UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2040

Expanding the Chlamydiae tree

Insights into genome diversity and evolution

JENNAH E. DHARAMSHI

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2021

ISSN 1651-6214
ISBN 978-91-513-1203-3
urn:nbn:se:uu:diva-439996

Dissertation presented at Uppsala University to be publicly examined in A1:111a, Biomedical Centre (BMC), Husargatan 3, Uppsala, Tuesday, 8 June 2021 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Prof. Dr. Alexander Probst (Faculty of Chemistry, University of Duisburg-Essen).

Abstract

Dharamshi, J. E. 2021. Expanding the Chlamydiae tree. Insights into genome diversity and evolution. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2040. 87 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1203-3.

Chlamydiae is a phylum of obligate intracellular bacteria. They have a conserved lifecycle and infect eukaryotic hosts, ranging from animals to amoeba. Chlamydiae includes pathogens, and is well-studied from a medical perspective. However, the vast majority of chlamydiae diversity exists in environmental samples as part of the uncultivated microbial majority.

Exploration of microbial diversity in anoxic deep marine sediments revealed diverse chlamydiae with high relative abundances. Using genome-resolved metagenomics various marine sediment chlamydiae genomes were obtained, which significantly expanded genomic sampling of Chlamydiae diversity. These genomes formed several new clades in phylogenomic analyses, and included Chlamydiaceae relatives. Despite endosymbiosis-associated genomic features, hosts were not identified, suggesting chlamydiae with alternate lifestyles.

Genomic investigation of Anoxychlamydiales, newly described here, uncovered genes for hydrogen metabolism and anaerobiosis, suggesting they engage in syntrophic interactions. Anaerobic metabolism is found across modern eukaryotes, and syntrophic hydrogen exchange is central in many hypotheses for eukaryotic evolution, but its origin is unknown. Chlamydial and eukaryotic homologs were the closest relatives in several of these gene phylogenies, providing evidence for a chlamydial contribution of these genes during eukaryotic evolution.

Gene-tree aware ancestral-state-reconstruction revealed a fermentative, mobile, facultatively anaerobic Chlamydiae ancestor, which was capable of endosymbiosis. Examination of Chlamydiae gene content evolution indicated complex dynamics, with a central role of horizontal gene transfer in major evolutionary transitions, related to energy metabolism and aerobiosis. Furthermore, chlamydiae have evolved through genome expansion in addition to gene loss, counter to many other obligate endosymbionts.

Sponge microbiome-associated chlamydiae were found in high relative abundance in some sponge species. Genome-resolved metagenomics identified diverse, yet co-associating chlamydial lineages, with distinctive genetic repertoires, including unexpected degradative and biosynthetic potential. Biosynthetic gene clusters were found across Chlamydiae, suggestive of secondary metabolite production and host-defence roles. Surveying environmental prevalence indicated wider associations between chlamydiae and marine invertebrates.

Finally, a wide-scale assessment of chlamydiae genetic contributions to eukaryotic evolution was performed. Over 100 distinct Chlamydiae-eukaryotic clades were identified in phylogenies across shared protein families. Although patterns are complex and direction of transfers often unclear, our results indicate larger avenues of chlamydial gene exchange with both plastid-bearing eukaryotes, and the last eukaryotic common ancestor.

In summary, in this thesis, cultivation-independent methods and evolutionary-driven investigations were used to expand the Chlamydiae tree, and to provide new insights into genomic diversity and evolution of the phylum.

Keywords: PVC superphylum, Chlamydiae, chlamydia, intracellular, symbiosis, endosymbiont, pathogen, marine sediment, sponge microbiome, metagenomics, uncultured microbial diversity, phylogenomics, microbial evolution, eukaryote evolution

Jennah E. Dharamshi, Department of Cell and Molecular Biology, Molecular Evolution, Box 596, Uppsala University, SE-752 37 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1203-3
urn:nbn:se:uu:diva-439996 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-439996)

To my wonderful friends and family
wherever you are in the world.

To the microbes that rule us all.

And to the 4 letters that make
TTGATCTTCGAG possible

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Dharamshi, J. E.**, Tamarit, D.*, Eme, L.*, Stairs, C. W., Martijn, J., Homa, F., Jørgensen, S. L., Spang, A., Ettema, T. J. G. (2020) Marine Sediments illuminate Chlamydiae diversity and evolution. *Current Biology*, 30(6):1032–1048 e7

II Stairs, C. W.*, **Dharamshi, J. E.***, Tamarit, D, Eme, L., Jørgensen, S. L., Spang, A., Ettema, T. J. G. (2020) Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. *Science Advances*, 6(35): eabb7258

III **Dharamshi, J.E.***, Köstlbacher, S.*, Schön, M. E., Collingro A., Ettema, T. J. G*., Horn, M* (2021) Gain of symbiotic traits underpins evolutionary transitions across the phylum Chlamydiae. *Manuscript*

IV **Dharamshi, J.E.,** Gaarselv N. G., Steffen K., Martin, T., Sipkema, D., Ettema, T.J.G. (2021) Marine sponges harbour novel and diverse chlamydial lineages. *Manuscript*

V Tamarit, D.*, **Dharamshi, J.E.***, Eme L., Ettema, T. J. G. (2021) Chlamydial genetic contribution to eukaryotic evolution. *Manuscript*

(*) Equal contribution

Contents

Abbreviations

ASR	Ancestral state reconstruction
ATP	Adenosine triphosphate
COG	Clusters of orthologous groups
CPR	Candidate phyla radiation
DNA	Deoxyribonucleic acid
DPANN	Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota
EB	Elementary body
ETC	Electron transport chain
FCB	Fibrobacteres, Chlorobi, and Bacteroidetes
Ga	Billion years
GC	Guanine-Cytosine
HGT	Horizontal gene transfer
LBA	Long branch attraction
LECA	Last eukaryotic common ancestor
LUCA	Last universal common ancestor
MAG	Metagenome-assembled genome
MAT	Ménage -à -trois
Mb	Mega base pairs
mbsf	Meters below sea floor
ML	Maximum-likelihood
MRO	Mitochondrion-related organelle
MSA	Multiple sequence alignment
NTT	Nucleotide transporter
OTU	Operational taxonomic unit
PVC	Planctomycetes, Verrucomicrobia, and Chlamydiae
RB	Reticulate body
SAG	Single amplified genome
SSU rRNA	Small subunit ribosomal ribonucleic acid
T3SS	Type III secretion system
TACK	Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota
TCA	Tricarboxylic acid

Introduction to this thesis

Illustrated on the cover of this thesis is an artistic representation of the tree of life. Underpinning this tree are the genomes of organisms that exist today. Our knowledge of life's evolutionary history is grounded and built from this genetic data. I find it fascinating that we can use this information to not only inspect an organism's present lifestyle and ecology but also to probe its past. However, our view on this tree of life is far from a complete picture as many branches are still hidden in the dark. This thesis brings some of these branches to light by expanding Chlamydiae genomic diversity. Although these chlamydiae are uncultivated, we could gain insight into their lifestyles and evolution by exploring their genomes.

In this thesis I will describe our efforts to further understand chlamydiae diversity and evolution. Using cultivation-independent approaches we have discovered previously unknown chlamydiae groups, thereby illuminating chlamydial branches on the tree of life. We have been able to learn about the lives of these chlamydiae from diverse environments and identify unexpected metabolic genes. We also reconstructed the ancestors of chlamydiae and studied their evolution. Our findings also challenge whether all chlamydiae share the same lifestyle, while conversely strengthening their label as symbionts. With this expanded chlamydiae repertoire we were also able to identify chlamydial genetic contributions to eukaryotic evolution.

These findings are outlined in the summary text preceding the five papers that compose this thesis. Due to format constraints some supplementary material is not shown here, and can instead be found using the electronic links. Also provided in the following summary text is an overview of historical and current knowledge of chlamydial ecology, diversity, and cell biology, alongside prior research findings on chlamydiae and eukaryote evolution.

While the focus of this thesis is Chlamydiae, it also intersects with other diverse topics that warrant introduction. In the following summary text I have also outlined background information that is of relevance across the five papers. This includes background information on the tree of life and efforts to improve our sampling of it, in addition to historical events of relevance for eukaryotic and chlamydial evolution. The evolutionary history of chlamydiae and their gene content evolution was examined in several papers. An introduction to microbial genome evolution, with a focus on horizontal gene transfer and genome reduction, is thus also provided. Cultivated and well-studied chlamydiae are obligate intracellular endosymbionts. To place

chlamydiae within the wider context of symbioses, an overview of symbiosis, host-microbe interactions, and endosymbionts is also provided. Methods for exploring microbial diversity and inferring evolutionary history were fundamental for this thesis, and key methodologies and considerations will likewise be introduced. My goal here was to give the reader information on related methods that come up throughout the papers and to frame them in a wider context.

On reflection, much of the research presented in this thesis was driven by curiosity-based science. We kept happening upon new questions and avenues to pursue related to chlamydiae, and soon my doctoral research was fully infected. Many of the papers in this thesis focus on new insights and present exceptions to what was previously known about the phylum Chlamydiae. I find it thrilling to think about how many unknowns there are and how much there is left to discover in microbial diversity and evolution. This thesis has also resulted in many question marks, and I am excited to see what future investigations uncover. One thing that I have noticed recurring is that exceptions to the rule still seem to be the rule in microbiology.

I would also like to point out that the presented work stems from collaborative efforts. The journey to get to this thesis would not have been possible without the talent of the co-authors involved in the different papers, and the great mutualistic interactions with them. Now without further ado, herein I present to you a thesis focused on expanding the Chlamydiae tree and learning more about the genome diversity and evolution of this underexplored phylum. I am optimistic that you will likewise be infected with curiosity to learn more about chlamydiae, and I hope you enjoy reading it as much as I enjoyed the journey here.

1 The evolving tree of life

"The story so far:
In the beginning the Universe was created. This has made a lot of people very angry and has been widely regarded as a bad move."
– Douglas Adams, *The Restaurant at the End of the Universe*

1.1 A (very) brief history of life on earth

The Earth was formed ~ 4.5 billion years (Ga) ago, and relatively shortly thereafter life on Earth evolved (*1*). Physical evidence of life from carbon isotope signatures, microfossils, and stromatolites date the evolution of life, and thus the Last Universal Common Ancestor (LUCA), to over 3.5 Ga (*2-5*). All cellular life today, as far as we are currently aware, has descended from a common ancestor that gave rise to the three domains of life: Bacteria, Archaea, and Eukarya. It is unknown whether Bacteria or Archaea evolved first or if the root of life lies between these two domains. However, it is clear that prokaryotes (Bacteria and Archaea) evolved first and eukaryotes (Eukarya) later (*6*).

Global anoxia persisted until the rise of oxygen in the atmosphere during the great oxidation event (GOE) 2.3 - 2.4 Ga (*7, 8*), with current estimates for permanent atmospheric oxygenation at 2.22 Ga (*9*). Here, the cyanobacterial invention of oxygenic photosynthesis led to a transition from a reducing to an oxidizing atmosphere (*7, 8*). Strict anaerobes were forced to retreat to anoxic environments, which would have been common as the deep ocean remained starved of oxygen until 0.5 to 1 Ga (*10*).

Eukaryotes are thought to have evolved in the span after the GOE, from the merger of an archaeon and bacterium 1.2 to 2 Ga (*1*). There are many hypotheses for the specific events that led to the evolution of eukaryotes (*i.e.*, eukaryogenesis). Several prominent hypotheses posit that eukaryotes evolved from syntrophic interactions between at least two partners—an archaeal host and bacterial symbiont (that would become the mitochondrion)—mediated by hydrogen exchange (*11-15*). Recent analyses suggest Asgard archaea and Alphaproteobacteria as the closest modern relatives of this archaeon and bacterium, respectively (*16, 17*). Some modern eukaryotes that live in anoxic environments produce hydrogen, using divergent mitochondria termed mitochondrion-related organelles (MROs) (*18, 19*). Despite diverse anaerobic

eukaryotes, and the proposed importance of hydrogen during eukaryogenesis, the evolutionary origins of this metabolism are unclear.

Another key "symbiogenesis" event would occur between 1.1 and 1.7 Ga (*1*). Here, the eukaryotic ancestor of the primary plastid-bearing lineage Archaeplastida (which includes plants and green algae) took up a cyanobacterial endosymbiont, that would become the chloroplast. Multicellular eukaryotes only later rose to prominence, alongside a rise in oxygen to more modern levels 0.6 Ga (*7*). This coincided with the oxygenation of much of the deep ocean and the evolution of the first animals. This was closely followed by the Cambrian explosion 0.541 Ga (*5*). Plants and metazoans (*i.e.*, animals and their protist relatives) then colonized terrestrial habitats and a few hundred million years of evolution later here we are today. But for the vast majority of life's history microbes have reigned king, and still today there is rarely an environment untouched by their presence. In essence, despite the multitude of multicellular life, we still live in a microbial world.

1.2 Our microbial world

> "The role of the infinitely small in nature is infinitely great"
> – Louis Pasteur

Microbes represent the unseen majority, contributing a substantial portion of global biomass and overshadowing that from animals (*20*). Much of this microbial biomass is found in the relatively unexplored deep subsurface, which is estimated to include 90% of bacteria with many living in biofilms (*20, 21*). There are predicted to be as many as 1 trillion species among these microbial cells, indicating the myriad of microbial lineages awaiting characterization and classification (*22, 23*). It is strange to think based on what we know now, and how pervasive microbial life is, that for much of human history we were naïve to their presence.

The early adopters of microbiology worked under very different constraints than practitioners today, limited to using visual cues for microbial classification. Carl Linnaeus, who invented modern taxonomy (*i.e.*, genus and species names), barely bothered with classifying microbes, relegating the lot to the single genus *Chaos* (meaning formless) (*24*). Then came Charles Darwin and the theory of evolution (*25*). "I think" a thesis involving evolution isn't allowed to be complete without quoting Darwin? Although not as well recognized, Darwin did apply his evolutionary theory to microbial life (*26*). In particular, he used microbes to underscore the point that evolution was not a progression from simple to complex.

The unity of life was appreciated early on, based on synonymous biochemistry found in both unicellular and multicellular organisms, famously

epitomized in the 1947 quote by microbiologist Albert Kluyver: "From elephant to butyric acid bacterium – it is all the same" (*24*). With the discovery of DNA so could begin the evolutionary classification of life.

1.3 New views on an old tree

> "The incredible diversity of life on this planet, most of which is microbial, can only be understood in an evolutionary framework."
> – Carl Woese

The use of the SSU rRNA gene as a marker for inferring evolutionary relationships, pioneered by Carl Woese and George Fox, was transformational in microbiology (*27, 28*). Their work to build an evolutionary tree of life led to the discovery of Archaea and the reclassification of cellular life into three domains. The sequencing of SSU rRNA gene sequences directly from environmental samples altered the face of microbial ecology, by revealing widespread uncultivated microbial diversity (*29, 30*). The vast majority of microbes remain uncultivated and dominate Earth's diverse microbiomes (*31, 32*). Over the last decade, the rise in culture-independent methods has allowed for their genomic exploration. This has led to the characterization of many previously unidentified phyla from a wide variety of environments, ranging from the human mouth to the deep subsurface (*33*). Uncultivated lineages have greatly expanded our knowledge of microbial diversity and provided new vistas on the growing tree of life (*34*). It is now clear that multicellular eukaryotic life is in the minority, far surpassed by the microbial majority.

Life is organized into the following taxonomic hierarchy: domain, phylum, class, order, genus, species, and strain. Some bacterial and archaeal phyla also associate with each other in superphyla. There are currently 27 proposed phyla in Archaea, and four superphyla: Euryarchaeota, TACK, Asgard, and DPANN (*35*). A total of 92 named bacterial phyla were included in recent reconstructions of the tree of life, though there are likely more (*34*). Bacterial superphyla include CPR, Terrabacteria, FCB, and PVC. The phylum Proteobacteria is also of particular note, as it includes a number of well-sampled classes. The current picture of the eukaryotic tree of life resolves seven higher-level supergroups and other orphan lineages (*e.g.*, Excavates) with unclear phylogenetic placement (*36*). Plastid-bearing eukaryotes include those with primary plastids (*i.e.*, Archaeplastida) and those with secondary or tertiary plastids derived from uptake of other eukaryotes (*37*).

Despite our increased knowledge of the tree of life, there have been biases in sampling, and unrepresented groups still abound. It is likely that currently unidentified microbial groups play important ecological and biogeochemical roles and will help to further inform us about evolutionary history.

1.4 Microbial taxonomy in the *Candidatus* era

As knowledge of microbial diversity initially stemmed from cultivated organisms, so too did systems for classification and taxonomy. The International Code of Nomenclature of Prokaryotes (ICNP) presents the hierarchical conventions and nomenclature rules for naming archaeal and bacterial species (*38, 39*). However, the rules for introducing or revising nomenclature are stringent and conditional on obtaining an organism in axenic culture. However, the vast majority of life cannot yet be cultivated, which has led to a wild-west situation for naming and classification (*39*).

Uncultivated taxa are currently accommodated by the ICNP through the designation *Candidatus*. However, such names do not have nomenclature standing and can be overwritten by the characterization of representative isolates. Furthermore, official *Candidatus* designation requires formal description, a situation not well-suited to taxonomic classification in projects where thousands of genomes can be obtained (*40*). This has necessitated the naming of uncultivated taxa without validation, resulting in the proliferation of synonyms and inconsistencies. Solutions have been suggested by the microbial research community to address the naming of uncultivated groups (*38, 41*). These fall into two broad ways forward: for the ICNP to recognize DNA sequences as type material and allow the formal classification of organisms not yet in axenic culture, or for a separate nomenclature code to be developed for uncultured Archaea and Bacteria.

In the meantime, efforts to reorganize microbial taxonomy using a genome-based phylogenetic approach have been undertaken. The GTDB database presents a revision of prokaryotic taxonomy. In this framework taxonomic ranks are normalized based on relative evolutionary divergence (*42*). Such efforts are needed for the efficient classification of genomes obtained by cultivation-independent means, and for resolving larger taxonomic inconsistencies. However, this has resulted in massive changes to existing taxonomy, which have not been without controversy. This has included the renaming of groups validly classified under the ICNP, the adoption of misnomers, and a lack of continuity, as the database is continually updated and taxonomic names changed accordingly (*39*). In addition, there have been complaints that name changes have been taken without appropriate consultation of experts working with the groups, thereby causing issues for future data analysis and interpretation (*39*).

2 Microbial genome evolution

2.1 Microbial genome evolution

Like all life, microbes evolve through time by accumulating changes in their genomes that are passed on to their descendants, and through which evolutionary processes such as selection and genetic drift can act. Advantageous genomic changes, that confer a survival or reproductive advantage, can undergo strong positive selection and quickly become fixed in a population. On the other hand, disadvantageous genomic changes are subject to negative purifying selection and can quickly be lost. Nevertheless, genetic drift can still result in the fixation of slightly deleterious changes. Genetic drift refers to the random sampling process of standing genetic variation, which occurs in a population through the birth and death of individuals (*43*). The effects of genetic drift are minimized in sufficiently large populations. However, microbes can undergo environmental isolation (for example through niche specialization) and be subject to fluctuations in population size. Population bottlenecks are particularly common for pathogens, resulting in a higher rate of evolution through genetic drift (*43*).

Genomic changes in an individual can arise through mutation, genome rearrangements, and the acquisition of exogenous DNA. Though mutation is the ultimate source of genetic variation (*44*). Gene mutations can occur at non-synonymous or synonymous nucleotide positions, altering or maintaining the encoded amino acid sequence, respectively. Non-synonymous mutations are relatively scarce in comparison to synonymous mutations, as the former result in changes to proteins, which are readily subject to selection, while the latter can be silent (*43*). Genetic rearrangements of varying lengths, such as deletions, duplications, insertions, inversions, and translocation events can also disrupt genes, affect their expression, and lead to the loss of genomic fragments (*45*).

Over time the accumulation of mutations or gene rearrangements can lead to the evolution of new gene functions and *de novo* genes. New gene functions can evolve through mutation, the fusion/fission of genes or protein domains, and gene duplication (*46*). In addition, novel genes can evolve *de novo* from previously non-coding stretches of DNA (Figure 1). Taxonomically restricted (or orphan) genes are found across the tree of life and do not appear to have homologs in other groups, and may represent *de novo* genes (*46, 47*).

Nevertheless, some apparent *de novo* genes could be the result of homology detection failure due to divergence (*48*).

Gene content innovation can also come from external sources, such as mobile elements and horizontally transferred genes (*45*). Genes can thus have complex evolutionary histories and their host genomes represent a mosaic of gene origins (Figure 1). Genes related through a common ancestral gene are termed gene homologs. Orthologs are homologous genes that have evolved vertically through speciation events and are thus expected to have conserved functions (*49*) (Figure 1). Homologous genes related by duplication events are termed paralogs and those that have been acquired by HGT are termed xenologs (Figure 1) (*49*). True gene orthologs, which have completely avoided gene duplications and transfers, are rare. Orthologous groups are therefore used to refer to sets of gene homologs that have evolved from a common ancestral gene at a given speciation event, for example in a phylum ancestor, regardless of subsequent events (*49*) (Figure 1).

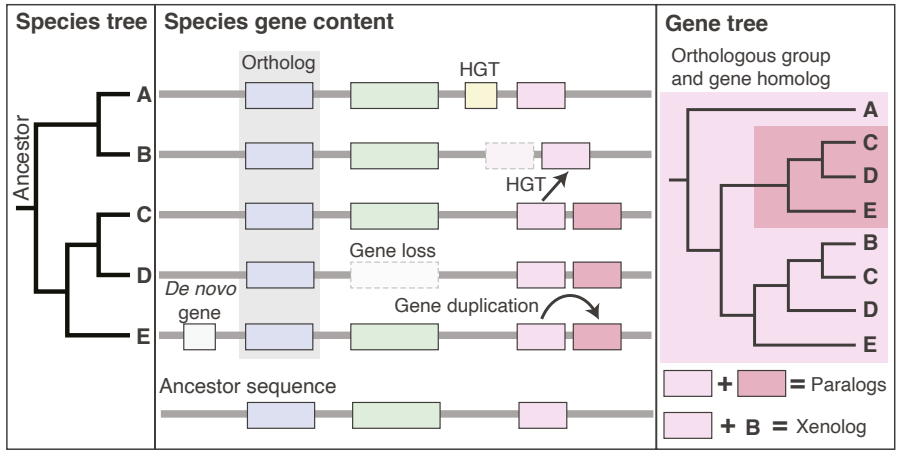

Figure 1. An outline of gene events leading to gene content evolution, and the resulting patterns in a gene tree relative to species tree.

2.2 HGT and the tangled tree

The exchange of genetic material is a major driving force in microbial genome evolution. HGT is important for ecological diversification by allowing organisms to more readily adapt to new environments and accelerating metabolic innovation (*50*). HGT is pervasive, occurring within and between all major divisions of life. HGT is increasingly recognized as playing a major role in both unicellular and multicellular eukaryote evolution and there are now many robust examples (*51, 52*). HGT is particularly widespread between eukaryotic hosts and their bacterial symbionts, with genes derived from HGT

providing host protection, altering their nutrition, and allowing adaptation to extreme environments (*53*). These HGTs can also include endosymbiotic gene transfer. These are transfers that occur from an endosymbiont, or from other organisms that have aided in its initial establishment, to the host nuclear genome (*50, 52*). This process has played a central role in the evolution of eukaryotic organelles (*e.g.*, mitochondria and plastids) (*52*).

HGTs can result in either the acquisition of new genetic content or the replacement of existing homologous stretches of DNA through recombination (Figure 1) (*54*). HGTs are more common between closely related organisms, decreasing in frequency with more distant evolutionary relationships (*50*). Regardless, "highways" (*i.e.*, increased rates) of HGT can be observed occurring between organisms that are more distantly related, yet reside in similar environments (*55*). Microbes in shared environments exchange genes more often, due both to proximity and the adaptive potential of gene gains from the same ecological niche. Genes that form functional units tend to be transferred together, such as subunits of larger complexes, operons, and gene clusters (*56*).

This large network of HGT between microbes can be seen as a tangled tree, obscuring evolutionary relationships between organisms (*57*). However, certain genes tend to undergo HGT more often than others, where genes encoding metabolic proteins are most commonly transferred. HGT events can be detected through phylogenetic conflict, where branching patterns for a gene or protein are not consistent with species relationships (*50*). Essential genes related to central cellular processes (*e.g.*, DNA replication, transcription, and translation) tend to be vertically inherited, since replacements or mutations can drastically reduce fitness. These functionally conserved genes, which are present in a group irrespective of their ecological niche, are referred to as the core genome (*58*). Species, and even closely related strains, can have large variation and diversity in their accessory (non-core) genome (*50*). Unlike pure laboratory cultures, wild microbial populations can be highly heterogeneous, with various genes found in low frequency (*58*). HGT is the main contributor to this flexible gene content, which is often related to environment-specific ecological adaptations.

2.3 Genome streamlining and gene loss

There are huge variations in genome size across the tree of life. In prokaryotes genome size can range from less than 0.12 Mb, as in the leafhopper insect symbiont *Candidatus* Nasuia deltocephalinicola (*59*), to more than 14 Mb, as is the case for *Sorangium cellulosum* (*60*). Gene loss is an important evolutionary process that occurs (i) through loss of a genomic region, or (ii) through loss of function mutations, and subsequent pseudogenization (*61*)

(Figure 1). Over time non-essential genes, under neutral selection, are thus lost through these processes (*62*).

Streamlining refers to the minimization of cellular complexity and size. Two main paths can lead to this pattern, selection favoring reduction in free-living organisms, and loss through the neutral process of genetic drift in symbionts (*62, 63*). Host-associated microbes, and in particular endosymbionts, have the smallest genomes apart from viruses and microbial-derived organelles (*63*). Obligate endosymbionts undergo relaxed selection, resulting in neutral genome reduction, that is driven either by the loss of DNA repair genes or Muller's ratchet (*63*). In the latter case, small population sizes and isolation of endosymbionts lead to increased genetic drift and a lack of recombination. Hence, mildly deleterious mutations can accumulate and the rate of sequence evolution, including gene loss, is accelerated in comparison to free-living relatives (*64*). Endosymbiotic genome reduction is an ongoing process with a continuum of reductive signatures found, and yet retention of core informational genes and genes for host-interaction or provision (*65*).

Under growth-limiting circumstances, where effective population size is large, streamlining can also be favored by selection in free-living organisms (*62*). Such genome reduction is more common in nutrient-poor environments where selection favors energy conservation, such as the deep subsurface (*e.g.*, marine sediments) (*66*) and oligotrophic marine water. Here, having even slightly lower energy needs can help an organism to outcompete others. For example, alphaproteobacterial members of the order Pelagibacteriales (formerly SAR11 clade) are the most abundant group in the world's oceans, yet they have small streamlined genomes (*67*). Free-living organisms can even have genome sizes approaching 1 Mb (*63*).

Where a gene function is dispensable, such as when a nutrient or compound can be obtained from the external environment, there is a selective advantage to gene loss (*62*). However, such gene losses can result in dependence on the production of these metabolites from co-occurring microbes. This is the basis of the reductive evolution theory termed the Black Queen Hypothesis, which suggests that the availability of public goods by "helpers" will result in adaptive gene loss in microbes that can benefit, leading to interdependent microbial communities (*68*). Auxotrophy in conditionally essential biosynthetic genes has been experimentally shown to result in a growth advantage (*69*). The exploration of uncultivated microbial diversity has revealed that genome streamlining and auxotrophy are commonplace and widespread across diverse groups (*62*). This could help to explain the interconnectivity of microbial communities, based on underlying metabolic exchange networks (*70*), and the emergent difficulty in obtaining pure culture isolates. This can lead to the evolution of obligate cross-feeding interactions, including primary metabolites (*e.g.*, nucleotides, amino acids, carbon sources, vitamins, etc.), which are cemented through gene loss (*70*).

3 Symbioses

"Life did not take over the globe by combat, but by networking."
– Lynn Margulis,
Microcosmos: Four Billion Years of Microbial Evolution

3.1 Symbiotic interactions

Symbiosis is a broadly defined term that encompasses all close interactions between two or more organisms that are sustained over time. Symbioses are ubiquitous and have large environmental, ecological, and evolutionary impacts. The types of symbiotic interactions sit along the mutualism-parasitism continuum. These include (i) mutualism where both partners benefit, (ii) commensalism where one partner benefits with no discernable effect to the other, and (iii) parasitism where one partner benefits but at a fitness cost to the other (Figure 2). Pathogens are organisms that cause virulence through a host-parasite interaction upon infection of a host (*71*). However, in reality whether an organism is a pathogen is dependent on ecological context and there are no clear-cut genetic distinctions between pathogenic and non-pathogenic organisms (*71*). Even tightly-linked intracellular mutualists can have antagonistic interactions with their hosts and become "opportunistic pathogens" (*72*). For example, the mutualistic coral symbiont *Symbiodinium*, a dinoflagellate algae that provides photosynthesis-derived compounds to its host, has also been found to parasitize the coral under heat stress conditions (*73*).

Symbioses occur between organisms across the tree of life. These interactions can be facultative or obligate for each partner. Microbial symbioses with multicellular hosts are understudied, while the extent of microbe-microbe symbioses is becoming clearer with cultivation-independent approaches. For example, symbiosis appears to be widespread across the more recently described DPANN archaea and CPR bacteria (*33, 74*).

Microbes often engage in metabolic cooperation. Syntropy refers to a subset of microbial symbioses, based on the obligate and mutualistic exchange of metabolites (*75*). Some key examples of syntropy involve the exchange of hydrogen under anaerobic conditions, such as between fermentative syntrophic bacteria and hydrogenotrophic methanogenic archaea, which produce and consume hydrogen, respectively (*76*). These interactions can also occur intracellularly, such as between anaerobic ciliates, whose MROs

produce hydrogen, and their archaeal symbionts that consume it (*77*). Such syntrophic interactions have been experimentally shown to emerge where variations in auxotrophy exist (*e.g.*, for amino acids) (*78*). This illustrates the likely commonality and importance of syntropy in diverse biospheres.

Specific molecular mechanisms help to facilitate symbiotic interactions. Prominent among these are the various secretion systems, whose functions include: (i) facilitating cell adhesion, (ii) delivering proteins and DNA extracellularly or to other target cells, (iii) injecting effectors and virulence factors during host infection, and (iv) excreting toxic compounds such as antibiotics (*79*). Secretion systems are involved in symbiotic interactions across the parasite-mutualism continuum and also play roles in competition and biofilm formation. Interestingly, the T3SS evolved from flagella and they share conserved core machinery (*80*). Flagella and chemotaxis are commonly used by symbionts that need to seek out their interacting partner (*81*). Symbionts often encode transporter genes for acquiring metabolites from other cells or the external environment, such as vitamins and amino acids. These include nucleotide transporters (NTTs), which are found across a wide range of organisms and can be used to transport ATP by parasites who scavenge host chemical energy reserves (*82*). Some intracellular bacterial symbionts also encode eukaryotic-like protein domains, which are thought to be used for controlling and infecting their host (*81*).

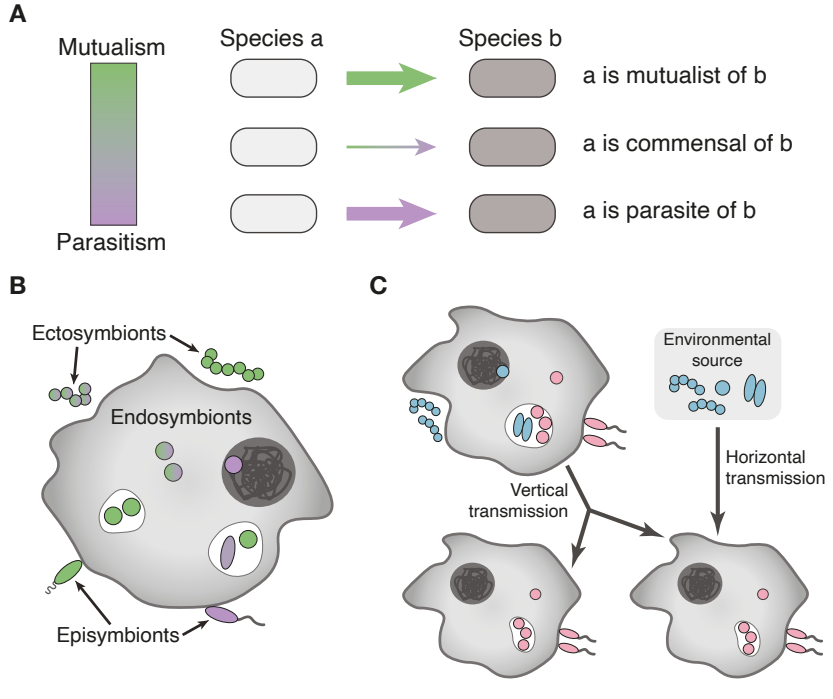

Figure 2. **A**. Symbiotic interactions along the mutualism-parasitism continuum **B**. Host-symbiont localization **C**. Vertical and horizontal symbiont transmission.

3.2 Host-microbe symbioses

In a host-microbe symbiosis, the host is typically designated as the larger organism and the smaller the symbiont. Symbionts can be either intracellular (endosymbionts) or extracellular (ectosymbionts), and in the latter can be loosely or tightly associated with the host (such as episymbionts) (Figure 2).

Symbionts are transmitted to their hosts through vertical or horizontal transmission, or mixed modes (*83*) (Figure 2). Horizontally transmitted symbionts are taken up from environmental sources, while vertically transmitted symbionts are inherited by offspring (*83*). Transmission mode has important implications for the evolutionary trajectories of a symbiosis. Vertically transmitted symbionts tend to have sustained host association, which can lead to co-diversification (*84*). Symbionts can participate in nutritional symbioses and aid host adaptation to different ecological niches (*85*). Symbionts can also provide host protection from toxic compounds, predation, and parasites. These defensive symbioses are often facilitated by chemical defense through symbiont production of secondary metabolites (*86*).

Microbial symbionts can influence the ecology, physiology, health, and behavior of their hosts. Together these microbial assemblages are referred to as the host microbiome. Marine animal microbiomes, including those of sponges and corals, are also important for the functioning of marine ecosystems such as coral reefs (*87*). They also play a role in stress tolerance and adaptation in the face of climate and other anthropogenic changes (*87*). Metabolic cross-feeding can also emerge within microbiomes and yield products or traits of value for the host (*88*). There is a clear importance of microbiomes for many animals, and an overarching paradigm of their benefit. However, it is important to note that not all depend on symbionts. Some animal lineages appear to be microbiome-free (*e.g.*, caterpillars) and others have low abundant microbiomes (*e.g.*, some birds) (*89*). Microbes identified in these organisms could represent transient microbial associations, perhaps from food consumption or other environmental sources and even parasites, rather than members of a resident-established microbiome (*89*).

3.3 Intracellular symbionts and endosymbiosis

Endosymbiosis refers to symbioses where one organism lives inside the cells of the other. Endosymbionts can reside in the host cytoplasm, inside host vacuoles, or even inside the host nucleus (*90*) (Figure 2). Vertically inherited endosymbionts tend to have highly reduced genomes, to have more stable host relationships, and to more often result in host-beneficial endosymbiosis where they provide a fitness advantage to the host (*91*). Over long evolutionary timescales interactions between a host and endosymbiont can become closely intertwined, where they become completely dependent on each other and

unable to survive without the other (*65*). There are even Russian-doll-like examples of endosymbionts living inside other endosymbionts. For example, mealybugs engage in a three-way symbiosis. They have a bacterial endosymbiont *Tremblaya princeps*, which in turn has its own bacterial endosymbiont *Moranella endobia* (*92*).

Many horizontally transmitted endosymbionts are mutualists or commensal to their hosts. However, this group does include prolific parasites, including pathogens that cause common diseases in humans and other animals. For instance, members of the genera *Legionella*, *Mycobacterium*, and *Chlamydia* are transmitted horizontally between hosts and are responsible for the atypical pneumonia Legionnaires' disease, tuberculosis and leprosy, and the eye disease trachoma and sexually transmitted infections, respectively (*84*). Single-cell eukaryotes, such as amoebae, often act as reservoirs of horizontally transmitted endosymbionts and have been described as “training grounds” for bacterial pathogens, though they can also host diverse beneficial symbionts (*93*). Amoeba is a non-taxonomic designation referring to microbial eukaryotes that have an “amoebal” form and that prey on other microbes using phagocytosis. Amoeba-resisting bacteria are horizontally transmitted endosymbionts that can escape predation. Many of these amoeba-infecting endosymbionts have evolved convergent features. Their lifestyles in amoeba are suggested to select for virulence traits that aid in infecting diverse animal cells (*84*). Interestingly, the presence of members of the Legionellales order has recently been detected in evolutionarily distant amoebal protist lineages, suggesting host lifestyle rather than taxonomy mediating their host range (*94*). Amoebal endosymbionts also tend to have larger genomes than other endosymbionts. This is thought to be due to larger within-host population sizes that allow for recombination, through HGT facilitated by co-infecting endosymbionts, and DNA from digested prey (*95*).

Endosymbiont genomes can become extremely reduced over time. In some cases only retaining genes for a few key functions, resulting in a transition from an endosymbiont to an organelle. Eukaryotic evolution has been impacted by the acquisition of endosymbionts through symbiogenesis. The majority of mitochondria and MROs, which evolved from a proteobacterial ancestor, and plastids, which evolved from a cyanobacterial ancestor, retain a remnant genome clarifying their symbiont past (*96*). Mitochondria can generally provide energy to their eukaryote host by respiring oxygen and most plastids can capture energy for their host through photosynthesis. The mitochondria of many eukaryotes that live in anaerobic conditions have undergone reductive evolution and lost the ability for respiration (*18, 19*). They instead conserve energy using fermentation (*18, 19*).

4 Chlamydiae

4.1 The PVC superphylum

The PVC superphylum is a group of bacterial phyla with shared ancestry consistently supported in reconstructions of species relationships (Figure 3). The superphylum is named PVC after the three founding members of the group (*i.e.*, Planctomycetes, Verrucomicrobia, and Chlamydiae) (*97*). This superphylum is of research importance from multiple perspectives, which include ecology, biotechnology, medicine, symbiosis, and evolutionary cell biology (*97-99*).

The PVC superphylum now also includes the phyla Lentisphaerae and Kirimatiellaeota (*99*), whose initial cultivated members were isolated from seawater and microbial mats, respectively (*100, 101*) (Figure 3). PVCs may also include the candidate phylum *Candidatus* Omnitrophica (initially described as OP3 (*102*)). Thus far, this group lacks a cultivated representative and based on environmental surveys appears to be composed of anaerobes (*103*). Although *Candidatus* Omnitrophica often affiliates with the PVC superphylum in phylogenetic analyses, its exact evolutionary relationship among PVC members is unclear (*34, 40, 104*). The phylum *Candidatus* Poribacteria, a group found as part of the sponge microbiome, was initially classified as a PVC member (*97, 105*). However, later phylogenomic analyses have refuted this as it was found to affiliate with non-PVC phyla (*34, 106*).

Several groups within PVCs play a central role in global biogeochemical cycles. Some members of the Planctomycetes are unique in performing anaerobic ammonia oxidation ("anammox"), combining ammonia and nitrite directly to form N_2 gas in an intracellular membrane-bound cell compartment termed the anammoxosome (*107, 108*). First identified in wastewater sludge, these anammox bacteria are responsible for a large proportion of global nitrogen turnover and have industrial potential through the removal of ammonium. Verrucomicrobia includes aerobic methanotrophs that can oxidize methane under extremely acidic pH conditions, and that contribute to combating climate change by reducing atmospheric methane emissions (*109, 110*). PVC members may also be important in the discovery of novel natural products. For example, members of the Planctomycetes found in the microbiomes of sponges and macroalgae are an untapped source of bioactive compounds and secondary metabolites (*111, 112*).

There has been much interest and debate in the study of PVCs from an evolutionary cell biology perspective (*99, 113*). For example, Planctomycetes and Chlamydiae lack central bacterial cell division genes (*e.g.*, FtsZ). Recent work has illuminated the "chlamydial anomaly", which asked why chlamydiae were sensitive to antibiotics that impact peptidoglycan synthesis, when peptidoglycan could not be detected (*114*). Using more advanced detection methods, both Chlamydiae (*115, 116*) and Planctomycetes (*117, 118*) have now been found to have peptidoglycan as part of their cell wall. The difficulty in initial detection in chlamydiae could be explained by the synthesis of peptidoglycan only during cell division (*119*). Chlamydiae appear to have polarized cell division and divide in a process more akin to budding than binary fission, where a ring of peptidoglycan is formed at the division plane (*119, 120*). Planctomycetes also have variations in cell division, including both binary fission and polar budding (*121*).

Planctomycetes were initially proposed to have an endomembrane system, but that has now been refuted and shown to be extreme invaginations of a single continuous cell membrane (*112*). Anammox Planctomycetes do have an intracellular lipid-bound compartment (*i.e.*, the anammoxosome), but this finding is in line with the now recognized widespread presence of organelles in bacteria (*122*). Despite this resolved anomaly, PVCs continue to surprise. The Planctomycetes member *Candidatus* Uab amorphum was recently shown to perform phagocytosis-like engulfment of bacteria and picoeukaryotes (*123*), with implications for understanding prokaryotic cellular complexity.

All PVC phyla have been found in host-associated environments (*99*). Kirimatiellaeota is found in animal intestinal tracts (*101*), and the Planctomycete *Akkermansia muciniphila* in the human gut microbiome (*124*). Verrucomicrobia also interact with eukaryotes (*125*), and include an extremely reduced ciliate endosymbiont with a genome size of 0.16 Mb (*126*). However, Chlamydiae are the most prolific as symbionts. Despite diverging from other PVCs 1-2 Ga (*1, 104*), all known members are symbionts.

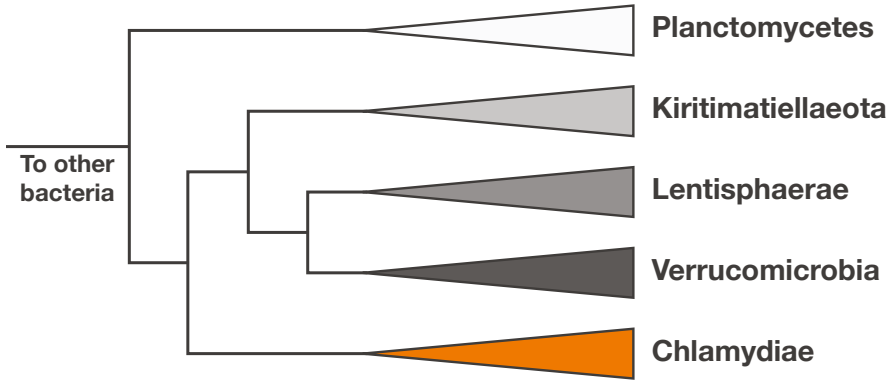

Figure 3. Evolutionary relationships between phyla within the PVC superphylum.

4.2 A historical perspective

There has been a century-long investigation into the phylum Chlamydiae (*127*). Even beyond that, descriptions of eye infections resembling those caused by *Chlamydia* date back thousands of years (*128*). Chlamydiae were first identified in 1907 in Java as the infectious agent responsible for the eye disease trachoma (*129*). They were given the name "Chlamydozoa" after the Greek word "khlamus", which means a mantle or cloak, as the *Chlamydia*-filled vacuoles had been mistaken for "mantled protozoans" (*128*).

Indeed, for many years chlamydiae remained cloaked in mystery, due to their reliance on a eukaryotic host and resulting difficulties in cultivation. In fact, chlamydiae were for decades thought to be viral or viral-like in nature due to their small size and obligate intracellular replication. In 1966 Moulder was able to demonstrate that chlamydiae were gram-negative bacteria and not viruses, using evidence such as the presence of both DNA and RNA, ribosomes, cell division, and cellular integrity throughout their lifecycle (*130*). Moulder also proposed the "energy parasite" hypothesis and suggested that chlamydiae scavenged ATP and other energy-rich metabolites from their eukaryotic hosts (*131*). Decades later he would be proven correct with the discovery of NTTs that could import ATP from the host cytosol (*132, 133*). Although energy parasites of their hosts, chlamydiae can also generate ATP.

The genome of *Chlamydia trachomatis* was sequenced in 1998, shortly after the first bacterial genome (*134*). This revealed a small genome, yet with unexpected metabolic genes. This included near-complete central metabolic pathways such as glycolysis, the TCA cycle, and an electron transport chain (ETC) (*127, 134*). As of early 2021, just over 20 years after the first chlamydial genome was sequenced, there are over 600 chlamydiae genome assemblies available on NCBI. Chlamydial pathogens have spurred a strong interest in studying chlamydial biology, and today the Chlamydia Basic Research Society (CBRS) brings together hundreds of researchers. Despite the large amount of sequence data, and a booming research community, there have been comparatively few research groups active in studying chlamydiae outside of the medically and zoologically relevant pathogens.

Yet, other chlamydiae are pervasive in the environment and their genomic and ecological diversity is massively underexplored (*104*). These poorly studied groups likely represent an important frontier for gaining a full picture of the ecological and environmental impacts of chlamydiae. Moreover, chlamydiae are relevant for understanding various evolutionary questions related to symbiosis, pathogenesis, evolutionary cell biology, relationships within PVCs, and their impact on eukaryote evolution.

4.3 The chlamydial lifecycle

All well-studied chlamydiae share a conserved biphasic lifecycle as obligate intracellular symbionts of eukaryotes. Despite this, there are large variations in genome size and metabolism across chlamydiae, with hosts ranging from microbial eukaryotes to animals. The chlamydial developmental lifecycle hinges on the transition between two functionally and morphologically distinct cell types: an extracellular non-dividing stage, known as elementary bodies (EBs), and an intracellular dividing stage, known as reticulate bodies (RBs) (*127, 135, 136*) (Figure 4). The mechanisms of the chlamydial lifecycle are particularly well-studied in human pathogens of the Chlamydiaceae family (*136*). Across all chlamydiae, the cycle begins when EBs encounter a potential host cell (Figure 4). EBs gain host entry through endocytosis into a membrane-bound vacuole, termed the inclusion. Here, the smaller EBs differentiate into larger RBs, in some chlamydiae growing in size from 0.3 μm to 1 μm (*135*). The RBs can divide throughout the intracellular portion of the lifecycle. RBs then differentiate back to EBs asynchronously before host release through lysis or extrusion (Figure 4). Extracellular and infectious, EBs can once again ebb into the cycle upon host contact.

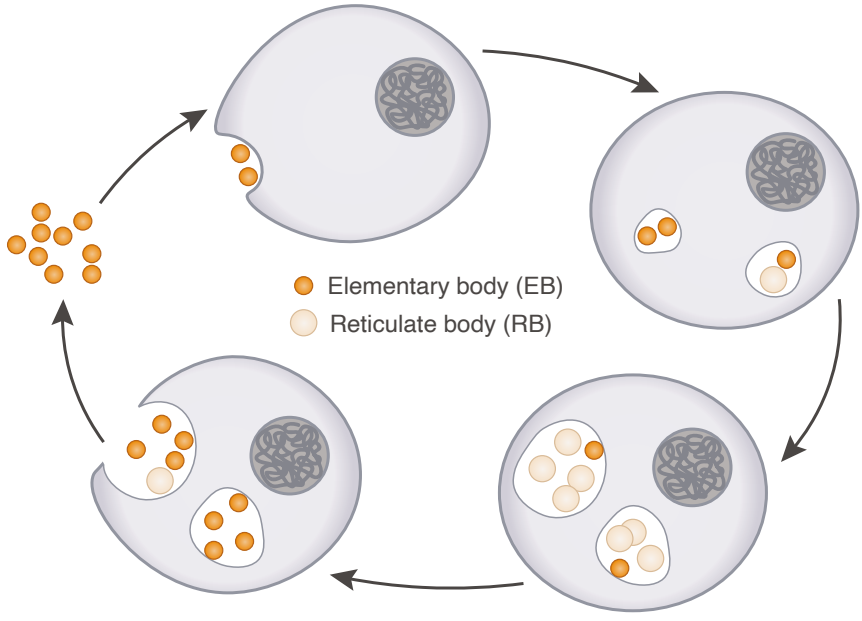

Figure 4. Overview of the conserved chlamydial biphasic lifecycle. Dividing intracellular reticulate bodies (RBs) and non-dividing elementary bodies (EBs) are shown infecting a eukaryotic host.

EBs are visually distinctive due to their highly condensed nucleoid, with DNA compacted using conserved histone-like proteins (*137, 138*). EBs are highly resistant to both osmotic and physical stress (*139*). Their ability to withstand harsh environmental conditions is attributed to their unique rigid cell wall, which is stabilized by disulfide cross-linking of outer membrane proteins (*127*). During differentiation to RBs crosslinking is reduced allowing for the membrane fluidity necessary for cell division (*136*). Eukaryotic lipids, such as cholesterol, are acquired from the host and incorporated into the chlamydial cell membrane (*140*). Inside the host, chlamydial localization varies, with some forming one large inclusion rather than several. Others are found in the cytoplasm, and some reside in the host nucleus (*90*). Under conditions of environmental stress—such as nutrient deprivation, exposure to antimicrobials, the host immunological response, and iron starvation—the chlamydial developmental cycle can be arrested and RBs enter a reversible state of "persistence" (*136*). Here, RBs transition to larger aberrant forms which do not divide, but persist intracellularly until conditions improve.

The transition from EBs to RBs is mediated by the global transcription factor EUO, which is conserved across Chlamydiae. Throughout the lifecycle, the T3SS mediates host interaction and manipulation through the secretion of effectors (*141*). It was initially thought that EB DNA compaction resulted in complete transcriptional shutdown and that they had a "spore-like" primary function in extracellular survival while awaiting a host encounter. But EBs are not spores, and they are now known to remain metabolically active outside the host. Chlamydial EBs can maintain infectivity for prolonged time periods, have respiratory activity outside the host, and can survive extracellularly when provided with different carbon and energy sources (*142, 143*). Although chlamydiae have not been observed dividing outside of a host, they are more adapted to host-free survival than initial observations implied. Chlamydiae that infect microbial eukaryotes may need to survive in the environment while waiting for a new host encounter, and host-free activity has been observed after nearly a month (*144*). Chlamydiae have stage-specific requirements for metabolites. RBs scavenge ATP from the host, and EBs synthesize their own ATP (*145*).

EB formation and endosymbiosis with eukaryotic hosts has been found in phylogenetically diverse chlamydial groups (*146*). Furthermore, genomes from uncultivated chlamydiae encode key components of this lifecycle, such as NTTs, a T3SS, and the master regulator EUO (*147*). Thus far, attempts to grow chlamydiae axenically have been unsuccessful. It is surprising, that an entire phylum of bacteria may have maintained the same lifestyle throughout evolutionary time. However, members of the Chlamydiaceae family, who all infect animal hosts, have by far been the most well-studied group.

4.4 The notorious Chlamydiaceae pathogens

The Chlamydiaceae or "pathogenic chlamydiae" include well-known pathogens with relevance for both human and livestock health, zoonoses, and conservation (*128, 136, 148*) (Figure 5). The most infamous member of the Chlamydiaceae is *Chlamydia trachomatis,* which causes the common sexually transmitted infection. Based on incidence rates 120 million new cases of *C. trachomatis* were estimated in just 2016 (*149*). However, this species is also responsible for causing trachoma. This neglected disease is the leading cause of preventable blindness (*150, 151*). *Chlamydophila pneumoniae* is also a prevalent human respiratory pathogen, estimated to be the cause of approximately 10 % of community-acquired pneumonia cases, and is also associated with chronic lung and heart conditions (*136*).

Several Chlamydiaceae species can also cause zoonotic disease in humans (*128*). This includes *C. psittaci*, a bird pathogen that causes avian chlamydiosis, *C. abortus*, which causes spontaneous abortions in ovine species, and *C. felis*, which can cause both conjunctivitis and respiratory infections in cats. In general, Chlamydiaceae species can infect a wide range of animals, including domesticated mammals, koalas and other marsupials, and more recently genomes have been obtained from reptiles (*128, 152*). Chlamydiaceae are microaerophiles often found in oxygen-poor tissues, and are able to divide under anaerobic conditions (*153*). Gene content and gene order are highly conserved in Chlamydiaceae, with recombination between strains common (*128*). Both intra- and interspecies HGT has been documented in Chlamydiaceae, yet gene transfers with other groups are rare (*140*).

For much of the history of investigation into Chlamydiae, they have been composed of only the orphan genus *Chlamydia*. Then considered an anomalous and highly derived group of pathogens distantly related to other bacteria. This remained the status quo until a combination of serendipity and curiosity-driven science drastically changed the landscape.

4.5 Discovery and rise of environmental chlamydiae

> "Here we report the preliminary characterization of a chlamydia-like organism which first appeared as a laboratory contaminant of unknown origin in our cell cultures." – Kahane *et al.*, 1993

In the 1990s, Kahane *et al*. decided to investigate a contaminant of mysterious origin that had appeared in their human cell line cultures. This organism "Z" was obligately intracellular, could grow in a variety of cell types, and had a developmental cycle closely resembling *Chlamydia*, yet was more distantly related on an SSU rRNA gene-level (*154, 155*). *Simkania negevensis* Z

represented the first described non-Chlamydiaceae member of the Chlamydiae phylum, and until recently it would remain the sole genome representative of the Simkaniaceae family. Soon, the amoeba-infecting *Parachlamydia acanthamoeba* (the first Parachlamydiaceae family member) was identified, which shared similar physiology (*156*).

Over the next few decades additional chlamydiae from outside the Chlamydiaceae were discovered in association with diverse eukaryotic hosts, ranging from protists to animals (*104, 157-159*). Due to historical precedence, they are collectively referred to as environmental chlamydiae or chlamydia-like organisms (CLOs). The first environmental chlamydiae genome was sequenced in 2004 and revealed a much larger genome size in relation to pathogenic chlamydiae (*i.e.*, the Chlamydiaceae) (*160*).

With the continued sequencing of new genomes, comparative genomics revealed striking similarities among chlamydiae in gene content related to their biphasic lifecycle and eukaryotic host association. In contrast, environmental chlamydiae were found to have extensive differences, with greater metabolic versatility encoded in their larger genomes (*127, 157, 158, 160, 161*). For example, they were found to have a complete TCA cycle, more ETC complexes, and could use unphosphorylated glucose in glycolysis (*127*). While many chlamydiae are auxotrophic for many cofactors, amino acids, and nucleotides, more extensive *de-novo* biosynthetic capabilities were found in some environmental groups (*162-164*). More recently, chlamydial lineages have been identified that encode genes for flagellar motility, and a CRISPR system (*164, 165*). Despite their large evolutionary distances, chlamydiae share a co-evolving plasmid, that may mediate gene transfer (*166*).

Interactions between environmental chlamydiae and their various hosts are often unclear, but some do appear to be capable of causing disease in animals. *Waddlia chondrophila* has been implicated in spontaneous abortion in mammals, including humans (*167*). Other environmental chlamydiae have been suggested as emerging or opportunistic pathogens, as they have been identified in association with respiratory tract infections and adverse pregnancy outcomes (*168, 169*). However, these groups are also found in healthy adults (*170*). These groups may simply be associated with amoeba found on various mucosal surfaces such as the nasal cavity. Members of the families Clavichlamydiaceae, Parilichlamydiaceae, and Piscichlamydiaceae are clear pathogens of fish and cause the disease epitheliocystis (*171*). Pariliclamydiaceae genome representatives have recently been sequenced. Phylogenomic analyses indicated a distant evolutionary relationship with Chlamydiaceae, yet their genomes closely resembled each other, a discrepancy suggested to be the result of convergent evolution (*172*).

Recent experimental evolution with chlamydiae found that the mode of host transmission caused shifts in symbiont-host interaction along the parasite-mutualist continuum. Chlamydiae that were continually transferred to new naïve host populations increased in parasitism and expression of genes

related to infectivity (*173*). Horizontal transmission between hosts is observed in most chlamydiae, but there are some exceptions. Species of the genus *Neochlamydia* appear to be vertically inherited mutualists of their amoebal hosts (*174*). These species protect hosts against co-infection with *Legionella pneumoniae*, a human pathogen that infects amoeba as an environmental reservoir (*175*). Indeed, other chlamydial groups have been found to provide amoeba with protection against *L. pneumoniae* (*176*). It seems likely that chlamydiae mutualists that engage in defensive symbioses are more widespread than currently recognized.

All cultured chlamydiae strictly divide intracellularly, and due to their dependence on eukaryotic hosts are inherently difficult to study. The retrieval of new species from environmental samples through co-cultivation has thus proven challenging (*104*). There are currently six chlamydial families with cultivated representatives (in cells of their respective hosts), and thus only few models available for studying basic chlamydial biology (*104*) (Figure 5).

4.6 Chlamydiae diversity in the environment

In recent years, cultivation-independent surveys, using the SSU rRNA gene, have revealed that the vast majority of diversity within Chlamydiae remains unsampled (*104, 177, 178*). These surveys found that chlamydiae are ubiquitous in nature, with widespread occurrence across a wide range of both environmental and host microbiomes. In particular, marine water, freshwater, wastewater, sediments, soil, and plants have been identified as environmental reservoirs of novel chlamydial diversity. Chlamydiae were also found to have high prevalence in various animal and environmental samples (*104*), including wild populations of social amoeba (*179*). Chlamydiae have also been found with high relative abundances in some animal microbiomes, including those of corals, sponges, and the enigmatic gutless worm Xenoturbella (*180-182*).

The most recent count put the number of family-level chlamydial lineages based on SSU rRNA gene amplicon surveys at 1,157 (*104*). A surprisingly high number, especially given the fact that many of the primers used for surveys of microbial diversity have mismatches to the chlamydial SSU rRNA gene, and thus poorly capture chlamydiae (*104, 177*). A recent investigation of bacterial diversity in amplicon and metagenomic data found that Chlamydiae were among the most understudied phyla, with the majority of their diversity still to be uncovered in metagenomes (*178*). Over the past few years culture-independent methods, such as metagenomics, have led to the sequencing of novel chlamydial genomes (*183*). All of the articles referenced in this next section were published after the outset of this thesis, indicating the rapid pace of chlamydiae discovery.

Cultivation-independent methods have now been used to acquire chlamydial genomes directly from host-associated environments, such as

snakes, ticks, and fish (*152, 172, 184*). From the perspective of environmental samples, chlamydial SAGs with divergent traits have now been retrieved from marine water (*164*). Chlamydiae metagenome-assembled genomes (MAGs) have also been recovered in recent large-scale genome-resolved metagenomic investigations, some of which were recently compiled and examined (*147*). Intriguingly, a few investigations of groundwater from the deep terrestrial subsurface identified Chlamydiae alongside presumed symbionts, such as CPR and DPANN members (*185, 186*). Chlamydiae were also identified in soil mini-metagenomes, alongside numerous giant viruses that suggest the presence of eukaryotic hosts (*187*). Meanwhile, chlamydiae MAGs with high relative abundance were recently identified in Antarctic soils that appear to be organoheterotrophs (*188*). Chlamydiae may even be associated with algae (which has important implications for section 4.8) (*189*). Nevertheless, despite these advances it is evident that much remains to be explored within chlamydiae and that most diversity is yet to have genomic representation.

4.7 Phylogenomic and taxonomic considerations

There are currently 9 recognized families in the phylum Chlamydiae, all belonging to the same class Chlamydiia. Chlamydiaceae (the sole member of the order Chlamydiales), Parachlamydiaceae, Waddliaceae, Criblamydiaceae, Simkaniaceae, and Rhabdochlamydiaceae (who are assigned to the now paraphyletic Parachlamydiales order), and the order-less families Piscichlamydiaceae, and Parilichlamydiaceae (*127, 158, 159*) (Figure 5). However, chlamydial taxonomy has been the source of much debate.

Exemplifying this is the case of the Chlamydiaceae. The family was split into two genera, *Chlamydia* and *Chlamydophila*, based on percent relatedness and phylogenetic analyses of rRNA genes (*190*). This designation was not well accepted by the chlamydial field. Based on this fact, in addition to genomic conservation and shared ecology among Chlamydiaceae, it was emended back to *Chlamydia* (*191, 192*). Complicating chlamydial taxonomy is relatively poor taxon sampling outside of the Chlamydiaceae. This has left species relationships within the phylum difficult to untangle and higher taxonomic ranks not well resolved (*104*).

The Parachlamydiales were distinguished from the Chlamydiales based on shared synapomorphies, specifically indels in conserved genes and signature proteins (*168*). Supporting this, most earlier species phylogenies with Chlamydiaceae members, and the limited sampling of environmental lineages, identified a divergence between these two groups. This suggested that "environmental chlamydiae" had evolved from a common ancestor, and "pathogenic chlamydiae" from another (*161, 193*). "Taxogenomic" marker proteins were identified that reliably inferred this topology, to be used for future taxonomic classification of novel chlamydiae (*193*).

These views on chlamydial phylogeny led to proposals that the Chlamydiae ancestor resembled protist-infecting chlamydiae, with the patterns seen in other lineages driven by differential gene loss as a result of endosymbiosis. In particular, it was suggested that Chlamydiaceae had evolved by massive gene loss of the metabolic diversity encoded by protist-infecting chlamydiae (*194*). However, a reconstruction of PVC gene content evolution, with limited chlamydial sampling, found that the Chlamydiae ancestor had already undergone genome contraction (*195*). Furthermore, they did not find elevated rates of gene loss within Chlamydiae, but rather low rates of gene gain (*195*).

With the genomes obtained using cultivation-independent methods, more comprehensive reconstructions of chlamydial species relationships could be performed. These confirmed the separation of Chlamydiales and Parachlamydiales and led to the proposal of additional clades of high taxonomic rank (*147*). However, they did not account for fast-evolving and compositionally biased lineages. Both of which are common among symbionts and could result in phylogenetic artifacts.

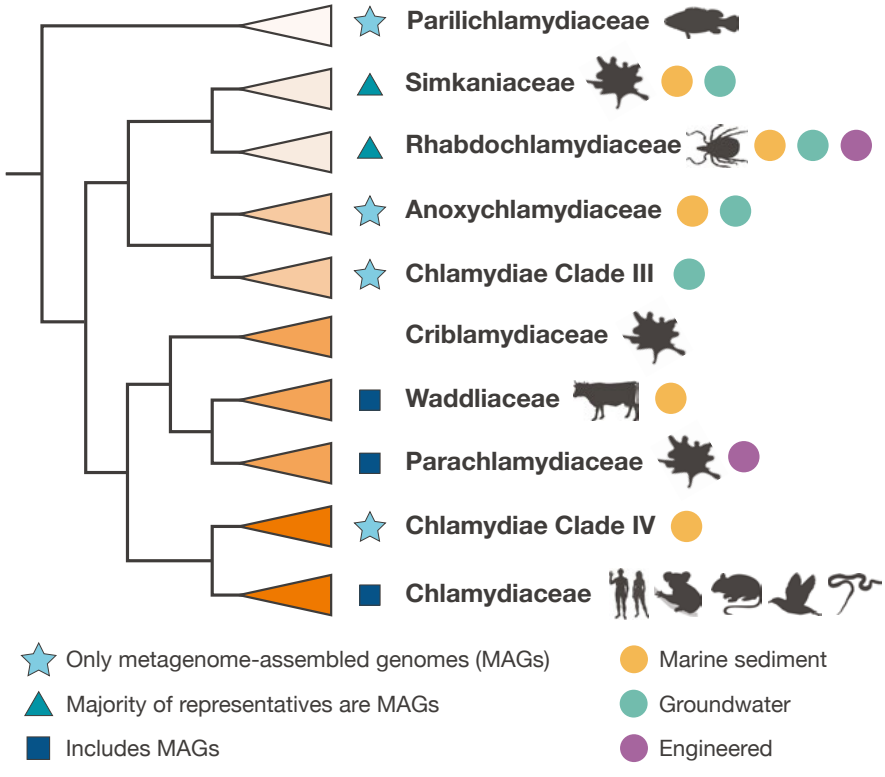

Figure 5. Species relationships of chlamydiae families with published genome representatives. The environments or hosts that members were sequenced from is indicated, alongside the availability of MAGs. Simkaniaceae and Rhabdochlamydiaceae are also referred to as Chlamydiae Clade I and II, respectively.

Further complicating the matter are recent efforts to establish a standardized taxonomy for bacteria using genome phylogeny (*42, 104*). The taxonomic names suggested for chlamydial groups in this standardized view are in further conflict, even at the phylum level. The database stemming from these efforts (GTDB) initially classified Chlamydiae as a phylum. Members were then demoted to a class of Verrucomicrobia, but have now once again achieved phylum rank. However, they are now inexplicably referred to as the phylum "Verrucomicrobia_A" instead of Chlamydiae. Throughout this thesis Chlamydiae will be referred to as their original widely accepted name.

4.8 Chlamydiae and eukaryotic evolution

Chlamydiae and eukaryotes both evolved over a wide time span, at some point in the last approximately 1-2 Ga (*1*). We do not know if ancestral chlamydiae already interacted with early eukaryotes. But the conservation of endosymbiosis with eukaryotes across extant phylum members is suggestive of it. Interestingly, a large number of chlamydial genes share homology with sequences from eukaryotes. It is not surprising that HGT has occurred between chlamydiae and their eukaryotic hosts, given their interaction over long-scale evolutionary time. However, a number of these proteins are found conserved in diverse chlamydiae and eukaryotes, suggesting ancient rather than modern transfer events. Most prominent, are possible HGTs between ancient chlamydiae and the ancestor of Archaeplastida (*i.e.,* plants and green algae).

An affiliation between a large number of chlamydial and plant homologs was already noted in the first chlamydiae genome (*134*). It was later found that most of the plant proteins with chlamydial homologs were targeted to the chloroplast (*196*). NTTs are key for transporting ATP across the chloroplast membrane and mediating plastid-host energy integration. These ATP-transporting NTTs were found to be of likely chlamydial origin. This lead to the hypothesis that chlamydiae played a role in the initial endosymbiotic establishment of the chloroplast cyanobacterial ancestor (*197*). With the identification of additional genes, an ancient tripartite symbiosis was proposed to have facilitated these transfers (*198*). Namely between the plastid-cyanobacterial ancestor, the Archaeplastida ancestor, and an ancestral chlamydiae. This was suggested as a more likely evolutionary scenario than multiple independent HGTs (*198*). The list of putatively transferred genes expanded with additional sequenced diversity and phylogenetic analyses, resulting in several dozen candidates (*199, 200*).

Then came the ménage -à -trois (MAT) hypothesis (*201*). The MAT hypothesis is based on an elaborate synthesis of biochemical fluxes and metabolic integration between the cyanobacterial plastid-ancestor, plastid-bearing host ancestor, and an ancient chlamydial parasite. The hypothesis hinges on the use of chlamydial genes related to glycogen metabolism for host

polysaccharide storage of cyanobacterial-derived carbon from photosynthesis (*201*). These glycogen-related proteins are secreted by extant chlamydiae into the host cytosol during infection. However, extant chlamydiae infecting plastid-bearing eukaryotes have thus far not been identified. Further updates and fine-tuning of the MAT hypothesis were later made (*202, 203*). However, the hypothesis has not been without controversy. Recent phylogenetic investigations of glycogen metabolism proteins did not find compelling evidence for chlamydial ancestry of these genes in Archaeplastida (*204*). This led the study authors to suggest that chlamydiae did not play a role in the establishment of the primary plastid endosymbiosis. Nevertheless, there does appear to be an affiliation of homologs from chlamydiae and plastid-bearing eukaryotes in phylogenetic analyses, representing a large proportion of the non-cyanobacterial chloroplast proteome (*205*). It remains to be seen whether these HGTs were indeed facilitated by overlapping endosymbiosis events.

Potential HGT events between chlamydiae and eukaryotes have also been found that trace back to other periods during eukaryotic evolution. HGTs from chlamydiae to Euglenophytes, which harbor a secondarily acquired plastid, include most components of a SUF system for iron-sulfur (FeS) cluster assembly (*206*). Several tRNAs likewise seem to have been transferred from chlamydiae to the mitochondria of specific plastid-bearing eukaryotes and an association with the origin of vascular plants suggested (*207*). Phylogenetic analyses of a bacterial-like tRNA-guanine transglycosylase found in some microbial eukaryotes was likewise indicative of chlamydiae HGT (*208*).

Some more general associations between eukaryotes and chlamydiae were identified early on, including homology between chlamydial histone-like proteins and histone H1 in eukaryotes (*209*). Interestingly some chlamydiae (*e.g.*, *Simkania negevensis*), mitochondria, and chloroplasts share a conserved group 1 intron in their 23S rRNA gene (*210*). A recent study inferred a late timing of mitochondrial acquisition based on the short stem length of last eukaryotic common ancestor (LECA) families of Alphaproteobacteria descent (*211*). Interestingly, a similarly timed acquisition of LECA families from Verrucomicrobia/Chlamydiae was likewise inferred.

5 Thesis aims

The aims of this thesis center around Chlamydiae with a particular focus on phylum genomic diversity and evolution. The key aims of this thesis fall into several broader categories and are as follows, to:

Chlamydial environmental diversity

- Gain a wider perspective on the environmental distribution and relative abundance of chlamydiae across various biomes (papers I, IV).
- Explore microbial diversity in anoxic marine sediments (paper I) and in the microbiomes of three sponge species (paper IV).
- Expand Chlamydiae genomic diversity from marine sediments and the sponge microbiome using genome-resolved metagenomics (papers I, IV).

Chlamydiae species relationships and taxonomy

- Employ phylogenomics to reconstruct Chlamydiae species relationships and to delineate previously unidentified clades (papers I, III, IV).
- Redefine chlamydiae taxonomy based on improved taxon sampling and phylogenomic methods accounting for potential biases (papers I, III, IV).

Ecology and environmental impacts of uncultivated chlamydiae

- Compare host-associated features, metabolism, and lifestyles of newly identified lineages with previously available taxa (papers I, II, IV).
- Investigate unique gene content that could indicate divergent lifestyle traits and ecological impacts (papers I, II, IV).
- Infer the evolutionary history of identified genes of interest and whether they have undergone HGT to or from chlamydiae (papers I, II, III, IV,V).

Ancestral state reconstruction of the Chlamydiae phylum

- Reconstruct Chlamydiae gene content evolution and ancestral states of key chlamydial and PVC ancestors (paper III).
- Examine evolutionary events, such as gene gains from HGT and losses, coinciding with major chlamydial transitions and divergences (paper III).

Chlamydiae and eukaryotic evolution

- Identify potential genetic contributions between chlamydiae and eukaryotes and how this has impacted their evolution (paper II, V).

Methods for exploring microbial diversity and inferring evolutionary history were central for fulfilling the above-stated aims. The next sections will outline general background information about such methodological considerations pertaining to the core of this thesis.

6 Exploring microbial diversity

6.1 Cultivation-dependence and the rise of omics

There are clear advantages of being able to grow an organism in pure culture under laboratory conditions. Cultivation allows an organism's growth dynamics, metabolism, response to environmental conditions, and cellular features to be more easily studied. It also simplifies the challenge of assessing gene functions, and obtaining large amounts of cellular components (*e.g.*, DNA, RNA, and proteins) for molecular analyses. However, the microbial diversity we are able to cultivate in the lab does not equate to the rich diversity present in nature. Staley and Konopka exemplified this in 1985 with their "Great Plate Count Anomaly" experiment. Here, they observed that the number of microbes that grew as colonies on standard culture plates paled in comparison to the number of microbes seen under the microscope (*212*).

Most microbial groups lack cultivated representatives, including 81% at genera-level (*32*). But using cultivation-independent approaches a large number of novel phyla have been discovered in recent years. There are some suggestions that a saturation point is being reached in discovering new phyla (*33*). Yet, much remains to be identified at finer taxonomic scales. With a myriad of environments left to be explored there may well be numerous phyla and groups with evolutionary and ecological importance awaiting discovery.

The large discrepancy in cultivated microbial groups can be explained by several factors. First, there has been a cultivation bias towards microbes with direct societal impacts. This includes microbes that are pathogens, that benefit livestock and crops, and that have applications in biotechnology. Second, many microbes depend on community interactions and metabolites (Zengler and Zaramela 2018). These conditions can be difficult to reproduce and maintain in artificial settings. Thirdly, barriers exist for growing some microbial groups in isolated pure cultures, due to obligate symbioses. Finally, some microbes simply grow slowly, on timescales we are unaccustomed to relative to their speedy cousins (*e.g.*, the lab happy *E. coli* with a division time of 20 minutes). Microbial cells in the deep subsurface are estimated to divide extremely slowly with biomass turnover on the order of hundreds to thousands of years (Hoehler and Jorgensen 2013). For example, it took 12 years to obtain an enrichment culture of the marine-sediment residing *Candidatus* Prometheoarchaeum syntrophicum, the first cultured member of the Asgard superphylum (Imachi, et al. 2020).

Nevertheless, microorganisms once thought to be intractable to cultivation have been obtained in recent years using traditional methods. This includes previously uncultured members of the human microbiome and dozens of novel Planctomycetes members (*121, 213*). New phyla can also still be discovered through standard cultivation, such as the eukaryotic phylum Hemimastigophora (*214*). Culturing methods have undergone a renaissance in recent years and innovations are helping to correct the cultivation divide, with many recent success stories (*215*). Despite these advances and the importance of culturing, cultivation-independent approaches are necessary for exploring microbial diversity and for gaining a larger picture of microbial ecology and evolution.

There has been a rapid jump in technological capability since the sequencing of the first microbial genome 25 years ago. Hundreds of genomes can now be obtained from individual environmental samples, most commonly using second-generation short-read technologies. Further advancements are now being pioneered through long-read sequencing platforms, the third generation of sequencing technology (*216*). A growing toolbox of omics methods is now available, with many applicable to uncultivated groups (*217*). Cultivation-independent omics can be applied both on a single-cell level and on a meta-community level. Methods include those for assessing microbial diversity in samples (*e.g.*, amplicon sequencing, single-cell genomics, and metagenomics), and those for assessing ongoing metabolic processes and gene expression (*e.g.*, single-cell- and meta- transcriptomics, proteomics, and metabolomics). Genomes and transcriptomes can also be assessed in parallel allowing a direct connection between genomic content and gene expression (*218*). Genome-resolved metagenomics can be used to directly obtain the genomes of organisms from environmental samples, represented by MAGs.

Cultivation-independent approaches to study organisms extend beyond the omics. Microbial cells can be visualized directly in environmental samples using fluorescence in-situ hybridization (Pernthaler, et al. 2002). Furthermore, metabolic activities can be assessed to a single-cell level, using tools such as nanoSIMS and Raman spectroscopy (Musat, et al. 2012). By connecting data from a microbial community across various perspectives a larger-scale picture can be built than would be possible through cultivation. Underpinning it all is a view widened by molecular data. The rise in omics techniques has been unprecedented in increasing understanding of the microbial world and will continue to be a driving force in future exploration. The following sections will focus on methods and considerations related to the cultivation-independent methods employed in this thesis. Chiefly, environmental sampling, amplicon sequencing, metagenomics, and genome-resolved metagenomics (Figure 6).

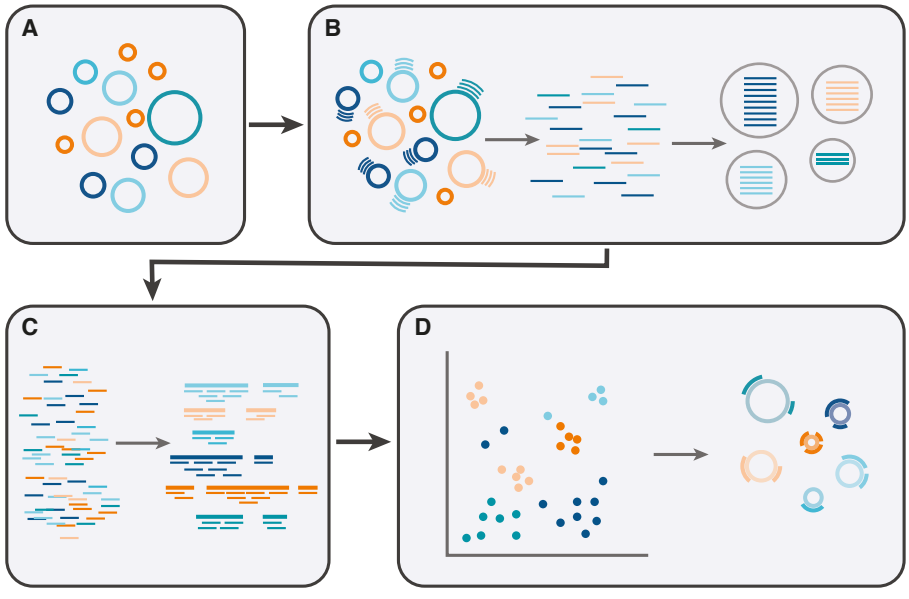

Figure 6. Steps in a typical workflow for genome-resolved metagenomics. These steps include **A**. obtaining DNA from environmental samples, **B**. assessing the microbial community using amplicons, **C**. sequencing and assembling a metagenome, and **D**. binning contigs into MAGs, using differential coverage information if available.

6.2 Environmental sampling

The first step for investigating microbial diversity is to obtain representative samples and high-quality DNA (Figure 6). Samples with lower microbial diversity and well-defined communities can be easier to subject to downstream methods. These will likely result in more complete and contiguous genomes. Sediment and soil samples can be difficult to work with since they are often complex and contain thousands of species with high strain diversity (*219*). While more complex from a sampling standpoint, extreme environments such as hydrothermal vents and hot springs can have less complex microbial communities and be more amenable to omics techniques. Sample complexity can be reduced by size filtering to enrich for specific populations, such as viruses or eukaryotes (*220, 221*). Furthermore, culture-based enrichment can also be used to increase the relative abundance of organisms of interest (*222*). Nevertheless, it is sometimes a fact that samples are suboptimal and high sequencing depth is needed to capture low-abundant microbes of interest. Obtaining several related samples (*e.g.*, at different time-points or in neighboring locations) can also be advantageous for later assembling MAGs. In all cases, a sterile workflow is important for avoiding sample contamination, in particular for low cell-count samples such as those from the deep subsurface (*223*). Collecting sample metadata is essential for

reproducibility, and environmental and chemical parameters can allow for connections with ecological and biogeochemical processes.

There are several factors to consider when extracting DNA from environmental samples. Downstream analyses can be impaired by compounds in samples. For example, humic acids and salts can impair DNA extraction and the enzymes used for DNA amplification (*224*). Cell extractions can help remove these compounds, although this comes at the cost of introducing bias against cells more tightly associated with the sample substrate. Biases can also be introduced through the DNA extraction method employed. Even subtle differences in DNA extraction can lead to large differences in measured microbial composition (*225*). In particular, lysis susceptibility can vary among microbial cells. Lysis methods include mechanical force, temperature changes, sonication, and chemical or enzymatic digestion. Mechanical force-based methods, such as bead beating, have been shown to increase the efficiency of extraction from hard-to-lyse microbial cells (*226*). However, these aggressive methods also shear DNA into shorter fragments, which is problematic for obtaining high-molecular-weight DNA needed for long-read sequencing. Overall, the DNA extraction method used should be the one best suited for the specific question and environmental samples. Yet it is important to remain aware of potential biases that could be introduced at this stage.

6.3 Amplicon sequencing

Amplicons involve PCR amplification and sequencing of a gene or genome segment of interest from organisms present in a sample (Figure 6). Amplicon sequencing can be used to assess any genomic region where primers can be designed. It can be used to detect specific metabolic genes in environmental samples, for example, the *mcrA* gene that is indicative of methanogenic archaea (*227*). However, amplicon sequencing is most commonly used to examine sample microbial diversity. This can be directed towards specific taxonomic groups of interest or more generally on a universal level. A good universal marker gene for assessing microbial diversity has the following properties: (i) it is found in a single copy, (ii) is present across the group of interest, (iii) has undergone minimal HGT, and (iv) has both fast- and slow-evolving sites for distinction and classification at various taxonomic levels.

The SSU rRNA gene meets all of these requirements and is the most widely used for assessing microbial communities. The SSU rRNA gene was used to build the initial three-domain tree of life and has led to the cultivation-independent discovery of many microbial lineages (*27, 30, 228*). The SSU rRNA gene is vertically inherited and highly conserved across all domains of life. It retains enough phylogenetic signal for distinguishing distantly related groups, yet has both conserved and variable regions for distinction between more recently diverged lineages. Taxonomy can be assigned based on

comparison to reference sequences and more distantly related sequences can be taxonomically classified using a lowest common ancestor approach. Like any universal marker, the SSU rRNA gene does have inherent issues. The rRNA gene operon is found in varying copy numbers across organisms. This results in biases in relative abundance of different groups from a microbial community (*229*). Since amplified SSU rRNA gene fragments are often short, there can be insufficient phylogenetic signal for determining exact evolutionary relationships. Although the SSU rRNA gene is arguably not the "best" marker, its strength lies in its wide—really universal—use by the microbial research community. It has been extensively studied and comprehensive databases allow for easy use and accurate classification.

In general, amplicon sequencing approaches are subject to methodological biases (*230*). Sequence errors can be introduced by DNA polymerase, and chimeric sequences formed during PCR amplification. Chimeric sequences should be less abundant than parent sequences. They can be removed by comparison with reference databases or through partial matches to more abundant sequences. Amplicon sequences are often grouped based on sequence identity into operational taxonomic units (OTUs). This results in easier taxonomic classification and accounts for small sequence errors. Species-level OTUs (97 %) are most common, but different boundaries can be used that correspond to various taxonomic levels (*231*). All amplicon sequencing methods are inherently biased by their reliance on primers. Primer mismatches to a target sequence will result in a lack or lower degree of amplification. There is no region in the SSU rRNA gene that is 100 % identical across all domains. Primer redundancy can increase the coverage of taxonomic groups, but there are limits and some groups will always be missed. Entire phyla are missed by some of the most commonly used universal primers (*232*). This includes phyla belonging to the CPR, and Chlamydiae (*177, 233*). Certain taxonomic groups, such as CPR bacteria, have particularly divergent SSU rRNA genes that have introns, which resulted in blind spots in our view of microbial diversity until direct sequencing of environmental samples (*234*).

Near full-length SSU rRNA genes and even full rRNA gene operons can now be obtained using long-read amplicon sequencing (*235-237*). This improves their resolution in phylogenetic trees and aids in identifying and classifying novel groups. Full-length rRNA gene sequences can now also be obtained without primers, though this requires large amounts of high-quality RNA (*238*). The use of amplicon sequencing in large-scale projects, such as the Earth Microbiome Project, has given us a wider picture of global microbial ecology (*239*). While of value and importance in and of itself, amplicon sequencing is also commonly used for the initial assessment of a microbial community. Samples of interest that are identified can then be subjected to further study, such as metagenomics.

6.4 Metagenomics

Metagenomics is the sequencing of all genetic material obtained from an environmental sample (Figure 6). Through this random "shotgun" sequencing, a global picture of sample diversity is obtained without needing to amplify a particular gene. The number of reads sequenced from a given organisms' genome corresponds to its relative abundance in an environmental sample (*i.e.*, read coverage). Metagenomes give a more accurate picture of a microbial community, even though biases can still be introduced during sampling and DNA extraction. Metagenomic workflows and currently available tools have been well-reviewed and are particular to the project and question being investigated (*240-243*).

It has been nearly two decades since the first metagenomes were obtained (*244, 245*). There have been drastic improvements to sequence read throughput since that time (*246, 247*). Illumina-based sequencing platforms are currently the most widely used for obtaining metagenomics reads. These platforms produce hundreds of millions of short reads with a low error rate and by deep sequencing low-abundant community members can be assessed. However, short reads are difficult to assemble and the large amount of data generated can likewise hamper the assembly process. In this context, assembly refers to the combination of sequence reads into longer stretches of contiguous DNA called contigs. It is rare to obtain single assembled contigs that represent complete genomes using short reads. Nevertheless, there have been recent improvements in short-read metagenome assembly (*248*).

Pacific Biosciences (PacBio) and Oxford Nanopore platforms produce long sequence reads at high-throughput, and allow complete genomes to be sequenced from organisms in pure culture (*249, 250*). Long-read metagenomic sequencing has thus far been minimal, but recent examples have resulted in thousands of high-quality genomes (*251*). Though in their infancy, these methods will no doubt become standard. A major disadvantage of long-read technologies is the high error rate, but this is continually being improved upon. Long-reads can be pre-corrected prior to assembly and corresponding short-reads used for error correction. In addition, short and long reads can be hybrid assembled, which has been shown to improve contiguity and the quality of downstream MAGs (*252, 253*).

There are currently a large number of assemblers available (*254*). These primarily fall into two larger categories based on underlying general principle, namely overlap-layout consensus (OLC) and De Bruijn graph assemblers (*255, 256*). In OLC methods, every read is compared to every other read to identify overlaps given a specified overlap length. The final contig sequence is then determined by the consensus (*i.e.*, most represented nucleotide at each position) of the overlapping reads. OLC methods are well-suited for longer reads but struggle with the sheer number and short size of reads generated in the majority of metagenomics studies. De Bruijn graph assemblers are faster

and less memory intensive since they avoid read alignment. Instead, reads are split into shorter fragments, termed k-mers, of a given length. Assembly graphs are then built from all possible overlapping k-mer pairs by n-1 nucleotides (where n is the k-mer length). A contig sequence is then reconstructed by connecting and traversing the graph of k-mers into a path.

Metagenome assemblies are often highly fragmented due to the complex problem of reconstructing many genomes at once based on only short fragments. Where there is not enough information to resolve an assembly graph ambiguity contiguity is broken. Certain genomic regions are particularly difficult to assemble, such as repetitive regions. Repeats are found in multiple locations throughout the genome and often cannot be spanned by short reads. The presence of strain variation in a sample can also hamper assembly in a similar way, due to the presence of sequence reads with only small differences. Sequence errors can also cause assembly issues, which can be amplified by excessive coverage (*248*). This is why the most abundant organisms in a sample can sometimes be poorly assembled.

Once assembled, metagenomes can be used directly to assess community-wide functions and microbial diversity. Such gene-centric metagenomics has been used to gain broad-scale insights, such as microbial drivers of biogeochemical processes. For example, the TARA oceans expedition surveyed core microbial functionalities in global ocean metagenomes and examined changes with environmental factors and geography (*257*). Microbial diversity in a metagenome can be assessed using a variety of genes, including the SSU rRNA gene. But with contig sequences now available, it is also possible to use concatenated protein phylogenies that provide better resolution. Ribosomal proteins are often found organized in a gene operon, and have been commonly used to assess relationships of uncultivated lineages (*34, 185, 258*). At least some ribosomal proteins are likely assembled together on a single contig that would have originated from the same genome. Ribosomal protein orthologous can be aligned to reference taxa, concatenated, and used to reconstruct a phylogenetic tree. This precludes the need for contigs to be classified together into MAGs, and gives a more accurate picture of metagenomic microbial diversity than a single gene (*17, 259*).

6.5 Obtaining MAGs

Circular closed genomes can be obtained on a single contig directly from short-read metagenomes, but this is rare. As of late 2019 only 59 bacterial and 3 archaeal such circular chromosomes had been sequenced, mostly with smaller chromosomes (*260*). Contigs that originate from the same genome can be grouped or "binned" together to reconstruct MAGs (Figure 6). Such genome-resolved metagenomics can then be used for evolutionary, functional, and ecological investigations. Obtaining MAGs of high quality is not always

a simple matter, but improvements in binning algorithms have led to projects reconstructing up to thousands of MAGs (*40, 185, 186, 234, 261*).

Contigs can be binned into MAGs using either taxonomy-dependent (supervised) or unsupervised methods (*242, 243*). Supervised binning works by reference-guidance from previously available genomes. However, this approach does not work for obtaining MAGs from novel groups. Unsupervised methods are now more common. These use similarities in statistical properties of contigs, such as nucleotide composition and read coverage, to group contigs into bins. Nucleotide composition is generally conserved and unique across a prokaryotic genome (*262-265*). This includes both GC-content and frequency profiles of oligonucleotides. There are trade-offs to the oligonucleotide length used. Specificity increases with length, but there are greater computational demands from the number of frequencies to calculate. Tetranucleotide frequency (four nucleotides) is most often used. Since tetramer profiles are distinctive for a genome, they can be used to bin contigs with similar composition together that likely originated from the same chromosome (*242, 243*). However, it is difficult to separate closely related species with similar genomic signatures, using nucleotide composition alone.

On average, a similar depth of read coverage is expected across contigs originating from the same genome (*242, 243*). This sequence coverage information is also typically incorporated into contig binning. However, contigs from different organisms can have the same coverage (*i.e.*, relative abundance). When multiple samples are considered this is much less likely to be the case. Differential coverage binning incorporates information from multiple related samples where an organism is found with different abundances (*266, 267*). Contigs originating from the same genome are expected to have similar sequence read coverage profiles across samples and are binned together. Combining differential coverage with composition increases contig classification accuracy, thereby improving MAG recovery. The samples used must overlap in their resident microbiome and have differences in relative abundance. For example, samples collected close together spatially or temporally (*e.g.*, different depths or time series). Biases in DNA extraction can also be taken advantage of to obtain metagenomes with different coverage profiles. Numerous automated tools now exist that can bin contigs using both composition and coverage, in addition to tools that can collate results from multiple binning methods (*242, 268, 269*).

Specific genomic regions are difficult to classify into bins due to deviations in composition. These poorly binned regions tend to be biased towards rRNA gene operons, tRNA genes, and recently acquired genomic material (*e.g.*, mobile elements, prophages, plasmids, and recent HGTs) (*270*). Microdiversity is also an issue when binning MAGs. Closely related strains exhibit similar composition profiles and often co-occur. This can result in bins with a large amount of heterogeneity, where contigs from different co-occurring strains have been binned together. In general, contigs can be binned

incorrectly resulting in MAGs with contamination. In this context contamination refers to sequences that originate from another organism.

Bin refinement through manual curation is needed to ensure high-quality and accurate MAGs, and to prevent contaminating sequences from polluting further analyses and databases. MAGs can be manually refined by examining contig composition, coverage, and read-pair linkage (*271, 272*). Contigs that have divergent profiles and which may have been misclassified can then be removed. Genomes derived from closely related organisms that were binned together can also be manually separated. MAG genome quality can be assessed using single-copy marker genes. The percentage found indicates completeness, and the percentage found in multiple copies indicates redundancy and possible contamination. Marker gene sets are not universal, some groups lack certain markers or have multiple copies, but these methods give a general indication of MAG quality. Community standards (MIMAG) now also exist for determining MAG quality (*273*). It is also important to note that a MAG represents a "population-level genome". In essence a MAG is a consensus of closely related strains present in a given microbial community.

Once high-quality MAGs have been sequenced genes can be identified and translated into their respective protein sequences. These can then be annotated, often using automated tools that assign putative functions by reference database sequence comparison (*243*). Conserved protein domains can also be identified to assist in functional prediction. Proteins can also be assigned to protein families, such as clusters of orthologous groups (COGs), which are found across the tree of life (*274*). Reconstructing the metabolism of MAGs, although painstaking, can be aided by tools that assign proteins to pathways and functional modules from databases. Together, comparative genomics and metabolic reconstruction can be powerful for providing insight into the lifestyles, functions, ecological impacts, and interactions of microbial populations represented by MAGs. New metabolism, ecology, and evolutionary insights can also be discovered using MAGs from uncultivated lineages as a focal point. For example, methane metabolism outside of the Euryarchaeota was first identified in Bathyarchaeota MAGs (*275*). While the first genomes of Asgard archaea, the closest evolutionary relatives of eukaryotes, were MAGs obtained from deep marine sediments (*16, 276*).

It remains an issue that many proteins are poorly annotated or lack known functions. Automated annotation pipelines are necessary with the current scale of data, but they can introduce or further propagate erroneous annotations (*277*). Furthermore, a large proportion of gene content is unique to individual species and thus lacks annotation. Continued investigation and untangling of gene functions is of large importance. Phylogenetic analyses can be used to further understanding of a genes' evolutionary history (*e.g.*, if it was an HGT from another group) (*278*). In addition, phylogenies can help to deduce putative functions of less well-characterized proteins, by examining their placement in trees alongside proteins with known well-annotated functions.

7 Inferring evolutionary history

7.1 Inferring phylogenetic trees

Using phylogenetic trees to infer evolutionary history underpins many investigations in biology. Phylogenetics is important for understanding the origin of genes and functions, the emergence of cellular features and metabolism, evolutionary transitions, and for reconstructing the tree of life. Phylogenetic trees also aid in assigning taxonomy and determining the species relationships of newly identified groups. The practicalities of and methods for inferring phylogenetic trees have been well-reviewed (*279-281*). The following sections will give an overview of phylogenetic tree reconstruction, phylogenomics, biases and artefacts, and ancestral state reconstruction.

In essence, a phylogenetic tree is a hypothesis about evolutionary history. In theory, any feature can be used for reconstructing a tree. For example, character traits from morphological data are used when studying fossils. With the availability of sequence data molecular phylogenetics is typically used and trees reconstructed from nucleotide or protein sequences. Some genes follow organismal relationships having evolved vertically through speciation, and can be used to infer species histories. Many others have undergone extensive loss, duplications, and HGTs. Phylogenetic reconstructions of these sequences thus represent the evolutionary history of that particular gene and not the vertical history of speciation that led to the current organisms encoding it.

Phylogenies reconstruct evolutionary history using models of evolution that outline the substitution process of the given characters. As sequences diverge mutations accumulate at the nucleotide level, which are reflected in corresponding amino acids if they occur in coding regions at non-synonymous sites. More closely related sequences are expected to have accumulated fewer substitutions than more distantly related sequences (given a constant rate of substitution). Sequences are aligned into a multiple sequence alignment (MSA) by identifying homologous sites that derive from the same ancestral site while accounting for insertions and deletions that have likewise occurred. MSAs are then usually trimmed to remove poorly aligned sites and to minimize alignment errors (*281*). In addition to an alignment, a substitution model of evolution needs to be provided for tree inference. Nucleotides and amino acids have specific equilibrium frequencies across a given set of sequences, different relative substitution rates (*i.e.*, exchangeabilities), and variation in these properties across and within sites. Ideally, sequence

evolution would be modelled by allowing exchangeabilities, frequencies, and rates to freely vary across and within sites. However, such substitution models are at this current time computationally highly demanding.

The most basic substitution models work under the assumption that substitution rates between characters are all equally probable and occur at the same rate. In reality, certain character changes are more likely. For example, purine nucleotides are more frequently exchanged for another purine than for pyrimidines, and amino acids are more likely to be exchanged with those that share physicochemical properties or whose codon sequences overlap. The General Time Reversible (GTR) model relaxes all substitution constraints and allows exchangeabilities to be free parameters (*282*). However, GTR is computationally intensive, especially for protein sequences. Thus, empirical substitution models are often used in such cases. These models are constructed by estimating parameters for a large-scale dataset of sequence alignments. These pre-defined matrices of substitution rates and equilibrium frequencies, such as the LG matrix (*283*), can then be applied to model the evolution of other sequences. However, these models alone do not take into account differences in the rate of evolution across sites (*e.g.*, slow-evolving versus fast-evolving sites). This can be accounted for by applying a multiplication factor that varies according to a distribution. A gamma distribution is most commonly used, with sites binned into categories with equal rate probabilities (*284*). A generalized FreeRate distribution can also be used to estimate rate differences, by building a rate distribution directly from the data (*285*).

However, this still assumes that all sites evolve according to the same underlying substitution matrix. This is not realistic from a biological standpoint as sites have site-specific preferences. Take the example of an enzyme that binds a negatively charged molecule. Sites corresponding to the enzyme's active site may be more constrained to positively charged amino acids. Mixture models are more realistic models of evolution that allow sites to evolve under heterogeneous substitution processes. They use the same relative substitution rates, but with unique equilibrium frequency profiles, thus creating a set of different substitution matrices. In the CAT model, equilibrium frequency profiles are determined directly from the data (*286*). However, the CAT model is currently only tractable within a Bayesian framework. For ML phylogenies, mixture models with set empirically determined frequency profiles can instead be used (*e.g.*, the C10 to C60 models) (*287*).

A phylogeny can be inferred using different tree inference methods, which include ML and Bayesian (*280*). ML methods find the tree that maximizes likelihood: the probability of observing the given alignment with the selected parameters (*i.e.*, the tree topology, its branch lengths, and the substitution model) (*279, 280, 288*). The landscape of all possible trees is often too large to fully explore, as the number of possible trees increases dramatically with sequence number. For example, a tree with 5 taxa has 15 alternative topologies, but one with 10 taxa has over 2 million, plus all possible branch

lengths. Heuristic algorithms are used to probe tree space and to find peaks that maximize the likelihood, hopefully identifying the global optimum (*280*).

Bayesian methods are iterative and make use of a "prior" – in the case of phylogenetics the prior probability distribution of all possible trees. This prior is updated to a posterior probability distribution based on the observed data (*i.e.*, the probability that each tree produced the observed data multiplied by the prior probability of that tree), which then becomes the new prior for the next iteration (*289*). The posterior probability distribution cannot be computed directly since both branch lengths and model parameters are continuous variables. The distribution is thus approximated by sampling the tree space using Markov chain Monte Carlo (MCMC) algorithms (*289, 290*). MCMC chains tend to move "uphill" (towards solutions with high posterior probabilities) in tree and parameter space and can get stuck around local optima. To increase the chance of finding the global optimum multiple chains are run, which converge when sampling the same optimum.

In Bayesian phylogenies posterior probability indicates branch support, as it represents the percentage of sampled trees that a particular branch occurs in. For ML trees branch support is often inferred by measuring their robustness to site resampling with bootstrapping. In Felsenstein's bootstrap proportions (FBP) alignment sites are resampled with replacement to obtain pseudo-alignments (*291*). Trees for each pseudo-alignment are then inferred using the same inference method as the original tree. Branch support is measured as the proportion of pseudo-trees in which each branch from the original reference tree is reconstructed. However, branch support can be low when phylogenetically unstable "rogue taxa" are included. Removing rogue taxa that move across branches in trees can improve phylogenetic accuracy (*292*). In addition, other approaches to phylogenetic bootstraps have been proposed to tackle these situations, such as the transfer bootstrap expectation (TBE) (*293*). TBE likewise estimates branch robustness in bootstrap trees but does so using a gradual "transfer" distance (*i.e.*, counting the number of taxa that would need to be transferred to either side of the branch). Consequently, this metric indicates high support for branches that are essentially (even if not identically) recovered in bootstrap trees. The TBE value can be seen at the fraction of stable taxa that do not move across a branch.

Inferring these nonparametric bootstrap trees can be computationally restrictive. Bootstrap approximation methods with parametric sampling are thus often used to obtain clade support, such as the ultrafast bootstrap approximation approach (*294*). Recently, a new posterior mean site frequency (PMSF) model has been developed, which provides a rapid approximation of profile mixture models given an input tree (*295*). PMSF thus allows for the computational and time feasibility to infer nonparametric bootstraps using these more complex substitution models with large concatenated alignments. Another option for assessing branch support is to use single branch tests, such as the SH-like approximate likelihood ratio test (*296*).

7.2 Phylogenomics

In the context of this thesis, phylogenomics refers to the use of genome-scale data (*i.e.*, multiple genes) to infer evolutionary history. Considerations for inferring phylogenomic trees have been the subject of a number of informative reviews (*281, 297-300*). Genes can have complex evolutionary histories, with gene-specific events obscuring underlying species relationships. Thus phylogenies of individual genes often do not mirror species evolution. Furthermore, single-gene trees typically fail to resolve deep relationships due to limited data. By inferring phylogenomic trees using a supermatrix of concatenated genes (or proteins), more information is incorporated thereby improving phylogenetic signal. Additionally, phylogenetic signal is averaged over a larger set of genes, providing a buffer to conflict introduced by divergent histories of individual genes. Nevertheless, using genes that have evolved through speciation events, and thus represent the vertical inheritance of organisms, is important for reconstructing accurate species trees.

Gene orthologs used in species reconstructions are often referred to as marker genes. Ideally, they are conserved for the reconstruction of deep evolutionary events, and found in a single copy across considered taxa. This can be more difficult when using incomplete data, as is often the case with genomes from uncultivated lineages. Marker genes can either be identified *de-novo* for a given dataset or by using a pre-defined set of reference genes to identify orthologs (*281*). Supermatrix approaches to phylogenomics combine data from multiple orthologous genes by concatenating them into a longer alignment. Inferring individual alignments for each gene separately prior to concatenation helps with identifying true homologous sites.

Concatenation can mitigate the effects of including a non-orthologous gene. However, their inclusion can still lead to error, particularly for short ancestral branches (*299*). Quality control of included genes is needed to prevent such data errors. By generating and inspecting individual gene trees prior to concatenation outliers can be identified. Such curation helps for detecting paralogous genes, recent duplications, and HGTs (*299*). Nevertheless, hidden paralogy can still be an issue (*301*). In this case an ancestral polymorphism can lead to gene trees discordant with the species tree, despite their appearance as orthologs in extant lineages (*302*). Error caused by the inclusion of genes that violate orthology assumptions, alongside additional errors discussed below, can result in erroneous species trees (*299, 301, 302*).

7.3 Phylogenetic artefacts

Random stochastic error (*i.e.*, sampling error) can result from insufficient phylogenetic information across the sites of a finite dataset, which can be a limiting factor in resolving trees for single genes (*297*). In species trees,

stochastic error is generally accounted for by concatenating genes into longer alignments, which increases the number of positions and thus information. However, concatenation can also magnify the effects of systematic error and result in the incorrect species tree being statistically supported (*299*). Systematic error stems from violations of assumptions in the models of evolution used for phylogenetic tree reconstruction (*299, 303, 304*). Often these arise from models assuming homogenous processes in the evolution of sequences, when in fact the process is heterogeneous. For example, evolutionary rates can be heterogeneous and vary through time at the same position in an alignment (heterotachy) (*305*). Such within-site variation is typically not accounted for due to computational limitations. Fast-evolving sites and taxa, and those divergent in their character composition, are more likely to violate model assumptions and introduce systematic error.

Long-branch attraction (LBA) artefacts result in long distantly-related branches grouping together in phylogenetic trees (*306*). LBA artefacts arise from underestimating the number of substitutions in sequences that have undergone a large amount of change and hence have convergent substitutions. This blurred phylogenetic signal can lead to lineages being incorrectly inferred as more closely related than they are in reality. Fast-evolving lineages and those that have anciently diverged are more often affected, since multiple substitutions are more likely to have occurred at the same site. Many endosymbionts are fast-evolving and elevated substitution rates in some groups have been verified through experimental evolution (*307, 308*). Other sources of systematic error are sequence composition biases. Models assume that sequence composition is homogeneous across taxa when it can vary. This can result in the artefactual grouping of unrelated lineages based on similarities in composition patterns (*300*). Endosymbionts are commonly AT-rich, which can cause them to be artefactually grouped together (*307, 309*). This has also been found in salt-adapted organisms, through acidified amino acid composition, such as Haloarchaea and Methanonatronarchaeia (*310*).

Using more realistic heterogeneous models that better fit the data can help to alleviate both LBA and compositional bias by preventing model misspecifications (*295, 303, 311*). However, this may not always be sufficient. Long-branching or compositionally biased taxa can also be removed from reconstructions, but this is not possible if they are of interest in a given study. Data transformations and site removal can also be used to alter the data so that it no longer violates models while preserving underlying phylogenetic signal. This can involve the removal of fast-evolving or heterogeneous sites and sequence recoding (*303, 312*). For example, heterogeneous sites can be identified and removed by their contribution to a protein alignments $\chi 2$-score, and hence bias relative to the overall distribution of amino acids (*17, 313*). Site removal and recoding decrease the number of informative sites and phylogenetic signal. However, across a large concatenated alignment the effect can be small. Transformed alignments can still violate models and

correct tree inference is not guaranteed, but they can lessen the degree of violation and thus help to reduce artefacts.

Taxon selection is also an important consideration for inferring species trees. Due to computational limitations, the number of taxa included needs to be carefully considered. Yet, the selection of taxa should include a broad sampling of representatives. Rich taxon sampling can also help to break long branches and counteract LBA (*297, 304*). However, it may not always be possible to increase taxon sampling. Due to the specific histories of speciation and extinction, some clades are simply taxon-poor and other groups are poorly sampled. Missing data can also be an issue (*297, 299, 314*). It is of particular concern when inferring species phylogenies with MAGs and SAGs that are highly incomplete. Missing data can result in taxa being inaccurately placed due to a lack of informative sites. In addition, it can exacerbate systematic error by reducing the detection of multiple substitutions. Using a more complete narrow selection of genes instead of a wider set with high incompleteness can improve phylogenetic reconstructions (*314*).

7.4 Ancestral state reconstruction

Ancestral state reconstruction (ASR) refers to the reconstruction of ancestral states based on characteristics of modern individuals or populations (*315*). ASR can be used to recover all kinds of biological character states of ancestors, such as amino acid and nucleotide sequences, genome composition including gene order, phenotypes and morphological traits, and even geographic range (*315*). Reconstruction methods rely on two main data inputs: a rooted species phylogeny to indicate how taxa are related by descent and a profile of character states across the given species. ASR can be used to infer ancestral gene content and to follow gene family evolution across a given set of species, as applied in this thesis. This allows gene-specific events such as duplications, losses, transfers, and originations (*de novo* genes or outside transfers) to be recovered. Reconstructions of microbial genome evolution are routinely performed using a numerical profile of gene family abundances in extant taxa, and a species tree (*316*) (Figure 7). However, such reconstruction methods are unaware of the underlying gene tree phylogeny. This can lead to the incorrect inference of events and does not allow for the direction of transfers to be established.

Gene-tree-aware ASR approaches incorporate gene trees and reconcile them with the given species tree (*317*) (Figure 7). By incorporating gene trees, information is gained about the specific evolutionary history of that gene in the context of the species tree (Figure 7). Gene-tree-aware approaches work using amalgamation, where all gene tree reconciliation scenarios with a given species tree are explored using a limited starting set of trees for each gene (for example bootstrap trees). Some tools incorporate gene tree uncertainty and

directly estimate rates of evolutionary events (losses, transfers, and duplications) for each gene (*318, 319*). Statistical support for events is given by the proportion of reconciled trees. HGT from unsampled and extinct lineages can also be accounted for (*320*).

Since gene-tree-aware ASR is based on phylogenetic results it is not immune from reconstruction artefacts in the underlying phylogenetic trees. These methods do not account for species tree uncertainty and an accurate inference is needed. Many gene families will have short alignments, and the resulting trees can suffer from sampling error. Gene trees are also considered separately when some genes co-evolve. For example, those found in operons and that function in complexes and which tend to be transferred together. Although transfers are not allowed to occur with ancestors, they can still be time-inconsistent when using an undated tree. ASR methods can infer gene "originations", but not whether they represent transfers from outside taxa or *de novo* gene births. Despite these assumptions and potential issues, ASR is becoming more accurate, particularly when gene trees are taken into account. ASR has recently been used to identify HGT among fungi, reconstruct early archaeal ancestors, and to probe the influx of bacterial gene content during a methanogen-to-halophile transition (*310, 317, 321*). Using ASR we can gain a glimpse into the genomic toolkit and lifestyles of microbial ancestors.

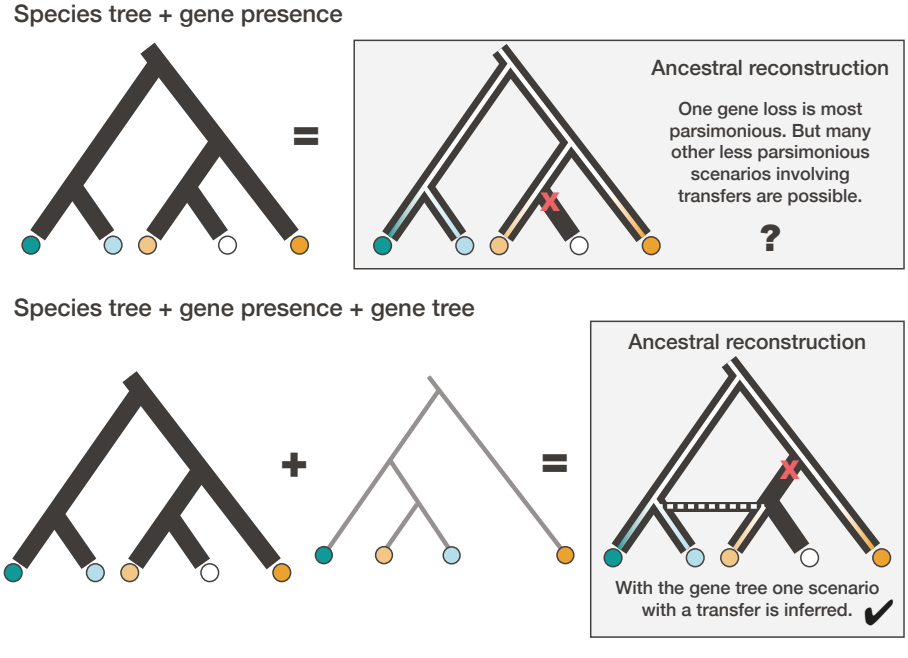

Figure 7. Given a species tree (black), and presence and absence patterns of a gene across taxa (circles), alternative ancestral reconstructions of the evolutionary history of a gene could be inferred. By including gene tree information (grey tree), this is narrowed down to one option involving a transfer event.

8 Main findings

Chlamydial environmental diversity
A wider perspective was gained on the environmental distribution and relative abundance of Chlamydiae. Chlamydiae are diverse and have high relative abundance in both marine sediments and the sponge-associated microbiome (papers I, IV). These chlamydial lineages are phylogenetically diverse, spanning known groups, but also including previously unidentified lineages (papers I, IV). Relatives of the sponge microbiome-associated chlamydiae are found in marine ecosystems, including other invertebrates (paper IV). Generally, diverse and abundant chlamydiae are also found in various additional environments including freshwater, groundwater, salt marshes, and wastewater (paper I). Using metagenomics, genomic diversity within the Chlamydiae phylum was expanded by nearly doubling sequenced species, with 42 high-quality chlamydial MAGs (papers I, IV).

Chlamydiae species relationships and taxonomy
Phylogenomic methods that accounted for potential artefacts were employed to determine species relationships within Chlamydiae (papers I, III, IV). Novel chlamydial clades were identified that corresponded to higher-level taxonomic ranks (papers I). Prominent among these were the Anoxychlamydiales and Chlamydiae Clade IV (paper I). Moreover, previously poorly sampled groups gained additional genomic representation (papers I, IV). In addition, these species trees showed that the animal pathogens Chlamydiaceae did not diverge early and that they share a common ancestor with species from marine environments (papers I, IV). The phylogenetic position of several long-branching and orphan taxa was also resolved, challenging prior conceptions about their early-diverging nature (papers I, III, IV). The majority of well-studied chlamydiae, and most who are cultivated, formed one half of the chlamydial species tree, while the second half was primarily composed of uncultivated lineages (paper I, III). This hints at unexplored environmental roles and biology of chlamydiae.

Ecology and environmental impacts of uncultivated chlamydiae
All chlamydiae genomes were found to have genes for traits associated with symbiosis and with the typical chlamydial biphasic lifecycle (paper I, III). This included various secretion systems (*e.g.*, the T3SS), NTTs and a likely ATP transporter, and proteins involved in EB formation (paper I). However,

unexpected genomic features were also identified. Chlamydiae Clade IV members encode flagellar genes (paper I), and several groups have the genetic potential to use additional carbon sources and to degrade diverse compounds (paper I, IV). The more widespread presence of biosynthesis-related genes was also found, including for *de novo* biosynthesis of amino acids, nucleotides, and cofactors (paper I, IV). Biosynthetic gene clusters for producing secondary metabolites were also newly identified in Chlamydiae, which may indicate defensive symbioses with host eukaryotes (paper IV). On the other hand, likely eukaryotic hosts were not found in anoxic marine sediment metagenomes (paper I). It is thus possible that some chlamydial lineages are not obligately associated with eukaryotes and can instead engage in symbiotic interactions with other members of their microbial communities (paper I). The Anoxychlamydiales, for example, appear to be fermentative anaerobes that may engage in syntropy (paper II). Anoxychlamydiales genomes encoded genes for anaerobiosis-associated metabolism including oxygen-sensitive proteins for producing hydrogen (paper II).

Ancestral state reconstruction of the Chlamydiae phylum

Genes related to symbiotic interactions and the chlamydial biphasic lifecycle were found to have already been present in the last Chlamydiae common ancestor (paper III). This suggests that the ancestor of Chlamydiae was already able to interact with and infect eukaryotic hosts, supporting the long-term conservation of the characteristic chlamydial lifestyle (paper III). The ancestor of Chlamydiae was reconstructed as a motile, fermentative, facultative anaerobe, indicating that it could move between oxic and anoxic environments (paper III). Chlamydiae gene content evolution was found to have been marked by extensive gene gain through interspecies transfer and HGT (paper III). In particular, amoeba-associated chlamydiae appear to have undergone genome expansion counter to many other endosymbionts (paper III).

Chlamydiae and eukaryotic evolution

The metabolism found in Anoxychlamydiales mirrored that of anaerobic eukaryotes (paper II). Phylogenetic analyses showed that several key proteins related to hydrogen production may have been donated from an Anoxychlamydiales ancestor to eukaryotes during eukaryotic evolution (paper II). Large-scale investigation of chlamydial genetic contributions to eukaryotes identified two overarching patterns of putative gene exchange (paper V). Genes that indicate chlamydial HGTs with LECA, and chlamydial HGTs with the ancestor of plastid-bearing eukaryotes (paper V). However, phylogenetic patterns were complex, and further analyses are necessary to better resolve the evolutionary history of these genes (paper V).

9 Paper summaries

Paper I. Diverse chlamydiae from marine sediments

Project goals
This study focused on examining newly identified chlamydiae from anoxic deep marine sediments. SSU rRNA gene amplicon sequencing was used to assess the relative abundance and diversity of these chlamydiae, and genome-resolved metagenomics to obtain sequenced representatives. Comparative genomics was then employed to investigate the lifestyles and metabolism of newly identified lineages relative to those previously sequenced.

Key results
Chlamydiae were identified at various depths (0.1 to 9.4 mbsf) in marine sediments from the Arctic Mid-Ocean Ridge. Based on SSU rRNA gene amplicons chlamydiae were diverse and had particularly high relative abundance in anoxic layers. These marine sediment lineages both span previously identified groups and form new clades.

Metagenome sequencing resulted in the retrieval of 24 MAGs with high completeness. In phylogenomic analyses of Chlamydiae species relationships seven clades of high taxonomic rank were resolved with distinct patterns in gene content. Three of these groups, Anoxychlamydiales, and Chlamydiae Clades II and IV (CC-II and CC-IV), were solely or primarily composed of newly identified lineages.

Chlamydiaceae and CC-IV share a common ancestor, revealing that Chlamydiaceae pathogenicity likely evolved after their divergence. They also share a small set of conserved protein families, that appear to be gene inventions with potential importance in their lifestyles. CC-IV also encode flagella, which had only rarely been found in chlamydiae previously.

Based on their gene content relative to known groups, marine sediment chlamydiae may use additional carbon sources and have less auxotrophy for nucleotide and amino acid biosynthesis. Despite this, all marine sediment chlamydiae encode genomic features consistent with a symbiotic lifestyle. This includes genes related to EB formation, a T3SS system among other secretion systems, and multiple NTTs, including a likely ATP transporter.

Despite gene features consistent with host association, we were unable to identify eukaryotic hosts that could explain chlamydial diversity and relative

abundance in these marine sediments. Furthermore, these lineages appeared to be actively replicating. In addition, we noted that diverse and abundant chlamydiae are found in various other environments.

Significance
Previously unidentified, diverse, relatively abundant, and active chlamydiae are found in anoxic marine sediments that may play previously overlooked ecological roles in this environment. The Chlamydiaceae have marine sediment relatives and all chlamydiae encode genetic features consistent with symbiosis. Nevertheless, some chlamydiae may not be obligate intracellular symbionts of eukaryotes.

Paper II. Discovery of anaerobic chlamydiae

Project goals
In paper I, nearly half of the MAGs consistently formed a well-supported clade of high taxonomic rank in phylogenomic analyses. This study focused on characterizing this newly identified marine sediment lineage, the Anoxychlamydiales. Here, their metabolism was reconstructed and the evolutionary origins of these genes in Anoxychlamydiales were further examined.

Key results
Anoxychlamydiales had high relative abundance, and were dominant members of the microbial community in some anoxic marine sediment metagenomes, indicated their potential importance. Examination of gene content unique to the Anoxychlamydiales among chlamydiae resulted in the identification of metabolic genes associated with anaerobiosis and strongly supportive of an anaerobic lifestyle.

Anoxychlamydiales appear to be hydrogen and acetate producing fermentative anaerobes. They lack a complete respiratory chain and TCA cycle and instead encode genes for conserving energy through glycolysis, the arginine deiminase pathway, and pyruvate fermentation to acetate. They also encode a trimeric electron-confurcating [FeFe]-hydrogenase to regenerate reducing equivalents for this fermentative metabolism and the three hydrogenase maturase genes necessary for generating its unique iron cluster. Other genes consistent with anaerobiosis were also identified, such as a ferrous iron transport system.

In phylogenetic reconstructions many of the anaerobiosis-associated genes were found to be most closely related to homologs from diverse anaerobic prokaryotes. However, in the case of the three hydrogenase maturase proteins and pyruvate:ferredoxin oxidoreductase, sequences from Anoxychlamydiales

branched sister to eukaryotic homologs with high support. Furthermore, tree topologies were indicative of vertical inheritance in Anoxychlamydiales, while patterns among eukaryotic sequences supported both organismal relationships and HGT. These findings are consistent with an early acquisition of these proteins from a chlamydial ancestor during eukaryotic evolution, potentially before LECA or eukaryotic diversification.

Significance
Some chlamydiae are anaerobes and may live in syntropy with hydrogen or acetate-consuming prokaryotes instead of eukaryotic hosts. Moreover, chlamydiae may have previously unrecognized impacts on marine sediment ecology and biogeochemistry. In addition, a chlamydial ancestor may have contributed hydrogen-metabolism-related genes to eukaryotes, supporting a mosaic evolutionary origin of eukaryotic metabolism.

Paper III. Ancestral state reconstruction of the Chlamydiae phylum

Project goals
In this study, gene-tree-aware ancestral state reconstruction was used to examine gene content evolution in Chlamydiae. Genome dynamics during chlamydial evolution and routes of internal transfers were examined. Gene gains were further investigated and putative HGTs identified. The metabolism and lifestyles of key chlamydial and PVC ancestors were also inferred based on reconstructed gene content.

Key results
Phylogenomic analyses were used to reconstruct a robust species tree of Chlamydiae using an expanded dataset of diverse lineages, many sequenced through cultivation-independent approaches. Updates to chlamydial taxonomy were proposed due to consistent evolutionary relationships.

Complex gene event dynamics were found in the ancestral reconstruction of Chlamydiae gene content evolution. This included substantial losses in the last common ancestor of Chlamydiae and the ancestors of several families. Rates of interspecies gene transfer varied with clear networks of transfer within and between chlamydial families, suggesting overlapping environmental niches. Chlamydiaceae and Anoxychlamydiales were conspicuously minimal in this network indicating niche specialization.

In particular, we noted that some chlamydiae ancestors have also evolved through gene content expansion, counter to many other endosymbionts. Through HGT, the ancestor of amoeba-infecting chlamydiae gained several

ETC complexes related to the establishment of a proton motive force. HGTs also occurred to other chlamydiae ancestors, often coinciding with evolutionary transitions in lifestyle. Reconstruction of the presence of genes encoding proteins with varying oxygen tolerance indicated a divide within chlamydiae.

In addition, the ancestor of the Chlamydiae phylum was reconstructed as a facultative anaerobe, with fermentative metabolism, and a flagellum. Since its divergence from other PVCs this ancestor lost hydrogen metabolism and amino acid biosynthesis capabilities. It also gained an extensive gene suite related to host interaction, and for acquiring energy-rich molecules and metabolites from exogenous sources.

Significance

Based on the reconstruction of genes present in the last common ancestor of Chlamydiae, we propose it was already capable of endosymbiosis in eukaryotic hosts. This suggests the conservation of the chlamydial intracellular lifestyle over at least a billion years of evolutionary history. Despite this long-term association, chlamydial gene content has not degraded and some have expanded in gene content size. Furthermore, we propose that the chlamydial biphasic lifecycle emerged from life stage transitions of their facultatively anaerobic ancestor between oxic and anoxic environments.

Paper IV. Sponge microbiome-associated chlamydiae

Project goals

Chlamydiae were previously identified in the microbiomes of the sponge species *Halichondria panicea*, *Haliclona oculata*, and *Haliclona xena* (*181*). In this study, we reassessed this finding using SSU rRNA gene amplicon sequencing and genome-resolved metagenomics. Phylogenomics and comparative genomic analyses were used to characterize these chlamydial lineages. Their gene content was further examined for indications of mutualism or parasitism and their environmental distribution was investigated.

Key results

A high relative abundance of Chlamydiae in these sponge species was confirmed based on amplicons. However, this was only found in samples collected at the same date and location as in the prior study (*181*). This indicates transient chlamydial associations or increases in relative abundance.

We obtained 18 high-quality draft chlamydial genomes that affiliated with four distinct families in reconstructions of Chlamydiae species relationships, two of which had few previous representatives. Patterns in central metabolic

gene content were similar between sponge microbiome-associated chlamydiae and other members of their respective families.

Sponge microbiome-associated chlamydiae genomes were enriched in genes related to fermentation and the degradation of diverse compounds, including genes found in other sponge microbiome members. Some also encode genes that are typically eukaryotic. Most surprisingly, polyketide synthase and non-ribosomal peptide synthetase genes were identified, which are central to the biosynthesis of secondary metabolites. Further examination revealed the widespread presence of biosynthetic gene clusters related to secondary metabolite production, including antibiotics, across the Chlamydiae phylum.

It is unclear if these sponge microbiome-associated chlamydiae are sponge symbionts, or rather if they are symbionts of another eukaryote associated with the sponge. Investigations of environmental prevalence showed that relatives of these chlamydial lineages are found in other marine invertebrates and related ecosystems, and that sponge-associated chlamydiae are distinct from other well-studied chlamydial groups.

Significance

Chlamydiae were newly revealing as potential producers of secondary metabolites. Given the intracellular nature of chlamydiae we were surprised to find these genes. Based on this we propose that some chlamydiae could act as defensive mutualistic symbionts of the intracellular niche of their hosts. Overall, these findings suggest chlamydial impacts on marine invertebrates and ecosystems that warrant further scrutiny.

Paper V. Chlamydiae and eukaryotic evolution

Project goals

In this study, chlamydiae genetic contributions to eukaryotic evolution were assessed using the recently increased sampling of chlamydiae and eukaryotic genomic diversity. LUCA protein family orthologous groups were found in common between chlamydiae and eukaryotes. Phylogenetic analyses were then used to systematically screen for clades composed of chlamydiae and eukaryotic sequences in these protein family trees. Patterns of inheritance and chlamydial and eukaryotic groups represented in these clades were then manually examined. Potential HGT events were identified and examined for functional and evolutionary factors favoring HGT between these groups.

Key results

A total of 1466 protein families were found in common between chlamydiae and eukaryotes. Phylogenetic analyses and manual screening of these protein

families resulted in the identification of 107 potential HGT events between chlamydiae and eukaryotes.

In these resulting trees, patterns in affiliation between chlamydial and eukaryotic sequences were complex, and the direction of transfer was often unclear. Nevertheless, two major avenues of potential HGT were identified, between chlamydiae and the ancestor of Archaeplastida, and chlamydiae and LECA. Indications of additional, more recent, avenues of HGT between chlamydiae and eukaryotes were also found.

Many of the identified genes were related to metabolism. In addition, we identified several likely HGTs in protein families with functions related to central informational processes.

Significance

Our results are consistent with a genetic contribution from chlamydiae to plastid-bearing eukaryotes. Although, we do not find additional support for the MAT hypothesis. In addition, these results are supportive of potential genetic contributions from chlamydiae to LECA. Further phylogenetic analysis is necessary to clarify and establish these genetic affiliations. Gene exchange between these two groups may have been facilitated by the presence of chlamydiae in the eukaryotic intracellular niche.

10 Concluding remarks and future perspectives

In summary, the work presented in this thesis shows that there is greater environmental, genomic, and metabolic diversity within the Chlamydiae phylum than previously thought. Supported by cultivation-independent methods, knowledge has been gained about several previously unidentified chlamydial groups. This updated view on the Chlamydiae phylum is now less biased by a dependence on co-cultivation.

Overall, this thesis further argues that chlamydiae are of research interest beyond a medical perspective and that there is much insight to be had from studying environmental lineages. We have shown how chlamydiae are relevant from environmental, ecological, biotechnological, and evolutionary perspectives. Further study is certain to bring additional surprising findings.

Despite the increase in sequenced chlamydiae presented here, there is still much work to be done. In this thesis chlamydial genomic diversity was increased through the investigation of only two environments. Given the phylogenetic diversity known to exist based on environmental surveys, future analyses are highly likely to identify new groups. Such investigations are needed for a comprehensive understanding of chlamydial biology and evolutionary history. Importantly, future studies are also necessary to confirm genetic inferences made about chlamydial lifestyles and metabolisms. Ideally, representatives of newly identified chlamydial groups would be obtained in culture. Co-culturing with various eukaryotic hosts could be attempted. In the case of the putatively syntrophic Anoxychlamydiales, anoxic enrichment may also be possible. However, cultivation may prove challenging, and no host-free cultures of chlamydiae have previously been obtained.

In any case, further analysis of environmental samples would help to prove or disprove several key proposals presented in this thesis. In particular, cultivation-free microscopy approaches, such as fluorescence in situ hybridization, could be used to visualize chlamydiae directly. Doing so would help us to determine if these chlamydiae have eukaryotic hosts, who these hosts or other interaction partners are, where any interactions lie along the pathogen-mutualist continuum and could confirm if they undergo the chlamydial biphasic lifestyle. In addition, using other omics techniques, such as transcriptomics, would help to determine whether genes are expressed and how this is affected by changes in environmental parameters.

By further expanding the Chlamydiae tree, exciting insights into chlamydiae genome diversity and evolution are sure to follow.

Popular science summary

Most species on Earth are microbes that are invisible to us and can only be seen using a microscope. These microorganisms are important for our health, industries, and environment. But we are unable to grow most of them in a lab environment and the majority remain undiscovered. One of these underexplored groups is Chlamydiae. Chlamydiae are bacteria, that alongside archaea are referred to as prokaryotes. We humans on the other hand are eukaryotes, a group that includes all life visible to the naked eye, such as animals, plants, and fungi, but also many microbes like amoeba. Chlamydiae are infamous for their most notorious member, *Chlamydia trachomatis*, which causes the sexually transmitted infection. Different types of chlamydiae are responsible for other diseases in humans and other animals too, including some types of pneumonia and eye infections. Some chlamydiae even infect microbial eukaryotes, such as amoeba. However, not all chlamydiae are harmful pathogens. Some form symbioses that instead benefit their hosts.

Symbiosis is when two or more species form long-term relationships with each other. It occurs between species across the tree of life. When one partner in a symbiosis causes harm to the other and gains benefits, it is referred to as a parasite or pathogen. On the other hand, when both partners benefit, both are called mutualists. These symbioses can occur between larger organisms, known as hosts, and smaller organisms called symbionts. All studied chlamydiae live inside the cells of their host eukaryotes as symbionts.

Symbiosis has been important for evolution. For instance, eukaryotes evolved around two billion years ago from a symbiosis between an archaeal cell and a bacterial cell. This is thought to have taken place in environments with little oxygen, with one partner providing hydrogen that the other consumed. The bacterial partner would later evolve into the mitochondrion, the powerhouse of the cell. Today, the mitochondria in some eukaryotes can produce hydrogen, instead of oxygen like ours do. Symbiosis also led to the evolution of all plant and algal life. Over one billion years ago, a symbiosis between a photosynthetic bacterium and a eukaryote led to the evolution of organisms with chloroplasts, which harness the sun's energy using photosynthesis. In both key symbioses, the partner organisms' genomes or collection of genetic material were expanded with genes transferred by other species. But the donors of many of these genes remain a mystery. By providing genes Chlamydiae may have played a role in eukaryote evolution.

While much is known about chlamydial animal pathogens, our understanding of chlamydiae on a wider scale is still limited because most of

their diversity has not been explored. But lots of chlamydiae infect microbial eukaryotes instead of animals, and many more have unknown lifestyles. In addition, there has been a lack of research into chlamydiae evolutionary history and their contributions to eukaryote evolution. This thesis addresses these gaps through research described in five papers. These papers focus on newly discovered chlamydiae from the ocean seafloor (paper I), newly discovered chlamydiae in the microbiomes of sponges (paper IV), chlamydiae evolution and their ancestors (paper III), chlamydiae that can live without oxygen and contributed these genes to eukaryotes (paper II), and genes that chlamydiae have generally donated to eukaryotes throughout their evolution (paper V).

Paper I focuses on newly discovered chlamydiae that our team found in the ocean seafloor. While exploring microbial diversity in the seafloor 3000 m deep in the North Atlantic Ocean we happened to find chlamydiae. To our surprise, these chlamydiae were dominant members of the microbial community. They were also extremely diverse, including groups that had not been discovered before. Several of the chlamydiae found even share a common ancestor with chlamydiae that infect humans and other animals. This tells us that chlamydiae animal pathogens have distant relatives in the ocean seafloor. When we examined the genomes of these chlamydiae we found genes that pointed to a symbiont lifestyle. But we were not able to find host eukaryotes in this environment. This led us to propose that some chlamydiae can live without a host.

In Paper II, we took a closer look at one of these groups that was extremely abundant in the seafloor. In their genomes we found genes for producing hydrogen and energy without oxygen. Based on what was previously known about chlamydiae this was a highly unexpected lifestyle. These groups may form symbioses with other prokaryotes where they exchange hydrogen, which is an essential gas in the metabolism of microbes that impact the environment. When we looked at the evolution of the hydrogen-related chlamydial genes, we found that they were most closely related to these genes in eukaryotes. Our findings suggest that chlamydiae donated these key genes to eukaryotes during their evolution. This could have been important for the initial symbiosis that led to the evolution of eukaryotes. Today, these genes allow some eukaryotes to produce hydrogen and live without oxygen.

Paper III outlines our reconstruction of the ancestor of Chlamydiae and the path of evolution to the different chlamydial groups existing today. We found that the ancestor of all chlamydiae was able to infect host eukaryotes and live inside their cells. This ancestor could also live and move between environments with and without oxygen. We expected to find that genome size had decreased during chlamydiae evolution since many symbionts that live inside host eukaryotes have smaller genomes. Since these symbionts can gain nutrients from their hosts, over time they tend to lose the genes for producing those nutrients themselves. To our surprise, some chlamydiae genomes have

instead increased in size despite being symbionts. Some chlamydiae groups have also gained genes related to producing energy and living with oxygen.

In paper IV, we focused on abundant chlamydiae we found in the microbiomes of several sponge species. Sponges are animals that play an important role in the health of ocean and freshwater ecosystems. They often live in coral reefs and feed by filtering small particles from water. Like many other animals, sponges have microbiomes made up of their associated microbes, and which play roles in their nutrition and health. It was not known if the chlamydiae found in the sponge microbiomes are parasites or mutualists. So, we sequenced the genomes of these chlamydiae to find more clues. Based on an analysis of their genes, we found that these chlamydiae can produce and degrade more chemical compounds than expected. This included genes for making antibiotics. This led us to propose that these chlamydiae are producing antibiotics to defend their hosts against attack from pathogens. Chlamydiae may also be important in other marine ecosystems. When we surveyed various environments we found closely related chlamydiae associated with other marine invertebrate animals. Chlamydiae may be important symbionts that had previously gone unnoticed in these environments.

Paper V highlights our search for genes that both chlamydiae and eukaryotes share in common. To see if chlamydiae had transferred genes to eukaryotes we reconstructing the evolutionary history of each gene. When we looked at the resulting patterns, we found that chlamydiae may have transferred genes that aided in the evolution of plants and algae. We also found that chlamydiae may have transferred other genes throughout eukaryote evolution. This includes central genes found today across all eukaryotic life. But the patterns we saw were complicated. In many cases, the direction of gene donation was unclear, with some genes potentially transferred from eukaryotes to chlamydiae instead.

The papers covered in this thesis highlight the discovery of new chlamydial groups of likely importance for ecology in ocean environments –from the bottom of the seafloor to coral reefs. Even though these new chlamydial groups have different lifestyles, symbiosis with eukaryotes appears to have persisted for over a billion years. During this time, chlamydiae ancestors may also have helped eukaryotes to live without oxygen and plants and algae to evolve, by transferring genes. Chlamydiae are more than just animal pathogens, and relatives of the medically relevant groups are important members of diverse ecosystems. Overall, this thesis suggests unexpected roles and impacts of chlamydiae in various environments. Future research is needed to further untangle the true extent of chlamydial diversity and to uncover their influence in both eukaryote evolution, and the lives of eukaryotes today as symbionts.

Svensk sammanfattning

De flesta arter som finns här på jorden är osynliga för oss och kan bara ses med ett mikroskop. Dessa mikroorganismer är viktiga för vår hälsa, inom industrin och för miljön. De flesta av alla mikrober går inte att odla i labb och majoriteten är fortfarande oupptäckta. En av dessa mindre utforskade mikrobiella grupperna heter Chlamydiae. Chlamydiae är bakterier som tillsammans med Archaea ingår i en större grupp som kallas prokaryoter. Vi människor tillhör istället en annan grupp som kallas eukaryoter, vilken inkluderar allt liv synligt för blotta ögat, såsom djur, växter och svampar, men också många mikrober såsom amöbor. Chlamydiae har generellt sett ett dåligt rykte på grund av sina mest ökända medlemmar, så som Chlamydia trachomatis som orsakar en vanlig sexuellt överförbar infektion. Olika typer av Chlamydiae kan leda till andra sjukdomar hos både människor och djur, till exempel vissa typer av lunginflammation eller ögoninfektioner. Vissa arter av Chlamydiae infekterar till och med mikrobiella eukaryoter, såsom amöbor. Däremot är inte alla Chlamydiae skadliga patogener, utan vissa kan ingå i symbioser som istället är till fördel för sin värdorganism.

En symbios är när två arter ingår i ett långsiktigt förhållande. Det sker mellan arter längsmed hela livets träd. När en part i en symbios är skadlig för den andra men själv drar nytta kallas den för parasit eller patogen. Om istället båda parterna drar nytta av symbiosen kallas de för mutualister. Dessa symbioser kan ske mellan större organismer (värdar), och mindre organismer (symbionter). All studerad Chlamydiae lever inuti cellerna hos sina värdeukaryoter som symbionter. Symbioser har varit viktiga för evolutionen. Eukaryoter utvecklades för ungefär två miljarder år sedan från en symbios mellan en arkée och en bakterie. Detta antas ha skett i syrefattig miljö, där ena parten skapade väte som den andra konsumerade. Den bakteriella symbionten utvecklades senare till något som kallas en mitokondrie, som fungerar som den eukaryota cellens kraftverk. Idag kan mitokondrierna hos vissa eukaryoter producera väte i stället för syre som våra mitokondrier producerar. Symbios ledde också till utvecklingen av allt växt- och algliv. För över en miljard år sedan ledde en symbios mellan en fotosyntetisk bakterie och en eukaryot till utvecklingen av organismer med kloroplaster, som utnyttjar solens energi med hjälp av fotosyntes. I båda av dessa viktiga symbioser expanderades partnernas arvsmassa med gener som överförts från andra arter. Men donatorerna av många av dessa gener är fortfarande ett mysterium. Genom att tillhandahålla gener kan Chlamydiae ha spelat en roll i eukaryoters utveckling.

Även om mycket är känt om de typer av Chlamydiae som agerar som patogener för djur, är vår generella förståelse av Chlamydiae fortfarande begränsad eftersom det mesta av deras mångfald ännu inte undersökts. Många typer av Chlamydiae infekterar till exempel mikrobiella eukaryoter istället för djur, och många fler har okända livsstilar. Dessutom har det undersökts väldigt lite kring den evolutionära historien hos Chlamydiae och deras bidrag till eukaryoters evolution. Denna avhandling behandlar dessa luckor genom forskning som beskrivs i fem artiklar. Dessa artiklar fokuserar på nyupptäckta Chlamydiae från havsbottnen (artikel I), nyupptäckta Chlamydiae i svampdjurs mikrobiom (artikel IV), evolutionen hos Chlamydiae och deras förfäder (artikel III), Chlamydiae som kan leva utan syre och som bidrog med dessa gener till eukaryoter (artikel II), samt gener som Chlamydiae generellt har donerat till eukaryoter under loppet av deras evolution (artikel V).

Artikel I fokuserar på nyligen upptäckta Chlamydiae som vår grupp hittade i havsbottnen. När vi utforskade mikrobiell mångfald i havsbottnen 3000 m under ytan i Nordatlanten fann vi av en slump Chlamydiae. Förvånansvärt nog dominerades antalet medlemmar i detta mikrobiella samhälle av Chlamydiae och vi såg en stor variation av sådana arter som inte tidigare upptäckts. Även om de påträffade arterna av Chlamydiae hittades i en helt annan miljö än de som infekterar djur, delar dessa grupper ändå en gemensam förfader. När vi undersökte arvsmassan hos dessa Chlamydiae fann vi gener som pekade på en symbiont livsstil, men vi kunde inte hitta någon potentiell eukaryotvärd. Detta fick oss att föreslå att vissa arter inom Chlamydiae kan leva utan en värd.

I artikel II tittade vi närmare på en av dessa grupper inom Chlamydiae som fanns i överflöd i havsbottnen. När vi tittade på deras arvsmassa hittade vi gener för att producera väte och energi, samt gener som gör det möjligt att leva utan syre. Baserat på vad som tidigare var känt om Chlamydiae var detta en mycket oväntad livsstil. Denna grupp inom Chlamydiae skulle möjligtvis kunna ingå en symbios med andra prokaryoter där de producerar väte åt den andra parten, vilket är en viktig gas i metabolismen hos mikrober som påverkar miljön. När vi undersökte evolutionen av dessa väterelaterade gener fann vi att kopiorna i eukaryoter verkar kunna ha kommit från Chlamydiae. Denna upptäckt tyder på att Chlamydiae donerade dessa gener till eukaryoter under deras utveckling. Detta kan ha varit viktigt för den initiala symbios som ledde till evolutionen av eukaryoter.

Artikel III beskriver vår rekonstruktion av förfäderna till Chlamydiae och den väg evolutionen tagit som har lett till de grupper som finns idag. Vi fann att förfadern till alla Chlamydiae kunde infektera och leva i eukaryota värdceller. Denna förfader kunde också leva i och röra sig mellan miljöer både med och utan syre. Vi förväntade oss att storleken av dess arvsmassa hade minskat under evolutionen av Chlamydiae då många symbionter som lever i värdeukaryoter har mindre arvsmassa. Eftersom dessa symbionter kan få näringsämnen från sina värdar tenderar de att över tid också förlora de gener som behövs för att själva kunna producera dessa näringsämnen. Till vår

förvåning har vissa arters arvsmassa faktiskt ökat i storlek även fast de är symbionter. Vissa grupper inom Chlamydiae har också fått gener relaterade till att producera energi och leva med syre.

I artikel IV fokuserade vi på Chlamydiae som vi hittade i mångfald i mikrobiomer hos flera arter av svampdjur. Svampdjur, även kallade tvättsvampar, är som namnet antyder djur, och de spelar en viktig roll för hälsan för havs- och sötvattensekosystem. De lever ofta i korallrev och livnär sig på små partiklar de filtrerar ut från vattnet runtomkring. Likt många andra djur har svampdjur också mikrobiom som består av deras associerade mikrober och som spelar roll i deras näring och hälsa. Det var inte känt ifall de Chlamydiae som hittades i svampdjurens mikrobiom är parasiter eller mutualister, så vi sekvenserade arvsmassan hos dessa Chlamydiae för att hitta fler ledtrådar. Baserat på en analys av deras gener fann vi att dessa Chlamydiae kan producera och bryta ner fler kemiska föreningar än väntat. Detta inkluderade gener för att göra antibiotika. Detta fick oss att föreslå att dessa Chlamydiae producerar antibiotika för att försvara sina värdar mot angrepp från patogener. Chlamydiae kan också vara viktiga i andra marina ekosystem. När vi undersökte olika miljöer hittade vi nära besläktade Chlamydiae associerade med andra marina ryggradslösa djur. Chlamydiae kan vara viktiga symbionter som tidigare gått obemärkt förbi i dessa miljöer.

Artikel V belyser vår sökning efter gener som Chlamydiae och eukaryoter har gemensamt. För att se ifall Chlamydiae överfört gener till eukaryoter rekonstruerade vi varje gens evolutionära historia. Från de resulterande mönstren fann vi att Chlamydiae kan ha överfört gener som hjälpt till i växters och algers evolution. Vi fann också att Chlamydiae kan ha överfört andra gener under loppet av evolutionen av eukaryoter. Detta inkluderar centrala gener som idag finns hos allt eukaryotiskt liv. Dock var de mönster vi såg komplicerade. I många fall var riktningen för donationen av gener oklar, med vissa gener som potentiellt istället överförts från eukaryoter till Chlamydiae.

De artiklar som behandlas i denna avhandling belyser upptäckten av nya grupper inom Chlamydiae som är av sannolik betydelse för ekologi i havsmiljöer - från djupt nere i havsbottnen till korallrev. Även om dessa nya grupper inom Chlamydiae har olika livsstilar så verkar symbioser med eukaryoter ha fortlevt i över en miljard år. Under denna tid kan förfäder till Chlamydiae genom överföring av gener också ha hjälpt eukaryoter att leva utan syre samt hjälpt växter och alger att utvecklas. Chlamydiae är mer än bara patogener för djur, och faktum är att släktingar till de medicinskt relevanta grupperna är viktiga medlemmar i olika ekosystem. Sammantaget antyder denna avhandling till oväntade roller och effekter av Chlamydiae i olika miljöer. Framtida forskning behövs för att ytterligare reda ut den verkliga omfattningen av mångfalden hos Chlamydiae och för att avgöra deras inflytande i både evolutionen av eukaryoter och eukaryoternas nutida liv som symbionter.

Acknowledgements

Throughout these past five years of working towards my PhD, I have been incredibly lucky to have a fantastic support network of mentors and friends. I am truly thankful to all of you, and could not have done this without your help and guidance, both socially and in science.

First and foremost, I would like to thank my supervisor **Thijs** for taking me on this journey. It's been an exciting, fun, and sometimes difficult ride, that's been made much smoother by your encouragement and advice. I appreciate our great discussions, and how you've helped me grow in independence. I feel fortunate to have had your support and am deeply grateful for these PhD years. It has been an absolute pleasure being a member of your team!

At different stages during my PhD I have also been supported by inspiring co-supervisors from whom I have learned so much. **Anja**, thank you for talking me into pursuing a PhD in the first place. It was your curiosity that led to my first project, and I strive to emulate your passion and excitement for science and taking a closer look! **Courtney**, thank you for pushing me to continue improving, for your enthusiasm, and your thoughtfulness. Your excellent office company brightened every day at the lab. **Dani**, thank you for your kindness, positivity, and willingness to lend a helping hand. I've really appreciated how we could chat about science and life for hours. To all my supervisors, thank you for your knowledge and for taking the time to pass on a little bit to me!

During my doctoral studies I've been fortunate to have shared the experience with some amazing fellow PhD students. **Eva**, **Max**, **Joran**, **Anders**, **Henning**, **Jolanda**, and **Laura W**. it has been wonderful going on this adventure with you! Thanks for sharing the ups, downs, and for all the fun in-between. **Disa**, you are missed and know you'd be proud. I would also like to thank all of the other current and former members of the lab who've provided me with mentorship and feedback, **Laura E.**, **Lionel**, **Jimmy**, **Kasia, Will,** and **Patricia**, your knowledge and insight have been instrumental, and I very much appreciate having had your support. To **Lina**, **Claudia**, and **Anna-Maria**, thanks for your all-knowing expertise and help in the lab. To **Tom** and **Martha**, even though it didn't end up in the thesis, thanks for the Iceland sampling trip escapade, and your all-around help! To **Felix**, thank you for keeping the computers running, and stopping me from panicking when they got angry. To all the other past members of the lab, including **Mahwash**, **Dries**, **Jonathan**, **Jun-Hoe**, **Fabian**, **Julian**, and **Anna K**., thank you for the

lunches, the afterworks, the Svenska fikas, and your all around enjoyable company! **Natalia** and **Christina**, I feel lucky that I got to mentor such great students, thank you for letting me learn with you! Also, thank you **Anders** and **Max**, (and **Martha**, the only person I know who can sample sediment cores on crutches) for the great times leading the functional genomics course sampling lab together. To the other current members of the lab in Wageningen, **Guillaume**, **Burak**, **Kassiani**, **Eric**, and **Hans**, thanks for being great virtual company. For those I haven't met in person yet, I hope I can visit soon and change that! In general, to everyone I met at WUR, during my brief stint in **Wageningen**, thanks for the microbes, Belgian beer sessions, and bike rides!

I've also been lucky to work with exceptional collaborators on various projects. **Stephan**, it has been fantastic working with you, both online and in person, these past few years. Learning and exploring with you has been a highlight of the PhD. Got chlamydia? **Matthias**, **Astrid**, and **Paul**, thank you for making me part of the chlamydiae family, and for taking me under your wing at conferences. **Detmer**, it's been great learning about sponges, thanks for all the helpful advice. **Karin**, thank you for getting me into sponges in the first place, your enthusiasm is infectious! **Steffen**, thank you for a great time in Bergen, and for staying up too late extracting DNA with me.

I'm also incredibly thankful to the great group of people at Molecular Evolution. **Siv**, **Lisa**, **Jan**, and **Christian**, thank you for acting as mentors and for all the insightful questions and input that helped me along the way. I'd also like to thank all the other members of Molevo, past and present, for the fun throughout the years, from Swedish smörgåsbord, to Spanish cheese, to BBQs, to eating around the world, to eating insects (we like food don't we?). Thank you **Andrea**, **Alejandro**, **Anna O.**, **Ellie**, **Emil**, **Erik**, **Gäelle**, **Guilhe**, **Julia**, **Jun-Hoe**, **Karl**, **Kristina**, **Mayank**, **Wei**, and **Zeynep**! Also, thank you **Ellie** and **Dani**, for starting Darwin at the Museum with me, and for sharing your enthusiasm for science and outreach! I'm also grateful to everyone at ICM, for the great retreats, beer clubs, and science. To **Staffan** and all the other giardians (**Laura R.**, **Sascha**, **Jana**, **Feifei**, **Asgeir**, **Showgy**, and **Jon)**, thanks for taking me in and letting me share in the diplomonad fun for a bit! **Konrad** and **Laura L**. it was great being on the ICM board with you. **Emil**, thanks for always organizing the best parties and for teaching me things about ribosomes (which is definitely what you work with). **Christoffer**, thank you for the morning swimming and lovely chats. **Javier**, it's been quite the journey, and I feel so happy to have had your friendship, and your amazing laugh.

A big thank you to **Alannah**, **Andrea**, **Anna O.**, **Courtney**, **Dani**, **David**, **Eva**, **Javier**, **Max**, **Patricia**, and **Stephan**, for your feedback, proofreading, and translating of various parts of this thesis. **Eva** and **Andrea**, it was incredibly fun living with you both at different points during the PhD, thank you for the laughs and hugs! **David**, thank you for supporting and believing in me, for holding your thumbs for me, and for keeping me sane during this isolating corona-time (and for helping me Swedish!).

I may have come to Uppsala for the science, but what kept me coming back was the truly fantastic people. To **all my friends** who have made life here so fun (even in the long dark winters), tusen tack! It's been a joy spending time and having laughs with you at all the fikas, picnics, nations, pub quizzes, wine Wednesdays, kayaking, and midsummers dancing like a frog. Here's to all the adventures we've been on together, both here and in Stockholm, Gotland, Kiruna, Gothenburg, Cesky Krumlov, Paris, Montreal, Vermont, Niagara, Greenbelt, Reykjavik, Helsinki, Tallinn, Seattle, Charlotte, Leipzig, and Frankfurt (by accident). Also, great to have all the awesome MEMEs along! All of this has come together to make the last five years truly memorable. To everyone who has been a part of these years in Uppsala and who helped make the PhD time the constant adventure that it was, thank you! To all those further afield, I've missed you, and can't wait until we can celebrate together again!

On that note, I would also like to thank all of my friends and family back in Canada, for being there for me now and throughout the years. To my parents (**Farid** and **Yvonne**), thank you for putting up with my childhood "but why's?", and instilling in me a zest to find out more. Your enthusiasm and encouragement for me to follow my dreams is how I'm here today, and I'm grateful from the bottom of my heart. To my sisters (**Alannah** and **Farrah**), thank you for being such great early debate partners, and for always being there for me. To the rest of my family, thanks for the community, can't wait for the non-Zoom famjam soon! To the **Markham**, and **Western** crews, I'm so happy to be able to count on your continued friendship, here's to all the adventures life has for us in the years to come!

Finally, to anyone reading this, thank you for taking the time to do so, and I hope you enjoy!

Tack så mycket allihopa!!! ☺

References

1. H. C. Betts *et al.*, Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol* **2**, 1556-1562 (2018).
2. M. T. Rosing, 13C-Depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from west greenland. *Science* **283**, 674-676 (1999).
3. A. P. Nutman, V. C. Bennett, C. R. L. Friend, M. J. Van Kranendonk, A. R. Chivas, Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* **537**, 535-+ (2016).
4. M. S. Dodd *et al.*, Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* **543**, 60-+ (2017).
5. A. H. Knoll, M. A. Nowak, The timetable of evolution. *Sci Adv* **3**, e1603076 (2017).
6. L. Eme, S. C. Sharpe, M. W. Brown, A. J. Roger, On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb Perspect Biol* **6**, (2014).
7. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307-315 (2014).
8. N. J. Planavsky, S. A. Crowe, M. Fakhraee, Evolution of the structure and impact of Earth's biosphere. *Nat Rev Earth Environ* **2**, 123–139 (2021).
9. S. W. Poulton *et al.*, A 200-million-year delay in permanent atmospheric oxygenation. *Nature*, (2021).
10. D. E. Canfield, A new model for Proterozoic ocean chemistry. *Nature* **396**, 450-453 (1998).
11. W. Martin, M. Müller, The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37-41 (1998).
12. D. Moreira, P. Lopez-Garcia, Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* **47**, 517-530 (1998).
13. A. Spang *et al.*, Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol* **4**, 1138-1148 (2019).
14. H. Imachi *et al.*, Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519-+ (2020).
15. P. Lopez-Garcia, D. Moreira, The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat Microbiol* **5**, 655-667 (2020).
16. A. Spang *et al.*, Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-179 (2015).
17. J. Martijn, J. Vosseberg, L. Guy, P. Offre, T. J. G. Ettema, Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101-105 (2018).
18. M. Muller *et al.*, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* **76**, 444-495 (2012).

19. C. W. Stairs, M. M. Leger, A. J. Roger, Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Philos T R Soc B* **370**, (2015).
20. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proc Natl Acad Sci U S A* **115**, 6506-6511 (2018).
21. H. C. Flemming, S. Wuertz, Bacteria and archaea on Earth and their abundance in biofilms. *Nature Reviews Microbiology* **17**, 247-260 (2019).
22. K. J. Locey, J. T. Lennon, Scaling laws predict global microbial diversity. *Proc Natl Acad Sci U S A* **113**, 5970-5975 (2016).
23. J. T. Lennon, K. J. Locey, More support for Earth's massive microbiome. *Biol Direct* **15**, 5 (2020).
24. N. Lane, The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'. *Philos T R Soc B* **370**, (2015).
25. C. Darwin, *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. (John Murray, London, 1859).
26. M. A. O'Malley, What did Darwin say about microbes, and how did microbiology respond? *Trends Microbiol* **17**, 341-347 (2009).
27. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088-5090 (1977).
28. C. R. Woese, O. Kandler, M. L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576-4579 (1990).
29. D. J. Lane *et al.*, Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* **82**, 6955-6959 (1985).
30. N. R. Pace, A molecular view of microbial diversity and the biosphere. *Science* **276**, 734-740 (1997).
31. M. S. Rappe, S. J. Giovannoni, The uncultured microbial majority. *Annu Rev Microbiol* **57**, 369-394 (2003).
32. K. G. Lloyd, A. D. Steen, J. Ladau, J. Yin, L. Crosby, Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *Msystems* **3**, (2018).
33. C. J. Castelle, J. F. Banfield, Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* **172**, 1181-1197 (2018).
34. L. A. Hug *et al.*, A new view of the tree of life. *Nat Microbiol* **1**, 16048 (2016).
35. B. J. Baker *et al.*, Diversity, ecology and evolution of Archaea. *Nat Microbiol* **5**, 887-900 (2020).
36. F. Burki, A. J. Roger, M. W. Brown, A. G. B. Simpson, The New Tree of Eukaryotes. *Trends Ecol Evol* **35**, 43-55 (2020).
37. S. J. Sibbald, J. M. Archibald, Genomic Insights into Plastid Evolution. *Genome Biology and Evolution* **12**, 978-990 (2020).
38. A. E. Murray *et al.*, Roadmap for naming uncultivated Archaea and Bacteria. *Nat Microbiol* **5**, 987-994 (2020).
39. R. A. Sanford, K. G. Lloyd, K. T. Konstantinidis, F. E. Loffler, Microbial Taxonomy Run Amok. *Trends Microbiol*, (2021).
40. D. H. Parks *et al.*, Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533-1542 (2017).
41. K. T. Konstantinidis, R. Rossello-Mora, R. Amann, Uncultivated microbes in need of their own taxonomy. *Isme Journal* **11**, 2399-2406 (2017).
42. D. H. Parks *et al.*, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004 (2018).

43. X. Didelot, A. S. Walker, T. E. Peto, D. W. Crook, D. J. Wilson, Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* **14**, 150-162 (2016).
44. R. Hershberg, Mutation--The Engine of Evolution: Studying Mutation and Its Role in the Evolution of Bacteria. *Cold Spring Harb Perspect Biol* **7**, a018077 (2015).
45. E. Darmon, D. R. Leach, Bacterial genome instability. *Microbiol Mol Biol Rev* **78**, 1-39 (2014).
46. D. I. Andersson, J. Jerlstrom-Hultqvist, J. Nasvall, Evolution of new functions de novo and from preexisting genes. *Cold Spring Harb Perspect Biol* **7**, (2015).
47. S. B. Van Oss, A. R. Carvunis, De novo gene birth. *PLoS Genet* **15**, e1008160 (2019).
48. C. M. Weisman, A. W. Murray, S. R. Eddy, Many, but not all, lineage-specific genes can be explained by homology detection failure. *Plos Biology* **18**, (2020).
49. T. Gabaldon, E. V. Koonin, Functional and evolutionary implications of gene orthology. *Nat Rev Genet* **14**, 360-366 (2013).
50. S. M. Soucy, J. Huang, J. P. Gogarten, Horizontal gene transfer: building the web of life. *Nat Rev Genet* **16**, 472-482 (2015).
51. L. Boto, Horizontal gene transfer in the acquisition of novel traits by metazoans. *P Roy Soc B-Biol Sci* **281**, (2014).
52. S. J. Sibbald, L. Eme, J. M. Archibald, A. J. Roger, Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. *Trends Parasitol* **36**, 927-941 (2020).
53. F. Husnik, J. P. McCutcheon, Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* **16**, 67-79 (2018).
54. X. Didelot, M. C. Maiden, Impact of recombination on bacterial evolution. *Trends Microbiol* **18**, 315-322 (2010).
55. R. G. Beiko, T. J. Harlow, M. A. Ragan, Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* **102**, 14332-14337 (2005).
56. C. Pal, B. Papp, M. J. Lercher, Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37**, 1372-1375 (2005).
57. W. F. Doolittle, Phylogenetic classification and the universal tree. *Science* **284**, 2124-2129 (1999).
58. O. X. Cordero, M. F. Polz, Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12**, 263-273 (2014).
59. G. M. Bennett, N. A. Moran, Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol* **5**, 1675-1688 (2013).
60. K. Han *et al.*, Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. *Sci Rep* **3**, 2101 (2013).
61. R. Albalat, C. Canestro, Evolution by gene loss. *Nature Reviews Genetics* **17**, 379-391 (2016).
62. S. J. Giovannoni, J. Cameron Thrash, B. Temperton, Implications of streamlining theory for microbial ecology. *ISME J* **8**, 1553-1565 (2014).
63. D. J. Martinez-Cano *et al.*, Evolution of small prokaryotic genomes. *Frontiers in Microbiology* **5**, (2015).
64. N. A. Moran, Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **93**, 2873-2878 (1996).
65. N. A. Moran, G. M. Bennett, The Tiniest Tiny Genomes. *Annual Review of Microbiology, Vol 68* **68**, 195-215 (2014).
66. W. D. Orsi, Ecology and evolution of seafloor and subseafloor microbial communities. *Nat Rev Microbiol*, 671–683 (2018).

67. S. J. Giovannoni, SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Ann Rev Mar Sci* **9**, 231-255 (2017).
68. J. J. Morris, R. E. Lenski, E. R. Zinserc, The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* **3**, e00036-00012 (2012).
69. G. D'Souza *et al.*, Less Is More: Selective Advantages Can Explain the Prevalent Loss of Biosynthetic Genes in Bacteria. *Evolution* **68**, 2559-2570 (2014).
70. G. D'Souza *et al.*, Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat Prod Rep* **35**, 455-488 (2018).
71. P. O. Methot, S. Alizon, What is a pathogen? Toward a process view of host-parasite interactions. *Virulence* **5**, 775-785 (2014).
72. P. J. Keeling, J. P. McCutcheon, Endosymbiosis: The feeling is not mutual. *Journal of Theoretical Biology* **434**, 75-79 (2017).
73. D. M. Baker, C. J. Freeman, J. C. Y. Wong, M. L. Fogel, N. Knowlton, Climate change promotes parasitism in a coral symbiosis. *ISME J* **12**, 921-930 (2018).
74. N. Dombrowski, J. H. Lee, T. A. Williams, P. Offre, A. Spang, Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *Fems Microbiol Lett* **366**, (2019).
75. B. E. Morris, R. Henneberger, H. Huber, C. Moissl-Eichinger, Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* **37**, 384-406 (2013).
76. A. Kouzuma, S. Kato, K. Watanabe, Microbial interspecies interactions: recent findings in syntrophic consortia. *Front Microbiol* **6**, 477 (2015).
77. A. E. Lind *et al.*, Genomes of two archaeal endosymbionts show convergent adaptations to an intracellular lifestyle. *Isme Journal* **12**, 2655-2667 (2018).
78. M. T. Mee, J. J. Collins, G. M. Church, H. H. Wang, Syntrophic exchange in synthetic microbial communities. *Proc Natl Acad Sci U S A* **111**, E2149-2156 (2014).
79. T. R. Costa *et al.*, Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* **13**, 343-359 (2015).
80. A. Diepold, J. P. Armitage, Type III secretion systems: the bacterial flagellum and the injectisome. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20150020 (2015).
81. J. B. Raina, V. Fernandez, B. Lambert, R. Stocker, J. R. Seymour, The role of microbial motility and chemotaxis in symbiosis. *Nature Reviews Microbiology* **17**, 284-294 (2019).
82. P. Major, T. M. Embley, T. A. Williams, Phylogenetic Diversity of NTT Nucleotide Transport Proteins in Free-Living and Parasitic Bacteria and Eukaryotes. *Genome Biol Evol* **9**, 480-487 (2017).
83. M. Bright, S. Bulgheresi, A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* **8**, 218-230 (2010).
84. Y. J. Shi *et al.*, The Ecology and Evolution of Amoeba-Bacterium Interactions. *Appl Environ Microb* **87**, (2021).
85. S. Sudakaran, C. Kost, M. Kaltenpoth, Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends Microbiol* **25**, 375-390 (2017).
86. E. B. Van Arnam, C. R. Currie, J. Clardy, Defense contracts: molecular protection in insect-microbe symbioses. *Chem Soc Rev* **47**, 1638-1651 (2018).
87. A. Apprill, Marine Animal Microbiomes: Toward Understanding Host-Microbiome Interactions in a Changing Ocean. *Front Mar Sci* **4**, (2017).
88. A. E. Douglas, The microbial exometabolome: ecological resource and architect of microbial communities. *Philos T R Soc B* **375**, (2020).

89. T. J. Hammer, J. G. Sanders, N. Fierer, Not all animals need a microbiome. *Fems Microbiol Lett* **366**, (2019).
90. F. Schulz, M. Horn, Intranuclear bacteria: inside the cellular control center of eukaryotes. *Trends Cell Biol* **25**, 339-346 (2015).
91. J. P. McCutcheon, B. M. Boyd, C. Dale, The Life of an Insect Endosymbiont from the Cradle to the Grave. *Current Biology* **29**, R485-R495 (2019).
92. F. Husnik *et al.*, Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**, 1567-1578 (2013).
93. E. C. M. Nowack, M. Melkonian, Endosymbiotic associations within protists. *Philos T R Soc B* **365**, 699-712 (2010).
94. M. D. Solbach, M. Bonkowski, K. Dumack, Novel Endosymbionts in Rhizarian Amoebae Imply Universal Infection of Unrelated Free-Living Amoebae by Legionellales. *Front Cell Infect Microbiol* **11**, 642216 (2021).
95. C. Moliner, P. E. Fournier, D. Raoult, Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* **34**, 281-294 (2010).
96. J. M. Archibald, Endosymbiosis and Eukaryotic Cell Evolution. *Current Biology* **25**, R911-R921 (2015).
97. M. Wagner, M. Horn, The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* **17**, 241-249 (2006).
98. D. P. Devos, N. L. Ward, Mind the PVCs. *Environmental Microbiology* **16**, 1217-1221 (2014).
99. E. Rivas-Marin, D. P. Devos, The Paradigms They Are a-Changin': past, present and future of PVC bacteria research. *Anton Leeuw Int J G* **111**, 785-799 (2018).
100. J. C. Cho, K. L. Vergin, R. M. Morris, S. J. Giovannoni, Lentisphaera araneosa gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. *Environ Microbiol* **6**, 611-621 (2004).
101. S. Spring *et al.*, Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum. *ISME J* **10**, 2801-2816 (2016).
102. P. Hugenholtz, B. M. Goebel, N. R. Pace, Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**, 4765-4774 (1998).
103. J. Glockner *et al.*, Phylogenetic diversity and metagenomics of candidate division OP3. *Environ Microbiol* **12**, 1218-1229 (2010).
104. A. Collingro, S. Kostlbacher, M. Horn, Chlamydiae in the Environment. *Trends Microbiol* **28**, 877-888 (2020).
105. L. Fieseler, M. Horn, M. Wagner, U. Hentschel, Discovery of the novel candidate phylum "Poribacteria" in marine sponges. *Appl Environ Microb* **70**, 3724-3732 (2004).
106. J. Kamke *et al.*, The candidate phylum Poribacteria by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *Plos One* **9**, e87353 (2014).
107. M. Strous *et al.*, Missing lithotroph identified as new planctomycete. *Nature* **400**, 446-449 (1999).
108. J. G. Kuenen, Anammox bacteria: from discovery to application. *Nature Reviews Microbiology* **6**, 320-326 (2008).

109. P. F. Dunfield *et al.*, Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**, 879-882 (2007).
110. A. Pol *et al.*, Methanotrophy below pH 1 by a new Verrucomicrobia species. *Nature* **450**, 874-878 (2007).
111. A. P. Graca, R. Calisto, O. M. Lage, Planctomycetes as Novel Source of Bioactive Molecules. *Frontiers in Microbiology* **7**, (2016).
112. S. Wiegand, M. Jogler, C. Jogler, On the maverick Planctomycetes. *Fems Microbiology Reviews* **42**, 739-760 (2018).
113. E. Rivas-Marin, I. Canosa, D. P. Devos, Evolutionary Cell Biology of Division Mode in the Bacterial Planctomycetes-Verrucomicrobia-Chlamydiae Superphylum. *Frontiers in Microbiology* **7**, (2016).
114. J. W. Moulder, Why Is Chlamydia Sensitive to Penicillin in the Absence of Peptidoglycan. *Infect Agent Dis* **2**, 87-99 (1993).
115. M. Pilhofer *et al.*, Discovery of chlamydial peptidoglycan reveals bacteria with murein sacculi but without FtsZ. *Nat Commun* **4**, 2856 (2013).
116. G. W. Liechti *et al.*, A new metabolic cell-wall labelling method reveals peptidoglycan in Chlamydia trachomatis. *Nature* **506**, 507-510 (2014).
117. M. C. F. van Teeseling *et al.*, Anammox Planctomycetes have a peptidoglycan cell wall. *Nature Communications* **6**, (2015).
118. O. Jeske *et al.*, Planctomycetes do possess a peptidoglycan cell wall. *Nat Commun* **6**, 7116 (2015).
119. G. Liechti *et al.*, Pathogenic Chlamydia Lack a Classical Sacculus but Synthesize a Narrow, Mid-cell Peptidoglycan Ring, Regulated by MreB, for Cell Division. *Plos Pathogens* **12**, (2016).
120. Y. Abdelrahman, S. P. Ouellette, R. J. Belland, J. V. Cox, Polarized Cell Division of Chlamydia trachomatis. *Plos Pathogens* **12**, (2016).
121. S. Wiegand *et al.*, Cultivation and functional characterization of 79 planctomycetes uncovers their unique biology. *Nat Microbiol* **5**, 126-140 (2020).
122. C. Greening, T. Lithgow, Formation and function of bacterial organelles. *Nature Reviews Microbiology* **18**, 677-689 (2020).
123. T. Shiratori, S. Suzuki, Y. Kakizawa, K. I. Ishida, Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nat Commun* **10**, 5529 (2019).
124. M. Derrien, M. C. Collado, K. Ben-Amor, S. Salminen, W. M. de Vos, The mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. *Appl Environ Microb* **74**, 1646-1648 (2008).
125. M. Sait *et al.*, Genomic and Experimental Evidence Suggests that Verrucomicrobium spinosum Interacts with Eukaryotes. *Front Microbiol* **2**, 211 (2011).
126. V. Serra *et al.*, Morphology, ultrastructure, genomics, and phylogeny of Euplotes vanleeuwenhoeki sp. nov. and its ultra-reduced endosymbiont "Candidatus Pinguicoccus supinus" sp. nov. *Sci Rep-Uk* **10**, (2020).
127. A. Omsland, B. S. Sixt, M. Horn, T. Hackstadt, Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. *FEMS Microbiol Rev* **38**, 779-801 (2014).
128. A. Nunes, J. P. Gomes, Evolution, phylogeny, and molecular epidemiology of Chlamydia. *Infect Genet Evol* **23**, 49-64 (2014).
129. L. Halberstädter, S. von Prowazek, Uber Zelleinschlüsse parasitärer Natur beim Trachom. *Arbeiten aus dem Kaiserlichen Gesundheitsamte* Berlin **26**, 44–47 (1907).
130. J. W. Moulder, The relation of the psittacosis group (Chlamydiae) to bacteria and viruses. *Annu Rev Microbiol* **20**, 107–130 (1966).

131. J. W. Moulder, *The biochemistry of intracellular parasitism*. (University of Chicago Press, Chicago, IL, USA, 1962).
132. J. Tjaden *et al.*, Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy. *J Bacteriol* **181**, 1196-1202 (1999).
133. S. Schmitz-Esser *et al.*, ATP/ADP Translocases: a Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae. *J Bacteriol* **186**, 683-691 (2004).
134. R. S. Stephens *et al.*, Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science* **282**, 754-759 (1998).
135. Y. M. Abdelrahman, R. J. Belland, The chlamydial developmental cycle. *FEMS Microbiol Rev* **29**, 949-959 (2005).
136. C. Elwell, K. Mirrashidi, J. Engel, Chlamydia cell biology and pathogenesis. *Nat Rev Microbiol* **14**, 385-400 (2016).
137. C. E. Barry, S. F. Hayes, T. Hackstadt, Nucleoid condensation in Escherichia coli that express a chlamydial histone homolog. *Science* **256**, 377-379 (1992).
138. G. Christiansen, L. B. Pedersen, J. E. Koehler, A. G.Lundemose, S. Birkelund, Interaction between the Chlamydia trachomatis Histone Hl-Like Protein(Hcl) and DNA. *J Bacteriol* **175**, 1785-1795 (1993).
139. T. Hackstadt, W. J. Todd, H. D. Caldwell, Disulfide-mediated interactions of the chlamydial major outer membrane protein: role in the differentiation of chlamydiae? *J Bacteriol* **161**, 25-31 (1985).
140. A. Gitsels, S. Van Lent, N. Sanders, D. Vanrompay, Chlamydia: what is on the outside does matter. *Critical Reviews in Microbiology* **46**, 100-119 (2020).
141. J. C. Ferrell, K. A. Fields, A working model for the type III secretion mechanism in Chlamydia. *Microbes Infect* **18**, 84-92 (2016).
142. B. S. Sixt *et al.*, Metabolic features of Protochlamydia amoebophila elementary bodies--a link between activity and infectivity in Chlamydiae. *PLoS Pathog* **9**, e1003553 (2013).
143. S. Grieshaber *et al.*, Impact of Active Metabolism on Chlamydia trachomatis Elementary Body Transcript Profile and Infectivity. *J Bacteriol* **200**, (2018).
144. S. Haider *et al.*, Raman microspectroscopy reveals long-term extracellular activity of chlamydiae. *Molecular Microbiology* **77**, 687-700 (2010).
145. A. Omsland, J. Sager, V. Nair, D. E. Sturdevant, T. Hackstadt, Developmental stage-specific metabolic and transcriptional activity of Chlamydia trachomatis in an axenic medium. *Proceedings of the National Academy of Sciences* **110**, 1970-1970 (2013).
146. M. Pilhofer *et al.*, Architecture and host interface of environmental chlamydiae revealed by electron cryotomography. *Environ Microbiol* **16**, 417-429 (2014).
147. T. Pillonel, C. Bertelli, G. Greub, Environmental metagenomic assemblies reveal seven new highly divergent chlamydial lineages and hallmarks of a conserved intracellular lifestyle. *Front Microbiol* **9**, 79 (2018).
148. N. L. Bachmann, A. Polkinghorne, P. Timms, Chlamydia genomics: providing novel insights into chlamydial biology. *Trends Microbiol* **22**, 464-472 (2014).
149. J. Rowley *et al.*, Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. *B World Health Organ* **97**, 548-+ (2019).
150. G. Vogel, Infectious diseases - Tackling neglected diseases could offer more bang for the buck. *Science* **311**, 592-593 (2006).
151. H. R. Wright, A. Turner, H. R. Taylor, Trachoma. *Lancet* **371**, 1945-1954 (2008).

152. A. Taylor-Brown, L. Spang, N. Borel, A. Polkinghorne, Culture-independent metagenomics supports discovery of uncultivable bacteria within the genus Chlamydia. *Sci Rep* **7**, 10661 (2017).
153. I. M. Sigar *et al.*, Comparison of In Vitro Chlamydia muridarum Infection Under Aerobic and Anaerobic Conditions. *Curr Microbiol* **77**, 1580-1589 (2020).
154. S. Kahane, R. Gonen, C. Sayada, J. Elion, M. G. Friedman, Description and partial characterization of a new Chlamydia-like microorganism. *Fems Microbiol Lett* **109**, 329-333 (1993).
155. S. Kahane, E. Metzer, M. G. Friedman, Evidence That the Novel Microorganism-Z May Belong to a New Genus in the Family Chlamydiaceae. *Fems Microbiol Lett* **126**, 203-207 (1995).
156. R. Amann *et al.*, Obligate Intracellular Bacterial Parasites of Acanthamoebae Related to Chlamydia spp. *Appl Environ Microb* **63**, 115–121 (1997).
157. M. Horn, Chlamydiae as symbionts in eukaryotes. *Annu Rev Microbiol* **62**, 113-131 (2008).
158. A. Taylor-Brown, L. Vaughan, G. Greub, P. Timms, A. 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. *Pathog Dis* **73**, (2015).
159. F. Bayramova, N. Jacquier, G. Greub, Insight in the biology of Chlamydia-related bacteria. *Microbes and Infection* **20**, 432-440 (2018).
160. M. Horn *et al.*, Illuminating the evolutionary history of chlamydiae. *Science* **304**, 728-730 (2004).
161. A. Collingro *et al.*, Unity in variety--the pan-genome of the Chlamydiae. *Mol Biol Evol* **28**, 3253-3270 (2011).
162. C. Bertelli *et al.*, The Waddlia genome: a window into chlamydial biology. *Plos One* **5**, e10890 (2010).
163. C. Bertelli *et al.*, Sequencing and characterizing the genome of Estrella lausannensis as an undergraduate project: training students and biological insights. *Front Microbiol* **6**, 101 (2015).
164. A. Collingro *et al.*, Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. *ISME J* **11**, 2334-2344 (2017).
165. C. Bertelli *et al.*, CRISPR System Acquisition and Evolution of an Obligate Intracellular Chlamydia-Related Bacterium. *Genome Biol Evol* **8**, 2376-2386 (2016).
166. S. Kostlbacher, A. Collingro, T. Halter, D. Domman, M. Horn, Coevolving Plasmids Drive Gene Flow and Genome Plasticity in Host-Associated Intracellular Bacteria. *Curr Biol* **31**, 346-357 e343 (2021).
167. D. Baud, V. Thomas, A. Arafa, L. Regan, G. Greub, Waddlia chondrophila, a Potential Agent of Human Fetal Death. *Emerging Infectious Diseases* **13**, 1239-1243 (2007).
168. R. S. Gupta, S. Naushad, C. Chokshi, E. Griffiths, M. Adeolu, 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 (2015).
169. N. Borel, A. Polkinghorne, A. Pospischil, A Review on Chlamydial Diseases in Animals: Still a Challenge for Pathologists? *Vet Pathol* **55**, 374-390 (2018).

170. D. Corsaro, D. Venditti, Detection of novel Chlamydiae and Legionellales from human nasal samples of healthy volunteers. *Folia Microbiol (Praha)* **60**, 325-334 (2015).
171. M. I. Blandford, A. Taylor-Brown, T. A. Schlacher, B. Nowak, A. Polkinghorne, Epitheliocystis in fish: An emerging aquaculture disease with a global impact. *Transbound Emerg Dis* **65**, 1436-1446 (2018).
172. A. Taylor-Brown *et al.*, Metagenomic analysis of fish-associated Ca. Parilichlamydiaceae reveals striking metabolic similarities to the terrestrial Chlamydiaceae. *Genome Biol Evol*, 2587–2595 (2018).
173. P. Herrera *et al.*, Molecular causes of an evolutionary shift along the parasitism-mutualism continuum in a bacterial symbiont. *P Natl Acad Sci USA* **117**, 21658-21666 (2020).
174. M. Okude *et al.*, Distribution of amoebal endosymbiotic environmental chlamydia Neochlamydia S13 via amoebal cytokinesis. *Microbiol Immunol*, (2020).
175. C. Maita *et al.*, Amoebal endosymbiont Neochlamydia protects host amoebae against Legionella pneumophila infection by preventing Legionella entry. *Microbes and Infection* **20**, 236-244 (2018).
176. L. Konig *et al.*, Symbiont-Mediated Defense against Legionella pneumophila in Amoebae. *Mbio* **10**, (2019).
177. I. Lagkouvardos *et al.*, Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the Chlamydiae. *ISME J* **8**, 115-125 (2014).
178. F. Schulz *et al.*, Towards a balanced view of the bacterial tree of life. *Microbiome* **5**, 140 (2017).
179. T. S. Haselkorn *et al.*, Novel Chlamydiae and Amoebophilus endosymbionts are prevalent in wild isolates of the model social amoebae Dictyostelium discoideum. *bioRxiv* **2020.09.30.320895**, (2020).
180. O. Israelsson, Chlamydial symbionts in the enigmatic Xenoturbella (Deuterostomia). *J Invertebr Pathol* **96**, 213-220 (2007).
181. M. A. Naim *et al.*, Host-specific microbial communities in three sympatric North Sea sponges. *FEMS Microbiol Ecol* **90**, 390-403 (2014).
182. D. B. Goldsmith *et al.*, Comparison of microbiomes of cold-water corals Primnoa pacifica and Primnoa resedaeformis, with possible link between microbiome composition and host genotype. *Sci Rep* **8**, 12383 (2018).
183. A. Taylor-Brown, D. Madden, A. Polkinghorne, Culture-independent approaches to chlamydial genomics. *Microb Genom* **4**, (2018).
184. T. Pillonel *et al.*, Sequencing the obligate intracellular Rhabdochlamydia helvetica within its tick host Ixodes ricinus to investigate their symbiotic relationship. *Genome Biol Evol*, 1334-1344 (2019).
185. K. Anantharaman *et al.*, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219 (2016).
186. A. J. Probst *et al.*, Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* **3**, 328-336 (2018).
187. L. V. Alteio *et al.*, Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. *Msystems* **5**, (2020).
188. M. Ortiz *et al.*, A genome compendium reveals diverse metabolic adaptations of Antarctic soil microorganisms. *bioRxiv* **2020.08.06.239558**, (2020).
189. J. Maire *et al.*, Intracellular bacteria are common and taxonomically diverse in cultured and in hospite algal endosymbionts of coral reefs. *ISME J*, (2021).

190. K. D. E. Everett, R. M. Bush, A. A. Andersen, 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. *International Journal of Systematic Bacteriology* **49**, 415-440 (1999).
191. R. S. Stephens, G. Myers, M. Eppinger, P. M. Bavoil, Divergence without difference: phylogenetics and taxonomy of Chlamydia resolved. *Fems Immunol Med Mic* **55**, 115-119 (2009).
192. K. Sachse *et al.*, Emendation of the family Chlamydiaceae: proposal of a single genus, Chlamydia, to include all currently recognized species. *Syst Appl Microbiol* **38**, 99-103 (2015).
193. T. Pillonel, C. Bertelli, N. Salamin, G. Greub, Taxogenomics of the order Chlamydiales. *Int J Syst Evol Microbiol* **65**, 1381-1393 (2015).
194. A. Subtil, A. Collingro, M. Horn, Tracing the primordial Chlamydiae: extinct parasites of plants? *Trends Plant Sci* **19**, 36-43 (2014).
195. O. K. Kamneva, S. J. Knight, D. A. Liberles, N. L. Ward, Analysis of genome content evolution in pvc bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. *Genome Biol Evol* **4**, 1375-1390 (2012).
196. F. S. L. Brinkman *et al.*, Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res* **12**, 1159-1167 (2002).
197. G. Greub, D. Raoult, History of the ADP/ATP-Translocase-Encoding Gene, a Parasitism Gene Transferred from a Chlamydiales Ancestor to Plants 1 Billion Years Ago. *Appl Environ Microb* **69**, 5530-5535 (2003).
198. J. Huang, J. P. Gogarten, Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biology* **8**, (2007).
199. B. Becker, K. Hoef-Emden, M. Melkonian, Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evol Biol* **8**, 203 (2008).
200. A. Moustafa, A. Reyes-Prieto, D. Bhattacharya, Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *Plos One* **3**, e2205 (2008).
201. S. G. Ball *et al.*, Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? *Plant Cell* **25**, 7-21 (2013).
202. F. Facchinelli, C. Colleoni, S. G. Ball, A. P. Weber, Chlamydia, cyanobiont, or host: who was on top in the menage a trois? *Trends Plant Sci* **18**, 673-679 (2013).
203. U. Cenci *et al.*, Biotic Host-Pathogen Interactions As Major Drivers of Plastid Endosymbiosis. *Trends Plant Sci* **22**, 316-328 (2017).
204. D. Domman, M. Horn, T. M. Embley, T. A. Williams, Plastid establishment did not require a chlamydial partner. *Nat Commun* **6**, 6421 (2015).
205. H. Qiu *et al.*, Assessing the bacterial contribution to the plastid proteome. *Trends Plant Sci* **18**, 680-687 (2013).
206. A. M. G. Novak Vanclova *et al.*, Metabolic quirks and the colourful history of the Euglena gracilis secondary plastid. *New Phytol* **225**, 1578-1592 (2020).
207. N. Knie, M. Polsakiewicz, V. Knoop, Horizontal gene transfer of chlamydial-like tRNA genes into early vascular plant mitochondria. *Mol Biol Evol* **32**, 629-634 (2015).
208. S. Manna, A. Harman, Horizontal gene transfer of a Chlamydial tRNA-guanine transglycosylase gene to eukaryotic microbes. *Mol Phylogenet Evol* **94**, 392-396 (2016).

209. T. Hackstadt, W. Baehr, Y. Ying, Chlamydia trachomatis developmentally regulated proteinis homologous to eukaryotic histone H1. *PNAS* **88**, 3937-3941 (1991).
210. K. D. Everett, S. Kahane, R. M. Bush, M. G. Friedman, An unspliced group I intron in 23S rRNA links Chlamydiales, chloroplasts, and mitochondria. *J Bacteriol* **181**, 4734-4740 (1999).
211. A. A. Pittis, T. Gabaldon, Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101-104 (2016).
212. J. T. Staley, A. Konopka, Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**, 321-346 (1985).
213. H. P. Browne *et al.*, Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543-+ (2016).
214. G. Lax *et al.*, Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature* **564**, 410-414 (2018).
215. W. H. Lewis, G. Tahon, P. Geesink, D. Z. Sousa, T. J. G. Ettema, Innovations to culturing the uncultured microbial majority. *Nat Rev Microbiol*, (2020).
216. E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, The Third Revolution in Sequencing Technology. *Trends Genet* **34**, 666-681 (2018).
217. E. A. Franzosa *et al.*, Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology* **13**, 360-372 (2015).
218. I. C. Macaulay *et al.*, G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* **12**, 519-522 (2015).
219. V. Torsvik, L. Ovreas, T. F. Thingstad, Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296**, 1064-1066 (2002).
220. S. Davila-Ramos *et al.*, A Review on Viral Metagenomics in Extreme Environments. *Front Microbiol* **10**, 2403 (2019).
221. A. Obiol *et al.*, A metagenomic assessment of microbial eukaryotic diversity in the global ocean. *Mol Ecol Resour* **20**, (2020).
222. T. O. Delmont *et al.*, Reconstructing rare soil microbial genomes using in situ enrichments and metagenomics. *Frontiers in Microbiology* **6**, (2015).
223. Y. Suzuki *et al.*, Deep microbial proliferation at the basalt interface in 33.5-104 million-year-old oceanic crust. *Commun Biol* **3**, (2020).
224. C. Schrader, A. Schielke, L. Ellerbroek, R. Johne, PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* **113**, 1014-1026 (2012).
225. T. O. Delmont *et al.*, Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* **77**, 1315-1324 (2011).
226. M. Albertsen, S. M. Karst, A. S. Ziegler, R. H. Kirkegaard, P. H. Nielsen, Back to Basics - The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. *Plos One* **10**, (2015).
227. L. M. Steinberg, J. M. Regan, Phylogenetic comparison of the methanogenic communities from an acidic, oligotrophic fen and an anaerobic digester treating municipal wastewater sludge. *Appl Environ Microbiol* **74**, 6663-6671 (2008).
228. N. R. Pace, Mapping the Tree of Life: Progress and Prospects. *Microbiol Mol Biol R* **73**, 565-576 (2009).
229. L. D. Crosby, C. S. Criddle, Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *Biotechniques* **34**, 790-+ (2003).
230. F. Bonk, D. Popp, H. Harms, F. Centler, PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. *J Microbiol Methods* **153**, 139-147 (2018).

231. P. Yarza *et al.*, Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**, 635-645 (2014).
232. A. Klindworth *et al.*, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**, (2013).
233. E. A. Eloe-Fadrosh, N. N. Ivanova, T. Woyke, N. C. Kyrpides, Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat Microbiol* **1**, 15032 (2016).
234. C. T. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208-U173 (2015).
235. J. Martijn *et al.*, Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. *Environmental Microbiology* **21**, 2485-2498 (2019).
236. M. Jamy *et al.*, Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Mol Ecol Resour* **20**, 429-443 (2020).
237. S. M. Karst *et al.*, High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nature Methods* **18**, 165-+ (2021).
238. S. M. Karst *et al.*, Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat Biotechnol* **36**, 190-195 (2018).
239. L. R. Thompson *et al.*, A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457-463 (2017).
240. V. Kunin, A. Copeland, A. Lapidus, K. Mavromatis, P. Hugenholtz, A bioinformatician's guide to metagenomics. *Microbiol Mol Biol Rev* **72**, 557-578, Table of Contents (2008).
241. C. Quince, A. W. Walker, J. T. Simpson, N. J. Loman, N. Segata, Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* **35**, 833-844 (2017).
242. F. P. Breitwieser, J. Lu, S. L. Salzberg, A review of methods and databases for metagenomic classification and assembly. *Brief Bioinform* **20**, 1125-1136 (2019).
243. R. Bharti, D. G. Grimm, Current challenges and best-practice protocols for microbiome analysis. *Brief Bioinform* **22**, 178-193 (2021).
244. G. W. Tyson *et al.*, Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43 (2004).
245. J. C. Venter *et al.*, Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74 (2004).
246. S. Goodwin, J. D. McPherson, W. R. McCombie, Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333-351 (2016).
247. J. M. Heather, B. Chain, The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8 (2016).
248. M. Ayling, M. D. Clark, R. M. Leggett, New approaches for metagenome assembly with short reads. *Brief Bioinform* **21**, 584-594 (2020).
249. A. Rhoads, K. F. Au, PacBio Sequencing and Its Applications. *Genom Proteom Bioinf* **13**, 278-289 (2015).
250. N. J. Loman, J. Quick, J. T. Simpson, A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods* **12**, 733-U751 (2015).

251. C. M. Singleton *et al.*, Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nature Communications* **12**, (2021).
252. D. Bertrand *et al.*, Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat Biotechnol* **37**, 937-944 (2019).
253. W. A. Overholt *et al.*, Inclusion of Oxford Nanopore long reads improves all microbial and viral metagenome-assembled genomes from a complex aquifer system. *Environmental Microbiology* **22**, 4000-4013 (2020).
254. J. Vollmers, S. Wiegand, A. K. Kaster, Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! *Plos One* **12**, (2017).
255. J. R. Miller, S. Koren, G. Sutton, Assembly algorithms for next-generation sequencing data. *Genomics* **95**, 315-327 (2010).
256. Z. Li *et al.*, Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Brief Funct Genomics* **11**, 25-37 (2012).
257. S. Sunagawa *et al.*, Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
258. M. Nomura, E. A. Morgan, Genetics of bacterial ribosomes. *Annu Rev Genet* **11**, 297-347 (1977).
259. C. J. Castelle *et al.*, Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* **25**, 690-701 (2015).
260. L. X. Chen, K. Anantharaman, A. Shaiber, A. M. Eren, J. F. Banfield, Accurate and complete genomes from metagenomes. *Genome Res* **30**, 315-333 (2020).
261. B. J. Tully, E. D. Graham, J. F. Heidelberg, The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* **5**, (2018).
262. D. T. Pride, R. J. Meinersmann, T. M. Wassenaar, M. J. Blaser, Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Res* **13**, 145-158 (2003).
263. H. Teeling, A. Meyerdierks, M. Bauer, R. Amann, F. O. Glockner, Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* **6**, 938-947 (2004).
264. J. Bohlin, E. Skjerve, D. W. Ussery, Investigations of oligonucleotide usage variance within and between prokaryotes. *Plos Comput Biol* **4**, (2008).
265. G. J. Dick *et al.*, Community-wide analysis of microbial genome sequence signatures. *Genome Biology* **10**, (2009).
266. M. Albertsen *et al.*, Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* **31**, 533-+ (2013).
267. J. Alneberg *et al.*, Binning metagenomic contigs by coverage and composition. *Nature Methods* **11**, 1144-1146 (2014).
268. C. M. K. Sieber *et al.*, Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**, 836-+ (2018).
269. G. V. Uritskiy, J. DiRuggiero, J. Taylor, MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, (2018).
270. W. C. Nelson, B. J. Tully, J. M. Mobberley, Biases in genome reconstruction from metagenomic data. *PeerJ* **8**, e10119 (2020).
271. A. M. Eren *et al.*, Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).

272. S. M. Karst, R. H. Kirkegaard, M. Albertsen, mmgenome: a toolbox for reproducible genome extraction from metagenomes. *bioRxiv*, (2016).
273. R. M. Bowers *et al.*, Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-731 (2017).
274. M. Y. Galperin *et al.*, COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* **49**, D274-D281 (2021).
275. P. N. Evans *et al.*, Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434-438 (2015).
276. K. Zaremba-Niedzwiedzka *et al.*, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358 (2017).
277. E. J. Richardson, M. Watson, The automatic annotation of bacterial genomes. *Brief Bioinform* **14**, 1-12 (2013).
278. M. Ravenhall, N. Skunca, F. Lassalle, C. Dessimoz, Inferring Horizontal Gene Transfer. *Plos Comput Biol* **11**, (2015).
279. M. Holder, P. O. Lewis, Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* **4**, 275-284 (2003).
280. Z. Yang, B. Rannala, Molecular phylogenetics: principles and practice. *Nat Rev Genet* **13**, 303-314 (2012).
281. P. Kapli, Z. Yang, M. J. Telford, Phylogenetic tree building in the genomic age. *Nat Rev Genet* **21**, 428-444 (2020).
282. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17**, 57-86 (1986).
283. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307-1320 (2008).
284. Z. H. Yang, Maximum-Likelihood Phylogenetic Estimation from DNA-Sequences with Variable Rates over Sites - Approximate Methods. *Journal of Molecular Evolution* **39**, 306-314 (1994).
285. J. Soubrier *et al.*, The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution* **29**, 3345-3358 (2012).
286. N. Lartillot, H. Philippe, A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**, 1095-1109 (2004).
287. S. Quang le, O. Gascuel, N. Lartillot, Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317-2323 (2008).
288. J. Felsenstein, Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading. *Syst Zool* **27**, 401-410 (1978).
289. F. F. Nascimento, M. dos Reis, Z. H. Yang, A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution* **1**, 1446-1454 (2017).
290. W. K. Hastings, Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97-& (1970).
291. J. Felsenstein, Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**, 783-791 (1985).
292. A. J. Aberer, D. Krompass, A. Stamatakis, Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst Biol* **62**, 162-166 (2013).
293. F. Lemoine *et al.*, Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452-456 (2018).

294. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* **30**, 1188-1195 (2013).
295. H. C. Wang, B. Q. Minh, E. Susko, A. J. Roger, Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst Biol* **67**, 216-235 (2018).
296. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).
297. H. Philippe, F. Delsuc, H. Brinkmann, N. Lartillot, Phylogenomics. *Annual Review of Ecology Evolution and Systematics* **36**, 541-562 (2005).
298. F. Delsuc, H. Brinkmann, H. Philippe, Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* **6**, 361-375 (2005).
299. H. Philippe *et al.*, Pitfalls in supermatrix phylogenomics. *Eur J Taxon* **283**, 1-25 (2017).
300. A. D. Young, J. P. Gillung, Phylogenomics - principles, opportunities and pitfalls of big-data phylogenetics. *Syst Entomol* **45**, 225-247 (2020).
301. N. Galtier, V. Daubin, Dealing with incongruence in phylogenomic analyses. *Philos Trans R Soc Lond B Biol Sci* **363**, 4023-4029 (2008).
302. A. Som, Causes, consequences and solutions of phylogenetic incongruence. *Brief Bioinform* **16**, 536-548 (2015).
303. N. Rodriguez-Ezpeleta *et al.*, Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol* **56**, 389-399 (2007).
304. H. Philippe *et al.*, Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol* **9**, e1000602 (2011).
305. P. Lopez, D. Casane, H. Philippe, Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution* **19**, 1-7 (2002).
306. E. Susko, A. J. Roger, Long Branch Attraction Biases in Phylogenetics. *Syst Biol*, (2021).
307. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* **10**, 13-26 (2012).
308. M. Gerth *et al.*, Rapid molecular evolution of Spiroplasma symbionts of Drosophila. *Microb Genom* **7**, (2021).
309. F. Husnik, T. Chrudimsky, V. Hypsa, Multiple origins of endosymbiosis within the Enterobacteriaceae (gamma-Proteobacteria): convergence of complex phylogenetic approaches. *Bmc Biol* **9**, (2011).
310. J. Martijn *et al.*, Hikarchaeia demonstrate an intermediate stage in the methanogen-to-halophile transition. *Nat Commun* **11**, 5490 (2020).
311. P. G. Foster, Modeling compositional heterogeneity. *Systematic Biology* **53**, 485-495 (2004).
312. E. Susko, A. J. Roger, On reduced amino acid alphabets for phylogenetic inference. *Mol Biol Evol* **24**, 2139-2150 (2007).
313. J. Viklund, T. J. Ettema, S. G. Andersson, Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29**, 599-615 (2012).
314. B. Roure, D. Baurain, H. Philippe, Impact of Missing Data on Phylogenies Inferred from Empirical Phylogenomic Data Sets. *Molecular Biology and Evolution* **30**, 197-214 (2013).
315. J. B. Joy, R. H. Liang, R. M. McCloskey, T. Nguyen, A. F. Y. Poon, Ancestral Reconstruction. *Plos Comput Biol* **12**, (2016).
316. M. Csuros, Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910-1912 (2010).

317. G. J. Szollosi, A. A. Davin, E. Tannier, V. Daubin, B. Boussau, Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos T R Soc B* **370**, (2015).
318. G. J. Szollosi, W. Rosikiewicz, B. Boussau, E. Tannier, V. Daubin, Efficient Exploration of the Space of Reconciled Gene Trees. *Systematic Biology* **62**, 901-912 (2013).
319. E. Jacox, C. Chauve, G. J. Szollosi, Y. Ponty, C. Scornavacca, ecceTERA: comprehensive gene tree-species tree reconciliation using parsimony. *Bioinformatics* **32**, 2056-2058 (2016).
320. G. J. Szollosi, E. Tannier, N. Lartillot, V. Daubin, Lateral Gene Transfer from the Dead. *Systematic Biology* **62**, 386-397 (2013).
321. T. A. Williams *et al.*, Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci U S A* **114**, E4602-E4611 (2017).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2040

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-439996

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2021