# THE ANATOMICAL AND PHYLOGENETIC NATURE OF THE FIRST MITOCHONDRIAL ANCESTOR

by

Sergio Andrés Muñoz-Gómez

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
November 2018

For my ever-loving parents.

# TABLE OF CONTENTS

**TABLE OF CONTENTS** ..... **iii**

**LIST OF TABLES** ..... **vii**

**LIST OF FIGURES** ..... **viii**

**ABSTRACT** ..... **x**

**LIST OF ABBREVIATIONS USED** ..... **xi**

**ACKNOWLEDGEMENTS** ..... **xv**

**CHAPTER 1: INTRODUCTION** ..... **1**

**CHAPTER 2: THE EVOLUTIONARY HISTORY OF THE MICOS COMPLEX AND ITS INTERACTION PARTNERS** ..... **7**

**2.1. INTRODUCTION** ..... 7

**2.2. MATERIALS AND METHODS** ..... 12

2.2.1. Databases ..... 12

2.2.2. Similarity searches ..... 12

2.2.3. Structural annotation ..... 13

2.2.4. Phylogenetic inference ..... 13

**2.3. RESULTS AND DISCUSSION** ..... 14

2.3.1. MICOS is an ancient multi-protein complex of mitochondria ..... 14

2.3.2. The presence of MICOS correlates with the presence of cristae ..... 18

2.3.3. Cristae morphology is not correlated with MICOS composition ..... 19

2.3.4. Gene duplications have given rise to some MICOS subunits ..... 20

2.3.5. The phylogenetic distribution of MICOS interactors ..... 21

2.3.6. MICOS has alphaproteobacterial roots ..... 27

2.3.7. Alphaproteobacterial Mic60 might interact with the product of its highly syntenic neighboring gene ..... 35

2.3.8. Alphaproteobacterial Mic60 evolved from a distant homolog in *Beta-*, and *Gammaproteobacteria* ..... 36

**2.4. CONCLUSIONS: A HYPOTHESIS FOR THE ORIGIN OF MITOCHONDRIAL CRISTAE** ..... 36

**CHAPTER 3: WHAT IS THE FUNCTION OF MIC60 IN THE *ALPHAPROTEOBACTERIA*? ..... 38**

**3.1. INTRODUCTION ..... 38**

**3.2. MATERIALS AND METHODS** ..... 43

3.2.1. Bacterial strains, media and growth conditions ..... 43

3.2.2. DNA methods and plasmid construction ..... 46

3.2.3. Construction of *R. sphaeroides* knock-out mutants ..... 54

3.2.4. Construction of *R. palustris* knock-out mutants ..... 55

3.2.5. Construction of *S. cerevisiae* strains ..... 55

3.2.6. Growth curve analyses of *R. sphaeroides* and *R. palustris* ..... 55

3.2.7. OD spectra ..... 56

3.2.8. Transmission electron microscopy ..... 56

3.2.9. Yeast spot dilution assay ..... 56

**3.3. PRELIMINARY RESULTS** ..... 57

3.3.1. Construction of Δ*mic60 and* Δ*orf52* knock-out mutant strains for *R. sphaeroides* and *R. palustris* ..... 57

3.3.2. Decreasing light intensity increases growth differences between WT and knock-out mutants ..... 62

3.3.3. Optical density (OD) spectra of whole cells do not differ between WT and knock-out mutants ..... 65

3.3.4. Knock-out mutants might display some ultrastructural defects ..... 69

3.3.5. Heterologous expression of Mic60 homologs in yeast does not recover growth ..... 74

**3.4. DISCUSSION** ..... 78

**3.5. CONCLUSIONS AND FUTURE DIRECTIONS** ..... 83

**CHAPTER 4: THE PHYLOGENY OF THE *ALPHAPROTEOBACTERIA* ..... 85**

**4.1. INTRODUCTION** ..... 85

**4.2. METHODS** ..... 87

4.2.1. Genome sequencing ..... 87

4.2.2. Genome assembly and annotation ..... 88

4.2.3. Taxon and gene selection ..... 89

4.2.4. Removal of compositionally-biased and fast-evolving sites ..... 90

4.2.5. Data recoding ..... 90

4.2.6. Phylogenetic inference ..... 91

4.2.7. Other analyses ..... 92

**4.3. RESULTS** ..... 93

4.3.1. The genomes and phylogenetic positions of three novel endosymbiotic alphaproteobacteria (*Rickettsiales* and *Holosporales*) ..... 93

4.3.2. Compositional heterogeneity appears to be a major confounding factor affecting phylogenetic inference of the *Alphaproteobacteria* ..... 96

4.3.3. The *Holosporales* is unrelated to the *Rickettsiales* and is instead most likely derived within the *Rhodospirillales* ..... 102

4.3.4. The *Rhodospirillales* is a diverse order and comprises five well-supported families ..... 106

4.3.5. The *Geminicoccaceae* might be basal to all other free-living alphaproteobacteria (the *Caulobacteridae*) ..... 107

4.3.6. Other deep relationships in the *Alphaproteobacteria* (*Pelagibacterales*, *Rickettsiales*, alphaproteobacterium sp. HIMB59) ..... 108

**4.4. DISCUSSION** ..... 108

**4.5. CONCLUSIONS** ..... 114

**CHAPTER 5: THE PHYLOGENETIC ROOTS OF MITOCHONDRIA ..... 115**

**5.1. INTRODUCTION** ..... 115

**5.2. METHODS** ..... 118

5.2.1. Metagenomics ..... 118

5.2.2. Gene marker and taxon selection ..... 120
5.2.3. Phylogenetic analyses ..... 121
**5.3. PRELIMINARY RESULTS AND DISCUSSION** ..... 122
5.3.1. Diverse alphaproteobacteria from Llamara and Alchichica ..... 122
5.3.2. Novel and disparate alphaproteobacterial lineages ..... 125
5.3.3. Mitochondria branch as sister to the *Alphaproteobacteria* ..... 127
5.3.4. Congruence between phylogenies ..... 131
**5.4. DISCUSSION AND CONCLUSIONS** ..... 131
**CHAPTER 6: CONCLUSIONS ..... 134**
**REFERENCES ..... 142**
**APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 2 ..... 170**
**APPENDIX B. SUPPLEMENTARY MATERIAL FOR CHAPTER 4 ..... 174**
**APPENDIX C. SUPPLEMENTARY MATERIAL FOR CHAPTER 5 ..... 196**
**APPENDIX D. COPYRIGHT PERMISSIONS ..... 204**

# LIST OF TABLES

**Table 3.1.** List of strains used in this study ..... 45

**Table 3.2.** List of plasmids used in this study ..... 47

**Table 3.3.** List of primers and their sequences used in this study ..... 49

**Table 4.1.** Genome features for the three novel rickettsialeans sequences in this study ..... 95

## LIST OF FIGURES

**Figure 1.1.** The origin and early evolution of mitochondria and eukaryotes. ..... 3

**Figure 2.1.** MICOS forms CSs and CJs to maintain and stabilize cristae in the mitochondria of *S. cerevisiae*. ..... 11

**Figure 2.2.** Distribution of MICOS subunits across the eukaryote evolutionary tree. ..... 17

**Figure 2.3.** Phylogenetic distribution of MICOS-interacting proteins. ..... 24

**Figure 2.4.** Distribution of Mic60 homologs in the UniProtKB database. ..... 28

**Figure 2.5.** Phylogenetic distribution of Mic60 in the *Alphaproteobacteria* and its conserved overall structure. ..... 31

**Figure 2.6.** Conservation of the LBS in between the coiled coils and the Mitofilin domain *sensu stricto* of Mic60. ..... 34

**Figure 3.1.** Members of the *Alphaproteobacteria* develop morphologically diverse ICMs to carry out different bioenergetic functions. ..... 41

**Figure 3.2.** PCR assays that confirm the disruption of the *mic60* and *orf52* genes in *R. sphaeroides*. ..... 59

**Figure 3.3.** PCR assays that confirm the disruption of the *mic60* and *orf52* genes in *R. palustris*. ..... 61

**Figure 3.4.** Photosynthetic growth curves of *R. sphaeroides* and *R. palustris* WT and Δ*mic60* and Δ*orf52* strains. ..... 64

**Figure 3.5.** OD spectra of photosynthetically grown *R. sphaeroides* and *R. palustris* WT and Δ*mic60* and Δ*orf52* strains. ..... 68

**Figure 3.6.** TEM micrographs of *R. sphaeroides* WT, and Δ*mic60* and Δ*orf52* strains grown chemoheterotrophically under oxia and darkness. ..... 71

**Figure 3.7.** TEM micrographs of *R. sphaeroides* WT, and Δ*mic60* and Δ*orf52* strains grown photoheterotrophically under anoxia and high light. ..... 73

**Figure 3.8.** Spot dilution assays of yeast strains to test the functional complementation of different *MIC60* gene constructs heterologously expressed in a *MIC60*Δ strain. ..... 77

**Figure 3.9.** The function of MICOS is probably conserved between mitochondria and alphaproteobacteria. ..... 82

**Figure 4.1.** Compositional heterogeneity in the *Alphaproteobacteria* is a major factor that confounds phylogenetic inference. ..... 98

**Figure 4.2.** Decreasing compositional heterogeneity by removing compositionally-biased sites disrupts the clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales*. ..... 101

**Figure 4.3.** The *Holosporales* branches in a derived position within the *Rhodospirillales* when compositional heterogeneity is reduced and the long-branching *Rickettsiales* and *Pelagibacterales* are removed. ..... 105

**Figure 4.4.** A higher-level classification scheme for the *Alphaproteobacteria* and the *Magnetococcia* class. nov. ..... 113

**Figure 5.1.** Alphaproteobacterial MAGs reconstructed from metagenomes from thermophilic microbial mats in the Llamara saltern (Chile) and microbialites in the Alchichica lake (Mexico). ..... 124

**Figure 5.2.** A phylogenetic tree inferred under the LG+PMSF(C60)+F+G4 model with the reference alphaproteobacteria and the most novel MAGs. ..... 126

**Figure 5.3.** A phylogenetic tree inferred under the LG+PMSF(C60)+F+G4 model after removing 60% of the most compositionally biased sites according to the ʐ metric (a total of 11,855 amino acid sites analyzed). ..... 129

**Figure 5.4.** Branch support variation, SH-aLRT (%) and UFBoot2+NNI (%) as the most compositionally biased sites according to ʐ metric and the $\chi^2$ method are removed. ..... 130

**Figure 6.1.** Two alternative scenarios for the evolutionary relationship between mitochondrial cristae and alphaproteobacterial ICMs. ..... 139

# ABSTRACT

Mitochondria are cell organelles that evolved by endosymbiosis. Even though the nature of the last mitochondrial common ancestor is well constrained, little is known about the nature of the first mitochondrial common ancestor. The goal of this thesis was to shed light on the anatomy and genealogy of the first mitochondrial ancestor.

Mitochondria compartmentalizes aerobic respiration inside membrane invaginations called cristae. The development of mitochondrial cristae involves the MICOS (Mitochondrial Contact Site and Cristae Organizing System) complex. I show that the core of MICOS, Mic60, is conserved across eukaryotes and was inherited from the alphaproteobacterial ancestor of mitochondria. Alphaproteobacterial Mic60 retains the same structure, domains and even a functionally critical amphipathic helix as eukaryotic Mic60. These observations revive the hypothesis that cristae might have evolved from intracytoplasmic membranes (ICMs). To experimentally test this hypothesis, I knocked out *mic60* and *orf52* (a gene immediately downstream of *mic60*) in two different ICM-developing alphaproteobacteria. Growth curves and transmission electron microscopy of these mutants show some differences relative to the wild type but do not provide a conclusive answer. However, in combination with additional experimental evidence provided by others, a stronger case is made for the conserved function of Mic60 in alphaproteobacteria supporting the hypothesis that cristae evolved from ICMs.

To better understand the specific evolutionary relationship between mitochondria and their closest relatives, the *Alphaproteobacteria*, I first inferred a robust phylogeny of the *Alphaproteobacteria* using a diverse set of strategies aimed at reducing systematic errors. I found that the parasitic *Rickettsiales* and *Holosporales*, previously thought to be most closely related to each other and to mitochondria, evolved independently from different ancestors. I then attempted to infer the phylogenetic placement of mitochondria by reconstructing novel alphaproteobacterial genomes from environmental metagenomes, including an expanded set of marker genes, and using improved methods for phylogenetic inference. Preliminary results suggest that mitochondria are sister to all the *Alphaproteobacteria*.

Together, these results suggest that the first mitochondrial ancestor was a free-living close relative of the modern *Alphaproteobacteria* that developed ICMs. This view is in opposition to a widely held view that mitochondria evolved from intracellular parasites.

# LIST OF ABBREVIATIONS USED

AOX: alternative oxidase

ATP: adenosine triphosphate

BAM: β-barrel assembly machinery

Bch: bacteriochlorophyll

BLAST: basic local alignment search tool

CJ: crista junction

CM: crista membrane

CS: contact site

Dex: dextrose

DNA: desoxyribonucleic acid

Dox: doxycycline

DUF: domain of unknown function

ER: endoplasmic reticulum

ERMIONE: ER-mitochondria organizing network

ETC: electron transport chain

EtOH: ethanol

FECA: first eukaryote common ancestor

FLP: flippase

FRT: FLP recognition target

gDNA: genomic DNA

Gly: glycerol

Gm: gentamycin

HMM: hidden Markov model

IBM: inner boundary membrane

ICM: intracytoplasmic membrane

ICMJ: ICM junction

IMS: intermmebrane space

JGI: Joint Genome Institute

Lac: lactate

LB: lysogeny broth

LBA: long-branch attraction

LBS: lipid binding site

LECA: last eukaryote common ancestor

Leu: leucine

LH1: light harvesting complex 1

LH2: light harvesting complex 2

MAG: metagenome-assembled genome

MCMC: Markov chain Monte Carlo

MIA: mitochondrial IMS assembly

MICOS: mitochondrial contact site and cristae organizing system

MIM: mitochondrial inner membrane

MitoCOG: mitochondria culusters of orthologous genes

ML: maximum likelihood

MMETSP: Marine Microbial Eukaryote Transcriptome Sequencing Project

MOM: mitochondrial outer membrane

MOPS: 3-(n-morpholino)propanesulfonic acid

mtDNA: mitocondrial DNA

NCBI: National Center for Biotechnology Information

NDCH2: node-discrete composition heterogeneity 2 model

OD: optical density

OM: outer membrane

PCR: polymerase chain reaction

PMSF: posterior mean site frequency profiles

RC: reaction center

RLB: Rhodobacter lysogeny broth

RNA: ribonucleic acid

rpm: revolutions per minute

rRNA: ribosomal RNA

SAM: sorting and assembly machinery

SAR: Stramenopila, Alveolata, Rhizaria

SC: synthetic complete

SH-aLRT: Shimodaira–Hasegawa approximate likelihood-ratio test

SSU rRNA: small subunit rRNA

TEM: transmission electron microscopy/microscope

TOM: translocase of the outer membrane

TPI: triose phosphate isomerase

TPR: tetratricopeptide repeat

UFBoot: ultra-fast bootstrap

UFBoot2+NNI: ultra-fast bootstrap2 and nearest neighbor interchange search

UPGMA: Unweighted Pair Group Method with Arithmetic Mean

Ura: uracil

VDAC: voltage-dependent anion channel

WT: wild type

# ACKNOWLEDGEMENTS

For the Stoic, virtue is all that matters. I will take some few words here to thank those who made me a more virtuous person, both epistemically and in personal character.

Fortuna has been good to me in life. She has given me loving, supporting and always-encouraging parents, Mariana and Edgar. I also felt the support of others in my family, my sister Sarita and my aunts Natalia and Ana Lucía. I am most thankful to them. Fortuna has also given me great friends and colleagues that have given meaning to my life. Most of my colleagues have served as role models. I have attempted to emulate some of their virtues and strengths and will aim to continue to do so. Many of them were teachers by example.

From Claudio Slamovits, who gave me the opportunity to come to the CGEB when I was an eager undergraduate, I learned to be more patient, tolerant and to put things in perspective. From Andrew Roger, who adopted me as one of his students, I learned kindness, politeness and tactfulness. I am most thankful to my two supervisors for their time and patience, and the knowledge and wisdom they shared with me. From John Archibald, I learned serenity, discipline, scholarship, and diplomacy. Claudio, Andrew and John have inspired me to become a better person and researcher. They made me feel welcome and part of the bigger CGEB family. I am also thankful to the other members of my committee, Melanie Dobson and Ford Doolittle, who offered helpful advice when most needed.

I thank Dayana Salas-Leiva and Ángela P. Fuentes-Pardo for their friendship, companionship and support. Jeremy Wideman for his friendship, advice, support and academic discussions. Sebastian Hess, for his friendship and for serving as example and role model for a passionate and rigorous naturalist. My mentor Michelle Leger who embraced me when I knew little and with much patience taught me how to do experimental work. My students Keira Durnin, Louise Connell and Kassandra Kennedy, who taught me more than I taught them. Friends and colleagues with whom I shared most cherished times: Susana Breglia, Pía Elustondo, Donika Shala, Ronie Haro, Jan and Sophie de Vries, Anika Schnake, Cameron Grisdale, Gina Filloramo, Jon Jerlström Hultqvist, Lucie Gallot-Lavallée, Yana Eglit, Gordon Lax, Pak Poon, and Renny Lee. Other friends who kept me company in the distance: Wendy Valencia, Camilo Calderón, Juan Pablo Narváez, Fabian Mejía, and Laura Toro. I'd also like to thank my undergraduate mentors who gave me the strong pillars upon which to build both academically and personally: John Jairo Ramírez, Ricardo Callejas, Edwin Patiño and Juan Manuel Daza. Katherine was the greatest Stoic challenge I've had. I also thank many other colleagues and friends who helped throughout these five years: Bruce Curtis, Wanda Danilchuk, Marlena Dlutek, Laura Eme, Courtney Stairs, and Tommy Harding. I leave being most thankful for the people, the experiences and the personal teachings I've had in Halifax. My Ph.D. work was funded by a Killam Pre-doctoral Scholarship and a Nova Scotia Graduate Scholarship.

# CHAPTER 1: INTRODUCTION

*"Midi-chlorians* [mitochondria] *are a microscopic life form that resides within all living cells."* —Qui-Gon Jinn, to Anakin Skywalker (Star Wars: Episode I *The Phantom Menace*, 1999).

*"Without the midi-chlorians* [mitochondria]*, life could not exist, and we would have no knowledge of the Force. They continually speak to us, telling us the will of the Force. When you learn to quiet your mind, you'll hear them speaking to you."* —Qui-Gon Jinn, to Anakin Skywalker (Star Wars: Episode I *The Phantom Menace*, 1999).

Mitochondria are cell organelles inferred to have been present in the eukaryote cenancestor, or the last common ancestor of all eukaryotes (Roger, Muñoz-Gómez, & Kamikawa, 2017). Unlike most other cell organelles also known to have been ancestral to eukaryotes (e.g., the endoplasmic reticulum, the Golgi apparatus, the nucleus, the cytoskeleton, etc.), the mitochondrion evolved from an endosymbiont most closely related to modern alphaproteobacteria (Andersson et al., 1998; Dayhoff & Schwartz, 1981; Yang, Oyaizu, Oyaizu, Olsen, & Woese, 1985). Aerobic eukaryotes use their mitochondria as bioenergetic organelles to produce most of the ATP of the cell through aerobic respiration, i.e., by coupling electron transport to chemiosmosis in the presence of oxygen. In contrast, most anaerobic eukaryotes use mitochondria to produce ATP fermentatively (Müller et al., 2012; Stairs, Leger, & Roger, 2015). In some anaerobic eukaryotes, mitochondria do not produce ATP (e.g., Tovar et al., 2003), and it is now also clear that some other anaerobic eukaryotes have lost their mitochondria altogether (Karnkowska et al., 2016).

The origin and early evolution of mitochondria can be subdivided into three main stages or reference points to ease its analysis and discussion (Fig. 1.1; Roger et al., 2017). First, we have the first mitochondrial ancestor (Fig. 1.1). This is the last common ancestor between modern alphaproteobacteria and mitochondria and corresponds to an alphaproteobacterial cell in phenotype. Last, we have the last mitochondrial ancestor (or the mitochondrial cenancestor; Fig. 1.1). This is the last common ancestor of all modern mitochondria in diverse eukaryote groups and corresponds to an organelle that was already highly specialized to aerobic respiration and highly integrated with its host cell

(and therefore was not too different from modern mitochondria). Between the first and the last mitochondrial common ancestors, there must have been diverse lineages of proto-mitochondria (Fig. 1.1). Proto-mitochondria are therefore all intermediate or transitional forms that evolved, and diversified, after the first but before the last mitochondrial ancestor. These encompass the very early endosymbiotic alphaproteobacteria that were barely reduced, as well as the more reduced proto-organelles; they are all long extinct (Roger et al., 2017).

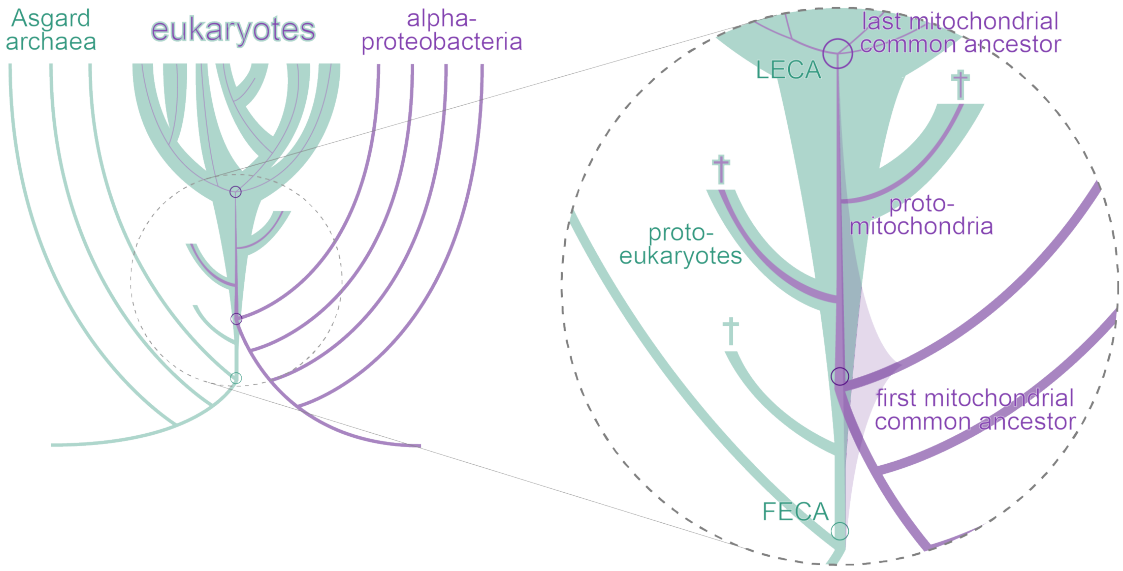



**Figure 1.1.** The origin and early evolution of mitochondria and eukaryotes. Mitochondria evolved from an endosymbiotic alphaproteobacterium (purple) within a host cell that was most closely related to modern Asgard archaea (green). The earliest ancestor of mitochondria (that is not also an ancestor of an extant alphaproteobacterium) is the first mitochondrial ancestor. Proto-mitochondria evolved from this first presumed alphaproteobacterial endosymbiont and comprise all transitional forms of mitochondria before the last mitochondrial ancestor (i.e., the mitochondrion of the last eukaryotic common ancestor; LECA). The timing of the mitochondrial endosymbiosis is uncertain (indicated by a purple shadow along the proto-eukaryotic stem) but postdates the first eukaryote common ancestor (FECA) and predates LECA. As far as we know, all transitional proto-eukaryotes between FECA and LECA went extinct (indicated by crosses). The complexity of the proto-eukaryotic genome and proteome gradually increased during eukaryogenesis (increasingly wider green branches), but the mitochondrial endosymbiont's genome and proteome were reduced, as the organelle incorporated proteins of host and foreign origin (progressively thinner purple branches for the mitochondrial endosymbiont contribution) (Roger et al., 2017). Note that this diagram deceptively represents evolutionary history in a linear and progressive way. More accurately, evolutionary history should be metaphorically represented as a dense bush in which most lineages do not make it to the present day, but a few side branches do.

The last mitochondrial common ancestor can be reconstructed by comparing modern mitochondria (e.g., Gabaldón & Huynen, 2003, 2007; Wang & Wu, 2014). Proto-mitochondria, on the other hand, are much more difficult to reconstruct because of the massive extinction of transitional forms, and because the genetic and phenotypic gap between the first and last mitochondrial ancestors is huge. Finally, the first mitochondrial ancestor could also be reconstructed by comparing modern mitochondria and alphaproteobacteria (e.g., Wang & Wu, 2014). However, the great diversity of the *Alphaproteobacteria* and the uncertainty of whether there is any alphaproteobacterial subgroup most closely related to mitochondria (see Chapter 5), makes the task very challenging. This thesis is primarily concerned with the nature of the first mitochondrial common ancestor.

Over the years, a considerable number of phenotypes have been proposed for the first mitochondrial common ancestor with the overarching goal of elucidating what the first endosymbiotic alphaproteobacterium was doing for its host. Generally, these proposals have been made by taking into consideration (a) the modern capabilities of mitochondria, (b) the phylogenetic affiliations of mitochondria, (c) the nature of the host cell, and (d) the environmental setting of the primordial endosymbiosis. Since each of these criteria entails varying degrees of uncertainty, all the resulting scenarios are necessarily very speculative.

Among the proposed ideas, several envision an aerobic respiring heterotroph (Margulis, 1970). The endosymbiont would have provided an advantage to its host, usually an anaerobic fermenter, by (implausibly) secreting ATP, removing fermentation waste (Margulis, 1970; Speijer, 2017; Whatley, John, & Whatley, 1979), serving as a methane sink (López-García & Moreira, 2015; Moreira & López-García, 1998), or removing toxic oxygen from within (Andersson & Kurland, 1999). A (likely versatile) photosynthetic bacterium has also been invoked as the first mitochondrial ancestor. It would have been useful to its host by allowing it to move to aerobic niches (Woese, 1977), recycling (oxidizing) reduced sulfur produced by host respiration (Searcy, 1992), releasing hydrogen from fermentation (Martin & Müller, 1998), or secreting organic photosynthate (intermediate anabolites) (Cavalier-Smith, 2006; Fenchel & Bernard, 1993b). Some have thought that the endosymbiont provided a combination of benefits to its host (e.g., iron-sulfur clusters and other metabolites resulting from the metabolism of glycerol, fatty acids, amino acids, etc.; Gabaldón & Huynen, 2003, 2007b). Moreover, some non-

mutualistic scenarios suggest that the first mitochondrial ancestor was a periplasmic predator (like *Bdellovibrio* or *Micavibrio*) that aggressively invaded its host (Davidov, Huchon, Koval, & Jurkevitch, 2006; Davidov & Jurkevitch, 2009), or a facultative intracellular parasite of eukaryotes like some rickettsialeans (e.g., Sassera et al., 2011; Wang & Wu, 2014).

The elusive nature of the first mitochondrial ancestor means that all proposed proximate selective forces for the initial endosymbiosis remain speculative. The ultimate selective advantage that mitochondria provided to their hosts, however, was undoubtedly greater efficiency in harnessing energy through aerobic respiration—all other mitochondrial functions are secondary and support aerobic respiration, and some of these have become secondarily essential to their eukaryote hosts (e.g., iron-sulfur cluster biosynthesis). The capability to more fully oxidize food through aerobic respiration was probably a new physiological property brought to the host by the first mitochondrial ancestor; many Asgard archaea appear to be fermenters (see e.g., Liu et al., 2018; Seitz, Lazar, Hinrichs, Teske, & Baker, 2016; Sousa, Neukirchen, Allen, Lane, & Martin, 2016). Transforming a bacterium into an organelle (i.e., symbiogenesis or 'organellogenesis') was effectively a process of enslavement by the host (Cavalier-Smith, 2006, 2007, 2013). The endosymbiont lost its autonomy and eventually specialized as an ATP-producing organelle. Therefore, a major driving force we can invoke to understand most of this transformation is an increased control on the nascent organelle by the proto-eukaryotic host (e.g., see Gross & Bhattacharya, 2009, 2011). The transition from the first mitochondrial ancestor, across proto-mitochondria, to the last mitochondrial ancestor required many evolutionary changes. Some of these likely happened before another one could evolve, although many changes probably overlapped temporarily (e.g., see Cavalier-Smith, 2007; Roger et al., 2017). Given the number and complexity of the changes involved, the transformation of the first into the last mitochondrial ancestor was necessarily gradual and produced many transitional forms long extinct (i.e., proto-mitochondria).

Attaining a robust phylogenetic placement of mitochondria in relation to their bacterial relatives is perhaps the best measure to constrain thoughts on the nature of the first mitochondrial ancestor. For example, a scenario in which mitochondria branch as sister to or within the *Rickettsiales* would favour a facultative intracellular energy parasite that invaded ancestral proto-eukaryotic cells (or meta-archaeal cells) as the first

mitochondrial ancestor (Wang & Wu, 2014). This might also be consistent with a flagellated cell with a repertoire of terminal cytochromes adapted to hypoxia; a bacterium not too different from endosymbiotic holosporaceans of protists (Sassera et al., 2011; Wang & Wu, 2014, 2015). On the other hand, a sister group to rickettsialeans, or a more basal position among alphaproteobacteria (e.g., Martijn, Vosseberg, Guy, Offre, & Ettema, 2018), would mean that the first mitochondrial ancestor was a free-living alphaproteobacterium. If so, chances are that it was a purple nonsulfur bacterium (i.e., an anoxygenic photosynthesizer) because this physiology might well have been ancestral to all alphaproteobacteria (Muñoz-Gómez, Wideman, Roger, & Slamovits, 2017). Future endeavors into inferring the phylogeny of Asgard archaea (the closest relatives of the proto-eukaryote host cell) and their metabolic diversity, as well as explorations on the diversity and phylogeny of 'environmental' alphaproteobacteria might help us to further refine the issue of the nature of the first mitochondrial ancestor.

This thesis is about mitochondria, but more specifically about their ancestry and origin. Its primary concern is the nature of the first mitochondrial common ancestor. As the title of this thesis makes clear, my scientific work in the last five years has largely focused on two aspects of mitochondria: (1) their internal structure, or anatomy, and (2) the phylogenetic relationships with their closest modern relatives, or their genealogy. The second chapter provides the first empirical evidence, based on gene similarities, for the hypothesis that mitochondrial cristae are transformed intracytoplasmic membranes. The third chapter summarizes experiments in alphaproteobacteria that attempted to provide functional evidence for the same hypothesis, meaning that mitochondrial cristae and alphaproteobacterial intracytoplasmic membranes share a common ancestry (i.e., are homologous). In the fourth chapter, I switch to large-scale phylogenetics, or phylogenomics, to arrive at a robust phylogram (a hypothesis of phylogeny) for the *Alphaproteobacteria*, the eubacterial group that encompasses the closest modern relatives of mitochondria. Having a relatively robust phylogram for the *Alphaproteobacteria* as a reference, in the fifth chapter I have attempted to phylogenetically place mitochondria among the alphaproteobacteria. Finally, in the sixth chapter I conclude this thesis by discussing my emerging, likely temporary, views on the nature of the first mitochondrial ancestor, and what directions can be taken in the future to make further progress.

# CHAPTER 2: THE EVOLUTIONARY HISTORY OF THE MICOS COMPLEX AND ITS INTERACTION PARTNERS

## 2.1. INTRODUCTION

Aerobic eukaryotes use their mitochondria as bioenergetic organelles to produce most of the ATP of the cell. They do so by coupling electron transport to chemiosmosis in the presence of oxygen. The bioenergetic function of mitochondria is carefully regulated by mitochondrial structure (Mannella, Lederer, & Jafri, 2013; Westermann, 2012). At the morphological level, mitochondrial shape is controlled by processes of fission (i.e., division) and fusion, as well as the interaction of the mitochondrial envelope with the cytoskeleton and diverse cellular membranes (e.g., the endoplasmic reticulum). The dynamic interaction among these three processes regulates the distribution, size and overall shape of mitochondria (e.g., punctate or reticulate) (Okamoto & Shaw, 2005). Whereas mitochondrial division and movement along the cytoskeleton distribute mitochondria to sites of high energy demand within the cell, mitochondrial fusion homogenizes their contents and confers a buffering capacity (Youle & Bliek, 2012). Together, these aspects of mitochondrial morphology and behaviour optimize the bioenergetic function of mitochondria in response to cellular needs (Galloway, Lee, & Yoon, 2012; Westermann, 2012). For instance, a mitochondrial network distributed along the cell cortex is usually found in metabolically active and respiring yeast cells, whereas punctate mitochondria are observed in resting cells (Egner, Jakobs, & Hell, 2002; Hoffmann & Avers, 1973). Overall mitochondrial structure is therefore adaptively coupled to energy production (Liesa & Shirihai, 2013; Mishra & Chan, 2016).

At the finer level of mitochondrial internal architecture, a tight coupling between function and structure is also observed (Mannella et al., 2001; Mannella, 2006; Mannella et al., 2013). The mitochondrial envelope comprises the mitochondrial outer membrane (MOM) and the mitochondrial inner membrane (MIM) (see Fig. 2.1). The MIM harbours a complex morphology and invaginates at tubular structures called crista junctions (CJs) to give rise to specialized sub-compartments called mitochondrial cristae, the structural hallmarks of mitochondria (Daems & Wisse, 1966; Mannella, Marko, Penczek, Barnard, & Frank, 1994; Perkins et al., 1997). It is believed that these sub-compartments limit the diffusion of molecules between the intra-cristal space and the intermembrane space,

thereby localizing proton gradients (Mannella, 2000; Strauss, Hofhaus, Schröder, & Kühlbrandt, 2008; Williams, 2000), concentrating metabolites (Mannella et al., 2001; Mannella et al., 2013), and preventing the release of signaling molecules (e.g., cytochrome *c* during apoptosis) (Cogliati et al., 2013; Frezza et al., 2006; Olichon et al., 2003; Varanita et al., 2015). Furthermore, cristae have been described as also having dynamic behaviour that depends on the fusion and fission of crista membranes (CM) in response to the metabolic state of mitochondria (Mannella, 2006). This new view of mitochondrial cristae, in which the MIM regulates the formation of CJs to create respiratory sub-compartments, contrasts with the orthodox textbook model where cristae are static and random infoldings of the MIM (Cogliati, Enriquez, & Scorrano, 2016). Cristae therefore constitute developmentally modulated sub-compartments that optimize the efficiency of aerobic respiration (Cogliati et al., 2016; Jayashankar, Mueller, & Rafelski, 2016).

The MIM exhibits a high degree of lateral heterogeneity, as it is compositionally, structurally, and functionally subdivided into two distinct domains: the inner boundary membrane (IBM) and the crista membrane (CM) (Vogel, Bornhövd, Neupert, & Reichert, 2006; Wurm & Jakobs, 2006) (Fig. 2.1A). Processes such as protein and lipid import, and solute transport occur at the IBM, whereas respiratory electron transport and ATP synthesis occur at the CM. CJs are highly negatively curved slot- or neck-like tubular structures at which the transition from the IBM to the CM occurs (Daems & Wisse, 1966; Perkins et al., 1997). Moreover, there exist numerous contact sites (CSs) between the MOM and IBM required for such diverse processes as protein import, lipid transfer, mitochondrial fusion and cristae stability (Harner et al., 2011; Horvath et al., 2015). The development of cristae from the IBM is thus dependent on the formation of both CJs and CSs, which usually overlap spatially at the mitochondrial envelope (Harner et al., 2011; Hoppins et al., 2011; Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012).

Several molecular factors are responsible for the biogenesis and morphology of mitochondrial cristae (Zick, Rabl, & Reichert, 2009). These include the MICOS (MItochondrial Contact Site and Cristae Organizing System) complex, the protein Mgm1, the subunits *e* and *g* of the $F_1F_O$-ATP synthase complex, as well as cardiolipin and the respiratory supercomplexes. The MICOS complex is responsible for creating CJs and CSs for cristae at the mitochondrial envelope (Fig. 2.1; see below; Rampelt, Zerbes, van

der Laan, & Pfanner, 2016). Mgm1 is a dynamin involved in mitochondrial fusion which has also been proposed to engage in homotypic interactions at CJs (Frezza et al., 2006; Meeusen et al., 2006). Subunits *e* and *g* of the ATP synthase complex, on the contrary, are required for the proper morphology of cristae by introducing positive curvature at crista rims through the mediation of ATP synthase complex dimerization/oligomerization (Davies, Anselmi, Wittig, Faraldo-Gómez, & Kühlbrandt, 2012; Paumard et al., 2002; Strauss et al., 2008). Moreover, the MIM lipid cardiolipin is required for the proper assembly, stability, and functioning of MIM complexes (Friedman, Mourier, Yamada, McCaffery, & Nunnari, 2015; Paradies, Paradies, De Benedictis, Ruggiero, & Petrosillo, 2014), whereas the respiratory supercomplexes represent the main constituents of CMs, as their absence in *Saccharomyces cerevisiae rho*$^{0}$ mutants leads to the complete disappearance of cristae (Friedman et al., 2015; Hoppins et al., 2011). Notwithstanding these multiple factors, MICOS has emerged as the main factor responsible for the formation of CJs, and therefore the maintenance and stability of mitochondrial cristae (Pfanner et al., 2014; van der Laan, Bohnert, Wiedemann, & Pfanner, 2012; Zerbes, van der Klei, et al., 2012).

MICOS is a large hetero-oligomeric protein complex that localizes at CJs and CSs (Harner et al., 2011; Hoppins et al., 2011; Pfanner et al., 2014; von der Malsburg et al., 2011), and that has been characterized both structurally and functionally in *Saccharomyces cerevisiae* (Harner et al., 2011; Hoppins et al., 2011; von der Malsburg et al., 2011) and *Homo sapiens* (John et al., 2005; Ott, Dorsch, Fraunholz, Straub, & Kozjak-Pavlovic, 2015) (Fig. 2.1A). The disruption of MICOS results in the virtual absence of CJs and the accumulation of internal stacked CM vesicles detached from the IBM. This mutant ultrastructure leads to impaired respiration and altered mtDNA inheritance, demonstrating the tight control that mitochondrial structure exerts on function (Harner et al., 2011; Hess et al., 2009; Hoppins et al., 2011; Itoh, Tamura, Iijima, & Sesaki, 2013; John et al., 2005; Rabl et al., 2009; von der Malsburg et al., 2011).

In *S. cerevisiae*, MICOS is composed of six subunits and assembles hierarchically into the subcomplexes Mic60-Mic19 and Mic12-Mic10-Mic26-Mic27 (Friedman et al., 2015; Itoh et al., 2013; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015; Zerbes, Höß, Pfanner, van der Laan, & Bohnert, 2016) (Fig. 2.1A). Mic60 and Mic10 are the two most functionally important MICOS subunits, as their disruption leads to the most severe

mutant phenotypes. Mic60 is a multifunctional protein that serves as the MICOS core subunit (John et al., 2005; Rabl et al., 2009). Mic60 determines the sites of MICOS assembly at CJs/CSs (Friedman et al., 2015), establishes CSs through its interaction with Sam50 (Tob55) and Tom40 at the MOM (Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012), and contributes to CJ formation by bending the MIM (Hessenberger et al., 2017). Mic19 regulates MICOS assembly by serving as an adaptor protein between the two MICOS subcomplexes at the intermembrane space (Friedman et al., 2015). Mic10 is a main morphogenetic factor of MICOS. Analogous to reticulons in the ER (Voeltz, Prinz, Shibata, Rist, & Rapoport, 2006) and CURT proteins in thylakoids (Armbruster et al., 2013), Mic10 forms multimers that introduce strong negative curvature at CJs to pinch off the IBM into cristae (Barbot et al., 2015; Bohnert et al., 2015). Mic26 and Mic27 (previously known as Aim37 in *S. cerevisiae*) are homologous proteins that bind to cardiolipin and might be involved in the stability of MICOS (including the higher-order multimerization of Mic10; Zerbes et al., 2016) and the distribution of cardiolipin in the MIM (Friedman et al., 2015; Koob & Reichert, 2014). Finally, Mic12 appears to be an additional stabilizing factor of MICOS (Friedman et al., 2015), which also serves as a MIM linker between the two MICOS subcomplexes (Zerbes et al., 2016). The MICOS complex therefore combines the diverse (non-redundant) functions of its subunits to create CJs and CSs and consequently to develop and maintain mitochondrial cristae.

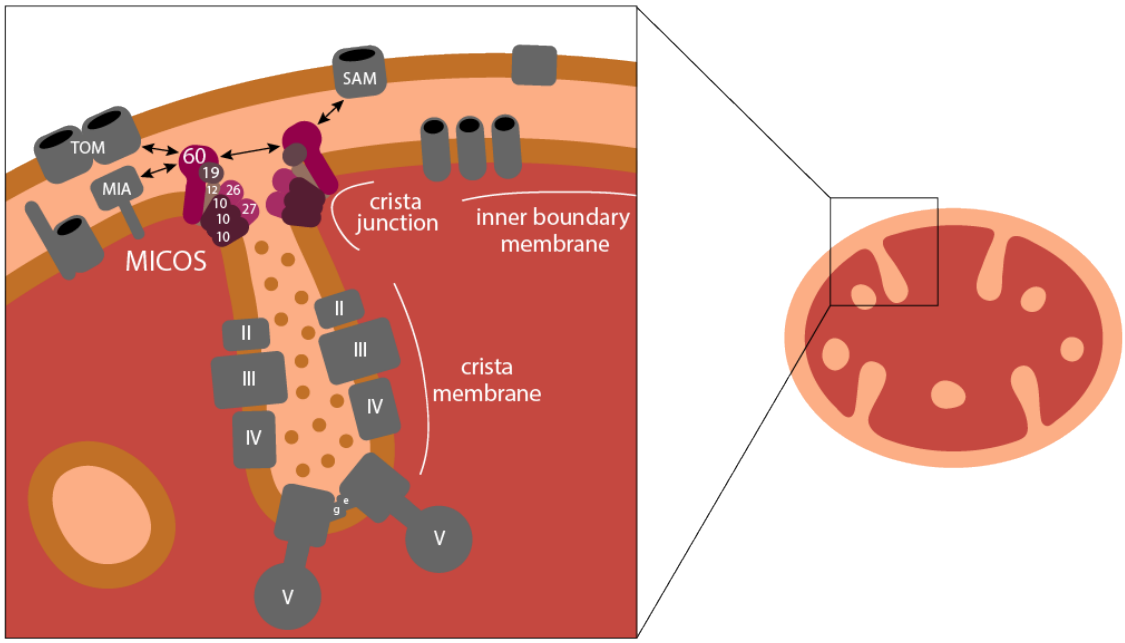



**Figure 2.1.** MICOS forms contact sites (CSs) and crista junctions (CJs) to maintain and stabilize cristae in the mitochondria of *S. cerevisiae* (Harner et al., 2011; Hoppins et al., 2011; Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012). MICOS is composed of the Mic60–Mic19 subcomplex which establishes CSs with the MOM and marks the sites of CJ formation, and the Mic12–Mic10–Mic26–Mic27 subcomplex which differentiates and bends the MIM at CJs (Friedman et al., 2015; Pfanner et al., 2014; Zerbes et al., 2016). The central MICOS subunit, Mic60, introduces membrane curvature at CJs, contacts the outer membrane through its interactions with the TOM and SAM complexes (additional interactions have also been reported with the MOM proteins porin and Ugo1), and further interacts with Mia40 to aid in the oxidative import of mitochondrial proteins (Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012).

The MICOS complex was found in 2011 (Harner et al., 2011; Hoppins et al., 2011; von der Malsburg et al., 2011; but Mic60 was discovered earlier in 1994; Icho et al., 1994) and major progress has been made in understanding its structure and function since then. By 2015, however, little to nothing was known about the evolutionary history of MICOS, other than that the complex was present both in yeast and some animals (e.g., *Homo sapiens* and *Caenorhabditis elegans*). Unravelling the evolutionary history of MICOS could shed light onto the origin and evolution of mitochondrial cristae, and by extension also mitochondria as organelles. This chapter describes the evolutionary history of MICOS and many of its interactors.

## 2.2. MATERIALS AND METHODS

### 2.2.1. Databases

Local custom databases were built from predicted proteomes downloaded from NCBI GenBank, the Joint Genome Institute (JGI), and the Broad Institute. The transcriptomes for a diverse set of eukaryotes were downloaded from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; Keeling et al., 2014).

### 2.2.2. Similarity searches

To find homologs of MICOS subunits across eukaryotes and prokaryotes, different similarity search methods were combined. Protein sequences of experimentally characterized MICOS subunits in *S. cerevisiae* and *H. sapiens* were first used as queries in reciprocal BLAST searches (Altschul, Gish, Miller, Myers, & Lipman, 1990) against protein sequences of other animals and fungi (in a local database). A reciprocal BLAST search relies on, first, a BLAST search against the proteome of a target species, and, second, a reverse BLAST search using the best hit from the first BLAST search against the proteome of the source species. If the reverse BLAST search recovers the query used from the source species as the best hit, then the hit in the target species is retained and considered a candidate homolog. The candidate homologs were further confirmed by searching for domains and motifs known in the query proteins using hmmscan in the Pfam server (https://pfam.xfam.org/). The same candidate homologs were then aligned and profile hidden Markov models (profile HMMs) were inferred using the HMMER suite (Eddy, 1998). The profile HMMs for each MICOS subunit were then used to search, using hmmsearch in the HMMER suite, for more distant candidate homologs against a

local custom database of diverse eukaryotic proteomes predicted from genomes (see above). All hits with E-values lower than 1E-02 were kept and then also confirmed by searching for known domains and motifs. New profile HMMs were built from a broader and more diverse set of candidate homologs, and hmmsearch searches were repeated. PSI-BLAST and BLASTP searches using diverse queries were also performed against subsets of the NCBI GenBank non-redundant (nr) database to increase the confidence in our candidate homologs.

The curated UniProtKB database was also searched using the new profile HMMs, as well as Pfam profile HMMs, for all MICOS subunits. A third, more inclusive, profile HMM for Mic60 was built using additional alphaproteobacterial sequences found after searching the UniProtKB database, and the search was then repeated to more comprehensively recover divergent alphaproteobacterial Mic60 homologs. A set of transcriptomes from diverse eukaryotes was also searched using the profile HMMs of the Mic60 and Mic10 subunits to better cover major eukaryote groups underrepresented by genomes.

### 2.2.3. Structural annotation

Alphaproteobacterial Mic60 homolog candidates identified in our similarity searches were structurally annotated using diverse predictors. The conserved C-terminal Mitofilin domain was annotated based on the program hmmscan in the HMMER suite (Eddy, 1998). Transmembrane domains were predicted using both the programs TMHMM (Krogh, Larsson, von Heijne, & Sonnhammer, 2001) and TMpred (Hofmann, 1993). Coiled-coil regions were annotated using the program COILS (Lupas, Van Dyke, & Stock, 1991).

### 2.2.4. Phylogenetic inference

Candidate homologs for the Mic26 subunit were aligned using MUSCLE v3.6 (Edgar, 2004) and manually inspected in MESQUITE (Maddison & Maddison, 2014). After inferring a preliminary phylogenetic tree, species-specific duplicates (*Danio rerio* and *Allomyces macrogynus*), as well as divergent (*Salpingoeca rosetta*) and highly gappy (*Schizosaccharomyces pombe*, *Monosiga brevicollis*, *Capsaspora owczarzaki*) sequences were removed. The best-fitting evolutionary model to the alignment was estimated using ProtTest v2.4 (Abascal, Zardoya, & Posada, 2005), and phylogenetic

analyses were performed using MrBayes v3.2.1 (Ronquist & Huelsenbeck, 2003), PhyML v2.4.4 (Guindon et al., 2010), and RAxML v2.2.3 (Stamatakis, 2014).

## 2.3. RESULTS AND DISCUSSION

### 2.3.1. MICOS is an ancient multi-protein complex of mitochondria

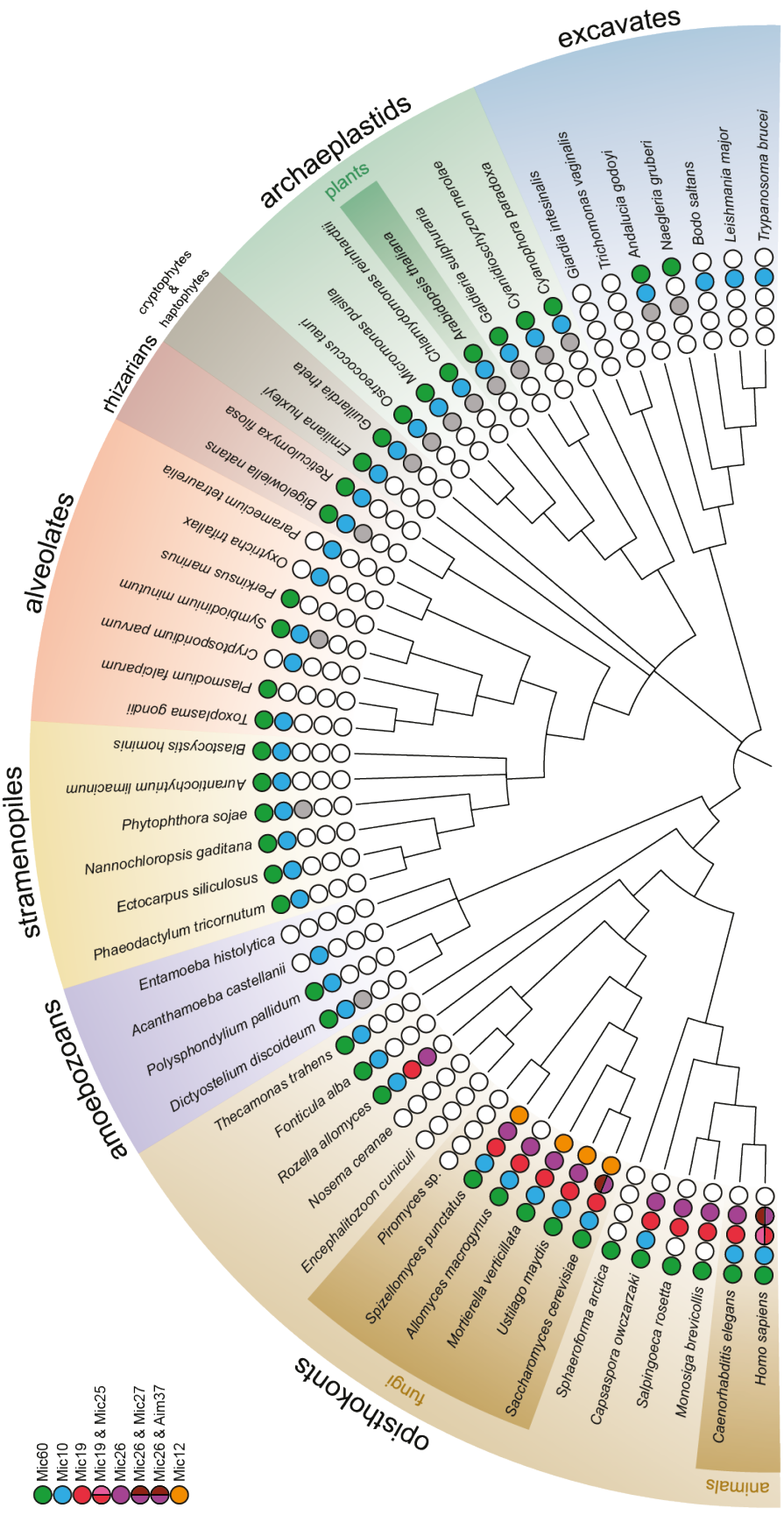
The composition of the MICOS complex had been investigated in the fungus *S. cerevisiae* (e.g., Harner et al., 2011; Hoppins et al., 2011; von der Malsburg et al., 2011), and the animals *C. elegans* (e.g., Head, Zulaika, Ryazantsev, & van der Bliek, 2011) and *H. sapiens* (e.g., An et al., 2012; Darshi et al., 2011). Even though fungi and animals are ancient lineages that diverged more than 902 Ma (mega-annum or millions of years ago; Eme, Sharpe, Brown, & Roger, 2014), they are sister taxa and are relatively closely related in the tree of all eukaryotes (Burki, 2014). Our first aim was to investigate the phylogenetic distribution across eukaryotes of MICOS subunits that had been characterized in both animals and fungi. To achieve this, we used a combination of similarity search methods aimed at identifying remote similarity that included BLAST, PSI-BLAST, and profile HMM searches (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

Two MICOS subunits, namely Mic60 and Mic10, were found in all major groups across the eukaryote tree (Fig. 2.2; Muñoz-Gómez et al., 2015, and see also Huynen, Mühlmeister, Gotthardt, Guerrero-Castillo, & Brandt, 2016). The evolutionary conservation of these two subunits agrees with the functional data derived from experiments in *S. cerevisiae* and *H. sapiens*. These two subunits were thought to be the core functional components of MICOS and are now known to be the central structural components of the two major MICOS subcomplexes (Friedman et al., 2015; Guarani et al., 2015; Huynen et al., 2016). Disrupting Mic60 or Mic10 leads to the most severe phenotypes, which include the near complete loss of CJs and the accumulation of crista membrane stacks in the mitochondrial matrix (e.g., Harner et al., 2011). Because these two core subunits are central to the formation of CJs, our finding suggests that CJs are widespread in aerobic mitochondria, even though they have only been characterized in animals, fungi and amoebae (Deng et al., 1999; Mannella et al., 1994; Nicastro, Frangakis, Typke, & Baumeister, 2000). The conservation of both Mic60 and Mic10 also suggests that all eukaryotes with aerobic mitochondria have the same general pathway

for the formation and maintenance of cristae by means of CJs (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

Another MICOS subunit, Mic19, also appears to have a broad distribution among eukaryotes (Fig. 2.2; Huynen et al., 2016; Muñoz-Gómez et al., 2015). This certainly agress with the experimental evidence which shows that Mic19 is among the most functionally important MICOS subunit after Mic60 and Mic10 (e.g., see Harner et al., 2011). Unfortunately, this subunit appears to be less conserved in sequence and therefore its presence in eukaryotes other than opisthokonts (animals, fungi and their closest relatives) was much more uncertain (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). The Mic26 subunit has a more restricted distribution in eukaryotes and appears to be found only in opisthokonts (Fig. 2.2; Muñoz-Gómez et al., 2015). Mic12 appears to be restricted to fungi (Fig. 2.2; Muñoz-Gómez et al., 2015). However, a later study which used profile-profile HMM searches showed that Mic12 has homologs in animals (QIL1 in *H. sapiens*) and plants (Huynen et al., 2016). Mic25 and Mic27 appear to have even more restricted phylogenetic distributions, being found only among vertebrates within animals (Fig. 2.2; Muñoz-Gómez et al., 2015). Similarly, Aim37 (later renamed as Mic27 as well) is found only within fungi, and is specifically restricted to the Saccharomycetales (Fig. 2.2; Muñoz-Gómez et al., 2015).

**Figure 2.2.** Distribution of MICOS subunits across the eukaryote evolutionary tree. Consensus evolutionary tree of eukaryotes (Adl et al., 2012). Major eukaryote groups are represented by species for which whole-genome data are available. The classical multicellular lineages (animals, fungi, and plants) are highlighted within these major groups. MICOS subunits were identified by a combination of BLAST, PSI-BLAST and profile HMM searching methods. Colored circles at tips indicate presence of MICOS subunits. Mic60, Mic10, Mic19, Mic25, Mic26, and Aim37/Mic27 are indicated by green, blue, red, red-pink, purple, and purple-brown circles, respectively. Gray circles indicate potential Mic19 orthologs in non-opisthokont lineages. The Mic12 subunit was more recently also found in animals and plants by Huynen et al., (2016). See also Fig. S2.1 for the evolutionary relationships between the Mic26, Mic27 and Aim37 subunits.


excavates
archaeplastids
plants
cryptophytes & haptophytes
rhizarians
alveolates
stramenopiles
amoebozoans
opisthokonts
fungi
animals
Trypanosoma brucei
Leishmania major
Bodo saltans
Naegleria gruberi
Andalucia godoyi
Trichomonas vaginalis
Giardia intestinalis
Cyanidioschyzon merolae
Galdieria sulphuraria
Arabidopsis thaliana
Chlamydomonas reinhardtii
Micromonas pusilla
Ostreococcus tauri
Guillardia theta
Emiliana huxleyi
Reticulomyxa filosa
Bigelowiella natans
Paramecium tetraurelia
Oxytricha trifallax
Perkinsus marinus
Symbiodinium minutum
Cryptosporidium parvum
Plasmodium falciparum
Toxoplasma gondii
Blastocystis hominis
Aurantiochytrium limacinum
Phytophthora sojae
Nannochloropsis gaditana
Ectocarpus siliculosus
Phaeodactylum tricornutum
Entamoeba histolytica
Acanthamoeba castellanii
Polysphondylium pallidum
Dictyostelium discoideum
Thecamonas trahens
Fonticula alba
Rozella allomyces
Nosema ceranae
Encephalitozoon cuniculi
Piromyces sp.
Spizellomyces punctatus
Allomyces macrogynus
Mortierella verticillata
Ustilago maydis
Saccharomyces cerevisiae
Sphaeroforma arctica
Capsaspora owczarzaki
Salpingoeca rosetta
Monosiga brevicollis
Caenorhabditis elegans
Homo sapiens
Mic60
Mic10
Mic19
Mic19 & Mic25
Mic26
Mic26 & Mic27
Mic26 & Aim37
Mic12

The MICOS complex is therefore an ancient multi-protein complex that is inferred to have been present in the mitochondrion of the eukaryote cenancestor, or the last eukaryote common ancestor from which all modern eukaryotes evolved (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). This cenancestral MICOS complex was composed of at least four subunits, namely Mic60, Mic10, Mic19 and Mic12 (see Fig. 2.2; but also Huynen et al., 2016; Muñoz-Gómez et al., 2015); note that the inference that Mic12 was ancestral to eukaryotes is based on the findings of Huynen et al., 2016. Other MICOS subunits, such as Mic25, Mic26, and Mic27 are derived and represent recent acquisitions of only some eukaryote groups (e.g., animals or fungi, or both; Fig. 2.2).

### 2.3.2. The presence of MICOS correlates with the presence of cristae

Despite the wide phylogenetic distribution of the MICOS complex across eukaryotes, there are some species for which no MICOS subunits could be found in their predicted proteomes (from sequenced genomes). These include, for example, the fungi *Piromyces* sp., *Encephalitozoon cuniculi*, and *Nosema ceranae*, the amoebozoan *Entamoeba histolytica*, and the excavates *Giardia intestinalis* and *Trichomonas vaginalis* (Fig. 2.2; Huynen et al., 2016; Muñoz-Gómez et al., 2015). These are all species known to have adapted to anoxic or hypoxic environments and in which mitochondria are functionally and structurally divergent and often reduced. It is therefore not a coincidence that mitochondria in these species, also called hydrogenosomes or mitosomes sometimes (or mitochondrion-related organelles), bear no cristae (Stairs et al., 2015). The absence of mitochondrial cristae in these species is thus accompanied by the lack of a MICOS complex (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

Other eukaryotes that also have divergent and structurally simplified mitochondria are *Blastocystis* and *Cryptosporidium*. However, these two eukaryotes have some components of a respiratory chain (electron transport chain, ETC) such as complex I and II, and an alternative oxidase (AOX) (Mogi & Kita, 2010; Zierdt, Donnolley, Muller, & Constantopoulos, 1988), and also develop cristae (e.g., *Blastocystis* and *Cryptosporidium muris*; Uni, Iseki, Maekawa, Moriya, & Takada, 1987; Zierdt, Donnolley, Muller, & Constantopoulos, 1988) or more random MIM folds (e.g., *Cryptosporidium parvum*; Keithly, Langreth, Buttle, & Mannella, 2005). Interestingly, we could detect MICOS subunits in these eukaryotes, such as both Mic60 and Mic10 in *Blastocystis*, and

Mic10 in *Cryptosporidium* (Fig. 2.2; Muñoz-Gómez et al., 2015). This finding suggests that mitochondria with partial respiratory chains might still require a MICOS complex, or some of its subunits, to aid in the development of cristae (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

### 2.3.3. Cristae morphology is not correlated with MICOS composition

The shape or form of mitochondrial cristae is relatively well conserved within major eukaryote groups and has therefore been used as a taxonomic character (Taylor, 1976). Three broad crista shapes are known for most eukaryotes: lamellar, tubular and discoidal (paddle-like), even though these can be further subdivided and more precisely defined (Seravin, 1993). However, the molecular bases for the shape (or morphology) of cristae have largely remained unknown. Given the causal role that MICOS plays in the development of mitochondria cristae, we asked whether there were any correlations between crista shape and the composition of the MICOS complex across eukaryotes.

Although there is a good correspondence between the presence or absence of cristae and MICOS, it is not obvious that the phylogenetic distribution of certain MICOS subunits is correlated with crista shapes (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). For example, plants (or archaeplastids) have lamellar, or also called flat, cristae, whereas alveolates and heterokonts (or stramenopiles) have tubular cristae, but in all three lineages, there are species that have the same core MICOS components (Mic60, Mic10, and Mic19; Fig. 2.2). Similarly, animals and fungi also have lamellar cristae despite their MICOS complexes containing more subunits than those of plants (Fig. 2.2). Although euglenozoans appear to have lost Mic60, the heteroloboseans *Naegleria gruberi* and *Percolomonas* sp. retain Mic60, indicating that the defining discoidal cristae of discicristates (i.e., eukaryotes with discoidal cristae) are not correlated with a specific known MICOS subunit composition (Fig. 2.2). Crista shape is therefore unlikely caused by the presence or the specific composition of the MICOS complex (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

This conclusion makes sense in the light of recent studies that show the important role that the $F_1F_O$-ATP synthase complex plays in shaping cristae. In animals, fungi and land plants, the ATP synthase complexes form V-shaped dimers (with different angles between the monomers), and these dimers themselves assemble into straight rows that introduce strong curvature at crista rims (Davies et al., 2012; Kühlbrandt, 2015; Strauss

et al., 2008). In ciliates, the ATP synthase forms U-shaped dimers which then assemble into helical rows that are responsible for the helically tubular shape of cristae in ciliate mitochondria (Allen, 1995; Mühleip et al., 2016). In euglenozoans, which have discoidal (paddle-like) cristae, V-shaped ATP synthase dimers assemble into short helical rows at the ridges of cristae (Mühleip, Dewar, Schnaufer, Kühlbrandt, & Davies, 2017).

### 2.3.4. Gene duplications have given rise to some MICOS subunits

Some MICOS subunits share sequence similarity and also have domains or motifs in common. For example, Mic26, Mic27 and Aim37 are obviously related and all three have an ApoO (apolipoprotein O) domain. This domain was subsequently shown to be able to bind to cardiolipin, a lipid only found in the MIM (Koob & Reichert, 2014). Phylogenetic analyses of this protein family reveal that Mic26 and Mic27 in animals, and Mic26 and Aim37 in fungi, originated from independent gene duplications that occurred at the origin of the Vertebrata and the Saccharomycetales, respectively (Fig. S2.1; Muñoz-Gómez et al., 2015). Consequently, Mic26 and Mic27 are paralogs in vertebrates, and Mic26 and Aim37 are paralogs in fungi. The ancestral MICOS subunit from which Mic26, Mic27 and Aim37 evolved is here named Mic26 because this was the first protein to have been named for this protein family (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). The evolutionary relationships among these MICOS subunits in different eukaryotes is therefore not easily described, and different conflicting nomenclatures for these genes have been proposed (e.g., Huynen et al., 2016; Muñoz-Gómez et al., 2015; Ott et al., 2015; Pfanner et al., 2014; Rampelt et al., 2016). Here I have abandoned our initial nomenclatural proposal, namely renaming Mic27 for Mic28 in the Saccharomycetales (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015), and have instead adopted the functional nomenclature suggested by yeast cell biologists where Aim37 is named Mic27 in the Saccharomycetales (Rampelt et al., 2016). This compromise, however, is far from optimal but allows for easier discussion of functional results.

Mic25 is another MICOS subunit that is similar to Mic19. Both proteins have a DUF737 domain of unknown function with coiled-coils followed by a CHCH ([coiled coil 1]-[helix 1]-[coiled coil 2]-[helix 2]) domain. Even though Mic19 is relatively widespread among eukaryotes, Mic25 is phylogenetically restricted to the Vertebrata. Therefore, we infer that Mic19 and Mic25 are paralogs that evolved before the diversification of modern vertebrates (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

Unfortunately, considerable sequence divergence among these proteins and their short lengths preclude rigorous phylogenetic analyses.

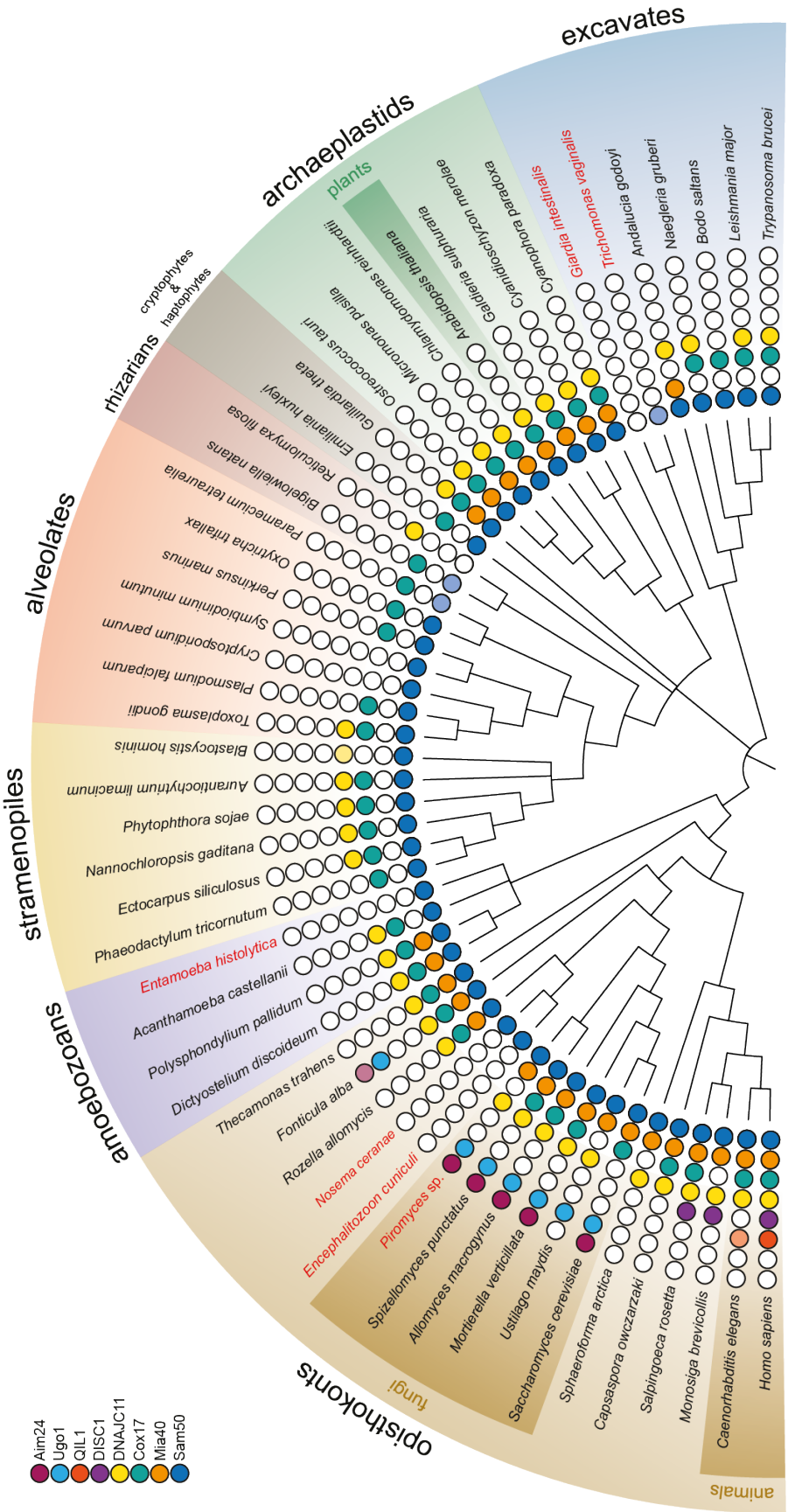
### 2.3.5. The phylogenetic distribution of MICOS interactors

MICOS, or its subunits, have been reported to interact with several proteins at the mitochondrial envelope in both *S. cerevisiae* and *H. sapiens*. In *S. cerevisiae*, for example, MICOS interacts with Tom40 of the TOM (Translocase of the Outer Mitochondrial membrane) complex, Sam50 of the SAM (Sorting and Assembly Machinery) complex, VDAC (Voltage-Dependent Anion Channel), Mia40 (Mitochondrial intermembrane space import and assembly protein 40), and Ugo1 (Mitochondrial fusion and transport protein) (Harner et al., 2011; Hoppins et al., 2011; Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012). Other proteins that have been shown to physically interact with MICOS are DNAJC11 (DnaJ heat shock protein family Hsp40 member C11), DISC-1 (Disrupted in schizophrenia 1), and QIL1 (later found to be a *bona fide* MICOS subunit, Mic12) in *H. sapiens* (Guarani et al., 2015; Park et al., 2010; Xie, Marusich, Souda, Whitelegge, & Capaldi, 2007), and Aim24 (Altered inheritance of mitochondria protein 24), and Cox17 (cytochrome c oxidase copper chaperone) in *S. cerevisiae* (Chojnacka, Gornicka, Oeljeklaus, Warscheid, & Chacinska, 2015; Harner et al., 2014). Indeed, the multiplicity of interactions that MICOS has at the mitochondrial envelope has led to the view that MICOS also functions as the protein scaffold of a larger network of protein complexes termed ERMIONE (ER-mitochondria organizing network) that controls mitochondrial function and biogenesis in *S. cerevisiae* (van der Laan et al., 2012; Wideman & Muñoz-Gómez, 2016).

In order to infer whether these interactions are ancestral or derived features in the evolutionary history of MICOS, we investigated the phylogenetic distribution of these MICOS-interacting proteins (Fig. 2.3; Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015). We show that Sam50, Mia40, Cox17, and DNAJC11 are widely distributed among eukaryotic diversity, suggesting their ancestral nature (Fig. 2.3). The distribution of Sam50 is largely congruent with that of other mitochondrial β-barrels, Tom40 and VDAC (Wideman, Gawryluk, Gray, & Dacks, 2013). Mia40 is also widespread, but absent from most acristate eukaryotes, as well as from members of the SAR clade (i.e., stramenopiles, alveolates, and rhizarians) and discicristates (e.g., *Naegleria gruberi*, *Bodo saltans*, *Trypanosoma brucei* and *Leishmania major*; Fig. 2.3). Cox17 and

DNAJC11 are similarly widespread but show more irregular distributions (Fig. 2.3). On the other hand, Aim24, Ugo1, QIL1, and DISC1 have more restricted phylogenetic distributions (Fig. 2.3). Both Aim24 and Ugo1 are specific to the Holomycota (fungi and their amoeboid relatives, e.g., *Fonticula* and nucleariids), and are only absent from the divergent microsporidians and Cryptomycota (*Rozella allomycis*). DISC1 appears to be present among animals and some of their single-celled relatives (e.g., choanoflagellates), whereas QIL1 is only found among animals (Fig. 2.3). Interestingly, with the exception of *Piromyces* sp., lineages that lack MICOS (i.e., microsporidians, *Entamoeba histolytica*, *Giardia intestinalis*, and *Trichomonas vaginalis*) also lack all MICOS-interacting proteins (with the exception of the ubiquitous and essential Sam50) (Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015).

**Figure 2.3.** Phylogenetic distribution of MICOS-interacting proteins. Similarity searches were performed as previously described (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). Species in red are those that have lost MICOS. Light color circles indicate potential orthologs with weaker sequence similarity. In the case of Sam50, highly divergent ciliate candidate orthologs were found using PSI-BLAST with the closest available Sam50 gene sequence (e.g., *Chromera velia* Cvel_14064). Although we could not detect a Sam50 ortholog in *Trichomonas vaginalis* with our bioinformatics methods, its presence in *T. vaginalis* hydrogenosomes is supported by experimental data previously reported (Kay, Lawler, & Kerr, 2013).


excavates
archaeplastids
plants
rhizarians
cryptophytes & haptophytes
alveolates
stramenopiles
amoebozoans
opisthokonts
fungi
animals
Trypanosoma brucei
Leishmania major
Bodo saltans
Naegleria gruberi
Andalucia godoyi
Trichomonas vaginalis
Giardia intestinalis
Cyanophora paradoxa
Cyanidioschyzon merolae
Galdieria sulphuraria
Arabidopsis thaliana
Chlamydomonas reinhardtii
Micromonas pusilla
Ostreococcus tauri
Guillardia theta
Emiliania huxleyi
Reticulomyxa filosa
Bigelowiella natans
Paramecium tetraurelia
Oxyfticha trifallax
Perkinsus marinus
Symbiodinium minutum
Cryptosporidium parvum
Plasmodium falciparum
Toxoplasma gondii
Blastocystis hominis
Aureococchytrium limacinum
Phytophthora sojae
Nannochloropsis gaditana
Ectocarpus siliculosus
Phaeodactylum tricornutum
Entamoeba histolytica
Acanthamoeba castellanii
Polysphondylium pallidum
Dictyostelium discoideum
Thecamonas trahens
Fonticula alba
Rozella allomycis
Nosema ceranae
Encephalitozoon cuniculi
Piromyces sp.
Spizellomyces punctatus
Allomyces macrogynus
Mortierella verticillata
Ustilago maydis
Saccharomyces cerevisiae
Sphaeroforma arctica
Capsaspora owczarzaki
Salpingoeca rosetta
Monosiga brevicollis
Caenorhabditis elegans
Homo sapiens
Aim24
Ugo1
QIL1
DISC1
DNAJC11
Cox17
Mia40
Sam50

The phylogenetic co-occurrence of MICOS and its interactors allow us to make some inferences about the functional evolution of the MICOS complex (Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015). Sam50, of the SAM complex, is the only MICOS interactor that is also found in alphaproteobacteria among those MICOS interactors that are widespread and therefore ancestral to eukaryotes (i.e., Sam50, Tom40, VDAC, Mia40, Cox17 and DNAJC11; Fig. 2.3). The interaction of MICOS with both SAM and TOM is presumed to bring together both translocases for the efficient transfer of β-barrel proteins from one complex to the other (Körner et al., 2012; Ott et al., 2012; Zerbes, Bohnert, et al., 2012). The homolog of Sam50 in alphaproteobacteria is BamA, the central channel of the BAM complex that is required for β-barrel assembly at the outer membrane of the alphaproteobacterial envelope (Paschen et al., 2003; Voulhoux, Bos, Geurtsen, Mols, & Tommassen, 2003). This inferred ancestral interaction is certainly consistent with the experimental evidence which indicates that MICOS interacts with the SAM complex in quite disparate eukaryotes, namely, animals (e.g., *H. sapiens*), fungi (e.g., *S. cerevisiae*) and euglenozoans (e.g., *T. brucei*; Hassan Hashimi, personal communication). (See Chapter 3 for a more detailed discussion of the relevance of this ancestral MICOS interaction.)

The TOM complex and the MIA (Mitochondrial Intermembrane space Assemby) pathway are also inferred to have been ancestral to eukaryotes based on their widespread distributions (e.g., see Fig. 2.3). In yeast mitochondria, the interaction of MICOS with TOM and Mia40 positions both complexes in close proximity for the correct oxidative folding of translocated proteins (Varabyova et al., 2013; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012). Most recently, an interaction between MICOS and an analogue of Mia40 has been discovered in *T. brucei* (Hassan Hashimi, personal communication). The interaction of MICOS with the VDAC porin is probably also ancestral, and it is hypothesized to concentrate VDAC porins in the vicinity of CJs to increase the diffusion of metabolites into the intracristal space (Hoppins et al., 2011; van der Laan et al., 2012). Other potential ancestral MICOS interactors are DNAJC11 and Cox17 but the functions of these interactions are more difficult to conceive (Chojnacka et al., 2015; Xie et al., 2007)—the DnaJ domain, but not the DUF3395 domain, of DNAJC11 is also found among prokaryotes.

Other MICOS-interacting proteins and functions can be inferred to have evolved more recently. For example, a MICOS connection with the mitochondrial fusion machinery

might predate the diversification of animals and fungi (i.e., opisthokonts), because even though Ugo1 is restricted to fungi and their closest relatives (Fig. 2.3), a potential homolog of Ugo1 was recently reported in humans (Janer et al., 2016). Similarly, the MICOS-stabilizing factor QIL1 in humans is a distant homolog of the fungal MICOS subunit Mic12 (Huynen et al., 2016). Another MICOS-stabilizing factor, the protein Aim24, appears to be specific to fungi. These more derived MICOS-interacting proteins point to the inherent evolvability of MICOS and suggest that numerous other interactions likely evolved in other understudied eukaryote lineages (as it has been more recently described in the euglenozoan *T. brucei*; Hassan Hashimi, personal communication).

After our study had been published (Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015), several other MICOS interactors were reported. These include the nucleoid-associated proteins TFAM, TFB2M and TFB1M (Yang et al., 2015), a putative homolog of the fusion adapter protein Ugo1 (SLC25A46) (Janer et al., 2016; see above), and the cardiolipin-synthesizing enzymes PGS1 and CSL1 (Serricchio, Vissa, Kim, Yip, & McQuibban, 2018). All these interactors were reported for MICOS in *H. sapiens* and interact with either Mic60 or both Mic60 and Mic19. Furthermore, a clear physical interaction between Mic10 and the dimeric $F_1F_O$-ATP synthase complex have been reported in *S. cerevisiae* (see Rampelt et al., 2017); subunits *e* and *g* of the ATP synthase complex, and which are reposnible for its dimerization, are likely ancestral to eukaryotes (Kühlbrandt, 2015). Future work should aim at elucidating the phylogenetic distributions of these proteins and confirming their interactions with MICOS in eukaryotes other than animals or fungi.

### 2.3.6. MICOS has alphaproteobacterial roots

The inference that MICOS was present in the mitochondrion of the eukaryote cenancestor (the cenancestral mitochondrion) encouraged us to search for MICOS subunits among prokaryotes (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). We performed profile HMM searches against the UniProtKB database and found candidate homologs for Mic60, the largest and most central MICOS subunit, in prokaryotes (Fig. 2.4). However, we could not find candidate homologs for any other MICOS subunit among prokaryotes.

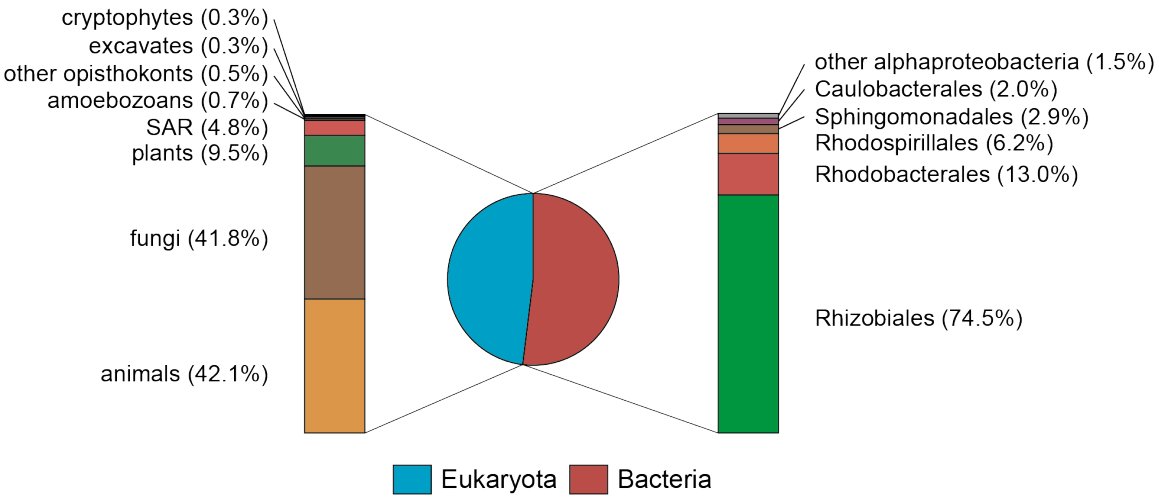
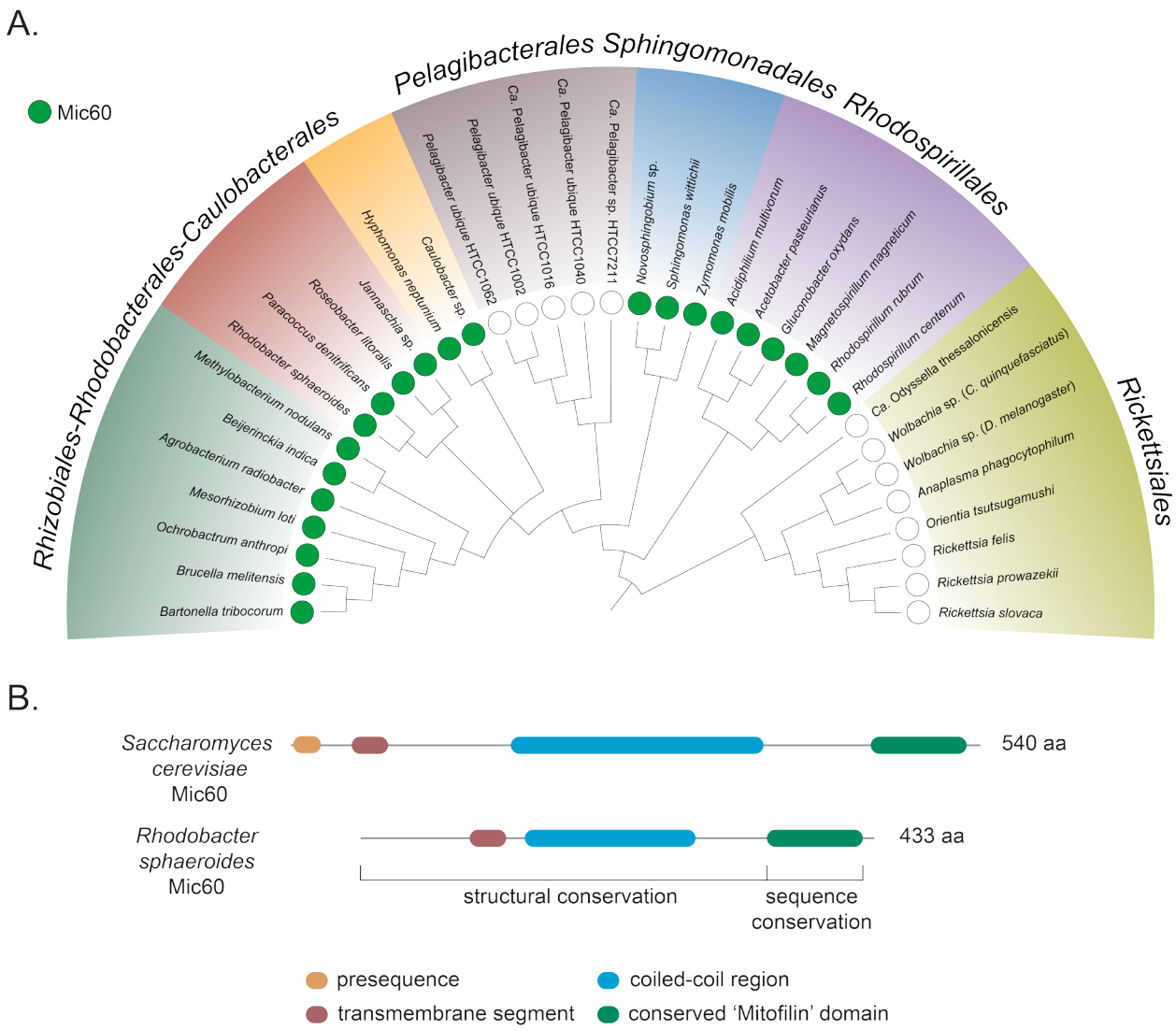



**Figure 2.4.** Distribution of Mic60 homologs in the UniProtKB database. Hidden Markov Model (HMM) profiles built from a representative set of both eukaryotic and prokaryotic Mic60 homologs were used to search the curated database UniProtKB as of June 2015. 608 Mic60 homologs from representatives of each major eukaryote supergroup were detected. Among the non-eukaryotic Mic60 homologs detected, all of them (663) belong to the alphaproteobacteria. Three prokaryotic Mic60 homologs detected outside alphaproteobacteria were clear false positives (data not shown) as they had very low scores and E-values, and were more similar to other characterized domains.

Inspecting the taxonomic distribution of the prokaryotic Mic60 homologs reveals that they are restricted to the *Alphaproteobacteria*, the bacterial group from which mitochondria evolved or that is most closely related to mitochondria (Fig. 2.4; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). Within the *Alphaproteobacteria*, the Mic60 homolog is widespread but appears to be absent from the *Pelagibacterales* (or SAR11 clade) and the *Rickettsiales* (Fig. 2.4). Both groups exhibit small genomes with a reduced gene content as the outcome of adaptation to the surface ocean or an intracellular environment, respectively. However, the *Pelagibacterales* is a derived order within the *Alphaproteobacteria*, whereas the *Rickettsiales* is sister to all other alphaproteobacterial orders (see Chapter 4). Therefore, it is uncertain whether the Mic60 homolog evolved before the diversification of all modern alphaproteobacteria (and was lost in the *Rickettsiales*), or if it first evolved in all other alphaproteobacteria after their divergence from the *Rickettsiales* order. The central MICOS component, Mic60, thus appears to have first evolved in the *Alphaprotebacteria* (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015).

We then examined the structure of the alphaproteobacterial Mic60 homolog in order to gain insights into its function. Our bioinformatic predictions of the domains, motifs, and secondary structures for alphaproteobacterial Mic60 homologs suggest that the general structure of Mic60 is conserved both in eukaryotes and alphaproteobacteria (Fig. 2.5). In eukaryotes, Mic60 consists of an N-terminal presequence (a targeting signal to the mitochondrion), followed by a transmembrane domain, then a middle coiled-coil region, and finally a C-terminal Mitofilin signature domain. The alphaproteobacterial Mic60 homolog has basically the same predicted structure, with the exception of the N-terminal presequence which is missing, as expected (Fig. 2.5).

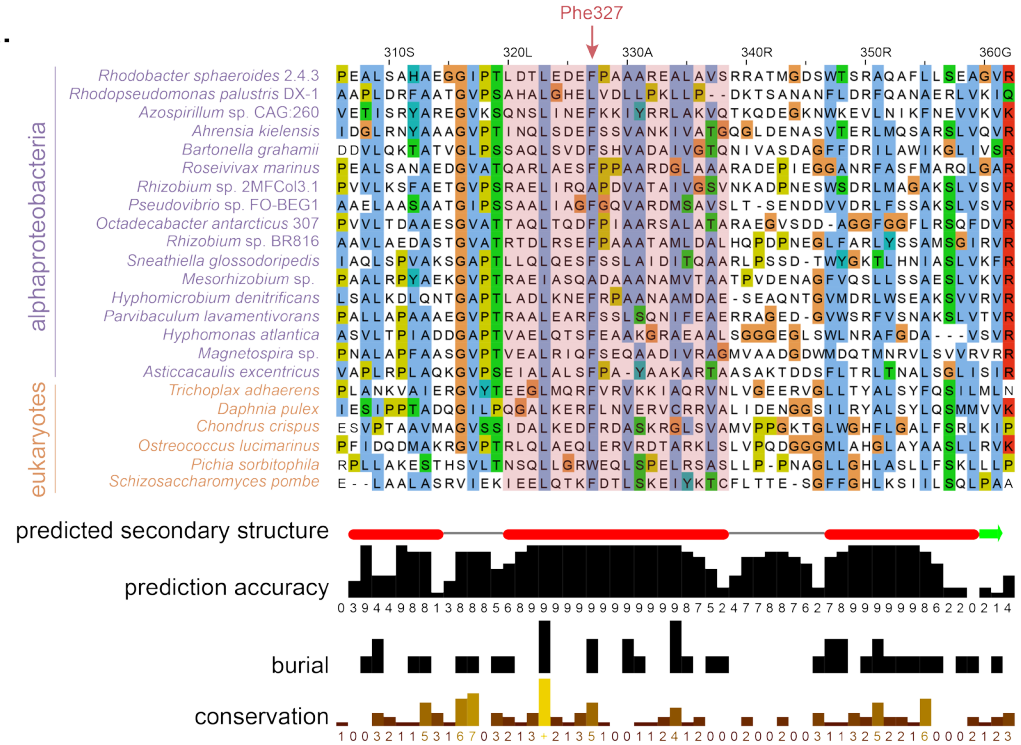
**Figure 2.5.** Phylogenetic distribution of Mic60 in the *Alphaproteobacteria* and its conserved overall structure. **A.** Phylogenetic distribution of Mic60 homologs in the *Alphaproteobacteria*. A consensus phylogenetic tree of the *Alphaproteobacteria*, with a representative set of species from the major groups, is derived from the literaturature. Alphaproteobacterial homologs of Mic60 were identified by HMM searches. **B.** Conserved Mic60 protein domain architecture in eukaryotes and prokaryotes. Structural motif predictions were carried out using TMHMM (Krogh et al., 2001), TMpred (Hofmann, 1993), and COILS (Lupas et al., 1991).


A.
Mic60
Rhizobiales-Rhodobacterales-Caulobacterales
Pelagibacterales
Sphingomonadales
Rhodospirillales
Rickettsiales
Bartonella tribocorum
Brucella melitensis
Ochrobactrum anthropi
Mesorhizobium loti
Agrobacterium radiobacter
Beijerinckia indica
Methylobacterium nodulans
Rhodobacter sphaeroides
Paracoccus denitrificans
Roseobacter litoralis
Jannaschia sp.
Hyphomonas neptunium
Caulobacter sp.
Pelagibacter ubique HTCC1062
Pelagibacter ubique HTCC1002
Ca. Pelagibacter ubique HTCC1016
Ca. Pelagibacter ubique HTCC1040
Ca. Pelagibacter sp. HTCC7211
Novosphingobium sp.
Sphingomonas wittichii
Zymomonas mobilis
Acidiphilium multivorum
Acetobacter pasteurianus
Gluconobacter oxydans
Magnetospirillum magneticum
Rhodospirillum rubrum
Rhodospirillum centenum
Ca. Odyssella thessalonicensis
Wolbachia sp. (C. quinquefasciatus)
Wolbachia sp. (D. melanogaster)
Anaplasma phagocytophilum
Orientia tsutsugamushi
Rickettsia felis
Rickettsia prowazekii
Rickettsia slovaca
B.
Saccharomyces cerevisiae Mic60
540 aa
Rhodobacter sphaeroides Mic60
433 aa
structural conservation
sequence conservation
presequence
transmembrane segment
coiled-coil region
conserved ‘Mitofilin’ domain

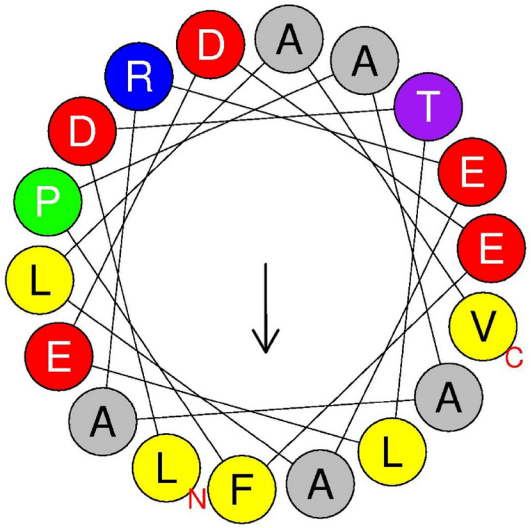
Mitochondrial Mic60 has recently been shown to also be able to deform the MIM at crista junctions (Hessenberger et al., 2017). This membrane-deforming capability depends on a lipid-binding site (LBS) that is found in between the middle coiled-coils and the C-terminal mitofilin domain *sensu stricto* (43 amino acids); the Pfam Mitofilin domain (PF09731), *sensu lato*, is defined differently and has an average size of 364.5 amino acids. This LBS corresponds to an amphipathic helix that presumably inserts itself into the MIM (Hessenberger et al., 2017). The LBS is extremely important for the function of Mic60 because its disruption leads to the loss of membrane binding and deformation, and also leads to phenotypes quite similar to those obtained when the whole Mic60 gene is removed (Hessenberger et al., 2017). I have found that this LBS and its corresponding amphipathic helix are well conserved not only across eukaryotes, but also in alphaproteobacteria (Fig. 2.6). Even the most critical amino acid within this LBS is conserved in some alphaproteobacteria relative to eukaryotes (Phe573 in the yeast *Chaetomium thermophilum* and Phe327 in the purple nonsulfur bacterium *Rhodobacter sphaeroides*; Fig. 2.6A). This observation constitutes important evidence for the functional conservation of Mic60 in both mitochondria and alphaproteobacteria.

**Figure 2.6.** Evolutionary conservation across mitochondria and alphaproteobacteria of the lipid binding site (LBS) that is found in between the coiled coils and the Mitofilin domain *sensu stricto* of Mic60. **A.** A predicted helix (highlighted in pink) that encompasses the LBS and the functionally critical amino acid Phe327 is predicted with high confidence and is also relatively well conserved across eukaryotes and alphaproteobacteria. **B.** The putative helix is predicted to be amphipathic in the purple nonsulfur bacterium *Rhodobacter sphaeroides* according to the HELIQUEST web server (Gautier, Douguet, Antonny, & Drin, 2008).

A.


Phe327
310S
320L
330A
340R
350R
360G
alphaproteobacteria
eukaryotes
Rhodobacter sphaeroides 2.4.3
Rhodopseudomonas palustris DX-1
Azospirillum sp. CAG:260
Ahrensia kielensis
Bartonella grahamii
Roseivivax marinus
Rhizobium sp. 2MFCol3.1
Pseudovibrio sp. FO-BEG1
Octadecabacter antarcticus 307
Rhizobium sp. BR816
Sneathiella glossodoripedis
Mesorhizobium sp.
Hyphomicrobium denitrificans
Parvibaculum lavamentivorans
Hyphomonas atlantica
Magnetospira sp.
Asticcacaulis excentricus
Trichoplax adhaerens
Daphnia pulex
Chondrus crispus
Ostreococcus lucimarinus
Pichia sorbitophila
Schizosaccharomyces pombe
predicted secondary structure
prediction accuracy
burial
conservation

B.


D
A
A
R
T
D
E
P
E
L
V
C
E
A
A
L
L
N
F
A

### 2.3.7. Alphaproteobacterial Mic60 might interact with the product of its highly syntenic neighboring gene

Eukaryotic MICOS is composed of at least four subunits (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015, and see above). However, only Mic60, the central scaffolding subunit of eukaryotic MICOS, has a homolog among the alphaproteobacteria (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). It is therefore conceivable that alphaproteobacterial Mic60 has an unknown interaction partner to accomplish its function in alphaproteobacterial envelopes. A phylogenetic profile of the genomic context of *mic60* across alphaproteobacteria reveals that *mic60* is associated with genes involved in the heme biosynthesis pathway (see Huynen et al., 2016 and Fig. S2.2). In most alphaproteobacterial genomes *mic60* is downstream of *hemC* (hydroxymethylbilane synthase; HMBS) and *hemD* (uroporphyrinogen-III synthase; UROS), and upstream of a hypothetical protein-coding gene sometimes misannotated as *hemY* (Fig. S2.2). All four genes have the same orientation and are usually tightly clustered with little intergenic space in between them, which suggests that they are co-transcribed as part of the same operon (Fig. S2.2). This is consistent with the regulatory requirements of Mic60 according to its hypothesized function (see below and Chapter 3).

Although, *hemC* and *hemD* are genuine enzymes of the heme biosynthesis pathway, the gene downstream of *mic60* clearly does not encode a heme biosynthetic enzyme. This protein is usually misannotated as HemY because it contains a well conserved hemY_N domain (PF07219) at its N-terminus. Confusingly, this domain is unrelated to genuine *hemY* (protoporphyrinogen IX oxidase), which is rather uncommon among alphaproteobacteria. Instead of *hemY*, most alphaproteobacteria use the product of *hemJ* to synthesize protoporphyrin IX (very few alphaproteobacteria have a genuine *hemY* gene homologous to that of *Escherichia coli* and *Bacillus subtilis*) (Kobayashi, Masuda, Tajima, Wada, & Sato, 2014). Here, the protein encoded by the gene downstream of *mic60* will be referred to as Orf52 based on its predicted size of 52 kDa in *R. sphaeroides* 2.4.1. Like alphaproteobacterial Mic60, Orf52 is an integral membrane protein, but it possesses two, instead of one, transmembrane segments at its N-terminus, and seems to expose its bulk to the periplasmic space. Moreover, Orf52 contains several tetratricopeptide repeat (TPR) motifs, which are usually involved in protein-protein interactions and found in proteins that are part of multiprotein complexes.

The motif order and composition of both Mic60 and Orf52 indicates that these proteins have structural roles in alphaproteobacterial envelopes (see Chapter 3 for a more focused discussion).

### 2.3.8. Alphaproteobacterial Mic60 evolved from a distant homolog in *Beta-*, and *Gammaproteobacteria*

The hypothetical protein interactor of alphaproteobacterial Mic60, namely Orf52, has a phylogenetic profile that extends to the beta-, and gammaproteobacteria. In these groups, orf52 is tightly clustered with a gene annotated as *hemX*, as well as *hemD* and *hemC* (Fig. S2.2). The synteny between the *hemC-hemD-hemX-orf52* gene cluster in beta-, and gammaproteobacteria, and the *hemC-hemD-mic60-orf52* gene cluster in alphaproteobacteria suggests that *mic60* and *hemX* are evolutionarily related, as argued by Huynen and colleagues (Huynen et al., 2016). Moreover, *hemX* is phylogenetically restricted to the beta-, and gammaproteobacteria (PF04375), thus complementing the distribution of *mic60* to the alpha-, beta- and gammaproteobacterial clade. This distribution corresponds to that of Orf52, which is also phylogenetically restricted to the alpha-, beta- and gammaproteobacterial clade (PF07219), unlike the *hemC* (PF01379) and *hemD* (PF02602) genes which are widespread among eubacteria. Similarly to Mic60, HemX has an N-terminal transmembrane segment and a central coiled-coil region (Huynen et al., 2016). However, HemX lacks a mitofilin signature domain, which is the only region sufficiently conserved at the sequence level between eukaryotic and alphaproteobacterial Mic60 (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). It is therefore probable that alphaproteobacterial Mic60 evolved from HemX which performs a more general role in the biogenesis of the beta- and gammaproteobacterial cell envelopes (in *E. coli*, HemX appears to be found as homo-oligomers; Stenberg et al., 2005).

## 2.4. CONCLUSIONS: A HYPOTHESIS FOR THE ORIGIN OF MITOCHONDRIAL CRISTAE

The origin of Mic60 in the *Alphaproteobacteria* raises the question of the function of this protein in a prokaryotic cell. The simplest hypothesis is that the function of Mic60 has remained conserved between alphaproteobacteria and mitochondria. However, functional shifts are known to be common in evolutionary history, especially when

considering huge time spans. Despite this, the hypothesis for the functional conservation between mitochondrial and alphaproteobacterial Mic60 makes predictions that can be tested.

The evolutionary analyses presented in this chapter have already provided evidence for the idea that the function of Mic60 in alphaproteobacteria might be similar to that of Mic60 in mitochondria. First is the presence of a mitofilin signature domain in alphaproteobacterial Mic60 (Muñoz-Gómez et al. 2015; Huynen et al. 2016). Second is the predicted structural conservation (same motif order and composition) between alphaproteobacterial and eukaryotic Mic60 (Muñoz-Gómez et al. 2015; Huynen et al. 2016). And third is the conservation of an amphipathic helix that serves as a lipid-binding site inside the mitofilin domain (mitofilin *sensu lato*) or next to it (mitofilin *sensu stricto*).

To make sense of the function of Mic60 in alphaproteobacterial cells, we need to bear in mind that many alphaproteobacteria develop intracytoplasmic membranes, or ICMs, that are used for bioenergetic purposes (e.g., photosynthesis, nitrification or methanotrophy; see Chapter 3 for details). These ICMs might resemble some crista shapes and had previously been thought to be the evolutionary precursors of cristae based on their superficial similarity (e.g., Stewart & Mattox, 1980, 1984, and see the introduction of Chapter 3). Little evidence was available, though, to take this idea seriously. Our discovery that MICOS has alphaproteobacterial roots provides new impetus to reconsider the idea that mitochondrial cristae evolved from alphaproteobacterial ICMs, and that both structures are therefore homologous. The next chapter develops in detail a hypothesis for the function of Mic60 in alphaproteobacterial envelopes and summarizes the current evidence that supports it.

# CHAPTER 3: WHAT IS THE FUNCTION OF MIC60 IN THE *ALPHAPROTEOBACTERIA*?

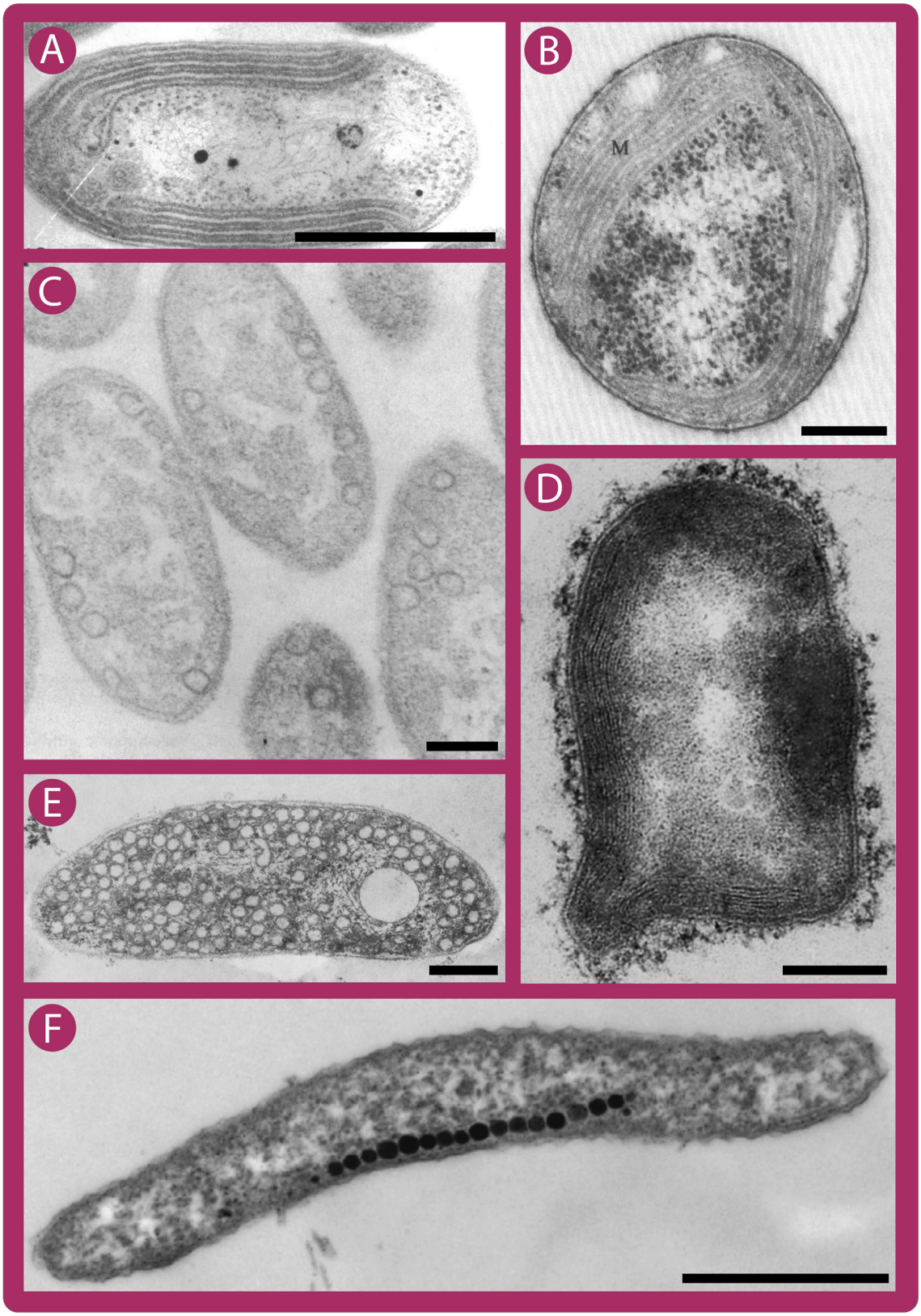
## 3.1. INTRODUCTION

At least three physiological "groups" in the *Alphaproteobacteria* develop extensive intracytoplasmic membrane (ICM) systems: the anoxygenic photosynthesizers (purple nonsulfur bacteria; a strictly physiological, nonphylogenetic term), the methanotrophs, and the nitrifiers (nitrite-oxidizing alphaproteobacteria) (Drews, 1992; Niederman, 2006; Pinevich, 1997) (Fig. 3.1). Most purple alphaproteobacteria (*Rhodopirillales*, *Sphingomonadales*, *Rhodobacterales*, and *Rhizobiales*) develop ICMs in the presence of light and the absence of oxygen (Drews, 1992; Drews & Golecki, 1995; Niederman, 2006) (Fig. 3.1E and F). Their ICMs house the photosynthetic apparatus and electron transport chain, which is generally composed of light-harvesting complexes 1 and 2 (LH1 and LH2), a type II reaction center (RC), a cytochrome $bc_1$, a periplasmic cytochrome $c_2$, and an ATP synthase. By means of cytochrome $bc_1$, which is also shared with the respiratory chain, the photosynthetic chain creates a proton motive force across the ICM that is harvested by the ATP synthase to produce ATP (Niederman, 2006). Methanotrophic alphaproteobacteria (*Methylocystaceae* and *Beijerinckiaceae*) use methane as a carbon and energy source (Tamas, Smirnova, He, & Dunfield, 2014) (Fig. 3.1B). The aerobic development of methanotrophic ICMs is strongly dependent on the expression of a particulate methane monooxygenase, the defining enzyme of methanotrophy (Hanson & Hanson, 1996; Niederman, 2006). The enzymes methanol dehydrogenase and formaldehyde dehydrogenase, which complete the oxidation of methane to formate, are also ICM-associated, being located in the intra-ICM space and the ICM, respectively (Brantner, Remsen, Owen, Buchholz, & Perille Collins, 2002; Zahn, Bergmann, Boyd, Kunz, & DiSpirito, 2001). These three ICM enzymes are coupled to the methanotrophic respiratory electron transport chain and ultimately the conversion of energy to ATP (DiSpirito, Kunz, Choi, & Zahn, 2004). The nitrifying alphaproteobacteria (the *Nitrobacter* genus) aerobically oxidize nitrite as their energy source (Ward, Arp, & Klotz, 2011) (Fig. 3.1A). The enzyme nitrite oxidoreductase (NOR or NXR), which oxidizes nitrite into nitrate, is also membrane-bound and localizes at the cytoplasmic face of ICM membranes (E Spieck, Ehrich, Aamand, & Bock, 1998). The

reducing power derived from nitrite oxidation is subsequently funneled through the respiratory electron chain of nitrifiers to produce ATP (Yamanaka & Fukumori, 1988).

The common feature of all the ICMs in these physiologically diverse groups of *Alphaproteobacteria* is their bioenergetic function. They all carry out processes that are tightly linked to the production of energy (ATP or reducing equivalents) (Drews & Golecki, 1995). Primarily, ICMs increase the surface area to accommodate increasing numbers of components that are connected to an electron transport chain, and ultimately ATP synthesis. All ICM types, as well, sustain electron transport chains that share at least two components with a typical aerobic respiratory chain: a cytochrome $bc_1$ complex, and the ATP synthase complex (Choi et al., 2003; DiSpirito et al., 2004; Spieck, Aamand, Bartosch, & Bock, 1996; Zahn et al., 2001); ICMs that develop aerobically (methanotrophic, nitrifying and some photosynthetic) might also harbor an oxygen-reducing cytochrome $aa_3$ complex. In purple alphaproteobacteria, ICMs provide a larger surface area for capturing light through photosynthetic pigments, as well as for protein complexes of the photosynthetic machinery. In nitrifying and methanotrophic alphaproteobacteria, ICM systems harbor the enzymatic machinery required for chemolithotrophy, which is considered to give a relatively low bioenergetic yield. In summary, the development of ICMs in *Alphaproteobacteria* likely compartmentalizes their bioenergetic metabolism, thereby concentrating the enzymatic reactions that occur within the intra-ICM space and increasing overall bioenergetic efficiency. In this sense, ICMs are functionally analogous to the respiratory cristae of mitochondria.

**Figure 3.1.** Members of the *Alphaproteobacteria* develop morphologically diverse ICMs to carry out different bioenergetic functions. **A.** *Nitrobacter winogradski* NB-255, a nitrifying alphaproteobacterium from the order *Rhizobiales* that develops flat ICMs at the cell periphery (scale bar: 1 μm; modified from Watson & Mandel, 1971). **B.** *Methylosinus trichosporium* OB3b, a methanotrophic alphaproteobacterium from the order *Rhizobiales* that develops flat ICMs at the cell periphery (scale bar: 0.2 μm; modified from Scott, Brannan, & Higgins, 1981). **C.** *Erythrobacter* sp. OCh114, an aerobic anoxygenic photosynthesizer from the order *Rhodobacterales* that develops few irregular and vesicular ICMs (scale bar: 0.2 μm; modified from Iba, Takamiya, Toh, & Nishimura, 1988). **D.** *Rhodoblastus sphagnicola*, an anaerobic anoxygenic photosynthesizer from the order *Rhizobiales* that develops flat ICMs at the cell periphery (scale bar: 0.5 μm; modified from Kulichevskaya, Guzev, Gorlenko, Liesack, & Dedysh, 2006). **E.** *Rhodobacter capsulatus* (formerly *Rhodobacter capsulata*), an anaerobic anoxygenic photosynthesizer from the order *Rhodobacterales* that develops abundant vesicular ICMs (scale bar: 0.2 μm; modified from Kaufmann, Reidl, Golecki, Garcia, & Drews, 1982). **F.** *Magnetospirillum gryphiswaldense*, a magnetotactic alphaproteobacterium from the order *Rhodospirillales* that develops ICMs in the form of magnetosomes to align itself along a geomagnetic field (scale bar: ~0.5 μm; modified from Schüler, 2008).


A
B
M
C
D
E
F

Historically, early transmission electron microscopic investigations revealed superficial similarities between the structure of mitochondrial cristae and the intracytoplasmic membranes (ICMs) of some bacteria (Munn, 1974). However, no explicit statements were made with regard to the evolutionary connection between these two kinds of structures. When the phylogenetic affiliation between mitochondria and purple nonsulfur bacteria (today's photosynthetic alphaproteobacteria) was first recognized based on the comparative biochemistry of the respiratory chain, mitochondrial cristae were interpreted as having evolved post-endosymbiotically in their specialization as respiratory organelles (John & Whatley, 1975). It was not until 1980 that Stewart and Mattox (based on parsimony and morphological considerations) explicitly proposed that mitochondrial cristae were derived from purple nonsulfur bacterial ICMs, and therefore had a pre-endosymbiotic origin (the homology hypothesis). Stewart and Mattox drew on the pioneering phylogenetic work of Dayhoff and Schwartz on organelle origins (Schwartz & Dayhoff, 1978) to suggest that mitochondria evolved polyphyletically from distinct endosymbiotic events with two purple nonsulfur bacteria, each with a different ICM morphology (Stewart & Mattox, 1980, 1984). Since then, Cavalier-Smith (1981, 1983a, 1983b) has also discussed these ideas in different contexts, and has recently argued in more detail that cristae evolved from vesicular ICMs that were present in an anoxygenic photosynthetic alphaproteobacterium (Cavalier-Smith, 2002, 2006, 2007). Others have also recently noted superficial morphological similarities between cristae and the ICMs developed by phylogenetically-derived alphaproteobacterial methanotrophs (Degli Esposti, 2014; López-García & Moreira, 2006; Moreira & López-García, 1998).

Other than these accounts, the evolutionary origin of mitochondrial cristae is seldom discussed in a literature that increasingly focuses on genomes and metabolism (e.g., Burger, Gray, Forget, & Lang, 2013; Müller et al., 2012). A major reason for this is that there was, until recently, no evidence for the evolutionary connection between cristae and ICMs beyond superficial morphological resemblance. We, and others, have shown that MICOS is an ancient mitochondrial protein complex of eukaryotes, and that its origin traces back to the *Alphaproteobacteria*, the bacterial progenitors of mitochondria (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015; Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015). In Chapter 2, I presented three main arguments that support the hypothesis that ICMs and cristae share a common ancestry, namely that alphaproteobacterial Mic60 has (1) a conserved mitofilin domain, (2) a conserved predicted general structure, and (3) even a critical amphipathic helix that

serves as a lipid-binding site to introduce membrane curvature. These observations tested the initial predictions of the more than three-decade-old “homology hypothesis” and revive the idea that mitochondrial cristae have a pre-endosymbiotic origin. The next logical steps are thus to test the ‘homology hypothesis’ using an experimental approach. The functional dissection of Mic60 (and its potential interactor Orf52) in a model alphaproteobacterium such as *Rhodobacter sphaeroides* or *Rhodopseudomonas palustris* is required to further test the homology hypothesis. In this chapter, I present preliminary experimental evidence aimed at testing the homology hypothesis and compile further evidence in support of the homology hypothesis between alphaproteobacterial ICMs and mitochondrial cristae.

## 3.2. MATERIALS AND METHODS

### 3.2.1. Bacterial strains, media and growth conditions

*Rhodopseudomonas palustris* TIE-1, a purple nonsulfur bacterium that develops lamellar ICMs, was kindly provided by Dianne K. Newman (California Institute of Technology) (Jiao, Kappler, Croal, & Newman, 2005). *Rhodobacter sphaeroides* 2.4.1, a purple nonsulfur bacterium that develops vesicular ICMs, was kindly provided by J. Thomas Beatty (University of British Columbia). *Escherichia coli* S17-1- λpir and DH5α-λpir were also gifts by the Beatty lab (University of British Columbia). *Escherichia coli* BW29427-λpir-RP4 was a gift by the Newman lab (California Institute of Technology). *S. cerevisiae* strains were provided by Melanie Dobson (Dalhousie University) and obtained from the Yeast Gene Deletion Strain Collection at Dalhousie University. See Table 3.1 for a list of strains used.

*Rhodopseudomonas palustris* was grown both chemo- and photo-heterotrophically at 30°C in YPS rich medium (Welander et al., 2012) (0.3% yeast extract, 0.3% peptone, 10 mM succinate, 100 mM MOPS, pH 7.0). *Rhodopseudomonas palustris* was grown photoheterotrophically at 30°C on FEM minimal medium (Ehrenreich & Widdel, 1994) (for 1L, 10 mL 100X basal medium stock, 22 mL 1M $NaHCO_3$, 1 mL 1M $NaHCO_3$, 1 mL 0.1 mg/mL vitamin B12, 1 mL trace element solution, 10 mL vitamin solution, and 10 mL 1M sodium acetate, pH 7.0). Basal medium stock solution: for 1L, 0.3 g $NH_4Cl$, 0.5 g $KH_2PO_4$, 0.1 g $CaCl_2{\cdot}2H_2O$, and 0.5 g $MgSO_4{\cdot}7H_2O$. Trace element solution: for 1 L, 3 g $Na_2$-EDTA$\cdot 2H_2O$, 1.1 g $FeSO_4{\cdot}7H_2O$, 190 mg $CoCl_2.6H_2O$, 42 mg $ZnCl_2$, 24 mg $NiCl_2.6H_2O$, 18 mg $Na_2MoO_4{\cdot}2H_2O$, 300 mg $H_3BO_4$, 2 mg $CuCl{\cdot}2H_2O$, and 50 mg

$MnCl_2 \cdot 4H_2O$. Vitamin solution: 4 mg 4-aminobenzoic acid, 1 mg D(+)biotin, 10 mg nicotinic acid, 5 mg Ca-(+)pantothenate, 10 mg pyridoxamine dihydrochloride, 10 mg thiamine chloride, and 50 mg riboflavin.

*Rhodobacter sphaeroides* was grown chemoheterotrophically at 30°C in LB (0.1% tryptone, 0.05% yeast extract, 0.05% NaCl) or RLB rich media (Jun, Saer, Madden, & Beatty, 2014) (0.1% tryptone, 0.05% yeast extract, 0.05% NaCl, 810 µM $MgCl_2$ and 510 µM $CaCl_2$). *Rhodobacter sphaeroides* was grown photoheterotrophically at 30°C on RCVBN minimal medium (Beatty & Gest, 1981; Wall, Weaver, & Gest, 1975; Weaver, Wall, & Gest, 1975) (for 1L, 10 mL 10% $(NH_4)_2SO_4$, 40 mL 10% DL malic acid, pH 6.8, 50 mL super salts solution, 0.1 mL 150 ug/mL biotin, 0.1 mL 10 mg/mL niacin, 15 mL 0.64M $KPO_4$ PH 6.8, and 884.8 mL $ddH_2O$). Super salts solution: for 1L, 40 mL 1% EDTA, 20 mL 20% $MgSO_4 \cdot 7H_2O$, 20 mL 7.5% $CaCl_2 \cdot 2H_2O$, 20 mL trace elements solution, 48 mL 0.5% $FeSO_4 \cdot 7H_2O$, 20 mL 0.1% thiamine-HCl and 832 mL $ddH_2O$. Trace elements solution: for 1L, 0.795 g $MnSO_4 \cdot H_2O$, 1.4 g $H_3BO_3$, 0.020 g $Cu(NO_3)_2 \cdot 3H_2O$, 0.120 g $ZnSO_4 \cdot 7H_2O$, 0.375 g $Na_2MoO_4 \cdot 2H_2O$, and 500 mL $ddH_2O$.

Yeast strains were grown at 30°C on rich YPAD medium (1% yeast extract, 2% peptone, 2% dextrose, and 0.004% adenine), or SC (0.5% $(NH_4)_2SO_4$, 0.17% yeast nitrogen base without amino acids, 0.17% SC-His-Ura-Leu-Trp mixture, 0.002% L-tryptophan, 0.002% L-histidine, and 0.002% uracil or 0.01% L-leucine), with either fermentable or non-fermentable carbon sources (2% dextrose, 2% lactate+0.05% dextrose, 3% glycerol+0.05% dextrose, 2% ethanol+0.05% dextrose, 3% glycerol+2% ethanol+0.05% dextrose). The pH was adjusted to 5.8 with concentrated NaOH. For selection of yeast strains containing the KanMX4 cassette, G418 (200 µg/mL) was added to the medium. For selection of yeast strains with *URA*3-marked plasmids (i.e., pCM189), L- uracil was omitted from the medium, and for *LEU*2-marked plasmids (i.e., pYX242), L-leucine was omitted. All media were solidified with 2% agar.

**Table 3.1.** List of strains used in this study.

| Strain | Genotype | Source |
|---|---|---|
| *Escherichia coli* TOP10 | F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str$^R$) *end*A1 *nup*G | Laboratory stock. |
| *Escherichia coli* S17-1-λpir-RP4 | TpR SmR *rec*A, *thi*, *pro*, *hsd*R-M+RP4: 2-Tc:Mu: Km Tn7 λpir. | J. Thomas Beatty (UBC). |
| *Escherichia coli* DH5α | F⁻ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17($r_k^-$, $m_k^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ⁻ | Laboratory stock. |
| *Escherichia coli* DH5α-λpir | sup E44, Δ*lac*U169 (Φ*lac*ZΔM15), *rec*A1, *end*A1, *hsd*R17, thi-1, *gyr*A96, *rel*A1, λpir phage lysogen. | J. Thomas Beatty (University of British Columbia). |
| *Escherichia coli* BW29427-λpir-RP4 | RP4-2(TetS, kan1360::FRT), thrB1004, lacZ58(de).(M15), dapA1341::[erm pir+], rpsL(strR),thi-, hsdS-, pro- | K. A. Datsenko and B. L. Wanner, unpublished data. |
| *Rhodobacter sphaeroides* 2.4.1 | Wild type. | J. Thomas Beatty (University of British Columbia). Originally isolated by W. R. Sistrom, University of Oregon, Portland. |
| *Rhodobacter sphaeroides* Δ*mic60* | ΔRSP6207 | This thesis. |
| *Rhodobacter sphaeroides* Δ*orf52* | ΔRSP1508 | This thesis. |
| *Rhodopseudomonas palustris* TIE-1 | Wild type. Isolate from Woods Hole, MA, USA. | Jiao et al., (2005) |
| *Rhodopseudomonas palustris* TIE-1 Δ*mic60* | ΔRS01285 | This thesis |
| *Rhodopseudomonas palustris* TIE-1 Δ*orf52* | ΔRS01290 | This thesis. |
| *Saccharomyces cerevisiae* BY4741a *mic60*Δ | *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mic60*Δ::KANMX4 | Yeast Viable Haploid Gene Collection. |
| *Saccharomyces cerevisiae* W303/1a | *MAT*a *leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 [cir⁺]* | Melanie Dobson. |
| *Saccharomyces cerevisiae* W303/1a *mic60*Δ | *MAT*a *leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 [cir⁺] mic60*Δ::KANMX4 | This thesis. |

### 3.2.2. DNA methods and plasmid construction

Total genomic DNA (gDNA) was extracted from *R. sphaeroides*, *R. palustris* and *S. cerevisiae* strains using either the ZR Bacterial/Fungal DNA Miniprep Kit (Zymo Research) with a BIO101/Savant FastPrep FP120 high-speed bead beater and a 30 min incubation at 60°C with 20 uL of proteinase K (20 mg/mL), or the Epicentre MasterPure DNA Purification Kit (Epicentre Biotechnologies). To make yeast plasmids containing the *AgMIC60* construct, a cDNA library of the jakobid *Andalucia godoyi* was used as template in PCR assays. All plasmid constructs were built with the Gibson Assembly Master Mix (New England Biolabs) or the NEB Builder HiFi DNA Assembly (New England Biolabs). To make the empty pCM189 and pYX242 plasmids that served as negative controls, the *ScMIC60* gene was removed by digesting with the restriction endonucleases BamHI and and NotI, and the digested plasmid was then religated using the CloneJET PCR Cloning Kit (Thermo Scientific). Plasmids were isolated from their bacterial hosts using the AxyPrep Plasmid Miniprep Kit (Axygen). All plasmid constructs were confirmed by Sanger sequencing using both forward and reverse primers. Table 3.2 lists all plasmids built for this thesis. All plasmids and their sequences are available from the author upon request. Table 3.3 lists all primers used to build the plasmids in Table 3.2.

**Table 3.2.** List of plasmids used in this study.

| Plasmid | Description | Source |
|---|---|---|
| pZDJ | Suicide plasmid used to knock-out genes in *R. sphaeroides*. | Brimacombe et al., (2013) |
| pZDJ-ΔRSP6207 | Suicide plasmid used to the *mic60* gene in *R. sphaeroides*. | This thesis. |
| pZDJ-ΔRSP1508 | Suicide plasmid used to knock-out the *orf52* gene in *R. sphaeroides*. | This thesis. |
| PJQ200SK | Mobilizable suicide vector; *sacB* $Gm^r$ | Quandt & Hynes, (1993) |
| PJQ200SK-ΔRS01285 | Suicide plasmid used to the *mic60* gene in *R. palustris*. | This thesis. |
| PJQ200SK-ΔRS01290 | Suicide plasmid used to knock-out the *orf52* gene in *R. palustris*. | This thesis. |
| pCM189 | Doxycycline-regulatable expression vector used as a negative control in *S. cerevisiae*. Single copy number plasmid (CEN, YCp). *URA3*. | Garí, Piedrafita, Aldea, & Herrero, (1997) |
| pCM189-trunc*MIC60* | Regulatable expression vector used to express trMic60 in *S. cerevisiae*. | This thesis. |
| pCM189-Rs*MIC60* | Regulatable expression vector used to express RsMic60 in *S. cerevisiae*. | This thesis. |
| pCM189-ScRs*MIC60* | Regulatable expression vector used to express ScRsMic60 in *S. cerevisiae*. | This thesis. |
| pCM189-Ag*MIC60* | Regulatable expression vector used to express AgMic60 in *S. cerevisiae*. | This thesis. |
| pCM189-Sc*MIC60* | Regulatable expression vector used to express ScMic60 in *S. cerevisiae*. | This thesis. |
| pYX242 | TPI constitutive expression vector used as a negative control in *S. cerevisiae*. Multicopy (2 μ) plasmid (episomal, YEp). *LEU*2. | Novagen. |
| pYX242-trunc*MIC60* | Constitutive expression vector used to express trMic60 in *S. cerevisiae*. | This thesis. |

| | | |
|---|---|---|
| pYX242-*RsMIC60* | Constitutive expression vector used to express RsMic60 in *S. cerevisiae*. | This thesis. |
| pYX242-*ScRsMIC60* | Constitutive expression vector used to express ScRsMic60 in *S. cerevisiae*. | This thesis. |
| pYX242-*AgMIC60* | Constitutive expression vector used to express AgMic60 in *S. cerevisiae*. | This thesis. |
| pYX242-*ScMIC60* | Constitutive expression vector used to express ScMic60 in *S. cerevisiae*. | This thesis. |

**Table 3.3.** List of primers and their sequences used in this study.

| Plasmid | Primer | Sequence (5'-3') | PCR template |
|---|---|---|---|
| pZDJ-ΔRSP6207 | 5'-MIT_T_F | TGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAAAACCGGGGTCGAGGGCTA | *R. sphaeroides* 2.4.1 WT gDNA |
| | 5'-MIT_R | CGCCGTGCCTTCCAGTTTTT | *R. sphaeroides* 2.4.1 WT gDNA |
| | kanFRT_T_F | CGGTCACAAAACACCTGAAAAACTGGAAGGCACGGCGATGATTCCGGGGATCCGTCGACC | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | kanFRT_T_R | CCAAAGCATGGAACCGGCCTCACTTCCCTTCCGCGGCTGCTGTAGGCTGGAGCTGCTTC | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | 3'-MIT_F | GGCCGGTTCCATGCTTTG | *R. sphaeroides* 2.4.1 WT gDNA |
| | 3'-MIT_T_R | AACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCGAGTGCTTGGGGTTGATCTC | *R. sphaeroides* 2.4.1 WT gDNA |
| pZDJ-ΔRSP1508 | 5'hemY_F | TGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGACGCGCTGATGGCCGAGAT | *R. sphaeroides* 2.4.1 WT gDNA |
| | 5'hemY_R | AATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAATCATGGAACCGGCCTCACT | *R. sphaeroides* 2.4.1 WT gDNA |
| | kanFRT_F | ATTCCGGGGATCCGTCGACC | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | kanFRT_R | TGTAGGCTGGAGCTGCTTC | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | 3'hemY_F | TCTAGAAAGTATAGGAACTTCGAAGCAGCTCCAGCCTACAGTTGAGCCCAGCACAAAATAGG | *R. sphaeroides* 2.4.1 WT gDNA |
| | 3'hemY_R | AACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCCCGAAATGATCCCCGACAGA | *R. sphaeroides* 2.4.1 WT gDNA |
| PJQ200SK-ΔRS01285 | KO-RpMic60-1 | AGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAGCTCAGGATCAAGATCGG | *R. palustris* TIE-1 WT gDNA |
| | KO-RpMic60-2 | GGTCGACGGATCCCCGGAATCATCCTTGGATTTTCCTCG | *R. palustris* TIE-1 WT gDNA |
| | KO-RpMic60-3 | ACGAGGAAAATCCAAGGATGATTCCGGGGATCCGTCGA | pKD13 plasmid DNA or KanFRT cassette PCR product |

| | | | |
|---|---|---|---|
| | KO-RpMic60-4 | TATGGCGACGGTTTTTGCAGTGTAGGCTGGAGCTGCTTC | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | KO-RpMic60-5 | CGAAGCAGCTCCAGCCTACACTGCAAAAACCGTCGCCA | *R. palustris* TIE-1 WT gDNA |
| | KO-RpMic60-6 | AGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAG GACTTGGTGTCCTTGCGG | *R. palustris* TIE-1 WT gDNA |
| PJQ200SK-ΔRS01290 | KO-RpHemY-1 | AGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGC TCTAGATTCCAGGCCTTGCTGAAG3' | *R. palustris* TIE-1 WT gDNA |
| | KO-RpHemY-2 | GGTCGACGGATCCCCGGAATCATGGGCAGACCTATGGC | *R. palustris* TIE-1 WT gDNA |
| | KO-RpHemY-3 | TCGCCATAGGTCTGCCCATGATTCCGGGGATCCGTCGA | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | KO-RpHemY-4 | CAGCGCGGCGGCCGGTAACCTGTAGGCTGGAGCTGCTTC | pKD13 plasmid DNA or KanFRT cassette PCR product |
| | KO-RpHemY-5 | CGAAGCAGCTCCAGCCTACAGGTTACCGGCCGCCGCGC | *R. palustris* TIE-1 WT gDNA |
| | KO-RpHemY-6 | AGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAG TTCTATTTGGCCCCCTGCCCCC | *R. palustris* TIE-1 WT gDNA |
| *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* | OUT2_F | CTACTTGACGCAGAGGCAGA | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | OUT2_R | GCTGGAGAGCCTTGCCTTTT | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | INT1_F | TTCGAAGCAGCTCCAGCCTAC | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | INT1_R | TGGAGAGCCTTGCCTTTTCC | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | INT2_F | TCGATGCCGAGAAAGGTGAG | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | INT2_R | TGTAGGCTGGAGCTGCTTCG | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | RSP6207_F | TGAGAAGAAACGGGACGGTG | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | RSP6207_R | ATTGCCTCAATGCGTTTGCG | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | RSP1508_F | GCTCTCCGCCAAGATGAAGT | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |

| | | | |
|---|---|---|---|
| | RSP1508_R | GGCTCAACCGCTTTTTCGTC | *R. sphaeroides* WT, Δ*mic60*, and Δ*orf52* gDNA |
| *R. palustris* WT, Δ*mic60*, and Δ*orf52* | KOX_F1 | CGAAAAAGCGCTGCTCGAAG | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOX_R1 | GCGCGATGATCACGAGAAAC | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOX_R2 | GAACCTGCGTGCAATCCATC | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOX_F3 | TCGCCTTCTTGACGAGTTCT | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOX_R3 | CGAGCGACTTGGTGTCCTT | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOY_F1 | GTCCCATCAATTTGCCACCG | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOY_R1 | GCCGATGTGGTTTTGTCGAA | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| | KOY_R2 | GAACCTGCGTGCAATCCATC | *R. palustris* WT, Δ*mic60*, and Δ*orf52* gDNA |
| *S. cerevisiae* W303/1a *mic60*Δ | A | TATAGTACGTAGTCGGCAAAGAACC | *S. cerevisiae* W303/1a *mic60*Δ gDNA |
| | kanB | CTGCAGCGAGGAGCCGTAAT | *S. cerevisiae* W303/1a *mic60*Δ gDNA |
| | kanC | TGATTTTGATGACGAGCGTAAT | *S. cerevisiae* W303/1a *mic60*Δ gDNA |
| | D | TATTTGAACCTGTATTTGTCGTTGA | *S. cerevisiae* W303/1a *mic60*Δ gDNA |
| pCM189-*RsMIC60* | Preseq_tail_F | AAATACACACACTAAATTACCGGATCAATTCGGGGGATCCATGATGCTAAGAACTACTGCCT | *S. cerevisiae* S288C WT gDNA |
| | Preseq_tail_R | GTTTCTTCTCACCAGCTTCGCGGGACTCCGGTTCTGACATTAGACCCCTCCTTAGCACAA | *S. cerevisiae* S288C WT gDNA |
| | RSP_6207_F | ATGTCAGAACCGGAGTCCCG | *R. sphaeroides* 2.4.1 WT gDNA |

| | | | |
|---|---|---|---|
| | RSP_6207_tail_R | ATTACATGATGCGGCCCTCCTGCAGGGCCCTAGCGGCCGCTCACTTCCCTTCCGCGGC | *R. sphaeroides* 2.4.1 WT gDNA |
| pCM189-*ScRsMIC60* | N-yeast_tail_F | AAATACACACACTAAATTACCGGATCAATTCGGGGGATCCATGATGCTAAGAACTACTGCC | *S. cerevisiae* S288C WT gDNA |
| | N-yeast_R | TTTGTCTTTGATAATTTTATTAAATTCT | *S. cerevisiae* S288C WT gDNA |
| | C-rhodo_tail_F | CACTCAAGTCAAAGAATTTAATAAAATTATCAAAGACAAAATCGCCGAAGCCTCCGG | *R. sphaeroides* 2.4.1 WT gDNA |
| | C-rhodo_tail_R | ATTACATGATGCGGCCCTCCTGCAGGGCCCTAGCGGCCGCTCACTTCCCTTCCGCGG | *R. sphaeroides* 2.4.1 WT gDNA |
| pCM189-*AgMIC60* | pCM189-AgMic60_F | AAATACACACACTAAATTACCGGATCAATTCGGGGGATCCATGTACGGAAACAGGCTTGCC | *A. godoyi* cDNA |
| | pCM189-AgMic60_R | ATTACATGATGCGGCCCTCCTGCAGGGCCCTAGCGGCCGCTTAAACGTCGGATTCCTCGAA | *A. godoyi* cDNA |
| pCM189-trunc*MIC60* | Preseq_tail_F | AAATACACACACTAAATTACCGGATCAATTCGGGGGATCCATGATGCTAAGAACTACTGCCT | *S. cerevisiae* S288C WT gDNA |
| | yeast_Mic60_trunc_R | AATTACATGATGCGGCCCTCCTGCAGGGCCCTAGCGGCCGCTTATTTGTCTTTGATAATTTTATTAAATTCT | *S. cerevisiae* S288C WT gDNA |
| pYX242-*RsMIC60* | pXYRsMic60_F | AACTACAAAAAACACATACAGGAATTCACCATGGCAATGCTAAGAACTACTGCCTCACG | pCM189-*RsMIC60* plasmid DNA |
| | pXYRsMic60_R | TCCTTATTCAGTTAGCTAGCTGAGCTCGAGTCACTTCCCTTCCGCGGC | pCM189-*RsMIC60* plasmid DNA |
| pYX242-*ScRsMIC60* | pXYScRsMic60_F | AACTACAAAAAACACATACAGGAATTCACCATGGCAATGCTAAGAACTACTGCCTCACG | pCM189-*ScRsMIC60* plasmid DNA |
| | pXYScRsMic60_R | TCCTTATTCAGTTAGCTAGCTGAGCTCGAGTCACTTCCCTTCCGCGGC | pCM189-*ScRsMIC60* plasmid DNA |
| pYX242-*AgMIC60* | pXYAgMic60_F | AACTACAAAAAACACATACAGGAATTCACCATGGCAATGTACGGAAACAGGCTTGCC | pCM189-*AgMIC60* plasmid DNA |
| | pXYAgMic60_R | TCCTTATTCAGTTAGCTAGCTGAGCTCGAGTTAAACGTCGGATTCCTCGAATG | pCM189-*AgMIC60* plasmid DNA |
| pYX242-trunc*MIC60* | pXYScMic60Trunc_F | AACTACAAAAAACACATACAGGAATTCACCATGGCAATGCTAAGAACTACTGCCTCACG | pCM189-trunc*MIC60* plasmid DNA |

| | | | |
|---|---|---|---|
| | pXYScMic60Trunc_R | TCCTTATTCAGTTAGCTAGCTGAGCTCGAGTTATTTGTCTTTGATAATTTTATTAAATTCT | pCM189-trunc*MIC60* plasmid DNA |

### 3.2.3. Construction of *R. sphaeroides* knock-out mutants

In order to knock-out genes in *R. sphaeroides*, a knock-out genetic construct was assembled into the suicide plasmid vector pZDJ (Brimacombe et al., 2013) with either the Gibson Assembly Cloning Kit (New England Biolabs) or the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). A suicide vector cannot replicate in its final host, i.e., *R. sphaeroides*, and therefore has to integrate into the host genome for the host bacterium to survive antibiotic selection (Jaschke, Saer, Noll, & Beatty, 2011). The antibiotic resistance cassette flanked by FLP recognition target (FRT) sites used to interrupt the genes to be knocked out corresponds to that used in the Keio collection (Baba et al., 2006). The resulting suicide vector with the knock-out construct was cloned into *E. coli* S17-1 λ-pir-RP4 which can replicate the suicide plasmid. *E. coli* S17-1 λ-pir was then conjugated with *R. sphaeroides* 2.4.1. Briefly, donor and recipient cells grown to stationary phase were mixed in a 1:2 volume ratio and pelleted by centrifugation for 1 min at 4,500-6,000 g, and then washed twice with antibiotic-free minimal medium (e.g., RCVBN). After the last wash, the pellet was resuspended in 50 uL of RCVBN, and 10 uL aliquots were spotted onto antibiotic-free RCVBN solid medium. The plates were incubated at 30° C overnight to allow conjugation to take place. Afterwards, an emulsion was made from the several inoculation spots and streaked onto antibiotic-containing RCVBN+Gm solid medium. *R. sphaeroides* exconjugants were then successively streaked onto new LB+Gm solid medium until no *E. coli* S17-1 λ-pir-RP4 contamination remained. The resulting pure *R. sphaeroides* exconjugants contained the pZDJ plasmid, with the knock-out gene construct, integrated in the chromosome by a first crossing over (recombination) event. In order to induce a second crossing over to excise the pZDJ suicide plasmid from the host chromosome, colonies were picked and grown on liquid LB with no antibiotic selection until late stationary phase (about two/three days). The cultures were then streaked on LB+10% sucrose solid medium which allows for counter-selecting of those colonies that have lost the integrated pZDJ plasmid. The counter-selection relies on the *sacB* gene carried by the suicide plasmid (Jaschke et al., 2011). In order to induce a third crossing over between the FRT sites of the kanamycin cassette, cells were grown in liquid LB without any antibiotic selection until late stationary phase and then streaked on LB solid medium. Resultant colonies were then screened through PCR assays.

### 3.2.4. Construction of *R. palustris* knock-out mutants

In order to make knock-out strains for *R. palustris*, the same general protocol used for *R. sphaeroides* was followed (Welander et al., 2012). The suicide plasmid used for *R. palustris* was pJQ200SK (Quandt & Hynes, 1993) and the host strain was *E. coli* BW29427-λpir-RP4 which requires diaminopimelic acid (DAP; 300 μM) for growth. This auxotrophy allows to easily remove the plasmid-donor bacterium from the medium after conjugation.

### 3.2.5. Construction of *S. cerevisiae* strains

Genomic DNA from *S. cerevisiae* strain BY4741a with a knock-out of the *MIC60* gene (i.e., the *mic60*Δ::*KANMX4* cassette) was used as a template to amplify through PCR the *mic60*::KanMX4 cassette with flaking genomic sequences. Primers A and D were used (see Table 3.3). The resulting PCR product was later used to transform *S. cerevisiae* strain W303/1a to G418-resistance. A standard lithium-acetate transformation protocol (Gietz & Woods, 2006) was used to introduce the gene deletion *mic60*::KanMX4 cassette, and also to transform yeast with a series of autonomously replicating plasmid constructs (i.e., the pYX242 and pCM189 plasmid series; see Table 3.2). The gene deletion was confirmed on selected transformants through PCR using primers A, D, KanB and kanC (Table 3.3).

### 3.2.6. Growth curve analyses of *R. sphaeroides* and *R. palustris*

Growth curves were performed for *R. sphaeroides* and *R. palustris* under photosynthetic conditions (i.e., anoxia and light). Three culture tubes (readable on the spectrophotometer) were inoculated with single colonies of WT, Δ*mic60* and Δ*orf52* strains from fresh plates, i.e., three-day and five-day old plates for *R. sphaeroides* and *R. palustris* respectively. These culture tubes were filled with medium to the top, closed with a rubber stopper, and carefully sealed with parafilm so no air would enter the tube. Cultures were grown to 0.8 $OD_{660}$ and then used to inoculate the experiment tubes to 0.1 $OD_{660}$. After inoculation, experiment tubes (i.e., three biological replicates) were first incubated in the dark at 30°C for about 12 h, and then transferred to light. Growth in the experiment tubes was monitored under low light at (45 μmol $s^{-1}$ $m^{-2}$; 9 W $m^{-2}$) and low light intensity (115 μmol $s^{-1}$ $m^{-2}$; 24 W $m^{-2}$) for several days. $OD_{660}$ measurements were taken at regular intervals and tubes were inverted to promote growth.

### 3.2.7. OD spectra

After growth curves for both *R. sphaeroides* and *R. palustris* were completed, cultures were transferred to 96-well plates that could be read by the BioTek Epoch microplate spectrophotometer. Absorbance spectra of whole cells in media were then taken from wavelengths of 300 to 990 nm in steps of 5 nm.

### 3.2.8. Transmission electron microscopy

Cells of *R. sphaeroides* WT, Δ*mic60* and Δ*orf52* strains were grown chemoheterotrophically (oxia and darkness) and photoheterotrophically (anoxia and light) to early stationary phase and fixed using the Karnovsky's fixative (Karnovsky, 1965). Cultures (5 mL) were harvested by centrifugation at 1,000 g for 10 min and at 4°C, the supernatant was discarded, and the pellet was fully resuspended in 2 mL of the Karnovsky's fixative (pH 6.8; Karnovsky, 1965) by vortexing. Specimen preparation, sectioning, and staining were carried out by Mary Anne Trevors at the Electron Microscopy Facility of the Faculty of Medicine, Dalhousie University. Cells in fixative were incubated at room temperature for 30 min and collected by centrifugation, then resuspended in 10% sucrose in 0.1 M cacodylate buffer. Cells were then collected by centrifugation and the resuspension in sucrose was repeated three times. The pellet was then resuspended in 1 mL of 2% osmium tetroxide. The final solution was prepared for imaging on a copper mesh grid. Cells were imaged by TEM with electron accelerating voltage of 80 kV. The sections were viewed using a JEOL JEM 1230 transmission electron microscope at 80 kV, and images were captured using a Hamamatsu ORCA-HR digital camera.

### 3.2.9. Yeast spot dilution assay

Yeast strains, inoculated from a single colony, were grown overnight on 1 mL of appropriate liquid medium (SC-Ura+2% dextrose or SC-Leu+2% dextrose) at 30°C and 220 rpm. Cells were collected by centrifugation at 4,000 rpm for 2 min and the supernatant was carefully removed. Cells were then washed twice with 800 µL dd$H_2O$. Afterwards, the cells were brought to the same $OD_{600}$ and ten-fold serial dilutions were prepared in a 96-well plate. $OD_{600}$ measurements were done in a Spectronic 20 spectrophotometer. After cells were diluted in series, a pinner was used to transfer them, or spot, on the appropriate solid media. Plates were incubated at 30°C. Doxycycline was

added to the plates to a final concentration of 2 μg/mL to repress expression from the pCM189 plasmid.

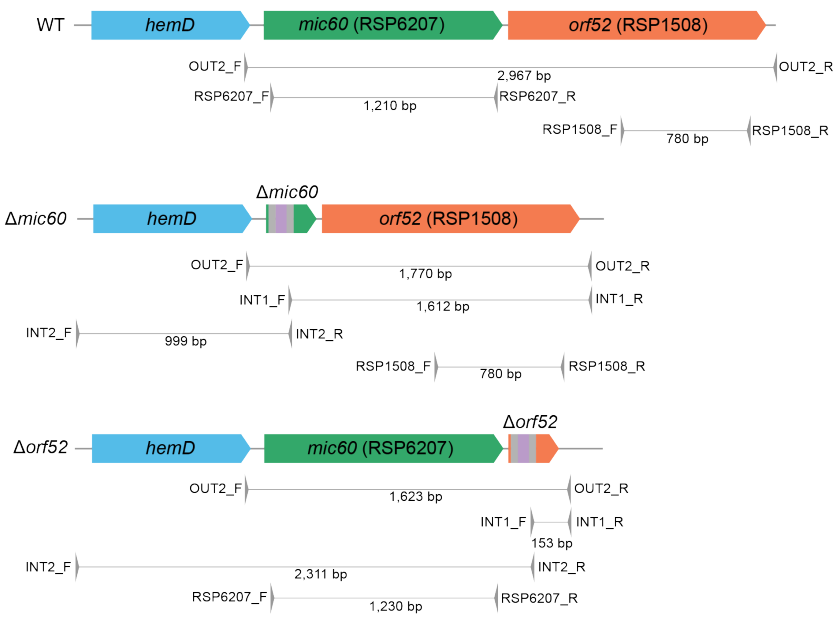
## 3.3. PRELIMINARY RESULTS

### 3.3.1. Construction of Δ*mic60 and* Δ*orf52* knock-out mutant strains for *R. sphaeroides* and *R. palustris*

Disrupting the *mic60* gene and its immediately downstream neighbor gene *orf52* from purple nonsulfur bacteria represents a straightforward approach to study the function of these genes in ICM-developing alphaproteobacteria. However, knocking out genes in *R. sphaeroides* and *R. palustris* is not as simple as in *E. coli* (see Materials and Methods above; Jaschke et al., 2011). These purple nonsulfur bacteria cannot be transformed and need to be conjugated with an *E. coli* strain to introduce a suicide plasmid containing the null allele that will eventually recombine with the host genome to replace the endogenous gene copy. To confirm that the *mic60* and *orf52* genes had been successfully disrupted in both *R. sphaeroides* and *R. palustris*, I tested these strains through multiple PCR assays using diverse combinations of external and internal primers covering the loci of interest (Fig. 3.2 and 3.3). For both *R. spaheroides* and *R. palustris* six primer pairs were tested. In the case of *R. sphaeroides*, the resultant Δ*mic60* and Δ*orf52* knock-out mutants only retained small scar peptides in place of the disrupted genes. In the case of *R. palustris*, on the other hand, the resultant knock-out mutants contained kanamycin-resistance gene cassettes in place of the disrupted genes. The PCR confirmation experiments all agreed with the theoretical expectations (Fig. 3.2A and 3.3A) and with each other (Fig. 3.2B and 3.2B). These analyses confirmed that Δ*mic60* and Δ*orf52* strains were successfully created for both *R. sphaeroides* and *R. palustris*.

**Figure 3.2.** PCR assays confirm the disruption of the *mic60* and *orf52* genes in *R. sphaeroides*. **A.** Genomic contexts of the loci in question for the WT, Δ*mic60* and Δ*orf52* strains. Primer pairs and the sizes expected for the amplified products are shown below each schematic gene arrangement. **B.** Agarose gel electrophosis assays of PCR-amplified products using internal and external primers to confirm the Δ*mic60* and Δ*orf52* strains.

A.


WT
hemD
mic60 (RSP6207)
orf52 (RSP1508)
OUT2_F
2,967 bp
OUT2_R
RSP6207_F
1,210 bp
RSP6207_R
RSP1508_F
780 bp
RSP1508_R
Δmic60
Δmic60
hemD
orf52 (RSP1508)
OUT2_F
1,770 bp
OUT2_R
INT1_F
1,612 bp
INT1_R
INT2_F
999 bp
INT2_R
RSP1508_F
780 bp
RSP1508_R
Δorf52
Δorf52
hemD
mic60 (RSP6207)
OUT2_F
1,623 bp
OUT2_R
INT1_F
153 bp
INT1_R
INT2_F
2,311 bp
INT2_R
RSP6207_F
1,230 bp
RSP6207_R

B.

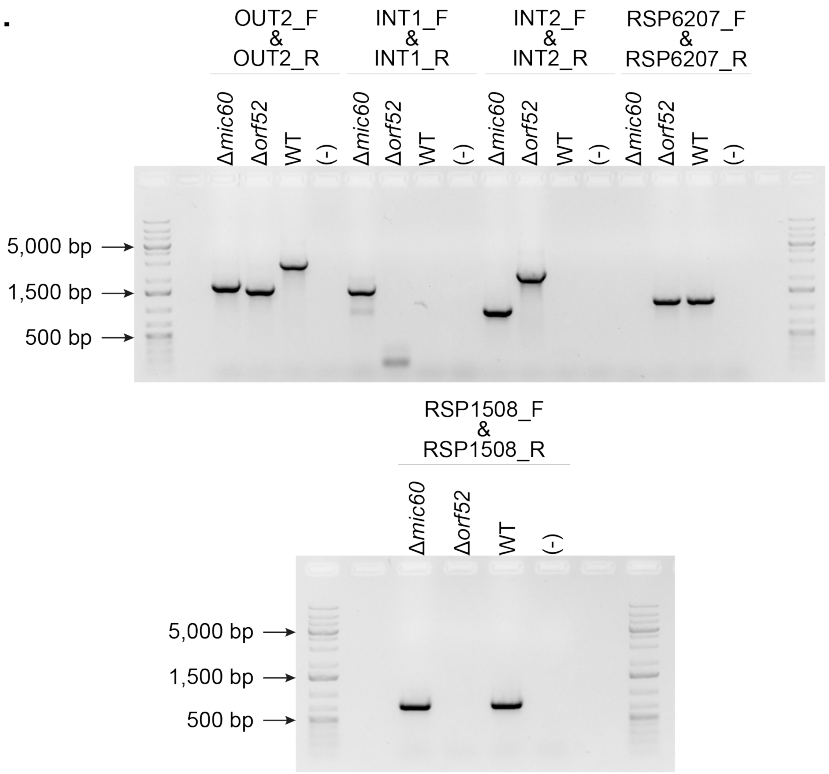
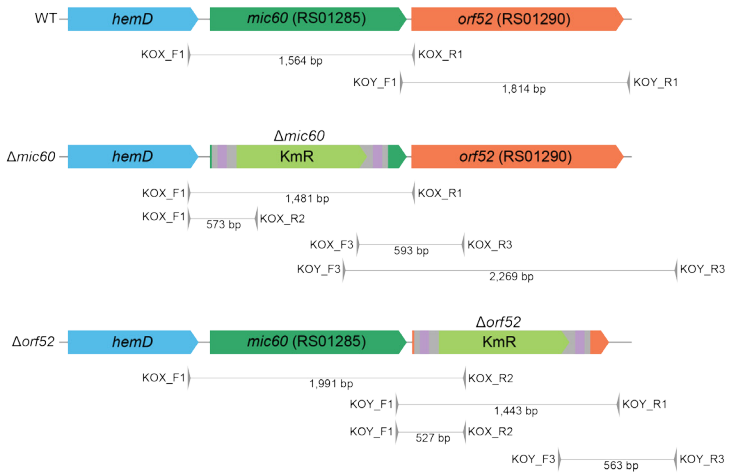
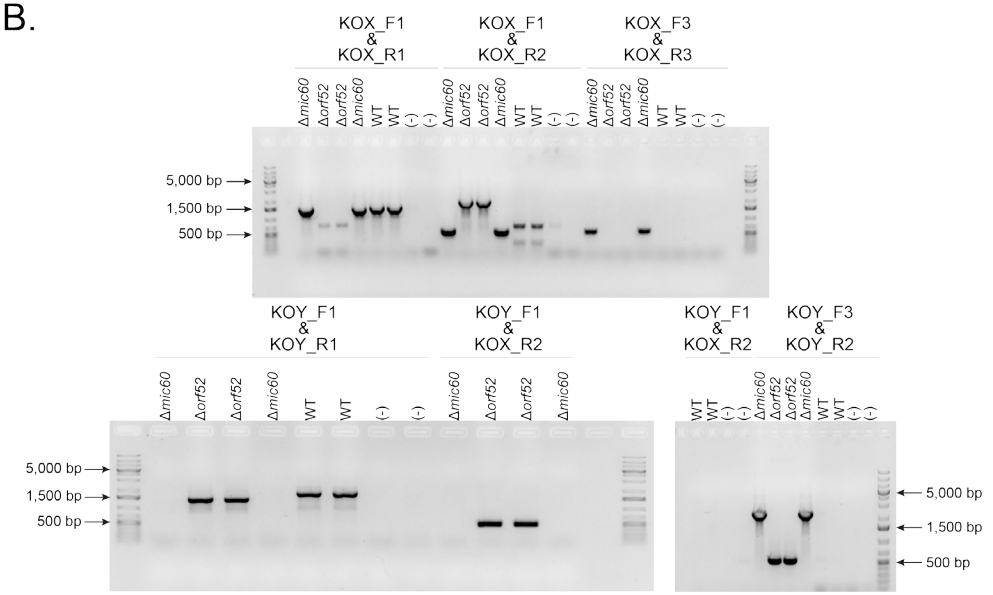

OUT2_F & OUT2_R
INT1_F & INT1_R
INT2_F & INT2_R
RSP6207_F & RSP6207_R
Δmic60
Δorf52
WT
(-)
5,000 bp
1,500 bp
500 bp
RSP1508_F & RSP1508_R
Δmic60
Δorf52
WT
(-)
5,000 bp
1,500 bp
500 bp

**Figure 3.3.** PCR assays confirm the disruption of the *mic60* and *orf52* genes in *R. palustris*. **A.** Genomic contexts of the loci in question for the WT, Δ*mic60* and Δ*orf52* strains. Primer pairs and the sizes expected for the amplified products are shown below each schematic gene arrangement. **B.** Agarose gel electrophosis assays of PCR-amplified products using internal and external primers to confirm the Δ*mic60* and Δ*orf52* strains.

A.


WT
hemD
mic60 (RS01285)
orf52 (RS01290)
KOX_F1
1,564 bp
KOX_R1
KOY_F1
1,814 bp
KOY_R1
Δmic60
Δmic60
hemD
KmR
orf52 (RS01290)
KOX_F1
1,481 bp
KOX_R1
KOX_F1
573 bp
KOX_R2
KOX_F3
593 bp
KOX_R3
KOY_F3
2,269 bp
KOY_R3
Δorf52
Δorf52
hemD
mic60 (RS01285)
KmR
KOX_F1
1,991 bp
KOX_R2
KOY_F1
1,443 bp
KOY_R1
KOY_F1
527 bp
KOX_R2
KOY_F3
563 bp
KOY_R3

B.


KOX_F1 & KOX_R1
KOX_F1 & KOX_R2
KOX_F3 & KOX_R3
Δmic60
Δorf52
Δorf52
Δmic60
WT
WT
(-)
(-)
5,000 bp
1,500 bp
500 bp
KOY_F1 & KOY_R1
KOY_F1 & KOY_R2
KOY_F1 & KOX_R2
KOY_F3 & KOY_R2
Δmic60
Δorf52
Δorf52
Δmic60
WT
WT
(-)
(-)
5,000 bp
1,500 bp
500 bp

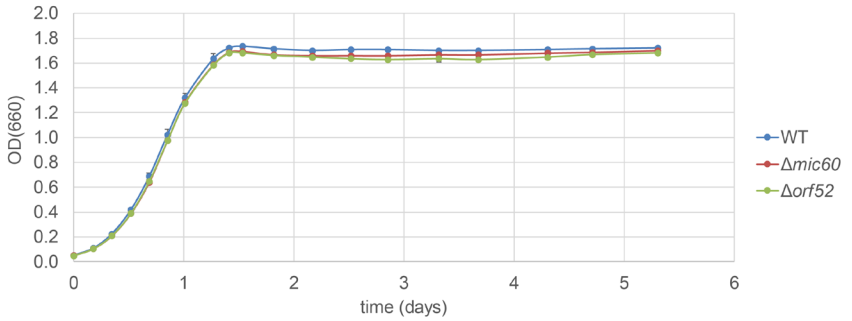
### 3.3.2. Decreasing light intensity increases growth differences between WT and knock-out mutants

As a first step to assess phenotypic differences between the WT and knock-out mutant strains, photosynthetic growth (i.e., under anoxia and light) in minimal medium was followed until late stationary phase. Under photosynthetic growth conditions, purple bacteria develop extensive ICM systems to house the photosynthetic complexes and capture increased amounts of light (Drews, 1992). Minimal medium, as opposed to rich medium, was used to exacerbate potential phenotypic differences between the strains (Beatty & Gest, 1981; Ehrenreich & Widdel, 1994). Two different light regimes were assessed: (1) high light intensity (115 µmol $s^{-1}$ $m^{-2}$ ≅ 24 W $m^{-2}$), and (2) low light intensity (45 µmol $s^{-1}$ $m^{-2}$ ≅ 9 W $m^{-2}$).

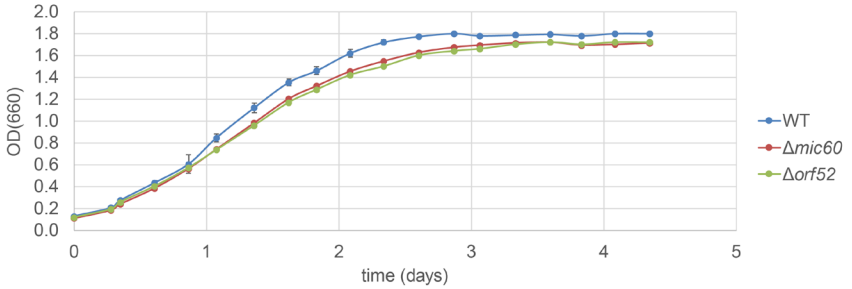
*R. sphaeroides* knock-out mutants, Δ*mic60* and Δ*orf52*, showed little differences in growth relative to the WT strain under the high-light regime (Fig. 3.4A). All three strains had relatively short exponential phases and reach stationary phase ($OD_{660}$ ≅ 1.700) in about 1.5 days. At late exponential phase the WT strain reached higher $OD_{660}$ values than the mutants, and these later remain relatively constant throughout all of the stationary phase. These differences between the WT and the knock-out mutant strains, although small, are statistically significant (see Fig. 3.4 and its short error bars). It is also interesting to note that the Δ*mic60* strain reached slightly higher $OD_{660}$ values than the Δ*orf52* strain (Fig 3.4A). In contrast, under the low-light regime, the growth difference between the WT and the mutant strains was more pronounced (Fig. 3.4B). All three strains grew slower in the low-light than in the high-light regime. The exponential phase was twice as long, and the stationary phase was only reached after about 3 days of growth (Fig. 3.4B). The WT reaches higher $OD_{660}$ values in the exponential phase which also remained higher during the stationary phase. A very small difference, but still significant, was seen between the Δ*mic60* and the Δ*orf52* strains (Fig. 3.4B).

**Figure 3.4.** Photosynthetic growth curves of *R. sphaeroides* and *R. palustris* WT and Δ*mic60* and Δ*orf52* strains. **A, B.** Photosynthetic growth of *R. sphaeroides* strains in RCVBN medium under high light and under low light, respectively. **C, D.** Photosynthetic growth of *R. palustris* strains in FEM medium under high light and in YPS medium under low light, respectively. Error bars represent one standard deviation calculated based on three biological replicates.

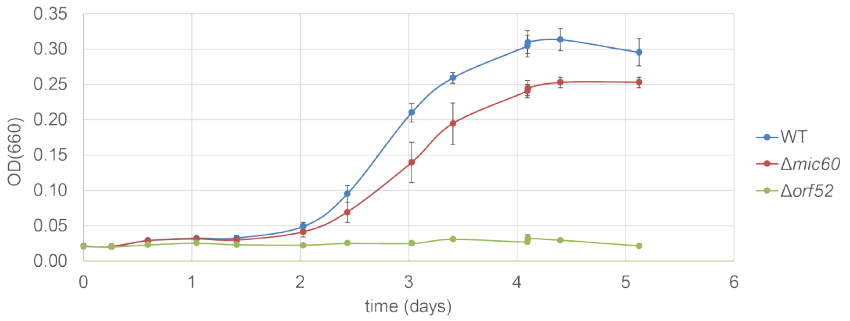
A.


R. sphaeroides growth at high light (RCVBN)
OD(660)
time (days)
WT
Δmic60
Δorf52

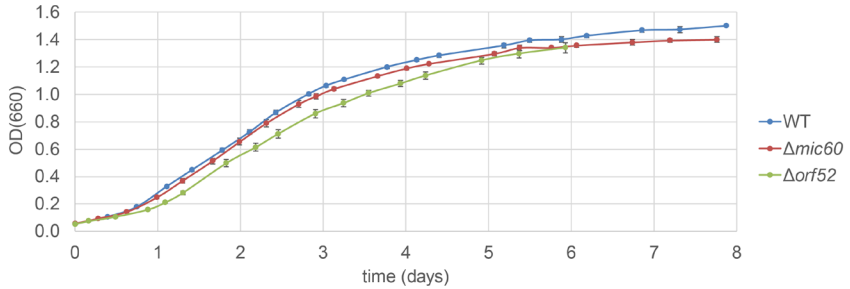
B.


R. sphaeroides growth at low light (RCVBN)
OD(660)
time (days)
WT
Δmic60
Δorf52

C.


R. palustris growth at high light (FEM)
OD(660)
time (days)
WT
Δmic60
Δorf52

D.


R. palustris growth at low light (YPS)
OD(660)
time (days)
WT
Δmic60
Δorf52

*R. palustris* knock-out mutants, Δ*mic60* and Δ*orf52*, showed much clearer growth differences relative to the WT and to those seen among *R. sphaeroides* strains (Fig. 3.4). Under the high-light regime, the Δ*mic60* knock-out mutant grew considerably more slowly relative to the WT, whereas the Δ*orf52* knock-out mutant showed no growth (Fig. 3.4A). Under a low-light regime, the experiment could not be carried out because all three strains showed little or no growth. To circumvent this, the strains were grown in rich-medium YPS when subjected to the low-light regime (Fig. 3.4B). All three strains grew much faster and reached higher $OD_{660}$ values, as expected (Fig. 3.4B). As in the high-light regime, and in *R. sphaeroides* strains, the WT grew faster, followed by Δ*mic60* and then Δ*orf52*. The differences in the growth rates between the three strains, however, became much smaller when grown in the YPS medium but were still statistically significant (Fig. 3.4B).

The growth curves presented here under two different light regimes for two disparate purple bacteria show that even though the differences in growth rates were small, they were accentuated under lower light intensities. Low-light conditions induce the expression of the light-harvesting 2 antenna complexes and also lead to increased amounts of ICMs (Drews & Golecki, 1995). Here is thus a potential connection between the function of *mic60* and *orf52* and the development of ICMs under low light conditions.

### 3.3.3. Optical density (OD) spectra of whole cells do not differ between WT and knock-out mutants

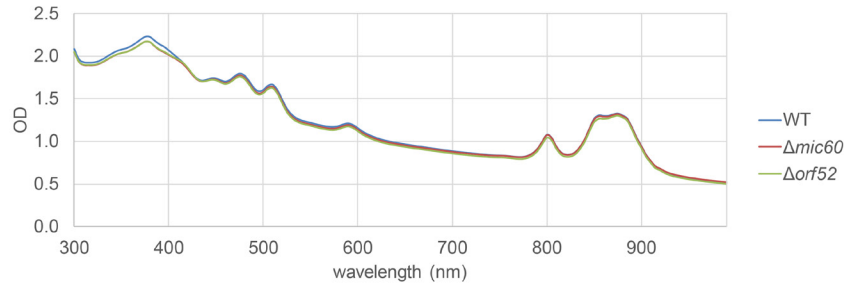
Photosynthetic ICMs house the photosynthetic machinery that is composed of the light-harvesting complexes 1 and 2 (LH1 and LH2), a type II reaction center (RC), a cytochrome $bc_1$, a periplasmic cytochrome $c_2$, and an ATP synthase (Drews, 1992; Drews & Golecki, 1995). The type II reaction center is tightly associated with the light-harvesting complex 1, whereas the light-harvesting complex 2 is found as an independent complex (Drews & Golecki, 1995). These two protein complexes also turn out to be the major factors determining the overall morphology of ICMs (Chandler, Gumbart, Stack, Chipot, & Schulten, 2009; Chandler, Hsin, Harrison, Gumbart, & Schulten, 2008; Qian, Bullough, & Hunter, 2008). If Mic60 or Orf52 has an effect on the development or morphology of ICMs (e.g., by facilitating protein secretion and insertion into the ICM), it is then possible that their absence might affect the relative proportion of the RC-LH1 (or B875 for the wavelength at which its non-covalently bound Bchl *a*

absorbs) to the LH2 (or B800-850) complexes. I therefore investigated this possibility by determining the optical density (OD) spectra of whole cells of *R. sphaeroides* and *R. palustris* under both high-light and low-light regimes.

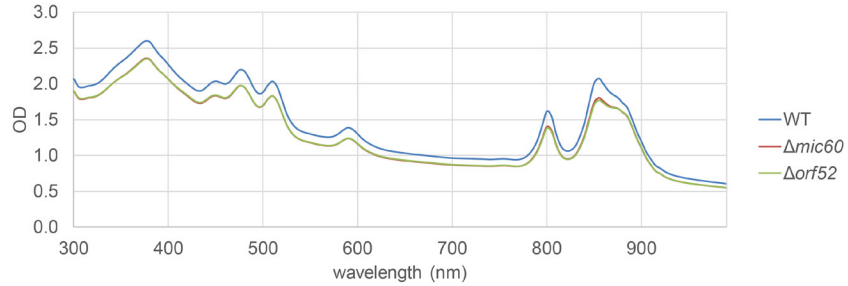
In the OD spectra of whole cells of purple bacteria, the peaks at 800 and 850 nm correspond to the LH2 complexes, whereas the peak at 875 nm corresponds to the RC-LH1 complexes (Hunter, 1995). All OD spectra show these three peaks (Fig. 3.3 and 3.4), but in *R. palustris* the 875 nm is less obvious. For *R. sphaeroides*, it can be seen how lower light intensity increases the relative height of the B850 peak which corresponds to the LH2 complex. This is because as light intensity decreases, higher amounts of LH2 complexes are synthesized for capturing more light to be transferred to the RC-LH1 complex. In *R. palustris*, however, this is not so clear. For both *R. sphaeroides* and *R. palustris* there are no clear differences in the shapes of the OD spectra between WT and mutants; the height and sharpness of the peaks are mostly a consequence of the concentration of the sample measured (Myers, Curtis, & Curtis, 2013). This suggests that disrupting either the *mic60* or *orf52* genes does not affect the relative proportion of the RC-LH1 to the LH2 complexes and therefore Mic60 and Orf52 do not have a direct effect on the major constituents of ICMs. Future efforts will need to be focused on determining OD or absorbance (of isolated chromatophores or protein complexes) spectra for *R. palustris* grown in rich media (e.g., FEM) where it can reach higher cell densities. Testing light intensities different from the ones tried here might reveal differences between the WT and knock-out mutant strains.

**Figure 3.5.** OD spectra of photosynthetically grown *R. sphaeroides* and *R. palustris* WT and Δ*mic60* and Δ*orf52* strains. **A, B.** OD spectra of photosynthetically grown *R. sphaeroides* strains in RCVBN medium under high light and under low light, respectively. **C, D.** OD spectra of photosynthetically grown *R. palustris* strains in FEM medium under high light and in YPS medium under low light. OD spectra were determined from 300 to 990 nm and in steps of 5 nm.

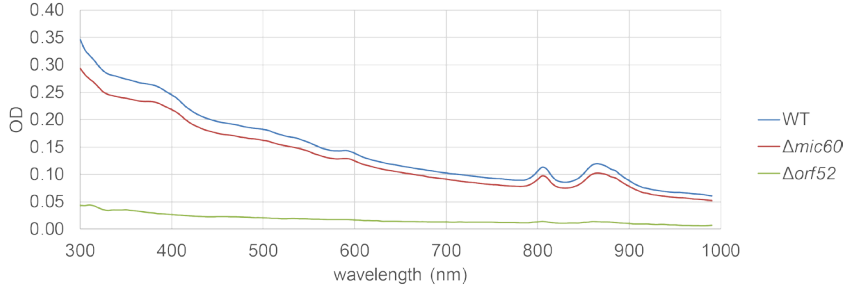
A.


OD spectra of R. sphaeroides at high light (RCVBN)
OD
wavelength (nm)
WT
Δmic60
Δorf52

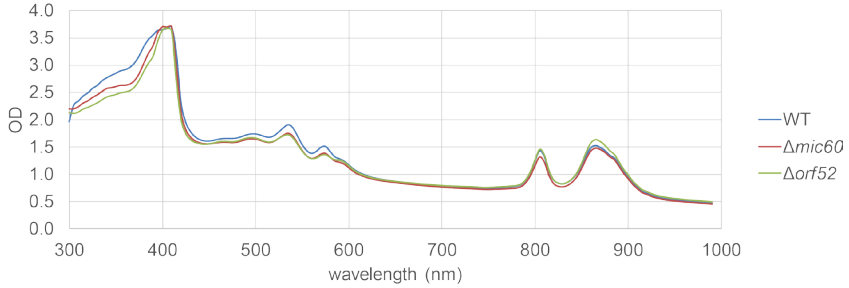
B.


OD spectra of R. sphaeroides at low light (RCVBN)
OD
wavelength (nm)
WT
Δmic60
Δorf52

C.


OD spectra of R. palustris at high light (FEM)
OD
wavelength (nm)
WT
Δmic60
Δorf52

D.


OD spectra of R. palustris at low light (YPS)
OD
wavelength (nm)
WT
Δmic60
Δorf52

### 3.3.4. Knock-out mutants might display some ultrastructural defects

Strains were grown to stationary phase, fixed with the Karnovky's fixative, processed and then viewed under the TEM. After several attempts, only one batch of *R. spaheroides* samples were properly fixed for TEM imaging, whereas properly fixed samples for *R. palustris* could not be obtained. When the strains were grown under oxic and dark conditions, ICMs were not observed (Fig. 3.6). This was expected because oxygen concentration and light are the main environmental factors responsible for inducing the expression of the the RC-LH1 and LH2 complexes and therefore ICMs (Drews, 1992). The Δ*mic60* and Δ*orf52* knock-out mutants did not differ anatomically from the WT strain under these conditions (Fig. 3.6). When grown under anoxia and light, ICMs were observed (Fig. 3.7). Under these conditions, there appear to be some consistent anatomical differences between the WT strain and the mutants, specifically the Δ*mic60* mutant (Fig. 3.7). The Δ*mic60* mutant displayed tubular ICMs alongside the typical vesicular ICMs of *R. sphaeroides* (Fig. 3.7). In the Δ*orf52* mutant, these tubular ICMs were also observed but appeared to be less common (Fig. 3.7). The tubular ICMs observed here resemble those seen in $LH2^-$ mutants (Kiley, Varga, & Kaplan, 1988) and also those seen in cultures grown in the presence of nitrate (Sabaty, Gagnon, & Verméglio, 1994). It has been suggested that these tubular ICMs are correlated with a decrease in the LH1/LH2 ratio of ICMs (Kiley et al., 1988; Verméglio, Joliot, & Joliot, 1995). This would be, however, in conflict with the results from the OD spectra that suggest that the LH1/LH2 ratio between WT and mutants did not differ (see Fig. 3.5). It is important to note that these observations are preliminary and need to be reproduced in the future. Different growth conditions, (e.g., light intensity and oxygen concetration) need to be tested. Future efforts should also focus on quantitatively assessing the number of tubular ICMs relative to vesicular ICMs seen per section.

**Figure 3.6.** TEM micrographs of *R. sphaeroides* WT, and Δ*mic60* and Δ*orf52* strains grown chemoheterotrophically under oxia and darkness. **A, B.** Cells of *R. sphaeroides* WT under oxia and darkness. **C, D.** Cells of *R. sphaeroides* Δ*mic60* under oxia and darkness. **E, F.** Cells of *R. sphaeroides* Δ*orf52* under oxia and darkness.

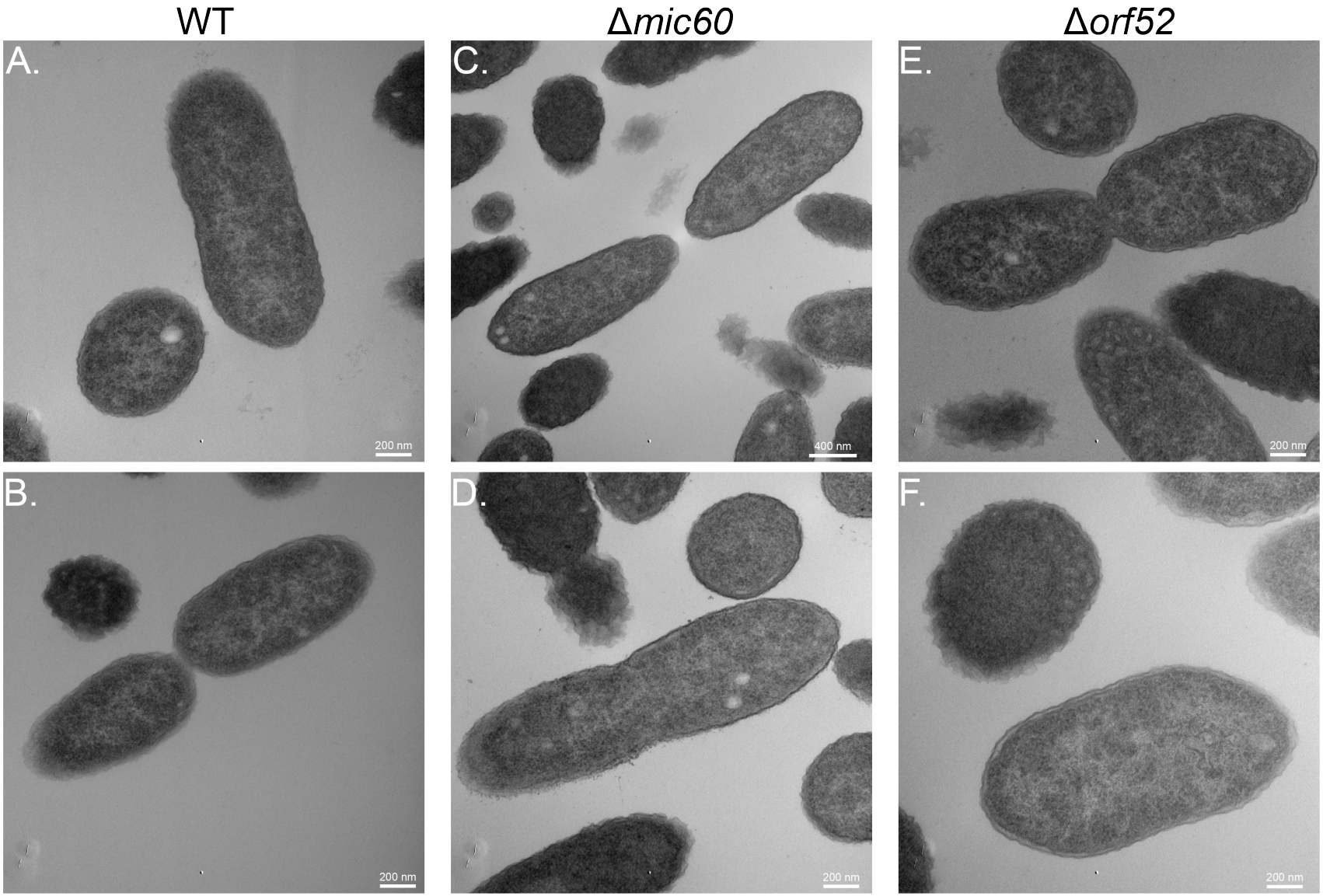
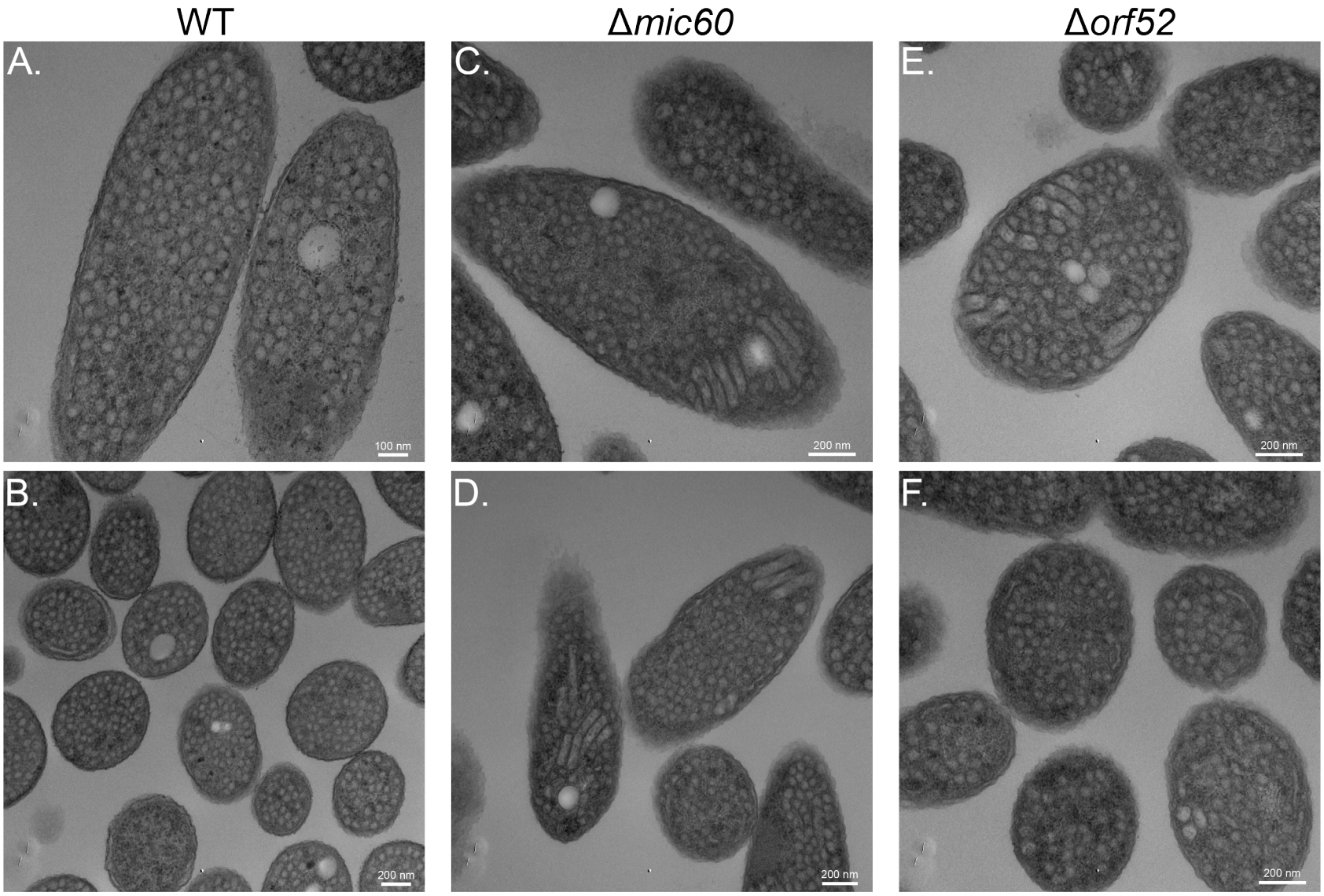

WT
Δmic60
Δorf52
A.
B.
C.
D.
E.
F.
200 nm
200 nm
400 nm
200 nm
200 nm
200 nm

**Figure 3.7.** TEM micrographs of *R. sphaeroides* WT, and Δ*mic60* and Δ*orf52* strains grown photoheterotrophically under anoxia and high light. **A, B.** Cells of *R. sphaeroides* WT under anoxia and high light. **C, D.** Cells of *R. sphaeroides* Δ*mic60* under anoxia and high light. **E, F.** Cells of *R. sphaeroides* Δ*orf52* under anoxia and high light.


WT
Δmic60
Δorf52
A.
B.
C.
D.
E.
F.
100 nm
200 nm
200 nm
200 nm
200 nm
200 nm

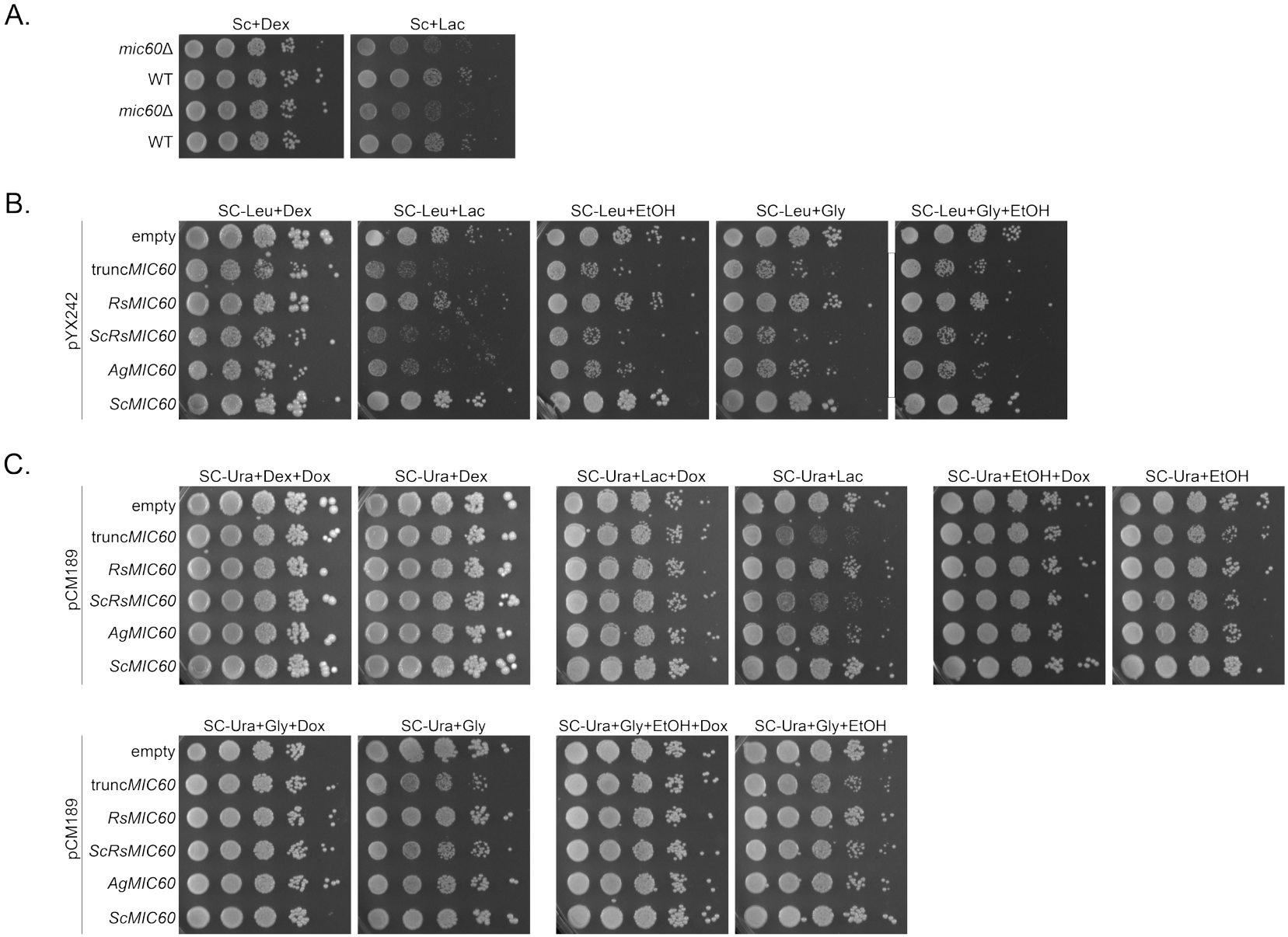
### 3.3.5. Heterologous expression of Mic60 homologs in yeast does not recover growth

An alternative approach to assessing the function of the alphaproteobacterial *mic60* homolog was to see if it could substitute for a eukaryotic *mic60* homolog in a model eukaryotic organism. To assess this, a series of plasmids were created that could express different versions of the *MIC60* gene in *S. cerevisiae* yeast lacking its own endogenous copy of the *MIC60* gene (*mic60*Δ). Two expression vectors were used, pCM189 and pYX242. pCM189 allows expression of the inserted gene to be repressed using doxycycline (a tetracycline derivative; Garí et al., 1997), whereas pYX242 provides constitutive expression of the inserted gene under the control of the *TPI* (triose phosphate isomerase) promoter. The different *MIC60* constructs were designed so as to provide different degrees of functional complementation as expected from the evolutionary distance between the source of the gene construct and yeast. The endogenous yeast *MIC60* homolog was used as a positive control, and empty pCM189 and pYX242 plasmids were used as negative controls. In order of expected degree of functional complementation, the following constructs were assayed: (1) the *MIC60* homolog of the jakobid *Andalucia godoyi* (*AgMIC60*), (2) a chimeric *MIC60* comprising the N-terminus of yeast and the C-terminus (the PFAM mitofilin domain) of the purple bacterium *R. sphaeroides* (*ScRsMIC60*), (3) the *MIC60* homolog of *R. sphaeroides* (*RsMIC60*), and (4) a truncated *MIC60* containing only the first N-terminal half of the yeast *MIC60* homolog (trunc*MIC60*). All of these constructs contained the endogenous yeast N-terminal presequence responsible for targeting pre-Mic60 to the yeast mitochondrion, except *AgMIC60* which uses the *A. godoyi* presequence.

To determine whether absence of the *MIC60* gene affected growth, WT and *mic60*Δ yeast strains were serially diluted and plated on fermentable (i.e., dextrose as carbon source) and non-fermentable (i.e., ethanol, glycerol or lactate as carbon sources) solid media and growth was assessed for several days (Fig. 3.8A). The *mic60*Δ strain showed no growth difference relative to the WT strain when grown on a fermentable medium (Fig. 3.8A). When both strains are grown on a non-fermentable medium, which exacerbates respiratory phenotypes, a slight but notable difference in the growth rate was now observed (Fig. 3.8A). Yeast *mic60*Δ strains transformed with plasmids encoding various constructs of the *MIC60* gene (see above) were then assayed for growth after spotting serial dilutions of the cultures on both fermentable and non-

fermentable solid media. When the *MIC60* constructs were expressed from the *TPI* constitutive promoter of the pYX242 plasmid, a similar pattern was seen for the positive control (*ScMIC60*) relative to the negative control (Fig. 3.8B). Furthermore, the expression of the trunc*MIC60*, *ScRsMIC60*, and *AgMIC60* gene constructs decreased growth relative to both the positive and negative controls (*ScMIC60* and empty, respectively). This implies that the expressed heterologous gene products are toxic and might be explained by the fact that Mic60 is a highly pleiotropic protein with many interactors and thus having no Mic60 might provide a higher fitness over expression of a misfit protein. Indeed, the constitutive expression of the trunc*MIC60*, *ScRsMIC60*, and *AgMIC60* gene constructs in fementable media (SC-Leu+Dex; Fig. 3.8B), demonstrates that these gene products are toxic even in conditions under which aerobic respiration is not required for growth. The *RsMIC60* construct, on the contrary, did not decrease growth as much relative to both controls. It is difficult to interpret this result, but it appears to be consistent with the findings of Tarasenko et al., (2017). Similar growth patterns are seen when expression from the pCM189-based plasmids was activated by the absence of doxycycline (Fig. 3.8C). In these experiments, however, the differences between the positive and negative controls were not appreciable. The causes are unknown. The addition of doxycycline to the medium represses transcription of the *MIC60* construct genes and no growth differences among the strains were observed (Fig. 3.8C). The growth patterns observed when the *MIC60* constructs are expressed from the doxycycline-regulatable promoter of pCM189 thus largely corroborate those seen when they were expressed from the constitutive *TPI* promoter of pYX242. In summary, three gene constructs appear to be toxic when expressed, whereas the *RsMIC60* gene construct confered no or a very slight growth advantage over the negative control (as also seen by Tarasenko et al., 2017).

**Figure 3.8.** Spot dilution assays of yeast strains to test the functional complementation of different *MIC60* gene constructs heterologously expressed in *mic60*Δ yeast. **A.** Relative growth rates of WT and *MIC60*Δ strains in fermentable and non-fermentable media. **B.** Relative growth of *mic60*Δ strains transformed with a series of *URA*3-marked pCM189 plasmids that express different *MIC60* gene contructs (trunc*MIC60*, *RsMIC60*, *ScRsMIC60*, *AgMIC60*, and *ScMIC60*) upon the absence of doxycycline on fermentable and non-fermentable media. **C.** Relative growth rates of *mic60*Δ strains transformed with a series of *LEU*2-marked pYX242 plasmids that constitutively express different *MIC60* gene contructs (trunc*MIC60*, *RsMIC60*, *ScRsMIC60*, *AgMIC60*, and *ScMIC60*) on fermentable and non-fermentable media. Synthetic complete (SC) medium lacking uracil (-Ura) or L-leucine (-Leu); Dex: dextrose; Lac: lactate; EtOH: ethanol; Gly: glycerol; Dox: doxycycline. Yeast were cultured at 30°C for four days.


A.
Sc+Dex
Sc+Lac
mic60Δ
WT
mic60Δ
WT
B.
SC-Leu+Dex
SC-Leu+Lac
SC-Leu+EtOH
SC-Leu+Gly
SC-Leu+Gly+EtOH
pYX242
empty
truncMIC60
RsMIC60
ScRsMIC60
AgMIC60
ScMIC60
C.
SC-Ura+Dex+Dox
SC-Ura+Dex
SC-Ura+Lac+Dox
SC-Ura+Lac
SC-Ura+EtOH+Dox
SC-Ura+EtOH
pCM189
empty
truncMIC60
RsMIC60
ScRsMIC60
AgMIC60
ScMIC60
SC-Ura+Gly+Dox
SC-Ura+Gly
SC-Ura+Gly+EtOH+Dox
SC-Ura+Gly+EtOH
pCM189
empty
truncMIC60
RsMIC60
ScRsMIC60
AgMIC60
ScMIC60

## 3.4. DISCUSSION

A method to successfully knock-out genes in both *R. sphaeroides* and *R. palustris* was devised. These knock-out strains did not exhibit obvious phenotypic differences at the colony level when grown on solid or in liquid media. However, a more careful examination of their growth rates did reveal some interesting differences. Even though the differences were small, they are consistent and statistically significant (Fig. 3.4). These differences also became more pronounced when light intensity was decreased. OD spectra of whole cells did not reveal obvious differences in regard to the relative proportions of RC-LH1 to LH2 complexes. This suggests that the knock-out mutants were not affected in their ability to synthesize and insert these complexes into their ICMs. However, preliminary TEM imaging of Δ*mic60* and Δ*orf52* mutants suggested an increased number of tubular ICMs alongside vesicular ICMs, relative to the WT strain. These differences have to be quantified at the population level in future studies.

I also found that different *MIC60* constructs heterologously expressed in a *mic60*Δ yeast cells were not able to functionally complement growth. On the contrary, it appears that the constructs were toxic, and this might reflect the specific co-evolved features that Mic60 has with its interaction partners. Indeed, expression of the human Mic60 protein is not able to functionally complement *mic60*Δ yeast strains (Ralf M. Zerbes, personal communication). It is therefore curious, and contrary to expectations, that the *RsMIC60* gene construct showed no toxicity and even slightly recovered the growth phenotype of the *mic60*Δ yeast. This is partially in agreement with the results of Tarasenko et al., (2017) who reported a slight improvement in the growth rate of the functionally complemented *mic60*Δ yeast strain with a *RsMIC60* gene construct.

Even though the results presented here are inconclusive, there is additional functional or experimental evidence for a possible role of *mic60* and *orf52* in the development of ICMs. I have compiled several lines of evidence that are consistent with the idea that Mic60 in alphaproteobacteria performs the same general function that it does in mitochondria. First is the presence of a Mitofilin signature domain in alphaproteobacterial Mic60 (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). Second is the predicted structural conservation (same motif order and composition) between alphaproteobacterial and eukaryotic Mic60 (Huynen et al., 2016; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015). Third is the residue- and secondary

structure-level conservation of the most functionally important motif in Mic60, an amphipathic helix that serves as a lipid-binding motif (see Chapter 2). Fourth is the expression profile of alphaproteobacterial Mic60 which reveals that it is overexpressed under conditions that promote ICM development in *Rhodobacter sphaeroides* (Callister et al., 2006). Fifth is the localization of alphaproteobacterial Mic60 to ICMs, as revealed by proteomic studies of isolated ICMs (i.e., chromatophores) of *Rhodobacter sphaeroides* (D'Amici, Rinalducci, Murgiano, Italiano, & Zolla, 2010; Jackson, Lewis, Tucker, Hunter, & Dickman, 2012), *Rhodospirillum rubrum* (Selao et al., 2011), and *Rhodopseudomonas palustris* (Fejes, Yi, Goodlett, & Beatty, 2003); these three purple bacteria belong to three separate orders, the *Rhodobacterales*, *Rhizobiales* and *Rhodospirillales*, respectively. And sixth is the association of a paralog of Mic60 with magnetosome membranes (Lohße et al., 2011), and its functional involvement in the development of magnetosomes among alphaproteobacterial magnetospirilla (Lohße et al., 2014)—this *mic60* paralog in the magnetospirilla was first reported by Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015.

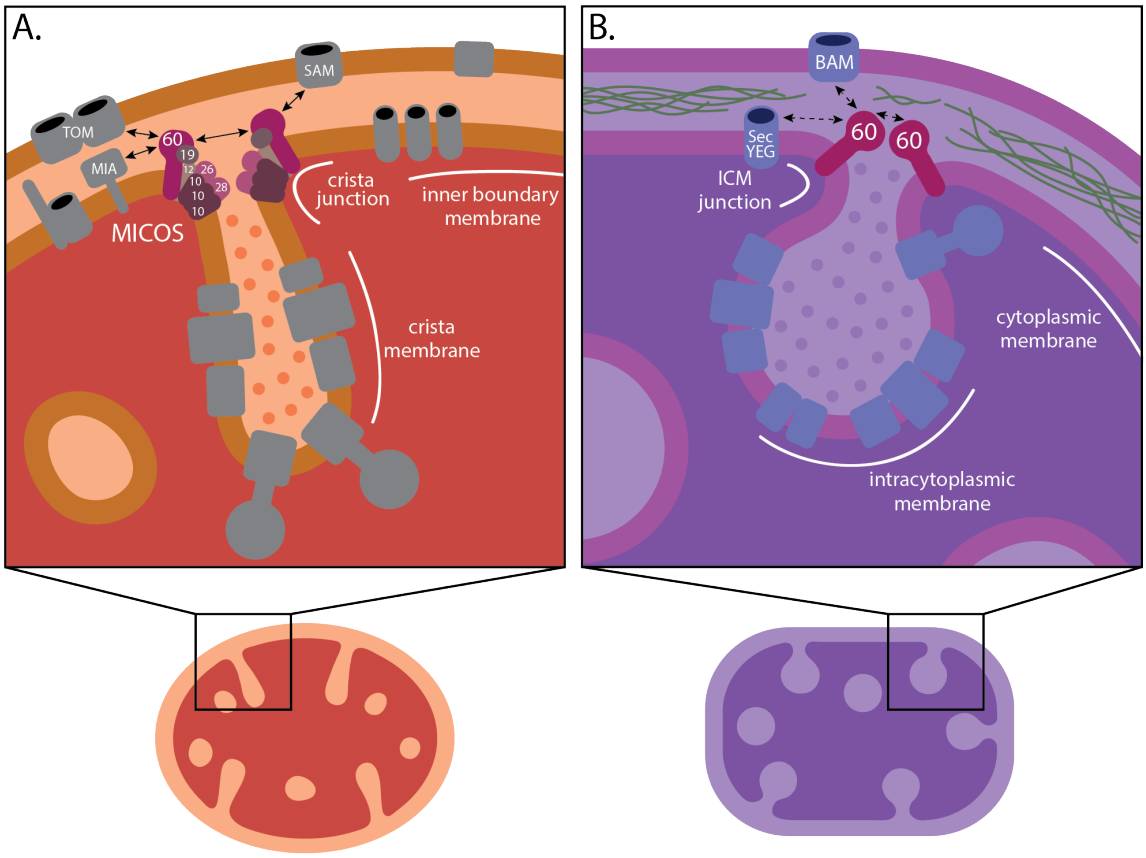
Furthermore, similarly to Mic60, Orf52 has also been found to be enriched in isolated ICMs from purple bacteria (Fejes et al., 2003; Jackson et al., 2012; Zeng et al., 2007), and is considerably more abundant under photosynthetic conditions in *R. sphaeroides* (Arai, Roh, & Kaplan, 2008). The genomic linkage, expression pattern, membrane localization, and motif order and structure of Mic60 and Orf52 suggest that these proteins might interact with each other to play a role in the development of ICMs, or more generally, in the biogenesis of alphaproteobacterial envelopes.

The strongest and most direct experimental evidence for the conserved function of Mic60 in both alphaproteobacteria and mitochondria was provided by Tarasenko et al., (2017). They showed not only that the *R. sphaeroides* homolog of Mic60 can functionally complement a yeast mutant strain by rescuing its growth phenotype, but also that the herelogously-expressed *RsMIC60* gene is able to recover CJs and restore the overall morphology of yeast mitochondria to some degree (Tarasenko et al., 2017). Furthermore, Tarasenko et al., (2017) also found that the heterologous expression of *RsMIC60* in *E. coli* can create invaginations of its cytoplasmic membrane that in some instances can resemble the vesicular ICMs of *R. sphaeroides*. Most convincingly, the same authors showed that the *in vitro* incubation of isolated *Rs*Mic60 can deform vesicular liposomes into tubules, thus demonstrating the deforming capabilities of

*Rs*Mic60 (Tarasenko et al., 2017). This last experiment agrees well with the finding that the lipid-binding and membrane-deforming aphipathic helix of Mic60 (Hessenberger et al., 2017) is also conserved among the alphaproteobacteria (see Chapter 2).

In order to create cristae, eukaryotic Mic60 engages in homotypic interactions with itself at CJs, and in heterotypic interactions with Sam50 and Tom40 at CSs between MIM and MOM (John et al., 2005; Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011) (Fig. 3.9A). The structural conservation of alphaproteobacerial Mic60 relative to eukaryotic Mic60 suggests that alphaproteobacterial Mic60 possesses all the elements required for the formation of ICM junctions (ICMJs) and CSs at the cytoplasmic membranes of alphaproteobacteria (Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015; Muñoz-Gómez, Slamovits, Dacks, & Wideman, 2015). A main function of alphaproteobacterial Mic60 could then be the formation of ICMJs. These junctions between ICMs and the cytoplasmic membrane of alphaproteobacteria are functionally analogous to CJs in mitochondria and likely play a role in restricting the diffusion of metabolites between the ICM lumen and the periplasm (Fig. 3.9B). Interestingly, the development of vesicular ICMs in *R. sphaeroides* and *Rhodospirillum rubrum* necessarily requires the formation of tubular regions between the ICM and the cytoplasmic membrane (Scheuring et al., 2014; Tucker et al., 2010). The lamellar ICMs of *Blastochloris viridis* have been shown to be connected to one another and the cytoplasmic membrane by tubular structures similar to mitochondrial CJs (Konorty et al., 2009; Konorty, Kahana, Linaroudis, Minsky, & Medalia, 2008). Moreover, ICMJs can also aid in the differentiation of the ICM from the cytoplasmic membrane by limiting the lateral diffusion of protein complexes between these two membrane domains (Fig. 3.9B). Similarly to CMs in mitochondria, ICMs clearly differ in composition, function and structure from the cytoplasmic membrane of alphaproteobacteria (Niederman, 2006) (Fig. 3.9). By creating ICMJs, alphaproteobacterial Mic60 might compartmentalize the diverse bioenergetic metabolisms that take place in alphaproteobacterial ICMs (Fig. 3.9B).

**Figure 3.9.** The function of MICOS is probably conserved between mitochondria and alphaproteobacteria. **A.** In the mitochondrion of *S. cerevisiae*, MICOS forms CSs and CJs to maintain and stabilize cristae (Harner et al., 2011; Hoppins et al., 2011; Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012). MICOS is composed of the Mic60–Mic19 subcomplex which establishes CSs with the MOM and marks the sites of CJ formation, and the Mic12–Mic10–Mic26–Mic28 subcomplex which differentiates and bends the MIM at CJs (Friedman et al., 2015; Muñoz-Gómez, Slamovits, Dacks, Baier, et al., 2015; Pfanner et al., 2014; Zerbes et al., 2016). The central MICOS subunit, Mic60, contacts the outer membrane through its interactions with the TOM and SAM complexes (additional interactions have also been reported with the MOM proteins porin and Ugo1), and further interacts with Mia40 to aid in the oxidative import of mitochondrial proteins (Körner et al., 2012; Ott et al., 2012; von der Malsburg et al., 2011; Zerbes, Bohnert, et al., 2012). **B.** In alphaproteobacteria, Mic60 is presumably involved in the formation of ICMJs and CSs to stabilize bioenergetic ICMs. Alphaproteobacterial Mic60 likely uses its conserved mitofilin domain to interact with the BAM complex, the bacterial homolog of SAM. Alphaproteobacterial Mic60 might therefore mark the sites of ICM invagination and keep ICMs anchored to the alphaproteobacterial envelope at CSs. Alternatively, alphaproteobacterial Mic60 might bring together protein translocases (e.g., SecYEG and BAM) for proper envelope biogenesis (the so called Bayer's junctions) at sites of murein hypotrophy. The compartmentalization of both mitochondrial cristae and alphaproteobacterial ICMs is achieved by narrow tubules (i.e., CJS and ICMJs) likely made and stabilized by Mic60. Mitochondrial cristae and alphaproteobacterial ICMs are therefore functionally analogous. They both constitute specialized sub-compartments that optimize the efficiency of energy transduction by concentrating bioenergetic metabolism. Thick arrows indicate physical interactions between protein partners. Dashed arrows indicate hypothesized physical interactions between protein partners based on the known function of their homologs.


A.
SAM
TOM
MIA
60
19
12
26
10
28
10
10
MICOS
crista
junction
inner boundary
membrane
crista
membrane
B.
BAM
Sec
YEG
60
60
ICM
junction
cytoplasmic
membrane
intracytoplasmic
membrane

Synergistically with the formation of ICMJs, alphaproteobacterial Mic60 might also be involved in the formation of CSs between the cytoplasmic membrane and the OM of the alphaproteobacterial envelope. This function would be analogous to the formation of CSs at the mitochondrial envelope, which requires Sam50 at the MOM as an interaction partner of eukaryotic Mic60 (Körner et al., 2012; Ott et al., 2012). In support of this hypothesis, Sam50 is homologous to the ubiquitous outer membrane (OM) protein BamA of Gram-negative bacteria (Heinz & Lithgow, 2014). In addition, alphaproteobacterial BamA homologs harbor N-terminal POTRA domains responsible for their interactions with diverse protein partners (Simmerman, Dave, & Bruce, 2014), consistent with the interaction of Mic60 with the POTRA domain of Sam50 at mitochondrial CSs (Bohnert et al., 2012; Körner et al., 2012; Zerbes, Bohnert, et al., 2012). CSs formed by Mic60 could therefore anchor ICMJs to the alphaproteobacterial envelope and stabilize ICMs (Fig. 3.9B). The maintenance and stability of alphaproteobacterial ICMs might therefore rely on the formation of CSs and ICMJs by Mic60 at the alphaproteobacterial envelope (Fig. 3.9B).

Alternatively, alphaMic60 might play a more general role as the molecular basis of the CSs formed between the cytoplasmic membrane and the OM (also called Bayer's junctions) required for the proper biogenesis of the alphaproteobacterial envelope (Bayer, 1991). Under this scenario, alphaproteobacterial Mic60 would function in facilitating protein export by bringing together translocases of the cytoplasmic membrane (e.g., SecYEG) in contact with BamA or other OM complexes (Fig. 3.9B). In modern mitochondria, Mic60 works similarly by interacting with the TOM (Tom40), MIA (Mia40) and SAM (Sam50) complexes to aid in protein import (Ott et al., 2012; von der Malsburg et al., 2011) (Fig. 3.9A).

## 3.5. CONCLUSIONS AND FUTURE DIRECTIONS

Comparative and experimental evidence is converging in support of the hypothesis that mitochondrial cristae evolved from ICMs. However, we need more direct evidence for the involvement of Mic60 in the development of alphaproteobacterial ICMs. I have created knock-out mutant and overexpression strains for two purple nonsulfur bacteria that develop different kinds of ICMs, *R. sphaeroides* (*Rhodobacterales*) which develops vesicular ICMs, and *R. palustris* (*Rhizobiales*) which develops lamellar ICMs. The overexpression of Mic60 in these purple bacteria (for which the strains have been made

using the pIND4 plasmid, Ind et al., 2009; and pSRK plasmid series, Khan, Gaines, Roop, & Farrand, 2008) for *R. sphaeroides* and *R. palustris*, respectively, is expected to change the shape of their ICMs, most probably by inducing their branching as has been observed in mitochondria. Detailed tomographic analyses of these strains might be able to reveal the extent to which the absence or overexpression of Mic60 and Orf52 affect the anatomy of these purple nonsulfur bacteria. Finding out what interaction partners Mic60 has at the envelope of *R. sphaeroides* and *R. palustris* might also give us important clues in regard to its function. For example, the distant homolog of Mic60 in the Gammaproteobacteria, HemX, appears to physically interact with Orf52 (‘HemY’) and with the cell division protein FtsB (Babu et al., 2018). Proximity labelling using fusion proteins with a promiscuous biotin ligase like BioID2 (Kim et al., 2016) is a promising method that could shed light on local microenvironments, in alphaproteobacterial envelopes, where Mic60 and Orf52 perform their functions.

# CHAPTER 4: THE PHYLOGENY OF THE *ALPHAPROTEOBACTERIA*

## 4.1. INTRODUCTION

The *Alphaproteobacteria* is an extraordinarily diverse and disparate group of bacteria and well-known to most biologists for also encompassing the mitochondrial lineage (Roger et al., 2017; Williams, Sobral, & Dickerman, 2007). The *Alphaproteobacteria* has massively diversified since its origin, giving rise to, for example, some of the most abundant (e.g., *Pelagibacter ubique*) and metabolically versatile (e.g., *Rhodobacter sphaeroides*) cells on Earth (Giovannoni, 2017; Madigan, Jung, & Madigan, 2009). The basic structure of the tree of the *Alphaproteobacteria* has largely been revealed through the analyses of 16S rRNA genes and several conserved proteins (Brindefalk, Ettema, Viklund, Thollesson, & Andersson, 2011; Fitzpatrick, Creevey, & McInerney, 2006; Garrity, Brenner, Krieg, & Staley, 2005; Georgiades, Madoui, Le, Robert, & Raoult, 2011; Lee et al., 2005; Luo, 2015; Rosenberg, DeLong, Lory, Stackebrandt, & Thompson, 2014; Thrash et al., 2011; Williams et al., 2007). Today, eight major orders are well recognized, namely the *Caulobacterales*, *Rhizobiales*, *Rhodobacterales*, *Pelagibacterales*, *Sphingomonadales*, *Rhodospirillales, Holosporales* and *Rickettsiales* (the latter two formerly grouped into the *Rickettsiales sensu lato*), and their interrelationships have also recently become better understood (Rodríguez-Ezpeleta & Embley, 2012; Viklund, Ettema, & Andersson, 2012; Viklund, Martijn, Ettema, & Andersson, 2013; Wang & Wu, 2014, 2015). These eight orders were grouped into two subclasses by Ferla *et al*. (2013): the subclass *Rickettsiidae* comprising the order *Rickettsiales* and *Pelagibacterales*, and the subclass *Caulobacteridae* comprising all other orders.

The great diversity of the *Alphaproteobacteria* itself presents a challenge to deciphering the deepest divergences within the group. Such diversity encompasses a broad spectrum of genome (nucleotide) and proteome (amino acid) compositions (e.g., the A+T%-rich *Pelagibacterales versus* the G+C%-rich *Acetobacteraceae*) and evolutionary rates (e.g., the fast-evolving *Pelagibacteriales*, *Rickettsiales* or *Holosporales versus* many slow-evolving species in the *Rhodospirillales*) (Ettema & Andersson, 2009). This diversity may lead to pervasive artefacts when inferring the phylogeny of the

*Alphaproteobacteria*, e.g., long-branch attraction (LBA) between the *Rickettsiales* and *Pelagibacterales*, especially when including mitochondria (Rodríguez-Ezpeleta & Embley, 2012; Viklund et al., 2012, 2013). Moreover, there are still important unknowns about the deep phylogeny of *Alphaproteobacteria* (see Ferla, Thrash, Giovannoni, & Patrick, 2013; Williams et al., 2007), for example, the divergence order among the *Rhizobiales*, *Rhodobacterales* and *Caulobacterales* (e.g., Williams et al., 2007), the monophyly of the *Pelagibacterales* (e.g., Viklund et al., 2013) and the *Rhodospirillales* (e.g., Ferla et al., 2013), and the precise placement of the *Rickettsiales* and its relationship to the *Holosporales* (e.g., Martijn et al., 2018; Wang & Wu, 2015).

Systematic errors stemming from using simple (not realistic enough) evolutionary models are perhaps the major confounding and limiting factor to inferring deep evolutionary relationships; the number of taxa and genes (or sites) can also be important factors. Previous multi-gene tree studies of the *Alphaproteobacteria* were compromised by at least one of these problems, namely, simple or unrealistic evolutionary models (because more sophisticated models were not available at the time; e.g., Williams et al., 2007), poor taxon sampling (because the focus was too narrow or few genomes were available; e.g., Georgiades et al., 2011; Martijn et al., 2015; Williams et al., 2007) or a small number of genes (because the focus was mitochondria; e.g., Martijn et al., 2018; Rodríguez-Ezpeleta & Embley, 2012; Wang & Wu, 2015). The most recent study on the phylogeny of the *Alphaproteobacteria*, and mitochondria, attempted to counter systematic errors (or phylogenetic artefacts) by reducing amino acid compositional heterogeneity (Martijn et al., 2018). Even though this tree has a weakly supported backbone, it suggests that the *Pelagibacterales*, *Rickettsiales* and *Holosporales*, which have compositionally biased genomes, are not each other's closest relatives (Martijn et al., 2018). A resolved and robust phylogeny of the *Alphaproteobacteria* is fundamental to addressing questions such as how streamlined bacteria, intracellular parasitic bacteria, or mitochondria evolved from their alphaproteobacterial ancestors. Therefore, a systematic study of the different biases affecting the phylogeny of the *Alphaproteobacteria*, and its underlying data, is much needed.

Here, we revised the phylogeny of the *Alphaproteobacteria* by using a large dataset of 200 conserved single-copy genes and employing carefully designed strategies aimed at alleviating phylogenetic artefacts. We showed that amino acid compositional heterogeneity, and more generally long-branch attraction, were major confounding

factors in estimating phylogenies for the *Alphaproteobacteria*. In order to counter these biases, we used novel dataset-specific profile mixture models and recoding schemes (both specifically designed to ameliorate compositional heterogeneity), and removed sites, genes and taxa that were compositionally biased. We also present three draft genomes for endosymbiotic alphaproteobacteria belonging to the *Rickettsiales* and *Holosporales*: (1) an undescribed midichloriacean endosymbiont of *Peranema trichophorum*, (2) an undescribed rickettsiacean endosymbiont of *Stachyamoeba lipophora*, and (3) the holosporalean '*Candidatus* Finniella inopinata', an endosymbiont of the rhizarian amoeboflagellate *Viridiraptor invadens* (Hess, Suthaus, & Melkonian, 2015). Our results provide the first strong evidence that the *Holosporales* are unrelated to the *Rickettsiales* and originated instead from within the *Rhodospirillales*. We incorporate these and other insights regarding the deep phylogeny of the *Alphaproteobacteria* into an updated taxonomy.

## 4.2. METHODS

### 4.2.1. Genome sequencing

Axenic cultures of *Viridiraptor invadens* strain Virl02, the host of '*Candidatus* Finniella inopinata', were grown on the filamentous green alga *Zygnema pseudogedeanum* strain CCAC 0199 as described in Hess & Melkonian, (2013). Once the algal food was depleted, *Viridiraptor* cells were harvested by filtration through a cell strainer (mesh size 40 µm to remove algal cell walls) and centrifugation (~1,000 g for 15 min). For short-read sequencing, DNA extraction of total gDNA was carried out with the ZR Fungal/Bacterial DNA MicroPrep Kit (Zymo Research) using a BIO101/Savant FastPrep FP120 high-speed bead beater and 20 µL of proteinase K (20 mg/mL). A sequencing library was made using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs). DNA sequencing libraries were sequenced with an Illumina MiSeq instrument (Dalhousie University; Canada). For long-read sequencing, DNA extraction was performed using a CTAB and phenol-chloroform method. Total gDNA was further cleaned through a QIAGEN Genomic-Tip 20/G. A sequencing library was made using the Nanopore Ligation Sequencing Kit 1D (SQK-LSK108). Sequencing was done on a portable MinION instrument (Oxford Nanopore Technologies).

*Peranema trichophorum* strain CCAP 1260/1B was obtained from the Culture Collection of Algae and Protozoa (CCAP, Oban, Scotland) and grown in liquid Knop medium plus

egg yolk crystals. Total gDNA was extracted following Lang and Burger (2007). A sequencing library was made using a TruSeq DNA Library Prep Kit (Illumina). DNA sequencing libraries were sequenced with an Illumina MiSeq instrument (Genome Quebec Innovation Centre; Canada).

*Stachyamoeba lipophora* strain ATCC 50324 cells feeding on *Escherichia coli* were harvested and these were broken up with pestle and mortar in the presence of glass beads (< 450 µm diameter). Total gDNA was extracted using the QIAGEN Genomic G20 Kit. A sequencing library was made using a TruSeq DNA Library Prep Kit (Illumina). DNA sequencing libraries were sequenced with an Illumina MiSeq instrument (Genome Quebec Innovation Centre; Canada).

### 4.2.2. Genome assembly and annotation

Short sequencing reads produced in an Illumina MiSeq for *Viridiraptor invadens*, *Peranema trichophorum*, and *Stachyamoeba lipophora* were first assessed with FASTQC v0.11.6 and then, based on its reports, trimmed with Trimmomatic v0.32 (Bolger, Lohse, & Usadel, 2014) using the options: HEADCROP:16 LEADING:30 TRAILING:30 MINLEN:36. Illumina adapters were similarly removed with Trimmomatic v0.32 using the option ILLUMINACLIP. Long sequencing reads produced in an Nanopore MinION instrument from *Viridiraptor invadens* were basecalled with Albacore v2.1.7, adapters were removed with Porechop v0.2.3, lambda phage reads were removed with NanoLyse v0.5.1, quality filtering was done with NanoFilt v2.0.0 (with the options '--headcrop 50 -q 8 -l 1000'), and identity filtering against the high-quality short Illumina reads was done with Filtlong v0.2.0 (and the options '--keep_percent 90 --trim --split 500 --length_weight 10 --min_length 1000'). Statistics were calculated throughout the read processing workflow with NanoStat v0.8.1 and NanoPlot v1.9.1. A hybrid co-assembly of both processed Illumina short reads and Nanopore long reads from *Viridiraptor invadens* was done with SPAdes v3.6.2 (Bankevich et al., 2012). Assemblies of the Illumina short reads from *Peranema trichophorum* and *Stachyamobea lipophora* were separately done with SPAdes v3.6.2 (Bankevich et al., 2012). The resulting assemblies for both *Viridiraptor invadens* and *Peranema trichophorum* were later separately processed with the Anvi'o v2.4.0 pipeline (Eren et al., 2015) and refined genome bins corresponding to '*Candidatus* Finniella inopinata' and the *Peranema*-associated rickettsialean were isolated primarily based on tetranucleotide sequence

composition and taxonomic affiliation of its contigs. A single contig corresponding to the genome of the *Stachyamoeba*-associated rickettsialean was obtained from its assembly and this was circularized by collapsing the overlapping ends of the contig. Gene prediction and genome annotation was carried out with Prokka v.1.13 (see Table 4.1).

### 4.2.3. Taxon and gene selection

The selection of 120 taxa was largely based on the phylogenetically diverse set of alphaproteobacteria determined by Wang and Wu (2015). To this set of taxa, recently sequenced and divergent unaffiliated alphaproteobacteria were added, as well as those claimed to constitute novel order-level taxa. Some other groups, like the *Pelagibacterales*, *Rhodospirillales* and the *Holosporales*, were expanded to better represent their diversity (see Fig. S4.1).

A set of 200 marker genes (54,400 sites; 9.03% missing data, see Fig. S4.1) defined by Phyla-AMPHORA was used (Wang & Wu, 2013). The genes are single-copy and predominantly vertically-inherited as assessed by congruence among them (Wang & Wu, 2013). Another smaller dataset of 40 compositionally-homogenous genes (5,570 sites; 5.98% missing data) was built by selecting the least compositionally heterogeneous genes from the larger 200 gene set according to compositional homogeneity tests performed in P4 (Table S4.1). This was done as an alternative way to overcome the strong compositional heterogeneity observed in datasets for the *Alphaproteobacteria* with a broad selection of taxa. In brief, the P4 tests rely on simulations based on a provided tree (here inferred for each gene under the model LG4X+F in IQ-TREE) and a model (LG+F+G4 available in P4) to obtain proper null distributions to which to compare the $X^2$ statistic. Most standard tests for compositional homogeneity (those that do not rely on simulating the data on a given tree) ignore correlation due to phylogenetic relatedness, and can suffer from a high probability of false negatives (Foster, 2004).

Variations of our full set were made to specifically assess the placement of each long-branching group individually. In other words, each group with comparatively long branches (the *Rickettsiales*, *Pelagibacterales*, *Holosporales*, and alphaproteobacterium sp. HIMB59) was analyzed in isolation, i.e., in the absence of other long-branching taxa. This was done with the purpose of removing the potential artefactual attraction among these groups. Taxon removal was done in addition to compositionally biased site

removal and data recoding into reduced character-state alphabets (for a summary of the different methodological strategies employed see Fig. S4.2).

### 4.2.4. Removal of compositionally-biased and fast-evolving sites

In an effort to reduce artefacts in phylogenetic inference from our dataset (which might stem from extreme divergence in the evolution of the *Alphaproteobacteria*), we removed sites estimated to be highly compositionally heterogeneous or fast evolving. The compositional heterogeneity of a site was estimated by using a metric intended to measure the degree of disparity between the most A+T%-rich taxa and all others. Taxa were ordered from lowest to highest proteome GARP:FIMNKY ratios; 'GARP' amino acids are encoded by G+C%-rich codons, whereas 'FIMNKY' amino acids are encoded by A+T%-rich codons. The resulting plot was visually inspected and a GARP:FIMNKY ratio cutoff of 1.06 (which represented a discontinuity or gap in the distribution) was chosen to divide the dataset into low GARP:FMINKY (or A+T%-rich) and higher GARP:FIMNKY (or 'G+C%-rich') taxa (Fig. S4.3). Next, we determined the degree of compositional bias per site (ʒ) for the frequencies of both FIMNKY and GARP amino acids between the A+T%-rich and all other ('G+C%-rich') alphaproteobacteria. To calculate this metric for each site the following formula was used:

$$ʒ = (\pi FIMNKY_{\text{A+T\%-rich}} - \pi FIMNKY_{\text{G+C\%-rich}}) + (\pi GARP_{\text{G+C\%-rich}} - \pi GARP_{\text{A+T\%-rich}})$$

where $\pi FIMNKY$ and $\pi GARP$ are the sum of the frequencies for FIMNKY and GARP amino acids at a site, respectively, for either 'A+T%-rich' or 'G+C%-rich' taxa. According to this metric, higher values measure a greater disparity between A+T%-rich alphaproteobacteria and all others; a measure of compositional heterogeneity or bias per site. The most compositionally heterogeneous sites according to ʒ were progressively removed using the software SiteStripper (Verbruggen, 2018) in increments of 10%. We also progressively removed the fastest evolving sites in increments of 10%. Conditional mean site rates were estimated under the LG+C60+F+R6 model in IQ-TREE v1.5.5 using the '-wsr' flag (Nguyen, Schmidt, von Haeseler, & Minh, 2015).

### 4.2.5. Data recoding

Our datasets were recoded into four- and six-character state amino acid alphabets using dataset-specific recoding schemes aimed at minimizing compositional heterogeneity in the data (Susko & Roger, 2007). The program  minmax-chisq, which implements the

methods of Susko and Roger (2007), was used to find the best recoding schemes—please see Fig. 4.3, S4.12 and S4.14 legends for the specific recoding schemes used for each dataset. The approach uses the chi-squared ($X^2$) statistic for a test of homogeneity of frequencies as a criterion function for determining the best recoding schemes. Let $\pi_i$ denote the frequency of bin $i$ for the recoding scheme currently under consideration. For instance, suppose the amino acids were recoded into four bins,

RNCM EHIPTWV ADQLKS GFY.

Then $\pi_4$would be the frequency with which the amino acids G, F or Y were observed. Let $\pi_{is}$ be the frequency of bin $i$ for the $s$th taxa. Then the $X^2$ statistic for the null hypothesis that the frequencies are constant, over taxa, against the unrestricted hypothesis is

$$t_s = \sum_{is} (\pi_{is} - \pi_i)^2 / \pi_i$$

The $X^2$ statistic provides a measure of how different the frequencies for the $s$th taxa are from the average frequencies. The maximum $t_s$ over $s$ is taken as an overall measure of how heterogeneous the frequencies are for a given recoding scheme. The minmax-chisq program searches through recoding schemes, moving amino acids from one bin to another, to try to minimize the $max\ t_s$ (Susko & Roger, 2007).

### 4.2.6. Phylogenetic inference

The inference of phylogenies was primarily done under the maximum likelihood (ML) framework and using IQ-TREE v1.5.5 (Minh, Nguyen, & von Haeseler, 2013; Nguyen et al., 2015). We first inferred guide trees (for a PMSF analysis) with a model that comprises the LG empirical matrix, with empirical frequencies estimated from the data (F), six rates for the FreeRate model to account for rate heterogeneity across sites (R6), and a mixture model with 60 amino acid profiles (C60) to account for compositional heterogeneity across sites—LG+C60+F+R6. Because the computational power and time required to properly explore the whole tree space (given such a big dataset and complex model) was too high, constrained tree searches were employed to obtain these initial guide trees (see Fig. S4.1 for the constraint tree). Many shallow nodes were constrained if they received maximum UFBoot and SH-aLRT support in a LG+PMSF(C60)+F+R6 analysis. All deep nodes, those relevant to the questions addressed here, were left unconstrained (Fig. S4.1). The guide trees were then used together with a dataset-

specific mixture model ES60 to estimate site-specific amino acid profiles, or a PMSF (Posterior Mean Site Frequency Profiles) model, that best account for compositional heterogeneity across sites (Wang, Minh, Susko, & Roger, 2018). The dataset-specific empirical mixture model ES60 also has 60 categories but, unlike the general C60, was directly estimated from our large dataset of 200 genes and 120 alphaproteobacteria using the methods described by Susko, Lincker, & Roger, (2018). Final trees were inferred using the LG+PMSF(ES60)+F+R6 model and a fully unconstrained tree search. Those datasets that produced the most novel topologies under ML were further analyzed under a Bayesian framework using PhyloBayes MPI v1.7 and the CAT-Poisson+Γ4 model (Lartillot, Lepage, & Blanquart, 2009; Lartillot & Philippe, 2004). This model allows for a very large number of classes to account for compositional heterogeneity across sites and, unlike in the more complex CAT-GTR+Γ4 model, also allows for convergence to be more easily achieved between MCMC chains. PhyloBayes MCMC chains were run for more than 10,000 cycles until convergence between the chains was achieved and the largest discrepancy (i.e., maxdiff parameter) was < 0.1. A consensus tree was generated from two PhyloBayes MCMC chains using a burn-in of 500 trees and sub-sampling every 10 trees.

Phylogenetic analyses of recoded datasets into four-character state alphabets were analyzed using IQ-TREE v1.5.5 and the model GTR+ES60S4+F+R6. ES60S4 is an adaptation of the dataset-specific empirical mixture model ES60 to four-character states. It is obtained by adding the frequencies of the amino acids that belong to each bin in the dataset-specific four-character state scheme S4 (see Data Recoding for details). Phylogenetic analyses of recoded datasets into six-character state alphabets were analyzed using PhyloBayes MPI v1.7 and the CAT-Poisson+Γ4 model. Maximum-likelihood analyses with a six-state recoding scheme could not be performed because IQ-TREE currently only supports amino acid datasets recoded into four-character states.

### 4.2.7. Other analyses

The 16S rRNA genes of '*Candidatus* Finniella inopinata', and the presumed endosymbionts of *Peranema trichophorum* and *Stachyamoeba lipophora* were identified with RNAmmer 1.2 server and BLAST searches. A set of 16S rRNA genes for diverse rickettsialeans and holosporaleans, and other alphaproteobacteria as outgroup, were retrieved from NCBI GenBank. The selection was based on Hess et al., (2016), Szokoli

et al., (2016) and Wang and Wu (2015). Environmental sequences for uncultured and undescribed rickettsialeans were retrieved by keeping the 50 best hits resulting from a BLAST search of our three novel 16S rRNA genes against the NCBI GenBank non-redundant (nr) database. The sequences were aligned with the SILVA aligner SINA v1.2.11 and all-gap sites were later removed. Phylogenetic analyses on this alignment were performed on IQ-TREE v1.5.5 using the GTR+F+R8 model.

A UPGMA (average-linkage) clustering of amino acid compositions based on the 200 gene set for the *Alphaproteobacteria* was built in MEGA 7 (Kumar, Stecher, & Tamura, 2016) from a matrix of Euclidean distances between amino acid compositions of sequences exported from the phylogenetic software P4 (http://p4.nhm.ac.uk/index.html).

## 4.3. RESULTS

### 4.3.1. The genomes and phylogenetic positions of three novel endosymbiotic alphaproteobacteria (*Rickettsiales* and *Holosporales*)

We sequenced the genomes of the novel holosporalean '*Candidatus* Finniella inopinata', an endosymbiont of the rhizarian amoeboflagellate *Viridiraptor invadens* (Hess et al., 2015), and two undescribed rickettsialeans, one associated with the heterolobosean amoeba *Stachyamoeba lipophora* and the other with the euglenoid flagellate *Peranema trichophorum*. The three genomes are small with a reduced gene number and high A+T% content, strongly suggesting an endosymbiotic lifestyle (Table 4.1). Comparisons of their rRNA genes show that these genomes are truly novel, being considerably divergent from other described alphaproteobacteria. As of February 2018, the closest 16S rRNA gene to that of the *Stachyamoeba*-associated rickettsialean belongs to *Rickettsia massiliae* str. AZT80, with only 88% identity. On the other hand, the closest 16S rRNA gene to that of the *Peranema*-associated rickettsialean belongs to an endosymbiont of *Acanthamoeba* sp. UWC8, which is only 92% identical. Phylogenetic analysis of both the 16S rRNA gene and the 200-gene set confirm that each species belongs to different families and orders within the *Alphaproteobacteria* (Fig. S4.4 and S4.5). '*Candidatus* Finniella inopinata' belongs to the recently described '*Candidatus* Paracaedibacteraceae' in the *Holosporales* (Hess et al., 2015), whereas the *Stachyamoeba*-associated rickettsialean belongs to the *Rickettsiaceae*, and the

*Peranema*-associated rickettsialean belongs to the '*Candidatus* Midichloriaceae', in the *Rickettsiales*.

**Table 4.1**. Genome features for the three novel rickettsialeans sequenced in this study.

| Species | *'Candidatus* Finniella inopinata' | *Stachyamoeba*-associated rickettsialean | *Peranema*-associated rickettsialean |
|---|---|---|---|
| Genome size | 1,792,168 bp | 1,738,386 bp | 1,375,759 bp |
| N50 | 174,737 bp | 1,738,386 bp | 28,559 bp |
| Contig number | 28 | 1 | 125 |
| Gene number[1] | 1,741 | 1,588 | 1,223 |
| A+T% content | 56.58% | 67.01% | 59.13% |
| Family | *Paracaedibacteraeae* | *Rickettsiaceae* | *'Candidatus* Midicloriaceae' |
| Order | *Holosporales* | *Rickettsiales* | *Rickettsiales* |
| Completeness[2] | 94.96% | 97.12% (=100%) | 92.08% |
| Redundancy[2] | 0.0% | 0.0% | 2.1% |

[1] as predicted by prokka v.1.13 (rRNA genes were searched with BAST).

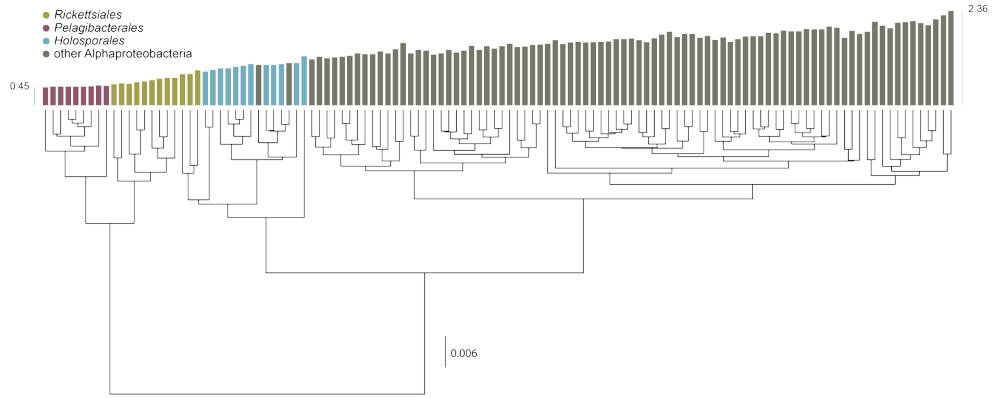
[2] as estimated by anvi'o v.2.4.0 using the Campbell et al., 2013 gene marker set.

### 4.3.2. Compositional heterogeneity appears to be a major confounding factor affecting phylogenetic inference of the *Alphaproteobacteria*

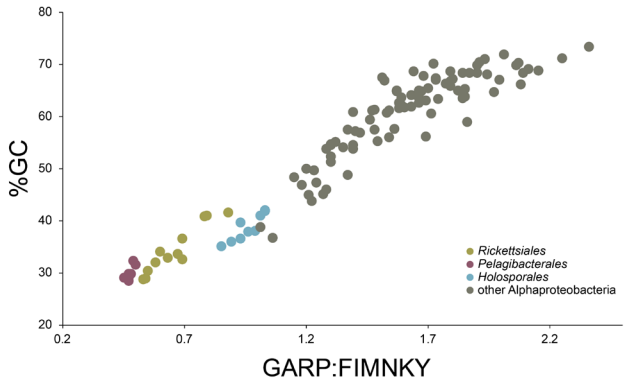
The average-linkage clustering of amino acid compositions shows that the *Rickettsiales*, *Pelagibacterales* and *Holosporales* are clearly distinct from other alphaproteobacteria. This indicates that these three taxa have divergent proteome amino acid compositions (Fig. 4.1A). These taxa also have the lowest GARP:FIMNKY ratios in all the *Alphaproteobacteria* (Fig. 4.1A); the *Pelagibacterales* being the most divergent, followed by the *Rickettsiales* and then the *Holosporales*. Such biased amino acid compositions appear to be the consequence of genome nucleotide compositions that are strongly biased towards high A+T%—a scatter plot of genome G+C% and proteome GARP:FIMNKY ratios shows a similar clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales* (Fig. 4.1B). This compositional similarity in the proteomes of the *Rickettsiales*, *Pelagibacterales* and *Holosporales*, which also turn out to be the longest-branching alphaproteobacterial groups in previously published phylogenies (e.g., Wang & Wu, 2015), could be the outcome of either a shared evolutionary history (i.e., the groups are most closely related to one another), or alternatively, evolutionary convergence (e.g., because of similar lifestyles or evolutionary trends toward small cell and genome sizes).

**Figure 4.1**. Compositional heterogeneity in the *Alphaproteobacteria* is a major factor that confounds phylogenetic inference. There are great disparities between the genome G+C% content and amino acid compositions of the *Rickettsiales*, *Pelagibacterales* and *Holosporales* and those of all other alphaproteobacteria. **A.** A UPGMA (average-linkage) clustering of amino acid compositions (based on the 200 gene set for the *Alphaproteobacteria*) shows that the *Rickettsiales* (brown), *Pelagibacterales* (gray), and *Holosporales* (light blue) all have very similar proteome amino acid compositions. At the tips of the tree, GARP:FIMNKY ratio values are shown as bars. **B.** A scatterplot depicting the strong correlation between G+C% (nucleotide compositions) and GARP:FIMNKY ratios (amino acid composition) for the 120 taxa in the *Alphaproteobacteria* shows a similar clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales*.

A.


Rickettsiales
Pelagibacterales
Holosporales
other Alphaproteobacteria
0.45
2.36
0.006

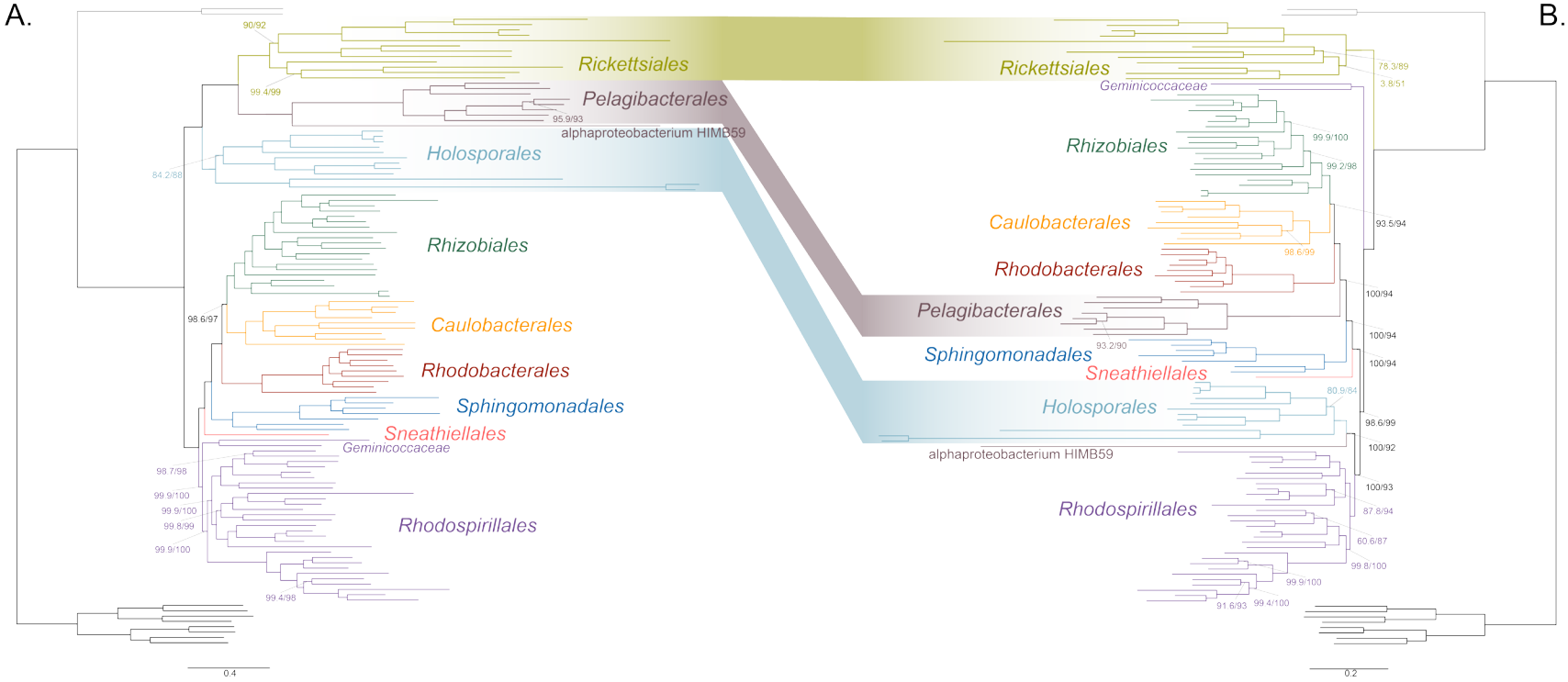
B.


%GC
GARP:FIMNKY
80
70
60
50
40
30
20
0.2
0.7
1.2
1.7
2.2
Rickettsiales
Pelagibacterales
Holosporales
other Alphaproteobacteria

As a first step to discriminate between these two alternatives, we tried the site-heterogenous model LG+PMSF(ES60)+F+R6 on our 200-gene dataset for the *Alphaproteobacteria*. The resulting tree united the *Rickettsiales*, *Pelagibacterales* and *Holosporales* in a fully supported clade (Fig. 4.2A). However, the clustering of these three groups is suggestive of a phylogenetic artefact (e.g., long-branch attraction or LBA); indeed, such a pattern resembles the one seen in the tree of proteome amino acid compositions (see Fig. 4.1A). This is because the three groups have the longest branches in the *Alphaproteobacteria* tree and have compositionally biased and fast-evolving genomes (see Fig. 4.2). If evolutionary convergence in amino acid compositions is confounding phylogenetic inference for the *Alphaproteobacteria*, methods aimed at reducing compositional heterogeneity might disrupt the clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales*.

To further test whether the clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales* is real or artefactual we tried different strategies to reduce the compositional heterogeneity of our dataset (see Fig. S4.2 for the diverse strategies employed). When removing the 50% most compositionally biased (heterogeneous) sites according to ʐ, the clustering between the *Rickettsiales*, *Pelagibacterales* and *Holosporales* is disrupted (Fig. 4.2B). The new more derived placements for the *Pelagibacterales* and *Holosporales* are well supported (further described below), and support tends to increase as compositionally biased sites are removed (Fig. S4.8). Furthermore, when each of these long-branching taxa is analyzed in isolation (i.e., in the absence of the other two), and compositional heterogeneity is decreased, new phylogenetic patterns emerge that are incompatible, or in conflict, with their clustering (Fig. S4.10-S4.13). In other words, removing the most compositionally biased sites, recoding the data into reduced character-state alphabets to minimize compositional bias, or using only the most compositionally homogeneous genes, converge to very similar phylogenetic patterns for the *Alphaproteobacteria* that are however incompatible with the clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales* (e.g., Fig. S4.10-S4.13, and 3). On the other hand, removing fast-evolving sites does not disrupt the clustering of these three long-branching groups (Fig. S4.7), suggesting that high evolutionary rates per site are not a major confounding factor when inferring the phylogeny of the *Alphaproteobacteria*.

**Figure 4.2.** Decreasing compositional heterogeneity by removing compositionally-biased sites disrupts the clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales*. All branch support values are 100% SH-aLRT and 100% UFBoot unless annotated. **A.** A maximum-likelihood tree inferred under the LG+PMSF(ES60)+F+R6 model and from the intact dataset which is highly compositionally heterogeneous. The three long-branching orders, the *Rickettsiales Pelagibacterales* and *Holosporales*, that have similar amino acid compositions form a clade. **B.** A maximum-likelihood tree inferred under the LG+PMSF(ES60)+F+R6 model and from a dataset whose compositional heterogeneity has been decreased by removing 50% of the most biased sites according to $\chi$. In this phylogeny the clustering of the *Rickettsiales*, *Pelagibacterales* and *Holosporales* is disrupted. The *Pelagibacterales* is sister to the *Rhodobacterales*, *Caulobacterales* and *Rhizobiales*. The *Holosporales* becomes sister to the *Rhodospirillales*. The *Rickettsiales* retains its basal position as sister to the *Caulobacteridae*. See Fig. S4.6 for the Bayesian consensus trees inferred in PhyloBayes MPI v1.7 under the CAT-Poisson+$\Gamma$4 model.


A.
B.
Rickettsiales
Pelagibacterales
alphaproteobacterium HIMB59
Holosporales
Rhizobiales
Caulobacterales
Rhodobacterales
Sphingomonadales
Sneathiellales
Geminicoccaceae
Rhodospirillales
90/92
99.4/99
95.9/93
84.2/88
98.6/97
98.7/98
99.9/100
99.9/100
99.8/99
99.9/100
99.4/98
0.4
78.3/89
3.8/51
99.9/100
99.2/98
93.5/94
98.6/99
100/94
100/94
100/94
93.2/90
80.9/84
98.6/99
100/92
100/93
87.8/94
60.6/87
99.8/100
99.9/100
91.6/93
99.4/100
0.2

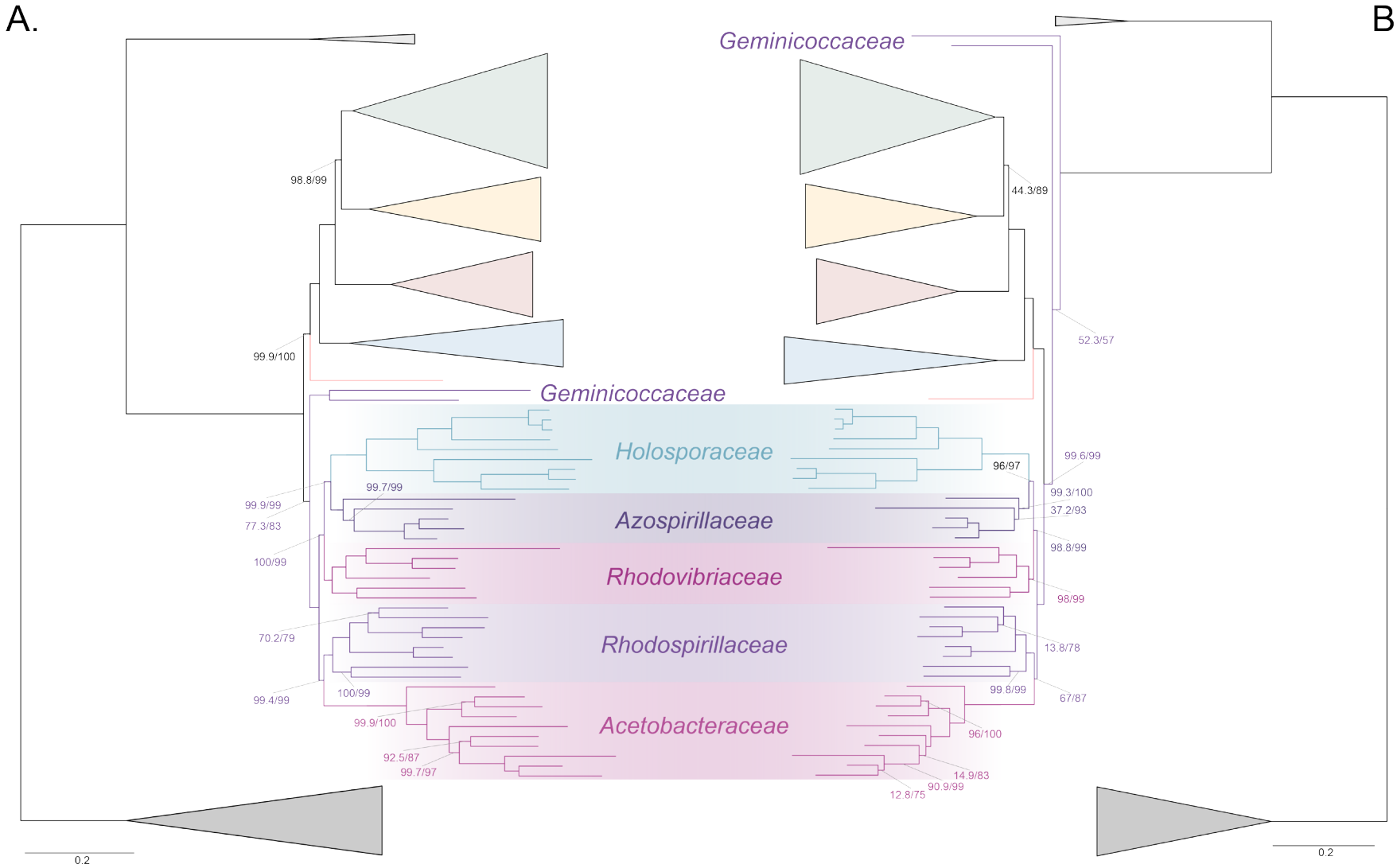
### 4.3.3. The *Holosporales* is unrelated to the *Rickettsiales* and is instead most likely derived within the *Rhodospirillales*

The *Holosporales* has traditionally been considered part of the *Rickettsiales sensu lato* because it appears as sister to the *Rickettsiales* in many trees (e.g., Hess et al., 2015; Montagna et al., 2013; Santos & Massard, 2014). It is exclusively composed of endosymbiotic bacteria living within diverse eukaryotes, and such a lifestyle is shared with all other members of the *Rickettsiales*. When we decrease, and then account for compositional heterogeneity, we recover tree topologies in which the *Holosporales* moves away from the *Rickettsiales* (e.g., Fig. 4.2B and S4.13A). For example, the *Holosporales* becomes sister to all free-living alphaproteobacteria (the *Caulobacteridae*) when only the 40 most homogeneous genes are used (Fig. S4.13A) or when 10% of the most compositionally-biased sites are removed (Fig. S4.8). When compositional heterogeneity is further decreased by removing 50% of the most compositionally-biased sites, the *Holosporales* becomes sister to the *Rhodospirillales* (Fig. 4.2B and S4.8; and see also Fig. S4.10B and S4.13B).

Similarly, when the long-branching *Rickettsiales* and *Pelagibacterales* groups (plus the extremely long-branching genera *Holospora* and '*Candidatus* Hepatobacter') are removed, after compositional heterogeneity had been decreased through site removal, the *Holosporales* move to a much more derived position well within the *Rhodospirillales* (Fig. 4.3A). If the very compositionally-biased and fast-evolving *Holospora* and '*Candidatus* Hepatobacter' are left in, the *Holosporales* is pulled away from this derived position and the whole clade moves closer to the base of the tree (data not shown). The same behaviour is seen when these same taxa are removed, and the data are then recoded into four- or six-character states (Fig. 4.3B and S4.14). Specifically, the *Holosporales* now consistently branches as sister to a subgroup of rhodospirillaleans that includes, among others, the epibiotic predator *Micavibrio aeruginosavorus* and the purple nonsulfur bacterium *Rhodocista centenaria* (the *Azospirillaceae*, see below). This new placement of the *Holosporales* has nearly full support under both maximum likelihood and Bayesian inference (e.g., >95% UFBoot; see Fig. 4.3).Thus, three different analyses independently converge to the same pattern and support a derived origin of the *Holosporales* within the *Rhodospirillales*: (1) removal of compositionally biased sites (Fig. 4.3A), (2) data recoding into four-character states using the dataset-specific scheme S4 (Fig. 4.3B), and (3) data recoding into six-character states using the

dataset-specific scheme S6 (Fig. S4.14); each of these strategies had to be combined with the removal of the *Pelagibacterales* and *Rickettsiales* to recover this phylogenetic position for the *Holosporales*.

**Figure 4.3.** The *Holosporales* branches in a derived position within the *Rhodospirillales* when compositional heterogeneity is reduced and the long-branching *Rickettsiales* and *Pelagibacterales* are removed. Branch support values are 100% SH-aLRT and 100% UFBoot unless annotated. **A.** A maximum-likelihood tree, inferred under the LG+PMSF(ES60)+F+R6 model, to place the *Holosporales* in the absence of the *Rickettsiales* and the *Pelagibacterales* and when compositional heterogeneity has been decreased by removing 50% of the most biased sites. The *Holosporales* is sister to the *Azospirillaceae* fam. nov. within the *Rhodospirillales*. **B.** A maximum-likelihood tree, inferred under the GTR+ES60S4+F+R6 model, to place the *Holosporales* in the absence of the *Rickettsiales* and the *Pelagibacterales* and when the data have been recoded into a four-character state alphabet (the dataset-specific recoding scheme S4: ARNDQEILKSTV GHY CMFP W) to reduce compositional heterogeneity. This phylogeny shows a pattern that matches that inferred when compositional heterogeneity has been alleviated through site removal. See Fig. S4.9 for the Bayesian consensus trees inferred in PhyloBayes MPI v1.7 and under the CAT-Poisson+Γ4 model.


A.
B.
98.8/99
99.9/100
Geminicoccaceae
Holosporaceae
Azospirillaceae
Rhodovibriaceae
Rhodospirillaceae
Acetobacteraceae
99.9/99
77.3/83
99.7/99
100/99
70.2/79
99.4/99
100/99
99.9/100
92.5/87
99.7/97
0.2
Geminicoccaceae
44.3/89
52.3/57
96/97
99.6/99
99.3/100
37.2/93
98.8/99
98/99
13.8/78
99.8/99
67/87
96/100
14.9/83
90.9/99
12.8/75
0.2

A fourth independent analysis further supports a derived placement of the *Holosporales* nested within the *Rhodospirillales*. Bayesian inference using the CAT-Poisson+Γ4 model, on a dataset whose compositional heterogeneity had been decreased by removing 50% of the most compositionally biased sites but for which no taxa had been removed, also recovered the *Holosporales* as sister to the *Azospirillaceae* (see Fig. S4.6).

### 4.3.4. The *Rhodospirillales* is a diverse order and comprises five well-supported families

The *Rhodospirillales* is an ancient and highly diversified group, but unfortunately this is rarely obvious from published phylogenies because most studies only include a few species for this order (Ferla et al., 2013; Georgiades et al., 2011; Williams et al., 2007). We have included a total of 31 *Rhodospirillales* taxa to better cover its diversity. Such broad sampling reveals trees with five clear subgroups within the *Rhodospirillales* that are well-supported in most of our analyses (e.g., Fig. 4.2B and 4.3). First is the *Acetobacteraceae* which comprises acetic acid, acidophilic, and photosynthesizing (bacteriochlorophyll-containing) bacteria. The *Acetobacteraceae* is strongly supported and relatively divergent from all other families within the *Rhodospirillales*. Sister to the *Acetobacteraceae* is another subgroup that comprises many photosynthesizing bacteria, including the type species for the *Rhodospirillales*, *Rhodospirillum rubrum*, as well as the magnetotactic bacterial genera *Magnetospirillum*, *Magnetovibrio* and *Magnetospira* (Fig. 4.3). This subgroup best corresponds to the poorly defined and paraphyletic *Rhodospirillaceae* family. We amend the *Rhodospirillaceae* taxon and restrict it to the clade most closely related to the *Acetobacteraceae*. As described above, when artefacts are accounted for, the *Holosporales* most likely branches within the *Rhodospirillales* and therefore we suggest that the *Holosporales sensu* Szokoli *et al*. (2016) be lowered in rank to the family *Holosporaceae*, which is sister to the *Azospirillaceae* (Fig. 4.3). The *Azospirillaceae* **fam. nov.** contains the purple bacterium *Rhodocista centenaria* and the epibiotic (neither periplasmic nor intracellular) predator *Micavibrio aeruginosavorus*, among others (e.g., see Fig. S4.5). The *Holosporaceae* and the *Azospirillaceae* clades appear to be sister to the *Rhodovibriaceae* **fam. nov.** (Fig. 4.3), a well-supported group that comprises the purple nonsulfur bacterium *Rhodovibrio salinarum*, the aerobic heterotroph *Kiloniella laminariae*, and the marine bacterioplankter '*Candidatus* Puniceispirillum marinum' (or the SAR116 clade). Each of these subgroups and their

interrelationships—with the exception of the *Holosporaceae* that branches within the *Rhodospirillales* only after compositional heterogeneity is countered—are strongly supported in nearly all of our analyses (e.g., see Fig. 4.2B and 4.3).

### 4.3.5. The *Geminicoccaceae* might be basal to all other free-living alphaproteobacteria (the *Caulobacteridae*)

The *Geminicocacceae* is a recently proposed family within the *Rhodospirillales* (Proença et al., 2017). It is currently represented by only two genera, *Geminicoccus* and *Arboriscoccus* (Foesel, Gössner, Drake, & Schramm, 2007; Proença et al., 2017). In most of our trees, however, *Tistrella mobilis* is often sister to *Geminicoccus roseus* with full statistical support (e.g., Fig. 4.2B and 4.3A, but see Fig. S4.9 for an exception) and we therefore consider it to be part of the *Geminococcaceae*. Interestingly, the *Geminicoccaceae* tends to have two alternative stable positions in our analyses, either as sister to all other families of the *Rhodospirillales* (e.g., Fig. 4.2A and 4.3A), or alternatively, as sister to all other orders of the *Caulobacteridae* (i.e., representing the most basal lineage of free-living alphaproteobacteria; Fig. 4.2B and 4.3B, or Fig. S4.11C and S4.12C). Our analyses designed to alleviate compositional heterogeneity, specifically site removal and recoding (without taxon removal), favor the latter position for the *Geminicoccaceae* (Fig. 4.2B and 4.3B). Moreover, as compositionally-biased sites are progressively removed, support for the affiliation of the *Geminicoccaceae* with the *Rhodospirillales* decreases, and after 50% of the sites have been removed, the *Geminicoccaceae* emerges as sister to all other free-living alphaproteobacteria with strong support (>95% UFBoot; Fig. S4.8). In further agreement with this trend, the much simpler model LG4X places the *Geminicocacceae* in a derived position as sister to the *Acetobacteraceae* (data not shown), but as model complexity increases, and compositional heterogeneity is reduced, the *Geminicoccaceae* moves closer to the base of the *Alphaproteobacteria* (Fig. 4.2A and 4.3A). Such a placement suggests that the *Geminicoccaceae* may be a novel and independent order-level lineage in the *Alphaproteobacteria*. However, because of the uncertainty in our results we opt here for conservatively keeping the *Geminicoccaceae* as the sixth family of the *Rhodospirillales* (Fig. 4.3A).

### 4.3.6. Other deep relationships in the *Alphaproteobacteria* (*Pelagibacterales*, *Rickettsiales*, alphaproteobacterium sp. HIMB59)

The clustering of the *Pelagibacterales* (formerly the SAR11 clade) with the *Rickettsiales* and *Holosporales* is more easily disrupted than that of the *Holosporales*. The removal of compositionally-biased sites (from 30% on; 16,320 out of 54,400 sites; Fig. S4.8), data recoding into four-character states (Fig. S4.12D), and a set of the most compositionally homogeneous genes (Fig. S4.13D) all support a derived placement of the *Pelagibacterales* as sister to the *Rhodobacterales*, *Caulobacterales* and *Rhizobiales* (in agreement with Martijn et al., 2018; Rodríguez-Ezpeleta & Embley, 2012; Viklund et al., 2012, 2013). The *Caulobacterales* is sister to the *Rhizobiales*, and the *Rhodobacterales* sister to both (e.g., Fig. 4.2B and 4.3). This is consistent throughout most of our results and such interrelationships become very robustly supported as compositional heterogeneity is increasingly alleviated (Fig. S4.8). The placement of the *Rickettsiales* as sister to the *Caulobacteridae* (i.e., all other alphaproteobacteria) remains stable across different analyses (Fig. 4.2B, S4.10C, S4.11C, S4.12C and S4.13D); this is also true when the other long-branching taxa, the *Pelagibacterales* and *Holosporales*, are removed. Yet, the interrelationships inside the *Rickettsiales* order remain uncertain; the '*Candidatus* Midichloriaceae' becomes sister to the *Anaplasmataceae* when fast sites are removed (Fig. S4.7), but to the *Rickettsiaceae* when compositionally-biased sites are removed (Fig. S4.8). The placement of alphaproteobacterium sp. HIMB59 is entirely uncertain (e.g., see Fig. 4.2B and S4.11E); it does not have a stable position in the *Alphaproteobacteria* across our diverse analyses. This is consistent with previous reports that suggest that alphaproteobacterium sp. HIMB59 is not closely related to the *Pelagibacterales* (Martijn et al., 2018; Viklund et al., 2013).

## 4.4. DISCUSSION

We have employed a diverse set of strategies to investigate the phylogenetic signal contained within 200 genes for the *Alphaproteobacteria*. Specifically, such strategies were primarily aimed at reducing amino acid compositional heterogeneity among taxa—a phenomenon that permeates our dataset (Fig. 4.1). Compositional heterogeneity is a clear violation of the phylogenetic models used in our, and previous, analyses, and known to cause phylogenetic artefacts (Foster, 2004). In the absence of more sophisticated models for inferring deep phylogeny, the only way to counter artefacts

caused by compositional heterogeneity is by removing compositionally-biased sites or taxa, or recoding amino acids into reduced alphabets. A combination of these strategies reveals that the *Rickettsiales sensu lato* (i.e., the *Rickettsiales* and *Holosporales*) is polyphyletic. Our analyses suggest that the *Holosporales* is derived within the *Rhodospirillales*, and that therefore this taxon should be lowered in rank and renamed the *Holosporaceae* family (see Fig. 4.2B and 4.3). The same methods suggest that the *Rhodospirillales* might indeed be a paraphyletic order and that the *Geminicoccaceae* could be a separate lineage that is sister to the *Caulobacteridae* (e.g., Fig. 4.2B). These two results, combined with our broader sampling, reorganize the internal phylogenetic structure of the *Rhodospirillales* and show that its diversity can be grouped into at least five well-supported major families (Fig. 4.3).

In 16S rRNA gene trees, the *Holosporales* has most often been allied to the *Rickettsiales* (Hess et al., 2015; Montagna et al., 2013). The apparent diversity of this group has quickly increased in recent years as more and more intracellular bacteria living within protists have been described. An endosymbiotic lifestyle is shared by all members of the *Holosporales* and is also shared with all those that belong to the *Rickettsiales*. Thus, it had been reasonable to accept their shared ancestry as suggested by some 16S rRNA gene trees (e.g., Hess et al., 2015; Montagna et al., 2013; Santos & Massard, 2014). Apparent strong support for the monophyly of the *Rickettsiales* and the *Holosporales* recently came from some multi-gene trees by Wang and Wu (2014, 2015) who expanded sampling for the *Holosporales*. However, an alternative placement for the *Holosporales* as sister to the *Caulobacteridae* has been reported by Ferla et al. (2013) based on rRNA genes, by Georgiades et al., (2011) based on 65 genes, by Schulz et al., (2015) based on 139 genes, as well as by Wang and Wu (2015) based on 26, 29, or 200 genes (see the supplementary information in Wang and Wu, 2015). This placement was acknowledged by Szokoli et al. (2015), who formally established the order *Holosporales*. Most recently, Martijn *et al*., (2018), who used strategies to reduce compositional heterogeneity, and similarly to Wang and Wu (2015), recovered a number of placements for the *Holosporales* within the *Alphaproteobacteria*; however these different placements for the *Holosporales* were poorly supported. Here we provide strong evidence for the hypothesis that the *Holosporales* is not related to the *Rickettsiales*, as suggested earlier (Ferla et al., 2013; Georgiades et al., 2011; Szokoli et al., 2016). The *Rickettsiales sensu lato* is polyphyletic. We show that the *Holosporales* is artefactually attracted to the *Rickettsiales*

(e.g., Fig. 4.2A), but as compositional bias is increasingly alleviated (through site removal and recoding), they move further away from them (Fig. 4.2B). The *Holosporales* is placed within the *Rhodopirillales* as sister to the family *Azospirillaceae* (Fig. 4.3). The similar lifestyles of the *Holosporales* and *Rickettsiales*, as well as other features like the presence of an ATP/ADP translocase (Wang & Wu, 2014), are therefore likely the outcome of convergent evolution.

A derived origin of the *Holosporales* has important implications for understanding the origin of mitochondria and the nature of their ancestor. Wang and Wu (2014, 2015) proposed that mitochondria are phylogenetically embedded within the *Rickettsiales sensu lato*. In their trees, mitochondria were sister to a clade formed by the *Rickettsiaceae*, *Anaplasmataceae* and '*Candidatus* Midichloriaceae', and the *Holosporales* was itself sister to all of them. This phylogenetic placement for mitochondria suggested that the ancestor of mitochondria was an intracellular parasite (Wang and Wu, 2014). But if the *Holosporales* is a derived group of rhodospirillaleans as shown here (see Fig. 4.3), then the argument that mitochondria necessarily evolved from parasitic alphaproteobacteria no longer holds. While the sisterhood of mitochondria and the *Rickettsiales sensu stricto* is still a possibility, such a relationship does not imply that the two groups shared a parasitic common ancestor (i.e., a parasitic ancestry for mitochondria). The most recent analyses done by Martijn *et al*. (2018) suggest that mitochondria are sister to all known alphaproteobacteria, also suggesting their non-parasitic ancestry. This study, and that of Martijn et al., thus complement each other and support the view that mitochondria most likely evolved from ancestral free-living alphaproteobacteria (*contra* Sassera et al., 2011; Wang & Wu, 2014, 2015).

The order *Rhodospirillales* is quite diverse and includes many purple nonsulfur bacteria as well as all magnetotactic bacteria within the *Alphaproteobacteria*. The *Rhodospirillales* is sister to all other orders in the *Caulobacteridae*, and has historically been subdivided into two families: the *Rhodospirillaceae* and the *Acetobacteraceae*. Recently, a new family, the *Geminicoccaceae*, was established for the *Rhodospirillales* (Proença et al., 2017). However, some of our analyses suggest that the *Geminicoccaceae* might be sister to all other *Caulobacteridae* (e.g., Fig. 4.2B and 4.3B). This phylogenetic pattern, therefore, suggests that the *Rhodospirillales* may be a paraphyletic order. The placement of the *Geminicoccaceae* as sister to the *Caulobacteridae* needs to be further tested once more sequenced diversity for this group

becomes available; if it were to be confirmed, the *Geminicoccaceae* should be elevated to the order level. Whereas the *Acetobacteraceae* is phylogenetically well-defined, there has been considerable uncertainty about the *Rhodospirillaceae* (e.g., Ferla et al., 2013), primarily because of poor sampling and a lack of resolution provided by the 16S rRNA gene. We subdivide the *Rhodospirillaceae sensu lato* into three subgroups (Fig. 4.3). We restrict the *Rhodospirillaceae sensu stricto* to the subgroup that is sister to the *Acetobacteracae* (Fig. 4.3). The other two subgroups are the *Rhodovibriaceae* and the *Azospirillaceae*; the latter is sister to the *Holosporaceae* (Fig. 4.3).

Based on our fairly robust phylogenetic patterns, we have updated the higher-level taxonomy of the *Alphaproteobacteria* (Fig. 4.4). We exclude the *Magnetococcales* from the *Alphaproteobacteria* class because of its divergent nature (e.g., see Fig. 1 in Esser, Martin, & Dagan, 2007 which shows that many of *Magnetococcus*' genes are more similar to those of beta-, and gammaproteobacteria). In agreement with its intermediate phylogenetic placement, we propose the *Magnetococcia* **class. nov.** At the highest level we define the *Alphaproteobacteria* class as comprising two subclasses *sensu* Ferla et al. (2013), the *Rickettsidae* and the *Caulobacteridae*. The former contains the *Rickettsiales*, and the latter contains all other orders, which are primarily and ancestrally free-living alphaproteobacteria. The order *Rickettsiales* comprises three families as previously defined, the *Rickettsiaceae*, the *Anaplasmataceae*, and the '*Candidatus* Midichloriaceae'. On the other hand, the *Caulobacteridae* is composed of seven phylogenetically well-supported orders: the *Rhodospirillales*, *Sneathiellales*, *Sphingomonadales*, *Pelagibacterales*, *Rhodobacterales*, *Caulobacterales* and *Rhizobiales*. Among the many species claimed to represent new order-level lineages on the basis of 16S rRNA gene trees (Cho & Giovannoni, 2003; Harbison, Price, Flythe, & Bräuer, 2017; Kurahashi, Fukunaga, Harayama, & Yokota, 2008; Kwon, Lee, Yang, & Kim, 2005; Wiese, Thiel, Gärtner, Schmaljohann, & Imhoff, 2009), only *Sneathiella* deserves order-level status (e.g., see Fig. 4.2 and S4.5; Kurahashi et al., 2008), since all others have derived placements in our trees and those published by others (Bazylinski et al., 2013; Harbison et al., 2017; Venkata Ramana et al., 2013; Williams, Lefèvre, Zhao, Beveridge, & Bazylinski, 2012). The *Rhodospirillales* order comprises six families, three of which are new, namely the *Holosporaceae*, *Azospirillaceae* and *Rhodovibriaceae* (Fig. 4.4). This new higher-level classification of the *Alphaproteobacteria* updates and expands those presented by Ferla et al. (2013), the 'Bergey's Manual of Systematics of

Archaea and Bacteria' (Garrity et al., 2005; Whitman, 2015), and 'The Prokaryotes' (Rosenberg et al., 2014).

**Class 1. *Alphaproteobacteria*** Garrity et al. 2005

- **Subclass 1. *Rickettsidae*** Ferla et al. 2013 **emend.** Muñoz-Gómez et al. 2018
  - **Order 1. *Rickettsiales*** Gieszczykiewicz 1939 **emend.** Dumler et al. 2001
    - **Family 1. *Anaplasmataceae*** Philip 1957
    - **Family 2. '*Candidatus* Midichloriaceae'** Montagna et al. 2013
    - **Family 3. *Rickettsiaceae*** Pinkerton 1936
- **Subclass 2. *Caulobacteridae*** Ferla et al. 2013 **emend.** Muñoz-Gómez et al. 2018
  - **Order 1. *Rhodospirillales*** Pfennig and Truper 1971
    - **Family 1. *Acetobacteriacae*** ex Henrici 1939 Gillis and De Ley 1980
    - **Family 2. *Rhodospirillaceae*** Pfennig and Truper 1971 **emend.** Muñoz-Gómez et al. 2018
    - **Family 3. *Azospirillaceae* fam. nov.** Muñoz-Gómez et al. 2018
    - **Family 4. *Holosporaceae*** Szokoli et al. 2016
    - **Family 5. *Rhodovibriaceae* fam. nov.** Muñoz-Gómez et al. 2018
    - **Family 6. *Geminicoccaceae*** Proenca et al. 2017
  - **Order 2. *Sneathiellales*** Kurahashi et al. 2008
  - **Order 3. *Sphingomonadales*** Yabuuchi et al. 1990
    - **Family 1. *Kordiimonadaceae*** Xu et al. 2014
    - **Family 2. *Sphingomonadaceae*** Kosako et al. 2000
    - **Family 3. *Erythrobacteraceae*** Lee et al. 2005 emend. Xu et al. 2009
  - **Order 4. *Pelagibacterales*** Ferla et al. 2013 (invalidly published)
    - **Family 1. *Pelagibacteraceae*** Trash et al. 2011 (invalidly published)
  - **Order 5. *Rhodobacterales*** Garrity et al. 2006
    - **Family 1. *Rhodobacteraceae*** Garrity et al. 2006
  - **Order 6. *Caulobacterales*** Henrici and Johnson 1935
    - **Family 1. *Parvularculaceae*** Cavalcanti and Thompson 2014
    - **Family 2. *Hyphomonadaceae*** Lee et al. 2005 emend. Abraham and Rhode 2014
    - **Family 3. *Caulobacteraceae*** Henrici and Johnson 1935
  - **Order 7. *Rhizobiales*** Kuykendall 2006
    - **Family *Bartonellaceae*** Gieszczykiewicz 1939
    - **Family *Beijerinckiaceae*** Garrity et al. 2006
    - **Family *Bradyrhizobiaceae*** Garrity et al. 2006
    - **Family *Brucellaceae*** Breed et al. 1957
    - **Family *Methylobacteriaceae*** Garrity et al. 2006
    - **Family *Methylocystaceae*** Bowman 2006
    - **Family *Rhizobiaceae*** Conn 1938
    - **Family *Hyphomicrobiaceae*** Babudieri 1950
    - **Family *Phyllobacteriaceae*** Mergaert and Swings 2006
    - **Family *Rhodobiaceae*** Garrity et al. 2006
    - **Family *Cohaesibacteraceae*** Hwang and Cho 2008 emend. Gallego et al. 2010
    - **Family *Xanthobacteraceae*** Lee et al. 2005

**Class 2. *Magnetococcia* class. nov.** Muñoz-Gómez et al. 2018

- **Order 1. *Magnetococcales*** Bazylinki et al. 2013
  - **Family 1. *Magnetococcaceae*** Bazylinki et al. 2013

**Figure 4.4.** A proposed higher-level classification scheme for the *Alphaproteobacteria* and the *Magnetococcia* class. nov.

## 4.5. CONCLUSIONS

A combination of methods to decrease compositional heterogeneity has been necessary to disrupt artefacts that arise when inferring the phylogeny of the *Alphaproteobacteria*. This is an example of the complex nature of the historical signal contained in modern genomes and the limitations of our current evolutionary models to capture these signals. A robust phylogeny of the *Alphaproteobacteria* is a precondition for placing the mitochondrial lineage. This is because including mitochondria certainly exacerbates the already strong biases in the data, and therefore represents additional sources for artefact in phylogeny inference (as seen in Wang and Wu, 2015 where the *Holosporales* is attracted by both mitochondria and the *Rickettsiales*). Future endeavors should take into account the new phylogenetic framework developed here. The incorporation of not-yet-cultivated 'environmental' diversity recovered from metagenomes will certainly expand the known diversity of the *Alphaproteobacteria* and improve phylogenetic inference.

# CHAPTER 5: THE PHYLOGENETIC ROOTS OF MITOCHONDRIA

## 5.1. INTRODUCTION

The nature of the first mitochondrial ancestor is of great importance to our understanding of the symbiotic basis (or the proximate selective force) that first established mitochondria inside their host cells. We can gain insight into this question by reconstructing the last common ancestor of alphaproteobacteria and the mitochondrial lineage (e.g., see Wang & Wu, 2015). But to be able to do this we need to confidently place the mitochondrial lineage in a phylogenetic context. Unfortunately, even though we have known for many years that mitochondria have a clear alphaproteobacterial affiliation (e.g., (Andersson et al., 1998; Dayhoff & Schwartz, 1981; Wang & Wu, 2015; Williams, Sobral, & Dickerman, 2007; Yang et al., 1985), it has been more elusive to find a precise alphaproteobacterial lineage that is reliably most closely related to mitochondria.

The search for the closest bacterial relatives to mitochondria has been done using different kinds of data, methods and rationales. Early comparative physiology pointed to aerobic bacteria as the progenitors of (classical aerobic) mitochondria (Margulis, 1970; Sagan, 1967). Detailed comparative biochemistry later suggested that purple nonsulfur bacteria and their non-photosynthetic relatives, or what we call today *Alphaproteobacteria*, had core respiratory chains more similar to those of mitochondria (e.g., sensitive to antimycin and rotenone) (John & Whatley, 1978; John & Whatley, 1975). Later on, comparative analyses of wheat mitochondrial SSU rRNA oligonucleotides supported the idea that mitochondria have an endosymbiotic origin from prokaryotes (Bonen, Cunningham, Gray, & Doolittle, 1977). Indeed, the first phylogenetic trees made from gene sequences (nucleus-encoded mitochondrial cytochrome *c*) showed that mitochondria are more related to purple nonsulfur bacteria (Dayhoff & Schwartz, 1981; Schwartz & Dayhoff, 1978). The same result was later confirmed by the phylogenetic analysis of the first full-length SSU rRNA sequences from many mitochondria (Yang et al., 1985). The availability of gene sequences for both mitochondria and alphaproteobacteria was too poor at the time, though, to determine more precise phylogenetic affinities. More and more sequence data for single genes later accumulated and new trees eventually suggested a closer affinity between

intracellular parasitic alphaproteobacteria like *Rickettsia* (order *Rickettsiales*) and mitochondria (Gupta, 1995; Olsen, Woese, & Overbeek, 1994; Viale & Arakaki, 1994).

The phylogenetic relationship between mitochondria and the *Rickettsiales* has also been recovered using concatenations of many genes from complete genomes (Andersson et al., 1998; Fitzpatrick, Creevey, & McInerney, 2006; Sassera et al., 2011; Williams et al., 2007; Wu et al., 2004). Estimating trees from multi-gene datasets aims to make statistically sound inferences by increasing the amount of data analyzed. When genomes from the order *Pelagibacterales* (also known as SAR11; some of the most abundant oceanic bacteria) were incorporated into phylogenetic analyses, they branched either as sister to mitochondria and the *Rickettsiales* (Williams et al., 2007), or as the closest sisters to mitochondria (Georgiades et al., 2011; Thrash et al., 2011). But later analyses refuted these phylogenetic hypotheses as artefacts caused by the convergent similarities observed in the amino acid composition of proteins encoded by mitochondrial genomes and the streamlined genomes of pelagibacteraleans (see Chapter 4; Brindefalk, Ettema, Viklund, Thollesson, & Andersson, 2011; Rodríguez-Ezpeleta & Embley, 2012; Viklund, Ettema, & Andersson, 2012; Viklund, Martijn, Ettema, & Andersson, 2013). Some recent phylogenetic analyses have relied on both mitochondrion-encoded, and nucleus-encoded mitochondrial genes (Wang & Wu, 2014, 2015a). These analyses, which have also expanded taxon sampling by including novel (and comparatively slow-evolving) genomes from amoeba endosymbionts (family *Holosporaceae*), suggest that mitochondria are embedded within the *Rickettsiales*, being sister to the *Rickettsiaceae*, *Anasplamataceae* and *Midichloriaceae* (Wang & Wu, 2014, 2015a), and are in line with earlier suggestions (Emelyanov, 2001a, 2001b).

The genomes of mitochondria, pelagibacteraleans and rickettsialeans have evolved similarly, even though they might not be phylogenetically related. All three have evolved fast and in a reductive fashion by losing many genes, and thus have ended up being heavily compositionally biased (in their genes and proteins, e.g., A+T% rich) relative to all other (alpha)proteobacteria (Rodríguez-Ezpeleta & Embley, 2012). This has led to many concerns about their possible artefactual phylogenetic attraction in trees (a.k.a., long-branch attraction). When the model of protein evolution cannot accurately explain the complex evolutionary history of the data, the wrong topologies might end up being strongly supported (this is a systematic error). This means that the consistent affiliation of mitochondria to rickettsialeans in multi-gene trees could be the outcome of overall

convergent evolution in the nucleotide or amino acid composition of genes and proteins, and not of real history.

In an attempt to overcome the potential pitfalls of concatenated multi-gene tree inference, 'phylome' (Abhishek, Bavishi, Bavishi, & Choudhary, 2011; Atteia et al., 2009; Esser et al., 2004; Georgiades & Raoult, 2011; Thiergart, Landan, Schenk, Dagan, & Martin, 2012) or network (Esser et al., 2004; M. Wu et al., 2004) analyses have been used to investigate the phylogenetic signal across mitochondrial genes. The 'phylome' method relies on analyzing mitochondrial genes (both mitochondrion-, and nucleus-encoded) individually (rather than together) and then looking at what alphaproteobacteria are the closest (most related, e.g., Abhishek et al., 2011; Atteia et al., 2009; Esser et al., 2004; Georgiades & Raoult, 2011) or less distant (most similar in trees, e.g., Thiergart et al., 2012 or simple BLAST searches, e.g., Abhishek et al., 2011) to mitochondria for every gene. Network analyses have been performed on concatenated multi-gene datasets, but instead of producing a strictly branching tree, they output a network that displays the degree of incongruence or non-tree signal in the dataset. These approaches have suggested that there is a conflicting signal among mitochondrial genes. For example, it was shown that many mitochondrial genes tend to be phylogenetically closer or are more similar to genes of *Rhodospirillum* (*Rhodospirillales*) (Thiergart et al., 2012) or *Ochrobactrum* (*Rhizobiales*) (Abhishek et al., 2011), rather than to rickettsialean genes. These results in the form of a mixed (or weak) signal could be explained by: (a) erased signal (ancient origin), (b) mosaic evolution (unequal, including fast, rates of gene evolution), and (c) a chimeric ancestry (alphaproteobacteria have genes of disparate ancestries because they have and still exchange genes laterally). Unfortunately, even though these methods serve as exploratory tools to investigate the nature of the data at hand, they not only suffer from systematic errors (model misspecification), but they are also particularly prone to stochastic errors (limited data).

Multi-gene phylogenetic analyses have become the most popular method to infer phylogenies. But the genes used have to be carefully chosen. These genes should have slow evolutionary rates (typically those nucleus-encoded), not be compositionally biased, and be robust to lateral transfer among alphaproteobacteria. Thus, a close examination of the data, e.g., through 'phylome' or network analyses, can aid in choosing the right genes. Also, and most importantly, finding novel 'environmental' alphaproteobacteria,

through metagenomics, that might be most closely related to mitochondria could solve the problem. Similar efforts have recently found the closest prokaryotic relatives known to both the nucleocytoplasm of eukaryotes (Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017) and plastids (Ponce-Toledo et al., 2017). A metagenomic approach to uncover novel environmental diversity for the *Alphaproteobacteria* is perhaps the most promising avenue we have at hand to shed light on the perennial question of the phylogenetic placement and nature of the first mitochondrial ancestor.

Indeed, Martijn et al., (2018) did just that for mitochondria. They increased the sampling for the *Alphaproteobacteria* by reconstructing metagenome-assembled genomes (or MAGs) from ocean metagenomes (the TARA Oceans Project) (see Chapter 4). They then combined strategies to decrease the compositional heterogeneity in their datasets (24 and 29 mitochondrion-encoded and nucleus-encoded genes, respectively) and found that mitochondria are not related to the *Rickettsiales* but instead branch as sister to all other alphaproteobacteria. However, the study by Martijn et al., (2018) relied on a small number of genes (5,320 sites) and only focused on novel marine alphaproteobacteria.

Chapter 4 was concerned with inferring a robust and reference phylogeny for the *Alphaproteobacteria*. This chapter presents preliminary results that place the mitochondrial lineage in the phylogeny of the *Alphaproteobacteria*. This was done by increasing the number of suitable genes for phylogenetic inference, expanding the diversity of novel alphaproteobacterial genomes by surveying environmental metagenomes, and using sophisticated methods that reduce compositional heterogeneity.

## 5.2. METHODS

### 5.2.1. Metagenomics

Novel metagenomes from thermophilic microbial mats in the Llamara saltern (Chile; Saghaï et al., 2017) and microbialites in the Alchichica lake (Mexico; Saghaï et al., 2015) were sequenced by our collaborators Purificación López-García and David Moreira at Université Paris-Sud.

Illumina short raw reads (150 bp long) were assessed with FastQC. Adapters were later removed with Trimmomatic v0.38 (Bolger et al., 2014). Datasets that were unique to a sample were individually assembled, whereas datasets that represented technical

replicates were co-assembled using metaSPAdes (Bankevich et al., 2012). All contigs smaller than 2,500 bp in the assemblies were later discarded. The reads from all datasets were mapped onto each assembly using Bowtie2 (Langmead & Salzberg, 2012). Contigs were binned using MaxBin v2.2.2 (Wu, Simmons, & Singer, 2016) which relies on differential coverage across samples, tetranucleotide composition and single-copy gene markers. The completeness and contamination of the bins reported by MaxBin v2.2.2 were assessed with CheckM v1.0.12 (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015). All genome bins that were phylogenetically affiliated to the *Alphaproteobacteria* based on the manual examination of the CheckM reference genome tree (itself based on the concatenation of 43 gene markers) were retained. Reads from all datasets were later mapped onto the alphaproteobacterial genome bins with Bowtie2. All paired and unpaired reads that mapped were subsequently co-assembled with metaSPAdes. The resulting co-assembly was processed through the Anvi'o workflow (Eren et al., 2015). In brief, the Anvi'o metagenomic workflow allowed us to map Illumina short reads to the co-assembly with Bowtie2 to obtain Illumina read coverage values, assign the taxonomy to each contig based on DIAMOND (Buchfink, Xie, & Huson, 2015) searches against the NCBI nr database using predicted contig genes as queries, and estimate the completeness and contamination of bins predicted by CONCOCT2 (Alneberg et al., 2014) based on single-copy gene markers. Anvi-interactive super-imposed all these layers of information on the metagenome scaffolds which were later visually inspected to manually refine the alphaproteobacterial genome bins.

In addition to alphaproteobacterial MAGs reconstructed from the Llamara saltern (46 MAGs) and Alchichica lake (69 MAGs) metagenomes, additional alphaproteobacterial MAGs reconstructed from metagenomes of the Caspian Sea (10 MAGs; Mehrshad, Amoozegar, Ghai, Fazeli, & Rodriguez-Valera, 2016), the Mediterranean Sea (16 MAGs; Tully, Sachdeva, Graham, & Heidelberg, 2017), worldwide oceans (210 MAGs; Delmont et al., 2018) and aquifers (40 MAGs; Anantharaman et al., 2016) were added to our database. The alphaproteobacterial MAGs from these processed metagenomes were chosen based on the author's annotations or my own analyses with CheckM.

### 5.2.2. Gene marker and taxon selection

A dataset that comprised both mitochondrion- and nucleus-encoded genes was assembled. Eukaryotes included in this dataset where chosen based on the availability of both mitochondrial and nuclear genomes or transcriptomes. For the mitochondrion-encoded genes, the MitoCOGs (Kannan, Rogozin, & Koonin, 2014) database was used as a guide and those MitoCOGs that were widespread among eukaryotes were kept. BLAST searches (Altschul et al., 1990) against a database that comprised a broad and diverse set of eubacteria (107 representatives from 27 cultured phyla) were done to retrieve homologs, these homologs were then aligned with MAFFT, and trees were later inferred with IQ-TREE. Those gene markers whose trees showed a sister relationship between eukaryotes and alphaproteobacteria were kept. For the nucleus-encoded genes, BLAST searches of all proteins contained in the proteomes of 13 representative eukaryotes were conducted against a database of 176 prokaryotes (136 eubacteria and 40 archaebacteria). BLAST hits were clustered into homologous families with a custom Perl script, aligned with MAFFT (L-INS-I; Katoh, Kuma, Toh, & Miyata, 2005), trimmed with BMGE (Criscuolo & Gribaldo, 2010), and phylogenetic trees were inferred under the LG model in RAxML v8 (Stamatakis, 2014) for each homologous family. Phylogenetic trees were then sorted based on the criterion that eukaryotes form a clade with alphaproteobacteria. Manual inspection of single gene trees then followed to remove paralogs and single contaminants. The foregoing steps to identify nucleus-encoded genes of mitochondrial origin were performed by my collaborator Laura Eme at Uppsala University. To these, I then added the gene markers independently identified by Wang & Wu, (2015), Derelle et al., (2015), and Leigh, (2009). I then removed all redundant gene markers relative to Eme's gene markers through BLAST searches, and redid BLAST searches against my own local database that comprised a broad and diverse set of eubacteria. Homologs were then retrieved, aligned with MAFFT (L-INS-i) and trees were inferred under the LG4X in IQ-TREE v1.6.5 (Nguyen et al., 2015). The gene markers whose trees showed a sister relationship between eukaryotes and alphaproteobacteria were kept. The final dataset comprised 113 genes in total, of which 45 were mitochondrion-encoded and 68 nucleus-encoded. If a gene had at least one homolog encoded in a mitochondrial genome it was considered mitochondrion-encoded. Many of the nucleus-encoded gene markers have mitochondrial functions. Curiously, some mitochondrion-encoded genes did not have a clear alphaproteobacterial origin; these

were discarded. A last BLAST search using alphaproteobacterial queries from the 113 gene markers determined above was performed against a database that comprised the reference alphaproteobacteria and novel MAGs. Homologs were retrieved, aligned with MAFFT (L-INS-i), trimmed with ZORRO (Wu, Chatterji, & Eisen, 2012; and a custom Perl script) and trees were inferred with IQ-TREE v1.6.5. to remove paralogs and unusually divergent sequences. Single gene files were then concatenated with SequenceMatrix v1.8 (Vaidya, Lohman, & Meier, 2011). The dataset was filtered using PhyloMCOA (de Vienne, Ollier, & Aguileta, 2012) to remove outlier sequences (distance='nodal'; bvalue=50); however no taxon or gene outliers were detected. Afterwards, sites that contained more 20% gaps were removed with the tool pxclsq from the phyx tools suite (Brown, Walker, & Smith, 2017). The final dataset contained 29,640 amino acid sites.

Another dataset that comprised only reference alphaproteobacterial genomes and novel MAGs was built. Reference alphaproteobacteria were selected as previously described (see Chapter 4). A set of 200 gene markers (54,400 sites) defined by Phyla-AMPHORA was used (as in Chapter 4; Wang & Wu, 2013). The genes are single-copy and predominantly vertically-inherited as assessed by congruence among them (Wang & Wu, 2013).

Most MAGs were removed after inspecting the CheckM reference genome tree. Those kept were chosen because they appear to represent novel lineages (i.e., maximized phylogenetic diversity) and had shorter branches than close relatives. After adding these MAGs into the Phyla-AMPHORA dataset (200 single-copy conserved genes), some more MAGs were further removed using the same criteria. This was done with the aim of reducing taxa for computational efficiency.

### 5.2.3. Phylogenetic analyses

Phylogenetic inference was done using IQ-TREE v1.6.5 under the LG+PMSF(C60)+F+G4 model using a previously inferred guide tree under the same model (i.e., LG+PMSF(C60)+F+G4) which had itself been inferred based on an LG4X+F guide tree (Minh et al., 2013; Nguyen et al., 2015). The PMSF model estimates site-specific amino acid profiles that account for compositional heterogeneity across sites under the maximum-likelihood framework (Wang et al., 2018). Data was recoded using the dataset-specific recoding scheme S4 (ANQESWY CGIPV RHLKM DFT) estimated

by the minmax-chisq program (see Susko & Roger, 2007, and Methods in Chapter 4 for further details). Phylogenetic analyses of datasets recoded into four-character state alphabets were analyzed using IQ-TREE v1.6.5 and the model GTR+C60S4+F+G4. C60S4 is an adaptation of the empirical mixture model C60 to four-character states. It is obtained by adding the frequencies of the amino acids that belong to each bin in the dataset-specific four-character state scheme S4 (see Data Recoding in the Methods of Chapter 4 for details). Site removal to reduce the compositional heterogeneity of the dataset was performed using two different methods, the ʐ metric (see Methods in Chapter 4) and the $\chi^2$ method (Viklund et al., 2012). The most compositionally heterogeneous sites according to ʐ were progressively removed using the software SiteStripper (Verbruggen, 2018) in increments of 10% up to 60%. For further details on how the ʐ metric measures compositional heterogeneity see the Methods in Chapter 4. The $\chi^2$ trimmer removes sites that contribute the most to the $\chi^2$ statistic (Martijn et al., 2018; Viklund et al., 2012). This is done by recalculating the $\chi^2$ statistic after removing the site in question and then comparing the new $\chi^2$ to the older.

## 5.3. PRELIMINARY RESULTS AND DISCUSSION

### 5.3.1. Diverse alphaproteobacteria from Llamara and Alchichica

The Atacama desert in the north of Chile is one of the driest places in the world. I investigated metagenomes sequenced from several microbial mats from a warm and saline shallow pond that differ in important physicochemical factors (oxygen, temperature and salinity; Saghaï et al., 2017). Similarly, I also investigated metagenomes sequenced from microbialites found in Alchichica lake, an alkaline (pH 9) and high-altitude lake (2,300 m above sea level) in east-central Mexico (Saghaï et al., 2015). The microbial composition of both environments had previously been studied using 16S/18S rRNA meta-barcoding approaches (Saghaï et al., 2017, 2015), and revealed that eubacteria, including alphaproteobacteria, were highly abundant and diverse.

I found a diverse set of alphaproteobacterial MAGs in the metagenomes of thermophilic microbial mats from the Llamara saltern in Chile (46 alphaproteobacterial MAGs) and the microbialites from the Alchichica lake in Mexico (69 alphaproteobacterial MAGs) (Fig. 5.1). Most of the MAGs recovered were above 80% complete and had little contamination after being manually refined in Anvi'o (Fig. 5.1). The MAGs were on

average only 0.8% contaminated, and in most cases most of this contamination is caused by strain heterogeneity, i.e., contamination by a closely related strain (as assessed by CheckM; Fig. 5.1). The genome sizes for the recovered MAGs ranged from 0.91 to 4.85 Mbp (Fig. 5.1). This suggests that the MAGs recovered are not limited to bacteria with small genomes, but that the methods are also able to recover larger genomes. The number of scaffolds per MAG recovered varied considerably. Some MAGs contained as few as 10 scaffolds, whereas other had up to 543. The G+C% content appears to be constant across the MAGs, with an average of 63%, but some taxa affiliated with the *Rhodospirillales* and the *Rickettsiales* appear to have G+C% contents below 50%. CheckM was able to assign most MAGs to known alphaproteobacterial orders; those that could not be affiliated to any taxonomic group in the *Alphaproteobacteria* represent novel or divergent members of known groups, as the tree suggests (see Fig. 5.1). It is also important to note that many of the new alphaproteobacterial MAGs are enriched in bacteriochlorophyll biosynthesis genes, suggesting that many of the MAGs come from anoxygenic photosynthetic bacteria (purple nonsulfur bacteria) in the microbial mats sampled (Fig. 5.1).

These novel MAGs expand the known genomic and phylogenetic diversity of the *Alphaproteobacteria*, and in combination with MAGs reconstructed from other environments by others, promise to improve our understanding of the evolutionary history of the *Alphaproteobacteria*. Moreover, these MAGs might also improve phylogenetic inference when attempting to find the closest modern relatives of mitochondria.

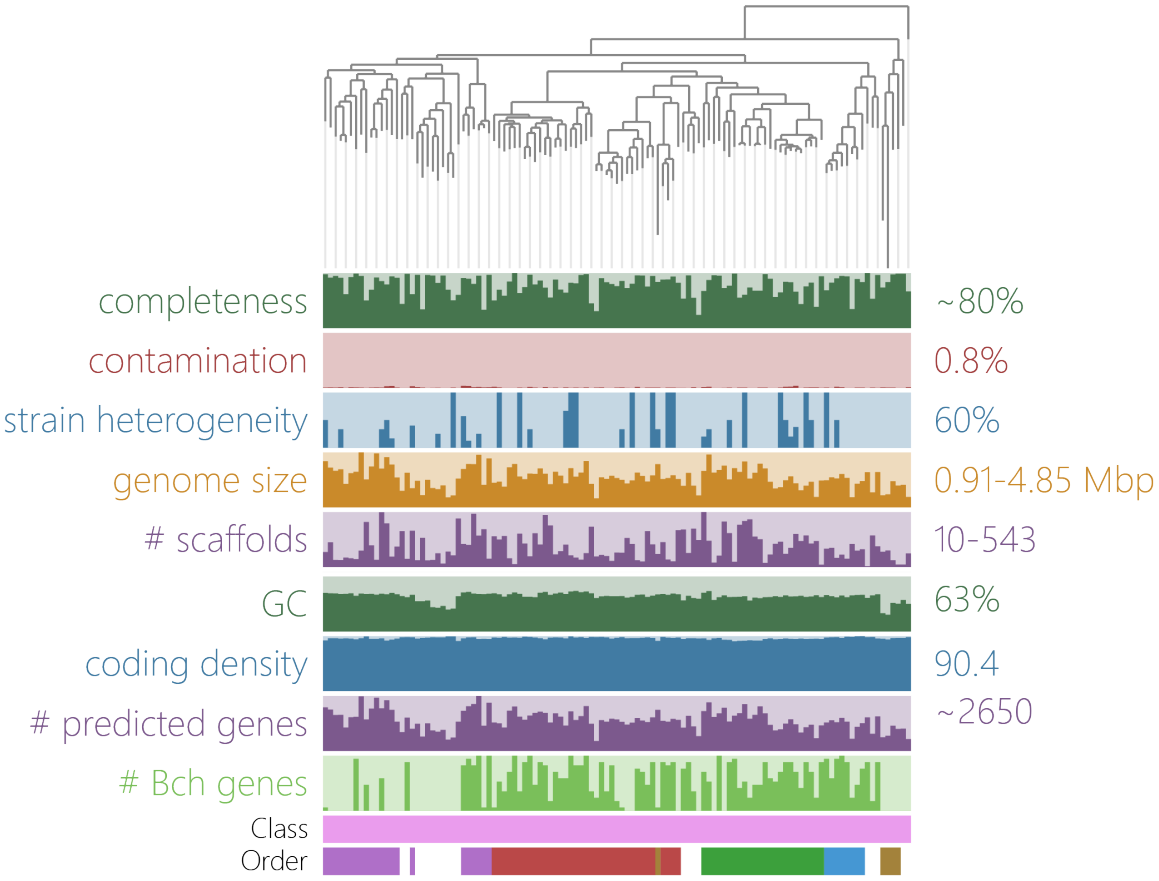



**Figure 5.1**. Alphaproteobacterial MAGs reconstructed from metagenomes from thermophilic microbial mats in the Llamara saltern (Chile) and microbialites in the Alchichica lake (Mexico). Different features for the reconstructed MAGs calculated by CheckM are shown in colored sections below a phylogenetic tree that includes all MAGS and was inferred in IQ-TREE under the LG4X model (43 marker genes). Bch: bacteriochlorophyll. See Fig. S5.1 for the abundances of each MAG across metagenomic samples.

### 5.3.2. Novel and disparate alphaproteobacterial lineages

It is expected that surveying metagenomes from novel environments would reveal a great diversity of new lineages in comparison to those that have been isolated and grown in the laboratory (e.g., Delmont et al., 2018). The most novel MAGs found in my preliminary survey (see Methods above) and a reference set of cultured and described alphaproteobacteria (see Methods and Chapter 4) were added to a dataset of 200 single-copy conserved genes and a phylogenetic tree was inferred under the site-heterogeneous model LG+PMSF(C60)+F+G4 in IQ-TREE v1.6.5. The resulting exploratory tree, which displays artefactual patterns (e.g., 'basal' placements for the *Pelagibacterales* and the *Holosporales*; see Chapter 4), appears to uncover two completely novel clades (labelled 'novel clade A' and 'novel clade B' in Fig. 5.2) that branch in between the *Magnetococcia* (*Magnetococcales*) and the *Alphaproteobacteria.* Whereas, 'novel clade A' has extremely long branches, 'novel clade B' has branches comparable in length to those seen in free-living members of other alphaproteobacterial orders. The same tree also suggests that most novel MAGs (shown with terminal circles in Fig. 5.2) group within or basal to known clades. For example, some MAGs represent novel and disparate members of the *Sneathiellales*, whereas other MAGs represent basal members of the *Rhodobacterales*, *Pelagibacterales*, *Rickettsiales* and *Geminicoccaceae*. It is also interesting to note that many of the novel MAGs basal to the *Rickettsiales* and *Pelagibacterales* have short branches, which might eventually improve phylogenetic tree inference after more careful attempts are made to reduce compositional heterogeneity. The novel diversity reported here expands upon that reported by Martijn et al., (2018) who focused solely on marine alphaproteobacteria. Here, I have not only sampled the same marine environments that Martijn et al. (2018) sampled, but I have also sampled novel freshwater environments and microbial mats.

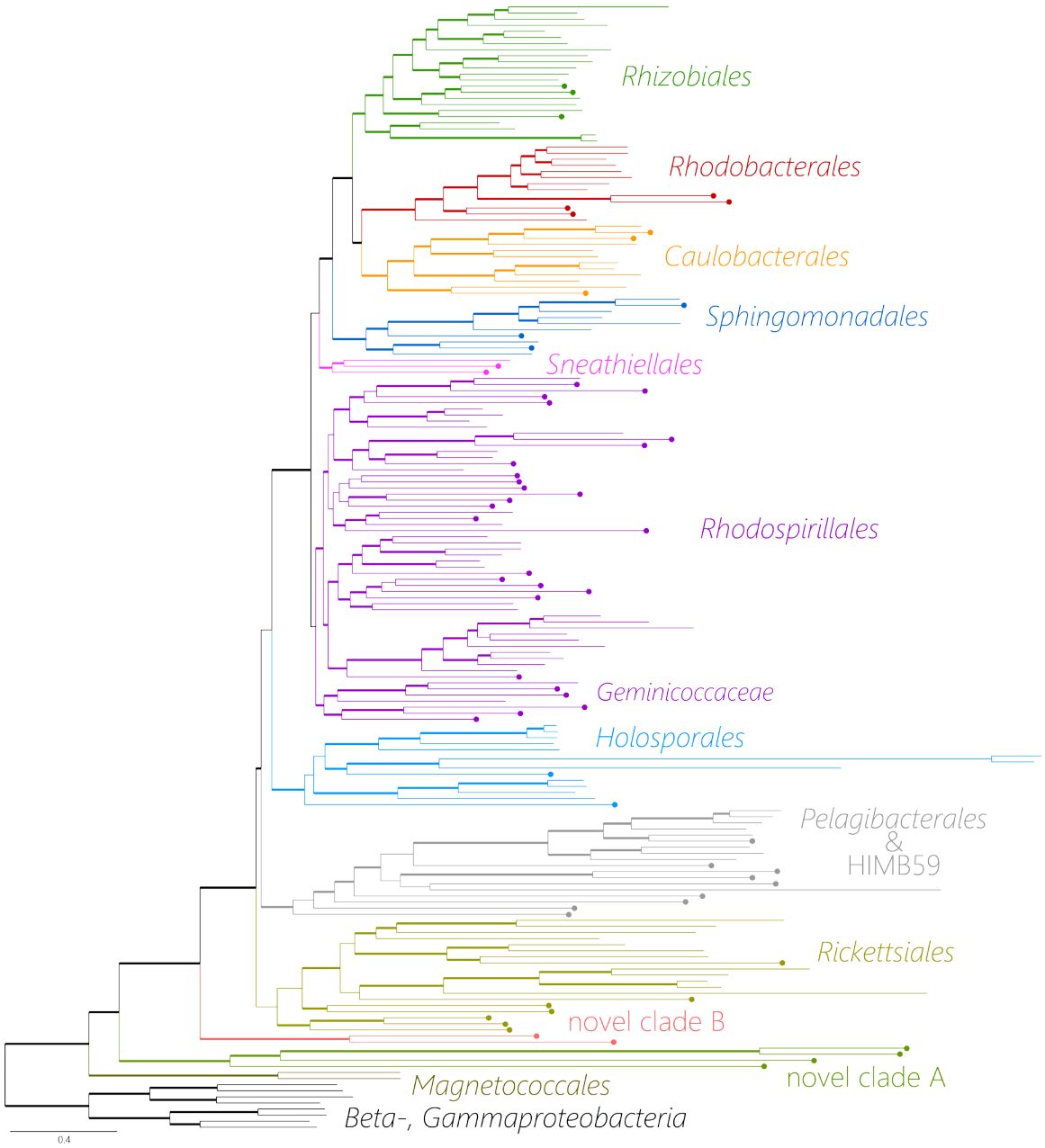



**Figure 5.2**. A phylogenetic tree inferred under the LG+PMSF(C60)+F+G4 model with the reference alphaproteobacteria and the most novel MAGs (113 marker genes of alphaproteobacterial origin in eukaryotes; 29,640 amino acid sites). Novel MAGs are indicated by terminal branch filled circles. Thick branches represent branch support values higher than 90% SH-aLRT and 90% UFBoot2+NNI. See Fig. S5.2 for a labeled tree.

### 5.3.3. Mitochondria branch as sister to the *Alphaproteobacteria*

Mitochondria evolved from within or are most closely related to modern alphaproteobacteria (e.g., Martijn et al., 2018; Wang & Wu, 2015). However, the exact relationship of mitochondria to the *Alphaproteobacteria* or any of its subgroups remains a topic of controversy (Martijn et al., 2018; Roger et al., 2017). I have therefore attempted to phylogenetically place the mitochondrial lineage in the *Alphaproteobacteria* phylogeny by relying on (1) an expanded set of novel alphaproteobacterial genomes (see above), (2) an increased number of marker genes, and (3) a combination of diverse methods aimed at disrupting artefactual patterns common in phylogenetic inference.

Heterogeneity in the amino acid composition across taxa can strongly bias phylogenetic inference (e.g., see Chapter 4; Foster, 2004). Therefore, as a first step to tackle the problem, I attempted to reduce compositional heterogeneity amongst taxa by removing compositionally-biased sites (according to the $z$ metric and the $\chi^2$ method) from a multi-protein dataset that comprised 113 genes of alphaproteobacterial origin (of which 45 are mitochondrion-encoded and 68 are nucleus-encoded). This dataset comprised a subset of the most phylogenetically disparate reference taxa and MAGs to ease computation. Exploratory trees were inferred in IQ-TREE v1.6.5 under the site-heterogeneous LG+PMSF(C60)+F+G4 model from a dataset from which the most compositionally-biased sites were progressively removed in increments of 10%.

The exploratory trees reveal that mitochondria consistently branch as sister to the *Alphaproteobacteria* (Fig. 5.3). Trees inferred both under the simple LG4X model (data not shown) and the most complex LG+PMSF(C60)+F+G4 (e.g., Fig. 5.3) show this pattern; a tree inferred from a dataset recoded into the four-character state scheme S4 also shows the same pattern (Fig. S5.4). Similarly, mitochondria remain sister to the *Alphaproteobacteria* across the spectrum of compositionally-biased site removal (Fig. 5.4). Indeed, support for this placement increases and then reaches stability after removing the first 10% most compositionally-biased sites (Fig. 5.4). These preliminary results thus corroborate the recent findings of Martijn et al., (2018). However, the evidence I present is stronger. Martijn et al., (2018) could only recover mitochondria as sister to the *Alphaproteobacteria* after removing 20% of the most compositionally-biased sites (1,329 amino acid sites; according to $\chi^2$ method) from their mitochondrion-encoded gene dataset which itself only comprised 6,649 amino acid sites. In contrast, my much

larger dataset, that combines both mitochondrion- and nucleus-encoded genes (113 genes in total), recovers the same pattern even with the simple LG4X model, which does not model compositional heterogeneity across sites unlike the LG+PMSF(C60)+F+G4 model. The results presented here, even though preliminary in nature, appear to be more robust than those presented by Martijn et al., (2018).

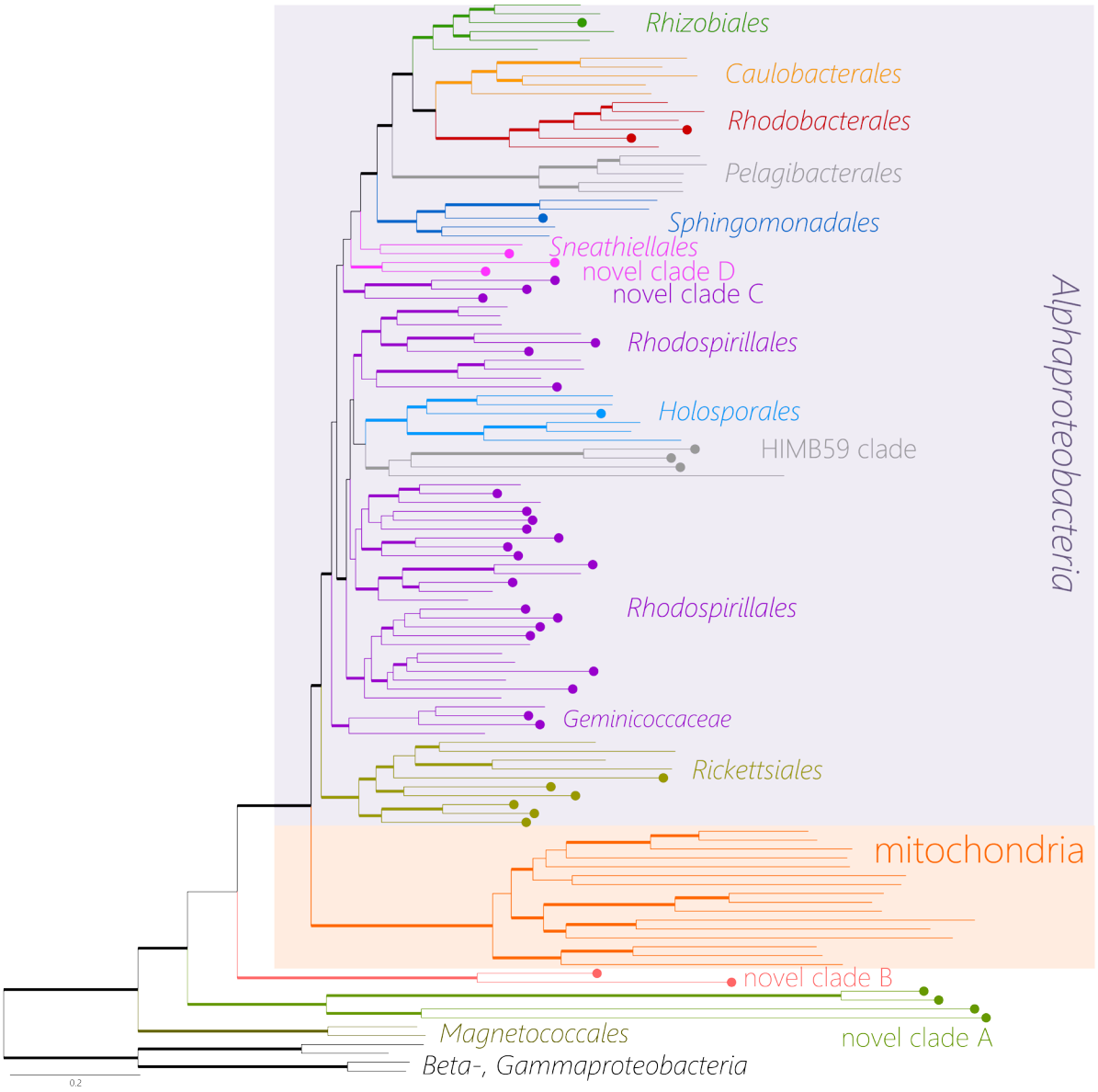



**Figure 5.3.** A phylogenetic tree inferred under the LG+PMSF(C60)+F+G4 model after removing 60% of the most compositionally biased sites according to the ʐ metric (a total of 11,855 amino acid sites analyzed from the 113-gene dataset). Novel MAGs are indicated by terminal branch filled circles. Thick branches represent branch support values higher than 90% SH-aLRT and 90% UFBoot2+NNI. See Fig. S5.3 for a labeled tree.

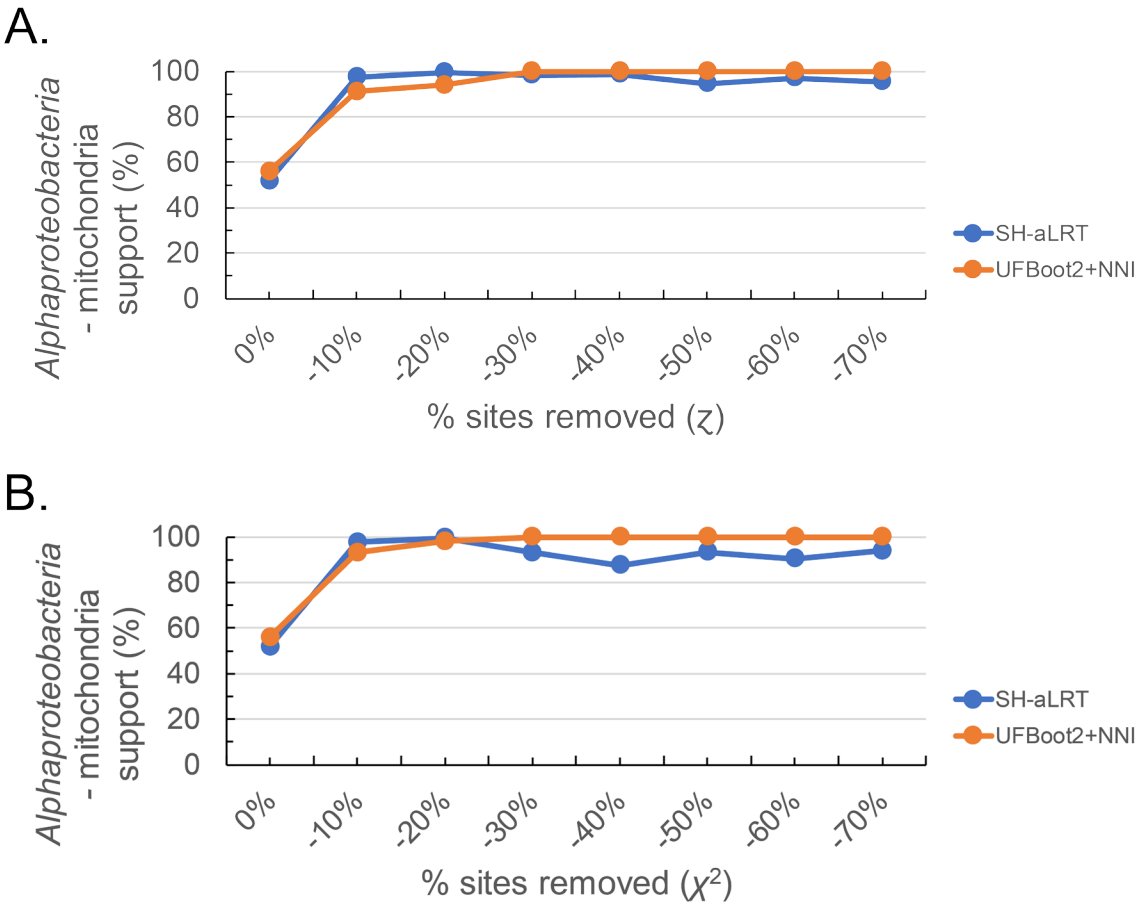



**Figure 5.4.** Branch support variation, SH-aLRT (%) and UFBoot2+NNI (%) as the most compositionally biased sites according to z metric and the $\chi^2$ method are removed. **A.** Branch support variation as compositionally-biased sites according to the z metric are removed in increments of 10%. **B.** Branch support variation as compositionally biased sites according to the $\chi^2$ method are removed in increments of 10%.

### 5.3.4. Congruence between phylogenies

Some other interesting patterns emerge from these preliminary analyses. Some new alphaproteobacterial clades appear to emerge and some of the artefactual patterns seen before removing compositionally-biased sites are disrupted. For example, two clades 'detach' from their previous affiliations to the *Sneathiellales* and the *Rhodospirillales* and now branch separately (labelled as novel clade C and D in Fig. 5.3). Moreover, the HIMB59 clade, which is sometimes allied to the *Pelagibacterales* (e.g., Grote et al., 2012), no longer groups with them and instead emerges as a sister group to the *Holosporaceae* within the *Rhodospirillales*. Several other patterns corroborate those seen in the reference phylogeny of the *Alphaproteobacteria* inferred in Chapter 4. For example, the relative branching order of higher-level groupings, like the *Rhizobiales*, *Caulobacterales*, *Rhodobacterales*, *Sphingomonadales*, *Sneathiellales*, *Rhodospirillales* and *Rickettsiales* remains stable. The independent origins of the *Rickettsiales* and *Holosporales* (shown in Chapter 4) is corroborated as the *Holosporales* now branches within the *Rhodospirillales*. The *Pelagibacterales* moves to a much more derived position as sister to the *Rhizobiales*, *Caulobacterales* and *Rhodobacterales*, as previously shown in Chapter 4 and by others (Martijn et al., 2018; Rodríguez-Ezpeleta & Embley, 2012; Viklund et al., 2012, 2013). The *Geminicocacceae* moves to a position where it is sister to all other ancestrally free-living alphaproteobacteria, and the *Rickettsiales* remains as sister to both. Despite these points of agreement between the phylogeny inferred here based on 113 genes of alphaproteobacterial origin in eukaryotes and the reference phylogeny of the *Alphaproteobacteria* based on 200 single-copy genes shown in Chapter 4, some conflicts do emerge. These include the specific interrelationship among the *Rhizobiales*, *Caulobacterales* and *Rhodobacterales*, and the interrelationships among the major families within the *Rhodospirillales*, including the position of the *Holosporaceae* and the HIMB59 clade. This incongruency likely stems from the different sets of genes analyzed and is currently the target of an ongoing investigation.

## 5.4. DISCUSSION AND CONCLUSIONS

In order to robustly place mitochondria in relation to the *Alphaproteobacteria*, my colleagues and I have (1) assembled novel alphaproteobaterial MAGs from diverse environments which include worldwide oceans, microbial mats, microbialites, and aquifers, (2) analyzed an expanded eukaryotic gene marker set of alphaproteobacterial

origin that includes both mitochondrion-, and nucleus-encoded genes, and (3) used novel methods for phylogenetic inference (e.g., site removal based on the $z$ metric, and the PMSF model in IQ-TREE). In comparison to the most recent study of Martijn et al., (2018) that focused only on marine metagenomes and a smaller set of marker genes (24 genes and 6,649 amino acid sites), I, with the help of my colleagues, have considerably expanded the data available to infer the phylogenetic origin of mitochondria. My preliminary results suggest that mitochondria are sister to the *Alphaproteobacteria*. This phylogenetic pattern is robust, being consistent across all or most of our analyses. Similarly, many aspects of the reference phylogeny for the *Alphaproteobacteria* presented in Chapter 4 are corroborated based on an independent and only partially overlapping dataset (113 eukaryotic genes of alphaproteobacterial origin). The few topological conflicts found will require further investigation to be fully explained.

The whole of the *Alphaproteobacteria* is extraordinarily diverse. The sisterhood between mitochondria and the *Alphaproteobacteria* therefore barely constrains our inferences about the nature of the first mitochondrial ancestor. This contrasts with the nucleocytoplasm of eukaryotes and plastids, for which recent metagenomic efforts have found that they are most closely related to less diversified or disparate (and therefore relatively more homogenous) groups, the Asgard archaea (or even the Heimdallarchaeota) for eukaryotes (Zaremba-Niedzwiedzka et al., 2017) and the cyanobacterium *Gloeomargarita lithophora* for plastids (Ponce-Toledo et al., 2017). Future efforts might still discover a new less diverse or disparate alphaproteobacterial lineage that is most closely related to mitochondria through either improved phylogenetic methods or increased sampling of new genomes. Such a discovery will certainly further constrain inferences on the nature of the first mitochondrial ancestor.

If this placement of mitochondria relative to the *Alphaproteobacteria* holds in future, several inferences can be made to make sense of this pattern. The signal in the surviving data (in the form of useful genes for phylogenetic inference) might just be insufficient. Enough time has erased the information required to extract real historical patterns. Even though the origin of mitochondria is more recent or at least as old as the divergence of proto-eukaryotes from their closest archaeal relatives, mitochondrial genes might have suffered even more extreme divergence regimes. It is also possible that there is no species-poor surviving group that is most closely related to mitochondria

other than the *Alphaproteobacteria* as a whole. This could be because species or lineage extinction is very common in evolutionary history.

The phylogenetic placement of mitochondria as sister to the *Alphaproteobacteria* might in principle suggest that the first mitochondrial ancestor was a free-living (alpha)proteobacterium. This is because mitochondria are no longer considered to be affiliated to the *Rickettsiales*, a classical group of intracellular parasites of eukaryotes. It is also because the alphaproteobacterial cenancestor, which is just one node away from the divergence point of mitochondria and the *Alphaproteobacteria*, is inferred to have been free-living. The suggestion that the first mitochondrial ancestor was free-living, now independently supported by the study of Martijn et al., (2018) and the preliminary analyses presented here, contrasts with a widely held view that had mitochondria branching within intracellular parasitic (or endosymbiotic more generally) bacteria, namely the *Rickettsiales* and the *Holosporales* (Sassera et al., 2011; Wang & Wu, 2014, 2015).

The results presented in this chapter are preliminary and thus different and improved data and methods from the ones used here might provide an alternative answer. Future efforts to account for compositional heterogeneity across both sites and taxa will become necessary. This could be done by (1) recoding data and using an across-taxa heterogenous model, (2) removing compositionally-biased sites and using an across-taxa heterogenous model (e.g., the NDCH2 model in P4; Foster, 2004), (3) using models that can account for both compositional heterogeneity across both sites and taxa (e.g., CAT-BP, Blanquart & Lartillot, 2008, or a newly developed model by Ed Susko and Andrew J. Roger, personal communication), or (4) partitioning the data according to heterotachous (e.g., across-taxon rate heterogeneity) classes and analyzing each partition with an across-site heterogeneous model. These strategies constitute some of the next steps in my own research.

# CHAPTER 6: CONCLUSIONS

'*I'm not afraid of death … I'm afraid of time*' —Prof. Brand (Interstellar, 2014)

In the preceding chapters I dealt with questions that could ultimately shed light on the nature of the first mitochondrial ancestor. This is a difficult problem and I do not claim to have substantially improved our knowledge on the topic. The difficulties are manifold. But perhaps the major hurdle is time. Time has erased most of the evidence evolutionary biologists rely on to reconstruct the past. Time has seen ancestral species diversify and diverge to almost beyond recognition (e.g., our oldest ancestor was a bacterium) but has also seen most species go extinct in the history of our planet. But time has not been all powerful. Stasis has preserved some clues for us. I have attempted to study and make sense of some features among modern biodiversity that have survived the rigors of time. All this with the aim of making inferences and reconstructing an event that happened more than two billion years ago.

To be able to reconstruct a major evolutionary transition, such as the origin of mitochondria, one needs to understand the start and end points of the transition (Roger et al., 2017). The end point of our problem at hand is the last mitochondrial ancestor or the mitochondrial cenancestor. Although we will keep learning about it as more disparate mitochondria are studied, our knowledge about the last mitochondrial ancestor is sufficiently precise and constrained by the diversity of modern mitochondria. On the other hand, we have less confidence about the start point or the first mitochondrial common ancestor. This has been largely because of the uncertainty of the placement of mitochondria within the *Alphaproteobacteria* and also because of the great diversity and disparity seen among modern alphaproteobacteria. My work throughout the last five years has been able to shed some light on the question of the nature of the first mitochondrial ancestor.

The results of my research allow me to make some conclusions about the nature of the first mitochondrial ancestor. The first one is that it now seems most probable that the first mitochondrial ancestor was a free-living alphaproteobacterium—I called this ancestor an 'alphaproteobacterium' because of the close patristic distance (or branch length) between mitochondria and the modern *Alphaproteobacteria*. This conclusion is primarily supported by the latest phylogenetic evidence based on multi-gene trees (Martijn et al., 2018; Chapters 4 and 5). There are two main arguments that support this conclusion.

First, the *Rickettsiales* and *Holosporales*, both comprised by intracellular parasites of eukaryotes and thought to be sister groups, are distantly unrelated to each other and evolved convergently from quite different ancestors (see Chapter 4). Thus, mitochondria are no longer phylogenetically embedded in a larger group comprised exclusively by intracellular parasites (the *Rickettsiales* and *Holosporales*) as previously thought (Wang & Wu, 2014, 2015). Second, the mitochondrial lineage itself appears to be unrelated to the *Rickettsiales* and instead branches as sister to the *Alphaproteobacteria* as a whole (see Chapter 5). The conclusion that the first mitochondrial ancestor was free-living contrasts with a long-held view that mitochondria could have evolved from intracellular parasites, perhaps related to the *Rickettsiales* (e.g., Andersson et al., 1998; Ball, Bhattacharya, & Weber, 2016; Fitzpatrick et al., 2006; Sassera et al., 2011; Wang & Wu, 2014, 2015).

However, there is an assumption that underlies the conclusion that the first mitochondrial ancestor was free-living. The assumption is that the ancestors of the *Alphaproteobacteria* and of the larger clade that comprises both the *Alphaproteobacteria* and mitochondria was free-living. This assumption, although intuitive, cannot be confidently defended. My preliminary trees reveal at least two novel MAG lineages that branch in between the *Magnetococcia* and the *Alphaproteobacteria*. Even though both linages are currently represented by few MAGs, their reconstructed genomes appear to be small, perhaps suggesting a parasitic, intracellular or symbiotic lifestyle (the genome size ranges for novel clades A and B are 0.95-3.02 Mbp and 0.92-1.57 Mbp, respectively). However, it is more probable that symbiotic lineages, parasitic or otherwise, have evolved repeatedly from free-living ancestors than the opposite (see Chapter 4 for an example). Future studies on the biology and diversity of these new lineages will clarify the issue.

Phylogenetics is not the only tool that can help us reconstruct the nature of the first mitochondrial ancestor. A second conclusion that I believe can be made is that the first mitochondrial ancestor was anatomically a complex alphaproteobacterium. More specifically, the first mitochondrial common ancestor could develop intracytoplasmic membranes (ICMs) for bioenergetic purposes (Muñoz-Gómez et al., 2017). This inference is primarily supported by comparative genomics and has also more recently been aided by experimental evidence (see Chapter 3). The first mitochondrial ancestor certainly had Mic60 with a signature Mitofilin domain (Muñoz-Gómez, Slamovits, Dacks,

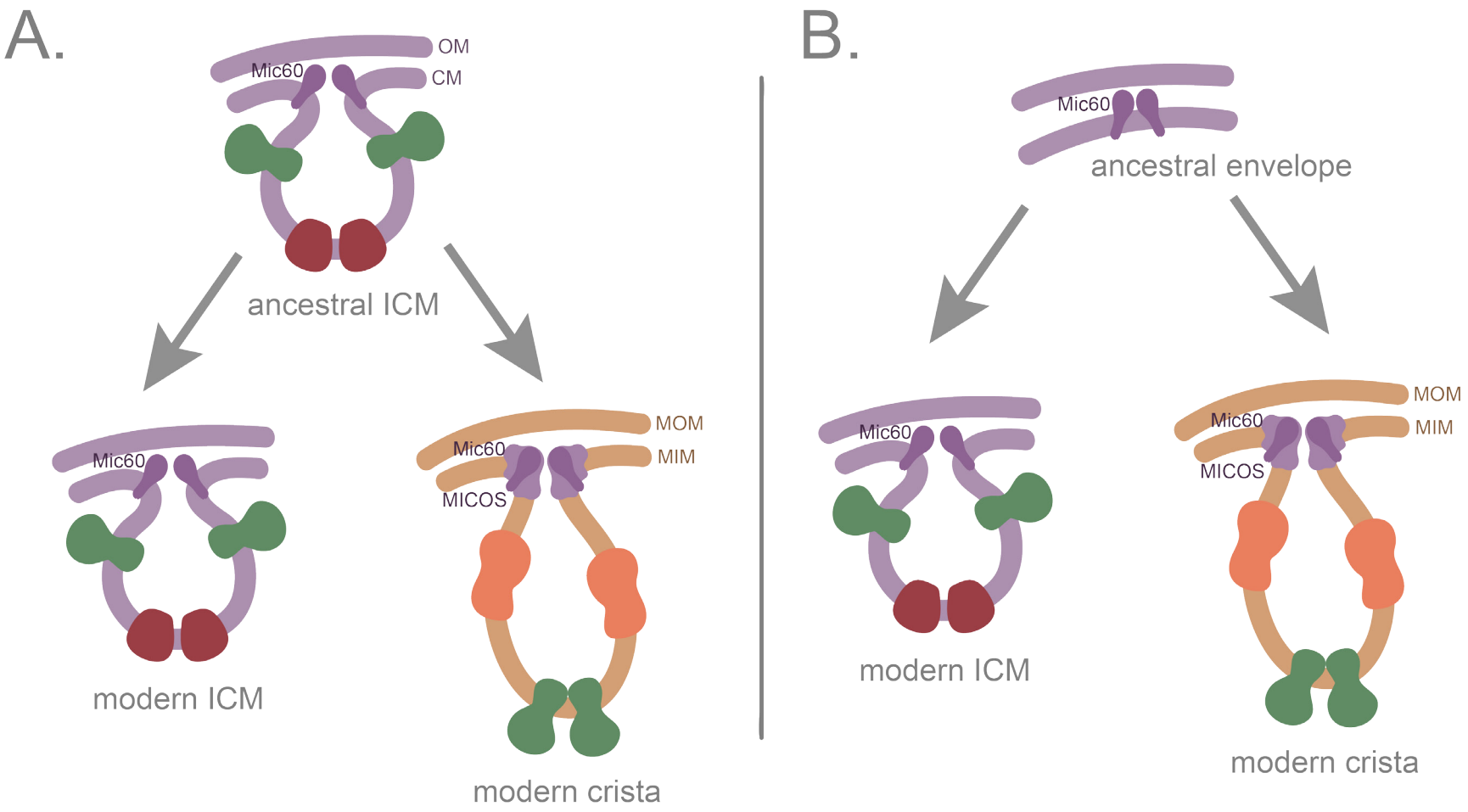
Baier, et al., 2015). In modern mitochondria, this protein is part of the MICOS complex which is largely responsible for the development and maintenance of cristae, i.e., by making CSs and CJs. The question I attempted to provide an answer to was: What roles did Mic60 have in the alphaproteobacterial ancestor of mitochondria? The answer to this question would help us understand whether the first mitochondrial ancestor developed bioenergetic ICMs.

I have compiled several lines of evidence, both bioinformatic and experimental, that are consistent with the idea that Mic60 in alphaproteobacteria performs the same general function that it does in mitochondria (see discussion in Chapter 3). The evidence that I have provided is, however, still circumstantial, and more direct evidence is needed. Tarasenko et al., (2017) provided more direct, experimental, evidence for a conserved function of Mic60 in alphaproteobacteria relative to mitochondria. More direct evidence can still be acquired by studying the function of Mic60 in alphaproteobacteria—a research direction that now different groups are undertaking. My own research, and that of others, has therefore provided support for the hypothesis that mitochondrial cristae evolved via the evolutionary transformation of alphaproteobacterial ICMs.

The simplest interpretation of the available evidence is that there has been historical continuity between alphaproteobacterial ICMs and mitochondrial cristae. That is, cristae and ICMs are homologous and both evolved from ICMs present in the last common ancestor between mitochondria and its sister group, the *Alphaproteobacteria* (i.e., the first mitochondrial ancestor)—cristae have a pre-endosymbiotic origin. This view implies that Mic60 has always had the same function in the development of ICMs. If this hypothesis turns out to be correct, bioenergetic ICMs might have pre-adapted the first mitochondrial ancestor to become an efficient bioenergetic or respiratory organelle (Muñoz-Gómez et al., 2017). However, there is another possibility that is also compatible with the evidence. This second scenario postulates that cristae and ICMs are deeply homologous and both evolved independently from an ancestor that did not develop any structures that resembled either cristae or ICMs—cristae have a post-endosymbiotic origin. This means that cristae and ICMs are largely convergent structures but share a crucial factor that underlies their development. This view implies that Mic60, which originally might have played a more general role in envelope development, was independently recruited to function in the development of both ICMs and cristae (Muñoz-Gómez et al., 2017).

Because both views are compatible with the available evidence, it is not easy to discriminate between the two evolutionary hypotheses. Cristae have been lost, or drastically reduced, numerous times in mitochondria that specialized to anaerobic niches throughout eukaryote evolution (Stairs et al., 2015)—cristae were present in the last mitochondrial ancestor. Similarly, the distribution of bioenergetic ICMs across the *Alphaproteobacteria* is patchy and this could be the outcome of multiple independent losses or convergent origins of ICMs (Muñoz-Gómez et al., 2017). It is thus uncertain how many times ICMs have evolved in the *Alphaproteobacteria*. These observations suggest that both cristae and ICMs can be lost as adaptations to specific environments. However, the diversity of the *Alphaproteobacteria* is so great that it is easy to conceive of independent origins of different ICM types. Many alphaproteobacteria that do not appear to develop ICMs also have a conserved Mic60 encoded in their genomes (Muñoz-Gómez et al., 2017). Only a detailed mechanistic comparison of the function of Mic60 across disparate alphaproteobacteria that develop different ICM types, and those that do not, could reveal clues as to whether the different ICM types seen among alphaproteobacteria are homologous or not (and therefore whether they have a single or multiple independent origins). These considerations might help one weigh the evidence to decide whether cristae and ICMs are homologous or analogous (i.e., deeply homologous).

**Figure 6.1.** Two alternative scenarios for the evolutionary relationship between mitochondrial cristae and alphaproteobacterial ICMs**. A.** Cristae and ICMs are homologous and the former evolved by the evolutionary transformation of the latter. Cristae would then constitute a specialized type of ICMs, respiratory ICMs, in a divergent lineage of the *Alphaproteobacteria*, i.e., mitochondria. **B.** Cristae and ICMs are deeply homologous, having evolved independently because of similar adaptive reasons, i.e., increased bioenergetic surface, but also having independently recruited the same developmental factor, i.e, Mic60.


A.
OM
Mic60
CM
ancestral ICM
Mic60
modern ICM
MOM
Mic60
MIM
MICOS
modern crista
B.
Mic60
ancestral envelope
Mic60
modern ICM
MOM
Mic60
MIM
MICOS
modern crista

If the first mitochondrial ancestor was an alphaproteobacterium that developed bioenergetic ICMs, what kind of ICMs did it develop? There are no examples of respiratory ICMs functionally analogous to cristae among modern alphaproteobacteria. The most common and widespread ICM type belongs to anoxygenic photosynthetic alphaproteobacteria (see Chapter 3). Other ICMs, like nitrifying, methanotrophic, or magnetosomes, have a very restricted and derived phylogenetic distribution (Muñoz-Gómez et al., 2017). They are unlikely to have been present in the ancestor of the *Alphaproteobacteria* or mitochondria. On the contrary, there is some evidence that anoxygenic photosynthesis could have been ancestral to the *Alphaproteobacteria* (Muñoz-Gómez et al., 2017). However, the evidence is controversial. In favour of the view that photosynthesis is ancestral to the *Alphaproteobacteria*, we have the widespread distribution of photosynthesis among the *Alphaproteobacteria*. Every major alphaproteobacterial order, with the exception of the parasitic *Rickettsiales*, contains photosynthetic members, i.e., the *Rhodospirialles*, *Sphingomonadales*, *Rhodobacterales*, *Caulobacterales* and *Rhizobiales*. There is also congruency at higher levels (i.e., major orders are recovered as monophyletic) for the vertical inheritance of photosynthetic genes in the *Alphaproteobacteria* (my own personal observations but see also Koblížek, Zeng, Horák, & Oborník, 2013; Swingley, Blankenship, & Raymond, 2009). Against, there is conflict at lower levels, i.e., within orders (Brinkmann, Göker, Koblížek, Wagner-Döbler, & Petersen, 2018), and the observation that some aerobic photosynthesizers carry their photosynthetic genes in plasmids (Petersen et al., 2012; Petersen, Frank, Göker, & Pradella, 2013; although these look more like chromosomes than plasmids, see Petersen et al., 2013). Systematic investigations to address the issue of the vertical inheritance of anoxygenic photosynthesis in the *Alphaproteobacteria* will be needed.

The presence of ICMs in the ancestor of mitochondria is much more compatible with the view that this ancestor was free-living. To my knowledge, there are no examples of intracellular bacteria, either parasitic, mutualistic, or commensalistic, that develop ICMs. The two main conclusions of my thesis, therefore, agree with each other. The nature of the first mitochondrial ancestor, however, remains poorly constrained. The idea that this ancestor was a free-living alphaproteobacterium that developed ICMs for photosynthesis is attractive and deserves more attention and future investigation. Indeed, some authors have speculated on the selective advantages that such a photosynthetic endosymbiont could have provided to a heterotrophic host (Cavalier-Smith, 2006, 2007; Fenchel &

Bernard, 1993a, 1993b; Woese, 1977). This is a small step in constraining the nature of the first mitochondrial ancestor but it is a step forward. Some important questions that remain about the nature of the first mitochondrial ancestor are: Did this ancestor develop magnetosomes for magnetotaxis? Was it flagellated? Did it have a repertoire of anaerobic energy enzymes? What did its respiratory chain look like (e.g., did it have low oxygen cytochromes)? Future comparative genomics within an updated phylogenetic framework for the *Alphaproteobacteria* will help elucidate the phenotype of the alphaproteobacterial cenancestor and that of the first mitochondrial ancestor.

# REFERENCES


Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, *21*(9), 2104–2105. https://doi.org/10.1093/bioinformatics/bti263

Abhishek, A., Bavishi, A., Bavishi, A., & Choudhary, M. (2011). Bacterial genome chimaerism and the origin of mitochondria. *Canadian Journal of Microbiology*, *57*(1), 49–61. https://doi.org/10.1139/W10-099

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., … Spiegel, F. W. (2012). The Revised Classification of Eukaryotes. *Journal of Eukaryotic Microbiology*, *59*(5), 429–514. https://doi.org/10.1111/j.1550-7408.2012.00644.x

Allen, R. D. (1995). Membrane tubulation and proton pumps. *Protoplasma*, *189*(1–2), 1–8. https://doi.org/10.1007/BF01280286

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., … Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

An, J., Shi, J., He, Q., Lui, K., Liu, Y., Huang, Y., & Sheikh, M. S. (2012). CHCM1/CHCHD6, novel mitochondrial protein linked to regulation of mitofilin and mitochondrial cristae morphology. *The Journal of Biological Chemistry*, *287*(10), 7411–7426. https://doi.org/10.1074/jbc.M111.277103

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., … Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*, *7*, 13219. https://doi.org/10.1038/ncomms13219

Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pont|[eacute]|n, T., Alsmark, U. C. M., Podowski, R. M., … Kurland, C. G. (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature*, *396*(6707), 133–140. https://doi.org/10.1038/24094

Andersson, S. G., & Kurland, C. G. (1999). Origins of mitochondria and hydrogenosomes. *Current Opinion in Microbiology*, *2*(5), 535–541. https://doi.org/10.1016/S1369-5274(99)00013-2

Arai, H., Roh, J. H., & Kaplan, S. (2008). Transcriptome dynamics during the transition from anaerobic photosynthesis to aerobic respiration in Rhodobacter sphaeroides 2.4.1. *Journal of Bacteriology*, *190*(1), 286–299. https://doi.org/10.1128/JB.01375-07

Armbruster, U., Labs, M., Pribil, M., Viola, S., Xu, W., Scharfenberg, M., … Leister, D. (2013). Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. *The Plant Cell*, *25*(7), 2661–2678. https://doi.org/10.1105/tpc.113.113118

Atteia, A., Adrait, A., Brugière, S., Tardif, M., van Lis, R., Deusch, O., … Rolland, N. (2009). A proteomic survey of Chlamydomonas reinhardtii mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. *Molecular Biology and Evolution*, *26*(7), 1533–1548. https://doi.org/10.1093/molbev/msp068

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., … Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/msb4100050

Babu, M., Bundalovic-Torma, C., Calmettes, C., Phanse, S., Zhang, Q., Jiang, Y., … Emili, A. (2018). Global landscape of cell envelope protein complexes in Escherichia coli. *Nature Biotechnology*, *36*(1), 103–112. https://doi.org/10.1038/nbt.4024

Ball, S. G., Bhattacharya, D., & Weber, A. P. M. (2016). EVOLUTION. Pathogen to powerhouse. *Science (New York, N.Y.)*, *351*(6274), 659–660. https://doi.org/10.1126/science.aad8864

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barbot, M., Jans, D. C., Schulz, C., Denkert, N., Kroppen, B., Hoppert, M., … Meinecke, M. (2015). Mic10 Oligomerizes to Bend Mitochondrial Inner Membranes at Cristae Junctions. *Cell Metabolism*, *21*(5), 756–763. https://doi.org/10.1016/j.cmet.2015.04.006

Bayer, M. E. (1991). Zones of membrane adhesion in the cryofixed envelope of Escherichia coli. *Journal of Structural Biology*, *107*(3), 268–280. https://doi.org/10.1016/1047-8477(91)90052-X

Bazylinski, D. A., Williams, T. J., Lefèvre, C. T., Trubitsyn, D., Fang, J., Beveridge, T. J., … Simpson, B. (2013). Magnetovibrio blakemorei gen. nov., sp. nov., a magnetotactic bacterium (Alphaproteobacteria: Rhodospirillaceae) isolated from a salt marsh. *International Journal of Systematic and Evolutionary Microbiology*, *63*(5), 1824–1833. https://doi.org/10.1099/ijs.0.044453-0

Beatty, J. T., & Gest, H. (1981). Generation of succinyl-coenzyme A in photosynthetic bacteria. *Archives of Microbiology*, *129*(5), 335–340. https://doi.org/10.1007/BF00406457

Blanquart, S., & Lartillot, N. (2008). A site- and time-heterogeneous model of amino acid replacement. *Molecular Biology and Evolution*, *25*(5), 842–858. https://doi.org/10.1093/molbev/msn018

Bohnert, M., Wenz, L.-S., Zerbes, R. M., Horvath, S. E., Stroud, D. A., von der Malsburg, K., … van der Laan, M. (2012). Role of mitochondrial inner membrane organizing system in protein biogenesis of the mitochondrial outer membrane. *Molecular Biology of the Cell*, *23*(20), 3948–3956. https://doi.org/10.1091/mbc.E12-04-0295

Bohnert, M., Zerbes, R. M., Davies, K. M., Mühleip, A. W., Rampelt, H., Horvath, S. E., … van der Laan, M. (2015). Central Role of Mic10 in the Mitochondrial Contact Site and Cristae Organizing System. *Cell Metabolism*, *21*(5), 747–755. https://doi.org/10.1016/j.cmet.2015.04.007

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonen, L., Cunningham, R. S., Gray, M. W., & Doolittle, W. F. (1977). Wheat embryo mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature. *Nucleic Acids Research*, *4*(3), 663–671.

Brantner, C. A., Remsen, C. C., Owen, H. A., Buchholz, L. A., & Perille Collins, M. L. (2002). Intracellular localization of the particulate methane monooxygenase and methanol dehydrogenase in Methylomicrobium album BG8. *Archives of Microbiology*, *178*(1), 59–64. https://doi.org/10.1007/s00203-002-0426-2

Brimacombe, C. A., Stevens, A., Jun, D., Mercer, R., Lang, A. S., & Beatty, J. T. (2013). Quorum-sensing regulation of a capsular polysaccharide receptor for the Rhodobacter capsulatus gene transfer agent (RcGTA). *Molecular Microbiology*, *87*(4), 802–817. https://doi.org/10.1111/mmi.12132

Brindefalk, B., Ettema, T. J. G., Viklund, J., Thollesson, M., & Andersson, S. G. E. (2011). A Phylometagenomic Exploration of Oceanic Alphaproteobacteria Reveals Mitochondrial Relatives Unrelated to the SAR11 Clade. *PLoS ONE*, *6*(9), e24457. https://doi.org/10.1371/journal.pone.0024457

Brinkmann, H., Göker, M., Koblížek, M., Wagner-Döbler, I., & Petersen, J. (2018). Horizontal operon transfer, plasmids, and the evolution of photosynthesis in Rhodobacteraceae. *The ISME Journal*, *12*(8), 1994. https://doi.org/10.1038/s41396-018-0150-9

Brown, J. W., Walker, J. F., & Smith, S. A. (2017). Phyx: phylogenetic tools for unix. *Bioinformatics*, *33*(12), 1886–1888. https://doi.org/10.1093/bioinformatics/btx063

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Burger, G., Gray, M. W., Forget, L., & Lang, B. F. (2013). Strikingly Bacteria-Like and Gene-Rich Mitochondrial Genomes throughout Jakobid Protists. *Genome Biology and Evolution*, *5*(2), 418–438. https://doi.org/10.1093/gbe/evt008

Burki, F. (2014). The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harbor Perspectives in Biology*, *6*(5), a016147. https://doi.org/10.1101/cshperspect.a016147

Callister, S. J., Nicora, C. D., Zeng, X., Roh, J. H., Dominguez, M. A., Tavano, C. L., … Lipton, M. S. (2006). Comparison of aerobic and photosynthetic Rhodobacter sphaeroides 2.4.1 proteomes. *Journal of Microbiological Methods*, *67*(3), 424–436. https://doi.org/10.1016/j.mimet.2006.04.021

Cavalier-Smith, T. (1981). Eukaryote kingdoms: Seven or nine? *Biosystems*, *14*(3–4), 461–481. https://doi.org/10.1016/0303-2647(81)90050-2

Cavalier-Smith, T. (1983). Endosymbiotic origin of the mitochondrial envelope. *Endocytobiology II*, 265-279.

Cavalier-Smith, T. (2002). The Phagotrophic Origin of Eukaryotes and Phylogenetic Classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology*, *52*(2), 297–354.

Cavalier-Smith, Thomas. (1983). A 6-kingdom classification and a unified phylogeny. In W. Schwemmler & H. E. A. Schenk (Eds.), *Endocytobiology II* (pp. 1027–1034). Berlin: de Gruyter.

Cavalier-Smith, Thomas. (2006). Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1596), 1943–1952. https://doi.org/10.1098/rspb.2006.3531

Cavalier-Smith, Thomas. (2007). The Chimaeric Origin of Mitochondria: Photosynthetic Cell Enslavement, Gene-Transfer Pressure, and Compartmentation Efficiency. In D. W. F. Martin & D. M. Müller (Eds.), *Origin of Mitochondria and Hydrogenosomes* (pp. 161–199). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-38502-8_8

Cavalier-Smith, Thomas. (2013). Symbiogenesis: Mechanisms, Evolutionary Consequences, and Systematic Implications. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 145–172. https://doi.org/10.1146/annurev-ecolsys-110411-160320

Chandler, D. E., Gumbart, J., Stack, J. D., Chipot, C., & Schulten, K. (2009). Membrane Curvature Induced by Aggregates of LH2s and Monomeric LH1s. *Biophysical Journal*, *97*(11), 2978–2984. https://doi.org/10.1016/j.bpj.2009.09.007

Chandler, D. E., Hsin, J., Harrison, C. B., Gumbart, J., & Schulten, K. (2008). Intrinsic Curvature Properties of Photosynthetic Proteins in Chromatophores. *Biophysical Journal*, *95*(6), 2822–2836. https://doi.org/10.1529/biophysj.108.132852

Cho, J.-C., & Giovannoni, S. J. (2003). Parvularcula bermudensis gen. nov., sp. nov., a marine bacterium that forms a deep branch in the alpha-Proteobacteria. *International Journal of Systematic and Evolutionary Microbiology*, *53*(Pt 4), 1031–1036. https://doi.org/10.1099/ijs.0.02566-0

Choi, D.-W., Kunz, R. C., Boyd, E. S., Semrau, J. D., Antholine, W. E., Han, J.-I., … DiSpirito, A. A. (2003). The membrane-associated methane monooxygenase (pMMO) and pMMO-NADH:quinone oxidoreductase complex from Methylococcus capsulatus Bath. *Journal of Bacteriology*, *185*(19), 5755–5764.

Chojnacka, M., Gornicka, A., Oeljeklaus, S., Warscheid, B., & Chacinska, A. (2015). Cox17 is an auxiliary factor involved in the control of the mitochondrial contact site and cristae organizing system. *Journal of Biological Chemistry*, jbc.M115.645069. https://doi.org/10.1074/jbc.M115.645069

Cogliati, S., Enriquez, J. A., & Scorrano, L. (2016). Mitochondrial Cristae: Where Beauty Meets Functionality. *Trends in Biochemical Sciences*, *41*(3), 261–273. https://doi.org/10.1016/j.tibs.2016.01.001

Cogliati, S., Frezza, C., Soriano, M. E., Varanita, T., Quintana-Cabrera, R., Corrado, M., … Scorrano, L. (2013). Mitochondrial Cristae Shape Determines Respiratory Chain Supercomplexes Assembly and Respiratory Efficiency. *Cell*, *155*(1), 160–171. https://doi.org/10.1016/j.cell.2013.08.032

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, *10*, 210. https://doi.org/10.1186/1471-2148-10-210

Daems, W. T., & Wisse, E. (1966). Shape and attachment of the cristae mitochondriales in mouse hepatic cell mitochondria. *Journal of Ultrastructure Research*, *16*(1–2), 123–140. https://doi.org/10.1016/S0022-5320(66)80027-8

D'Amici, G. M., Rinalducci, S., Murgiano, L., Italiano, F., & Zolla, L. (2010). Oligomeric Characterization of the Photosynthetic Apparatus of Rhodobacter sphaeroides R26.1 by Nondenaturing Electrophoresis Methods. *Journal of Proteome Research*, *9*(1), 192–203. https://doi.org/10.1021/pr9005052

Darshi, M., Mendiola, V. L., Mackey, M. R., Murphy, A. N., Koller, A., Perkins, G. A., … Taylor, S. S. (2011). ChChd3, an inner mitochondrial membrane protein, is essential for maintaining crista integrity and mitochondrial function. *The Journal of Biological Chemistry*, *286*(4), 2918–2932. https://doi.org/10.1074/jbc.M110.171975

Davidov, Y., Huchon, D., Koval, S. F., & Jurkevitch, E. (2006). A new alpha-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. *Environmental Microbiology*, *8*(12), 2179–2188. https://doi.org/10.1111/j.1462-2920.2006.01101.x

Davidov, Y., & Jurkevitch, E. (2009). Predation between prokaryotes and the origin of eukaryotes. *BioEssays*, *31*(7), 748–757. https://doi.org/10.1002/bies.200900018

Davies, K. M., Anselmi, C., Wittig, I., Faraldo-Gómez, J. D., & Kühlbrandt, W. (2012). Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. *Proceedings of the National Academy of Sciences of the*

*United States of America*, *109*(34), 13602–13607. https://doi.org/10.1073/pnas.1204593109

Dayhoff, M. O., & Schwartz, R. M. (1981). Evidence on the Origin of Eukaryotic Mitochondria from Protein and Nucleic Acid Sequences. *Annals of the New York Academy of Sciences*, *361*(1), 92–104. https://doi.org/10.1111/j.1749-6632.1981.tb54359.x

de Vienne, D. M., Ollier, S., & Aguileta, G. (2012). Phylo-MCOA: a fast and efficient method to detect outlier genes and species in phylogenomics using multiple co-inertia analysis. *Molecular Biology and Evolution*, *29*(6), 1587–1598. https://doi.org/10.1093/molbev/msr317

Degli Esposti, M. (2014). Bioenergetic Evolution in Proteobacteria and Mitochondria. *Genome Biology and Evolution*, *6*(12), 3238–3251. https://doi.org/10.1093/gbe/evu257

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., … Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, *3*(7), 804. https://doi.org/10.1038/s41564-018-0176-9

Deng, Y., Marko, M., Buttle, K. F., Leith, A., Mieczkowski, M., & Mannella, C. A. (1999). Cubic membrane structure in amoeba (Chaos carolinensis) mitochondria determined by electron microscopic tomography. *Journal of Structural Biology*, *127*(3), 231–239. https://doi.org/10.1006/jsbi.1999.4147

Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č., … Eliáš, M. (2015). Bacterial proteins pinpoint a single eukaryotic root. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), E693-699. https://doi.org/10.1073/pnas.1420657112

DiSpirito, A. A., Kunz, R. C., Choi, D.-W., & Zahn, J. A. (2004). Respiration in Methanotrophs. In D. Zannoni (Ed.), *Respiration in Archaea and Bacteria* (pp. 149–168). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4020-3163-2_7

Drews, G. (1992). Intracytoplasmic membranes in bacterial cells: organization, function and biosynthesis. In S. Mohan, C. S. Dow, & J. A. Cole (Eds.), *Prokaryotic Structure and Function: A New Perspective : Forty-seventh Symposium of the Society for General Microbiology Held at the University of Edinburgh, April 1991*. Cambridge University Press.

Drews, Gerhart, & Golecki, J. R. (1995). Structure, Molecular Organization, and Biosynthesis of Membranes of Purple Bacteria. In R. E. Blankenship, M. T. Madigan, & C. E. Bauer (Eds.), *Anoxygenic Photosynthetic Bacteria* (pp. 231–257). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/0-306-47954-0_12

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763.

Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Egner, A., Jakobs, S., & Hell, S. W. (2002). Fast 100-nm resolution three-dimensional microscope reveals structural plasticity of mitochondria in live yeast. *Proceedings of the National Academy of Sciences*, *99*(6), 3370–3375. https://doi.org/10.1073/pnas.052545099

Ehrenreich, A., & Widdel, F. (1994). Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. *Appl. Environ. Microbiol.*, *60*(12), 4517–4526.

Eme, L., Sharpe, S. C., Brown, M. W., & Roger, A. J. (2014). On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. *Cold Spring Harbor Perspectives in Biology*, *6*(8), a016139. https://doi.org/10.1101/cshperspect.a016139

Emelyanov, V. V. (2001a). Evolutionary relationship of Rickettsiae and mitochondria. *FEBS Letters*, *501*(1), 11–18. https://doi.org/10.1016/S0014-5793(01)02618-7

Emelyanov, V. V. (2001b). Rickettsiaceae, Rickettsia-Like Endosymbionts, and the Origin of Mitochondria. *Bioscience Reports*, *21*(1), 1–17. https://doi.org/10.1023/A:1010409415723

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319

Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., … Martin, W. (2004). A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular Biology and Evolution*, *21*(9), 1643–1660. https://doi.org/10.1093/molbev/msh160

Esser, C., Martin, W., & Dagan, T. (2007). The Origin of Mitochondria in Light of a Fluid Prokaryotic Chromosome Model. *Biology Letters*, *3*(2), 180–184. https://doi.org/10.1098/rsbl.2006.0582

Ettema, T. J. G., & Andersson, S. G. E. (2009). The alpha-proteobacteria: the Darwin finches of the bacterial world. *Biology Letters*, *5*(3), 429–432. https://doi.org/10.1098/rsbl.2008.0793

Fejes, A. P., Yi, E. C., Goodlett, D. R., & Beatty, J. T. (2003). Shotgun proteomic analysis of a chromatophore-enriched preparation from the purple phototrophic bacterium Rhodopseudomonas palustris. *Photosynthesis Research*, *78*(3), 195–203. https://doi.org/10.1023/B:PRES.0000006752.81486.74

Fenchel, T., & Bernard, C. (1993a). A purple protist. *Nature*, *362*(6418), 300–300. https://doi.org/10.1038/362300a0

Fenchel, T., & Bernard, C. (1993b). Endosymbiotic purple non-sulphur bacteria in an anaerobic ciliated protozoon. *FEMS Microbiology Letters*, *110*(1), 21–25.

Ferla, M. P., Thrash, J. C., Giovannoni, S. J., & Patrick, W. M. (2013). New rRNA Gene-Based Phylogenies of the Alphaproteobacteria Provide Perspective on Major Groups, Mitochondrial Ancestry and Phylogenetic Instability. *PLoS ONE*, *8*(12), e83383. https://doi.org/10.1371/journal.pone.0083383

Fitzpatrick, D. A., Creevey, C. J., & McInerney, J. O. (2006). Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Molecular Biology and Evolution*, *23*(1), 74–85. https://doi.org/10.1093/molbev/msj009

Foesel, B. U., Gössner, A. S., Drake, H. L., & Schramm, A. (2007). Geminicoccus roseus gen. nov., sp. nov., an aerobic phototrophic Alphaproteobacterium isolated from a marine aquaculture biofilter. *Systematic and Applied Microbiology*, *30*(8), 581–586. https://doi.org/10.1016/j.syapm.2007.05.005

Foster, P. G. (2004). Modeling compositional heterogeneity. *Systematic Biology*, *53*(3), 485–495.

Frezza, C., Cipolat, S., Martins de Brito, O., Micaroni, M., Beznoussenko, G. V., Rudka, T., … Scorrano, L. (2006). OPA1 Controls Apoptotic Cristae Remodeling Independently from Mitochondrial Fusion. *Cell*, *126*(1), 177–189. https://doi.org/10.1016/j.cell.2006.06.025

Friedman, J. R., Mourier, A., Yamada, J., McCaffery, J. M., & Nunnari, J. (2015). MICOS coordinates with respiratory complexes and lipids to establish mitochondrial inner membrane architecture. *ELife*, *4*, e07739. https://doi.org/10.7554/eLife.07739

Gabaldón, T., & Huynen, M. A. (2003). Reconstruction of the Proto-Mitochondrial Metabolism. *Science*, *301*(5633), 609–609. https://doi.org/10.1126/science.1085463

Gabaldón, T., & Huynen, M. A. (2007a). From Endosymbiont to Host-Controlled Organelle: The Hijacking of Mitochondrial Protein Synthesis and Metabolism. *PLOS Computational Biology*, *3*(11), e219. https://doi.org/10.1371/journal.pcbi.0030219

Gabaldón, T., & Huynen, M. A. (2007b). Reconstruction of ancestral proteomes. In D. A. Liberles (Ed.), *Ancestral Sequence Reconstruction*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199299188.001.0001/acprof-9780199299188-chapter-12

Galloway, C. A., Lee, H., & Yoon, Y. (2012). Mitochondrial morphology—emerging role in bioenergetics. *Free Radical Biology and Medicine*, *53*(12), 2218–2228. https://doi.org/10.1016/j.freeradbiomed.2012.09.035

Garí, E., Piedrafita, L., Aldea, M., & Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in

Saccharomyces cerevisiae. *Yeast (Chichester, England)*, *13*(9), 837–848. https://doi.org/10.1002/(SICI)1097-0061(199707)13:9<837::AID-YEA145>3.0.CO;2-T

Garrity, G., Brenner, D. J., Krieg, N. R., & Staley, J. T. (Eds.). (2005). *Bergey's Manual® of Systematic Bacteriology: Volume Two: The Proteobacteria (Part C)* (2nd ed.). Springer US. Retrieved from //www.springer.com/us/book/9780387241456

Gautier, R., Douguet, D., Antonny, B., & Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics (Oxford, England)*, *24*(18), 2101–2102. https://doi.org/10.1093/bioinformatics/btn392

Georgiades, K., Madoui, M.-A., Le, P., Robert, C., & Raoult, D. (2011). Phylogenomic Analysis of Odyssella thessalonicensis Fortifies the Common Origin of Rickettsiales, Pelagibacter ubique and Reclimonas americana Mitochondrion. *PLoS ONE*, *6*(9), e24857. https://doi.org/10.1371/journal.pone.0024857

Georgiades, K., & Raoult, D. (2011). The rhizome of Reclinomonas americana, Homo sapiens, Pediculus humanus and Saccharomyces cerevisiae mitochondria. *Biology Direct*, *6*(1), 55. https://doi.org/10.1186/1745-6150-6-55

Gietz, R. D., & Woods, R. A. (2006). Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods in Molecular Biology (Clifton, N.J.)*, *313*, 107–120. https://doi.org/10.1385/1-59259-958-3:107

Giovannoni, S. J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science*, *9*, 231–255. https://doi.org/10.1146/annurev-marine-010814-015934

Gross, J., & Bhattacharya, D. (2009). Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective. *Nature Reviews Genetics*, *10*(7), 495–505. https://doi.org/10.1038/nrg2610

Gross, J., & Bhattacharya, D. (2011). Endosymbiont or host: who drove mitochondrial and plastid evolution? *Biology Direct*, *6*(1), 12. https://doi.org/10.1186/1745-6150-6-12

Grote, J., Thrash, J. C., Huggett, M. J., Landry, Z. C., Carini, P., Giovannoni, S. J., & Rappé, M. S. (2012). Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade. *MBio*, *3*(5), e00252-12. https://doi.org/10.1128/mBio.00252-12

Guarani, V., McNeill, E. M., Paulo, J. A., Huttlin, E. L., Fröhlich, F., Gygi, S. P., … Harper, J. W. (2015). QIL1 is a novel mitochondrial protein required for MICOS complex stability and cristae morphology. *ELife*, *4*, e06265. https://doi.org/10.7554/eLife.06265

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gupta, R. S. (1995). Evolution of the chaperonin families (HSP60, HSP 10 and TCP-1) of proteins and the origin of eukaryotic cells. *Molecular Microbiology*, *15*(1), 1–11. https://doi.org/10.1111/j.1365-2958.1995.tb02216.x

Hanson, R. S., & Hanson, T. E. (1996). Methanotrophic bacteria. *Microbiological Reviews*, *60*(2), 439–471.

Harbison, A. B., Price, L. E., Flythe, M. D., & Bräuer, S. L. (2017). Micropepsis pineolensis gen. nov., sp. nov., a mildly acidophilic alphaproteobacterium isolated from a poor fen, and proposal of Micropepsaceae fam. nov. within Micropepsales ord. nov. *International Journal of Systematic and Evolutionary Microbiology*, *67*(4), 839–844. https://doi.org/10.1099/ijsem.0.001681

Harner, M. E., Unger, A.-K., Izawa, T., Walther, D. M., Özbalci, C., Geimer, S., … Neupert, W. (2014). Aim24 and MICOS modulate respiratory function, tafazzin-related cardiolipin modification and mitochondrial architecture. *ELife*, *3*, e01684. https://doi.org/10.7554/eLife.01684

Harner, M., Körner, C., Walther, D., Mokranjac, D., Kaesmacher, J., Welsch, U., … Neupert, W. (2011). The mitochondrial contact site complex, a determinant of mitochondrial architecture. *The EMBO Journal*, *30*(21), 4356–4370. https://doi.org/10.1038/emboj.2011.379

Head, B. P., Zulaika, M., Ryazantsev, S., & van der Bliek, A. M. (2011). A novel mitochondrial outer membrane protein, MOMA-1, that affects cristae morphology in Caenorhabditis elegans. *Molecular Biology of the Cell*, *22*(6), 831–841. https://doi.org/10.1091/mbc.E10-07-0600

Heinz, E., & Lithgow, T. (2014). A comprehensive analysis of the Omp85/TpsB protein superfamily structural diversity, taxonomic occurrence, and evolution. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00370

Hess, D. C., Myers, C. L., Huttenhower, C., Hibbs, M. A., Hayes, A. P., Paw, J., … Caudy, A. A. (2009). Computationally Driven, Quantitative Experiments Discover Genes Required for Mitochondrial Biogenesis. *PLoS Genet*, *5*(3), e1000407. https://doi.org/10.1371/journal.pgen.1000407

Hess, S., & Melkonian, M. (2013). The mystery of clade X: Orciraptor gen. nov. and Viridiraptor gen. nov. are highly specialised, algivorous amoeboflagellates (Glissomonadida, Cercozoa). *Protist*, *164*(5), 706–747. https://doi.org/10.1016/j.protis.2013.07.003

Hess, S., Suthaus, A., & Melkonian, M. (2015). “Candidatus Finniella” (Rickettsiales, Alphaproteobacteria), Novel Endosymbionts of Viridiraptorid Amoeboflagellates (Cercozoa, Rhizaria). *Applied and Environmental Microbiology*, *82*(2), 659–670. https://doi.org/10.1128/AEM.02680-15

Hessenberger, M., Zerbes, R. M., Rampelt, H., Kunz, S., Xavier, A. H., Purfürst, B., … Daumke, O. (2017). Regulated membrane remodeling by Mic60 controls formation of mitochondrial crista junctions. *Nature Communications*, *8*, 15258. https://doi.org/10.1038/ncomms15258

Hoffmann, H.-P., & Avers, C. J. (1973). Mitochondrion of Yeast: Ultrastructural Evidence for One Giant, Branched Organelle per Cell. *Science*, *181*(4101), 749–751. https://doi.org/10.1126/science.181.4101.749

Hofmann. (1993). TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, *374*(166). Retrieved from http://www.ch.embnet.org/documentation/mfc-35.pdf

Hoppins, S., Collins, S. R., Cassidy-Stone, A., Hummel, E., Devay, R. M., Lackner, L. L., … Nunnari, J. (2011). A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *The Journal of Cell Biology*, *195*(2), 323–340. https://doi.org/10.1083/jcb.201107053

Horvath, S. E., Rampelt, H., Oeljeklaus, S., Warscheid, B., van der Laan, M., & Pfanner, N. (2015). Role of membrane contact sites in protein import into mitochondria. *Protein Science*, *24*(3), 277–297. https://doi.org/10.1002/pro.2625

Hunter, C. N. (1995). Genetic Manipulation of the Antenna Complexes of Purple Bacteria. In R. E. Blankenship, M. T. Madigan, & C. E. Bauer (Eds.), *Anoxygenic Photosynthetic Bacteria* (pp. 473–501). Dordrecht: Springer Netherlands. https://doi.org/10.1007/0-306-47954-0_22

Huynen, M. A., Mühlmeister, M., Gotthardt, K., Guerrero-Castillo, S., & Brandt, U. (2016). Evolution and structural organization of the mitochondrial contact site (MICOS) complex and the mitochondrial intermembrane space bridging (MIB) complex. *Biochimica Et Biophysica Acta*, *1863*(1), 91–101. https://doi.org/10.1016/j.bbamcr.2015.10.009

Iba, K., Takamiya, K., Toh, Y., & Nishimura, M. (1988). Roles of bacteriochlorophyll and carotenoid synthesis in formation of intracytoplasmic membrane systems and pigment-protein complexes in an aerobic photosynthetic bacterium, Erythrobacter sp. strain OCh114. *Journal of Bacteriology*, *170*(4), 1843–1847.

Icho, T., Ikeda, T., Matsumoto, Y., Hanaoka, F., Kaji, K., & Tsuchida, N. (1994). A novel human gene that is preferentially transcribed in heart muscle. *Gene*, *144*(2), 301–306.

Ind, A. C., Porter, S. L., Brown, M. T., Byles, E. D., de Beyer, J. A., Godfrey, S. A., & Armitage, J. P. (2009). Inducible-Expression Plasmid for Rhodobacter sphaeroides and Paracoccus denitrificans. *Applied and Environmental Microbiology*, *75*(20), 6613–6615. https://doi.org/10.1128/AEM.01587-09

Itoh, K., Tamura, Y., Iijima, M., & Sesaki, H. (2013). Effects of Fcj1-Mos1 and mitochondrial division on aggregation of mitochondrial DNA nucleoids and organelle morphology. *Molecular Biology of the Cell*, *24*(12), 1842–1851. https://doi.org/10.1091/mbc.E13-03-0125

Jackson, P. J., Lewis, H. J., Tucker, J. D., Hunter, C. N., & Dickman, M. J. (2012). Quantitative proteomic analysis of intracytoplasmic membrane development in

Rhodobacter sphaeroides. *Molecular Microbiology*, *84*(6), 1062–1078. https://doi.org/10.1111/j.1365-2958.2012.08074.x

Janer, A., Prudent, J., Paupe, V., Fahiminiya, S., Majewski, J., Sgarioto, N., … Shoubridge, E. A. (2016). SLC25A46 is required for mitochondrial lipid homeostasis and cristae maintenance and is responsible for Leigh syndrome. *EMBO Molecular Medicine*, *8*(9), 1019–1038. https://doi.org/10.15252/emmm.201506159

Jaschke, P. R., Saer, R. G., Noll, S., & Beatty, J. T. (2011). Modification of the genome of Rhodobacter sphaeroides and construction of synthetic operons. *Methods in Enzymology*, *497*, 519–538. https://doi.org/10.1016/B978-0-12-385075-1.00023-8

Jayashankar, V., Mueller, I. A., & Rafelski, S. M. (2016). Shaping the multi-scale architecture of mitochondria. *Current Opinion in Cell Biology*, *38*, 45–51. https://doi.org/10.1016/j.ceb.2016.02.006

Jiao, Y., Kappler, A., Croal, L. R., & Newman, D. K. (2005). Isolation and Characterization of a Genetically Tractable Photoautotrophic Fe(II)-Oxidizing Bacterium, Rhodopseudomonas palustris Strain TIE-1. *Applied and Environmental Microbiology*, *71*(8), 4487–4496. https://doi.org/10.1128/AEM.71.8.4487-4496.2005

John, G. B., Shang, Y., Li, L., Renken, C., Mannella, C. A., Selker, J. M. L., … Zha, J. (2005). The mitochondrial inner membrane protein mitofilin controls cristae morphology. *Molecular Biology of the Cell*, *16*(3), 1543–1554. https://doi.org/10.1091/mbc.E04-08-0697

John, P., & Whatley, F. R. (1978). Paracoccus denitrificans Davis (Micrococcus denitrificans Beijerinck) as a Mitochondrion. *Advances in Botanical Research*, *4*, 51–115. https://doi.org/10.1016/S0065-2296(08)60369-9

John, Philip, & Whatley, F. R. (1975). Paracoccus denitrificans and the evolutionary origin of the mitochondrion. *Nature*, *254*(5500), 495–498. https://doi.org/10.1038/254495a0

Jun, D., Saer, R. G., Madden, J. D., & Beatty, J. T. (2014). Use of new strains of Rhodobacter sphaeroides and a modified simple culture medium to increase yield and facilitate purification of the reaction centre. *Photosynthesis Research*, *120*(1), 197–205. https://doi.org/10.1007/s11120-013-9866-6

Kannan, S., Rogozin, I. B., & Koonin, E. V. (2014). MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. *BMC Evolutionary Biology*, *14*, 237. https://doi.org/10.1186/s12862-014-0237-5

Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S. C., Petrželková, R., Eme, L., … Hampl, V. (2016). A Eukaryote without a Mitochondrial Organelle. *Current Biology*, *26*(10), 1274–1284. https://doi.org/10.1016/j.cub.2016.03.053

Karnovsky, M. (1965). A formaldehyde-glutaraldehyde fixative of high osmolarity for use in electron microscopy. *J Cell Biol*, *27*, 137A–138A.

Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, *33*(2), 511–518. https://doi.org/10.1093/nar/gki198

Kaufmann, N., Reidl, H.-H., Golecki, J. R., Garcia, A. F., & Drews, G. (1982). Differentiation of the membrane system in cells of Rhodopseudomonas capsulata after transition from chemotrophic to phototrophic growth conditions. *Archives of Microbiology*, *131*(4), 313–322. https://doi.org/10.1007/BF00411178

Kay, C. J., Lawler, K., & Kerr, I. D. (2013). Analysis of the Sam50 translocase of excavate organisms supports evolution of divergent organelles from a common endosymbiotic event. *Bioscience Reports*, *33*(6). https://doi.org/10.1042/BSR20130049

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., … Worden, A. Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol*, *12*(6), e1001889. https://doi.org/10.1371/journal.pbio.1001889

Keithly, J. S., Langreth, S. G., Buttle, K. F., & Mannella, C. A. (2005). Electron tomographic and ultrastructural analysis of the Cryptosporidium parvum relict mitochondrion, its associated membranes, and organelles. *The Journal of Eukaryotic Microbiology*, *52*(2), 132–140. https://doi.org/10.1111/j.1550-7408.2005.04-3317.x

Khan, S. R., Gaines, J., Roop, R. M., & Farrand, S. K. (2008). Broad-Host-Range Expression Vectors with Tightly Regulated Promoters and Their Use To Examine the Influence of TraR and TraM Expression on Ti Plasmid Quorum Sensing. *Applied and Environmental Microbiology*, *74*(16), 5053–5062. https://doi.org/10.1128/AEM.01098-08

Kiley, P. J., Varga, A., & Kaplan, S. (1988). Physiological and structural analysis of light-harvesting mutants of Rhodobacter sphaeroides. *Journal of Bacteriology*, *170*(3), 1103–1115.

Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., & Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell*, *27*(8), 1188–1196. https://doi.org/10.1091/mbc.E15-12-0844

Kobayashi, K., Masuda, T., Tajima, N., Wada, H., & Sato, N. (2014). Molecular Phylogeny and Intricate Evolutionary History of the Three Isofunctional Enzymes Involved in the Oxidation of Protoporphyrinogen IX. *Genome Biology and Evolution*, *6*(8), 2141–2155. https://doi.org/10.1093/gbe/evu170

Koblížek, M., Zeng, Y., Horák, A., & Oborník, M. (2013). Regressive Evolution of Photosynthesis in the Roseobacter Clade. In J. T. Beatty (Ed.), *Advances in*

*Botanical Research* (Vol. 66, pp. 385–405). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123979230000138

Konorty, M., Brumfeld, V., Vermeglio, A., Kahana, N., Medalia, O., & Minsky, A. (2009). Photosynthetic System in Blastochloris viridis Revisited. *Biochemistry*, *48*(22), 4753–4761. https://doi.org/10.1021/bi900267r

Konorty, M., Kahana, N., Linaroudis, A., Minsky, A., & Medalia, O. (2008). Structural analysis of photosynthetic membranes by cryo-electron tomography of intact Rhodopseudomonas viridis cells. *Journal of Structural Biology*, *161*(3), 393–400. https://doi.org/10.1016/j.jsb.2007.09.014

Koob, S., & Reichert, A. S. (2014). Novel intracellular functions of apolipoproteins: the ApoO protein family as constituents of the Mitofilin/MINOS complex determines cristae morphology in mitochondria. *Biological Chemistry*, *395*(3), 285–296. https://doi.org/10.1515/hsz-2013-0274

Körner, C., Barrera, M., Dukanovic, J., Eydt, K., Harner, M., Rabl, R., … Reichert, A. S. (2012). The C-terminal domain of Fcj1 is required for formation of crista junctions and interacts with the TOB/SAM complex in mitochondria. *Molecular Biology of the Cell*, *23*(11), 2143–2155. https://doi.org/10.1091/mbc.E11-10-0831

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kühlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. *BMC Biology*, *13*(1), 89. https://doi.org/10.1186/s12915-015-0201-x

Kulichevskaya, I. S., Guzev, V. S., Gorlenko, V. M., Liesack, W., & Dedysh, S. N. (2006). Rhodoblastus sphagnicola sp. nov., a novel acidophilic purple non-sulfur bacterium from Sphagnum peat bog. *International Journal of Systematic and Evolutionary Microbiology*, *56*(6), 1397–1402. https://doi.org/10.1099/ijs.0.63962-0

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Kurahashi, M., Fukunaga, Y., Harayama, S., & Yokota, A. (2008). Sneathiella glossodoripedis sp. nov., a marine alphaproteobacterium isolated from the nudibranch Glossodoris cincta, and proposal of Sneathiellales ord. nov. and Sneathiellaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, *58*(Pt 3), 548–552. https://doi.org/10.1099/ijs.0.65328-0

Kwon, K. K., Lee, H.-S., Yang, S. H., & Kim, S.-J. (2005). Kordiimonas gwangyangensis gen. nov., sp. nov., a marine bacterium isolated from marine sediments that forms a distinct phyletic lineage (Kordiimonadales ord. nov.) in the "Alphaproteobacteria." *International Journal of Systematic and Evolutionary Microbiology*, *55*(5), 2033–2037. https://doi.org/10.1099/ijs.0.63684-0

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lartillot, N., Lepage, T., & Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, *25*(17), 2286–2288. https://doi.org/10.1093/bioinformatics/btp368

Lartillot, N., & Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, *21*(6), 1095–1109. https://doi.org/10.1093/molbev/msh112

Lee, K.-B., Liu, C.-T., Anzai, Y., Kim, H., Aono, T., & Oyaizu, H. (2005). The hierarchical system of the “Alphaproteobacteria”: description of Hyphomonadaceae fam. nov., Xanthobacteraceae fam. nov. and Erythrobacteraceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, *55*(Pt 5), 1907–1919. https://doi.org/10.1099/ijs.0.63663-0

Leigh, J. W. (2009). *Congruence in phylogenomic data: Exploring artefacts in deep eukaryotic phylogeny*. Dalhousie University, Halifax, Canada.

Liesa, M., & Shirihai, O. S. (2013). Mitochondrial Dynamics in the Regulation of Nutrient Utilization and Energy Expenditure. *Cell Metabolism*, *17*(4), 491–506. https://doi.org/10.1016/j.cmet.2013.03.002

Liu, Y., Zhou, Z., Pan, J., Baker, B. J., Gu, J.-D., & Li, M. (2018). Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *The ISME Journal*, *12*(4), 1021–1031. https://doi.org/10.1038/s41396-018-0060-x

Lohße, A., Borg, S., Raschdorf, O., Kolinko, I., Tompa, É., Pósfai, M., … Schüler, D. (2014). Genetic Dissection of the mamAB and mms6 Operons Reveals a Gene Set Essential for Magnetosome Biogenesis in Magnetospirillum gryphiswaldense. *Journal of Bacteriology*, *196*(14), 2658–2669. https://doi.org/10.1128/JB.01716-14

Lohße, A., Ullrich, S., Katzmann, E., Borg, S., Wanner, G., Richter, M., … Schüler, D. (2011). Functional Analysis of the Magnetosome Island in Magnetospirillum gryphiswaldense: The mamAB Operon Is Sufficient for Magnetite Biomineralization. *PLoS ONE*, *6*(10), e25561. https://doi.org/10.1371/journal.pone.0025561

López-García, P., & Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. *BioEssays*, *28*(5), 525–533. https://doi.org/10.1002/bies.20413

López-García, P., & Moreira, D. (2015). Open questions on the origin of eukaryotes. *Trends in Ecology & Evolution*, *30*(11), 697–708. https://doi.org/10.1016/j.tree.2015.09.005

Luo, H. (2015). Evolutionary origin of a streamlined marine bacterioplankton lineage. *The ISME Journal*, *9*(6), 1423–1433. https://doi.org/10.1038/ismej.2014.227

Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science (New York, N.Y.)*, *252*(5009), 1162–1164. https://doi.org/10.1126/science.252.5009.1162

Maddison, W. P., & Maddison, D. R. (2014). Mesquite: a modular system for evolutionary analysis (Version 3.01). Retrieved from http://mesquiteproject.org

Madigan, M. T., Jung, D. O., & Madigan, M. T. (2009). An Overview of Purple Bacteria: Systematics, Physiology, and Habitats. In *The Purple Phototrophic Bacteria* (pp. 1–15). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-8815-5_1

Mannella, C. A., Marko, M., Penczek, P., Barnard, D., & Frank, J. (1994). The internal compartmentation of rat-liver mitochondria: tomographic study using the high-voltage transmission electron microscope. *Microscopy Research and Technique*, *27*(4), 278–283. https://doi.org/10.1002/jemt.1070270403

Mannella, C. A., Pfeiffer, D. R., Bradshaw, P. C., Moraru, I. I., Slepchenko, B., Loew, L. M., … Marko, M. (2001). Topology of the mitochondrial inner membrane: dynamics and bioenergetic implications. *IUBMB Life*, *52*(3–5), 93–100. https://doi.org/10.1080/15216540152845885

Mannella, Carmen A. (2000). Introduction: Our Changing Views of Mitochondria. *Journal of Bioenergetics and Biomembranes*, *32*(1), 1–4. https://doi.org/10.1023/A:1005562109678

Mannella, Carmen A. (2006). The relevance of mitochondrial membrane topology to mitochondrial function. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1762*(2), 140–147. https://doi.org/10.1016/j.bbadis.2005.07.001

Mannella, Carmen A., Lederer, W. J., & Jafri, M. S. (2013). The connection between inner membrane topology and mitochondrial function. *Journal of Molecular and Cellular Cardiology*, *62*, 51–57. https://doi.org/10.1016/j.yjmcc.2013.05.001

Margulis, L. (1970). *Origin of eukaryotic cells*. New Haven, CT: Yale University Press.

Martijn, J., Schulz, F., Zaremba-Niedzwiedzka, K., Viklund, J., Stepanauskas, R., Andersson, S. G. E., … Ettema, T. J. G. (2015). Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into Rickettsiaceae evolution. *The ISME Journal*. https://doi.org/10.1038/ismej.2015.46

Martijn, J., Vosseberg, J., Guy, L., Offre, P., & Ettema, T. J. G. (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*, *557*(7703), 101–105. https://doi.org/10.1038/s41586-018-0059-5

Martin, W., & Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature*, *392*(6671), 37–41. https://doi.org/10.1038/32096

Meeusen, S., DeVay, R., Block, J., Cassidy-Stone, A., Wayson, S., McCaffery, J. M., & Nunnari, J. (2006). Mitochondrial Inner-Membrane Fusion and Crista Maintenance Requires the Dynamin-Related GTPase Mgm1. *Cell*, *127*(2), 383–395. https://doi.org/10.1016/j.cell.2006.09.021

Mehrshad, M., Amoozegar, M. A., Ghai, R., Fazeli, S. A. S., & Rodriguez-Valera, F. (2016). Genome Reconstruction from Metagenomic Data Sets Reveals Novel Microbes in the Brackish Waters of the Caspian Sea. *Appl. Environ. Microbiol.*, *82*(5), 1599–1612. https://doi.org/10.1128/AEM.03381-15

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Mishra, P., & Chan, D. C. (2016). Metabolic regulation of mitochondrial dynamics. *The Journal of Cell Biology*, *212*(4), 379–387. https://doi.org/10.1083/jcb.201511036

Mogi, T., & Kita, K. (2010). Diversity in mitochondrial metabolic pathways in parasitic protists Plasmodium and Cryptosporidium. *Parasitology International*, *59*(3), 305–312. https://doi.org/10.1016/j.parint.2010.04.005

Montagna, M., Sassera, D., Epis, S., Bazzocchi, C., Vannini, C., Lo, N., … Bandi, C. (2013). "Candidatus Midichloriaceae" fam. nov. (Rickettsiales), an Ecologically Widespread Clade of Intracellular Alphaproteobacteria. *Applied and Environmental Microbiology*, *79*(10), 3241–3248. https://doi.org/10.1128/AEM.03971-12

Moreira, D., & López-García, P. (1998). Symbiosis Between Methanogenic Archaea and δ-Proteobacteria as the Origin of Eukaryotes: The Syntrophic Hypothesis. *Journal of Molecular Evolution*, *47*(5), 517–530. https://doi.org/10.1007/PL00006408

Mühleip, A. W., Dewar, C. E., Schnaufer, A., Kühlbrandt, W., & Davies, K. M. (2017). In situ structure of trypanosomal ATP synthase dimer reveals a unique arrangement of catalytic subunits. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(5), 992–997. https://doi.org/10.1073/pnas.1612386114

Mühleip, A. W., Joos, F., Wigge, C., Frangakis, A. S., Kühlbrandt, W., & Davies, K. M. (2016). Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria. *Proceedings of the National Academy of Sciences*, 201525430. https://doi.org/10.1073/pnas.1525430113

Müller, M., Mentel, M., Hellemond, J. J. van, Henze, K., Woehle, C., Gould, S. B., … Martin, W. F. (2012). Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. *Microbiology and Molecular Biology Reviews*, *76*(2), 444–495. https://doi.org/10.1128/MMBR.05024-11

Munn, E. A. (1974). *The structure of mitochondria*. Academic Press.

Muñoz-Gómez, S. A., Slamovits, C. H., Dacks, J. B., Baier, K. A., Spencer, K. D., & Wideman, J. G. (2015). Ancient Homology of the Mitochondrial Contact Site and Cristae Organizing System Points to an Endosymbiotic Origin of Mitochondrial Cristae. *Current Biology*, *25*(11), 1489–1495. https://doi.org/10.1016/j.cub.2015.04.006

Muñoz-Gómez, S. A., Slamovits, C. H., Dacks, J. B., & Wideman, J. G. (2015). The evolution of MICOS: Ancestral and derived functions and interactions. *Communicative & Integrative Biology*, *8*(6), e1094593. https://doi.org/10.1080/19420889.2015.1094593

Muñoz-Gómez, S. A., Wideman, J. G., Roger, A. J., & Slamovits, C. H. (2017). The Origin of Mitochondrial Cristae from Alphaproteobacteria. *Molecular Biology and Evolution*, *34*(4), 943–956. https://doi.org/10.1093/molbev/msw298

Myers, J. A., Curtis, B. S., & Curtis, W. R. (2013). Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophysics*, *6*(1), 4. https://doi.org/10.1186/2046-1682-6-4

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nicastro, D., Frangakis, A. S., Typke, D., & Baumeister, W. (2000). Cryo-electron tomography of neurospora mitochondria. *Journal of Structural Biology*, *129*(1), 48–56. https://doi.org/10.1006/jsbi.1999.4204

Niederman, R. A. (2006). Structure, Function and Formation of Bacterial Intracytoplasmic Membranes. In J. M. Shively (Ed.), *Complex Intracellular Structures in Prokaryotes* (pp. 193–227). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/7171_025

Okamoto, K., & Shaw, J. M. (2005). Mitochondrial Morphology and Dynamics in Yeast and Multicellular Eukaryotes. *Annual Review of Genetics*, *39*(1), 503–536. https://doi.org/10.1146/annurev.genet.38.072902.093019

Olichon, A., Baricault, L., Gas, N., Guillou, E., Valette, A., Belenguer, P., & Lenaers, G. (2003). Loss of OPA1 Perturbates the Mitochondrial Inner Membrane Structure and Integrity, Leading to Cytochrome c Release and Apoptosis. *Journal of Biological Chemistry*, *278*(10), 7743–7746. https://doi.org/10.1074/jbc.C200677200

Olsen, G. J., Woese, C. R., & Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *Journal of Bacteriology*, *176*(1), 1–6. https://doi.org/10.1128/jb.176.1.1-6.1994

Ott, C., Dorsch, E., Fraunholz, M., Straub, S., & Kozjak-Pavlovic, V. (2015). Detailed Analysis of the Human Mitochondrial Contact Site Complex Indicate a Hierarchy of Subunits. *PLoS ONE*, *10*(3), e0120213. https://doi.org/10.1371/journal.pone.0120213

Ott, C., Ross, K., Straub, S., Thiede, B., Götz, M., Goosmann, C., … Kozjak-Pavlovic, V. (2012). Sam50 functions in mitochondrial intermembrane space bridging and biogenesis of respiratory complexes. *Molecular and Cellular Biology*, *32*(6), 1173–1188. https://doi.org/10.1128/MCB.06388-11

Paradies, G., Paradies, V., De Benedictis, V., Ruggiero, F. M., & Petrosillo, G. (2014). Functional role of cardiolipin in mitochondrial bioenergetics. *Biochimica Et Biophysica Acta*, *1837*(4), 408–417. https://doi.org/10.1016/j.bbabio.2013.10.006

Park, Y.-U., Jeong, J., Lee, H., Mun, J. Y., Kim, J.-H., Lee, J. S., … Park, S. K. (2010). Disrupted-in-schizophrenia 1 (DISC1) plays essential roles in mitochondria in collaboration with Mitofilin. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17785–17790. https://doi.org/10.1073/pnas.1004361107

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Paschen, S. A., Waizenegger, T., Stan, T., Preuss, M., Cyrklaff, M., Hell, K., … Neupert, W. (2003). Evolutionary conservation of biogenesis of β-barrel membrane proteins. *Nature*, *426*(6968), 862–866. https://doi.org/10.1038/nature02208

Paumard, P., Vaillier, J., Coulary, B., Schaeffer, J., Soubannier, V., Mueller, D. M., … Velours, J. (2002). The ATP synthase is involved in generating mitochondrial cristae morphology. *The EMBO Journal*, *21*(3), 221–230. https://doi.org/10.1093/emboj/21.3.221

Perkins, G., Renken, C., Martone, M. E., Young, S. J., Ellisman, M., & Frey, T. (1997). Electron Tomography of Neuronal Mitochondria: Three-Dimensional Structure and Organization of Cristae and Membrane Contacts. *Journal of Structural Biology*, *119*(3), 260–272. https://doi.org/10.1006/jsbi.1997.3885

Petersen, J., Brinkmann, H., Bunk, B., Michael, V., Päuker, O., & Pradella, S. (2012). Think pink: photosynthesis, plasmids and the Roseobacter clade. *Environmental Microbiology*, *14*(10), 2661–2672. https://doi.org/10.1111/j.1462-2920.2012.02806.x

Petersen, J., Frank, O., Göker, M., & Pradella, S. (2013). Extrachromosomal, extraordinary and essential—the plasmids of the Roseobacter clade. *Applied Microbiology and Biotechnology*, *97*(7), 2805–2815. https://doi.org/10.1007/s00253-013-4746-8

Pfanner, N., Laan, M. van der, Amati, P., Capaldi, R. A., Caudy, A. A., Chacinska, A., … Nunnari, J. (2014). Uniform nomenclature for the mitochondrial contact site and cristae organizing system. *The Journal of Cell Biology*, *204*(7), 1083–1086. https://doi.org/10.1083/jcb.201401006

Pinevich, A. V. (1997). Intracytoplasmic membrane structures in bacteria. *Endocytobiosis and Cell Research*, *12*(1), 9–40.

Ponce-Toledo, R. I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K., & Moreira, D. (2017). An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Current Biology*, *27*(3), 386–391. https://doi.org/10.1016/j.cub.2016.11.056

Proença, D. N., Whitman, W. B., Varghese, N., Shapiro, N., Woyke, T., Kyrpides, N. C., & Morais, P. V. (2017). Arboriscoccus pini gen. nov., sp. nov., an endophyte from a pine tree of the class Alphaproteobacteria, emended description of Geminicoccus roseus, and proposal of Geminicoccaceae fam. nov. *Systematic and Applied Microbiology*. https://doi.org/10.1016/j.syapm.2017.11.006

Qian, P., Bullough, P. A., & Hunter, C. N. (2008). Three-dimensional Reconstruction of a Membrane-bending Complex THE RC-LH1-PufX CORE DIMER OF RHODOBACTER SPHAEROIDES. *Journal of Biological Chemistry*, *283*(20), 14002–14011. https://doi.org/10.1074/jbc.M800625200

Quandt, J., & Hynes, M. F. (1993). Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene*, *127*(1), 15–21. https://doi.org/10.1016/0378-1119(93)90611-6

Rabl, R., Soubannier, V., Scholz, R., Vogel, F., Mendl, N., Vasiljev-Neumeyer, A., … Reichert, A. S. (2009). Formation of cristae and crista junctions in mitochondria depends on antagonism between Fcj1 and Su e/g. *The Journal of Cell Biology*, *185*(6), 1047–1063. https://doi.org/10.1083/jcb.200811099

Rampelt, H., Bohnert, M., Zerbes, R. M., Horvath, S. E., Warscheid, B., Pfanner, N., & van der Laan, M. (2017). Mic10, a Core Subunit of the Mitochondrial Contact Site and Cristae Organizing System, Interacts with the Dimeric F1Fo-ATP Synthase. *Journal of Molecular Biology*, *429*(8), 1162–1170. https://doi.org/10.1016/j.jmb.2017.03.006

Rampelt, H., Zerbes, R. M., van der Laan, M., & Pfanner, N. (2016). Role of the mitochondrial contact site and cristae organizing system in membrane architecture and dynamics. *Biochimica Et Biophysica Acta*. https://doi.org/10.1016/j.bbamcr.2016.05.020

Rodríguez-Ezpeleta, N., & Embley, T. M. (2012). The SAR11 Group of Alpha-Proteobacteria Is Not Related to the Origin of Mitochondria. *PLoS ONE*, *7*(1), e30520. https://doi.org/10.1371/journal.pone.0030520

Roger, A. J., Muñoz-Gómez, S. A., & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Current Biology*, *27*(21), R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E., & Thompson, F. (Eds.). (2014). *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (4th ed.). Berlin Heidelberg: Springer-Verlag. Retrieved from //www.springer.com/gp/book/9783642301964

Sabaty, M., Gagnon, J., & Verméglio, A. (1994). Induction by nitrate of cytoplasmic and periplasmic proteins in the photodenitrifier Rhodobacter sphaeroides forma sp.

denitrificans under anaerobic or aerobic condition. *Archives of Microbiology*, *162*(5), 335–343.

Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, *14*(3), 225-IN6. https://doi.org/10.1016/0022-5193(67)90079-3

Saghaï, A., Gutiérrez-Preciado, A., Deschamps, P., Moreira, D., Bertolino, P., Ragon, M., & López-García, P. (2017). Unveiling microbial interactions in stratified mat communities from a warm saline shallow pond. *Environmental Microbiology*, *19*(6), 2405–2421. https://doi.org/10.1111/1462-2920.13754

Saghaï, A., Zivanovic, Y., Zeyen, N., Moreira, D., Benzerara, K., Deschamps, P., … López-García, P. (2015). Metagenome-based diversity analyses suggest a significant contribution of non-cyanobacterial lineages to carbonate precipitation in modern microbialites. *Frontiers in Microbiology*, *6*, 797. https://doi.org/10.3389/fmicb.2015.00797

Santos, H. A., & Massard, C. L. (2014). The Family Holosporaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 237–246). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_264

Sassera, D., Lo, N., Epis, S., D'Auria, G., Montagna, M., Comandatore, F., … Bandi, C. (2011). Phylogenomic evidence for the presence of a flagellum and cbb(3) oxidase in the free-living mitochondrial ancestor. *Molecular Biology and Evolution*, *28*(12), 3285–3296. https://doi.org/10.1093/molbev/msr159

Scheuring, S., Nevo, R., Liu, L.-N., Mangenot, S., Charuvi, D., Boudier, T., … Reich, Z. (2014). The architecture of Rhodobacter sphaeroides chromatophores. *Biochimica Et Biophysica Acta*, *1837*(8), 1263–1270. https://doi.org/10.1016/j.bbabio.2014.03.011

Schüler, D. (2008). Genetics and cell biology of magnetosome formation in magnetotactic bacteria. *FEMS Microbiology Reviews*, *32*(4), 654–672. https://doi.org/10.1111/j.1574-6976.2008.00116.x

Schwartz, R. M., & Dayhoff, M. O. (1978). Origins of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts. *Science*, *199*(4327), 395–403. https://doi.org/10.1126/science.202030

Scott, D., Brannan, J., & Higgins, I. J. (1981). The Effect of Growth Conditions on Intracytoplasmic Membranes and Methane Mono-oxygenase Activities in Methylosinus trichosporium OB3b. *Microbiology*, *125*(1), 63–72. https://doi.org/10.1099/00221287-125-1-63

Searcy, D. G. (1992). Origins of mitochondria and chloroplasts from sulphur-based symbioses. In H. Hartman & K. Matsuno (Eds.), *The Origin and Evolution of the Cell* (pp. 47–78). Singapore: World Scientific.

Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P., & Baker, B. J. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with

pathways for acetogenesis and sulfur reduction. *The ISME Journal*, *10*(7), 1696–1705. https://doi.org/10.1038/ismej.2015.233

Selao, T. T., Branca, R., Chae, P. S., Lehtiö, J., Gellman, S. H., Rasmussen, S. G. F., … Norén, A. (2011). Identification of Chromatophore Membrane Protein Complexes Formed under Different Nitrogen Availability Conditions in Rhodospirillum rubrum. *Journal of Proteome Research*, *10*(6), 2703–2714. https://doi.org/10.1021/pr100838x

Seravin, L. N. (1993). [The basic types and forms of the fine structure of mitochondrial cristae: the degree of their evolutionary stability (capacity for morphological transformations)]. *Tsitologiia*, *35*(4), 3–34.

Serricchio, M., Vissa, A., Kim, P. K., Yip, C. M., & McQuibban, G. A. (2018). Cardiolipin synthesizing enzymes form a complex that interacts with cardiolipin-dependent membrane organizing proteins. *Biochimica Et Biophysica Acta*, *1863*(4), 447–457. https://doi.org/10.1016/j.bbalip.2018.01.007

Simmerman, R. F., Dave, A. M., & Bruce, B. D. (2014). Structure and function of POTRA domains of Omp85/TPS superfamily. *International Review of Cell and Molecular Biology*, *308*, 1–34. https://doi.org/10.1016/B978-0-12-800097-7.00001-4

Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N., & Martin, W. F. (2016). Lokiarchaeon is hydrogen dependent. *Nature Microbiology*, *1*, 16034. https://doi.org/10.1038/nmicrobiol.2016.34

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., … Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, *521*(7551), 173–179. https://doi.org/10.1038/nature14447

Speijer, D. (2017). Alternating terminal electron-acceptors at the basis of symbiogenesis: How oxygen ignited eukaryotic evolution. *BioEssays*, *39*(2), 1600174. https://doi.org/10.1002/bies.201600174

Spieck, E, Ehrich, S., Aamand, J., & Bock, E. (1998). Isolation and immunocytochemical location of the nitrite-oxidizing system in nitrospira moscoviensis. *Archives of Microbiology*, *169*(3), 225–230.

Spieck, Eva, Aamand, J., Bartosch, S., & Bock, E. (1996). Immunocytochemical detection and location of the membrane-bound nitrite oxidoreductase in cells of Nitrobacter and Nitrospira. *FEMS Microbiology Letters*, *139*(1), 71–76. https://doi.org/10.1016/0378-1097(96)00123-1

Stairs, C. W., Leger, M. M., & Roger, A. J. (2015). Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Phil. Trans. R. Soc. B*, *370*(1678), 20140326. https://doi.org/10.1098/rstb.2014.0326

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stenberg, F., Chovanec, P., Maslen, S. L., Robinson, C. V., Ilag, L. L., Heijne, G. von, & Daley, D. O. (2005). Protein Complexes of the Escherichia coli Cell Envelope. *Journal of Biological Chemistry*, *280*(41), 34409–34419. https://doi.org/10.1074/jbc.M506479200

Stewart, K. D., & Mattox, K. (1980). Phylogeny of phytopflagellates. In E. R. Cox (Ed.), *Phytoflagellates* (Vol. 2, pp. 433–462). New York: Elsevier/North-Holland.

Stewart, K. D., & Mattox, K. R. (1984). The case for a polyphyletic origin of mitochondria: Morphological and molecular comparisons. *Journal of Molecular Evolution*, *21*(1), 54–57. https://doi.org/10.1007/BF02100627

Strauss, M., Hofhaus, G., Schröder, R. R., & Kühlbrandt, W. (2008). Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. *The EMBO Journal*, *27*(7), 1154–1160. https://doi.org/10.1038/emboj.2008.35

Susko, E., Lincker, L., & Roger, A. J. (2018). Accelerated Estimation of Frequency Classes in Site-heterogeneous Profile Mixture Models. *Molecular Biology and Evolution*, *may026*. https://doi.org/10.1093/molbev/msy026

Susko, E., & Roger, A. J. (2007). On reduced amino acid alphabets for phylogenetic inference. *Molecular Biology and Evolution*, *24*(9), 2139–2150. https://doi.org/10.1093/molbev/msm144

Swingley, W. D., Blankenship, R. E., & Raymond, J. (2009). Evolutionary Relationships Among Purple Photosynthetic Bacteria and the Origin of Proteobacterial Photosynthetic Systems. In C. N. Hunter, F. Daldal, M. C. Thurnauer, & J. T. Beatty (Eds.), *The Purple Phototrophic Bacteria* (pp. 17–29). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8815-5_2

Szokoli, F., Castelli, M., Sabaneyeva, E., Schrallhammer, M., Krenek, S., Doak, T. G., … Petroni, G. (2016). Disentangling the Taxonomy of Rickettsiales and Description of Two Novel Symbionts (“Candidatus Bealeia paramacronuclearis” and “Candidatus Fokinia cryptica”) Sharing the Cytoplasm of the Ciliate Protist Paramecium biaurelia. *Applied and Environmental Microbiology*, *82*(24), 7236–7247. https://doi.org/10.1128/AEM.02284-16

Tamas, I., Smirnova, A. V., He, Z., & Dunfield, P. F. (2014). The (d)evolution of methanotrophy in the Beijerinckiaceae—a comparative genomics analysis. *The ISME Journal*, *8*(2), 369–382. https://doi.org/10.1038/ismej.2013.145

Tarasenko, D., Barbot, M., Jans, D. C., Kroppen, B., Sadowski, B., Heim, G., … Meinecke, M. (2017). The MICOS component Mic60 displays a conserved membrane-bending activity that is necessary for normal cristae morphology. *J Cell Biol*, jcb.201609046. https://doi.org/10.1083/jcb.201609046

Taylor, F. J. R. (1976). Flagellate Phylogeny: A Study in Conflicts*. *The Journal of Protozoology*, *23*(1), 28–40. https://doi.org/10.1111/j.1550-7408.1976.tb05242.x

Thiergart, T., Landan, G., Schenk, M., Dagan, T., & Martin, W. F. (2012). An Evolutionary Network of Genes Present in the Eukaryote Common Ancestor

Polls Genomes on Eukaryotic and Mitochondrial Origin. *Genome Biology and Evolution*, *4*(4), 466–485. https://doi.org/10.1093/gbe/evs018

Thrash, J. C., Boyd, A., Huggett, M. J., Grote, J., Carini, P., Yoder, R. J., … Giovannoni, S. J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Scientific Reports*, *1*, 13. https://doi.org/10.1038/srep00013

Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M., … Lucocq, J. M. (2003). Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. *Nature*, *426*(6963), 172–176. https://doi.org/10.1038/nature01945

Tucker, J. D., Siebert, C. A., Escalante, M., Adams, P. G., Olsen, J. D., Otto, C., … Hunter, C. N. (2010). Membrane invagination in Rhodobacter sphaeroides is initiated at curved regions of the cytoplasmic membrane, then forms both budded and fully detached spherical vesicles. *Molecular Microbiology*, *76*(4), 833–847. https://doi.org/10.1111/j.1365-2958.2010.07153.x

Tully, B. J., Sachdeva, R., Graham, E. D., & Heidelberg, J. F. (2017). 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ*, *5*, e3558. https://doi.org/10.7717/peerj.3558

Uni, S., Iseki, M., Maekawa, T., Moriya, K., & Takada, S. (1987). Ultrastructure of Cryptosporidium muris (strain RN 66) parasitizing the murine stomach. *Parasitology Research*, *74*(2), 123–132.

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171–180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

van der Laan, M., Bohnert, M., Wiedemann, N., & Pfanner, N. (2012). Role of MINOS in mitochondrial membrane architecture and biogenesis. *Trends in Cell Biology*, *22*(4), 185–192. https://doi.org/10.1016/j.tcb.2012.01.004

Varabyova, A., Topf, U., Kwiatkowska, P., Wrobel, L., Kaus-Drobek, M., & Chacinska, A. (2013). Mia40 and MINOS act in parallel with Ccs1 in the biogenesis of mitochondrial Sod1. *The FEBS Journal*. https://doi.org/10.1111/febs.12409

Varanita, T., Soriano, M. E., Romanello, V., Zaglia, T., Quintana-Cabrera, R., Semenzato, M., … Scorrano, L. (2015). The Opa1-Dependent Mitochondrial Cristae Remodeling Pathway Controls Atrophic, Apoptotic, and Ischemic Tissue Damage. *Cell Metabolism*, *21*(6), 834–844. https://doi.org/10.1016/j.cmet.2015.05.007

Venkata Ramana, V., Kalyana Chakravarthy, S., Ramaprasad, E. V. V., Thiel, V., Imhoff, J. F., Sasikala, C., & Ramana, C. V. (2013). Emended description of the genus Rhodothalassium Imhoff et al., 1998 and proposal of Rhodothalassiaceae fam. nov. and Rhodothalassiales ord. nov. *Systematic and Applied Microbiology*, *36*(1), 28–32. https://doi.org/10.1016/j.syapm.2012.09.003

Verbruggen, H. (2018). SiteStripper version 1.03. http://www.phycoweb.net/software.

Verméglio, A., Joliot, P., & Joliot, A. (1995). Organization of Electron Transfer Components and Supercomplexes. In R. E. Blankenship, M. T. Madigan, & C. E. Bauer (Eds.), *Anoxygenic Photosynthetic Bacteria* (pp. 279–295). Dordrecht: Springer Netherlands. https://doi.org/10.1007/0-306-47954-0_14

Viale, A. M., & Arakaki, A. K. (1994). The chaperone connection to the origins of the eukaryotic organelles. *FEBS Letters*, *341*(2), 146–151. https://doi.org/10.1016/0014-5793(94)80446-X

Viklund, J., Ettema, T. J. G., & Andersson, S. G. E. (2012). Independent Genome Reduction and Phylogenetic Reclassification of the Oceanic SAR11 Clade. *Molecular Biology and Evolution*, *29*(2), 599–615. https://doi.org/10.1093/molbev/msr203

Viklund, J., Martijn, J., Ettema, T. J. G., & Andersson, S. G. E. (2013). Comparative and phylogenomic evidence that the alphaproteobacterium HIMB59 is not a member of the oceanic SAR11 clade. *PloS One*, *8*(11), e78858. https://doi.org/10.1371/journal.pone.0078858

Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M., & Rapoport, T. A. (2006). A Class of Membrane Proteins Shaping the Tubular Endoplasmic Reticulum. *Cell*, *124*(3), 573–586. https://doi.org/10.1016/j.cell.2005.11.047

Vogel, F., Bornhövd, C., Neupert, W., & Reichert, A. S. (2006). Dynamic subcompartmentalization of the mitochondrial inner membrane. *The Journal of Cell Biology*, *175*(2), 237–247. https://doi.org/10.1083/jcb.200605138

von der Malsburg, K., Müller, J. M., Bohnert, M., Oeljeklaus, S., Kwiatkowska, P., Becker, T., … van der Laan, M. (2011). Dual role of mitofilin in mitochondrial membrane organization and protein biogenesis. *Developmental Cell*, *21*(4), 694–707. https://doi.org/10.1016/j.devcel.2011.08.026

Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M., & Tommassen, J. (2003). Role of a Highly Conserved Bacterial Protein in Outer Membrane Protein Assembly. *Science*, *299*(5604), 262–265. https://doi.org/10.1126/science.1078973

Wall, J. D., Weaver, P. F., & Gest, H. (1975). Gene transfer agents, bacteriophages, and bacteriocins of Rhodopseudomonas capsulata. *Archives of Microbiology*, *105*(1), 217–224. https://doi.org/10.1007/BF00447140

Wang, H.-C., Minh, B. Q., Susko, E., & Roger, A. J. (2018). Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology*, *67*(2), 216–235. https://doi.org/10.1093/sysbio/syx068

Wang, Z., & Wu, M. (2013). A Phylum-Level Bacterial Phylogenetic Marker Database. *Molecular Biology and Evolution*, *30*(6), 1258–1262. https://doi.org/10.1093/molbev/mst059

Wang, Z., & Wu, M. (2014). Phylogenomic Reconstruction Indicates Mitochondrial Ancestor Was an Energy Parasite. *PLoS ONE*, *9*(10), e110685. https://doi.org/10.1371/journal.pone.0110685

Wang, Z., & Wu, M. (2015). An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Scientific Reports*, *5*, 7949. https://doi.org/10.1038/srep07949

Ward, B. B., Arp, D. J., & Klotz, M. G. (2011). *Nitrification*. American Society for Microbiology Press.

Watson, S. W., & Mandel, M. (1971). Comparison of the Morphology and Deoxyribonucleic Acid Composition of 27 Strains of Nitrifying Bacteria. *Journal of Bacteriology*, *107*(2), 563–569.

Weaver, P. F., Wall, J. D., & Gest, H. (1975). Characterization of Rhodopseudomonas capsulata. *Archives of Microbiology*, *105*(1), 207–216. https://doi.org/10.1007/BF00447139

Welander, P. V., Doughty, D. M., Wu, C.-H., Mehay, S., Summons, R. E., & Newman, D. K. (2012). Identification and characterization of Rhodopseudomonas palustris TIE-1 hopanoid biosynthesis mutants. *Geobiology*, *10*(2), 163–177. https://doi.org/10.1111/j.1472-4669.2011.00314.x

Westermann, B. (2012). Bioenergetic role of mitochondrial fusion and fission. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1817*(10), 1833–1838. https://doi.org/10.1016/j.bbabio.2012.02.033

Whatley, J. M., John, P., & Whatley, F. R. (1979). From Extracellular to Intracellular: The Establishment of Mitochondria and Chloroplasts. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *204*(1155), 165–187. https://doi.org/10.1098/rspb.1979.0020

Whitman, W. B. (Ed.). (2015). *Bergey's Manual of Systematics of Archaea and Bacteria*. Wiley. Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/9781118960608

Wideman, J. G., Gawryluk, R. M. R., Gray, M. W., & Dacks, J. B. (2013). The ancient and widespread nature of the ER-mitochondria encounter structure. *Molecular Biology and Evolution*, *30*(9), 2044–2049. https://doi.org/10.1093/molbev/mst120

Wideman, J. G., & Muñoz-Gómez, S. A. (2016). The evolution of ERMIONE in mitochondrial biogenesis and lipid homeostasis: An evolutionary view from comparative cell biology. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. https://doi.org/10.1016/j.bbalip.2016.01.015

Wiese, J., Thiel, V., Gärtner, A., Schmaljohann, R., & Imhoff, J. F. (2009). Kiloniella laminariae gen. nov., sp. nov., an alphaproteobacterium from the marine macroalga Laminaria saccharina. *International Journal of Systematic and Evolutionary Microbiology*, *59*(2), 350–356. https://doi.org/10.1099/ijs.0.001651-0

Williams, K. P., Sobral, B. W., & Dickerman, A. W. (2007). A robust species tree for the alphaproteobacteria. *Journal of Bacteriology*, *189*(13), 4578–4586. https://doi.org/10.1128/JB.00269-07

Williams, R. J. P. (2000). Mitochondria and chloroplasts: localized and delocalized bioenergetic transduction. *Trends in Biochemical Sciences*, *25*(10), 479. https://doi.org/10.1016/S0968-0004(00)01675-3

Williams, T. J., Lefèvre, C. T., Zhao, W., Beveridge, T. J., & Bazylinski, D. A. (2012). Magnetospira thiophila gen. nov., sp. nov., a marine magnetotactic bacterium that represents a novel lineage within the Rhodospirillaceae (Alphaproteobacteria). *International Journal of Systematic and Evolutionary Microbiology*, *62*(10), 2443–2450. https://doi.org/10.1099/ijs.0.037697-0

Woese, C. R. (1977). Endosymbionts and mitochondrial origins. *Journal of Molecular Evolution*, *10*(2), 93–96. https://doi.org/10.1007/BF01751802

Wu, M., Chatterji, S., & Eisen, J. A. (2012). Accounting For Alignment Uncertainty in Phylogenomics. *PLOS ONE*, *7*(1), e30288. https://doi.org/10.1371/journal.pone.0030288

Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., … Eisen, J. A. (2004). Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology*, *2*(3), E69. https://doi.org/10.1371/journal.pbio.0020069

Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, *32*(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638

Wurm, C. A., & Jakobs, S. (2006). Differential protein distributions define two sub-compartments of the mitochondrial inner membrane in yeast. *FEBS Letters*, *580*(24), 5628–5634. https://doi.org/10.1016/j.febslet.2006.09.012

Xie, J., Marusich, M. F., Souda, P., Whitelegge, J., & Capaldi, R. A. (2007). The mitochondrial inner membrane protein mitofilin exists as a complex with SAM50, metaxins 1 and 2, coiled-coil-helix coiled-coil-helix domain-containing protein 3 and 6 and DnaJC11. *FEBS Letters*, *581*(18), 3545–3549. https://doi.org/10.1016/j.febslet.2007.06.052

Yamanaka, T., & Fukumori, Y. (1988). The nitrite oxidizing system of Nitrobacter winogradskyi. *FEMS Microbiology Reviews*, *4*(4), 259–270. https://doi.org/10.1111/j.1574-6968.1988.tb02746.x

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J., & Woese, C. R. (1985). Mitochondrial origins. *Proceedings of the National Academy of Sciences*, *82*(13), 4443–4447.

Yang, R.-F., Sun, L.-H., Zhang, R., Zhang, Y., Luo, Y.-X., Zheng, W., … Liu, D.-P. (2015). Suppression of Mic60 compromises mitochondrial transcription and oxidative phosphorylation. *Scientific Reports*, *5*. https://doi.org/10.1038/srep07990

Youle, R. J., & Bliek, A. M. van der. (2012). Mitochondrial Fission, Fusion, and Stress. *Science*, *337*(6098), 1062–1065. https://doi.org/10.1126/science.1219855

Zahn, J. A., Bergmann, D. J., Boyd, J. M., Kunz, R. C., & DiSpirito, A. A. (2001). Membrane-Associated Quinoprotein Formaldehyde Dehydrogenase from Methylococcus capsulatus Bath. *Journal of Bacteriology*, *183*(23), 6832–6840. https://doi.org/10.1128/JB.183.23.6832-6840.2001

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., … Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, *541*(7637), 353–358. https://doi.org/10.1038/nature21031

Zeng, X., Roh, J. H., Callister, S. J., Tavano, C. L., Donohue, T. J., Lipton, M. S., & Kaplan, S. (2007). Proteomic Characterization of the Rhodobacter sphaeroides 2.4.1 Photosynthetic Membrane: Identification of New Proteins. *Journal of Bacteriology*, *189*(20), 7464–7474. https://doi.org/10.1128/JB.00946-07

Zerbes, R. M., Bohnert, M., Stroud, D. A., von der Malsburg, K., Kram, A., Oeljeklaus, S., … van der Laan, M. (2012). Role of MINOS in mitochondrial membrane architecture: cristae morphology and outer membrane interactions differentially depend on mitofilin domains. *Journal of Molecular Biology*, *422*(2), 183–191. https://doi.org/10.1016/j.jmb.2012.05.004

Zerbes, R. M., Höß, P., Pfanner, N., van der Laan, M., & Bohnert, M. (2016). Distinct roles of Mic12 and Mic27 in the mitochondrial contact site and cristae organizing system. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2016.02.031

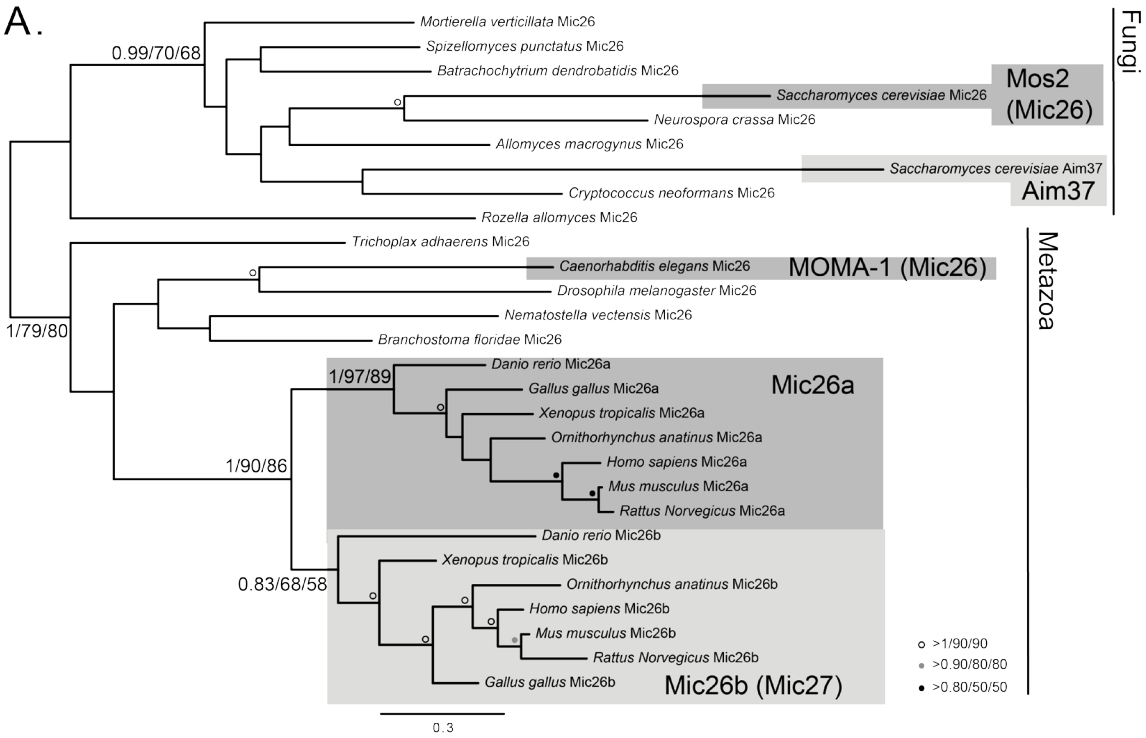
Zerbes, R. M., van der Klei, I. J., Veenhuis, M., Pfanner, N., van der Laan, M., & Bohnert, M. (2012). Mitofilin complexes: conserved organizers of mitochondrial membrane architecture. *Biological Chemistry*, *393*(11), 1247–1261. https://doi.org/10.1515/hsz-2012-0239

Zick, M., Rabl, R., & Reichert, A. S. (2009). Cristae formation-linking ultrastructure and function of mitochondria. *Biochimica et Biophysica Acta*, *1793*(1), 5–19. https://doi.org/10.1016/j.bbamcr.2008.06.013

Zierdt, C. H., Donnolley, C. T., Muller, J., & Constantopoulos, G. (1988). Biochemical and ultrastructural study of Blastocystis hominis. *Journal of Clinical Microbiology*, *26*(5), 965–970.

# APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 2

**Figure S2.1.** Mic26 evolution in Metazoa and Fungi. Lineage-specific duplications of Mic26 in the Vertebrata and the Saccharomycetales. **A.** Phylogenetic analysis of Mic26 and related proteins from several species of Metazoa and Fungi demonstrating that Aim37 in *S. cerevisiae* is not orthologous to the Mic27 proteins found in Vertebrata. **B.** Phylogenetic analysis of Mic26 and Aim37 proteins from invertebrates and Fungi demonstrating that Aim37 resulted from a gene duplication early in the Saccharomycetales. Node support is shown in the following order: MrBayes(PP)/PhyML(BS)/RAxML(BS) or as symbolized in inset.


A.
Mortierella verticillata Mic26
Spizellomyces punctatus Mic26
Batrachochytrium dendrobatidis Mic26
0.99/70/68
Saccharomyces cerevisiae Mic26
Mos2 (Mic26)
Neurospora crassa Mic26
Allomyces macrogynus Mic26
Saccharomyces cerevisiae Aim37
Aim37
Cryptococcus neoformans Mic26
Rozella allomyces Mic26
Fungi
Trichoplax adhaerens Mic26
Caenorhabditis elegans Mic26
MOMA-1 (Mic26)
Drosophila melanogaster Mic26
Nematostella vectensis Mic26
1/79/80
Branchostoma floridae Mic26
Danio rerio Mic26a
1/97/89
Gallus gallus Mic26a
Mic26a
Xenopus tropicalis Mic26a
Ornithorhynchus anatinus Mic26a
1/90/86
Homo sapiens Mic26a
Mus musculus Mic26a
Rattus Norvegicus Mic26a
Metazoa
Danio rerio Mic26b
Xenopus tropicalis Mic26b
0.83/68/58
Ornithorhynchus anatinus Mic26b
Homo sapiens Mic26b
Mus musculus Mic26b
Rattus Norvegicus Mic26b
Gallus gallus Mic26b
Mic26b (Mic27)
>1/90/90
>0.90/80/80
>0.80/50/50
0.3

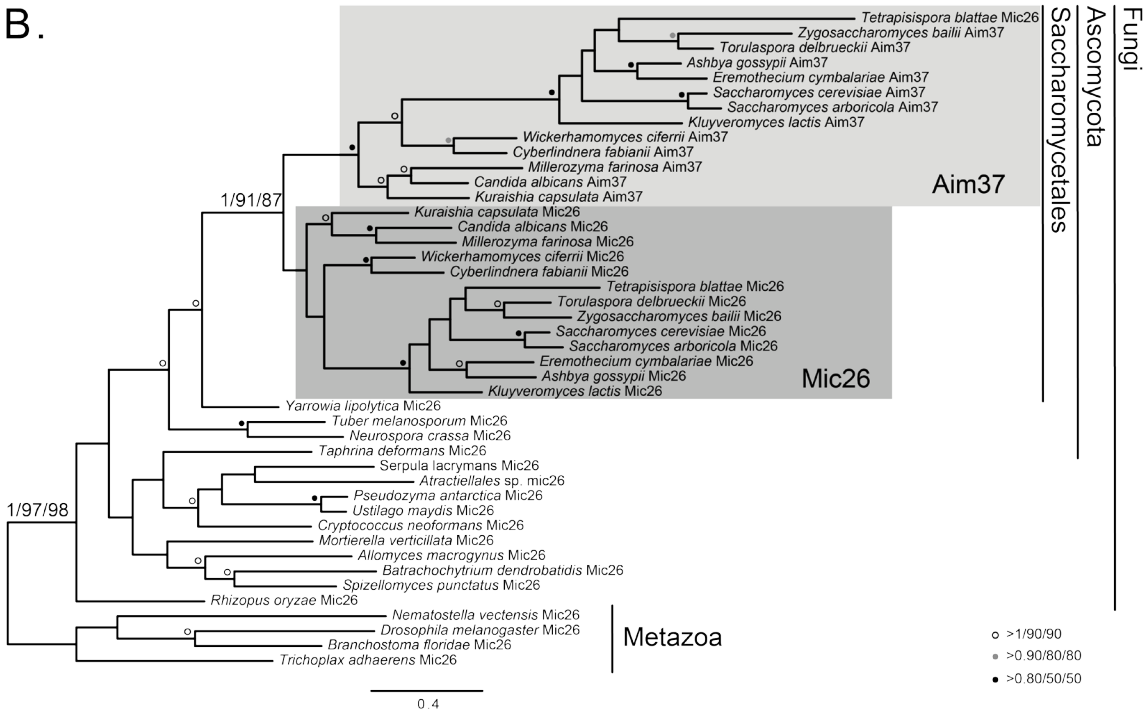
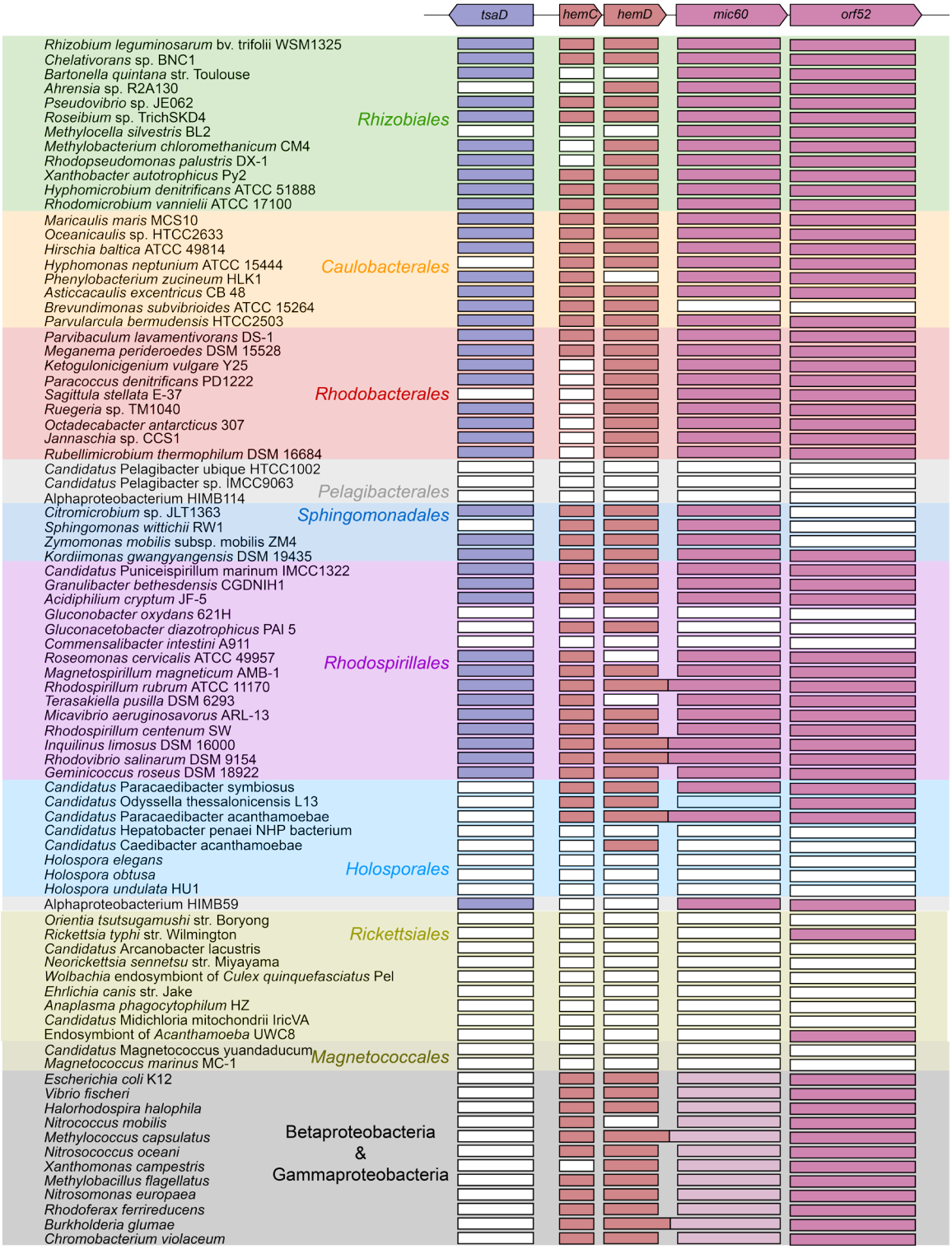

B.
Tetrapisispora blattae Mic26
Zygosaccharomyces bailii Aim37
Torulaspora delbrueckii Aim37
Ashbya gossypii Aim37
Eremothecium cymbalariae Aim37
Saccharomyces cerevisiae Aim37
Saccharomyces arboricola Aim37
Kluyveromyces lactis Aim37
Wickerhamomyces ciferrii Aim37
Cyberlindnera fabianii Aim37
Millerozyma farinosa Aim37
Candida albicans Aim37
Kuraishia capsulata Aim37
Aim37
1/91/87
Kuraishia capsulata Mic26
Candida albicans Mic26
Millerozyma farinosa Mic26
Wickerhamomyces ciferrii Mic26
Cyberlindnera fabianii Mic26
Tetrapisispora blattae Mic26
Torulaspora delbrueckii Mic26
Zygosaccharomyces bailii Mic26
Saccharomyces cerevisiae Mic26
Saccharomyces arboricola Mic26
Eremothecium cymbalariae Mic26
Ashbya gossypii Mic26
Kluyveromyces lactis Mic26
Mic26
Saccharomycetales
Ascomycota
Fungi
Yarrowia lipolytica Mic26
Tuber melanosporum Mic26
Neurospora crassa Mic26
Taphrina deformans Mic26
Serpula lacrymans Mic26
Atractiellales sp. mic26
Pseudozyma antarctica Mic26
1/97/98
Ustilago maydis Mic26
Cryptococcus neoformans Mic26
Mortierella verticillata Mic26
Allomyces macrogynus Mic26
Batrachochytrium dendrobatidis Mic26
Spizellomyces punctatus Mic26
Rhizopus oryzae Mic26
Nematostella vectensis Mic26
Drosophila melanogaster Mic26
Branchostoma floridae Mic26
Trichoplax adhaerens Mic26
Metazoa
>1/90/90
>0.90/80/80
>0.80/50/50
0.4

**Figure S2.2.** The phylogenetic distribution of the *hemC-hemD-mic60-orf52* gene cluster in the *Alpha*-, *Beta*-, and *Gammaproteobacteria*. Three successive rounds of profile HMM searches were performed with different customized profiles built with more inclusive sets of taxa to identify divergent α-proteobacterial *mic60* homologs. Homology assessments were finally made based on profile HMM search results and the genomic context of the candidate *mic60* homologs. *orf52* homologs were identified using the HemY_N (PF07219) Pfam profile. Neighboring genes surrounding the *mic60-orf52* were identified using the NCBI graphical genome browser. *tsaD* is a gene upstream, and in opposite direction, of the *hemC-hemD-mic60-orf52* gene cluster that encodes for a tRNA N6-adenosine threonylcarbamoyltransferase.


tsaD
hemC
hemD
mic60
orf52
Rhizobium leguminosarum bv. trifolii WSM1325
Chelativorans sp. BNC1
Bartonella quintana str. Toulouse
Ahrensia sp. R2A130
Pseudovibrio sp. JE062
Roseibium sp. TrichSKD4
Methylocella silvestris BL2
Methylobacterium chloromethanicum CM4
Rhodopseudomonas palustris DX-1
Xanthobacter autotrophicus Py2
Hyphomicrobium denitrificans ATCC 51888
Rhodomicrobium vannielii ATCC 17100
Rhizobiales
Maricaulis maris MCS10
Oceanicaulis sp. HTCC2633
Hirschia baltica ATCC 49814
Hyphomonas neptunium ATCC 15444
Phenylobacterium zucineum HLK1
Asticcacaulis excentricus CB 48
Brevundimonas subvibrioides ATCC 15264
Parvularcula bermudensis HTCC2503
Caulobacterales
Parvibaculum lavamentivorans DS-1
Meganema perideroedes DSM 15528
Ketogulonicigenium vulgare Y25
Paracoccus denitrificans PD1222
Sagittula stellata E-37
Ruegeria sp. TM1040
Octadecabacter antarcticus 307
Jannaschia sp. CCS1
Rubellimicrobium thermophilum DSM 16684
Rhodobacterales
Candidatus Pelagibacter ubique HTCC1002
Candidatus Pelagibacter sp. IMCC9063
Alphaproteobacterium HIMB114
Pelagibacterales
Citromicrobium sp. JLT1363
Sphingomonas wittichii RW1
Zymomonas mobilis subsp. mobilis ZM4
Kordiimonas gwangyangensis DSM 19435
Sphingomonadales
Candidatus Puniceispirillum marinum IMCC1322
Granulibacter bethesdensis CGDNIH1
Acidiphilium cryptum JF-5
Gluconobacter oxydans 621H
Gluconacetobacter diazotrophicus PAl 5
Commensalibacter intestini A911
Roseomonas cervicalis ATCC 49957
Magnetospirillum magneticum AMB-1
Rhodospirillum rubrum ATCC 11170
Terasakiella pusilla DSM 6293
Micavibrio aeruginosavorus ARL-13
Rhodospirillum centenum SW
Inquilinus limosus DSM 16000
Rhodovibrio salinarum DSM 9154
Geminicoccus roseus DSM 18922
Rhodospirillales
Candidatus Paracaedibacter symbiosus
Candidatus Odyssella thessalonicensis L13
Candidatus Paracaedibacter acanthamoebae
Candidatus Hepatobacter penaei NHP bacterium
Candidatus Caedibacter acanthamoebae
Holospora elegans
Holospora obtusa
Holospora undulata HU1
Holosporales
Alphaproteobacterium HIMB59
Orientia tsutsugamushi str. Boryong
Rickettsia typhi str. Wilmington
Candidatus Arcanobacter lacustris
Neorickettsia sennetsu str. Miyayama
Wolbachia endosymbiont of Culex quinquefasciatus Pel
Ehrlichia canis str. Jake
Anaplasma phagocytophilum HZ
Candidatus Midichloria mitochondrii IricVA
Endosymbiont of Acanthamoeba UWC8
Rickettsiales
Candidatus Magnetococcus yuandaducum
Magnetococcus marinus MC-1
Magnetococcales
Escherichia coli K12
Vibrio fischeri
Halorhodospira halophila
Nitrococcus mobilis
Methylococcus capsulatus
Nitrosococcus oceani
Xanthomonas campestris
Methylobacillus flagellatus
Nitrosomonas europaea
Rhodoferax ferrireducens
Burkholderia glumae
Chromobacterium violaceum
Betaproteobacteria
&
Gammaproteobacteria

# APPENDIX B. SUPPLEMENTARY MATERIAL FOR CHAPTER 4


**Figure S4.1.** Constraint tree, used for IQ-TREE analyses, labeled with taxon names and also degree of missing data per taxon. *Rhizobiales* (green), *Caulobacterales* (orange), *Rhodobacterales* (red), *Sphingomonadales* (blue), *Sneathiellales* (pink), *Rhodospirillales* (purple), *Holosporales* (light blue), *Pelagibacterales* and alphaproteobacterium sp. HIMB59 (maroon), *Rickettsiales* (brown green), *Magnetococcales* (gray), *Beta-* and *Gammeproteobacteria* (black).

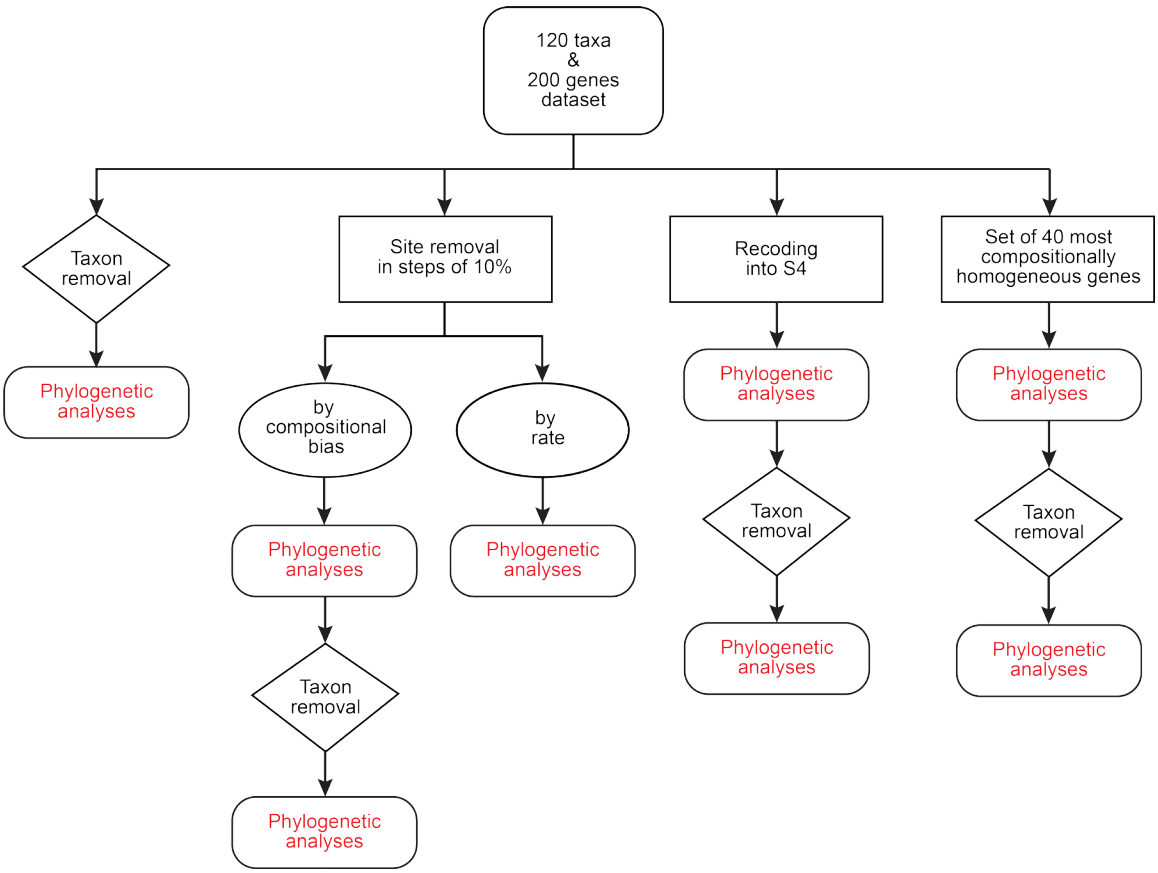



**Figure S4.2.** A diagram of the strategies employed in this study.

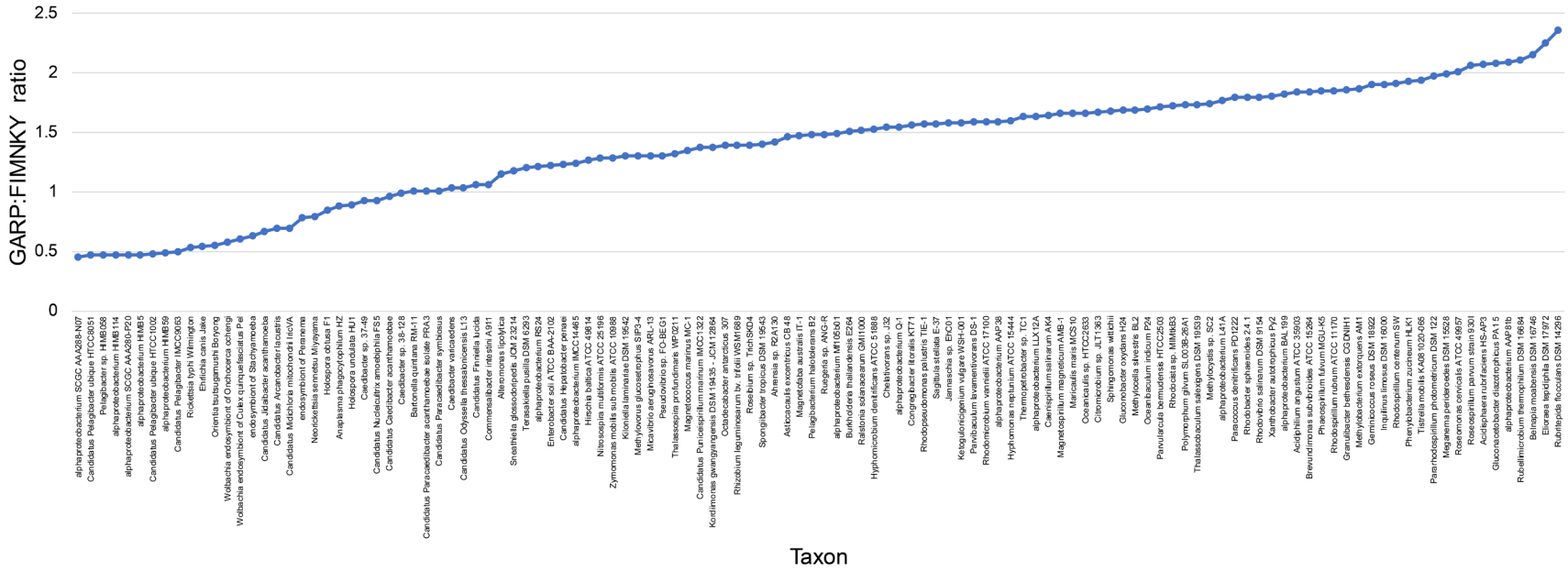
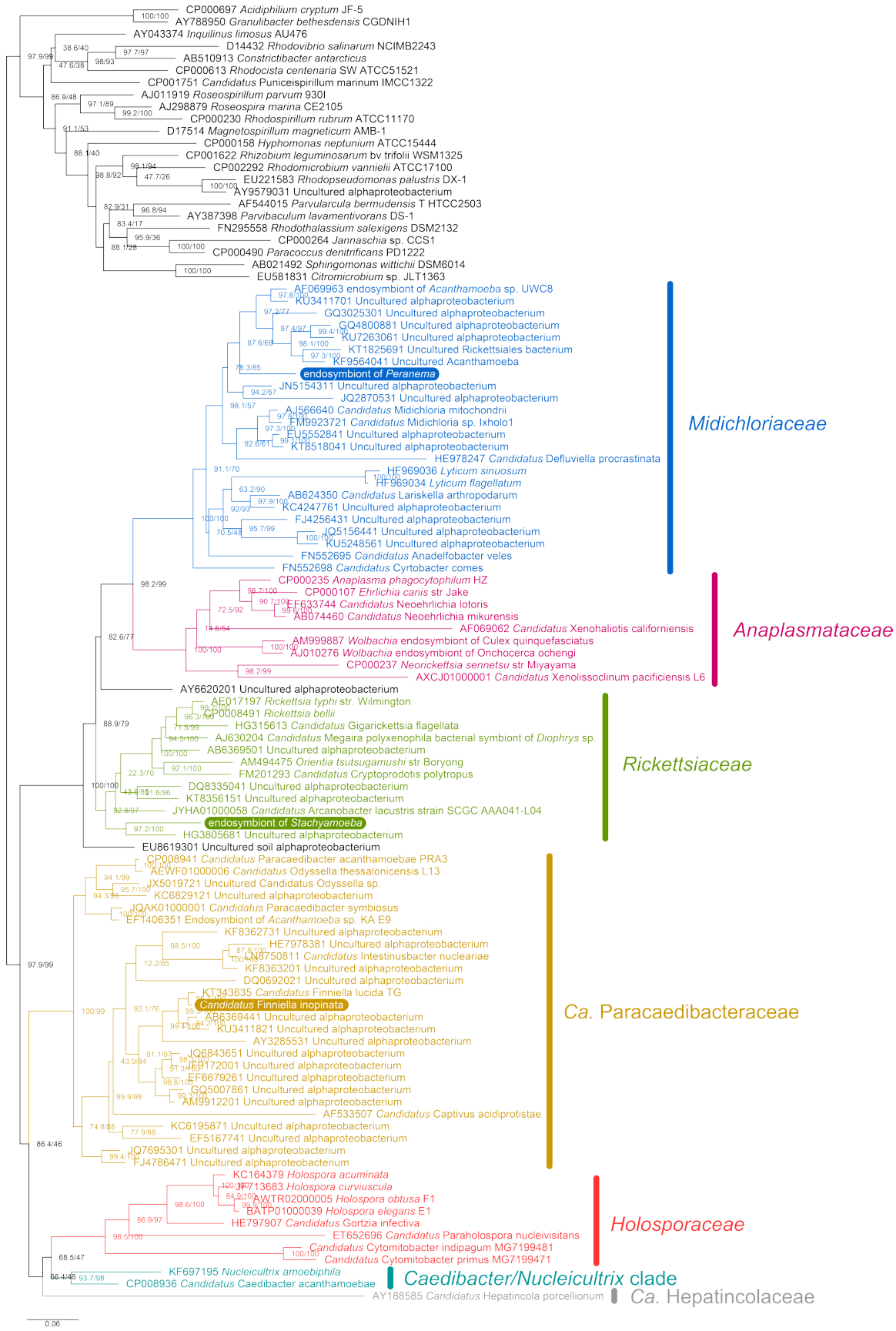



**Figure S4.3.** GARP:FIMNKY ratios across the proteomes of the 120 alphaproteobacteria used in this study.

**Figure S4.4.** A 16S rRNA gene maximum-likelihood tree of the *Rickettsiales* and *Holosporales* that phylogenetically places the three endosymbionts whose genomes were sequenced in this study: (1) '*Candidatus* Finniella inopinata' endosymbiont of *Viridiraptor invadens* strain Virl02, (2) an alphaproteobacterium associated with *Peranema trichophorum* strain CCAP 1260/1B, and (3) an alphaproteobacterium associated with *Stachyamoeba lipophora* strain ATCC 50324. Branch support values are SH-aLRT and UFBoot.


CP000697 Acidiphilium cryptum JF-5
AY788950 Granulibacter bethesdensis CGDNIH1
AY043374 Inquilinus limosus AU476
D14432 Rhodovibrio salinarum NCIMB2243
AB510913 Constrictibacter antarcticus
CP000613 Rhodocista centenaria SW ATCC51521
CP001751 Candidatus Puniceispirillum marinum IMCC1322
AJ011919 Roseospirillum parvum 930I
AJ298879 Roseospira marina CE2105
CP000230 Rhodospirillum rubrum ATCC11170
D17514 Magnetospirillum magneticum AMB-1
CP000158 Hyphomonas neptunium ATCC15444
CP001622 Rhizobium leguminosarum bv trifolii WSM1325
CP002292 Rhodomicrobium vannielii ATCC17100
EU221583 Rhodopseudomonas palustris DX-1
AY9579031 Uncultured alphaproteobacterium
AF544015 Parvularcula bermudensis T HTCC2503
AY387398 Parvibaculum lavamentivorans DS-1
FN295558 Rhodothalassium salexigens DSM2132
CP000264 Jannaschia sp. CCS1
CP000490 Paracoccus denitrificans PD1222
AB021492 Sphingomonas wittichii DSM6014
EU581831 Citromicrobium sp. JLT1363
AF069963 endosymbiont of Acanthamoeba sp. UWC8
KU3411701 Uncultured alphaproteobacterium
GQ3025301 Uncultured alphaproteobacterium
GQ4800881 Uncultured alphaproteobacterium
KU7263061 Uncultured alphaproteobacterium
KT1825691 Uncultured Rickettsiales bacterium
KF9564041 Uncultured Acanthamoeba
endosymbiont of Peranema
JN5154311 Uncultured alphaproteobacterium
JQ2870531 Uncultured alphaproteobacterium
AJ566640 Candidatus Midichloria mitochondrii
FM9923721 Candidatus Midichloria sp. Ixholo1
EU5552841 Uncultured alphaproteobacterium
KT8518041 Uncultured alphaproteobacterium
HE978247 Candidatus Defluviella procrastinata
HF969036 Lyticum sinuosum
HF969034 Lyticum flagellatum
AB624350 Candidatus Lariskella arthropodarum
KC4247761 Uncultured alphaproteobacterium
FJ4256431 Uncultured alphaproteobacterium
JQ5156441 Uncultured alphaproteobacterium
KU5248561 Uncultured alphaproteobacterium
FN552695 Candidatus Anadelfobacter veles
FN552698 Candidatus Cyrtobacter comes
Midichloriaceae
CP000235 Anaplasma phagocytophilum HZ
CP000107 Ehrlichia canis str Jake
EF633744 Candidatus Neoehrlichia lotoris
AB074460 Candidatus Neoehrlichia mikurensis
AF069062 Candidatus Xenohaliotis californiensis
AM999887 Wolbachia endosymbiont of Culex quinquefasciatus
AJ010276 Wolbachia endosymbiont of Onchocerca ochengi
CP000237 Neorickettsia sennetsu str Miyayama
AXCJ01000001 Candidatus Xenolissoclinum pacificiensis L6
Anaplasmataceae
AY6620201 Uncultured alphaproteobacterium
AE017197 Rickettsia typhi str. Wilmington
CP0008491 Rickettsia bellii
HG315613 Candidatus Gigarickettsia flagellata
AJ630204 Candidatus Megaira polyxenophila bacterial symbiont of Diophrys sp.
AB6369501 Uncultured alphaproteobacterium
AM494475 Orientia tsutsugamushi str Boryong
FM201293 Candidatus Cryptoprodotis polytropus
DQ8335041 Uncultured alphaproteobacterium
KT8356151 Uncultured alphaproteobacterium
JYHA01000058 Candidatus Arcanobacter lacustris strain SCGC AAA041-L04
endosymbiont of Stachyamoeba
HG3805681 Uncultured alphaproteobacterium
Rickettsiaceae
EU8619301 Uncultured soil alphaproteobacterium
CP008941 Candidatus Paracaedibacter acanthamoebae PRA3
AEWF01000006 Candidatus Odyssella thessalonicensis L13
JX5019721 Uncultured Candidatus Odyssella sp.
KC6829121 Uncultured alphaproteobacterium
JQAK01000001 Candidatus Paracaedibacter symbiosus
EF1406351 Endosymbiont of Acanthamoeba sp. KA E9
KF8362731 Uncultured alphaproteobacterium
HE7978381 Uncultured alphaproteobacterium
AM8750811 Candidatus Intestinusbacter nuclearis
KF8363201 Uncultured alphaproteobacterium
DQ0692021 Uncultured alphaproteobacterium
KT343635 Candidatus Finniella lucida TG
Candidatus Finniella inopinata
AB6369441 Uncultured alphaproteobacterium
KU3411821 Uncultured alphaproteobacterium
AY3285531 Uncultured alphaproteobacterium
JQ6843651 Uncultured alphaproteobacterium
KF9172001 Uncultured alphaproteobacterium
EF6679261 Uncultured alphaproteobacterium
GQ5007861 Uncultured alphaproteobacterium
AM9912201 Uncultured alphaproteobacterium
AF533507 Candidatus Captivus acidiprotistae
KC6195871 Uncultured alphaproteobacterium
EF5167741 Uncultured alphaproteobacterium
JQ7695301 Uncultured alphaproteobacterium
FJ4786471 Uncultured alphaproteobacterium
Ca. Paracaedibacteraceae
KC164379 Holospora acuminata
JF713683 Holospora curviuscula
AWTR02000005 Holospora obtusa F1
BATP01000039 Holospora elegans E1
HE797907 Candidatus Gortzia infectiva
ET652696 Candidatus Paraholospora nucleivisitans
Candidatus Cytomitobacter indipagum MG7199481
Candidatus Cytomitobacter primus MG7199471
Holosporaceae
KF697195 Nucleicultrix amoebiphila
CP008936 Candidatus Caedibacter acanthamoebae
Caedibacter/Nucleicultrix clade
AY188585 Candidatus Hepatincola porcellionum
Ca. Hepatincolaceae
0.06

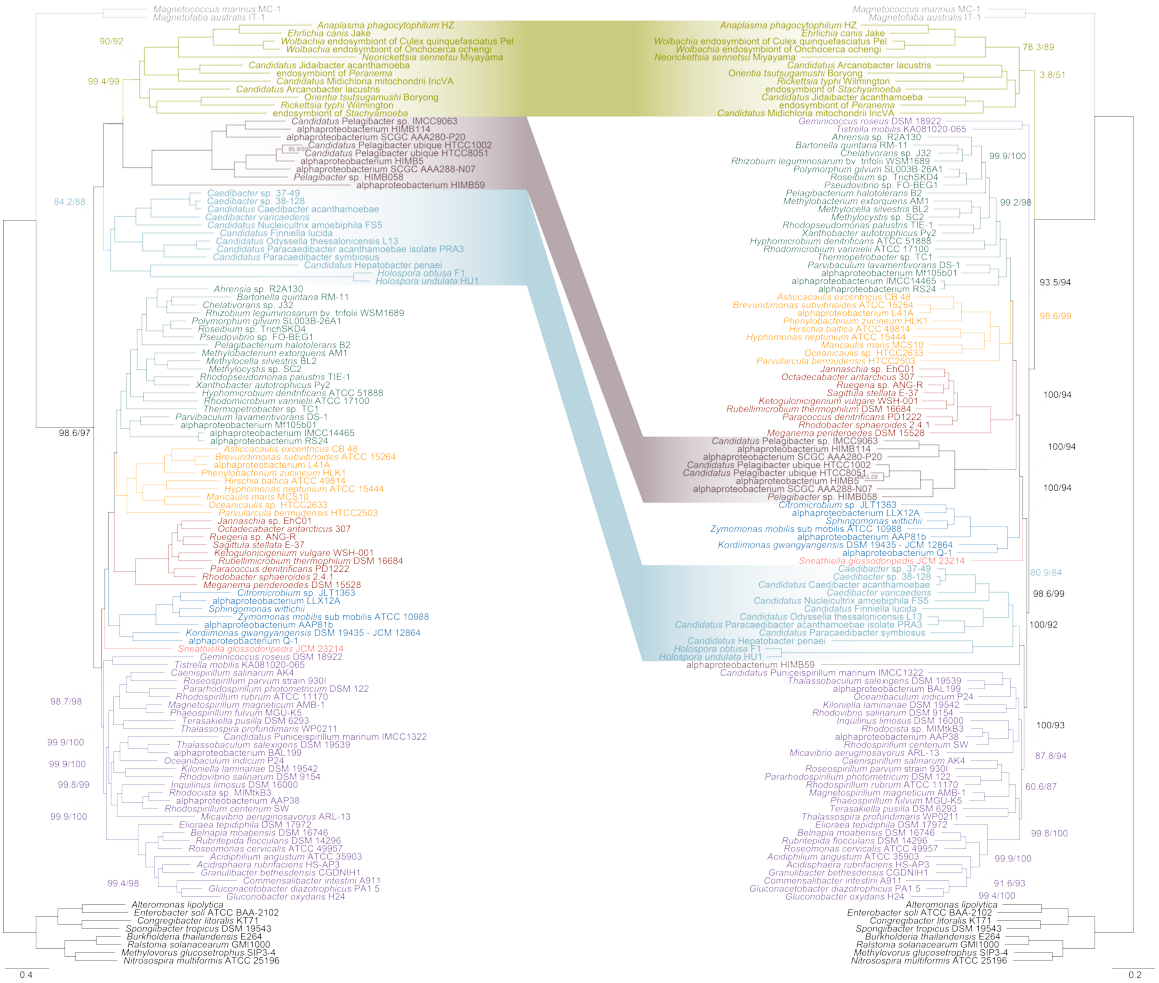
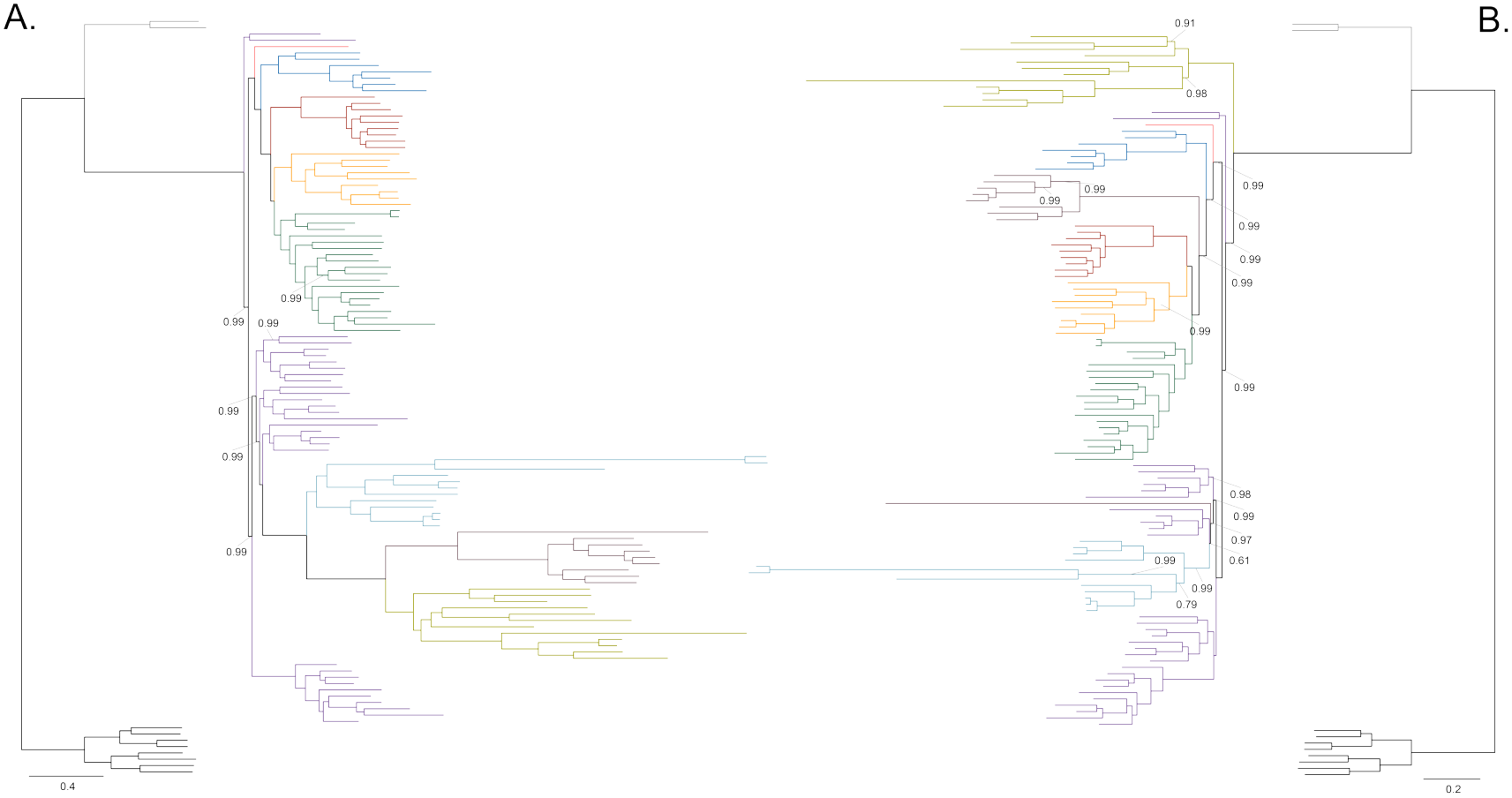


**Figure S4.5.** A labeled version of Figure 4.2. Branch support values are 100% SH-aLRT and 100% UFBoot unless annotated. *Rhizobiales* (green), *Caulobacterales* (orange), *Rhodobacterales* (red), *Sphingomonadales* (blue), *Sneathiellales* (pink), *Rhodospirillales* (purple), *Holosporales* (light blue), *Pelagibacterales* and alphaproteobacterium sp. HIMB59 (maroon), *Rickettsiales* (brown green), *Magnetococcales* (gray), *Beta-* and *Gammeproteobacteria* (black).

**Figure S4.6.** Bayesian consensus trees inferred with PhyloBayes MPI v1.7 and the CAT-Poisson+Γ4 model. Branch support values are 1.0 posterior probabilities unless annotated. **A.** Bayesian consensus phylogram inferred from the full dataset which is highly compositionally heterogeneous. **B.** Bayesian consensus phylogram inferred from a dataset whose compositional heterogeneity has been decreased by removing 50% of the most biased sites according to ζ. See Figs. 2A and 2B for the most likely trees inferred in IQ-TREE v1.5.5 and the LG+PMSF(C60)+F+R6 model. *Rhizobiales* (green), *Caulobacterales* (orange), *Rhodobacterales* (red), *Sphingomonadales* (blue), *Sneathiellales* (pink), *Rhodospirillales* (purple), *Holosporales* (light blue), *Pelagibacterales* and alphaproteobacterium sp. HIMB59 (maroon), *Rickettsiales* (brown green), *Magnetococcales* (gray), *Beta*- and *Gammeproteobacteria* (black).


A.
B.
0.99
0.99
0.99
0.99
0.99
0.99
0.4
0.91
0.98
0.99
0.99
0.99
0.99
0.99
0.99
0.99
0.98
0.99
0.97
0.61
0.99
0.99
0.79
0.2

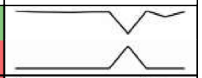
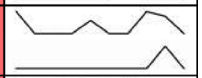
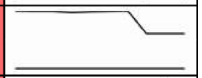
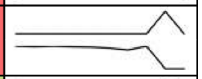
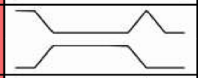
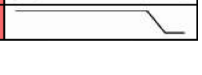
| Site removal by rate (LG+PMSF(C60)F+R6) | 0% (54,400) | -10% (48,960) | -20% (43,520) | -30% (38,080) | -40% (32,640) | -50% (27,200) | -60% (21,760) | -70% (16,320) | -80% (16,320) | -90% (5,440) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Anaplasmataceae* + *Midichloriaceae* | 93 | 90 | 88 | 87 | 90 | 89 | 0 | 94 | 69 | 89 |  |
| *Rickettsiaceae* + *Midichloriaceae* | 0 | 0 | 0 | 0 | 0 | 0 | 73 | 0 | 0 | 0 | |
| *Caulobacteridae* - *Holosporales* | 100 | 0 | 0 | 0 | 58 | 0 | 0 | 96 | 77 | 0 |  |
| *Rhodospirillales* + *Holosporales* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 95 | 0 | |
| *Rhodospirillales* + *Geminicoccaceae* | 100 | 98 | 98 | 94 | 97 | 98 | 99 | 0 | 0 | 0 |  |
| *Caulobacteridae* - *Geminicoccaceae* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| *Caulobacterales* + *Rhodobacterales* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 51 | 0 |  |
| *Rhizobiales* + *Caulobacterales* | 97 | 95 | 97 | 95 | 93 | 90 | 81 | 98 | 0 | 0 | |
| *Rickettsiales* + *Pelagibacterales* | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 70 |  |
| *Rhodob.* + *Caulob.* + *Rhizob.* + *Pelagibacterales* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| *Rhodospirillaceae* + *Azosp.* + *Rhodov.* | 100 | 97 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |  |
| *Acetobacteraceae* + *Rhodospirillaceae* | 0 | 0 | 100 | 100 | 100 | 100 | 97 | 0 | 0 | 0 | |
| *Azospirillaceae* + *Rhodovibraceae* | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 0 | 0 |  |

**Figure S4.7.** Ultrafast bootstrap (UFBoot) variation as the fastest sites are progressively removed in steps of 10%.

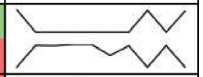
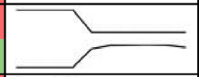
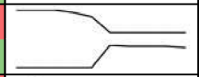
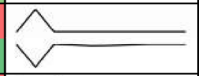
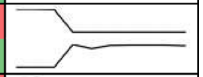
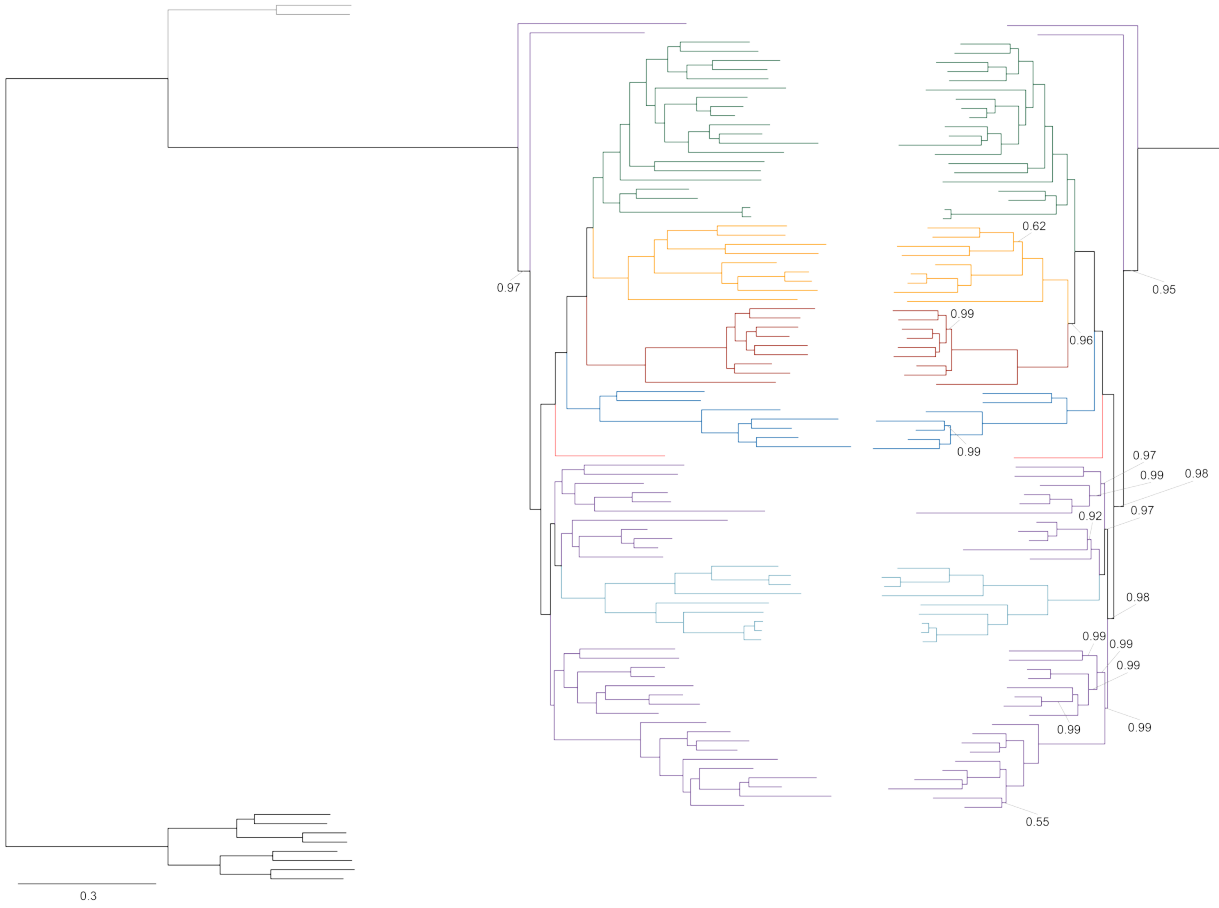
| Site removal by composition (LG+PMSF(C60)F+R6) | 0% (54,400) | -10% (48,960) | -20% (43,520) | -30% (38,080) | -40% (32,640) | -50% (27,200) | -60% (21,760) | -70% (16,320) | -80% (16,320) | -90% (5,440) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Anaplasmataceae* + *Midichloriaceae* | 93 | 0 | 0 | 0 | 0 | 0 | 0 | 88 | 0 | 86 |  |
| *Rickettsiaceae* + *Midichloriaceae* | 0 | 92 | 91 | 97 | 97 | 51 | 91 | 0 | 92 | 0 | |
| *Caulobacteridae* - *Holosporales* | 100 | 100 | 100 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |  |
| *Rhodospirillales* + *Holosporales* | 0 | 0 | 0 | 0 | 79 | 93 | 94 | 96 | 94 | 85 | |
| *Rhodospirillales* + *Geminicoccaceae* | 100 | 100 | 100 | 88 | 69 | 0 | 0 | 0 | 0 | 0 |  |
| *Caulobacteridae* - *Geminicoccaceae* | 0 | 0 | 0 | 0 | 0 | 99 | 93 | 93 | 93 | 87 | |
| *Caulobacterales* + *Rhodobacterales* | 0 | 75 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  |
| *Rhizobiales* + *Caulobacterales* | 97 | 0 | 100 | 96 | 92 | 94 | 97 | 99 | 100 | 99 | |
| *Rickettsiales* + *Pelagibacterales* | 100 | 100 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  |
| *Rhodob.* + *Caulob.* + *Rhizob.* + *Pelagibacterales* | 0 | 0 | 0 | 99 | 81 | 94 | 96 | 96 | 96 | 93 | |
| *Rhodospirillaceae* + *Azosp.* + *Rhodov.* | 100 | 91 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  |
| *Acetobacteraceae* + *Rhodospirillaceae* | 0 | 0 | 55 | 100 | 100 | 100 | 100 | 100 | 100 | 99 | |
| *Azospirillaceae* + *Rhodovibraceae* | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 95 | 100 | 97 |  |

**Figure S4.8.** Ultrafast bootstrap (UFBoot) variation as compositionally biased sites, according to ʒ, are progressively removed in steps of 10%.

**Figure S4.9.** Bayesian consensus trees inferred with PhyloBayes MPI v1.7 and the CAT-Poisson+Γ4 model. Branch support values are 1.0 posterior probabilities unless annotated. **A.** Bayesian consensus phylogram inferred to place the *Holosporales* in the absence of the *Rickettsiales* and the *Pelagibacterales* and when compositional heterogeneity has been decreased by removing 50% of the most biased sites according to ζ. **B.** Bayesian consensus phylogram inferred to place the *Holosporales* in the absence of the *Rickettsiales* and the *Pelagibacterales* and when the data have been recoded into a four-character state alphabet (the dataset-specific recoding scheme S4: ARNDQEILKSTV GHY CMFP W) to reduce compositional heterogeneity. See Figs. 2A and 2B for the most likely trees inferred in IQ-TREE v1.5.5 and the LG+PMSF(C60)+F+R6 and GTR+ES60S4+F+R6 models, respectively. *Rhizobiales* (green), *Caulobacterales* (orange), *Rhodobacterales* (red), *Sphingomonadales* (blue), *Sneathiellales* (pink), *Rhodospirillales* (purple), *Holosporales* (light blue), *Magnetococcales* (gray), *Beta-* and *Gammeproteobacteria* (black).

A.


0.97
0.62
0.99
0.96
0.99
0.95
0.97
0.99
0.98
0.92
0.97
0.98
0.99
0.99
0.99
0.99
0.99
0.55
0.3

B.

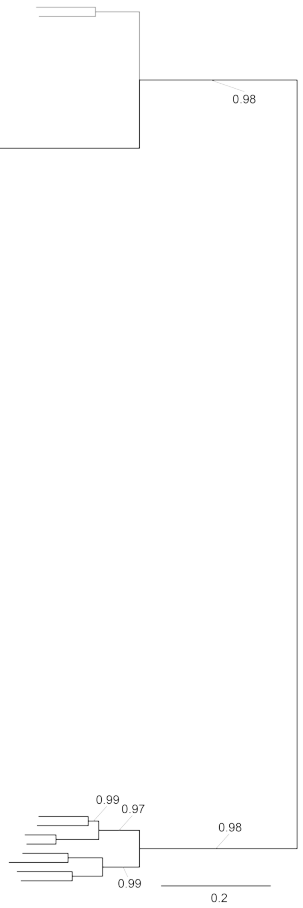

0.98
0.99
0.99
0.97
0.98
0.99
0.2

**Figure S4.10.** Maximum-likelihood phylograms obtained from the full dataset and after removing long-branching taxa. **A.** A phylogram for which all long-branching taxa are included. **B.** A phylogram to assess the placement of the *Holosporales*. **C.** A phylogram to assess the placement of the *Rickettsiales*. **D.** A phylogram to assess the placement of the *Pelagibacterales*. **E.** A phylogram to assess the placement of alphaproteobacterium sp. HIMB59.

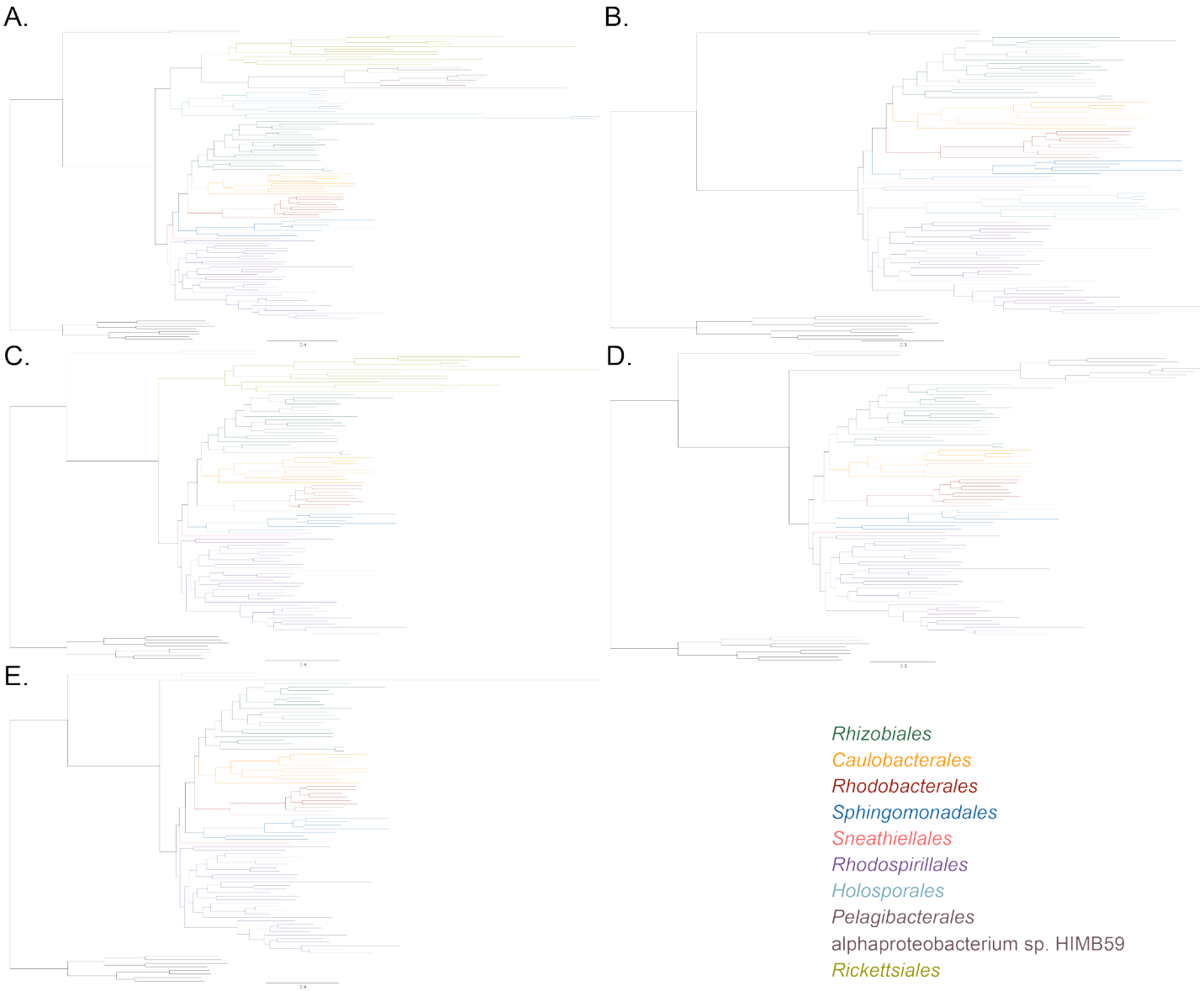

A.
B.
C.
D.
E.
Rhizobiales
Caulobacterales
Rhodobacterales
Sphingomonadales
Sneathiellales
Rhodospirillales
Holosporales
Pelagibacterales
alphaproteobacterium sp. HIMB59
Rickettsiales

**Figure S4.11.** Maximum-likelihood phylograms obtained after removing 50% of the most compositionally biased sites and removing long-branching taxa. **A.** A phylogram for which all long-branching taxa are included. **B.** A phylogram to assess the placement of the *Holosporales*. **C.** A phylogram to assess the placement of the *Rickettsiales*. **D.** A phylogram to assess the placement of the *Pelagibacterales*. **E.** A phylogram to assess the placement of alphaproteobacterium sp. HIMB59.

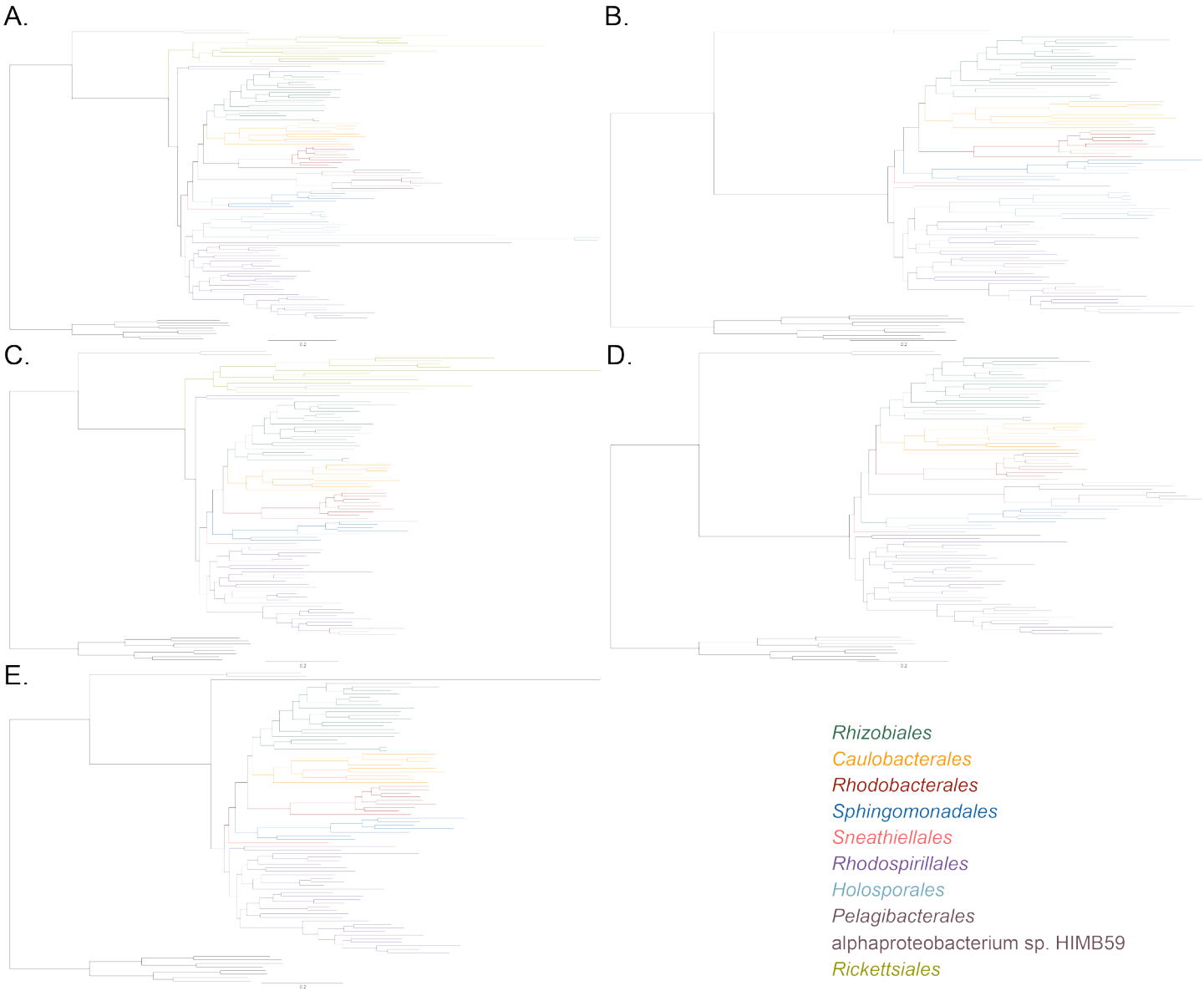

A.
B.
C.
D.
E.
Rhizobiales
Caulobacterales
Rhodobacterales
Sphingomonadales
Sneathiellales
Rhodospirillales
Holosporales
Pelagibacterales
alphaproteobacterium sp. HIMB59
Rickettsiales

**Figure S4.12.** Maximum-likelihood phylograms obtained after recoding the data into S4 and removing long-branching taxa. **A.** A phylogram for which all long-branching taxa are included (recoding scheme: RNCM EHIPTWV ADQLKS GFY). **B.** A phylogram to assess the placement of the *Holosporales* (recoding scheme: ARNDQEILKSTV GHY CMFP W). **C.** A phylogram to assess the placement of the *Rickettsiales* (recoding scheme: PY RNMF GHLKTW ADCQEISV). **D.** A phylogram to assess the placement of the *Pelagibacterales*. **E.** A phylogram to assess the placement of alphaproteobacterium sp. HIMB59 (recoding scheme: RLKMT ANDQEIPSV CW GHFY).

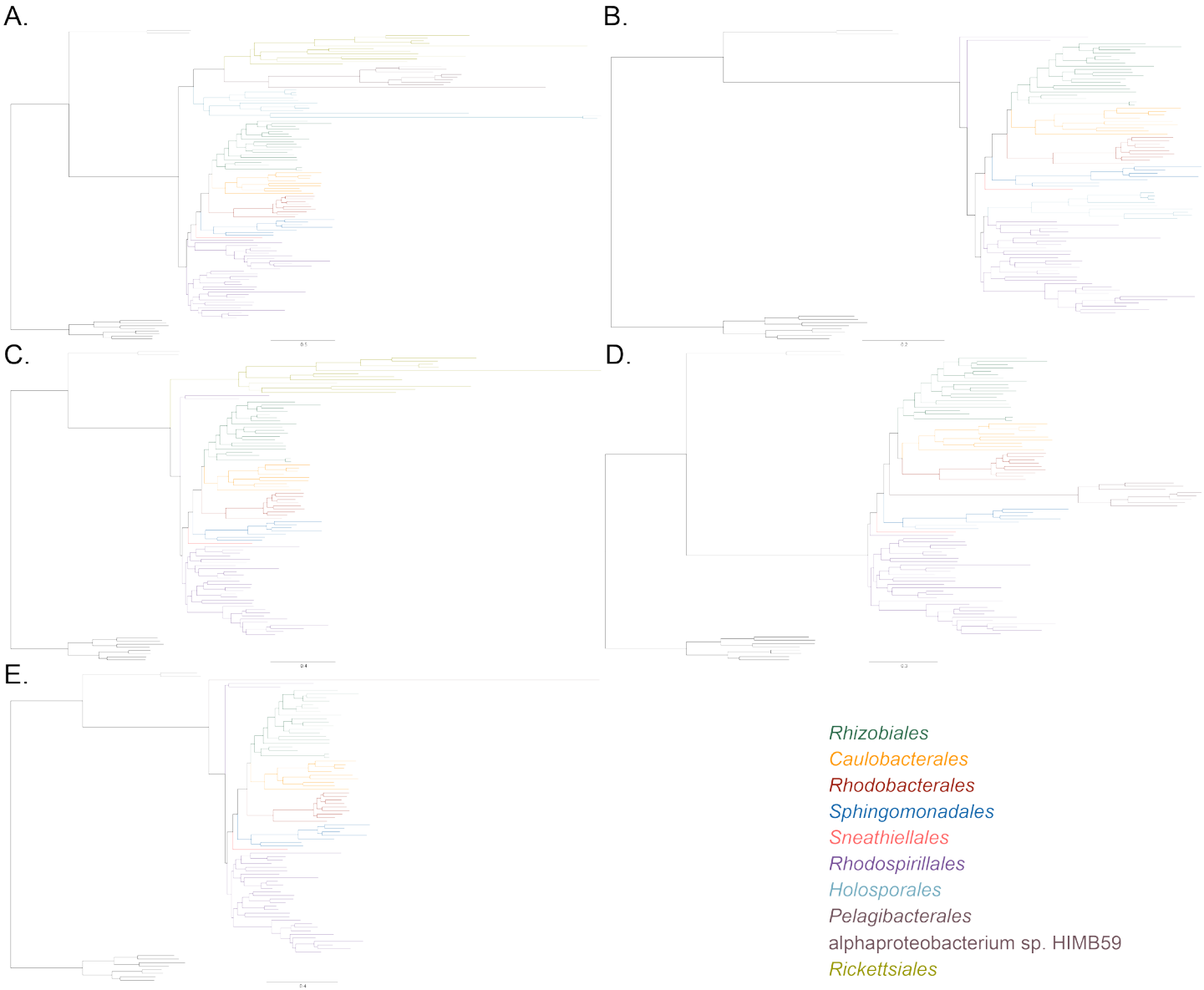
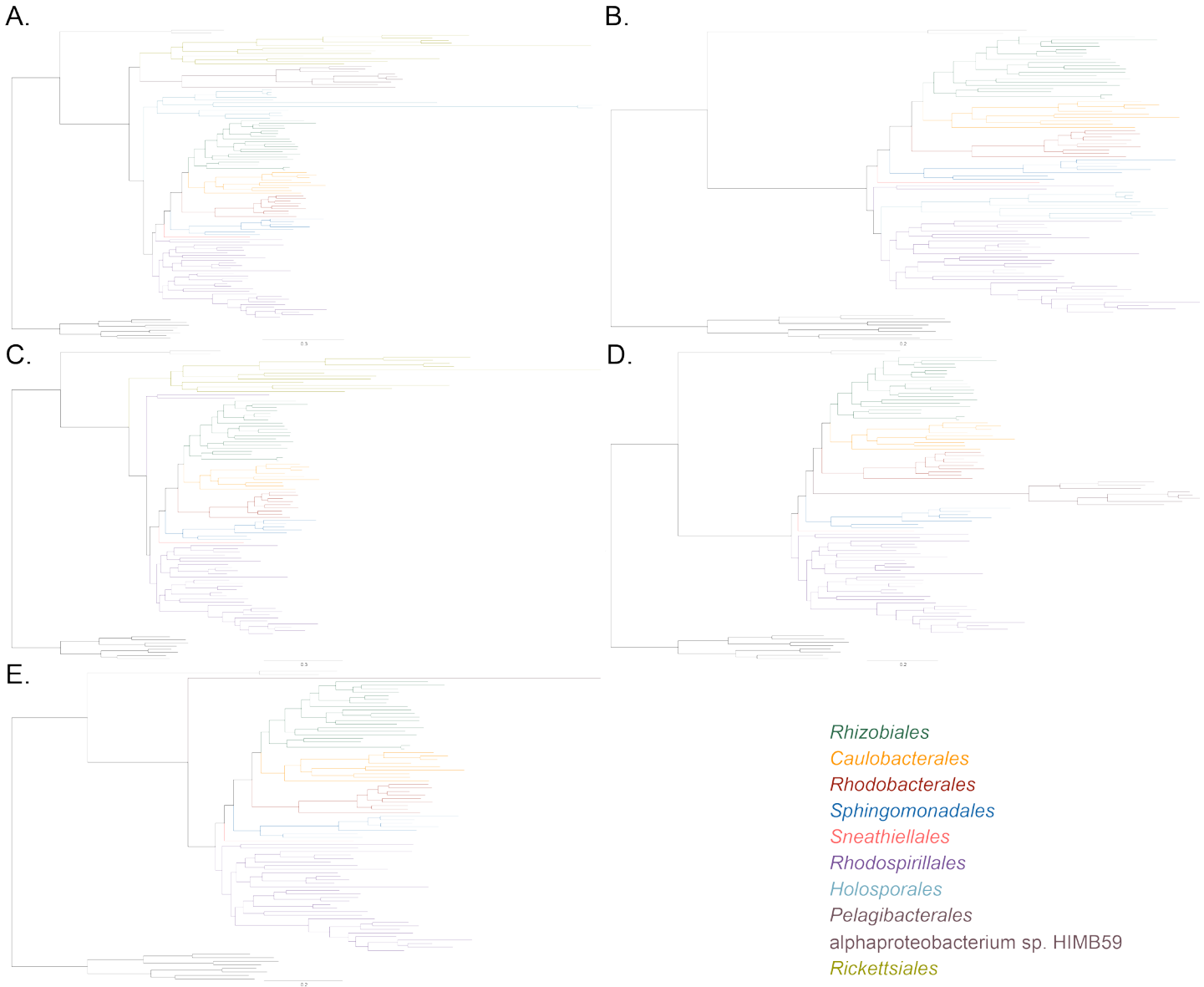

A.
B.
C.
D.
E.
Rhizobiales
Caulobacterales
Rhodobacterales
Sphingomonadales
Sneathiellales
Rhodospirillales
Holosporales
Pelagibacterales
alphaproteobacterium sp. HIMB59
Rickettsiales

**Figure S4.13.** Maximum-likelihood phylograms obtained from the 40 most compositionally homogeneous genes and removing long-branching taxa. **A.** A phylogram for which all long-branching taxa are included. **B.** A phylogram to assess the placement of the *Holosporales*. **C.** A phylogram to assess the placement of the *Rickettsiales*. **D.** A phylogram to assess the placement of the *Pelagibacterales*. **E.** A phylogram to assess the placement of alphaproteobacterium sp. HIMB59.


A.
B.
C.
D.
E.
Rhizobiales
Caulobacterales
Rhodobacterales
Sphingomonadales
Sneathiellales
Rhodospirillales
Holosporales
Pelagibacterales
alphaproteobacterium sp. HIMB59
Rickettsiales

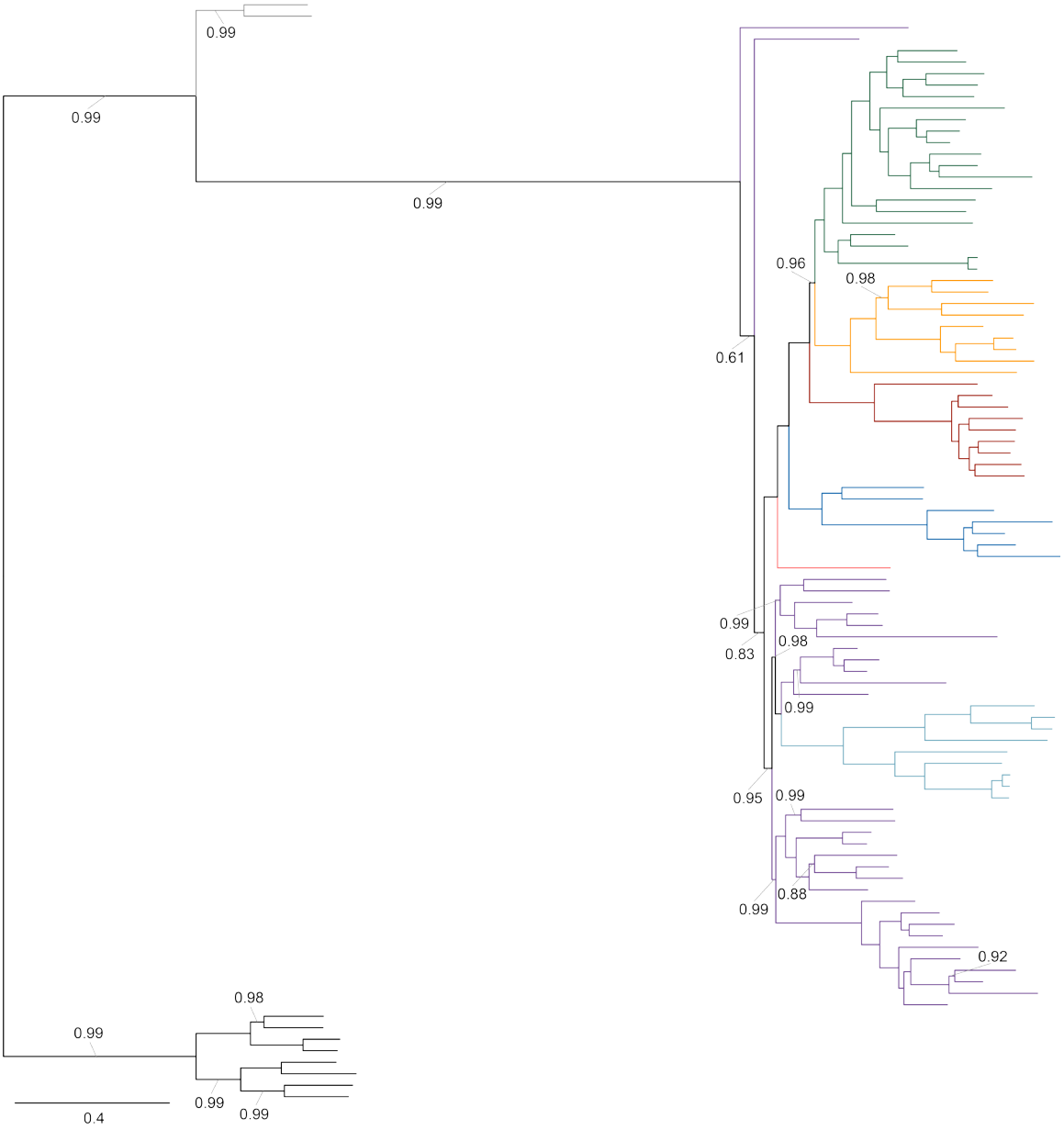



**Figure S4.14.** Bayesian consensus phylogram inferred to place the *Holosporales* in the absence of the *Rickettsiales* and the *Pelagibacterales* and when the data have been recoded into a six-character state alphabet (the dataset-specific recoding scheme S6: AQEHISV RKMT PY DCLF NG W) to reduce compositional heterogeneity. Branch support values are 1.0 posterior probabilities unless annotated. *Rhizobiales* (green), *Caulobacterales* (orange), *Rhodobacterales* (red), *Sphingomonadales* (blue), *Sneathiellales* (pink), *Rhodospirillales* (purple), *Holosporales* (light blue), *Magnetococcales* (gray), *Beta-* and *Gammeproteobacteria* (black).

**Table S4.1.** A list of the least compositionally heterogeneous genes out of the 200 single-copy and vertically-inherited genes used in this study.

| # | p-value | chi-square | Marker | Annotation | Length |
|---|---|---|---|---|---|
| 1 | 1.0000 | 1.0000 | Alpha.18 | rpsL 30S ribosomal protein S12 | 123 |
| 2 | 1.0000 | 1.0000 | Alpha.66 | recA recombinase A | 332 |
| 3 | 0.9700 | 1.0000 | Alpha.227 | rpsQ 30S ribosomal protein S17 | 76 |
| 4 | 0.9400 | 1.0000 | Alpha.74 | rplT ribosomal protein L20 | 118 |
| 5 | 0.8600 | 1.0000 | Alpha.34 | rplN 50S ribosomal protein L14 | 122 |
| 6 | 0.8600 | 1.0000 | Alpha.60 | rpsI 30S ribosomal protein S9 | 132 |
| 7 | 0.7900 | 1.0000 | Alpha.77 | rpsS 30S ribosomal protein S19 | 92 |
| 8 | 0.7500 | 1.0000 | Alpha.118 | hypothetical protein | 165 |
| 9 | 0.7400 | 1.0000 | Alpha.69 | 30S ribosomal protein S14 | 101 |
| 10 | 0.7400 | 1.0000 | Alpha.107 | 30S ribosomal protein S11 | 122 |
| 11 | 0.7300 | 1.0000 | Alpha.256 | rpsT 30S ribosomal protein S20 | 87 |
| 12 | 0.6800 | 1.0000 | Alpha.136 | 30S ribosomal protein S10 | 101 |
| 13 | 0.6700 | 1.0000 | Alpha.31 | rplK 50S ribosomal protein L11 | 142 |
| 14 | 0.6700 | 1.0000 | Alpha.36 | rpsD 30S ribosomal protein S4 | 205 |
| 15 | 0.6100 | 1.0000 | Alpha.128 | preprotein translocase subunit YajC | 88 |
| 16 | 0.5500 | 1.0000 | Alpha.65 | rpsM 30S ribosomal protein S13 | 122 |
| 17 | 0.5400 | 1.0000 | Alpha.123 | nuoK2 NADH quinone oxidoreductase subunit 11 (chain K) | 102 |
| 18 | 0.4700 | 1.0000 | Alpha.42 | NAD(P)H-quinone oxidoreductase subunit 3 | 120 |
| 19 | 0.4500 | 1.0000 | Alpha.79 | Ribosomal protein L19 | 121 |
| 20 | 0.4400 | 1.0000 | Alpha.70 | rpsH 30S ribosomal protein S8 | 132 |
| 21 | 0.4200 | 1.0000 | Alpha.68 | rpmA 50S ribosomal protein L27 | 84 |
| 22 | 0.3600 | 1.0000 | Alpha.99 | rplM 50S ribosomal protein L13 | 154 |
| 23 | 0.3600 | 1.0000 | Alpha.172 | ATP-dependent HslUV protease peptidase subunit HslV | 176 |
| 24 | 0.3300 | 1.0000 | Alpha.9 | glutaredoxin-like protein grla | 106 |
| 25 | 0.3200 | 1.0000 | Alpha.37 | rpsC 30S ribosomal protein S3 | 213 |
| 26 | 0.2900 | 1.0000 | Alpha.233 | transcriptional regulator | 163 |
| 27 | 0.2800 | 1.0000 | Alpha.12 | 30S ribosomal protein S7 | 156 |
| 28 | 0.2700 | 1.0000 | Alpha.147 | rpsR 30S ribosomal protein S18 | 73 |
| 29 | 0.2600 | 1.0000 | Alpha.67 | DNA-directed RNA polymerase subunit alpha | 331 |
| 30 | 0.2400 | 1.0000 | Alpha.210 | 30S ribosomal protein S6 | 105 |
| 31 | 0.2100 | 1.0000 | Alpha.117 | rpsP 30S ribosomal protein S16 | 103 |
| 32 | 0.2100 | 1.0000 | Alpha.235 | ppa inorganic pyrophosphatase | 174 |
| 33 | 0.1600 | 1.0000 | Alpha.56 | hypothetical protein | 213 |
| 34 | 0.1200 | 1.0000 | Alpha.93 | rplQ 50S ribosomal protein L17 | 124 |
| 35 | 0.1000 | 1.0000 | Alpha.170 | nucleoside diphosphate kinase | 140 |
| 36 | 0.1000 | 1.0000 | Alpha.215 | iojap protein family | 103 |
| 37 | 0.0900 | 1.0000 | Alpha.39 | rplE 50S ribosomal protein L5 | 179 |
| 38 | 0.0900 | 1.0000 | Alpha.73 | rplV 50S ribosomal protein L22 | 116 |
| 39 | 0.0600 | 1.0000 | Alpha.41 | infC translation initiation factor IF-3 | 165 |
| 40 | 0.0600 | 1.0000 | Alpha.149 | 30S ribosomal protein S15 | 89 |

# APPENDIX C. SUPPLEMENTARY MATERIAL FOR CHAPTER 5

**Figure S5.1.** Abundances estimated as mean coverage values for each alphaproteobacterial MAG reconstructed across metagenomic samples for both the Llamara saltern and the Alchichica lake. For Llamara MAGs, their abundances were estimated only for metagenomic samples coming from the Llamara saltern. For Alchichica MAGs, their abundances were estimated only for metagenomic samples coming from the Alchichica lake. See Fig. 5.1 for further detailes about these MAGs.

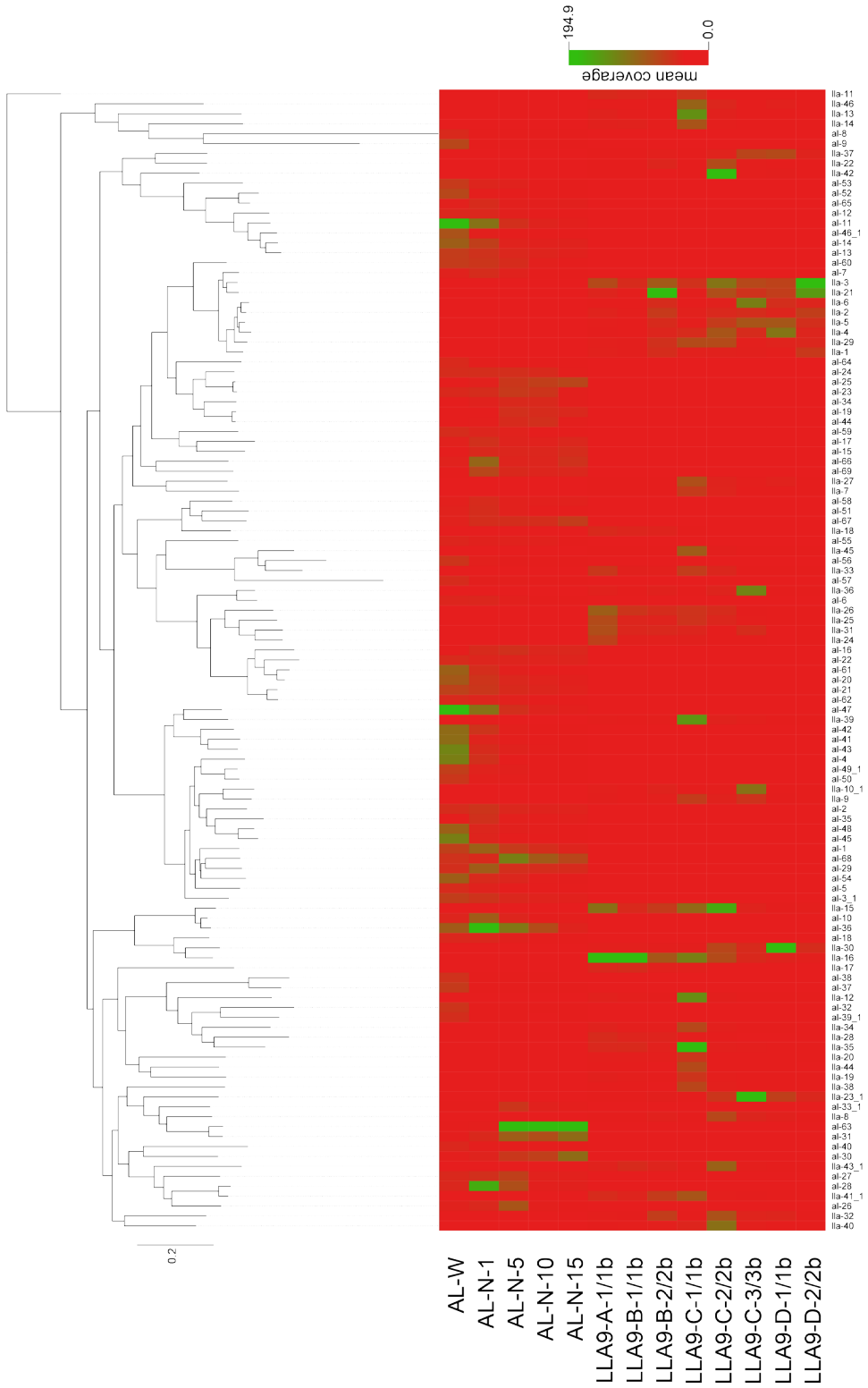

mean coverage
194.9
0.0
0.2
AL-W
AL-N-1
AL-N-5
AL-N-10
AL-N-15
LLA9-A-1/1b
LLA9-B-1/1b
LLA9-B-2/2b
LLA9-C-1/1b
LLA9-C-2/2b
LLA9-C-3/3b
LLA9-D-1/1b
LLA9-D-2/2b

**Figure S5.2.** Labeled phylogenetic tree for Fig. 5.2. See legend of Fig. 5.2 for details. Branch support values are SH-aLRT (%) and FBoot2+NNI (%).

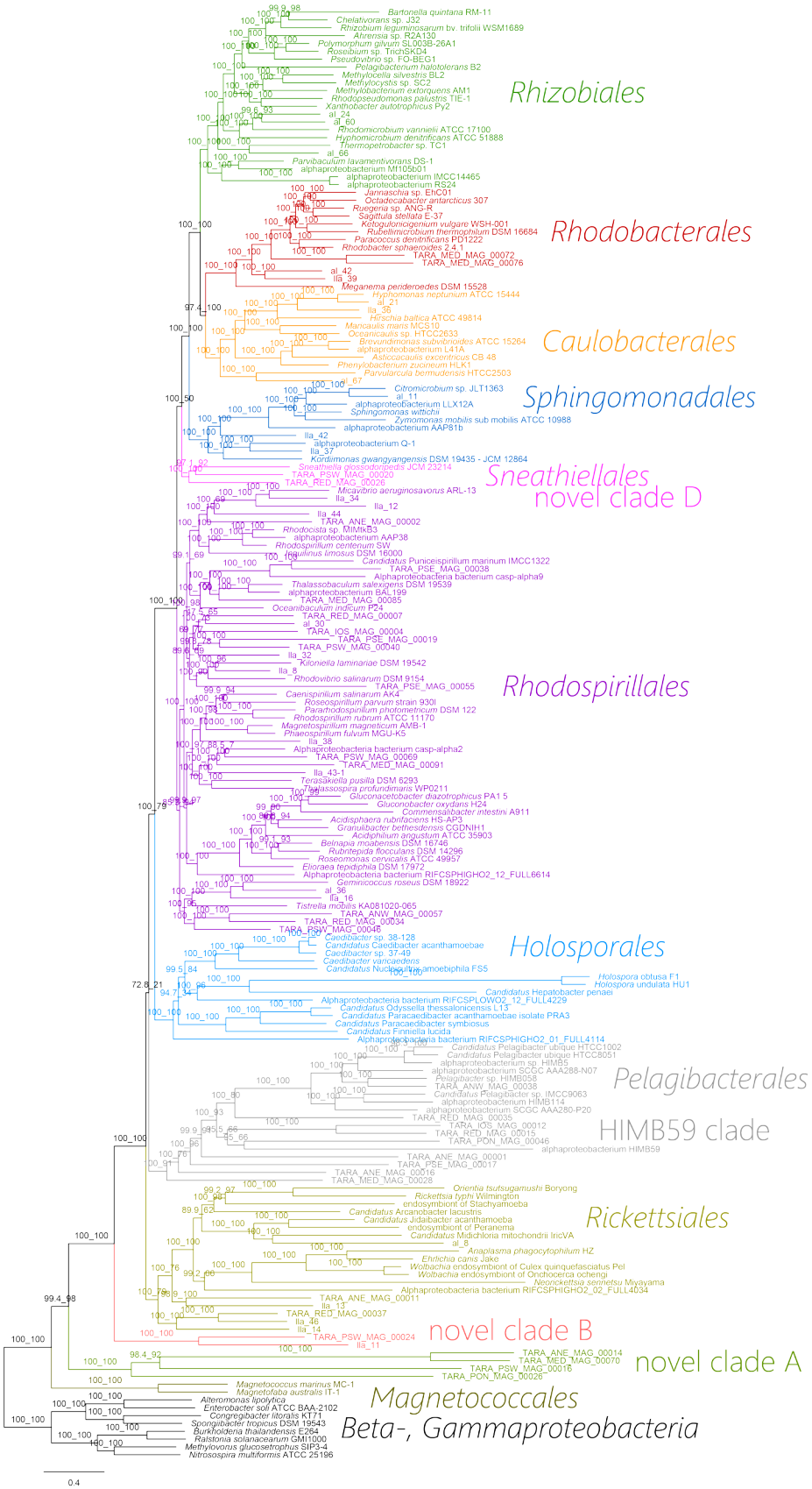

Rhizobiales
Rhodobacterales
Caulobacterales
Sphingomonadales
Sneathiellales
novel clade D
Rhodospirillales
Holosporales
Pelagibacterales
HIMB59 clade
Rickettsiales
novel clade B
novel clade A
Magnetococcales
Beta-, Gammaproteobacteria
0.4

**Figure S5.3.** Labeled phylogenetic tree for Fig. 5.3. See legend of Fig. 5.3 for details. Branch support values are SH-aLRT (%) and FBoot2+NNI (%).

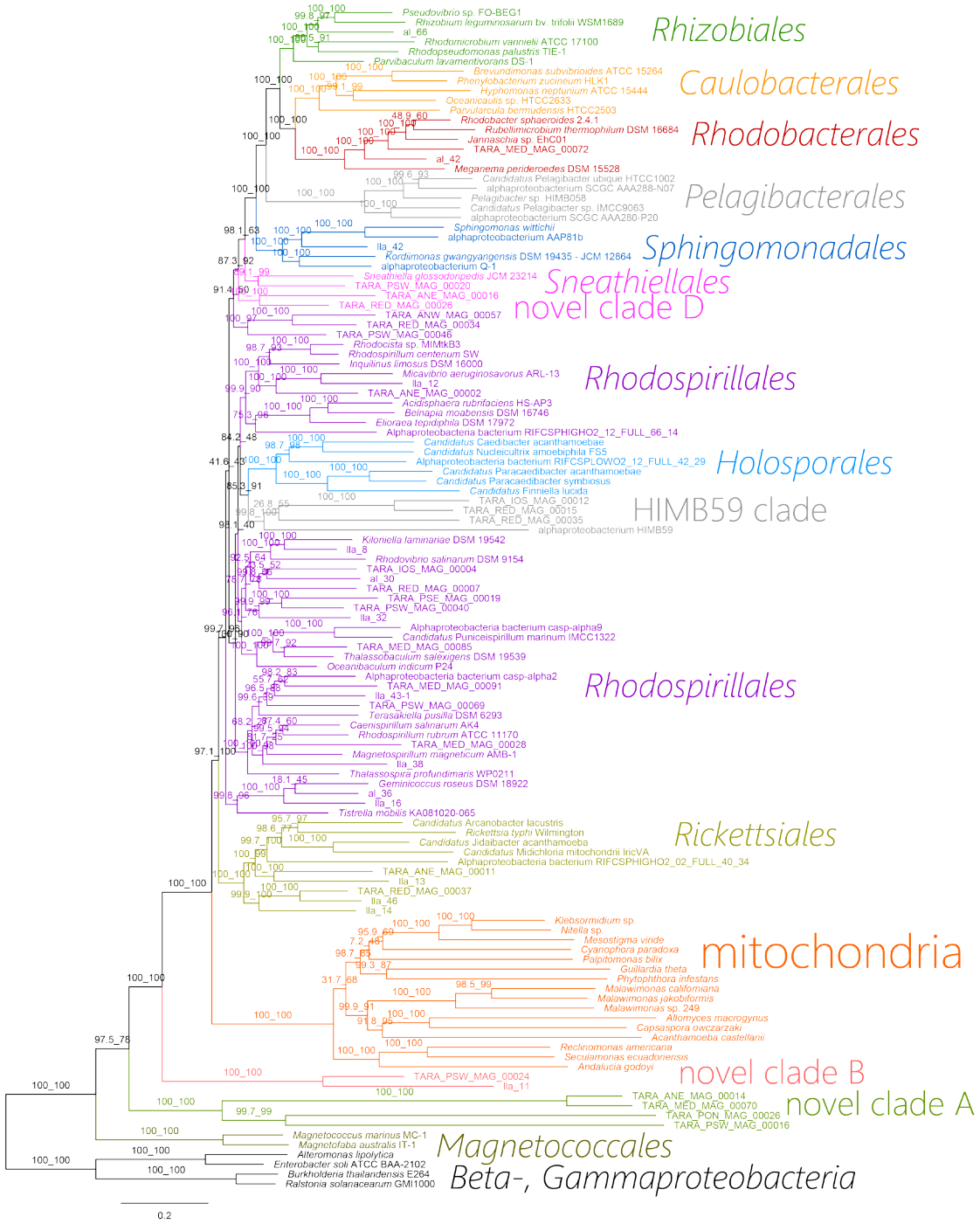
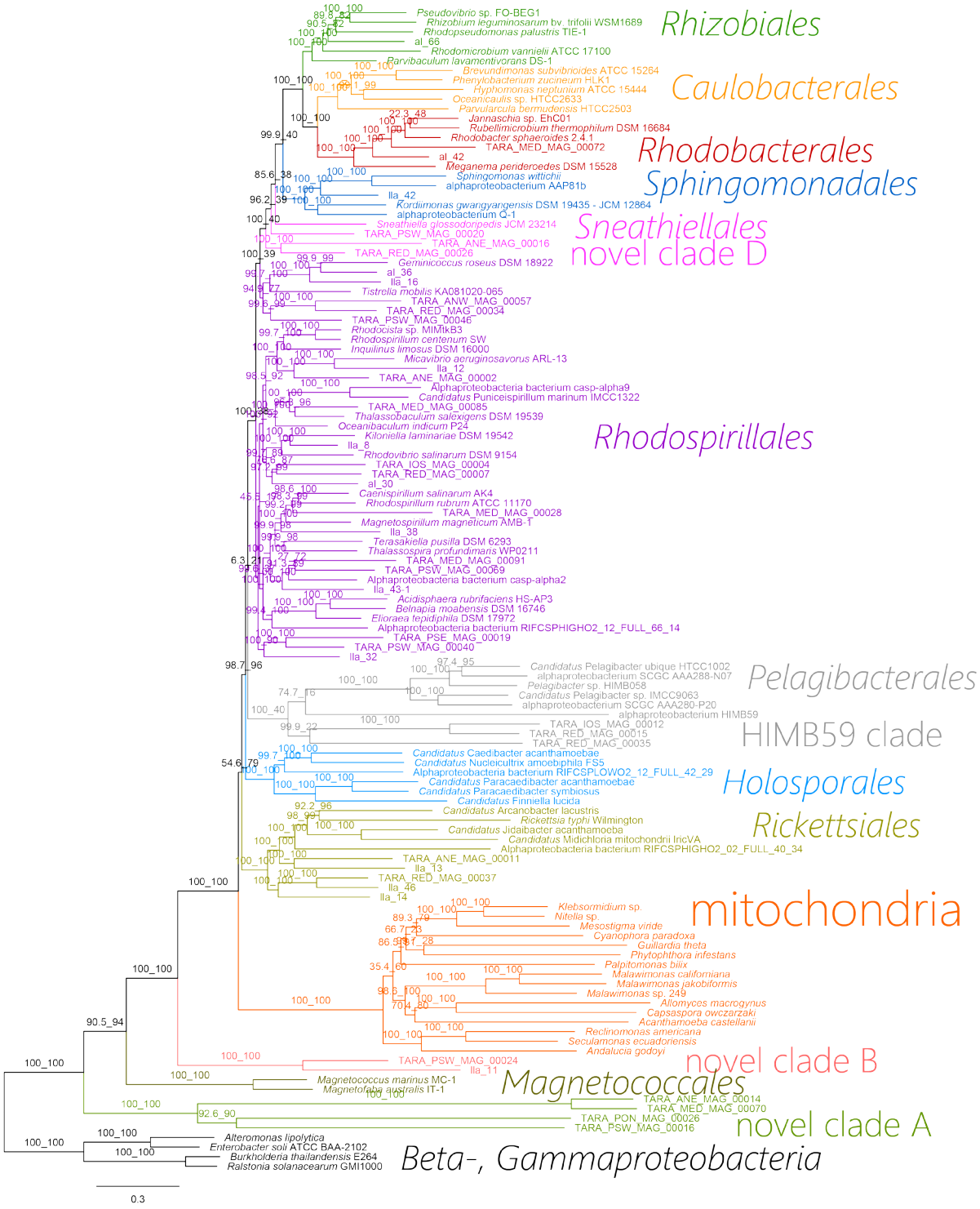

Rhizobiales
Caulobacterales
Rhodobacterales
Pelagibacterales
Sphingomonadales
Sneathiellales
novel clade D
Rhodospirillales
Holosporales
HIMB59 clade
Rhodospirillales
Rickettsiales
mitochondria
novel clade B
novel clade A
Magnetococcales
Beta-, Gammaproteobacteria
Pseudovibrio sp. FO-BEG1
Rhizobium leguminosarum bv. trifolii WSM1689
al_66
Rhodomicrobium vannielii ATCC 17100
Rhodopseudomonas palustris TIE-1
Parvibaculum lavamentivorans DS-1
Brevundimonas subvibrioides ATCC 15264
Phenylobacterium zucineum HLK1
Hyphomonas neptunium ATCC 15444
Oceanicaulis sp. HTCC2633
Parvularcula bermudensis HTCC2503
Rhodobacter sphaeroides 2.4.1
Rubellimicrobium thermophilum DSM 16684
Jannaschia sp. EhC01
TARA_MED_MAG_00072
al_42
Meganema perideroedes DSM 15528
Candidatus Pelagibacter ubique HTCC1002
alphaproteobacterium SCGC AAA288-N07
Pelagibacter sp. HIMB058
Candidatus Pelagibacter sp. IMCC9063
alphaproteobacterium SCGC AAA280-P20
Sphingomonas wittichii
alphaproteobacterium AAP81b
lla_42
Kordimonas gwangyangensis DSM 19435 - JCM 12864
alphaproteobacterium Q-1
Sneathiella glossodoripedis JCM 23214
TARA_PSW_MAG_00020
TARA_ANE_MAG_00016
TARA_RED_MAG_00026
TARA_ANW_MAG_00057
TARA_RED_MAG_00034
TARA_PSW_MAG_00046
Rhodocista sp. MIMtkB3
Rhodospirillum centenum SW
Inquilinus limosus DSM 16000
Micavibrio aeruginosavorus ARL-13
lla_12
TARA_ANE_MAG_00002
Acidisphaera rubrifaciens HS-AP3
Belnapia moabensis DSM 16746
Elioraea tepidiphila DSM 17972
Alphaproteobacteria bacterium RIFCSPHIGHO2_12_FULL_66_14
Candidatus Caedibacter acanthamoebae
Candidatus Nucleicultrix amoebiphila FS5
Alphaproteobacteria bacterium RIFCSPLOWO2_12_FULL_42_29
Candidatus Paracaedibacter acanthamoebae
Candidatus Paracaedibacter symbiosus
Candidatus Finniella lucida
TARA_IOS_MAG_00012
TARA_RED_MAG_00015
TARA_RED_MAG_00035
alphaproteobacterium HIMB59
Kiloniella laminariae DSM 19542
lla_8
Rhodovibrio salinarum DSM 9154
TARA_IOS_MAG_00004
al_30
TARA_RED_MAG_00007
TARA_PSE_MAG_00019
TARA_PSW_MAG_00040
lla_32
Alphaproteobacteria bacterium casp-alpha9
Candidatus Puniceispirillum marinum IMCC1322
TARA_MED_MAG_00085
Thalassobaculum salexigens DSM 19539
Oceanibaculum indicum P24
Alphaproteobacteria bacterium casp-alpha2
TARA_MED_MAG_00091
lla_43-1
TARA_PSW_MAG_00069
Terasakiella pusilla DSM 6293
Caenispirillum salinarum AK4
Rhodospirillum rubrum ATCC 11170
TARA_MED_MAG_00028
Magnetospirillum magneticum AMB-1
lla_38
Thalassospira profundimaris WP0211
Geminicoccus roseus DSM 18922
al_36
lla_16
Tistrella mobilis KA081020-065
Candidatus Arcanobacter lacustris
Rickettsia typhi Wilmington
Candidatus Jidaibacter acanthamoeba
Candidatus Midichloria mitochondrii IricVA
Alphaproteobacteria bacterium RIFCSPHIGHO2_02_FULL_40_34
TARA_ANE_MAG_00011
lla_13
TARA_RED_MAG_00037
lla_46
lla_14
Klebsormidium sp.
Nitella sp.
Mesostigma viride
Cyanophora paradoxa
Paipitomonas bilix
Guillardia theta
Phytophthora infestans
Malawimonas californiana
Malawimonas jakobiformis
Malawimonas sp. 249
Allomyces macrogynus
Capsaspora owczarzaki
Acanthamoeba castellanii
Reclinomonas americana
Seculamonas ecuadoriensis
Andalucia godoyi
TARA_PSW_MAG_00024
lla_11
TARA_ANE_MAG_00014
TARA_MED_MAG_00070
TARA_PON_MAG_00026
TARA_PSW_MAG_00016
Magnetococcus marinus MC-1
Magnetofaba australis IT-1
Alteromonas lipolytica
Enterobacter soli ATCC BAA-2102
Burkholderia thailandensis E264
Ralstonia solanacearum GMI1000
0.2

**Figure S5.4.** A phylogenetic tree inferred under the GTR+C60S4+F+G4 model from the 113-gene dataset (29,640 amino acid sites) recoded into the dataset-specific four-character recoding scheme S4 (ANQESWY CGIPV RHLKM DFT). Novel MAGs are indicated by terminal branch filled circles. Branch support values are SH-aLRT (%) and FBoot2+NNI (%).


Rhizobiales
Caulobacterales
Rhodobacterales
Sphingomonadales
Sneathiellales
novel clade D
Rhodospirillales
Pelagibacterales
HIMB59 clade
Holosporales
Rickettsiales
mitochondria
novel clade B
Magnetococcales
novel clade A
Beta-, Gammaproteobacteria
Pseudovibrio sp. FO-BEG1
Rhizobium leguminosarum bv. trifolii WSM1689
Rhodopseudomonas palustris TIE-1
al_66
Rhodomicrobium vannielii ATCC 17100
Parvibaculum lavamentivorans DS-1
Brevundimonas subvibrioides ATCC 15264
Phenylobacterium zucineum HLK1
Hyphomonas neptunium ATCC 15444
Oceanicaulis sp. HTCC2633
Parvularcula bermudensis HTCC2503
Jannaschia sp. EhC01
Rubellimicrobium thermophilum DSM 16684
Rhodobacter sphaeroides 2.4.1
TARA_MED_MAG_00072
al_42
Meganema perideroedes DSM 15528
Sphingomonas wittichii
alphaproteobacterium AAP81b
lla_42
Kordiimonas gwangyangensis DSM 19435 - JCM 12864
alphaproteobacterium Q-1
Sneathiella glossodoripedis JCM 23214
TARA_PSW_MAG_00020
TARA_ANE_MAG_00016
TARA_RED_MAG_00026
Geminicoccus roseus DSM 18922
al_36
lla_16
Tistrella mobilis KA081020-065
TARA_ANW_MAG_00057
TARA_RED_MAG_00034
TARA_PSW_MAG_00046
Rhodocista sp. MIMtkB3
Rhodospirillum centenum SW
Inquilinus limosus DSM 16000
Micavibrio aeruginosavorus ARL-13
lla_12
TARA_ANE_MAG_00002
Alphaproteobacteria bacterium casp-alpha9
Candidatus Puniceispirillum marinum IMCC1322
TARA_MED_MAG_00085
Thalassobaculum salexigens DSM 19539
Oceanibaculum indicum P24
Kiloniella laminariae DSM 19542
lla_8
Rhodovibrio salinarum DSM 9154
TARA_IOS_MAG_00004
TARA_RED_MAG_00007
al_30
Caenispirillum salinarum AK4
Rhodospirillum rubrum ATCC 11170
TARA_MED_MAG_00028
Magnetospirillum magneticum AMB-1
lla_38
Terasakiella pusilla DSM 6293
Thalassospira profundimaris WP0211
TARA_MED_MAG_00091
TARA_PSW_MAG_00089
Alphaproteobacteria bacterium casp-alpha2
lla_43_1
Acidisphaera rubrifaciens HS-AP3
Belnapia moabensis DSM 16746
Elioraea tepidiphila DSM 17972
Alphaproteobacteria bacterium RIFCSPHIGHO2_12_FULL_66_14
TARA_PSE_MAG_00019
TARA_PSW_MAG_00040
lla_32
Candidatus Pelagibacter ubique HTCC1002
alphaproteobacterium SCGC AAA288-N07
Pelagibacter sp. HIMB058
Candidatus Pelagibacter sp. IMCC9063
alphaproteobacterium SCGC AAA280-P20
alphaproteobacterium HIMB59
TARA_IOS_MAG_00012
TARA_RED_MAG_00015
TARA_RED_MAG_00035
Candidatus Caedibacter acanthamoebae
Candidatus Nucleicultrix amoebiphila FS5
Alphaproteobacteria bacterium RIFCSPLOWO2_12_FULL_42_29
Candidatus Paracaedibacter acanthamoebae
Candidatus Paracaedibacter symbiosus
Candidatus Finniella lucida
Candidatus Arcanobacter lacustris
Rickettsia typhi Wilmington
Candidatus Jidaibacter acanthamoeba
Candidatus Midichloria mitochondrii IricVA
Alphaproteobacteria bacterium RIFCSPHIGHO2_02_FULL_40_34
TARA_ANE_MAG_00011
lla_13
TARA_RED_MAG_00037
lla_46
lla_14
Klebsormidium sp.
Nitella sp.
Mesostigma viride
Cyanophora paradoxa
Guillardia theta
Phytophthora infestans
Palpitomonas bilix
Malawimonas californiana
Malawimonas jakobiformis
Malawimonas sp. 249
Allomyces macrogynus
Capsaspora owczarzaki
Acanthamoeba castellanii
Reclinomonas americana
Seculamonas ecuadoriensis
Andalucia godoyi
TARA_PSW_MAG_00024
lla_11
Magnetococcus marinus MC-1
Magnetofaba australis IT-1
TARA_ANE_MAG_00014
TARA_MED_MAG_00070
TARA_PON_MAG_00026
TARA_PSW_MAG_00016
Alteromonas lipolytica
Enterobacter soli ATCC BAA-2102
Burkholderia thailandensis E264
Ralstonia solanacearum GMI1000
0.3

# APPENDIX D. COPYRIGHT PERMISSIONS

Chapter 1 is partly based on the following published manuscript:

> Roger, A. J., **Muñoz-Gómez, S. A.**, & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. Current Biology, 27(21), R1177–R1192.

Chapter 2 is based on the following published manuscripts:

> **Muñoz-Gómez, S. A.**, Slamovits, C. H., Dacks, J. B., Baier, K. A., Spencer, K. D., & Wideman, J. G. (2015). Ancient Homology of the Mitochondrial Contact Site and Cristae Organizing System Points to an Endosymbiotic Origin of Mitochondrial Cristae. Current Biology, 25(11), 1489–1495.
> **Muñoz-Gómez, S. A.**, Slamovits, C. H., Dacks, J. B., & Wideman, J. G. (2015). The evolution of MICOS: Ancestral and derived functions and interactions. Communicative & Integrative Biology, 8(6), e1094593.
> **Muñoz-Gómez, S. A.**, Wideman, J. G., Roger, A. J., & Slamovits, C. H. (2017). The Origin of Mitochondrial Cristae from Alphaproteobacteria. Molecular Biology and Evolution, 34(4), 943–956.

Chapter 3 takes arguments and ideas from the following published manuscript:

> **Muñoz-Gómez, S. A.**, Wideman, J. G., Roger, A. J., & Slamovits, C. H. (2017). The Origin of Mitochondrial Cristae from Alphaproteobacteria. Molecular Biology and Evolution, 34(4), 943–956.

Chapter 4 is based on the submitted manuscript:

> **Muñoz-Gómez, S. A.**, Hess, S., Burger, G., Lang, F., Susko, E., Slamovits, C. H. & Roger, A. J. (2018). An updated phylogeny of the *Alphaproteobacteria* reveals that the parasitic *Rickettsiales* and *Holosporales* have independent origins. Submitted.

The introduction of Chapter 5 is partly based on the following published manuscript:

> Roger, A. J., **Muñoz-Gómez, S. A.**, & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. Current Biology, 27(21), R1177–R1192.


| | |
|---|---|
| **Title:** | Ancient Homology of the Mitochondrial Contact Site and Cristae Organizing System Points to an Endosymbiotic Origin of Mitochondrial Cristae |
| **Author:** | Sergio A. Muñoz-Gómez,Claudio H. Slamovits,Joel B. Dacks,Kaitlyn A. Baier,Katelyn D. Spencer,Jeremy G. Wideman |
| **Publication:** | Current Biology |
| **Publisher:** | Elsevier |
| **Date:** | 1 June 2015 |

## OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Dec 05, 2018

This Agreement between Mr. Sergio Munoz ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

| | |
|---|---|
| License Number | 4482840794561 |
| License date | Dec 05, 2018 |
| Licensed content publisher | Oxford University Press |
| Licensed content publication | Molecular Biology and Evolution |
| Licensed content title | The Origin of Mitochondrial Cristae from Alphaproteobacteria |
| Licensed content author | Muñoz-Gómez, Sergio A.; Wideman, Jeremy G. |
| Licensed content date | Jan 12, 2017 |
| Type of Use | Thesis/Dissertation |
| Institution name | |
| Title of your work | The anatomical and phylogenetic nature of the first mitochondrial ancestor |
| Publisher of your work | Dalhousie University |
| Expected publication date | Dec 2018 |
| Permissions cost | 0.00 CAD |
| Value added tax | 0.00 CAD |
| Total | 0.00 CAD |
| Title | The anatomical and phylogenetic nature of the first mitochondrial ancestor |
| Institution name | Dalhousie University |
| Expected presentation date | Dec 2018 |
| Portions | Figure 1, Figure 2 |
| Requestor Location | Mr. Sergio Munoz<br>5850 College Street<br>8B1 - Sir Charles Tupper Building<br>Dalhousie University<br>Halifax, NS B3H 4R2<br>Canada<br>Attn: Mr. Sergio Munoz |
| Publisher Tax ID | GB125506730 |
| Billing Type | Invoice |
| Billing Address | Mr. Sergio Munoz<br>5850 College Street<br>8B1 - Sir Charles Tupper Building<br>Dalhousie University<br>Halifax, NS B3H 4R2<br>Canada<br>Attn: Mr. Sergio Munoz |
| Total | 0.00 CAD |

Terms and Conditions

**STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL**

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:
v1.4

## OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Dec 05, 2018

This Agreement between Mr. Sergio Munoz ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

| | |
|---|---|
| License Number | 4482841062567 |
| License date | Dec 05, 2018 |
| Licensed content publisher | Oxford University Press |
| Licensed content publication | Molecular Biology and Evolution |
| Licensed content title | The Origin of Mitochondrial Cristae from Alphaproteobacteria |
| Licensed content author | Muñoz-Gómez, Sergio A.; Wideman, Jeremy G. |
| Licensed content date | Jan 12, 2017 |
| Type of Use | Thesis/Dissertation |
| Institution name | |
| Title of your work | The anatomical and phylogenetic nature of the first mitochondrial ancestor |
| Publisher of your work | Dalhousie University |
| Expected publication date | Dec 2018 |
| Permissions cost | 0.00 CAD |
| Value added tax | 0.00 CAD |
| Total | 0.00 CAD |
| Title | The anatomical and phylogenetic nature of the first mitochondrial ancestor |
| Institution name | Dalhousie University |
| Expected presentation date | Dec 2018 |
| Portions | Text extracts from pages 943-952. |
| Requestor Location | Mr. Sergio Munoz<br>5850 College Street<br>8B1 - Sir Charles Tupper Building<br>Dalhousie University<br>Halifax, NS B3H 4R2<br>Canada<br>Attn: Mr. Sergio Munoz |
| Publisher Tax ID | GB125506730 |
| Billing Type | Invoice |
| Billing Address | Mr. Sergio Munoz<br>5850 College Street<br>8B1 - Sir Charles Tupper Building<br>Dalhousie University<br>Halifax, NS B3H 4R2<br>Canada<br>Attn: Mr. Sergio Munoz |
| Total | 0.00 CAD |

Terms and Conditions

**STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL**

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:
v1.4

Journal
Communicative & Integrative Biology ›
Volume 8, 2015 - Issue 6

Open access

947 Views | 8 CrossRef citations to date | 3 Altmetric

Article Addendum

# The evolution of MICOS: Ancestral and derived functions and interactions

Sergio A Muñoz-Gómez, Claudio H Slamovits, Joel B Dacks & Jeremy G Wideman

Article: e1094593 | Received 19 Aug 2015, Accepted 08 Sep 2015, Accepted author version posted online: 12 Oct 2015, Published online: 12 Oct 2015

Download citation https://doi.org/10.1080/19420889.2015.1094593 Check for updates