Utilising genomics of novel uncultivable Chlamydiae to further understand chlamydial diversity, pathogenesis and evolution

Taylor-Brown, Alyce

Taylor-Brown, A. (2018). Utilising genomics of novel uncultivable Chlamydiae to further understand chlamydial diversity, pathogenesis and evolution [University of the Sunshine Coast, Queensland].
https://doi.org/10.25907/00246
Document Type: Thesis

Utilising genomics of novel uncultivable *Chlamydiae* to further understand chlamydial diversity, pathogenesis and evolution

Alyce Taylor-Brown

B. BiomedSci (Hons)

Submitted for the degree of Doctor of Philosophy

to the

Faculty of Science, Health, Engineering and Education,

University of the Sunshine Coast

December 2017

Supervisors:

Prof. Adam Polkinghorne (University of the Sunshine Coast)

Prof. Peter Timms (University of the Sunshine Coast)

Prof. Barbara Nowak (University of Tasmania)

Table of contents

Utilising genomics of novel uncultivable *Chlamydiae* to further understand chlamydial diversity, pathogenesis and evolution

Table of contents 1

Abstract 9

List of abbreviations 11

List of publications and conference presentations 13

Acknowledgements 15

CHAPTER 1: Literature review and introduction 17

1. *Chlamydia* and *Chlamydia*-related bacteria 17

Figure 1: Phylogenetic tree of the Chlamydiales including abundance of each clade in public databases 17

1.1. Unique biological features of chlamydiae (EBs, RBs and the inclusion) 18

Chlamydial developmental cycle 18

Figure 2: Chlamydial developmental cycle. 18

Type three secretion system 20

The chlamydial inclusion 20

1.2. Host diversity and disease spectrum of the *Chlamydiaceae:* a broad overview 21

Chlamydial pathogens of humans and zoonotic pathogens 21

Chlamydial pathogens of birds 23

Chlamydial pathogens of reptiles and amphibians 24

Chlamydial pathogens of other mammals 25

Emerging chlamydial pathogens and their hosts 27

1.3. Phylogenetic and ecological diversity of *Chlamydia*-related bacteria 28

Figure 3: Significant events in the detection, isolation and identification of CRBs over the last 20 years 29

Chlamydial endosymbionts of amoebae (and their pathogenic potential) 30

Chlamydial endosymbionts of insects 32

Chlamydial pathogens of fish 33

CRBs as human and mammalian pathogens 35

Uncultured and unclassified clades represent a wealth of uncovered diversity within the Chlamydiae 38

2. Genomics of *Chlamydiae*..... 39

2.1. Chlamydia genomes 39

Figure 4: Shared genomic synteny within *Chlamydia* genomes 39

Size and reduced metabolic capacity..... 40

Table 1: Genome characteristics of members of the *Chlamydiales*..... 42

Plasticity zone 43

Figure 5: Diversity in the chlamydial plasticity zon 43

Peptidoglycan genes..... 44

Chlamydia-specific virulence genes..... 45

Genome-based phylogenetic analyses 46

2.2. CRB genomes and the pan-*Chlamydiales* genome 46

Size and expanded metabolic capacity..... 47

Pan-*Chlamydiales* genome..... 48

Figure 6: Pan- and core-*Chlamydiales* genome 49

Family, genus and species-specific genes 49

2.3. Genetic exchange 51

Recombination..... 51

Genomic islands 52

Extrachromosomal elements 54

Figure 7: ORF arrangement in the *Chlamydia* plasmid 55

3. Culture-independent approaches for genomic characterisation of *Chlamydiae*..... 57

3.1. Culture-independent genomics of *Chlamydiae*..... 58

Immunomagnetic-separation – multiple displacement amplification..... 58

RNA 'bait'-DNA hybridisation 60

Metagenomic approaches 60

Table 2: Summary of culture-independent genome sequencing methodologies used for *Chlamydiae*..... 62

4. Lessons from the literature and concluding remarks..... 65

5. References 67

CHAPTER 2: Aims, significance and project development 88

References 93

CHAPTER 3: Characterisation of *Chlamydia pneumoniae* and other novel chlamydial infections in captive snakes (Veterinary Microbiology 178, 2015) 97

3.1 Abstract 98

3.2 Introduction 99

3.3 Methods 100

3.3.1 Snake collections 100

3.3.2 Sample collection and DNA extraction 100

3.3.3 Chlamydial screening 100

3.3.4 Phylogenetic analysis 101

3.4 Results 101

3.4.1 Detection of chlamydiae in captive snakes 101

3.4.2 Chlamydial identity by PCR and sequencing 102

3.4.3 Phylogenetic analysis 103

3.4.4 Genotype diversity 104

3.5 Discussion 104

Declarations 107

Chapter 3 Tables & Figures 108

Table 3.1a: Presence of chlamydial infection in captive snakes from six collections in Switzerland. 108

Table 3.1b: Presence of chlamydial infections in six collections of captive snakes per anatomical site and diagnostic method. 109

Figure 3.1: Phylogenetic tree depicting relationships between novel genotypes identified in this study and related chlamydial sequences obtained from Genbank. 114

References 115

CHAPTER 4: Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes (BMC Genomics 17, 2016) .. 117

4.1 Abstract 118

4.2 Background 119
4.3 Results and discussion 120
4.3.1 Description of "Candidatus Chlamydia sanzinia", sp. nov. 120
4.3.2. Metagenome reconstruction and chlamydial genome assembly 120
4.3.3 Ca. Chlamydia sanzinia 2742-308 forms a distinct clade with C. pneumoniae and C. pecorum 121
4.3.4 The genome of Ca. Chlamydia sanzinia 2742-308 harbours a 7.5 Kbp cryptic plasmid 122
4.3.5 The genome of Ca. Chlamydia sanzinia 2742-308 does not appear to contain a plasticity zone 122
4.3.6 An expanded repertoire of polymorphic membrane proteins are located on "pmp islands" for Ca. Chlamydia sanzinia 2742-308 124
4.3.7 Ca. Chlamydia sanzinia 2742-308 displays genomic hallmarks of a pathogen 125
4.4 Conclusion 126
4.5 Methods 127
4.5.1 DNA extraction and microbial DNA enrichment 127
4.5.2 Genome sequencing and assembly 127
4.5.3 Genome annotation and analysis 127
4.5.4 Phylogenetic analysis 128
4.5.5 Inclusion membrane protein prediction 128
Declarations 128
Chapter 4 Tables & Figures 130
Table 4.1: Comparative analysis of chlamydial genome features. 130
Figure 4.1: *Chlamydia* species core genome phylogenetic tree. 132
Figure 4.3: Lack of the plasticity zone in *Chlamydia sp.* 2742-308 133
References 134
CHAPTER 5: Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus *Chlamydia* (Scientific Reports 7, 2017) 138
5.1 Abstract 139
5.2 Introduction 140
5.3 Results and discussion 141
5.3.1 Snake choana and cloaca metagenome assembly and microbial composition 141
5.3.2 Chlamydial genome construction from a snake choana metagenome 142

5.3.3 Phylogenomic analysis of *Ca.* Chlamydia corallus within the *Chlamydiaceae* 143
5.3.4 Plasmid-based diversity within chlamydial species infecting snakes 144
5.3.5 Genetic differences within the plasticity zone of *Ca.* Chlamydia corallus 146
5.4 Conclusions 148
5.5 Materials and methods 149
5.5.1 Sample preparation 149
5.5.2 Ethics approval and consent to participate 149
5.5.3 Metagenome assembly and analysis 149
5.5.4 Phylogenetic analysis 150
5.5.5 Data availability 150
Declarations 151
Chapter 5 Tables & Figures 152
Table 5.1: Microbial composition of choana and cloaca samples from captive snakes 152
Table 5.2: Genome characteristics of Uncultured *Chlamydia sp.* G3/2741-324 compared to closely-related chlamydial species. 153
Table 5.3: Characteristics of chlamydial plasmids obtained from snake choana and cloaca metagenomes 154
Figure 5.1: Phylogenetic position of the novel taxon, *Ca.* Chlamydia corallus G3/2742-324 within the *Chlamydiaceae*. 155
Figure 5.2: Plasmid sequence-based phylogenetic analysis of all samples sequenced in this study and other chlamydial plasmid sequences. 156
Figure 5.3: Comparison of plasticity zone regions encoded by *Ca.* Chlamydia corallus and related chlamydial species. 157
References 158
CHAPTER 6: Culture-independent genomics of a novel chlamydial pathogen of fish provides new insight into host specific adaptations utilized by these intracellular bacteria (Environmental Microbiology 19, 2017) 162
6.1 Abstract 163
6.2 Introduction 164
6.3 Results and discussion 166
6.3.1 Detection of epitheliocystis in cultured *Epinephelus coioides* 166

6.3.2 Morphological description of *Ca.* Similichlamydia epinephelii cysts in E. coioides 168
6.3.3 Identification of the chlamydial agent of epitheliocystis in E. coioides gill samples by PCR and ISH . 168
6.3.4 Proposal of sp. nov. *Ca.* Similichlamydia epinephelii 169
6.3.5 Genomic analysis of *Ca.* Similichlamydia epinephelii 170
6.3.6 Updated phylogenomics of *Ca.* S. epinephelii within the order *Chlamydiales* 172
6.3.7 The genome of Ca. S. epinephelii provides evidence for the ancient acquisition of a chlamydial Type III secretion system (T3SS) 173
6.3.8 Genomic evidence of the pathogenic potential of *Ca.* S. epinephelii 174
6.3.9 Predicting novel pathogenic proteins in the genome of *Ca.* S. epinephelii 176
6.4 Conclusion 176
6.5 Experimental Procedures 177
6.5.1 Ethics statement 177
6.5.2 Sample collection 177
6.5.3 DNA extraction 178
6.5.4 PCR, sequencing and phylogenetic analysis 178
6.5.5 Histology and in situ hybridisation 179
6.5.6 Genome sequencing and assembly 180
6.5.7 Genome analysis 180
Declarations 181
Chapter 6 Tables & Figures 182
Figure 1: Epitheliocystis in Orange-spotted grouper (*E. coioides*) gills stained with hematoxylin and eosin (a, b) and chlamydial localisation to cysts shown by in situ hybridisation (c, d, e). 185
Figure 2: Pan-genome analysis. 186
Figure 3: Phylogenetic tree constructed from concatenated amino acid sequences corresponding to sixty-four single copy orthologs in all publicly available chlamydial genomes/assemblies. 187
Figure 4: Conservation of three out of the four chlamydial T3SS clusters. 188
References 189
CHAPTER 7: Comparative genomics of the *Candidatus* Parilichlamydiaceae reveals evidence of convergent evolution and genome reduction in the phylum *Chlamydiae* ... 196
7.1 Abstract 197

7.2 Introduction 198
7.3 Materials and methods 200
7.3.1 Sample collection and DNA treatment 200
7.3.2 Metagenome sequencing and assembly 201
7.3.3 Metagenome annotation and comparative genomic analysis 201
7.4 Results and Discussion 202
7.4.1 *Ca.* Parilichlamydiaceae species possess hallmarks of a highly reduced obligate intracellular bacterial genome 202
7.4.2 The minimal metabolic capacity of *Ca.* Parilichlamydiaceae is an example of convergent evolution 203
7.4.3 Nucleotide metabolism and acquisition 205
7.4.4 The updated core *Chlamydiales* genome 211
7.5 Limitations and future directions 213
7.6 Conclusions 215
Declarations 216
Chapter 7 Tables & Figures 217
Figure 7.1: Relationship between chromosome size and number of coding regions in *Chlamydiae* with reference to phylogenetic relationships and ecological diversity. 219
Figure 7.2: Distribution of COGs across *Ca.* Parilichlamydiaceae, CRBs, *Chlamydiaceae* and small genome bacteria used in this analysis 220
Figure 7.3: Major metabolic pathways inferred from genome sequences from three species of the *Ca.* Parilichlamydiaceae. 221
Figure 7.4: *Chlamydiales* core and pan-genome analysis. 222
References 223
CHAPTER 8: General discussion 231
8.1 Summary of the major findings 231
8.2 Chlamydial taxonomic diversity revealed in this study 232
8.2.1 Epidemiological survey of chlamydiae in snakes revealed unexpected levels of chlamydial diversity and a putative reptile-only lineage 232
Figure 1: Annotated *Chlamydiaceae* plasmid sequence-based phylogenetic tree. 236

8.2.2 Novel *Ca.* Parilichlamydiaceae species as host-specific pathogens that broaden the diversity of this phylum 237

Figure 2: The expansion of the *Ca.* Parilichlamydiaceae family over the last four years. 238

8.2.3 Is there a limit to chlamydial diversity? 240

Figure 3: Schematic diagram representing the increase in the number of chlamydial species detected and described, in relation to time, number of hosts sampled (above the line) and advances in molecular methods (below the line). 241

Figure 4: *Chlamydiales* phylogenetic tree highlighting the hypothetical phylogenetic positions yet to be determined. (grey dashed lines) 245

8.3 Chlamydial biological diversity revealed in this study 247

8.3.1 Chlamydia species infecting closely related hosts do not always share genetic traits 247

8.3.2 Convergent evolution of chlamydial genomes is influenced by host restriction 249

8.3.3 Can we predict biology based on host or taxonomy? Can we extrapolate this prediction to the most recent ancestor of the extant members of the phylum? 251

Figure 4: The relationship between chromosome size and number of coding regions in the *Chlamydiae*. 253

Table 1: Predicted genome characteristics of the chlamydial ancestor 251

8.4 Metagenomics as a research tool for Chlamydiae 255

8.5 Limitations and future directions 257

8.5.1 Technical limitations of the metagenomic studies 257

8.5.2 Future directions for chlamydial diversity and genomics studies 259

Table 2: Advantages and disadvantages of aspects of DNA preparation methods for culture-independent metagenomic sequencing 259

8.6 Conclusions 261

References 264

Abstract

The phylum *Chlamydiae* comprises a diverse group of obligate intracellular bacteria that share a unique biphasic developmental cycle. While many of these bacteria are considered pathogens of humans and animals, particularly in the well-described *Chlamydiaceae* family, other more distantly related members can be detected in the environment. The limiting factor to studying these diverse *Chlamydia*-related bacteria is their requirement to be cultured in a host cell, which for many novel and/or *Candidatus* species, is lacking.

To gain insight into the biology and epidemiology of novel bacteria in the phylum *Chlamydiae*, my project aimed to develop a culture-independent genome sequencing (CIGS) work-flow to characterise the genomes of novel uncultivated chlamydiae in clinical samples. I addressed these aims by applying our CIGS method to two examples encompassing the taxonomic and biological diversity within the phylum, including (a) novel species in the genus *Chlamydia* from reptiles; and (b) deeply rooted chlamydiae that had previously been identified infecting fish.

CIGS analysis of *Chlamydia*-positive samples from captive snakes revealed an unexpected level of diversity present including novel *Chlamydia pneumoniae* strains and two proposed novel *Chlamydia* species, *Candidatus (Ca.)* Chlamydia sanzinia and *Ca.* Chlamydia corallus. Plasmid sequences from metagenomes also suggest additional diversity in snake hosts, which could reflect host species diversity.

The discovery of another gill-associated chlamydiae from a novel fish host, *Ca.* Similichlamydia epinephilii in Orange-spotted grouper, further highlighted the fact that chlamydial diversity extends through the far-reaches of the phylum and that species in this deep-branching clade appear to be largely host specific. These *Ca.* Parilichlamydiaceae genomes are remarkably smaller than other *Chlamydia*-related bacterial genomes, and are functionally more similar to their *Chlamydiaceae* counterparts despite their earlier divergence. Together, this data suggests that convergent evolution has shaped the biology and pathogenesis of this clade.

This work has revealed that CIGS is an effective method to further characterise chlamydial diversity whilst providing a unique opportunity to understand the biology of novel uncultivated chlamydial species without the need for cultivation.

List of abbreviations

EB, RB	Elementary body, Reticulate body
CRB	*Chlamydia*-related bacteria
Ca.	Candidatus
TEM	Transmission electron microscopy
ATP, ADP	Adenosine triphosphate, Adenosine diphosphate
DNA, RNA	Deoxyribonucleic acid, Ribonucleic acid
rRNA	Ribosomal RNA
PCR	Polymerase chain reaction
IHC	Immunohistochemistry
OTU	Operational taxonomic unit
Bp, Mbp, Kbp	Base pairs, Megabase pairs, Kilobase pairs
CDS	Coding sequence
PZ	Plasticity zone
HGT, LGT	Horizontal gene transfer, Lateral gene transfer
T3SS	Type Three Secretion System
SNP	Single nucleotide polymorphism
GI	Genomic island
ORF	Open reading frame
IMS	Immunomagnetic separation
WGA	Whole genome amplification

MDA	Multiple displacement amplification
LPS	Lipolysaccharide
IgG	Immunoglobulin G
Ct	Threshold cycle
IFU	Inclusion forming units

List of publications and conference presentations

Publications

1. Alyce Taylor-Brown, Lloyd Vaughan, Gilbert Greub, Peter Timms, Adam Polkinghorne: Twenty years of research into *Chlamydia*-like organisms: A revolution in our understanding of the biology and pathogenicity of members of the phylum *Chlamydiae*. *Pathogens and Disease* 04/2015; 73(1). DOI:10.1093/femspd/ftu009

2. Alyce Taylor-Brown, Simon Rüegg, Adam Polkinghorne, Nicole Borel: Characterisation of *Chlamydia pneumoniae* and other novel chlamydial infections in captive snakes. *Veterinary Microbiology* 04/2015; 178(1-2). DOI:10.1016/j.vetmic.2015.04.021

3. Alyce Taylor-Brown, Nathan L. Bachmann, Nicole Borel, Adam Polkinghorne: Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes. *BMC Genomics* 12/2016; 17(1). DOI:10.1186/s12864-016-3055-x

4. Alyce Taylor-Brown, Trestan Pillonel, Andrew Bridle, Weihong Qi, Nathan L. Bachmann, Terry L. Miller, Gilbert Greub, Barbara Nowak, Helena M.B.Seth-Smith, Lloyd Vaughan, Adam Polkinghorne: Culture-independent genomics of a novel chlamydial pathogen of fish provides new insight into host-specific adaptations utilized by these intracellular bacteria. *Environmental Microbiology* 05/2017; 19(5). DOI: 10.1111/1462-2920.13694

5. Alyce Taylor-Brown and Adam Polkinghorne: New and emerging chlamydial infections of creatures great and small. *New Microbes and New Infections* 07/2017; 18 DOI: 10.1016/j.nmni.2017.04.004

6. Alyce Taylor-Brown, Labolina Spang, Nicole Borel, Adam Polkinghorne: Culture-independent metagenomics supports pathogen discovery for uncultivable bacteria within the genus *Chlamydia*. *Scientific Reports* 7/2017; DOI:10.1038/s41598-017-10757-5

7. Alyce Taylor-Brown, Danielle Madden, Adam Polkinghorne: Culture-independent approaches for chlamydial genomics and metagenomics. *Microbial genomics* (Submitted Oct 2017)

8. Alyce Taylor-Brown, Trestan Pillonel, Gilbert Greub, Barbara Nowak, Lloyd Vaughan, Adam Polkinghorne: Convergent evolution of chlamydial genomes is influenced by host restriction. *ISMEJ* (To be submitted Dec 2017)

Conference Presentations – Oral presentation

Nov 2014 *Australian Chlamydia Conference*

Molecular characterisation of novel chlamydial infections in captive snakes in Switzerland

Sep 2015 *European Meeting on Animal Chlamydioses*

Culture-independent genome sequencing and analysis of the chlamydial agent of epitheliocystis in Yellowtail Kingfish

Jul 2016 *Australian Society for Microbiology Annual Meeting*

Culture-independent genomic characterisation of Candidatus Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes

Sep 2016 *European Society for Chlamydia Research meeting*

Culture-independent genomics of a novel chlamydial pathogen of fish provides insight into the evolution of pathogenicity in the Chlamydiae

Nov 2017 *Australian Chlamydia Conference*

Convergent evolution of chlamydial genomes is influenced by host restriction

Conference presentations – Poster presentation

Jan 2015 *Chlamydia Basic Research Society Meeting*

Characterisation of novel chlamydial infections in snakes (Presented by Nicole Borel)

Jul 2015 *Australian Society for Microbiology Annual Meeting*

Culture-independent genome sequencing and analysis of the chlamydial agent of epitheliocystis in Yellowtail Kingfish

Jul 2016 *Australian Society for Microbiology Annual Meeting*

Culture-independent genomic characterisation of Candidatus Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes

Sep 2016 *European Society for Chlamydia Research meeting*

Culture-independent genomic characterisation of Candidatus Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes

Sep 2017 *Wildlife Diseases Association Conference*

Genome level comparison of Chlamydia pecorum present in both 'northern' and 'southern' koala populations (First author and presented by Alistair Legione)

Acknowledgements

First I'd like to acknowledge financial assistance from the USC Research Training Program Scholarship (Australian Government Research Training Program Scholarship) without which I would not have been able to undertake such a time-demanding feat. I'd like to thank the USC for providing the opportunity, infrastructure and administrative support for me to complete the project, and the Centre for Animal Health Innovation, Faculty of Science, Health, Education and Engineering and School of Science and Engineering, for providing funding for travel and lab consumables.

I am thankful to my principal supervisor, Adam Polkinghorne, for taking me on for not one but two PhDs after seeing potential in me during honours and having faith in me after the first PhD didn't go to plan. As a sidenote, I am grateful for the post-doc that was hired to essentially do that project, spurring me to embark on this one. From a bumpy honours year with very brief meetings and a false start at a PhD, Adam has become a great mentor, friend and someone I look forward to working with again in my research career. Adam has always had my best interests at heart and has pushed me to think critically, read widely, collaborate "up" and get out of my comfort zone. He has fostered my sense of ambition and given me the space to become an independent researcher by allowing me to form my own hypotheses, develop experimental plans, engage with collaborators and write grant applications. He has given me abundant opportunities to publish, present, travel, collaborate, teach and mentor and for that I am truly grateful. My co-supervisors, Peter Timms and Barbara Nowak have both guided me in ways that complement Adam's supervision. Peter always has the "bigger picture" in mind and his years of experience in the *Chlamydia* field have helped shaped the way I think about chlamydial biology, infection and disease. He has always challenged my explanations and questioned "why?" to push me to think harder and more deeply. Although being further away, Barbara has had a keen interest in my project and my

development as a researcher. I appreciate Barbara's direct approach and admire her capacity and drive.

Major thanks also goes to past and present members of the *Chlamydia* research group. Namely, Marina Mathew for being the first influence in and outside the lab in the IHBI days; Martina Jelocnik, who has been there all the way through, one of the most intelligent and supportive people I know, whose passion I admire and whose charisma I envy; and Nathan Bachmann for patiently transferring his bioinformatics skills to me. I am grateful to other members of the team that have provided generous support, knowledge, feedback and friendship over the years.

I'm grateful to all my co-authors: Nicole Borel, Simon Ruegg, Labolina Spang, Lloyd Vaughan, Helena Seth-Smith, Weihong Qi, Gilbert Greub, Trestan Pillonel, Terry Miller, Andrew Bridle and Danielle Madden for their ongoing contributions to my projects, papers, and my personal and professional development. I'm thankful to the numerous reviewers for providing feedback on my papers and inevitably improving them, especially ones that provided negative comments that provoked me to write more clearly.

Last, I'd like to acknowledge my friends and family for their support, patience and encouragement: my parents for encouraging me to study and allowing me to be independent; my friends for putting up with me being late or absent and in particular Emma for impromptu grammar lessons; and my partner Brent for always believing in me, being my rock, being a fantastic listener, doing the dishes, encouraging me to push on and supporting me even when I was completely absorbed by my project.

CHAPTER 1: Literature review and introduction

1. *Chlamydia* and *Chlamydia*-related bacteria

The *Chlamydiae* are an assemblage of bacteria that are united by their obligate intracellular lifestyle and unique developmental cycle (Figure 1; described in more detail in section 1.1.1). The Phylum *Chlamydiae* is composed of a single order, *Chlamydiales* which comprises several diverse families described in sections 1.1.2 and 1.1.3. While updating chlamydial taxonomy has not been without its challenges (1, 2), the recently accepted (i) taxonomic system for *Chlamydiaceae* (3) and; (ii) the availability of expanded gene sets encompassing phylogenetic markers capable of differentiating taxa at the species, genus and family level (4), provides the basis for classification and terminology used throughout this chapter.

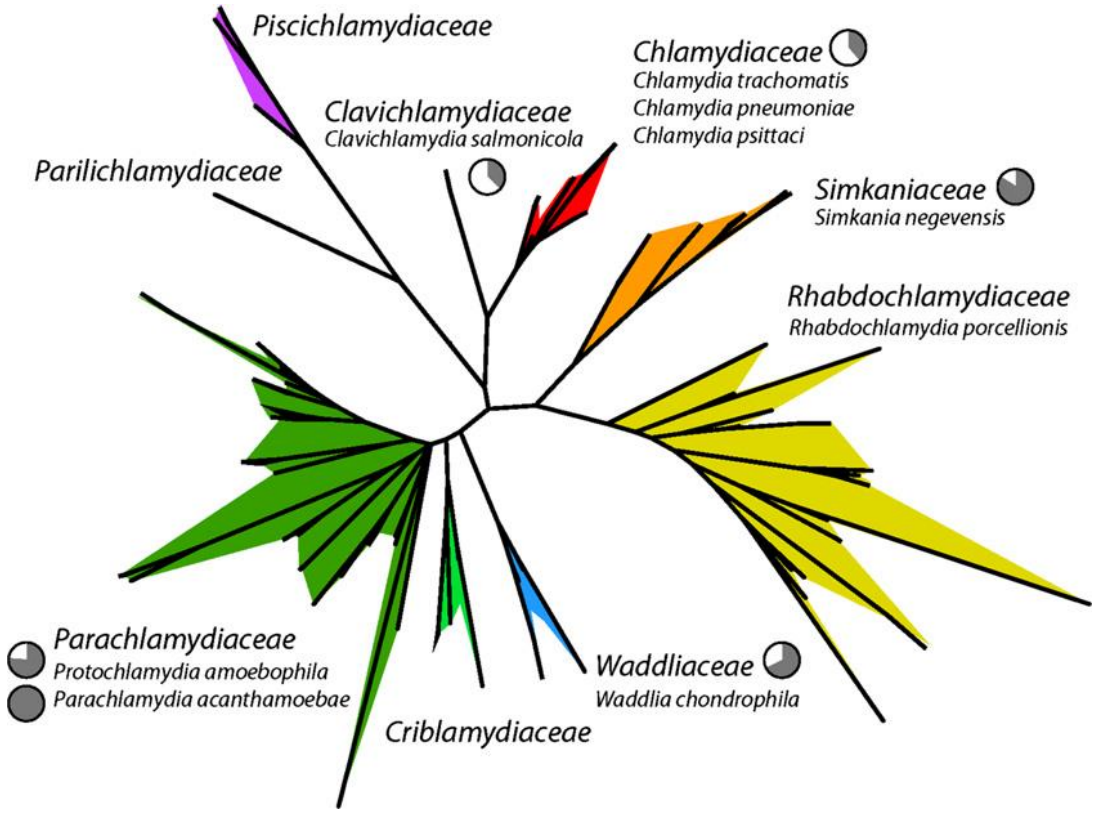


Figure 1: Phylogenetic tree of the Chlamydiales *including abundance of each clade in public databases. Pie charts represent genome sizes relative to that of* Parachlamydia acanthamoebae *(5).*

Our knowledge of the diversity and significance of members of the *Chlamydiaceae,* comprising many well recognised pathogens of humans and animals, has been well established thanks to more than 50 years of intensive biomedical research. The genus *Chlamydia* remains the most widely studied of the *Chlamydiae*, and until the 1990s, it was thought to be the only genus in the only family of this order (1, 6-8). However, research over the last 20 years has revealed that this family only represents the 'tip of the iceberg' in terms of diversity within the phylum *Chlamydiae*. In this, we are referring to the recent explosion of the description of eight additional families (collectively referred to as *Chlamydia*-related bacteria; CRBs) of genetically related obligate intracellular bacteria that infect a wide range of hosts.

The strict intracellular nature is shared by all organisms in the phylum, and has prevented culturing many of the species that infect 'non-model' organisms. Recent advances in culture-independent, high throughput, genome and metagenomic sequencing have allowed researchers to gain insight into the unique biological properties that contribute to the pathogenesis of these bacteria in the absence of *in vitro* characterisation.

1.1. Unique biological features of chlamydiae (EBs, RBs and the inclusion)

Members of the *Chlamydiae* are characterised by their biphasic life cycle and possession of a number of unique proteins that permit interactions with their host. This section discusses the fundamentals of chlamydial biology, and other biological features such as aspects of metabolism and virulence of the *Chlamydiae* are covered from a genomic standpoint in later sections of this chapter.

Chlamydial developmental cycle

Early Transmission Electron Microscopy (TEM) studies showed that one of the defining features of the *Chlamydiae* is its unique biphasic lifecycle. The chlamydial developmental cycle begins with endocytosis of the infectious elementary bodies (EBs) into the host cell (Figure 2).

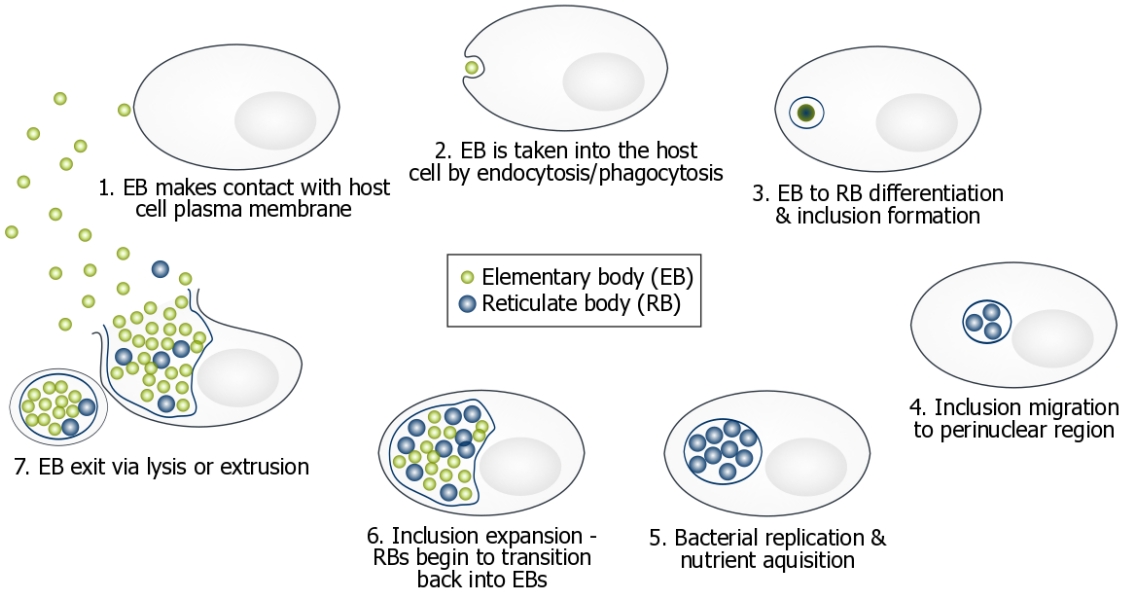


Figure 2: Chlamydial developmental cycle; infectious EBs are endocytosed into the host cell where they differentiate to RBs within an inclusion. RBs replicate, expanding the inclusion before de-differentiation back to EBs. The EBs burst the inclusion and exit the host cell. (10)

Here, they reside within a membrane-bound cytoplasmic vacuole, termed the inclusion (section 1.1.3), which facilitates differentiation from the EB to the reticulate body (RB). The RBs replicate by binary fission within the inclusion until they asynchronously convert back to EBs, before lysing the inclusion and host cell upon exit and perpetuating the infectious cycle. This typically takes place over 48-72 hours, with conversion to RBs and back taking place between 2-12 hours and 18-48 hours, respectively (Figure 2) (9).

This developmental cycle and cell morphologies are well-characterised in *Chlamydia spp.,* but significant deviations are observed among the CRBs, perhaps as an adaptation to diverse host cell niches. EBs are generally smaller than RBs (0.3mm and 1.0mm, respectively in *C. trachomatis*), with a condensed nucleoid, mediated by a bacterial histone-like protein (11, 12). This was thought to confer a "transcriptional shutdown", and led to the impression that EBs are "metabolically inert". This was in part due to their resistance to harsh conditions, attributed to the cell wall stability conferred by disulphide cross-linking of cysteine-rich outer membrane proteins complex, which is relaxed (reduced) in the RB form. This paradigm has since been revised as there is a multitude of

evidence that EBs are in fact capable of synthesising ATP and proteins, have transcriptional and translational capabilities, and are enriched in proteins involved in central carbon and glucose metabolism (5, 13-15), which are thought to induce a burst of metabolic activity upon host cell entry. RBs on the other hand, highly express proteins involved in ATP generation, protein synthesis and nutrient transport. In CRB EBs the major outer membrane is replaced by a porin family protein (16, 17), which confers its rigid structure. CRB RBs are more variable, with intermediate forms, elongate RBs and head-and-tail cells described for these bacteria. Developmental cycles take up to 7 days, depending on the chlamydial species and host cell line in which the chlamydiae were cultivated in (8, 18, 19).

Type three secretion system

Whilst not unique to chlamydiae, the type three secretion system (T3SS) is fundamental in establishing chlamydial infection. The chlamydial T3SS is a needle-like injector embedded into the chlamydial membrane that functions as a channel to facilitate transport of molecules into the host cell cytosol (20). Throughout the developmental cycle, different aspects of the T3SS are expressed and various molecules (effectors) are secreted carry out processes that modulate and exploit the host cell such as cytoskeleton remodelling, centrosome tethering and disruption of normal pathways (21, 22). The T3SS is highly conserved throughout the phylum confirming its essential role in chlamydial pathogenesis.

The chlamydial inclusion

The chlamydial inclusion is initially derived from host cell plasma membrane proteins. Once inside the vacuole, chamydiae recruit lipids, sphingomyelin and cholesterol, in part by intercepting Golgi-derived vesicles (23, 24). Acquisition of nutrients by chlamydiae is also largely mediated by inclusion membrane proteins (Incs) which are inserted into the inclusion membrane upon invasion. Incs are considered chlamydial effectors as they are targeted to the inclusion

membrane or host cell cytoplasm by the T3SS (22). These proteins most likely also play roles in maintaining the structure of the inclusion (24).

1.2. Host diversity and disease spectrum of the Chlamydiaceae: a broad overview

Bacteria in the family *Chlamydiaceae* are globally significant, widely distributed human and animal pathogens. Until recently, the *Chlamydiaceae* was comprised of nine taxonomically recognised, well-defined species belonging to the genus *Chlamydia*: *C. trachomatis, C. muridarum, C. suis, C. psittaci, C. abortus, C. caviae, C. felis, C. pneumoniae* and *C. pecorum* (1). This view has begun to rapidly change with the discovery and description of two novel chlamydial species and a Candidatus species in avian hosts: *C. avium*, *C. gallinacea* (3) and *Candidatus* C. ibidis (25) from domestic and wild birds.

Chlamydial pathogens of humans and zoonotic pathogens

Few species in the genus are commonly reported in humans, with *C. trachomatis* considered the main human pathogen. Other species such as *C. psittaci* are considered zoonotic pathogens, where inoculation risk is limited to direct contact with infectious material, and disease is often only established in favourable conditions such as immune compromise.

It is thought that *C. trachomatis* pre-dates civilisation, then co-evolved with humans over at least 700 million years (26), adapting primarily to the conjunctival tissue in the eye where it causes trachoma. From here, it disseminated to the urogenital tract where it infects the mucosal lining and can be spread through secretions. Trachoma is the leading infectious cause of preventable blindness worldwide (27), and reproductive disease caused by the same bacteria is the leading bacterial cause of sexually transmitted infections worldwide (28). Despite being easily managed by antibiotic therapy, the damage caused by scarring at both the ocular and reproductive sites of infection are irreversible. Repeated infection with *C. trachomatis* leads to inversion of the eyelids, causing the eyelashes to scratch and scar the cornea. Additionally, the genital infection is often

asymptomatic, and 10% of cases of *C. trachomatis* urogenital infections progress to pelvic inflammatory disease (29), which ultimately can result in infertility. *C. trachomatis* can be classified into one of several serovars, which can be grouped into two distinct biovars. Serovars A-C are ocular-only strains and serovars D-K are urogenital strains: these two groups comprise the trachoma biovar. The second biovar, lymphogranuloma venerum (LGV), encompasses serovars L1-L3, which are associated with reproductive disease. Initially, strains of *C. trachomatis* were classified on the basis of antibody reactivity to the major outer membrane protein (MOMP). Interestingly, the phylogenetic relationships between the serovars have only been resolved with the aid of comparative genomics (30). The main biological difference between the ocular and urogenital serovars lies in the tryptophan synthase operon, which is incomplete in the urogenital strains (31-33). This biological difference is hypothesised to be attributed to the presence of indole-producing bacteria present in the vaginal microbiome, hence compensating for the genetic inability to synthesise tryptophan (31).

C. pneumoniae is one of the oldest chlamydial species and has one of the broadest host ranges in the genus. In humans, it is most commonly associated with community-acquired pneumonia, with an incidence rate around 10% (34). *C. pneumoniae* may also have roles in other non-infectious cardiovascular and neurological diseases such as bronchitis and chronic obstructive pulmonary disease, atherosclerosis, Alzheimer's disease, stroke and migraine (35). The genomes of the human strains are highly conserved compared to animal strains, suggesting that *C. pneumoniae* originated in animals and crossed the host barrier at least once over the course of its evolution (36-38).

From a public health perspective, one of the most important human chlamydial pathogens is *C. psittaci*; traditionally recognised as an avian pathogen, potential cause of zoonotic disease and notifiable in most countries. Most human psittacosis cases, which manifest as atypical pneumonia, are confined to isolated outbreaks in groups closely associated with birds for work or recreation such as bird keepers and poultry workers (39-45). So far, evidence for human to human

transmission is limited to three European cases, suggesting it is extremely rare (44, 45). However, recent comparative genomic analysis of human-isolated *C. psittaci* strains suggest that only indirect contact/environmental exposure is necessary to opportunistically infect and cause symptoms in humans (43, 46).

While it is primarily a pathogen causing abortion in small ruminants, *C. abortus* poses a zoonotic risk to pregnant women in contact with infected animals during the birthing process (47), as well as laboratory workers exposed to the bacteria (48). As yet, no human-isolated genomes of *C. abortus* have been characterised.

There is also growing molecular evidence that *C. suis,* a chlamydial species endemic to pigs, can infect humans (49). Recently, *C. suis* was detected in the air, on surfaces and from the conjunctivae of two people at a slaughterhouse in Belgium (50). At the same time, a broad study of chlamydial agents in Nepal detected *C. suis* DNA in the eyes of patients with trachoma (51). *C. suis* is the only known chlamydial species to naturally harbour a tetracycline resistance gene cassette (52, 53), potentially intensifying the zoonotic threat already posed by this species.

Chlamydial pathogens of birds

The description of avian chlamydiosis dates back to the 1890's, when an outbreak of flu-like symptoms transmitted from birds to humans was thought to be caused by a *Salmonella* bacterium (54). Over the next few decades, following similar pandemics, and the development of cell culture methods, the aetiological agent of psittacosis (so named due to its apparent origin in Psittacine birds; ornithosis in non-Psittacines), was classified as a virus due to the intracytoplasmic vacuoles containing microbial cells (54), and finally as a bacteria closely related to the aetiological agents of the lymphogranuloma venereum and trachoma inclusion conjunctivitis (ie, *C. trachomatis*). Nowadays, *C. psittaci* is widely recognised as pathogen endemic to parrots but is also common in poultry species such as chickens, turkeys and ducks. Other agents of avian chlamydiosis are now being described, which have similar presentations but potentially stricter host ranges.

The aetiological agent of psittacosis, *C. psittaci*, is hypothesised to have emerged from South American (55) or Australian parrots (43). Bird to bird transmission was shown to be rapid in experimentally infected ducklings and their mock-infected counterparts (56), and it is likely that *C. psittaci* can be transmitted between avian and mammalian hosts, evidenced by the fact that *C. psittaci* has been sporadically described in several livestock species (49, 57-59), as well as humans.

As mentioned earlier, the discovery and description of three new taxa in avian hosts has challenged our understanding of the diversity of relationships between chlamydiae and their hosts. *C. avium* comprises strains from pigeons and psittacine birds, *C. gallinacea* comprises strains from poultry (60) and *Ca.* C. ibidis was isolated from the digestive tracts of feral Ibises in France (25). Although these novel species were only initially described in Europe, *C. avium* and *C. gallinacea* are now considered to be widespread in both European and Asian countries (61, 62).

Chlamydial pathogens of reptiles and amphibians

Chlamydiosis has been described in both free-ranging and captive reptilian and amphibian hosts including puff adders, boas, chameleons, crocodiles, frogs and tortoises (63-68). The infection typically manifests as inflammatory lesions (granulomatous inflammation) in spleen, heart, lung and liver (65, 69). Cases of proliferative pneumonia, (67, 70), lesions in the gastrointestinal tract, wasting disease, necrotising mycocarditis, necrotising enteritis and splenitis (67, 71-73) have also been reported.

Although some studies identified *C. psittaci* as the aetiological agent (68, 74), *C. pneumoniae* is the most widespread chlamydiosis agent in reptiles. Other studies identified *Chlamydia*-related bacteria (63), or novel lineages within the *Chlamydiaceae*.

Chlamydial pathogens of other mammals

Most veterinary chlamydia research efforts centre on agriculturally and/or economically important, domesticated animals. Other mammalian infections that are given particular research attention are those that pose zoonotic threats, whilst others are important infection models.

C. pecorum is a major threat to the long-term survival of the koala, in which it has two manifestations reminiscent of *C. trachomatis* infections in humans. *C. pecorum* infection can lead to infertility as a result of prolonged reproductive disease, and blindness as a sequellae from untreated conjunctivitis (75). Epidemiological studies have revealed a complex relationship between chlamydial infection and disease, with analysis of the relationship between *C. pecorum* infectious load and disease showing that a number of asymptomatically infected animals shed more *Chlamydia* than those with inactive but chronic chlamydial disease (76).

There is also a role for *C. pecorum* in a plethora of diseases including encephalomyelitis (from which it was first described (77)), polyarthritis, and conjunctivitis in ruminants such as cattle, pigs, sheep and goats (78-81). The endemic nature of *C. pecorum* infections in livestock in many populations suggests it resides within a local reservoir, such as the gastrointestinal tract, from which it can opportunistically infect other anatomical sites. This is presumed to occur via the bloodstream, but evidence for this is lacking.

Recent molecular and serological studies have shown that the host range of *C. pecorum* does not remain limited to domesticated/farmed animals: other wildlife including mammals and birds have also been reported to harbour *C. pecorum* (82-86), which could represent natural/wild reservoirs of *C. pecorum*.

Along with humans, reptiles and amphibians as mentioned above, *C. pneumoniae* readily infects marsupials such as the koalas and brushtail possum (67), where it is mostly asymptomatic but can cause respiratory disorders. Interestingly, comparative genomics studies place the animal

strains of *C. pneumoniae* (horse and koala isolates) as "evolutionarily basal" to the human strains, while only the animal strains harbour a chlamydial plasmid (36, 37). Major genetic differences are also seen in the plasticity zone, a region surrounding the replication terminus that is subject to high levels of variation between species and strains, with humans *C. pneumoniae* strains harbouring a purine ribonucleotide synthesis operon, which is absent from the animal strains (discussed in more detail in section 2.1). Horses are also hosts for *C. pneumoniae*, which is again associated with respiratory disease that mirrors that seen in humans (87, 88). Additionally, there is rare but convincing evidence that *C. psittaci* can cause reproductive failure in horses (89, 90), where it is presumed to disseminate from the respiratory tract.

Ovine sporadic enzootic abortion attributed to *C. abortus* is widespread throughout Europe, North Africa and the U.S.A, but interestingly, it is absent from Australia and New Zealand. Although a live attenuated vaccine against *C. abortus* is available, it can be expensive to vaccinate whole flocks.

C. suis is endemic in pigs, where it is associated with a range of pathologies such as conjunctivitis, enteritis, pneumonia and reproductive disorders (49). Tetracycline resistant strains emerged in the late 1990's, posing a threat not only to the farming industry but to public health in general.

C. muridarum, *C. caviae* and *C. felis* are pathogens of mice, guinea pigs and cats, respectively. *C. muridarum* infects the genital tract, upper respiratory tract and ocular membranes of mice and as such is extensively used as a model to study chlamydial reproductive and ocular disease as a counterpart to its close relative, *C. trachomatis*. Such studies have been fundamental to our understanding of chlamydial pathogenesis by providing insight into host cell manipulation strategies employed by *Chlamydia* to establish infection, as well as the immunological factors associated with response to infection and subsequent contribution to disease sequellae. *C. caviae* and *C. felis* were initially thought to be *C. psittaci* isolates, prior to the advent of 16S rRNA

sequencing, and manifest as conjunctivitis or keratitis in guinea pigs and cats, respectively. Due to *C. felis* vaccine availability, prevalence of feline conjunctivitis is low, also reducing the zoonotic risk to humans. Like *C. muridarum*, *C. caviae* is widely used as a chlamydial infection model (91).

Emerging chlamydial pathogens and their hosts

The intensification of farming and widespread popularity of companion animals have presented ongoing opportunities for infectious diseases to emerge (92). Wider sampling, improvements in surveillance and diagnostics and advances in molecular methods have contributed to a wider recognition of the emergence of these pathogens (92, 93), and revealed a breadth of novel families and chlamydial hosts within the phylum/family. It is clear that much of the diversity, as well as opportunities to answer evolutionary questions, lies in chlamydial infections of wildlife, namely wild animals closely related to their domesticated counterparts, and/or those residing in sympatric areas. For instance, native animals are often overlooked as hosts for chlamydial disease, but in fact these ecologically interesting animals could hold keys to understanding disease pathogenesis and bacterial evolution. Unfortunately, a caveat for these studies lies in the trade-off between invasive sampling techniques and wealth of information they can provide. Researchers are thus largely confined to opportunistic sampling upon presentation of individual animals to wildlife hospitals.

As expected, birds, mammals and reptiles harbour much of the untapped diversity within *Chlamydia*. Several novel unclassified chlamydial species were recently reported in faecal specimens in wild sea birds such as penguins and gulls (94-97). These taxa are closely related despite their geographically distinct locations. Likewise, phylogenetic analysis of 16S rRNA and *ompA* sequences from lung, liver, spleen and brain from crocodiles showed three genotypes within a novel clade related to C. *caviae* and *C. felis* (98). A separate study identified a 55% prevalence of *Chlamydiaceae* in Australian crocodiles with conjunctivitis and/or pharyngitis syndrome, but no sequencing was conducted to further identify the agent/s (99).

A novel chlamydial lineage has been described in female roe deer from France, that comprises both vaginal and faecal samples and is most closely related to *C. trachomatis* and *C. suis* (100). The pathogenic potential is unknown but these novel taxa highlights how little we know about emerging chlamydial infections in wild animals.

Most recently, following a metagenomic study that identified *Chlamydiae* in bat faeces (101) a screening study reported novel *Chlamydia* lineages in bat droppings in Finland (102). Interestingly, insects collected at feeding sites did not harbour species in the *Chlamydiaceae,* suggesting that the chlamydial insect endosymbiont lineages may have a stricter host preference, and that the *Chlamydiaceae* lineages are not diet-borne.

1.3. Phylogenetic and ecological diversity of Chlamydia-related bacteria

As one can see from the overview detailed in section 1.2, the host range within the *Chlamydiaceae* family alone is very broad, covering mammals, birds, reptiles and amphibians, with wider sampling and advances in diagnostic and molecular techniques leading to discoveries of novel lineages in recent years (Figure 3). Further, our understanding of the true taxonomic and host diversity within the extended phylum is seemingly still incomplete.

It has been estimated that the members of the genus *Chlamydia* (*Chlamydiaceae* family) and the CRBs (*Parachlamydiaceae* family) diverged more than 700 million years ago, from a last common ancestor that also resided within a host cell (103, 104). This realisation has meant that molecular and cell biology studies of CRBs should be viewed as an opportunity to 'look into the window of the past' for members of the *Chlamydiaceae*.

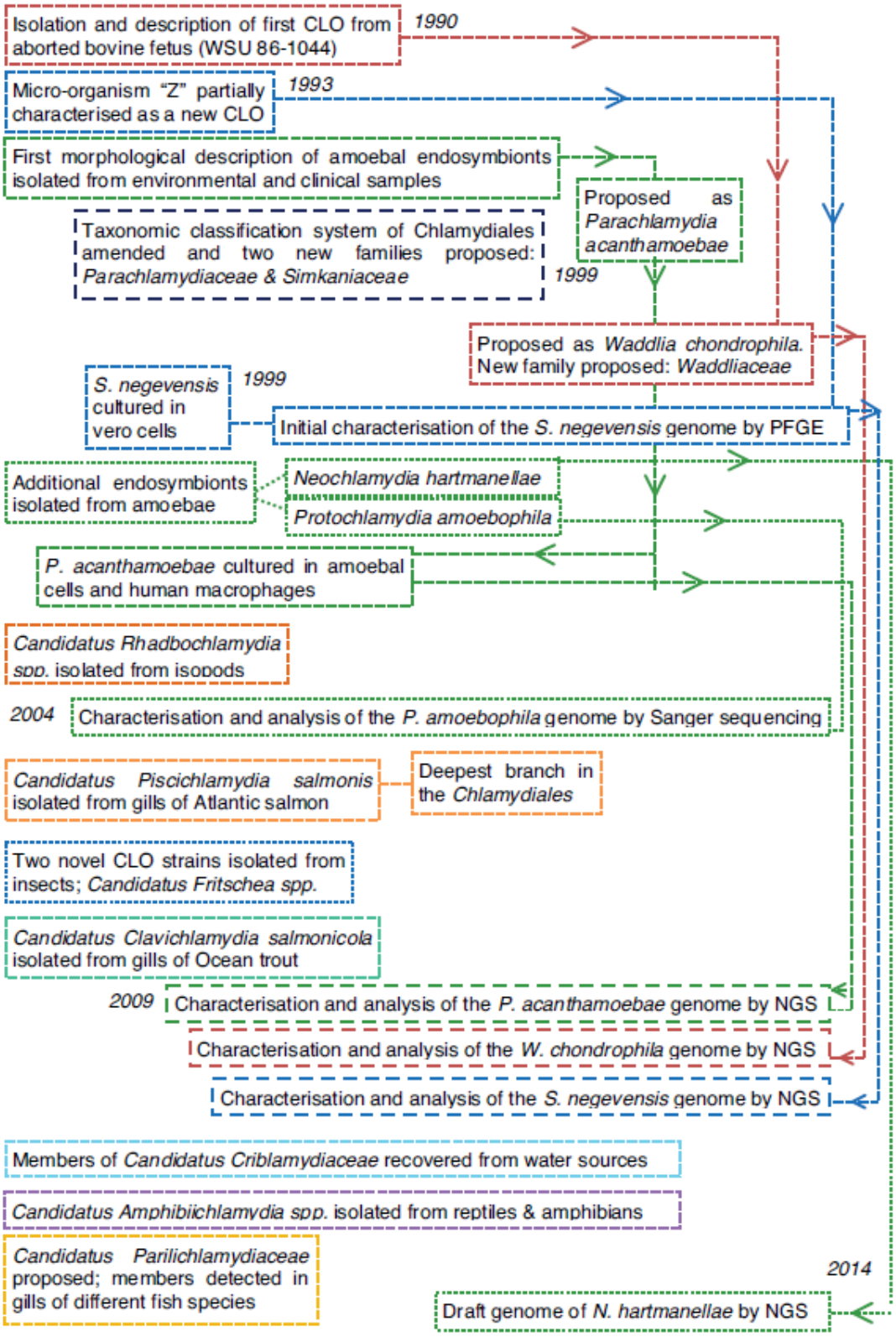


Figure 3: Significant events in the detection, isolation and identification of CRBs (at the time of preparation (2014) referred to as CLOs) over the last 20 years. Genome sequencing events are also included.

The isolation of an obligate intracellular bacterium from an aborted bovine foetus (*Waddlia chondrophila* WSU-86-1044) and subsequently a cell culture contaminant (*Simkania negevensis* Z) with morphological features characteristic of *Chlamydiae* changed the landscape of chlamydial research, as the realisation hit that the phylum was in fact more diverse than previously thought. Given the isolation of a vast number of CRBs from environmental sources (water, soil) and traditional vectors (protozoa, insects, bats), there is no doubt that these sources have the capacity to act as reservoirs and/or vectors of potentially 'environmental pathogenic' chlamydiae.

Chlamydial endosymbionts of amoebae (and their pathogenic potential)

Free living amoebae are ubiquitous in the environment, inhabiting many environmental niches such as soil, water, dust, air and guano. Their ability to survive in such harsh conditions is due to their life cycle, which involves a replicative form (trophozoite) and dormant form (cyst); this biology means that they are resistant to many biocides (105). This inherent resistance thus further protects the bacteria. Amoebal species and their endosymbionts are thus important in public health due to the proclivity of several species to cause, or carry microbes that cause, disease in humans (106, 107). The chlamydial species described below are amoeba-resisting bacteria, meaning that not only to they resist digestion by the amoeba (as a food source), they utilise their amoebal host cytoplasm as the environment in which they replicate. True endosymbionts do this without lysing the cell and hence maintain a stable host-parasite ratio, although this is often dependent on environmental conditions. The role of these amoeba-resisting bacteria in human and animal disease is controversial (106, 107).

Using the approach taken by Kahane *et al.* (1993, 1999) for *S. negevensis,* Everett *et al.* (1999) demonstrated 86– 87% 16S rRNA nucleotide similarities with other chlamydiae, proposing *Parachlamydia acanthamoebae* as the type species of the genus *Parachlamydia* in the family *Parachlamydiaceae.* The discovery of this bacterium suggested for the first time that amoebae could potentially act as environmental reservoirs for pathogens belonging to the *Chlamydiae* phylum.

Members of the *Parachlamydiaceae* appear to be ubiquitous among amoebae, with amoebal hosts initially including *Acanthamoeba castellanii, A. polyphaga* (109), *Naegleria spp*, and *Hartmanella sp.*, (110). *Parachlamydiaceae*-containing amoebae have been retrieved from activated sludge (111), hot springs (112) and soil (113), while clinical specimens such as corneal and nasal samples have been found to harbour *Parachlamydiaceae* (113, 114).

Two members of the *Parachlamydiaceae*, *Ca.* Metachlamydia lacustris (115) and *Ca.* Mesochlamydia elodeae (116) were initially isolated from aquatic amoebae and demonstrate high host specificity in contrast to other *Parachlamydiaceae.* Interestingly, these species co-inhabit the amoebae with members of either alpha-proteobacteria or betaproteobacteria (115, 117), and such close association could offer opportunities for genetic exchange. The genomes of this clade would further clarify whether this has taken place.

Water sources have otherwise proven to be a rich source of CRB diversity, with two novel chlamydial species described in 2006, *Criblamydia sequanensis* and 2011, *Estrella lausannensis*, both part of a novel family, the *Criblamydiaceae* (97, 98). *E. lausannensis* and *C. sequanensis* have been studied using amoebal co-culture. *E. lausannensis,* was cultured in *Dictyostelium discoideum*, a genetically tractable soil amoebae (118), potentially providing a novel model system for studying CRB biology, given the huge numbers of mutants available. Neither species has been successfully cultured in a mammalian cell line, suggesting they are indeed not capable of causing disease in humans, but *E. lausennensis* is capable of replicating in fish cell lines, albeit poorly (119). *Vermamoeba vermiformis* harbouring a novel *Chlamydiales* species, *Ca.* Rubidus massiliensis, was recently isolated in a hospital water supply in Marseille, France (120).

Another marine amoeba, *Vexillifera sp.* was found to harbour novel lineages of chlamydiae within and outside the *Simkaniaceae,* in a marine lake, Lago di Paola. A novel species *Ca.* Neptunochlamydia vexilliferae, was described as an endosymbiont in this amoebal host with abundance dependent on the picoplankton abundance (121). Similarly, drinking water supplies and

wastewater in Israel also tested positive for *S. negevensis* antigens, often in association with amoebic antigens (122).

Chlamydial endosymbionts of insects

Arthropods such as ticks, fleas and other insects are emerging as another potential rich reservoir of CRB diversity.

Ticks are well-known vectors for other pathogens (123) and as such their role as chlamydial vectors has now been investigated in Switzerland, Algeria and Australia, with studies reporting prevalence rates of 6.4-45% in pooled ticks with individual prevalence of 0.89-28.4% (124-127). So far, the chlamydial prevalence is highest in adult ticks, perhaps corresponding to the duration of feeding (123), or a sampling bias. The four studies report primarily *Rhabdochlamydiaceae*, *Parachlamydiaceae, and Simkaniaceae,* with *Rhabdochlamydiaceae* DNA being most abundant, within cohorts and individuals (average $2x10^4$ copies/μL compared to ≤ $3x10^2$) (125). The ticks themselves appear to be unaffected by their chlamydial infection, suggesting that ticks may be "opportunistic reservoirs" for CRBs (123). No study has yet detected the same chlamydial species in a tick and its host, however a recent study examining skin biopsies from tick bites identified tick-related chlamydial sequences in the biopsies but not controls (126). This work also implies that there may be other unknown intermediate hosts, amplifiers or reservoirs that facilitate animal-animal or animal-human transmission

Chlamydial endosymbionts of arthropods were first identified by TEM in 1995 within bacteriocytes in the gut of the whitefly, *Bemisia tabaci*. Later, genetic analysis confirmed this bacteria and a symbiont of *Eriococus spurius* were indeed two novel species within the *Simkaniaceae*, comprising a novel genus, *Ca.* Fritschea: *Ca.* F. eriococci and *Ca.* F. bemisiae (128, 129). These data not only further expand the host diversity displayed by members of the *Chlamydiales*, but also further complicate our understanding of the complex host-symbiont relationship.

Currently, two species within the Rhabdochlamydia genus have been proposed, although database and phylogenetic analyses show that *Ca.* Rhabdochlamydiaceae is putatively the largest family in the *Chlamydiae* (5). *Ca.* R. porcellionis was detected in the hepatopancreas of a woodlouse, and interestingly, display a distinctive morphology of rod shaped elementary bodies with translucent oblong structures (130). Similarly, *Ca.* R. crassificans was proposed as a novel pathogen of the cockroach, where it was identified in fat body and ovary (131). The implications of these two species in public health remains to be explored. Of note is the fact that several human screening studies have detected *Ca.* Rhabdochlamydia species distinct from the two described in insects (114, 132-134), suggesting that either a) species in this clade are specific to their hosts, and may not be able to cross host barriers, and b) this family may exhibit additional diversity with respect to phylogenetic diversity and host range and/or c) *Ca.* Rhabdochlamydiaceae hosts have been vastly under-sampled..

Chlamydial pathogens of fish

Analogous to several chlamydial species described above associated with respiratory symptoms, the novel species belonging to diverse families described below are commonly associated with the fish gill disease, epitheliocystis. These infections provide some of the best evidence in favour of the CRBs as pathogens, which is controversial given the isolation of several species solely from environmental sources.

Like other chlamydiae, lesions (cysts) are formed by aggregations of infectious EBs within epithelial cells, which, in fish gills, lead to mechanical obstruction of the lamellae. To date, epitheliocystis has been reported in over 90 species of fish globally (135) in both wild and cultured fish from marine and freshwater environments (136, 137). Of biggest concern is the impact of mortalities caused by chlamydial epitheliocystis in cultured fish (138-143), which appear to be most affected during the larval or juvenile stages with mortalities reaching 100%, leading to economic losses.

Two species are currently proposed in the *Ca.* Piscichlamydiaceae family, which to date is the most deeply branching lineage in the *Chlamydiae*: *Ca.* P. salmonis, affecting Salmonids (138, 144, 145) and *Ca.* P. cyprinus affecting Carp (146). Both species have been found in mixed infections either with each other (144, 147), or with other chlamydial epitheliocystis agents (148). Interestingly, neither species has been described in fish from the Southern hemisphere, perhaps suggesting that geographically distinct lineages exist within these clades, or, again, that under-sampling has led to this perceived specificity.

Ca. Parilichlamydiaceae is a recently described family of chlamydial pathogens of fish and represents the largest fish-associated chlamydial family, closely related to *Ca.* Piscichlamydiaceae. It comprises three genera (*Ca.* Parilichlamydia, *Ca.* Similichlamydia and *Ca.* Actinochlamydia) that encompass five Candidatus species. Members of this clade have been detected in African (149), Australian (140, 141, 150), Scandinavian (151), Mediterranean (152) and Swiss waters (147). Additional chlamydial epitheliocystis agents include two members of the family *Simkaniaceae* (136, 153) and a single *Ca.* Clavichlamydia species, *Ca.* C. salmonicola, which, interestingly, is phylogenetically closer to the *Chlamydiaceae* than the other fish-associated lineages (137, 142). The geographical and host species diversity of the *Ca.* Parilichlamydiaceae clade is truly impressive and raises considerable interest on the genomic background for this feat.

An additional CRB, *Ca.* Renichlamydia lutjani (*Ca.* Rhabdochlamydiaceae), was identified in a blue-striped snapper from Hawaii (154) and its role as a purely epitheliocystis agent is questionable due to its detection in internal organs as opposed to the gills or skin. On the other hand, few epitheliocystis studies have included internal organs in the analysis, an aspect needing closer attention in future as this would have implications for disease progression and dissemination.

In addition to these characterised organisms, short *Chlamydiales* 16S rRNA sequences have also been detected in association with cases of epitheliocystis in a diverse range of fish species, including Leopard shark (155), Eagle ray (156), Leafy sea dragon and Silver perch (157).

CRBs as human and mammalian pathogens

A major driver of research interest in CRBs has been the repeated association between these bacteria and a range of different diseases in humans and animals, highlighting their potential as emerging bacterial pathogens but also illustrating that direct evidence for the majority of CRBs as pathogens is lacking.

S. negevensis has not only been isolated from environmental niches and reservoirs, but has also been associated with respiratory disease in humans by both serological and molecular methods. Detection of IgG antibodies to *S. negevensis* indicated a previous infection in 37–62% of pneumonia patients in two separate cohorts, with a small proportion of those indicating a current acute infection (158-161) observed a 25% prevalence rate in bronchiolitis patients by culture and/or PCR. Further, the drinking water was postulated to be the source of infection in a cohort of children with pneumonia, based on immunoassay and culture, as well as *S. negevensis* 16S rRNA sequence amplification from 76% of nasopharyngeal swabs and corresponding drinking water (162). Other studies have found a range of prevalence rates in healthy populations (163, 164), suggesting the potential opportunistic nature of this organism. *In vitro* studies have also demonstrated successful *Simkania* growth in several epithelial cell types in which an inflammatory response was also observed (165). More recently, *S. negevensis* antibodies were also detected in association with gastrointestinal symptoms (166).

The pathogenicity of *P. acanthamoebae* was first suspected when the bacteria was isolated from an amoeba recovered from the water of a humidifier implied in an outbreak of fever (167). Since then, a growing body of evidence has suggested that these CRBs may be pathogenic to humans and a variety of animal hosts. Like *S. negevensis*, in humans, molecular and serological studies have

linked infections of *Parachlamydia spp.* to respiratory disease (168-171). The permissivity of pneumocytes, lung fibroblasts and macrophages to *P. acanthamoebae* (172, 173) has also been demonstrated, adding support for the pathogenic potential of these bacteria and also indicating a potential route of dissemination through the body. Further strengthening the argument for *Parachlamydia* as a respiratory pathogen, both a murine and bovine model of parachlamydial respiratory disease (174, 175) have been established, fulfilling the third and fourth of Koch's postulates.

Parachlamydia spp. may also be linked to adverse pregnancy outcomes in ruminants (176, 177) and, moreover, in humans (133, 178). The prevalence of *Parachlamydia spp.* associated with abortion in cattle has been studied extensively in Europe, with initial Swiss studies showing a prevalence of over 60% in placental lesions by immunohistochemistry (176), while further studies have demonstrated slightly lower prevalence rates (179, 180). In Scotland, prevalence has been reported at around 20%, with a higher prevalence by PCR detection (181). Blumer *et al.* (2011) also found the presence of CRBs as mixed infections with *Chlamydiaceae* or other CRBs in ruminants.

Considering *Parachlamydia* has been isolated from both ruminant foetal tissues and human respiratory samples, this bacterium potentially poses a zoonotic threat, particularly in individuals who have contact with livestock. Interestingly, in a study of healthy individuals, detection of *Parachlamydia sp.* was associated with interaction with farm animals (133), supporting a potential role of this bacteria as a zoonotic agent. Further, *Parachlamydia* and other CRBs have been isolated from cattle drinking water (183), suggesting a possible source of infection and mode of transmission for this pathogen to cattle and potentially humans. Maternal-foetal transmission of *Parachlamydia* has also been demonstrated in a case study and was postulated to be a result of zoonotic transmission (178).

As previously mentioned, *W. chondrophila* was originally isolated from an aborted bovine foetus (8) and has since been described as an abortigenic agent in a number of studies on adverse

pregnancy outcomes in cattle throughout Europe, using both molecular and serological methods (176, 184). The disparity between reports from different authors highlights the prospect of an unknown determinant of susceptibility to infection and/or progression to disease, a common trend throughout the phylum. In humans, *W. chondrophila* has been reported in association with miscarriage and other adverse pregnancy outcomes in up to 30% of cohorts studied (386 and 438 women in two separate studies) (185, 186). *W. chondrophila* was also shown to multiply inside endometrial cells (187), further confirming its pathogenicity toward humans. At 96 h post-infection, the bacteria transform into persistent enlarged aberrant bodies that could be linked to recurrent episodes of miscarriage. Studies that show some evidence of acute or previous *Waddlia* infection in women who have miscarried suggests a possible reactivation of a latent asymptomatic infection, further strengthening this argument (185). While initial studies focused on cervicovaginal swabs, a recent study detected *W. chondrophila* in a placenta from miscarriage by both PCR and IHC (188), providing convincing evidence of a pathogenic role of *Waddlia* in abortion. What does remain to be established is the route of entry and transmission of this pathogen, as well as the underlying mechanism of pathogenesis. Studies of well water suggest that this could be one potential reservoir (189), while other authors hypothesize that routes of entry could be sexual transmission or via the bloodstream following a respiratory infection (185), or acquired following contact with animals (178). Most recently, *Waddlia* EBs were isolated from bovine placenta 6 months after experimental infection of pregnant cattle. Interestingly, *W. chondrophila* DNA was detected in the lung of the same animal, giving further strength to the notion that this anatomical site could be a local reservoir for *W. chondrophila* within the host (190).

Additionally, two novel *Waddlia* species have been isolated from or detected in the urine (191) or tissues (192) of bats in Malaysia and Mexico, respectively, expanding the diversity, pathogenic potential and zoonotic risk of members of *Waddliaceae*.

Two chlamydial species belong to a family closely related to the *Chlamydiaceae* were found to be highly prevalent in several amphibian hosts. In captive salamander and newt populations, disease manifested as anorexia, lethargy, and oedema, with high levels of leukocyte infiltration in liver sections (193). The aetiological agent was determined to be a novel species of *Chlamydiales*, *Ca.* Amphibiichlamydia salamandre, in the absence of any other commonly described viral, fungal bacterial pathogens. Conversely, in a population of bullfrogs from the Netherlands, another novel taxon, *Ca.* Amphibiichlamydia ranarum was identified in 71% of animals, without any associated morbidity (194). These two studies highlight again how little is known about chlamydial diversity throughout the animal kingdom, whilst also suggesting that certain clades may have preferences for certain hosts, although much further study is required to confirm this hypothesis.

Uncultured and unclassified clades represent a wealth of uncovered diversity within the Chlamydiae

Beyond the use of 16S rRNA sequencing as the first tool for the characterisation, classification and phylogenetic analysis of the recently described CRBs, researchers are beginning to make use of the ever-expanding metagenomic and amplicon datasets that are available in diverse databases to discover new CRBs.

Recently, Lagkouvardos et al. (2014) used advanced, high stringency bioinformatic analyses to interrogate over 22,000 high quality, non-redundant chlamydial 16S rRNA gene sequences found in diverse databases. Using near-full-length sequences and a conservative clustering approach, 17 family-level lineages supported by two or more isolates were identified, and this number increased to 28 when considering families represented by only one isolate. Similar analysis of the V4–V6 region of the 16S rRNA gene in over 12 000 amplicons resulted in 181 putative families each supported by at least two isolates. At the species level, a potential 1161–2276 OTUs were represented, depending on the bioinformatic method used (195). Ecological analyses revealed the majority of OTUs belong to marine and freshwater environments and only a small proportion (2%) arose from terrestrial environments (195).

2. Genomics of *Chlamydiae*

The application of genome sequencing revolutionised our understanding of chlamydial biology, especially since genetic manipulation systems were lacking, and are still in their infancy (196). Chlamydial genomics studies hence circumvent several of the issues faced in understanding chlamydial biology.

Sequencing of *C. trachomatis* paved the way for chlamydial genomics (197). Shortly after its characterisation, the genomes of *C. pneumoniae*, *C. muridarum and C. caviae*, were all characterised allowing comparisons between the former "*Chlamydia*" and "*Chlamydophila*" genera (198-201). These comparisons were the first to offer explanations of tissue tropism for different species, which later expanded into strain-level tissue or organo-tropisms, with *C. pneumoniae* genomes being sequenced from human artherosclerotic lesions, respiratory samples and koala conjunctival tissue, (35, 36, 202-204) and *C. trachomatis* strains sequenced from both the ocular and genital sites (205-207). Interestingly, the chlamydiae are unique in that the addition of more strains for comparative genomics does not seem to add much to the pan-genome. Rather, sequencing of large numbers of isolates, such as for the human pathogens *C. psittaci* and *C. trachomatis*, and agriculturally important *C. pecorum* and *C. abortus* offers opportunities to investigate the widespread global diversity and intraspecies evolution of these human and veterinary pathogens, as well as comparisons of endemic and epidemic strains.

Sequencing of CRB genomes followed, about 14 years after their discovery, enabling phylum-wide comparisons to accelerate our understanding of chlamydial biology. Interestingly, all CRB genomes characterised to date were sequenced from amoebal endosymbionts, isolated using amoebal co-culture, whilst none have been characterised from animal hosts.

2.1. Chlamydia *genomes*

Chlamydia genomes display a striking level of gene order conservation (shared synteny) across the species, with the major gene order rearrangements seen between the former "*Chlamydia*" and "*Chlamydophila*" genera, ie. *C. muridarum* and *Ca.* C. ibidis. (208). Deviations from this are mainly seen in the plasticity zone – a recombination "hotspot" situated near the replication termination region – and families of membrane proteins (199). Chlamydial genome sequences now encompass all recognised species, and this section aims to summarise studies that have attempted to describe the genomic features underpinning the unique biology of chlamydiae.

Size and reduced metabolic capacity

Chlamydiaceae genomes are small at around 1.04-1.24 Mbp (Table 1), encoding roughly 898 to 1097 genes, with the coding density around 90%. The small size and high coding density is common among obligate intracellular pathogens such as *Mycoplasma, Rickettsia* and *Buchnera* (209, 210). The reductive evolution characterised by irreversible gene loss exhibited by *Chlamydia sp*. has been associated with metabolic parasitism during adaptation to an intracellular habitat (211).

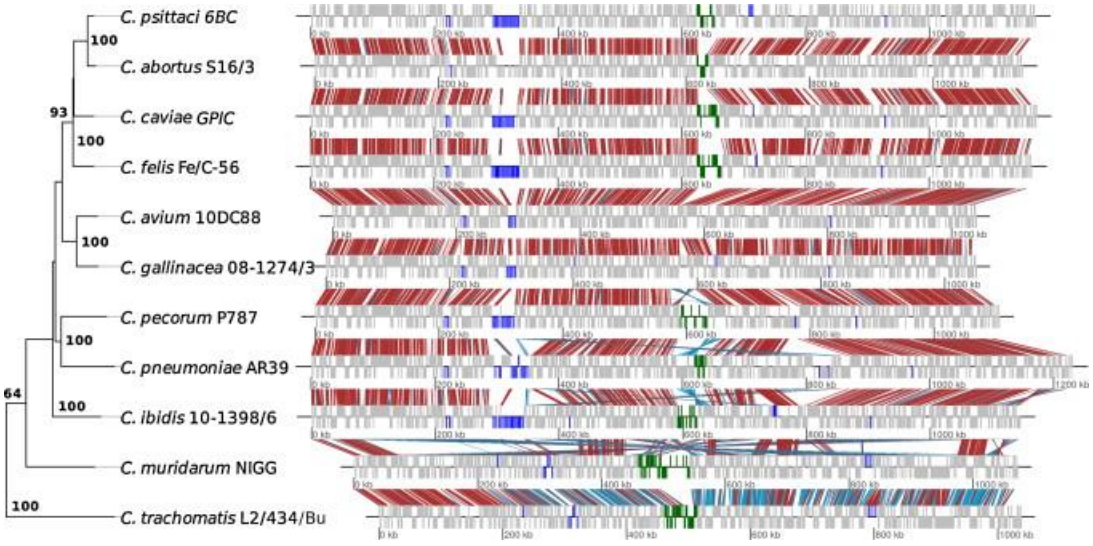


Figure 4: Shared genomic synteny within Chlamydia *genomes (Knittler et al., 2014).*

A large comparative genomics study of bacteria with different lifestyles revealed a convergent evolution of obligate intracellular bacteria, independent of phyla (212). This niche restriction is

said to be accompanied by less efficient purifying selection, which may lead to the accumulation of slightly deleterious mutations, explained by Muller's ratchet theory (213), with higher intraspecies dN/dS ratios than inter-species ratios (214).

The smaller genomes (compared to free-living bacteria) are characterised by limited metabolic capacity. Chlamydiae were previously referred to as "energy parasites" (215), as they obtain ATP from their host, a finding that was confirmed by initial genome sequencing efforts that revealed high concentrations of ATP-binding cassette transporters, including several copies of an ATP/ADP carrier protein (216). However, subsequent analyses also suggested that they can generate ATP from the glycolysis pathway and oxidative phosphorylation (197, 198, 217). Other transporters are common in the chlamydia genomes, for molecules such as amino acids, peptides and metal ions, and these are thought to replace genes required for *de novo* synthesis of such products (197-199, 201, 218).

Similarly, genes encoding essential functions in aerobic respiration are present in *Chlamydia* genomes, and electron flux may be supported by pyruvate, succinate, glycerol-3-phosphate and NADH dehydrogenases, NADH-ubiquinone oxidoreductase and cytochrome oxidase (197, 198). The TCA cycle however, is incomplete as it lacks three key components: citrate synthase, aconitase and isocitrate dehydrogenase, consistent with early observations that acetyl-coA generated by glycolysis or fatty acid degradation cannot enter the TCA cycle (220).

Table 1: *Genome characteristics of members of the Chlamydiales*

Species, strain	Chromosome size (Mbp)	Plasmid size (Kbp)	Predicted CDSs	Hypothetical proteins	% CDS unknown function	% GC	Reference(s)
Chlamydia-related bacteria							
Simkania negevensis "Z"	2.5	132	2519	1374	55%	38	(Collingro, et al., 2011)
Waddlia chondrophila WSU 86-1044	2.1	15.5	1934	784	41%	43	(Bertelli, et al., 2010)
Parachlamydia acanthamoebae UV-7	3.1	Np	2788	1812	65%	39	(Collingro, et al., 2011)
Protochlamydia amoebophila UWE25	2.4	Np	1986	1245	63%	35	(Horn et al., 2004)
Protochlamydia naegleriophila KNIC	2.8	145	2415	1095	45%	43	(Bertelli et al., 2016)
Neochlamydia sp. TUME1	2.5	Np	2043	1323	65%	38	(Ishida et al., 2014)
Rubidus massiliensis Rubidis	2.5	39, 80	2446	991	41%	33	(Bou Khalil et al., 2016)
Criblamydia sequanensis CRIB-18	2.9	89.5	2426	641	26%	38	(Bertelli, et al., 2014)
Estrella lausannensis CRIB-30	2.8	9.1	2213	677	31%	48	(Bertelli, et al., 2015)
Chlamydiaceae							
C. trachomatis A/HAR-13	1	7.5^	911	294	32%	41	(Stephens, et al., 1998)
C. pneumoniae AR39	1.2	Np^	1112	537	48%	40	(Read et al., 200)

Plasticity zone

The chlamydial plasticity zone (PZ) is a region of extensive variation between chlamydial genomes (199), which, while highly variable, generally harbours (i) acetyl coA chains (*accBC*), (ii) cytotoxin genes/adherence factor (*tox*), (iii) phospholipase D (*pld*), (iv) membrane attack complex/perforin (*MACPF*), (v) tryptophan biosynthesis operon (*trpABFCDR*, *kynU*, *prsA*) and (vi) purine biosynthesis genes (*guaAB-add*) (196, 219). Until recently, iterations of the chlamydial PZ had been described in all *Chlamydia* genomes sequenced to date, ranging from ~12 Kbp to ~86 Kbp, with 11 to 48 genes (196, 219).

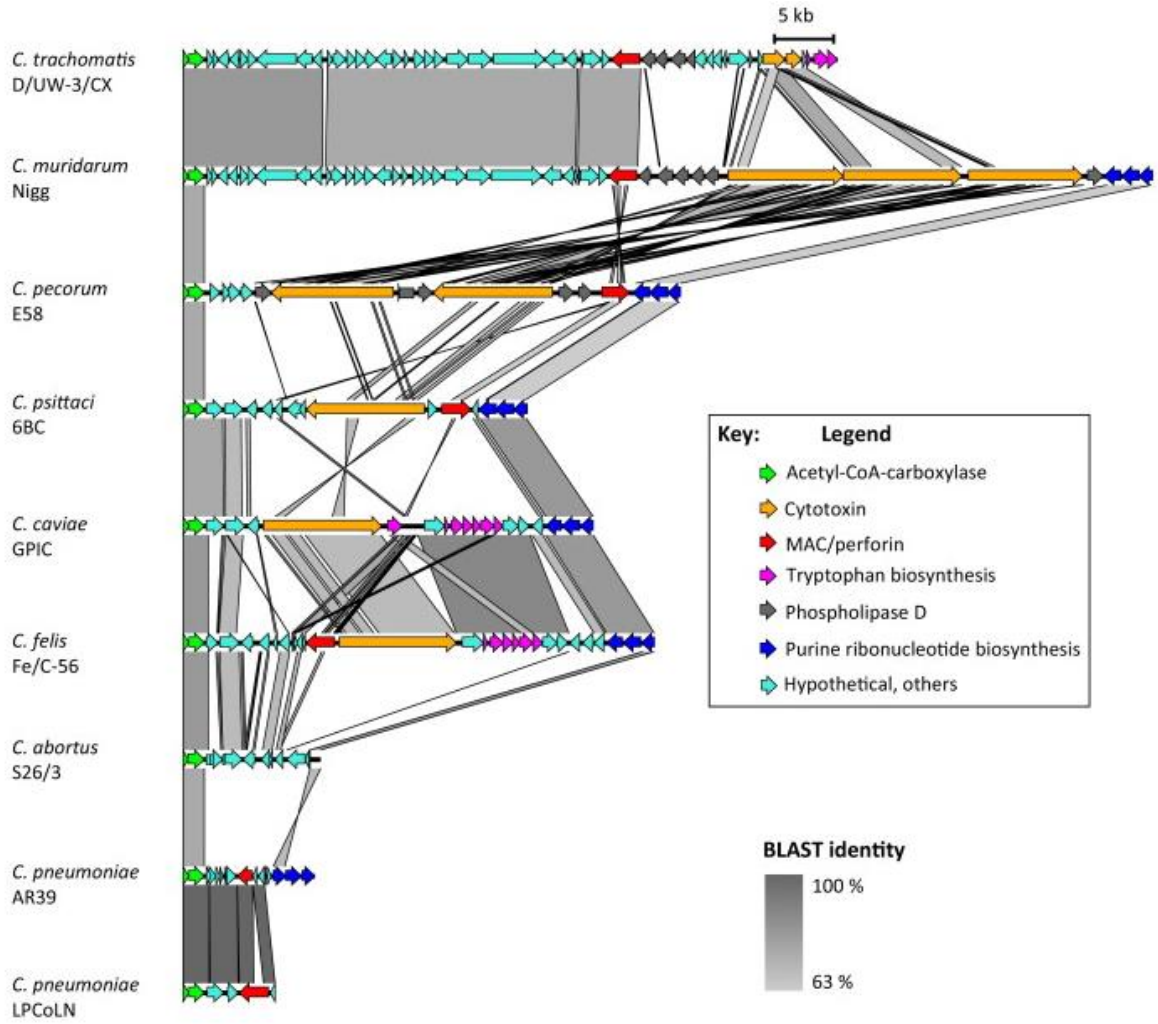


Figure 5: Diversity in the chlamydial plasticity zone (Bachmann et al., 2014)

Several studies have implicated the PZ genes in species-specific adaptation, tissue tropism and virulence. Read *et al.* speculated that differences observed in the *tox* gene between the 6BC clade and other clades of *C. psittaci* could account for host range expansion, which it potentially acquired through a single HGT event prior to clonal expansion. Similarly, Jelocnik *et al.* showed that only pig strains of *C. pecorum* have three copies of *tox*, compared to one full length and one pseudogene in cow, sheep and koala strains, potentially suggesting host adaptation mechanisms. The gene has been shown to be expressed in early *C. trachomatis* and *C. muridarum* studies, which showed that *tox*-carrying strains had cytopathic effects in HeLa cells (222). A recent *C. trachomatis* study implicated the MACPF domain family protein and the two PZ-encoded PLD genes in tissue tropism based on phylogeny (223), backing up a study by (224), that showed MACPF is secreted by *C. trachomatis* and is associated with PLD localisation to intra-inclusion lipid droplet formation.

Peptidoglycan genes

Peptidoglycan is an essential structural component in the cell walls of most if not all bacteria (225), providing a barrier around the cell membrane and a stable anchor for membrane transporters and other integral membrane proteins (226), which are abundant in chlamydiae. Previous dogma in the *Chlamydia* field held that, despite being sensitive to beta-lactam antibiotics, *Chlamydia* species lacked peptidoglycan, because all attempts to purify it had been unsuccessful (227). With the advent of genomics, researchers were finally able to understand the "chlamydial peptidoglycan anomaly". The *C. trachomatis* genome and transcriptome respectively revealed the presence of genes encoding the entire pathway for peptidoglycan biosynthesis and assembly (21, 197, 228), and developmental-stage dependent expression with peptidoglycan genes upregulated during the transition phase between EBs and RBs (21). Genomic studies were then followed up by the development of a new cell wall labelling method (229), the use of which has found that peptidoglycan is present as low levels in the chlamydial cell wall.

Chlamydia-specific virulence genes

Aside from the type three secretion system proteins, which are discussed in more detail in 2.2.2, comparative genomics studies have identified two families of membrane proteins with considerable research interest due to their high level of diversity within and between species.

(i) Polymorphic membrane proteins

Polymorphic membrane proteins (*pmps*) are integral membrane proteins unique to the *Chlamydiae* that confer antigenic diversity. Pmps have been described as chlamydial virulence factors, as well as used as vaccine candidates (230). As the name suggests, Pmps display varying levels of sequence homology but are united by their GGAIX(n)FXXNX(n) structure (231). Interestingly, they were initially identified by immunogenicity assays in *C. abortus* (232), but ensuing genomics studies identified nine to 21 genes in each species, with a recent gene duplication event said to account for the larger numbers of *Pmp*G derivatives in *C. pneumoniae* (35, 196, 219). Pmps constitute 3-5% of the coding capacity, further highlighting their importance in chlamydial biology despite massive genome reduction.

(ii) Inclusion membrane proteins

Inclusion membrane proteins are effectors secreted into the inclusion membrane upon invasion to maintain inclusion membrane structure and modulate host cell processes (24). The transmembrane helix (*tmh*) family, to which the inclusion membrane proteins (Incs) belong, is another highly variable protein family unique to the *Chlamydiaceae*. Like the pmps, Incs vary in number and length, and exhibit far lower levels of amino acid similarity than that observed across the full genome. Up to 107 Incs have been described in chlamydial genomes, but their lack of sequence homology can make them difficult to detect *in silico*.

Genome-based phylogenetic analyses

Ongoing genomics studies which build on previous single-or multi-gene typing studies are slowly unravelling the global population framework of several *Chlamydia* species (30, 43, 221, 233-235). Genome-based phylogenetic analyses have revealed greater strain-level diversity than analogous single or multi-gene phylogenies. Perhaps some of the strongest evidence in favour of utilising whole genome comparative analysis for phylogenetics studies is a recent *C. trachomatis* study that showed the whole genome phylogeny is not entirely congruent with previously described *ompA* phylogenies, and in fact, varies across the species. The primary factor that generates this *ompA* gene bias in phylogenetics is that it is under positive selective pressure and has been subject to recombination (30, 236).

The lower genetic diversity seen in *C. abortus* and other species could reflect the high level of clonality within the species, or evolutionary bottlenecks (233). Conversely, long branch lengths within the *C. suis* whole genome phylogeny indicate that a high level of diversity is maintained in population studies (234). The high level of diversity seen in *C. pecorum*, *C. psittaci* may represent the global spread of these organisms with their hosts based on natural segregation and anthropogenic factors such as trade and farming, as well as high levels of recombination (200, 237). Genome-wide phylogenetic analysis of *C. psittaci* first revealed the presence of a clonal, highly virulent "6BC" clade, originating in parrots, phylogenetically separate from other clades due to recombination between the clades, but not within (55). This phylogenetic separation could not be resolved or explained by single- or multi-gene epidemiological studies alone.

*2.2. CRB genomes and the pan-*Chlamydiales *genome*

The first CRB genome was sequenced in 2004, *P. amoebophila* (104), presenting the first opportunity for researchers to pry open the genetic secrets of CRBs. Genomes for the majority of the CRBs are lacking, however, the expanding number of available chlamydial genome sequences has allowed researchers to explore the evolutionary history of the *Chlamydiae* by tracing the

acquisition of particular genes. Distribution of the coding sequences (CDSs) in the *P. amoebophila* UWE25 genome suggest gene acquisition over a number of lateral gene transfer events with a last common ancestor that was intracellular but less dependent on host metabolism (104). Some virulence factors are conserved throughout the phylum, while others seem unique to certain families.

Size and expanded metabolic capacity

In contrast to the compact, highly conserved genomes exhibited by members of the *Chlamydiaceae*, CRB genomes are at least twice the size (2.1-3.1 Mbp), encoding up to three times the number of genes (1986-2788), highlighting their lack of genome reduction compared to the *Chlamydiaceae (*Table 1). In parallel with their phylogenetic distance from the *Chlamydiaceae* and indicative of their biological differences, CRB genomes are far more diverse, as a reflection of their preference for, primarily, an amoebal cell niche, which would provide a far less homeostatic environment compared to multicellular eukaryotes, with far fewer available nutrients. Retention of large gene sets that are shared between CRB genomes may also allow these species to jump between not only amoebal hosts but also higher eukaryote hosts, if opportunity arises or if conditions become unfit for replication.

Further, the high level of genomic synteny observed in the *Chlamydiaceae* is not observed in other families, although, this could be a symptom of the lack of genomic data for strains and species of CRBs, or a reflection of the monophyletic nature of the *Chlamydiaceae*. Nonetheless, there are a large number of genome rearrangements between species such as *P. amoebophila* and *P. naegleriophila*.

As one example of this broader host range and larger genome size, CRBs possess a higher biosynthetic capacity than traditional chlamydiae (238, 239), suggesting a less reliant dependency on the host cell. For instance, several amino acid synthesis pathways are far more complete in the

CRB genomes than in the *Chlamydiaceae* (240). Genes involved in menaquinone production for oxidative phosphorylation are also present, further differentiating CRB metabolic capacity from *Chlamydiaceae*.

Pan-Chlamydiales *genome*

Like their *Chlamydiaceae* brethren, over half of the CDSs in the CRB genomes have an as yet unknown function. Collingro*, et al.* (2011) identified 560 "core genes" common across the *Chlamydiaceae* and the CRBs with genomes available at the times (representing *Chlamydiacae*, *Parachlamydiaceae*, *Waddliaceae* and *Simkaniaceae*). Psomopoulos *et al.,* (2012) identified members of 304 protein families to belong to the 'core' gene set of the *Chlamydiae*, with the addition of more species and strains in while larger genomes correlated with a higher number of unique genes (Table 3). More recently, with the addition of *Criblamydiaceae* members, Pillonel *et al* established 450 core genes in the *Chlamydiales*. It appears this number will continue to decrease as more CRB genomes are characterised, particularly those that form the deeply branching families such as *Ca.* Piscichlamydiace and *Ca.* Parilichlamydiaceae. Among the conserved genes are (i) housekeeping genes that are presumably involved in the highly conserved intracellular lifestyle and unique developmental cycle (240), (ii) genetic processing loci (241) and (iii) members of all 100 clusters of orthologs, which are conserved among all intracellular bacteria.

The type three secretion system (T3SS) is a gene cassette common to several Gram-negative bacteria that confers the ability to sense eukaryotic cells and secretes effector proteins in order to fuse with the host cell membrane, and thus infect the cell via a needle-like injection mechanism. These structures were first observed as projections in the *C. caviae* cell membrane (242), and the genes encoding it were later confirmed to be situated on four distinct loci throughout the genome (20, 243). A high number of structural and chaperone components of the T3SS are conserved between the CRBs and the *Chlamydiaceae,* including inner membrane proteins and needle formation proteins (240, 244).

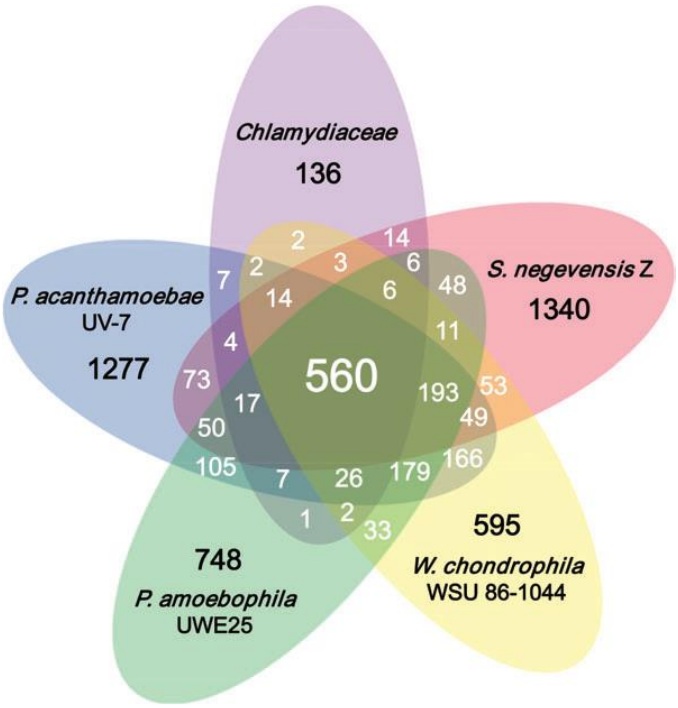


*Figure 6: Pan- and core-*Chlamydiales *genome (Collingro et al., 2011)*

However, the *Chlamydiaceae* genomes also encode some flagellar proteins that are missing from the CRB genomes which could have implications for intracellular survival (240). Despite many structural components of the T3SS being encoded by the CRBs, many effectors recognized in the *Chlamydiaceae* are not. Of particular interest are the Inc proteins, of which just three are conserved throughout the *Chlamydiae* (240) and TARP, a translocated actin-recruiting protein implicated in chlamydial invasion (245), which is absent from the CRB genomes, suggesting a host cell entry mechanism distinct from that of the *Chlamydiaceae*. The observed conservation of T3SS structural proteins coupled with the lack of homologous effector proteins indicates that all chlamydiae share some mechanisms for host cell entry but differ greatly in their host survival mechanisms, perhaps reflecting the varying pressures placed on these organisms by their diverse range of host cells.

Family, genus and species-specific genes

A primary difference between the *Chlamydiaceae* and the CRB families is the monophyly of the *Chlamydiaceae*, which currently allows assumptions to be made about the whole family based on

the single genus. As mentioned, the CRB families are far more diverse, each comprising several genera. Interestingly but not surprisingly, genus-specific features have emerged throughout these families, suggesting ongoing evolution toward genus- and species-specific host adaptation, tissue tropism or pathogenicity.

One of the more intriguing sets of species-specific genes is the methyl-accepting chemotaxis proteins and accompanying genes which make up an intact chemotaxis system similar to that of *Escherichia coli*. The system is thought to be fully functional due to both the lack of mutations and the expression of these genes during growth of *P. acanthamoebae,* however, it is unclear what the exact role of this system is in the absence of a flagellar system. Sequence homology suggests it may have been acquired by LGT from *Cyanobacteria* and/or *Clostridia* (240).

Within the same family, two strains of *P. naegleriophila* were recently described as having both a CRISPR locus and an upstream *cas* operon. The CRISPR locus was identified in *Neochlamydia* strains in the absence of the *cas* genes, and the gene arrangement surrounding these loci is conserved in the *Parachlamydiaceae*, suggesting that this system was acquired after the divergence of *P. naegleriophila* and *P. amoebophila*, potentially from *Anaeromyxobacter dehalogenans* or *Rhodothermus marinus* (246).

The chlamydial outer membrane is crucial for host cell adhesion and invasion. The *Chlamydiaceae* possess a unique outer membrane complex, the main component of which is the major outer membrane protein (MOMP). *S. negevensis* encodes 35 MOMP-like proteins, while *W. chondrophila* encodes 11 (238, 240). *Parachlamydia spp.* and *Neochlamydia spp.* are almost devoid of MOMP-like proteins, while *Protochlamydia spp.* appears to have replaced them with structurally similar porin proteins (17, 247). Similarly, a lower number of Pmps are encoded in all CRB genomes (240, 248).

Paralogs are reasonably rare within the *Chlamydiae*, and indeed within the genomes of obligate intracellular bacteria with highly reduced genomes. In the *S. negevensis* genome, Domman *et al.*

described 277 expansion events, comprising many small gene families which are the result of many small-scale transfers or duplications. In the *Parachlamydiaceae*, fewer expansion events have led to larger expansions in fewer gene families: four large gene families were identified between *Neochlamydia sp.* and *Protochlamydia sp.* Most of these expanded sets of genes display motifs involved in protein-protein interaction and/or are common to eukaryotic organisms (249). Specifically, many have domains involved in eukaryotic ubiquitin-related enzymes, suggesting these gene families may play roles in protein degradation or alteration, and hence be involved modulating the host cell. Lastly, many genus- or species- specific genes are encoded on very unique plasmids, or are hypothetical proteins with unknown function (discussed in more detail in the following section).

2.3. Genetic exchange

Recombination

Due to its obligate intracellular nature, *Chlamydia sp.* were long thought not to recombine their genomes at either the strain or species level. The first evidence for chlamydial recombination actually came prior to genome sequencing, when Hayes *et al.* described "hybrid" LGV *C. trachomatis* strains in African patients that did not react with any of the serovar-specific monoclonal antibodies used for serotyping (236). This evidence, coupled with the sequencing, cloning and expression of *RecA* in *C. trachomatis* (250) and subsequent identification of components of the Rec pathway (*RecBCDF*) in *C. trachomatis* genomes (197), solidified the idea that *Chlamydia* genomes can and do recombine.

Our understanding of chlamydial recombination was then further advanced by the use of wide scale comparative genomics. Millman *et al.* (2001) showed that the different phylogenetic signal exhibited by *ompA* compared to other gene loci suggested divergence of several *C. trachomatis* strains was driven by recombination. Their analysis also supported interspecies recombination in *ompA* between *C. muridarum* (MoPn/NiggII; mouse strain) and *C. pneumoniae* (N16; horse strain).

Gomes *et al.* built on previous studies that hinted at a rate of evolution of membrane proteins suggestive of horizontal DNA transfer, and used reference and clinical isolates to demonstrate that *C. trachomatis* undergoes extensive recombination throughout the chromosome, and identified breakpoint regions in the clinical isolates. More recent larger-scale genomic studies showed that 26% of the SNPs identified in *C. trachomatis* genomes sequenced at the time were acquired independently, and on physically linked blocks, most likely as a result of homologous recombination (30, 252).

Genome-wide recombination is also prevalent in *C. psittaci*, where members of the 6BC clade display identical recombination profiles (43). *C. suis* exhibits one of the highest recombination rates in the genus, probably due to/biased by the horizontal transfer of the *tet*C cassette (52), and most of the recombination events in this species are between the two main clades. In both species, recombination, while less frequent, contributes more to genome diversification than mutation (52, 55). Conversely, *C. abortus* strains sequenced to date, which cover a range of geographical locations and tissue types, exhibit a relatively low level of SNP density and hence only a few recombination events (233). Other examples of recombination throughout the phylum are given in the following section.

Genomic islands

Genomic islands (GIs) are regions of the chromosome that display different G+C% and codon usage signatures compared to the rest of the genome, indicative of horizontal transfer from a donor DNA sequence (253). Other hallmarks of GIs include flanking regions containing direct repeats or tRNAs, high abundance of hypothetical proteins or abundance of pathogenic proteins, and evidence of mobile elements or insertion sequences. Genomic islands are also often pathogenicity islands, and are seen throughout the bacterial kingdom (254). Acquisition of GIs facilitates transfer of DNA regions that confer evolutionary advantages to survival in specific host niches. As discussed above, due to the intracellular nature of chlamydiae, there are few

opportunities for direct interaction with other bacteria in the host cell cytoplasm, however, ancient chlamydiae may have been more susceptible to DNA exchange, perhaps in an amoebal or multicellular eukaryotic host.

A major advance in our understanding of genetic exchange in chlamydiae came from a *C. suis* study that found novel insertion sequence elements flanking the tetracycline resistance cassette (*tet*C) unique to this species (255), which in resistant strains is fused with an *inv*-like gene at a recombination hotspot, resulting in interruption of the gene. Based on sequence analysis, authors postulate that the *tet*C island was transferred from an *Aeromonas salmonicida* plasmid (255), a bacterial fish pathogen potentially acquired by pig strains of *C. suis* through feed, or from a Betaproteobacteria species (52). The subsequent intra-chlamydial transfer was shown to be replicated in *in vitro* experiments where the *tet*C cassette and flanking DNA was transferred to *C. trachomatis* and *C. muridarum* strains, but not to *C. caviae* strains (53).

The first sequenced CRB genome gave us another opportunity to examine potential GIs throughout the phylum. Horn *et al.* (2004) first described a 100 Kbp genomic region in *P. amoebophila* UWE25 with a higher than average G+C% encoding a type IV secretion system and several transposases indicative of a lateral gene transfer event that occurred after the divergence of the *Chlamydiaceae* and *Parachlamydiaceae* lineages. This region was later confirmed as a genomic island based on flanking tRNAs and direct repeats and the presence of phage-related proteins and other mobility genes (256). Smaller genomic islands are exhibited on the chromosomes of its closest relatives: *P. naegleriophila* strain KNic and strain Diamant, respectively harbour a 37 Kbp and 15 Kbp GI, encoding genes for DNA damage and repair, potassium transporter and high number of hypothetical proteins (246).

Extrachromosomal elements

(i) Chlamydia bacteriophages

Chlamydial bacteriophages were first discovered as a single-stranded circular DNA bacteriophage (Chp1) of the family *Microviridae* in *C. psittaci* (219). Subsequently, 5 distinct phages have been characterised in *C. abortus*, *C. pecorum*, *C. pneumoniae* and *C. caviae* (257). Their genomes are over 4.5 Kbp, with G+C% similar to that of chlamydial genomes, and are predicted to encode 8 or 9 ORFs (257).

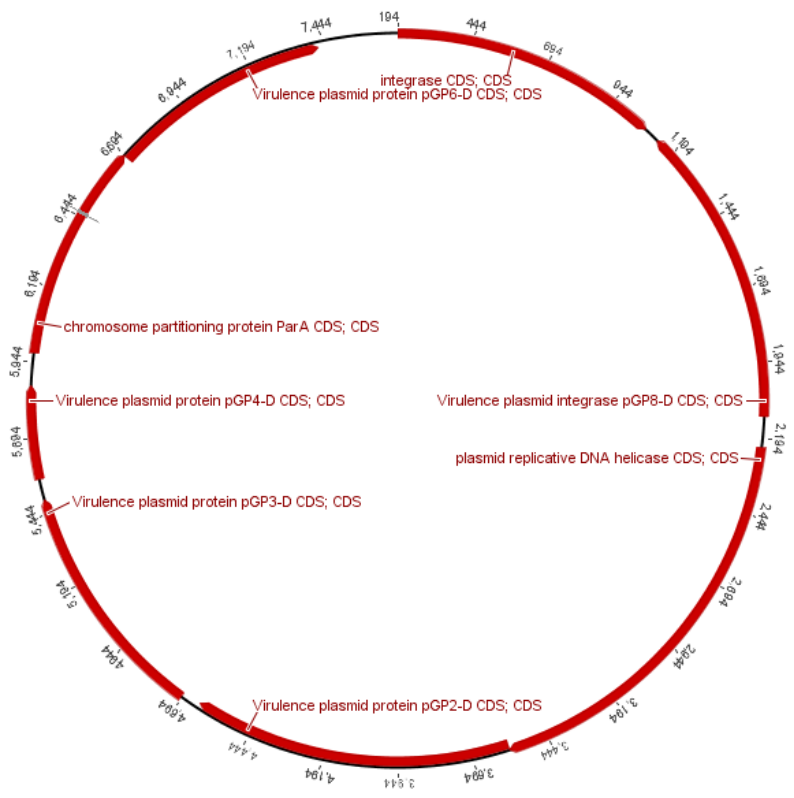
Phage genome remnants have been found in the chromosomes of phage-free chlamydial species/strains (201), suggesting that the phage was previously integrated into the chromosome, within the PZ, and lost during the speciation process. Some researchers have shown that phage actually delays peak infection in guinea pigs (258).

(ii) Chlamydial cryptic plasmid

Almost all *Chlamydia* species, but not all strains, are known to carry a plasmid (259), which is suggested to have co-evolved with the chromosome, based on congruent plasmid and chromosome phylogenies. *Chlamydia* plasmids are nonconjugative and nonintegrative but do carry virulence factors that have been used as vaccine candidates. *Chlamydia* plasmids are highly conserved in the genus and highly syntenic, normally organised with eight open reading frames, encoding for genes involved in plasmid maintenance and glycogen synthesis (260), summarised in Figure 7.

There is abundant evidence that the chlamydial plasmid is associated with glycogen accumulation in the host cell cytosol as a result of positive and negative transcriptional regulation by Pgp4 and Pgp5, respectively (261, 262). The down-regulation of *glgA*, a glycogen biosynthesis gene encoded on the chromosome, was seen in plasmid-negative *C. trachomatis* strains (263).

Figure 7: ORF arrangement in the Chlamydia *plasmid*



The presence of a plasmid has been suggested to contribute to the pathogenicity or tissue tropism of the chlamydial species (264, 265), but mixed findings from *in vitro* and *in vivo* studies highlight how little we know about the chlamydial plasmid and its role in pathogenicity. Several studies suggest that the plasmid is not required for growth *in vitro*, with no difference in morphology and similar susceptibility to antimicrobials. Meanwhile, *in vivo* studies demonstrated plasmid deficiency in either *C. muridarum* or *C. trachomatis* attenuates upper genital tract pathology in mice (266, 267) and ocular pathology in non-human primates (268). This plasmid-dependent pathogenicity is thought to be mediated by *pgp3* (264). Other studies showed that plasmid-deficient strains had high infective dose values, or that plasmid-free *C. muridarum* could induce tubal inflammation in mice (264).

Plasmid genome dynamics and our understanding of the true function of the chlamydial plasmid are further confounded by the fact that not all strains carry the plasmid, yet can still cause disease. For example, several genitotropic strains of *C. trachomatis* lack a plasmid, while others have a deletion in one of the potentially redundant replication proteins, which actually may result in

false negatives as a nucleic acids amplification test for *C. trachomatis* is designed in that region (269). The human strains of *C. pneumoniae* are plasmid-free, compared to the animal strains (36). *C. pecorum* plasmids are prevalent in koalas and livestock (55.8% and 38.6%, respectively), but the sporadic bovine encephalomyelitis strains were free of the plasmid (221). These findings suggest a possible role in host adaptation or restriction rather than virulence. Further, *C. abortus* and *Ca.* C. ibidis both lack a plasmid, and the *C. suis* plasmid is rare and truncated, suggesting it may be in the process of being inactivated or lost (196).

(iii) CRB plasmids

In stark contrast to the *Chlamydia* plasmid, but in line with the huge degree of genomic diversity throughout the phylum, CRBs harbour plasmids ranging in size and function. Unsurprisingly, the plasmids of each species carry genes unique to each that may be involved in species-specific host adaptation and provided clues to the evolution of the *Chlamydiae*. The main plasmid features (size, no. predicted coding regions, G+C%) are summarised in Table 1.

The *S. negevensis* and *P. naegleriophila* plasmids are the largest described so far at 132 Kbp, and 145 Kbp respectively, and are the only examples of a mobilisable F-type conjugative plasmids in the phylum (240, 246). These plasmids carry modules for plasmid stability, replication and propagation and host adaptation. Homologs of some of the plasmid-encoded elements are found on the chromosomes of other CRBs: parts of the *tra* subset of the type IV secretion system are both encoded by the *P. acanthamoebae and P. amoebophila* genomes, the latter of which encodes the operon on a genomic island and a *Mip* homolog is found on the *Simkania* plasmid, which is located on the chromosome in other species (240). The *P. naegleriophila* KNIC and *Rubidus masilliensis tra* regions are highly collinear, perhaps reflecting their phylogenetic relatedness (246, 270). The *tra* region exhibited by these CRBs, which commonly infect amoebae, appears to be a mosaic of F-type and Ti-type, the latter of which is a feature of *Agrobacterium tumefaciens*, a plant pathogen. These findings led to the hypothesis that a "megaplasmid", largely resembling the *S. negevensis* plasmid,

was acquired by the ancient chlamydial ancestor, potentially from *Rickettsiales*, before losing elements via gene loss or chromosomal integration (240).

The *W. chondrophila* plasmid displays a high level of strain-level diversity, with the 2032/99 strain encoding nine out of the 22 predicted genes from the WSU 86-1044 15.6 Kbp plasmid on its chromosome (238). The genes are encoded at different sites, suggesting integration, which is also evidenced by the fact that numerous small sequences, including transposases are shared between the WSU 86-1044 chromosome and plasmid, and also between the WSU-86-1044 plasmid and *P. amoebophila* genome.

A common feature of the CRB plasmids is their function as putative detoxification systems, presumably assisting their intra-amoebal lifestyle. The *C. sequannensis* plasmid encodes several cation efflux systems including one for arsenite, while one of the two *R. massiliensis* plasmids encodes resistance or efflux genes for antibiotics, tellurite and copper (270). Toxins are also commonly encoded on CRB plasmids. The *E. lausannesis* and second *R. masilliensis* plasmid encode a *RelF*/*MazF*, the endoribonuclease component of the two-part *RelEF*/*MazEF* toxin-antitoxin system, which is located on the *W. chondrophila* chromosome (238, 244, 270). This system is involved in plasmid maintenance during cell division.

3. Culture-independent approaches for genomic characterisation of *Chlamydiae*

Microbial community profiling and ecology analysis is proving to be a useful tool for uncovering novel diverse microbial taxa in the far reaches of our biosphere, many of which may have industrial uses or are human or animal pathogens. Initial microbial diversity studies involved culture-dependent methods, which are laborious and not always successful. Advances sequencing methodologies such as 454 pyrosequencing led to the understanding that there were many more microbes that were unable to be cultivated in the lab than those that were. The profiling tools quickly switched to culture-independent molecular methods, commonly using conserved regions of the 16S rRNA gene to characterise diversity in environments such as soil, sediment and water,

as well as human gut, skin and oral microbiomes. These studies revealed that cultivable bacteria only represent ~1% of the number of bacteria in a given sample (271). In the chlamydial context, as has already been discussed, these molecular methods have been instrumental in our understanding of the extent of chlamydial diversity (1, 134, 195).

These "gene-centric" methods come with limitations, such as primer bias and restriction to previously identified taxa. To overcome this, and better characterise the true diversity of these niches and others, and in particular gain a deeper insight into the viral and fungal diversity in addition to bacterial and archaeal diversity, there has been a shift toward deep metagenomic sequencing in which the entire DNA extract is subject to shotgun sequencing.

Advancing the field even further are depletion or enrichment methods for certain components of the DNA sample. Commercial kits have been developed that enable selection or depletion of specific DNA based on methylation properties or restriction enzymes while other methods involve separating target DNA from background DNA by immunomagnetic separation and fluorescence-activated cell sorting. These methods are often followed by whole genome enrichment to yield higher levels of the target bacterial DNA appropriate for genome sequencing.

3.1. Culture-independent genomics of Chlamydiae

Clinical samples often contain low numbers of pathogens, and to obtain sufficient material for WGS, intracellular bacteria such as chlamydia, *in vitro* culture is required. Due to the laborious nature of cultivating chlamydiae in host cells, the propensity for microbes to mutate after several passages (272) and the need for high throughput approaches for large-scale genomics analysis to "catch up" with the genomics studies of other bacterial groups, culture-independent molecular approaches are an attractive alternative gaining recognition and use throughout the chlamydial field.

Immunomagnetic-separation – multiple displacement amplification

Seth-Smith and colleagues broke new ground in 2013 when they sequenced novel *C. trachomatis* strains using an immunomagnetic separation (IMS) approach to select for chlamydial EBs in the swab samples. They coupled this with whole genome amplification (WGA) by multiple displacement amplification (MDA) to further amplify all DNA in the sample. Using primary mouse antibodies directed at chlamydial LPS, and secondary IgG conjugated to magnetic beads, the chlamydial EBs are separated from the sample material (273). Host DNA is removed by DNase treatment prior to MDA and sequencing. This method builds on previously described protocols for IMS for diagnostic purposes such as *Listeria* detection in food sources, and *C. trachomatis* in urine samples, and would also be suitable for use from cultured samples. Interestingly, around the same time, a single *C. trachomatis* genome was reported from a clinical swab (Ct value 25) subject to MDA alone (274). This yielded low coverage depth over the entire chlamydial chromosome, high coverage of the plasmid, high coverage of several vaginal bacterial species, and an over-abundance of human reads. This study highlighted the usefulness of WGA for metagenome analysis and metagenome-assembled genome assembly

Putman et al then applied the IMS-MDA method to 10 *C. trahomatis* strains directly from endo-cervical swabs. While lower IFU loosely correlated with more unsolved bases, the authors showed this method was highly sensitive, with whole genomes obtained from samples containing as little as 4 IFU per swab prior to enrichment (275). This study had two main findings that may have been overlooked had these samples been prepared and sequenced according to traditional methods (culture following by EB purification). In one sample, unresolved bases reflected a within-host clonal population, and in another, clonal variation at two nucleotide positions separated by 150,000 bp was suggestive of active recombination in response to selective pressures not present *in vitro* (based on culture and sequencing of the same sample).

RNA 'bait'-DNA hybridisation

C. trachomatis has also been the subject of another culture-independent approach which uses biotinylated RNA probes ('baits') to hybridise chlamydial DNA away from a complex DNA mixture (276); a method that has successfully yielded viral genomes from mixed samples. The 120-mer custom RNA bait sets are designed to span the entire chromosome and hybridisation occurs after library preparation. Like IMS-MDA, this method employs magnetic separation using streptavidin-coated beads to separate the "captured" (hybridised) DNA, prior to genome sequencing. The authors used this methods for both swab and urine samples, with a higher sensitivity from urine samples, potentially due to lower background DNA. Over 95% of the genome was covered by as little as ~1% of the reads in samples with as little as $3.9x10^4$ chlamydial genome copies, however clinical samples were sequenced twice, while cultured samples were sequenced once to obtain the same genome coverage, with far higher depth.

C. pecorum genomic studies using the same method have revealed that, as previously suspected, animals may be colonised with more than one strain at the same anatomical site (277). These whole genomes, obtained from koala ocular and urogenital swabs and bovine joints and ocular swabs, could be separated from each other based on abundance, and were phylogenetically distinct, separated by up to over 6000 SNPs. The presence of "minor strains" in clades dominated by Australian livestock supports ongoing exposure and potential cross-transmission of livestock strains into koalas from sheep and cattle. This finding would not have been possible if sequencing was conducted on cultured samples.

Metagenomic approaches

Whilst not published for any chlamydial species prior to this PhD project, a host DNA methylated DNA depletion approach, coupled with MDA as above, has gained popularity in circumstances where the pathogen is either a) unknown, b) uncultivable, c) in low abundance

and/or d) potentially novel. This method, although the most expensive of the three detailed here, yields the most information with regard to the entire metagenomic content at the site of infection.

This approach was used to characterise the genome of a plant pathogen, *Ca.* Liberibacter asiaticus (278), and subsequently the genomes of a novel genus of fish pathogens from Sea bream (279). Like the other methods, sequencing results are heavily dependent on the level and proportions of chlamydial/bacterial DNA in the starting sample, and in this case, bacterial DNA was further enriched by the use of laser-capture microdissection of cysts from the gill tissue.

A final approach, not included in Table 2, is the use of multiplexed microdroplet PCR which employs 500 overlapping 1-1.3 Kbp amplicons (280). Like IMS-MDA and bait-probe hybridisation, this method is heavily dependent on knowledge about the target species for specific assay design. It also requires very specific downstream bioinformatics methods. Nonetheless, it is applicable for *C. trachomatis* phylogenetic and typing studies, and is capable of discriminating between strains in a sample, however, areas of recombination are hard to resolve. This study also acts as a reminder that DNA extraction method is paramount to good quality sequencing, especially in low-abundant samples.

Table 2: *Summary of culture-independent genome sequencing methodologies used for* Chlamydiae

Category	**Targeted genome capture**		**Non-targeted (meta)genome capture**
Method	**IMS-MDA**	**Probe-bait hybridisation**	**Depletion-enrichment**
Molecular basis	Anti-LPS antibody binding coupled with whole genome amplification	Biotinylated RNA probe hybridisation to pathogen genome 'bait'	Depletion of host methylated DNA coupled with whole genome amplification
Sample collection	Swab stored in transport media	Swab, tissue, urine	Swab, tissue
Application	Initial sample, prior to DNA extraction	After total DNA extraction	After total DNA extraction
Novel species discovery	No[1]	No	Yes
Variant detection	Yes?	Yes	Yes
Multiple strains/ co-infection	Yes	Yes	Yes
Cost	$5-10 per sample	$?	$60-70 per sample

Speed (excl. DNA extraction and sequencing)	~5hr	24hr hybridisation	1-2 days (~ 8hrs hands-on time)
Sensitivity (for 100% genome coverage)	4 IFUs/swab (Putman et al., 2013) $1.1x10^5$ genome copies (post-IMS) (Seth-Smith et al., 2013)	~1000 genome copies/ swab (Bachmann et al., 2014) $3.3x10^5$, $5.5x10^6$ (urine, swab) (Christiansen et al., 2014)	Not reported
Specificity[#]	Up to 95% (Putman et al., 2013)		Not applicable >99%[2]
Limiting factors(s)/ requirement(s)	- Reference genome, - Suitable characterised surface-exposed antigen - Sample must be preserved in transport media	- Reference genome	- Relative abundance of microbial and eukaryotic DNA in sample

Species applied to	*C. trachomatis*	*Wolbachia sp., C. trachomatis, C. pecorum*	*Ca.* Liberibacter asiasticus, *Ca.* Ichthyocystis spp.
References	Seth-Smith et al., 2013, Putman et al., 2013	Melnikov et al., 2011*; Christiansen et al., 2014;	Zheng et al., 2014; Seth-Smith et al., 2016*;

*Not chlamydial genomics studies

[#] (highest % of non-chlamydial reads still allowing 90-100% chlamydial genome coverage)

[1] Only if chosen antigen is present in novel species and is recognised by the antibody

[2] Achieved by two sequencing runs totalling $7.8x10^7$ reads from ***Ca.*** Liberibacter asiaticus; 175pg/uL ***Ca.*** L. asiaticus DNA

4. Lessons from the literature and concluding remarks

Despite historical and ongoing research efforts to characterise the diversity of the *Chlamydiales*, much of the diversity – even within the *Chlamydiaceae* – remains to be uncovered, let alone cultivated and characterised. Indeed, analysis of sequence databases reveals a plethora of uncultivated, unclassified chlamydial taxa that would "fill in" the *Chlamydiales* phylogenetic tree (195) highlighting how little we know about the true ecological, biological and genomic diversity of the *Chlamydiae*. Given recent descriptions of novel chlamydial taxa from avian, piscine and reptilian hosts, it follows that these hosts would be ideal reservoirs to investigate and characterise novel species (25, 60, 64, 67, 134, 136, 140, 141, 281).

Investigations into chlamydial infections of reptiles have revealed an array of aetiological agents (*C. psittaci*, *C. pneumoniae* and uncultured *Chlamydiae*) associated with various disease presentations (granulomatous inflammation, wasting disease, myocarditis etc) (63-65, 73, 282). Most reports to date have been case studies, so little is known about chlamydial species and strains present in captive populations. No reptile-associated chlamydial lineages have been described, and no chlamydial genomes have been characterised from any reptilian host. However, it is suspected that *C. pneumoniae* may have originated in amphibians or reptiles (36, 37), and characterisation of closely-related species may provide clues to the evolution of *Chlamydia* and its adaptation to different host niches.

Further, chlamydial infections of fish have been described frequently across the globe, usually in aquaculture settings, presenting an economically significant threat (135). Interestingly, these species appear specific to their fish hosts with chlamydial diversity increasing with every additional fish host that is investigated. Prior to the commencement of this project, nine species of *Chlamydiae* in five families outside the *Chlamydiaceae* were molecularly described in ten diverse fish hosts (135-138, 140-142, 148-149). More lineages were added to these families over the last three years, and others are represented by partial sequences only, while some studies did not

conduct molecular analyses (146, 283-285). Hence, the true phylogenetic diversity of these bacteria is likely underestimated, suggesting molecular screening studies should continue. Despite our increased understanding of these agents, in no case have Koch's postulates been fulfilled, as all attempts to culture the agent(s) *in vitro* have proven unsuccessful (135, 284, 286), and no animal infection models have been established for this infection. Researchers have instead relied on fulfilling Fredericks and Relman's postulates (287). This lack of culture system has been a barrier to understanding the biology and pathogenicity of these unique bacteria. Importantly, 16S rRNA sequencing and phylogenetic analysis place two of these families – *Ca.* Piscichlamydiaceae and *Ca.* Parilichlamydiaceae – at the root of the *Chlamydiales* phylogenetic tree, but genomic studies are needed to confirm their relationship with most recent common ancestors of this order, as well as to assess the pathogenic potential of these novel chlamydiae.

To date, 16S rRNA gene sequencing has been the most widely used tool for chlamydial phylogenetic analyses, with nearly all genetic information on novel species, such as those from piscine and reptilian hosts, limited to sequencing of near full-length fragments of this gene. Whilst this tool is useful for pathogen detection, phylogenetic analysis and epidemiological studies, more sophisticated tools are needed to understand the virulence and metabolic capacity of novel chlamydial pathogens. For most bacteria, this would be easily achieved by axenic culture, biochemical assays and genomic characterisation from pure culture. However, for novel species, *in vitro* culture systems are not readily available or can be laborious and costly to establish. With deep sequencing becoming faster, more affordable and higher throughput (288), research groups within the chlamydia field have recently developed several methods of deep sequencing from clinical samples to gain insight into the biology of these species (275-277, 289), in the absence of an established culture step that has thus far hampered genomic characterisation of novel pathogens throughout the *Chlamydiae* and in other intracellular bacteria. These methods have provided unprecedented ability to a) capture a high level of diversity in less time than traditional gene screening methods, and b) assemble whole genomes from complex samples,

without culture. The application of these methods to novel uncultivated chlamydial species is anticipated to provide insight into the unique host-pathogen interactions exhibited by chlamydiae in novel infections for which no culture system exists.

5. References

1. Everett KD, Bush RM, Andersen AA. 1999. Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol 49 Pt 2:415-440.

2. Schachter J, Stephens RS, Timms P, Kuo C, Bavoil PM, Birkelund S, Boman J, Caldwell H, Campbell LA, Chernesky M, Christiansen G, Clarke IN, Gaydos C, Grayston JT, Hackstadt T, Hsia R, Kaltenboeck B, Leinonnen M, Ojcius D, McClarty G, Orfila J, Peeling R, Puolakkainen M, Quinn TC, Rank RG, Raulston J, Ridgeway GL, Saikku P, Stamm WE, Taylor-Robinson DT, Wang SP, Wyrick PB. 2001. Radical changes to chlamydial taxonomy are not necessary just yet. Int J Syst Evol Microbiol 51:249; author reply 251-243.

3. Sachse K, Bavoil PM, Kaltenboeck B, Stephens RS, Kuo CC, Rossello-Mora R, Horn M. 2015. Emendation of the family *Chlamydiaceae*: proposal of a single genus, *Chlamydia*, to include all currently recognized species. Syst Appl Microbiol 38:99-103.

4. Pillonel T, Bertelli C, Salamin N, Greub G. 2015. Taxogenomics of the order *Chlamydiales*. Int J Syst Evol Microbiol 65:1381-1393.

5. Omsland A, Sixt BS, Horn M, Hackstadt T. 2014. Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. FEMS Microbiol Rev 38:779-801.

6. Kahane S, Metzer E, Friedman MG. 1995. Evidence that the novel microorganism 'Z' may belong to a new genus in the family *Chlamydiaceae*. FEMS Microbiol Lett 126:203-207.

7. Rurangirwa FR, Dilbeck PM, Crawford TB, McGuire TC, McElwain TF. 1999. Analysis of the 16S rRNA gene of micro-organism WSU 86-1044 from an aborted bovine foetus reveals that it is a member of the order *Chlamydiales*: proposal of *Waddliaceae* fam. nov., *Waddlia chondrophila* gen. nov., sp. nov. Int J Syst Bacteriol 49 Pt 2:577-581.

8. Dilbeck PM, Evermann JF, Crawford TB, Ward AC, Leathers CW, Holland CJ, Mebus CA, Logan LL, Rurangirwa FR, McGuire TC. 1990. Isolation of a previously undescribed rickettsia from an aborted bovine fetus. J Clin Microbiol 28:814-816.

9. Abdelrahman YM, Belland RJ. 2005. The chlamydial developmental cycle. FEMS Microbiol Rev 29:949-959.

10. McCafferty. https://chem.duke.edu/labs/mccafferty/bacterial-pathogenesis.

11. Hackstadt T, Baehr W, Ying Y. 1991. *Chlamydia trachomatis* developmentally regulated protein is homologous to eukaryotic histone H1. Proc Natl Acad Sci U S A 88:3937-3941.

12. Moulder JW. 1991. Interaction of chlamydiae and host cells in vitro. Microbiol Rev 55:143-190.

13. Haider S, Wagner M, Schmid MC, Sixt BS, Christian JG, Hacker G, Pichler P, Mechtler K, Muller A, Baranyi C, Toenshoff ER, Montanaro J, Horn M. 2010. Raman microspectroscopy reveals long-term extracellular activity of *Chlamydiae*. Mol Microbiol 77:687-700.

14. Saka HA, Thompson JW, Chen YS, Kumar Y, Dubois LG, Moseley MA, Valdivia RH. 2011. Quantitative proteomics reveals metabolic and pathogenic properties of *Chlamydia trachomatis* developmental forms. Mol Microbiol 82:1185-1203.

15. Sixt BS, Heinz C, Pichler P, Heinz E, Montanaro J, Op den Camp HJ, Ammerer G, Mechtler K, Wagner M, Horn M. 2011. Proteomic analysis reveals a virtually complete set of proteins for translation and energy generation in elementary bodies of the amoeba symbiont *Protochlamydia amoebophila*. Proteomics 11:1868-1892.

16. Aistleitner K, Heinz C, Hormann A, Heinz E, Montanaro J, Schulz F, Maier E, Pichler P, Benz R, Horn M. 2013. Identification and characterization of a novel porin family highlights a major difference in the outer membrane of chlamydial symbionts and pathogens. PLoS One 8:e55010.

17. Heinz E, Tischler P, Rattei T, Myers G, Wagner M, Horn M. 2009. Comprehensive in silico prediction and analysis of chlamydial outer membrane proteins reflects evolution and life style of the *Chlamydiae*. BMC Genomics 10:634.

18. Kahane S, Everett KD, Kimmel N, Friedman MG. 1999. Simkania negevensis strain ZT: growth, antigenic and genome characteristics. Int J Syst Bacteriol 49 Pt 2:815-820.

19. Collingro A, Toenshoff ER, Taylor MW, Fritsche TR, Wagner M, Horn M. 2005. '*Candidatus* Protochlamydia amoebophila', an endosymbiont of *Acanthamoeba spp*. Int J Syst Evol Microbiol 55:1863-1866.

20. Betts-Hampikian HJ, Fields KA. 2010. The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. Front Microbiol 1:114.

21. Belland RJ, Zhong G, Crane DD, Hogan D, Sturdevant D, Sharma J, Beatty WL, Caldwell HD. 2003. Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proc Natl Acad Sci U S A 100:8478-8483.

22. Valdivia RH. 2008. Chlamydia effector proteins and new insights into chlamydial cellular microbiology. Curr Opin Microbiol 11:53-59.

23. Rockey DD, Scidmore MA, Bannantine JP, Brown WJ. 2002. Proteins in the chlamydial inclusion membrane. Microbes Infect 4:333-340.

24. Mital J, Miller NJ, Dorward DW, Dooley CA, Hackstadt T. 2013. Role for chlamydial inclusion membrane proteins in inclusion membrane structure and biogenesis. PLoS One 8:e63426.

25. Vorimore F, Hsia RC, Huot-Creasy H, Bastian S, Deruyter L, Passet A, Sachse K, Bavoil P, Myers G, Laroucau K. 2013. Isolation of a New *Chlamydia* species from the Feral Sacred Ibis (*Threskiornis aethiopicus*): *Ca.* Chlamydia ibidis. PLoS One 8:e74823.

26. Clarke IN. 2011. Evolution of *Chlamydia trachomatis*. Ann N Y Acad Sci 1230:E11-18.

27. Taylor HR, Burton MJ, Haddad D, West S, Wright H. 2014. Trachoma. Lancet doi:10.1016/S0140-6736(13)62182-0.

28. CDC. 2016. https://www.cdc.gov/std/stats15/chlamydia.htm.

29. Oakeshott P, Kerry S, Aghaizu A, Atherton H, Hay S, Taylor-Robinson D, Simms I, Hay P. 2010. Randomised controlled trial of screening for *Chlamydia trachomatis* to prevent pelvic inflammatory disease: the POPI (prevention of pelvic infection) trial. BMJ 340:c1642.

30. Harris SR, Clarke IN, Seth-Smith HM, Solomon AW, Cutcliffe LT, Marsh P, Skilton RJ, Holland MJ, Mabey D, Peeling RW, Lewis DA, Spratt BG, Unemo M, Persson K, Bjartling C, Brunham R, de Vries HJ, Morre SA, Speksnijder A, Bebear CM, Clerc M, de Barbeyrac B, Parkhill J, Thomson NR. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet 44:413-419, S411.

31. Fehlner-Gardiner C, Roshick C, Carlson JH, Hughes S, Belland RJ, Caldwell HD, McClarty G. 2002. Molecular basis defining human *Chlamydia trachomatis* tissue tropism. A possible role for tryptophan synthase. J Biol Chem 277:26893-26903.

32. Caldwell HD, Wood H, Crane D, Bailey R, Jones RB, Mabey D, Maclean I, Mohammed Z, Peeling R, Roshick C, Schachter J, Solomon AW, Stamm WE, Suchland RJ, Taylor L, West SK, Quinn TC, Belland RJ, McClarty G. 2003. Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. J Clin Invest 111:1757-1769.

33. Carlson JH, Porcella SF, McClarty G, Caldwell HD. 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. Infect Immun 73:6407-6418.

34. Teh B, Grayson ML, Johnson PD, Charles PG. 2012. Doxycycline vs. macrolides in combination therapy for treatment of community-acquired pneumonia. Clin Microbiol Infect 18:E71-73.

35. Roulis E, Polkinghorne A, Timms P. 2013. *Chlamydia pneumoniae*: modern insights into an ancient pathogen. Trends Microbiol 21:120-128.

36. Myers GS, Mathews SA, Eppinger M, Mitchell C, O'Brien KK, White OR, Benahmed F, Brunham RC, Read TD, Ravel J, Bavoil PM, Timms P. 2009. Evidence that human *Chlamydia pneumoniae* was zoonotically acquired. J Bacteriol 191:7225-7233.

37. Rattei T, Ott S, Gutacker M, Rupp J, Maass M, Schreiber S, Solbach W, Wirth T, Gieffers J. 2007. Genetic diversity of the obligate intracellular bacterium *Chlamydophila pneumoniae* by genome-wide analysis of single nucleotide polymorphisms: evidence for highly clonal population structure. BMC Genomics 8:355.

38. Mitchell CM, Hutton S, Myers GS, Brunham R, Timms P. 2010. *Chlamydia pneumoniae* is genetically diverse in animals and appears to have crossed the host barrier to humans on (at least) two occasions. PLoS Pathog 6:e1000903.

39. Laroucau K, Aaziz R, Meurice L, Servas V, Chossat I, Royer H, de Barbeyrac B, Vaillant V, Moyen JL, Meziani F, Sachse K, Rolland P. 2015. Outbreak of psittacosis in a group of women exposed to *Chlamydia psittaci*-infected chickens. Euro Surveill 20.

40. C DEB, Dehollogne C, Dumont A, Spierenburg M, Heijne M, Gyssens I, J VDH, Vanrompay D. 2016. Managing a cluster outbreak of psittacosis in Belgium linked to a pet shop visit in The Netherlands. Epidemiol Infect 144:1710-1716.

41. Ling Y, Chen H, Chen X, Yang X, Yang J, Bavoil PM, He C. 2015. Epidemiology of *Chlamydia psittaci* Infection in Racing Pigeons and Pigeon Fanciers in Beijing, China. Zoonoses Public Health 62:401-406.

42. Spoorenberg SM, Bos WJ, van Hannen EJ, Dijkstra F, Heddema ER, van Velzen-Blad H, Heijligenberg R, Grutters JC, de Jongh BM, Ovidius study g. 2016. *Chlamydia psittaci*: a relevant cause of community-acquired pneumonia in two Dutch hospitals. Neth J Med 74:75-81.

43. Branley J, Bachmann NL, Jelocnik M, Myers GS, Polkinghorne A. 2016. Australian human and parrot *Chlamydia psittaci* strains cluster within the highly virulent 6BC clade of this important zoonotic pathogen. Sci Rep 6:30019.

44. Wallensten A, Fredlund H, Runehagen A. 2014. Multiple human-to-human transmission from a severe case of psittacosis, Sweden, January-February 2013. Euro Surveill 19.

45. McGuigan CC, McIntyre PG, Templeton K. 2012. Psittacosis outbreak in Tayside, Scotland, December 2011 to February 2012. Euro Surveill 17.

46. Branley JM, Roy B, Dwyer DE, Sorrell TC. 2008. Real-time PCR detection and quantitation of *Chlamydophila psittaci* in human and avian specimens from a veterinary clinic cluster. Eur J Clin Microbiol Infect Dis 27:269-273.

47. Rohde G, Straube E, Essig A, Reinhold P, Sachse K. 2010. Chlamydial zoonoses. Dtsch Arztebl Int 107:174-180.

48. Ortega N, Caro MR, Gallego MC, Murcia-Belmonte A, Alvarez D, Del Rio L, Cuello F, Buendia AJ, Salinas J. 2015. Isolation of *Chlamydia abortus* from a laboratory worker diagnosed with atypical pneumonia. Ir Vet J 69:8.

49. Schautteet K, Vanrompay D. 2011. *Chlamydiaceae* infections in pig. Vet Res 42:29.

50. De Puysseleyr K, De Puysseleyr L, Dhondt H, Geens T, Braeckman L, Morre SA, Cox E, Vanrompay D. 2014. Evaluation of the presence and zoonotic transmission of *Chlamydia suis* in a pig slaughterhouse. BMC Infect Dis 14:560.

51. Dean D, Rothschild J, Ruettger A, Kandel RP, Sachse K. 2013. Zoonotic *Chlamydiaceae* species associated with trachoma, Nepal. Emerg Infect Dis 19:1948-1955.

52. Joseph SJ, Marti H, Didelot X, Read TD, Dean D. 2016. Tetracycline Selective Pressure and Homologous Recombination Shape the Evolution of *Chlamydia suis*: A Recently Identified Zoonotic Pathogen. Genome Biol Evol 8:2613-2623.

53. Suchland RJ, Sandoz KM, Jeffrey BM, Stamm WE, Rockey DD. 2009. Horizontal transfer of tetracycline resistance among *Chlamydia spp*. in vitro. Antimicrob Agents Chemother 53:4604-4611.

54. Vanrompay D, Ducatelle R, Haesebrouck F. 1995. *Chlamydia psittaci* infections: a review with emphasis on avian chlamydiosis. Vet Microbiol 45:93-119.

55. Read TD, Joseph SJ, Didelot X, Liang B, Patel L, Dean D. 2013. Comparative analysis of *Chlamydia psittaci* genomes reveals the recent emergence of a pathogenic lineage with a broad host range. MBio 4.

56. Thierry S, Vorimore F, Rossignol C, Scharf S, Sachse K, Berthon P, Durand B, Virlogeux-Payant I, Borel N, Laroucau K. 2016. Oral Uptake of *Chlamydia psittaci* by Ducklings Results in Systemic Dissemination. Plos One 11.

57. Kauffold J, Wehrend A. 2014. [Reproductive disorders in the female pig: causes, manifestation, diagnostics and approach in herd health care]. Tierarztl Prax Ausg G Grosstiere Nutztiere 42:179-186; quiz 187.

58. Van Loo H, Pardon B, De Schutter P, De Bleecker K, Vanrompay D, Deprez P, Maris J. 2014. Detection of *Chlamydia psittaci* in Belgian cattle with signs of respiratory disease and milk drop syndrome. Vet Rec 175:562.

59. Burrell MH, Chalmers WS, Kewley DR. 1986. Isolation of *Chlamydia psittaci* from the respiratory tract and conjunctivae of thoroughbred horses. Vet Rec 119:302-303.

60. Sachse K, Laroucau K, Riege K, Wehner S, Dilcher M, Creasy HH, Weidmann M, Myers G, Vorimore F, Vicari N, Magnino S, Liebler-Tenorio E, Ruettger A, Bavoil PM, Hufert FT, Rossello-Mora R, Marz M. 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 37:79-88.

61. Zocevic A, Vorimore F, Marhold C, Horvatek D, Wang D, Slavec B, Prentza Z, Stavianis G, Prukner-Radovcic E, Dovc A, Siarkou VI, Laroucau K. 2012. Molecular characterization of atypical *Chlamydia* and evidence of their dissemination in different European and Asian chicken flocks by specific real-time PCR. Environ Microbiol 14:2212-2222.

62. Guo W, Li J, Kaltenboeck B, Gong J, Fan W, Wang C. 2016. *Chlamydia gallinacea*, not *C. psittaci*, is the endemic chlamydial species in chicken (*Gallus gallus*). Sci Rep 6:19638.

63. Soldati G, Lu ZH, Vaughan L, Polkinghorne A, Zimmermann DR, Huder JB, Pospischil A. 2004. Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. Vet Pathol 41:388-397.

64. Hotzel H, Blahak S, Diller R, Sachse K. 2005. Evidence of infection in tortoises by Chlamydia-like organisms that are genetically distinct from known *Chlamydiaceae* species. Vet Res Commun 29 Suppl 1:71-80.

65. Homer BL, Jacobson ER, Schumacher J, Scherba G. 1994. Chlamydiosis in mariculture-reared green sea turtles (*Chelonia mydas*). Vet Pathol 31:1-7.

66. Blumer C, Zimmermann DR, Weilenmann R, Vaughan L, Pospischil A. 2007. *Chlamydiae* in free-ranging and captive frogs in Switzerland. Vet Pathol 44:144-150.

67. Bodetti TJ, Jacobson E, Wan C, Hafner L, Pospischil A, Rose K, Timms P. 2002. Molecular evidence to support the expansion of the hostrange of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. Syst Appl Microbiol 25:146-152.

68. Huchzermeyer FW, Langelet E, Putterill JF. 2008. An outbreak of chlamydiosis in farmed Indopacific crocodiles (*Crocodylus porosus*). J S Afr Vet Assoc 79:99-100.

69. Howerth EW. 1984. Pathology of naturally occurring chlamydiosis in African clawed frogs (*Xenopus laevis*). Vet Pathol 21:28-32.

70. Berger L, Volp K, Mathews S, Speare R, Timms P. 1999. *Chlamydia pneumoniae* in a free-ranging giant barred frog (*Mixophyes iteratus*) from Australia. J Clin Microbiol 37:2378-2380.

71. Jacobson E, Origgi F, Heard D, Detrisac C. 2002. Immunohistochemical staining of chlamydial antigen in emerald tree boas (*Corallus caninus*). J Vet Diagn Invest 14:487-494.

72. Jacobson ER, Heard D, Andersen A. 2004. Identification of *Chlamydophila pneumoniae* in an emerald tree boa, *Corallus caninus*. J Vet Diagn Invest 16:153-154.

73. Cope I, Wheelhouse N, Pocknell A, Dagleish M, Summers B. 2014. An unusual presentation of *Chlamydia pneumoniae* infection in a Royal Python (*Python regius*). Veterinary Record Case Reports 2.

74. Robertson T, Bibby S, O'Rourke D, Belfiore T, Agnew-Crumpton R, Noormohammadi AH. 2010. Identification of Chlamydial species in crocodiles and chickens by PCR-HRM curve analysis. Vet Microbiol 145:373-379.

75. Polkinghorne A, Hanger J, Timms P. 2013. Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. Vet Microbiol 165:214-223.

76. Wan C, Loader J, Hanger J, Beagley K, Timms P, Polkinghorne A. 2011. Using quantitative polymerase chain reaction to correlate *Chlamydia pecorum* infectious load with ocular, urinary and reproductive tract disease in the koala (*Phascolarctos cinereus*). Aust Vet J 89:409-412.

77. Fukushi H, Hirai K. 1992. Proposal of *Chlamydia pecorum* sp. nov. for *Chlamydia* strains derived from ruminants. Int J Syst Bacteriol 42:306-308.

78. Kaltenboeck B, Heinen E, Schneider R, Wittenbrink MM, Schmeer N. 2009. *OmpA* and antigenic diversity of bovine *Chlamydophila pecorum* strains. Vet Microbiol 135:175-180.

79. Reinhold P, Sachse K, Kaltenboeck B. 2011. *Chlamydiaceae* in cattle: commensals, trigger organisms, or pathogens? Vet J 189:257-267.

80. Walker E, Lee EJ, Timms P, Polkinghorne A. 2015. *Chlamydia pecorum* infections in sheep and cattle: A common and under-recognised infectious disease with significant impact on animal health. Vet J 206:252-260.

81. Giannitti F, Anderson M, Miller M, Rowe J, Sverlow K, Vasquez M, Canton G. 2016. *Chlamydia pecorum*: fetal and placental lesions in sporadic caprine abortion. J Vet Diagn Invest 28:184-189.

82. Greco G, Corrente M, Buonavoglia D, Campanile G, Di Palo R, Martella V, Bellacicco AL, D'Abramo M, Buonavoglia C. 2008. Epizootic abortion related to infections by *Chlamydophila abortus* and *Chlamydophila pecorum* in water buffalo (*Bubalus bubalis*). Theriogenology 69:1061-1069.

83. Chahota R, Gupta S, Bhardwaj B, Malik P, Verma S, Sharma AM. 2015. Seroprevalence studies on animal chlamydiosis amongst ruminants in five states of India. Vet World 8:72-75.

84. Di Francesco A, Baldelli R, Donati M, Cotti C, Bassi P, Delogu M. 2013. Evidence for *Chlamydiaceae* and *Parachlamydiaceae* in a wild boar (*Sus scrofa*) population in Italy. Vet Ital 49:119-122.

85. Frutos MC, Venezuela F, Kiguen X, Re V, Cuffini C. 2012. Detection of the *ompA* gene of *Chlamydophila pecorum* in captive birds in Argentina. Rev Argent Microbiol 44:65-68.

86. Tanaka C, Miyazawa T, Watarai M, Ishiguro N. 2005. Bacteriological survey of feces from feral pigeons in Japan. J Vet Med Sci 67:951-953.

87. Di Francesco A, Donati M, Mattioli L, Naldi M, Salvatore D, Poglayen G, Cevenini R, Baldelli R. 2006. *Chlamydophila pneumoniae* in horses: a seroepidemiological survey in Italy. New Microbiol 29:303-305.

88. Theegarten D, Sachse K, Mentrup B, Fey K, Hotzel H, Anhenn O. 2008. Chlamydophila spp. infection in horses with recurrent airway obstruction: similarities to human chronic obstructive disease. Respir Res 9:14.

89. Szeredi L, Hotzel H, Sachse K. 2005. High prevalence of chlamydial (*Chlamydophila psittaci*) infection in fetal membranes of aborted equine fetuses. Vet Res Commun 29 Suppl 1:37-49.

90. Jelocnik M, Branley J, Heller J, Raidal S, Alderson S, Galea F, Gabor M, Polkinghome A. 2017. Multilocus sequence typing identifies an avian-like *Chlamydia psittaci* strain involved in equine placentitis and associated with subsequent human psittacosis. Emerging Microbes & Infections 6.

91. Saikku P, Laitinen K, Leinonen M. 1998. Animal models for *Chlamydia pneumoniae* infection. Atherosclerosis 140 Suppl 1:S17-19.

92. Cantas L, Suer K. 2014. Review: the important bacterial zoonoses in "one health" concept. Front Public Health 2:144.

93. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. Nature 451:990-993.

94. Aaziz R, Gourlay P, Vorimore F, Sachse K, Siarkou VI, Laroucau K. 2015. *Chlamydiaceae* in North Atlantic Seabirds Admitted to a Wildlife Rescue Center in Western France. Appl Environ Microbiol 81:4581-4590.

95. Isaksson J, Christerson L, Blomqvist M, Wille M, Alladio LA, Sachse K, Olsen B, González-Acuña D, Herrmann B. 2015. *Chlamydiaceae*-like bacterium, but no *Chlamydia psittaci*, in sea birds from Antarctica. Polar Biology 38:1931-1936.

96. Blomqvist M, Christerson L, Waldenstrom J, Herrmann B, Olsen B. 2012. *Chlamydia psittaci* in Swedish wetland birds: a risk to zoonotic infection? Avian Dis 56:737-740.

97. Christerson L, Blomqvist M, Grannas K, Thollesson M, Laroucau K, Waldenstrom J, Eliasson I, Olsen B, Herrmann B. 2010. A novel *Chlamydiaceae*-like bacterium found in faecal specimens from sea birds from the Bering Sea. Environ Microbiol Rep 2:605-610.

98. Sariya L, Kladmanee K, Bhusri B, Thaijongrak P, Tonchiangsai K, Chaichoun K, Ratanakorn P. 2015. Molecular evidence for genetic distinctions between *Chlamydiaceae* detected in Siamese crocodiles (*Crocodylus siamensis*) and known *Chlamydiaceae* species. Jpn J Vet Res 63:5-14.

99. Shilton CM, Jerrett IV, Davis S, Walsh S, Benedict S, Isberg SR, Webb GJ, Manolis C, Hyndman TH, Phalen D, Brown GP, Melville L. 2016. Diagnostic investigation of new disease syndromes in farmed Australian saltwater crocodiles (*Crocodylus porosus*) reveals associations with herpesviral infection. J Vet Diagn Invest 28:279-290.

100. Aaziz R, Vorimore F, Verheyden H, Picot D, Bertin C, Ruettger A, Sachse K, Laroucau K. 2015. Detection of atypical *Chlamydiaceae* in roe deer (*Capreolus capreolus*). Vet Microbiol 181:318-322.

101. Veikkolainen V, Vesterinen EJ, Lilley TM, Pulliainen AT. 2014. Bats as reservoir hosts of human bacterial pathogen, *Bartonella mayotimonensis*. Emerg Infect Dis 20:960-967.

102. Hokynar K, Vesterinen EJ, Lilley TM, Pulliainen AT, Korhonen SJ, Paavonen J, Puolakkainen M. 2017. Molecular Evidence of *Chlamydia*-Like Organisms in the Feces of *Myotis daubentonii* Bats. Appl Environ Microbiol 83.

103. 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 69:5530-5535.

104. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, Brandt P, Nyakatura GJ, Droege M, Frishman D, Rattei T, Mewes HW, Wagner M. 2004. Illuminating the evolutionary history of chlamydiae. Science 304:728-730.

105. Schuster FL, Visvesvara GS. 2004. Free-living amoebae as opportunistic and non-opportunistic pathogens of humans and animals. Int J Parasitol 34:1001-1027.

106. Greub G, Raoult D. 2004. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17:413-433.

107. Horn M, Wagner M. 2004. Bacterial endosymbionts of free-living amoebae. J Eukaryot Microbiol 51:509-514.

108. Kahane S, Gonen R, Sayada C, Elion J, Friedman MG. 1993. Description and partial characterization of a new *Chlamydia*-like microorganism. FEMS Microbiol Lett 109:329-333.

109. Greub G, Raoult D. 2002. Crescent bodies of *Parachlamydia acanthamoeba* and its life cycle within Acanthamoeba polyphaga: an electron micrograph study. Appl Environ Microbiol 68:3076-3084.

110. Horn M, Wagner M, Muller KD, Schmid EN, Fritsche TR, Schleifer KH, Michel R. 2000. *Neochlamydia hartmannellae* gen. nov., sp. nov. (*Parachlamydiaceae*), an endoparasite of the amoeba *Hartmannella vermiformis*. Microbiology 146 (Pt 5):1231-1239.

111. Collingro A, Poppert S, Heinz E, Schmitz-Esser S, Essig A, Schweikert M, Wagner M, Horn M. 2005. Recovery of an environmental *Chlamydia* strain from activated sludge by co-cultivation with *Acanthamoeba sp.* Microbiology 151:301-309.

112. Sampo A, Matsuo J, Yamane C, Yagita K, Nakamura S, Shouji N, Hayashi Y, Yamazaki T, Yoshida M, Kobayashi M, Ishida K, Yamaguchi H. 2014. High-temperature adapted primitive *Protochlamydia* found in *Acanthamoeba* isolated from a hot spring can grow in immortalized human epithelial HEp-2 cells. Environ Microbiol 16:486-497.

113. Fritsche TR, Gautom RK, Seyedirashti S, Bergeron DL, Lindquist TD. 1993. Occurrence of bacterial endosymbionts in *Acanthamoeba spp*. isolated from corneal and environmental specimens and contact lenses. J Clin Microbiol 31:1122-1126.

114. Corsaro D, Venditti D, Valassina M. 2002. New parachlamydial 16S rDNA phylotypes detected in human clinical samples. Res Microbiol 153:563-567.

115. Corsaro D, Michel R, Walochnik J, Muller KD, Greub G. 2010. *Saccamoeba lacustris*, sp. nov. (Amoebozoa: Lobosea: Hartmannellidae), a new lobose amoeba, parasitized by the novel chlamydia '*Candidatus* Metachlamydia lacustris' (*Chlamydiae: Parachlamydiaceae*). Eur J Protistol 46:86-95.

116. Corsaro D, Muller KD, Wingender J, Michel R. 2013. "*Candidatus* Mesochlamydia elodeae" (*Chlamydiae: Parachlamydiaceae*), a novel chlamydia parasite of free-living amoebae. Parasitol Res 112:829-838.

117. Corsaro D, Pages GS, Catalan V, Loret JF, Greub G. 2010. Biodiversity of amoebae and amoeba-associated bacteria in water treatment plants. Int J Hyg Environ Health 213:158-166.

118. Lienard J, Croxatto A, Prod'hom G, Greub G. 2011. *Estrella lausannensis*, a new star in the *Chlamydiales* order. Microbes Infect 13:1232-1241.

119. Kebbi-Beghdadi C, Fatton M, Greub G. 2015. Permissivity of insect cells to *Waddlia chondrophila*, *Estrella lausannensis* and *Parachlamydia acanthamoebae*. Microbes Infect 17:749-754.

120. Pagnier I, Valles C, Raoult D, La Scola B. 2015. Isolation of *Vermamoeba vermiformis* and associated bacteria in hospital water. Microb Pathog 80:14-20.

121. Pizzetti I, Schulz F, Tyml T, Fuchs BM, Amann R, Horn M, Fazi S. 2016. Chlamydial seasonal dynamics and isolation of '*Candidatus* Neptunochlamydia vexilliferae' from a Tyrrhenian coastal lake. Environ Microbiol 18:2405-2417.

122. Kahane S, Platzner N, Dvoskin B, Itzhaki A, Friedman MG. 2004. Evidence for the presence of *Simkania negevensis* in drinking water and in reclaimed wastewater in Israel. Appl Environ Microbiol 70:3346-3351.

123. Karbowiak G, Biernat B, Szewczyk T, Sytykiewicz H. 2015. The role of particular tick developmental stages in the circulation of tick-borne pathogens affecting humans in Central Europe. 1. The general pattern. Ann Parasitol 61:221-228.

124. Croxatto A, Rieille N, Kernif T, Bitam I, Aeby S, Peter O, Greub G. 2014. Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks Tick Borne Dis 5:359-365.

125. Pilloux L, Aeby S, Gaumann R, Burri C, Beuret C, Greub G. 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 81:8177-8182.

126. Hokynar K, Sormunen JJ, Vesterinen EJ, Partio EK, Lilley T, Timonen V, Panelius J, Ranki A, Puolakkainen M. 2016. *Chlamydia*-Like Organisms (CLOs) in Finnish Ixodes ricinus Ticks and Human Skin. Microorganisms 4.

127. Burnard D, Weaver H, Gillett A, Loader J, Flanagan C, Polkinghorne A. 2017. Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasit Vectors 10:46.

128. Everett KD, Thao M, Horn M, Dyszynski GE, Baumann P. 2005. Novel chlamydiae in whiteflies and scale insects: endosymbionts 'Candidatus Fritschea bemisiae' strain Falk and '*Candidatus* Fritschea eriococci' strain Elm. Int J Syst Evol Microbiol 55:1581-1587.

129. Thao ML, Baumann L, Hess JM, Falk BW, Ng JC, Gullan PJ, Baumann P. 2003. Phylogenetic evidence for two new insect-associated *Chlamydia* of the family *Simkaniaceae*. Curr Microbiol 47:46-50.

130. Kostanjsek R, Strus J, Drobne D, Avgustin G. 2004. '*Candidatus* Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). Int J Syst Evol Microbiol 54:543-549.

131. Corsaro D, Thomas V, Goy G, Venditti D, Radek R, Greub G. 2007. '*Candidatus* Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach *Blatta orientalis* (Insecta: Blattodea). Syst Appl Microbiol 30:221-228.

132. Hornung S, Thuong BC, Gyger J, Kebbi-Beghdadi C, Vasilevsky S, Greub G, Baud D. 2015. Role of *Chlamydia trachomatis* and emerging *Chlamydia*-related bacteria in ectopic pregnancy in Vietnam. Epidemiol Infect 143:2635-2638.

133. Baud D, Kebbi C, Kulling JP, Greub G. 2009. Seroprevalence of different *Chlamydia*-like organisms in an asymptomatic population. Clin Microbiol Infect 15 Suppl 2:213-215.

134. Corsaro D, Valassina M, Venditti D. 2003. Increasing diversity within *Chlamydiae*. Crit Rev Microbiol 29:37-78.

135. Stride MC, Polkinghome A, Nowak BF. 2014. Chlamydial infections of fish: diverse pathogens and emerging causes of disease in aquaculture species. Vet Microbiol 171:258-266.

136. Fehr A, Walther E, Schmidt-Posthaus H, Nufer L, Wilson A, Svercel M, Richter D, Segner H, Pospischil A, Vaughan L. 2013. *Candidatus* Syngnamydia venezia, a novel member of the phylum *Chlamydiae* from the broad nosed pipefish, *Syngnathus typhle*. PLoS One 8:e70853.

137. Karlsen M, Nylund A, Watanabe K, Helvik JV, Nylund S, Plarre H. 2008. Characterization of *'Candidatus* Clavochlamydia salmonicola': an intracellular bacterium infecting salmonid fish. Environ Microbiol 10:208-218.

138. Draghi A, 2nd, Popov VL, Kahl MM, Stanton JB, Brown CC, Tsongalis GJ, West AB, Frasca S, Jr. 2004. Characterization of "*Candidatus* piscichlamydia salmonis" (order *Chlamydiales*), a chlamydia-like bacterium associated with epitheliocystis in farmed Atlantic salmon (*Salmo salar*). J Clin Microbiol 42:5286-5297.

140. Stride MC, Polkinghorne A, Miller TL, Groff JM, Lapatra SE, Nowak BF. 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 79:1590-1597.

141. Stride MC, Polkinghorne A, Powell MD, Nowak BF. 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:e82889.

142. Mitchell SO, Steinum T, Rodger H, Holland C, Falk K, Colquhoun DJ. 2010. Epitheliocystis in Atlantic salmon, *Salmo salar* L., farmed in fresh water in Ireland is associated with *'Candidatus* Clavochlamydia salmonicola' infection. J Fish Dis 33:665-673.

143. Qiao C, Li J, Wei XH, Wang JL, Wang YF, Liu XS. 2014. SRP gene is required for Helicoverpa armigera prophenoloxidase activation and nodulation response. Dev Comp Immunol 44:94-99.

144. Schmidt-Posthaus H, Polkinghorne A, Nufer L, Schifferli A, Zimmermann DR, Segner H, Steiner P, Vaughan L. 2012. A natural freshwater origin for two chlamydial species, Candidatus Piscichlamydia salmonis and Candidatus Clavochlamydia salmonicola, causing mixed infections in wild brown trout (*Salmo trutta*). Environ Microbiol 14:2048-2057.

145. Draghi A, 2nd, Bebak J, Daniels S, Tulman ER, Geary SJ, West AB, Popov VL, Frasca S, Jr. 2010. Identification of *'Candidatus* Piscichlamydia salmonis' in Arctic charr *Salvelinus alpinus* during a survey of charr production facilities in North America. Dis Aquat Organ 89:39-49.

146. Kim DJ, Park JH, Seok SH, Cho SA, Baek MW, Lee HY, Park JH. 2005. Epitheliocystis in carp (*Cyprinus carpio*) in South Korea. J Vet Med Sci 67:119-120.

147. Guevara Soto M, Vidondo B, Vaughan L, Seth-Smith HM, Nufer L, Segner H, Rubin JF, Schmidt-Posthaus H. 2016. The emergence of epitheliocystis in the upper Rhone region: evidence for *Chlamydiae* in wild and farmed salmonid populations. Arch Microbiol 198:315-324.

148. 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 8:1672-1689.

149. Steigen A, Nylund A, Karlsbakk E, Akoll P, Fiksdal IU, Nylund S, Odong R, Plarre H, Semyalo R, Skar C, Watanabe K. 2013. *'Cand.* Actinochlamydia clariae' gen. nov., sp. nov., a unique intracellular bacterium causing epitheliocystis in catfish (*Clarias gariepinus*) in Uganda. PLoS One 8:e66840.

150. 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 79:4914-4920.

151. Steigen A, Karlsbakk E, Plarre H, Watanabe K, Overgard AC, Brevik O, Nylund A. 2015. A new intracellular bacterium, *Candidatus* Similichlamydia labri sp. nov. (*Chlamydiaceae*) producing epitheliocysts in ballan wrasse, *Labrus bergylta* (Pisces, Labridae). Arch Microbiol 197:311-318.

152. Seth-Smith HMB, Katharios P, Dourala N, Mateos JM, Fehr AGJ, Nufer L, Ruetten M, Guevara Soto M, Vaughan L. 2017. *Ca.* Similichlamydia in Epitheliocystis Co-infection of Gilthead Seabream Gills: Unique Morphological Features of a Deep Branching Chlamydial Family. Front Microbiol 8:508.

153. Nylund S, Steigen A, Karlsbakk E, Plarre H, Andersen L, Karlsen M, Watanabe K, Nylund A. 2015. Characterization of *'Candidatus* Syngnamydia salmonis' (*Chlamydiales, Simkaniaceae*), a bacterium associated with epitheliocystis in Atlantic salmon (*Salmo salar* L.). Arch Microbiol 197:17-25.

154. Corsaro D, Work TM. 2012. Candidatus Renichlamydia lutjani, a Gram-negative bacterium in internal organs of blue-striped snapper *Lutjanus kasmira* from Hawaii. Dis Aquat Organ 98:249-254.

155. Polkinghorne A, Schmidt-Posthaus H, Meijer A, Lehner A, Vaughan L. 2010. Novel *Chlamydiales* associated with epitheliocystis in a leopard shark *Triakis semifasciata*. Dis Aquat Organ 91:75-81.

156. Camus A, Soto E, Berliner A, Clauss T, Sanchez S. 2013. Epitheliocystis hyperinfection in captive spotted eagle rays Aetobatus narinari associated with a novel Chlamydiales 16S rDNA signature sequence. Dis Aquat Organ 104:13-21.

157. Meijer A, Roholl PJ, Ossewaarde JM, Jones B, Nowak BF. 2006. Molecular evidence for association of chlamydiales bacteria with epitheliocystis in leafy seadragon (*Phycodurus eques*), silver perch (*Bidyanus bidyanus*), and barramundi (*Lates calcarifer*). Appl Environ Microbiol 72:284-290.

158. Lieberman D, Dvoskin B, Lieberman DV, Kahane S, Friedman MG. 2002. Serological evidence of acute infection with the *Chlamydia*-like microorganism *Simkania negevensis* (Z) in acute exacerbation of chronic obstructive pulmonary disease. Eur J Clin Microbiol Infect Dis 21:307-309.

159. Lieberman D, Kahane S, Lieberman D, Friedman MG. 1997. Pneumonia with serological evidence of acute infection with the *Chlamydia*-like microorganism "Z". Am J Respir Crit Care Med 156:578-582.

160. Friedman MG, Kahane S, Dvoskin B, Hartley JW. 2006. Detection of *Simkania negevensis* by culture, PCR, and serology in respiratory tract infection in Cornwall, UK. J Clin Pathol 59:331-333.

161. Kahane S, Greenberg D, Friedman MG, Haikin H, Dagan R. 1998. High prevalence of "Simkania Z," a novel *Chlamydia*-like bacterium, in infants with acute bronchiolitis. J Infect Dis 177:1425-1429.

162. Kahane S, Greenberg D, Newman N, Dvoskin B, Friedman MG. 2007. Domestic water supplies as a possible source of infection with *Simkania*. J Infect 54:75-81.

163. Friedman MG, Galil A, Greenberg S, Kahane S. 1999. Seroprevalence of IgG antibodies to the *Chlamydia*-like microorganism *'Simkania* Z' by ELISA. Epidemiol Infect 122:117-123.

164. Husain S, Kahane S, Friedman MG, Paterson DL, Studer S, McCurry KR, Wolf DG, Zeevi A, Pilewski J, Greenberg D. 2007. *Simkania negevensis* in bronchoalveolar lavage of lung transplant recipients: a possible association with acute rejection. Transplantation 83:138-143.

165. Kahane S, Fruchter D, Dvoskin B, Friedman MG. 2007. Versatility of *Simkania negevensis* infection *in vitro* and induction of host cell inflammatory cytokine response. J Infect 55:e13-21.

166. Donati M, Fiani N, Di Francesco A, Di Paolo M, Vici M, Cevenini R. 2013. IgG and IgA response to *Simkania negevensis* in sera of patients with respiratory and gastrointestinal symptoms. New Microbiol 36:303-306.

167. Birtles RJ, Rowbotham TJ, Storey C, Marrie TJ, Raoult D. 1997. *Chlamydia*-like obligate parasite of free-living amoebae. Lancet 349:925-926.

168. Lamoth F, Jaton K, Vaudaux B, Greub G. 2011. *Parachlamydia* and *Ca.* Rhabdochlamydia: emerging agents of community-acquired respiratory infections in children. Clin Infect Dis 53:500-501.

169. Greub G. 2009. *Parachlamydia acanthamoebae*, an emerging agent of pneumonia. Clin Microbiol Infect 15:18-28.

170. Greub G, Boyadjiev I, La Scola B, Raoult D, Martin C. 2003. Serological hint suggesting that *Parachlamydiaceae* are agents of pneumonia in polytraumatized intensive care patients. Ann N Y Acad Sci 990:311-319.

171. Greub G, Berger P, Papazian L, Raoult D. 2003. Parachlamydiaceae as rare agents of pneumonia. Emerg Infect Dis 9:755-756.

172. Casson N, Medico N, Bille J, Greub G. 2006. *Parachlamydia acanthamoebae* enters and multiplies within pneumocytes and lung fibroblasts. Microbes Infect 8:1294-1300.

173. Greub G, Mege JL, Raoult D. 2003. *Parachlamydia acanthamoebae* enters and multiplies within human macrophages and induces their apoptosis [corrected]. Infect Immun 71:5979-5985.

174. Lohr M, Prohl A, Ostermann C, Liebler-Tenorio E, Schroedl W, Aeby S, Greub G, Reinhold P. 2014. A bovine model of a respiratory *Parachlamydia acanthamoebae* infection. Pathog Dis doi:10.1111/2049-632X.12201.

175. Casson N, Entenza JM, Borel N, Pospischil A, Greub G. 2008. Murine model of pneumonia caused by *Parachlamydia acanthamoebae*. Microb Pathog 45:92-97.

176. Borel N, Ruhl S, Casson N, Kaiser C, Pospischil A, Greub G. 2007. *Parachlamydia spp*. and related *Chlamydia*-like organisms and bovine abortion. Emerg Infect Dis 13:1904-1907.

177. Barkallah M, Gharbi Y, Hassena AB, Slima AB, Mallek Z, Gautier M, Greub G, Gdoura R, Fendri I. 2014. Survey of infectious etiologies of bovine abortion during mid- to late gestation in dairy herds. PLoS One 9:e91549.

178. Baud D, Goy G, Gerber S, Vial Y, Hohlfeld P, Greub G. 2009. Evidence of maternal-fetal transmission of *Parachlamydia acanthamoebae*. Emerg Infect Dis 15:120-121.

179. Ruhl S, Casson N, Kaiser C, Thoma R, Pospischil A, Greub G, Borel N. 2009. Evidence for *Parachlamydia* in bovine abortion. Vet Microbiol 135:169-174.

180. Ruhl S, Goy G, Casson N, Thoma R, Pospischil A, Greub G, Borel N. 2008. *Parachlamydia acanthamoebae* infection and abortion in small ruminants. Emerg Infect Dis 14:1966-1968.

181. Wheelhouse N, Howie F, Gidlow J, Greub G, Dagleish M, Longbottom D. 2012. Involvement of *Parachlamydia* in bovine abortions in Scotland. Vet J 193:586-588.

182. Blumer S, Greub G, Waldvogel A, Hassig M, Thoma R, Tschuor A, Pospischil A, Borel N. 2011. *Waddlia, Parachlamydia* and *Chlamydiaceae* in bovine abortion. Vet Microbiol 152:385-393.

183. Wheelhouse N, Sait M, Gidlow J, Deuchande R, Borel N, Baily J, Caldow G, Longbottom D. 2011. Molecular detection of *Chlamydia*-like organisms in cattle drinking water. Vet Microbiol 152:196-199.

184. Dilbeck-Robertson P, McAllister MM, Bradway D, Evermann JF. 2003. Results of a new serologic test suggest an association of *Waddlia chondrophila* with bovine abortion. J Vet Diagn Invest 15:568-569.

185. Baud D, Goy G, Osterheld MC, Croxatto A, Borel N, Vial Y, Pospischil A, Greub G. 2014. Role of *Waddlia chondrophila* placental infection in miscarriage. Emerg Infect Dis 20:460-464.

186. Baud D, Thomas V, Arafa A, Regan L, Greub G. 2007. *Waddlia chondrophila*, a potential agent of human fetal death. Emerg Infect Dis 13:1239-1243.

187. Kebbi-Beghdadi C, Cisse O, Greub G. 2011. Permissivity of Vero cells, human pneumocytes and human endometrial cells to *Waddlia chondrophila*. Microbes Infect 13:566-574.

188. Baud D, Goy G, Osterheld MC, Borel N, Vial Y, Pospischil A, Greub G. 2011. *Waddlia chondrophila*: from bovine abortion to human miscarriage. Clin Infect Dis 52:1469-1471.

189. Codony F, Fittipaldi M, Lopez E, Morato J, Agusti G. 2012. Well water as a possible source of *Waddlia chondrophila* infections. Microbes Environ 27:529-532.

190. Wheelhouse N, Flockhart A, Aitchison K, Livingstone M, Finlayson J, Flachon V, Sellal E, Dagleish MP, Longbottom D. 2016. Experimental challenge of pregnant cattle with the putative abortifacient *Waddlia chondrophila*. Sci Rep 6:37150.

191. Chua PK, Corkill JE, Hooi PS, Cheng SC, Winstanley C, Hart CA. 2005. Isolation of *Waddlia malaysiensis*, a novel intracellular bacterium, from fruit bat (*Eonycteris spelaea*). Emerg Infect Dis 11:271-277.

192. Pierle SA, Morales CO, Martinez LP, Ceballos NA, Rivero JJ, Diaz OL, Brayton KA, Setien AA. 2015. Novel *Waddlia* Intracellular Bacterium in *Artibeus intermedius* Fruit Bats, Mexico. Emerg Infect Dis 21:2161-2163.

193. Martel A, Adriaensen C, Bogaerts S, Ducatelle R, Favoreel H, Crameri S, Hyatt AD, Haesebrouck F, Pasmans F. 2012. Novel *Chlamydiaceae* disease in captive salamanders. Emerg Infect Dis 18:1020-1022.

194. Martel A, Adriaensen C, Sharifian-Fard M, Vandewoestyne M, Deforce D, Favoreel H, Bergen K, Spitzen-van der Sluijs A, Devisscher S, Adriaens T, Louette G, Baert K, Hyatt A, Crameri S, Haesebrouck F, Pasmans F. 2013. The novel *'Candidatus* Amphibiichlamydia ranarum' is highly prevalent in invasive exotic bullfrogs (*Lithobates catesbeianus*). Environ Microbiol Rep 5:105-108.

195. Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M. 2014. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. ISME J 8:115-125.

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

197. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754-759.

198. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW, Stephens RS. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat Genet 21:385-389.

199. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28:1397-1406.

200. Voigt A, Schofl G, Saluz HP. 2012. The *Chlamydia psittaci* genome: a comparative analysis of intracellular pathogens. PLoS One 7:e35097.

201. Read TD, Myers GS, Brunham RC, Nelson WC, Paulsen IT, Heidelberg J, Holtzapple E, Khouri H, Federova NB, Carty HA, Umayam LA, Haft DH, Peterson J, Beanan MJ, White O, Salzberg SL, Hsia RC, McClarty G, Rank RG, Bavoil PM, Fraser CM. 2003. Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the *Chlamydiaceae*. Nucleic Acids Res 31:2134-2147.

202. Roulis E, Bachmann N, Humphrys M, Myers G, Huston W, Polkinghorne A, Timms P. 2015. Phylogenetic analysis of human *Chlamydia pneumoniae* strains reveals a distinct Australian indigenous clade that predates European exploration of the continent. BMC Genomics 16:1094.

203. Roulis E, Bachmann NL, Myers GS, Huston W, Summersgill J, Hudson A, Dreses-Werringloer U, Polkinghorne A, Timms P. 2015. Comparative genomic analysis of human *Chlamydia pneumoniae* isolates from respiratory, brain and cardiac tissues. Genomics 106:373-383.

204. Weinmaier T, Hoser J, Eck S, Kaufhold I, Shima K, Strom TM, Rattei T, Rupp J. 2015. Genomic factors related to tissue tropism in *Chlamydia pneumoniae* infection. BMC Genomics 16:268.

205. Borges V, Pinheiro M, Vieira L, Sampaio DA, Nunes A, Borrego MJ, Gomes JP. 2014. Complete Genome Sequence of *Chlamydia trachomatis* Ocular Serovar C Strain TW-3. Genome Announc 2.

206. Unemo M, Seth-Smith HM, Cutcliffe LT, Skilton RJ, Barlow D, Goulding D, Persson K, Harris SR, Kelly A, Bjartling C, Fredlund H, Olcen P, Thomson NR, Clarke IN. 2010. The Swedish new variant of *Chlamydia trachomatis*: genome sequence, morphology, cell tropism and phenotypic characterization. Microbiology 156:1394-1404.

207. Thomson NR, Holden MT, Carder C, Lennard N, Lockey SJ, Marsh P, Skipp P, O'Connor CD, Goodhead I, Norbertzcak H, Harris B, Ormond D, Rance R, Quail MA, Parkhill J, Stephens RS, Clarke IN. 2008. *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. Genome Res 18:161-171.

208. Knittler MR, Sachse K. 2015. *Chlamydia psittaci*: update on an underestimated zoonotic agent. Pathog Dis 73:1-15.

209. Gil R, Silva FJ, Pereto J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68:518-537.

210. Sakharkar KR, Dhar PK, Chow VT. 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. Int J Syst Evol Microbiol 54:1937-1941.

211. Zomorodipour A, Andersson SG. 1999. Obligate intracellular parasites: Rickettsia prowazekii and Chlamydia trachomatis. FEBS Lett 452:11-15.

212. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4:13.

213. Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.

214. Borges V, Nunes A, Ferreira R, Borrego MJ, Gomes JP. 2012. Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation. J Bacteriol 194:6143-6153.

215. McClarty G. 1994. *Chlamydiae* and the biochemistry of intracellular parasitism. Trends Microbiol 2:157-164.

216. Hatch TP, Al-Hossainy E, Silverman JA. 1982. Adenine nucleotide and lysine transport in *Chlamydia psittaci*. J Bacteriol 150:662-670.

217. Iliffe-Lee ER, McClarty G. 1999. Glucose metabolism in *Chlamydia trachomatis*: the 'energy parasite' hypothesis revisited. Mol Microbiol 33:177-187.

218. Vandahl BB, Birkelund S, Christiansen G. 2004. Genome and proteome analysis of *Chlamydia*. Proteomics 4:2831-2842.

219. Nunes A, Gomes JP. 2014. Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. Infect Genet Evol 23:49-64.

220. Weiss E. 1967. Transaminase activity and other enzymatic reactions involving pyruvate and glutamate in *Chlamydia* (psittacosis-trachoma group). J Bacteriol 93:177-184.

221. Jelocnik M, Bachmann NL, Kaltenboeck B, Waugh C, Woolford L, Speight KN, Gillett A, Higgins DP, Flanagan C, Myers GS, Timms P, Polkinghorne A. 2015. Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas. BMC Genomics 16:893.

222. Belland RJ, Scidmore MA, Crane DD, Hogan DM, Whitmire W, McClarty G, Caldwell HD. 2001. *Chlamydia trachomatis* cytotoxicity associated with complete and partial cytotoxin genes. Proc Natl Acad Sci U S A 98:13984-13989.

223. Ferreira R, Antelo M, Nunes A, Borges V, Damiao V, Borrego MJ, Gomes JP. 2014. In silico scrutiny of genes revealing phylogenetic congruence with clinical prevalence or tropism properties of *Chlamydia trachomatis* strains. G3 (Bethesda) 5:9-19.

224. Taylor LD, Nelson DE, Dorward DW, Whitmire WM, Caldwell HD. 2010. Biological characterization of *Chlamydia trachomatis* plasticity zone MACPF domain family protein CT153. Infect Immun 78:2691-2699.

225. Egan AJ, Vollmer W. 2013. The physiology of bacterial cell division. Ann N Y Acad Sci 1277:8-28.

226. Mattei PJ, Neves D, Dessen A. 2010. Bridging cell wall biosynthesis and bacterial morphogenesis. Curr Opin Struct Biol 20:749-755.

227. Hatch TP. 1996. Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae? J Bacteriol 178:1-5.

228. Albrecht M, Sharma CM, Reinhardt R, Vogel J, Rudel T. 2010. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res 38:868-877.

229. Liechti GW, Kuru E, Hall E, Kalinda A, Brun YV, VanNieuwenhze M, Maurelli AT. 2014. A new metabolic cell-wall labelling method reveals peptidoglycan in *Chlamydia trachomatis*. Nature 506:507-510.

230. Vasilevsky S, Stojanov M, Greub G, Baud D. 2016. Chlamydial polymorphic membrane proteins: regulation, function and potential vaccine candidates. Virulence 7:11-22.

231. Henderson IR, Lam AC. 2001. Polymorphic proteins of *Chlamydia spp*.--autotransporters beyond the Proteobacteria. Trends Microbiol 9:573-578.

232. Longbottom D, Russell M, Jones GE, Lainson FA, Herring AJ. 1996. Identification of a multigene family coding for the 90 kDa proteins of the ovine abortion subtype of *Chlamydia psittaci*. FEMS Microbiol Lett 142:277-281.

233. Seth-Smith HMB, Buso LS, Livingstone M, Sait M, Harris SR, Aitchison KD, Vretou E, Siarkou VI, Laroucau K, Sachse K, Longbottom D, Thomson NR. 2017. European *Chlamydia abortus* livestock isolate genomes reveal unusual stability and limited diversity, reflected in geographical signatures. BMC Genomics 18:344.

234. Seth-Smith HM, Wanninger S, Bachmann N, Marti H, Qi W, Donati M, di Francesco A, Polkinghorne A, Borel N. 2017. The *Chlamydia suis* Genome Exhibits High Levels of Diversity, Plasticity, and Mobile Antibiotic Resistance: Comparative Genomics of a Recent Livestock Cohort Shows Influence of Treatment Regimes. Genome Biol Evol 9:750-760.

235. Hadfield J, Harris SR, Seth-Smith HMB, Parmar S, Andersson P, Giffard PM, Schachter J, Moncada J, Ellison L, Vaulet MLG, Fermepin MR, Radebe F, Mendoza S, Ouburg S, Morre SA, Sachse K, Puolakkainen M, Korhonen SJ, Sonnex C, Wiggins R, Jalal H, Brunelli T, Casprini P, Pitt R, Ison C, Savicheva A, Shipitsyna E, Hadad R, Kari L, Burton MJ, Mabey D, Solomon AW, Lewis D, Marsh P, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2017. Comprehensive global genome dynamics of *Chlamydia trachomatis* show ancient diversification followed by contemporary mixing and recent lineage expansion. Genome Res doi:10.1101/gr.212647.116.

236. Hayes LJ, Yearsley P, Treharne JD, Ballard RA, Fehler GH, Ward ME. 1994. Evidence for naturally occurring recombination in the gene encoding the major outer membrane protein of lymphogranuloma venereum isolates of *Chlamydia trachomatis*. Infect Immun 62:5659-5663.

237. Bachmann NL, Fraser TA, Bertelli C, Jelocnik M, Gillett A, Funnell O, Flanagan C, Myers GS, Timms P, Polkinghorne A. 2014. Comparative genomics of koala, cattle and sheep strains of *Chlamydia pecorum*. BMC Genomics 15:667.

238. Bertelli C, Collyn F, Croxatto A, Ruckert C, Polkinghorne A, Kebbi-Beghdadi C, Goesmann A, Vaughan L, Greub G. 2010. The *Waddlia* genome: a window into chlamydial biology. PLoS One 5:e10890.

239. Greub G, Kebbi-Beghdadi C, Bertelli C, Collyn F, Riederer BM, Yersin C, Croxatto A, Raoult D. 2009. High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PLoS One 4:e8423.

240. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, Friedman MG, Rattei T, Myers GS, Horn M. 2011. Unity in variety--the pan-genome of the *Chlamydiae*. Mol Biol Evol 28:3253-3270.

241. Psomopoulos FE, Siarkou VI, Papanikolaou N, Iliopoulos I, Tsaftaris AS, Promponas VJ, Ouzounis CA. 2012. The *Chlamydiales* pangenome revisited: structural stability and functional coherence. Genes (Basel) 3:291-319.

242. Hsia RC, Pannekoek Y, Ingerowski E, Bavoil PM. 1997. Type III secretion genes identify a putative virulence locus of *Chlamydia*. Mol Microbiol 25:351-359.

243. Hefty PS, Stephens RS. 2007. Chlamydial type III secretion system is encoded on ten operons preceded by sigma 70-like promoter elements. J Bacteriol 189:198-206.

244. Bertelli C, Aeby S, Chassot B, Clulow J, Hilfiker O, Rappo S, Ritzmann S, Schumacher P, Terrettaz C, Benaglio P, Falquet L, Farinelli L, Gharib WH, Goesmann A, Harshman K, Linke B, Miyazaki R, Rivolta C, Robinson-Rechavi M, van der Meer JR, Greub G. 2015. Sequencing and characterizing the genome of *Estrella lausannensis* as an undergraduate project: training students and biological insights. Front Microbiol 6:101.

245. Lane BJ, Mutchler C, Al Khodor S, Grieshaber SS, Carabeo RA. 2008. Chlamydial entry involves TARP binding of guanine nucleotide exchange factors. PLoS Pathog 4:e1000014.

246. Bertelli C, Cisse OH, Rusconi B, Kebbi-Beghdadi C, Croxatto A, Goesmann A, Collyn F, Greub G. 2016. CRISPR System Acquisition and Evolution of an Obligate Intracellular *Chlamydia*-Related Bacterium. Genome Biol Evol 8:2376-2386.

247. Heinz E, Rockey DD, Montanaro J, Aistleitner K, Wagner M, Horn M. 2010. Inclusion membrane proteins of *Protochlamydia amoebophila* UWE25 reveal a conserved mechanism for host cell interaction among the *Chlamydiae*. J Bacteriol 192:5093-5102.

248. Ishida K, Sekizuka T, Hayashida K, Matsuo J, Takeuchi F, Kuroda M, Nakamura S, Yamazaki T, Yoshida M, Takahashi K, Nagai H, Sugimoto C, Yamaguchi H. 2014. Amoebal endosymbiont *Neochlamydia* genome sequence illuminates the bacterial role in the defense of the host amoebae against *Legionella pneumophila*. PLoS One 9:e95166.

249. Domman D, Collingro A, Lagkouvardos I, Gehre L, Weinmaier T, Rattei T, Subtil A, Horn M. 2014. Massive expansion of ubiquitination-related gene families within the *Chlamydiae*. Mol Biol Evol doi:10.1093/molbev/msu227.

250. Hintz NJ, Ennis DG, Liu WF, Larsen SH. 1995. The recA gene of *Chlamydia trachomatis*: cloning, sequence, and characterization in *Escherichia coli*. FEMS Microbiol Lett 127:175-180.

251. Millman KL, Tavare S, Dean D. 2001. Recombination in the ompA gene but not the *omcB* gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. J Bacteriol 183:5997-6008.

252. Joseph SJ, Read TD. 2012. Genome-wide recombination in *Chlamydia trachomatis*. Nat Genet 44:364-366.

253. Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol 2:414-424.

254. Che D, Hasan MS, Chen B. 2014. Identifying pathogenicity islands in bacterial pathogenomics using computational approaches. Pathogens 3:36-56.

255. 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 48:3989-3995.

256. Greub G, Collyn F, Guy L, Roten CA. 2004. A genomic island present along the bacterial chromosome of the *Parachlamydiaceae* UWE25, an obligate amoebal endosymbiont, encodes a potentially functional F-like conjugative DNA transfer system. BMC Microbiol 4:48.

257. Sliwa-Dominiak J, Suszynska E, Pawlikowska M, Deptula W. 2013. Chlamydia bacteriophages. Arch Microbiol 195:765-771.

258. Rank RG, Bowlin AK, Cane S, Shou H, Liu Z, Nagarajan UM, Bavoil PM. 2009. Effect of Chlamydiaphage phiCPG1 on the course of conjunctival infection with "*Chlamydia caviae*" in guinea pigs. Infect Immun 77:1216-1221.

259. Jelocnik M, Bachmann NL, Seth-Smith H, Thomson NR, Timms P, Polkinghorne AM. 2016. Molecular characterisation of the *Chlamydia pecorum* plasmid from porcine, ovine, bovine, and koala strains indicates plasmid-strain co-evolution. PeerJ 4:e1661.

260. Song L, Carlson JH, Whitmire WM, Kari L, Virtaneva K, Sturdevant DE, Watkins H, Zhou B, Sturdevant GL, Porcella SF, McClarty G, Caldwell HD. 2013. *Chlamydia trachomatis* Plasmid-Encoded Pgp4 Is a Transcriptional Regulator of Virulence-Associated Genes. Infection and Immunity 81:636-644.

261. Carlson JH, Whitmire WM, Crane DD, Wicke L, Virtaneva K, Sturdevant DE, Kupko JJ, 3rd, Porcella SF, Martinez-Orengo N, Heinzen RA, Kari L, Caldwell HD. 2008. The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor. Infect Immun 76:2273-2283.

262. Matsumoto A, Izutsu H, Miyashita N, Ohuchi M. 1998. Plaque formation by and plaque cloning of *Chlamydia trachomatis* biovar trachoma. J Clin Microbiol 36:3013-3019.

263. Rockey DD. 2011. Unraveling the basic biology and clinical significance of the chlamydial plasmid. J Exp Med 208:2159-2162.

264. Zhong G. 2016. Chlamydial Plasmid-Dependent Pathogenicity. Trends in Microbiology doi:http://dx.doi.org/10.1016/j.tim.2016.09.006.

265. Chen J, Yang Z, Sun X, Tang L, Ding Y, Xue M, Zhou Z, Baseman J, Zhong G. 2015. Intrauterine infection with plasmid-free Chlamydia muridarum reveals a critical role of the plasmid

in chlamydial ascension and establishes a model for evaluating plasmid-independent pathogenicity. Infection and Immunity 83:2583-2592.

266. Lei L, Chen J, Hou S, Ding Y, Yang Z, Zeng H, Baseman J, Zhong G. 2014. Reduced live organism recovery and lack of hydrosalpinx in mice infected with plasmid-free *Chlamydia muridarum*. Infect Immun 82:983-992.

267. Olivares-Zavaleta N, Whitmire W, Gardner D, Caldwell HD. 2010. Immunization with the attenuated plasmidless *Chlamydia trachomatis* L2(25667R) strain provides partial protection in a murine model of female genitourinary tract infection. Vaccine 28:1454-1462.

268. Kari L, Whitmire WM, Olivares-Zavaleta N, Goheen MM, Taylor LD, Carlson JH, Sturdevant GL, Lu C, Bakios LE, Randall LB, Parnell MJ, Zhong G, Caldwell HD. 2011. A live-attenuated chlamydial vaccine protects against trachoma in nonhuman primates. J Exp Med 208:2217-2223.

269. Seth-Smith HM, Harris SR, Persson K, Marsh P, Barron A, Bignell A, Bjartling C, Clark L, Cutcliffe LT, Lambden PR, Lennard N, Lockey SJ, Quail MA, Salim O, Skilton RJ, Wang Y, Holland MJ, Parkhill J, Thomson NR, Clarke IN. 2009. Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. BMC Genomics 10:239.

270. Bou Khalil JY, Benamar S, Baudoin JP, Croce O, Blanc-Tailleur C, Pagnier I, Raoult D, La Scola B. 2016. Developmental Cycle and Genome Analysis of "*Rubidus massiliensis*," a New *Vermamoeba vermiformis* Pathogen. Front Cell Infect Microbiol 6:31.

271. Hugenholtz P. 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol 3:REVIEWS0003.

272. McCurdy K, Joyce J, Hamilton S, Nevins C, Sosna W, Puricelli K, Rayner JO. 2011. Differential accumulation of genetic and phenotypic changes in Venezuelan equine encephalitis virus and Japanese encephalitis virus following passage *in vitro* and *in vivo*. Virology 415:20-29.

273. Seth-Smith HM, Harris SR, Scott P, Parmar S, Marsh P, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2013. Generating whole bacterial genome sequences of low-abundance species from complex samples with IMS-MDA. Nat Protoc 8:2404-2412.

274. 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:E405-408.

275. 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 159:2109-2117.

276. Christiansen MT, Brown AC, Kundu S, Tutill HJ, Williams R, Brown JR, Holdstock J, Holland MJ, Stevenson S, Dave J, Tong CY, Einer-Jensen K, Depledge DP, Breuer J. 2014. Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. BMC Infect Dis 14:591.

277. Bachmann NL, Sullivan MJ, Jelocnik M, Myers GS, Timms P, Polkinghorne A. 2015. Culture-independent genome sequencing of clinical samples reveals an unexpected heterogeneity of infections by *Chlamydia pecorum*. J Clin Microbiol 53:1573-1581.

278. Zheng Z, Deng X, Chen J. 2014. Whole-Genome Sequence of "*Candidatus* Liberibacter asiaticus" from Guangdong, China. Genome Announc 2.

279. Seth-Smith HM, Dourala N, Fehr A, Qi W, Katharios P, Ruetten M, Mateos JM, Nufer L, Weilenmann R, Ziegler U, Thomson NR, Schlapbach R, Vaughan L. 2016. Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular beta-proteobacteria. ISME J 10:1791-1803.

280. Joseph SJ, Li B, Ghonasgi T, Haase CP, Qin ZS, Dean D, Read TD. 2014. Direct amplification, sequencing and profiling of *Chlamydia trachomatis* strains in single and mixed infection clinical samples. PLoS One 9:e99290.

281. Corsaro D, Venditti D. 2004. Emerging chlamydial infections. Crit Rev Microbiol 30:75-106.

282. Huchzermeyer FW, Gerdes GH, Foggin CM, Huchzermeyer KD, Limper LC. 1994. Hepatitis in farmed hatchling Nile crocodiles (*Crocodylus niloticus*) due to chlamydial infection. J S Afr Vet Assoc 65:20-22.

283. Groff JM, LaPatra SE, Munn RJ, Anderson ML, Osburn BI. 1996. Epitheliocystis infection in cultured white sturgeon (*Acipenser transmontanus*): antigenic and ultrastructural similarities of the causative agent to the chlamydiae. J Vet Diagn Invest 8:172-180.

284. Katharios P, Papadaki M, Papandroulakis N, Divanach P. 2008. Severe mortality in mesocosm-reared sharpsnout sea bream *Diplodus puntazzo* larvae due to epitheliocystis infection. Dis Aquat Organ 82:55-60.

285. Szakolczai J, Vetesi F, Pitz SR. 1999. Epitheliocystis disease in cultured pacu (*Piaractus mesopotamicus*) in Brazil. Acta Vet Hung 47:311-318.

286. Hoffman GL, Dunbar CE, Wolf K, Zwillenberg LO. 1969. Epitheliocystis, a new infectious disease of the bluegill (*Lepomis macrochirus*). Antonie Van Leeuwenhoek 35:146-158.

287. Fredricks DN, Relman DA. 1996. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol Rev 9:18-33.

288. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13:601-612.

289. Seth-Smith HM, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C, Parmar S, Pitt R, Baker S, Ison CA, Marsh P, Jalal H, Lewis DA, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2013. Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. Genome Res 23:855-866.

CHAPTER 2: Aims, significance and project development

The work presented in this thesis describes the genomic characterisation of several novel chlamydial species infecting reptile and fish hosts with the general aims of advancing our understanding of the taxonomic and biological diversity with the phylum *Chlamydiae*. To achieve this aim, this project had the following objectives:

Objective 1: to demonstrate the use of molecular methods for discovery of novel uncultivatable pathogens in the phylum *Chlamydiae*;

Objective 2: to apply molecular methods to understand the molecular epidemiology of novel chlamydial pathogens in the phylum *Chlamydiae*;

Objective 3: to bioinformatically characterise the virulence and metabolic capacities of novel chlamydial pathogens in the phylum *Chlamydiae*.

The work presented in this thesis was performed in the laboratory of Prof. Adam Polkinghorne in the Centre for Animal Health Innovation in the Faculty of Science, Health, Engineering and Education at the University of the Sunshine Coast, Australia, with aspects of the projects conducted with various collaborators from Australian and international institutes, as listed in each research chapter. Prof Polkinghorne is a leader in *Chlamydia* genomics with a keen interest and background in novel chlamydial infections in other animal hosts and the pathogenic mechanisms driving these infections.

Until recently, our knowledge of the host range and diversity of members of the *Chlamydiaceae* was thought to be near-complete (1). Advances in molecular diagnostics have led to the identification and description of an expanded cohort of new taxa within this family from avian, reptilian and other hosts, highlighting how much we still have to learn about the diversity, biology and pathogenicity of the *Chlamydiaceae* as a whole (2-8). Further, the identification of novel families outside the *Chlamydiaceae*, collectively termed *Chlamydia*-related bacteria (CRBs), has both clouded

and expanded our knowledge of chlamydial diversity and evolution (1, 9, 10), as explained in chapter 1.

Ongoing genomics studies within the *Chlamydiaceae* have been paramount to understanding the genetic basis of many of the pathogenic mechanisms employed by chlamydiae [eg. the type three secretion system (T3SS), inclusion membrane proteins and polymorphic membrane proteins (11-14)], as well as offering opportunities to investigate the widespread global diversity and intraspecies evolution of human and veterinary pathogens (15-20). Over the 10 years since the first CRB genome was published, to the commencement of this project, chlamydial genomic studies have expanded in parallel with phylum diversity expansion (21-23). The inclusion of more, diverse, distant chlamydiae in comparative genomics studies has allowed the description of pan- (all genes), core- (shared genes) and accessory (unique genes) genomes within the phylum, families and species (21, 24-26). These studies have been slowly "narrowing in" on the genes required for the unique biphasic intracellular life cycle exhibited by the chlamydiae, but until now, no genomic representatives of any deeply-rooted chlamydiae have been included in these analyses.

The limiting factor to sequencing more, diverse *Chlamydia*-related bacteria, is their requirement to be cultured in a host cell, and for many novel and/or *Candidatus* species, this is lacking. Hence, I have developed a workflow using DNA depletion and enrichment techniques and deep metagenomic sequencing (ie. culture-independent genome sequencing) to characterise the genomes of novel uncultivated chlamydiae, straight from the sample material. Aspects of this method have been developed and used previously for other uncultivated bacteria and for chlamydiae (27, 28). This method addresses several limitations encountered in novel chlamydial infections, such as a low abundance of the pathogen, lack of a reference genome and lack of an *in vitro* culture system.

I addressed the aims by applying our culture-independent genomics method to two species as examples of the taxonomic and biological diversity within the phylum. The first example is of

snakes infected by novel *Chlamydiaceae* species, and the second is of chlamydial infections of fish, which are associated with deeply rooted chlamydiae in the *Ca.* Parilichlamydiaceae family.

The discovery and description of three new taxa in avian hosts prior to the commencement of this project highlighted that significant additional diversity may exist in the 'well-characterised' genus *Chlamydia* (5, 6), and other studies suggested that reptiles may also harbour an as yet untapped diversity of chlamydial species (2, 3, 29-31). In Objectives 1 and 2 of this project, I sought to characterise an outbreak of chlamydial infections of captive snakes in Switzerland. An initial epidemiological study was conducted of six private collections following unexpected mortalities (**Chapter 3**). The study showed that one of the mortalities was attributed to *C. pneumoniae* infection (32). Based on previous studies looking at chlamydial diversity in reptiles, and preliminary data giving inconclusive species identification, I hypothesised that these animals would harbour existing and putative novel chlamydial species, and provide greater insight into the phylogenetic relationships and host tropisms of existing and novel species. Our subsequent screening showed a high prevalence of chlamydial infection in asymptomatic animals, and 16S rRNA sequencing not only revealed *C. pneumoniae* positive animals, but also novel chlamydial species present in these snake populations. Based on these results, as a part of the third objective of this project, I sought to characterise the genomes of novel chlamydial species associated with infections in snakes to identify a number of hallmarks of chlamydial biology in these species, including virulence factors and membrane proteins. These results are described in **Chapter 4** (*Ca.* C. sanzinia) and **Chapter 5** (*Ca.* C. corallus), which were published in the journals, BMC Genomics and Scientific Reports, respectively. The work in these chapters was conducted in collaboration with the Institute of Veterinary Pathology and Section for Epidemiology at the University of Zürich.

The second example utilises genomics of novel members of the deeply-rooting chlamydial family to understand chlamydial pathogenesis and evolution at the phylum-level. The work in

Chapters 6 and 7 follow on from previous work conducted primarily in Prof. Barbara Nowak's laboratory, in which three novel chlamydial species were discovered in three novel fish hosts of aquaculture interest in southern Australia (South Australia and Tasmania) (33-35). These species formed a novel family, *Candidatus* Parilichlamydiaceae, distantly related to the *Chlamydiaceae* and closely related to the *Ca.* Piscichlamydiaceae, the most deeply rooting chlamydial family described several years prior (36, 37). On this basis, I hypothesised that the genomes of these novel chlamydiae would (a) be closer in composition to the most recent common ancestors of this monomorphic order, the *Chlamydiales*, than they would be to the high adapted reduced genomes of the *Chlamydiaceae*; and (b) provide clues to the evolution of pathogenic mechanisms within the broader phylum *Chlamydiae*.

For **Chapter 6**, I characterised a novel chlamydial species (*Ca.* S. epinephelii) infecting cultured Orange-spotted grouper (*Epinephelus coioides*) in North Queensland. This work was conducted in collaboration with researchers at James Cook University who, using histology, found evidence for epitheliocystis in fish experiencing respiratory distress. Order *Chlamydiales*-specific PCR and FISH were then used to confirm the chlamydial association with the cysts. Our epidemiological survey showed that epitheliocystis prevalence and severity varied over the year. This was the first description of a chlamydial species in grouper, and the first report of epitheliocystis in Queensland. Given the as yet uncultivated nature of these pathogens, and with recent advances in culture-independent and metagenomic methods, I utilised genomic data to investigate the pathogenic potential of this novel uncultured epitheliocystis agent. This was the first published chlamydial genome from fish and confirmed that many of the chlamydial hallmarks, particularly virulence mechanisms such as the T3SS were acquired by the chlamydial ancestor. The chapter is presented in the form of a paper that was published in Environmental Microbiology.

Given the host-specific nature of the *Ca.* Parilichlamydiaceae members, I wanted to further investigate genomic differences throughout the family and phylum. **Chapter 7** details this

comparative genomic study of the deeply-rooting *Ca.* Parilichlamydiaceae family. Members of this family continue to be described in novel fish hosts (unpublished data), re-iterating the idea that members of this family are specific to their hosts. This comparative genomics work revealed that members of this family have highly reduced genomes with limited metabolic capacity, which functionally share more similarities with the *Chlamydiaceae* than the CRBs. This work is prepared in the format of a manuscript submitted for publication in ISME J.

This thesis contains eight chapters, which are structured as follows:

In **Chapter 1,** I review the literature and provide a general introduction to the phylum *Chlamydiae*. The first section summarises the taxonomic and biological diversity within the phylum prior to undertaking this project. The second section focuses on chlamydial genomics and its application to our understanding of chlamydial biology.

Chapter 2 outlines the aims and development of the project and the structure of the resulting thesis, with summaries of each chapter, their significance and their relationship to each other.

In **Chapter 3**, I detail the investigation of novel chlamydial infections in captive snakes in Switzerland, which led to genomic descriptions of two novel taxa. This initial study was published in Veterinary Microbiology.

Chapter 4 provides the genomic characterisation of a novel uncultured chlamydiae from the first reptilian host, *Ca*. C. sanzinia, which was published in BMC Genomics.

Chapter 5 provides genomic characterisation of another novel uncultivated chlamydial species found in the snake collections, *Ca.* C. corallus. This work was published in Scientific Reports.

Chapter 6 describes the molecular and genomic characterisation of a novel *Ca.* Parilichlamydiaceae species in a novel fish host in Queensland, published in Environmental Microbiology.

In **Chapter 7** I then conducted comparative genomics on members of the *Ca.* Parilichlamydiaceae family.

Chapter 8 is a general discussion of the findings of this research, and their contribution to the chlamydial field.

References

1. Everett KD, Bush RM, Andersen AA. 1999. Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol 49 Pt 2:415-440.

2. Bodetti TJ, Jacobson E, Wan C, Hafner L, Pospischil A, Rose K, Timms P. 2002. Molecular evidence to support the expansion of the hostrange of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. Syst Appl Microbiol 25:146-152.

3. Hotzel H, Blahak S, Diller R, Sachse K. 2005. Evidence of infection in tortoises by *Chlamydia*-like organisms that are genetically distinct from known *Chlamydiaceae* species. Vet Res Commun 29 Suppl 1:71-80.

4. Sachse K, Vretou E, Livingstone M, Borel N, Pospischil A, Longbottom D. 2009. Recent developments in the laboratory diagnosis of chlamydial infections. Vet Microbiol 135:2-21.

5. Vorimore F, Hsia RC, Huot-Creasy H, Bastian S, Deruyter L, Passet A, Sachse K, Bavoil P, Myers G, Laroucau K. 2013. Isolation of a New *Chlamydia* species from the Feral Sacred Ibis (*Threskiornis aethiopicus*): Chlamydia ibidis. PLoS One 8:e74823.

6. Sachse K, Laroucau K, Riege K, Wehner S, Dilcher M, Creasy HH, Weidmann M, Myers G, Vorimore F, Vicari N, Magnino S, Liebler-Tenorio E, Ruettger A, Bavoil PM, Hufert FT, Rossello-Mora R, Marz M. 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 37:79-88.

7. Aaziz R, Vorimore F, Verheyden H, Picot D, Bertin C, Ruettger A, Sachse K, Laroucau K. 2015. Detection of atypical *Chlamydiaceae* in roe deer (*Capreolus capreolus*). Vet Microbiol 181:318-322.

8. Isaksson J, Christerson L, Blomqvist M, Wille M, Alladio LA, Sachse K, Olsen B, González-Acuña D, Herrmann B. 2015. *Chlamydiaceae*-like bacterium, but no *Chlamydia psittaci*, in sea birds from Antarctica. Polar Biology 38:1931-1936.

9. Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M. 2014. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. ISME J 8:115-125.

10. Corsaro D, Valassina M, Venditti D. 2003. Increasing diversity within *Chlamydiae*. Crit Rev Microbiol 29:37-78.

11. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754-759.

12. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW, Stephens RS. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat Genet 21:385-389.

13. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28:1397-1406.

14. Voigt A, Schofl G, Saluz HP. 2012. The *Chlamydia psittaci* genome: a comparative analysis of intracellular pathogens. PLoS One 7:e35097.

15. Harris SR, Clarke IN, Seth-Smith HM, Solomon AW, Cutcliffe LT, Marsh P, Skilton RJ, Holland MJ, Mabey D, Peeling RW, Lewis DA, Spratt BG, Unemo M, Persson K, Bjartling C, Brunham R, de Vries HJ, Morre SA, Speksnijder A, Bebear CM, Clerc M, de Barbeyrac B, Parkhill J, Thomson NR. 2012. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. Nat Genet 44:413-419, S411.

16. Read TD, Joseph SJ, Didelot X, Liang B, Patel L, Dean D. 2013. Comparative analysis of *Chlamydia psittaci* genomes reveals the recent emergence of a pathogenic lineage with a broad host range. MBio 4.

17. Bachmann NL, Fraser TA, Bertelli C, Jelocnik M, Gillett A, Funnell O, Flanagan C, Myers GS, Timms P, Polkinghorne A. 2014. Comparative genomics of koala, cattle and sheep strains of *Chlamydia pecorum*. BMC Genomics 15:667.

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

19. Hadfield J, Harris SR, Seth-Smith HMB, Parmar S, Andersson P, Giffard PM, Schachter J, Moncada J, Ellison L, Vaulet MLG, Fermepin MR, Radebe F, Mendoza S, Ouburg S, Morre SA, Sachse K, Puolakkainen M, Korhonen SJ, Sonnex C, Wiggins R, Jalal H, Brunelli T, Casprini P, Pitt R, Ison C, Savicheva A, Shipitsyna E, Hadad R, Kari L, Burton MJ, Mabey D, Solomon AW, Lewis D, Marsh P, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2017. Comprehensive global

genome dynamics of *Chlamydia trachomatis* show ancient diversification followed by contemporary mixing and recent lineage expansion. Genome Res doi:10.1101/gr.212647.116.

20. Seth-Smith HMB, Buso LS, Livingstone M, Sait M, Harris SR, Aitchison KD, Vretou E, Siarkou VI, Laroucau K, Sachse K, Longbottom D, Thomson NR. 2017. European *Chlamydia abortus* livestock isolate genomes reveal unusual stability and limited diversity, reflected in geographical signatures. BMC Genomics 18:344.

21. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, Friedman MG, Rattei T, Myers GS, Horn M. 2011. Unity in variety--the pan-genome of the *Chlamydiae*. Mol Biol Evol 28:3253-3270.

22. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, Brandt P, Nyakatura GJ, Droege M, Frishman D, Rattei T, Mewes HW, Wagner M. 2004. Illuminating the evolutionary history of chlamydiae. Science 304:728-730.

23. Bertelli C, Aeby S, Chassot B, Clulow J, Hilfiker O, Rappo S, Ritzmann S, Schumacher P, Terrettaz C, Benaglio P, Falquet L, Farinelli L, Gharib WH, Goesmann A, Harshman K, Linke B, Miyazaki R, Rivolta C, Robinson-Rechavi M, van der Meer JR, Greub G. 2015. Sequencing and characterizing the genome of *Estrella lausannensis* as an undergraduate project: training students and biological insights. Front Microbiol 6:101.

24. Psomopoulos FE, Siarkou VI, Papanikolaou N, Iliopoulos I, Tsaftaris AS, Promponas VJ, Ouzounis CA. 2012. The *Chlamydiales* pangenome revisited: structural stability and functional coherence. Genes (Basel) 3:291-319.

25. Domman D, Collingro A, Lagkouvardos I, Gehre L, Weinmaier T, Rattei T, Subtil A, Horn M. 2014. Massive expansion of ubiquitination-related gene families within the *Chlamydiae*. Mol Biol Evol doi:10.1093/molbev/msu227.

26. Pillonel T, Bertelli C, Salamin N, Greub G. 2015. Taxogenomics of the order *Chlamydiales*. Int J Syst Evol Microbiol 65:1381-1393.

27. Zheng Z, Deng X, Chen J. 2014. Whole-Genome Sequence of "*Candidatus* Liberibacter asiaticus" from Guangdong, China. Genome Announc 2.

28. Seth-Smith HM, Dourala N, Fehr A, Qi W, Katharios P, Ruetten M, Mateos JM, Nufer L, Weilenmann R, Ziegler U, Thomson NR, Schlapbach R, Vaughan L. 2016. Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular beta-proteobacteria. ISME J 10:1791-1803.

29. Soldati G, Lu ZH, Vaughan L, Polkinghorne A, Zimmermann DR, Huder JB, Pospischil A. 2004. Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. Vet Pathol 41:388-397.

30. Homer BL, Jacobson ER, Schumacher J, Scherba G. 1994. Chlamydiosis in mariculture-reared green sea turtles (*Chelonia mydas*). Vet Pathol 31:1-7.

31. Huchzermeyer FW, Gerdes GH, Foggin CM, Huchzermeyer KD, Limper LC. 1994. Hepatitis in farmed hatchling Nile crocodiles (*Crocodylus niloticus*) due to chlamydial infection. J S Afr Vet Assoc 65:20-22.

32. Ruegg SR, Regenscheit N, Origgi FC, Kaiser C, Borel N. 2015. Detection of *Chlamydia pneumoniae* in a collection of captive snakes and response to treatment with marbofloxacin. Vet J 205:424-426.

33. Stride MC, Polkinghorne A, Powell MD, Nowak BF. 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:e82889.

34. 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 79:4914-4920.

35. Stride MC, Polkinghorne A, Miller TL, Groff JM, Lapatra SE, Nowak BF. 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 79:1590-1597.

36. Draghi A, 2nd, Bebak J, Daniels S, Tulman ER, Geary SJ, West AB, Popov VL, Frasca S, Jr. 2010. Identification of *'Candidatus* Piscichlamydia salmonis' in Arctic charr *Salvelinus alpinus* during a survey of charr production facilities in North America. Dis Aquat Organ 89:39-49.

37. Draghi A, 2nd, Popov VL, Kahl MM, Stanton JB, Brown CC, Tsongalis GJ, West AB, Frasca S, Jr. 2004. Characterization of "*Candidatus* piscichlamydia salmonis" (order *Chlamydiales*), a chlamydia-like bacterium associated with epitheliocystis in farmed Atlantic salmon (*Salmo salar*). J Clin Microbiol 42:5286-5297.

CHAPTER 3: Characterisation of *Chlamydia pneumoniae* and other novel chlamydial infections in captive snakes (Veterinary Microbiology 178, 2015)

Alyce Taylor-Brown[a*], Simon Rüegg[b*] Adam Polkinghorne[a] and Nicole Borel[c]

[a] Faculty of Health, Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland 4556, Australia

[b] Section of Veterinary Epidemiology, University of Zurich, Winterthurerstrasse 270, 8057 Zurich, Switzerland

[c] Institute of Veterinary Pathology, University of Zurich, Winterthurerstrasse 268, 8057 Zurich, Switzerland

* Both authors equally contributed to the study

Corresponding author:

Nicole Borel, Prof., DVM, Dipl. ECVP, FVH Pathology

Institute of Veterinary Pathology, Vetsuisse Faculty

University of Zurich, Winterthurerstrasse 268

CH-8057 Zurich, Switzerland

Tel: +41-44-635-8563

Fax: +41-44-635-8934

Email: n.borel@access.uzh.ch

Keywords: *Chlamydia pneumoniae*, *Chlamydia*-like, reptile, phylogenetic analysis, asymptomatic

3.1 Abstract

Chlamydiosis has been described in both free-ranging and captive reptiles. The infection usually manifests as granulomatous inflammation in inner organs such as spleen, heart, lung and liver but might also occur in asymptomatic reptiles. The aim of this study was to investigate and characterise *Chlamydia pneumoniae* and potential other novel chlamydial infections in the choana and cloaca samples of 137 clinically healthy captive snakes from six private collections. Forty eight samples from 29 animals were found to be positive by a *Chlamydiaceae* family-specific qPCR. By *Chlamydia* species-specific ArrayTube Microarray, 43 samples were positive, with 36 of these being identified as *C. pneumoniae*. The prevalence of *Chlamydia* ranged from 5 to 33%. PCR and sequencing of the *Chlamydiales* 16S rRNA signature sequence of 21 *Chlamydia* positive samples revealed the presence of seven novel 16S rRNA genotypes. BLAST-n and phylogenetic analysis of the near-full length 16S rRNA gene sequence of each of these novel 16S rRNA sequences revealed that five genotypes share closest sequence identity to 16S rRNA sequences from *C. pneumoniae* (98.6-99.2%), suggesting that these sequences are novel *C. pneumoniae* strains. One genotype is 96.9% similar to *C. pneumoniae* strains suggesting it may originate from a yet undescribed chlamydial species within the genus *Chlamydia*. The seventh genotype shares 99.0% nucleotide identity with *C. muridarum*. This study further highlights the broad host range for *C. pneumoniae* and suggests that reptiles may still contain a significant and largely uncharacterised level of chlamydial genetic diversity that requires further investigation.

3.2 Introduction

The *Chlamydiae* are a unique phylum of intracellular bacteria that are ubiquitous in the environment and are able to infect and cause disease in a wide range of hosts. In recent years there has been an expansion of this phylum and, as such, it now consists of nine families including a range of novel "*Chlamydia*-like" organisms that share obligate intracellular lifestyles and a similar developmental cycle but otherwise a diverse range of morphologies, host specificities and disease outcomes (Taylor-Brown et al., 2014).

Chlamydiosis has been described in both free-ranging and captive reptilian hosts including puff adders, boas, chameleons, crocodiles, frogs and tortoises (Homer et al., 1994; Bodetti et al., 2002; Jacobson et al., 2002; Soldati et al., 2004; Hotzel et al., 2005; Huchzermeyer et al., 2008). The infection typically manifests as inflammatory lesions in affected organs (granulomatous inflammation) (Howerth, 1984; Homer et al., 1994). Organs commonly affected are the spleen, heart, lung and liver, and diagnosis is usually on the basis of visualisation of the inclusions by light or electron microscopy. In many cases, proliferative pneumonia was observed prior to euthanasia or natural death of reptiles (Bodetti et al., 2002), and limited cases have reported inflammation or lesions in the gastrointestinal tract, wasting disease, necrotising mycocarditis, necrotising enteritis and splenitis (Bodetti et al., 2002; Jacobson et al., 2002; Jacobson et al., 2004; Cope et al., 2014).

Although some studies identified *Chlamydia psittaci* (Huchzermeyer et al., 1994; Robertson et al., 2010) or novel *Chlamydia*-like organisms (Soldati et al., 2004) as the aetiological agent in reptilian chlamydiosis, *Chlamydia pneumoniae* appears to the most widespread (Bodetti et al., 2002; Vlahovic et al., 2006). However, while human and marsupial *C. pneumoniae* strains are well characterised (Roulis et al., 2013), reptilian strains are not.

Detection of asymptomatic carriers in association with a lethal case of granulomatous inflammation of the heart, liver and splenopancreas of a horned viper (*Vipera ammodytes ammodytes*)

caused by *C. pneumoniae* prompted the investigation and characterisation of the diversity of chlamydiae strains circulating in clinically inconspicuous captive snakes in Switzerland.

3.3 Methods

3.3.1 Snake collections

The six private collections (arbitrarily named collections 1 to 6), included in this study, contributed samples from 51, 11, 17, 17, 17 and 24 exotic snakes, respectively. Collections 1 and 4 provided samples from snakes belonging to the families *Viperidae, Colubridae, Pythonidae* and *Elapidae*. Collections 5 and 6 included snakes belonging to *Boidae* and *Pythonidae* families, while collection 6 also included Vipers. Collections 2 and 3 provided samples from Vipers only. Owners 1, 2 and 3 trade snakes while the other owners and their collections have no contact. One snake, in collection 1, was sampled at two time points (Table 3.2).

3.3.2 Sample collection and DNA extraction

An arbitrary selection of 137 snakes from six collections (272 samples in total) was sampled by swabbing the cloaca and choana and extracting DNA according to Borel et al. (2008).

3.3.3 Chlamydial screening

A *Chlamydiaceae* family-specific qPCR targeting the 23S rRNA gene was carried out in 96-well microtitre plates on all samples (n=272), and positive samples were then subjected to an ArrayTube Microarray (AT) targeting 23S rRNA (described in Ehricht et al. (2006)). The AT Microarray chip carried 36 *Chlamydiales* probes, five positive controls and one internal staining control. Sample DNA was amplified and biotin labelled prior to hybridisation according to Borel et al. (2008). Hybridisation patterns were assigned on the basis of the most intense signal, with the provision that all other probes of the same species were among the 10 most intense signals.

A PCR targeting the 298 bp *Chlamydiales* signature sequence of the 16S rRNA gene was conducted on the 23 AT-positive samples using 16SIGF and 16SIGR primers (Everett et al., 1999),

followed by amplicon sequencing. One sample corresponding to each genotype was chosen for near full length (approximately 1400 bp) 16S rRNA sequencing using the 16SIGF primer paired with the 16SB1 primer (Hosokawa et al., 2006), followed by amplicon sequencing.

3.3.4 Phylogenetic analysis

Resulting sequences were first compared to chlamydial sequences previously deposited in Genbank using BLAST (Altschul et al., 1990), then further analysis was conducted in Geneious v.7.1 (Kearse et al., 2012). Sequences were aligned using the Geneious algorithm, and phylogenetic trees were constructed using MrBayes, with a HKY85 substitution model.

3.4 Results

A total of 137 snakes from six collections were tested at both the cloaca and choana to ascertain the presence and diversity of chlamydial infections in captive snakes in six collections in Switzerland.

3.4.1 Detection of chlamydiae in captive snakes

The presence of chlamydiae in this population of snakes ranged from 5.9% to 33.3% in the six collections (Table 3.1a). Initially, a *Chlamydiaceae* specific qPCR and *Chlamydia* species-specific ArrayTube microarray (AT) were conducted to screen for the presence of chlamydiae. By qPCR, 48 samples from 29 animals were positive for *Chlamydiaceae*, and of these, the chlamydial species could be identified in 45 samples by ArrayTube microarray (Table 3.1b). By *Chlamydiaceae* qPCR, four snakes were positive at the choana only, six snakes were positive at the cloaca only and 19 snakes were positive at both sites. By ArrayTube, six animals were positive at the choana only, seven were positive at the cloaca only and 15 were positive at both sites (Table 3.1b). By *Chlamydia* ArrayTube microarray, the hybridisation pattern indicated clear species identity of *C. pneumoniae* at the cloaca of six animals, the choana of four animals and at both sites in 13 animals, *C. muridarum* at both sites from one snake and the genus *Chlamydia* (chlamydial species could not be determined)

in three samples from two animals. An additional four samples from four animals (two choana and two cloaca samples) had inconclusive or uncharacteristic AT hybridisation patterns (Table 3.1b). *C. pneumoniae* was detected in all six collections, and it was the sole *Chlamydia* detected in collections 3, 4 and 6. No chlamydia was detected in snakes belonging to the *Elapidae*.

3.4.2 Chlamydial identity by PCR and sequencing

Further molecular characterization of chlamydial strains circulating in these six collections was performed. A 279-286 bp sequence could be fully resolved from 21 of the 23 AT-positive samples following direct sequencing of the PCR product in both directions. BLAST-n analysis revealed that for 20 of these sequences, closest sequence similarity (95.0-99.0% nucleotide identity) could be found to one of two 'Uncultured *Chlamydia*' sequences (accession numbers GQ507439.1 and GQ507442.1), *C. pneumoniae* strains or a chlamydial sequence isolated from a tortoise (accession number AY845424.1 (Table 3.2)). The last sequence shared 99.0% nucleotide identity with *C. muridarum* (Table 3.2). No sequences shared 100% nucleotide similarity with any sequences in the database.

Multiple sequence alignment of these short novel sequences against other representative 16S rRNA sequences from species in each chlamydial family lead to the designation of seven unique genotypes from the 21 PCR amplified sequences (genotypes 1-7).

Two unique sequences (genotypes 1 and 2) were found in seven samples each, differing from each other by one base. These sequences were both most similar to 'Uncultured *Chlamydia*' isolates (97.5-98.5%) and *C. pneumoniae* (97.9-98.5%). A third novel sequence (genotype 3) was unique to one sample, and shared 96.4% nucleotide similarity with the same uncultured isolates and *C. pneumoniae,* differing by up to 11 bases. Two unique sequences (genotypes 4 and 5) were also found in one sample each. Genotype 4 shares 92.8% nucleotide identity with a '*Chlamydiales* bacterium' isolated from a tortoise; the ArrayTube signal for this sample was characteristic for *Chlamydia* genus. Genotype 5 is 99.5% identical to *C. muridarum* (one SNP), as predicted by the ArrayTube analysis.

The remaining two genotypes (genotypes 6 and 7) are shared between two samples each, differ from each other by six nucleotides and differ from other C*hlamydiaceae* sequences by up to 29 bases each. These unique genotypes individually share closest similarity to the 'uncultured *Chlamydia*' isolates (96.9%) and *C. pneumoniae* (96.9%) as above.

3.4.3 Phylogenetic analysis

These novel 16S rRNA sequences were further analysed using Geneious software. Sequences were aligned with a selection of available chlamydial sequences deposited in GenBank. Phylogenetic analysis of the unique 16S rRNA sequences identified in these samples against other 16S rRNA sequences obtained from the order *Chlamydiales* revealed a distinct clade of five of the unique 16S rRNA sequences identified in this study (genotypes 1, 2, 3, 6 and 7) within the broader clade formed by member species of the family *Chlamydiaceae* (data not shown). The remaining two unique sequences cluster with a *Chlamydiales* sequence obtained from a tortoise and a *C. muridarum* isolate, as expected from BLAST results and sequence analysis (data not shown).

To further understand the phylogenetic position of these novel chlamydiales, near full length (1354-1437 bp) 16S rRNA gene sequences were amplified from one sample corresponding to each genotype (n=7). By Geneious analysis, genotypes 1, 2, 3, 6 and 7 were 99.0, 99.1, 99.2, 98.6 and 99.1% identical to *C. pneumoniae* strain LPCoLN (CP001713.1), respectively, with 10-17 SNPs across 1420-1437 nucleotides, confirming their identity as novel *C. pneumoniae* strains. Genotype 4 was 96.9% identical across 1357 nucleotides (up to 41 differences), suggesting it is a novel species in the genus *Chlamydia*, closely related to *C. pneumoniae*. Genotype 5 shared 99.8% identity with *C. muridarum* (2 SNPs).

These sequence differences are reflected in the phylogenetic tree where genotype 4 branches away from the *C. pneumoniae* clade, separated from *C. pneumoniae* isolates by ≥ 39 nucleotides, whilst the remaining genotypes branch most closely with *C. pneumoniae* isolates, with fewer SNPs (Figure 3.1). By maximum likelihood methods, genotype 6 also diverges before the remaining genotypes,

however bootstrapping for this method was lower overall, and tree topology was not reflective of whole genome phylogeny (Chapter 1, Figure 4). As predicted by the ArrayTube results and initial sequencing, sample 2742-399 (genotype 5) clusters most closely with *C. muridarum*.

3.4.4 Genotype diversity

Of the novel chlamydial 16S rRNA sequences detected in this study, genotypes 1 and 2 were the most common genotypes found: genotype 1 was found in multiple animals from two collections (1 and 3), while genotype 2 was identified in three different snake species (*Vipera latastei*, *Python regius*, *Elaphe guttata* across three collections (2, 5 and 7) (Table 2). Each of the other novel 16S rRNA genotypes was restricted to a single collection.

Adding more confidence to the detection of these novel 16S rRNA genotypes, paired sampling of the choana and cloaca (n=3) revealed the presence of the same 16S rRNA genotype in an individual animal (Table 2). Equally, two snakes in collection 1 that were sampled at different time points were infected with the same genotype. Over half of the novel 16S rRNA genotypes (1, 2, 6 and 7) could be found at either cloaca or choana sample whilst Genotypes 3, 4 and 5 were restricted to either choana or cloaca sample.

In terms of the diversity of chlamydial strains, collection 5 was revealed to be the most diverse with three novel chlamydial 16S rRNA genotypes and one 16S rRNA sequence consistent with *C. muridarum* (Table 2). Owner 6 houses snakes infected with two genotypes, while one genotype per owner was detected in the remaining collections.

3.5 Discussion

In order to verify the presence and assess the diversity of clinically unsuspicious chlamydial infections in snakes, 137 animals from 6 collections were investigated for the presence of *Chlamydia sp*. Forty eight samples from 29 snakes were found to be positive for *Chlamydia sp*., and of these, 23 were identified as *C. pneumoniae*. This is the first report of *C. pneumoniae* and novel chlamydiae

in clinically unsuspicious captive snakes in Switzerland. A recent study from Argentina also described the presence of *C. pneumoniae* in cloaca samples from otherwise unsuspicious captive snakes, using a different detection method (Frutos et al., 2014).

The molecular positivity of *Chlamydia sp.* among six collections ranged from one animal to 11 animals, as is listed in Table 3.1a, with the overall positivity rate being 21.7%. These numbers are slightly lower than prevalence rates of chlamydiosis seen in other reptiles in previous opportunistic studies (Bodetti et al., 2002; Soldati et al., 2004; Robertson et al., 2010; Frutos et al., 2014). The main difference between those and the current study is that the previous studies have involved clinically diseased animals, while the animals in this study were healthy, thus a) they were not presenting clinical signs at the time of sampling and b) their inner organs were not available for analysis. Differences also lie in the methods used to determine positivity, ie, IHC and PCR or PCR alone, as well as the gene amplified in the assays.

Following PCR and sequencing, genotypes were designated on the basis of 16S rRNA sequence similarity. Seven genotypes were found among six collections in Switzerland. Collection 5 contained the highest level of genotype diversity (genotypes 2, 3, 4, 5 and 7). Aside from one sequence (a *C. muridarum* sequence), all sequences shared 96-99% similarity with *C. pneumoniae*. Due to a lack of full length chlamydial 16S rRNA sequences from reptiles, no further comparisons are able to be made within this host range. Nevertheless, phylogenetically, these novel genotypes form a clade separate from animal and human *C. pneumoniae* strains, supporting previous suggestions that human *C. pneumonia*e isolates have diverged from reptilian and amphibian *C. pneumoniae* isolates (Mitchell et al., 2010). The full length 16S rRNA gene of a novel genotype (genotype 4), is 96.9% and 96.8% identical to *C. pneumoniae* strains LPCoLN and TW-183, respectively, placing it as a novel species within the *Chlamydiaceae* (Everett et al., 1999). It was found in one sample only and shared 93.6% nucleotide identity across 279 nucleotides with a previous isolate from a tortoise. Additional research efforts such as whole genome sequencing and *in vitro* cell culture are required

to further understand the phylogenetic position of this genotype as well as biological differences from *C. pneumoniae*. While the chlamydial 16S rRNA signature sequence has proven useful as a genotyping tool in this study, the near full length sequence is necessary to confirm identity and phylogenetic relationships.

It remains unknown exactly why some of the hybridisation patterns were uncharacteristic for *C. pneumoniae*. One sample with a weak *C. pneumoniae* signal was genotyped as genotype 2, which, although across the near full length 16S rRNA gene is 99.2% similar to *C. pneumoniae* LPCoLN, is 99.0% similar to the human strains of *C. pneumoniae,* which the ArrayTube probes are designed against, providing one possible explanation for the weak signal (Ehricht et al., 2006). For the two choana samples that were positive by qPCR but unidentifiable by ArrayTube (2742-323 and 2742-371), 16S rRNA sequencing failed. This could have been due to low chlamydial load (Ct values of ≥ 33), or a mixed sample inhibiting PCR amplification and sequencing, although this was not identified in the present study.

No genotype was associated to any taxon of snake or collection. The present study highlights the genetic diversity of chlamydial organisms infecting clinically unsuspicious captive snakes in Switzerland, which is comparable to that of a previous study that identified five *C. pneumoniae* genotypes among eight reptilian and amphibian host species in geographically distinct locations (Bodetti et al., 2002). The main difference between these two studies is the typing scheme applied. It is also possible that the entire population diversity was not captured in this study as only 21 samples from 18 animals were successfully genotyped. Bodetti et al. (2002) also identified mixed infections with multiple *Chlamydia* species, a phenomenon that was not observed in this study.

While chlamydiosis has been described in a number of reptiles and amphibians, different chlamydial species have been implicated (Berger et al., 1999; Soldati et al., 2004; Blumer et al., 2007; Robertson et al., 2010; Martel et al., 2013). *C. pneumoniae*, on the other hand, has been described more widely in reptiles such as snakes, iguanas and chameleons (Bodetti et al., 2002; Jacobson et

al., 2002; Jacobson et al., 2004). In the current study of captive animals, we found *C. pneumoniae* to be the most common aetiological agent of chlamydiosis in reptiles, with novel chlamydiae related to *C. pneumoniae* also able to infect snakes. As yet the implications for progression for disease in these collections and wider are unknown; further work will aim to isolate these strains for *in vitro* characterisation to investigate the pathogenic potential of *C. pneumoniae* and novel chlamydiae in reptiles.

Detection of *Chlamydiae* in these asymptomatic captive snakes broadened the host range of reptilian species known to be susceptible to chlamydial infection, and with the increasing popularity of keeping reptiles and amphibians as pets, these findings are important in terms of considering the potential zoonotic risk. The same 16S rRNA genotypes were identified in snakes belonging to two of three owners who trade snakes, which could suggest a possible transmission network. Not only is *C. pneumoniae* a common cause of pneumonia in humans, it has also been associated with atherosclerosis, coronary heart disease and Alzeihmer's disease (Roulis et al., 2013).

Declarations

Acknowledgements

The authors are grateful to Carmen Kaiser, Institute of Veterinary Pathology, Vetsuisse Faculty, University of Zurich, for the excellent technical help. This work was partially supported by the Swiss Association for wildlife and zoo animals to NB and SR.

Conflict of interest statement

The authors declare that they have no conflict of interest with respect to the research, authorship, and/or publication of this article.

Author contributions

A.T.B conducted lab work, data analysis and wrote the manuscript. S.R coordinated sample collection and reviewed the manuscript. A.P managed the project and reviewed the manuscript. N.B devised and managed the project and reviewed the manuscript.

Chapter 3 Tables & Figures

Table 3.1a: *Presence of chlamydial infection in captive snakes from six collections in Switzerland.*

Collection	Number of snakes sampled	Number of samples	Number of samples positive by qPCR	Number of samples positive by AT
1	51	98	18	17
2	11	22	5	3
3	17	34	9	9
4	17	34	1	1
5	17	36	11	11
6	24	48	4	4
Total	137	272	48	45

Table 3.1b: *Presence of chlamydial infections in six collections of captive snakes per anatomical site and diagnostic method. This table highlights the positivity levels at different anatomical sites and highlights the discrepancies between AT and qPCR results. Footnotes detail samples that had inconsistent or conflicting results.*

		Collection 1	**Collection 2**	**Collection 3**	**Collection 4**	**Collection 5**	**Collection 6**
Number animals positive by qPCR at choana only		**1**	**1**	**0**	**0**	**1**	**1**
Number animals positive by ArrayTube at choana only		**2**	**3**	**0**	**0**	**0** [4, 5]	**1**
Identity by ArrayTube	*C. pneumoniae*	2	1				1
	Chlamydia/Chlamydophila sp.						
	Other		2 [2,3]				
Number animals positive by qPCR at cloaca only		**3**	**0**	**1**	**1**	**0**	**1**
Number animals positive by ArrayTube at cloaca only		**3**	**0**	**1**	**1**	**1**	**1**
Identity by ArrayTube	*C. pneumoniae*	2		1	1	1	1
	Chlamydia/Chlamydophila sp.	1					
	Other						
Number animals positive by qPCR at both sites		**7**	**2**	**4**	**0**	**5**	**1**
Number animals positive by ArrayTube at both sites		**6** [1]	**0**	**4**	**0**	**4**	**1**

Identity by ArrayTube	*C. pneumoniae*	6		4		2	1
	Chlamydia/Chlamydophila sp.					1	
	Other					1 [6]	
Total animals positive by qPCR at at least one site		**11**	**3**	**5**	**1**	**6**	**3**
Total samples positive by qPCR		18	5	9	1	11	4
Total animals positive by ArrayTube at at least one site		**11**	**3**	**5**	**1**	**5**	**4**
Total samples positive by ArrayTube		17	3	9	1	9	4

[1] 0-11-CL: negative by AT but qPCR positive

[2] 2464-201: weak *C. pneumoniae* signal

[3] 2464-204: unidentifiable signal

[4] 2742-323: unidentifiable signal

[5] 2742-371: unidentifiable signal

[6] 2742-399: *C. muridarum* by AT

Table 3.2. *Genotype designation and distribution for sequences obtained from cloaca and choana swabs from captive snakes in Switzerland.*

Owner	Sample	Anatomical site	Snake species	ArrayTube identity	qPCR Ct value	Genotype	Closest BLAST hit (ID%) (194–198 bp)
1	23430-12	Choana	*Vipera a. ammodytes*	*C. pneumoniae*	25.9	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	23430-09	Cloaca	*Vipera a. ammodytes*	*C. pneumoniae*	30.4	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	1679-8[a]	Cloaca	*Vipera a. ammodytes*	*Chlamydia sp.*	33.1	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	2464-262[a]	Cloaca	*Vipera a. ammodytes*	*C. pneumoniae*	33.6	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
2	2464-202	Choana	*Vipera latastei*	*C. pneumoniae*[b]	29.6	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	2464-203[1]	Choana	*Vipera latastei*	*C. pneumoniae*	31.4	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	2464-204[1]	Cloaca	*Vipera latastei*	inconclusive	32.2	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
3	2464-236	Cloaca	*Montivipera wagneri*	*C. pneumoniae*	30.8	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	2464-246	Cloaca	*Montivipera raddei*	*C. pneumoniae*	32.8	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)

Owner	Sample	Anatomical site	Snake species	ArrayTube identity	qPCR Ct value	Genotype	Closest BLAST hit (ID%) (194–198 bp)
	2464-255	Choana	*Montivipera albizona*	*C. pneumoniae*	26.9	1	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
4	2741-436	Choana	*Eunectes notaeus*	*C. pneumoniae*	28.7	7	Uncultured *Chlamydia* isolate 2040.3 (97%)
	2742-308	Choana	*Sanzania madagascariensis volontany*	*Chlamydia sp.*	21.0	4	*Chlamydiales* bacterium V1242-01 (95%)
	2742-324	Choana	*Corallus batesii*	*C. pneumoniae*	28.6	3	Uncultured *Chlamydia* isolates 2040.3 or EBB (96%)
5	2742-332	Choana	*Corallus caninus*	*C. pneumoniae*	24.6	7	Uncultured *Chlamydia* isolates 2040.3 or EBB (97%)
	2742-387[2]	Choana	*Python regius 2*	*C. pneumoniae*	20.8	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	2742-388[2]	Cloaca	*Python regius 2*	*C. pneumoniae*	26.2	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	2742-399	Cloaca	*Liasisboa/Bothrochilus boa*	*C. muridarum*	26.8	5	*Chlamydia muridarum* (99%)
6	0661-435[3]	Choana	*Atheris squamiger*	*C. pneumoniae*	30.4	6	Uncultured *Chlamydia* isolate 2040.3 (97%)

Owner	Sample	Anatomical site	Snake species	ArrayTube identity	qPCR Ct value	Genotype	Closest BLAST hit (ID%) (194–198 bp)
	0661-436[3]	Cloaca	*Atheris squamiger*	*C. pneumoniae*	30.4	6	Uncultured *Chlamydia* isolate 2040.3 (97%)
	0661-437	Cloaca	*Python regius*	*C. pneumoniae*	29.8	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)
	0661-448	Choana	*Elaphe guttata*	*C. pneumoniae*	31.9	2	Uncultured *Chlamydia* isolates 2040.3 or EBB (98%)

[1,2,3] Samples from different anatomical sites from the same animals (1, 2, 3) at the same timepoints.

[a] Samples from the same animal at different timepoints.

[b] Questionable positive AT result.

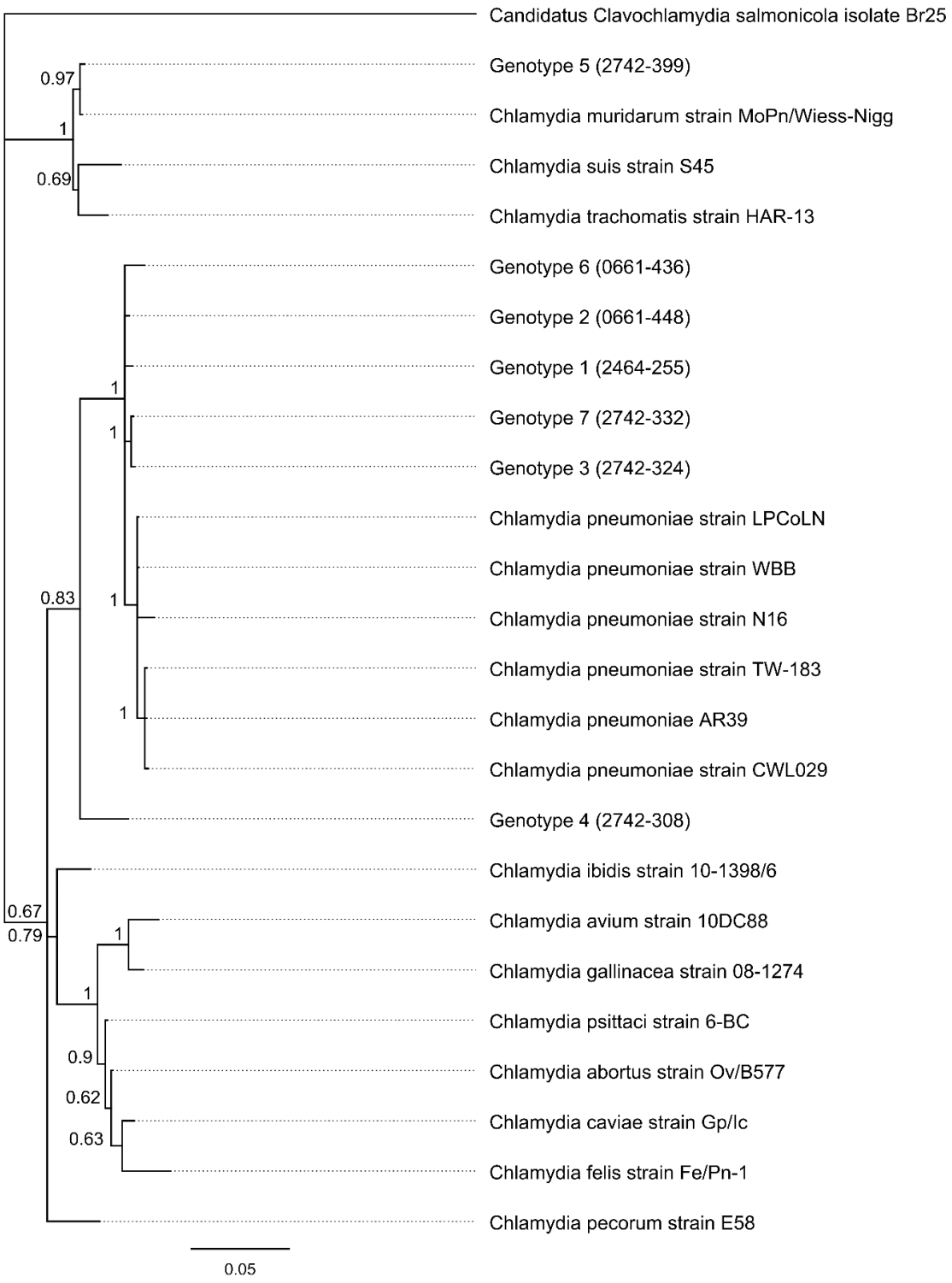


Figure 3.1: *Phylogenetic tree depicting relationships between novel genotypes identified in this study and related chlamydial 16S rRNA gene sequences obtained from Genbank. MrBayes phylogenetic tree constructed in Geneious v7.1. Posterior probabilities are shown on each node.*

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.

Berger, L., Volp, K., Mathews, S., Speare, R., Timms, P., 1999. *Chlamydia pneumoniae* in a free-ranging giant barred frog (*Mixophyes iteratus*) from Australia. J Clin Microbiol 37, 2378-2380.

Blumer, C., Zimmermann, D.R., Weilenmann, R., Vaughan, L., Pospischil, A., 2007. *Chlamydiae* in free-ranging and captive frogs in Switzerland. Vet Pathol 44, 144-150.

Bodetti, T.J., Jacobson, E., Wan, C., Hafner, L., Pospischil, A., Rose, K., Timms, P., 2002. Molecular evidence to support the expansion of the hostrange of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. Syst Appl Microbiol 25, 146-152.

Borel, N., Kempf, E., Hotzel, H., Schubert, E., Torgerson, P., Slickers, P., Ehricht, R., Tasara, T., Pospischil, A., Sachse, K., 2008. Direct identification of chlamydiae from clinical samples using a DNA microarray assay: a validation study. Mol Cell Probes 22, 55-64.

Cope, I., Wheelhouse, N., Pocknell, A., Dagleish, M., Summers, B., 2014. An unusual presentation of *Chlamydia pneumoniae* infection in a Royal Python (*Python regius*). Veterinary Record Case Reports 2.

Ehricht, R., Slickers, P., Goellner, S., Hotzel, H., Sachse, K., 2006. Optimized DNA microarray assay allows detection and genotyping of single PCR-amplifiable target copies. Mol Cell Probes 20, 60-63.

Everett, K.D., Bush, R.M., Andersen, A.A., 1999. Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol 49 Pt 2, 415-440.

Frutos, M.C., Monetti, M.S., Re, V.E., Cuffini, C.G., 2014. Molecular evidence of *Chlamydophila pneumoniae* infection in reptiles in Argentina. Rev Argent Microbiol 46, 45-48.

Homer, B.L., Jacobson, E.R., Schumacher, J., Scherba, G., 1994. Chlamydiosis in mariculture-reared green sea turtles (*Chelonia mydas*). Vet Pathol 31, 1-7.

Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M., Fukatsu, T., 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol 4, e337.

Hotzel, H., Blahak, S., Diller, R., Sachse, K., 2005. *Evidence of infection in tortoises by Chlamydia-like organisms that are genetically distinct from known Chlamydiaceae species.* Vet Res Commun 29 Suppl 1, 71-80.

Howerth, E.W., 1984. Pathology of naturally occurring chlamydiosis in African clawed frogs (*Xenopus laevis*). Vet Pathol 21, 28-32.

Huchzermeyer, F.W., Gerdes, G.H., Foggin, C.M., Huchzermeyer, K.D., Limper, L.C., 1994. Hepatitis in farmed hatchling Nile crocodiles (*Crocodylus niloticus*) due to chlamydial infection. J S Afr Vet Assoc 65, 20-22.

Huchzermeyer, F.W., Langelet, E., Putterill, J.F., 2008. An outbreak of chlamydiosis in farmed Indopacific crocodiles (*Crocodylus porosus*). J S Afr Vet Assoc 79, 99-100.

Jacobson, E., Origgi, F., Heard, D., Detrisac, C., 2002. Immunohistochemical staining of chlamydial antigen in emerald tree boas (*Corallus caninus*). J Vet Diagn Invest 14, 487-494.

Jacobson, E.R., Heard, D., Andersen, A., 2004. Identification of *Chlamydophila pneumoniae* in an emerald tree boa, *Corallus caninus*. J Vet Diagn Invest 16, 153-154.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

Martel, A., Adriaensen, C., Sharifian-Fard, M., Vandewoestyne, M., Deforce, D., Favoreel, H., Bergen, K., Spitzen-van der Sluijs, A., Devisscher, S., Adriaens, T., Louette, G., Baert, K., Hyatt, A., Crameri, S., Haesebrouck, F., Pasmans, F., 2013. The novel '*Candidatus* Amphibiichlamydia ranarum' is highly prevalent in invasive exotic bullfrogs (*Lithobates catesbeianus*). Environ Microbiol Rep 5, 105-108.

Mitchell, C.M., Hutton, S., Myers, G.S., Brunham, R., Timms, P., 2010. *Chlamydia pneumoniae* is genetically diverse in animals and appears to have crossed the host barrier to humans on (at least) two occasions. PLoS Pathog 6.

Robertson, T., Bibby, S., O'Rourke, D., Belfiore, T., Agnew-Crumpton, R., Noormohammadi, A.H., 2010. Identification of Chlamydial species in crocodiles and chickens by PCR-HRM curve analysis. Vet Microbiol 145, 373-379.

Roulis, E., Polkinghorne, A., Timms, P., 2013. *Chlamydia pneumoniae*: modern insights into an ancient pathogen. Trends Microbiol 21, 120-128.

Soldati, G., Lu, Z.H., Vaughan, L., Polkinghorne, A., Zimmermann, D.R., Huder, J.B., Pospischil, A., 2004. Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. Vet Pathol 41, 388-397.

Taylor-Brown, A., Vaughan, L., Greub, G., Timms, P., Polkinghorne, A., 2014. Twenty years of research into *Chlamydia*-like organisms: a revolution in our understanding of the biology and pathogenicity of members of the phylum *Chlamydiae*. Pathogens and Disease (doi: 10.1093/femspd/ftu009)

Vlahovic, K., Dovc, A., Lasta, P., 2006. Zoonotic aspects of animal chlamydioses - A review. Veterinary Archives 76, 259-274.

CHAPTER 4: Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes (BMC Genomics 17, 2016)

Alyce Taylor-Brown[a], Nathan L Bachmann[a], Nicole Borel[b], Adam Polkinghorne[a*]

Centre for Animal Health Innovation, Faculty of Science, Health, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland 4556, Australia[a];

Institute of Veterinary Pathology, University of Zurich, Winterthurerstrasse 268, 8057 Zurich, Switzerland[b]

* Address correspondence to Adam Polkinghorne, apolking@usc.edu.au

Keywords*: Chlamydia*, culture-independent sequencing, genomics, reptile

4.1 Abstract

Background

Recent molecular studies have revealed considerably more diversity in the phylum *Chlamydiae* than was previously thought. Evidence is growing that many of these novel chlamydiae may be important pathogens in humans and animals. A significant barrier to characterising these novel chlamydiae is the requirement for culturing. We recently identified a range of novel uncultured chlamydiae in captive snakes in Switzerland, however, nothing is known about their biology. Using a metagenomics approach, the aim of this study was to characterise the genome of a novel chlamydial taxon (*Candidatus* Chlamydia sanzinia 2742-308) from the choana of a captive snake, to expand our knowledge of the additional diversity and biology of the genus *Chlamydia*.

Results

We identified two chlamydial genomic contigs: a 1,113,073 bp contig, and a 7,504 bp contig, representing the chromosome and plasmid of *Ca.* Chlamydia sanzinia 2742-308, respectively. The 998 predicted coding regions include an expanded repertoire of outer membrane proteins (pmp and omp genes), some of which exhibited frameshift mutations, as well as several chlamydial virulence factors such as the translocating actin-recruitment phosphoprotein (*Tarp*) and macrophage inhibition potentiator (*Mip*). A suite of putative inclusion membrane proteins were also predicted. Notably, no evidence of a traditional chlamydial plasticity zone was identified. Phylogenetically, *Ca.* Chlamydia sanzinia forms a clade with *C. pneumoniae* and *C. pecorum*, distinct from former "*Chlamydophila*" species.

Conclusions

Genomic characterisation of a novel uncultured chlamydiae from the first reptilian host has expanded our understanding of the diversity and biology of a genus that was thought to be the most well-characterised in this unique phylum. It is anticipated that this method will be suitable for characterisation of other novel chlamydiae.

4.2 Background

The *Chlamydiae* are a phylum of intracellular bacteria that are characterised by their unique biphasic lifecycle [1-3]. While they are ubiquitous in the environment [4], a significant number are also associated with disease in a wide range of hosts [2]. The traditional family in this phylum, *Chlamydiaceae*, consists of a single genus, *Chlamydia*, and includes important human and animal pathogens such as *Chlamydia trachomatis* and *Chlamydia psittaci*. Even though this is the best-understood family in the phylum, two new species, and one *Candidatus* species, in the genus *Chlamydia*, were recently described from birds [5-6], highlighting how little we still know about the full diversity of these obligate intracellular pathogens. Aside from human and mammalian hosts [7], chlamydiosis has been reported in both free-ranging and captive reptilian hosts including several snake species, turtles, tortoises and crocodiles, among others [8-11]. The most common species found in these hosts to date is *Chlamydia pneumoniae*, however, a recent study also revealed 16S rRNA sequences corresponding to potentially novel *Chlamydia* species [11].

Barriers for characterising the biology of these unique intracellular parasites lie in the fact they require a host cell to undergo replication. For many novel species, *in vitro* culture systems are not available. With deep sequencing becoming faster, more affordable and higher throughput [12], groups within the chlamydia field have recently developed several methods of deep sequencing from clinical samples to gain insight into the biology of these species [13-17]. These alternative methods bypass the labour-intensive and costly culture step that has hampered genomic characterisation of novel pathogens.

In the current study, we have utilised a culture-independent method to sequence the genome of a previously uncharacterised and uncultivable new member of the genus *Chlamydia*. The subsequent comparative genomics enabled by this approach allows for fast and effective identification of a number of hallmarks of chlamydial biology in these novel species, including virulence factors and membrane proteins.

4.3 Results and discussion

4.3.1 Description of "Candidatus Chlamydia sanzinia", sp. nov.

We propose a novel species with Candidatus status in the genus *Chlamydia,* based on genomic DNA sequence similarity in the absence of maintenance in culture: *Ca.* Chlamydia sanzinia (*Sanzinia,* pertaining to the host genus name). *Ca.* C. sanzinia shares < 97% 16S rRNA nucleotide identity with other *Chlamydia spp*. Additionally, it shares < 96%, <98%, <96%, <95% and <95% nucleotide sequence identity with other *Chlamydia spp*. for *rpoN, ftsK, pepF*, *adk* and *hemL*, respectively; classifying it as a new species using the scheme recommended by Pillonel *et al* [18]. This novel species was detected in the choana of a captive asymptomatic Madagascar tree boa (*Sanzinia madagascariensis volontany*) in a private collection in Switzerland.

4.3.2. Metagenome reconstruction and chlamydial genome assembly

5,561,445 paired reads were obtained from the treated DNA sample following shotgun sequencing on an Illumina MiSeq. As no reference genome was available for this putative novel species, reads were trimmed prior to *de novo* assembly into 261,306 contigs.

BLASTn analysis revealed two contigs that were suspected to be of chlamydial origin: a 1,113,073 bp chromosomal contig with 81 % nucleotide identity with *C. pneumoniae* LPCoLN (CP001713.1), and a 7,504 bp plasmid contig with 77% nucleotide identity with the plasmid of *C. pneumoniae* LPCoLN (CP001714.1).

329,886 reads mapped to the chromosomal contig, accounting for ~6% of the reads. The mean read coverage across the genome was ~44x, with at least 10x read coverage at every base with the exception of some of the predicted polymorphic membrane proteins. Interestingly, the mean coverage of the plasmid was estimated to be ~1888x, with the plasmid reads accounting for ~37% of the total reads.

Despite treatment of the DNA to enrich for microbial DNA, at least 227,252 contigs are believed to be host-derived, based on BLASTn analysis against two available snake genomes (*Python bivittatus*; AEQU00000000.2 and *Pantherophis guttatus;* JTLQ00000000.1), although the majority of these contigs are very short. While the method used to enrich for microbial DNA relies on depletion of methylated DNA, there is evidence to suggest that vertebrate mitochondrial DNA may not be methylated at many regions [19]. This would explain the high coverage (~1904x) obtained for the mitochondrial genome obtained in this metagenome, accounting for ~5% of the reads.

The metagenome was also screened for host-associated microbiota to assess the proportion of reads devoted to other bacterial species. Full or partial 16S rRNA sequences were detected for five non-chlamydial bacteria in the sample, with significantly lower coverage than the chlamydial genome.

Automated annotation of the chlamydial chromosome and plasmid by RAST followed by manual annotation in Artemis resulted in prediction of 998 coding regions, demonstrating a similarly reduced gene content to other members of the *Chlamydiaceae* (Table 4.1). Roughly one third of the genome consists of hypothetical proteins (314). The two chlamydial genomic contigs of *Ca.* Chlamydia sanzinia, had G+C contents of 38.5% and 32.3%, respectively (Table 4.1) [5, 6]. The chromosome was shown to be able to be circularised *in silico* by read mapping across the contig break.

4.3.3 Ca. Chlamydia sanzinia 2742-308 forms a distinct clade with C. pneumoniae and C. pecorum

Using the novel sequence data for this unique chlamydial taxon, we extracted the core genome for comparative phylogenetic analyses to show that *Ca.* Chlamydia sanzinia 2742-308 is most closely related to *C. pneumoniae* (LPCoLN strain), with an average nucleotide identity of 76.9%, in agreement with the 16S rRNA sequence identity previously described for this putative novel species [11]. In our phylogenetic tree construction, this taxon shares the same minor clade with *C.*

pneumoniae and *C. pecorum*, within a major clade comprised of former "*Chlamydophila*" *spp.* and the novel avian *Chlamydia spp* and distinct from *C. muridarum, C. suis* and *C. trachomatis* (Figure 4.1). Branch lengths indicate that *Ca.* Chlamydia sanzinia and *C. pneumoniae* have evolved at a similar rate after divergence from a common ancestor.

4.3.4 The genome of Ca. Chlamydia sanzinia 2742-308 harbours a 7.5 Kbp cryptic plasmid

Chlamydial plasmids are nonconjugative, nonintegrative and found in most chlamydial species [19]. While we are still understanding the function of this plasmid, studies have suggested roles in regulation of the developmental cycle and in influencing tissue tropism and disease outcome [20-22]. *De novo* assembly of the trimmed reads produced an extra-chromosomal contig resembling the chlamydial plasmid, the gene content and arrangement of which are typical of other chlamydial plasmids (Figure 4.2b). The *Ca.* Chlamydia sanzinia 2742-308 plasmid proteins share 42.7 to 87.2% amino acid similarity with *Chlamydia spp.* plasmid proteins. The phylogenetic tree constructed from a concatenated alignment of the conserved proteins, reflects these relationships (Figure 4.2a), which mirror that of the chromosome. The nucleotide sequence identity across the plasmid was much lower, ranging from 31 to 69%, although much of this variation can be attributed to genes absent from some of the plasmids (Table 4.1, Figure 4.2b).

It is also interesting to note that coverage of the plasmid contig was almost 43 times that of the chromosome (~1,888x vs ~44x and accounting for ~37% of the reads), suggesting a plasmid copy number of up to 43 per chromosome. This seems unlikely given previous descriptions of plasmid copy numbers in the range of two to ten [23-25]. An alternative explanation for the high coverage is that the MDA process preferentially amplifies the plasmid. Other studies have also shown that plasmid copy numbers change throughout the developmental cycle [23-25].

4.3.5 The genome of Ca. Chlamydia sanzinia 2742-308 does not appear to contain a plasticity zone

The chlamydial plasticity zone (PZ) is a region of extensive variation between chlamydial genomes [2, 7], which, while highly variable, generally harbours (i) acetyl coA carboxylase chains

(*acc*BC), (ii) cytotoxin genes/adherence factor, (iii) phospholipase D (*pld*), (iv) membrane attack complex/perforin (*MACPF*), (v) tryptophan biosynthesis operon (*trp*ABFCDR, *kyn*U, *prs*A) and (vi) purine biosynthesis genes (*gua*AB-*add*) [2, 7]. Iterations of the chlamydial PZ have been described in the genomes of all species of *Chlamydia* to date ranging from ~12 kbp to ~86 kbp, with 11 to 48 genes [2, 7].

Despite rigorous homology and conserved domain searches through all 261,306 contigs, the only features of a chlamydial PZ in *Ca.*Chlamydia sanzinia 2742-308 are *acc*B & *acc*C (Cs308_0799 & Cs308_0800) (Figure 4.3). While this region is very small, it is similar to that of its closest relative, *C. pneumoniae* LPCoLN and to that of *C. avium* 10DC88, both of which only possess *acc*BC and *MACPF*. Two hypothetical proteins have weak sequence similarity to *C. psittaci* adherence factor (*tox*) (Cs308_0802 and Cs308_0803). The finding of neither a *MACPF* or plasticity zone-*pld* is consistent given the evidence that MACPF may assist PLDs in lipid acquisition and processing [2], in which case the absence of a MACPF may depend on the absence of a PLD. The lack of a complete *trp* operon, a feature shared with its closest relative, *C. pneumoniae*, as well as other *Chlamydia spp.* except *C. pecorum*, *C. caviae* and *C. felis,* suggests a different pathway for synthesising tryptophan. Interestingly, an aromatic amino acid synthase, which has been described as an alternative pathway for *trp* synthesis in *C. pneumoniae* [26], also appears to be absent. The lack or truncation of the *guaAB-add* operon for purine biosynthesis is common to *C. trachomatis*, *C. abortus*, and some strains of *C. pneumoniae* and *C. psittaci*. The genome of *Ca.* Chlamydia sanzinia does however appear to possess an AMP nucleosidase (Cs308_0522) which plays a role in purine nucleoside salvage.

We considered the possibility that the plasticity zone was either not sequenced or not assembled, given that multiple displacement amplification may introduce an amplification bias [27], as seen for the plasmid. However, overlapping read mapping to the joined ends of the chromosomal contig demonstrates that we have obtained the complete genome for this bacterium.

The absence of an obvious chlamydial plasticity zone characteristic to other members in the genus, as such, sets this putative novel species apart from its known closest phylogenetic relatives. Strain diversity studies in other chlamydial species suggest that further investigation could elucidate remnants of a PZ if other strains of this novel taxon were to be characterised.

4.3.6 An expanded repertoire of polymorphic membrane proteins are located on "pmp islands" for Ca. Chlamydia sanzinia 2742-308

Chlamydial membrane proteins are thought to play an important role in host-parasite interaction throughout the chlamydial developmental cycle. Given their antigenic properties, many have also been the target of extensive vaccine studies [28]. Polymorphic membrane proteins (Pmps) are one such family of membrane protein, unique to chlamydiae, that are highly variable but are united by their GGA (I, L, V) and FxxN tetrapeptide motifs [28-29]. They also function as autotransporters in the Type V secretion system [29]. Investigation of the genome of *Ca.* Chlamydia sanzinia 2742-308 revealed not only the presence of homologues of previously described *pmp* genes, exhibiting the aforementioned motifs, but also the presence of an expanded group of membrane proteins annotated as *omp5,* an outer membrane protein (Cs308_0070, 0072, 0079, 0084, 0085, Cs308_674-677). These protein encoding genes share significant sequence homology with either *pmp* or *omp* proteins from *C. pneumoniae*, *C. pecorum*, *C. psittaci* and *C. abortus* (43-62% amino acid similarity). This family of membrane protein genes, together with the *pmp*s, are arranged in four clusters, not dissimilar to the *pmp* distribution in other chlamydial genomes.

Although twelve *omp5* genes were annotated by automated methods, further analysis determined three of these (Cs308_0074, Cs308_0076 and Cs38_0077) to be fragments of two *pmpG* pseudogenes, (Cs308_0073 and Cs308_0075), with truncations attributed to frameshift mutations. The fragments together contain the repeat motifs, middle domain protein and autotransporter domains characteristic to *pmp*s. Frameshifts are common to *pmp* encoding genes, and it has been suggested that these mutations promote antigenic diversity [26].

Pmp genes account for around 4 to 5% of the coding regions in closely related chlamydial species (up to 22% for some strains of *C. pneumoniae* [26]). The predicted *Pmp*-encoding genes in this genome account for 3.1%; with the inclusion of the additional *omp5* genes these membrane proteins together account for approximately 5% of the coding regions of this genome, suggesting the expanded *pmp* and *omp* repertoire confers characteristics advantageous to survival inside the host cells infected by this novel chlamydial agent.

4.3.7 Ca. Chlamydia sanzinia 2742-308 displays genomic hallmarks of a pathogen

The *Chlamydiae* possess a number of mechanisms to exploit the host cell machinery in order to survive and replicate inside a host cell. One of the ways in which chlamydiae achieve this is by the secretion of "effectors", virulence factors which function to influence host signalling, cleave host proteins and suppress host defences. Homologues of several virulence factors secreted by the Type Three Secretion System (T3SS) are present in the genome of *Ca.* Chlamydia sanzinia 2742-308: namely, a homolog of translocated actin-recruiting phosphoprotein (*Tarp*) (Cs308_0200), macrophage inhibition potentiator (*mip*) (Cs308_0291), tail-specific protease (*tsp*) (Cs308_0179) and serine/threonine protein kinase (*Pkn*D) (Cs308_0709) [30], while genes encoding the chlamydial outer protein, CopB, and a chlamydial protein associating with death domains (CADD), appear to be absent [30].

Another chlamydial virulence factor, chlamydial protease-like activity factor (CPAF) (Cs308_0639), secreted by the *sec*-dependent Type II secretion pathway [31], was also identified in this novel genome.

The inclusion membrane proteins (Incs) are another group of membrane proteins unique to the *Chlamydiae,* exposed to the host cytosol and hypothesised to be involved in inclusion membrane biogenesis [32, 33]. They share little sequence similarity but instead possess a 40-60 amino acid bi-lobed hydrophobic structure [32]. In addition to two copies of *Inc*A and single copy each of *Inc*B and *Inc*C (Cs308_0059, Cs308_0863, Cs_0864, Cs_0885), an additional 41 putative Inc proteins

were predicted *in silico* in the *Ca.* Chlamydia sanzinia 2742-308 genome. This is within the range described for other *Chlamydia spp,* which are predicted to encode as few as 36 Inc proteins (*C. trachomatis*) or up to as many as 107 (*C. pneumoniae*) [32], providing further evidence that these proteins are integral to inclusion development and host interaction across the genus [2].

4.4 Conclusion

For many novel species within and outside of the *Chlamydiae*, culture-independent sequencing from clinical samples provides a unique opportunity to understand the biology of the species for which there is a) no established culture system or b) no reference genome. While our group and others have recently used several genome sequencing methods to broaden our knowledge of previously described species in the genus *Chlamydia* [13-17], the current study suggests that a shotgun deep sequencing approach is better suited to novel species. For instance, using "bait" probes designed from a reference genome risks overlooking previously undetected or undescribed features, such as a plasmid [25]. Likewise, using only homology or binning approaches for metagenomes from clinical samples such as this may also omit accessory proteins or extra-chromosomal sequences that either share no homology to known proteins or are lacking in conserved phylogenetic markers. The use of MDA for whole genome amplification prior to genome sequencing does come with its own limitations however, such as such as preferential amplification of certain genomic regions or particular bacteria in a microbial community [27, 34]. The amplification skew is unpredictable and appears to be heavily dependent on both the complexity of the sample and the MDA protocol applied [27]. These issues appear to be able to be partially overcome by combining MDA with other depletion or enrichment methods [15]. Nonetheless, these tools are particularly useful for uncultivable intracellular organisms such as members of the phylum *Chlamydiae,* and can likely be incredibly valuable for further characterisation of a range of novel chlamydiae reported within and outside the genus *Chlamydia*.

This method also circumvents sequencing of genetic changes that are often acquired during passaging.

We have used this method to sequence and characterise a novel uncultured bacterium in the genus *Chlamydia* from the first reptilian host, expanding our understanding of the diversity and biology of a genus that was thought to be the most well-characterised in this biologically unique phylum.

4.5 Methods

4.5.1 DNA extraction and microbial DNA enrichment

Genomic DNA was extracted from a swab from the choana of a Madagascar tree boa (*Sanzinia madagascariensis volontany*) as per [11]. Total genomic DNA was then subject to host methylated-DNA depletion using the NEBNext Microbiome DNA Enrichment kit (New England BioLabs, USA), according to manufacturer's instructions. The selectively enriched DNA was purified by ethanol precipitation before being subject to multiple displacement amplification using the Qiagen Repli-G mini kit (Qiagen, Germany), according to manufacturer's instructions.

4.5.2 Genome sequencing and assembly

Sequencing was carried out on an Illumina MiSeq at the Australian Genome Research Facility, Walter & Eliza Hall Institute, Parkville, Australia using 150 bp paired end reads. Read quality was assessed with FastQC v0.11.2 and trimmed using Trimmomatic v.035 [35]. Trimmed reads were *de novo* assembled using SPAdes v3.1.1 [36], with kmer values of 51, 71, 91, 101 and 127 in both multi-cell and single-cell mode. QUAST [37] was used to assess the quality of the assemblies.

4.5.3 Genome annotation and analysis

Resulting contigs were subject to a BLASTx search against an in-house chlamydial protein database. Contigs longer than 1000 bp with hits with e-values ≤ 0.005 and identity values ≥ 60%

were subsequently also manually compared to the NCBI database. Two chlamydial contigs were uploaded to RAST [38] for automated annotation. Additional coding regions were identified and annotated using Artemis [39]. The chromosomal contig was split to resemble the genome architecture of *C. trachomatis,* and reads were mapped back to the assembly to assess genome coverage. A 60 bp region with <10x coverage was removed from the 5' end of the contig, resulting in reads overlapping the the 5' and 3' ends of the contig. A similar method was also applied to the plasmid contig; a 124bp region was removed from the 5' end. The nucleotide sequence of the genome of uncultured *Chlamydia sp.* 2742-308 was deposited in Genbank under accession numbers CP014639 (chromosome) and CP014640 (plasmid).

To identify the origins of non-chlamydial contigs, Metaxa [40] was employed to detect ribosomal RNA sequences of prokaryotic and eukaryotic origins. These results informed downstream read-mapping to assess the proportion of reads belonging to different origins using Geneious [41], BWA aligner [42], and BEDTools [43].

4.5.4 Phylogenetic analysis

To assess the genome-wide phylogenetic relationships among *Chlamydia spp*., the core genome was extracted using the LS-BSR package [44]. A phylogenetic tree was constructed from the computed alignment using FastTree [45] and visualised in Geneious [41]. For the plasmid phylogeny, each gene was extracted and translated *in silico*, prior to concatenation. The phylogenetic tree was then constructed from the resulting alignment.

4.5.5 Inclusion membrane protein prediction

The amino acid sequences for all 314 hypothetical proteins were subject to transmembrane helix prediction using TMHMM [46]. Hydropathy plots were visualised to identify characteristic bi-lobed hydrophobic domains.

Declarations

Ethics approval and consent to participate

The collection and molecular analysis of the snake sample was approved and performed in accordance with the relevant guidelines and regulations of the Veterinary Office of Canton Zurich (authorization no. ZH010/15).

Availability of data and material

The nucleotide sequence of the genome of uncultured *Chlamydia sp.* 2742-308 was deposited in Genbank under accession numbers CP014639 (chromosome) and CP014640 (plasmid); bioproject PRJNA312988.

Competing interests

The authors declare no competing financial interests.

Authors' contributions

A.T.B performed the experiments and analysis, and wrote the main manuscript. N.L.B contributed to analysis and prepared Figure 4.3. All authors reviewed the manuscript.

Acknowledgements

We thank Simon Rüegg for sample collection Carmen Kaiser for DNA extraction and initial PCR screening. We thank Helena Seth-Smith for her advice on the preparation of our samples for culture-independent genome sequencing as well as downstream analysis.

Chapter 4 Tables & Figures

Table 4.13: *Comparative analysis of chlamydial genome features. One strain representative for each species was analysed.*

Species, strain (accession number)	***Uncultured Chlamydia sanzinia 2742-308 (CP014639)***	***C. pneumoniae LPCoLN (CP006571.1)***	***C. pecorum MC/ MarsBar (NZ_CM002310.1)***	***C. psittaci 6BC (NC_017287.1)***	***C. felis Fe/C-56 (AP006861.1)***	***C. caviae GPIC (AE015925.1)***	***C. abortus S26/3 (NC_004552.2)***	***C. avium 10DC88 (CP006571.1)***	***C. gallinacea 08-1274/3 (AWUS00000000.1)***	***C. ibidis 10-1398/6 (APJW00000000.1)***	***C. muridarum Nigg (NC_002620.2)***	***C. trachomatis A/HAR-13 (NC_007429.1)***	***C. suis MD56 (CM002267.1)***
Chromosome length (Mbp)	1.11	1.24	1.11	1.17	1.17	1.17	1.14	1.04	1.05	1.15	1.07	1.04	1.07
GC content (%)	38.5	40.5	41.1	39.1	39.4	39.2	39.9	36.9	37.9	38.3	40.3	41.3	42.0
No. CDSs	998	1,097	945	1,003	1,005	998	964	940	898	949	904	911	915
Hypothetical proteins	314	426	297	337	324	376	219	242	207	235	353	294	218
Plasmid length (Kbp) (No. ORFS)	7.5 (8)	7.5[a] (8)	7.5[a] (8)	7.5 (8)	7.5 (8)	7.9 (7)	Np	7.1 (7)	7.0 (7)	Np	7.5 (8)	7.5[b] (8)	5.9 (6)

Np; No plasmid

[a] Plasmid not present in all strains

[b] Plasmid length of *C. trachomatis* L2b/CS784/08

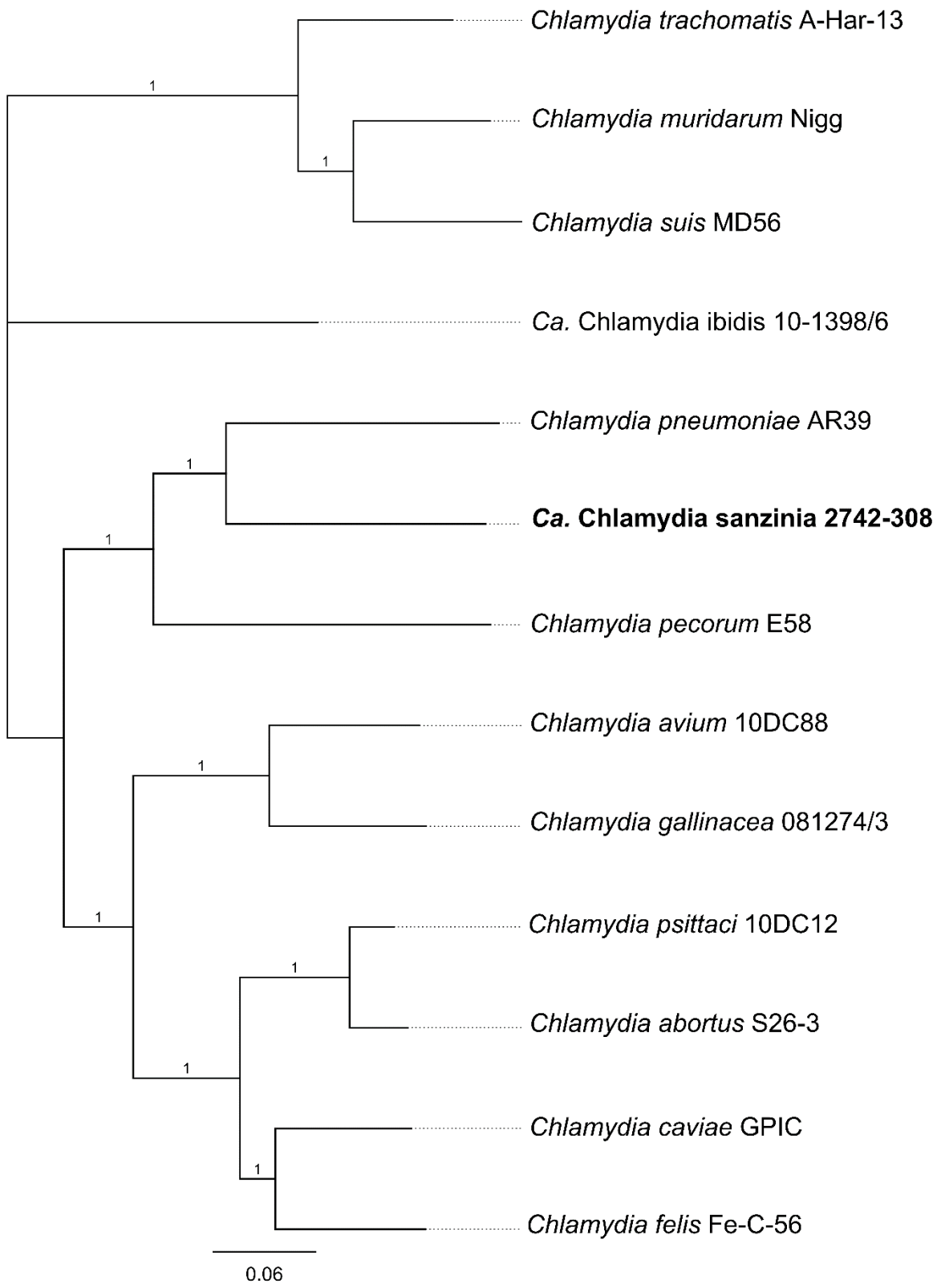


Figure 4.1: Chlamydia *species core genome phylogenetic tree. The core genome was extracted using the LS-BSR package and phylogenetic tree constructed using FastTree. Numbers on the branches indicate support values. Scale bar indicates nucleotide substitutions per site.*

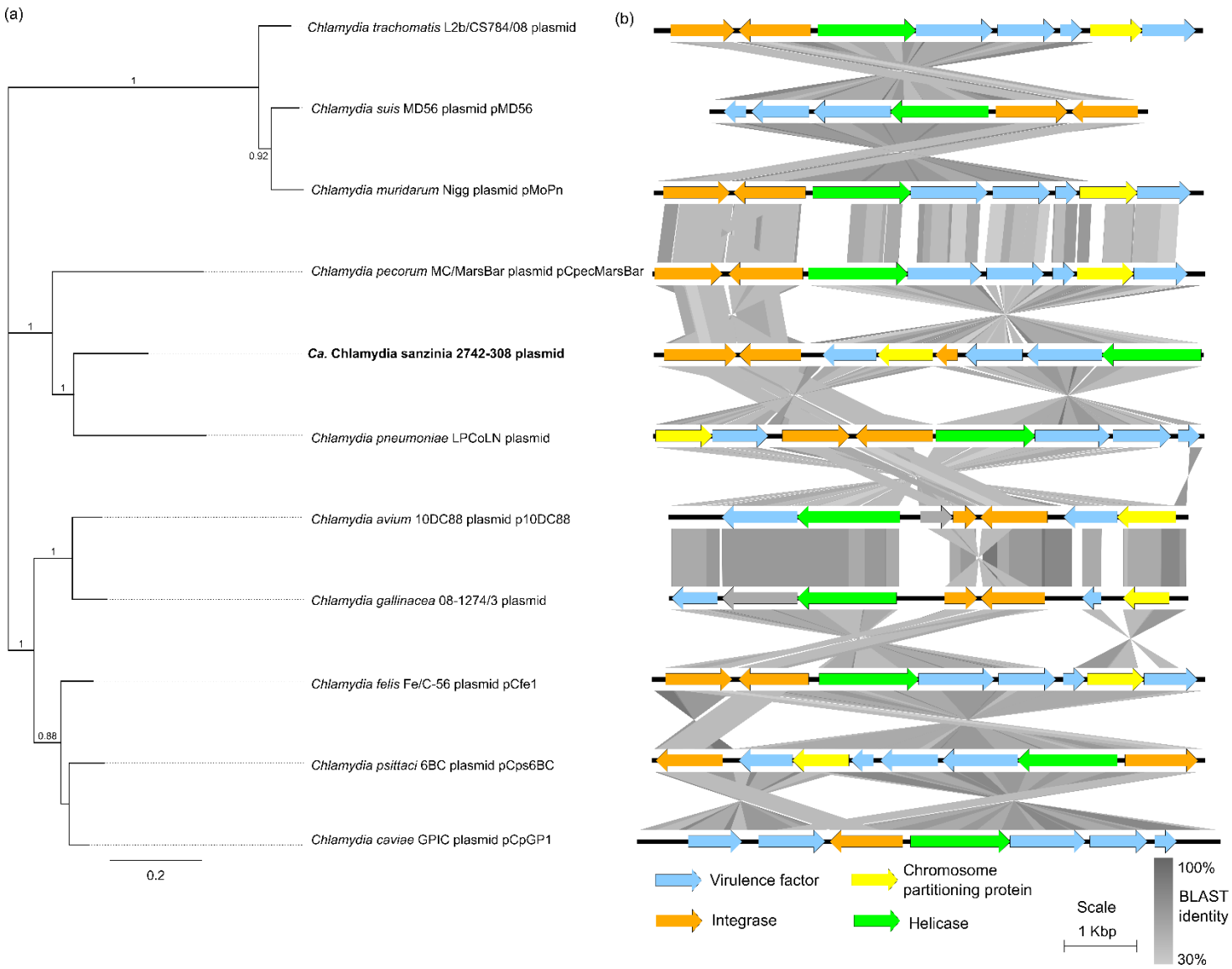


Figure 4.2: *Chlamydial plasmid phylogeny and arrangement. Chlamydial plasmid proteins were extracted from each sequence, concatenated and aligned prior to Maximum Likelihood phylogenetic tree construction using the FastTree algorithm in Geneious (a);* Chlamydia *plasmid nucleotide sequences were compared via tBLASTx analysis and their arrangement plotted in EasyFig (b). Block arrows represent proteins, coded by colour and grey shading represents sequence homology.*

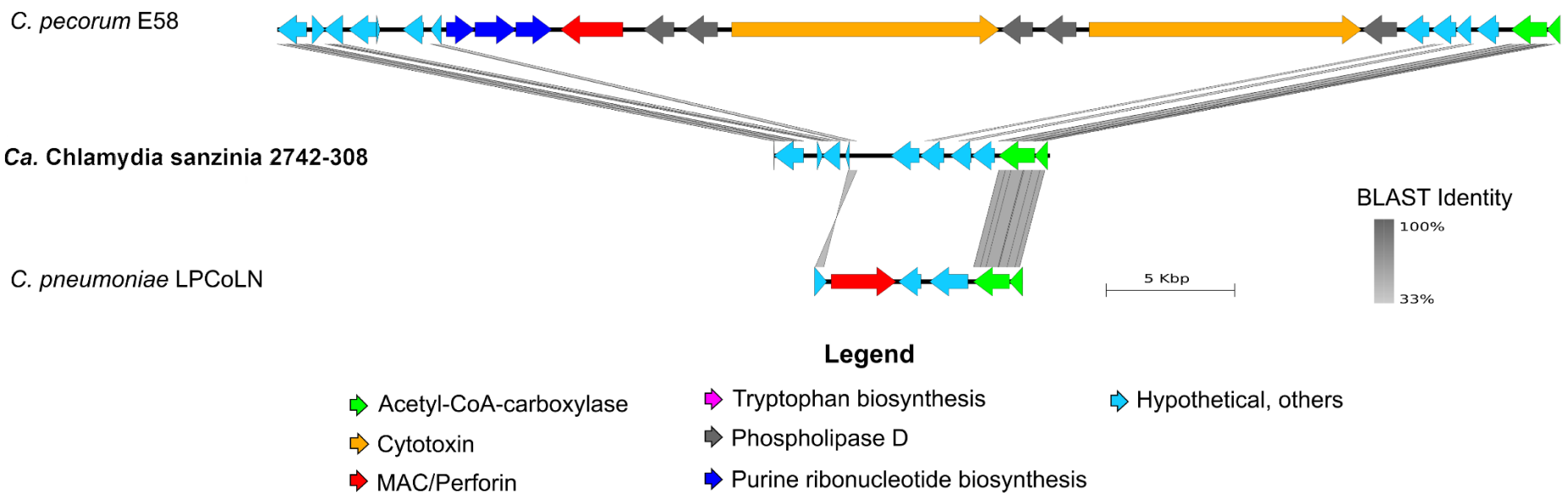


Figure 4.3: *Lack of the plasticity zone in* Chlamydia sp. *2742-308. The region of the genome encoding the plasticity zone in* C. pecorum *and* C. pneumoniae *were compared to that of* Chlamydia sp. *2742-308 via tBLASTx analysis and their arrangement plotted in EasyFig. Block arrows represent proteins, coded by colour and grey shading represents sequence homology.*

References

1. Abdelrahman YM, Belland RJ: The chlamydial developmental cycle. *FEMS microbiology reviews* 2005, 29(5):949-959.

2. Nunes A, Gomes JP: Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. *Infection, genetics and evolution* 2014, 23:49-64.

3. Taylor-Brown A, Vaughan L, Greub G, Timms P, Polkinghorne A: Twenty years of research into *Chlamydia*-like organisms: a revolution in our understanding of the biology and pathogenicity of members of the phylum *Chlamydiae*. *Pathogens and disease* 2015, 73(1):1-15.

4. Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M: Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. *The ISME journal* 2014, 8(1):115-125.

5. Vorimore F, Hsia RC, Huot-Creasy H, Bastian S, Deruyter L, Passet A, Sachse K, Bavoil P, Myers G, Laroucau K: Isolation of a new *Chlamydia* species from the Feral Sacred Ibis (*Threskiornis aethiopicus*): Chlamydia ibidis. *PloS one* 2013, 8(9):e74823.

6 Sachse K, Bavoil PM, Kaltenboeck B, Stephens RS, Kuo CC, Rossello-Mora R, Horn M: Emendation of the family *Chlamydiaceae*: proposal of a single genus, *Chlamydia*, to include all currently recognized species. *Systematic and applied microbiology* 2015, 38(2):99-103.

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

8. Bodetti TJ, Jacobson E, Wan C, Hafner L, Pospischil A, Rose K, Timms P: Molecular evidence to support the expansion of the hostrange of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. *Systematic and applied microbiology* 2002, 25(1):146-152.

9. Soldati G, Lu ZH, Vaughan L, Polkinghorne A, Zimmermann DR, Huder JB, Pospischil A: Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. *Veterinary pathology* 2004, 41(4):388-397.

10. Huchzermeyer FW, Langelet E, Putterill JF: An outbreak of chlamydiosis in farmed Indopacific crocodiles (*Crocodylus porosus*). *Journal of the South African Veterinary Association* 2008, 79(2):99-100.

11. Taylor-Brown A, Ruegg S, Polkinghorne A, Borel N: Characterisation of *Chlamydia pneumoniae* and other novel chlamydial infections in captive snakes. *Veterinary microbiology* 2015, 178(1-2):88-93.

12. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW: Transforming clinical microbiology with bacterial genome sequencing. *Nature reviews Genetics* 2012, 13(9):601-612.

13. Putman TE, Suchland RJ, Ivanovitch JD, Rockey DD: Culture-independent sequence analysis of *Chlamydia trachomatis* in urogenital specimens identifies regions of recombination and in-patient sequence mutations. *Microbiology* 2013, 159(10):2109-2117.

14. Seth-Smith HM, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C *et al*: Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. *Genome research* 2013, 23(5):855-866.

15. Seth-Smith HM, Harris SR, Scott P, Parmar S, Marsh P, Unemo M, Clarke IN, Parkhill J, Thomson NR: Generating whole bacterial genome sequences of low-abundance species from complex samples with IMS-MDA. *Nature protocols* 2013, 8(12):2404-2412.

16. Christiansen MT, Brown AC, Kundu S, Tutill HJ, Williams R, Brown JR, Holdstock J, Holland MJ, Stevenson S, Dave J *et al*: Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. *BMC infectious diseases* 2014, 14:591.

17. Bachmann NL, Sullivan MJ, Jelocnik M, Myers GS, Timms P, Polkinghorne A: Culture-independent genome sequencing of clinical samples reveals an unexpected heterogeneity of infections by *Chlamydia pecorum*. *Journal of clinical microbiology* 2015, 53(5):1573-1581.

18. Pillonel T, Bertelli C, Salamin N, Greub G: Taxogenomics of the order Chlamydiales. *International Journal of Systematic and Evolutionary Int J Syst Evol Microbiology* 2015, 65(4):1381-1393.

19. Hong EE, Okitsu CY, Smith AD, Hsieh CL: Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial DNA. *Molecular and cellular biology* 2013, 33(14):2683-2690.

20. Rockey DD: Unraveling the basic biology and clinical significance of the chlamydial plasmid. *The Journal of experimental medicine* 2011, 208(11):2159-2162.

21. Song L, Carlson JH, Whitmire WM, Kari L, Virtaneva K, Sturdevant DE, Watkins H, Zhou B, Sturdevant GL, Porcella SF *et al*: *Chlamydia trachomatis* plasmid-encoded Pgp4 is a transcriptional regulator of virulence-associated genes. *Infection and immunity* 2013, 81(3):636-644.

22. Yang C, Starr T, Song L, Carlson JH, Sturdevant GL, Beare PA, Whitmire WM, Caldwell HD: Chlamydial Lytic Exit from Host Cells Is Plasmid Regulated. *MBio* 2015, 6(6):e01648-01615.

23. Pickett MA, Everson JS, Pead PJ, Clarke IN: The plasmids of *Chlamydia trachomatis* and *Chlamydophila pneumoniae* (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents. *Microbiology* 2005, 151(Pt 3):893-903.

24. Ferreira R, Borges V, Nunes A, Borrego MJ, Gomes JP: Assessment of the load and transcriptional dynamics of *Chlamydia trachomatis* plasmid according to strains' tissue tropism. *Microbiological research* 2013, 168(6):333-339.

25. Jelocnik M, Bachmann NL, Kaltenboeck B, Waugh C, Woolford L, Speight KN, Gillett A, Higgins DP, Flanagan C, Myers GS *et al*: Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas. *BMC genomics* 2015, 16:893.

26. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW, Stephens RS: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nature genetics* 1999, 21(4):385-389.

27. Yilmaz S, Allgaier M, Hugenholtz P: Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* 2010, 7(12):943-944.

28. Vasilevsky S, Stojanov M, Greub G, Baud D: Chlamydial polymorphic membrane proteins: regulation, function and potential vaccine candidates. *Virulence* 2016, 7(1):11-22.

29. Henderson IR, Lam AC: Polymorphic proteins of *Chlamydia spp*.--autotransporters beyond the *Proteobacteria*. *Trends in microbiology* 2001, 9(12):573-578.

30. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S *et al*: Unity in variety--the pan-genome of the *Chlamydiae*. *Molecular biology and evolution* 2011, 28(12):3253-3270.

31. Chen D, Lei L, Lu C, Flores R, DeLisa MP, Roberts TC, Romesberg FE, Zhong G: Secretion of the chlamydial virulence factor CPAF requires the Sec-dependent pathway. *Microbiology* 2010, 156(10):3031-3040

32. Lutter EI, Martens C, Hackstadt T: Evolution and conservation of predicted inclusion membrane proteins in chlamydiae. *Comparative and functional genomics* 2012, 2012:362104.

33. Mital J, Miller NJ, Dorward DW, Dooley CA, Hackstadt T: Role for chlamydial inclusion membrane proteins in inclusion membrane structure and biogenesis. *PloS one* 2013, 8(5):e63426.

34. Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, Holland T, Cotton D, Hauser L, Keller M: Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Applied & Environmental Microbiology* 2007, 73(10):3205-3214.

35. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30(15):2114-2120.36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al*: SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 2012, 19(5):455-477.

37. Gurevich A, Saveliev V, Vyahhi N, Tesler G: QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013, 29(8):1072-1075.

38. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: The RAST Server: rapid annotations using subsystems technology. *BMC genomics* 2008, 9:75.

39. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: Artemis: sequence visualization and annotation. *Bioinformatics* 2000, 16(10):944-945.

40. Bengtsson J, Eriksson KM, Hartmann M, Wang Z, Shenoy BD, Grelet GA, Abarenkov K, Petri A, Rosenblad MA, Nilsson RH: Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie van Leeuwenhoek* 2011, 100(3):471-475.

41. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, 28(12):1647-1649.

42. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010, 26(5):589-595.

43. Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26(6):841-842.

44. Sahl JW, Caporaso JG, Rasko DA, Keim P: The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2014, 2:e332.

45. Price MN, Dehal PS, Arkin AP: FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution* 2009, 26(7):1641-1650.

46. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 2001, 305(3):567-580.

CHAPTER 5: Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus *Chlamydia* (Scientific Reports 7, 2017)

Alyce Taylor-Brown[1], Labolina Spang[1], Nicole Borel[2], Adam Polkinghorne[1#]

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

[2] Institute of Veterinary Pathology, University of Zurich, Zurich, CH-8057, Switzerland

Corresponding author: apolking@usc.edu.au

5.1 Abstract

Advances in culture-independent methods have meant that we can more readily detect and diagnose emerging infectious disease threats in humans and animals. Metagenomics is fast becoming a popular tool for detection and characterisation of novel bacterial pathogens in their environment, and is particularly useful for obligate intracellular bacteria such as *Chlamydiae* that require labour-intensive culturing. We have used this tool to investigate the microbial metagenomes of *Chlamydia*-positive cloaca and choana samples from snakes.

The microbial complexity within these anatomical sites meant that despite previous detection of chlamydial 16S rRNA sequences by single-gene broad-range PCR, only a chlamydial plasmid could be detected in all samples, and a chlamydial chromosome in one sample. Comparative genomic analysis of the latter revealed it represented a novel taxon, *Ca.* Chlamydia corallus, with genetic differences in regards to purine and pyrimidine metabolism. Utilising statistical methods to relate plasmid phylogeny to the phylogeny of chromosomal sequences showed that the samples also contain additional novel strains of *Ca.* C. corallus and two putative novel species in the genus *Chlamydia*.

This study highlights the value of metagenomics methods for rapid novel bacterial discovery and the insights it can provide into the biology of uncultivable intracellular bacteria such as *Chlamydiae*.

5.2 Introduction

Recent advances in culture-independent molecular methods and diagnostics, coupled with an increased breadth and depth of sampling, have played a significant role in detecting emerging disease threats in humans and animals (1-4). This method is advantageous over traditional single-gene pathogen detection methods in which nucleotide differences alone are, in some cases, not powerful enough to distinguish between closely related species (5-7). Metagenomic sequencing also addresses challenges encountered in a clinical setting: a) if the putative pathogen is novel or unknown, b) if there are no established culture systems, or c) if culture is time-consuming and laborious (eg. for obligate intracellular pathogens such as *Chlamydiae)*.

We recently used metagenomics to sequence and characterise the genome of a putative novel Candidatus species, *Ca.* Chlamydia sanzinia (8), originating from a diverse group of chlamydial strains circulating among clinically healthy, captive snakes in Switzerland (9). This species is closely related to *Chlamydia pneumoniae* and its genome encodes several chlamydial virulence markers such as a type three secretion system, translocated actin-recruiting phosphoprotein (Tarp) and chlamydial protease-like activity factor (CPAF) (8). Elsewhere, chlamydiosis has been described in both wild and captive reptiles, including crocodiles, lizards and snakes, in broad geographical locations, with the impact of infection ranging from asymptomatic infections to severe disease (9-16). Little or nothing is otherwise known about the the biological diversity of chlamydiae infecting these hosts.

Further, few studies have used metagenomics to describe the metagenome and microbiota of wild or non-model vertebrates, with most studies focussing on mammalian species and agriculturally important animals (17, 18). Recently, groups have used culture-dependent and independent methods to characterise the microbiota of several anatomical sites in various reptile hosts, to uncover the diversity and function within these communities, and their potential impact on animal and human health (19-23). We therefore aimed to assess the microbial diversity in

Chlamydia-positive cloaca and choana samples from captive, asymptomatic snakes. In doing so, we also showed that metagenomics analysis is not only useful for novel chlamydial species discovery but that it also reveals key genomic differences between a novel chlamydial taxon and established species.

5.3 Results and discussion

5.3.1 Snake choana and cloaca metagenome assembly and microbial composition

135,167,964 reads were obtained across two cloaca (G1/1679-8 and G2/2464-204) and three choana (G3/2742-324, G6/0661-435, and G7/2741-436) samples in which novel *C. pneumoniae* strains were detected by 16S rRNA sequencing. Reads were trimmed for quality and adapter sequence prior to *de novo* assembly and metagenome binning using SPAdes and MaxBin, respectively. 27,763-378,622 contigs were obtained for the samples (76-10,516 contigs over 1,000 bp).

Metagenomic assessment revealed a high level of complexity in the samples, with such deep sequencing allowing us to simultaneously uncover the microbiota of these sites, putative novel bacteria and microbial eukaryotes residing in the choana and cloaca of these snakes. The cloaca samples harboured up to five bacterial species, and the choana samples up to two species based on recovery of full 16S rRNA sequences and partial or complete bacterial genomes (Table 4,1).

Many of the BLAST hits of microbiota species were known members of reptilian, piscine or mammalian microbiomes, for example *Achromobacter sp.* and *Luteimonas sp.* in the respiratory tract (21, 22) and *Serratia marcescens, Pseudomonas aeruginosa* and *Salmonella enterica* in the cloaca (23). Interestingly, *Chitinophageaceae* appeared to dominate the choana samples, which has not been described before. This discrepancy is most likely due to the different detection methods used (19-23). It is unclear what role these bacteria are playing at these sites: *S. enterica* has been repeatedly described as a reptile pathogen and such a high level of abundance may provide evidence for this

(Table 5.1). As the name suggests, the *Chitinophagaceae* members in the choana samples (which are 98.5% identical to each other) are rich in chitinases. The role for these bacteria and their enzymes in the choana/oral cavity are unknown, but they may contribute to the digestion of the exoskeleton of animals ingested as part of the snake's diet. Interestingly, previous studies of reptile microbiota did not detect any *Chlamydia* species (19-23). This highlights a strength of metagenomics and strongly suggests that it may be a pathogen rather than a commensal, however, additional *in vitro* and *in vivo* studies are obviously necessary to confirm this. We also detected rRNA sequences from a flagellate, *Monocercomonas coluborum,* in a choana sample (G1/1679-8), and a fungal species related to *Sporothrix schenckii* in a cloaca sample (G7/2741-436). Further metagenomic sequencing would clarify the presence, abundance and roles of these and other species in the choana and cloaca of snakes.

Despite methylated DNA depletion prior to MDA and sequencing, a host mitochondrial genome was recovered from each sample. It has been shown that mitochondrial DNA may not be methylated in all species (24), and this is reflected in our data by the presence of these sequences at differing levels of coverage in each sample, combined with fragmented mitochondrial genomes for two species. The mitochondrial genomes were obtained on a single contig or over up to six contigs, with the read coverage ranging from 43x to 38,621x, accounting for ~0.07% to ~31% of the reads.

5.3.2 Chlamydial genome construction from a snake choana metagenome

Given the fact that more than one bacterial genome was present in these metagenomes, a full chlamydial genome could only be recovered from a single sample (G3/2472-324), the characteristics of which are summarised in Table 5.2 in comparison with other chlamydial genomes. The single metagenome containing chlamydial chromosomal contigs contained 4,445 contigs over 1,000 bp, seven of which were chlamydial and divided between six predicted chromosomal contigs and one predicted plasmid contig. The combined chromosomal contigs total 1,196,452 bp in

length and were predicted to encode 1,076 genes. The GC content of 39.33% was comparable to other chlamydial genomes (Table 5.2). The plasmid contig was 7,522 bp, harboured the typical eight open reading frames and has a lower GC content than the chromosome (32.0%), as is expected for plasmids. The average read coverage across the chromosome and plasmid were ~110x and ~40,134x, respectively.

For the remaining four samples, despite lacking a chlamydial chromosome, a plasmid sequence could be recovered, presumably since chlamydial plasmids are found at copy numbers of two to ten times that of the chromosome (25, 26) and because MDA may preferentially amplify the plasmid DNA (8). The plasmid sequences ranged from 7,210 to 7,534 bp, with 5-14,471x average read coverage (Table 3).

5.3.3 Phylogenomic analysis of Ca. Chlamydia corallus within the Chlamydiaceae

To assess the genetic relationship of Ca. Chlamydia corallus to other chlamydial species, we utilised the classification scheme published by Pillonel *et al.*, 2015 (5). Sequence homology within the 16S rRNA gene placed this novel taxon as a member of the order *Chlamydiales,* (99.2% identical to *C. pneumoniae* LPCoLN [9]). Sequence analysis of the additional genes show that this sample is closely related to *C. pneumoniae* but is sufficiently genetically different (ANI 90.16-90.31% with *C. pneumoniae* AR39 and LPCoLN) that, according to this scheme, G3/2742-324 should be classified as a novel species within the genus *Chlamydia*. To visualise the genetic relationships between this putative new chlamydial species with other members of the genus *Chlamydia*, a phylogenetic tree was constructed from a concatenation of the eleven gene alignments (5). The resultant phylogenetic tree re-iterates the distinct lineage formed by G3/2742-324, in a major cluster with *Ca.* Chlamydia sanzinia, *C. pneumoniae* and *C. pecorum* (Figure 5.1).

Based on the nucleotide identities of the analysed genes highlighted in the classification scheme (5), G3/2742-324 should be classified as a novel Candidatus species within the genus *Chlamydia*. We propose for it the name *Candidatus* Chlamydia corallus (strain G3/2742-324), so

named for the genus of the Amazon Basin emerald tree boa (*Corallus batesii*) in which it was detected. *Ca.* Chlamydia corallus was detected in the choana of a clinically healthy, captive snake in Switzerland.

5.3.4 Plasmid-based diversity within chlamydial species infecting snakes

Although culture-independent genome sequencing failed to resolve whole genome sequences for the all samples in this study, an extra-chromosomal plasmid was detected in all five metagenomes. An approximately 7.5 Kbp contig from G1/1679-8, G2/2464-204, G3/2742-324, G7/2741-436, and an approximately 7.2 Kbp contig from G6/0661-435, showed a BLASTn hit against the *C. pneumoniae* LPCoLN plasmid. The nucleotide sequence identity between these sequences and *C. pneumoniae* LPCoLN plasmid was 78.76-86.09% and among each other was 77.45-96.8%.

Almost all chlamydial species, but not all strains, are known to carry a plasmid, and the nucleotide and amino acid sequences are highly conserved between species (26). The presence of a plasmid has been suggested to contribute to the pathogenicity or tissue tropism of the chlamydial species (27, 28), and plasmid proteins are used for diagnostic targets and vaccine candidates. The chlamydial plasmid is normally organised with eight open reading frames (ORFs), encoding for genes involved in plasmid maintenance and glycogen synthesis (29). Alignment of the plasmid sequences revealed conservation of these ORFs, with the exception of a gap in the coding region for helicase in the sequence for G6/0661-435, which resulted in a partial plasmid sequence.

Previous research has also shown that there is a co-evolution between the chromosome and plasmid sequences for the chlamydial species (26, 30), so in the absence of chlamydial chromosomal genetic data for these additional samples, we performed phylogenetic analysis on the nucleotide sequences across the plasmid in order to assess the genetic relationship of all strains obtained from the metagenomes in this study. In agreement with the tree constructed from its chromosomal loci, phylogenetic analysis revealed that *Ca.* Chlamydia corallus clusters with but is

genetically distinct from *C. pneumoniae*. (Figure 5.2). Plasmid sequences from G6/0661-435 and G7/2741-436 can also be found in this clade. Notably, G1/1679-8 and G2/2464-204 form two additional branches distinct from *Ca.* C. corallus, *C. pneumoniae* and *Ca.* C. sanzinia, sharing 84.09% of their nucleotides with each other and 72.27% to 83.51% with these three species.

As no typing scheme exists to distinguish species based on plasmid sequence analysis or phylogeny, we used linear regression analysis to assess the relationships between chromosome and plasmid pair-wise patristic distances (sum of branch lengths) within the *Chlamydiaceae*. Based on the phylogenetic markers used in this study, at the strain level (eg. *C. pneumoniae* LPCoLN and N16; *C. pecorum* MC/MarsBar and L1), patristic distances for the chromosome are 0, while for the plasmid they are 0.01-0.02. For closely related species pairs such as *C. caviae* and *C. felis* or *C. suis* and *C. trachomatis*, chromosomal and plasmid patristic distances are 0.12-0.16 and 0.20-0.22, respectively. For more distantly related species pairs such as *C. trachomatis* and *C. psittaci*, or *C. pecorum* and *C. muridarum*, chromosomal and plasmid patristic distances are higher: 0.28-0.33 and 0.52-0.57. For the sequences obtained in this study, branch lengths between G6/0661-435 and G7/2741-436, and between these two samples and G3/2742-324, are equivalent to those at the strain level (0.0), as are their extrapolated chromosomal patristic distances based on the curve. On the other hand, the plasmid and extrapolated chromosomal patristic distances between G1/1679-8, G2/2464-204 and *C. pneumoniae* of 0.18-0.21 and 0.11-0.13, respectively, are slightly lower than those of *C. caviae* and *C. felis,* but not as close as that of *C. psittaci* and *C. abortus,* highlighting their relatedness to each other (Figure 5.2). Meanwhile, their branch lengths with most other members of the genus is 0.36-0.41, which is comparable to most other pair-wise distances, thus may represent two distinct novel species.

These data also fit with initial testing results, in which G1/1679-8 and G2/2464-204 could not be definitively assigned to a species based on a *Chlamydiaceae* ArrayTube assay designed to detect established species (8) (chapter 3). G1/1679-8 was identified as a *Chlamydia* species, and

G2/2464-204 did not yield a conclusive result (Table 3) (9). G3/2742-324, G6/0661-435 and G7/2741-436 were designated as *C. pneumoniae,* which is in line with their close plasmid nucleotide identity. This suggests the assay is less specific when taxa are so closely related, but is robust enough to detect novel species.

Given the above, the phylogenetic position of G1/1679-8 and G2/2464-204 among other species and their evolutionary distances from other species and strains provide strong evidence of additional species-level diversity within the *Chlamydiaceae*. These data provide a) evidence that, for some taxa, 16S rRNA sequencing is not sufficient to speciate, b) validation of the use of genome sequencing to further investigate genetic diversity within and/or between populations, and c) evidence for the use of plasmid sequence to assess diversity and phylogeny of novel chlamydial species for which plasmids are ubiquitous.

5.3.5 Genetic differences within the plasticity zone of Ca. Chlamydia corallus

In order to further characterise the genome of the novel species, *Ca.* Chlamydia corallus, in comparison to other chlamydial species, the plasticity zone (PZ) region was analysed. The plasticity zone is a unique region within the *Chlamydia* genome that has been associated with host adaptation for some chlamydial species (6, 31). The well-known variability between the chlamydial species within this region makes it an appropriate target for understanding the factors that might have influenced the tissue tropism of *Ca.* Chlamydia corallus.

The plasticity zone of *Ca.* Chlamydia corallus is approximately 13,700 bp in size and composed of genes required for several biochemical pathways such as Acetyl-CoA-carboxylase (*accBC)*, purine and pyrimidine synthesis genes (*guaABadd*) and the MAC/perforin gene, as seen in Figure 5.3. When compared with other chlamydial species, the plasticity zone harboured by *Ca.* Chlamydia corallus is structurally most similar to the human-isolated strain of *C. pneumoniae* AR39 (Figure 5.3). Both species have a slightly smaller plasticity zone than other species, due to the absence of any cytotoxin, which is present in *C. psittaci* and duplicated in *C. pecorum* (32, 33). The

main difference between the PZs of *Ca.* C. corallus and *C. pneumoniae* appears to be fragmented or truncated hypothetical proteins in AR39, and the absence of a putative lipoprotein in *Ca.* Chlamydia corallus, which is present in both strains of *C. pneumoniae* (Figure 5.3) (34).

The plasticity zone of *Ca.* Chlamydia corallus was genetically variable from *Ca.* Chlamydia sanzinia (Figure 5.3). For instance, the MAC/perforin complex gene was not detected in the plasticity zone of *Ca.* Chlamydia sanzinia (8). The function of the MAC/perforin gene in the chlamydial species is unknown, but has been suggested to contribute to the pathogenesis of these species (35). Additional differences in the PZs of *Ca.* Chlamydia sanzinia and *Ca.* Chlamydia corallus lie in the purine ribonucleotide biosynthesis pathways, as highlighted in Figure 5.3. The purine ribonucleotide biosynthesis (*guaABadd*) cluster, detected in *Ca.* Chlamydia corallus, plays a critical role in both *de novo* and salvage pathways for purine synthesis in prokaryotes (36). This cluster is present in some chlamydial species (33, 34, 37), but was never detected in the other recently described snake chlamydia, *Ca.* Chlamydia sanzinia (8). Chlamydial species that do not encode for this gene cluster are most likely able to synthesise purines through alternative pathways (38). Its absence in other bacterial species, such as *Helicobacter pylori*, has been found to have an effect on their rate of growth (36). The absence of the *guaABadd* genes in several of the chlamydial species, however, indicates that these genes are not needed by the chlamydial species; its absence in *Ca.* Chlamydia sanzinia also suggests that these genes are not necessary for the chlamydial species to establish infection in snakes (8, 37).

No tryptophan operon (*trpAB)* was detected in the plasticity zone or other genomic regions of *Ca.* C. corallus. Tryptophan is a necessary amino acid for chlamydial growth (39), however, host cell defence mechanisms against chlamydial infections exist in which interferon gamma (IFN-γ) production depletes intracellular tryptophan stores (40). Certain strains of *C. trachomatis* encode for an intact tryptophan operon (*trpAB*), which is absent or incomplete in other chlamydial species (37), suggesting that not all chlamydial species are able to synthesise tryptophan. For urogenital

strains of *C. trachomatis*, the vaginal microbiota is believed to provide indole, allowing for synthesis of tryptophan (39). The absence of a tryptophan operon would suggest that *Ca.* C. corallus either has alternative pathways for synthesising tryptophan or is possibly completely auxotrophic for tryptophan. Notably, neither *Ca.* C. corallus nor Ca. C. sanzinia encode for an aromatic amino acid hydroxylase, which has been suggested to contribute to tryptophan metabolism in the absence of *trpAB*. As has been suggested for *C. trachomatis* (39), the diverse microbiota in these snakes may provide nutrients or substrates for chlamydial synthesis of amino acids. Metagenomic mining revealed several tryptophan synthesis pathway or rescue genes encoded by the bacteria present in the cloaca and choana samples, for example, tryptophan synthase, tryptophanase and indole-3-glycerol phosphatase were detected among the samples, encoded by *Achromobacter sp.*, *Serratia marcescens*. *Clostridium sp., Salmonella enterica* and *Chitinophagaceae*. Previous studies also describe the presence of indole-producing bacteria at these sites (21-23).

5.4 Conclusions

In the current study, we have used culture-independent metagenome analysis to investigate the microbial metagenome of snake choana and cloaca samples. In doing so, we have shown that this method provides a wealth of biological information for novel species discovery through microbial community profiling, and have described the presence of highly abundant bacterial species at these sites, some of which have not previously been described. The animal and public health implications of these findings are unknown, but the repeated observations of human pathogens in the microbiota of snakes (21,22) and other reptiles warrants further investigation.

The metagenomic method used is particularly useful for characterising novel species or strains with no reference genome such as novel uncultivable bacteria (eg. members of the phylum *Chlamydiae*). The complexity within these anatomical sites meant that despite previous detection of chlamydial 16S rRNA sequences by PCR, only a chlamydial plasmid could be detected in all samples, and a single chlamydial chromosome. Nonetheless, comparative analysis of the novel

chlamydial species with other *Chlamydia sp.* revealed genetic differences in regards to purine and pyrimidine metabolism. The detection of chlamydial plasmids in all samples, which was only possible using this method, highlights additional diversity within the *Chlamydiae*, which appears to be a growing trend with increased breadth and depth of sampling and advances in molecular techniques. Further studies, such as metatranscriptomic analysis would better elucidate the complex role of the microbiota on chlamydial pathogenesis and vice versa.

5.5 Materials and methods

5.5.1 Sample preparation

Suspected novel genotypes (n=5) of *C. pneumoniae* were recently detected in collections of captive snakes in Switzerland (8). Clinical swabs were taken from either the cloana or choana of clinically healthy snakes, and DNA was extracted as previously described (8). All samples were subjected to host methylated DNA depletion prior to multiple displacement amplification, as previously described (9).

5.5.2 Ethics approval and consent to participate

The collection and molecular analysis of the snake samples was approved and performed in accordance with the relevant guidelines and regulations of the Veterinary Office of Canton Zurich (authorization no. ZH010/15).

5.5.3 Metagenome assembly and analysis

Deep sequencing was carried out on an Illumina NextSeq at the Australian Genome Research Facility using 150 bp paired-end reads. Read quality was assessed through FastQC v.011.2 and reads were trimmed for adaptors and quality using Trimmomatic v.305 (41). Reads were assembled into contigs using SPAdes v.3.1.1 in metagenome mode with default kmer values (21, 33, 55) (42). Each assembly was assessed through QUAST (43). To obtain chlamydial contigs from the assembled metagenome, contigs were subject to BLAST analysis against an in-house

chlamydial genome database, and subsequent analysis against the NCBI nucleotide database. Contigs with hits against chlamydial sequences were automatically annotated using RAST (44) and manually curated in Artemis (45).

Metaxa was employed initially to assess the species richness within the resulting metagenomes, detecting ribosomal RNA subunits of various origins (46). MaxBin was used to construct partial or complete draft genomes for the microbial species detected in the samples and determine genome completeness for each assembly (47).

Burrows-Wheeler aligner, SAMtools and BEDtools were used to map reads and assess read coverage across the various metagenomic components (48-50).

5.5.4 Phylogenetic analysis

The genetic relationships of the novel species described in this study to other chlamydial species was assessed using the classification system published by Pillonel *et al.,* (5). Individual genes were extracted from the assembled genome and established chlamydial species, including *Simkania negevensis* as an out group. Extracted genes were concatenated and aligned using MAFFT (51) and a phylogenetic tree based on the resulting alignment was constructed using FastTree (52); both were run in Geneious v7.1 (53).

Plasmid phylogeny was performed based on the alignment of nucleotide sequences using MAFFT (51) and tree construction using FastTree (52). In order to include plasmid sequences from all possible species, nucleotide sequences were re-ordered and large gaps were removed so that each resulting plasmid sequence was 5,522-6,170 bp, thus comparable to *C. suis* plasmid which lacks the parA and pgp-6 genes.

5.5.5 Data availability

The metagenomic sequence data obtained for the Ca. Chlamydia corallus chromosome and plasmid was deposited in Genbank under accession numbers as part of bioproject PRJNA312988.

Declarations

Acknowledgements

We thank Simon Rüegg for sample collection and Carmen Kaiser for DNA extraction and initial PCR screening. We thank Helena Seth-Smith for her advice on the preparation of our samples for culture-independent genome sequencing as well as downstream analysis. This work was funded by a University of the Sunshine Coast Faculty Research Grant.

Author contributions

A.T.B managed the project, conducted the laboratory experiments and bioinformatics analysis, and prepared the manuscript; S.S. conducted bioinformatics analysis, prepared the figures and manuscript; N.B. provided samples and reviewed the manuscript; A.P. managed the project and reviewed the manuscript.

Additional Information

The authors declare no competing financial interests.

Table 5.4: *Microbial composition of choana and cloaca samples from captive snakes*

	G1/1679-8	**G2/2464-204**	**G3/2742-324**	**G6/0661-435**	**G7/2741-436**
Host	*Vipera a. ammodytes*	*Vipera latastei*	*Corallus batesii*	*Atheris squamiger*	*Eunectes notaeus*
(Family)	(Viperidae)	(Viperidae)	(Boidae)	(Viperidae)	(Boidae)
Anatomical site	Cloaca	Cloaca	Choana	Choana	Choana
No. partial or complete bacterial genomes detected	4	1	3	0	2
No. 16S rRNA sequences detected	5	1	2	0^	2
No. microbial eukaryotes detected	1	0	0	0	1
Most abundant bacterial taxon	*Serratia*	*Salmonella enterica*	*Chitinophageaceae*		*Chitinophageaceae*
(coverage of 16S rRNA sequence)	*marcescens* (~125x)	(72x)	*bacterium* (~702x)	N.a	*bacterium* (~20x)
(Phylum)	(Proteobacteria)	(Proteobacteria)	(Bacteroidetes)		(Bacteroidetes)

^ Only contigs ≥ 1,000 bp were considered

N.a; not applicable

Table 5.2: *Genome characteristics of Uncultured Chlamydia sp. G3/2741-324 compared to closely-related chlamydial species.*

	Ca. **Chlamydia corallus G3/2742-324**	***Ca.*** **C. sanzinia G4/2742-308 (CP014639)**	***C. pneumoniae*** **LPCoLN (CP006571.1)**	***C. pecorum*** **MC/Marsbar (NZ_CM002310.1)**	***C. trachomatis*** **A/HAR-13 (NC_007429.1)**
Host	Emerald tree boa (*Corallus batesii*)	Madagascar tree boa (*Sanzinia madagascariensis*)	Koala (*Phascolarctos cinereus*)	Koala (*Phascolarctos cinereus*)	Human *(Homo sapiens)*
Chromosome length (bp)	1,196,452	1,113,233	1,241,024	1,230,439	1,044,459
GC content (%)	39.9	38.5	40.6	40.6	41.2
No. CDS	1,076	998	1,095	1,116	954
No. hypothetical proteins	356	314	426	297	294
Plasmid length (bp)	7,621	7,504	7,530^	7,547^	7,510^
No. CDS on plasmid	8	8	8	8	8

^ Not present in all strains

Table 5.3: *Characteristics of chlamydial plasmids obtained from snake choana and cloaca metagenomes*

	G1/1679-8	**G2/2464-204**	**G3/2742-324**	**G6/0661-435**	**G7/2741-436**
No. contigs	1	1	1	4	1
Length (bp)	7,534	7,518	7,522	7,210[1]	7,522
Mean read coverage	~3,243x	~389x	~40,132x	~5x	~14,471x
No. ORFs	8	8	8	8	8
BLAST hit (% nucleotide ID)	*C. pneumoniae* LPCoLN plasmid (81%)	*C. pneumoniae* LPCoLN plasmid (88%)	*C. pneumoniae* LPCoLN plasmid (88%)	*C. pneumoniae* LPCoLN plasmid (86%)	*C. pneumoniae* LPCoLN plasmid (87%)
ArrayTube result	*Chlamydia sp.*	Inconclusive	*C. pneumoniae*	*C. pneumoniae*	*C. pneumoniae*

[1] Possibly incomplete sequence; truncated helicase gene predicted over ends of contig.

[2] Nucleotide identity (%) of near-full length 16S rRNA sequence.

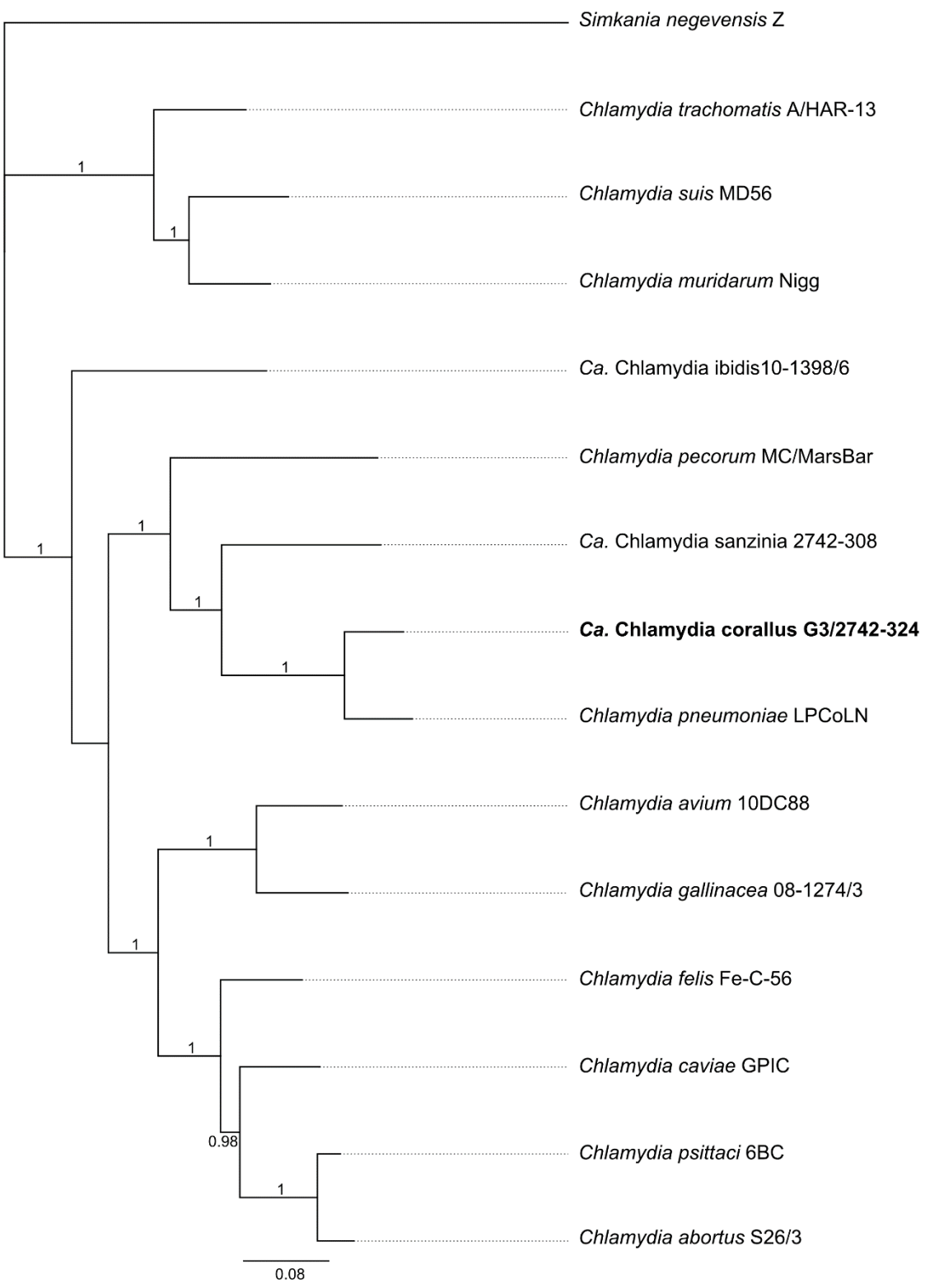


Figure 5.1: *Phylogenetic position of the novel taxon, Ca. Chlamydia corallus G3/2742-324 within the Chlamydiaceae. Phylogenetically informative markers genes were retrieved from each genome, concatenated and aligned using MAFFT, prior to tree construction using FastTree. Numbers on the branches indicate support values.*

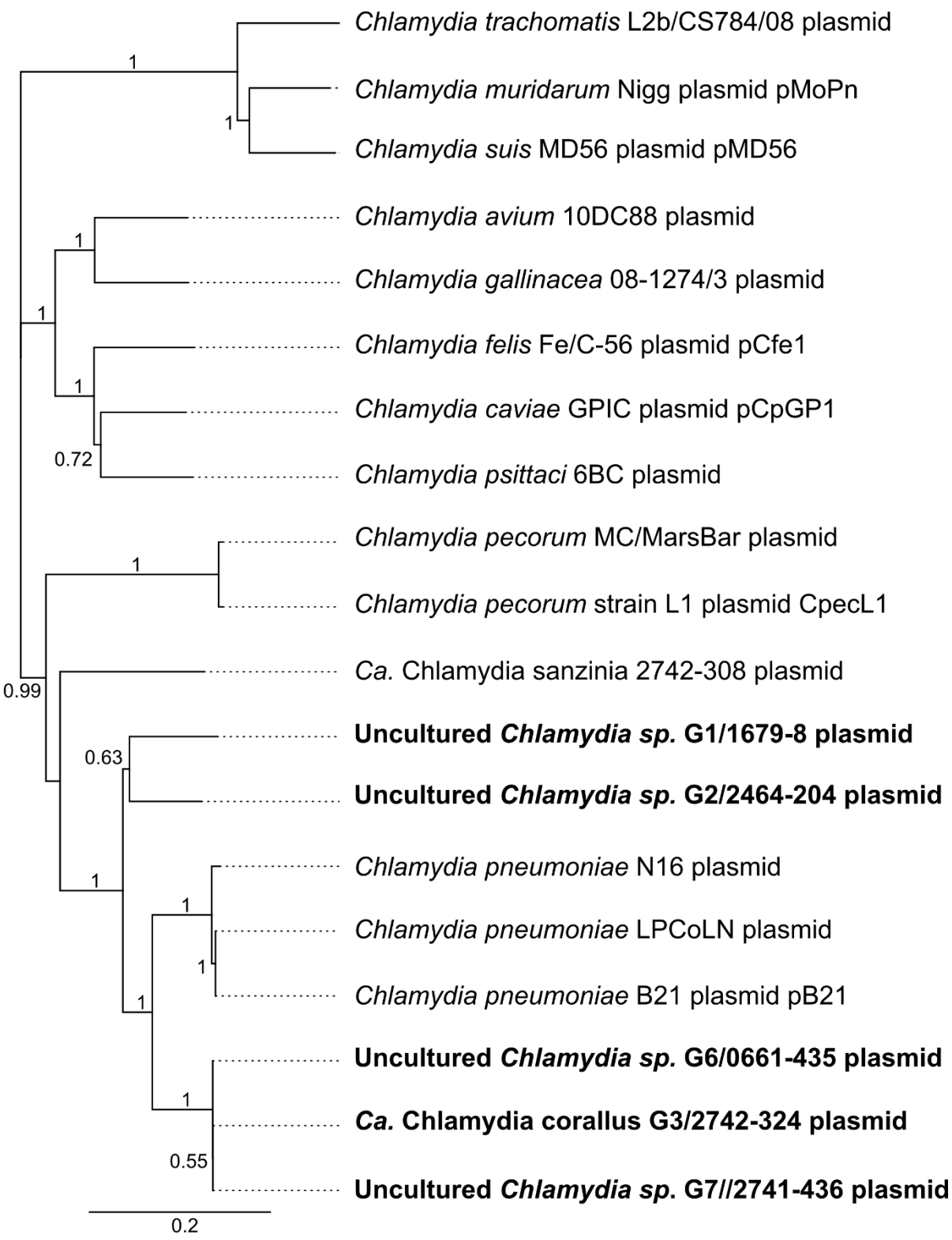


Figure 5.2: *Plasmid sequence-based phylogenetic analysis of all samples sequenced in this study and other chlamydial plasmid sequences. Nucleotide sequences were aligned using MAFFT and tree was constructed using FastTree. Support values are shown on the branches.*

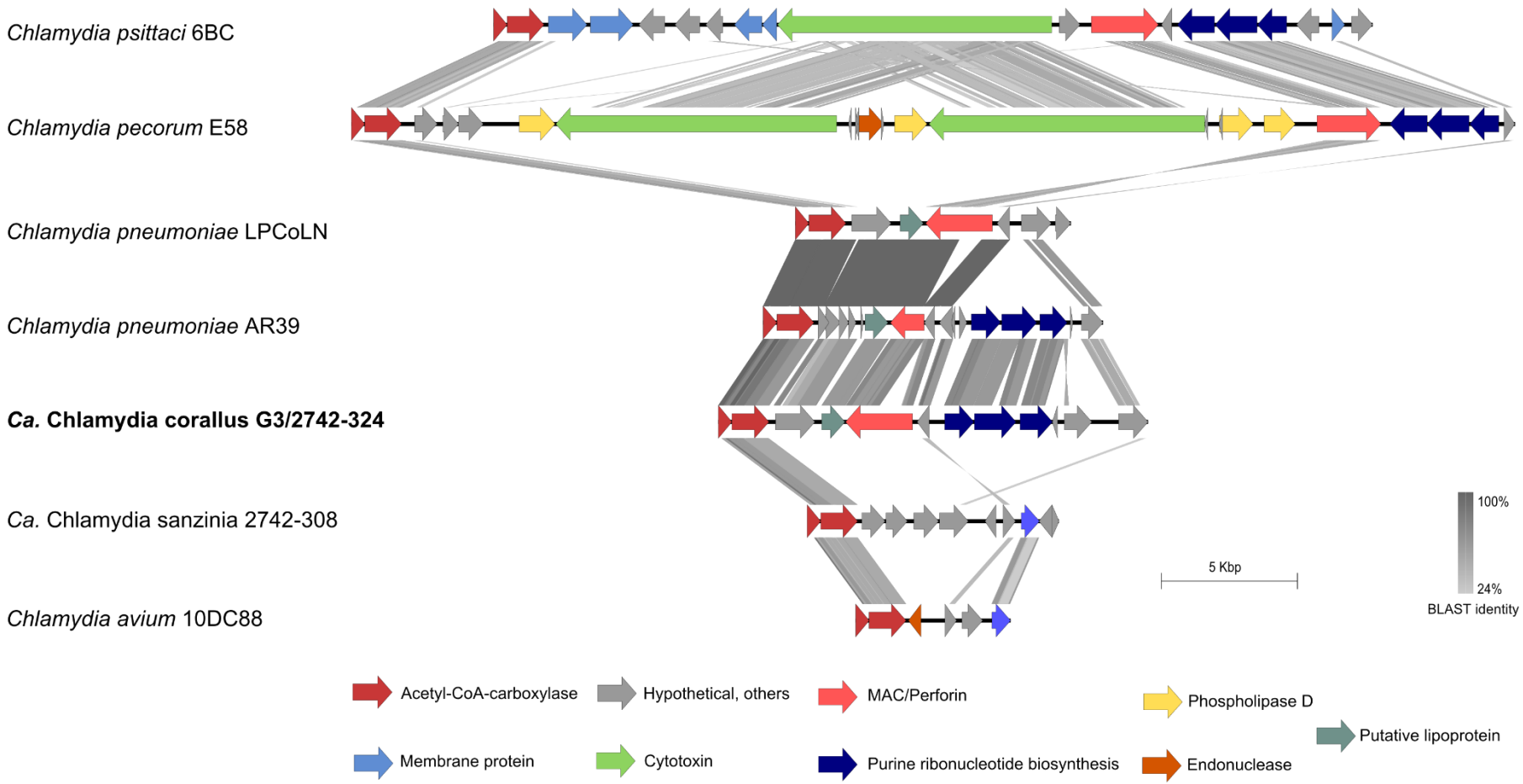


Figure 5.3: *Comparison of plasticity zone regions encoded by Ca. Chlamydia corallus and related chlamydial species. Figure constructed using EasyFig (50). Grey shading represents tBLASTx matches (see BLAST identity scale). Coloured arrows represent coding regions (see legend).*

References

1. Miller R.R., Montoya V., Gardy J.L., Patrick D.M., Tang P. Metagenomics for pathogen detection in public health. *Genome Med* **5**:81, gm485; 10.1186/gm485 (2013).

2. Lu X., Zhang X.X., Wang Z., Huang K., Wang Y., *et al.* Bacterial pathogens and community composition in advanced sewage treatment systems revealed by metagenomics analysis based on high-throughput sequencing. *PLoS One* **10**, 0125549; 0125549; 10.1371/journal.pone.0125549 (2015).

3. Nieuwenhuijse D.F., Koopmans M.P. Metagenomic Sequencing for Surveillance of Food- and Waterborne Viral Diseases. *Front Microbiol* **8**, 230; 10.3389/fmicb.2017.00230 (2017).

4. Taylor-Brown, A., Polkinghorne, A. New and emerging chlamydial infections of creatures great and small, in press. *New Microbes New Infect* 10.1016/j.nmni.2017.04.004 (2017).

5. Pillonel T., Bertelli C., Salamin N., Greub G. Taxogenomics of the order *Chlamydiales*. I*nt J Syst Evol Microbiol* **65**, 1381-1393 (2015).

6. Bachmann N.L., Polkinghorne A., Timms P. *Chlamydia* genomics: providing novel insights into chlamydial biology. *Trends Microbiol* **22**, 464-472. (2014).

7. Rajendhran J., Gunasekaran P. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res* **166,** 99-110 (2011).

8. Taylor-Brown A., Bachmann N.L., Borel N., Polkinghorne A. Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes. *BMC Genomics* **17**, 710; 10.1186/s12864-016-3055-x (2016).

9. Taylor-Brown A., Rüegg S., Polkinghorne A., Borel N. Characterisation of *Chlamydia pneumoniae* and other novel chlamydial infections in captive snakes. *Vet Microbiol* **178**, 88-93 (2015).

10. Bodetti T.J., Jacobson E., Wan C., Hafner L., Pospischil A., *et al.* Molecular Evidence to Support the Expansion of the Hostrange of *Chlamydophila pneumoniae* to Include Reptiles as Well as Humans, Horses, Koalas and Amphibians. *Syst Appl Microbiol* **25**, 146-152 (2002).

11. Jacobson E.R., Heard D., Andersen A. Identification of *Chlamydophila pneumoniae* in an Emerald Tree Boa, *Corallus caninus*. *J Vet Diagn Invest* **16**, 153-154 (2004).

12. Soldati G., Lu Z.H., Vaughan L., Polkinghorne A., Zimmermann D.R., *et al.* Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. *Vet Pathol* **41**, 388-397 (2004).

13. Huchzermeyer F.W., Langelet E., Putterill J.F. An outbreak of chlamydiosis in farmed Indopacific crocodiles (*Crocodylus porosus*). *J S Afr Vet Assoc* **79**, 99-100. (2008)

14. Frutos M.C., Monetti M.S., Re V.E., Cuffini C.G. Molecular evidence of *Chlamydophila pneumoniae* infection in reptiles in Argentina. *Rev Argent Microbiol* **46**, 45-48 (2014).

15. Rüegg S.R., Regenscheit N., Origgi F.C., Kaiser C., Borel N. Detection of *Chlamydia pneumoniae* in a collection of captive snakes and response to treatment with marbofloxacin. *Vet J* **205**, 424-426. (2015).

16. Shilton C.M., Jerrett I.V., Davis S., Walsh S., Benedict S., *et al.* Diagnostic investigation of new disease syndromes in farmed Australian saltwater crocodiles (*Crocodylus porosus*) reveals associations with herpesviral infection. *J Vet Diagn Invest* **28**, 279-290 (2016).

17. Hanning I, Diaz-Sanchez S. 2015. The functionality of the gastrointestinal microbiome in non-human animals. *Microbiome* **3**, 51; 10.1186/s40168-015-0113-6 (2015).

18. Yoon S.S., Kim E.K., Lee W.J. Functional genomic and metagenomic approaches to understanding gut microbiota-animal mutualism. *Curr Opin Microbiol* **24**, 38-46. (2015).

19. Costello EK, Gordon JI, Secor SM, Knight R. Postprandial remodeling of the gut microbiota in Burmese pythons. *ISME J* **4**, 1375-1385. (2010).

20. Keenan S.W., Engel A.S., Elsey R.M. The alligator gut microbiome and implications for archosaur symbioses. *Sci Rep* **3**, 2877; 10.1038/srep02877 (2013).

21. Schmidt V., Marschang R.E., Abbas M.D., Ball I., Szabo I., *et al.* Detection of pathogens in Boidae and Pythonidae with and without respiratory disease. *Vet Rec* **172**, 236; 10.1136/vr.100972 (2013).

22. Plenz B., Schmidt V., Grosse-Herrenthey A., Kruger M., Pees M. Characterisation of the aerobic bacterial flora of boid snakes: application of MALDI-TOF mass spectrometry. *Vet Rec* **176**, 285; 10.1136/vr.102580 (2015).

23. McLaughlin R.W., Cochran P.A., Dowd S.E. Metagenomic analysis of the gut microbiota of the Timber Rattlesnake, *Crotalus horridus*. *Mol Biol Rep* **42**, 1187-1195 (2015).

24. Hong E.E., Okitsu C.Y., Smith A.D., Hsieh C.L. Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial DNA. *Mol Cell Biol* **33**, 2683-2690 (2013).

25. Pickett M.A., Everson J.S., Pead P.J., Clarke I.N. The plasmids of *Chlamydia trachomatis* and *Chlamydophila pneumoniae* (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents. *Microbiology* **151**, 893-903 (2005).

26. Jelocnik M., Bachmann N.L., Seth-Smith H., Thomson N.R., Timms P., *et al.* Molecular characterisation of the *Chlamydia pecorum* plasmid from porcine, ovine, bovine, and koala strains indicates plasmid-strain co-evolution. *PeerJ* **4**, e1661; 10.7717/peerj.1661 (2016).

27. Zhong G. Chlamydial Plasmid-Dependent Pathogenicity. *Trends Microbiol* **25**, 141-152 (2016).

28. Chen J., Yang Z., Sun X., Tang L., Ding Y., *et al.* Intrauterine infection with plasmid-free *Chlamydia muridarum* reveals a critical role of the plasmid in chlamydial ascension and establishes a model for evaluating plasmid-independent pathogenicity. *Infect Immun* **83**, 2583-2592 2015).

29. Song L., Carlson J.H., Whitmire W.M., Kari L., Virtaneva K., *et al. Chlamydia trachomatis* Plasmid-Encoded Pgp4 Is a Transcriptional Regulator of Virulence-Associated Genes. *Infect Immun* **81**, 636-644 (2013).

30. Seth-Smith H.M., Harris S.R., Persson K., Marsh P., Barron A., *et al.* Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. *BMC Genomics* **10**, 239; 10.1186/1471-2164-10-239 (2009).

31. Roulis E., Bachmann N.L., Myers G.S., Huston W., Summersgill J., *et al.* Comparative genomic analysis of human *Chlamydia pneumoniae* isolates from respiratory, brain and cardiac tissues. *Genomics* **106**, 373-383 (2015).

32. Jelocnik M., Bachmann N.L., Kaltenboeck B., Waugh C., Woolford L., *et al.* Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas. *BMC Genomics* **16**, 893; 10.1186/s12864-015-2053-8 . (2015).

33. Voigt A., Schöfl G., Saluz H.P. The *Chlamydia psittaci* Genome: A Comparative Analysis of Intracellular Pathogens. *PLoS ONE* **7**, e35097; 10.1371/journal.pone.0035097 (2012).

34. Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., *et al.* Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res* **28**, 397-1406. (2000).

35. Taylor L.D., Nelson D.E., Dorward D.W., Whitmire W.M., Caldwell H.D. Biological Characterization of *Chlamydia trachomatis* Plasticity Zone MACPF Domain Family Protein CT153. *Infect Immun* **78**, 2691-2699 (2010).

36. Liechti G., Goldberg J.B. *Helicobacter pylori* Relies Primarily on the Purine Salvage Pathway for Purine Nucleotide Biosynthesis. *J Bacteriol* **194**, 839-854 (2012).

37. Mitchell C.M., Hovis K.M., Bavoil P.M., Myers G.S., Carrasco J.A., *et al.* Comparison of koala LPCoLN and human strains of *Chlamydia pneumoniae* highlights extended genetic diversity in the species. *BMC Genomics* **11**, 1-10 (2010).

38. Gil R., Silva F.J., Peretó J., Moya A. Determination of the Core of a Minimal Bacterial Gene Set. *Microbiol Mol Biol Rev* **68**, 518-537 (2004).

39. Ziklo N., Huston W.M., Hocking J.S., Timms P. *Chlamydia trachomatis* Genital Tract Infections: When Host Immune Response and the Microbiome Collide. *Trends Microbiol* **24**, 750-765 (2016).

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

41. Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114-2120 (2014).

42. Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477 (2012).

43. Gurevich A., Saveliev V., Vyahhi N., Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072-1075 (2013).

44. Aziz R.K., Bartels D., Best A.A., DeJongh M., Disz T., *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9,** 75; 10.1186/1471-2164-9-75 (2008).

45. Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16,** 944-945 (2000).

46. Bengtsson J., Eriksson K.M., Hartmann M., Wang Z., Shenoy B.D., *et al.* Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit

(12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie van Leeuwenhoek* **100** (2011).

47. Wu Y-W., Tang Y-H., Tringe S.G., Simmons B.A., Singer S.W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 1-18 (2014).

48. Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).

49. Quinlan A.R., Hall I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

50. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

51. Katoh K., Misawa K., Kuma K.Ä., Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30,** 3059-3066 (2002).

52. Price M.N., Dehal P.S., Arkin A.P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Genome Biol Evol* **26**, 1641-1650 (2009).

53. Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).

54. Sullivan M.J., Petty N.K., Beatson S.A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009-1010 (2011).

CHAPTER 6: Culture-independent genomics of a novel chlamydial pathogen of fish provides new insight into host specific adaptations utilized by these intracellular bacteria (Environmental Microbiology 19, 2017)

Running title: Novel chlamydial epitheliocystis agent in Grouper

Alyce Taylor-Brown[1], Trestan Pillonel[2], Andrew Bridle[3], Weihong Qi[4], Nathan Bachmann[1#a], Terrence L. Miller[5#], Gilbert Greub[2], Barbara Nowak[3], Helena M.B. Seth-Smith[4,6#], Lloyd Vaughan[6], Adam Polkinghorne[1]*

[1] Centre for Animal Health Innovation, Faculty of Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland, 4556, Australia

[2] Institute of Microbiology, University of Lausanne, CH-1011 Lausanne, Switzerland

[3] Institute of Marine and Antarctic Studies, University of Tasmania, Newnham, Tasmania, 7248, Australia

[4] Functional Genomics Centre, University of Zurich, CH-8057 Zurich, Switzerland

[5] Centre for Sustainable Tropical Fisheries and Aquaculture, College of Marine and Environmental Sciences, James Cook University, Cairns, Queensland 4878, Australia

[6] Institute of Veterinary Pathology, University of Zurich, CH-8057 Zurich, Switzerland

* Correspondence: Adam Polkinghorne; Centre for Animal Health Innovation, Faculty of Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland, 4556, Australia; apolking@usc.edu.au, +617 5456 5578

6.1 Abstract

Several *Chlamydiales* families are associated with epitheliocystis, a common condition of the fish gill epithelium. These families share common ancestors with the *Chlamydiaceae* and environmental *Chlamydiae*. Due to the lack of culture systems, little is known about the biology of these chlamydial fish pathogens.

We investigated epitheliocystis in cultured Orange-spotted grouper (*Epinephelus coioides*) from North Queensland, Australia. Basophilic inclusions were present in the gills of 22/31 fish and the presence of the chlamydial pathogen in the cysts was confirmed by *in situ* hybridisation. Giant grouper (*Epinephelus lanceolatus*) cultured in the same systems were epitheliocystis free. 16S rRNA gene sequencing revealed a novel member of the *Candidatus* Parilichlamydiaceae: *Ca.* Similichlamydia epinephelii.

Using metagenomic approaches, we obtained an estimated 68% of the chlamydial genome, revealing that this novel chlamydial pathogen shares a number of key pathogenic hallmarks with the *Chlamydiaceae,* including an intact Type III Secretion system and several chlamydial virulence factors. This provides additional evidence that these pathogenic mechanisms were acquired early in the evolution of this unique bacterial phylum.

The identification and genomic characterisation of *Ca.* S. epinephelii provides new opportunities to study the biology of distantly-related chlamydial pathogens while shining a new light on the evolution of pathogenicity of the *Chlamydiaceae*.

6.2 Introduction

The *Chlamydiae* are a phylum of obligate intracellular bacteria ubiquitous in the environment and associated with disease in a wide array of hosts. While our knowledge of the *Chlamydiaceae* family of human and veterinary pathogens is well established, the phylum has expanded drastically over the last twenty years (Taylor-Brown *et al.*, 2015). These more recently described families are collectively referred to as "*Chlamydia*-related bacterias" (CRBs) or environmental *Chlamydiae*; the latter referring to the fact a large number of these species have been detected in or isolated from environmental sources such as water (Thomas *et al.*, 2006; Lienard *et al.*, 2011), free-living amoebae (Amann *et al.*, 1997; Everett *et al.*, 1999; Fritsche *et al.*, 2000) and insects (Kostanjsek *et al.*, 2004; Everett *et al.*, 2005). Metagenomic studies suggest that chlamydial sequences deposited in public databases are actually predominantly recovered from aquatic environments (Lagkouvardos *et al.*, 2014). These novel species are deeply rooted within the chlamydial phylogenetic tree, representing a wealth of untapped diversity and biological significance.

Although there is an increasing body of evidence, the pathogenic potential of many CRBs is still unclear, certainly in comparison to the "traditional *Chlamydiae*", ie. the *Chlamydiaceae* family. Molecular and sero- epidemiological surveys have detected CRBs in association with respiratory disease in humans (Kahane *et al.*, 1998; Friedman *et al.*, 1999) and livestock (Wheelhouse *et al.*, 2013) and reproductive disease in both humans and animals (Borel *et al.*, 2007; Baud *et al.*, 2011; Barkallah *et al.*, 2014). Further, murine respiratory disease models have recently been established for *Waddlia chondrophila* (Pilloux *et al.*, 2016) and *Parachlamydia acanthamoebae* (Casson *et al.*, 2008), as has a bovine model of *P. acanthamoebae* respiratory disease (Lohr *et al.*, 2014) and a murine model of *W. chondrophila* genital infection (Vasilevsky *et al.*, 2015), confirming Koch's postulates for these causative agents of disease. Questions nevertheless remain about whether these organisms are truly pathogens of their varied host species.

Perhaps the strongest evidence for the pathogenic potential of CRBs lies in organisms from the deepest branches of chlamydial phylogeny, occurring in a wide range of taxonomically and geographically diverse fish hosts and associated with the gill disease, epitheliocystis (Stride *et al.*, 2014). Originally described in salmonids of the genus *Salmo* in the Northern hemisphere, the *Candidatus* family Piscichlamydiaceae (Draghi *et al.*, 2004) signalled a new and deep rooted clade of the phylum *Chlamydiae*. Description of a closely related *Candidatus* family, Parilichlamydiaceae in Yellowtail kingfish (*Seriola lalandi*), followed. Members of this clade have been detected in African (Steigen *et al.*, 2013), Australian (Stride *et al.*, 2013b; Stride *et al.*, 2013a), Scandinavian (Steigen *et al.*, 2015), Mediterranean (Seth-Smith *et al.*, 2016) and Swiss waters (Guevara Soto *et al.*, 2016c). Additional chlamydial epitheliocystis agents include two members of the family *Simkaniaceae* (Fehr *et al.*, 2013; Nylund *et al.*, 2015) and the closest marine relative to the *Chlamydiaceae*, Ca. Clavichlamydia salmonis (Karlsen *et al.*, 2008), the latter sharing many features common to their closest land-based relatives (Karlsen *et al.*, 2008; Schmidt-Posthaus *et al.*, 2012). The geographical and host species diversity of the *Ca.* Parilichlamydiaceae clade is truly impressive and raises considerable interest on the genomic background for this feat.

Despite our increased understanding of these epitheliocystis agents, in no case have Koch's postulates been fulfilled, as all attempts to culture the agent(s) *in vitro* have proved unsuccessful (Hoffman *et al.*, 1969; Bradley *et al.*, 1988; Chen *et al.*, 2003; Stride *et al.*, 2014; Katharios *et al.*, 2015; Seth-Smith *et al.*, 2016). Researchers have instead relied on fulfilling Fredericks and Relman's postulates (Fredricks and Relman, 1996). Given the diversity in the causal agents, it is not surprising that the disease manifestations also vary, ranging from little or no host tissue reaction to a moderate host response and epithelial cell proliferation. The physiological responses are varied and include laboured/rapid respiration, lethargy, flared opercula and weak swimming behaviour, with white to yellow nodules often visible on the gill epithelium. Epitheliocystis is of significant interest and increasing burden in the aquaculture industry, with high mortality rates in some

populations and substantial economic losses (Bradley *et al.*, 1988; Draghi *et al.*, 2004; Katharios *et al.*, 2008; Stride *et al.*, 2014).

Orange-spotted grouper (*Epinephelus coioides*) and Giant grouper (*E. lanceolatus*) are emerging tropical aquaculture species, with the value of individual fish ranging up to US$169 per kilo for live Giant grouper (Sadovy and Vincent, 2002). Unfortunately, grouper are susceptible to a number of infectious diseases of bacterial, viral, and parasitic origins (Nagasawa, 2004; Harikrishnan *et al.*, 2011; Lafferty *et al.*, 2015). Intensification of farming appears to exacerbate disease risk; high stocking densities in particular are proposed to increase disease susceptibility and mortality due to disease is a major production constraint (Nagasawa, 2004; Nowak and LaPatra, 2006) (Nagasawa, 2004; Harikrishnan *et al.*, 2011). Epitheliocystis has until now not been reported in any species of Serranidae/Epinephelinae.

We aimed to characterise a putative novel chlamydial epitheliocystis agent detected in Orange-spotted grouper in North Queensland, Australia. Given the as yet uncultivated nature of these pathogens, and with recent advances in culture-independent and metagenomic methods (Seth-Smith *et al.*, 2013; Bachmann *et al.*, 2015; Katharios *et al.*, 2015; Qi *et al.*, 2016; Seth-Smith *et al.*, 2016; Taylor-Brown *et al.*, 2016), we utilised genomic data to investigate the pathogenic potential of this novel uncultured epitheliocystis agent.

6.3 Results and discussion

6.3.1 Detection of epitheliocystis in cultured Epinephelus coioides

Three different novel chlamydial species have been described in association with epitheliocystis from farmed Australian fish species. *Ca.* Parilichlamydia carangidicola was detected in five Yellowtail Kingfish cohorts between 2002 and 2014 (Stride *et al.*, 2013c) in association with differing levels of disease severity and infection rates. *Ca.* Similichlamydia latridicola was described in wild and farmed Striped trumpeter (*Latris lineata*) (Stride *et al.*, 2013b) and *Ca.* Similichlamydia laticola in seven cohorts of Barramundi (*Lates calcarifer*) (Stride *et al.*, 2013a). To date, genomics has

been carried out on beta- and gamma-Proteobacterial epitheliocystis agents (Katharios *et al.*, 2015; Qi *et al.*, 2016; Seth-Smith *et al.*, 2016), but not on chlamydial fish pathogens.

In the current study, an epidemiological survey was carried out over 12 months to assess the disease dynamics throughout ten groups of farmed *E. coioides* and *E. lanceolatus* in Queensland, Australia following the observation of putative epitheliocystis in *E. coioides* upon examination for gill flukes. Table 6.1 details the cohorts sampled in this study: 53 fish were sampled from 10 cohorts (31 *E. coioides* and 22 *E. lanceolatus*) between March 2014 and February 2015. All fish sampled were juveniles except for the broodstock samples from Cairns (EC-May-14).

Histopathology showed that disease prevalence in *E. coioides* increased from initial sampling in March 2014 (Autumn) (83.33%; n=6) to October 2014 (Spring) when it peaked at 100% of all gills sampled (n=6). Overall, 22 out of 31 *E. coioides* sampled were found to have epitheliocysts following histopathological examination. Affected fish had an average of up to 2.1 cysts per gill filament from four out of the six cohorts of *E. coioides* (Table 6.1). Interestingly, by February 2015 (Summer), there was no histological evidence of epitheliocystis, although only three fish were sampled at this time point. Some studies of epitheliocystis have suggested that cyclic changes in temperature contribute to disease severity and prevalence, with greater epitheliocystis prevalence observed in warmer months (Nowak and Clark, 1999) which may more be a reflection of seasonal changes rather than water temperatures per se (Guevara Soto *et al.*, 2016b). Groupers have been shown to produce innate and adaptive immune activation in response to infection with other intracellular pathogens (Huang *et al.*, 2011; Wei *et al.*, 2012; Lai *et al.*, 2014). Three *E. coioides* broodstock (EC-May-14) sampled at a different facility were negative for epitheliocystis, suggesting that the bacterium could be introduced to the ponds or JCU facility via seawater exchange (see methods) or that there is age-dependent effect on disease prevalence and severity.

Interestingly, the *E. lanceolatus* that were sampled from the same ponds showed no clinical or histopathological evidence of epitheliocystis (Table 6.2), suggesting either that the agent is species-specific or that *E. lanceolatus* may be less susceptible to bacterial infection in general than

E. coioides. This phenomenon was also observed in a mass mortality event on a Brazilian fish farm, where only the pacus (*Piaractus mesopotamicus*) in a mixed pond succumbed to epitheliocystis (Szakolczai *et al.*, 1999). It is also unclear what role, if any, co-infections with other epitheliocystis agents (Schmidt-Posthaus *et al.*, 2012; Guevara Soto *et al.*, 2016a; Guevara Soto *et al.*, 2016c; Seth-Smith *et al.*, 2016) or parasites, such as ciliates (Fehr *et al.*, 2013), amoebae (Clark and Nowak, 1999), or gill flukes (this study), may have on the dynamics of chlamydial epitheliocystis infection.

6.3.2 Morphological description of Ca. Similichlamydia epinephelii cysts in E. coioides

Staining of infected gill sections with haemotoxylin and eosin showed round to oblong basophilic inclusions ranging from 10 μm to 100 μm in diameter (Figure 6.1a, b). Cysts were spread sporadically throughout the filaments and not confined to a particular location in the lamellae as is often described (Figure 6.1b), with the number of cysts ranging from 0 to 35 per filament in the positive *E. coioides* cohorts. While most cysts were enclosed in hyaline capsules, some cysts were walled off by eosinophilic capsules (Figure 6.2b), presumably derived from layers of epithelial cells. In several cases, two cysts were observed occupying the same interlamellar space. The cyst morphology observed in *E. coioides* gills most resemble epitheliocystis seen in two other Australian fish species, Striped trumpeter and Barramundi, caused by *Ca.* Similichlamydia laticola and *Ca.* S. latridicola (Lai *et al.*, 2013; Stride *et al.*, 2013b; Stride *et al.*, 2013a), and epitheliocystis lesions caused by *Ca.* Similichlamydia sp. in Ballan wrasse from Norway (Steigen *et al.*, 2015).

Host response was variable as shown by the "walling off" of the cysts and large variation in the hyperplastic response of the epithelium (Figure 6.1a, b). There was limited evidence of host response such as leucocyte infiltration and epithelial hyperplasia in this population, and lamellar fusion was uncommon.

6.3.3 Identification of the chlamydial agent of epitheliocystis in E. coioides gill samples by PCR and ISH

To identify the aetiological agent of the epitheliocystis detected in the *E. coioides* gill samples, *Chlamydiales*-specific PCR targeting the signature sequence of the 16S rRNA gene was

performed on four samples from the EC-May-Jul-14 cohort, which were positive for *Chlamydiales* DNA. Direct sequencing of one amplicon and subsequent BLAST analysis revealed 95-96% nucleotide identity to uncultured *Chlamydiales* detected in Wrasse species (Accession numbers KC469567-469570). Screening for other epitheliocystis agents was not performed in this investigation.

Subsequently, sequencing of a ~1,400 bp 16S rRNA gene amplicon from two of these PCR positive samples confirmed these *Chlamydiae* to be a novel species in the *Ca.* Similichlamydia genus. The sequences obtained from grouper gills are 99.6% identical to each other (ie. 4 SNPs difference) and share the highest nucleotide identity with *Candidatus* species in the *Ca.* Similichlamydia genus; *Ca.* S. laticola (95.4-96%), *Ca.* S. latridicola (95.8-96.1%), *Ca.* S. labri (95.8-95.9%) and other unassigned similichlamydial sequences (95.3-96.2%). The phylogenetic tree clearly demonstrates the close relationship between the sequences obtained from grouper gills and members of the *Ca.* Parilichlamydiaceae family.

Following our demonstration of PCR positivity for *Chlamydiae*, we targeted a pan-*Chlamydiales* anti-sense probe against the 16S rRNA gene to detect DNA or mRNA associated with the cysts by *in situ* hybridisation. Using this probe, hybridisation signals localised to all cysts (Figure 6.1d), while no hybridisation was seen using the non-sense probe (scrambled sequence) on a serial section from the same gill arch (Figure 6.1e). As a positive control, the anti-sense probe was also successfully hybridised to cysts in an archival Yellowtail kingfish epitheliocystis gill section (Stride *et al.*, 2014) under the same conditions (data not shown). Lighter stained signals (Figure 6.1d) likely come from cysts that contain fewer bacteria than the darkly stained cysts, possibly reflecting different stages of the cyst development.

6.3.4 Proposal of sp. nov. Ca. Similichlamydia epinephelii

Based on the above findings, we propose that the causative agent of epitheliocystis described in the Orange-spotted grouper belongs to a novel species in the family

Parilichlamydiaceae and hence, assign the name, *Candidatus* Similichlamydia epinephelii (L. mas. n., *Epinepehelus*, the genus to which the fish host belongs). The 16S rRNA gene of two samples of *Ca.* S. epinephelii has been sequenced to date, and the sequences are 99.7% identical to each other, and only 95.4% to 96.1% similar to characterised *Ca.* Similichlamydia *spp* (accession numbers KX880946 and KX880947).

The novel bacteria is associated with epitheliocystis disease, evidenced by hybridisation of a *Chlamydiales* 16S rRNA probe (probe sequence: 5'DIG-ATGTAYTACTAACCCTTCCGCCACTA-3'DIG) to the cysts in the gill epithelium. Pathogenicity of this species is also evidenced by virulence factors present in the draft genome. The chlamydial inclusions are reminiscent of other species in the *Ca.* Parilichlamydiaceae family. They appear round to oblong to irregular, with a basophilic, granular appearance under light microscopy.

The genome has also been sequenced using metagenomics and the contigs deposited under Bioproject PRJNA343727.

6.3.5 Genomic analysis of Ca. Similichlamydia epinephelii

Virtually no information exists on the biology of CRBs belonging to the family *Ca.* Parilichlamydiaceae, largely due to the absence of a culture system for members of these distantly related and recently described *Chlamydiae*. In an effort to provide the first insight into the biology of *Ca.* Similichlamydia, as a representative member of the *Ca.* Parilichlamydiaceae, we performed a metagenomics analysis of a *Ca.* Simichlamydia epinephelii PCR positive tissue sample from an infected *E. coioides* described in this study.

5,561,445 paired reads were obtained from a methylated DNA-depleted DNA extract following shotgun sequencing on an Illumina MiSeq. As no reference genome was available for this putative novel species, reads were trimmed prior to *de novo* assembly into 78,820 contigs. A series of filtering steps were applied to the assembled contigs (see Experimental procedures). 169 contigs harbouring at least one gene with greater sequence identity to chlamydial genes than other

bacterial species were retained. This resulted in a partial draft genome of 981,542 nucleotides, estimated to represent approximately 68.2% of the *Ca.* S. epinephelii genome, demonstrated by the presence of 73 of 107 conserved marker genes (Wu *et al.*, 2014). However, upon manual examination, we detected an additional 21 genes with partial sequences, bringing the total to 94. Taking these estimates, and based on similar coding density and the assumption that these markers are distributed evenly throughout the genome (Wu *et al.*, 2014), we predict that the full *Ca.* S. epinephelii genome would be approximately 1,117,929 - 1,439,211 bp, and encode 1037-1280 genes. The obvious disadvantage to this homology-based method is that contigs containing only novel proteins would be discarded due to a lack of sequence similarity to previously characterised chlamydial proteins. However, given the complexity of the microbial community in the gill, owing to its constant exposure to the aquatic environment, coupled with a similar G+C content of the pathogen and host genome (Han and Zhao, 2008), this method provided the most confidence in eliminating false positive contigs. In the absence of a culture system, it is anticipated that future sequencing efforts for other members of this family will confirm genome completeness and the presence of family-, genus- and species-specific proteins.

The G+C content of this partial genome is 39.65% which is comparable to that of the *Chlamydiales*, and in particular, the CRBs (Horn et al., 2004; Greub *et al.*, 2009; Bertelli *et al.*, 2010; Collingro *et al.*, 2011; Bertelli *et al.*, 2014; Bertelli *et al.*, 2015). Automated annotation using RAST coupled with manual annotation in Artemis resulted in the prediction of 943 coding sequences (CDSs), 300 of these being hypothetical proteins. These metrics are comparable to the genomes of members of the *Chlamydiales*, which range from 1.0-3.1 Mbp in length, with the number of coding regions being proportional to genome size (Taylor-Brown *et al.*, 2015). The predicted coding density is ~90%, reflecting the compact nature of the genome and indicative of their obligate intracellular lifestyle. The genome statistics mentioned above are summarised in Table 5.3.

Comparison of coding regions revealed that *Ca.* S. epinephelii shares most homologs (n=395) with *Protochlamydia sp.* C2 (Figure 6.2a), but a comparable number are shared with other members of the *Parachlamydiaceae*. A smaller subset of homologs are shared with the *Chlamydiaceae* (n=331-350). One notable set of genes shared with the latter include the presence of hallmark regulatory genes of the chlamydial developmental cycle such as the early upstream open reading frame *(euo*). While we were unable to characterise the developmental cycle of this novel pathogen using microscopic methods, this suggests that this pathogen may share similar developmental stages to species in the *Chlamydiaceae* and other chlamydial families.

312 orthologous groups were found to be common among *Ca.* Parilichlamydiaceae, *Simkaniaceae*, *Chlamydiaceae*, *Parachlamydiaceae* and *Waddliaceae* (Figure 6.2b) by cluster analysis. This is a much smaller number than reported in previous pan-genome studies (Collingro *et al.*, 2011; Gupta *et al.*, 2015; Pillonel *et al.*, 2015), but reflects the evolutionary distance between this species and other members of the phylum. Given the draft nature of the *Ca.* S. epinephelii genome, it is likely that this gene set is an underestimate and could be considered the preliminary "draft core genome" of the *Chlamydiales*. Among these are genes involved in DNA replication and repair, transcription and translation, general secretion pathway proteins and some transporter proteins. Conversely, 239 proteins are unique to *Ca.* S. epinephelii when compared to these groups. This mostly comprises hypothetical proteins and is again comparably smaller than the number of unique genes encoded by other CRBs, yet higher than the *Chlamydiaceae*, and suggests that we have not yet captured the true level of uniqueness of this distant clade.

6.3.6 Updated phylogenomics of Ca. S. epinephelii within the order Chlamydiales

We employed sixty-four single copy orthologs conserved among the *Planctomycetes/ Verrucomicrobia/ Chlamydiales* superphylum to reconstruct the phylogeny of the order *Chlamydiales*. Phylogeny obtained on the basis of the concatenated amino acid sequences confirmed *Ca.* S

epinephelii to be the most distantly related species to chlamydial species whose genomes are available (Figure 6.3).

The concatenated amino acid tree topology is analogous to the 16S rRNA gene phylogeny and is congruent with genome-wide trees depicted in recent pan-*Chlamydiales* genome studies (Collingro *et al.*, 2011; Psomopoulos *et al.*, 2012; Gupta *et al.*, 2015; Pillonel *et al.*, 2015) with the addition of this newly described taxon. Further, the phylogenetic placement based on genomic information supports a third clade with the *Chlamydiales*, as proposed by Gupta *et al.* (Gupta *et al.*, 2015). Additional genomic resources for the *Ca.* Parilichlamydiaceae will further clarify the phylogenetic placement of this clade.

6.3.7 The genome of Ca. S. epinephelii provides evidence for the ancient acquisition of a chlamydial Type III secretion system (T3SS)

A key chlamydial virulence mechanism employed by *Chlamydiae* is the Type III secretion system, T3SS, an ancient, highly conserved piece of molecular machinery that facilitates the translocation of effectors into the eukaryotic host cell through a needle-like "injectisome" to mediate bacterial survival and replication (Hueck, 1998; Horn et al., 2004; Coburn *et al.*, 2007; Peters et al., 2007). T3SS components can be divided into "apparatus" proteins that account for the structure of the transmembrane needle-like system, and "effector" proteins that are secreted into the host cytosol to mediate cell adhesion, entry, and invasion. Distinct clusters of T3SS proteins have been described, (Hueck, 1998; Read *et al.*, 2000; Hefty and Stephens, 2007; Peters et al., 2007; Betts-Hampikian and Fields, 2010) which appear to be highly conserved throughout the *Chlamydiae*.

The draft genome assembly of *Ca.* S. epinephelii reveals that most of the structural/apparatus components of the T3SS are intact**.** Furthermore, we observed the previously described conservation of the T3SS gene arrangement, as 3 of the 4 clusters were successfully assembled. These regions are syntenic with other chlamydial species analysed in this study (Figure

6.4; (Bertelli *et al.*, 2015)). Amino acid sequence identity for these proteins ranged from 26.2 to 69.5% between *Ca.* S. epinephelii and the CRBs, as could be expected for conserved proteins that are thought to be essential for the intracellular lifestyle of the *Chlamydiae*. This is comparable to the sequence conservation observed between other families, eg. *Parachlamydiaceae* and *Criblamydiaceae* share 61.8% amino acid identity for SctV. SctU and SctV (cluster 1) and SctR and SctS (cluster 2) appear to be the most highly conserved across the *Chlamydiales*. Cluster 4 could not be identified; it may be split over several contigs.

Homologs of some T3SS chaperones appear to be present, namely *syc*E (GCCT14_06130, 06150, 04180), *sct*G (GCCT14_04160), and *sct*O (GCCT14_04080). The heat shock protein family chaperones are also conserved in *Ca.* S. epinephelii, including two genomic copies each of *gro*ES and *gro*EL (hsp60 family), and one copy of *dna*K (hsp70 family), which share between 32.9 and 49.0% amino acid identity with other chlamydial species. Additionally, general secretion pathway proteins were identified in the genome, which could contribute to the stability, interaction and transport of secreted proteins.

While the structural and regulatory components of the T3SS are well conserved throughout the *Chlamydiae*, effectors are not, probably representing a major difference in host cell modulation mechanisms once host cell entry has taken place. We predicted 123 putative T3SS effectors (EffectiveDB; Jehl *et al.*, 2011), 65 of which were hypothetical proteins. These numbers are comparable to other pathogenic *Chlamydiae*. This suggests that *Ca.* S. epinephelii may possess several novel species-specific proteins which are potentially secreted into the host cell (Small *et al.*, 2004).

6.3.8 Genomic evidence of the pathogenic potential of Ca. S. epinephelii

In order to further establish the pathogenic potential of *Ca.* S. epinephelii, we conducted a genomic survey for a range of known chlamydial virulence factors.

Members of the *Chlamydiales* are known to encode a range of proteases that are employed to cleave host proteins during attachment and invasion of the host cell. One such protease is the chlamydial protease-like activity factor (CPAF), which is capable of degrading a range of host cell proteins involved in cytoskeleton stability and antigen presentation, for example, major histocompatibility complex (Zhong *et al.*, 2001; Chen *et al.*, 2010). A homolog of CPAF was identified (GCCT14_00610) that possesses an S41 peptidase domain and was most similar to that of *E. laussanensis* (21.9% amino acid identity). CPAF requires the *sec*-dependent pathway for secretion and cleavage (Chen *et al.*, 2010), and parts of this translocase complex including chaperones were identified in the *Ca.* S. epinephelii assembly (GCCT14_00180, 05190, 02900, 07980, 01650, 06730, 0001, 8720), suggesting CPAF could indeed be secreted by *Ca.* S. epinephelii. A putative homolog of tail-specific protease (*tsp*) (GCCT14_08300), which is implicated in blocking host transcriptional control and cytokine production (Lad *et al.*, 2007) is also present in the genome, containing the characteristic PDZ domain. The presence of these proteases and some of their chaperone provides evidence of mechanisms by which *Ca.* S. epinephelii may modulate the host cell and evade the fish host immune response upon infection.

Three copies of hemolysin (GCCT14_01140, 01150 and 08420), a secreted protein that lyses red blood cells, were annotated automatically in the *Ca.* S. epinephelii genome. Based on the protein domains identified (DUF21), these genes may actually encode for transporters involved in sodium and/or magnesium uptake. In fact, one of the hemolysins (GCCT14_08420) lies downstream of a magnesium and cobalt efflux protein. Related to hemolysin, *Ca.* S. epinephelii encodes a RTX (repeats in toxin) homolog (GCCT14_01940), containing the glycine-rich COG2931 domain. Although this cytolytic toxin family is described throughout Gram-negative bacteria (Coote, 1992), the copy in *Ca.* S. epinephelii exhibits less than 10% amino acid identity to the *rtx*A gene in *Ca.* Protochlamydia amoebophila UWE25. The high level of sequence divergence may hint at the advantage conferred by high levels of genetic variation. While this protein was predicted to be secreted, it is unclear whether it is functional, given that a) RTX toxins require

post-translational modification to become active and b) no other genes within the operon (*rtx*CBD) were identified (Lally *et al.*, 1999).

6.3.9 Predicting novel pathogenic proteins in the genome of Ca. S. epinephelii

In order to assess the novel aspects of the pathogenic processes utilised by Ca. S. epinephelii, we predicted putative novel pathogenic proteins based on similarity to known virulence factors.

278 proteins were determined "pathogenic" using the MP3 hybrid model (Gupta *et al.*, 2014), while 235 proteins were considered "virulent" by VirulentPred (Garg and Gupta, 2008), making up around 30% of the coding regions. This is comparable to other emerging CRB pathogens, as well as representatives of the *Chlamydiaceae*. 214 proteins were predicted to have roles in virulence by both programs, 139 of which are hypothetical proteins. This analysis indicates that *Ca.* S. epinephelii may harbour several novel pathogenesis mechanisms.

Further, a number of these predicted pathogenic proteins are co-located within what could either be virulence operons, or pathogenicity islands (PAIs). Several hallmarks of PAIs such as the presence of previously-described pathogenic proteins, high numbers of hypothetical proteins, deviation from the chromosomal G+C content and flanking tRNAs were observed on some of these regions (Che *et al.*, 2014). In two cases, genes for competence, recombination and DNA repair are present up or downstream of the putative pathogenic proteins.

6.4 Conclusion

This study provides (i) the first description of epitheliocystis in species of Serranidae, (ii) the description of a novel *Ca.* Similichlamydia sp. and (iii) the first genomic insight into a member of the *Ca.* Parilichlamydiaceae.

The results demonstrate that, despite the fact that Koch's postulates remain unproven, the molecular genomic data presents strong evidence that *Ca.* S. epinephelii is indeed the causative agent of epitheliocystis in Orange-spotted grouper. The presence of several known pathogenic and

virulence proteins such as CPAF and putative cytolytic toxins suggests that *Ca.* S. epinephelii dedicates a large portion of its genome to pathogenic mechanisms, much of which remains undiscovered. The conservation of the T3SS provides further support for pathogenic potential of this novel organism as well as evidence for its ancient acquisition.

These findings also suggest that future comparative genomic investigations will be of benefit in the contexts of other members of the *Chamydiales* as well as other epitheliocystis agents. Importantly, in lieu of using culture methods to understand the biology and pathogenicity of distantly related *Chlamydiae*, genomic methods directly applied to animal samples can be used to gain insight into the pathogenicity of this and other related species.

6.5 Experimental Procedures

6.5.1 Ethics statement

The animal ethics application to collect gill samples from farmed Orange-spotted grouper and Giant grouper was reviewed and approved by the James Cook University Animal Ethics Committee (document approval number A2056). Samples were collected from two commercial aquaculture farms (Pejo enterprises and FinFish enterprises) under the Australian Code for the Care and Use of Animals for Scientific Purposes (Australian Research Council) and the Animal Care and Protection Act (Queensland Department of Agriculture and Fisheries).

6.5.2 Sample collection

Initial discovery of putative epitheliocystis in *Epinephelus coioides* came from individuals that were sourced from an open pond aquaculture farm near Innisfail, Queensland (30-35 practical salinity units [psu]) and held in a closed recirculating aquarium system at James Cook University (JCU) in Cairns (33-35 psu) for experimental gill fluke infection trials between May and July 2014. The open pond aquaculture farm is adjacent to a saline creek, with which it has regular water exchange. The closed system at JCU has 10% weekly water exchanges with seawater. None of the

water used for exchange is treated. Microscopic examination of gills from this cohort revealed the presence of epitheliocystis-like nodules, which were subsequently preserved in neutral buffered formalin for histological and DNA sequencing analyses. This prompted further investigation of previously preserved gill material from past experiments and prospective sampling of both *E. coioides* and *E. lanceolatus*, within the aquaria facility on the JCU campus and on the farm between March 2014 and February 2015. Details of the cohorts sampled are outlined in Table 6.2.

6.5.3 DNA extraction

Gill clippings were placed in DNA grade ethanol upon collection and stored at room temperature. Prior to DNA extraction, tissues were removed from ethanol, and incubated with proteinase K and ATL buffer (Qiagen) at 56°C overnight until lysed, with intermittent vortexing. RNA was depleted by digestion with RNase A (Life Technologies) for 10 min at 37°C. DNA was purified by column purification with the DNeasy Blood & Tissue DNA extraction kit (Qiagen), according to manufacturer's instructions, with a final elution volume of 50μL.

6.5.4 PCR, sequencing and phylogenetic analysis

An initial screen for *Chlamydiales* was conducted on DNA from one *E. coioides* sample with putative epitheliocystis. The PCR was targeted to the *Chlamydiales* signature sequence of the 16S rRNA gene, using the primers described in (Everett et al., 1999). PCR was conducted on 5 μL of extracted DNA, with final concentrations of 1X reaction buffer (Roche), 4.5uM $MgCl_2$, 0.1 uM each dNTP (Thermoscientific), 0.3 uM each primer (Sigma) and 1 unit of Taq. Taq polymerase was activated at 95°C for 10 mins, then 35 cycles of 95°C for 30 sec, 56°C for 30 sec and 72°C for 1 min 30 sec were carried out prior to final extension at 72°C for 7 mins. For positive samples, amplification of the near full-length 16S rRNA gene was conducted using the primers 16SIGF and 16SB1. Reaction mix was as per the above reaction, but with final concentration of 2.5μM $MgCl_2$. Cycling conditions were as above, but with 1 min denaturation and annealing at 55°C.

Purified amplicons were prepared using the Big Dye Terminator cycle sequencing kit (Applied Biosystems), followed by EDTA purification and sequenced at the Molecular Genetics Research Facility at Queensland University of Technology. Sequences were deposited in Genbank under accession numbers KX880946-KX880947.

6.5.5 Histology and in situ hybridisation

Gills that were not placed in ethanol were fixed in formalin for downstream processing. Samples were dehydrated through six ethanol washes (70% to 100%) for 8 hours, two xylene washes for 2 hours and embedded in paraffin for 2 hours. 6 μm sections were then fixed to polylysine coated slides at 37°C prior to haematoxylin and eosin staining or *in situ* hybridisation (ISH).

Slides were stained for routine histology. Briefly, sections were de-paraffinised through two xylene washes followed by rehydration through decremental ethanol washes. Slides were stained in Harris haemotoxylin for 5-8 min, rinsed, and counter-stained in eosin for 30 sec. Sections were dehydrated through incremental ethanol washes, then twice in xylene, and dried prior to addition of the coverslip which was mounted using DPX.

Briefly, unstained slides were de-paraffinised and de-hydrated through three graded washes each of xylene and ethanol. Proteins were degraded by Proteinase K, then incubated in glycine followed by PBS. Denaturation was performed for 1 hr at 90°C, followed by hybridisation of the probes targeted to the *Chlamydiales* 16S rRNA gene overnight at 55°C at a final concentration of 500 ng/ml in hybridisation buffer. The slides were then washed through a series of stringency washes prior to addition of blocking buffer. The slides were incubated with alkaline phosphatase-conjugated anti-DIG antibody for 2 hours at room temperature, then incubated with NBT-Bcip in a dark box overnight to develop. Slides were counter-stained and de-hydrated through ethanol washes prior to coverslip mounting and visualisation by light microscopy.

6.5.6 Genome sequencing and assembly

Extracted DNA from a selected positive sample was subject to microbial DNA enrichment by depletion of methylated DNA using the NEBNext Microbiome Enrich Kit (New England Biolabs, Massachusetts, United States of America), according to manufacturer's instructions. The resulting DNA was precipitated with ethanol and resuspended in 50 μLTE buffer prior to 5 μL undergoing multiple displacement amplification using the Repli-G kit (Qiagen, Germany). 2 μg of enriched DNA was sent to the Australian Genome Research Facility (AGRF, University of Queensland) for shotgun sequencing on an Illumina MiSeq, 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 phred 33 quality cutoff then *de novo* assembly using SPAdes v3.1.1 (Bankevich *et al.*, 2012) in single-cell mode using default k-mer values. Chlamydial contigs were differentiated from host and other bacterial contigs on the basis of translated nucleotide similarity by running BLASTx analysis against an in-house chlamydial protein database. Contigs longer than 1000 bp with hits with e-values ≤ 0.005 and identity values ≥ 20% were subsequently also manually compared to the NCBI database. Metagenomics binning was also conducted on the final contig set to assess the completeness of the genome (Wu *et al.*, 2014). The resulting genomic contigs were annotated using RAST (Aziz *et al.*, 2008) and PFAM (Finn *et al.*, 2014) with additional manual annotation in Artemis (Rutherford *et al.*, 2000).

Contigs corresponding to the *Ca.* Similichlamydia epinephelii assembly were submitted to Genbank under Bioproject PRJNA343727.

6.5.7 Genome analysis

The genomic data for *Ca.* Similichlamydia epinephelii was compared to publicly available chlamydial genomes. Proteins were clustered into orthologous groups using OrthoFinder with

default parameters (Emms and Kelly, 2015). Additional annotations were curated using Prokka (Seemann, 2014).

Sixty-four single copy orthologs shared among all chlamydial genomes were identified. Amino acid sequences were aligned using MAFFT (Katoh *et al.*, 2002) and the alignments were concatenated prior to phylogenetic tree construction using FastTree (Price *et al.*, 2009) with default parameters with an *Akkermansia mucinophila* (Verrucomicrobia) outgroup.

Amino acids of the all CDSs were submitted to EffectiveDB (Jehl *et al.*, 2011) for prediction of secreted proteins and proteins with eukaryotic-like domains, and to MP3 (Gupta *et al.*, 2014) and VirulentPred (Garg and Gupta, 2008) to predict pathogenic proteins.

Declarations

Acknowledgments and conflict of interest

We are thankful to Mel Leef, University of Tasmania, for tissue processing and histology and Nicholas Thomson, Sanger Institute, UK for advice on genomic analysis. Partial funding for this work was provided through the Queensland Smart Futures grant program to TLM. Dr Richard Knuckey and Adam Reynolds generously provided access to brood stock and farmed orange-spotted and giant grouper samples used in this study. We also thank Martina Jelocnik for ongoing technical support.

The authors have no conflict of interest to declare.

Author contributions

A.T.B conducted lab work, data analysis and wrote the manuscript. T.P, N.L.B and W.Q provided bioinformatics support. A.B provided laboratory support. T.M provided samples. All authors contributed to interpretation of results and reviewed the manuscript.

Chapter 6 Tables & Figures

Table 6.1: *Prevalence and severity of epitheliocystis in cohorts of two grouper species by histopathology.*

Cohort	Number fish sampled	Number fish samples positive	Prevalence	Average number of cysts per filament
EC-May-Jul-14	6	5	83.33%	0.399
EC-May-14	3	0	0.00%	0
EC-Aug-14	8	7	87.50%	0.808
EC-Oct-14	6	6	100.00%	2.1
EC-Nov-14	5	4	80.00%	0.871
EC-Feb-15	3	0	0.00%	0
EL-Mar-14	4	0	0.00%	0
EL-May-14	1	0	0.00%	0
EL-Aug-14	1	0	0.00%	0
EL-Feb-15	16	0	0.00%	0

EC; E. coioides, EL; E. lanceolatus.

Table 6.2: *Chronological sampling and geographical distribution of E. coioides and E. lanceolatus*

Species	Samples collected	Cohort name	Location	Specimen type	Number examined
Epinephelus coioides	May-July 2014	EC-May-Jul-14	Innisfail/JCU	Juveniles	6
Epinephelus lanceolatus	Mar 2014	EL-Mar-14	Innisfail/JCU	Juveniles	4
Epinephelus lanceolatus	May 2014	EL-May-14	Innisfail/JCU	Juveniles	1
Epinephelus coioides	May 2014	EC-May-14	Cairns	Broodstock	3
Epinephelus coioides	August 2014	EC-Aug-14	Innisfail/JCU	Juveniles	8
Epinephelus coioides	October 2014	EC-Oct-14	Innisfail/JCU	Juveniles	1
Epinephelus lanceolatus	August 2014	EL-Aug-14	Innisfail/JCU	Juveniles	6
Epinephelus coioides	November 2014	EC-Nov-14	Innisfail/JCU	Juveniles	5
Epinephelus coioides	February 2015	EC-Feb-15	Innisfail	Juveniles	3
Epinephelus lanceolatus	Febuary 2015	EL-Feb-15	Innisfail	Juveniles	16

JCU; James Cook University (Cairns campus)

Table 6.3: *Assembly metrics for the draft genome of Ca. S. epinephelii*

Assembly metric	***Ca.* S. epinephelii GCCT14**
No. contigs (≥1,000 bp)	169
Largest contig	34,955 bp
Total length (sum of contigs)	981,540 bp
%G+C	39.65%
N50	10,591 bp
L50	26
No. predicted CDSs	937
Predicted coding density	90%
No. rRNA operons	1

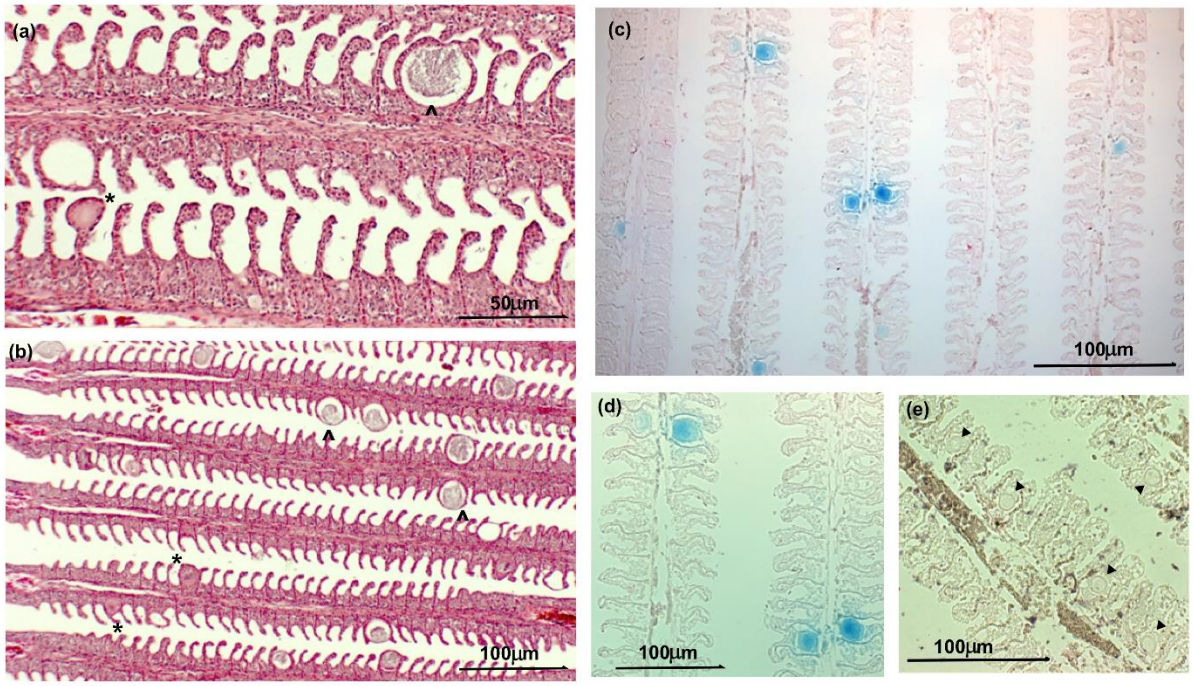


Figure 6.1: *Epitheliocystis in Orange-spotted grouper (*E. coioides*) gills stained with hematoxylin and eosin (a, b) and chlamydial localisation to cysts shown by in situ hybridisation (c, d, e). Basophilic cysts were spread sporadically throughout the gills and were encased in eosinophilic (*) or hyaline (^) capsules. A probe targeted to the 16S rRNA gene hybridised to cysts in epitheliocystis-positive gills (c, d), while the non-sense probe did not hybridise (black arrowheads) (e). Lighter stained signals most likely indicate cysts that contain fewer bacteria than the darkly stained cysts.*

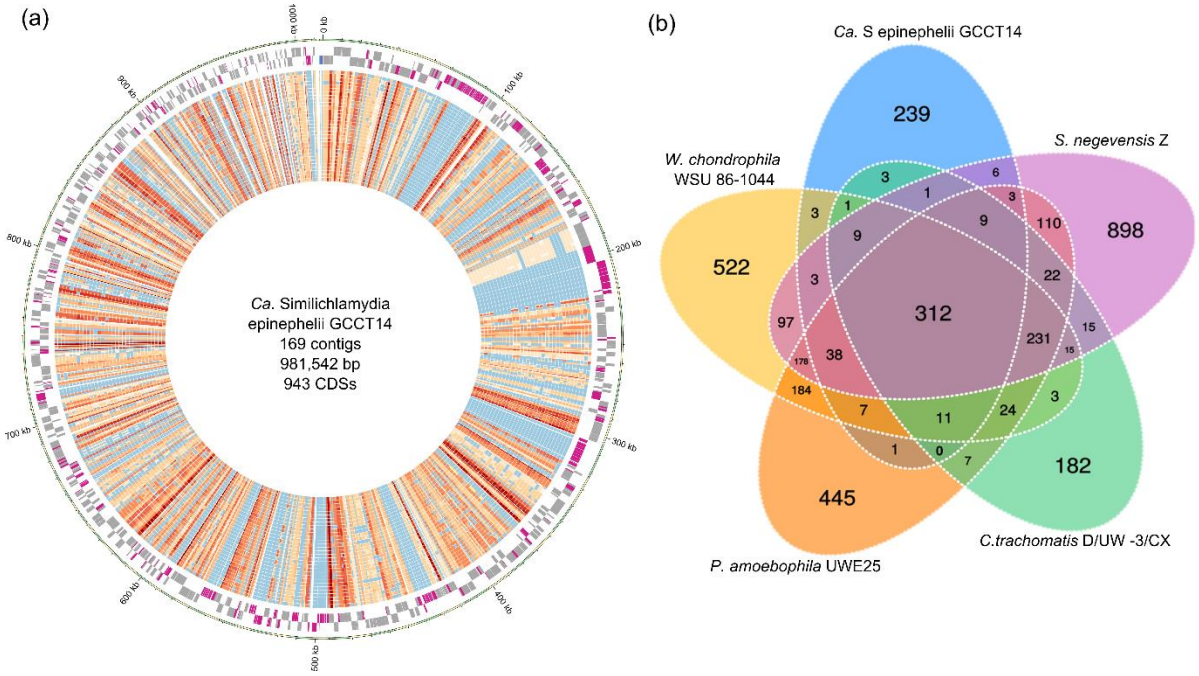


Figure 6.2: *Pan-genome analysis. CDSs were compared between the draft genome (68% complete) of* Ca. *S. epinephelii (outer rings) and other chlamydial genomes (inner shaded rings) (a); shading represents sequence conservation (red, orange, yellow: highly, moderately and somewhat conserved; blue: no homolog). Genes unique to* Ca. *S. epinephelii are coloured pink. Genomes used for analysis are listedthe order of most to least homologs with* Ca. *S. epinephelii, also corresponding to the order from the outermost to innermost ring. Orthologous groups were compared between* Ca. *S. epinephelii and representatives of four chlamydial families (b); unique and shared orthogroups are represented in the shaded ovals.*

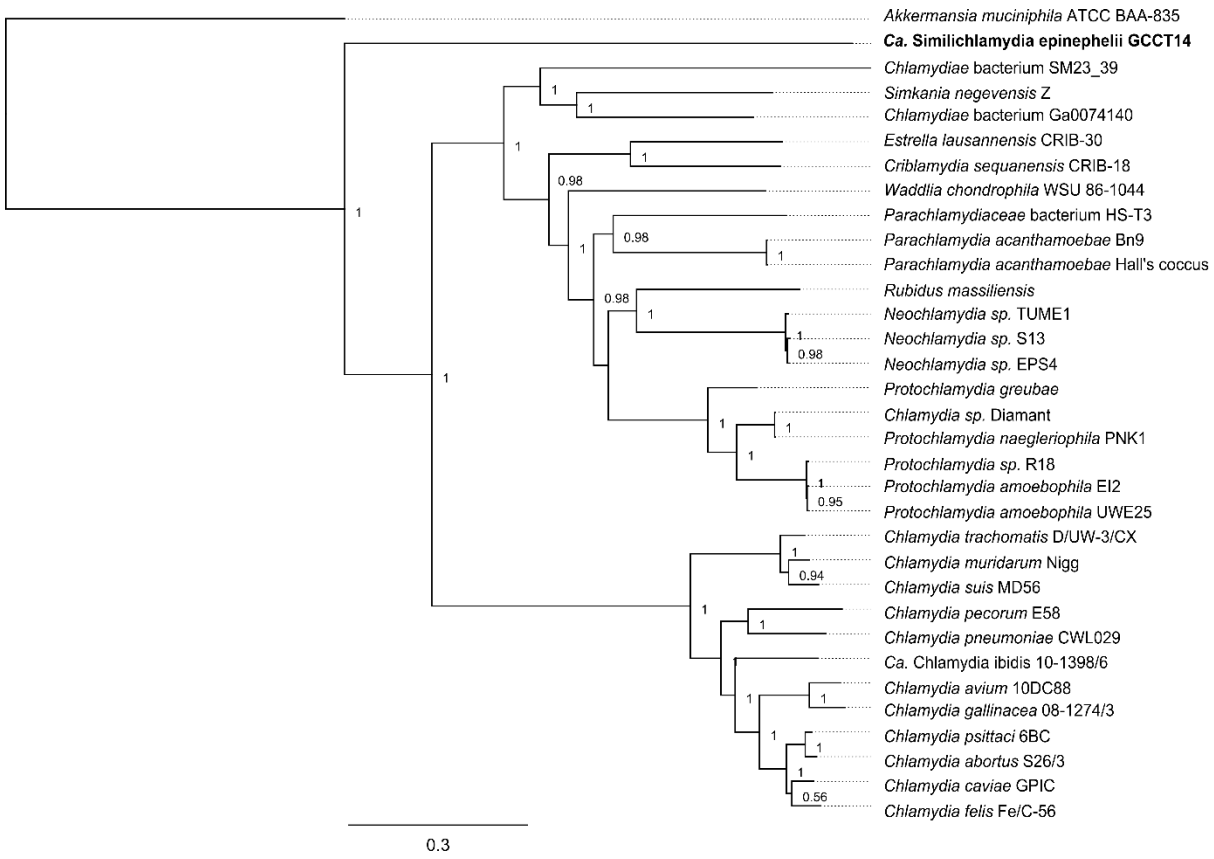


Figure 6.3: *Phylogenetic tree constructed from concatenated amino acid sequences corresponding to sixty-four single copy orthologs in all publicly available chlamydial genomes/assemblies.* Akkermansia mucinophila *was included as an outgroup species (*Verrucomicrobia *group). Tree constructed using FastTree based on an amino acid sequence alignment by MAFFT. Branch support values are shown.*

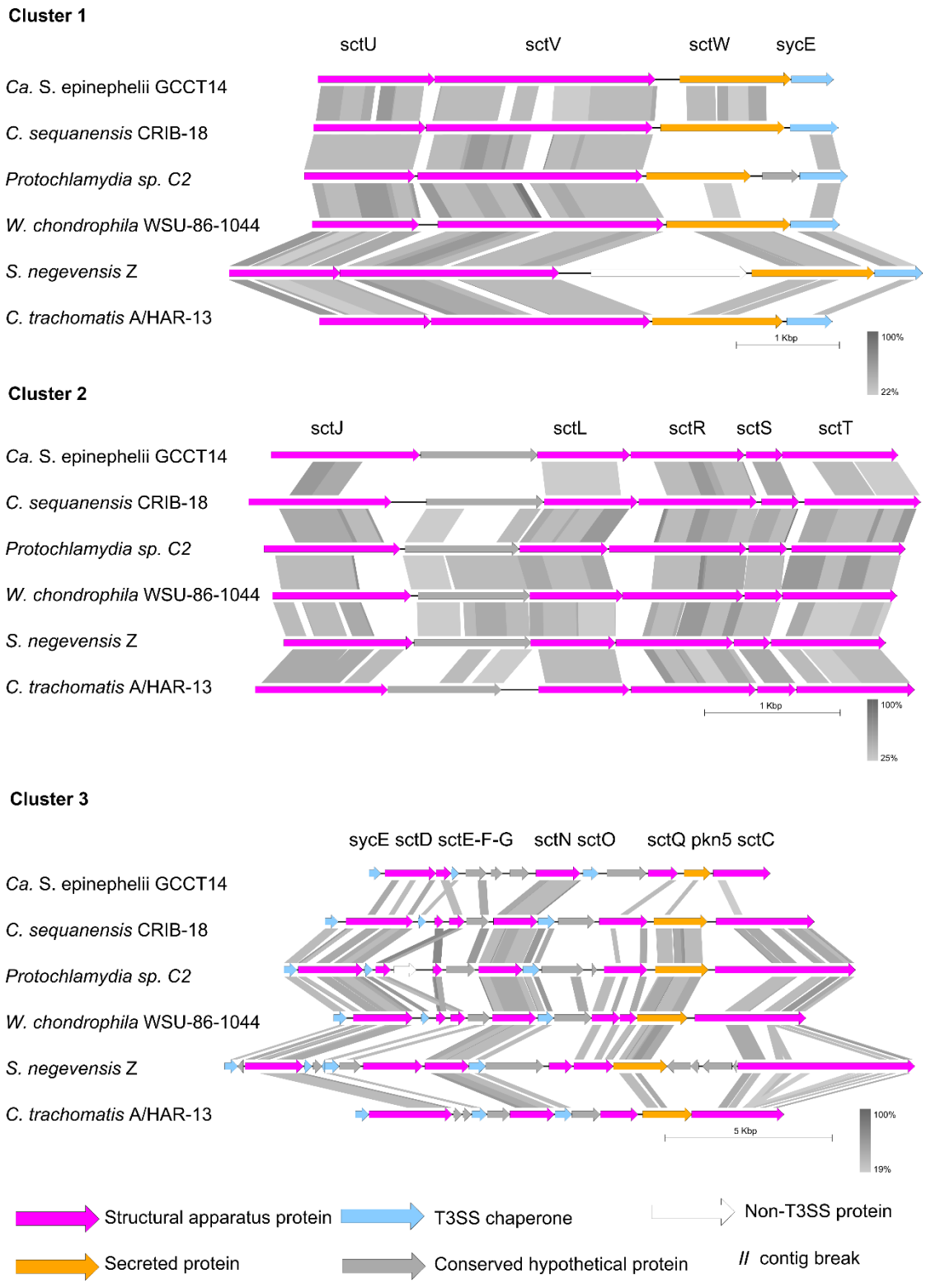


Figure 6.4: *Conservation of three out of the four chlamydial T3SS clusters. Arrows denote CDSs and are coloured based on function (see legend). Grey shading represents sequence similarity based on BLASTx analysis. Figure constructed using EasyFig (Sullivan et al., 2011).*

References

Amann, R., Springer, N., Schonhuber, W., Ludwig, W., Schmid, E.N., Muller, K.D., and Michel, R. (1997) Obligate intracellular bacterial parasites of acanthamoebae related to *Chlamydia spp*. *Appl Environ Microbiol* **63**: 115-121.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bachmann, N.L., Sullivan, M.J., Jelocnik, M., Myers, G.S., Timms, P., and Polkinghorne, A. (2015) Culture-independent genome sequencing of clinical samples reveals an unexpected heterogeneity of infections by *Chlamydia pecorum*. *J Clin Microbiol* **53**: 1573-1581.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455-477.

Barkallah, M., Gharbi, Y., Hassena, A.B., Slima, A.B., Mallek, Z., Gautier, M. et al. (2014) Survey of infectious etiologies of bovine abortion during mid- to late gestation in dairy herds. *PLoS One* **9**: e91549.

Baud, D., Goy, G., Osterheld, M.C., Borel, N., Vial, Y., Pospischil, A., and Greub, G. (2011) *Waddlia chondrophila*: from bovine abortion to human miscarriage. *Clin Infect Dis* **52**: 1469-1471.

Bertelli, C., Goesmann, A., and Greub, G. (2014) *Criblamydia sequanensis* Harbors a Megaplasmid Encoding Arsenite Resistance. *Genome Announc* **2**.

Bertelli, C., Collyn, F., Croxatto, A., Ruckert, C., Polkinghorne, A., Kebbi-Beghdadi, C. et al. (2010) The *Waddlia* genome: a window into chlamydial biology. *PLoS One* **5**: e10890.

Bertelli, C., Aeby, S., Chassot, B., Clulow, J., Hilfiker, O., Rappo, S. et al. (2015) Sequencing and characterizing the genome of *Estrella lausannensis* as an undergraduate project: training students and biological insights. *Front Microbiol* **6**: 101.

Betts-Hampikian, H.J., and Fields, K.A. (2010) The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. *Front Microbiol* **1**: 114.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Borel, N., Ruhl, S., Casson, N., Kaiser, C., Pospischil, A., and Greub, G. (2007) *Parachlamydia spp*. and related *Chlamydia*-like organisms and bovine abortion. *Emerg Infect Dis* **13**: 1904-1907.

Bradley, T.M., Newcomer, C.E., and Maxwell, K.O. (1988) Epitheliocystis associated with massive mortalities of cultured lake trout *Salvelinus namaycush*. *Diseases of Aquatic Organisms* **4**: 9-17.

Casson, N., Entenza, J.M., Borel, N., Pospischil, A., and Greub, G. (2008) Murine model of pneumonia caused by *Parachlamydia acanthamoebae*. *Microb Pathog* **45**: 92-97.

Che, D., Hasan, M.S., and Chen, B. (2014) Identifying pathogenicity islands in bacterial pathogenomics using computational approaches. *Pathogens* **3**: 36-56.

Chen, D., Lei, L., Lu, C., Flores, R., DeLisa, M.P., Roberts, T.C. et al. (2010) Secretion of the chlamydial virulence factor CPAF requires the *Sec*-dependent pathway. *Microbiology* **156**: 3031-3040.

Chen, Y., Yu, P., Luo, J., and Jiang, Y. (2003) Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome* **14**: 859-865.

Clark, A., and Nowak, B.F. (1999) Field investigations of amoebic gill disease in Atlantic salmon, *Salmo salar* L., in Tasmania. *Journal of Fish Diseases* **22**: 433-443.

Coburn, B., Sekirov, I., and Finlay, B.B. (2007) Type III secretion systems and disease. *Clin Microbiol Rev* **20**: 535-549.

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R.C. et al. (2011) Unity in variety--the pan-genome of the *Chlamydiae*. *Mol Biol Evol* **28**: 3253-3270.

Coote, J.G. (1992) Structural and functional relationships among the RTX toxin determinants of gram-negative bacteria. *FEMS Microbiol Rev* **8**: 137-161.

Draghi, A., 2nd, Popov, V.L., Kahl, M.M., Stanton, J.B., Brown, C.C., Tsongalis, G.J. et al. (2004) Characterization of "*Candidatus* piscichlamydia salmonis" (order *Chlamydiales*), a chlamydia-like bacterium associated with epitheliocystis in farmed Atlantic salmon (*Salmo salar*). *J Clin Microbiol* **42**: 5286-5297.

Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.

Everett, K.D., Bush, R.M., and Andersen, A.A. (1999) Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. *Int J Syst Bacteriol* **49 Pt 2**: 415-440.

Everett, K.D., Thao, M., Horn, M., Dyszynski, G.E., and Baumann, P. (2005) Novel chlamydiae in whiteflies and scale insects: endosymbionts '*Candidatus* Fritschea bemisiae' strain Falk and '*Candidatus* Fritschea eriococci' strain Elm. *Int J Syst Evol Microbiol* **55**: 1581-1587.

Fehr, A., Walther, E., Schmidt-Posthaus, H., Nufer, L., Wilson, A., Svercel, M. et al. (2013) *Candidatus* Syngnamydia venezia, a novel member of the phylum *Chlamydiae* from the broad nosed pipefish, *Syngnathus typhle*. *PLoS One* **8**: e70853.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. et al. (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222-230.

Fredricks, D.N., and Relman, D.A. (1996) Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* **9**: 18-33.

Friedman, M.G., Galil, A., Greenberg, S., and Kahane, S. (1999) Seroprevalence of IgG antibodies to the chlamydia-like microorganism *'Simkania Z'* by ELISA. *Epidemiol Infect* **122**: 117-123.

Fritsche, T.R., Horn, M., Wagner, M., Herwig, R.P., Schleifer, K.H., and Gautom, R.K. (2000) Phylogenetic diversity among geographically dispersed *Chlamydiales* endosymbionts

recovered from clinical and environmental isolates of *Acanthamoeba spp.* *Appl Environ Microbiol* **66**: 2613-2619.

Garg, A., and Gupta, D. (2008) VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinformatics* **9**: 62.

Greub, G., Kebbi-Beghdadi, C., Bertelli, C., Collyn, F., Riederer, B.M., Yersin, C. et al. (2009) High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. *PLoS One* **4**: e8423.

Guevara Soto, M., Vaughan, L., Segner, H., Wahli, T., Vidondo, B., and Schmidt-Posthaus, H. (2016a) Epitheliocystis Distribution and Characterization in Brown Trout (*Salmo trutta*) from the Headwaters of Two Major European Rivers, the Rhine and Rhone. *Front Physiol* **7**: 131.

Guevara Soto, M., Vidondo, B., Vaughan, L., Rubin, J.F., Segner, H., Samartin, S., and Schmidt-Posthaus, H. (2016b) Investigations into the temporal development of epitheliocystis infections in brown trout: a histological study. *J Fish Dis*.

Guevara Soto, M., Vidondo, B., Vaughan, L., Seth-Smith, H.M., Nufer, L., Segner, H. et al. (2016c) The emergence of epitheliocystis in the upper Rhone region: evidence for *Chlamydiae* in wild and farmed salmonid populations. *Arch Microbiol* **198**: 315-324.

Gupta, A., Kapil, R., Dhakan, D.B., and Sharma, V.K. (2014) MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. *PLoS One* **9**: e93907.

Gupta, R.S., Naushad, S., Chokshi, C., Griffiths, E., and 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**: 765-781.

Han, L., and Zhao, Z. (2008) Comparative analysis of CpG islands in four fish genomes. *Comp Funct Genomics*: 565631.

Harikrishnan, R., Balasundaram, C., and Heo, M.-S. (2011) Fish health aspects in grouper aquaculture. *Aquaculture* **320**: 1-21.

Hefty, P.S., and Stephens, R.S. (2007) Chlamydial type III secretion system is encoded on ten operons preceded by sigma 70-like promoter elements. *J Bacteriol* **189**: 198-206.

Hoffman, G.L., Dunbar, C.E., Wolf, K., and Zwillenberg, L.O. (1969) Epitheliocystis, a new infectious disease of the bluegill (*Lepomis macrochirus*). *Antonie Van Leeuwenhoek* **35**: 146-158.

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C.L., Purkhold, U., Fartmann, B., Brandt, P., Nyakatura, G.J., Droege, M., Frishman, D., Rattei, T., Mewes, H., Wagner, M. (2004) Illuminating the Evolutionary History of *Chlamydiae*. *Science*. **304:** 728-730

Huang, Y., Huang, X., Yan, Y., Cai, J., Ouyang, Z., Cui, H. et al. (2011) Transcriptome analysis of orange-spotted grouper (*Epinephelus coioides*) spleen in response to Singapore grouper iridovirus. *BMC Genomics* **12**: 556.

Hueck, C.J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**: 379-433.

Jehl, M.A., Arnold, R., and Rattei, T. (2011) Effective--a database of predicted secreted bacterial proteins. *Nucleic Acids Res* **39**: D591-595.

Kahane, S., Greenberg, D., Friedman, M.G., Haikin, H., and Dagan, R. (1998) High prevalence of "Simkania Z," a novel *Chlamydia*-like bacterium, in infants with acute bronchiolitis. *J Infect Dis* **177**: 1425-1429.

Karlsen, M., Nylund, A., Watanabe, K., Helvik, J.V., Nylund, S., and Plarre, H. (2008) Characterization of '*Candidatus* Clavochlamydia salmonicola': an intracellular bacterium infecting salmonid fish. *Environ Microbiol* **10**: 208-218.

Katharios, P., Papadaki, M., Papandroulakis, N., and Divanach, P. (2008) Severe mortality in mesocosm-reared sharpsnout sea bream *Diplodus puntazzo* larvae due to epitheliocystis infection. *Dis Aquat Organ* **82**: 55-60.

Katharios, P., Seth-Smith, H.M., Fehr, A., Mateos, J.M., Qi, W., Richter, D. 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**: 17609.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059-3066.

Kostanjsek, R., Strus, J., Drobne, D., and Avgustin, G. (2004) '*Candidatus* Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). *Int J Syst Evol Microbiol* **54**: 543-549.

Lad, S.P., Yang, G., Scott, D.A., Wang, G., Nair, P., Mathison, J. et al. (2007) Chlamydial CT441 is a PDZ domain-containing tail-specific protease that interferes with the NF-kappaB pathway of immune response. *J Bacteriol* **189**: 6619-6625.

Lafferty, K.D., Harvell, C.D., Conrad, J.M., Friedman, C.S., Kent, M.L., Kuris, A.M. et al. (2015) Infectious diseases affect marine fisheries and aquaculture economics. *Ann Rev Mar Sci* **7**: 471-496.

Lagkouvardos, I., Weinmaier, T., Lauro, F.M., Cavicchioli, R., Rattei, T., and Horn, M. (2014) Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. *ISME J* **8**: 115-125.

Lai, C.C., Crosbie, P.B.B., Battaglene, S.C., and Nowak, B.F. (2013) Effects of epitheliocystis on serum lysozyme activity and osmoregulation in cultured juvenile striped trumpeter, *Latris lineata* (Forster). *Aquaculture* **388–391**: 99-104.

Lai, Y.X., Jin, B.L., Xu, Y., Huang, L.J., Huang, R.Q., Zhang, Y. et al. (2014) Immune responses of orange-spotted grouper, *Epinephelus coioides*, against virus-like particles of betanodavirus produced in *Escherichia coli*. *Vet Immunol Immunopathol* **157**: 87-96.

Lally, E.T., Hill, R.B., Kieba, I.R., and Korostoff, J. (1999) The interaction between RTX toxins and target cells. *Trends Microbiol* **7**: 356-361.

Lienard, J., Croxatto, A., Prod'hom, G., and Greub, G. (2011) *Estrella lausannensis*, a new star in the *Chlamydiales* order. *Microbes Infect* **13**: 1232-1241.

Lohr, M., Prohl, A., Ostermann, C., Liebler-Tenorio, E., Schroedl, W., Aeby, S. et al. (2014) A bovine model of a respiratory *Parachlamydia acanthamoebae* infection. *Pathog Dis.*

Nagasawa, K. (2004) Diseases of cultured groupers. In. Nagasawa, K., and (eds.), E.R.C.-L. (eds). Southeast Asian Fisheries Development Center, Aquaculture Department, Iloilo, Philippines: Southeast Asian Fisheries Development Center (SEAFDEC), p. 81.

Nowak, B.F., and Clark, A. (1999) Prevalence of epitheliocystis in Atlantic salmon, *Salmo salar* L., farmed in Tasmania, Australia. *Journal of Fish Diseases* **22**: 73-78.

Nowak, B.F., and LaPatra, S.E. (2006) Epitheliocystis in fish. *J Fish Dis* **29**: 573-588.

Nylund, S., Steigen, A., Karlsbakk, E., Plarre, H., Andersen, L., Karlsen, M. et al. (2015) Characterization of '*Candidatus* Syngnamydia salmonis' (*Chlamydiales, Simkaniaceae*), a bacterium associated with epitheliocystis in Atlantic salmon (*Salmo salar* L.). *Arch Microbiol* **197**: 17-25.

Peters, J., Wilson, D.P., Myers, G., Timms, P., Bavoil, P.M. (2007) Type III secretion à la *Chlamydia*. *Trends in Microbiology* **16:** 241-251.

Pillonel, T., Bertelli, C., Salamin, N., and Greub, G. (2015) Taxogenomics of the order *Chlamydiales*. *Int J Syst Evol Microbiol* **65**: 1381-1393.

Pilloux, L., LeRoy, D., Brunel, C., Roger, T., and Greub, G. (2016) Mouse Model of Respiratory Tract Infection Induced by *Waddlia chondrophila*. *PLoS One* **11**: e0150909.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641-1650.

Psomopoulos, F.E., Siarkou, V.I., Papanikolaou, N., Iliopoulos, I., Tsaftaris, A.S., Promponas, V.J., and Ouzounis, C.A. (2012) The *Chlamydiales* pangenome revisited: structural stability and functional coherence. *Genes (Basel)* **3**: 291-319.

Qi, W., Vaughan, L., Katharios, P., Schlapbach, R., and Seth-Smith, H.M. (2016) Host-Associated Genomic Features of the Novel Uncultured Intracellular Pathogen *Ca.* Ichthyocystis Revealed by Direct Sequencing of Epitheliocysts. *Genome Biol Evol* **8**: 1672-1689.

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O. et al. (2000) Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res* **28**: 1397-1406.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Sadovy, Y.J., and Vincent, A.C.J. (2002) Ecological Issues and the Trades in Live Reef Fishes. In *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem*. Sale, P.F. (ed). San Diego: Academic Press, pp. 391-420.

Schmidt-Posthaus, H., Polkinghorne, A., Nufer, L., Schifferli, A., Zimmermann, D.R., Segner, H. et al. (2012) A natural freshwater origin for two chlamydial species, *Candidatus* Piscichlamydia salmonis and *Candidatus* Clavochlamydia salmonicola, causing mixed infections in wild brown trout (*Salmo trutta*). *Environ Microbiol* **14**: 2048-2057.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

Seth-Smith, H.M., Dourala, N., Fehr, A., Qi, W., Katharios, P., Ruetten, M. et al. (2016) Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular beta-proteobacteria. *ISME J* **10**: 1791-1803.

Seth-Smith, H.M., Harris, S.R., Skilton, R.J., Radebe, F.M., Golparian, D., Shipitsyna, E. et al. (2013) Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. *Genome Res* **23**: 855-866.

Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**: 1581-1590.

Steigen, A., Karlsbakk, E., Plarre, H., Watanabe, K., Overgard, A.C., Brevik, O., and Nylund, A. (2015) A new intracellular bacterium, *Candidatus* Similichlamydia labri sp. nov. (*Chlamydiaceae*) producing epitheliocysts in ballan wrasse, *Labrus bergylta* (Pisces, Labridae). *Arch Microbiol* **197**: 311-318.

Steigen, A., Nylund, A., Karlsbakk, E., Akoll, P., Fiksdal, I.U., Nylund, S. 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**: e66840.

Stride, M.C., Polkinghome, A., and Nowak, B.F. (2014) Chlamydial infections of fish: diverse pathogens and emerging causes of disease in aquaculture species. *Vet Microbiol* **171**: 258-266.

Stride, M.C., Polkinghorne, A., Powell, M.D., and Nowak, B.F. (2013a) "*Candidatus* Similichlamydia laticola", a novel *Chlamydia*-like agent of epitheliocystis in seven consecutive cohorts of farmed Australian barramundi, *Lates calcarifer* (Bloch). *PLoS One* **8**: e82889.

Stride, M.C., Polkinghorne, A., Miller, T.L., and Nowak, B.F. (2013b) 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**: 4914-4920.

Stride, M.C., Polkinghorne, A., Miller, T.L., Groff, J.M., Lapatra, S.E., and Nowak, B.F. (2013c) 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**: 1590-1597.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009-1010.

Szakolczai, J., Vetesi, F., and Pitz, S.R. (1999) Epitheliocystis disease in cultured pacu (*Piaractus mesopotamicus*) in Brazil. *Acta Vet Hung* **47**: 311-318.

Taylor-Brown, A., Bachmann, N.L., Borel, N., and Polkinghome, A. (2016) Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes. *BMC Genomics* **17**: 710-718

Taylor-Brown, A., Vaughan, L., Greub, G., Timms, P., and 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**: 1-15.

Thomas, V., Casson, N., and Greub, G. (2006) *Criblamydia sequanensis*, a new intracellular *Chlamydiales* isolated from Seine river water using amoebal co-culture. *Environ Microbiol* **8**: 2125-2135.

Vasilevsky, S., Gyger, J., Piersigilli, A., Pilloux, L., Greub, G., Stojanov, M., and Baud, D. (2015) *Waddlia chondrophila* induces systemic infection, organ pathology, and elicits Th1-associated humoral immunity in a murine model of genital infection. *Front Cell Infect Microbiol* **5**: 76.

Wei, S., Huang, Y., Cai, J., Huang, X., Fu, J., and Qin, Q. (2012) Molecular cloning and characterization of c-type lysozyme gene in orange-spotted grouper, *Epinephelus coioides*. *Fish Shellfish Immunol* **33**: 186-196.

Wheelhouse, N., Longbottom, D., and Willoughby, K. (2013) *Chlamydia* in cases of cattle pneumonia in Scotland. *Vet Rec* **172**: 110.

Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**: 26.

Zhong, G., Fan, P., Ji, H., Dong, F., and Huang, Y. (2001) Identification of a chlamydial protease-like activity factor responsible for the degradation of host transcription factors. *J Exp Med* **193**: 935-942.

CHAPTER 7: Comparative genomics of the *Candidatus* Parilichlamydiaceae reveals evidence of convergent evolution and genome reduction in the phylum *Chlamydiae*

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

[1] Centre for Animal Health Innovation, Faculty of Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland, 4556, Australia

[2] Institute of Microbiology, University of Lausanne and University Hospital Center, CH-1011 Lausanne, Switzerland

[3] Institute of Veterinary Pathology, University of Zurich, Zurich, CH-8057, Switzerland.

[4] Institute of Marine and Antarctic Studies, University of Tasmania, Newnham, Tasmania, 7248, Australia

* Correspondence: Adam Polkinghorne; Centre for Animal Health Innovation, Faculty of Science, Engineering and Education, University of the Sunshine Coast, Sippy Downs, Queensland, 4556, Australia; apolking@usc.edu.au, +617 5456 557

7.1 Abstract

Intracellular bacteria typically possess highly reduced, compact genomes, reflective of their need to sequester many essential nutrients from the host that they have lost the capability to synthesise themselves. The repeated evolution of this characteristic across diverse lineages resembles convergent evolution. *Chlamydiae* are one example of obligate intracellular bacteria with genomes ranging from 1.0 to 3.5 Mbp, and a host range spanning mammals, birds, fish, invertebrates and unicellular protists. The ecological and phylogenetic diversity of this phylum offers ongoing opportunities to study intracellular survival and metabolic pathways and adaptations. Of particular evolutionary significance are *Chlamydiae* belonging to the earliest diverging clade, found only in aquatic vertebrates. We assembled chlamydial genomes corresponding to three *Ca*. Parilichlamydiaceae species from gill metagenomes. This revealed highly reduced genomes similar in size to the terrestrial *Chlamydiaceae*, and in contrast to other *Chlamydia*-related bacteria (CRBs) that have a more cosmopolitan host range. We describe a reduction in genes encoding synthesis of nucleotides and amino acids, among other nutrients, and an enrichment of transport proteins. We suggest 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

7.2 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 phylum. A primary hallmark of obligate intracellular bacteria is their inability to replicate in a host-free environment, intimating that they depend on host-derived nutrient acquisition for replication and growth. These bacteria may be either pathogens or symbionts, the latter of which often contribute substrates for host metabolism (Douglas, 1998, Shigenobu et al., 2000), whilst intracellular pathogens have converged on similar mechanisms for intracellular entry and survival (Casadevall, 2008).

An obligate intracellular lifestyle is closely correlated with a reduced genome size compared to free-living organisms (Casadevall, 2008). Host-associated bacteria have lost many genes for functions that could be achieved through methods other than *de novo* synthesis such as biosynthesis from intermediates or uptake from the host via transport proteins (Moran and Plague, 2004, Moran, 2002, Bliska and Casadevall, 2009, Casadevall, 2008). Although which genes are lost is seemingly specific to different bacterial groups, the net result is a highly reduced, non-redundant genome, with a high coding percentage and reduced metabolic capacity compared to free-living bacteria (Mushegian, 1999, Moran, 2002, Konstantinidis and Tiedje, 2004, Sakharkar et al., 2004, Merhej et al., 2009). Although converging on the same ecological niche, these bacteria have taken different evolutionary paths to genomic minimalism (Moran, 2002).

It has been suggested that the initial transition to an intracellular habitat is accompanied by massive genomic deletions that are then fixed within the lineage (Moran and Mira, 2001, Moran, 2002). This is probably as a consequence of the increased impact of genetic drift resulting from the reduced effective population size of obligate intracellular bacteria (potentially from bottlenecks in bacterial numbers at the time of infection) favouring fixation of deleterious mutations, coupled

with intrinsic deletional bias (Mushegian, 1999, Mira et al., 2001, Sakharkar and Chow, 2005, Andersson and Kurland, 1998, Wolf and Koonin, 2013, Lynch, 2006, Kuo et al., 2009, Dutta and Paul, 2012, Batut et al., 2014). Reduced genomes are also generally lower in tRNAs and rRNAs and are biased toward a low G+C%, the latter of which is hypothesised to be due to either deletion of DNA repair genes (Moran et al., 2008, Nishida, 2012), mutational bias (Hershberg and Petrov, 2010) or selection (Hildebrand et al., 2010).

Chlamydiae are one such example of host-associated, obligate intracellular bacteria with minimal genomes. The well-characterised 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 to 1097 genes (Vorimore et al., 2013, Sachse et al., 2014, Taylor-Brown et al., 2016, Bachmann et al., 2014). With a few exceptions, species in this family are restricted to a particular group (e.g. mammals only or birds only), when compared to the host range of other related families in this phylum, which share a common ancestor with the *Chlamydiaceae* (*Chlamydia*-related bacteria; CRBs), recently proposed as the *Parachlamydiales* order [Gupta et al., 2016; Pillonel et al., 2018]. The CRB families will from here be referred to as *Parachlamydiales*, to delineate them from *Ca.* Parilichlamydiaceae (proposed at *Parilichlamydiales*). Members of these families possess slightly larger genomes (2.1-3.4 Mbp), which are suggested to reflect their adaptation to an amoebal cell niche, which may have played a roles as a "melting pot" for horizontal transfer among chlamydiae (Horn et al., 2004, Collingro et al., 2011, Taylor-Brown et al., 2015, Greub, 2009). *Parachlamydiales* exhibit a cosmopolitan distribution and broad host range spanning lower to higher eukaryotes, with infection often resulting in pathogenesis (Corsaro and Greub, 2006, Taylor-Brown et al., 2015, Lamoth and Greub, 2010). As our knowledge of the phylum expands, the ecological and phylogenetic diversity of the *Chlamydiae* offer ongoing opportunities to study intracellular survival mechanisms (Taylor-Brown et al., 2015, Konig et al., 2017).

We sought to utilise novel genome data from a distant fish-associated chlamydial family, *Ca.* Parilichlamydiaceae, to investigate chlamydial biology. Interestingly, different fish species

appear to harbour distinct chlamydial species. These bacteria are the most distantly related taxonomically described species in the phylum *Chlamydiae*, lying at the root of 16S rRNA-based phylogenies (Stride et al., 2013, Taylor-Brown et al., 2017). Remarkably, since their first description in Yellowtail kingfish (Stride et al., 2013a), bacteria belonging to two proposed genera (*Ca.* Similichlamydia & *Ca.* Parilichlamydia) in this novel family have been found in at least seven fish species (Stride et al., 2013b, Stride et al., 2013c, Steigen et al., 2013, Steigen et al., 2015, Seth-Smith et al., 2016, Guevara Soto et al., 2016, Taylor-Brown et al., 2017a) in the northern and southern hemispheres, leading to the suggestion that each fish species might be infected with its own co-evolved gill chlamydial pathogen (Stride et al., 2014). In the first study to provide insight into the biology of these distantly related and uncultivated chlamydiae, to our surprise, we discovered that the genome of the first species sequenced from the *Ca.* Parilichlamydiaceae, *Ca.* Similichlamydia epinephilii, was highly reduced and reminiscent of their terrestrial host-associated *Chlamydiaceae* relatives, despite their earlier divergence (Taylor Brown et al., 2017a). In a broader comparative genomics 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 species that are highly host-restricted.

7.3 Materials and methods

7.3.1 Sample collection and DNA treatment

Nucleic acid extracts from previously obtained epitheliocystis-positive gill samples from Yellowtail kingfish (n=5), Barramundi (n=2) and Striped Trumpeter (n= 3), were subject to DNA preparation and metagenome sequencing (Stride et al., 2013a, Stride et al., 2013b, Stride et al., 2013c). 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. The resulting DNA was precipitated with ethanol and

resuspended in 50 μL TE buffer prior to 5 μL undergoing multiple displacement amplification using the Repli-G MDA kit (Qiagen, Germany), to enrich the microbial DNA.

7.3.2 Metagenome sequencing and assembly

Two micrograms of enriched DNA was sent to the Australian Genome Research facility (AGRF) 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 to 2,5744,043 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 tetra-nucleotide frequencies, relative abundance and the presence of conserved bacterial genes by MaxBin v2.2.1 (Wu et al., 2016). In some cases, mixed bins occurred and these were re-binned and manually sorted using BLASTx analysis against an in-house chlamydial protein database. Contigs were also checked for G+C% vs coverage continuity, and in cases where a divergence in G+C% content coincided with a divergence in coverage, contigs were trimmed. Reads were mapped to the resulting bins and used for re-assembly using SPAdes. Not all gill metagenomes yielded complete chlamydial genomes so were omitted from downstream analysis. Only one sequence per species was used for comparative analysis.

7.3.3 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, (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 was compared with publicly available chlamydial genomes. 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). Amino acids of all the CDSs were submitted to KEGG BLASTKoala annotation for metabolic pathway construction, and compared to 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).

7.4 Results and Discussion

7.4.1 Ca. Parilichlamydiaceae species possess hallmarks of a highly reduced obligate intracellular bacterial genome

De novo assembly of gill metagenomes prepared from *Chlamydiae*-positive Yellowtail kingfish, Striped trumpeter and Barramundi (Table 7.1) yielded chlamydial assemblies consisting of seven, eight and 28 contigs for *Ca.* P. carangidicola, *Ca.* S. latridicola and *Ca.* S. laticola, respectively, the details of which are summarised in Table 7.2. The final draft assemblies totalled 771,491-881,979 bp, and were estimated to be 94.4% complete, based on the presence of 101 out of 107 phylogenetic markers, several of which are not conserved in other chlamydial genomes. 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%, as expected for bacteria that have undergone reductive evolution (Andersson and Kurland, 1998, Sakharkar and Chow, 2005, Merhej et al., 2009). We omitted from our analyses additional metagenomes that were sequenced but did not contain complete or sufficiently-covered chlamydial genomes. In these cases, non-chlamydial bacterial species were in higher abundance (based on read coverage) than the chlamydial sequences.

The G+C content of 37.5-43.7% within this family was comparable to other *Chlamydia*-related species and in the range generally exhibited by highly reduced genomes (Nishida, 2012, Merhej et al., 2009, Taylor-Brown et al., 2015). Mean read coverage across the metagenome-

assembled genomes also varied, with 49x, 46x and 692x for *Ca.* P. carangidicola, *Ca.* S. laticola and *Ca.* S. latridicola, respectively with the chlamydial genome accounting for 1-38% of the trimmed reads.

There is a positive relationship between genome size and number of coding regions in the *Chlamydiae*, with *Chlamydiaceae* and *Ca.* Parilichlamydiaceae possessing significantly smaller genomes compared to the *Parachlamydiales* (Figure 7.1a), despite their earlier divergence (Figure 7.1b). We consider this genome size convergence to be reflective of the predicted strict dependence on vertebrate hosts, compared to the *Parachlamydiales*, which retain the ability to naturally infect both unicellular and multicellular eukaryotes (Collingro et al., 2011), whereas members of the *Chlamydia* genus may not be successfully cultivated in *A. castellani* (Essig et al., 1997, Wirz et al., 2008), whilst no *Chlamydia* species have been successfully recovered from environmental samples by amoebal co-culture. This host range relationship may be exemplified in the *Ca.* Parilichlamydiaceae as, with a few exceptions, each species in this family has been detected in a single fish species. This relationship is consistent with previously described trends of high coding percentage, small genome size and low number of coding regions in obligate intracellular or host-associated bacteria (Konstantinidis and Tiedje, 2004, Merhej et al., 2009), such as well-characterised bacteria from diverse phyla that possess similarly small genomes, such as *Mycoplasma pneumoniae*, *Borellia burgdorferi* and *Buchnera aphidicola*. These species (pathogens and symbionts) were thus integrated into our analysis along with one representative each of *Parachlamydiales* (*P. amoebophila*) and *Chlamydiaceae* (*C. trachomatis*).

7.4.2 The minimal metabolic capacity of Ca. Parilichlamydiaceae is an example of convergent evolution

Several efforts have been made to characterise the minimal requirements for microbial life. These studies have coupled biochemical and genetic data from a range of bacteria from diverse niches and phyla and as such, several iterations of the "minimal gene set" have been proposed, with minor differences being attributed to the availability of genomes and basis for essentiality

(Mushegian, 1999, Koonin, 2000, Sakharkar et al., 2004, Gil et al., 2004, Hutchison et al., 2016, Ye et al., 2016). We have considered sets of conserved bacterial COGs to make generalised functional comparisons (Merhej et al., 2009) and the minimal bacterial gene set proposed by Gil et al., 2004 to assess the presence of essential bacterial genes given the unexpected small size of these novel genomes.

The gene repertoire of the *Ca.* Parilichlamydiaceae that could be assigned COGs is primarily comprised of genes for DNA replication and repair, translation and post-translational modification (information storage and processing) coupled with a low representation of genes for defence mechanisms, lipid metabolism and co-enzyme metabolism (Figure 7.2). By comparison, free-living bacterial genomes are rich in genes for amino acid metabolism and energy production (Figure 7.2).

Interestingly, of 100 COGs that were 'universally' lost by alpha- and gamma-proteobacterial mutualists and parasites (Merhej et al., 2009), six were retained by Parilichlamydiaceae species, while 12 were retained by *C. trachomatis* and 30 by *P. amoebophila*, suggesting that perhaps these are required for virulence or pathogenicity by these taxa. To further validate the genome reduction observed in the *Ca.* Parilichlamydiaceae, we considered a set of 208 orthologs conserved between *P. amoebophila, C. trachomatis, Legionella pneumophila, Coxiella burnetii, Escherichia coli* and *Rhodopirellula baltica* (Gimenez et al., 2011). We found 141 orthologs conserved in the *Ca.* Parilichlamydiaceae genomes, congruent with 141, 136 and 155 in *M. pneumoniae*, *B. burgdorferi* and *B. aphidicola*, respectively.

Essential genes for DNA and RNA metabolism and protein processing are highly conserved in *Ca.* Parilichlamydiaceae with a few exceptions (Gil et al., 2004). Three parts of the translation machinery involved in tRNA maturation and modification are missing in our draft genomes: two GTP-binding proteins (*mnmE* and *mnmG*) and dimethyladenosine transferase (*ksgA*). It is possible that other predicted GTP-binding proteins and rRNA methyltransferases may

account for the same functions. All translation factors are present, as are genes for protein processing and secretion and ribosomal proteins, including those involved in ribosome function. All aminoacyl-tRNA synthases are present except for glutaminyl tRNA synthase (*glnS*), which is also absent from other chlamydial genomes. Two of the four cellular transport proteins are 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 – are predicted in the draft *Ca.* Parilichlamydiaceae genomes. Some genes for energetic and intermediary metabolism that are also differentially present throughout the chlamydiae were absent, and are described in more detail below.

7.4.3 Nucleotide metabolism and acquisition

It has been well-documented that the *Chlamydiaceae* lack genes for purine and pyrimidine *de novo* synthesis (Nunes and Gomes, 2014, Bertelli et al., 2010, Bertelli et al., 2015, Stephens et al., 1998, Kalman et al., 1999, Voigt et al., 2012), and this is also seen in the *Ca.* Parilichlamydiaceae. Notably, the *Ca.* Parilichlamydiaceae appear to lack ribose-phosphate pyrophosphokinase (*prsA*), meaning phosphoribosyl pyrophosphate (PRPP), a precursor molecule, cannot be synthesised. *P. amoebophila* possesses *prs*A, as do the three species of small-genome bacteria used in our comparison, but *C. trachomatis* does not. Hence, none of the essential genes involved in purine *de novo* synthesis appear 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, eg. (hypoxanthine phosphoribosyltranferase*; hpt*). *Ca.* Parilichlamydiaceae do 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, however.

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 synthesised from dUMP via thymidylate synthetase, 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 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 et al., 2013a, Stride et al., 2013c), 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 (Stride et al., 2014, Nowak and LaPatra, 2006), 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).

Ca. Parilichlamydiaceae genomes are predicted to encode three nucleotide transporters/translocases (NTT or Npt) that may facilitate the uptake of host-excreted

compounds. Homologs of those transporters can be identified in all *Chlamydiae* genomes and several other bacterial phyla (Greub and Raoult, 2003, Heinz et al., 2014). They are each around 500 amino acids in length and contain an MFS domain and 10-12 transmembrane domains, providing further evidence that they are membrane-bound transport pumps. These NTTs share 12 to 62% amino acid similarity to the five functionally characterised 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*. 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.

Amino acid metabolism

CRBs are capable of synthesising several amino acids whilst *Chlamydiaceae* and other obligate intracellular bacteria are generally auxotrophic for these nutrients (Omsland et al., 2014, Bertelli et al., 2015, Bertelli et al., 2010, Gil et al., 2004). No genes for amino acid synthesis were included in the minimal gene set except for glycine hydroxymethyltransferase *glyA*, which catalyses serine to glycine interconversion and is encoded by the *Ca.* Parilichlamydiaceae (Gil et al., 2004. No other amino acid synthesis genes are encoded by *Ca.* Parilichlamydiaceae in a striking similarity to *M. 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 *C. trachomatis,* which is one of the only amino acids that some *Chlamydia spp.* can synthesise. Tryptophan metabolism is implicated in persistence

and tissue tropism (Akers and Tan, 2006) but interestingly cannot be synthesised by all species, as evidenced by the lack of an intact or partial tryptophan biosynthesis operon in our draft genomes. A tryptophan/tyrosine permease is predicted, which shares 30% amino acid identity to that of *Parachlamydia* spp., with the amino acid/polyamine transporter 2 domain predicted. This transporter may enable uptake and utilisation of tryptophan and related amino acids *in lieu* of biosynthesis.

Carbohydrate and energy metabolism

Ca. Parilichlamydiaceae encode the essential components of the glycolysis/gluconeogenesis pathway to produce pyruvate, ATP and NADH from glucose, or the reverse (Figure 7.3). Unlike other *Chlamydiae*, they are predicted to use phosphoglucomutase (*pgm*) to convert glucose 1-phosphate to glucose 6-phosphate, whereas *P. amoebophila* uses glucokinase (*gck*) to phosphorylate glucose.

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* and other small-genome bacteria.

Although pyruvate can be synthesised from phosphoenol-pyruvate by pyruvate kinase (*pyk*), *Ca.* Parilichlamydiaceace 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 (i) presence of a complete cycle in most *Parachlamydiales* while (ii) *Chlamydiaceae* lack essential genes (citrate synthase, aconitase and isocitrate dehydrogenase (*gltA*, *acnB* and *icd*), and 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 (Fraser et al., 1997, Himmelreich et al., 1996). 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 other parts of the electron transport chain generate sufficient energy to account for the loss of this cycle in bacteria with highly reduced genomes.

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-motive force generation (Figure 7.3). 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, not dissimilar to other small-genome bacteria.

Predicted carbon metabolism pathways in the *Ca.* Parilichlamydiaceae highly resemble that of the other species investigated, given that much of the carbon utilised 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 the other genomes analysed except for *B. burgdorferi*, 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 who reside in 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, like *B. aphidicola*, that reside inside a host-derived vacuole (Mushegian and Koonin, 1996, Gil et al., 2004). It could also be the case for *Chlamydiae* as they reside inside the inclusion, however lipids are still required for membrane biogenesis. *Ca.* Parilichlamydiaceae species appear to be lacking the enzymes deemed essential for glycerophospholipid biosynthesis (Gil et al., 2004), and several other chlamydiae are lacking the components to synthesise acetyl-coA, the starting molecule for fatty acid synthesis. This suggests that these pathways may be in decay in some species, or that host-derived CoA or intermediates are able to be utilised. Nonetheless, *Ca.* Parilichlamydiaceae encode two copies of CDP-diacylglycerol-glycerol-phosphatidyltransferase (*pgsA*), a long-chain-fatty-acid-[acyl-carrier-protein] ligase with a phosphopantetheine binding domain and acyl-coA synthetase (AMP-forming) domain, the latter of which may account for the lack of acyl-coA synthetase (*fadD*) in most *Chlamydiae*, and a fourth protein which is a distantly related homolog of an acyl-ACP-phospholipid O-acyltransferase. These genes, coupled with uptake of substrates by transporters described below, may accomplish fatty acid and lipid metabolism *in lieu* of other genes.

Parilichlamydiaceae genomes are enriched with membrane transporters

As compensation for their reduced genome size and biosynthetic capacities, *Ca.* Parilichlamydiaceae species possess a diverse repertoire of 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 each *Ca.* Parilichlamydiaceae species, while *OppBCDF* are also present at a single locus. This structure is most similar to that observed in the *Chlamydiaceae,* whereas in CRB 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 predicted in the *Opp* proteins may suggest that these genes function to transport several molecule classes.

The presence of operons for phospholipid, lipoprotein and metallic cation transport are again variable between *Parachlamydiales* and the *Chlamydiaceae*. For example, phospholipid transporters (*mlaDEF*) are encoded by *Parachlamydiales* only, while more lipopolysaccharide transporters are encoded by *Parachlamydiales* (*rfbAB* and *lptFGB*) than *Chlamydiaceae* (*lpt* only) but not by *Ca.* Parilichlamydiaceae. Lipoprotein transporters (*lol*) are encoded throughout the phylum. *Ca.* Parilichlamydiaceae encode several metallic cation transporters (zinc, manganese, iron); again, some may serve several functions. 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.

7.4.4 The updated core Chlamydiales genome

The reduced genome size of the *Ca.* Parilichlamydiaceae has a strong impact on the size of the *Chlamydiales* core genome: just 342 orthologous groups are shared between the *Ca.* Parilichlamydiaceae, *Parachlamydiaceae*, *Simkaniaceae*, *Waddliaceae*, *Criblamydiaceae* and *Chlamydiaceae* (Figure 6.4a). 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 60% and 80% for *Parachlamydiales* and *Chlamydiaceae* (Figure 7.4b). We anticipate that the number of core chlamydial genes will plateau once representatives of all uncharacterised families are sequenced. For example, no arthropod-associated chlamydial genomes have yet been sequenced despite their repeated observation and diversity among vectors such as ticks (Pilloux et al., 2015, Hokynar et al., 2016, Burnard et al., 2017, Croxatto et al., 2014).

The genomic data generated from such studies will hopefully help elucidate host-specific adaptation mechanisms and clarify the relationship between genome size, population size and host restriction.

Most core genes are involved in DNA replication and repair, transcription, translation, post-translational modification, inorganic ion transport and carbohydrate metabolism (Figure 7.4c). Orthologous groups shared between *Ca.* S. latridicola and at least one other family also fall into these categories, together with proteins for which a general function only could be predicted. In general, the largest orthologous groups exhibited throughout the phylum are shared between all chlamydial families and hence contribute to the core genome. These include various ABC transporters, ATP/ADP translocases, proteins with tetratricopeptide repeats, GroEL and GroES family chaperones, sodium/alanine symporter family proteins, and various proteases, including the tail-specific protease. Conversely, the largest protein families not present in the *Ca.* Parilichlamydiaceae genomes are mainly restricted to protein groups that are expanded from gene duplication events in certain species or families, for example the polymorphic membrane proteins of the *Chlamydia sp.* genomes (Nunes and Gomes, 2014, Bachmann et al., 2014, Vasilevsky et al., 2016), the outer membrane protein family of the *Waddliaceae* (Bertelli et al., 2010), and the and leucine rich repeat proteins (LRRs) present in *Protochlamydia amoeobophila* (Eugster et al., 2007) and *Neochlamydia spp.* genomes (Domman et al., 2014). Notably, several of the largest orthogroups missing from *Ca.* Parilichlamydiaceae genomes are also absent from *Chlamydiaceae* genomes, further suggesting several gene sets have been lost in parallel by both *Ca.* Parilichlamydiaceae and *Chlamydiaceae* over the course of adaptation to their respective multicellular eukaryotic hosts, whilst their presence is several CRB species highlights either a) their necessity in particular genomes, b) that they are yet to be lost from those species as they are still in the process of reductive evolution, or c) that they were acquired by the ancestor of the *Parachlamydiales* clade.

Our initial comparative analysis using one species representative for each chlamydial family revealed 256 proteins unique to *Ca.* Parilichlamydiaceae (Figure 7.4a). Further comparative analysis 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 1 or 2 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. We also observed several genes that *Ca.* Parilichlamydiaceae have in common with *A. mucinophila*, a member of the *Plantomycetes-Verrucomicrobia-Chlamydiae* superphylum, that are variably encoded throughout the *Chlamydiae*. Examples include *SecG*, which is encoded by CRBs but variable within *Chlamydia sp.* genomes, the *MlaDEF* genes, which again are mostly encoded only by the former "*Chlamydophila*" species, as well as CRBs, and tRNA synthetases for proline, asparagine and glycine.

7.5 Limitations and future directions

While we have presented a brief comparative analysis of the metabolic strategies employed by novel gill-associated CRBs in comparison to host-associated bacteria with reduced genomes, there are several shared and unique mechanisms still to be understood for these unique CRBs. We have not performed any analysis of biological features reported in other chlamydial genomics studies including virulence mechanisms, secretion systems, antimicrobial resistance, 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 et al., 2017a), whilst others should be the focus of further comparative studies upon completion of genome sequencing of other CRBs.

Notably, plasmids have been reported for all *Parachlamydiales* species, and likewise are harboured by most strains of most species of *Chlamydia*. As we sequenced entire metagenomes,

and plasmid sequences are preferentially amplified by MDA (Taylor-Brown et al., 2016, Taylor-Brown et al., 2017b) we concluded that the strains we sequenced do not contain plasmids, giving rise to three possible scenarios. The first is that members of this family do not harbour a plasmid at all. Indeed, it remains unclear what the exact role(s) of *Chlamydia* plasmids are, so it is entirely possible that the plasmid was lost over the course of their evolution. The second scenario is that the plasmid was acquired by the *Chlamydiaceae/ Parachlamydiales* ancestor after the divergence of the *Ca.* Parilichlamydiaceae clade. The third fourth scenario is that parts of the plasmid have been incorporated into chromosome. This is not evident from G+C% content or presence of plasmid-related genes on the chromosome. The third scenario is that plasmids may be present in this family but are not harboured by any of the strains we sequenced.

It is also unclear what role co-habitation 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., 2017b, Taylor-Brown et al., 2016).

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 characterised 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), we opted for a non-targeted deep-sequencing approach, which provided insight into the chlamydial agent of interest against a background of gill mirobiota, 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 quite confident, based on the presence of conserved and essential genes, the number of rRNAs and tRNAs, the GC skew and contig end read pairing, 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.

This 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, co-cultivation of *Chlamydiae* in free-living amoebae has been only successful with chlamydial species having 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 multiple phyla. As with the terrestrial vertebrate specialists, the *Chlamydiaceae*, all with a comparable genome size, the *Ca.* Parilichlamydiaceae may similarly have a restricted (largely marine) vertebrate host range. This hypothesis needs to be tested by examining other fish species sharing the habitats of *Ca.* Parilichlamydiaceae infected fish, as well as the widening range of potential invertebrate hosts, known to harbour *Chlamydiae* (Viver et al., 2017). 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.

7.6 Conclusions

In our comparative genomics analysis of gill-associated chlamydiae, rather than a larger genome representing a broader metabolic capacity than their terrestrial and amoebae-associated cousins, we characterised a highly reduced genome shaped by reductive evolution presumably over a long period of host adaptation to the gill niche. We predict that this reductive evolution is an example of convergent evolution, where the gill-associated chlamydiae have evolved a genomic repertoire similar in content to members of the *Chlamydiaceae*, despite their earlier divergence, whilst

retaining selected genes from their endosymbiont relatives, and possessing several lineage-specific genes that may have arisen via gene birth and gain. This genome reduction may be reflective of a) a long period of host adaptation to the gill niche, b) limited transmission routes, resulting in c) low population sizes, leading to c) genetic drift, favouring purifying selection. Genome sequencing of diverse novel species from a range of ecological niches will further elucidate these relationships.

Declarations

Acknowledgements

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.

Conflict of Interest

The authors have no conflicts of interest to declare.

Author 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.

Chapter 7 Tables & Figures

Table 7.5: *Details of gill samples utilised for metagenome sequencing and chlamydial genome assembly.*

Sample	**76ST10**	**Hat2**	**007YTK11**
Host	Striped trumpeter (*Latris lineata*)	Barramundi (*Lates calciferer*)	Yellowtail kingfish (*Seriola lalandi*)
Chlamydial species	*Ca.* Similichlamydia latridicola	*Ca.* Similichlamydia laticola	*Ca.* Parilichlamydia carangidicola
Collection date	Nov 2010	Aug 2012	May 2011
Collection location	Tasmania	South Australia	South Australia

Table 7.6: *Genome characteristics for members of the phylum Chlamydiae. Ranges of values are given for each family, with the number of species or strains used for comparison in brackets.*

Family (No. genomes used for comparison)	Chromosome size (Mbp)	No. predicted CDS	Plasmid size range (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	1986 - 3042	39-145 (40-160)	33.0 - 44.0	1 - 4	35-42	69.7[3] - 90.8
Waddliaceae (2)	2.116 - 2.141	1,934, 2,028	15.5[1] (22)	43	2	34-37	92.2 - 93
Simkaniaceae (1)	2.496	2519	132 (136)	38	1	35	91.3
Criblamydiaceae (2)	2.820 - 2.969	2213, 2426	9.1 (15), 89.5 (92)	38.0 - 48.0	1	40	86.0 - 89.7
Chlamydiaceae (5)	1.041 - 1.229	911 - 1112	7.5[2] (8)	37.4 - 41.6	1 - 2	37 - 39	89.0 - 91.2

[1] *W. chondrophila* plasmid in WSU 86-1044 strain only

[2] *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*

[3] 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%

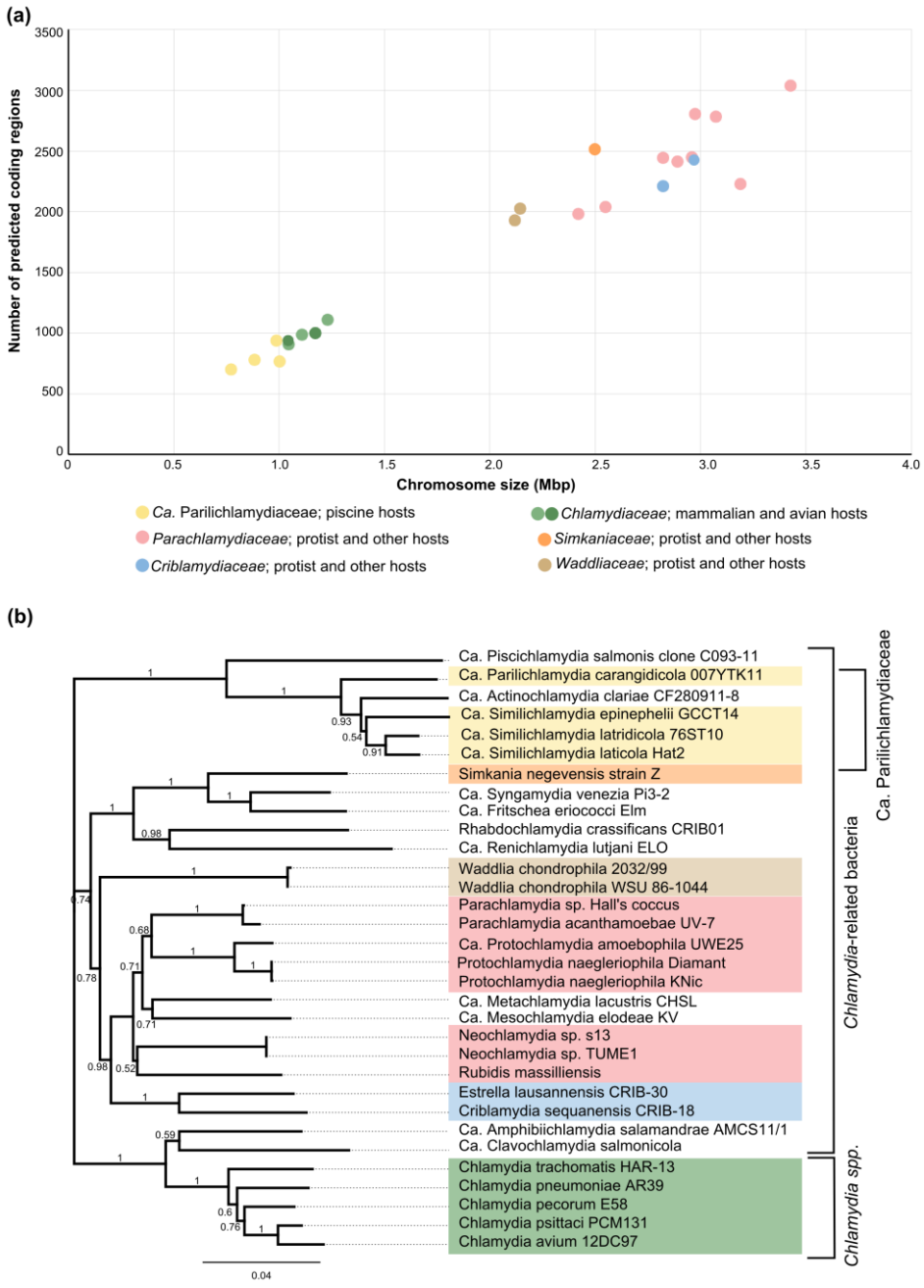


Figure 7.1: *Relationship between chromosome size and number of coding regions in* Chlamydiae *with reference to phylogenetic relationships and ecological diversity. (a) Relationship between chromosome size and number of coding regions in* Chlamydiae. *Species are coloured based on chlamydial family. (b) 16S rRNA phylogenetic tree depicting the relationship between the* Ca. *Parilichlamydiaceae (yellow background) and other classified chlamydial species. Species with genomic representatives have shaded backgrounds according to family, whilst for taxa without genomic representatives, only one genus-level taxon is shown for simplicity. Epitheliocystis agents are denoted by asterisks. The tree was constructed by FastTree based on a MAFFT alignment in Geneious.*

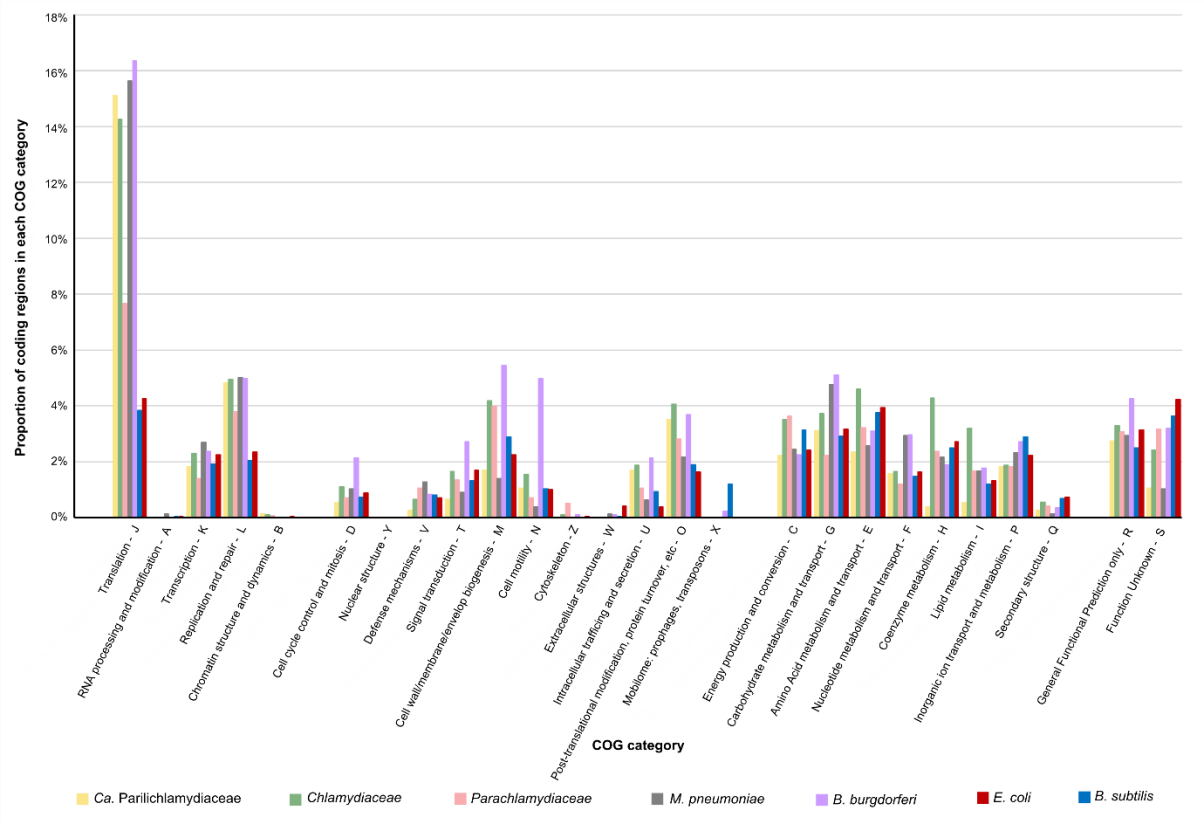


Figure 7.2: *Distribution of COGs across* Ca. *Parilichlamydiaceae, CRBs,* Chlamydiaceae *and small genome bacteria used in this analysis. COG categories are presented as a proportion of total number of coding regions and are grouped per general function from left to right: information storage and processing (J, A, K, L, B), cellular processes and signalling (D, Y, V, T, M, N, Z, W, U, O, X), metabolism (C, G, E, F, H, I, P, Q) and poorly characterised genes (R, S). See legend for species colours.*

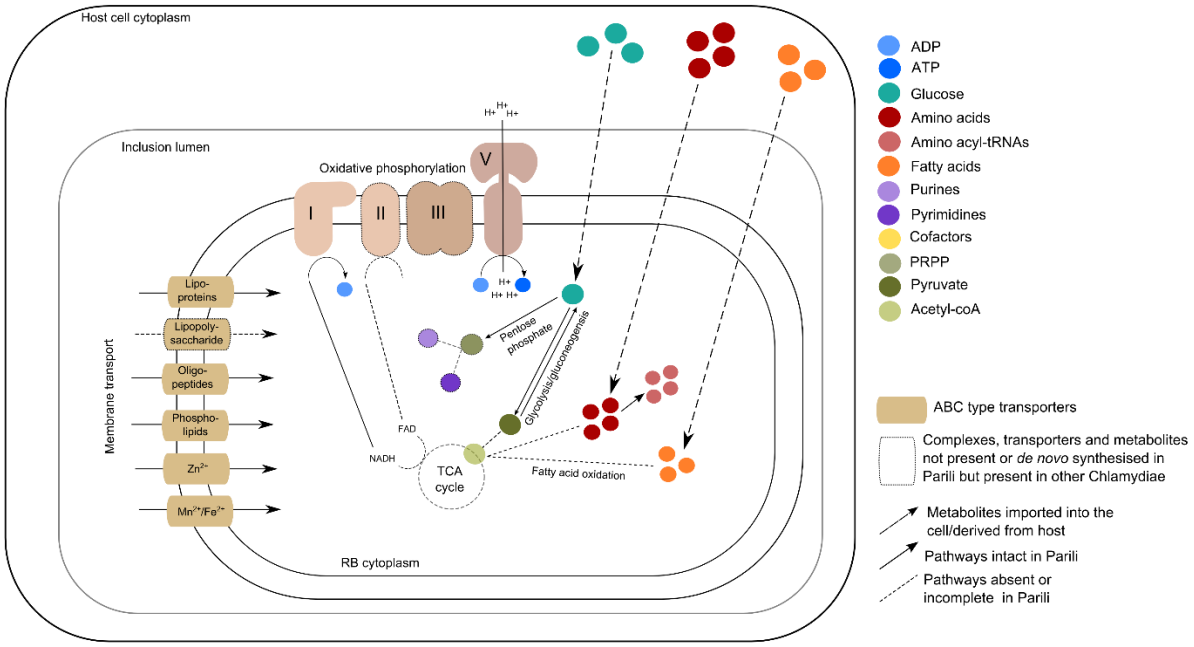


Figure 7.3: *Major metabolic pathways inferred from genome sequences from three species of the* Ca. *Parilichlamydiaceae. Schematic diagram of selected metabolic processes in the reduced genome of* Ca. *Parilichlamydiaceae species. The RB is represented within the inclusion in the host cell cytoplasm. Solid and dashed lines represent predicted or absent paths in the* Ca. *Parilichlamydiaceae, respectively. Substrates and transporters are coloured as per the legend.*

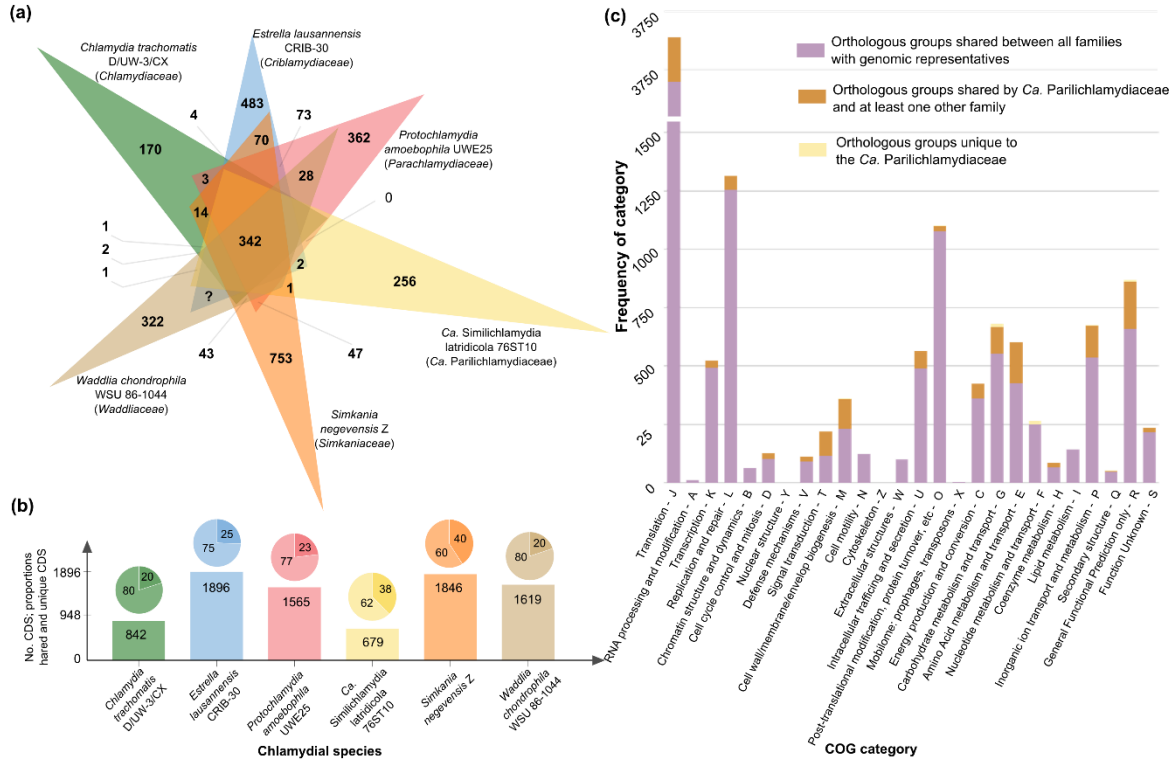


Figure 7.4: Chlamydiales *core and pan-genome analysis. Orthologous groups of proteins were compared between* Ca. *S. latridicola as a representative of the* Ca. *Parilichlamydiaceae, and representatives of five chlamydial families. Protein sequences were clustered using OrthoFinder. (a) Shared and unique orthologous groups are represented by the shaded triangles. (b) Number of coding regions for each genome (bar chart), and proportion (pie chart) of each genome that is unique (dark shading) and shared (light shading). (c) Distribution of Clusters of Orthologous Group categories (COGs) in 342 core orthologous groups. COGs for other orthogroups shared by* Ca. *Parilichlamydiaceae or unique to* Ca. *Parilichlamydiaceae are also shown. See legend for colour scheme.*

References

Akers, J. C. & Tan, M. (2006). Molecular mechanism of tryptophan-dependent transcriptional regulation in *Chlamydia trachomatis*. Journal of bacteriology 188, 4236-43.

Andersson, P., Klein, M., Lilliebridge, R. A. & Giffard, P. M. (2013). Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 19, E405-8.

Andersson, S. G. & Kurland, C. G. (1998). Reductive evolution of resident genomes. Trends in microbiology 6, 263-8.

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., Mcneil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. BMC genomics 9, 75.

Bachmann, N. L., Polkinghorne, A. & Timms, P. (2014). *Chlamydia* genomics: providing novel insights into chlamydial biology. Trends in microbiology 22, 464-472.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 19, 455-77.

Batut, B., Knibbe, C., Marais, G. & Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. Nature reviews. Microbiology 12, 841-50.

Bertelli, C., Aeby, S., Chassot, B., Clulow, J., Hilfiker, O., Rappo, S., Ritzmann, S., Schumacher, P., Terrettaz, C., Benaglio, P., Falquet, L., Farinelli, L., Gharib, W. H., Goesmann, A., Harshman, K., Linke, B., Miyazaki, R., Rivolta, C., Robinson-Rechavi, M., Van Der Meer, J. R. & Greub, G. (2015). Sequencing and characterizing the genome of Estrella lausannensis as an undergraduate project: training students and biological insights. Frontiers in microbiology 6, 101.

Bertelli, C., Cisse, O. H., Rusconi, B., Kebbi-Beghdadi, C., Croxatto, A., Goesmann, A., Collyn, F. & Greub, G. (2016). CRISPR System Acquisition and Evolution of an Obligate Intracellular *Chlamydia*-Related Bacterium. Genome biology and evolution 8, 2376-86.

Bertelli, C., Collyn, F., Croxatto, A., Ruckert, C., Polkinghorne, A., Kebbi-Beghdadi, C., Goesmann, A., Vaughan, L. & Greub, G. (2010). The *Waddlia* genome: a window into chlamydial biology. PloS one 5, e10890.

Bliska, J. B. & Casadevall, A. (2009). Intracellular pathogenic bacteria and fungi--a case of convergent evolution? Nature reviews. Microbiology 7, 165-71.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-20.

Burnard, D., Weaver, H., Gillett, A., Loader, J., Flanagan, C. & Polkinghorne, A. (2017). Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasit Vectors 10, 46.

Casadevall, A. (2008). Evolution of intracellular pathogens. Annual review of microbiology 62, 19-33.

Christiansen, M. T., Brown, A. C., Kundu, S., Tutill, H. J., Williams, R., Brown, J. R., Holdstock, J., Holland, M. J., Stevenson, S., Dave, J., Tong, C. Y., Einer-Jensen, K., Depledge, D. P. & Breuer, J. (2014). Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. BMC infectious diseases 14, 591.

Collingro, A., Kostlbacher, S., Mussmann, M., Stepanauskas, R., Hallam, S. J. & Horn, M. (2017). Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. The ISME journal.

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R. C., Read, T. D., Bavoil, P. M., Sachse, K., Kahane, S., Friedman, M. G., Rattei, T., Myers, G. S. & Horn, M. (2011). Unity in variety--the pan-genome of the *Chlamydiae*. Molecular biology and evolution 28, 3253-70.

Corsaro, D. & Greub, G. (2006). Pathogenic potential of novel *Chlamydiae* and diagnostic approaches to infections due to these obligate intracellular bacteria. Clinical microbiology reviews 19, 283-97.

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. Diseases of aquatic organisms 37, 61-72.

Croxatto, A., Rieille, N., Kernif, T., Bitam, I., Aeby, S., Peter, O. & Greub, G. (2014). Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks and tick-borne diseases 5, 359-65.

Domman, D., Collingro, A., Lagkouvardos, I., Gehre, L., Weinmaier, T., Rattei, T., Subtil, A. & Horn, M. (2014). Massive expansion of ubiquitination-related gene families within the *Chlamydiae*. Molecular biology and evolution.

Douglas, A. E. (1998). Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. Annu Rev Entomol 43, 17-37.

Dugan, J., Rockey, D. D., Jones, L. & Andersen, A. A. (2004). Tetracycline resistance in *Chlamydia* suis mediated by genomic islands inserted into the chlamydial inv-like gene. Antimicrobial agents and chemotherapy 48, 3989-95.

Dutta, C. & Paul, S. (2012). Microbial lifestyle and genome signatures. Curr Genomics 13, 153-62.

Emms, D. M. & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome biology 16, 157.

Essig, A., Heinemann, M., Simnacher, U. & Marre, R. (1997). Infection of Acanthamoeba castellanii by *Chlamydia pneumoniae*. Applied and environmental microbiology 63, 1396-9.

Eugster, M., Roten, C. A. & Greub, G. (2007). Analyses of six homologous proteins of *Protochlamydia amoebophila* UWE25 encoded by large GC-rich genes (lgr): a model of evolution and concatenation of leucine-rich repeats. BMC evolutionary biology 7, 231.

Evans, D. H., Piermarini, P. M. & Choe, K. P. (2005). The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. Physiol Rev 85, 97-177.

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A. & Eddy, S. R. (2015). HMMER web server: 2015 update. Nucleic acids research 43, W30-8.

Fisher, D. J., Fernandez, R. E. & Maurelli, A. T. (2013). *Chlamydia trachomatis* transports NAD via the Npt1 ATP/ADP translocase. Journal of bacteriology 195, 3381-6.

Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J. F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., Van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., Mcdonald, L., Artiach, P., Bowman, C., Garland, S., Fuji, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O. & Venter, J. C. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390, 580-6.

Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic acids research 43, D261-9.

Gil, R., Silva, F. J., Pereto, J. & Moya, A. (2004). Determination of the core of a minimal bacterial gene set. Microbiology and molecular biology reviews 68, 518-37.

Gimenez, G., Bertelli, C., Moliner, C., Robert, C., Raoult, D., Fournier, P. E. & Greub, G. (2011). Insight into cross-talk between intra-amoebal pathogens. BMC genomics 12, 542.

Greub, G. (2009). Parachlamydia acanthamoebae, an emerging agent of pneumonia. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 15, 18-28.

Greub, G., Kebbi-Beghdadi, C., Bertelli, C., Collyn, F., Riederer, B. M., Yersin, C., Croxatto, A. & Raoult, D. (2009). High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PloS one 4, 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. Applied and environmental microbiology 69, 5530-5.

Guevara Soto, M., Vidondo, B., Vaughan, L., Seth-Smith, H. M., Nufer, L., Segner, H., Rubin, J. F. & Schmidt-Posthaus, H. (2016). The emergence of epitheliocystis in the upper Rhone region: evidence for *Chlamydiae* in wild and farmed salmonid populations. Archives of microbiology 198, 315-24.

Heinz, E., Hacker, C., Dean, P., Mifsud, J., Goldberg, A. V., Williams, T. A., Nakjang, S., Gregory, A., Hirt, R. P., Lucocq, J. M., Kunji, E. R. & Embley, T. M. (2014). Plasma membrane-located purine nucleotide transport proteins are key components for host exploitation by microsporidian intracellular parasites. PLoS pathogens 10, e1004547.

Hershberg, R. & Petrov, D. A. (2010). Evidence that mutation is universally biased towards AT in bacteria. PLoS genetics 6, e1001115.

Hildebrand, F., Meyer, A. & Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. PLoS genetics 6, e1001107.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B. C. & Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic acids research 24, 4420-49.

Hokynar, K., Sormunen, J. J., Vesterinen, E. J., Partio, E. K., Lilley, T., Timonen, V., Panelius, J., Ranki, A. & Puolakkainen, M. (2016). *Chlamydia*-Like Organisms (CLOs) in Finnish *Ixodes ricinus* Ticks and Human Skin. Microorganisms 4.

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C. L., Purkhold, U., Fartmann, B., Brandt, P., Nyakatura, G. J., Droege, M., Frishman, D., Rattei, T., Mewes, H. W. & Wagner, M. (2004). Illuminating the evolutionary history of chlamydiae. Science 304, 728-30.

Hutchison, C. A., 3rd, Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z. Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G. & Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. Science 351, aad6253.

Ip, Y. K. & Chew, S. F. (2010). Ammonia production, excretion, toxicity, and defense in fish: a review. Frontiers in physiology 1, 134.

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. & Stephens, R. S. (1999). Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nature genetics 21, 385-9.

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28, 27-30.

Kanehisa, M., Sato, Y. & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. Journal of molecular biology 428, 726-731.

Konig, L., Siegl, A., Penz, T., Haider, S., Wentrup, C., Polzin, J., Mann, E., Schmitz-Esser, S., Domman, D. & Horn, M. (2017). Biphasic Metabolism and Host Interaction of a Chlamydial Symbiont. mSystems 2.

Konstantinidis, K. T. & Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. Proceedings of the National Academy of Sciences of the United States of America 101, 3160-5.

Koonin, E. V. (2000). How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet 1, 99-116.

Kuo, C. H., Moran, N. A. & Ochman, H. (2009). The consequences of genetic drift for bacterial genome complexity. Genome research 19, 1450-4.

Lai, C. C., Crosbie, P. B. B., Battaglene, S. C. & Nowak, B. F. (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 microbiology reviews 34, 260-80.

Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-95.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-9.

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. Annual review of microbiology 60, 327-49.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y. & Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic acids research 45, D200-D203.

Merhej, V., Royer-Carenzi, M., Pontarotti, P. & Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4, 13.

Mira, A., Ochman, H. & Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. Trends in genetics : TIG 17, 589-96.

Mojica, S., Huot Creasy, H., Daugherty, S., Read, T. D., Kim, T., Kaltenboeck, B., Bavoil, P. & Myers, G. S. (2011). Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58. Journal of bacteriology 193, 3690.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. Cell 108, 583-6.

Moran, N. A., Mccutcheon, J. P. & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42, 165-90.

Moran, N. A. & Mira, A. (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome biology 2.

Moran, N. A. & Plague, G. R. (2004). Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14, 627-33.

Moulder, J. W. (1974). Intracellular parasitism: life in an extreme environment. The Journal of infectious diseases 130, 300-6.

Mushegian, A. (1999). The minimal genome concept. Curr Opin Genet Dev 9, 709-14.

Mushegian, A. R. & Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proceedings of the National Academy of Sciences of the United States of America 93, 10268-73.

Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P. & Liebl, U. (2002). An alternative flavin-dependent mechanism for thymidylate synthesis. Science 297, 105-7.

Nishida, H. (2012). Evolution of genome base composition and genome size in bacteria. Frontiers in microbiology 3, 420.

Nowak, B. F. & Lapatra, S. E. (2006). Epitheliocystis in fish. Journal of fish diseases 29, 573-88.

Nunes, A. & Gomes, J. P. (2014). Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 23, 49-64.

Nylund, A., Kvenseth, A. M. & Isdal, E. (1998). A Morphological Study of the Epitheliocystis Agent in Farmed Atlantic Salmon. Journal of Aquatic Animal Health 10, 43-55.

Omsland, A., Sixt, B. S., Horn, M. & Hackstadt, T. (2014). Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. FEMS microbiology reviews 38, 779-801.

Pillonel, T., Bertelli, C., Salamin, N. & Greub, G. (2015). Taxogenomics of the order *Chlamydiales*. International journal of systematic and evolutionary microbiology 65, 1381-93.

Pilloux, L., Aeby, S., Gaumann, R., Burri, C., Beuret, C. & Greub, G. (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. Applied and environmental microbiology 81, 8177-82.

Psomopoulos, F. E., Siarkou, V. I., Papanikolaou, N., Iliopoulos, I., Tsaftaris, A. S., Promponas, V. J. & Ouzounis, C. A. (2012). The *Chlamydiales* pangenome revisited: structural stability and functional coherence. Genes 3, 291-319.

Putman, T. E., Suchland, R. J., Ivanovitch, J. D. & Rockey, D. D. (2013). Culture-independent sequence analysis of *Chlamydia trachomatis* in urogenital specimens identifies regions of recombination and in-patient sequence mutations. Microbiology 159, 2109-17.

Qi, W., Vaughan, L., Katharios, P., Schlapbach, R. & Seth-Smith, H. M. (2016). Host-Associated Genomic Features of the Novel Uncultured Intracellular Pathogen *Ca.* Ichthyocystis Revealed by Direct Sequencing of Epitheliocysts. Genome biology and evolution 8, 1672-89.

Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-2.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics 16, 944-5.

Sachse, K., Laroucau, K., Riege, K., Wehner, S., Dilcher, M., Creasy, H. H., Weidmann, M., Myers, G., Vorimore, F., Vicari, N., Magnino, S., Liebler-Tenorio, E., Ruettger, A., Bavoil, P. M., Hufert, F. T., Rossello-Mora, R. & Marz, M. (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. Systematic and applied microbiology 37, 79-88.

Sakharkar, K. R. & Chow, V. T. (2005). Strategies for genome reduction in microbial genomes. Genome informatics. International Conference on Genome Informatics 16, 69-75.

Sakharkar, K. R., Dhar, P. K. & Chow, V. T. (2004). Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. International journal of systematic and evolutionary microbiology 54, 1937-41.

Seth-Smith, H. M., Dourala, N., Fehr, A., Qi, W., Katharios, P., Ruetten, M., Mateos, J. M., Nufer, L., Weilenmann, R., Ziegler, U., Thomson, N. R., Schlapbach, R. & Vaughan, L. (2016). Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular beta-proteobacteria. The ISME journal 10, 1791-803.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407, 81-6.

Steigen, A., Karlsbakk, E., Plarre, H., Watanabe, K., Overgard, A. C., Brevik, O. & Nylund, A. (2015). A new intracellular bacterium, *Candidatus* Similichlamydia labri sp. nov. (*Chlamydiaceae*) producing epitheliocysts in ballan wrasse, *Labrus bergylta* (Pisces, Labridae). Archives of microbiology 197, 311-8.

Steigen, A., Nylund, A., Karlsbakk, E., Akoll, P., Fiksdal, I. U., Nylund, S., Odong, R., Plarre, H., Semyalo, R., Skar, C. & Watanabe, K. (2013). *'Cand.* Actinochlamydia clariae' gen. nov., sp. nov., a unique intracellular bacterium causing epitheliocystis in catfish (*Clarias gariepinus*) in Uganda. PloS one 8, e66840.

Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V. & Davis, R. W. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282, 754-9.

Stride, M. C., Polkinghorne, A., Miller, T. L., Groff, J. M., Lapatra, S. E. & Nowak, B. F. (2013a). 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*). Applied and environmental microbiology 79, 1590-7.

Stride, M. C., Polkinghorne, A., Miller, T. L. & Nowak, B. F. (2013b). 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). Applied and environmental microbiology 79, 4914-20.

Stride, M. C., Polkinghorne, A. & Nowak, B. F. (2014). Chlamydial infections of fish: diverse pathogens and emerging causes of disease in aquaculture species. Veterinary microbiology 170, 19-27.

Stride, M. C., Polkinghorne, A., Powell, M. D. & Nowak, B. F. (2013c). "*Candidatus* Similichlamydia laticola", a novel *Chlamydia*-like agent of epitheliocystis in seven consecutive cohorts of farmed Australian barramundi, *Lates calcarifer* (Bloch). PloS one 8, e82889.

Taylor-Brown, A., Bachmann, N. L., Borel, N. & Polkinghorne, A. (2016). Culture-independent genomic characterisation of *Candidatus* Chlamydia sanzinia, a novel uncultivated bacterium infecting snakes. BMC genomics 17, 710.

Taylor-Brown, A., Pillonel, T., Bridle, A., Qi, W., Bachmann, N. L., Miller, T. L., Greub, G., Nowak, B., Seth-Smith, H. M., Vaughan, L. & Polkinghorne, A. (2017a). Culture-independent genomics of a novel chlamydial pathogen of fish provides new insight into host-specific adaptations utilized by these intracellular bacteria. Environmental microbiology.

Taylor-Brown, A., Spang, L., Borel, N. & Polkinghorne, A. (2017b). Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus *Chlamydia*. Scientific reports in press.

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*. Pathogens and disease 73, 1-15.

Vasilevsky, S., Stojanov, M., Greub, G. & Baud, D. (2016). Chlamydial polymorphic membrane proteins: regulation, function and potential vaccine candidates. Virulence 7, 11-22.

Viver, T., Orellana, L. H., Hatt, J. K., Urdiain, M., Diaz, S., Richter, M., Anton, J., Avian, M., Amann, R., Konstantinidis, K. T. & Rossello-Mora, R. (2017). The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa. Environmental microbiology.

Voigt, A., Schofl, G., Heidrich, A., Sachse, K. & Saluz, H. P. (2011). Full-length de novo sequence of the *Chlamydophila psittaci* type strain, 6BC. Journal of bacteriology 193, 2662-3.

Voigt, A., Schofl, G. & Saluz, H. P. (2012). The *Chlamydia psittaci* genome: a comparative analysis of intracellular pathogens. PloS one 7, e35097.

Vorimore, F., Hsia, R. C., Huot-Creasy, H., Bastian, S., Deruyter, L., Passet, A., Sachse, K., Bavoil, P., Myers, G. & Laroucau, K. (2013). Isolation of a New *Chlamydia* species from the Feral Sacred Ibis (*Threskiornis aethiopicus*): *Chlamydia ibidis*. PloS one 8, e74823.

Wirz, M., Polkinghorne, A., Dumrese, C., Ziegler, U., Greub, G., Pospischil, A. & Vaughan, L. (2008). Predator or prey? *Chlamydophila abortus* infections of a free-living amoebae, *Acanthamoeba castellani* 9GU. Microbes and infection / Institut Pasteur 10, 591-7.

Wolf, Y. I. & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. Bioessays 35, 829-37.

Wu, Y. W., Simmons, B. A. & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605-7.

Ye, Y. N., Ma, B. G., Dong, C., Zhang, H., Chen, L. L. & Guo, F. B. (2016). A novel proposal of a simplified bacterial gene set and the neo-construction of a general minimized metabolic network. Scientific reports 6, 35082.

CHAPTER 8: General discussion

8.1 Summary of the major findings

The work presented in this thesis has contributed significantly to our understanding of chlamydial taxonomic and host diversity, pathogenesis and evolution.

The major findings of this work were:

1. Chlamydial diversity is far more extensive than previously thought, demonstrated by the detection of several novel *Chlamydiaceae* in reptilian hosts. Specifically, we found an unexpected level of diversity circulating in captive snakes, with novel *C. pneumoniae* strains and two novel *Chlamydia sp.* Plasmid sequences from metagenomes also suggest additional species-level diversity within the genus *Chlamydia* in snakes, further highlighting the potential of reptiles to be hosts of a large and untapped reservoir of diversity within this family.

2. The discovery of another gill-associated chlamydiae from a novel fish host, Orange-spotted grouper, further highlights that unchartered chlamydial diversity extends through the far-reaches of the phylum. Based on my studies and those describing additional novel chlamydial species in this deep-branching clade from each new fish host studied, I hypothesise that even greater diversity is waiting to be uncovered in fish, as well as other marine hosts.

3. As an example of a gill-associated chlamydiae, members of the *Ca.* Parilichlamydiaceae were unexpectedly found to have genomes that are remarkably smaller than other *Chlamydia*-related bacteria, and *Chlamydiaceae*. These species are functionally more similar to their *Chlamydiaceae* counterparts despite their phylogenetic separation, suggesting that convergent evolution has shaped the biology and pathogenesis of this family.

4. Culture-independent metagenomics was demonstrated to be an effective tool for novel pathogen detection and subsequent genomic characterisation. Regardless of the anatomical site or host that the sample was derived from, I showed that this approach was extremely powerful in

providing insight into the identity and biology of previously uncharacterised chlamydial species for which there is neither an established culture system nor reference genome.

8.2 Chlamydial taxonomic diversity revealed in this study

Improvements in surveillance and diagnostics have contributed to the detection and wider recognition of emerging infectious diseases, including chlamydiosis. Additionally, advances in culture-independent molecular techniques such as 16S rRNA gene sequencing and single-cell genomics are driving a deeper understanding of the true level of phylogenetic bacterial diversity, resulting in an abundance of uncultivated microbial species (1, 2). The characterisation of novel families, *Candidatus* families and unclassified, uncultured lineages from a very diverse range of hosts, habitats and niches including amoebae, insects, fish gills, jellyfish guts and aquatic environments (3-8) further underline the idea that the uncultivated taxa probably outweigh the cultivated and classified taxa in this phylum.

I discovered and characterised the genomes of two novel *Chlamydia* species in a) clinically healthy captive snakes that may represent a reptile-only lineage prior to the divergence of *C. pneumoniae* and, b) one novel *Ca.* Parilichlamydiaceae species infecting a novel fish host, representing yet another species-specific "fish chlamydia". In the following sections, I will discuss the chlamydial host and taxonomic diversity uncovered during my PhD, and explore the potential for even more diversity throughout the phylum.

8.2.1 Epidemiological survey of chlamydiae in snakes revealed unexpected levels of chlamydial diversity and a putative reptile-only lineage

I aimed to explore the diversity of uncultivated chlamydiae and discover novel pathogens in the phylum *Chlamydiae* to assess whether analysis of uncultured chlamydial organisms will provide new insight into the biology and evolution of these obligate intracellular pathogens.

A logical place to begin to explore the diversity of uncultured chlamydiae is to screen understudied hosts. Chlamydiosis has been described in both free-ranging and captive amphibian and reptilian hosts including puff adders, boas, chameleons, crocodiles, frogs and tortoises (9-14). While *C. pneumoniae* has been most commonly reported in association with infections in reptiles, several studies also suggested that uncultured *Chlamydia*-like organisms might also infect these cold-blooded hosts (12, 15), and *C. psittaci* has been detected in some reptiles (16).

A lethal case of granulomatous inflammation of the heart, liver and splenopancreas of a horned viper (*Vipera ammodytes ammodytes*) caused by *C. pneumoniae* (17) prompted us to investigate and characterise the diversity of chlamydiae circulating in captive snakes in Switzerland. As described in Chapter 3, qPCR-screening of 29 captive snakes revealed PCR positivity of 21.1% of these animals. These positive samples were then subject to *Chlamydiae* ArrayTube Microarray (AT) species identification (18). This assay carries species-, genus- and family-specific probes, with specific hybridisation patterns characteristic for each (19). Whilst most samples demonstrated hybridisation patterns consistent with *C. pneumoniae,* several could only be identified to the genus level, or were inconclusive altogether. These ambiguous signals prompted us to conduct 16S rRNA sequencing.

Despite this study being small, I observed a remarkable level of diversity with our 16S rRNA sequencing confirming seven novel chlamydial genotypes. Five of these in fact had nucleotide identities of >97.5% with *C. pneumoniae*, and phylogenetic analysis showed they formed their own subclade (Chapter 3), differing from each other by 10 to 17 SNPs. The remaining two genotypes were one *C. muridarum* genotype and one putative novel *Chlamydia sp.*, the latter aligning with earlier and ongoing studies that suggested an involvement of putative novel *Chlamydia sp.* in reptiles (12). The detection of the same genotype in both the cloaca and choana in the same animal and the repeated detection of the same bacteria at different sampling points, suggests that snakes experience prolonged, asymptomatic and systemic infections with these novel chlamydiae. A

similar trend was observed in a recent study where inner organs as well as swab samples from pond sliders free of clinical signs harboured the same *ompA* genotype (15).

Interestingly, while the sampling only entailed choana and cloaca swabs from captive snakes in Switzerland, the snake species are not endemic to that region, but rather are imported from geographically dispersed regions, raising questions over the origins of the bacteria in these hosts: does this diversity reflect global diversity or is this an artefact of the domestication and global trade of these animals from different parts of the world? Since my initial study, other groups around the world have reported similar prevalence and diversity of chlamydiae in a range of reptilian hosts. For instance, chlamydial prevalence was highest in Japanese zoo-held reptiles (over mammals and birds) with both *C. pneumoniae* and *C. psittaci* detected (20), mirroring findings from an earlier Argentinean study that found significant strain-level diversity of *C. pneumoniae* distinct from human strains in snakes and turtles (21). Neither of these studies were directly comparable to mine unfortunately, as I used the 16S rRNA gene and subsequent studies used *ompA* and *rpoB*, respectively. The most recent study in Poland detected several novel clades of *Chlamydia spp.* in not only pharyngeal and cloacal swabs, but also other tissues from tortoises and turtles (15), providing further evidence for this diverse order of animals as a suitable starting point for further investigations into chlamydial diversity. This also suggests that the diversity observed in my study may actually be representative of the global diversity, and the repeated detection of *C. pneumoniae* strains suggest that these are indeed of reptile/animal origin rather than examples of zooanthroponosis. More knowledge on the import and origin of animals in all studies is required to better grasp the origin(s), epidemiology and diversity of reptile-associated chlamydiae.

From the *Chlamydia*-positive choana and cloaca samples (except for the *C. muridarum*-positive sample) described in Chapter 3, I next focussed on metagenomic sequencing (detailed in Chapters 4 and 5) of six samples corresponding to the genotypes with either: a) 16S rRNA gene sequences corresponding to a novel species, b) conflicting 16S rRNA sequence and AT data in

terms of *C. pneumoniae* classification, and c) the highest chlamydial load. This ensured the best possibility of obtaining novel sequence data and good read coverage, to clarify the taxonomic position of these novel chlamydiae and to learn more about their biology in the absence of a cultivable isolate.

In doing so, the genomics analysis confirmed that two of the strains suspected to be related to *C. pneumoniae* were in fact novel chlamydial species, namely *Ca.* C. sanzinia, from a Madagascan tree boa (*Sanzinia madagascariensis volontany*) (Chapter 4) and *Ca.* C. corallus, from an Emerald tree boa (*Corallus batesii*) (Chapter 5). These species are phylogenetically situated between *C. pneumoniae* and *C. pecorum*, with *Ca.* C. corallus representing the most closely related species to *C. pneumoniae*, which, interestingly, has also been detected in reptiles. Four additional metagenomes contained plasmid sequences corresponding to two putative novel species and two strains of *Ca.* C. corallus that I then used to further explore the genetic diversity in these snake populations. Based on their branch lengths, *C. corallus* has seemingly evolved at a similar rate to *C. pneumoniae* since their divergence from their most recent common ancestor, to which *Ca.* C. sanzinia is also closely related (Chapters 4 and 5). This divergence appears quite recent relative to several other closely related species pairs, e.g. *C. avium* and *C. gallinacea.*

Chlamydial plasmid sequences are not an ideal phylogenetic marker because a) not all strains carry the plasmid (22-25), b) not all species carry the plasmid (26, 27) and c) we cannot predict the presence of plasmid in a novel uncultivated pathogen. However, due to the fact that chlamydial plasmids are non-conjugative and have co-evolved with the chromosome (28), the chromosome and plasmid sequence-based phylogenetic relationships (of the plasmid-carrying strains) mirror each other, with a linear relationship between chromosome and plasmid pair-wise patristic distances. This allowed us to estimate phylogenetic relationships from plasmid sequences alone, that were yielded in our metagenomes due to the high ratio of plasmid copy number to the chromosome (~2-10x) (24, 25)

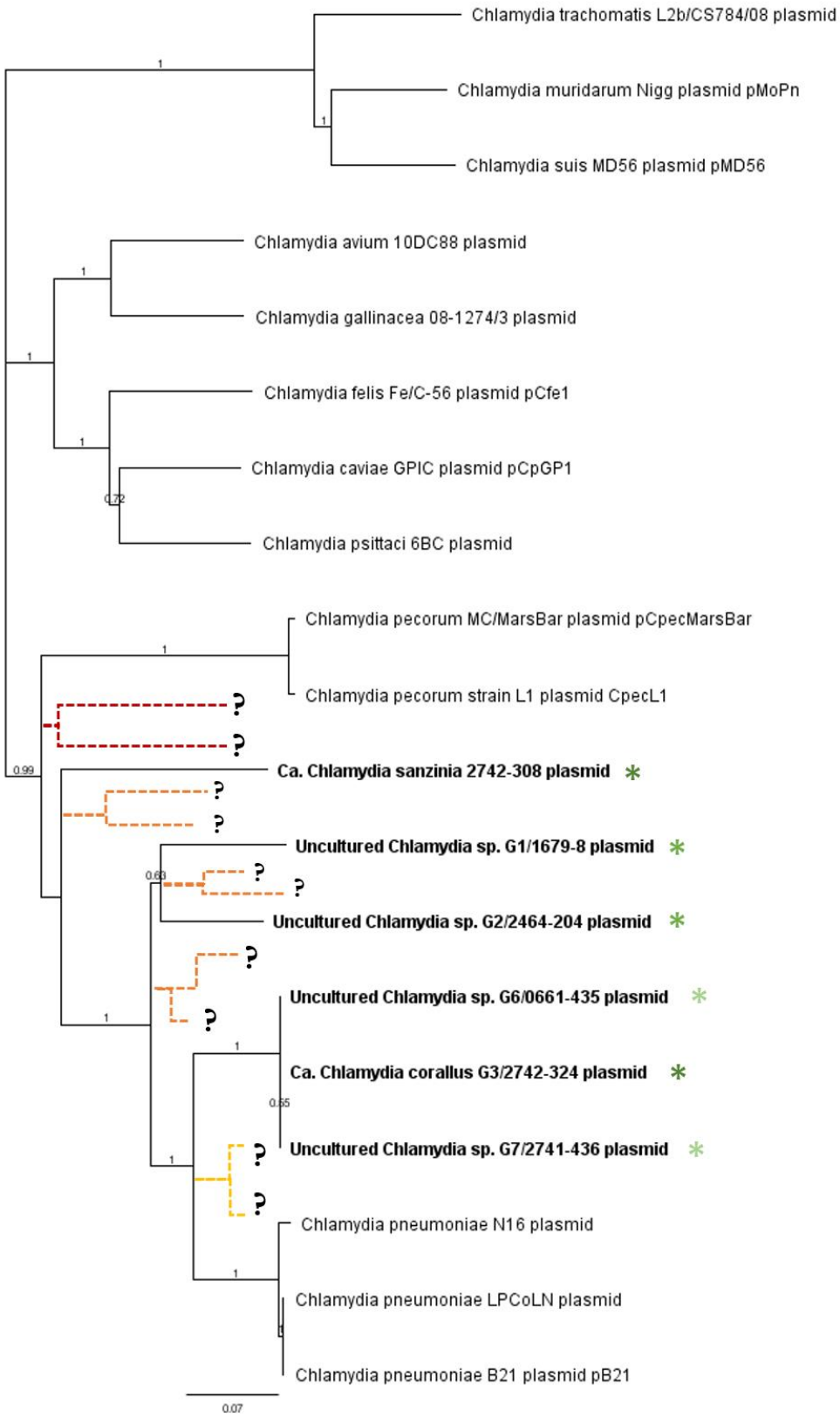


Figure 8.1: Annotated Chlamydiaceae *plasmid sequence-based phylogenetic tree. Dashed lines with question marks represent possibly "missing" taxa. Colours represent different potential divergence points. Asterisks denote novel reptilian* Chlamydia *species.*

and preferential amplification of low molecular weight DNA during the MDA step of the DNA preparation for CIGS. This showed an even higher level of species-level diversity than estimated by the 16S rRNA sequences, with two putative novel species sitting between *Ca.* C. sanzinia and *Ca.* C. corallus, the former of which precede the divergence of *Ca.* C. corallus and *C. pneumoniae*. All members (characterised so far) of this *Ca.* C. sanzinia-*Ca.* C. corallus-*C. pneumoniae* clade harbour a plasmid.

Recent studies have revealed that chlamydial diversity – even within the *Chlamydiaceae* – is vastly underestimated (3, 29-32). Hence, re-examination of phylogenetic relationships as new species are discovered and characterised raises questions over chlamydial evolution: what intermediate species are we missing, and what will *those* species tell us about chlamydial biology? Given a) the identification of two novel *Chlamydia* species confirmed by genome sequencing from two different snake species, b) the detection of two additional unique chlamydial plasmid sequences that may represent two novel species and c) detection of novel *C. pneumoniae* genotypes by 16S rRNA, *ompA* and *rpoB* sequencing, it follows that we may identify even more chlamydial diversity in reptilian hosts. Hence, I have speculated on the phylogenetic placement of putative novel taxa in Figure 8.1. There are over 3000 species of snakes worldwide (33), so we may only be scratching the surface of reptile-associated chlamydial diversity, and *Chlamydiaceae* diversity.

8.2.2 Novel Ca. Parilichlamydiaceae species as host-specific pathogens that broaden the diversity of this phylum

The *Ca.* Parilichlamydiaceae is a recently described family proposed in 2013, a year before commencement of this project. At that time, it consisted of four proposed species from four distinct hosts: Parilichlamydia carangidicola (Yellowtail kingfish), Actinochlamydia clariae (African catfish), Similichlamydia laticola (Barramundi) and Similichlamydia latridicola (Striped trumpeter) (34-37). The Parilichlamydiaceae family has since expanded dramatically (Figure 8.2) with two studies contributing novel taxa comprising *Ca.* Similichlamydia labri from Wrasse in Norway (38),

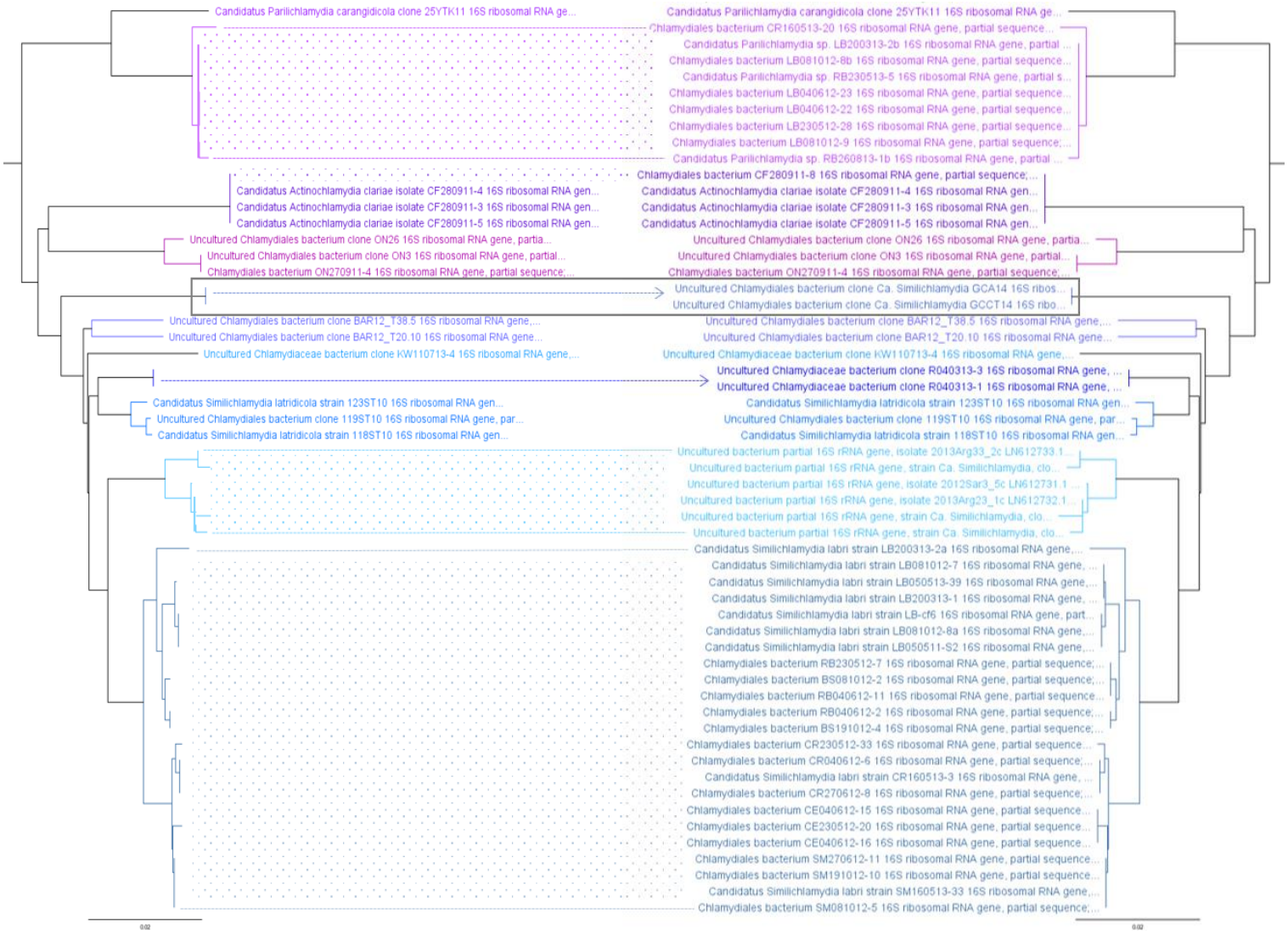


Figure 8.2: The expansion of the Parilichlamydiaceae family over the last four years. The Parilichlamydiaceae family prior to the beginning of this project (March 2014; left) and as updated to August 2017; right. Species-level taxa are denoted by different colours. The boxed species is Ca. Similichlamydia epinephelii, which I discovered during this project infecting Orange-spotted grouper.

a novel species-level *Ca.* Similichlamydia clade comprising sequences from Brown trout and Sea bream in Swiss and Mediterranean waters (39, 40). Finally, I characterised a novel species, *Ca.* Similichlamydia epinephelii, from Orange-spotted grouper in Australia (41).

With a few exceptions, the increasing abundance of CRBs found in association with previously un-investigated fish species affected by epitheliocystis has led to the suggestion that each individual fish species may be affected by novel, unique lineages of *Chlamydiae*. Remarkably, although this fish species (*E. coioides*) was cultured with another species (Giant grouper; *Epinephelus lanceolatus*), only the Orange-spotted grouper remained positive throughout the sampling period, by both molecular methods and histology (Chapter 6), providing further support for *Ca.* Parilichlamydia species as host-specific pathogens.

If member species in the *Ca.* Parilichlamydiaceae each infects a novel host, this would contrast with the other main described chlamydial family infecting fish, namely *Ca.* Piscichlamydiaceae (42). This latter family were first described about 10 years prior, in which a single species, *Ca.* P. salmonis, is known to infect several species (Salmon, Trout and Charr) (43-45), albeit these hosts all belonging to the Salmonidae family. Curiously, a single species in the *Ca.* Clavochlamydiaceae family, *Ca.* C. salmonicola, has a similar host range (45-47), despite being phylogenetically more closely related to the *Chlamydiaceae* than the *Ca.* Piscichlamydiaceae or *Ca.* Parilichlamydiaceae, which are sister clades forming the deepest branch in the *Chlamydiales* phylogenetic tree. Although the *Ca.* Parilichlamydiaceae family members have been described in a wide range of hosts, their discovery also cover a wide geographical range, bringing into question the nature of host-specificity in this clade. Species in this family have been described in waters around the South East and North East coasts of Australia, East Africa, Mediterranean, Alpine and Nordic Europe. It remains to be seen whether the "species-specificity" of this group of chlamydiae is a real phenomenon, or an artefact of sampling bias, which itself may be an artefact of species endemicity to particular regions. Regardless, the Parilichlamydiaceae (as well as the *Ca.*

Piscichlamydiaceae and *Ca.* Clavichlamydiaceae) are restricted to fish hosts only, spanning, at this point, Perciformes, Scorpaeniformes, Siluriformes, and Salmoniformes. Interestingly, no *Ca.* Piscichlamydiaceae or *Ca.* Clavochlamydiaceae have been detected in the Southern hemisphere, again suggesting at least a host family-level geographical specificity in the *Ca.* Parilichlamydiaceae. On this basis, I suspect we will most likely continue to find new species in new hosts and ascertain the true degree of natural host specificity if we a) sample different fish at different locations and b) sample sympatric animals (not just fish) c) sample fish that are cultured together, d) sample wild fish that are supplied to aquaculture or aquarium trade facilities, and e) sample wild animals that are in close proximity to those in (b) and; f) sample fish that do not present with epitheliocystis.

What further complicates the fish host-specificity of this *Ca.* Parilichlamydia-*Ca.* Piscichlamydia clade is the existence of other gill-associated chlamydiae throughout the phylum that are more closely related to other species that are insect or amoebal endosymbionts. For example, two species of *Ca.* Sygnamydia have been described in Salmon (48) and Broad-nosed pipefish (49), within the *Simkaniaceae* family. This family appears to be very diverse in terms of hosts and the associated biology. Hence, an understanding of host-specificity for these pathogens, in the absence of culture systems and a resulting inability to run experimental challenge trials, is only in its infancy.

8.2.3 Is there a limit to chlamydial diversity?

As discussed in Chapter 1, over the last 20 or so years, we have seen an explosion in the detection and description of novel chlamydial species from diverse families referred to as CRBs, due to the fact that many of these species have been isolated from diverse hosts or environments such as ticks, amoebae and water sources. The question still remains: is there a limit to chlamydial diversity?

Figure 8.3 provides a schematic overview of our increasing knowledge of the diversity of chlamydiae over time, which coincided with an increase in diversity as a function of a number of other variables on the x axis that cannot be specifically separated from each other.

First, we are screening a wider range of hosts, from amoebae to fish to wild mammals. Second, this knowledge has been helped along by advances in technology and methods we employ to detect these "novel" species. High throughput methods have allowed us to screen and detect more samples and more species at a time than previous culture-dependent methods, which were more laborious and less sensitive. Third, the increase in diversity has been dependent on the increased need for characterisation of pathogens as demand on sustainable food supply increases.

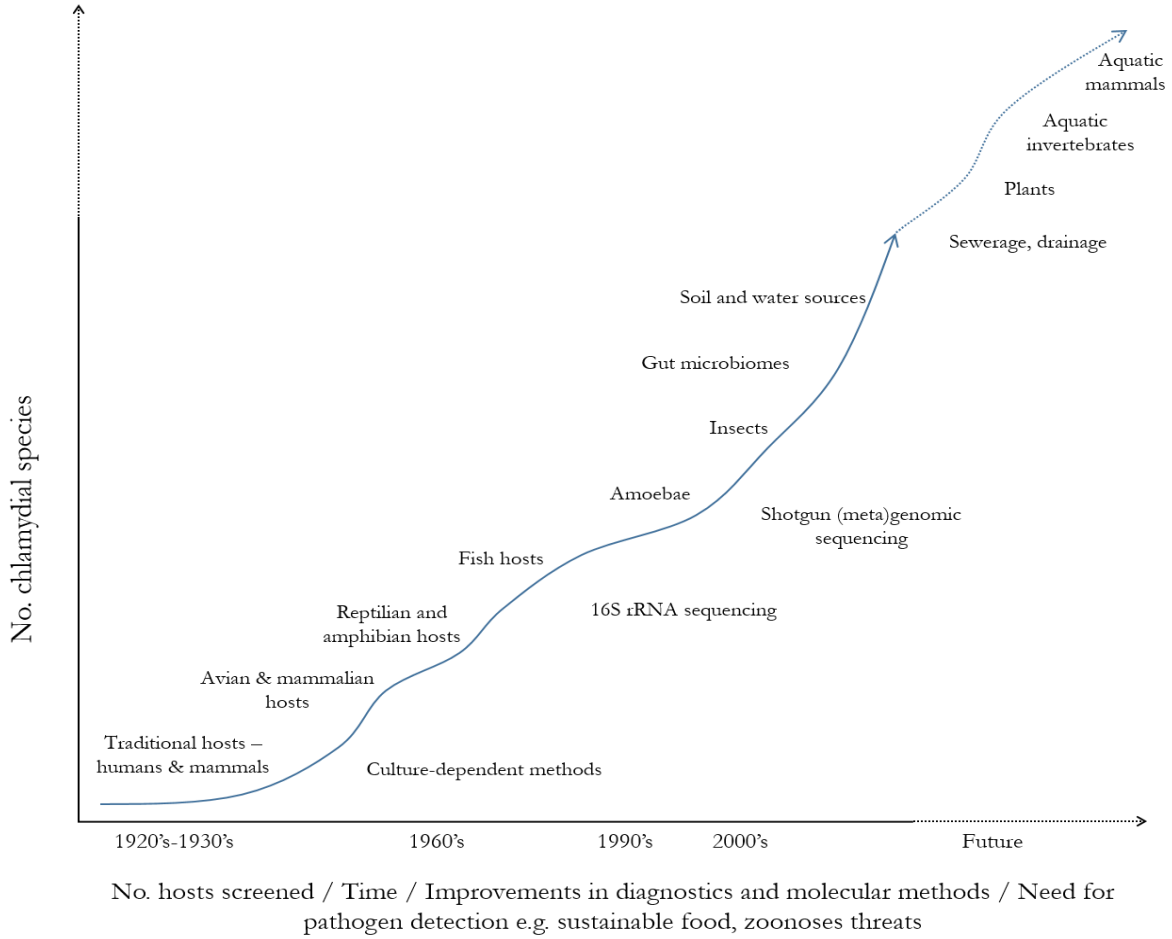


Figure 8.3: Schematic diagram representing the increase in the number of chlamydial species detected and described, in relation to time, number of hosts sampled (above the line) and advances in molecular methods (below

the line). The dashed axes and trend line represent potential future trends based on the "unsampled" hosts and niches.

Researchers were initially and still are interested primarily in chlamydial species that cause disease in humans and animals, such as *C. trachomatis* and *C. psittaci*, so screening was mainly restricted to those hosts. Now, with an increasing rate and risk of zoonotic disease as we encroach onto wildlife habitat, and with the increasing demand on commercial livestock and other sustainable food sources, it is imperative that we characterise the pathogens of these animals, and their potential for spill-over into humans. It follows that as these factors change over time, our awareness of chlamydial diversity will also increase, until it presumably plateaus – that is, if we ever discover the true range of chlamydial diversity. Based on literature searches and examination of public sequence repositories, in the following section, I have proposed a non-exhaustive list of hosts and niches yet to be screened for chlamydiae, which may provide further clues to the origin, evolution and unique biology of the chlamydiae:

1) More wild (terrestrial) mammals and birds

While the overall number of studies are limited compared to studies of humans and domesticated animals, there is growing evidence that chlamydial infections in wild mammals are more widespread than previously thought (50). Wild animals are important in terms of zoonotic transmission, food supply, habitat, reservoirs and vectors transmission.

Increasing reports of new chlamydiae, and new cases of "old" chlamydiae in new hosts (51), suggest that we have only scratched the surface of chlamydial diversity in wildlife. The proposal of three new species isolated from three distinct avian hosts in Europe (26, 32), together with reports of novel unclassified chlamydial species from faecal specimens in wild sea birds such as penguins and gulls [43-46], highlights how little we still know about chlamydial diversity and disease risk in the wild. In particular, for migratory animals such as birds, it is important that we

understand the carriage of the potentially pathogenic species across geographically diverse locations.

Studies investigating wild mammals, particularly those closely related to and/or in close proximity to domesticated livestock species have thus far revealed a breadth of diversity exhibited by the presence of *Chlamydiaceae* and CRBs in animals such as wild boar, water buffalo, yaks and deer (52-55). Further, marsupials other than the koala may represent wild reservoirs of chlamydial diversity but less is known about the prevalence and diversity of chlamydiae in these hosts, with low levels of *C. pecorum*, *C. pneumoniae* and CRBs all detected in bandicoots and possums, among other endangered species (56).

In addition to revealing more about chlamydial host and taxonomic diversity, these studies may also be of benefit to improving native animal conservation by understanding the risks for spill-over between wild and domesticated animals, and predicting new and emerging zoonotic disease threats for humans and animals. Contemporary spill-over studies are lacking, probably due to a lack of sampling of sympatric populations (50), but should provide some insight into a) origin and cross-host transmission of chlamydiae, b) public health initiatives potentially required and c) conservation strategies that could be employed.

2) Aquatic mammals e.g. seals, whales, sea leopards, otters

Published studies on chlamydial infections or diseases of aquatic mammals such as whales, dolphins, seals and otters, are sparse, and limited to potentially non-specific seroprevalence. For instance, antibodies to *C. psittaci* and *C. abortus* were detected in Steller sea lions and Hawaiian monk seals, respectively (57, 58). Further, the seroprevalence of *C. psittaci* was 83-85% in bottlenose dolphins in U.S.A (59). It is unclear from these studies if there were actual infections with *C. psittaci* or *C. abortus* or other close-related species that may assay cross-reactivity. The authors suggested that it was likely that the antibodies were produced in response to exposure from chlamydial shedding from sympatric birds, although the heightened immune response in

Chlamydia-seropositive animals suggests that the infection risk and chlamydial diversity harboured by these hosts is worthy of further study.

Beyond these studies, chlamydial sequences from the nasal cavities of seals and corresponding sea water have been deposited in Genbank, but no paper has yet been published. Interestingly, the majority of these sequences form two major novel clusters (data not shown), but more sequence information will prove vital to understanding the diversity and epidemiology of infecting strains in these hosts and ecological niches. Further investigation of these animals would be of interest to fish-associated chlamydiae studies from the perspective of their close proximity to wild fish, which may represent wild reservoirs of chlamydiae.

3) Terrestrial invertebrates, e.g. flies, beetles, bees, ants, mosquitoes, ticks

Insects are of particular interest given a) their biodiversity, b) their mobility and proximity to humans and other animals and their role as potential mechanical and biological vectors c) their long evolutionary history and d) their increasing introduction as a sustainable food source as the global population expands. Recent studies in arthropods have uncovered a richness of emerging groups of chlamydiae. Molecular studies in Europe, Africa and Australia have recently revealed that ticks harbour CRBs (60-63) that primarily belong to the *Ca.* Rhabdochlamydiaceae, *Parachlamydiaceae,* and *Simkaniaceae,* with *Ca.* Rhabdochlamydiaceae DNA being most abundant, suggesting that ticks are the natural hosts of the latter agents (62). A recent study examining skin biopsies from tick bites identified tick-related chlamydial sequences in the biopsies but not controls [65], highlighting the importance of studies to evaluate the role that these CRBs might have on human and animal health. Thus far, CRBs have been also been detected in cockroaches, white flies, fleas, lice and spiders by molecular methods (8, 64-67). The range of these infections has varied drastically, with suggestions of endosymbiont relationships to overt disease affecting digestive glands, so their impact remains undetermined, but the interest in their role as vectors seems

widespread. Further, their small size and the existence of well-characterised rearing techniques make them ideal models for studying chlamydial infection and transmission in the future.

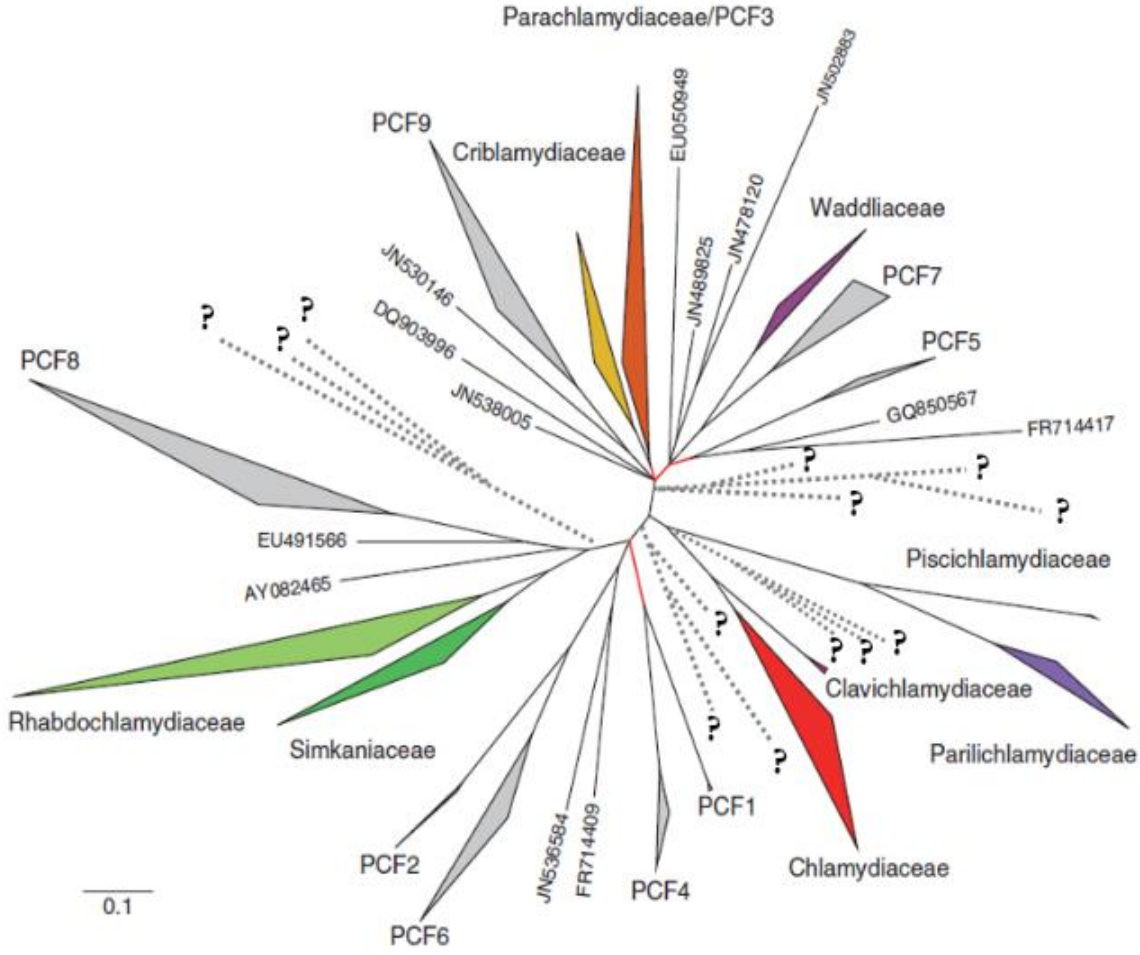


Figure 8.4: Chlamydiales *phylogenetic tree highlighting the hypothetical phylogenetic positions yet to be determined. (grey dashed lines). Adapted from Lagkourvardos et al, 2014.*

4) Aquatic invertebrates e.g. crabs, shrimp, corals, sponges and jellyfish

Aquatic arthropods, like their terrestrial cousins, are another example of potential exceptional host diversity that is underrepresented in terms of chlamydial infections. Screening these hosts for chlamydiae may prove fruitful given their abundance and proximity to fish, and may fill gaps in our knowledge about host specificity, given that there are several clades of chlamydiae that appear to be specific to other arthropods. Further, studies carried out in such

hosts prior to the advent of molecular methods did visualise cysts in these hosts but the phylogenetic classification of these species remains unknown (68-70).

Research groups have only in the last few years begun to characterise the vast biodiversity of our oceans, including the microbial diversity. Corals, which are integral parts of these ecosystems, contain "cell-associated microbial aggregates", which in a recent study employing specific and non-specific staining methods, were suggested to consist of chlamydiae (or rickettsiae), as pathogens or symbionts of at last one species (Acropora) of coral in the indo-pacific region (71). Analysis of an expressed sequence tag library to assess the phylogenomic classification of a deuterosome, *Xenoturbella westabladdi*, identified one chlamydial 16S rRNA sequence (72, 73), and proposed it as a novel *Simkaniaceae* species. Another novel species belonging to the same family was found in the gastric microbiome of the jellyfish, (74). Although the genome sequence was not fully characterised, the authors highlighted the genome size to be comparable to species infecting unicellular hosts and hypothesised a "nested symbioses" for this particular species.

Beyond host diversity lies the question of phylogenetic positions of putative novel species. Culture-independent molecular studies and more recent metagenomic investigations have revealed novel bacterial lineages that are phylogenetically situated between the originally described phyla and phyla with cultivated representatives (2). In a recent review, metagenome-derived sequences increased bacterial diversity by 70% (75). In the *Chlamydiae*, more specifically, nine highly-supported families were predicted based on currently available sequences from a range of environments, and primarily lie between currently described families (3). It follows that this tree will continue to expand upon further investigation of hosts and niches, as we approach saturation, and that novel lineages as further predicted in Figure 8.4 will be uncovered.

The last question is – why do we care? Thus far, microbial diversity is far richer and displays a "greater evolutionary depth" than the other kingdoms/divisions (1), representing the largest proportion of the earth's biodiversity. Given this level of diversity and abundance, microorganisms,

including chlamydiae, are important sources of knowledge about the strategies and limits of many forms of life, whilst representing potential new resources of biotechnological value. The American Society for Microbiology (2016)(76) has recommended several lines of study across several disciplines in order to investigate this diversity in our era of technology, highlighting the need to further define this diversity for applications to bioremediation, development of rapid detection techniques and food security, among others. Both culture collections and genomic studies will be critical in understanding the full extent of microbial diversity. *Chlamydiae*, as an ecologically diverse phylum with cultivated and uncultivated representatives, offer an ideal starting point for querying the environment novel eukaryote hosts for novel chlamydial species that may pose disease risks or answer evolutionary questions.

8.3 Chlamydial biological diversity revealed in this study

My studies to explore chlamydial diversity within the phylum also opened new opportunities to characterise the metabolic and pathogenic capacities of these novel chlamydiae *in lieu* of cultivation to conduct biochemical and cell biological characterisation. These genomic interrogations revealed that a) key biological differences can exist between two closely related novel chlamydial species occupying the same host niche; and b) breaking the trend of observations that CRBs have larger genomes than their traditional counterparts in the *Chlamydiaceae*, chlamydial pathogens of fish in the *Ca.* Parilichlamydiaceae have only small genomes with strong evidence of massive genome reduction similar to the *Chlamydiaceae*.

8.3.1 Chlamydia species infecting closely related hosts do not always share genetic traits

My comparative genomics analysis of *Ca.* C. sanzinia and *Ca.* C. corallus revealed that key genomic differences exist within the plasticity zone (PZ) of these related snake-associated *Chlamydia* species. The PZ is a "hotspot" for diversity between *Chlamydia* and other bacterial species, presumably due to its chromosomal location proximal to the replication terminus (77-79).

The PZ and the genes encoded in this region have been implicated in tissue tropism in *C. trachomatis* (80), where ocular and urogenital serovars are distinguished from each other, and *C. pneumoniae* (81), where non-respiratory isolates cluster together.

None of the putative virulence proteins such as MACPF or cytotoxin are encoded by *Ca.* sanzinia. This is in stark contrast to the PZ harboured by *Ca.* Chlamydia corallus, the arrangement of which is most similar to the human-isolated strain of *C. pneumoniae* AR39. Neither *Ca.* Chlamydia corallus nor *C. pneumoniae* AR39 encodes a cytotoxin, which is present in variable numbers in other species including *C. trachomatis*, *C. psittaci* and *C. pecorum* (82-84) however, each harbours a purine ribonucleotide biosynthesis operon, which is absent from *Ca.* C. sanzinia and the *C. pneumoniae* animal strain LPCoLN, which are the closest relatives of *Ca.* C sanzinia.

These results suggest that genes in this region may not be pertinent to adaption to particular hosts, demonstrated by the fact that the PZs in *Ca.* C. corallus and *Ca.* C. sanzinia have marked differences. Further genomic studies on a) additional strains of *Ca.* C. corallus and *Ca.* C. sanzinia, particularly from diverse hosts, if they exist, and b) additional species capable of infecting snakes, will help confirm this. In support of this, two koala pathogens, *C. pecorum* McMarsBar and *C. pneumoniae* LPCoLN, exhibit major differences in the plasticity zone demonstrated by the differential presence of *tox* and *guaAB-add* (24, 83, 85, 86) and likewise, and there is even greater disparity between the avian pathogens *C. psittaci*, *C. gallinacea* and *C. avium*, which, whilst they share a common ancestor, have markedly different gene content and arrangement in their plasticity zones (32). On the other hand, in the 6BC clade of *C. psittaci*, both the cytotoxin gene – which exhibits high sequence similarity to the EHEC adherence factor and clostridial large cytotoxins – and the *guaAB-add* cluster – implicated in purine ribonucleotide biosynthesis – are encoded, whereas members of the other clades don't encode these genes (77, 78, 82, 87). This clade has a broad host range and is highly virulent, associated with disease in birds and humans and more

recently, horses (88), and it has been suggested that the presence of *tox* in this clade only has facilitated the host range expansion in this lineage (87).

8.3.2 Convergent evolution of chlamydial genomes is influenced by host restriction

Among chlamydiae outside of the *Chlamydiaceae*, there is no region analogous to the PZ upon which to draw comparisons. Hence, I focused my analysis on the reduced metabolic capacity representing a high level of adaption to and dependence on the fish host, which I hypothesised to represent an example of convergent evolution with the *Chlamydiaceae*.

Compared to free-living organisms, an obligate intracellular lifestyle is closely correlated with a reduced genome size (89). Host-associated bacteria, and especially intracellular bacteria, have lost many anabolic genes (89-92), resulting in a highly reduced, non-redundant genome, with a high coding percentage and reduced metabolic capacity compared to free-living bacteria (91, 93-96). These bacteria – parasites and mutualists – offer an example of convergent evolution of diverse phyla toward the intracellular niche. I observed this on a phylum-scale within the *Chlamydiae*.

The four *Ca.* Parilichlamydiaceae species genomes I assembled from gill metagenomes were ~0.77-0.98 Mbp in size, predicted to encode 704 to 943 coding regions (Chapters 6 & 7). This is far smaller than other CRB genomes, which range from 2.1 to 3.5 Mbp with 1986 to 3042 predicted CDSs, and also smaller than the *Chlamydiaceae* genomes (~1.0-1.2 Mbp). The genome reduction in the *Chlamydiaceae* is explained by a reduction in genes for biosynthesis of many nutrients/metabolites/substrates that are presumably able to be imported from the host cell cytoplasm. These include nucleotides, amino acids and lipids, the pathways which are also incomplete in *Chlamydiaceae* species (78, 84, 97, 98) which are also highly adapted to higher eukaryote species, in many cases to a single host (51, 99). Such pathways are generally more complete in amoebae-associated CRB genomes, suggesting a lesser dependence on the host (100-103), or perhaps a certain level of flexibility in the range of hosts a chlamydial species can infect. For example, the *W. chondrophila* genome encodes genes for *de novo* synthesis of at least ten amino

acids (102) and the *P. amoebophila* genome encodes genes for the TCA cycle (103), whereas *C. trachomatis* only encodes genes for two amino acids and the TCA genes are incomplete (97).

With the exception of *W. chondrophila,* the CRBs for which we have genome sequences are associated with a range of amoebal hosts including *Acanthamoeba castellanii, Naegleria gruberi* and *Hartmanella* species (6, 7, 104-108). The relationship of the chlamydial species to the amoebal hosts is still under debate, with most evidence suggesting these species are amoebal endosymbionts (104, 105, 109). Other reports of these species in association with disease states in humans and animals conflict with this hypothesis (110-119). Previously, it has been proposed that the larger genome (2.1-3.5 Mbp) and increased metabolic capacity (more genes for de novo synthesis of nucleotides, lipids, amino acids and carbohydrates, compared to the *Chlamydiaceae*) is to account for a less stable environment in an amoebal host cell (100, 103). I extend this proposal by suggesting that the larger chlamydial genome, including both the chromosome and highly variable plasmid, possessed by the mostly amoebae-associated CRBs, allow for host jumping between multi-celled and single-celled hosts, whilst the reduced genomes of *Chlamydiaceae* and *Ca.* Parilichlamydiaceae do not allow for this flexibility, and are restricted to multi-celled hosts.

Hence, the *Chlamydiaceae* and CRB families have distinct evolutionary histories that have shaped their distinct biological characteristics and genomic dynamics. In other bacteria, acquisition of an intracellular life style is accompanied by a rapid genome shrinkage and loss of genes (91, 93). The genome size and content of *Ca.* Parilichlamydiaceae species indicate that chlamydial obligate intracellular parasitism must have occurred prior to the divergence of the three evolutionarily distinct groups (*Chlamydiaceae*, CRBs and the *Ca.* Parilichlamydiaceae-Piscichlamydiaceae clade). The divergence of *Chlamydiaceae* and *Parachlamydiaceae* is suggested to have occurred over 700 million years ago (103, 120). Phylogenetic analysis of the *Chlamydiae* including *Ca.* Parilichlamydiaceae suggests that the divergence of this family occurred prior to the divergence of

Chlamydiaceae and other CRBs, and that, *Ca.* Parilichlamydiaceae species have been restricted to the gill cell niche to which they are now highly adapted.

8.3.3 Can we predict biology based on host or taxonomy? Can we extrapolate this prediction to the most recent ancestor of the extant members of the phylum?

16S rRNA sequencing and phylogenetic data have provided the framework for characterising chlamydial biological diversity based on chlamydial taxonomic diversity. What will be perhaps far more interesting than the extent of taxonomic and ecological diversity is the biological diversity conferred by adaptation to different niches.

At the start of my PhD, I hypothesised that chlamydiae infecting fish possess large genomes (2-3.5 Mbp) akin to those harboured by CRBs, encoding a comparable number of genes involved in housekeeping, virulence and heightened metabolic capacity. This hypothesis was based on the evolutionary relationships inferred by 16S rRNA sequence data demonstrating that the *Ca.* Parilichlamydiaceae-Piscichlamydiaceae clade diverged prior to the divergence of the *Chlamydiaceae* and CRBs and branches deeper than the other families. The other factor contributing to this hypothesis is the evolutionary age and history of fish themselves, which are estimated to have been on earth for 400-500 million years (121, 122). What I did not take into account for my initial hypothesis was that *because* of the evolutionary history of fish, it is most likely that the chlamydial species associated with them, which are largely host-specific, have co-evolved with those hosts and hence undergone years of genome reduction to eliminate genome redundancy resulting in the draft genomes I have characterised. These results disproved my hypothesis and was the first hint that in *Chlamydiae*, taxonomy is not a good predictor of biology.

Instead, as touched on in the previous section, I propose that we can predict genome size and content, and hence biological features, based on the ability of the species to infect a wide or restricted range of host types, i.e. the ability to infect single-celled hosts such as amoebae in addition to multi-celled hosts including vertebrates and invertebrates, as is observed for several

CRBs such as *S. negevensis* and *Protochlamydia spp.* Other examples of order-level specificity have emerged throughout the phylum, such as arthropod-associated *Ca.* Rhabdochlamydia spp. Interestingly, attempts to culture *Ca.* Fritschea bemisiae, an insect-associated species, in *Acanthamoeba* failed despite successful initial infection (65), providing evidence in support of my hypothesis In further support, mammalian and fish cell lines were shown to be amenable to culture of amoeba-associated CRBs (123, 124), highlighting the enhanced host 'promiscuity' exhibited by these CRBs with larger genomes.

What may throw out this hypothesised relationship between host range and biology are the genomes of fish-associated chlamydiae that belong to families other than the *Ca.* Parilichlamydiaceae, *Ca.* Piscichlamydiaceae and *Ca.* Clavochlamydiaceae, such as *Ca.* Sygnamydia spp in *Simkaniaceae*, which is an increasingly ecologically heterogeneous family. My hypotheses for the roles of such taxa is that those species residing in other clades are not well-adapted to fish hosts, but instead are natural endosymbionts of protists, or aquatic invertebrates, like their close relatives, that have had the opportunity to cross the host barrier into fish, and because of their (hypothesised) genomic capacity for survival in single-celled eukaryotes, are able to colonise the new host. Of course, this is all hypothesis and will require a) genome sequences of these novel chlamydiae; and/or b) model systems to demonstrate their infectivity.

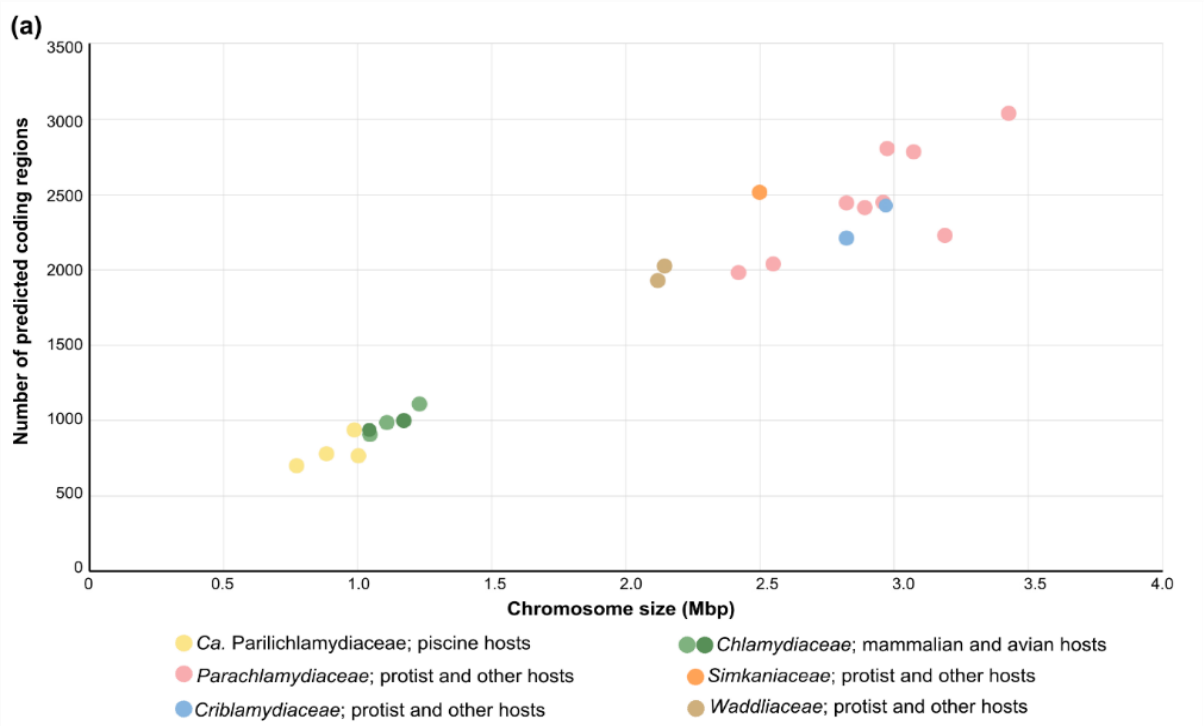


Figure 8.5: The relationship between chromosome size and number of coding regions in the Chlamydiae. *Species are coloured by family; see legend.*

Based on the genomic findings from my PhD project, my review of the literature and some assumptions mentioned above, I predict that the chlamydial ancestor had the following traits (also summarised in Table 8.1):

- Microbial lifestyle: I predict that the chlamydial ancestor was facultatively intracellular. This is based on the enhanced capacity for biosynthesis of nutrients by amoebae-associated chlamydiae (100-103, 125-129).

- Genome size: I suggest that the chlamydial ancestor had a chromosome of at least 3.5 Mbp, based on genomic data from species sequenced to date. This is still relatively small but reflects the proposed level of host-association. From here, through several series of cross-host transmission events and subsequent host adaptation, unnecessary genes were lost and others were duplicated.

Table 8.1: *Predicted genome characteristics of the chlamydial ancestor*

	Range – Chlamydiae	Predicted – chlamydial ancestor
Genome size	0.7- 3.5 Mbp	> 3.5 Mbp
No. CDS	704-	> 3,500
Plasmid size	7.5-150 Kbp	> 150 Kbp
No. CDS on plasmid	7-138	> 140
G+C %	38-42%	> 45%
No. rRNAs	1-4	> 3
No. tRNAs	34-42	< 60
Coding percentage	82-92%	< 85%

- Extra-chromosomal elements: In addition to the large chromosome, I hypothesise that the chlamydial ancestor possessed at least one conjugative >150 Kbp plasmid. Plasmids are reasonably ubiquitous throughout the phylum but are exceptionally variable (100-102, 127, 128, 130). Other researchers have suggested that the chlamydial ancestor harboured a mega-plasmid similar to the *S. negevensis* plasmid, from which the other chlamydial plasmids originated, before losing elements via gene loss or chromosomal integration (100).

- Virulence factors: The recent finding of flagellar components in marine chlamydiae (125), coupled with the flagella-derived T3SS conserved in all chlamydiae sequenced to date (131), leads me to propose that the chlamydial ancestor had flagella, (perhaps in addition to a T3SS).

Future studies in the same vein as my own PhD studies will be required to confirm these hypotheses.

8.4 Metagenomics as a research tool for Chlamydiae

Initial chlamydial "omics" studies resolved several previously held paradigms such as the "peptidoglycan anomaly" (97, 132-134), and allowed us to appreciate the genomic basis underpinning several aspects of the intracellular life style with specific adaptations to a range of hosts. While our group and others have recently used several genome sequencing methods to broaden our knowledge of previously described species in the genus *Chlamydia* [13–17], the current study suggests that a shotgun deep sequencing approach is better suited to novel species. For instance, using "bait" probes designed from a reference genome risks overlooking previously undetected or undescribed features, such as a plasmid [25], and is impossible if the species is novel. Further, a fast, cheap method developed for *C. trachomatis* that relies first on sample collection into appropriate media and second on antibody binding to a previously characterised surface-exposed antigen (135, 136), is also not applicable to novel species. On the other hand, a non-targeted approach, metagenomics, as applied throughout my PhD, has allowed for a less biased exploration of clinical samples for which we are yet to determine the aetiological agent, or for which the aetiological agent cannot be cultured.

Although part of the technology used – multiple displacement amplification – was initially developed to amplify plasmid and phage DNA (137), it has since been applied to a range of sample types including clinical samples and single cells (138-142). In my studies described in this thesis, we have successfully applied it here to a mixed clinical DNA extract from which the host methylated DNA has been depleted. The latter aspect of the DNA preparation is the most vital to these samples, due to the abundance of host DNA in most clinical specimens collected from a host. The combination of these approaches, as employed in similar studies conducted in parallel with mine (40, 143, 144), proved fruitful, as I successfully applied this technology and subsequent bioinformatics pipeline to a) two sample types: swabs and gill tissue, b) two diverse animal hosts: snakes and fish, and c) samples PCR-positive for either *Chlamydia sp.* and CRBs. This allowed me

to obtain and characterise two novel chlamydial species from snakes (Chapters 4 and 5), a novel species from fish (Chapter 6), and characterise the genomes of three previously-described fish-associated chlamydiae (Chapter 7), hence using "omics" studies to further characterise chlamydial diversity and biology.

Metagenomics also allowed me to confirm previous studies using PCR-based interrogation of single gene sequences. In the snake study, what I believed based on AT and 16S rRNA gene sequence data were novel *C. pneumoniae* strains were, in fact, based on chromosome and/or plasmid sequences, novel chlamydial species. The metagenomic data agreed with the AT results, in that the two putative novel species (G1/1679-8 and G2/2464-204) were assigned either to the *Chlamydia* genus level by AT or were inconclusive (Chapters 3 and 5), meaning that the AT assay is a good predictor of novel taxa. Other recent examples where genome sequencing (ie. applying culture-independent metagenomics to clinical samples) could provide phylogenetic clarification as well as additional biological information include a novel *C. psittaci*-like clade specific to crocodiles (145), a novel *C. suis*-related clade found in Roe deer (29) and novel *Chlamydiaceae*-like sequences detected in faeces from sea birds (31, 146). Contrastingly, the genomic data for the newly characterised fish-associated genomes derived from gill metagenomes confirmed their placement in a deeply-rooted clade, earlier postulated by 16S rRNA studies (Chapters 6 and 7).

The anatomical sites I have studied – choana and cloaca of snakes and gills of fish – are ripe with microbiota and other pathogens that, if in higher abundance than the target chlamydial bacteria, may mask complete sequence of the chlamydial genome. Whilst the masking of the chlamydial species is most likely due to the presence of other pathogens, these findings raise the possibility that some of the chlamydial species discovered in this project may be commensal organisms. Notably, there is ongoing speculation around this is in the veterinary chlamydia field, particularly in birds and ruminants. Sachse & Laroucau (2015) suggest that *C. gallinacea* may be a commensal in poultry given its high prevalence and lack of clinical signs, and Reinhold and

colleagues (2011) debated the nature of chlamydial species in livestock given the endemic yet opportunistic nature of these infections.

8.5 Limitations and future directions

8.5.1 Technical limitations of the metagenomic studies

The metagenomic approach taken, although fruitful, came with several limitations, some of which were not apparent prior to embarking on this project.

The use of MDA for whole genome amplification prior to genome sequencing comes with limitations such as preferential amplification of certain genomic regions or particular bacteria in a microbial community (149, 150). The amplification skew is unpredictable and appears to be heavily dependent on both the complexity of the sample and the MDA protocol applied (149)]. These issues appear to be able to be partially overcome by combining MDA with other depletion or enrichment methods (151, 152), which were not attempted in my study.

Expanding on the MDA limitations, a major aspect lacking from my metagenomics studies is the application of the depletion-enrichment method to a mock or spiked sample, in order to identify a threshold at which bacterial genome assembly from these metagenomes would be viable. The metagenomes obtained for both the snake choana and fish gill metagenomes contained hundreds of thousands of host-derived contigs. Importantly, the species richness within the sample itself, which is difficult to assess without running several individual specific PCR assays or conducting microbiome sequencing, seems to be the biggest factor in successful chlamydial genome assembly from these metagenomes.

In Table 2 I have outlined a number of depletion and enrichment combinations, their plausibility for clinical samples, and their pros and cons in terms of cost and effectiveness. Ideally, in the future, I could run a mock sample, in which I co-culture several bacteria with host cells in known relative concentrations, and conduct several DNA treatments in parallel prior to

sequencing and metagenomic assembly. To extend this, I could run serial dilutions of the starting concentrations of each bacteria, plus culture them in different relative concentrations, to obtain thresholds at which the method is effective. Next, I could trial two rounds of depletion, which is the key factor in this process, prior to MDA, whose main purpose is to generate enough DNA for library preparation, in order to increase the effectiveness of methylated DNA depletion. Thirdly, if a DNA sample was split in two, and depletion was conducted on the separate aliquots, then the aliquots were pooled, this may yield enough DNA for sequencing with a theoretical better bacterial genome sequencing outcome, although would be slightly more costly.

A final but important limitation of the use of metagenomics in my studies is that the genomes I have characterised are drafts assembled from metagenomes, meaning they may be incomplete. In the case of the *Ca.* C. sanzinia genome, its assembly into a single contig with overlapping ends evidenced its completeness, whilst *Ca.* C. corallus assembled into six contigs, with predicted contiguity, genomic synteny with closely related species and presence of conserved markers (153). For the fish-associated metagenomes, which assembled into six to 28 chlamydial contigs, the presence of essential genes supports the notion that we have assembled very near-complete genomes. However, we cannot account for regions such as genomic islands (if they exist) that may have assembled distinctly from other chromosomal regions, or plasmids, which exhibit GC contents different from the chromosome and may be differentially covered by reads. Only when we cultivate these organisms to obtain pure DNA can we be sure that our draft genomes are complete and not lacking extra-chromosomal elements. Further, when we have more genomic representatives of other CRBs, in particular for other insect-associated and other gill-associated chlamydiae, can we confidently assess the extent of host adaptation and convergent evolution explored here.

8.5.2 Future directions for chlamydial diversity and genomics studies

There were several aspects not addressed throughout my project that would be ideal topics to explore further or integrate into future studies, including further epidemiological investigations, wider comparative genomics studies and metatranscriptomic studies.

First, ongoing re-assessment of the way we describe the groups of families within the phylum eg. "*Chlamydia*-like, "*Chlamydia*-related", pathogenic, environmental, traditional is required. Not all terms adequately describe the groups of species, and as we detect and characterise more and more taxa, these groupings and descriptions will presumably change. The classification scheme appears to continually need revisions as well. Gupta et al. most recently took steps toward re-classification of the phylum (a topic that has been controversial over the years (108, 154, 155)) into three orders, pending genomic information for the *Ca.* Parilichlamydiaceae and *Ca.* Piscichlamydiaceae families as the most deeply branching families. By this scheme, these families would constitute the third order, *Parilichlamydiales*, in addition to *Parachlamydiales* (CRBs excluding *Ca.* Piscichlamydiaceae and *Ca.* Parilichlamydiaceae) and *Chlamydiales* (*Chlamydiaceae*). This scheme is not yet widely accepted but the ongoing characterisation of chlamydial genomes should clarify these groupings and I suspect this will change in the near future. The *Ca.* Parilichlamydiaceae genomic data generated from my project may spur taxonomists to revise these groupings, which I would be in favour of, based on the diversity among CRBs that is too high to continue to group them together.

The genomic information generated by this and similar studies for diverse members of this phylum can be used for design of multi-gene typing schemes such as MLST which are cheaper than WGS and hence can be applied to a larger group of samples. Such schemes would help to assess the population diversity for certain species and hosts, for instance in the Yellowtail kingfish cohorts which were infected with *Ca.* P. carangidicola over five years (36)(Chapter 7), and the Striped trumpeter and Trout studies which comprised wild and cultured samples (35, 157). This

information, in turn, could then be used to resolve a number of outstanding knowledge gaps over the reservoirs and/or transmission of these novel chlamydiae in aquaculture systems.

It also remains unclear what role microbe-microbe interactions have played in shaping the genomes and strain-level diversity of any number of these reptile-associated and gill-associated lineages. Other chlamydial species such as *C. trachomatis*, *C. psittaci* and *C. suis* exhibit high levels of recombination among strains or have acquired genes from other bacteria (88, 158, 159). Furthermore, recent genus-wide genomics studies revealed extensive levels of admixture between species, demonstrating different levels and origins of gene flow and recombination flux, particularly in *C. psittaci* (159, 160). It is possible that similar incidents have occurred over the evolution of both the *Chlamydia* species and *Ca.* Parilichlamydiaceae species characterised in my studies, especially if they are have co-evolved with their respective hosts and other resident microbes. Microbial communities have been reported to facilitate horizontal gene transfer between closely and distantly related bacteria (161). Reptiles such as snakes and crocodiles harbour a diverse oral and gastrointestinal flora as well many respiratory and gastrointestinal pathogens of bacterial and viral origin that may donate DNA to chlamydial genomes (162-164) or contribute to gene loss and decay by compensating for functions (165).

Next, metatranscriptomic characterisation of a) natural chlamydial infections of fish and snakes, b) natural mixed infections (in either host), c) individual infections *in vitro* and d) mixed infections *in vitro* would help clarify the cross-talk between bacterial species and between bacteria and the host cells in response to different naturally occurring or simulated environmental pressures. In both of my studies, host and environmental factors were not directly addressed, although were speculated on in each chapter. Metatranscriptomics, particularly for uncultivable bacterial pathogens is still in its infancy, however a recent study of *P. amoebophila* cultured in its natural host, *Acanthamoeba castellanii*, revealed the time and developmental stage-specific regulation and expression of the type three secretion system (166).

The final step to properly characterising novel bacteria and proving Koch's postulates for pathogenesis will be to culture these organisms. This will allow us to characterise ultrastructural and biochemical characteristics, and visualise interactions with the host cells, and confirm metabolic needs predicted by the genome. The genomic data generated for uncultivable chlamydiae can hence be used to inform culture attempts. Thus far for chlamydial epitheliocystis agents, all attempts to cultivate the bacteria *in vitro* have failed, mostly due to the use of incorrect (host-free) media (40, 167-169). Otherwise, fish cell lines have been shown to be susceptible to other CRBs (123) and a zebrafish model for *W. chondrophila* infection has been developed (170). Other reasons for culture fail may be due to improper sample storage or inhibition by other bacteria.

8.6 Conclusions

My work has made a significant contribution to broadening our understanding of chlamydial biological diversity by utilising powerful culture-independent techniques to characterise genomes of diverse, novel chlamydial species. The potential for these techniques is almost limitless given the growing number of studies providing additional evidence for uncultured chlamydial diversity, and the vast number and diversity of hosts and environments yet to be explored. The last 20 years of CRB research have totally transformed our understanding chlamydial biology. With this information and techniques in hand, the next 20 years will no doubt be a 'super-revolution'.

Table 8.2: *Advantages and disadvantages of aspects of DNA preparation methods for culture-independent metagenomic sequencing*

Method/combination	Advantages and disadvantages
DNA only	Overabundance of unwanted host DNA/reads – would need far deeper sequencing; No amplification bias.
DNA + MDA only	Bias for most abundant DNA species ie. host = overabundance as above for raw DNA, but more pronounced. Has been used successfully for *C. trachomatis* from swabs (156), but yielded very low coverage with very deep sequencing.
DNA + MDA + Depletion	Same issue as above – if amplification is biased toward most abundant, the depletion may be less effective. Has been used successfully on LCM samples (143), where most of the host DNA has already been removed
DNA + Depletion only	Ideal, as no amplification bias… BUT total amount not enough for library prep & sequencing – would need to repeat, hence increasing cost
DNA + Depletion + MDA*	MDA for purpose of reaching target DNA amount for library prep, with bonus of amplifying most abundant DNA remaining in supernatant (unmethylated DNA = bacterial DNA)
DNA + Depletion + MDA + Depletion + MDA	Might mitigate the issues accompanying each option (rel. abundance, amount) … BUT becomes more costly.
DNA + MDA + Depletion + MDA + Depletion	As above

DNA + carrier DNA	Not necessary!
DNA + MDA + carrier DNA	Same issues as for raw DNA
DNA + Depletion + carrier DNA	Plausible? Would mitigate amplification bias issues from DNA + Depletion + MDA, BUT sequencing depth may still be compromised due to the overabundance of host material.
DNA + Depletion + MDA + carrier DNA	Plausible… BUT unnecessary as the role of the carrier DNA is to reach a target DNA amount which is achieved by MDA – so they cancel each other out.

* Method used in this project

References

1. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765-4774.
2. Rappe MS, Giovannoni SJ. 2003. The uncultured microbial majority. Annu Rev Microbiol 57:369-394.
3. Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M. 2014. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. ISME J 8:115-125.
4. 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 73:1-15.
5. Corsaro D, Michel R, Walochnik J, Muller KD, Greub G. 2010. *Saccamoeba lacustris*, sp. nov. (Amoebozoa: Lobosea: Hartmannellidae), a new lobose amoeba, parasitized by the novel chlamydia 'Candidatus *Metachlamydia* lacustris' (*Chlamydiae*: *Parachlamydiaceae*). Eur J Protistol 46:86-95.
6. Amann R, Springer N, Schonhuber W, Ludwig W, Schmid EN, Muller KD, Michel R. 1997. Obligate intracellular bacterial parasites of acanthamoebae related to *Chlamydia spp*. Appl Environ Microbiol 63:115-121.
7. Birtles RJ, Rowbotham TJ, Storey C, Marrie TJ, Raoult D. 1997. *Chlamydia*-like obligate parasite of free-living amoebae. Lancet 349:925-926.
8. Corsaro D, Thomas V, Goy G, Venditti D, Radek R, Greub G. 2007. *'Candidatus* Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach Blatta orientalis (Insecta: Blattodea). Syst Appl Microbiol 30:221-228.
9. Homer BL, Jacobson ER, Schumacher J, Scherba G. 1994. Chlamydiosis in mariculture-reared green sea turtles (*Chelonia mydas*). Vet Pathol 31:1-7.
10. Bodetti TJ, Jacobson E, Wan C, Hafner L, Pospischil A, Rose K, Timms P. 2002. Molecular evidence to support the expansion of the hostrange of *Chlamydophila pneumoniae* to include reptiles as well as humans, horses, koalas and amphibians. Syst Appl Microbiol 25:146-152.
11. Jacobson E, Origgi F, Heard D, Detrisac C. 2002. Immunohistochemical staining of chlamydial antigen in emerald tree boas (*Corallus caninus*). J Vet Diagn Invest 14:487-494.
12. Soldati G, Lu ZH, Vaughan L, Polkinghorne A, Zimmermann DR, Huder JB, Pospischil A. 2004. Detection of mycobacteria and chlamydiae in granulomatous inflammation of reptiles: a retrospective study. Vet Pathol 41:388-397.
13. Hotzel H, Blahak S, Diller R, Sachse K. 2005. Evidence of infection in tortoises by *Chlamydia*-like organisms that are genetically distinct from known *Chlamydiaceae* species. Vet Res Commun 29 Suppl 1:71-80.
14. Huchzermeyer FW, Langelet E, Putterill JF. 2008. An outbreak of chlamydiosis in farmed Indopacific crocodiles (*Crocodylus porosus*). J S Afr Vet Assoc 79:99-100.
15. Mitura A, Niemczuk K, Zareba K, Zajac M, Laroucau K, Szymanska-Czerwinska M. 2017. Free-living and captive turtles and tortoises as carriers of new *Chlamydia spp*. PLoS One 12:e0185407.

16. Robertson T, Bibby S, O'Rourke D, Belfiore T, Agnew-Crumpton R, Noormohammadi AH. 2010. Identification of Chlamydial species in crocodiles and chickens by PCR-HRM curve analysis. Vet Microbiol 145:373-379.

17. Ruegg SR, Regenscheit N, Origgi FC, Kaiser C, Borel N. 2015. Detection of *Chlamydia pneumoniae* in a collection of captive snakes and response to treatment with marbofloxacin. Vet J 205:424-426.

18. Taylor-Brown A, Ruegg S, Polkinghorne A, Borel N. 2015. Characterisation of *Chlamydia pneumoniae* and other novel chlamydial infections in captive snakes. Vet Microbiol 178:88-93.

19. Ehricht R, Slickers P, Goellner S, Hotzel H, Sachse K. 2006. Optimized DNA microarray assay allows detection and genotyping of single PCR-amplifiable target copies. Mol Cell Probes 20:60-63.

20. Kabeya H, Sato S, Maruyama S. 2015. Prevalence and characterization of *Chlamydia* DNA in zoo animals in Japan. Microbiol Immunol 59:507-515.

21. Frutos MC, Monetti MS, Re VE, Cuffini CG. 2014. Molecular evidence of *Chlamydophila pneumoniae* infection in reptiles in Argentina. Rev Argent Microbiol 46:45-48.

22. Mitchell CM, Hutton S, Myers GS, Brunham R, Timms P. 2010. *Chlamydia pneumoniae* is genetically diverse in animals and appears to have crossed the host barrier to humans on (at least) two occasions. PLoS Pathog 6:e1000903.

23. Myers GS, Mathews SA, Eppinger M, Mitchell C, O'Brien KK, White OR, Benahmed F, Brunham RC, Read TD, Ravel J, Bavoil PM, Timms P. 2009. Evidence that human *Chlamydia pneumoniae* was zoonotically acquired. J Bacteriol 191:7225-7233.

24. Jelocnik M, Bachmann NL, Kaltenboeck B, Waugh C, Woolford L, Speight KN, Gillett A, Higgins DP, Flanagan C, Myers GS, Timms P, Polkinghorne A. 2015. Genetic diversity in the plasticity zone and the presence of the chlamydial plasmid differentiates *Chlamydia pecorum* strains from pigs, sheep, cattle, and koalas. BMC Genomics 16:893.

25. Pickett MA, Everson JS, Pead PJ, Clarke IN. 2005. The plasmids of *Chlamydia trachomatis* and *Chlamydophila pneumoniae* (N16): accurate determination of copy number and the paradoxical effect of plasmid-curing agents. Microbiology 151:893-903.

26. Vorimore F, Hsia RC, Huot-Creasy H, Bastian S, Deruyter L, Passet A, Sachse K, Bavoil P, Myers G, Laroucau K. 2013. Isolation of a New *Chlamydia* species from the Feral Sacred Ibis (Threskiornis aethiopicus): *Chlamydia ibidis*. PLoS One 8:e74823.

27. Thomson NR, Yeats C, Bell K, Holden MT, Bentley SD, Livingstone M, Cerdeno-Tarraga AM, Harris B, Doggett J, Ormond D, Mungall K, Clarke K, Feltwell T, Hance Z, Sanders M, Quail MA, Price C, Barrell BG, Parkhill J, Longbottom D. 2005. The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Res 15:629-640.

28. Seth-Smith HM, Harris SR, Persson K, Marsh P, Barron A, Bignell A, Bjartling C, Clark L, Cutcliffe LT, Lambden PR, Lennard N, Lockey SJ, Quail MA, Salim O, Skilton RJ, Wang Y, Holland MJ, Parkhill J, Thomson NR, Clarke IN. 2009. Co-evolution of genomes and plasmids within *Chlamydia trachomatis* and the emergence in Sweden of a new variant strain. BMC Genomics 10:239.

29. Aaziz R, Vorimore F, Verheyden H, Picot D, Bertin C, Ruettger A, Sachse K, Laroucau K. 2015. Detection of atypical *Chlamydiaceae* in roe deer (*Capreolus capreolus*). Vet Microbiol 181:318-322.

30. Aaziz R, Gourlay P, Vorimore F, Sachse K, Siarkou VI, Laroucau K. 2015. *Chlamydiaceae* in North Atlantic Seabirds Admitted to a Wildlife Rescue Center in Western France. Appl Environ Microbiol 81:4581-4590.

31. Isaksson J, Christerson L, Blomqvist M, Wille M, Alladio LA, Sachse K, Olsen B, González-Acuña D, Herrmann B. 2015. *Chlamydiaceae*-like bacterium, but no *Chlamydia psittaci*, in sea birds from Antarctica. Polar Biology 38:1931-1936.

32. Sachse K, Laroucau K, Riege K, Wehner S, Dilcher M, Creasy HH, Weidmann M, Myers G, Vorimore F, Vicari N, Magnino S, Liebler-Tenorio E, Ruettger A, Bavoil PM, Hufert FT, Rossello-Mora R, Marz M. 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 37:79-88.

33. Uetz P. 2017. The Reptile Database. http://reptile-database.reptarium.cz/advanced_search?taxon=snake&submit=Search. Accessed 25/11/17.

34. Stride MC, Polkinghorne A, Powell MD, Nowak BF. 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:e82889.

35. 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 79:4914-4920.

36. Stride MC, Polkinghorne A, Miller TL, Groff JM, Lapatra SE, Nowak BF. 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 79:1590-1597.

37. Steigen A, Nylund A, Karlsbakk E, Akoll P, Fiksdal IU, Nylund S, Odong R, Plarre H, Semyalo R, Skar C, Watanabe K. 2013. *'Cand.* Actinochlamydia clariae' gen. nov., sp. nov., a unique intracellular bacterium causing epitheliocystis in catfish (*Clarias gariepinus*) in Uganda. PLoS One 8:e66840.

38. Steigen A, Karlsbakk E, Plarre H, Watanabe K, Overgard AC, Brevik O, Nylund A. 2015. A new intracellular bacterium, *Candidatus* Similichlamydia labri sp. nov. (*Chlamydiaceae*) producing epitheliocysts in ballan wrasse, *Labrus bergylta* (Pisces, Labridae). Arch Microbiol 197:311-318.

39. Guevara Soto M, Vidondo B, Vaughan L, Seth-Smith HM, Nufer L, Segner H, Rubin JF, Schmidt-Posthaus H. 2016. The emergence of epitheliocystis in the upper Rhone region: evidence for *Chlamydiae* in wild and farmed salmonid populations. Arch Microbiol 198:315-324.

40. Seth-Smith HM, Dourala N, Fehr A, Qi W, Katharios P, Ruetten M, Mateos JM, Nufer L, Weilenmann R, Ziegler U, Thomson NR, Schlapbach R, Vaughan L. 2016. Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular beta-proteobacteria. ISME J 10:1791-1803.

41. Taylor-Brown A, Pillonel T, Bridle A, Qi W, Bachmann NL, Miller TL, Greub G, Nowak B, Seth-Smith HMB, Vaughan L, Polkinghorne A. 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 19:1899-1913.

42. Draghi A, 2nd, Popov VL, Kahl MM, Stanton JB, Brown CC, Tsongalis GJ, West AB, Frasca S, Jr. 2004. Characterization of "*Candidatus* piscichlamydia salmonis" (order *Chlamydiales*), a *Chlamydia*-like bacterium associated with epitheliocystis in farmed Atlantic salmon (*Salmo salar*). J Clin Microbiol 42:5286-5297.

43. Draghi A, 2nd, Bebak J, Daniels S, Tulman ER, Geary SJ, West AB, Popov VL, Frasca S, Jr. 2010. Identification of *'Candidatus* Piscichlamydia salmonis' in Arctic charr Salvelinus alpinus during a survey of charr production facilities in North America. Dis Aquat Organ 89:39-49.

44. Draghi A, 2nd, Bebak J, Popov VL, Noble AC, Geary SJ, West AB, Byrne P, Frasca S, Jr. 2007. Characterization of a Neochlamydia-like bacterium associated with epitheliocystis in cultured Arctic charr *Salvelinus alpinus*. Dis Aquat Organ 76:27-38.

45. Schmidt-Posthaus H, Polkinghorne A, Nufer L, Schifferli A, Zimmermann DR, Segner H, Steiner P, Vaughan L. 2012. A natural freshwater origin for two chlamydial species, *Candidatus* Piscichlamydia salmonis and *Candidatus* Clavochlamydia salmonicola, causing mixed infections in wild brown trout (*Salmo trutta*). Environ Microbiol 14:2048-2057.

46. Mitchell SO, Steinum T, Rodger H, Holland C, Falk K, Colquhoun DJ. 2010. Epitheliocystis in Atlantic salmon, Salmo salar L., farmed in fresh water in Ireland is associated with *'Candidatus* Clavochlamydia salmonicola' infection. J Fish Dis 33:665-673.

47. Karlsen M, Nylund A, Watanabe K, Helvik JV, Nylund S, Plarre H. 2008. Characterization of *'Candidatus* Clavochlamydia salmonicola': an intracellular bacterium infecting salmonid fish. Environ Microbiol 10:208-218.

48. Nylund S, Steigen A, Karlsbakk E, Plarre H, Andersen L, Karlsen M, Watanabe K, Nylund A. 2015. Characterization of *'Candidatus* Syngnamydia salmonis' (*Chlamydiales*, *Simkaniaceae*), a bacterium associated with epitheliocystis in Atlantic salmon (*Salmo salar* L.). Arch Microbiol 197:17-25.

49. Fehr A, Walther E, Schmidt-Posthaus H, Nufer L, Wilson A, Svercel M, Richter D, Segner H, Pospischil A, Vaughan L. 2013. *Candidatus* Syngnamydia venezia, a novel member of the phylum Chlamydiae from the broad nosed pipefish, *Syngnathus typhle*. PLoS One 8:e70853.

50. Burnard D, Polkinghorne A. 2016. Chlamydial infections in wildlife-conservation threats and/or reservoirs of 'spill-over' infections? Vet Microbiol 196:78-84.

51. Taylor-Brown A, Polkinghorne A. 2017. New and emerging chlamydial infections of creatures great and small. New Microbes New Infect 18:28-33.

52. Pospischil A, Kaiser C, Hofmann-Lehmann R, Lutz H, Hilbe M, Vaughan L, Borel N. 2012. Evidence for *Chlamydia* in wild mammals of the Serengeti. J Wildl Dis 48:1074-1078.

53. Regenscheit N, Holzwarth N, Greub G, Aeby S, Pospischil A, Borel N. 2012. Deer as a potential wildlife reservoir for *Parachlamydia* species. Vet J 193:589-592.

54. Qin SY, Huang SY, Yin MY, Tan QD, Liu GX, Zhou DH, Zhu XQ, Zhou JZ, Qian AD. 2015. Seroprevalence and risk factors of *Chlamydia abortus* infection in free-ranging white yaks in China. BMC Vet Res 11:8.

55. Greco G, Corrente M, Buonavoglia D, Campanile G, Di Palo R, Martella V, Bellacicco AL, D'Abramo M, Buonavoglia C. 2008. Epizootic abortion related to infections by *Chlamydophila abortus* and *Chlamydophila pecorum* in water buffalo (*Bubalus bubalis*). Theriogenology 69:1061-1069.

56. Burnard D, Huston WM, Webb JK, Jelocnik M, Reiss A, Gillett A, Fitzgibbon S, Carver S, Carrucan J, Flanagan C, Timms P, Polkinghorne A. 2017. Molecular evidence of *Chlamydia pecorum* and arthropod-associated *Chlamydiae* in an expanded range of marsupials. Sci Rep 7:12844.

57. Burek KA, Gulland FM, Sheffield G, Beckmen KB, Keyes E, Spraker TR, Smith AW, Skilling DE, Evermann JF, Stott JL, Saliki JT, Trites AW. 2005. Infectious disease and the decline of Steller sea lions (*Eumetopias jubatus*) in Alaska, USA: insights from serologic data. J Wildl Dis 41:512-524.

58. Aguirre AA, Keefe TJ, Reif JS, Kashinsky L, Yochem PK, Saliki JT, Stott JL, Goldstein T, Dubey JP, Braun R, Antonelis G. 2007. Infectious disease monitoring of the endangered Hawaiian monk seal. J Wildl Dis 43:229-241.

59. Bossart GD, Romano TA, Peden-Adams MM, Schaefer A, McCulloch S, Goldstein JD, Rice CD, Fair PA, Cray C, Reif JS. 2014. Clinicoimmunopathologic findings in Atlantic bottlenose dolphins Tursiops truncatus with positive *Chlamydiaceae* antibody titers. Dis Aquat Organ 108:71-81.

60. Burnard D, Weaver H, Gillett A, Loader J, Flanagan C, Polkinghorne A. 2017. Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasit Vectors 10:46.

61. Hokynar K, Sormunen JJ, Vesterinen EJ, Partio EK, Lilley T, Timonen V, Panelius J, Ranki A, Puolakkainen M. 2016. *Chlamydia*-Like Organisms (CLOs) in Finnish *Ixodes ricinus* Ticks and Human Skin. Microorganisms 4.

62. Pilloux L, Aeby S, Gaumann R, Burri C, Beuret C, Greub G. 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 81:8177-8182.

63. Croxatto A, Rieille N, Kernif T, Bitam I, Aeby S, Peter O, Greub G. 2014. Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks Tick Borne Dis 5:359-365.

64. Sixt BS, Kostanjsek R, Mustedanagic A, Toenshoff ER, Horn M. 2013. Developmental cycle and host interaction of *Rhabdochlamydia porcellionis*, an intracellular parasite of terrestrial isopods. Environ Microbiol 15:2980-2993.

65. Everett KD, Thao M, Horn M, Dyszynski GE, Baumann P. 2005. Novel chlamydiae in whiteflies and scale insects: endosymbionts *'Candidatus* Fritschea bemisiae' strain Falk and *'Candidatus* Fritschea eriococci' strain Elm. Int J Syst Evol Microbiol 55:1581-1587.

66. Kostanjsek R, Strus J, Drobne D, Avgustin G. 2004. *'Candidatus* Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). Int J Syst Evol Microbiol 54:543-549.

67. Vanthournout B, Hendrickx F. 2015. Endosymbiont dominated bacterial communities in a dwarf spider. PLoS One 10:e0117297.

68. Cajaraville MP, Angulo E. 1991. *Chlamydia*-like organisms in digestive and duct cells of mussels from the Basque coast. J Invertebr Pathol 58:381-386.

69. Morrison C, Shum G. 1983. Rickettsias in the kidney of the bay scallop, *Argopecten irradians* (Lamarck). Journal of Fish Diseases 6:537-541.

70. Renault T, Cochennec N. 1995. *Chlamydia*-like organisms in ctenidia and mantle cells of the Japanese oyster *Crassostrea ggias* from the French Atlantic coast. Diseases of aquatic organisms 23:153-159.

71. Work TM, Aeby GS. 2014. Microbial aggregates within tissues infect a diversity of corals throughout the Indo-Pacific. Marine Ecology Progress Series 500:1-9.

72. Israelsson O. 2007. Chlamydial symbionts in the enigmatic *Xenoturbella* (Deuterostomia). J Invertebr Pathol 96:213-220.

73. Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum *Xenoturbellida*. Nature 444:85-88.

74. Viver T, Orellana LH, Hatt JK, Urdiain M, Diaz S, Richter M, Anton J, Avian M, Amann R, Konstantinidis KT, Rossello-Mora R. 2017. The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa. Environ Microbiol doi:10.1111/1462-2920.13763.

75. Schulz F, Eloe-Fadrosh EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, Kyrpides NC, Woyke T. 2017. Towards a balanced view of the bacterial tree of life. Microbiome 5:140.

76. Microbiology ASf. 2016. A Working Document for Multi-Agency Consideration - The Unseen National Resource - RECOMMENDATIONS FOR MICROBIAL DIVERSITY RESEARCH. https://www.asm.org/index.php/public-affairs-report/114-unknown/unknown/5046-microbial-diversity-research-priorities-a-working-document-for-multi-agency-consideration. Accessed

77. Read TD, Myers GS, Brunham RC, Nelson WC, Paulsen IT, Heidelberg J, Holtzapple E, Khouri H, Federova NB, Carty HA, Umayam LA, Haft DH, Peterson J, Beanan MJ, White O, Salzberg SL, Hsia RC, McClarty G, Rank RG, Bavoil PM, Fraser CM. 2003. Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the *Chlamydiaceae*. Nucleic Acids Res 31:2134-2147.

78. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28:1397-1406.

79. Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol 1.

80. Carlson JH, Porcella SF, McClarty G, Caldwell HD. 2005. Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. Infect Immun 73:6407-6418.

81. Weinmaier T, Hoser J, Eck S, Kaufhold I, Shima K, Strom TM, Rattei T, Rupp J. 2015. Genomic factors related to tissue tropism in *Chlamydia pneumoniae* infection. BMC Genomics 16:268.

82. Voigt A, Schofl G, Saluz HP. 2012. The *Chlamydia psittaci* genome: a comparative analysis of intracellular pathogens. PLoS One 7:e35097.

83. Bachmann NL, Fraser TA, Bertelli C, Jelocnik M, Gillett A, Funnell O, Flanagan C, Myers GS, Timms P, Polkinghorne A. 2014. Comparative genomics of koala, cattle and sheep strains of *Chlamydia pecorum*. BMC Genomics 15:667.

84. Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW, Stephens RS. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat Genet 21:385-389.

85. Roulis E, Bachmann NL, Myers GS, Huston W, Summersgill J, Hudson A, Dreses-Werringloer U, Polkinghorne A, Timms P. 2015. Comparative genomic analysis of human *Chlamydia pneumoniae* isolates from respiratory, brain and cardiac tissues. Genomics 106:373-383.

86. Mitchell CM, Hovis KM, Bavoil PM, Myers GS, Carrasco JA, Timms P. 2010. Comparison of koala LPCoLN and human strains of *Chlamydia pneumoniae* highlights extended genetic diversity in the species. BMC Genomics 11:442.

87. Knittler MR, Sachse K. 2015. *Chlamydia psittaci*: update on an underestimated zoonotic agent. Pathog Dis 73:1-15.

88. Branley J, Bachmann NL, Jelocnik M, Myers GS, Polkinghorne A. 2016. Australian human and parrot *Chlamydia psittaci* strains cluster within the highly virulent 6BC clade of this important zoonotic pathogen. Sci Rep 6:30019.

89. Casadevall A. 2008. Evolution of intracellular pathogens. Annu Rev Microbiol 62:19-33.

90. Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14:627-633.

91. Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583-586.

92. Bliska JB, Casadevall A. 2009. Intracellular pathogenic bacteria and fungi--a case of convergent evolution? Nat Rev Microbiol 7:165-171.

93. Mushegian A. 1999. The minimal genome concept. Curr Opin Genet Dev 9:709-714.

94. Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A 101:3160-3165.

95. Sakharkar KR, Dhar PK, Chow VT. 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. Int J Syst Evol Microbiol 54:1937-1941.

96. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4:13.

97. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282:754-759.

98. Mojica S, Huot Creasy H, Daugherty S, Read TD, Kim T, Kaltenboeck B, Bavoil P, Myers GS. 2011. Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58. J Bacteriol 193:3690.

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

100. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, Friedman MG, Rattei T, Myers GS, Horn M. 2011. Unity in variety--the pan-genome of the *Chlamydiae*. Mol Biol Evol 28:3253-3270.

101. Bertelli C, Goesmann A, Greub G. 2014. *Criblamydia sequanensis* Harbors a Megaplasmid Encoding Arsenite Resistance. Genome Announc 2.

102. Bertelli C, Collyn F, Croxatto A, Ruckert C, Polkinghorne A, Kebbi-Beghdadi C, Goesmann A, Vaughan L, Greub G. 2010. The *Waddlia* genome: a window into chlamydial biology. PLoS One 5:e10890.

103. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, Brandt P, Nyakatura GJ, Droege M, Frishman D, Rattei T, Mewes HW, Wagner M. 2004. Illuminating the evolutionary history of chlamydiae. Science 304:728-730.

104. Greub G, Raoult D. 2004. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17:413-433.

105. Horn M, Wagner M. 2004. Bacterial endosymbionts of free-living amoebae. J Eukaryot Microbiol 51:509-514.

106. Schuster FL, Visvesvara GS. 2004. Free-living amoebae as opportunistic and non-opportunistic pathogens of humans and animals. Int J Parasitol 34:1001-1027.

107. Horn M, Wagner M, Muller KD, Schmid EN, Fritsche TR, Schleifer KH, Michel R. 2000. *Neochlamydia hartmannellae* gen. nov., sp. nov. (*Parachlamydiaceae*), an endoparasite of the amoeba *Hartmannella vermiformis*. Microbiology 146 (Pt 5):1231-1239.

108. Everett KD, Bush RM, Andersen AA. 1999. Emended description of the order *Chlamydiales*, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol 49 Pt 2:415-440.

109. Greub G, La Scola B, Raoult D. 2003. *Parachlamydia acanthamoeba* is endosymbiotic or lytic for Acanthamoeba polyphaga depending on the incubation temperature. Ann N Y Acad Sci 990:628-634.

110. Baud D, Goy G, Osterheld MC, Croxatto A, Borel N, Vial Y, Pospischil A, Greub G. 2014. Role of *Waddlia chondrophila* placental infection in miscarriage. Emerg Infect Dis 20:460-464.

111. Koschwanez M, Meli M, Vogtlin A, Greub G, Sidler X, Handke M, Sydler T, Kaiser C, Pospischil A, Borel N. 2012. *Chlamydiaceae* family, *Parachlamydia spp*., and *Waddlia spp*. in porcine abortion. J Vet Diagn Invest 24:833-839.

112. Blumer S, Greub G, Waldvogel A, Hassig M, Thoma R, Tschuor A, Pospischil A, Borel N. 2011. Waddlia, Parachlamydia and *Chlamydiaceae* in bovine abortion. Vet Microbiol 152:385-393.

113. Baud D, Goy G, Osterheld MC, Borel N, Vial Y, Pospischil A, Greub G. 2011. *Waddlia chondrophila*: from bovine abortion to human miscarriage. Clin Infect Dis 52:1469-1471.

114. Ruhl S, Goy G, Casson N, Thoma R, Pospischil A, Greub G, Borel N. 2008. *Parachlamydia acanthamoebae* infection and abortion in small ruminants. Emerg Infect Dis 14:1966-1968.

115. Borel N, Ruhl S, Casson N, Kaiser C, Pospischil A, Greub G. 2007. Parachlamydia spp. and related *Chlamydia*-like organisms and bovine abortion. Emerg Infect Dis 13:1904-1907.

116. Barkallah M, Fendri I, Dhieb A, Gharbi Y, Greub G, Gdoura R. 2013. First detection of *Waddlia chondrophila* in Africa using SYBR Green real-time PCR on veterinary samples. Vet Microbiol 164:101-107.

117. Niemi S, Greub G, Puolakkainen M. 2011. *Chlamydia*-related bacteria in respiratory samples in Finland. Microbes Infect 13:824-827.

118. Friedman MG, Kahane S, Dvoskin B, Hartley JW. 2006. Detection of *Simkania negevensis* by culture, PCR, and serology in respiratory tract infection in Cornwall, UK. J Clin Pathol 59:331-333.

119. Kahane S, Greenberg D, Friedman MG, Haikin H, Dagan R. 1998. High prevalence of "Simkania Z," a novel *Chlamydia*-like bacterium, in infants with acute bronchiolitis. J Infect Dis 177:1425-1429.

120. 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 69:5530-5535.

121. Shimeld SM, Donoghue PC. 2012. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development 139:2091-2099.

122. Morris SC, Caron JB. 2014. A primitive fish from the Cambrian of North America. Nature 512:419-422.

123. Kebbi-Beghdadi C, Batista C, Greub G. 2011. Permissivity of fish cell lines to three Chlamydia-related bacteria: *Waddlia chondrophila*, *Estrella lausannensis* and *Parachlamydia acanthamoebae*. FEMS Immunol Med Microbiol 63:339-345.

124. Kebbi-Beghdadi C, Cisse O, Greub G. 2011. Permissivity of Vero cells, human pneumocytes and human endometrial cells to *Waddlia chondrophila*. Microbes Infect 13:566-574.

125. Collingro A, Kostlbacher S, Mussmann M, Stepanauskas R, Hallam SJ, Horn M. 2017. Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. ISME J doi:10.1038/ismej.2017.95.

126. Domman D, Collingro A, Lagkouvardos I, Gehre L, Weinmaier T, Rattei T, Subtil A, Horn M. 2014. Massive expansion of ubiquitination-related gene families within the *Chlamydiae*. Mol Biol Evol doi:10.1093/molbev/msu227.

127. Bertelli C, Cisse OH, Rusconi B, Kebbi-Beghdadi C, Croxatto A, Goesmann A, Collyn F, Greub G. 2016. CRISPR System Acquisition and Evolution of an Obligate Intracellular *Chlamydia*-Related Bacterium. Genome Biol Evol 8:2376-2386.

128. Bertelli C, Aeby S, Chassot B, Clulow J, Hilfiker O, Rappo S, Ritzmann S, Schumacher P, Terrettaz C, Benaglio P, Falquet L, Farinelli L, Gharib WH, Goesmann A, Harshman K, Linke B, Miyazaki R, Rivolta C, Robinson-Rechavi M, van der Meer JR, Greub G. 2015. Sequencing and characterizing the genome of *Estrella lausannensis* as an undergraduate project: training students and biological insights. Front Microbiol 6:101.

129. Greub G, Kebbi-Beghdadi C, Bertelli C, Collyn F, Riederer BM, Yersin C, Croxatto A, Raoult D. 2009. High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PLoS One 4:e8423.

130. Bou Khalil JY, Benamar S, Baudoin JP, Croce O, Blanc-Tailleur C, Pagnier I, Raoult D, La Scola B. 2016. Developmental Cycle and Genome Analysis of "*Rubidus massiliensis*," a New *Vermamoeba vermiformis* Pathogen. Front Cell Infect Microbiol 6:31.

131. Betts-Hampikian HJ, Fields KA. 2010. The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. Front Microbiol 1:114.

132. Moulder JW. 1991. Interaction of chlamydiae and host cells in vitro. Microbiol Rev 55:143-190.

133. Albrecht M, Sharma CM, Reinhardt R, Vogel J, Rudel T. 2010. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res 38:868-877.

134. Hatch TP. 1996. Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae? J Bacteriol 178:1-5.

135. 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 159:2109-2117.

136. Seth-Smith HM, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C, Parmar S, Pitt R, Baker S, Ison CA, Marsh P, Jalal H, Lewis DA, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2013. Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. Genome Res 23:855-866.

137. Dean FB, Nelson JR, Giesler TL, Lasken RS. 2001. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11:1095-1099.

138. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS. 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 99:5261-5266.

139. McLean JS, Lombardo MJ, Badger JH, Edlund A, Novotny M, Yee-Greenbaum J, Vyahhi N, Hall AP, Yang Y, Dupont CL, Ziegler MG, Chitsaz H, Allen AE, Yooseph S, Tesler G, Pevzner PA, Friedman RM, Nealson KH, Venter JC, Lasken RS. 2013. Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. Proc Natl Acad Sci U S A 110:E2390-2399.

140. Lasken RS. 2012. Genomic sequencing of uncultured microorganisms from single cells. Nat Rev Microbiol 10:631-640.

141. Lasken RS. 2009. Genomic DNA amplification by the multiple displacement amplification (MDA) method. Biochem Soc Trans 37:450-453.

142. Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc Natl Acad Sci U S A 104:11889-11894.

143. 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 8:1672-1689.

144. Katharios P, Seth-Smith HM, Fehr A, Mateos JM, Qi W, Richter D, Nufer L, Ruetten M, Guevara Soto M, Ziegler U, Thomson NR, Schlapbach R, Vaughan L. 2015. Environmental marine pathogen isolation using mesocosm culture of sharpsnout seabream: striking genomic and morphological features of novel *Endozoicomonas sp.* Sci Rep 5:17609.

145. Sariya L, Kladmanee K, Bhusri B, Thaijongrak P, Tonchiangsai K, Chaichoun K, Ratanakorn P. 2015. Molecular evidence for genetic distinctions between *Chlamydiaceae* detected in Siamese crocodiles (*Crocodylus siamensis*) and known *Chlamydiaceae* species. Jpn J Vet Res 63:5-14.

146. Christerson L, Blomqvist M, Grannas K, Thollesson M, Laroucau K, Waldenstrom J, Eliasson I, Olsen B, Herrmann B. 2010. A novel *Chlamydiaceae*-like bacterium found in faecal specimens from sea birds from the Bering Sea. Environ Microbiol Rep 2:605-610.

147. Sachse K, Laroucau K. 2015. Two more species of *Chlamydia*-does it make a difference? Pathog Dis 73:1-3.

148. Reinhold P, Sachse K, Kaltenboeck B. 2011. *Chlamydiaceae* in cattle: commensals, trigger organisms, or pathogens? Vet J 189:257-267.

149. Yilmaz S, Allgaier M, Hugenholtz P. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat Methods 7:943-944.

150. Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, Holland T, Cotton D, Hauser L, Keller M. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl Environ Microbiol 73:3205-3214.

151. Seth-Smith HM, Harris SR, Scott P, Parmar S, Marsh P, Unemo M, Clarke IN, Parkhill J, Thomson NR. 2013. Generating whole bacterial genome sequences of low-abundance species from complex samples with IMS-MDA. Nat Protoc 8:2404-2412.

152. Christiansen MT, Brown AC, Kundu S, Tutill HJ, Williams R, Brown JR, Holdstock J, Holland MJ, Stevenson S, Dave J, Tong CY, Einer-Jensen K, Depledge DP, Breuer J. 2014. Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. BMC Infect Dis 14:591.

153. 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.

154. Schachter J, Stephens RS, Timms P, Kuo C, Bavoil PM, Birkelund S, Boman J, Caldwell H, Campbell LA, Chernesky M, Christiansen G, Clarke IN, Gaydos C, Grayston JT, Hackstadt T, Hsia R, Kaltenboeck B, Leinonnen M, Ojcius D, McClarty G, Orfila J, Peeling R, Puolakkainen M, Quinn TC, Rank RG, Raulston J, Ridgeway GL, Saikku P, Stamm WE, Taylor-Robinson DT, Wang SP, Wyrick PB. 2001. Radical changes to chlamydial taxonomy are not necessary just yet. Int J Syst Evol Microbiol 51:249; author reply 251-243.

155. Stephens RS, Myers G, Eppinger M, Bavoil PM. 2009. Divergence without difference: phylogenetics and taxonomy of *Chlamydia* resolved. FEMS Immunol Med Microbiol 55:115-119.

156. 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:E405-408.

157. Guevara Soto M, Vaughan L, Segner H, Wahli T, Vidondo B, Schmidt-Posthaus H. 2016. Epitheliocystis Distribution and Characterization in Brown Trout (*Salmo trutta*) from the Headwaters of Two Major European Rivers, the Rhine and Rhone. Front Physiol 7:131.

158. Joseph SJ, Marti H, Didelot X, Read TD, Dean D. 2016. Tetracycline Selective Pressure and Homologous Recombination Shape the Evolution of *Chlamydia suis*: A Recently Identified Zoonotic Pathogen. Genome Biol Evol 8:2613-2623.

159. Joseph SJ, Read TD. 2012. Genome-wide recombination in *Chlamydia trachomatis*. Nat Genet 44:364-366.

160. Joseph SJ, Marti H, Didelot X, Castillo-Ramirez S, Read TD, Dean D. 2015. *Chlamydiaceae* Genomics Reveals Interspecies Admixture and the Recent Evolution of *Chlamydia abortus* Infecting Lower Mammalian Species and Humans. Genome Biol Evol 7:3070-3084.

161. Stecher B, Maier L, Hardt WD. 2013. 'Blooming' in the gut: how dysbiosis might contribute to pathogen evolution. Nat Rev Microbiol 11:277-284.

162. Shilton CM, Jerrett IV, Davis S, Walsh S, Benedict S, Isberg SR, Webb GJ, Manolis C, Hyndman TH, Phalen D, Brown GP, Melville L. 2016. Diagnostic investigation of new

disease syndromes in farmed Australian saltwater crocodiles (*Crocodylus porosus*) reveals associations with herpesviral infection. J Vet Diagn Invest 28:279-290.

163. Costello EK, Gordon JI, Secor SM, Knight R. 2010. Postprandial remodeling of the gut microbiota in Burmese pythons. ISME J 4:1375-1385.

164. Keenan SW, Engel AS, Elsey RM. 2013. The alligator gut microbiome and implications for archosaur symbioses. Sci Rep 3:2877.

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

166. Konig L, Siegl A, Penz T, Haider S, Wentrup C, Polzin J, Mann E, Schmitz-Esser S, Domman D, Horn M. 2017. Biphasic Metabolism and Host Interaction of a Chlamydial Symbiont. mSystems 2.

167. Hoffman GL, Dunbar CE, Wolf K, Zwillenberg LO. 1969. Epitheliocystis, a new infectious disease of the bluegill (*Lepomis macrochirus*). Antonie Van Leeuwenhoek 35:146-158.

168. Kumar G, Mayrhofer R, Soliman H, El-Matbouli M. 2013. Novel Chlamydiales associated with epitheliocystis in grass carp (*Ctenopharyngodon idella*). Vet Rec 172:47.

169. Venizelos DT, Lou W, Charalampopoulos TT. 1996. Development of an algorithm for the calculation of the scattering properties of agglomerates. Appl Opt 35:542-548.

170. Fehr AG, Ruetten M, Seth-Smith HM, Nufer L, Voegtlin A, Lehner A, Greub G, Crosier PS, Neuhauss SC, Vaughan L. 2016. A Zebrafish Model for *Chlamydia* Infection with the Obligate Intracellular Pathogen *Waddlia chondrophila*. Front Microbiol 7:1829.