4, 2018

Data deposition: BioSamples SAMN08162581, SAMN08162582, SAMN08162583

Abstract

Chlamydiae are an example of obligate intracellular bacteria that possess highly reduced, compact genomes (1.0–3.5 Mbp), reflective of their abilities to sequester many essential nutrients from the host that they no longer need to synthesize themselves. The *Chlamydiae* is a phylum with a very wide host range spanning mammals, birds, fish, invertebrates, and unicellular protists. This ecological and phylogenetic diversity offers ongoing opportunities to study intracellular survival and metabolic pathways and adaptations. Of particular evolutionary significance are *Chlamydiae* from the recently proposed *Ca.* Parilichlamydiaceae, the earliest diverging clade in this phylum, species of which are found only in aquatic vertebrates. Gill extracts from three *Chlamydiales*-positive Australian aquaculture species (Yellowtail kingfish, Striped trumpeter, and Barramundi) were subject to DNA preparation to deplete host DNA and enrich microbial DNA, prior to metagenome sequencing. We assembled chlamydial genomes corresponding to three *Ca*. Parilichlamydiaceae species from gill metagenomes, and conducted functional genomics comparisons with diverse members of the phylum. This revealed highly reduced genomes more similar in size to the terrestrial *Chlamydiaceae*, standing in contrast to members of the *Chlamydiae* with a demonstrated cosmopolitan host range. We describe a reduction in genes encoding synthesis of nucleotides and amino acids, among other nutrients, and an enrichment of predicted transport proteins. *Ca*. Parilichlamydiaceae share 342 orthologs with other chlamydial families. We hypothesize that the genome reduction exhibited by *Ca.* Parilichlamydiaceae and *Chlamydiaceae* is an example of within-phylum convergent evolution. The factors driving these events remain to be elucidated.

Key words: *Chlamydiae*, *Chlamydia*-related bacteria, genomics, metagenomics, convergent evolution, intracellular bacteria, metabolism.

Introduction

Many bacteria have evolved to live in close association with other organisms. Among these are obligate intracellular bacteria that have adapted to thrive inside a eukaryotic host cell (Moulder 1974; Casadevall 2008; Omsland et al. 2014), a trait that is present in various bacterial phyla. *Chlamydiae* are one such example of host-associated, obligate intracellular bacteria with minimal genomes. The well-characterized family of human and animal pathogens, *Chlamydiaceae*, comprises 11 species and three *Candidatus* species, all with genomes in the range of 1.0–1.2 Mbp, encoding 898–1,097 genes (Vorimore et al. 2013; Bachmann et al. 2014; Sachse et al. 2014;

Taylor-Brown et al. 2016). With a few exceptions, species in this family are restricted to a particular host group (e.g., mammals only or birds only); however, we know very little about the host ranges of other families in a phylum which may comprise as many as 180 families or more (Lagkouvardos et al. 2014). Our efforts in this regard have been severely limited to those chlamydial families we could isolate for detailed study. The *Chlamydiaceae* were mostly first isolated from terrestrial vertebrates via egg yolk cultures which, apart from anything else, automatically select for bacteria able to survive and replicate at 39 °C. This stands in contrast to the marine environment, where temperatures above 28 °C can lead to widespread destruction, such as in coral bleaching events.

A highly successful alternative approach has been the use of amoebal coculture, commonly between 15 and 20 °C, which has led to the description of novel chlamydial families, commonly referred to as "*Chlamydia*-related bacteria" (CRBs). Recently grouped together in the *Parachlamydiales* order (Gupta et al. 2015; Pillonel et al. 2018), these species exhibit a cosmopolitan distribution and broad host range spanning unicellular and multicellular eukaryotes, with infection often resulting in disease in some hosts (Corsaro and Greub 2006; Lamoth and Greub 2010; Taylor-Brown et al. 2015). Members of these families possess genomes double the size of the *Chlamydiaceae* (2.1-3.4 Mbp), likely reflecting their adaptation to an amoebal cell niche, which may have played a role as a "melting pot" for horizontal transfer among *Chlamydiae* (Horn et al. 2004; Greub 2009; Collingro et al. 2011; Taylor-Brown et al. 2015). A key feature of these larger chlamydial genomes is evidence of a significantly expanded metabolic capacity compared with the *Chlamydiaceae*, (Horn et al. 2004; Bertelli et al. 2010; Collingro et al. 2011; Bertelli et al. 2015). To further expand our knowledge of the biological diversity in this phylum, encompassing its predicted wide ecological and phylogenetic diversity, we urgently need to gain information on families which, until now, have proven refractile to isolation attempts (Taylor-Brown et al. 2015; Konig et al. 2017). To this end, we sought to utilize novel genome data obtained by direct sequencing of an early-diverging gill-associated chlamydial family, the *Ca.* Parilichlamydiaceae.

The *Ca.* Parilichlamydiaceae family is associated with a highly prevalent and highly diverse gill disease, epitheliocystis, recently reviewed by Blandford et al. (2018). These bacteria are the most distantly related taxonomically described species in the phylum *Chlamydiae*, lying at the root of 16S rRNA-based phylogenies (Stride, Polkinghorne, Powell, et al. 2013; Taylor-Brown, Spang, et al. 2017). This observation has been further recently confirmed using genome-based phylogenies (Pillonel et al. 2018). Remarkably, since their first description in Yellowtail kingfish (Stride, Polkinghorne, Powell, et al. 2013), bacteria belonging to two proposed genera (*Ca.* Similichlamydia and *Ca.* Parilichlamydia) have been found in at least seven fish species (Steigen et al. 2013; Stride, Polkinghorne, Miller, et al. 2013; Stride, Polkinghorne, Miller, Nowak, 2013; Steigen et al. 2015; Guevara Soto et al. 2016; Seth-Smith et al. 2016; Taylor-Brown, Pillonel et al. 2017) in the northern and southern hemispheres, leading to the suggestion that each fish species might be infected with its own co-evolved chlamydial gill pathogen (Stride et al. 2014). In the first study to provide insight into the biology of these uncultivated chlamydiae, to our surprise, we discovered that the genome of the first species sequenced from the *Ca.* Parilichlamydiaceae, *Ca.* Similichlamydia epinephelii, was highly reduced and reminiscent of their terrestrial host-associated *Chlamydiaceae* relatives, despite their earlier divergence (Taylor-Brown, Pillonel et al. 2017).

In the present, broader comparative genomics study encompassing the draft genomes of several recently described species in the *Ca*. Parilichlamydiaceae sequenced in this study, we now suggest that these families have arrived at this genomic architecture by within-phylum convergent evolution as a result of increased genetic drift in these highly host-restricted species. We further describe key metabolic differences between these pathogens of marine vertebrates, the *Ca*. Parilichlamydiaceae (*Ca.* P. carangidicola, *Ca*. S. laticola, *Ca.* S. latridicola), their land vertebrate restricted cousins, the *Chlamydiaceae* (*Chlamydia trachomatis*) and the more ubiquitous or promiscuous *Parachlamydiales* (*Protochlamydia amoebophila*).

Materials and Methods

Sample Collection and DNA Treatment

As detailed in supplementary table S1, Supplementary Material online, nucleic acid extracts from previously obtained epitheliocystis-positive gill samples from Yellowtail kingfish ($n = 5$), Barramundi ($n = 2$), and Striped Trumpeter ($n = 3$), were prepared for metagenome sequencing (Stride, Polkinghorne, Powell, et al. 2013; Stride, Polkinghorne, Miller, et al. 2013; Stride, Polkinghorne, Miller, Nowak, 2013). Total nucleic acids were subject to depletion of methylated DNA using the NEBNext Microbiome Enrich Kit (New England Biolabs, Massachusetts, United States of America), according to manufacturer's instructions, as per Taylor-Brown et al. (2017). The resulting DNA was precipitated with ethanol and resuspended in 50 μL TE buffer prior to a 5 μL aliquot undergoing multiple displacement amplification using the Repli-G MDA kit (Qiagen, Germany), to enrich the microbial DNA.

Metagenome Sequencing and Assembly

Two micrograms of treated DNA were sent to the Australian Genome Research facility (AGRF, Parkville, Australia) for shotgun sequencing on either an Illumina HiSeq, or Illumina NextSeq, with 150 bp paired end reads. Resulting reads

were subject to quality assessment using FastQC prior to quality and adaptor trimming using Trimmomatic v0.33 (Bolger et al. 2014) with trimming parameters based on initial quality screening which differed per sample. 11,066,014–25,744,043 trimmed reads were then subject to de novo assembly using SPAdes v3.1.1 (Bankevich et al. 2012) in metagenome mode using default *k*-mer values. Chlamydial contigs were differentiated from other bacterial contigs and fragmented host contigs using metagenomic binning based on tetranucleotide frequencies, relative abundance and the presence of conserved bacterial genes by MaxBin v2.2.1 (Wu et al. 2016). In some cases, mixes of sequences from different origins occurred. They were rebinned and manually sorted using BlastX analysis against an in-house chlamydial protein database, and BlastX against the nr database. Contigs were also checked for G + C% versus coverage continuity, and in cases where a divergence in G + C% content coincided with a significant divergence in coverage, contigs were trimmed. Reads were mapped to the resulting bins and used for re-assembly with SPAdes to improve the assemblies. Only one sequence per species was used for comparative analysis.

The completeness of the three metagenomics bins was evaluated by identifying three different sets of highly conserved genes: 1) 107 nearly universal phylogenetic marker genes were identified with MaxBin v2.2.1 (Wu et al. 2016); 2) 200 genes that have been previously listed in minimal or core bacterial gene sets (Mushegian 1999; Koonin 2000; Gil et al. 2004; Sakharkar et al. 2004; Hutchison et al. 2016; Ye et al. 2016); 3) 208 genes conserved between *P. amoebophila*, *C. trachomatis*, *Legionella pneumophila*, *Coxiella burnetii*, *Escherichia coli*, and *Rhodopirellula baltica* (Gimenez et al. 2011). These genes were identified using MaxBin, BLASTp, BLASTkoala, CD-Search (Marchler-Bauer et al. 2017), and Hmmscan (Finn et al. 2015), with manual confirmation by inspection.

Metagenome Annotation and Comparative Genomic Analysis

The resulting genomic contigs were annotated using RAST (Aziz et al. 2008), CD-search (expect value threshold of 0.01) (Marchler-Bauer et al. 2017), Hmmscan (against the Pfam and TIGRFAM databases, using a gathering threshold and expect value threshold of 0.01) (Finn et al. 2015) with manual annotation curated in Artemis (Rutherford et al. 2000). The genomic data for the three *Ca.* Parilichlamydiaceae sequences were compared with 27 publicly available chlamydial genomes and one Verrucomicrobia genome (*Akkermansia muciniphila*), a free-living relative of the *Chlamydiae*, that was used as outgroup (supplementary table S2, Supplementary Material online). Proteins were clustered into orthologous groups using OrthoFinder version 0.4.0 with default parameters (Emms and Kelly 2015) and ascribed COG annotations (Galperin et al. 2015) using BLASTP version 2.3.0+ (Camacho et al. 2009) with an *e*-value cut-off of $1e-5$, a minimal query coverage of 50% and a minimal identity of 20% (supplementary table S3, Supplementary Material online). Amino acids of all the CDSs were submitted to KEGG BLASTKoala annotation for metabolic pathway construction, and compared with other bacterial pathways available in the database (Kanehisa and Goto 2000; Kanehisa et al. 2016). Burrows–Wheeler aligner, SAMtools, and BEDtools were used to map reads and assess read coverage across the metagenome-assembled genomes (Li et al. 2009; Li and Durbin 2010; Quinlan and Hall 2010).

A phylogenetic tree was also constructed based on a set of 29 conserved proteins: *pgk, argS, cafA, efp, sigA, sufB, sufS, secA, htrA, IleS, tuf, valS, pgsA, rpoA, cysS, priA, gmk, EngD, fusA, serS, rluD, pgi, gatB, rnhB, truB, rbfA, infB, nusA*, and *rpe*, as used previously (Taylor-Brown, Pillonel et al. 2017). These were identified in chlamydial genomes, as well as an outgroup species, *A. muciniphila*, aligned using MAFFT, and concatenated into a 16,244 amino acid alignment. This alignment was used to construct an approximate maximum likelihood phylogenetic tree using FastTree with a Jones–Taylor–Thorton model (Price et al. 2009) based on a MAFFT alignment with a BLOSUM62 scoring matrix (Katoh and Standley 2013) of 29 concatenated core proteins. Both MAFFT and FastTree were executed in Geneious (Kearse et al. 2012).

Results and Discussion

Ca. Parilichlamydiaceae Genome Features

Ten gill metagenomes from *Chlamydiae*-positive Yellowtail kingfish, Striped trumpeter and Barramundi were assembled de novo (supplementary table S1, Supplementary Material online). Three assemblies yielded nearly complete chlamydial genomes consisting of 7, 8, and 28 contigs for *Ca.* P. carangidicola, *Ca.* S. latridicola and *Ca.* S. laticola, respectively (table 1). The final draft assemblies totaled 771,491–881,979 bp. Automated and manual annotation resulted in 704, 768, and 782 predicted CDSs for *Ca.* P. carangidicola, *Ca.* S. laticola and *Ca.* S. latridicola, respectively, with coding percentages of 90.7–92.4% (tables 1 and 2). Of the remaining metagenomes, six did not yield complete genomes or genomes with sufficient coverage, and one did not yield any bacterial genomes. In the former examples, nonchlamydial bacterial species were in higher abundance (estimated based on read coverage) than the chlamydial sequences (supplementary table S1, Supplementary Material online).

The genomes of the three representatives of the *Parilichlamydiaceae* family exhibit G + C content of 37.5–43.7%, comparable to other *Chlamydia*-related species and in the range generally exhibited by highly reduced genomes (table 2; Merhej et al. 2009; Nishida 2012; Taylor-Brown et al. 2015). Further, the number of rRNAs (1 operon) and tRNAs (41–42) were similar to those found in other chlamydial genomes. Mean read coverage across the genomes also

Table 1
Details of Gill Samples Utilized for Metagenome Sequencing and Resulting Metagenome and Genome Assemblies

	Sample	76ST10	Hat2	007YTK11
Gill Sample Details	Host	Striped Trumpeter (*Latris lineata*)	Barramundi (*Lates calciferer*)	Yellowtail Kingfish (*Seriola lalandi*)
	Collection date	Nov 2010	Aug 2012	May 2011
	Collection location	Tasmania	South Australia	South Australia
	Chlamydial species	*Ca.* Similichlamydia latridicola	*Ca.* Similichlamydia laticola	*Ca.* Parilichlamydia carangidicola
Initial metagenome Assembly metrics	No. raw reads	16,776,972[a]	40,539,772[a]	29,405,932[b]
	No. trimmed reads	11,06,014[c]	25,744,042[d]	24,551,544
	No. contigs (≥0 bp)	58,390	153,109	369,007
	No. contigs (≥1,000 bp)	2,016	10,469	58,677
	Largest contig (bp)	604,417	265,296	687,426
	G+C content (%)	40.31	38.98	40.33
	N50	2,409	1,599	4,589
	L50	519	5,011	12,065
	No. unique 16S rRNA sequences	2	1	3
Chlamydial genome assembly metrics	No. contigs	8	28	7
	Largest contig (bp)	419,991	97,196	415,258
	Genome completeness[e]	94.4%	94.4%	94.4%
	G + C content (%)	37.45	43.73	38.83
	N50	150,576	37,594	415,258
	Mean read coverage	~692×	~46×	~49×
	% trimmed reads mapped	38%	1.7%	1%

[a]Obtained on an Illumina NextSeq (11 samples per lane/flow cell).
[b]Obtained on an Illumina HiSeq (9 samples per lane/flow cell).
[c]Reads from single lane used for assembly.
[d]Reads from two lanes used for assembly.
[e]Presence of conserved bacterial proteins.

Table 2
Genome Characteristics for Members of the Phylum *Chlamydiae*

Family (No. Genomes Used for Comparison)	Chromosome Size (Mbp)	No. Predicted CDS	Plasmid Size (kbp) (No. ORFs)	G + C Content (%)	No. rRNA Operons	No. tRNAs	Coding Percentage (%)
Ca. Parilichlamydiaceae *(4)*	0.771–0.981	704–943	No evidence	37.5–43.7	1	41–42	90.2–92.4
Parachlamydiaceae (9)	2.417–3.424	1,986–3,042	39–145 (40–160)	33.0–44.0	1–4	35–42	69.7[c]–90.8
Waddliaceae (2)	2.116–2.141	1,934, 2,028	15.5[a] (22)	43	2	34–37	92.2–93
Simkaniaceae (1)	2.496	2,519	132 (136)	38	1	35	91.3
Criblamydiaceae (2)	2.820–2.969	2,213, 2,426	9.1 (15), 89.5 (92)	38.0–48.0	1	40	86.0–89.7
Chlamydiaceae (5)	1.041–1.229	911–1,112	7.5[b] (8)	37.4–41.6	1–2	37–39	89.0–91.2

NOTE.—Ranges of values are given for each family, with the number of species or strains used for comparison in brackets.
[a]*W. chondrophila* plasmid in WSU 86-1044 strain only.
[b]*Chlamydia* plasmid not found in *C. abortus* or *Ca.* C. ibidis. Plasmid also not present in all strains of *C. pecorum*, *C. pneumoniae*, or *C. trachomatis*.
[c]Coding percentage of 69.7% and 71.3% in *Neochlamydia* sp. TUME1 and *Neochlamydia* sp. s13; remainder of genomes have a minimum coding percentage of 82.0%.

varied, with 49×, 46×, and 692× coverage for *Ca*. P. carangidicola, *Ca*. S. laticola and *Ca.* S. latridicola, respectively with the chlamydial genome accounting for 1–38% of the trimmed reads (table 1).

The *Ca*. Parilichlamydiaceae Possess Hallmarks of a Highly Reduced Obligate Intracellular Bacterial Genome

The completeness of these novel metagenome-assembled genomes was assessed based on the identification of three different sets of nearly universal bacterial genes and genes conserved in other chlamydial species.

First, binning analysis by Maxbin estimated the genomes to be 94.4% complete, based on the presence of 101 out of 107 phylogenetic marker genes, several of which are not conserved in other chlamydial genomes or bacteria with similarly small genome sizes (table 3). Second, we assessed the presence of 200 genes that have been previously listed in minimal or core bacterial genes (Gil et al. 2004; supplementary table S4, Supplementary Material online). Highly conserved genes

Table 3

Presence[a] of Sets of Universal Marker Genes in *Chlamydiae* and Other Small-genome Bacteria

Species	Core Minimal Bacterial Gene Set (Gil et al. 2004)	Core Minimal Bacterial Gene Set—Subset (Gil et al. 2004)	208 Orthologues in Amoebal & Related Pathogens (Gimenez et al. 2011)	100 COGs Conserved in 99% of Bacteria (Merhej et al. 2009)	100 COGs Lost from Intracellular Bacteria (Merhej et al. 2009)[b]	Single-copy Marker Genes Present in 95% of Bacteria (Dupont et al. 2011; Wu et al. 2014)
Ca. S. latridicola	185/200	27/36	141/208	99/100	6/100	101/107
P. amoebophila	186/200	24/36	208/208	99/100	30/100	106/107
C. trachomatis	183/200	23/36	208/208	100/100	12/100	106/107
M. pneumoniae		24/36	123/208	98/100		101/107
B. aphidicola		24/36	155/208	99/100		105/107
B. burgdorferi		20/36	136/208	97/100		105/107

[a]Presence includes putative non-orthologous gene displacements.
[b]Presence refers to how many of the "lost" genes are retained.

include genetic information processing genes and genes for core metabolic processes: 185/200 (92.5%) of these genes were identified in all three assemblies. We examined the subset of genes in the minimal gene set proposed by Gil et al. (2004) related to nucleotide, lipid and cofactor biosynthesis in an expanded set of bacteria and found 27/36 (75.0%) genes conserved, comparable to other chlamydiae and other bacteria (table 3, supplementary table S4, Supplementary Material online).

We considered a set of 100 Clusters of Orthologous Groups (COGs) shared by 99% of bacteria (Merhej et al. 2009): 99 were conserved in the three *Ca.* Parilichlamydiaceae assemblies and *P. amoebophila*, whilst all 100 were conserved in *C. trachomatis*. 97–100 COGs were conserved in other small-genome bacteria. We also considered a set of 208 orthologs conserved between *P. amoebophila, C. trachomatis*, *L. pneumophila*, *C. burnetii*, *E. coli*, and *R. baltica* (Gimenez et al. 2011), of which 141 were conserved in the *Ca.* Parilichlamydiaceae genomes, congruent with 123–155 in other bacterial species with reduced genomes (table 3).

Next, of 100 that were "universally lost" by alpha- and gamma-proteobacterial mutualists and parasites (Merhej et al. 2009), six were retained by *Ca.* Parilichlamydiaceae species, while 12 were retained by *C. trachomatis* and 30 by *P. amoebophila*, suggesting these may be required to infect the hosts of these species and may not be required by the host-restricted marine *Ca.* Parilichlamydiaceae.

There is a positive relationship between chlamydial genome size and number of coding regions (fig. 1*a*). All known members of the *Chlamydiaceae* and *Ca.* Parilichlamydiaceae, two distantly related families (fig. 1*b*), exhibit significantly smaller genomes compared with the *Parachlamydiales* (table 2, fig. 1*a*). We consider this genome size convergence to be reflective of the predicted strict dependence on vertebrate hosts and reduced host range of *Chlamydiaceae* and *Ca.* Parilichlamydiaceae species. *Parachlamydiales* species have the ability to infect both unicellular and multicellular eukaryotes (Collingro et al. 2011), whereas there is a lack of compelling evidence for members of the *Chlamydia* genus to infect and replicate in free-living amoeba in vitro (Wirz et al. 2008) nor have any *Chlamydia* species been successfully recovered from environmental samples by amoebal coculture. Similarly, all species of the *Ca.* Parilichlamydiaceae family (with a few exceptions) have been detected in a single fish species and have never been described in nonfish hosts (Blandford et al. 2018).

Chlamydia genomes display a striking level of gene order conservation (shared synteny) across different species, with the exception of the plasticity zone, a recombination "hotspot" situated near the replication termination region (Stephens et al. 1998; Kalman et al. 1999; Collingro et al. 2011). Shared synteny is often a hallmark of highly reduced, stabilized genomes (Mira et al. 2001) and although the fragmented nature of the *Ca.* Parilichlamydiaceae assemblies

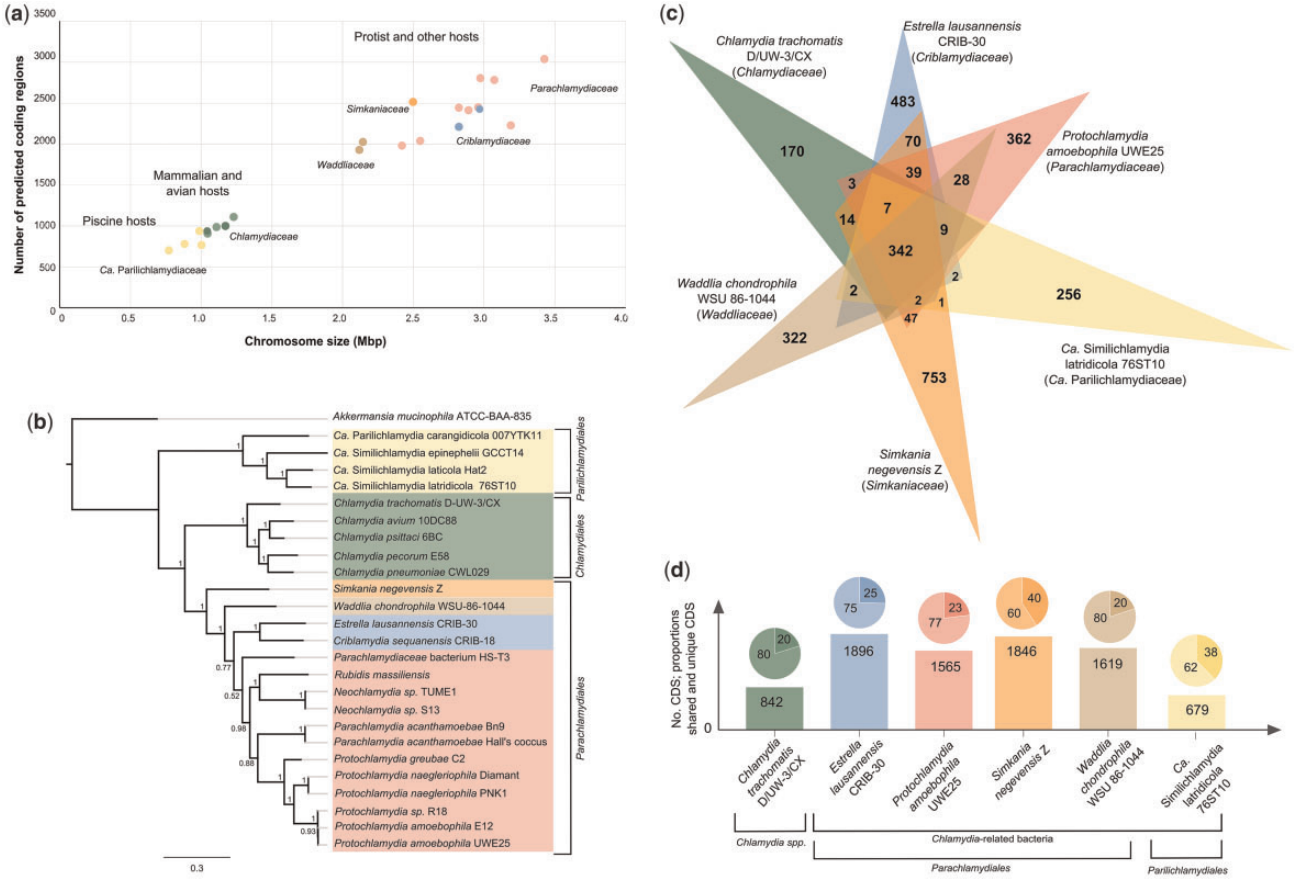

Fig. 1.—Genome dynamics of *Ca*. Parilichlamydiaceae within the *Chlamydiae*. (*a*) Relationship between chromosome size and number of coding regions in *Chlamydiae*. Species are colored based on chlamydial family. (*b*) Phylogenetic tree depicting the relationship between the *Ca*. Parilichlamydiaceae (yellow background) and other classified chlamydial species with genomic representatives. A free-living relative of the *Chlamydiae*, *Akkermansia muciniphila* was used as an outgroup species to root the tree. The tree was constructed by FastTree with a Jones–Taylor–Thorton model (Price et al. 2009) based on a MAFFT alignment with a BLOSUM62 scoring matrix (Katoh and Standley 2013) of 29 concatenated core proteins (*pgk, argS, cafA, efp, sigA, sufB, sufS, secA, htrA, IleS, tuf, valS, pgsA, rpoA, cysS, priA, gmk, EngD, fusA, serS, rluD, pgi, gatB, rnhB, truB, rbfA, infB, nusA*, and *rpe*). Both were executed in Geneious (Kearse et al. 2012). Numbers on the nodes indicate branch support values. (*c*) Orthologous groups shared and unique to representative species of six chlamydial families. Shared and unique orthologous groups are represented by the shaded triangles. Protein sequences were clustered using OrthoFinder. (*d*) Number of coding regions for each genome (bar chart), and proportion (pie chart) of each genome that is unique (dark shading) or shared with at least one other species (light shading).

makes it difficult to draw conclusions about chromosome-wide gene order, the lengths of the contigs are such that good evidence for locally syntenic regions between the three species described here for this family could be observed (fig. 2*a*). This high level of genomic synteny observed in the *Chlamydiaceae* and *Ca*. Parilichlamydiaceae families is not observed in the *Parachlamydiaceae* family. Indeed, a large number of genome rearrangements were observed between different species of the *Protochlamydia* genus such as *P. amoebophila* and *P. naegleriophila* (Collingro et al. 2011; Domman et al. 2014; Bertelli et al. 2016). Additionally, the level of shared synteny decreases with increased genetic distance (Bertelli et al. 2016), so unsurprisingly, there is very little synteny between *Ca*. Parilichlamydiaceae species and representatives of other chlamydial families (fig. 2*b* and *c*).

Updated *Chlamydiales* Core Genome and *Ca*. Parilichlamydiaceae-Specific Genes

The reduced genome size of the *Ca*. Parilichlamydiaceae has a strong impact on the size of the *Chlamydiales* core genome with just 342 orthologous groups shared between the *Ca*. Parilichlamydiaceae, *Parachlamydiaceae*, *Simkaniaceae*, *Waddliaceae*, *Criblamydiaceae*, and *Chlamydiaceae* families (fig. 1c, supplementary fig. S1a, table S6, Supplementary Material online). This figure is smaller than previously described core gene sets in which fewer, less diverse family representatives could be incorporated (Collingro et al. 2011; Psomopoulos et al. 2012; Pillonel et al. 2015).

The number of genes shared with at least one other species as a proportion of the total coding sequence is 62% in the *Ca*. Parilichlamydiaceae genomes, and this value varies between

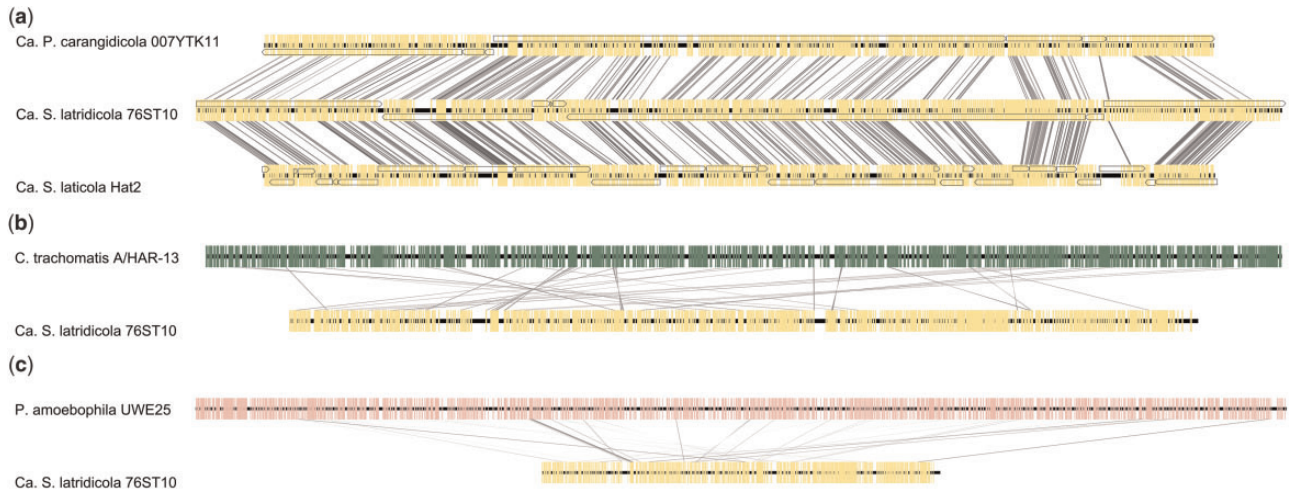

Fig. 2.—Genomic synteny within the *Ca.* Parilichlamydiaceae. Genome maps with tblastx analysis for (*a*) the three *Ca.* Parilichlamydiaceae species characterized in this study (contigs are denoted by unshaded pointed rectangles); (*b*) *Ca.* S latridicola and *Chlamydia trachomatis* and (*c*) *Ca.* S latridicola and *Protochlamydia amoebophila* (*b*). Coding regions are shaded vertical blocks along the black horizontal lines. Grey diagonal lines indicate position of orthologs between different species.

60% and 80% for *Parachlamydiales* and *Chlamydiaceae* (fig. 1*d*). We anticipate that the number of core chlamydial genes will begin to plateau once representatives of the many uncharacterized families are sequenced (Lagkouvardos et al. 2014). For example, despite their repeated observation and diversity among vectors such as ticks (Croxatto et al. 2014; Pilloux et al. 2015; Hokynar et al. 2016; Burnard et al. 2017). no arthropod-associated chlamydial genomes have yet been included in such analyses, although this family (*Ca.* Rhabdochlamydiaceae) are predicted to be the largest in the phylum. An expansion of such genomic studies will hopefully help elucidate host-specific adaptation mechanisms and clarify the relationship between genome size, population size, and host range.

Comparative analysis using one species representative for each chlamydial family revealed 256 proteins unique to *Ca.* S. latridicola (fig. 1*c*). Further comparative analyses including orthogroups present in all species in the family against genomes from other groups (e.g., *Parachlamydiales* only, *Chlamydiaceae* only), revealed 72 orthogroups shared by all species in the *Ca.* Parilichlamydiaceae and unique to that family. When one or two missing taxa were tolerated to account for the two genera in the family, this number increased to 126 or 147, respectively. Most of these were hypothetical proteins, with no COG or PFAM domain annotations. Interestingly, we also observed several genes that Parilichlamydiaceae have in common with one of the closest free-living relatives, *A. muciniphila* (*Verrucomicrobia*) that are coded with variable frequency throughout the rest of the phylum *Chlamydiae* (supplementary table S5, Supplementary Material online). Examples include 1) the protein translocation channel subunit *secG*, which is encoded by CRBs but has a variable presence within *Chlamydia* sp. genomes, 2) the *mlaDEF* genes, an ABC transporter complex involved in a phospholipid transport system, which again are mostly encoded only by the former "*Chlamydophila*" species, as well as CRBs, and; 3) tRNA synthetases for proline, asparagine, and glycine.

Large orthogroups missing from the *Ca.* Parilichlamydiaceae are lineage-specific proteins that have undergone expansions after divergence of a particular lineage or species, such as the polymorphic membrane proteins in *Chlamydia* spp. (Voigt et al. 2012; Vasilevsky et al. 2016) and the ubiquitin ligase-like protein expansion in the *Neochlamydia* spp. (Domman et al. 2014) (fig. 1*a*, supplementary table S7, Supplementary Material online). Notably, several of the largest orthogroups missing from *Ca.* Parilichlamydiaceae genomes are also absent from *Chlamydiaceae* genomes (supplementary fig. S1*b*, table S7, Supplementary Material online), including Rhs-like hypothetical proteins and a number of other hypothetical proteins. The absence of several gene sets from these two groups further confirms previous findings that many genes encoded only by *Parachlamydiales* were acquired by the ancestor of the *Parachlamydiales* clade or by HGT after speciation. Kamneva et al. (2012), in a wide-scale evolutionary genomic analysis of the *Planctomycetes-Verrucomicrobia-Chlamydiae* superphylum, suggested that the last common chlamydial ancestor was already characterized by a small genome, in comparison to the genome of the *Verrucomicrobia-Chlamydiae-Lentisphaerae* ancestor and that genome expansion has driven the evolution of the *Parachlamydiales* clade. We hypothesize that the low rates of gene birth, duplication and transfer seen for the *Chlamydia* species is analogous to that seen for the *Ca.* Parilichlamydiaceae as there is no evidence of recent HGT and no gene duplication.

The *Ca*. Parilichlamydiaceae Species Exhibit a Minimal Metabolic Capacity

Genetic Information Processing

All enzymes (13 genes; Gil et al. 2004) involved in basic replication machinery (e.g., DNA polymerases) are present in the *Ca*. Parilichlamydiaceae genomes, as are genes involved in DNA repair and modification. Likewise, all components of the basic transcription machinery (e.g., RNA polymerases) are present. The *Ca*. Parilichlamydiaceae genomes encode all 21 aminoacyl-tRNA synthesis genes except for Glutaminyl-tRNA synthase, which is also not encoded by *P. amoebophila* and *C. trachomatis*.

Three parts of the translation machinery involved in tRNA maturation and modification were missing in the three draft genomes: Two GTP-binding proteins (*mnmE* and *mnmG*) and dimethyladenosine transferase (*ksgA*). Two of the four cellular transport proteins were present: Low-affinity inorganic phosphate transporter (*pitA*) and the histidine-containing phosphocarrier protein (*ptsH*), both of which are conserved in other *Chlamydiae*. Neither of the phosphotransferase enzymes—PTS enzyme I (*ptsI*), which is present in other *Chlamydiae*, and PTS enzyme II (*ptsG*), which is not, were predicted in the draft *Ca*. Parilichlamydiaceae genomes (supplementary table S4, Supplementary Material online).

Nucleotide Metabolism and Acquisition

It has been well-documented that the *Chlamydiaceae* lack genes for purine and pyrimidine de novo synthesis (Stephens et al. 1998; Kalman et al. 1999; Bertelli et al. 2010; Voigt et al. 2012; Nunes and Gomes 2014; Bertelli et al. 2015), and this is also seen in the *Ca*. Parilichlamydiaceae. Notably, the *Ca*. Parilichlamydiaceae appears to lack ribose-phosphate pyrophosphokinase (*prsA*), meaning phosphoribosyl pyrophosphate (PRPP), a precursor molecule, cannot be synthesized. *Protochlamydia amoebophila* and *A. muciniphilia* both possess *prs*A, but *C. trachomatis* does not (supplementary table S4, Supplementary Material online), further supporting the previous hypothesis (Kamneva et al. 2012). Hence, none of the essential genes involved in purine de novo synthesis appears to be encoded by the *Ca*. Parilichlamydiaceae genomes, nor are genes for purine salvage, including several that were determined as part of the minimal gene set, for example, hypoxanthine phosphoribosyltranferase; *hpt* (supplementary fig. S2, table S4, Supplementary Material online). *Ca*. Parilichlamydiaceae do, however, encode adenylate, guanylate and nucleoside-diphosphate kinases (*adk*, *gmk*, and *ndk*), ribonucleotide reductase subunits (*nrdEF/AB*), and DNA and RNA polymerases (*polA* and *rpoABC*), meaning the pathways for purine and pyrimidine nucleic acid synthesis are intact (supplementary fig. S2, Supplementary Material online).

For pyrimidine metabolism, it does not appear that UTP can be aminated to CTP by species in the *Ca*. Parilichlamydiaceae due to the lack of CTP synthetase (*pyrG*), which is otherwise ubiquitously encoded throughout the phylum. We predict dTMP can be synthesized from dUMP via thymidylate synthetase (supplementary fig. S2, Supplementary Material online), the identity and predicted function of which is a flavin-dependent analog of the *thyA* methyltransferase, *thyX*. The origin of the *Ca*. Parilichlamydiaceae *thyX* proteins appear to be distinct from other chlamydial *thyX* sequences, which in previous studies were omitted from phylogenetic analysis due to their lack of sequence identity (Myllykallio et al. 2002). Rather, the *Ca*. Parilichlamydiaceae *thyX* sequences share up to 41% amino acid identity with Spirochaetes and Deinococci. The phylogenetic relationships between the thyX proteins is depicted in supplementary figure S3, Supplementary Material online, confirming the *Ca*. Parilichlamydiaceae proteins are from a different lineage to other chlamydial thyX proteins.

The reduction of genes for purine and pyrimidine metabolism in the draft *Ca*. Parilichlamydiaceae genomes suggests that the bacteria source at least some of those compounds from the host. Fish excrete small amounts of nitrogenous waste as purines, taurines, creatine and methylamines (Evans et al. 2005), so they could be obtained by the bacteria from gill epithelial cells (pavement cells; PVCs), which facilitate excretion of most of the nitrogenous waste in fish (Ip and Chew 2010). This is consistent with the observations in Striped trumpeter, where the infected cells were mostly PVCs as the cysts were present in the respiratory epithelium on the lamellae (Lai et al. 2013). While the infected cells were not identified in epitheliocystis cases from YTK or Barramundi (Stride, Polkinghorne, Powell, et al. 2013; Stride, Polkinghorne, Miller, Nowak, 2013), according to the position of these cells at the base of the gill lamella, they were either PVCs or chloride cells (mitochondria rich cells; MRCs). PVCs are the most common cells (90%) in the fish gill epithelium (Evans et al. 2005) and are the main type of cell reported to be infected during epitheliocystis (Nowak and LaPatra 2006; Stride et al. 2014), whilst MRCs were the main cell type infected in amberjack, *Seriola dumerili* (Crespo et al. 1999) and Atlantic salmon, *Salmo salar* (Nylund et al. 1998). The preference of *Ca*. Parilichlamydiaceae species for certain cells and the associated implications for chlamydial metabolism remain to be elucidated.

Amino Acid Metabolism

Parachlamydiales are capable of synthesizing several amino acids whilst *Chlamydiaceae* and other obligate intracellular bacteria are frequently auxotrophic for many of these nutrients (Gil et al. 2004; Bertelli et al. 2010; Omsland et al. 2014; Bertelli et al. 2015). No genes for amino acid synthesis were included in the minimal gene set except for glycine

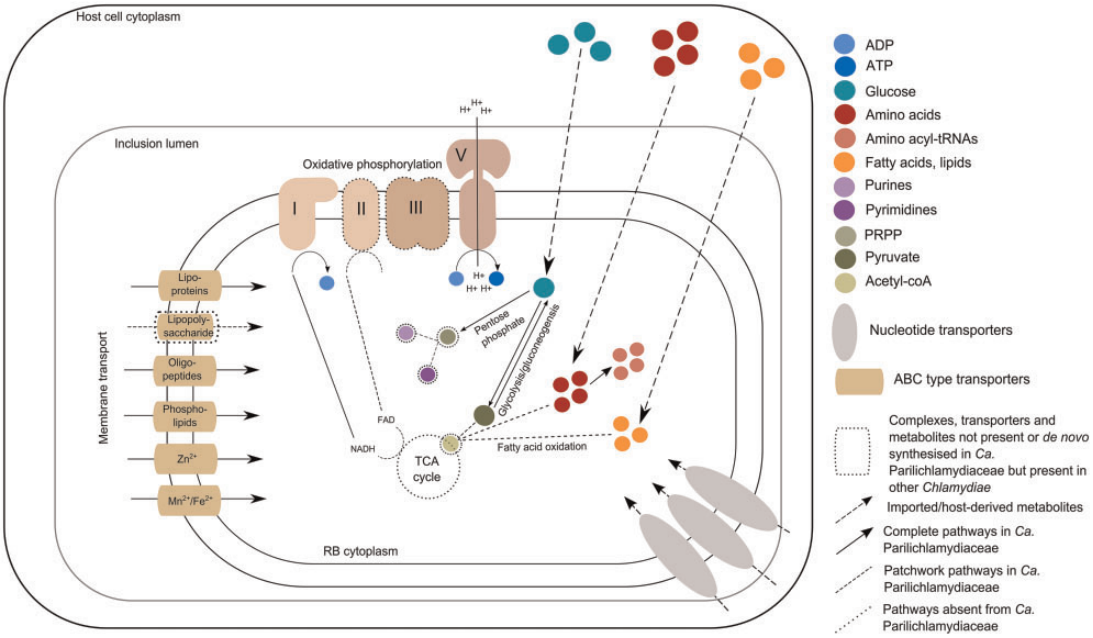

Fig. 3.—Major metabolic pathways inferred from genome sequences of three species of the *Ca.* Parilichlamydiaceae. Schematic diagram of selected metabolic processes in the reduced genomes of *Ca.* Parilichlamydiaceae species. The RB is represented within the inclusion in the host cell cytoplasm. Solid and dashed lines represent predicted or absent paths in the *Ca.* Parilichlamydiaceae, respectively. Substrates and transporters are colored as per the legend.

hydroxymethyltransferase (*glyA*), which catalyzes serine to glycine interconversion and is encoded by the *Ca.* Parilichlamydiaceae (Gil et al. 2004; supplementary table S4, Supplementary Material online). No other amino acid synthesis genes are encoded by *Ca.* Parilichlamydiaceae in a striking similarity to *Mycoplasma pneumoniae*, which instead encodes a number of transport systems with varying levels of substrate-specificity (Himmelreich et al. 1996).

One major difference between the *Chlamydiaceae* and *Ca.* Parilichlamydiaceae genomes is the presence of a tryptophan synthesis operon in some strains of *C. trachomatis*, which is one of the only amino acids that some chlamydial species can synthesize. Tryptophan metabolism is implicated in persistence and tissue tropism (Akers and Tan 2006) but interestingly cannot be synthesized by all species or strains, as evidenced by the lack of an intact tryptophan biosynthesis operon in our draft genomes. A tryptophan/tyrosine permease is predicted in the *Ca.* Parilichlamydiaceae genomes, which shares 30% amino acid identity to that of *Parachlamydia* spp., with the amino acid/polyamine transporter 2 domain predicted (PF03222, IPR018227). This transporter may enable uptake and utilization of tryptophan and related amino acids in lieu of biosynthesis.

Carbohydrate and Energy Metabolism

Ca. Parilichlamydiaceae encode the essential components of the glycolysis pathway to produce pyruvate, ATP and NADH from glucose, or the reverse, gluconeogenesis (fig. 3). Glucose-6-phosphate is most likely imported via a sugar-phosphate transporter (*UhpC*) (Schwoppe et al. 2002; Mehlitz et al. 2017), which is shared among other chlamydiae, as the starting molecule for carbohydrate metabolism. Unlike other *Chlamydiae*, *Ca.* Parilichlamydiaceae are predicted to use phosphoglucomutase (*pgm*) to convert glucose 1-phosphate to glucose 6-phosphate, whereas *P. amoebophila* uses glucokinase (*glk*) to phosphorylate glucose. *Akkermansia muciniphila*, one of the closest free-living bacterial relatives of *Chlamydiae*, encodes both enzymes. Phylogenetic analysis of the *A. muciniphila* and chlamydial enzymes showed that the *A. muciniphila glk* and *pgm* are not ancestral to chlamydial *glk* or *pgm* (supplementary fig. S4, Supplementary Material online), but rather, have been acquired by each chlamydial lineage separately after the divergence from the shared PVC ancestor.

The components of the pentose phosphate pathway are present and show similarities to both *Parachlamydiales* and *Chlamydiaceae*. *Ca.* Parilichlamydiaceae appear to rely on diphosphate-dependent phosphofructokinase (*pfpB*) to convert glucose 6-phosphate and fructose 2, 6-biphosphate, like *Parachlamydiales*, instead of the ATP-dependent phosphofructokinase 1 (*pfk*), used by *Chlamydiaceae*.

Along with glucose, host-derived pyruvate may also be imported into the chlamydial cell, as suggested by previous cell biology and genomic analyses. No specific transporters have yet been identified in chlamydiae (Zomorodipour and Andersson 1999; Konig et al. 2017); however, *Rickettsiae* encode auxin efflux carrier transporters which may achieve this

(Driscoll et al. 2017). Although pyruvate can be synthesized from phosphoenol-pyruvate by pyruvate kinase (*pyk*), *Ca*. Parilichlamydiaceae genomes appear to lack the pyruvate dehydrogenase subunits (*aceEF*) for acetyl-coA generation from pyruvate encoded by other *Chlamydiae*. Hence, genes for the TCA cycle, which are differentially encoded throughout the phylum, are also absent. The 1) presence of a complete cycle in most *Parachlamydiales* while 2) *Chlamydiaceae* lack essential genes (citrate synthase, aconitase and isocitrate dehydrogenase [*gltA*, *acnB*, and *icd*], and 3) additional genes are truncated in certain strains (Stephens et al. 1998; Kalman et al. 1999; Mojica et al. 2011; Voigt et al. 2011), further suggests that the TCA genes encoded by the last common ancestor of the *Parachlamydiales* and *Chlamydiaceae* were lost by both the *Chlamydiaceae* and the *Ca*. Parilichlamydiaceae. Further, acetyl-coA can be generated via the above pathway in *M. pneumoniae* which lacks other genes for a TCA, and *B. burgdorferi* is missing the same enzymes as *Ca*. Parilichlamydiaceae genomes (Himmelreich et al. 1996; Fraser et al. 1997). No TCA genes were included in the minimal gene set, as several genes are missing or not essential in the species analysed (Gil et al. 2004). The authors proposed that NAD+ could be yielded by the reduction of pyruvate by lactate dehydrogenase (*ldh*), which curiously is not encoded by any chlamydial species. It is likely that bacteria with highly reduced genomes that lost the TCA cycle rely more heavily on their host as energy source.

The oxidative phosphorylation pathway encoded by the *Ca.* Parilichlamydiaceae genomes is functionally similar to that of the *Chlamydiaceae*, with the expected set of V-type ATPases (subunits ABDEIK) encoded for proton/sodium gradient generation (fig. 3). Electrons are donated by NADH (complex I) and reduced by a set of Na+-translocating NADH-quinone reductases (subunits ABCDEFG). The succinate and fumarate dehydrogenases (complex II) appear to be absent from *Ca.* Parilichlamydiaceae genomes.

Predicted carbon metabolism pathways in the *Ca.* Parilichlamydiaceae highly resemble that of the other species investigated, given that much of the carbon utilized by *Ca.* Parilichlamydiaceae is derived from glycolysis and both the oxidative and reductive phases of the pentose phosphate pathway already detailed. Absent from the *Ca*. Parilichlamydiaceae genomes, but present in other *Chlamydiae* are several enzymes that participate in tetrahydrofolate (THF) and ammonia metabolism. An additional enzyme only present in *C. trachomatis* and *P. amoebophila*, L-serine dehydratase (*sdsL*), can also convert serine to ammonia, while methylenetetrahydrofolate dehydrogenase (*folD*) participates in methylene-THF formation. Again, this is pertinent to the *Ca.* Parilichlamydiaceae infecting gill epithelial cells which participate in nitrogenous compound exchange. Methylamines are one such by-product which may be able to be uniquely used by the *Ca.* Parilichlamydiaceae.

Lipid and Fatty Acid Metabolism

A minimal suite of genes for biosynthesis of lipids is proposed for bacterial endosymbionts that reside inside a host-derived vacuole (Mushegian and Koonin 1996; Gil et al. 2004), which is congruous with *Chlamydiae* residing within its inclusion. *Ca.* Parilichlamydiaceae species, and other *Chlamydiae*, appear to encode patchwork pathways for lipid and fatty acid metabolism (Gil et al. 2004; supplementary table S4, Supplementary Material online). For example, *Ca.* Parilichlamydiaceae lack the pyruvate oxidoreductases and dehydrogenases to synthesize acetyl-coA, the starting molecule for fatty acid synthesis, as do several other *Chlamydiae*. Further, only a few species in the *Criblamydiaceae* and *Waddliaceae* encode *fadD*, long-chain fatty acid-coA ligase (Bertelli et al. 2014; Bertelli et al. 2015), necessary for fatty acid biosynthesis, however, *Ca.* Parilichlamydiaceae do encode a long-chain-fatty-acid–[acyl-carrier-protein] ligase (*aas*) to cleave long-chain fatty acids. This protein has a phosphopantetheine binding domain (PF00550, IPR009081) and acyl-coA synthetase (AMP-forming) domain (PF00501, IPR000873), the latter of which may account for the lack of *fadD*. *Ca.* Parilichlamydiaceae encode two copies of CDP-diacylglycerol-glycerol-phosphatidyltransferase (*pgsA*), necessary for conversion of CDP-diacyl-glycerol to phosphatidyl-glycerophosphate; however, *Chlamydiae* do not possess the phosphatidylglycerophosphatase to convert this to phosphatidyl glycerol, the precursor to cardiolipin. This patchwork of genes, coupled with uptake of host-derived substrates and intermediates by predicted transporters described below, may accomplish fatty acid and lipid metabolism in lieu of other complete pathways.

Ca. *Parilichlamydiaceae Genomes are Rich in Membrane Transporters*

As compensation for their reduced genome size and biosynthetic capacities, *Ca.* Parilichlamydiaceae species possess a diverse repertoire of predicted membrane transporters as touched on in earlier sections. The most abundant of these are ABC-type transporters. For oligopeptide transport, up to six copies of *oppA* are encoded for in each *Ca.* Parilichlamydiaceae species, while *oppBCDF* are also present at a single locus. This gene structure is most similar to that observed in the *Chlamydiaceae*, whereas in *Parachlamydiales* genomes, only *oppABC* are consistently present and encoded at separate loci. Despite dipeptide (*dpp*), cationic peptide (*sap*), and nickel (*nik*) transporters being seemingly absent from these *Ca.* Parilichlamydiaceae genomes, the presence of *dpp*, *sap* and *nik* domains (cl01709, cl28564, cl26276) predicted in the *opp* proteins may suggest that these genes function to transport several molecule classes.

The presence of operons for predicted phospholipid, lipoprotein and metallic cation transport are again variable

between *Parachlamydiales* and the *Chlamydiaceae*. For example, predicted phospholipid transporters (*mlaDEF*) are encoded by *Parachlamydiales* only, whereas more predicted lipopolysaccharide transporters are encoded by *Parachlamydiales* (*rfbAB* and *lptFGB*) than *Chlamydiaceae* (*lpt* only) but not by *Ca.* Parilichlamydiaceae. Suspected lipoprotein transporters (*lol*) are predicted throughout the phylum. *Ca.* Parilichlamydiaceae encode several predicted metallic cation transporters (zinc, manganese, iron); again, some may serve several functions. Of these, only *znu* type zinc transporters are seen in the *Chlamydiaceae*. Neither methionine transporters (encoded by *Parachlamydiales* only) nor arginine transporters (encoded by *Chlamydiaceae* only) are encoded in the draft genomes of the *Ca.* Parilichlamydiaceae.

Ca. Parilichlamydiaceae genomes are predicted to encode three nucleotide transporters/translocases (NTT or Npt) that may facilitate the acquisition of energy from their hosts. Homologs of those transporters can be identified in all *Chlamydiae* genomes and several other bacterial phyla (Greub and Raoult 2003; Heinz et al. 2014). In *Ca.* Parilichlamydiaceae genomes, they are each around 500 amino acids in length and contain an MFS domain (cl21472) and 10–12 transmembrane domains, providing further evidence that they are membrane-bound transport pumps. These NTTs share 12–62% amino acid similarity to the five functionally characterized NTTs encoded by *P. amoebophila* and the two encoded by *C. trachomatis* (*P. amoebophila* NTT4 is highly divergent). *Ca.* Parilichlamydiaceae NTT1 shares significant sequence identity with Npt1 in *C. trachomatis* which has been shown to also import NAD, a function which is carried out by NTT4 in *P. amoebophila* (Tjaden et al. 1999; Haferkamp et al. 2004, 2006). For this reason, *C. trachomatis* NTT1 has been termed a "hybrid" class I/III antiporter (Fisher et al. 2013). Although substrate affinity cannot be predicted by sequence similarity, the *Ca.* Parilichlamydiaceae NTT1, or another homolog, may also function as a hybrid class I/III NAD/ATP antiporter. The presence of these transporters may be sufficient for nucleotide acquisition by *Ca.* Parilichlamydiaceae.

Limitations and Future Directions

While we have presented a brief analysis of the metabolic strategies employed by novel gill-associated chlamydiae in comparison with host-associated bacteria with reduced genomes, there are several shared and unique mechanisms still to be understood. We have not performed any analysis of biological features reported in other chlamydial genomics studies such as virulence mechanisms, secretion systems, antimicrobial resistance, outer membrane proteins, recombination and plasmids (Dugan et al. 2004; Horn et al. 2004; Greub et al. 2009; Bertelli et al. 2010; Collingro et al. 2011; Domman et al. 2014; Bertelli et al. 2016; Collingro et al. 2017). Some of these were described in our previous study (Taylor-Brown, Pillonel et al. 2017), whilst others should be the focus of further comparative studies upon completion of genome sequencing of other chlamydiae.

It is unclear what role cohabitation of the gill by *Chlamydiae* and other microbes could play in shaping these chlamydial genomes. Although not described in detail in this paper, our sequencing yielded highly diverse metagenomes generally containing at least two dominant bacterial species, as has been described previously (Andersson et al. 2013; Qi et al. 2016; Seth-Smith et al. 2016; Taylor-Brown et al. 2016; Taylor-Brown, Pillonel, et al. 2017). This is of course a pitfall of the depletion-enrichment technique which enriches the most abundant bacterial species in the sample. However, as in vitro culture systems remain elusive for these bacteria, and no reference genomes have been characterized upon which to base targeted genome sequencing methods, as has been performed recently for other chlamydial species (Putman et al. 2013; Christiansen et al. 2014; Taylor-Brown et al. 2018), we opted for a non-targeted deep-sequencing approach, which provided insight into the chlamydial agent of interest against a background of gill microflora, and highlights an area of further study. A major caveat of this method is that the inferences we have made throughout this analysis are based on draft genomes. However, we are confident, based on the presence of conserved and essential genes and the number of rRNAs and tRNAs, that we have assembled complete or near-complete chlamydial genomes, that will be able to be confirmed and improved once systems are established to culture these organisms.

These genome data could be used to inform future cultivation attempts (e.g., nutrient supplementation), as no systems yet exist to isolate these species. Of note, cocultivation of *Chlamydiae* in free-living amoebae has only been successful with chlamydial species possessing larger genomes, such as the *Waddliaceae* (Bertelli et al. 2010) and *Parachlamydiaceae* (Greub et al. 2009), which may be a prerequisite for a host range spanning both unicellular and multicellular species. As with the terrestrial vertebrate specialists, the *Chlamydiaceae*, all with a comparable genome size, the *Ca.* Parilichlamydiaceae may also have a restricted aquatic vertebrate host range. Most recently, *Ca.* Syngnamydia salmonis, a member of the *Simkaniaceae* family, was cocultivated in *Neoparamoeba perurans* (Nylund et al. 2018), suggesting that in fact some gill-associated chlamydiae can survive and replicate in a gill-associated amoebal host. However, no *Ca.* S. salmonis were isolated from a natural *N. peurans* infection. Characterization of the genome sequence of this and other species will help confirm the host range-genome size relationship exhibited by this phylum, as well as provide the first interfamily genomic comparisons between gill-associated chlamydiae.

Further examination of other fish species sharing the habitats of *Ca*. Parilichlamydiaceae infected fish, as well as the widening range of potential invertebrate hosts known to harbor *Chlamydiae* (Viver et al. 2017), will also be fundamental to understanding the host range and infection dynamics in gill-associated chlamydiae. Indeed, epitheliocystis-like lesions have also been described in major marine invertebrate taxa (sponges and corals), where they are known as CAMAs, cell associated microbial aggregates (Work and Aeby 2014; Bourne et al. 2016; Webster and Thomas 2016). Closely related gamma-proteobacteria of the genus *Endozoicomonas* have been found in invertebrates and also as pathogens causing epitheliocystis (Katharios et al. 2015). It will be intriguing to see whether members of the *Chlamydiae*, including the *Ca*. Parilichlamydiaceae, are also shared between invertebrates and fish. If so, this would open up new avenues for exploring pathogen transmission in marine ecosystems. Further, population diversity studies for these three species could also be useful to understand routes of transmission and could be achieved by further genome sequencing or the development and implementation of a suitable multi-gene typing scheme. Lastly, the ongoing description of novel epitheliocystis agents in new hosts and increasing number of available epitheliocystis-associated bacterial genome sequences from diverse phyla including *Chlamydiae*, *Betaproteobacteria*, and *Gammaproteobacteria* warrants a separate broad-scale comparative genomics study.

Conclusions

Comparative analyses of representatives of three new gill-associated *Ca.* Parilichlamydiaceae species show that members of this early diverging *Chlamydiae* clade have highly reduced genomes with limited metabolic capacity. The variation in genome size observed in those three clades could be the result of gene gains in the *Parachlamydiales* clade or convergent genome reduction in *Chlamydiales* and *Parilichlamydiales*. We hypothesize that the highly reduced genomes of the distantly related *Chlamydiaceae* and *Ca.* Parilichlamydiaceae spp. is associated with their limited host range as compared with amoeba-associated *Parachlamydiales* species. This genome reduction may be reflective of 1) a long period of host adaptation to the vertebrate gill niche, 2) limited transmission routes, resulting in 3) low population sizes, leading to 4) high level of genetic drift and fixation of deletions, favoring genome reduction. Genome sequencing of diverse novel chlamydial species from a range of ecological niches will further elucidate the cause of the wide variation of genome size observed within the *Chlamydiae* phylum.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors would like to acknowledge Megan Stride and Abigail Elizur for generously sharing DNA extracts, Helena Seth-Smith and Weihong Qi for discussions regarding appropriate DNA preparation and bioinformatics methods, Nathan Bachmann, Erin Price, and Derek Sarovich for bioinformatics assistance, and Peter Timms and Martina Jelocnik for ongoing chlamydial taxonomy and evolution discussions. This work was supported by a University of the Sunshine Coast Faculty Research Grant awarded to A.T.B. (grant number FoSHEE-HDR-035.06065).

Ethics Approval and Consent to Participate

DNA samples used for genome sequencing were obtained from previous studies with ethics approval as follows. Sampling of Yellowtail kingfish was conducted opportunistically and after commercial harvest. Animals were killed by commercial staff and subject to standard industry harvest practices. Barramundi samples were collected as a part of routine farm health monitoring and provided to the researchers as fixed samples. They were exempt from ethics approval by the University of Tasmania Animal Ethics Committee. Sampling of Striped trumpeter was approved by the University of Tasmania Animal Ethics Committee, project number AEC0009926.

Authors' Contributions

A.T.B. conducted lab work, data analysis, and wrote the manuscript. T.P. provided bioinformatics support and wrote the manuscript. All authors contributed to interpretation of results and reviewed the manuscript.

Availability of Data and Material

The genomic data generated during the current study for *Ca*. Parilichlamydia carangidicola, *Ca*. Similichlamydia laticola, and *Ca*. Similichlamydia latridicola have been submitted to the NCBI genome database, This project has been deposited at NCBI under the accessions SAMN081652581-SAMN08162583.

Literature Cited

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

Andersson P, Klein M, Lilliebridge RA, Giffard PM. 2013. Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. Clin Microbiol Infect. 19(9):E405–E408.

Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9(1):75.

Bachmann NL, Polkinghorne A, Timms P. 2014. *Chlamydia* genomics: providing novel insights into chlamydial biology. Trends Microbiol. 22(8):464–472.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.
Bertelli C, et al. 2010. The *Waddlia* genome: a window into chlamydial biology. PLoS One 5(5):e10890.
Bertelli C, et al. 2015. Sequencing and characterizing the genome of *Estrella lausannensis* as an undergraduate project: training students and biological insights. Front Microbiol. 6:101.
Bertelli C, et al. 2016. CRISPR system acquisition and evolution of an obligate intracellular *Chlamydia*-related bacterium. Genome Biol Evol. 8(8):2376–2386.
Bertelli C, Goesmann A, Greub G. 2014. Criblamydia sequanensis Harbors a Megaplasmid Encoding Arsenite Resistance. Genome Announc 2(5):e00949–14.
Blandford MI, Taylor-Brown A, Schlacher TA, Nowak B, Polkinghorne A. 2018. Epitheliocystis in fish: an emerging aquaculture disease with a global impact. Transbound Emerg Dis. doi: 10.1111/tbed.12908.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.
Bourne DG, Morrow KM, Webster NS. 2016. Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. Annu Rev Microbiol. 70:317–340.
Burnard D, et al. 2017. Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasit Vect. 10(1):46.
Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
Casadevall A. 2008. Evolution of intracellular pathogens. Annu Rev Microbiol. 62:19–33.
Christiansen MT, et al. 2014. Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. BMC Infect Dis. 14(1):591.
Collingro A, et al. 2011. Unity in variety – the pan-genome of the *Chlamydiae*. Mol Biol Evol. 28(12):3253–3270.
Collingro A, et al. 2017. Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. ISME J. 11(10):2334–2344.
Corsaro D, Greub G. 2006. Pathogenic potential of novel *Chlamydiae* and diagnostic approaches to infections due to these obligate intracellular bacteria. Clin Microbiol Rev. 19(2):283–297.
Crespo S, Zarza C, Padros F, Marin de Mateo M. 1999. Epitheliocystis agents in sea bream *Sparus aurata*: morphological evidence for two distinct chlamydia-like developmental cycles. Dis Aquat Organ. 37:61–72.
Croxatto A, et al. 2014. Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks Tick Borne Dis. 5(4):359–365.
Domman D, et al. 2014. Massive expansion of ubiquitination-related gene families within the *Chlamydiae*. Mol Biol Evol. 31(11):2890–904.
Driscoll TP, et al. 2017. Wholly *Rickettsia*! Reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. MBio 8(5):e00859–e00817.
Dugan J, Rockey DD, Jones L, Andersen AA. 2004. Tetracycline resistance in *Chlamydia suis* mediated by genomic islands inserted into the chlamydial inv-like gene. Antimicrob Agents Chemother. 48(10):3989–3995.
Dupont CL, et al. 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J. 6:1186–1199.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16(1):157.
Evans DH, Piermarini PM, Choe KP. 2005. The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. Physiol Rev. 85(1):97–177.
Finn RD, et al. 2015. HMMER web server: 2015 update. Nucleic Acids Res. 43(W1):W30–W38.
Fisher DJ, Fernandez RE, Maurelli AT. 2013. *Chlamydia trachomatis* transports NAD via the Npt1 ATP/ADP translocase. J Bacteriol. 195(15):3381–3386.
Fraser CM, et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390(6660):580–586.
Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(D1):D261–D269.
Gil R, Silva FJ, Pereto J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev. 68(3):518–537.
Gimenez G, et al. 2011. Insight into cross-talk between intra-amoebal pathogens. BMC Genomics. 12:542.
Greub G, et al. 2009. High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PLoS ONE 4(12):e8423.
Greub G, Raoult D. 2003. History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a *Chlamydiales* ancestor to plants 1 billion years ago. Appl Environ Microbiol. 69(9):5530–5535.
Greub G. 2009. *Parachlamydia acanthamoebae*, an emerging agent of pneumonia. Clin Microbiol Infect. 15(1):18–28.
Guevara Soto M, et al. 2016. The emergence of epitheliocystis in the upper Rhone region: evidence for *Chlamydiae* in wild and farmed salmonid populations. Arch Microbiol. 198(4):315–324.
Gupta RS, Naushad S, Chokshi C, Griffiths E, Adeolu M. 2015. A phylogenomic and molecular markers based analysis of the phylum *Chlamydiae*: proposal to divide the class Chlamydiia into two orders, *Chlamydiales* and *Parachlamydiales* ord. nov., and emended description of the class Chlamydiia. Antonie Van Leeuwenhoek. 108(3):765–781.
Haferkamp I, et al. 2004. A candidate NAD+ transporter in an intracellular bacterial symbiont related to *Chlamydiae*. Nature 432(7017):622–625.
Haferkamp I, et al. 2006. Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. Mol Microbiol. 60(6):1534–1545.
Heinz E, et al. 2014. Plasma membrane-located purine nucleotide transport proteins are key components for host exploitation by microsporidian intracellular parasites. PLoS Pathog. 10(12):e1004547.
Himmelreich R, et al. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res. 24(22):4420–4449.
Hokynar K, et al. 2016. C*hlamydia*-like organisms (CLOs) in finnish ixodes ricinus ticks and human skin. Microorganisms 4(3):28.
Horn M, et al. 2004. Illuminating the evolutionary history of chlamydiae. Science 304(5671):728–730.
Hutchison CA, et al. 2016. Design and synthesis of a minimal bacterial genome. Science 351(6280):aad6253.
Ip YK, Chew SF. 2010. Ammonia production, excretion, toxicity, and defense in fish: a review. Front Physiol. 1:134.
Kalman S, et al. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat Genet. 21(4):385–389.
Kamneva OK, Knight SJ, Liberles DA, Ward NL. 2012. Analysis of genome content evolution in PVC bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. Genome Biol Evol. 4(12):1375–1390.
Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1):27–30.
Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: kEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 428(4):726–731.
Katharios P, et al. 2015. Environmental marine pathogen isolation using mesocosm culture of sharpsnout seabream: striking genomic and

morphological features of novel *Endozoicomonas* sp. Sci Rep. 5(1):17609.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.
Konig L, et al. 2017. Biphasic metabolism and host interaction of a chlamydial symbiont. mSystems 2:e00202–16.
Koonin EV. 2000. How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet. 1:99–116.
Lagkouvardos I, et al. 2014. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. ISME J. 8(1):115–125.
Lai CC, Crosbie PBB, Battaglene SC, Nowak BF. 2013. Effects of epitheliocystis on serum lysozyme activity and osmoregulation in cultured juvenile striped trumpeter, *Latris lineata* (Forster). Aquaculture 388–391:99–104.
Lamoth F, Greub G. 2010. Amoebal pathogens as emerging causal agents of pneumonia. FEMS Microbiol Rev. 34(3):260–280.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26(5):589–595.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Marchler-Bauer A, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45(D1):D200–D203.
Mehlitz A, et al. 2017. Metabolic adaptation of *Chlamydia trachomatis* to mammalian host cells. Mol Microbiol. 103(6):1004–1019.
Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct. 4(1):13.
Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17(10):589–596.
Mojica S, et al. 2011. Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58. J Bacteriol. 193(14):3690.
Moulder JW. 1974. Intracellular parasitism: life in an extreme environment. J Infect Dis. 130(3):300–306.
Mushegian A. 1999. The minimal genome concept. Curr Opin Genet Dev. 9(6):709–714.
Mushegian AR, Koonin EV. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A. 93(19):10268–10273.
Myllykallio H, et al. 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. Science 297(5578):105–107.
Nishida H. 2012. Evolution of genome base composition and genome size in bacteria. Front Microbiol. 3:420.
Nowak BF, LaPatra SE. 2006. Epitheliocystis in fish. J Fish Dis. 29(10):573–588.
Nunes A, Gomes JP. 2014. Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. Infect Genet Evol. 23:49–64.
Nylund A, et al. 2018. Genotyping of *Candidatus* Syngnamydia salmonis (*Chlamydiales; Simkaniaceae*) co-cultured in *Paramoeba perurans* (*Amoebozoa; Paramoebidae*). Arch Microbiol 200(6):859–867.
Nylund A, Kvenseth AM, Isdal E. 1998. A morphological study of the epitheliocystis agent in farmed atlantic salmon. J Aquat Anim Health. 10(1):43–55.
Omsland A, Sixt BS, Horn M, Hackstadt T. 2014. Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. FEMS Microbiol Rev. 38(4):779–801.
Pillonel T, Bertelli C, Greub G. 2018. Environmental metagenomic assemblies reveal seven new highly divergent chlamydial lineages and hallmarks of a conserved intracellular lifestyle. Front Microbiol. 9(79):doi: 10.3389/fmicb.2018.00079. eCollection 2018.
Pillonel T, Bertelli C, Salamin N, Greub G. 2015. Taxogenomics of the order *Chlamydiales*. Int J Syst Evol Microbiol. 65(Pt 4):1381–1393.
Pilloux L, et al. 2015. The high prevalence and diversity of *Chlamydiales* DNA within Ixodes ricinus ticks suggest a role for ticks as reservoirs and vectors of *Chlamydia*-related bacteria. Appl Environ Microbiol. 81(23):8177–8182.
Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 26(7):1641–1650.
Psomopoulos FE, et al. 2012. The *Chlamydiales* pangenome revisited: structural stability and functional coherence. Genes (Basel) 3(2):291–319.
Putman TE, Suchland RJ, Ivanovitch JD, Rockey DD. 2013. Culture-independent sequence analysis of *Chlamydia trachomatis* in urogenital specimens identifies regions of recombination and in-patient sequence mutations. Microbiology 159(Pt_10):2109–2117.
Qi W, Vaughan L, Katharios P, Schlapbach R, Seth-Smith HM. 2016. Host-associated genomic features of the novel uncultured intracellular pathogen *Ca.* ichthyocystis revealed by direct sequencing of epitheliocysts. Genome Biol Evol. 8(6):1672–1689.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.
Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10):944–945.
Sachse K, et al. 2014. Evidence for the existence of two new members of the family *Chlamydiaceae* and proposal of *Chlamydia avium* sp. nov. and *Chlamydia gallinacea* sp. nov. Syst Appl Microbiol. 37(2):79–88.
Sakharkar KR, Dhar PK, Chow VT. 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. Int J Syst Evol Microbiol. 54(6):1937–1941.
Schwoppe C, Winkler HH, Neuhaus HE. 2002. Properties of the glucose-6-phosphate transporter from *Chlamydia pneumoniae* (HPTcp) and the glucose-6-phosphate sensor from *Escherichia coli* (UhpC). J Bacteriol. 184(8):2108–2115.
Seth-Smith HM, et al. 2016. Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular beta-proteobacteria. ISME J. 10(7):1791–1803.
Steigen A, et al. 2013. *Cand*. Actinochlamydia clariae' gen. nov., sp. nov., a unique intracellular bacterium causing epitheliocystis in catfish (*Clarias gariepinus*) in Uganda. PLoS One 8(6):e66840.
Steigen A, et al. 2015. A new intracellular bacterium, *Candidatus* Similichlamydia labri sp. nov. (*Chlamydiaceae*) producing epitheliocysts in ballan wrasse, *Labrus bergylta* (Pisces, Labridae). Arch Microbiol. 197(2):311–318.
Stephens RS, et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: *chlamydia trachomatis*. Science 282(5389):754–759.
Stride MC, Polkinghorne A, Powell MD, et al. 2013. "*Candidatus* Similichlamydia laticola", a novel *Chlamydia*-like agent of epitheliocystis in seven consecutive cohorts of farmed Australian barramundi, *Lates calcarifer* (Bloch). PLoS One 8(12):e82889.
Stride MC, Polkinghorne A, Miller TL, et al. 2013. Molecular characterization of "*Candidatus* Parilichlamydia carangidicola," a novel Chlamydia-like epitheliocystis agent in yellowtail kingfish, *Seriola lalandi* (Valenciennes), and the proposal of a new family, "*Candidatus* Parilichlamydiaceae" fam. nov. (order *Chlamydiales*). Appl Environ Microbiol. 79(5):1590–1597.
Stride MC, Polkinghorne A, Miller TL, Nowak BF. 2013. Molecular characterization of "*Candidatus* Similichlamydia latridicola" gen. nov., sp. nov. (*Chlamydiales*: "*Candidatus* Parilichlamydiaceae"), a novel *Chlamydia*-like epitheliocystis agent in the striped trumpeter, *Latris lineata* (Forster). Appl Environ Microbiol. 79(16):4914–4920.

Stride MC, Polkinghome A, Nowak BF. 2014. Chlamydial infections of fish: diverse pathogens and emerging causes of disease in aquaculture species. Vet Microbiol. 171(1–2):258–266.
Taylor-Brown A, Bachmann NL, Borel N, Polkinghorne A. 2016. Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes. BMC Genomics. 17(1):710.
Taylor-Brown A, Spang L, Borel N, Polkinghorne A. 2017. Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus *Chlamydia*. Sci Rep. 7:e10661
Taylor-Brown A, Pillonel T, et al. 2017. Culture-independent genomics of a novel chlamydial pathogen of fish provides new insight into host-specific adaptations utilized by these intracellular bacteria. Environ Microbiol. 19(5):1899–1913.
Taylor-Brown A, Madden D, Polkinghorne A. 2018. Culture-independent approaches to chlamydial genomics. Microb Genom. doi: 10.1099/mgen.0.000145.
Taylor-Brown A, Vaughan L, Greub G, Timms P, Polkinghorne A. 2015. Twenty years of research into *Chlamydia*-like organisms: a revolution in our understanding of the biology and pathogenicity of members of the phylum *Chlamydiae*. Pathog Dis. 73(1):1–15.
Tjaden J, et al. 1999. Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for transport of energy. J Bacteriol. 181(4):1196–1202.
Vasilevsky S, Stojanov M, Greub G, Baud D. 2016. Chlamydial polymorphic membrane proteins: regulation, function and potential vaccine candidates. Virulence 7(1):11–22.
Viver T, et al. 2017. The low diverse gastric microbiome of the jellyfish Cotylorhiza tuberculata is dominated by four novel taxa. Environ Microbiol 19(8):3039–3058.
Voigt A, Schöfl G, Heidrich A, Sachse K, Saluz HP. 2011. Full-length de novo sequence of the Chlamydophila psittaci type strain, 6BC. J Bacteriol. 193(10):2662–2663.
Voigt A, Schofl G, Saluz HP. 2012. The Chlamydia psittaci genome: a comparative analysis of intracellular pathogens. PLoS One 7(4):e35097.
Vorimore F, et al. 2013. Isolation of a new chlamydia species from the feral sacred ibis (*Threskiornis aethiopicus*): chlamydia ibidis. PLoS One 8(9):e74823.
Webster NS, Thomas T. 2016. The sponge hologenome. MBio 7(2):e00135–e00116.
Wirz M, et al. 2008. Predator or prey? Chlamydophila abortus infections of a free-living amoebae, *Acanthamoeba castellani* 9GU. Microbes Infect. 10(6):591–597.
Work TM, Aeby GS. 2014. Microbial aggregates within tissues infect a diversity of corals throughout the Indo-Pacific. Mar Ecol Prog Ser. 500:1–9.
Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32(4):605–607.
Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26.
Ye YN, et al. 2016. A novel proposal of a simplified bacterial gene set and the neo-construction of a general minimized metabolic network. Sci Rep. 6:35082.
Zomorodipour A, Andersson SG. 1999. Obligate intracellular parasites: *rickettsia prowazekii* and *Chlamydia trachomatis*. FEBS Lett. 452(1–2):11–15.

Associate editor: Tal Dagan