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

Alyce Taylor Brown1, Trestan Pillonel2, Gilbert Greub2, Lloyd Vaughan3,4, Barbara Nowak5, and Adam Polkinghorne1,*

1USC Animal Research Centre, Faculty of Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland, Australia
2Institute of Microbiology, University of Lausanne and University Hospital Center, Switzerland
3Institute of Veterinary Pathology, University of Zurich, Switzerland
4Pathovet AG, Tagelswangen, Switzerland
5Institute 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 Parachlamydiaceae 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 Chlamydia (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. carangidica, Ca. S. laticolca, Ca. S. latridicola), their land vertebrate restricted cousins, the Chlamydiaceae (Chlamydia trachomatis) and the more ubiquitous or promiscuous Parachlamydiaceae (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; Sakharov et al. 2004; Hutchison et al. 2016; Ye et al. 2016); 3) 208 genes conserved between P. amoebophila, C. trachomatis, Legionella pneumophila, Coxiella burnetii, and Rhodopirellula baltica (Gimenez et al. 2011). These genes were identified using MaxBin, BLASTp, BLASTkoala, CD-search (Aziz et al. 2008), CD-search (expect value threshold of 0.01) (Marchler-Bauer et al. 2017), and Hmmmscan (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), Hmmmscan (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, elfA, sigA, sufB, sufS, secA, htrA, ileS, tuf, valS, pgsA, 1poA, cysS, priA, gmk, EngD, fusA, serS, rluD, pgi, gatB, mhb, 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–Thornton 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. carangidica, 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. carangidica, 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. carangidica, Ca. S. latridicola and Ca. S. laticola, respectively (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
varied, with 49×, 46×, and 692× coverage for Ca. P. carangidica, 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 1
Details of Gill Samples Utilized for Metagenome Sequencing and Resulting Metagenome and Genome Assemblies

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gill Sample Details</th>
<th>Host</th>
<th>Collection date</th>
<th>Collection location</th>
<th>Chlamydial species</th>
<th>Initial metagenome Assembly metrics</th>
<th>Chlamydial genome assembly metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample 76ST10</td>
<td>Striped Trumpeter (Latris lineata)</td>
<td>Nov 2010</td>
<td>Tasmania</td>
<td>Ca. Similichlamydia latridicola</td>
<td>16,776,972×</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Hat2</td>
<td>Barramundi (Lates califerer)</td>
<td>Aug 2012</td>
<td>South Australia</td>
<td>Ca. Similichlamydia latricola</td>
<td>40,539,772×</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>007YTK11</td>
<td>Yellowtail Kingfish (Seriola lalandi)</td>
<td>May 2011</td>
<td>South Australia</td>
<td>Ca. Parilichlamydia carangidicola</td>
<td>29,405,932×</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gill Sample Details</th>
<th>Host</th>
<th>Collection date</th>
<th>Collection location</th>
<th>Chlamydial species</th>
<th>Initial metagenome Assembly metrics</th>
<th>Chlamydial genome assembly metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>76ST10</td>
<td>Striped Trumpeter (Latris lineata)</td>
<td>Nov 2010</td>
<td>Tasmania</td>
<td>Ca. Similichlamydia latridicola</td>
<td>16,776,972×</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Hat2</td>
<td>Barramundi (Lates califerer)</td>
<td>Aug 2012</td>
<td>South Australia</td>
<td>Ca. Similichlamydia latricola</td>
<td>40,539,772×</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>007YTK11</td>
<td>Yellowtail Kingfish (Seriola lalandi)</td>
<td>May 2011</td>
<td>South Australia</td>
<td>Ca. Parilichlamydia carangidicola</td>
<td>29,405,932×</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2
Genome Characteristics for Members of the Phylum Chlamydiae

<table>
<thead>
<tr>
<th>Family (No. Genomes Used for Comparison)</th>
<th>Chromosome Size (Mbp)</th>
<th>No. Predicted CDS</th>
<th>Plasmid Size (kbp)</th>
<th>G + C Content (%)</th>
<th>No. rRNA Operons</th>
<th>No. tRNAs</th>
<th>Coding Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca. Parilichlamydiaceae (4)</td>
<td>0.771 0.981</td>
<td>704 943</td>
<td>No evidence</td>
<td>37.5 43.7</td>
<td>1</td>
<td>41 42</td>
<td>90.2 92.4</td>
</tr>
<tr>
<td>Parachlamydiaceae (9)</td>
<td>2.417 3.424</td>
<td>1,986 3,042</td>
<td>39 145 (40 160)</td>
<td>33.0 44.0</td>
<td>14</td>
<td>35 42</td>
<td>69.7 90.8</td>
</tr>
<tr>
<td>Waddliaceae (2)</td>
<td>2.116 2.141</td>
<td>1,934 2,028</td>
<td>15.5 (22)</td>
<td>43</td>
<td>2</td>
<td>34 37</td>
<td>92.2 93</td>
</tr>
<tr>
<td>Simkaniaceae (1)</td>
<td>2.496</td>
<td>2,519</td>
<td>132 (136)</td>
<td>38</td>
<td>1</td>
<td>35</td>
<td>91.3</td>
</tr>
<tr>
<td>Criblamydiaceae (2)</td>
<td>2.820 2.969</td>
<td>2,213 2,426</td>
<td>9.1 (15) 89.5 (92)</td>
<td>38.0 48.0</td>
<td>1</td>
<td>40</td>
<td>86.0 89.7</td>
</tr>
<tr>
<td>Chlamydiaceae (5)</td>
<td>1.041 1.229</td>
<td>911 1,112</td>
<td>7.5 (8)</td>
<td>37.4 41.6</td>
<td>1 2</td>
<td>37 39</td>
<td>89.0 91.2</td>
</tr>
</tbody>
</table>

Nons. Ranges of values are given for each family, with the number of species or strains used for comparison in brackets.

*W. chondrophila* plasmid in WSU 86 1044 strain only.

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.

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%.
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. 1a). All known members of the Chlamydiaceae and Ca. Parilichlamydiaceae, two distantly related families (fig. 1b), exhibit significantly smaller genomes compared with the Parachlamydiaceae (table 2, fig. 1a). 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. Parachlamydiaceae 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 Chlamydiaceae 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 amoeba 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...
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. 2a). This high level of genomic synteny observed in the Chlamydiae and Akkermansia muciniphila was used as an outgroup species to root the tree. The tree was constructed by FastTree with a Jones Taylor Thornton 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, ropA, cysS, prA, gmk, EngD, fusA, serS, rluD, pgi, gagB, mbhB, trub, ribA, infB, nusA, and rpo). 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).

Updated Chlamydiae Core Genome and Ca. Parilichlamydiaceae-Specific Genes

The reduced genome size of the Ca. Parilichlamydiaceae has a strong impact on the size of the Chlamydiae 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. 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 Thornton 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, ropA, cysS, prA, gmk, EngD, fusA, serS, rluD, pgi, gagB, mbhB, trub, ribA, infB, nusA, and rpo). 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).
60% and 80% for Parachlamydiales and Chlamydiaceae (fig. 1d). 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. 1c). 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 “Chlamyophila” 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; Vasiilevsky et al. 2016) and the ubiquitin ligase-like protein expansion in the Neochlamydia spp. (Domman et al. 2014) (fig. 1a, 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. S1b, 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 (mmnE and mmnG) 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 Chlamydiae 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. P. amoebophila and A. muciniphila both possess prsA, 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 phosphoribosyltransferase; 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 polABC), 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 synthase (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 (Mylykallio 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), so they could be obtained by the bacteria as purines, taurines, creatine and methylamines (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

Parachlamydiaceae are capable of synthesizing several amino acids whilst Chlamydiae 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.
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 Chlamydiaceae, 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 Chlamydiaceae, 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 Parachlamydiaceae and Chlamydiaceae. Ca. Parilichlamydiaceae appear to rely on diphosphate-dependent phosphofructokinase (pfpB) to convert glucose 6-phosphate and fructose 2, 6-biphosphate, like Parachlamydiaceae, 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 phosphoenolpyruvate by pyruvate kinase (pyk), Ca. Parilichlamydiaceae genomes appear to lack the pyruvate dehydrogenase subunits (aceEF) for acetyl-CoA generation from pyruvate encoded by other Chlamydiaceae. 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 Parachlamydiaceae 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 Parachlamydiaceae 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 from 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 Chlamydiaceae 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 (fodD) participates in methylene-THF formation. Again, this is pertinent to the Ca. Parilichlamydiaceae infecting gut 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 Chlamydiaceae residing within its inclusion. Ca. Parilichlamydiaceae species, and other Chlamydiaceae, 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 Chlamydiaceae. 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-[acylcarrier-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 phosphatidylglycerophosphate; however, Chlamydiaceae do not possess the phosphatidylglycerophosphatase to convert this to phosphatidylglycerol, 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 bio-synthetic 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 Parachlamydiaceae 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 Parachlamydiaceae and the Chlamydiaceae. For example, predicted phospholipid transporters (mlaDEF) are encoded by Parachlamydiales only, whereas more predicted lipopolysaccharide transporters are encoded by Parachlamydiaceae (rfbAB and lptFGB) than Chlamydiaceae (lpt only) but not by Ca. Parilichlamydiaceae. Suspected lipoprotein transporters (llo) 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 Parachlamydiaceae 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 Chlamydia 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 Chlamydomonas (P. amoebophila NTT4 is highly divergent). Ca. Parilichlamydiaceae NTT1 shares significant sequence identity with Npt1 in Chlamydomonas 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, Chlamydomonas NTT1 has been termed a “hybrid” class VIII 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 VIII 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 Waddliaeae (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 Chlamydiae, 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 Simkania 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. perurans 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 Chlamydiaceae (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 Aebly 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 Chlamydiaceae, 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 Chlamydiaceae 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 Parachlamydiaceae clade or convergent genome reduction in Chlamydiaceae and Parilichlamydiaceae. 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 Parachlamydiaceae 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 Chlamydiaceae 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


Katoch K, Standley DM. 2013. MAFFT multiple sequence alignment soft
ware version 7: improvements in performance and usability. Mol Biol
Evol. 30(4):772 780.
desktop software platform for the organization and analysis of se
mydial symbiont. mSystems 2:e00202 16.
Koonin EV. 2000. How many genes can make a cell: the minimal gene set
Lagkouvardos I, et al. 2014. Integrating metagenomic and amplicon data
bases to resolve the phylogenetic and ecological diversity of the
Lai CC, Crosbie PBB, Battagline SC, Novak BF. 2013. Effects of epithe
lyicosistion on serum lysozyme activity and osmoregulation in cultured
juvenile striped trumpetfish, Latris lineata (Forster). Aquaculture 388
391:99 104.
Lamoth F, Greub G. 2010. Amoebal pathogens as emerging causal agents
Li H, Durbin R. 2010. Fast and accurate long read alignment with
Bioinformatics 25(16):2078 2079.
proteins by subfamily domain architectures. Nucleic Acids Res.
45(1):D200 D203.
parative genomic analysis reveals convergent evolution of specialized
Mira A, Ochman H, Moran NA. 2001. Deleterial bias and the evolution of
Mojca S, et al. 2011. Genome sequence of the obligate intracellular an
Moulder JW. 1974. Intracellular parasitism: life in an extreme environ-
9(6):709 714.
Mushegian AR, Koonin EV. 1996. A minimal gene set for cellular life de
rived by comparison of complete bacterial genomes. Proc Natl Acad
Nylund A, Gomes JP. 2014. Evolution, phylogeny, and molecular epidemi-
ology of Chlamydia. Infect Genet Evol. 23:49 64.
(Chlamydiaceae; Simkaniaceae) co cultured in Paramoeba perurans
epitheliocyst agent in farmed Atlantic salmon: J Aquat Anim Health.
10(1):43 55.
Omsland A, Sixt BS, Horn M, Hackstadt T. 2014. Chlamydial metabo-
лизm revisited: interspecies metabolic variability and developmental
Pillowen T, Bertelli C, Greub G. 2018. Environmental metagenomic assem-
bles reveal seven new highly divergent chlamydial lineages and
Pilloux L, et al. 2015. The high prevalence and diversity of Chlamydi-
dae DNA within bodes ricinius ticks suggest a role for ticks as reservoirs and
81(23):8177 8182.
Price MN, Dehal PS, Arkin DD. 2009. FastTree: computing large minimum
26(7):1641 1650.
Psomopoulos FE, et al. 2012. The Chlamydiaceae pangenome revisited:
Putman TE, Suchland RJ, Ivanovitch JD, Rockey DD. 2013. Culture inde-
pendent sequence analysis of Chlamydia trachomatis in urogenital spec-
imens identifies regions of recombination and in patient sequence
mutations. Microbiology 159(Pt 10):2109 2117.
associated genomic features of the novel uncultured intracellular path-
gen 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 com-
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.
obligatory intracellular parasites of humans: a comparative analysis. Int
phosphate transporter from Chlamydia pneumoniae (HPTcp) and the
184(8):2108 2115.
Seth Smith HM, et al. 2016. Emerging pathogens of gillhead seabream:
characterisation and genomic analysis of novel intracellular beta pro-
tobacteria. ISME J. 10(7):1791 1803.
a unique intracellular bacterium causing epitheliocysts in catfish
Steigen A, et al. 2015. A new intracellular bacterium, Candidatus Similichlamydia labri sp. nov. (Chlamydiaceae) producing epitheliocysts in
ballan wrasse, Labrus bergylta (Serranidae). Arch Microbiol.
197(2):311 318.
Stepsns RS, et al. 1998. Genome sequence of an obligate intracellular
cysts in seven consecutive cohorts of farmed Australian barramundi,
tion of “Candidatus Parichlamydia carangicola,” a novel Chlamydia like epitheliocyst agent in yellowtail kingfish, Seriola lalandi (Valenciennes), and the proposal of a new family,
acterization of “Candidatus Similichlamydia latricola” gen. nov., sp.
nov. (Chlamydiaceae: “Candidatus Parichlamydiaceae”), a novel
Chlamydia like epitheliocyst agent in the striped trumpetfish, Latris

Comparative genomics of Ca. Parilichlamydiaceae