ISOLATION AND CHARACTERIZATION OF *CANDIDATUS CHLORACIDOBACTERIUM THERMOPHILUM*

A Dissertation in
Biochemistry, Microbiology, and Molecular Biology
by
Amaya M. Garcia Costas

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

December 2010
The dissertation of Amaya M. Garcia Costas was reviewed and approved* by the following

**Donald A. Bryant**
Ernest C. Pollard Professor of Biotechnology and Professor of Biochemistry and Molecular Biology
Dissertation Advisor
Chair of Committee

**Jean E. Brenchley**
Professor of Biochemistry and Molecular Biology

**John H. Golbeck**
Professor of Biochemistry and Biophysics
Professor of Chemistry

**Jennifer L. Macalady**
Assistant Professor of Geosciences

**Stephan C. Schuster**
Professor of Biochemistry and Molecular Biology

**Scott B. Selleck**
Professor of Biochemistry and Molecular Biology
Head of the Department of Biochemistry and Molecular Biology

*Signatures are on file with the Graduate School
ABSTRACT
Historically, new microorganisms have been discovered either by setting up enrichment cultures in the laboratory that favor particular types of physiologies or by sequencing the amplified 16S ribosomal RNA gene from DNA extracted from environmental samples. These approaches have been successful in exposing the large variety of physiological capabilities (enrichment cultures) and vast phylogenetic diversity (16S rRNA surveys) of the microbial world. Yet they pose severe limitations to furthering our understanding of microbial diversity and to the discovery of microorganisms with novel physiologies. Another approach, metagenomics, has emerged as a powerful tool to study community composition, to define physiological capabilities of microbial communities and to discover novel microorganisms. A particularly fascinating group of microorganisms to study are the chlorophototrophs. Chlorophototrophs use membrane-embedded chlorophyll-protein complexes called reaction centers to harvest light energy and transform it into chemical energy. Prior to the work presented here, there were five known bacterial phyla that contained chlorophototrophs, including the *Cyanobacteria, Chlorobi, Chloroflexi, Firmicutes* and *Proteobacteria*. These distinct groups vary in their reaction centers, accessory pigments and carbon metabolism among other properties. Microbial mats in alkaline hot springs of Yellowstone National Park are dominated by chlorophototrophic microorganisms and provide an ideal setting for studying them. The mats from two particular springs, Octopus and Mushroom Springs, have been the subject of over three decades of research, and much is known about their geochemistry and microbial composition. Two phyla of chlorophototrophs, the *Cyanobacteria* and the *Chloroflexi* reside in these low-carbonate and low-sulfide mats. A metagenomic study conducted on these mats revealed the presence of a third type of chlorophototroph that had escaped detection by other approaches and that belongs to the phylum *Acidobacteria*, a phylum previously not known to contain chlorophototrophs. An analysis of the metagenome predicted this new chlorophototroph, *Candidatus Chloracidobacterium thermophilum*, to have Type-1 reaction centers, the Fenna Matthew Olson protein and chlorosomes as antenna structures, properties also found in the chlorophototrophic *Chlorobi*. Unlike the *Chlorobi*ales, however, *Cab. thermophilum* was predicted to be an aerobe. The present work describes the isolation of this new chlorophototroph in culture and its initial characterization. An enriched culture of *Cab. thermophilum* has been generated from a cyanobacterial enrichment culture cultivated from the mats of Octopus Spring that contained *Cab. thermophilum* as a minor component. Like other *Acidobacteria*, *Cab. thermophilum* is difficult to culture and has long generation times and fastidious growth requirements. In addition, the culture contains heterotrophic microorganisms that seem to be providing unidentified, essential growth factor(s). Physiological studies of this culture have confirmed the metagenomic predictions that *Cab. thermophilum* is an aerobic chlorophototroph that synthesizes chlorosomes containing bacteriochlorophyll c as antenna pigments. In order to understand the physiology of *Cab. thermophilum* further, its genome has been completely sequenced. It consists of two chromosomes, both of which harbor essential genes. The genome contains all of the genes required for phototrophy with chlorosomes as antenna structures but
lacks key genes of all known carbon fixation pathways, as well as genes for assimilatory nitrate and sulfate reduction. In addition, it lacks the biosynthetic pathways for the synthesis of the amino acids valine, isoleucine and leucine. These genomic analyses clearly define Cab. thermophilum as a chlorophototroph that is dependent on other members of the mat community for essential nutrients. The structure of the chlorosome antenna complex of Cab. thermophilum has been investigated and its protein, lipid, quinone and pigment composition elucidated. Although synthesized by an aerobe, the chlorosomes resemble those of other green bacteria in their general shape and by the presence of CsmA and CsmI-like proteins. However, they contain additional unique proteins and lipids. Moreover, the presence of the xanthophylls echinenone and canthaxanthin reflects the aerobic environment from which Cab. thermophilum was isolated. Like the chlorophototrophic Chlorobi, Cab. thermophilum synthesizes three types of chlorophylls, bacteriochlorophyll (BChl) c, chlorophyll (Chl) a and BChl a. High performance liquid chromatography analyses combined with mass spectrometry have revealed that Cab. thermophilum methylates its BChl c at the C-8 and C-12 positions and that its C-17 propionic group is esterified with a variety of isoprenoid and straight alkane moieties. The most abundant BChl c species, especially at high light intensities, has been found to be [8-iBu, 12-Et]-BChl c esterified with the unbranched C-18 alcohol, octadecanol. Interestingly, although Cab. thermophilum is an aerobe, its chlorosomes exhibit redox-dependent quenching of fluorescence emission. Lastly, carotenoid biosynthesis in Cab. thermophilum has been investigated. Three genes predicted to code for a lycopene cyclase, a ketolase, and a hydroxylase respectively have been expressed heterologously and their enzymatic activities confirmed.
TABLE OF CONTENTS

List of Tables ........................................................................................................... ix
List of Figures .......................................................................................................... x
List of Abbreviations ............................................................................................... xiv
Acknowledgements ................................................................................................. xv

CHAPTER 1. Introduction to microbial diversity, phototroph diversity and
Octopus and Mushroom Springs ................................................................................. 1
  1.1. Assessing microbial diversity ...................................................................... 2
    1.1.1. Why study microbial diversity ........................................................... 2
    1.1.2. The beginnings of our understanding of microbial diversity ................. 3
    1.1.3. The molecular revolution .................................................................. 6
    1.1.4. Genomics and Metagenomics ............................................................ 7
  1.2. Bacterial phototrophic diversity ................................................................. 10
    1.2.1. What is phototrophy? ........................................................................ 10
    1.2.2. Bacterial chlorophototrophs .............................................................. 12
  1.3. Octopus Spring ......................................................................................... 15
  1.4. Thesis goal and scope ............................................................................. 19
  1.5. References ............................................................................................... 21

CHAPTER 2. Isolation and Initial Characterization of *Candidatus*
Chloracidobacterium thermophilum ........................................................................... 36
Abstract .................................................................................................................. 37
  2.1. Introduction .............................................................................................. 38
  2.2. Methods ..................................................................................................... 41
    2.2.1. Culturing techniques ........................................................................ 41
    2.2.2. Physiological studies .......................................................................... 42
    2.2.3. Fluorescence *In Situ* Hybridization (FISH) ................................... 43
    2.2.4. Hexanol-treatment of *Candidatus*
        Chloracidobacterium thermophilum cells ............................................. 43
    2.2.5. Thin-section electron microscopy ...................................................... 43
  2.3. Results ........................................................................................................ 44
    2.3.1. Detection of *Candidatus*
        Chloracidobacterium thermophilum in the
        Cyanobacterial enrichment ................................................................. 44
    2.3.2. Isolation of *Candidatus* Chloracidobacterium
        thermophilum ....................................................................................... 45
    2.3.3. Physiological properties of *Candidatus*
        Chloracidobacterium thermophilum culture and
        further purification attempts ................................................................. 46
    2.3.4. *Candidatus* Chloracidobacterium thermophilum is
        a phototroph ......................................................................................... 48
2.3.5. Absorption spectra of hexanol-treated and untreated *Candidatus* Chloracidobacterium thermophilum cells 49
2.3.6. *Candidatus* Chloracidobacterium thermophilum synthesizes chlorosomes and a slime layer .......... 49
2.4. Discussion ........................................................................ 50
2.5. References ................................................................. 55

CHAPTER 3. Genome sequence and analyses of *Candidatus*
Chloracidobacterium thermophilum........................................ 76
Abstract .................................................................................. 77
3.1. Introduction .......................................................................... 78
3.2. Methods ............................................................................... 80
  3.2.1. Bacterial strains, culture conditions and DNA isolation ........................................... 80
  3.2.2. DNA sequencing and assembly ........................................................................ 80
  3.2.3. Sequence annotation and analyses ................................................................... 81
  3.2.4. Phylogenetic analyses ............................................................................... 81
3.3. Results and Discussion ....................................................... 82
  3.3.1. General genome features ............................................................................... 82
  3.3.2. Pigment biosynthesis and reaction center ....................................................... 83
  3.3.3. Electron transport .................................................................................. 84
  3.3.4. Metabolism and carbon acquisition .............................................................. 87
  3.3.5. Transposases and repetitive elements .......................................................... 89
  3.3.6. Comparison to *Candidatus* Chloracidobacterium thermophilum type M .................. 90
  3.3.7. Comparison to other Acidobacteria ................................................................ 93
  3.3.8. Other genes of interest ........................................................................ 94
  3.3.9. Summary ................................................................................ 94
3.4. References ........................................................................... 96

CHAPTER 4. Biochemical characterization of the chlorosomes of
*Candidatus* Chloracidobacterium thermophilum ....................... 122
Abstract ................................................................................ 123
4.1. Introduction ........................................................................ 124
4.2. Methods ............................................................................ 130
  4.2.1. Culture conditions ........................................................................ 130
  4.2.2. Chlorosome isolation ........................................................................ 130
  4.2.3. Electron microscopy ........................................................................ 131
  4.2.4. Cryoimaging ................................................................................ 131
  4.2.5. HPLC and Mass Spectrometry analyses of Bacteriochlorophylls ...................... 131
  4.2.6. Quinone and carotenoid analyses ....................................................... 132
  4.2.7. Lipid analyses ................................................................................ 133
  4.2.8. Polyacrylamide gel electrophoresis ....................................................... 133
4.2.9. Amino terminus sequencing and mass spectrometry analyses of chlorosome proteins.......................... 134
4.2.10. Absorbance and Fluorescence emission spectroscopy ............................................................. 134
4.3. Results ................................................................................................................................. 135
4.3.1. Chlorosome isolation and morphology ................................................................. 135
4.3.2. Bacteriochlorophyll c internal arrangement ............................................... 135
4.3.3. Bacteriochlorophyll c homologs ................................................................. 136
4.3.4. Bacteriochlorophyll α and Chlorophyll α analyses ........................................ 137
4.3.5. Carotenoids .............................................................................................................. 137
4.3.6. Menadione ................................................................................................................ 138
4.3.7. Lipids .......................................................................................................................... 139
4.3.8. SDS-PAGE and protein alignment ............................................................ 139
4.3.9. Spectroscopy ............................................................................................................ 140
4.4. Discussion .................................................................................................................... 141
4.5. References .................................................................................................................. 149

CHAPTER 5. Heterologous expression of Candidatus Chloracidobacterium thermophilum .......................................................... 180
Abstract .................................................................................................................................. 181
5.1. Introduction .................................................................................................................. 182
5.2. Methods ....................................................................................................................... 185
5.2.1. Molecular manipulations and cloning .............................................................. 185
5.2.2. In vitro assay of CrtO activity ............................................................................. 186
5.2.3. Protein induction conditions ................................................................................. 186
5.2.4. Pigment extraction and analysis ............................................................................. 187
5.2.5. Phylogenetic analyses ............................................................................................ 187
5.3. Results ........................................................................................................................... 187
5.3.1. Expression of crtYc-crtYd in lycopene-producing
Escherichia coli ............................................................................................................... 187
5.3.2. Expression of crtC in lycopene producing
Escherichia coli ............................................................................................................... 188
5.3.3. Expression of crtO in γ-carotene producing
Escherichia coli ............................................................................................................... 188
5.3.4. In vitro modification of β-carotene by Candidatus
Chloracidobacterium thermophilum CrtO expressed
in Escherichia coli .............................................................................................................. 188
5.3.5. Expression of cruF in lycopene-producing
Escherichia coli ............................................................................................................... 189
5.3.6. Phylogeny of carotenoid-biosynthesis enzymes in
Candidatus Chloracidobacterium thermophilum ............................................................. 189
5.4. Discussion .................................................................................................................... 190
5.5. References .................................................................................................................. 196
CHAPTER 6. Concluding remarks ................................................................. 219

APPENDIX A. ............................................................................................... 223
APPENDIX B................................................................. 225
APPENDIX C................................................................................................. 229
APPENDIX D. ................................................................................................. 230
LIST OF TABLES

CHAPTER 1
Table 1.1 Physical and chemical properties of Octopus Spring.......................... 29

CHAPTER 2
Table 2.1 Primers used in this study................................................................. 60

CHAPTER 3
Table 3.1 General genome features of Candidatus Chloracidobacterium Thermophilum and three additional available acidobacterial genomes .......... 103
Table 3.2 Energy harvesting genes in Candidatus Chloracidobacterium thermophilum .............................................................. 104
Table 3.3 Menaquinone genes in Candidatus Chloracidobacterium thermophilum .............................................................. 105
Table 3.4 Location and similarities of transposases found in Candidatus Chloracidobacterium thermophilum .............................................................. 106
Table 3.5 Genes found in Candidatus Chloracidobacterium thermophilum but not in Type M strain .............................................................. 107
Table 3.6 Genes found in Type M strain but not in Candidatus Chloracidobacterium thermophilum .............................................................. 109

CHAPTER 4
Table 4.1 Masses of Bacteriochlorophyll c homologs ........................................... 158
Table 4.2 Mass of carotenoid and quinone species ................................................. 159
Table 4.3 Properties of proteins in chlorosomes of Candidatus Chloracidobacterium thermophilum .............................................................. 160

CHAPTER 5
Table 5.1 Primers used in this study................................................................. 201
Table 5.2. Properties of plasmids used in this study .................................................. 202
LIST OF FIGURES

CHAPTER 1
Figure 1.1. Bacterial phylogenetic tree from 1987. .................................................. 30
Figure 1.2. Bacterial phylogenetic tree from 2003 ..................................................... 31
Figure 1.3. Bacteriorhodopsin diagram ................................................................. 32
Figure 1.4. Reaction center diagram ...................................................................... 33
Figure 1.5. Diversity of chlorophototrophs .............................................................. 34
Figure 1.6. Octopus Spring and its microbial community ........................................... 35

CHAPTER 2
Figure 2.1. Nucleotide alignment of the 16S rRNA genes of two
Candidatus Chloracidobacterium thermophilum populations .................................. 61
Figure 2.2. Amino acid alignment of the PscA protein of two Candidatus
Chloracidobacterium thermophilum populations ................................................... 62
Figure 2.3. Fluorescence in situ hybridization micrograph of
cyanobacterial enrichment .................................................................................. 63
Figure 2.4. Images of the cultures used and generated in this study ...................... 64
Figure 2.5. Absorption spectra of the cultures used and generated in this
study ..................................................................................................................... 65
Figure 2.6. Phase contrast micrographs of the cultures used and generated in this
study ..................................................................................................................... 66
Figure 2.7. Images of colonies formed in solid media by heterotrophic contaminants in the Candidatus Chloracidobacterium thermophilum
culture ................................................................................................................. 67
Figure 2.8. Growth cycle in the Candidatus Chloracidobacterium
thermophilum culture .......................................................................................... 68
Figure 2.9. Images of the different organisms found in the Candidatus
Chloracidobacterium thermophilum culture following density
centrifugation ....................................................................................................... 69
Figure 2.10. Phase contrast micrographs of the separated components of
the Candidatus Chloracidobacterium thermophilum culture ................................ 70
Figure 2.11. Qy peak absorption spectra of the Candidatus
Chloracidobacterium thermophilum cells ........................................................... 71
Figure 2.12. PCR amplification with primers specific for Candidatus
Chloracidobacterium thermophilum and Anoxybacillus spp ................................ 72
Figure 2.13. Absorption spectra of the Candidatus Chloracidobacterium
thermophilum culture .......................................................................................... 73
Figure 2.14. Fluorescence micrograph of autofluorescent Candidatus
Chloracidobacterium thermophilum cells ........................................................... 74
Figure 2.15. Thin-section electron micrograph of a Candidatus
Chloracidobacterium thermophilum cell ................................................................ 75
CHAPTER 3
Figure 3.1. Map of the two chromosomes of Candidatus Chloracidobacterium thermophilum. ................................................................. 110
Figure 3.2. Distribution of CDS belonging to different Cluster of Orthologous Groups on the two chromosomes of Candidatus Chloracidobacterium .................................................................................................................. 111
Figure 3.3. Neighbor-joining phylogenetic tree of B(Chl) synthases rooted with the Escherichia. coli gene ubiA. ........................................ 112
Figure 3.4. Representative nuo operons found in Candidatus Chloracidobacterium thermophilum ........................................................................... 113
Figure 3.5. Putative operons involved in the oxidation of menaquinol ................................................................. 114
Figure 3.6. KEGG diagram showing biochemical pathways involved in sulfur metabolism................................................................. 115
Figure 3.7. KEGG diagram showing biosynthetic pathways for the amino acids valine, leucine and isoleucine. ................................................................. 116
Figure 3.8. KEGG diagram showing ABC transporters found in Candidatus Chloracidobacterium thermophilum. ................................................................. 117
Figure 3.9. KEGG diagram showing biochemical pathways involved in nitrogen metabolism................................................................. 118
Figure 3.10. Representative CRISPR gene neighborhood ......................................................................................................................... 119
Figure 3.11. Comparison of the genome of the lab isolate with the genome of the dominant Candidatus Chloracidobacterium thermophilum species in the mat ............................................................................................................. 120
Figure 3.12. 16S neighbor-joining phylogenetic tree showing the relation of cultured and uncultured acidobacteria belonging to groups 1-6 and 8 ......................................................................................................................... 121

CHAPTER 4
Figure 4.1. Absorption spectrum of isolated chlorosomes from Candidatus Chloracidobacterium thermophilum. ................................................................. 161
Figure 4.2. Electron micrograph of negatively stained isolated chlorosomes ......................................................................................................................... 162
Figure 4.3. Bacteriochlorophyll c arrangement in the chlorosomes of Candidatus Chloracidobacterium thermophilum ................................................................. 163
Figure 4.4. RP-HPLC elution profile of BCHls extracted from Candidatus Chloracidobacterium thermophilum. ................................................................. 164
Figure 4.5. Molecular structure of the major species of BCHl c found in the chlorosomes of Cab. thermophilum ......................................................................................................................... 165
Figure 4.6. MS/MS profile of the major species of BCHl c found in the chlorosomes of Cab. thermophilum. ................................................................. 166
Figure 4.7. RP-HPLC elution profile of BCHls extracted from Candidatus Chloracidobacterium thermophilum cells that have been cultured under two different light intensities. ................................................................. 167
Figure 4.8. RP-HPLC elution profile and absorbance spectrum of Chl a and BChl a. ................................................................. 168
Figure 4.9. Carotenoids in Candidatus Chloracidobacterium thermophilum. ........................................................................ 169
Figure 4.10. Quinones in Candidatus Chloracidobacterium thermophilum. ................................................................. 170
Figure 4.11. GC elution profile of fatty acids extracted from Candidatus Chloracidobacterium thermophilum chlorosomes. 171
Figure 4.12. HPLC elution profile of total lipids extracted from Candidatus Chloracidobacterium thermophilum. .................. 172
Figure 4.13. SDS-PAGE analyses of protein composition of isolated chlorosomes from Chlorobaculum tepidum, Candidatus Chloracidobacterium thermophilum and Chloroflexus aurantiacus Y-396...................................................... 173
Figure 4.14. Alignment of CsmS and CsmT amino acid sequences from Cab. thermophilum. ........................................... 174
Figure 4.15. Hydropathy plots of CsmS, CsmT and CsmI. ......................... 175
Figure 4.16. Local protein alignment of CsmA sequences ............... 176
Figure 4.17. Alignment of Cabther_17410 ........................................... 177
Figure 4.18. Absorbance spectrum of Bacteriochlorophyll c in cells and chlorosomes of Candidatus Chloracidobacterium thermophilum. 178
Figure 4.19. Fluorescence spectrum of Bacteriochlorophyll c in Candidatus Chloracidobacterium thermophilum under oxidizing and reducing conditions .................................................. 179

CHAPTER 5
Figure 5.1. Proposed carotenoid biosynthetic pathway in Candidatus Chloracidobacterium thermophilum ........................................... 203
Figure 5.2. HPLC analyses of YcYd expressed in lycopene-producing Escherichia coli ................................................................. 204
Figure 5.3. Absorption spectra of carotenoids identified in figure 5.2 205
Figure 5.4. HPLC analyses of CrtC expressed in lycopene-producing Escherichia coli ................................................................. 206
Figure 5.5. Absorption spectra of carotenoids identified in figure 5.4 207
Figure 5.6. HPLC analyses of CrtO expressed in γ-carotene producing Escherichia coli ................................................................. 208
Figure 5.7. Absorption spectra of carotenoids identified in figure 5.6 209
Figure 5.8. HPLC analyses of in vitro CrtO activity on β-carotene 210
Figure 5.9. Absorption spectra of carotenoids identified in figure 5.8 211
Figure 5.10. HPLC analyses of CrtO expressed in γ-carotene producing Escherichia coli ................................................................. 212
Figure 5.11. Absorption spectra of carotenoids identified in figure 5.10 213
Figure 5.12. Alignment of 400 residues from the N-terminus of lycopene cyclases from the CruA family ................................................................. 214
Figure 5.13 Phylogenetic tree of CrtYcYd and CruA ................................................................. 215
Figure 5.14 Phylogenetic tree of CrtO ................................................................. 216
Figure 5.15 Phylogenetic tree of CrtC ................................................................. 217
Figure 5.16 Phylogenetic tree of CrtD ................................................................. 218

APPENDICES
Figure A.1 HPLC analyses of unknowns in spent media collected at various time points from a culture of Candidatus Chloracidobacterium thermophilum ................................................................. 223
Figure A.2. HPLC analyses of unknowns in spent media collected at various time points from a culture of Candidatus Chloracidobacterium thermophilum ................................................................. 224
Figure B.1. Images of cell suspensions of representatives of all green bacteria before and after exposure to high light intensities ................................................................. 225
Figure B.2. Photodegradation of cell suspensions from representatives from all three phyla of green bacteria ................................................................. 226
Figure B.3. Photodegradation Candidatus Chloracidobacterium thermophilum cell suspensions ................................................................. 227
Figure B.4. Photodegradation of chlorosome suspensions representatives from all three phyla of green bacteria ................................................................. 228
Figure C.1. In situ expression of the reaction center gene pscA from Candidatus Chloracidobacterium thermophilum during a diel cycle ................................................................. 229
Figure D.1. Neighbor-joining phylogenetic tree showing the relationships among various Candidatus Chloracidobacterium thermophilum pscA sequences ................................................................. 230
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACIII</td>
<td>alternative complex III</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>BChl</td>
<td>Bacteriochlorophyll</td>
</tr>
<tr>
<td>CAI</td>
<td>codon adaptation index</td>
</tr>
<tr>
<td>CBB</td>
<td>Calvin-Benson-Bassham</td>
</tr>
<tr>
<td>Chl</td>
<td>Chlorophyll</td>
</tr>
<tr>
<td>COG</td>
<td>cluster of orthologous groups</td>
</tr>
<tr>
<td>CRISPR</td>
<td>clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>DGGE</td>
<td>denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>DGTA</td>
<td>diacylglycerolhydroxymethyltrimethyl-β-alanine</td>
</tr>
<tr>
<td>DMAPP</td>
<td>dimethylallyl diphosphate</td>
</tr>
<tr>
<td>DMSP</td>
<td>3-dimethylsulphoniopropionate</td>
</tr>
<tr>
<td>DOXP</td>
<td>1-deoxy-d-xylulose-5-phosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FAP</td>
<td>filamentous anoxygenic phototroph</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescence in situ hybridization</td>
</tr>
<tr>
<td>FMO</td>
<td>Fenna-Matthews-Olson</td>
</tr>
<tr>
<td>GGPP</td>
<td>geranylgeranyl pyrophosphate</td>
</tr>
<tr>
<td>GSB</td>
<td>green sulfur bacteria</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-iperezineethanesulfonic acid</td>
</tr>
<tr>
<td>IMG</td>
<td>integrated microbial genomes</td>
</tr>
<tr>
<td>IPP</td>
<td>isopentenyl pyrophosphate</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>LB</td>
<td>Luria broth</td>
</tr>
<tr>
<td>LHC</td>
<td>light harvesting complex(es)</td>
</tr>
<tr>
<td>MVA</td>
<td>mevalonate</td>
</tr>
<tr>
<td>MS</td>
<td>Mushroom spring</td>
</tr>
<tr>
<td>MPTA</td>
<td>N, N-diethyl-N-[2-(4-methylphenoxy)ethyl]amine</td>
</tr>
<tr>
<td>Na₄EDTA</td>
<td>ethylenediaminetetraacetic acid tetrasodium salt</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>OS</td>
<td>Octopus spring</td>
</tr>
<tr>
<td>RDP</td>
<td>ribosomal database project</td>
</tr>
<tr>
<td>TCA</td>
<td>tricarboxylic acid</td>
</tr>
<tr>
<td>YNP</td>
<td>Yellowstone National Park</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

First of all I would like to acknowledge my advisor Dr. Don Bryant for taking me on, for sharing his passion for science and for pushing me to better my work and my critical thinking even when I resisted. I cannot express justly what a great honor it has been to be one of Don Bryant’s students.

The work presented in this thesis has been possible thanks to a large number of individuals. Several facilities housed under the Huck Institutes for the Life Sciences have helped with their analyses. I must acknowledge Deb Grove and Ashley Price in the Nucleic Acid Facility, Missy Hazen and Ruth Haldeman in the Electron Microscope Facility, James Miller, Heike Benz and Dr. Li in the Mass Proteomics Facility and Elaine Kunz and Nicole Zembower in the Cytometry Facility who provided expertise in their respective areas. David Ward and Chris Klatt at Montana State University-Bozeman and Kurt Toenjes and David Butler at Montana State University-Billings shared their lab facilities, office space and expertise while I was working on the thesis in Montana. Eric Boyd, a post-doc in John Peters lab at MSU-Bozeman, Scott Miller at University of Montana and Nicki Parenteau at NASA Ames Research Center gracefully shared DNA samples from hot springs throughout Yellowstone National Park. Past grad student Joel Graham provided valuable advice on carotenoids and HPLC analysis. Post-docs Rick Alvey, Kajetan Vogl and Yusuke Tsukatani have been a great resource for cloning experiments, anything green-sulfur and carotenoid, and providing a steady supply of Cab. thermophilum cells, respectively. The help of fellow grad student Zhenfeng Liu was essential in closing the genome. Our lab research manager, Gaozhong Shen, has cheerfully and patiently kept the lab going and has helped in too-many-to-count aspects of this thesis.

On a personal note, I must thank the entire department for providing such a wonderful working and learning atmosphere. Linda Kunes, the coordinator of the graduate program, has been kind, extremely helpful and encouraging throughout my stay at Penn State. My committee members, Drs. Brenchley, Golbeck, Schuster and Macalady, have provided respectful assistance and each of them has been a source of inspiration. The Bryant lab has been an amazing place to work at and I will be lucky to find such camaraderie and good-hearted hard-working people in my future endeavors. It has been a pleasure and an honor to be in the lab with all of them and I wish them much happiness and success. I must single out fellow grad students Zhenfeng Liu, Steve Romberger and Yu Xu who have shared the lab with me from the beginning, and post-doc Marcus Ludwig for his unique sense of humor. Special thanks to Rick Alvey and Kajetan Vogl for their friendship and never-ending patience and grace when answering my questions, and to Kajetan
Vogl for the always generous sharing of his house, even after Spain beat Germany on the Euro Cup and the World Cup!

Throughout the thesis I have had a super-special support team in the form of my two daughters, Rebeca and Ana, and my terrific husband, David. The pride and enthusiasm that Rebeca and Ana have taken on my work has been a source of constant encouragement; their happiness and liveliness have always provided a refreshing balance to my life. David is everything I ever dreamed of and more. His unconditional love and dedication have allowed me to achieve this doctorate without having to sacrifice our wonderful life. Thank you.
CHAPTER 1

Introduction to microbial diversity, phototroph diversity and Octopus

and Mushroom Springs
1.1. ASSESSING MICROBIAL DIVERSITY
1.1.1. Why study microbial diversity

When viewed through the human eye, it might seem that our planet is dominated by flowering plants, arthropods and mammals. However, the Earth is a microbial planet, and any understanding of living organisms will not be complete without a study of these small forms of life.

There are more microorganisms (total estimate of $5 \times 10^{30}$) on Earth than any other type of life form, and their total carbon mass is estimated to be equivalent to 60–100% of the estimated total carbon in plants (95). In addition, microorganisms are found in every one of Earth’s habitats, including the boiling hot springs of Yellowstone National Park, ice cores isolated from Antarctica, acid mine drainage runoff, agricultural and forested soils, inside and outside the bodies of most animals, and in the roots and leaves of plants; microorganisms are found up in the clouds and in the deep subsurface more than a mile down into the Earth’s core (3, 52, 59).

For most of our planet’s history, the biota on Earth has been solely composed of microorganisms. The current oldest evidence of life consists of fossilized microbial cells in 3.5 billion year old rocks, whereas the oldest evidence of eukaryotic life, microscopic eukaryotes, dates to 2.1 billion years ago. Multicellular metazoan life does not appear in the fossil record until about 600 million years ago (25). In the often-used exercise of compressing all of life’s history to a 24-hour period, life on Earth has been exclusively microbial for about 21 of those hours (71).
Furthermore, through processes such as nitrogen fixation and photosynthesis, microorganisms control the biosphere. It is estimated that 50% of the atmospheric oxygen is produced by marine phytoplankton and 60% of the fixed nitrogen is produced by microorganisms (74). In addition, we are now able to use diverse microbial metabolisms to clean up oil spills and soils contaminated with heavy metals and to generate biofuels. Microorganisms are also being used to cure diseases and the role of the human gut microbiota in disease is now beginning to be understood. Microorganisms were the first source of many current antibiotics and, undoubtedly, the search for the next generation of antimicrobial agents will involve further mining of secondary metabolites, biochemical pathways and genomes of yet-uncultured microorganisms. Because of their ubiquitous presence and their critical role in the environment and in our health it is imperative that we study microbial diversity.

1.1.2. The beginnings of our understanding of microbial diversity

The first documented instance of human awareness of the presence of microorganisms was in 1676 when Antoine van Leeuwenhoek presented his observations to the Royal Society of London of the “white matter” he had found in his teeth using a rudimentary lens he had himself built (6). The next two centuries saw significant advances in lens designs and microscopes, as well as in chemistry and physics (e.g., gas laws and discovery of many natural elements). These findings set the stage for the emergence of the field of microbiology, which began
in earnest in the late 18th century with Edward Jenner’s development of the smallpox vaccine (28).

The 19th century was a time for great microbiological advancements, such as Louis Pasteur’s germ theory of disease and Robert Koch’s advances in culturing and his postulates on infectious diseases; this century, however, also marks the beginning of our understanding of microbial diversity. During the 1870s and 1880s Ferdinand Cohn began classifying bacteria based on their shape, pigmentation, and abilities to ferment or to cause disease (28). At the same time two microbiologists, considered by many the fathers of environmental microbiology, were using enrichment cultures to isolate microorganisms from the environment. The first, Sergei Winogradsky, in Russia, was studying soil microorganisms and was able to isolate two different bacteria, from the *Nitrosomonas* and *Nitrobacter* genera, each responsible for one step of nitrification. He also discovered the process of chemoautotrophy and made significant contributions to our understanding of sulfur oxidation in *Beggiatoa* (69). Martin Beijerinck, a Dutch microbiologist, discovered the bacterial processes of nitrogen fixation and sulfate reduction, among many others (49). In addition, Beijerinck was the first professor of Microbiology at the Delft Polytechnical Institute and became the founder of the Delft School of Microbiology. It was in this Dutch school that the field of environmental microbiology was officially founded (49, 86).

A notable member of this school was Cornelius B. van Niel, a disciple of Albert Kluyver, who had himself been one of Beijerinck’s students. C. B. van Niel made significant contributions to our understanding of many different
microorganisms and bacterial processes, and he is well known for his studies on purple and green photosynthetic bacteria and insights into the nature of photosynthesis. However, van Niel made an equally significant contribution in his capacity of educator. In 1929 he moved to the U.S., taught at the Hopkins Marine Station of Stanford University, and founded a summer course in microbial diversity, which was attended by aspiring microbiologists from all over the country (7). Among his many students and postdoctoral associates was Roger Stanier, who studied *Cyanobacteria* and many other types of microorganisms and, who along with van Niel, promoted the use of the word “prokaryote” (36, 75). Another van Niel student was Ralph Wolfe who studied methanogenic bacteria (now *Archaea*) and pioneered the biochemical studies that deciphered the methanogenesis pathway (23). Ralph Wolfe and Roger Stanier were postdoctoral mentors to the author’s M. Sc. and Ph.D. thesis advisors, respectively, and many modern environmental microbiologists can likewise trace their lineage up to C. B. van Niel and the Delft School of Microbiology.

By the 1950s many of today’s well-known type strains, from fermentative to photosynthetic, from anaerobic to aerobic, heterotrophs and autotrophs, cellulose degraders, sulfur and nitrogen oxidizers and reducers, etc., had been isolated through the application of enrichment cultures and other cultivation techniques mostly pioneered by the Delft School of Microbiology.
1.1.3. The molecular revolution

In 1977, twenty-four years after the discovery of the structure of DNA, Carl Woese and G. Fox published a landmark paper in microbiology. By comparing nucleotide sequences of 16S ribosomal RNA, they showed that the prokaryotes encompass two different groups of microorganisms, the Bacteria and the Archaeabacteria (now Archaea) (96). With this molecular approach it was possible, for the first time, to classify all the known microorganisms within an evolutionary framework. Later, and using similar 16S sequencing technology, Woese proposed a phylogenetic tree for the Bacteria, which showed a possible evolutionary relationship among all of the bacterial clades known to date (97). Figure 1 shows this tree with the 11 different bacterial phyla known at the time.

Norman Pace pioneered the use of these molecular tools to assess microbial diversity in the environment in a culture-independent way. Beijenrick had presciently predicted that the enrichment cultures that were being employed to isolate microorganisms were selecting for those microorganisms that found the laboratory conditions advantageous and were leaving many other strains behind (8). By applying Carl Woese’s molecular methods to the study of environmental microbial communities, Pace was able to assess the microbial composition of a particular environment directly, bypassing the need for culturing those microorganisms in the laboratory (55). This new approach involved first extracting DNA from an environmental sample and then amplifying and sequencing the 5S and 16S RNA genes from all microorganisms present in that environment, regardless of whether these organisms grew in a laboratory culture or not. His early
studies on bacterial diversity in the Sargasso sea community and in various hot springs in Yellowstone National Park (8, 70), as well as subsequent surveys of many other environments, have revealed that the cultivation studies of the previous 150 years had yielded only about 1% of the total types of microorganisms present in Earth’s ecosystems (2, 56).

By incorporating data acquired through this technology, already-known bacterial groups have seen new members added to their lineage and new clades of previously unknown microorganisms have been discovered. Figure 2 shows a recent 16S phylogenetic tree of the Bacteria with 52 clades, many of which have no cultured representatives. Despite documented shortcomings of this molecular approach (60, 73, 78, 87, 88) and the development of more sophisticated methods (see below), 16S rRNA studies continue to be used as a first approach to identify the microbial composition of virtually any ecosystem.

1.1.4. Genomics and Metagenomics

DNA sequencing technologies greatly improved as a result of the much-publicized human genome sequencing project, ushering in the era of microbial genomics. The first completely sequenced bacterial genome was that of *Haemophilus influenzae* in 1995 (31), and the first archaeal genome, *Methanococcus jannaschii*, was completed in the next year (22). By 2000 there were over 30 prokaryotic genomes sequenced, with a reported rate at that time of about one genome sequenced every two months (32). As of the spring of 2010, the Genomes On-Line Database (http://www.genomesonline.org/) contained 5791
bacterial and 279 archaeal completely sequenced or “in progress” genomes, and bacterial genomes can now be sequenced in a few hours.

Just as Norman Pace applied Woese’s 16S sequencing approach from cultured isolates to the study of environmental samples, the field of genomics was adapted to study microbial communities in situ; the field of metagenomics was born (40). As with 16S environmental surveys, metagenomics involves extracting DNA from an environmental sample and sequencing it. Unlike 16S surveys, however, the goal in metagenomics is to identify functional genes present in a microbial community to establish the types of biochemical processes taking place, and when possible, the organisms responsible for those processes. In some instances, such as in simple communities or by sequencing extremely deeply in more complex communities, metagenomics might make it possible to obtain complete genome assemblies for individual organisms.

One of the critical limitations of metagenomics is that only fragments of DNA are being sequenced rather than continuous genomes. Consequently information regarding operon organization and regulation, for example, is not directly obtained with metagenomic analyses (41, 98). In addition, it is rare to encounter a DNA fragment linking phylogenetical and functional genes (the “who is there” and “what are they doing” genes). Instead, a metagenomic library will typically contain many DNA fragments each containing a few functional genes at best (41, 98). Despite these limitations, metagenomics is emerging as the tool of choice for understanding microbial species composition of a community and gathering an initial understanding of community dynamics.
The first studies of metagenomics focused on simple microbial communities such as the Sargasso Sea microbial community (10, 76), an acid mine drainage site (81), and whale carcass communities (80), although there have also been some early studies on complex soil communities (24, 64). Analyses of the Sargasso Sea and the acid mine drainage metagenomes became the poster children of the possibilities of this approach in the early 2000s. The Sargasso Sea metagenome linked a gene encoding bacteriorhodopsin to a γ-proteobacterium, a member of the SAR206 clade (10). SAR206 and other SAR clades, notably SAR11, have been shown to constitute a large percentage of the total biomass of the phytoplankton in oceans and, thanks to this metagenome study, one was able to begin to understand the physiology and ecology of this significant group of microorganisms (37,72). Because (1) the SAR microorganisms had yet to be cultured in a laboratory and (2) bacteriorhodopsin was at the time believed to only occur in the Archaeal group Halobacteria, a metagenomic approach was essential to the discovery of rhodopsin-like physiology in this important group of microorganisms.

In the analyses of the metagenome of the acid mine drainage community from Iron Mountain, CA the investigators were able to bin different DNA fragments according to G + C content and generate in doing so the nearly complete genomes of the two main species of this simple microbial community: the bacterial Leptospirillum group II and the archaeal Ferroplasma, as well as three other partial genomes of minor species (81). In addition, metabolic reconstructions from these genomes generated hypotheses regarding carbon and nitrogen acquisition in this
ecosystem as well as energy generation and survival strategies employed by the microorganisms in this extreme environment (81).

As with genomics, the field of metagenomics has grown exponentially from a handful of studies in the beginning of the decade to, as of the Fall of 2010, 119 metagenomic projects reported in the Integrated Microbial Genomes with Microbiome Samples (http://img.jgi.doe.gov/cgi-bin/m/main.cgi).

1.2. BACTERIAL PHOTOTROPHIC DIVERSITY
1.2.1. What is phototrophy?

Phototrophy is the process of harnessing light energy to generate chemical energy in the form of ATP. There are two known mechanisms in nature by which this is done. The first mechanism is based on retinal, which is associated with a rhodopsin-like, transmembrane protein, which acts as a light-activated proton gate (figure 1.3) (19, 52). Absorption of a photon by the retinal molecule causes it to undergo an isomerization reaction, converting a trans double bond into a cis double bond; this results in the release of a proton to the outside of the cytoplasmic membrane in the periplasm by the transmembrane protein. Dark reactions return the retinal to the all-trans configuration, and a proton is taken up by the protein from the cytoplasm in processes that involve conformational changes of the protein. The electrochemical gradient thus generated across the membrane is dissipated by ATP synthase producing ATP in the process (19, 52). This type of phototrophy is widespread within the Archaea and Bacteria, including the archaebal phylum Halobacteria, the α-Proteobacteria and γ-Proteobacteria, the Flavobacteria and others in the Bacterial clade (5, 26). Collectively, organisms that can generate
proton motive force using light and retinal proteins have been called “retinalophototrophs” to distinguish them from the chlorophyll (Chl)-based “chlorophototrophs” described below (19).

The second mechanism employs a Chlorophyll (Chl)-protein complex, called a reaction center, embedded in the cytoplasmic membrane. All reaction centers contain a special pair of (bacterio)-Chl molecules as well as bound cofactors that include additional Chls, quinones and, in the case of type I reaction centers, iron-sulfur clusters (figure 1.4) (38). Absorption of a photon by one of the special-pair Chls causes this molecule to move to an energetically excited state. This excited state Chl is a strong reductant and is able to rapidly transfer an electron to one of the cofactors bound in the reaction center, creating what is known as a “charge-separated state”, an oxidized Chl and a reduced electron acceptor (also a Chl or related molecule). From this primary acceptor, the electron is quickly transferred to terminal electron acceptors in the reaction centers (Q_B in type II reaction centers or a [4Fe-4S] cluster in type I reaction centers). Secondary electron transfer events subsequently occur via transmembrane oxidoreductase complexes, some of which translocate protons and generate an electrochemical proton gradient during electron transfer process. This electrochemical gradient is transformed to chemical energy by ATP synthase (12, 19, 52).

Reaction centers are often surrounded by accessory light-harvesting complexes that absorb and transfer excitation energy to the special pair Chls of reaction centers. Despite the similar functions within these antenna structures, there exists in nature a great diversity in structure and mechanism of these
complexes, each having different protein and pigment composition, ratios and structural arrangements. The known light harvesting complexes include the phycobilisome of the *Cyanobacteria*, the light harvesting complexes (LHC) I and II of the purple bacteria and LHC I of the *Chloroflexi*, chlorosomes in both the *Chlorobi* and *Chloroflexi* and the Fenna-Matthew-Olson (FMO) protein in the *Chlorobi* (figure 1.5).

### 1.2.2. Bacterial chlorophototrophs

Chlorophototrophs have thus far been found exclusively in the domain *Bacteria*, and within this domain, in just five different kingdoms/phyla: *Cyanobacteria*, *Chlorobi* (green sulfur bacteria), *Proteobacteria* (purple phototrophs), *Chloroflexi* (filamentous anoxygenic phototrophs), and *Firmicutes* (the Heliobacteria) (12, 19, 52). The chlorophototrophs found within each of these phyla generally form coherent groups in terms of the structure of their reaction centers, pigment composition, and carbon metabolism. With the exception of the *Cyanobacteria*, all of the other chlorophototrophic phyla also include chemotrophic microorganisms (20, 44, 52).

The *Cyanobacteria* are a morphologically diverse group that encompasses single-celled organisms as well as large (30-150 μm length) multicellular, filamentous organisms (93). They are unique among the chlorophototrophs in that (1) they usually utilize two different reaction centers, both a highly reducing type 1 heterodimeric reaction center (photosystem I) and a heterodimeric, type 2 reaction center (photosystem II), both containing Chl *a*; and (2) they evolve oxygen. In
addition, nearly all *Cyanobacteria* are autotrophs that fix carbon through the Calvin-Benson-Bassham (CBB) cycle (18). The only known exception is a yet-uncultured group of *Cyanobacteria* that includes microorganisms that lack genes coding for Calvin cycle and Krebs cycle enzymes and for photosystem II (15, 100). *Cyanobacteria*, formerly called blue-green algae, have been studied for over 200 years and there are now ca. 1000 different species described and numerous type strains.

The green sulfur bacteria (GSB) are strict anaerobes that employ a type 1 reaction center and two antenna structures, the BChl *a*-binding FMO protein and chlorosomes (membrane-enclosed sacs filled with BChl *c*, *d* or *e*) (34). They are autotrophs like the *Cyanobacteria*, but lack ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) and other CBB enzymes and fix CO₂ through the reverse tricarboxylic acid (TCA) cycle (34). The first member of the GSB to be described was *Chlorobium limicola* in 1906 (53) and *Chlorobaculum tepidum*, which is naturally transformable, has become the predominant model organism for this group (33). There are now about 20 described species of chlorophototrophic *Chlorobi*.

Within the *Proteobacteria*, the purple bacteria are a group of chlorophototrophs that utilize a type II reaction center with either BChl *a* or BChl *b* (44). Their antenna structures, LHC I and LHC II, contain BChl *a* or *b* and have been crystallized to a 4.8 and 2 Å resolution respectively (57, 65); additionally they have been the subject of numerous atomic force microscopy studies (77). Purple bacteria are autotrophic via the CBB cycle, although many can live mixotrophically.
using light as a source of energy and organic compounds as a source of carbon (12, 19). Purple bacteria are chlorophototrophic only under anaerobic conditions, but many can grow chemoheterotrophically under oxic conditions (79).

The kingdom/phylum Chloroflexi comprises two groups of chlorophototrophs, the green Chloroflexi and the red Chloroflexi (19, 21). Both groups synthesize type-2 reaction centers that contain BChl a and a LHC I-like antenna. The green Chloroflexi have an additional antenna structure, the chlorosome (see the Chlorobi above). Both types of organisms have the genes encoding the enzymes of the 3-hydroxypropionate cycle (48, 99), and some of the green Chloroflexi are capable of autotrophic growth. Oscillochloris trichoides, a green Chloroflexi, is unusual in having Rubisco and fixes carbon by the CBB pathway while oxidizing sulfide (46). Like the purple bacteria, most Chloroflexi can live chemoheterotrophically under oxic conditions and either mixotrophically or chlorophotoautotrophically under anoxic conditions. Chloroflexus aurantiacus was isolated from Octopus Spring in Yellowstone National Park and was the first member of this group to be described (58).

The Heliobacteria, belonging to the Firmicutes, are the only known chlorophototrophs that have a gram-positive cell wall and produce endospores (51). Like the Cyanobacteria and the Chlorobi, they utilize type-1 reaction centers, but synthesize BChl g as their major Chl pigment. There is no known antenna system for this group of chlorophototrophs (42). Heliobacteria are obligate anaerobes and heterotrophs and analyses of the only available genome, from Heliobacterium modesticum, have revealed no genes for carbon fixation enzymes (68).
Interestingly, the Heliobacteria have the ability to fix nitrogen (a property shared with the *Cyanobacteria*, purple bacteria and *Chlorobi*). The first phototrophic heliobacterium, *Heliobacterium chlorum*, was isolated in 1983 (35), but two other strains, *H. modesticaldum* and *H. mobilis*, have become the preferred models for these group of chlorophototrophs.

### 1.3. OCTOPUS AND MUSHROOM SPRINGS

Octopus Spring (OS) and Mushroom Spring (MS) are thermal features in Yellowstone National Park located in the central western section of the Park near the Great Fountain Geyser in the Lower Geyser Basin. As is the case for most hot springs in this location, OS and MS are alkaline, siliceous hot springs with the temperature of the source water being ca. 95 °C for OS and ca. 75 °C for MS and the pH ca. 8.2, respectively.

MS spring was first described by Thomas Brock in 1967 as

…a spring [that] is away from tourist areas along the Howard Eaton Trail about 0.25 km north of Great Fountain Geyser. Its source is a large pool about 10 by 15 m, and the outflow is through a single channel about 15 cm wide and 5 cm deep which flows west. The channel gradually widens and bends several times, and the water is forced to spread out by the extensive algal layers which develop, although there are always areas with well-drained channels. The effluent eventually widens out into a marsh …. (16)

and a micrograph of microorganisms from the outflow channel of this hot spring can already be found in his landmark paper “Life at High Temperatures” in 1967 (17).

OS was described in the original survey of hot springs conducted in 1883 by A. C. Peale (27). Remarkably, it has remained relatively stable since that first
description, and this stability, along with its location around a bend and out of the public view, has also made OS an ideal site for scientific studies. It consists of a slightly smaller pool than the one found in MS and is located on top of a small hill. OS has two narrow (~1 meter wide) outflow channels, the first channel flows downhill directly west whereas the other channel flows slightly southwest for about 10 meters before fanning out into a large marsh area of flowing hot spring water (figure 1.6).

MS and OS are virtually identical springs in terms of their geochemistry and biology. They are both characterized by high alkalinity, silica deposition and low concentrations of sulfide and phosphate (table 1.1). Both hot springs contain numerous thermophilic chemotrophs, most notably *Thermus aquaticus*, which was first isolated from MS (Brock 1969) and *Thermocrinis ruber* (literally “hot red hair”), isolated from the pink streamers found in the outflow channels of OS (43).

It is, however, the chlorophototrophic mats that develop between the 37°C to 70°C sections of the effluent channels that flow from these pools that have been most investigated. These mats consist of an upper oxic layer of about 1-2 mm where photosynthesis occurs and a lower anoxic layer where decomposition occurs.

The first studies of the microorganisms of the OS phototrophic mats were carried out by Thomas Brock and his students and date back to the early 1970’s. Through microscopy and cultivation studies, it was established that these mats contained two types of microorganisms, the cyanobacterium *Synechococcus lividus* and the filamentous anoxygenic phototroph (FAP) *Cfx. aurantiacus* (27, 50). The ecology, with special focus on photosynthesis and carbon metabolism, of these
microorganisms was extensively studied throughout the 70s and the 80s. $^{14}$C-radiolabeling studies revealed photosynthetic activity by *S. lividus* during the day with the accumulation of glycogen and the excretion of glycolate, especially at high light intensities, due to saturation of the Calvin cycle (9). In the evening several fermentation products such as acetate, propionate, lactate, butyrate, ethanol and valerate appeared as a result of fermentation by *S. lividus* of the accumulated glycogen (4).

*Cfx. aurantiacus* was shown to use, in a light-dependent manner, both the glycolate and the acetate produced by *S. lividus* (4, 9). The food chain was postulated to be completed mainly by methanogens; further $^{14}$C-radiolabeling studies showed that the main precursor to methanogenesis in OS was CO$_2$, probably because *Cfx. aurantiacus* was consuming most of the acetate (67). The first microelectrode studies of the mat also were performed during this period, and revealed the daily extreme variations in oxygen partial pressure, pH and light intensity that the members of this community experience (63).

In the 1990s 16S rRNA analyses of this microbial mat revealed numerous uncultured and previously unknown microorganisms, including 5 different types of cyanobacteria and 2 different types of FAPs (29, 89, 66). Interestingly, these molecular analyses also revealed that *S. lividus* was, at best, a minor component of the mat, and that culturing conditions, probably the pH, had selected for this organism over all of the other cyanobacteria in the mat. The uncultured groups of cyanobacteria had a curved rod-like morphology similar to that of thermophilic
Synechococcus and were termed Synechococcus types A, A’, A”, B, B’, respectively.

Through DGGE analyses of 16S amplicons, it was shown that each of these cyanobacterial ecotypes resides in a narrow environmental niche in the microbial mat defined by light and temperature. Synechococcus type A, for example, was found predominantly at higher temperatures, whereas types B and B’ were found to dominate at increasingly lower temperatures (30, 61, 66, 90, 91). By using molecular methods to monitor enrichments, dominant cyanobacterial strains in the mat now have been cultured and their physiologies studied in the laboratory (1, 47, 90). Moreover, their genomes have been sequenced, which have revealed further surprises such as the ability of these diverse Synechococcus spp. to fix nitrogen (11).

The two FAP components of the mat were also shown to consist of various ecotypes each that were equally adapted to specific temperature and light gradients. Although one of these FAPs was closely related to Cfx. aurantiacus (66, 91), the FAP species dominating the mat seemed to be not Cfx. aurantiacus-like but belonging to a more divergent, uncultured organism initially termed “organism C” (54). In 2000, FAPs that lack BChl c, also called red Chloroflexi, were isolated from other hot spring microbial mats and molecular analyses assigned them to a new genus within the Chloroflexi, Roseiflexus (13, 14, 39). In OS, organism “C” and its ecotypes were shown to be closely related to members of this new genus, in particular to Roseiflexus castenholzii (54). The major Roseiflexus strains from OS,
strains RS-1 and RS-2, have now been cultured (85), and the genome of Roseiflexus sp. RS-1 has been sequenced (48, 85).

Lipid isotopic fractionation studies have revealed that both Chloroflexus and Roseiflexus spp. in the mat have a heavier isotopic signature than would be expected by their previously predicted photoheterotrophic metabolism. Instead, the $^{13}\text{C}/^{12}\text{C}$ ratio in their lipids is in agreement with what would be expected if they were fixing carbon through the 3-hydroxypropionate pathway, as has been described for Cfx. aurantiacus (82-85). The genomes of these major FAP species from the mat confirm that both the Chloroflexus strains as well as the Roseiflexus strains contain genes coding for that entire 3-hydroxypropionate pathway, even though, to date, no Roseiflexus strain has been grown autotrophically (48). These genomic and stable isotope data in combination suggest that FAPs in OS exhibit multiple growth physiologies, shifting from heterotrophy to photoheterotroph/photomixotroph to photoautotrophy, while adapting to diurnal shifts in environmental conditions in the mat (84).

1.4. THESIS GOALS AND SCOPE

In 2003 the National Science Foundation awarded three Frontiers In Biological Research (FIBR) grants, one of them to a team led by David Ward of Montana State University. Ward and his team proposed to apply the then emerging tool of metagenomics to another habitat: chlorophototrophic microbial mats of hot springs in Yellowstone National Park, in particular the mats found in Octopus and Mushroom Springs. Analyses of this metagenome revealed a much more diverse
microbial community than anticipated even by 16S gene sequencing studies and predicted the presence of previously unknown microorganisms with novel physiologies.

The goals of the present work are to isolate in a laboratory culture one of those previously unknown members of the mat, Candidatus Chloracidobacterium thermophilum (hereafter Cab. thermophilum). This microorganism had escaped detection through decades of studies despite being the 4th most abundant member of the Octopus/Mushroom Springs communities and is predicted by the metagenomic analyses to be an aerobic chlorosome-containing chlorophototroph. In addition to cultivating Cab. thermophilum, this work intends to test these predictions made about the physiological capabilities of this interesting microorganism (see section 2.1 for a complete introduction to these predictions) and further investigate its physiology to better understand its role in the Octopus/Mushroom mat ecosystems.

Because of the elegant combination of metagenome-driven discovery of new microbial diversity with isolation in culture, in the tradition of the Delft school of Microbiology, the discovery and subsequent isolation of Cab. thermophilum presented in this work has become another landmark example of the power of metagenomic analyses (62).
1.5. REFERENCES:


molecular monitoring of enrichment cultures. Antonie Van Leeuwenhoek 71:143-150.


FIGURE LEGENDS

Figure 1.1. Bacterial phylogenetic tree from 1987. 16S phylogenetic tree showing the 11 Bacterial phyla known at the time. From Woese C. R. 1987. Bacterial evolution. Microbiol Rev 51:221-271

Figure 1.2. Bacterial phylogenetic tree from 2003. 16S gene phylogenetic tree showing in black wedges, the 11 original phyla described by Woese (see ref. for figure 1); in white wedges, 14 additional phyla with cultured representatives that have been described since 1987; and in gray wedges, 26 additional phyla that have no cultured representatives. From The uncultured microbial majority. Michael S. Rappé, Stephen J. Giovannoni. 2003. Ann Rev of Microbiol 57: 369–394.

Figure 1.3. Bacteriorhodopsin diagram. Ribbon diagram showing the light absorbing retinal (purple) from bacteriorhodopsin and the path of the hydrogen ion through residues in the transmembrane peptide. From Alberts, B. 2005. Molecular Biology of the Cell. Garland Science, London, UK.

Figure 1.4. Reaction center diagram. A. Ribbon diagram of the type II reaction center of Rhodobacter sphaeroides showing the two transmembrane peptides L and M and the additional peptide H. B. The cofactors that form this reaction center are shown in their respective locations with respect to the polypeptide ribbon diagram shown in A. The path of the electron is shown by a black arrow. The special pair of chlorophylls are shown in yellow (P\textsubscript{A} and P\textsubscript{B}) and the terminal electron acceptors are shown in cyan (Q\textsubscript{A} and Q\textsubscript{B}). From Jones, M. R. Biochemical Society Transactions. 2009. 37:400-407.

Figure 1.5. Diversity of chlorophototrophs. Diagram of the five known phyla of phototrophs and their properties such as the type of reaction center and antenna system they possess and their oxygen tolerance. The chlorophototrophs in the Proteobacteria and Chloroflexi clades can grow chemoheterotrophically in the presence of oxygen, but are phototrophs only under anaerobic conditions.

Figure 1.6. Octopus Spring and its microbial community. A. Photograph of Octopus Spring in Yellowstone National Park. B. Octopus Spring phototrophic mat. C. Phase contrast micrograph of the green layer from the phototrophic mat showing long filamentous microorganisms (from the Chloroflexi clade) and curved rods (Cyanobacteria from the Synechococcus genus). D. Fluorescence micrographs of the top mm of the green layer of the phototrophic mat showing autofluorescence by the cyanobacterial members of the mat. From Ward, D. M., M. J. Ferris, S. C. Nold, and M. M. Bateson. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 62:1353-1370.
Table 1.1. Physical and chemical properties of Octopus Spring*

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (°C)</td>
<td>91.2</td>
</tr>
<tr>
<td>pH</td>
<td>8.3</td>
</tr>
<tr>
<td>Cl (mg/L)</td>
<td>256</td>
</tr>
<tr>
<td>SO₄ (mg/L)</td>
<td>23</td>
</tr>
<tr>
<td>Na (mg/L)</td>
<td>321</td>
</tr>
<tr>
<td>K (mg/L)</td>
<td>16.6</td>
</tr>
<tr>
<td>Ca (mg/L)</td>
<td>0.8</td>
</tr>
<tr>
<td>Mg (mg/L)</td>
<td>0.0</td>
</tr>
<tr>
<td>Fe (mg/L)</td>
<td>0.01</td>
</tr>
<tr>
<td>Mn (mg/L)</td>
<td>0.006</td>
</tr>
<tr>
<td>Cr (mg/L)</td>
<td>0.010</td>
</tr>
<tr>
<td>Hg (µg/L)</td>
<td>0.15</td>
</tr>
<tr>
<td>As (mg/L)</td>
<td>1.28</td>
</tr>
<tr>
<td>Cu (mg/L)</td>
<td>0.002</td>
</tr>
<tr>
<td>Zn (mg/L)</td>
<td>0.003</td>
</tr>
<tr>
<td>H₂S (mg/L)</td>
<td>0.2</td>
</tr>
<tr>
<td>SiO₂ (mg/L)</td>
<td>254</td>
</tr>
<tr>
<td>HBO₃ (mg/L)</td>
<td>10.3</td>
</tr>
<tr>
<td>NH₃ (mg/L)</td>
<td>0.00</td>
</tr>
<tr>
<td>NO₂⁻ (mg/L)</td>
<td>0.01</td>
</tr>
<tr>
<td>NO₃⁻ (mg/L)</td>
<td>0.28</td>
</tr>
<tr>
<td>PO₄ (mg/L)</td>
<td>1.70</td>
</tr>
</tbody>
</table>

Figure 1.1
Figure 1.2
Figure 1.3
Figure 1.4
Figure 1.5

<table>
<thead>
<tr>
<th></th>
<th>Reaction centers</th>
<th>Antennae</th>
<th>Oxygen tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cyanobacteria</strong></td>
<td>RC2, RC1</td>
<td>Phycobilisomes</td>
<td>++</td>
</tr>
<tr>
<td><strong>Chlorobi</strong></td>
<td>RC1</td>
<td>Chlorosomes, FMO</td>
<td>--</td>
</tr>
<tr>
<td><strong>Proteobacteria</strong></td>
<td>RC2, RC1</td>
<td>LHC 1,2</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Chloroflexi</strong></td>
<td>RC2</td>
<td>LHC 1, Chlorosomes</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Firmicutes</strong></td>
<td>RC1</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Figure 1.6
CHAPTER 2

Isolation and Initial Characterization of *Candidatus* Chloracidobacterium thermophilum

**Publication:**
ABSTRACT
A metagenomic study conducted on the phototrophic mats of Octopus and Mushroom Springs in Yellowstone National Park predicted the presence in these mats of a chlorophototrophic microorganism that belongs to the bacterial phylum Acidobacteria. Genes associated with this microorganism included pscA, csmA, bchK and acsF which code for the apoprotein of a type-1 reaction center, the chlorosome baseplate protein, bacteriochlorophyll (BChl) c synthase and the oxygen-dependent magnesium-protoporphyrin IX monomethyl ester oxidative cyclase, respectively. The presence of these genes suggested that this organism is an aerobe that synthesizes a type-1 reaction center and uses chlorosomes for light-harvesting. Gene amplification and fluorescence microscopy have been used to detect this microorganism, Candidatus Chloracidobacterium (Cab.) thermophilum, in a cyanobacterial enrichment derived from the Octopus Spring chlorophototrophic mats. By using the herbicide atrazine as a photosystem II inhibitor and providing reduced carbon and nitrogen sources, the cyanobacterial component was eliminated from this enrichment and an actively growing culture of Cab. thermophilum was generated, albeit with heterotrophic contaminants. Like other Acidobacteria, Cab. thermophilum is difficult to culture and has long generation times. Our initial physiological, biochemical and microscopic analyses confirm that Cab. thermophilum is an aerobic chlorophototroph that synthesizes chlorosomes as antenna pigments.
2.1 INTRODUCTION

Type-1 reaction centers are found in three phototrophic clades *Heliobacteriales, Chlorobiales* (green sulfur bacteria) and *Cyanobacteria* (5,6,12). In the *Heliobacteriales* and *Chlorobiales*, they consist of a transmembrane homodimeric complex (PshA and PscA, respectively) that contains, among other prosthetic groups, the special pair of chlorophylls (Chls) and a [4Fe-4S] cluster, F_X, that acts as secondary or tertiary electron acceptor, and of an additional peptide (PshB and PscB, respectively) that harbors two more [4Fe-4S] clusters that act as the terminal electron acceptors during light induced excitation (15, 18, 27). Green sulfur bacteria (*Chlorobiales*) have two additional distinct polypeptides, PscC and PscD, auxiliary to the core reaction center. In these organisms, two PscC subunits function as electron donors to P840^+ (25), whereas the role of the PscD is uncertain at this time (15,18). In *Cyanobacteria* photosystem I is the heterodimeric type-1 reaction center, which comprises the PsaA and PsaB peptides as the core of the reaction center in addition to PsaC, housing the terminal iron sulfur clusters, as well as 9 other different accessory subunits with diverse roles (13).

To determine whether there were additional chlorophototrophs that use a type-1 reaction center, the PscA protein sequence from the green sulfur bacterium (GSB) *Chlorobaculum tepidum* was used as query of a metagenomic library constructed from DNA extracted from the well-characterized phototrophic mats of Octopus and Mushroom Springs in Yellowstone National Park (YNP) (4, 30, 31). This search retrieved two novel sequences that were compared phylogenetically to the sequences of other type-1 reaction center apoproteins (7). The first of these two novel sequences, termed Octopus Spring Green Sulfur Bacterium (OSGSB),
positioned at the base of the *Chlorobiales* clade and was most closely related to the PscA sequence of *Chloroherpeton thalassium*, a basal GSB. The second sequence, designated *Candidatus* Chloracidobacterium thermophilum (Cab. thermophilum), did not belong to any of the three known clades with type-1 reaction centers (7).

The Octopus/Mushroom Spring metagenomic library was constructed using plasmid vectors (4) and the plasmids that contained sequences encoding *pscA* of Cab. thermophilum were retrieved, and the genes surrounding *pscA* were sequenced. The genes coding for a protein homologous to PscB, the second subunit of type-1 reaction centers, and for the Fenna-Matthews-Olson (FMO) protein, the Bacteriochlorophyll (BChl) *a*-binding protein present in GSB (5, 6, 12), were found immediately downstream of *pscA* forming an apparent operon. Other genes in the plasmids, which were overlapping but non-identical, did not code for products involved in phototrophy. However, BLASTP searches using these genes as queries showed most of them to be most closely related to an acidobacterium, *Solibacter usitatus*, whose genome was the first from that phylum to be completely sequenced (32).

In addition to the metagenomic plasmid library from the Octopus and Mushroom Springs, a BAC clone library, constructed from DNA extracted from the phototrophic mat communities of these springs, was available (4). Assembled metagenomic contigs were mapped to these BAC clones and two contigs mapped to BAC clone M60-018 J19. One of these contigs contained the *pscA-pscB-fmoA* operon while the other one had the phylogenetic markers *recA* and 16S rRNA genes. Phylogenetic analyses of these two markers confirmed that the organism
from which these genes are derived belonged to the bacterial clade *Acidobacteria* (7), a kingdom/phylum that was not previously known to contain phototrophs (10).

In order to learn more about this putative novel phototroph, the contigs in the metagenomic library were binned according to G+C content and gene phylogeny. The contigs that were found to be associated with this microorganism contained several additional genes associated with phototrophy. For example, the BChl biosynthesis genes *bchY* and *bchK*, coding for a subunit of chlorophyllide *a* reductase and BChl *c* synthase, respectively, were found in these contigs as well as the *csmA* gene that codes for the BChl *a*-binding protein located in the baseplate of chlorosomes, antenna structures found in members of the *Chlorobiales* and the *Chloroflexi* (6, 12).

Other genes from these contigs hinted at the possibility that this organism was an aerobe. For example, the presence of *acsF*, which codes for the oxygen-dependent magnesium-protoporphyrin IX monomethyl ester oxidative cyclase, suggested that *Cab. thermophilum* synthesizes BChl aerobically; the detection of cytochrome oxidases hinted at the possibility of aerobic respiration.

In combination, these metagenomic analyses predicted the presence in the Octopus/Mushroom Springs chlorophototrophic mats of a chlorophototroph that synthesizes chlorosomes, FMO and homodimeric type-1 reaction centers, that belongs to the *Acidobacteria* and that it is most likely an aerobe (7). These predicted features represent a novel type of physiology since none of the microorganisms presently known that synthesize either chlorosomes, FMO, homodimeric type-1 reaction centers or a combination of these three are capable of
doing so in the presence of oxygen. Furthermore, there are no known chlorophototrophs in the phylum *Acidobacteria*.

The microbial mats of Octopus and Mushroom Springs have been well characterized through decades of research and there exist cultures derived of the dominant microorganisms in these ecosystems (1, chapter 1). This chapter describes the detection and isolation of *Cab. thermophilum* from one such culture, a three-year old cyanobacterial enrichment derived from the Octopus Spring (OS) phototrophic mats (1). It also presents the results from an initial characterization of this microorganism that confirmed the metagenomic findings that this microorganism is an aerobic chlorophototroph that synthesizes chlorosomes.

### 2.2 METHODS

**2.2.1 Culturing techniques.** Enrichments were grown in B-HEPES medium (28, http://www-cyanosite.bio.purdue.edu/media/Table/BG11.html) containing the following (g/L): NaNO₃ (1.5), MgSO₄·7H₂O (0.075), CaCl₂·2H₂O (0.0359), citric acid (0.0065), ethylenediaminetetraacetic acid tetrasodium salt (Na₄EDTA) (0.001), K₂HPO₄ (0.0459), Fe(NH₄)citrate (0.024), Na₂CO₃ (0.020), and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (1.1). Trace elements (explained below) were added and the pH was adjusted to 8.0 with 2M KOH. *Cab. thermophilum* cultures were cultured in Cab media which consists of B-HEPES medium described above supplemented with 1 mM ammonium chloride and a mixture of carbon sources added in the form of sodium salts (1 mM each): lactate, succinate, glycolate, butyrate, propionate, and acetate. Filtered sterilized vitamin
B₁₂ (100 μg/L) and a mixture of vitamins (explained below) were added after autoclaving. The final pH of this Cab medium was 8.0. The mixture of trace elements contained the following (final concentration mg/L): H₃BO₃ (2.86), MnCl₄·4H₂O (1.81), ZnSO₄·7H₂O (0.222), Na₂MoO₄·2H₂O (0.39), CuSO₄·5H₂O (0.079), Co(NO₃)₂· 6 H₂O (0.0494). The mixture of vitamins consisted of 10 μg /L of riboflavin and 100 μg/L of the following: thiamine· HCl, L-Ascorbic acid, D-Ca-pantothenate, folic acid, nicotinic acid, 4-aminobenzoic acid, pyridoxine·HCL, and lipoic acid. To eliminate the cyanobacteria in the original enrichment the modified B-HEPES medium initially also contained atrazine, 10 μM final concentration. To inhibit the growth of the heterotrophic contaminants, the modified B-HEPES medium was at times supplemented with bacitracin (1 μg/mL).

2.2.2 Physiological studies. To determine whether the growth of Cab. thermophilum was dependent on light, enrichments of Cab. thermophilum were subcultured in parallel under light and dark treatments for at least three serial transfers of one week incubation each. These cultures were monitored spectroscopically through the absorption maximum of the predicted main pigment, BChl c, at 745 nm, and genetically by PCR amplification of genes specific to Cab. thermophilum (Table 2.1).

For other physiological studies the Cab. thermophilum culture was routinely assessed microscopically and through BChl c absorbance as above. Absorbance spectra were taken in a GENESYS 10 spectrophotometer (Thermo Electron Corp., Rochester, NY).
2.2.3 Fluorescence In Situ Hybridization (FISH). For the FISH experiments, cells from the enrichment were fixed in 4% paraformaldehyde and spotted on Teflon-printed slides (Electron Microscope Scientific, Hatfield, PA, USA). Smears were dehydrated by immersion into a series of 50%, 80% and 90% (v/v) ethanol for 3 minutes each. Cells were then ready for fluorescence in situ hybridization as described by Pernthaler et al. (26). Oligonucleotide probes HoAc-1402-a-A-18 (20, 23), specific for the phylum Acidobacteria, and Eub 334 (2), as a control for all Bacteria, were used during hybridization. Both probes were purchased from Sigma-Genosys (St. Louis, MO, USA). Cells were visualized in an Olympus BX-60 epifluorescent upright microscope equipped with an UplandFL 100× oil objective, a mercury vapor lamp and Olympus filter cubes 41001 and 41002 (Olympus, Center Valley, PA, USA).

2.2.4 Hexanol-treatment of Candidatus Chloracidobacterium. thermophilum cells. Cab. thermophilum cells were incubated in hexanol-saturated Tris-HCl buffer and visualized in an Olympus BX-60 epifluorescence upright microscope equipped with an UplandFL 100× oil objective, a mercury vapor lamp and Olympus filter cubes 41001 (Olympus, Center Valley, PA).

2.2.5 Thin-section electron microscopy. For electron microscopy of thin-sectioned cells, harvested Cab. thermophilum cells were fixed with an osmium tetraoxide-potassium permanganate tandem fixation as described by Hohmann-Marriott et al (17). Following fixation, cells were stained en block with 1% (w/v) uracil acetate and dehydrated with 25%, 50%, 70%, 80%, 90%, 95% and 100% (v/v) ethanol and with acetonitrile. Dehydrated cells were embedded in Spurs resin
according to standard protocols and allowed to polymerize in a 60 °C oven overnight. Images were visualized in a JEOL 1200 EXII Transmission electron microscope (Peabody, MA).

2.3 RESULTS

2.3.1 Detection of *Candidatus* Chloracidobacterium thermophilum in the Cyanobacterial enrichment. PCR amplification of DNA extracted from the OS cyanobacterial enrichment with specific primers for *recA*, *16S rRNA*, *pscA* and *csmA* genes of Cab. thermophilum showed strong single amplicon bands as products. These bands were sequenced and compared to the respective metagenome genes to verify their identity. Comparative analyses showed that these gene products from the enrichment, although belonging to Cab. thermophilum, were derived from a different population than the one found in the metagenome (metagenome sequence hereafter designated Cab. thermophilum type M). The 16S rRNA genes, for example, differ by 3% in a total of 1450 nucleotides (figure 2.1) and the amino acid sequences of PscA between these two populations differ by 1.6% of the total residues (figure 2.2).

FISH analyses of the cyanobacterial enrichment culture revealed the presence of at least three different morphological and phylogenetic types of microorganisms (figure 2.3), one of them corresponding to an acidobacterium. In figure 2.3, the large, slightly curved rods represent *Synechococcus* spp. cells, the dominant organism in the OS phototrophic mats (30, 31). These cyanobacterial cells are autofluorescent and in the fluorescence micrograph appear orange due to the
combination of this autofluorescence and to hybridization to the universal bacterial probe, Eub 334. In addition to these curved rods, long, thin filamentous cells stained green with these universal probe. I was able to isolate these filaments in pure culture (figure 2.7) and by PCR amplification and sequencing of its 16S rRNA gene have verified that they belong to the genus *Meiothermus*, a thermophilic group of chemotrophs found in hot springs worldwide (8, 22, 29). The third type of morphology consisted of short rods, and these were stained both by the universal bacterial probe Eub 334 and, more importantly, by the acidobacterial specific probe HoAc-1402-a-A-18 and appear yellow in the fluorescence micrograph.

### 2.3.2 Isolation of *Candidatus Chloracidobacterium thermophilum.*

The addition of atrazine in combination with supplements of ammonia and a mixture of fermentation end-products to the cyanobacterial enrichment resulted in the complete elimination of *Synechococcus* spp. from the enrichment after ca. three serial transfers as evidenced visually (figure 2.4), spectroscopically (figure 2.5), and microscopically (figure 2.6). This new culture consisted of Cab. thermophilum as the only chlorophototroph and, for the most part, the most numerous member of the culture. Phase-contrast micrographs of the Cab. thermophilum cells confirmed the morphology predicted by the FISH analyses. Cab. thermophilum cells appear as rods about 2 µm in length and 0.75 µm in diameter (figure 2.6).

Isolation of Cab. thermophilum was strictly dependent on two factors: the addition of ammonia as a nitrogen source to the media and the addition of the mixture of reduced carbon sources, as outlined in the methods section. Other carbon
and nitrogen sources, such as glucose or casein amino acids, did not support growth of *Cab. thermophilum* in the presence of the heterotroph contaminants.

By plating this new enrichment in solid LB and *Cab* media heterotrophic contaminants also present in this culture have been isolated (figure 2.7). An organism that forms bright yellow colonies overnight when grown on LB was identified as an *Anoxybacillus* sp. through amplification and sequencing of its 16S rDNA gene. A second isolate forms dark pink colonies after three days of incubation in *Cab* media. This organism, also detected by the FISH studies described above, has a filamentous morphology; 16S rRNA amplification and sequencing confirmed that it belongs to the genus *Meiothermus*.

2.3.3 Physiological properties of *Cab. thermophilum* culture and further purification attempts. Under laboratory growth conditions *Cab. thermophilum* cells have a doubling time of ca. 20-30 hours and only undergo 4-5 cell divisions before entering stationary phase as evidenced by monitoring BCHl c absorbance (figure 2.8). Total absorbance by this pigment synthesized by *Cab. thermophilum* only doubles 3-4 times during the course of 5 days before reaching a plateau. Interestingly, *Cab. thermophilum* growth does not seem to occur during the initial 24 hours that follow inoculation into fresh medium as evidenced by the absence of actively dividing *Cab. thermophilum* cells during microscopic examination and by the lack of increased BCHl c absorbance (figure 2.8). The *Anoxybacillus* sp. present in the culture rapidly divide during this initial period, which causes cell densities of this organism in the culture to increase rapidly (figure 2.8). Following these first 24 hours *Anoxybacillus* sp. cells begin to form spores and this coincides
with an increase of the pH of the media and growth of Cab. thermophilum. This phenomenon can readily be observed as the culture changes color from white (Anoxybacillus sp.) to orange (Cab. thermophilum) (figure 2.8).

Efforts to purify Cab. thermophilum further from the other components of the enrichment have consistently failed. Cab. thermophilum does not grow on solid media, neither on plates or agar deeps, under the conditions tested. This is most probably due to the small number of cell divisions the cells undergo before reaching stationary phase.

In dilution to extinction experiments, the heterotrophic contaminants of the culture rapidly multiplied and overtook the enrichment. This might be due to the higher growth rate of the contaminants, with reports of a maximal growth rate of 1.87-1.84 h\(^{-1}\) for some Anoxybacillus species (9). Alternatively, Cab. thermophilum cells might require a large inoculum size in order to grow which would prevent growth under dilute conditions. Physiological studies conducted by Elizabeth Harvey as part of her undergraduate honors thesis research support this latter scenario (14). Ms. Harvey found that for optimum Cab. thermophilum growth, cultures should be diluted 1 to 5, which represents a BChl c relative absorbance of ca 0.030 at the Qy maxima.

Differential centrifugation separates the components of the culture as seen in figures 2.9 and 2.10 to ~95% homogeneity. However, using Cab. thermophilum cells isolated this way as inocula for new cultures did not result in growth. Similarly, attempts to eliminate contaminants with the antibiotic bacitracin also failed to produce an actively growing, axenic Cab. thermophilum culture. These
combined results seem to indicate the presence of a growth factor that the contaminants might be providing. However, addition of filter sterilized spent media of the Cab. thermophilum culture collected at different time points has also failed to stimulate growth.

2.3.4 Candidatus Chloracidobacterium thermophilum is a phototroph  To test whether Cab. thermophilum is an obligate phototroph, Cab. thermophilum enrichments were serially subcultured with parallel light and dark incubations. Although the cultures incubated in the presence of light exhibited BChl c absorbance following two incubation periods, no BChl c absorbance was observed in cultures subcultured in the dark (figure 2.11), indicating that BChl c was not being synthesized in the dark. PCR amplification with Cab. thermophilum specific 16S rRNA primers using DNA extracted from these serial light and dark treatments as template produced amplicons from cultures grown in the light but not in cells after the second serial subculture in the dark (figure 2.12). These data indicated that Cab. thermophilum DNA was below the detection limit when the enrichment culture was repeatedly subcultured and incubated in the dark. The same template DNAs, however, produced amplicons when Anoxybacillus-specific primers for the 16S rRNA were used (figure 2.12). These results indicate that, as predicted, growth of Anoxybacillus sp. appears to be independent of light conditions but growth of Cab. thermophilum is strongly stimulated by light and appears to be negligible in the dark under the chosen conditions, as evidenced by the absence of BChl c and detectable Cab. thermophilum DNA in the enrichment cultures produced after serial subculturing in the dark.
2.3.5 Absorption spectra of hexanol-treated and untreated *Candidatus Chloracidobacterium thermophilum* cells. BChl c molecules are normally aggregated in the chlorosomes and exhibit an absorbance maximum at ca. 745 nm in this aggregated form. 1-Hexanol treatment disrupts these aggregates and blue-shifts the absorbance to ca. 675 nm, characteristic of monomeric BChl c (see Discussion). The absorption spectra of Cab. thermophilum cells before and after incubation in Tris-HCl buffer saturated with 1-hexanol shows the expected blue-shift of the absorption spectrum. Prior to hexanol treatment, the absorbance maximum of Cab. thermophilum cells was 743 nm at the Qy peak, signifying aggregated BChl c. In cells treated with 1-hexanol, the Qy absorption maximum 675 nm (figure 2.13), indicative of monomeric BChl c. Moreover, following incubation in a 1-hexanol-saturated Tris-HCL buffer Cab. thermophilum cells autofluoresce (figure 2.14).

2.3.6 *Candidatus Chloracidobacterium thermophilum* synthesizes chlorosomes and produces a slime layer. Thin-section electron micrographs of Cab. thermophilum cells revealed electron-opaque, chlorosome-like structures attached to the inner surface of the cytoplasmic membrane (Figure 2.15). These structures were, on average, about 99.6 ± 19 nm long and 30.8 ± 2.0 nm wide. Interestingly, these structures did not appear at the poles of cells in any of our preparations. These micrographs also reveal a slime layer surrounding the Cab. thermophilum cell.
2.4 DISCUSSION

This chapter describes the isolation and initial characterization of Cab. thermophilum, an aerobic chlorophototroph that synthesizes chlorosomes, that is predicted to synthesize the FMO protein and type 1 reaction centers, and that belongs to the phylum Acidobacteria. The existence of this microorganism and its unusual physiology and novel phylogeny were first identified by bioinformatics analyses of the metagenomes of the OS/MS communities (7).

Cab. thermophilum was first isolated in culture as a minor component of a Synechococcus sp. enrichment from the phototrophic mat of OS. It was able to survive undetected in this oxic enrichment culture for more than three years until analytical PCR (results presented here) revealed the presence of a microorganism in the cyanobacterial enrichment culture, which had similar properties to the ones predicted by the metagenomic analyses. The presence of this microorganism in an oxic enrichment was surprising, because no other organism known at that time was able to grow or synthesize BChl c, chlorosomes, FMO protein or homodimeric reaction centers in the presence of oxygen.

Alignments of the amplified 16S rRNA gene and of the PscA amino acid sequences further revealed that this organism, although similar enough to the one described in the metagenome to be considered the same species, nevertheless represents a distinctive subpopulation. In this respect, Cab. thermophilum resembles other members of the OS/MS communities which have undergone adaptive radiation to fill temperature and light intensity niches and thus consist of several subpopulations (30, 31).
Because Cab. thermophilum is not predicted to synthesize type-2 reaction centers, the herbicide atrazine, which is a competitive inhibitor of quinone binding to the Q_B site of photosystem II (19), was used to inhibit the growth of the *Synechococcus* spp. present original cyanobacterial enrichment 1). Without *Synechococcus* spp. present in the culture, Cab. thermophilum required the addition of reduced carbon and nitrogen in order to grow in the laboratory. These growth requirements establish that Cab. thermophilum is a heterotroph, which is dependent on other community members for essential growth factors.

Cab. thermophilum resembles other *Acidobacteria* in its long generation times and requirement for stringent culture conditions (10, 32). Cell densities of Cab. thermophilum in culture are extremely low, which might explain its inability to form detectable colonies on solid media. Moreover, its growth seems to be dependent on the presence of heterotrophic contaminants. These low cell densities and dependency on other microorganisms seem to indicate there might be a limiting nutrient or condition. Efforts to identify what this limiting nutrient might be are further described in chapter 3 and Appendix A. Despite these efforts, it is reasonable to believe that Cab. thermophilum might not be able to reach high cell densities when removed from its natural habitat. For example, microorganisms belonging to the *Acidobacteria* and *Verrucomicrobia* isolated from soil samples formed microscopic colonies, only detectable with a stereomicroscope, after three months of incubation under CO_2 partial pressures designed to mimic the ones found in small soil micro-niches (see reference 27, chapter 3).
The growth of Cab. thermophilum in the dark was non-detectable, and cells of Cab. thermophilum were rapidly diluted out in serial subcultures that were incubated under dark incubation, whereas growth was consistent under light treatments. The light dependency of growth, in combination with the requirement for reduced carbon compounds described above, establish that Cab. thermophilum is a photoheterotroph. In this form of metabolism, the charge separation generated as a result of light absorption is used to generate energy in the form of ATP, while the reduced carbon assimilated is channeled into anaplerotic reactions to build biomass and to generate reductant for growth. Other known photoheterotrophs in nature are the *Heliobacteria*, which are obligate photoheterotrophs, many purple bacteria including aerobic anoxygenic phototrophs, and most phototrophic members of the *Chloroflexi*. Organisms belonging to these latter groups have the ability for both autotrophy and (photo)heterotrophy (21, 24). In some photoheterotrophs, the reducing power generated by the light-induced charge separation is used to fix nitrogen (16), among other activities, while other photoheterotrophs only use it for anabolic reactions (24). Cab. thermophilum lacks the genes for nitrogen fixation (see chapter 3), and, at this point we speculate that reducing equivalents are returned to the reaction center in cyclic electron transport generating a proton motive force for ATP production.

The presence of genes homologous to *bchK* and *csmA* associated with Cab. thermophilum (7) suggested that Cab. thermophilum synthesizes BChl c and chlorosomes, vesicle-like antenna structures that typically contain hundreds of thousands of BChl c molecules (12). This prediction was confirmed by electron
microscopy. Thin-section electron micrographs showed chlorosomes attached to the lateral cytoplasmic membrane surfaces of Cab. thermophilum cells.

The absorbance properties of Cab. thermophilum cells further confirmed the presence of chlorosomes in this organism. The combined in vivo absorbance spectra of Cab. thermophilum cells in Tris-HCl buffer and in the presence of 1-hexanol exhibited the predicted changes expected for aggregated BChl c molecules in chlorosomes. BChl c is special because it can self-aggregate to form large suprastructures inside chlorosomes. This aggregation is partly due to the interaction between the central magnesium of one BChl c molecule and the C-3\(^1\)-hydroxyl group of a nearby BChl c molecule. In addition BChl c lacks the C-13\(^2\)-methylcarboxyl group that other (B)Chl molecules have and that would prevent the BChl c molecules from coming into close proximity with each other due to steric hindrance. When aggregated into supramolecular structures, BChl c exhibits an absorbance maximum at 745-750 nm. The observed absorption maximum for Cab. thermophilum cells at ca.745 nm falls within this range.

1-Hexanol interferes with the liganding interactions between adjacent BChl c molecules without disturbing the integrity of the chlorosome envelope (3). Under these circumstances, the absorption spectrum observed is similar to that of monomeric BChl c rather than BChl c supramolecular structures, and the absorption maximum is blue-shifted to 670 nm. Cab. thermophilum cells that had been incubated in 1-hexanol-saturated Tris-HCl buffer exhibited this shift in the absorption maximum, further supporting the presence of BChl c aggregated in chlorosomes in this organism. Although chlorosomes have been found in two other
clades of phototrophs (6, 12), this is the first known instance of an aerobic microorganism that synthesizes these antenna structures.

In summary, an enrichment culture of Cab. thermophilum has been generated, and initial physiological studies of this culture confirm that this microorganism is an aerobic chlorophotoheterotroph that synthesizes chlorosomes. It is also predicted to synthesize a type-1 homodimeric reaction center and the FMO protein. Cab. thermophilum is the first phototrophic member of the phylum Acidobacteria, and its discovery expands the total number of known bacterial phyla with chlorophototrophic members from five to six (7).
2.5. REFERENCES:


18. **Jagannathan, B., and J. H. Golbeck.** 2008. Unifying principles in homodimeric type I photosynthetic reaction centers: properties of PscB and the F\textsubscript{A}, F\textsubscript{B} and F\textsubscript{X} iron-sulfur clusters in green sulfur bacteria. Biochim Biophys Acta 1777:1535-1544.


FIGURE LEGENDS

Figure 2.1. Nucleotide sequence alignment of the 16S rRNA genes of two *Candidatus Chloracidobacterium thermophilum* populations. Cab. thermophilum, the laboratory isolate and Cab. thermophilum type M, the metagenomic sequence.

Figure 2.2. Amino acid alignment of the PscA protein of two *Candidatus Chloracidobacterium thermophilum* populations. Cab. thermophilum, the laboratory isolate and Cab. thermophilum type M, the metagenomic sequence.

Figure 2.3. Fluorescent *in situ* hybridization micrograph of a cyanobacterial enrichment. This enrichment was generated from the phototrophic mat of Octopus Spring that had been stained with phylogenetic specific fluorescent probes. Green, probe Eub338 with a FITC (fluorescein isothiocyanate) chromophore that hybridizes with all Bacteria. Yellow, probe Eub338-FITC plus probe HoAc-1402-a-A-18 with a Cy3 chromophore that stains Acidobacteria. Please note that the *Synechococcus* spp. cells in the bottom right corner area are orange due to autofluorescence in combination with hybridization to probe Eub338-FITC.

Figure 2.4. Images of the cultures used and generated in this study. 1. The initial *Synechococcus* sp. Type B’ enrichment. 2. The Cab. thermophilum enrichment. 3. The Cab. thermophilum enrichment after three serial subcultures in the dark.

Figure 2.5. Absorption spectra of the cultures used and generated in this study. A. Absorption spectra of the cyanobacterial enrichment. Note the absorbance maximum at 745 nm characteristic of BChl c. B. Absorption spectrum of the Cab. thermophilum enrichment. Note the absence of absorbance at 682 nm characteristic of Chl a.

Figure 2.6. Phase contrast micrographs of the cultures used and generated in this study. A. Phase contrast micrograph of the *Synechococcus* sp. Type B’ enrichment. The large dark rods are *Synechococcus* spp., the most abundant organism in this culture. The long thin filaments represent *Meiothermus* sp. Note the small rod at the lower right corner, which is probably Cab. thermophilum. B. Phase contrast micrograph of the Cab. thermophilum culture.

Figure 2.7. Images of colonies formed on solid medium by heterotrophic contaminants in the *Candidatus Chloracidobacterium thermophilum* culture. A. *Anoxybacillus* sp. B. *Meiothermus* sp. Note the heterogeneity in colony pigmentation, size and morphology.

Figure 2.8. Growth cycle in the *Candidatus Chloracidobacterium thermophilum* culture. A. Overall cell density (determined from absorbance at 650 nm) and relative absorbance of BChl c (measured as the height of the Qy peak
of BChl c at 745 nm) during an incubation period. For this experiment cultures were transferred to fresh media on day one. **B.** Image of a Cab. thermophilum enrichment culture during a 5-day incubation period. Photographs were taken after days one, three and five after transfer of an inoculum to fresh media. Note that the pH increases during each incubation period.

**Figure 2.9.** Images of the different organisms found in the *Candidatus Chloracidobacterium thermophilum* culture following density centrifugation. **Top.** Image of the Cab. thermophilum culture immediately following centrifugation on a table-top microcentrifuge. Please note the brown Cab. thermophilum cells on top, red *Meiothermus* sp. cells in the middle and yellow *Anoxybacillus* sp. cells at the bottom. **Bottom.** Image of the same three components of the Cab. thermophilum culture but separated in different eppendorf tubes.

**Figure 2.10.** Phase contrast micrographs of the separated components of the *Candidatus Chloracidobacterium thermophilum* culture. **A.** *Anoxybacillus* sp. **B.** *Meiothermus* sp. **C.** Cab. thermophilum.

**Figure 2.11.** Absorption spectra of *Candidatus Chloracidobacterium thermophilum* cells. Absorption spectra of Cab. thermophilum cells that had been serially subcultured 3 times in the light or 3 times in the dark.

**Figure 2.12.** PCR amplification with primers specific for *Candidatus Chloracidobacterium thermophilum* and *Anoxybacillus* sp. The 16S rRNA gene from both organisms was amplified from DNA extracted from the Cab. thermophilum enrichment culture and serial subcultures incubated under light (L) and dark (D) treatments. The inoculum (I) used for these treatments is also shown as well as a negative control using DNA extracted from *Escherichia coli* cells.

**Figure 2.13.** Absorption spectra of *Candidatus Chloracidobacterium thermophilum* culture. Cells were incubated in Tris- HCl buffer (black line) and in 1-hexanol-saturated Tris- HCl buffer (gray line). Note the shift of the BChl c absorption maxima.

**Figure 2.14.** Fluorescence micrograph of autofluorescent *Candidatus Chloracidobacterium thermophilum* cells. Cells were incubated in Tris- HCl buffer saturated with 1-hexanol for 10 minutes prior to visualization. The 1-hexanol blue-shifts the absorbance of BChl c to ~675 nm and causes cells to become autofluorescent in the visible range.

**Figure 2.15.** Thin-section electron micrograph of a *Candidatus Chloracidobacterium thermophilum* cell. Chlorosomes can be seen as electron-opaque ovals appressed to the lateral inner leaflet of the cell membrane. Thin-sections were stained with 2% (w/v) uranyl acetate prior to visualization.
<table>
<thead>
<tr>
<th>Gene amplified</th>
<th>Description</th>
<th>Forward (5’ → 3’)</th>
<th>Reverse (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cab. thermophilum</em> 16S rRNA</td>
<td>Ribosomal RNA used as phylogenetic marker</td>
<td>GAT CCT GGC TCA GAA TC (1)</td>
<td>GGC TTA CAC AGG ATA CC</td>
</tr>
<tr>
<td><em>Anoxybacillus</em> sp. 16S rRNA</td>
<td>Ribosomal RNA used as phylogenetic marker</td>
<td>CGA AAG TCT GAC GGA GCA AC</td>
<td>TCA TCG TTT ACG GCG TGG AC</td>
</tr>
<tr>
<td>Bacterial 16S rRNA</td>
<td>Ribosomal RNA used as phylogenetic marker</td>
<td>AGA GTT TGA TCA TGG CTC AG (2)</td>
<td>GGT TAC CTT GTT ACG ACT T (2)</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> recA</td>
<td>DNA repair protein used as phylogenetic marker</td>
<td>GCA TGG AAG GCG CTG TAG CC</td>
<td>CAG ACA GCC TGC TTC CTA CC</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> pscA</td>
<td>Type 1 reaction center protein</td>
<td>GCT TTT CGT CCT ATG CCA ACG</td>
<td>AGG GGT CAA ACC AAC CGT AAG G</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> fmoA</td>
<td>BChl a-binding antenna protein</td>
<td>GGC AAG ATT ATG GTC GGA AAG</td>
<td>CCC TCA ATA GGA ACA CTA AAC CCC</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> csmA</td>
<td>BChl a binding protein located in the baseplate of chlorosomes</td>
<td>GTC GTG GAC TCC ATC GTG</td>
<td>TCC TTC GGC TGT GAA ACG</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> bchU</td>
<td>C-20 methyltransferase for BChl c biosynthesis</td>
<td>TTC TCT ACG AGT ATT TGC GTG</td>
<td>CGG TGG ATT TCC TCA TAA AAC G</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> acsF</td>
<td>Oxidative ring cyclase for BChl a and c and Chl a biosynthesis under oxic conditions</td>
<td>CAG GAT TTC TCC CAT CTC CCA C</td>
<td>CCA ATA AAG TCG CGG TGT TTG AT</td>
</tr>
</tbody>
</table>


Figure 2.3
Figure 2.4
Figure 2.5

A

![Absorbance spectrum of Chl a from Synechococcus and Bchl c from Cab. thermophilum at 682 nm and 745 nm wavelengths.]

B

![Absorbance spectrum of Bchl c from Cab. thermophilum at 745 nm wavelength.]

Absorbance

Wavelength (nm)
Figure 2.6

A

B
Figure 2.7
Figure 2.8

A

Absorbance

Overall cell density

Relative absorbance by Bchl c

Days

B

pH 8

Day one

Day three

Day five
Figure 2.9
Figure 2.11
Figure 2.12

L₁ D₁ L₂ D₂ I E. coli

Acidobacterial

Anoxybacillus sp.
Figure 2.13

Absorbance vs. Wavelength (nm)

- 675 nm
- 743 nm
Figure 2.14
Figure 2.15
CHAPTER 3

Sequencing and analysis of the complete genome of Candidatus Chloracidobacterium thermophilum

Publication:
ABSTRACT
As shown by 16S rRNA surveys, members of the poorly characterized bacterial kingdom/phylum Acidobacteria are ubiquitous in soils world-wide. The organisms in this kingdom/phylum have proven to be difficult to culture in the laboratory and, although at least 26 different acidobacterial clades are recognized, only four of them have cultured representatives, all of which are aerobic chemoheterotrophs. A recent study compared the genomes of three cultured members of the Acidobacteria. Analyses of these genomes hinted at possible roles in carbon and nitrogen cycling for these organisms. Our laboratory recently isolated a phototrophic Acidobacterium, Candidatus Chloracidobacterium thermophilum (Cab. thermophilum), expanding our knowledge of the physiological capabilities of the members of this clade. The complete genome of this phototrophic acidobacterium is presented here. The genome is divided into two chromosomes, both of which encode essential genes. The genome includes all of the genes required for chlorophyll-based phototrophy with chlorosomes as antenna structures, bacteriochlorophyll c as the main pigment, and a type-1 reaction center. It lacks key genes in all of the known carbon fixation pathways, as well as genes for assimilatory nitrate and sulfate reduction. In addition, it lacks the biosynthetic pathways for the synthesis of the amino acids valine, isoleucine and leucine. The genome does contain genes involved in the oxidation of reduced carbon compounds and genes that code for both the bc_1 complex and alternative complex III. Our genomic analyses clearly define Cab. thermophilum as an aerobic photoheterotroph that is dependent on other members of the mat community for essential nutrients and strongly suggest that this organism plays a very different ecological role than the ones predicted for the other members of the Acidobacteria. As is the case with secondary chromosomes in other bipartite genomes, the smaller chromosome has undergone more numerous rearrangements and might be less subject to purifying selection than the larger chromosome. This smaller chromosome also contains a large number of transposases that are homologous to transposases found in other members of the Octopus Spring phototrophic mats, from which this organism originates.
3.1. INTRODUCTION

The phylum *Acidobacteria* is a ubiquitous clade of microorganisms that through 16S rDNA environmental surveys have been found in diverse habitats, which include acid mine drainages (27), plant roots (29), forest soils (28), the deep sea (41), chemolithoautotrophic caves (35) and hot springs (2). These 16S rDNA studies have revealed this bacterial group to be numerous, in some cases constituting over 40% of the total sequences in their habitats (29, 42), and phylogenetically diverse, with reports of as many as 11 to 26 subdivisions (3, 55). Little is known, however, about the physiological capabilities of the *Acidobacteria* or what roles these organisms play in the environments that they inhabit. At the moment there are only a few characterized isolates and most of the proposed *Acidobacterial* subdivisions have no cultured representatives.

In general microorganisms belonging to this clade have proven difficult to culture, requiring stringent culture conditions such as low nutrient content, high partial pressures of CO₂, and long incubation times (15, 24). Based on the few culturable isolates, however, the *Acidobacteria* seem to be metabolically diverse. For example, *Acidobacterium capsulatum*, the first *Acidobacterium* characterized and for which the phylum is named, and *Geothrix fermentas* are both capable of iron reduction (9, 27). *Holophaga foetida*, isolated from hydrocarbon-rich anoxic sediment, is a homoacetogenic member of the *Acidobacteria* that degrades aromatic compounds (31). In addition, there has been a report linking acidobacterial sequences to methane metabolism through the use of a novel fluorescence sensor of metabolic active cells in a methane-rich culture (25).
**Candidatus** Chloracidobacterium thermophilum (Cab. thermophilum), a chlorophotoheterotroph recently isolated by our laboratory and which belongs to the Acidobacteria, expands the metabolic capabilities of this group to include chlorophototrophy (4). Like other acidobacteria, Cab. thermophilum is difficult to cultivate in the laboratory, and exhibits a doubling time of ca. 18 hours (Chapter 2). Furthermore, growth of Cab. thermophilum seems to depend on the heterotrophs also present in the enrichment culture, and efforts to grow it axenically have failed thus far.

In the past genomic approaches have proven to be a useful tool in deciphering culture conditions for difficult-to-grow microorganisms. A comparative analysis of the genomes of three members of the SAR11 clade, for example, revealed that these organisms lack the genes required for assimilatory sulfate reduction yet contain transporters and degradation pathways for the sulfur compounds 3-dimethylsulphoniopropionate (DMSP) and methionine (50). Subsequent addition of increasing concentrations of DMSP to the growth media of one of these strains, **Candidatus** Pelagibacter ubique, resulted in increased maximum cell density that was proportional to the DMSP concentration added (50). These findings indicated that sulfur had been a limiting nutrient in this culture and that **Cand.** Pelagibacter ubique had the ability to use DMSP as a sulfur source.

Genomic analyses may also offer clues to the metabolic roles organisms play in their natural habitats. The complete genome sequences of three Acidobacteria have recently been published and found to contain degradation pathways for complex carbohydrates such as cellulose and chitin and genes
involved in nitrate and nitrite reduction (52). It was suggested that the presence of these pathways might indicate that these three organisms could be playing a significant role in global carbon and nitrogen recycling in the habitats from which they were initially isolated.

This chapter reports the sequencing and analysis of the complete genomic sequence of Cab. thermophilum, which was undertaken with the goal, among others, of increasing our understanding of the physiological properties of this microorganism and its role in the phototrophic mat community from which it was isolated.

### 3.2 METHODS

#### 3.2.1 Bacterial strains, culture conditions and DNA isolation.

Cab. thermophilum was grown as previously described (4, Chapter 2). Cells were harvested by centrifugation and genomic DNA was isolated according to the Joint Genome Institute Bacterial DNA isolation CTAB protocol (http://my.jgi.doe.gov/general/).

#### 3.2.2 DNA sequencing and assembly.

Isolated DNA was sequenced in a second generation 454 pyrosequencer (GS-20 FLX; Roche) in the laboratory of Dr. Stephan C. Schuster at The Pennsylvania State University. A total of 426,814 reads were obtained and assembled into 81 contigs, with an average read depth of 24X, using the Newbler assembler. Gap closing was performed by Sanger sequencing of PCR products generated through combinatorial PCR, multiplex PCR and thermal asymmetric interlaced PCR (TAIL PCR) (32). Further assembly and visualization
of these sequences with the original contigs were carried out using the Phred/Phrap/Consed software package. (16, 20). To facilitate gap closure, contigs were aligned with paired-end sequences from environmental BAC clones and small clones that contain Cab. thermophilum DNA generated from a metagenomic library from Octopus Spring samples. These environmental BAC clones provided important scaffolding information that allowed predictions to be made to facilitate the design of the PCR reactions employed in gap closure.

3.2.3 Sequence annotation and analyses. The completely assembled genome was automatically annotated by computer by the Joint Genome Institute (http://www.jgi.doe.gov/). Manual curation of the automated annotations were completed on individual genes as required. Putative frameshifts were manually inspected and corrected, and where necessary, resequencing of PCR amplicons was employed to verify the sequence. The Integrated Microbial Genomes (IMG) website (34) and user-generated Perl-scripts were used for genome analyses and comparisons. Comparison of genome sequences of the lab isolate with contigs assembled from the metagenome of the major mat population (Klatt et al., in preparation) was carried out by using the Artemis Comparison Tool (6). Codon adaptation index (CAI) analyses were done through the EMBOSS software package (44).

3.2.4 Phylogenetic analyses. Phylogenetic trees were constructed using PHYLIP software (18) from protein or 16S rDNA sequence alignments generated using ClustalW or Greengenes NAST (12, 13), respectively and manually modified with BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). 16S rRNA sequences
from members of the *Acidobacteria* were obtained through the RDP site at Michigan State University (10). BchK, ChlG and BchG sequences were obtained from NCBI or IMG. Phylogenetic trees were visualized with TreeView (39).

**3.3 RESULTS AND DISCUSSION**

**3.3.1 General genome features.** The genome of Cab. thermophilum consists of two circular chromosomes (2,683,362 bp, and 1,012,010 bp, respectively) (figure 3.1), has a GC content of 61 mol%, and is predicted to encode 3,071 proteins (table 3.1). We investigated gene distribution between the two chromosomes to verify that the smaller chromosome was not a megaplasmid. Genes from all recognized clusters of orthologous groups (COGs) were found in both chromosomes (figure 3.2). More importantly, genes required for the essential property of chlorophototrophy, such as BChl biosynthesis genes, are encoded by both chromosomes (see below). Other general properties of the genome are listed in table 3.1.

The genome contains a single rRNA operon, which is located on the large chromosome. This finding of a single copy of this operon is consistent with the slow growth rate and K-selection lifestyle of Cab. thermophilum (26, 49). Under laboratory conditions it takes ~5 days to observe a few doublings of the population (Chapter 2). This finding is also consistent with observations for other *Acidobacteria* (15, 24, 28). Only one of the four completely sequenced genomes from the phylum *Acidobacteria* has more than one rRNA operon (table 3.1) and most, if not all, reports on isolates of this group of microorganisms allude to their
slow growth rates and low-density culture populations. (15, 24, 28, 52). As expected from the low number of rRNA operons, the genome also codes for a relatively small number (=44) of tRNAs (30).

3.3.2 Pigment biosynthesis and reaction center. As predicted by previous analyses (4), the genome of Cab. thermophilum contains complete pathways for the biosynthesis of BChl \(a\), BChl \(c\) and Chl \(a\), as well as genes for carotenoid biosynthesis, the synthesis of a type-1 reaction center, and the BChl \(a\)-binding Fenna-Matthews-Olson protein (FMO) (table 3.2). A phylogenetic analysis of the concatenated genes \(bchN/B/L\) and \(bchX/Y/Z\), coding for protochlorophyllide and chlorophyllide reductase, respectively, showed that these genes represent the earliest diverging branch of a clade that included both Cab. thermophilum and the purple bacteria (5). This position is consistent with the phylogenetic relationship of the Acidobacteria and Proteobacteria based on proteins shared among all organisms (8). Among BChl \(c\)-specific genes, \(bchU\) (BChl \(c\) C20 methyltransferase) is phylogenetically closest to the Chloroflexi (4), while the B(Chl) synthases BchK (BChl \(c\)), BchG (BChl \(a\)) and ChlG (Chl \(a\)) each represent an early diverging sequence within the Chlorobiales (figure 3.3). This mixed phylogeny, which includes genes with most similarity to those found in three other groups of chlorophototrophs, provides evidence for the important role of horizontal gene transfer in generating diversity among chlorophototrophs.

The genes encoding enzymes for carotenoid biosynthesis predict that Cab. thermophilum should be capable of synthesizing a variety of xanthophyll carotenoids, including ketocarotenoids (\(crtO\)), and hydroxylated, desaturated
carotenoids \((crtC, crtD)\) (table 3.2). Like the \textit{Cyanobacteria} and \textit{Chlorobiales}, \textit{Cab. thermophilum} contains the multi-step (CrtP, CrtQ, CrtH) pathway for the synthesis of all-\textit{trans}-lycopene from phytoene rather than the CrtI-dependent pathway found in the \textit{Chloroflexi} and most other prokaryotes \((5, 18)\). BLASTP analyses using the recently characterized CrtZ-ISO phytoene isomerase from \textit{Synechococcus elongatus} as query \((7)\) did not identify any matches to this protein even when e-values were relaxed to 1e-0. Interestingly, \textit{Cab. thermophilum} is predicted to have two different lycopene cyclases. It contains genes coding for the heterodimeric \(Y_cY_d\) lycopene cyclase and a CruA-type lycopene cyclase \((33)\). Additional information concerning the biochemistry of carotenoids in \textit{Cab. thermophilum} is presented in Chapter 5.

3.3.3 **Electron transport.** The genome of \textit{Cab. thermophilum} contains several putative operons that encode genes involved in electron transport. There is one complete set of genes encoding the 14 subunits that make up the type-1 NADH dehydrogenase (NADH:menaquinone oxidoreductase), or complex 1, and an additional nearly complete set of genes encoding 10 of the subunits. Some of the subunits are organized in apparent operons containing either 5 or 8 genes each, while other \textit{nuo} genes are scattered throughout the genome or occur in association with other operons (figure 3.4). An analysis of the \textit{nuo} genes in the three available genomes of \textit{Acidobacteria} revealed that these organisms also have two copies of most of the \textit{nuo} genes. In these organisms, these genes are organized in larger operons than in \textit{Cab. thermophilum}; in addition to a few isolated genes, 9 to 12 genes are found in each operon. A recent report on the genome of \textit{Desulfovibrio}
magneticus strain RS-1 revealed that this organism also has a nearly complete duplication of the nuo operon (38).

Cab. thermophilum synthesizes menaquinone-8 (Chapter 4), and orthologs were identified for all of the genes required for menaquinone biosynthesis via an alternative pathway first identified in Streptomyces coelicolor A3 (table 3.3, 22).

The menaquinol-oxidizing components of the electron transport chain appear to be highly branched in Cab. thermophilum, with as many as three possible alternatives. Firstly, there is a large operon on chromosome 2 with homologous genes to the Alternative Complex III (ACIII) found in Chloroflexus spp. and other bacteria, including two acidobacteria, Solibacter usitatus and Korebacter versatilis (54, IMG). In Cfx. aurantiacus, ACIII appears to be the sole mechanism for proton translocation across the membrane, as evidenced by biochemical and genomic analyses (54), and it is presumed that ACIII performs a function similar to the cytochrome bc1 complex in other bacteria by oxidizing quinol molecules and reducing cytochrome c. Initial biochemical analyses performed with ACIII in Cfx. aurantiacus support this role, as ACIII reduced auracyanin in the presence of menaquinol-4 (19). In Cab. thermophilum, this quinol oxidase may be coupled with a terminal cytochrome aa3 oxidase, which is encoded just downstream (table 2, figure 3.5). In addition, the putative ACIII operon also includes a gene annotated as SCO1/senC, that is also found in other ACIII operons and is involved in the biogenesis of respiratory and photosynthetic complexes (54).

In addition to ACIII, chromosome 1 encodes two operons that include genes encoding for cytochrome b subunits (petB) of the cytochrome bc1 complex and
Rieske Fe/S proteins (petA) (table 3.2, figure 3.5). In one of the operons there is also a gene with some sequence similarity to subunits of the cytochrome bd-type quinol oxidase that is usually associated with microaerophilic conditions. Because there is no petC in this genome, this may encode a replacement for cytochrome c1. BLASTP analyses of the two copies of petA and petB suggest that the two operons have different origins. The petA and petB located in the operon shown in the A panel of figure 3.5 are closely related to apparently orthologous genes in GSB, whereas the petA and petB genes shown in the B panel share greatest sequence similarity with genes found in other Acidobacteria.

Because Cab. thermophilum encounters a wide range of conditions throughout a diel cycle in the microbial mats in which it occurs, from oxygen supersaturation at midday to nearly anoxic conditions during the night, it is plausible that Cab. thermophilum uses both the ACIII and the cytochrome bc1 complexes under different oxygen conditions throughout the day. For example, biochemical studies on Cfx. aurantiacus revealed the existence of two ACIII systems: one that is preferentially used for respiration while the other is used during phototrophic growth (54). In addition, one of these complexes might not be a proton pump and could be used to uncouple electron transport and the generation of reducing equivalents from ATP synthesis. Gene expression analyses will help determine if and when these genes are transcribed, and biochemical analyses will be required to demonstrate when these complexes are active.

In addition to these complexes, there is a putative operon with genes that code for ferredoxin and ferredoxin:NADP+ oxidoreductase. It is unknown at this
time whether the products of these genes are involved in light-induced electron transport or are required for some other process. It is presently uncertain how electrons are returned to the reaction center during cyclic electron transport, but ferredoxin could certainly play a role in this process. There are also genes coding for subunits of additional oxidases, such a cbb3-type cytochrome c oxidase, which is located immediately downstream of the pscA, pscB and fmoA genes. This type of cytochrome oxidase has a higher affinity for oxygen, which could be important during transition periods when the mat becomes anoxic or nearly so.

3.3.4 Metabolism and carbon acquisition. Because Cab. thermophilum grows poorly in culture, and its growth is dependent on the presence of heterotrophic members of the mat, I looked for incomplete metabolic pathways that might provide insights into identifying media supplements to enhance growth and that might lead to an axenic culture. For example, the genome of Cab. thermophilum does not seem to encode for a sulfate reductase (figure 3.6), even though we have been culturing this organism for over two years in a medium with sulfate as the sole sulfur source. It does have an apparent pathway to use thiosulfate, and substituting thiosulfate for sulfate in the media led to a ~15% increase in growth yield. Other sulfur sources, such as cysteine, methionine or elemental sulfur, showed no effect on growth yields under the conditions tested.

The genome of Cab. thermophilum does not contain complete pathways for the biosynthesis of the branched chain amino acids valine, leucine and isoleucine (figure 3.7). In fact, the only genes in the genome belonging to these pathways are the ones involved in the terminal steps, the aminotransferases and aminoacyl tRNA
synthetases, for each of these three amino acids. Similar to some other
*Acidobacteria* (52), the genome apparently contains a complete set of genes for the
catabolism of branched amino acids. Interestingly, one of the few transporters
found in the genome is predicted to transport branched chain amino acids (figure
3.8). These three observations strongly suggest that Cab. thermophilum obtains
branched chain amino acids from its environment.

The addition of ammonia as a nitrogen source was required in order to
isolate Cab. thermophilum from a cyanobacterial enrichment and, as expected, the
genome does not encode nitrate reductase or nitrite reductase (figure 3.9). Similarly, no genes for nitrogen fixation were identified. A single gene,
Cabther_01630, is predicted to be the ammonia transporter for this organism.

The genome of Cab. thermophilum shows that this organism is a
heterotroph and does not have the capability to fix carbon dioxide and to grow
autotrophically. The genome lacks key genes in all known carbon fixation pathways
employed by other chlorophototrophs, including the Calvin-Benson-Bassham,
reverse TCA cycle and the hydroxypropionate pathways. The genome does encode
a complete set of enzymes for glycolysis as well as those for the oxidative TCA
cycle and for the oxidation of small carboxylic acids, such as acetate and lactate,
and alcohols such as ethanol. These findings support the observations made when
culturing this organism in the laboratory. Cab. thermophilum does not grow
autotrophically in the laboratory and the growth medium is routinely supplemented
with a mixture of fermentation end products such as acetate, propionate, lactate and
glycolate. Like the genomes of other *Acidobacteria*, it contains the *coxLS* operon
coding for the aerobic carbon monoxide dehydrogenase (52), which might protect cells and provide a source of electrons by oxidizing carbon monoxide to carbon dioxide.

3.3.5. Transposases and repetitive elements. The genome has a large number (n=31) of transposases and repeated sequences. Most of these are located on the smaller chromosome and are highly similar (74-91% identity) to transposases found in the two dominant cyanobacterial species in the mat, *Synechococcus* spp. type A and type B’ (table 3.4). There are also transposases homologous to those of *Methyloacidiphilum infernorum*, an aerobic methanotroph, although this organism so far has been isolated only from acidic hot springs, and has not been detected in Octopus Spring (46). Interestingly, most of the *Synechococcus*-like transposases are located in the smaller chromosome, whereas the *M. infernorum*-like transposases are more numerous on the larger chromosome. We have investigated the gene neighborhoods around the transposase genes to determine whether horizontal gene transfer has occurred among the microorganisms in this hot spring. Although the G + C content of the genes neighboring these elements differed significantly from that of the whole genome, we did not find any occurrence of other *Synechococcus* genes associated with these transposase genes. We have found, however, that ca. 50% of the DNA methylases found in the genome of Cab. thermophilum are highly homologous to DNA methylases in *Roseiflexus* spp. In a few cases, these methylases seemed to be associated with the *Synechococcus* spp. transposases (figure 3.10). A search in the genomes of both *Synechococcus* spp. type A and type B’ did not reveal any of the *Roseiflexus* spp. DNA methylases in these genomes,
which suggests that these genes might have been acquired directly from Roseiflexus spp. Like Cab. thermophilum, the Octopus Spring (OS) Synechococcus spp. Type A and B’ genomes contain an unusually large number of transposases (ca. 80 to 100 respectively) whereas other Synechococcus species have ca. 0 to 10 copies of these genes (21, IMG). In addition, both the Synechococcus spp. genomes and the genome of Cab. thermophilum contain Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) elements, although the elements from the different clades are not similar in sequence. The combined presence of all these genes suggests that phage infection might be a significant threat to the microorganisms of microbial mat communities of OS and MS, and that phage infection may promote gene transfer and exchange of DNA modification systems among these microorganisms. Although no phages have been isolated from OS, there have been several sequence-based studies documenting the phage population of this hot spring (40, 45) and the potential for horizontal gene transfer in this environment (45).

3.3.6. Comparison to Candidatus Chloracidobacterium thermophilum type M.

In addition to the genome sequence of the laboratory isolate (4), we have large assembled contigs generated from metagenomic libraries produced from Octopus and Mushroom Springs that belong to a closely related Cab. thermophilum population (hereafter referred as Cab. thermophilum type M; Klatt et al., in preparation). This sequence represents the consensus sequence of the major Cab. thermophilum population present in these two similar hot springs sampled at 60°C (Octopus Spring) and 65°C (Mushroom Spring) (51). A comparison of the complete genome sequence of our isolate and contigs from Cab. thermophilum type
M showed that gene order (*i.e.*, synteny) was much more highly conserved on the large chromosomes than on the smaller chromosomes of these two populations (figure 3.11). This phenomenon has been observed in other microorganisms with multipartite genomes. Two independent genomic comparisons of closely related bacteria of the genus *Burkholderia* and of three *Agrobacterium* biovars, both with genomes distributed on two chromosomes, showed that the large (primary) chromosomes for each of these genera were also largely syntenic, while the smaller (secondary) chromosomes had undergone significant rearrangements (23, 47). The authors of the *Agrobacterium* study proposed that the smaller chromosomes were initially large plasmids that eventually acquired essential genes from the main chromosome, rather than the smaller plasmid originating from a split of a larger ancestral chromosome. The two distinctive origins of replication in the *Cab. thermophilum* chromosomes support this hypothesis. The smaller chromosome has a *parAB* origin, usually associated with plasmids (1), whereas the large chromosome appears to have the more typical DnaA-dependent origin of replication (36). Several studies on the relative gene expression in the two chromosomes of *Vibrio cholera* and other *Vibrio* spp., all of which have multipartite genomes, have indicated a possible ecological advantage for possessing two chromosomes (14, 48). By hosting mostly nonessential genes that are transcribed later in the growth cycle, the secondary chromosome is less constrained and able to acquire and modify new genes that might confer advantages to the organism under changing environmental conditions. Indeed, the evolutionary rates of both nonsynonymous (*dN*) and synonymous (*dS*) substitutions in the smaller
chromosome of *Vibrio* and *Burkholderia* spp. were faster than in the larger chromosome, and there is reduced codon usage bias in the genes of the smaller chromosome in all known organisms with multipartite chromosomes (11). In *Cab.* thermophilum the difference in CAI between the chromosomes is negligible (average CAI = 0.692 and 0.691, respectively for the larger and smaller chromosome), which may indicate that this model might not apply to this organism. However, the lack of synteny between the small chromosome of our isolate compared to that of the metagenomic secondary chromosome, the larger number of transposases present in this chromosome, as well as the fact that most growth-related genes are located in the large chromosome of *Cab.* thermophilum (table 3.2) support a similar dynamic role of the smaller chromosome to those found in other multipartite genomes. Thus, it is possible that having two chromosomes might help *Cab.* thermophilum to compete in and adapt to changing environmental conditions.

Gene content comparisons among the two strains of *Cab.* thermophilum, our lab isolate and the strain assembled from the metagenome, did not reveal significant differences between these strains (tables 3.4 and 3.5), although these comparisons cannot be completed at this time because there are gaps in the metagenome sequences. As expected, the metagenome assembly is missing most transposase genes and repeated sequences. Despite these gaps, comparative analyses have shown that there are more putative response regulators present in the contigs from the metagenome that in the lab isolate sequence. Miller *et al.* (37) and our own functional gene survey (appendix D) have detected four *Cab.* thermophilum subpopulations in White Creek hot spring, each with a specialized temperature
niche. This finding suggests that there could be genetic differences, albeit unknown at this time, that allow these strains to grow under different temperature conditions.

3.3.7 Comparison to other Acidobacteria. Cab. thermophilum is the first known chlorophototroph in the phylum Acidobacteria. Its gene content differences compared to the other acidobacterial genomes seem to reflect the vastly different environments in which these organisms are found. For example, Cab. thermophilum lacks genes that code for orphan permeases that so far have been found only in acidobacterial genomes, along with a transcriptional activator (padR) associated with them. This transcriptional activator is responsible for activating a response to phenolic toxicity in other microorganisms, a response which presumably would not be necessary in a hot spring microbial mat. In addition, unlike other sequenced Acidobacteria, Cab. thermophilum lacks complete pathways for the breakdown of complex carbohydrates such as cellulose and chitin, although it does possess a gene coding for glucose amylase (Cabther_23150), which might be involved in the degradation of starch and glycogen. Lastly, it does not have genes that code for nitrate or nitrite reductase (table 3.1).

It has been suggested that Acidobacteria might play a role in carbon degradation and nitrogen cycling in soils, but the lack of complete degradation pathways for complex carbohydrates and the inability to reduce nitrate in Cab. thermophilum seems to indicate that the same role cannot be assigned to this hot spring acidobacterium. Similar to other Acidobacteria, Cab. thermophilum is surrounded by a slime layer. Although Cab. thermophilum does not seem to possess a complete cellulose biosynthesis pathway, its genome does contain genes involved
in extracellular secretions scattered across the genome, including a homolog of the \textit{epsD} gene, encoding a common extracellular protein.

Cab. thermophilum belongs to the subdivision 4 of the \textit{Acidobacteria} (figure 4.12). To our best knowledge, Cab. thermophilum is the only \textit{Acidobacterium} of this group that can presently be grown in laboratory culture.

\textbf{3.3.8 Other genes of interest.} Genes encoding enzymes for hopanoid biosynthesis are found in a putative operon on chromosome 2. Although hopanoids have been assumed to contribute to membrane stability in bacteria, much like sterols do in eukaryotic cells, a recent study in \textit{Rhodopseudomonas palustris} showed that some of these pentacyclic terpenoids also play a role in pH homeostasis in this organism (53). Given the pH extremes that Cab. thermophilum encounters in its natural habitat (43), where daytime pH is usually above 9 due to photosynthetic activity of the cyanobacteria, it is possible that the hopanoids might have a similar role in pH homeostasis in this organism.

The genome of Cab. thermophilum contains genes coding for several antioxidant enzymes such as catalase, superoxide dismutase, methionine sulfoxide reductase, rubrerythrin, peroxiredoxin and several peroxidases distributed on both chromosomes. As with pH, Cab. thermophilum must withstand extreme changes in oxygen conditions throughout a diel cycle (43), and these genes likely contribute to its tolerance of the very high, supersaturating oxygen partial pressures that occur during peak periods of cyanobacterial photosynthesis.

\textbf{3.3.9 Summary.} This chapter describes the completely sequenced genome of \textit{Candidatus Chloracidobacterium} thermophilum. In agreement with culture studies,
analysis of the genome reveals this organism to be a chlorophotoheterotroph that uses light energy to generate chemical energy for growth and that utilizes organic carbon compounds, especially fermentation end-products, as a carbon source.

Cab. thermophilum is difficult to grow in culture, and the genome provides several important clues for this. The organism lacks the capability to reduce carbon, sulfate, and nitrate, and lacks the ability to synthesize branched chain amino acids. Although supplementing the growth medium with these compounds has not resulted in either faster and/or more sustained growth or in an axenic culture, knowledge of the multiple auxotrophies of Cab. thermophilum has revealed very important insights into the ecological constraints that this microorganism must encounter in nature.

The genome sequence has provided evidence for further ecological constraints, for example in the form of phage attack, and suggests that horizontal gene transfer might be a mechanism by which members of the mat deal with phage attack. Lastly, analyses of the genome sequence have generated hypotheses concerning the synthesis of the abundant carotenoids that occur in this organism as well as about nature of the electron transport chains in Cab. thermophilum, which could provide flexibility in dealing with changing oxygen concentrations in the mat.
3.4. REFERENCES:


FIGURE LEGENDS

Figure 3.1. Map of the two chromosomes of Candidatus Chloracidobacterium thermophilum. The Cab. thermophilum genome consists of two circular chromosomes, 2,683,362 bp and 1,012,010 bp, respectively, and has a GC content of 61 mol%. From outside to the center: genes on forward strand (colored by COG categories), genes on reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black) GC content, GC skew. Figure generated through the IMG site (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi).

Figure 3.2. Distribution of CDS belonging to different Cluster of Orthologous Groups (COGs) on the two chromosomes of Candidatus Chloracidobacterium. Individual categories: Information storage and processing: (1) Translation, ribosomal structure and biogenesis; (2) Transcription; (3) Replication, recombination and repair. Cellular processes and signaling: (1) Cell cycle control, cell division, chromosome partioning; (2) Defense mechanisms; (3) Signal transduction mechanisms, (4) Cell wall/membrane/envelope biogenesis; (5) cell motility; (6) Intracellular trafficking, secretion; (7) Posttranslational modification, protein turnover, chaperones. Metabolism: (1) Energy production and conversion; Transport and metabolism of (2) Carbohydrates; (3) Amino acids; (4) Nucleotides; (5) Coenzymes; (6) Lipids; (7) Inorganic ions; (8) Secondary metabolites. Poorly characterized: (1) General function prediction only; (2) Function unknown.

Figure 3.3. Neighbor-joining phylogenetic tree of B(Chl) synthases rooted with the Escherichia coli gene ubiA. Bootstrap values after 100 iterations are shown.

Figure 3.4. Representative nuo operons found in Candidatus Chloracidobacterium thermophilum. The two nearly complete sets of nuo genes present in the genome of Cab. thermophilum are organized in the shown three small operons as well as additional scattered genes across the chromosomes.

Figure 3.5. Putative operons involved in the oxidation of menaquinol. A. Three genes encoding the corresponding subunits of a cytochrome bc1 complex located in chromosome 1. B. Two genes that encode for the corresponding subunits of a cytochrome bc1 complex located in chromosome 1. C. A putative operon that codes for the subunits of ACIII. A gene annotated as phzF, involved in phenazine biosynthesis and a gene annotated as SCO1/senC, involved in biogenesis of respiratory and photosynthetic apparatus are also shown.

Figure 3.6. KEGG diagram showing biochemical pathways involved in sulfur metabolism. Enzymes for each chemical reaction are shown in boxes by their Enzyme Commission (EC) number. Enzymes that are encoded in the genome of Cab. thermophilum are shown in green.

Figure 3.7. KEGG diagram showing biosynthetic pathways for the amino acids valine, leucine and isoleucine. Enzymes for each chemical reaction are shown in
boxes by their Enzyme Commission (EC) number. Enzymes that are encoded in the genome of Cab. thermophilum are shown in green.

Figure 3.8. KEGG diagram showing ABC transporters found in Candidatus Chloracidobacterium thermophilum. Enzymes for each chemical reaction are shown in boxes by their Enzyme Commission (EC) number. Enzymes that are encoded in the genome of Cab. thermophilum are shown in green.

Figure 3.9. KEGG diagram showing biochemical pathways involved in nitrogen metabolism. Enzymes for each chemical reaction are shown in boxes by their Enzyme Commission (EC) number. Enzymes that are encoded in the genome of Cab. thermophilum are shown in green.

Figure 3.10. Representative CRISPR gene neighborhood. The transposases shown are homologous to those found in Synechococcus spp. found in the OS chlorophhototrophic microbial mat. The DNA methylase shown is homologous to DNA methylases found in Roseiflexus spp.

Figure 3.11. Comparison of the genome of the lab isolate with the genome of the dominant Candidatus Chloracidobacterium thermophilum species in the mat. Seven and three contigs (blue bars at top) assembled from a metagenomic library produced from the phototrophic mats of Octopus and Mushroom Springs were aligned to the large and small chromosomes, respectively, of the genome of our lab isolate.

Figure 3.12. Neighbor-joining phylogenetic tree for 16S rRNA showing the relationship of cultured and uncultured acidobacteria belonging to groups 1-6 and 8. Branches without names indicate organisms that have not been cultured in the laboratory. *Denotes organisms with a sequenced genome.
Table 3.1 General genome features of *Candidatus* Chloracidobacterium thermophilum and three additional available acidobacterial genomes

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genome size</th>
<th>Coding sequences</th>
<th>% G + C</th>
<th>rRNA operons</th>
<th>Optimal temperature (°C)</th>
<th>Optimal pH</th>
<th>Environment</th>
<th>Nitrate reduction</th>
<th>Nitrite reduction</th>
<th>Cellulose degradation</th>
<th>Chitin degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cab. thermophilum</em></td>
<td>3,695,372</td>
<td>3,071</td>
<td>61.3</td>
<td>1</td>
<td>53</td>
<td>9</td>
<td>Hot spring</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Acidobacterium capsulatum</em></td>
<td>4,127,496</td>
<td>3,502</td>
<td>60.3</td>
<td>1</td>
<td>30</td>
<td>3.5</td>
<td>Acid mine drainage</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Korebacter versatilis</em></td>
<td>5,650,368</td>
<td>5,239</td>
<td>58.4</td>
<td>1</td>
<td>25</td>
<td>6</td>
<td>Clover pasture</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Solibacter usitatus</em></td>
<td>9,965,640</td>
<td>8,568</td>
<td>61.9</td>
<td>2</td>
<td>25</td>
<td>7</td>
<td>Clover pasture</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 3.2 Energy harvesting genes in *Candidatus* Chloracidobacterium thermophilum.

<table>
<thead>
<tr>
<th>LOCUS TAG</th>
<th>PRODUCT NAME</th>
<th>CHROMOSOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabther_12200</td>
<td>crtB  Phytoene synthase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_20890</td>
<td>crtP  Phytoene desaturase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_00940</td>
<td>crtH  Phytoene desaturase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_00480</td>
<td>crtQ  Phytoene desaturase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_06920</td>
<td>ctyC  Lycopene cyclase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_06910</td>
<td>ctyD  Lycopene cyclase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_12080</td>
<td>ctnA  Lycopene cyclase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_10330</td>
<td>cto  β-carotene ketolase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_05620</td>
<td>ctc  Hydroxyneoporphorene synthase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_20690</td>
<td>ctd  Carotenoid 3',4' desaturase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_21570</td>
<td>pscA  Type 1 Reaction center core protein</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_21560</td>
<td>pscB  Type 1 Reaction center core protein</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_03870</td>
<td>bchD  Mg-protoporphyrin IX methyl transferase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_23660</td>
<td>bciB  Divinyl reductase</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_07510</td>
<td>bchH  Mg-chelatase, subunit H</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_03870</td>
<td>bchD  Mg-chelatase, subunit D</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_23670</td>
<td>bchF  2-vinyl-bacteriochlorophyllide hydratase</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_09200</td>
<td>bchX  Chlorophyllide reductase, subunit X</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_15520</td>
<td>bchY  Chlorophyllide reductase, subunit Y</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_15080</td>
<td>bchZ  Chlorophyllide reductase, subunit Z</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_15070</td>
<td>bchC  2-desacetyl-2-hydroxyethyl bacteriochlorophyllide a dehydrogenase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_23210</td>
<td>bchG  Bacteriochlorophyll c synthetase</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_12210</td>
<td>bchK  Bacteriochlorophyll c synthetase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_15080</td>
<td>bchL  Light independent protochlorophyllide reductase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_09200</td>
<td>bchX  Chlorophyllide reductase, subunit X</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_15080</td>
<td>bchZ  Chlorophyllide reductase, subunit Z</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_25250</td>
<td>* Quinol:cytochrome c oxidoreductase monoheme cytochrome</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25260</td>
<td>** Quinol:cytochrome c oxidoreductase membrane protein</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25270</td>
<td>** Quinol:cytochrome c oxidoreductase quinone-binding subunit 1</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25280</td>
<td>** [4Fe-4S] containing molybdopterin oxidoreductase</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_09000</td>
<td>petA  Rieske [Fe-S] protein</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_08990</td>
<td>petB  Cytochrome b subunit of the bc complex</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_08980</td>
<td>cydD  Cytochrome bd-type quinol oxidase</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_07580</td>
<td>petA  Rieske [Fe-S] protein</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_07570</td>
<td>petB  Cytochrome b subunit of the bc complex</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_25210</td>
<td>coxA  Cytochrome c oxidase, subunit 1</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25210</td>
<td>coxB  Cytochrome c oxidase, subunit 2</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25190</td>
<td>coxC  Heme/copper type cytochrome/quinol oxidase, subunit 3</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25180</td>
<td>coxD* caa(3)-type oxidase, subunit 4</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25250</td>
<td>** Quinol:cytochrome c oxidoreductase monoheme cytochrome</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25260</td>
<td>** Quinol:cytochrome c oxidoreductase membrane protein</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25270</td>
<td>** Quinol:cytochrome c oxidoreductase quinone-binding subunit 1</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25280</td>
<td>** [4Fe-4S] containing molybdopterin oxidoreductase</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25290</td>
<td>** Quinol:cytochrome c oxidoreductase pentaheme cytochrome</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_21520</td>
<td>ccoP  Cytochrome c oxidase Cbb3-type, subunit 3</td>
<td>1</td>
</tr>
</tbody>
</table>

* Tentative gene name designation
** No standard gene designation available
Table 3.3 Menaquinone genes in *Candidatus* Chloracidobacterium thermophilum.

<table>
<thead>
<tr>
<th>Locus Tag* in <em>Streptomyces coelicolor</em> A3(2)</th>
<th>E value</th>
<th>Locus Tag* in <em>Cab. thermophilum</em></th>
<th>Annotation in <em>Cab. thermophilum</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>SCO4506</td>
<td>4^-26</td>
<td>Cabther_14820</td>
<td>Predicted periplasmic solute-binding protein</td>
</tr>
<tr>
<td>SCO4327</td>
<td>8^-15</td>
<td>Cabther_00730</td>
<td>Futalosine nucleosidase</td>
</tr>
<tr>
<td>SCO4550</td>
<td>6^-28</td>
<td>Cabther_29130</td>
<td>Radical SAM domain protein, CofH subfamily</td>
</tr>
<tr>
<td>SCO4326</td>
<td>4^-58</td>
<td>Cabther_09520</td>
<td>Predicted periplasmic solute-binding protein</td>
</tr>
<tr>
<td>SCO4556</td>
<td>5^-56</td>
<td>Cabther_05020</td>
<td>Ubiquinone/menaquinone biosynthesis methyltransferases</td>
</tr>
<tr>
<td>SCO4491</td>
<td>4^-46</td>
<td>Cabther_11640</td>
<td>Putative 4-hydroxybenzoate polyprenyltransferase</td>
</tr>
<tr>
<td>SCO4490</td>
<td>0</td>
<td>Cabther_16540</td>
<td>3-octaprenyl-4-hydroxybenzoate decarboxylase</td>
</tr>
<tr>
<td>SCO4492</td>
<td>4^-22</td>
<td>Cabther_16910</td>
<td>Polyprenyl P-hydroxybenzoate and phenylacrylic acid decarboxylases</td>
</tr>
</tbody>
</table>

*Locus Tags as they appear in the Integrated Microbial Genomes website (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi, ref. 34).*
Table 3.4 Location and similarities of transposases found in *Candidatus* Chloracidobacterium thermophilum.

<table>
<thead>
<tr>
<th>LOCUS TAG</th>
<th>PRODUCT NAME</th>
<th>AA LENGTH</th>
<th>CHROMOSOME</th>
<th>TOP BLASTP HIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabther_02830</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>444</td>
<td>1</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_02970</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>446</td>
<td>1</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_06970</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>446</td>
<td>1</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_07130</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>446</td>
<td>1</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_10830</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>446</td>
<td>1</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_04710</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>517</td>
<td>1</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_19090</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>380</td>
<td>1</td>
<td>Nitrosococcus ocean ATCC 19707</td>
</tr>
<tr>
<td>Cabther_02110</td>
<td>Transposase and inactivated derivatives</td>
<td>255</td>
<td>1</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_13840</td>
<td>Transposase and inactivated derivatives</td>
<td>246</td>
<td>1</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_16260</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>290</td>
<td>1</td>
<td>Xylella fastidiosa Dixon</td>
</tr>
<tr>
<td>Cabther_23000</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>446</td>
<td>2</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_25650</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>429</td>
<td>2</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_27990</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>517</td>
<td>2</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_23340</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>420</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_03550</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>420</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_26370</td>
<td>Transposase and inactivated derivatives</td>
<td>420</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_26390</td>
<td>Transposase and inactivated derivatives</td>
<td>420</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_26820</td>
<td>Transposase and inactivated derivatives</td>
<td>420</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_26080</td>
<td>Transposase and inactivated derivatives</td>
<td>420</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_28520</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>395</td>
<td>2</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_29480</td>
<td>Transposase and inactivated derivatives</td>
<td>338</td>
<td>2</td>
<td>Synechococcus JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_27440</td>
<td>Transposase and inactivated derivatives</td>
<td>341</td>
<td>2</td>
<td>Synechococcus JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_25790</td>
<td>Transposase and inactivated derivatives</td>
<td>278</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_30850</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>309</td>
<td>2</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
<tr>
<td>Cabther_26670</td>
<td>Transposase and inactivated derivatives</td>
<td>246</td>
<td>2</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_27690</td>
<td>Transposase and inactivated derivatives</td>
<td>205</td>
<td>2</td>
<td>Synechococcus sp. JA-2-3B'a</td>
</tr>
<tr>
<td>Cabther_23980</td>
<td>Transposase and inactivated derivatives</td>
<td>222</td>
<td>2</td>
<td>Synechococcus sp. JA-3-3Ab</td>
</tr>
<tr>
<td>Cabther_23730</td>
<td>Transposase, IS 605 OrfB family, central region</td>
<td>202</td>
<td>2</td>
<td>Methylacidiphylum infernorum V4</td>
</tr>
</tbody>
</table>
Table 3.5 Genes found in *Candidatus Chloracidobacterium thermophilum* but not in type M strain.

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Product Name</th>
<th>Number of aa</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabther_12350</td>
<td>3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase</td>
<td>352</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_10640</td>
<td>40-residue YVTN family beta-propeller repeat</td>
<td>371</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_00250</td>
<td>alanyl-tRNA synthetase</td>
<td>876</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_24190</td>
<td>Antitoxin of toxin-antitoxin stability system</td>
<td>76</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_27500</td>
<td>arsenical-resistance protein</td>
<td>378</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_00190</td>
<td>asparaginyl-tRNA synthetase</td>
<td>450</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_24650</td>
<td>Bacteriophage protein gp37</td>
<td>204</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_30220</td>
<td>BsuBI/PstI restriction endonuclease C-terminus</td>
<td>320</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_00190</td>
<td>asparaginyl-tRNA synthetase</td>
<td>450</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_24650</td>
<td>Bacteriophage protein gp37</td>
<td>204</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_08850</td>
<td>CAAX amino terminal protease family.</td>
<td>272</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_00190</td>
<td>asparaginyl-tRNA synthetase</td>
<td>450</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_24650</td>
<td>Bacteriophage protein gp37</td>
<td>204</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_08850</td>
<td>CAAX amino terminal protease family.</td>
<td>272</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_00190</td>
<td>asparaginyl-tRNA synthetase</td>
<td>450</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_24650</td>
<td>Bacteriophage protein gp37</td>
<td>204</td>
<td>2</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Count</td>
<td>Class</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Cabther_18890</td>
<td>Predicted flavoprotein</td>
<td>171</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_08800</td>
<td>Predicted membrane protein</td>
<td>204</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_18060</td>
<td>Predicted membrane protein</td>
<td>235</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_02630</td>
<td>Predicted NAD/FAD-dependent oxidoreductase</td>
<td>366</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_25860</td>
<td>Predicted nucleotidyltransferases</td>
<td>120</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_25760</td>
<td>Predicted nucleotidyltransferases</td>
<td>81</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_24910</td>
<td>Predicted phosphatase homologs to histone macroH2A1</td>
<td>446</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_15700</td>
<td>Predicted transcriptional regulator containing an HTH domain</td>
<td>488</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_09070</td>
<td>Restriction endonuclease</td>
<td>430</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_00340</td>
<td>Restriction endonuclease S subunits</td>
<td>420</td>
<td>2</td>
</tr>
<tr>
<td>Cabther_00270</td>
<td>Thioredoxin</td>
<td>108</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_02110</td>
<td>Uncharacterised protein family</td>
<td>255</td>
<td>1</td>
</tr>
<tr>
<td>Cabther_02280</td>
<td>Zn-dependent hydrolases, including glyoxylases</td>
<td>215</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3.6. Genes found in Type M strain but not in *Candidatus Chloracidobacterium thermophilum*.

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Product Name</th>
<th>Number of aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>MatcabDRAFT_06170</td>
<td>Adenine specific DNA methylase Mod</td>
<td>302</td>
</tr>
<tr>
<td>MatcabDRAFT_06180</td>
<td>Adenine specific DNA methylase Mod</td>
<td>927</td>
</tr>
<tr>
<td>MatcabDRAFT_24770</td>
<td>Adenylate kinase and related kinases</td>
<td>163</td>
</tr>
<tr>
<td>MatcabDRAFT_07440</td>
<td>Beta-lactamase class C and other penicillin binding proteins</td>
<td>373</td>
</tr>
<tr>
<td>MatcabDRAFT_10070</td>
<td>CAAX amino terminal protease family.</td>
<td>285</td>
</tr>
<tr>
<td>MatcabDRAFT_04500</td>
<td>Coenzyme PQQ synthesis protein D (PqqD).</td>
<td>87</td>
</tr>
<tr>
<td>MatcabDRAFT_16940</td>
<td>conserved hypothetical protein TIGR00278</td>
<td>91</td>
</tr>
<tr>
<td>MatcabDRAFT_30070</td>
<td>CRISPR-associated protein GSU0053</td>
<td>286</td>
</tr>
<tr>
<td>MatcabDRAFT_29750</td>
<td>CRISPR-associated protein, GSU0053 family, N-terminal domain</td>
<td>407</td>
</tr>
<tr>
<td>MatcabDRAFT_29760</td>
<td>CRISPR-associated protein, GSU0054 family</td>
<td>143</td>
</tr>
<tr>
<td>MatcabDRAFT_30080</td>
<td>CRISPR-associated protein, GSU0054 family</td>
<td>184</td>
</tr>
<tr>
<td>MatcabDRAFT_26230</td>
<td>CRISPR-associated protein, GSU0054 family, N-terminal domain</td>
<td>403</td>
</tr>
<tr>
<td>MatcabDRAFT_29760</td>
<td>CRISPR-associated protein, GSU0054 family, N-terminal domain</td>
<td>143</td>
</tr>
<tr>
<td>MatcabDRAFT_30080</td>
<td>CRISPR-associated protein, GSU0054 family</td>
<td>184</td>
</tr>
<tr>
<td>MatcabDRAFT_29760</td>
<td>CRISPR-associated protein, GSU0054 family, N-terminal domain</td>
<td>403</td>
</tr>
<tr>
<td>MatcabDRAFT_29760</td>
<td>CRISPR-associated protein, GSU0054 family, N-terminal domain</td>
<td>143</td>
</tr>
<tr>
<td>MatcabDRAFT_30080</td>
<td>CRISPR-associated protein, GSU0054 family</td>
<td>184</td>
</tr>
<tr>
<td>MatcabDRAFT_26230</td>
<td>Glycosyltransferase</td>
<td>403</td>
</tr>
<tr>
<td>MatcabDRAFT_24690</td>
<td>having DD or ED</td>
<td>230</td>
</tr>
<tr>
<td>MatcabDRAFT_30820</td>
<td>Laminin G domain.</td>
<td>784</td>
</tr>
<tr>
<td>MatcabDRAFT_24820</td>
<td>LSU ribosomal protein L14P</td>
<td>125</td>
</tr>
<tr>
<td>MatcabDRAFT_24840</td>
<td>LSU ribosomal protein L16P</td>
<td>145</td>
</tr>
<tr>
<td>MatcabDRAFT_24890</td>
<td>LSU ribosomal protein L18P</td>
<td>212</td>
</tr>
<tr>
<td>MatcabDRAFT_24800</td>
<td>LSU ribosomal protein L5P</td>
<td>168</td>
</tr>
<tr>
<td>MatcabDRAFT_29670</td>
<td>MerR family regulatory protein.</td>
<td>58</td>
</tr>
<tr>
<td>MatcabDRAFT_08410</td>
<td>Molecular chaperone</td>
<td>430</td>
</tr>
<tr>
<td>MatcabDRAFT_04480</td>
<td>Molybdenum cofactor biosynthesis enzyme</td>
<td>155</td>
</tr>
<tr>
<td>MatcabDRAFT_22000</td>
<td>Molybdopterin-guanine dinucleotide biosynthesis protein A</td>
<td>211</td>
</tr>
<tr>
<td>MatcabDRAFT_22380</td>
<td>Nucleoside-diphosphate-sugar pyrophosphorylase involved in</td>
<td></td>
</tr>
<tr>
<td>MatcabDRAFT_16300</td>
<td>PemK-like protein.</td>
<td>154</td>
</tr>
<tr>
<td>MatcabDRAFT_01450</td>
<td>Polyketide cyclase / dehydrase and lipid transport.</td>
<td>147</td>
</tr>
<tr>
<td>MatcabDRAFT_26970</td>
<td>Predicted CoA-binding protein</td>
<td>141</td>
</tr>
<tr>
<td>MatcabDRAFT_06120</td>
<td>Predicted dehydrogenases and related proteins</td>
<td>297</td>
</tr>
<tr>
<td>MatcabDRAFT_12070</td>
<td>Predicted helicase</td>
<td>752</td>
</tr>
<tr>
<td>MatcabDRAFT_02550</td>
<td>Predicted integral membrane protein</td>
<td>166</td>
</tr>
<tr>
<td>MatcabDRAFT_24780</td>
<td>Preprotein translocase subunit SecY</td>
<td>299</td>
</tr>
<tr>
<td>MatcabDRAFT_02940</td>
<td>Protoglobin.</td>
<td>158</td>
</tr>
<tr>
<td>MatcabDRAFT_02920</td>
<td>Restriction endonuclease</td>
<td>355</td>
</tr>
<tr>
<td>MatcabDRAFT_24830</td>
<td>ribosomal protein L29</td>
<td>64</td>
</tr>
<tr>
<td>MatcabDRAFT_24790</td>
<td>Ribosomal protein L6P/L9E</td>
<td>163</td>
</tr>
<tr>
<td>MatcabDRAFT_24850</td>
<td>SSU ribosomal protein S3P</td>
<td>198</td>
</tr>
<tr>
<td>MatcabDRAFT_24740</td>
<td>SSU ribosomal protein S4P</td>
<td>160</td>
</tr>
<tr>
<td>MatcabDRAFT_08400</td>
<td>Superfamily I DNA and RNA helicases</td>
<td>408</td>
</tr>
<tr>
<td>MatcabDRAFT_06140</td>
<td>Type III restriction enzyme, res subunit.</td>
<td>255</td>
</tr>
<tr>
<td>MatcabDRAFT_00380</td>
<td>Uncharacterized conserved protein related to C-terminal domain of</td>
<td></td>
</tr>
<tr>
<td>MatcabDRAFT_02130</td>
<td>Uncharacterized protein conserved in bacteria</td>
<td>64</td>
</tr>
</tbody>
</table>
Figure 3.1

2,683,362 bp

1,012,010 bp
Figure 3.2
Figure 3.3

Cab thermophilum ChlG
- Chlorocarpeton thalassium ChlG
- Chlorobium phaeobacteroides ChlG
- Chlorobium tepidum ChlG
- Roseobacter littoralis BChl G
- Rhodo bacter sphaeroides BChl G

Cab. thermophilum BchG
- Chlorocarpeton thalassium BchG
- Chlorobium tepidum BchG
- Chlorobium phaeobacteroides BchG

Cab. thermophilum BchK
- Chlorocarpeton thalassium BchK
- Chlorobium phaeobacteroides BchK
- Chlorobium tepidum BchK
- Chloroflexus aggregans BchK
- Chloroflexus aurantiacus BchK
Figure 3.4
Figure 3.5
Figure 3.6
Figure 3.7

VALINE, LEUCINE AND Isoleucine BIOSYNTHESIS

\[
\begin{align*}
\text{Valine, Leucine and Isoleucine biosynthesis} & \\
\text{Deoxyn} & \rightarrow 2\text{-methylmalate} \\
\text{1-MVA} & \rightarrow \text{2-Carboxyoxo} \\
\text{1-MVA} & \rightarrow \text{2-Hydroxyethyl} \\
\text{2-OME} & \rightarrow \text{Pyruvate} \\
\text{2-Methylmalate} & \rightarrow \text{Acetyl-CoA} \\
\end{align*}
\]
Figure 3.10

transposase

Site-specific DNA methylase

CRISPR element
associated protein

CRISPR element

transposase
Figure 3.11
Figure 3.12
CHAPTER 4

Biochemical characterization of the chlorosomes of Candidatus Chloracidobacterium thermophilum

Publications:


ABSTRACT
Chlorosomes are vesicle-like, light-harvesting organelles characterized by containing very large numbers of bacteriochlorophyll (BChl) c, d or e molecules. These antenna structures have traditionally been found in green chlorophototrophs belonging to the Chlorobiales and the Chloroflexi. Biochemical analyses were conducted on the BChls of Candidatus Chloracidobacterium thermophilum, an acidobacterium that also synthesizes chlorosomes. In this organism BChl c molecules in the chlorosome self-aggregate in a manner different from the concentric nanotubes observed in the chlorosomes of the green sulfur bacterium Chlorobaculum tepidum. Cab. thermophilum resembles Cba. tepidum in that it methylates its antenna BChls at the C-8 and C-12 positions, but it uses a variety of isoprenoid and straight chain alkyl alcohols as esterifying moieties, like the filamentous anoxygenic phototroph Chloroflexus aurantiacus. Unlike the carotenoids found in other green bacteria, the major carotenoids in Cab. thermophilum are the xanthophylls echinenone and canthaxanthin. These carotenoids, which are believed to confer enhanced protection against reactive oxygen species, may have been selected for by the highly oxic environment, from which Cab. thermophilum was isolated. Cab. thermophilum chlorosomes also contain significant amounts of menaquinone-8, which probably acts as a quencher of energy transfer under oxic conditions. As in other green bacteria, the chlorosomes of Cab. thermophilum contain homologs of CsmA and CsmI/CsmJ in its lipid envelope, as well as several unique polypeptides. Interestingly, although Cab. thermophilum is an aerobe, its chlorosomes exhibit redox-dependent quenching of fluorescence emission. The combined biochemical analyses presented here show that the chlorosomes of Cab. thermophilum possess some unique features but also share many of the fundamental properties of chlorosomes from other phyla despite being synthesized under oxic conditions.
4.1 INTRODUCTION

Chlorosomes are vesicle-like, light-harvesting organelles characterized by containing very large numbers of bacteriochlorophyll (BChl) c, d or e molecules (>200,000 per chlorosome) that self-aggregate into supramolecular structures (18, 19). Because of the large number of pigment molecules in a single chlorosome, chlorosomes exhibit a remarkable ability to harvest light, and thus, organisms that synthesize these organelles are often found in low-light environments (35). These antenna structures have been found in three groups of chlorophototrophs: the green sulfur bacteria (GSB; Chlorobiales); the filamentous anoxygenic phototrophs (FAPs) of the genera Chloroflexus, Chloronema and Oscillochloris (the green Chloroflexi); and one representative from the Acidobacteria, Candidatus Chloracidobacterium (Cab.) thermophilum (5, 6). Most studies on the properties of chlorosomes have been conducted with chlorosomes isolated from the model organisms, Chlorobaculum tepidum (GSB) and Cfx. aurantiacus (FAP) (11-15, 47, 56), although there have been a few studies on the chlorosomes of other green bacteria such as the FAPs Chlorothrix halophila (64) and Chloronema giganteum (21), and the GSB including Chlorobium phaeobacteroides and Prosthecochloris aestuarii (40).

The bacteriochlorophylls (BChls) in chlorosomes are unique among all other (bacterio)chlorins in that they can self-aggregate and form supramolecular structures without the aid of proteins. This property is due to the absence of the $13^2$-methylcarboxyl group, which would prevent hydrogen bonding between neighboring BChls, and to the presence of a $3^1$-hydroxyl group that can chelate the
central Mg atom of adjacent BChl molecules (18, 19). Excitation energy absorbed by a BChl moves quickly through these suprastructures due to strong exciton coupling of the BChls, and it is eventually transferred to the reaction center through BChl \(a\) bound to the baseplate protein CsmA via Förster-type energy transfer (45, 54, 65). In GSB there is an additional antenna system, the BChl \(a\)-containing Fenna-Matthews-Olson (FMO) protein, that receives excitations from CsmA and transfers them to the reaction center BChls (18, 41, 43).

Two basic models for the structural organization of BChls in chlorosomes had been proposed in the past: a lamellar model (46) and a cylinder/rod model (52, 57). By combining results from solid-state NMR and cryoelectron microscopy (cryo-EM) of chlorosomes of wild-type and a \(bchQ\ bchR\ bchU\) mutant of the \(Cba.\ tepidum\) that lacks all three bacteriochlorophyllide \(c\) methyltransferase activities, the structure of BChls in chlorosomes was determined. This study showed that the BChl molecules in chlorosomes are arranged in \(syn-anti\) monomer stacks that form concentric nanotubes (19). The organization of the BChls in the wild type and mutant strain differ in the orientation of the \(syn-anti\) monomer stacks relative to the long axis of the chlorosome (19).

The use of the mutant strain \(bchQ\ bchR\ bchU\) was critical to this study. In wild-type \(Cba.\ tepidum\), a variety of alkyl groups can be found in the C-8\(^2\) and C-12\(^1\) positions of BChl \(c\), and an individual chlorosome contains a mixture of BChl \(c\) molecules with different methylation patterns at these positions (23). This heterogeneity translates into structural disorder in the BChl aggregates that had previously precluded the elucidation of the overall organization of the BChl
suprastructure. The molecular heterogeneity generated by the alkyl groups and the C-3\(^1\) chirality was eliminated in the \textit{bchQ bchR bchU} mutant, which resulted in the increased structural order that was critical for the elucidation of the BChl suprastructure in the mutant.

In addition to these alkyl groups on the C-8\(^2\) and C-12\(^1\) carbons, BChls in chlorosomes are esterified with either an isoprenoid or an unbranched alcohol chain at the propionic acid at the C-17 position. In most GSB, for example, over 95% of the BChl \(c\) molecules are esterified with the C-15 unsaturated isoprenoid, farnesol (7, 40). The FAP \textit{Cfx. aurantiacus} differs from \textit{Cba. tepidum}, a GSB, in that the BChl \(c\) molecules in its chlorosomes are not methylated at the C-8\(^2\) and C-12\(^1\) positions (3). Moreover, the BChl \(c\) molecules in \textit{Cfx. aurantiacus} are esterified at the C-17 propionic acid with alcohols that include both isoprenoid and unbranched alkyl species (22, 29, 33).

The carotenoid and quinone content of chlorosomes also differs among chlorosome-containing phototrophs (18). The major carotenoid found in \textit{Cba. tepidum} chlorosomes is chlorobactene, an aromatic derivative of \(\gamma\)-carotene, whereas the brown-colored GSB, such as \textit{C. phaeobacteroides}, contain mostly isorenieratene and \(\beta\)-isorenieratene, both aromatic dicyclic carotenoids (25, 39, 60). This distribution is probably due to the different environments in which these two types of organisms are found, and perhaps the different BChls that they contain in their chlorosomes. The absorption spectra of isorenieratene and BChl \(e\), the Bchl found in brown species, overlap (38, 39) thus having isorenieratene in these organisms expands the range of wavelengths available for light absorption. In
addition to these main carotenoids, both green and brown-colored Chlorobiales also synthesize small amounts of glycosylated carotenoids (37, 38, 59). The chlorosomes of Cfx. aurantiacus, under photosynthetic anoxic conditions, contain mostly \( \gamma \)-carotene and \( \beta \)-carotene (48).

The chlorosomes of both Cba. tepidum and of Cfx. aurantiacus contain isoprenoid quinone species that probably function in redox-dependent fluorescence quenching (11, 12). In Cfx. aurantiacus the quinone is menaquinone-10, while in Cba. tepidum the major quinone in chlorosomes is the unique quinone, chlorobiumquinone (12, 18), although some menaquinone-7 is also occurs (13, 18).

All chlorosomes are surrounded by a protein-stabilized, monolayer lipid envelope membrane, and they contain wax esters, carotenoids, and quinones in addition to BChl \( c/d/e \) and BChl \( a \) (18). The monolipid membrane in the chlorosomes of Cba. tepidum consists of the glycolipids rhamnosylgalactosyldiacylglycerol (RGDG) and monogalactosyldiacylglycerol (MGDG), which together comprise 55% of the total polar lipids, and of the phospholipids (di)phosphatidylglycerol (PG and DPG) and phosphatidylethanolamine (PE), which comprise about 30% of the total polar lipids (55). In Cfx. aurantiacus, the glycolipid fraction consists mainly of MGDG and digalactosyl diacylglyceride and the phospholipid fraction consists of PG and phosphatidyl inositol (PI) (27). Interestingly, the phospholipid aminoglycosphingolipid (AGS) is found in Cba. tepidum and other GSB, but not in the FAPs, and seems to be co-localized with the Fenna-Matthew-Olson protein (FMO) (55). In addition to these polar lipids, the chlorosomes of some green
bacteria, in particular those strains isolated from thermophilic environments, also contain wax esters that may play a role in membrane or BChl c stabilization (18).

The roles and organization of the proteins in this membrane have been thoroughly studied in the chlorosomes of the GSB, *Cba. tepidum*. In this organism, ten different chlorosome proteins designated CsmA, CsmB, CsmC, CsmD, CsmE, CsmF, CsmH, CsmI, CsmJ and CsmX, have been identified (8-10, 66, 67). In a genetic study, nine mutants, each lacking one of the proteins embedded in the chlorosome envelope, were generated, shown to be viable, and were still able to synthesize chlorosomes (17). CsmA was the only essential protein; no viable mutant could be generated in which this protein was missing. CsmA is the most abundant of all chlorosome proteins (about ~2,700 molecules per chlorosome out of ~5,000 total protein molecules). It is a small polypeptide (about 5.7-6.2 kDa.), which binds one BChl a molecule and 1-2 carotenoids, and whose C-terminus is post-translationally cleaved during chlorosome biogenesis (18, 44). Cross-linking studies have shown that CsmA subunits form a paracrystalline layer, making the baseplate of the chlorosome, which is attached to the FMO layer that contacts the reaction centers of GSB (30, 31, 47). This location facilitates its role in transferring excitation energy from the chlorosomes to reaction centers through the BChl a–binding protein, FMO. CsmE is closely similar in sequence to CsmA, is also post-translationally cleaved and located in/near the baseplate, but CsmE is present in much lower amounts than CsmA (30).

CsmI, CsmJ and CsmX share homology with adrenodoxin-type ferredoxins and contain [2Fe-2S] clusters (18, 30, 66). Fluorescence emission studies of wild
type Cba. tepidum cells and csmI csmJ mutants have shown that these proteins are likely to play a role in redox quenching during exposure to oxic conditions, probably by transferring electrons to and from a quencher in the chlorosomes (30). CsmX is present at very low concentrations in chlorosomes, and it does not appear to play a significant role in redox quenching of energy transfer (30).

Other chlorosome proteins can be grouped into two families, CsmC/D/H and CsmB/F/H based on sequence similarity. Chlorosomes from mutant strains that in which all of the genes from either one of these two families were inactivated produced chlorosomes that differed from wild-type chlorosomes in total pigment content and overall shape. Based on these studies, these proteins are thought to influence the overall shape of chlorosomes, may influence carotenoid distribution/import into chlorosomes, and finally may determine the orientation of the BChl c monomer stacks relative to the long axis of the chlorosomes (32).

Although chlorosomes of a FAP, Cfx. aurantiacus, have a protein composition distinctly different from Cba. tepidum chlorosomes, most of its proteins can be assigned to one of the four families found in Cba. tepidum (18). Cfx. aurantiacus chlorosomes contain CsmA in the baseplate (18). CsmY is a [2Fe-2S] protein related to the CsmI/I/X family, and CsmM and CsmN are distantly related to CsmC/D, while CsmO is a member of the CsmB/F family. The only chlorosome protein in Cfx. aurantiacus without a direct homolog in Cba. tepidum is CsmP, which seems to be related to proteins found in Synechocystis sp. PCC 6803, Synechococcus sp. PCC 7002, and Halobacterium spp. (18).
The acidobacterium Cab. thermophilum is the first known organism to synthesize chlorosomes aerobically (6). Moreover, it is also the first chlorosome-synthesizing organism that does not belong to either the GSB or the FAPs. Because of these unique properties, I isolated and analysed chlorosomes from Cab. thermophilum with the goal of investigating whether they would have unusual properties that would enhance light harvesting under oxic conditions. This chapter presents these analyses and shows that, although Cab. thermophilum chlorosomes do have some novel properties, their properties are similar to the properties of chlorosomes found in members of both the *Chlorobiales* and the *Chloroflexi*.

4.2 METHODS

4.2.1 Culture conditions. Cab. thermophilum was cultured in a HEPES-buffered, minimal salts medium supplemented with reduced organic carbon sources and ammonia as N-source as previously described (6, Chapter 2). Cultures were incubated at 53°C with mild shaking and harvested after 5 days, when BChl c absorption was maximal. *Cba. tepidum* and *Cfx. aurantiacus* Y-396 were cultured under anoxic conditions as previously described (14, 34). *Escherichia coli* was grown anoxically in LB medium in screw-capped test tubes with no air space.

4.2.2 Chlorosome isolation. Chlorosomes from *Cba. tepidum* were isolated as described by Vassilieva et al. (67). For the isolation of chlorosomes from Cab. thermophilum and *Cfx. aurantiacus* Y-396, cells were harvested in chlorosome isolation buffer (CIB) containing 2.0 M NaSCN, 10 mM Tris-HCl, pH 8, 5 mM EDTA, 1.0 mM PMSF and 2.0 mM dithiothreitol (DTT), incubated with lysozyme.
(3 mg mL⁻¹) for 20 min and subjected to 3 passes through a chilled (~4 °C) French pressure cell at 138 MPa. Unbroken cells and debris were pelleted by low-speed centrifugation (2,000 × g), and the resulting supernatant was collected and brought to 30% sucrose (w/v). A sucrose step-gradient was generated with 45% (w/v) sucrose in CIB, followed by the low-speed supernatant fraction, 20% (w/v) sucrose (w/v) and 5% (w/v) sucrose, both in CIB. This gradient was ultracentrifuged for 18 h at 250,000 × g. The chlorosome fraction appeared as a dark brown layer at the top of the gradient. This band was collected and diluted three-fold in chlorosome buffer (10 mM K-phosphate, 150 mM NaCl, pH 7.5); chlorosomes were then pelleted by high-speed ultracentrifugation at 220,000 × g. This last step was repeated once more, and the resulting chlorosome pellet was resuspended, aliquoted and stored at 4°C for cryoimaging and electron microscopy or stored at –80°C for other biochemical analyses.

4.2.3 Electron microscopy. Isolated chlorosomes were negatively stained with 1% (w/v) uranyl acetate and visualized in a JEOL 1200 EXII Transmission electron microscope (Peabody, MA).

4.2.4 Cryoimaging. Cryoimaging and tomographic reconstructions were generated by Dr. Gert Oostergetel in the laboratory of Dr Egbert Boekema at, University of Groningen, The Netherlands.

4.2.5 HPLC and mass spectrometric analyses of bacteriochlorophylls. Pigments were extracted from cell and chlorosome samples by sonication in methanol:acetone (7:2 v/v). Cell debris was removed by centrifugation and the resulting supernatant was filtered through a 0.2-μm filter. Reversed-phase HPLC
(RP-HPLC) separation of the extracted pigments was carried out on a 25 cm \( \times \) 4.6 mm Discovery 5 \( \mu \)m C18 column (Supelco, Bellefonte, PA) as previously described \((15, 61)\). Elution of BChl \( c \) was monitored at 660 nm. For pigment purification for mass spectrometry analyses, pigments were separated on a semi-preparative 25 cm \( \times \) 10 mm Discovery 5 \( \mu \)m C18 column (Supelco, Bellefonte, PA) with a similar elution program as previously described \((15, 61)\). BChls were extracted in methanol:acetone (7:2, v/v), dried under a steady stream of N\(_2\), and resuspended in an equal volume of ether and methanol. Sterile deionized water was then added drop-wise until the ether phase separated. The ether phase containing the BChl was collected, dried under a steady stream of N\(_2\), resuspended in methanol:acetone (7:2, v/v), filtered through a 0.2-\( \mu \)m filter, and loaded onto the semi-preparative column for separation. Mass spectrometry measurements were carried out at the Mass Spectrometry Facility, Huck Institutes of the Life Sciences, at The Pennsylvania State University, University Park. To elucidate the different species of BChl \( c \) homologs, RP-HPLC was coupled with MS/MS as described \((2)\). Formic acid, which demetallated the BChls and increased detection sensitivity in the MS/MS analyses, was added post-column prior to the MS/MS analyses.

**4.2.6 Quinone and carotenoid analyses.** Quinones were extracted from cells and chlorosome fractions by repeated sonication in methanol:acetone (7:2). Cell debris was removed by centrifugation, and the resulting supernatant was dried under a stream of N\(_2\). The extract was resuspended in methanol and an equal volume of hexane was added. After mixing, an equal volume of 1.0 M NaCl was added, and the extract was shaken and centrifuged. BChl molecules precipitated between the
aqueous and organic phases; the hexane phase containing quinones was collected, dried under a stream of N₂, resuspended in methanol:acetone (7:2), and filtered through a 0.2-μm filter. RP-HPLC analysis of these extracted quinones was performed as described above; quinones were detected by their absorption at 270 nm.

For pigment purification for mass spectrometric analyses, pigments and quinones were separated on a semipreparative 25 cm × 10 mm Discovery 5 μm C18 column (Supelco, Bellefonte, PA) with a similar elution program as previously described (6, 15). Quinones and carotenoids were extracted in a methanol:acetone (7:2) solvent followed by hexane/NaCl extraction as described above. Mass spectrometric measurements were performed at the Mass Spectrometry Facility, Huck Institutes of the Life Sciences, at The Pennsylvania State University (University Park, PA).

4.2.7 Lipid analyses. Lipid analyses were conducted in the laboratory of Dr. Stefan Schouten in the Royal Netherlands Institute for Sea Research, The Netherlands.

4.2.8 Polyacrylamide gel electrophoresis. Chlorosome proteins were analyzed by electrophoresis on Tris-Tricine-buffered polyacrylamide gels in the presence of sodium dodecylsulfate (SDS) (53). The stacking gel was 4% T and 2.6% C and the resolving gel was 16% T and 3.3% C. Briefly, samples containing ca. 10 μg of BChl c were incubated at 65 °C in 1× loading buffer (0.1 M Tris-HCl buffer, pH 6.8, 24% (v/v) glycerol, 1% (w/v) SDS, 2% (v/v) mercaptoethanol, 0.02% (w/v) Coomassie blue for about 15 min and electrophoresed for 18 h at constant
amperage. Proteins were visualized through silver staining (4) or Coomassie blue staining.

4.2.9 Amino terminal sequencing and mass spectrometry analyses of chlorosome proteins. Chlorosome proteins separated by SDS-PAGE were transferred to a PDVF membrane in a BioRad Semidry Electroblot apparatus using a 25 mM Tris, 192 mM glycine 20% (v/v) methanol and 0.1% (w/v) SDS transfer buffer according to manufacturer instructions. Following electro-transfer, blots were rinsed with distilled water and transferred proteins were visualized with Ponceau stain (0.1% (w/v) Ponceau S in 5% (v/v) acetic acid). Bands were excised and sequenced at The Pennsylvania State University Protein Sequencing Facility, Hershey, PA. For tryptic peptide mass-fingerprinting analyses, SDS-PAGE gels were stained with Coomassie blue, and visualized bands were excised and analyzed at The Pennsylvania State University Protein Sequencing Facility, Hershey, PA.

4.2.10 Absorbance and Fluorescence emission spectroscopy. Room temperature absorbance spectra were measured with a GENESYS 10 spectrophotometer (Thermo Electron Corp., Rochester, NY). Fluorescence emission were measured with an SLM-Aminco Model 8000C spectrofluorometer, modified by On-Line Systems, Inc. for computerized data acquisition, and equipped with a liquid nitrogen Dewar for acquisition of spectra at 77 K. Data were analyzed using Igor-Pro (Wavemetrics, Inc. Lake Oswego, OR). For fluorescence emission measurements, cell and chlorosome fractions were thawed and diluted to a BChl c absorbance $Q_y = 0.1$ in 10 mM Tris-HCl buffer, pH 7.2. Except for measurements taken under oxidizing conditions, sodium dithionite (final concentration 25 mM)
was added and samples were incubated in the dark for ca. 2 h prior to measurements.

4.3. RESULTS

4.3.1 Chlorosome isolation and morphology. A preparation of chlorosomes was isolated by using methods similar to those used to isolate chlorosomes from *Cba. tepidum* (67), in which chlorosomes are released from the cell membrane by the addition of the strong chaotrope, 2.0 M NaSCN. An absorption spectrum of this preparation had maxima for BChl *c* at 461 nm and 743 nm, and a weak absorption could be seen at ~795 nm. (figure 4.1). Electron micrographs of the isolated, negatively stained chlorosomes revealed that they are rod-shaped ellipsoids. *Cab. thermophilum* chlorosomes exhibited more variation in length, with an average of 99.6 ± 19 nm long (range= 65 to 129 nm), than in diameter, 30.8 ±2.0 nm wide (range= 24 to 32 nm) (figure 4.2). Close inspection of the micrographs suggested that one surface was flattened, presumably representing the surface that includes the baseplate protein CsmA, which attaches chlorosomes to the FMO protein layer on the cell membrane. The other surface exhibited irregularly spaced ridges and grooves; this surface seems unlikely to be involved in attachment to FMO

4.3.2 Bacteriochlorophyll *c* internal arrangement. Collaborators Drs. Oostergetel and Boekema produced a high-resolution 3-D reconstruction model of a *Cab. thermophilum* chlorosome surface from a tilt-series of electron micrographs of negatively stained chlorosomes (figure 4.3A). Cryo-electron microscopy of unstained chlorosomes embedded in vitreous ice was also performed (Figure 4.3B).
The microscopic observations suggested that these chlorosomes are constructed from roughly ellipsoidal but irregularly spaced domains; the long axis of these ellipsoidal domains occurred at approx. 45° to the long axis of the chlorosome. These domains varied in size (ca. 15-20 nm × 25-30 nm), and occasionally seemed to be divided into two subdomains (figure 4.3). In cryo-EM images, undulating, arc-like layers with a spacing of ~2.1 to 2.3 nm were observed (see arrows in figure 4.3B); the observed spacing is very similar to the spacing of BCHl c layers in the chlorosomes of *Cba. tepidum* (19, 42). However, no evidence for the occurrence of concentric nanotubes of the type seen in *Cba. tepidum* was observed by cryo-electron microscopy. The lamellae appeared to be arranged in irregular, undulating arcs (figure 4.3), but the precise organization of the BCHl molecules in the supramolecular structure is currently unknown.

**4.3.3 Bacteriochlorophyll c homologs.** RP-HPLC analyses of BCHls extracted from whole cells and isolated chlorosomes revealed a complex mixture of BCHl homologs with the most prominent peak eluting at about 46.5 min (figure 4.4). The mass of this peak, as well as its MS/MS fragmentation pattern, identified this peak as [8-iBu, 12-Et]-BChl c esterified with the unbranched C-18 alcohol, octadecanol (stearol) (figure 4.5). The MS/MS spectrum showed a diagnostic ion peak at 630.4 a.m.u., corresponding to bacteriochlorophyllide c methylated with isobutyl and ethyl moieties at the C-8 and C-12 positions, and a fragmentation pattern characteristic of BCHl c (2) with ion peaks at 613 representing the loss of water at the C-3\(^1\) position and a peak at 585 representing the loss of a carbonyl group at the C-13\(^1\) position (figure 4.6). In addition to this main peak, LC-MS/MS analyses
identified several minor peaks, which were shown to be BChl c esterified with other alcohols, including the isoprenoids farnesol and geranylgeraniol (figure 4.4, table 4.1). These same analyses identified the mass of adjacent peaks as differing by 14 mass units, i.e., the mass of a methyl group (table 4.1). This is as expected for Bchl c homologs with different degrees of alkylation at C-8 and C-12. The genome of Cab. thermophilum encodes homologs of bchQ and bchR (Chapter 3), which in Cba. tepidum have been shown to encode the radical type, SAM-dependent methyltransferases that methylate the C-8 and C-12 positions, respectively (23). HPLC analyses of BChls extracted from cells grown under various light intensities showed that with decreasing light intensities the proportion of homologs esterified with isoprenoid tails (i.e., geranylgeraniol) increases (figure 4.7).

4.3.4 Bacteriochlorophyll a and Chlorophyll a analyses. Chl a and BChl a were also extracted from Cab. thermophilum cells. Chl a could be detected by its absorption maximum at 667 nm as well as its spectrum; and it co-eluted with Chl a\textsubscript{PD} from Cba. tepidum, indicating that it is esterified with Δ2,6-phytadienol and not phytol as in Synechococcus sp. PCC 7002. BChl a was detected by its absorption maximum at 770 nm and its characteristic absorption spectrum; it co-eluted with authentic BChl a\textsubscript{P} esterified with phytol from Cba. tepidum (figure 4.8). These data are in agreement with the genes found in the genome and with previous pigment analyses (6).

4.3.5 Carotenoids. RP-HPLC analyses of carotenoids extracted from whole cells and isolated chlorosomes revealed a prominent peak that elutes at 45.3 min, which has been identified as echinenone (figure 4.9). In addition to having the single
broad absorption band characteristic of ketocarotenoids with a maximum at 465 nm (figure 4.9), it coeluted with an authentic echinenone standard from *Synechococcus* sp. PCC 7002, and its mass (551 a.m.u.) corresponds to the mass of this carotenoid (table 4.2). In addition, the carotenoids lycopene (52.2 min), γ-carotene (53.7 min), β-carotene (55.3 min.), and canthaxanthin (30 min, 564 a.m.u.) were identified through coelution with known standards and/or by mass spectrometric analyses (figure 4.9).

4.3.6 Menaquinone. The redox-dependent fluorescence emission behavior of *Cab.* thermophilum and its chlorosomes (see below) suggested that a quenching agent was present in the chlorosomes. RP-HPLC analyses of quinones extracted from whole cells and isolated chlorosomes revealed a late-eluting (i.e., hydrophobic) quinone (61 min) with absorption properties similar to those of menaquinone (figure 4.10). This hydrophobic menaquinone eluted after an authentic menaquinone-8 standard, which was isolated from *E. coli* cells grown under anoxic conditions. However, this quinone differed by two mass units from the menaquinone-8 standard (table 4.2). This discrepancy might be the result of the menaquinone in *Cab.* thermophilum having a reduced double bond in one of its eight isoprenoid units. Although the physiological role of this reduction is not clear, menaquinones with reduced double bonds in one of the isoprenoid units have been found in many organisms, most notably in soil organisms (21, 26). An additional minor peak that eluted at 18 minutes and that had an absorption spectrum characteristic of quinones was also detected but not further characterized (figure 4.10).
4.3.7 Lipids. Fatty acid analyses revealed that Cab. thermophilum mainly synthesizes isoC15 and isoC16 fatty acids and, surprisingly, also small amounts of a C18 \( n \)-alkane (figure 4.11). The polar lipid fractions were mainly comprised of a betaine lipid, diacylglycerylhydroxymethyltrimethyl-β-alanine (DGTA), and the phospholipids phosphatidylethanolamine and phosphatidylethanolmethylamine (figure 4.12). Additional minor polar lipids were detected, and one of them was tentatively identified as phosphatidylcholine.

4.3.8 SDS-PAGE and protein alignment. SDS-PAGE analyses of the chlorosomes of Cab. thermophilum revealed the presence of four highly abundant, low molecular weight proteins (<15 kDa) and four proteins of lower abundance in the 20 to 25 kDa range (figure 4.13). Amino-terminal sequencing of the four highly abundant proteins identified them as CsmA (8.1 kDa) and three other proteins, which were designated as CsmR, CsmS and CsmT (table 4.3), none of which had significant homology to other known chlorosome proteins or to any protein sequences in the Genbank database. Two of these proteins, CsmS and CsmT, are of similar size (table 4.3) and share 44% amino acid identity (figure 4.14). Moreover, they are predicted to possess two transmembrane domains (figure 4.15). Interestingly the isoelectric point of these two peptides differs by 4 units (table 4.3).

Tryptic-peptide mass fingerprinting of individual bands from a Coomassie-blue-stained polyacrylamide gel confirmed these results and identified one of the low-abundance proteins as a homolog of CsmI, an adrenodoxin-like [2Fe-2S] protein also present in the chlorosome envelopes of Chlorobiales and Chloroflexi (table 4.3).
An alignment of the CsmA protein sequences from Cab. thermophilum and green bacteria from both the Chlorobiales and Chloroflexi showed that CsmA from Cab. thermophilum contains the completely conserved, critical histidine residue that is believed to bind BChl a (figure 4.15). This alignment also showed that the CsmA sequence from Cab. thermophilum is somewhat more similar to those of Chloroflexi than to those of Chlorobiales. A similar homology is found in the proteins with unknown function encoded by the genes that surround the csmA gene (figure 4.17).

4.3.9 Spectroscopy. The absorption spectrum of isolated chlorosomes exhibits maxima at 462 nm (Soret peak) and 743 nm (Q\textsubscript{y} peak) (figure 4.18). The Q\textsubscript{y} absorption band is narrow and resembles the Q\textsubscript{y} absorption spectra of BChl c aggregates in chlorosomes of Chloroflexi. In whole cells, the absorbance spectrum had maxima at 462 nm and 744 nm (figure 4.18), although these peak values varied by 1 to 2 nm depending on the physiological state of the cells.

The fluorescence emission spectrum of isolated chlorosomes exhibited two well separated peaks, with maxima at 752 nm and 796 nm, that correspond to BChl c and BChl a associated with CsmA, respectively (figure 4.19). In agreement with the absorption spectra, the fluorescence emission spectra of Cab. thermophilum cells and chlorosomes also resembled those of the Chloroflexi. Interestingly, the fluorescence emission amplitude in Cab. thermophilum cells and chlorosomes was enhanced by a factor of 17-fold by the addition of the reducing agent sodium dithionite (figure 4.19).
4.4. DISCUSSION

The data presented in this study describe the properties of the chlorosomes of the only known chlorophototrophic *Acidobacterium*, Cab. thermophilum. In general, these structures closely resemble those found in the chlorophototrophic *Chlorobi* and the green *Chloroflexi*, and an isolation procedure based on those employed in these organisms allowed chlorosomes to be isolated from Cab. thermophilum.

Chlorosomes from Cab. thermophilum exhibit a striking morphology that suggests that internal ellipsoidal domains, or perhaps undulating arcs, occur in the suprastructure produced from by BChl c molecules. This differs from the organization of the BChls in chlorosomes of *Cba. tepidum*, which are mostly arranged in concentric nanotubes (19). Although this is the first report describing this type of BChl substructure in chlorosomes, Cab. thermophilum may not be the only chlorosome-containing phototroph with this unusual BChl arrangement. A study of the surface topography of chlorosomes isolated from representative green bacteria from both the *Chlorobiales* and the *Chloroflexi* revealed two morphologies, described as “rough” and “smooth,” of chlorosome surfaces (36). Although the authors attributed these differences to the lipid and protein composition of the chlorosome membrane, our findings lead us to propose that the “rough chlorosomes” described in that report may have had BChl molecules arranged in ellipsoidal domains similar to the ones observed in the chlorosomes of Cab. thermophilum, and that these domains are responsible for the “rough” surface topology. If this is indeed the case, organisms from all three chlorosome-
synthesizing phyla exhibit this novel suprastructure. More interestingly, ultrastructural data on chlorosomes from members of all three phyla containing green bacteria has shown that the BChl molecules of chlorosomes are capable of forming distinctly different types of suprastructures, which include both concentric nanotubes, such as the ones found in *Cba. tepidum*, and ellipsoidal domains or “undulating lamellae” in the case of *Cab. thermophilum*. This latter result seems to support the lamellar model of BChl c organization originally proposed by Psencik et al. (46), just not for the chlorosomes the model was originally intended to describe.

*Cab. thermophilum* methylates its BChls at the C-8\(^2\) and C-12\(^1\) positions like the GSB. These methylations have been shown to make *Cba. tepidum* more competitive under light-limiting situations (23), and we presume that these methylations play a similar role in the chlorosomes of *Cab. thermophilum*. Unlike *Cba. tepidum*, however, *Cab. thermophilum* does not use an isoprenoid alcohol (farnesol) to esterify its major BChl. Instead, the data indicate that the straight-chain alcohol, octadecanol (stearol) is the major alcohol esterified to BChl c. In addition to this dominant BChl c species, *Cab. thermophilum* synthesizes in varying amounts Bchl c esterified with other alcohols, including geranylgeraniol. In this aspect, *Cab. thermophilum* resembles the green sulfur bacterium *C. phaeobacteroides* or green FAPs such as *Chloronema* sp., which also display a diversity of methylation patterns and esterifying alcohols, including those with straight chains (21, 40). In both *Cab. thermophilum* and *C. phaeobacteroides*, an increased synthesis of isoprenoid alcohols has been observed under light-limiting
conditions (figure 5.5, 40). It has been suggested that the limited availability of reducing power, rather than light intensity, might be responsible for the increased production of isoprenoid chains under light-limiting conditions, because the biosynthesis of reduced, straight-chain alcohols has a higher demand for reducing power than the biosynthesis of isoprenoid alcohols (40). Alternatively, in Cab. thermophilum it could be the case that, by synthesizing a saturated alkane alcohol chain at the higher light intensities which correlate with high oxygen partial pressures in its natural habitat (50), peroxidation of the double bonds in the isoprenoid chains is avoided (63).

Although the antenna BChls comprise most of the total BCHl content in organisms with chlorosomes (>97% in Cba. tepidum (18), GSB and FAPs synthesize additional (B)Chls (16). Small amounts of BCHl a are found in FMO and the reaction centers in Chlorobiales, and in the chlorosome baseplate protein of both GSB and FAPs. In addition, Chl a is found in the reaction center of GSB (5). Cab. thermophilum synthesizes both of these pigments, and our HPLC analyses show that they are identical in structure to the pigments of Cba. tepidum. Based upon their behavior in the HPLC analyses, they should have the same esterifying tails as those found in GSB, which esterify Chl a with Δ2,6-phytadienol and BCHl a with phytol. This similarity is not surprising, given that, as GSB (18, 28), Cab. thermophilum synthesizes the FMO protein (62) and is predicted to have a Type 1 reaction center.

The chlorosomes of Cab. thermophilum have a unique carotenoid composition. Whereas the chlorosomes of Chlorobiales mostly contain
chlorobactene or isorenieratene and the green FAPS synthesize γ-carotene as their major carotenoid, the dominant carotenoid in Cab. thermophilum is echinenone. Traces of lycopene, γ-carotene, β-carotene and canthaxanthin were also found. Both echinenone and canthaxanthin are xanthophylls and possess keto-groups on their rings. Xanthophyll carotenoids have been shown to play roles in photoprotection in other microorganisms (58). In particular, echinenone might be involved in protection against peroxide radicals (49). Because of the high light intensities and the oxic environment that Cab. thermophilum encounters in its natural habitat (50), these ketocarotenoids probably also play a photoprotective role in this microorganism. Cells grown in the presence of the lycopene cyclase inhibitor \(N,N\)-diethyl-\(N\)-[2-(4-methylphenoxy)ethyl]amine (MPTA) exhibit increased sensitivity to photodegradation (Appendix C).

Although Cab. thermophilum is an aerobe, the fluorescence emission of excited BCHls is redox-dependent and is severely quenched under oxic conditions (see below). In other organisms that exhibit this behavior, the quencher is a quinone (11-13, 61), and HPLC and mass spectrometric analyses revealed the presence of menaquinone-8 in the chlorosomes of Cab. thermophilum. Chlorobiumquinone, which is present in chlorosomes of \(Cba.\ tepidum\) (13) was not present in the chlorosomes of Cab. thermophilum, and other hydroxylated quinones, which have been postulated to be responsible for fluorescence quenching in \(Cfx.\ aurantiacus\) when added exogenously (61), were likewise not present. However, an additional unidentified, putative early-eluting quinone was present in small amounts.
The lipid composition of Cab. thermophilum cells and chlorosomes seems to reflect the alkaline hot spring environment from which this organism was isolated and is unique among chlorosome-containing chlorophototrophs. The major lipid, DGTA, in this organism is a betaine lipid, which contains nitrogen instead of phosphorus. Phosphate levels in Octopus Spring are naturally low, and chlorophototrophs in the mat community of this spring have adaptations to the low phosphate concentration (1). Although the phospholipid fraction is quite similar to that of chlorosomes from other green bacteria, the glycolipids typically found in the membranes of other chlorosomes, such as MGDG, were not present in the chlorosomes of Cab. thermophilum. The fatty acid fraction was also unique and included small amounts of a C18 n-alkane (50).

The lipid envelope of the chlorosomes of Cab. thermophilum contains a protein that belongs to the CsmA protein family. CsmA, the BChl a-binding, chlorosome baseplate protein, is the only chlorosome protein that is essential for viability in Cba. tepidum (and presumably in all other chlorosome-containing microorganisms), and it is the most abundant of all chlorosome proteins (16, 17). It is typically a small polypeptide (about 5.7-6.2 kDa) and is post-translationally processed by proteolytic cleavage of its C-terminus (8). The data show that CsmA in the chlorosomes of Cab. thermophilum has a similar molecular mass and probably plays a similar role as in other chlorosome-containing microorganisms. It contains the critical histidine residue for binding BChl a and, similar to other chlorosomes, it is the most abundant protein in the chlorosomes as seen in SDS-PAGE analyses (figure 4.2). However, it differs in that most of the protein does not
seem to be post-translationally modified. CsmA from Cab. thermophilum is most closely related in sequence to CsmA from *Cfx. aurantiacus* and *Cfx. aggregans* (figure 4.4). ORFs adjacent to the *csmA* gene in *Cab. thermophilum* are also most closely related to ORFs adjacent to *csmA* of *Cfx. aggregans*. These ORFs of unknown function flank the *csmA* genes in all other chlorosome-containing microorganisms, and in *Cba. tepidum* (and presumably other green bacteria) are essential for viability (17). The functions of these proteins are still unknown, however.

The chlorosomes of *Cab. thermophilum* also contained an Fe/S protein that is distantly related to CsmI in the green sulfur bacterium *Chloroherpeton thalassium*. This protein is predicted to contain an adrenodoxin-like [2Fe-2S] domain, and it is similar in size to other [2Fe-2S] proteins found in various chlorosomes. Chlorosomes from *Cab. thermophilum* have an EPR properties consistent with the presence of a [2Fe2S] cluster (S. Romberger, personal communication). It seems reasonable to hypothesize that this protein has a role in reduction of the oxidized quencher as is the case in *Cba. tepidum* (30). In addition to these CsmA and CsmI homologs, the chlorosomes of Cab. thermophilum contain three additional, abundant proteins, CsmR, CsmS and CsmT, which show no homology to other known chlorosome proteins, nor do they exhibit any particular sequence motif that might provide clues about their function. Two of them, CsmR and CsmS, are highly similar to each other in size and sequence, and undoubtedly these proteins arose through a gene duplication event. In addition, these two polypeptides are predicted to have two transmembrane α-helical domains and are
the first known chlorosome proteins predicted to have this property. Although these domains were predicted based on algorithms designed for a bilipid membrane, rather than the monolipid chlorosomal membrane, no other chlorosomal protein from other green bacteria has been predicted to contain such $\alpha$-helical domains. The difference in isoelectric point in these two peptides might be indicative of different functions or might indicate that these proteins form heterodimers through complementary charge compensation. Further work on these proteins will help elucidate their topological arrangement as well as their roles in the chlorosome envelope.

Cab. thermophilum chlorosomes have a narrow $Q_y$ absorption band that is indicative of similar packing of the BCHls in these chlorosomes to those of green FAPs, which also exhibit narrow $Q_y$ bands. Accordingly, the fluorescence emission spectrum of Cab. thermophilum chlorosomes is also more similar to that of Cfx. aurantiacus chlorosomes than to the spectra of chlorosomes from Cba. tepidum. Unlike Cfx. aurantiacus, however, both in whole cells and in isolated chlorosomes of Cab. thermophilum, fluorescence emission was enhanced by a factor of 17 by the addition of the reducing agent, sodium dithionite, showing that a redox-dependent quenching of energy transfer occurs in these chlorosomes. Cab. thermophilum is an aerobic microorganism that produces chlorosomes in the presence of oxygen. However, the fluorescence emission results imply that Cab. thermophilum chlorosomes do not transfer light energy efficiently under oxidizing conditions. These data clearly show that in this respect the chlorosomes of Cab. thermophilum resemble those of the strict anaerobe, Cba. tepidum. These chlorosomes also exhibit
enhanced fluorescence emission under reducing conditions after the addition of sodium dithionite. This redox-dependent fluorescence emission suggested that a quencher must be present in the chlorosomes, and this was tentatively identified as menaquinone-8.

The combined biochemical analyses of isolated chlorosomes from Cab. thermophilum presented in this chapter show that, although this is an aerobic organism, its chlorosomes are remarkably similar to those of other green bacteria, all of which produce and utilize these organelles under anoxic conditions. These chlorosomes from an aerobe share properties with those of the green sulfur bacteria (Chlorobiales) and the green FAPs (Chloroflexi), such as methylation patterns of the BChl c homologs and spectroscopic properties. However, their most significant adaptation to oxic conditions appears to be the presence of large amounts of ketocarotenoids. Although Cab. thermophilum is an aerobe, the redox dependence of the fluorescence emission data challenge the premise that these structures are used for light harvesting under highly oxic conditions. Recent in situ metatranscriptomic studies of samples from the phototrophic mats of Mushroom Spring, where Cab. thermophilum is found in nature, that were collected at different times of the day have revealed that, in fact, Cab. thermophilum is most transcriptionally actively at dusk, when the light intensity and oxygen partial pressures are low (Liu et al., in preparation). This finding is in agreement with the present studies of the chlorosomes— an antenna complex that can adapt and protect itself against excessive illumination and oxidation but that absorbs and transfers light energy at low illumination intensity or under reduced oxygen conditions.
4.5 REFERENCES


FIGURE LEGENDS

Figure 4.1. Absorption spectrum of isolated chlorosomes from Candidatus Chloracidobacterium thermophilum.

Figure 4.2. Electron micrograph of negatively stained isolated chlorosomes. Chlorosomes were stained with 1% (w/v) uranyl acetate. Magnification = 20,000X. Note the flat surface visible on some of the chlorosomes that is presumably the baseplate by which chlorosomes are attached to FMO layer associated with the cell membrane.

Figure 4.3. Tomographic reconstruction and cryo-electron microscopic images the chlorosomes of Candidatus Chloracidobacterium thermophilum. A. Stereo image of a negatively stained chlorosome as an iso-surface representation from a tomographic reconstruction. B. Cryo-electron microscopic image of isolated chlorosomes embedded in vitreous ice. Note the undulating lamellae. The arrows indicate the ~2.2-nm spacings between adjacent lamellae.

Figure 4.4. RP-HPLC elution profile of BCHls extracted from Candidatus Chloracidobacterium thermophilum. Profile was obtained by monitoring at 667 nm. See text for further details and table 4.1 for peak assignments.

Figure 4.5. Structure of the major species of BCHl c found in the chlorosomes of Cab. thermophilum: [8-iBu, 12-Et]-BChl c esterified with octadecanol.

Figure 4.6. MS/MS profile of the major species of BCHl c found in the chlorosomes of Cab. thermophilum. See text for details.

Figure 4.7. RP-HPLC elution profile of BCHls extracted from Candidatus Chloracidobacterium thermophilum cells that were grown under different light intensities. Top, 120 µmol photons m⁻² s⁻¹; Middle: 60 µmol photons m⁻² s⁻¹; Bottom: 30 µmol photons m⁻² s⁻¹. Please refer to figure 4.4 and table 4.1 for identity of the peaks.

Figure 4.8. RP-HPLC elution profile and absorbance spectrum of Chl a and BCHl a. A. Elution profile monitored at 667 nm. The elution peaks corresponding to BCHl a and Chl a are indicated. B. Absorbance spectrum of BCHl a. C. Absorbance spectrum of Chl a.

Figure 4.9. Carotenoids in Candidatus Chloracidobacterium thermophilum. A. HPLC elution profile monitored at 491 nm of carotenoids extracted from Cab. thermophilum. Peaks 1, 5-8 represent canthaxanthin (1), echinenone (5), lycopene (6), γ-carotene (7) and β-carotene (8). Peaks 2, 3 and 5 represent unidentified
carotenoids. B. Absorption spectrum of peak 1 (canthaxanthin). C. Absorption spectrum of peak 5 (echinenone).

**Figure 4.10. Quinones in Candidatus Chloracidobacterium thermophilum.** A. HPLC elution profile monitored at 297 nm of quinones extracted from Cab. thermophilum. Peaks 1 and 5 represent an unidentified quinone and menaquinone-8(H2), respectively. Peaks 2, 3, and 4 were identified as carotenoids by their absorption spectra. Other minor peaks represent BCHls or bacteriopheophytins. B. Absorption spectrum of peak 5 (menaquinone-8(H2)). Insert, absorption spectra of menaquinone 8 extracted from *E. coli*. C. Absorption spectra of peak 1.

**Figure 4.11. GC elution profile of fatty acids extracted from Candidatus Chloracidobacterium thermophilum chlorosomes.**

**Figure 4.12. HPLC elution profile of total lipids extracted from Candidatus Chloracidobacterium. thermophilum.**

**Figure 4.13. SDS-PAGE analyses showing the protein composition of isolated chlorosomes from Chlorobium tepidum, Candidatus Chloracidobacterium. thermophilum and Choroflexus aurantiacus Y-396.** Proteins were visualized by silver staining as described in the methods section of this chapter.

**Figure 4.14. Alignment of CsmS and CsmT amino acid sequences from Cab. thermophilum.**

**Figure 4.15. Hydropathy plots of CsmS, CsmT and CsmJ.** Protein sequences were analyzed by the Kyte-Doolittle scale and with a 10-residue window in the TopPred site from the Institut Pasteur (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred).

**Figure 4.16. Alignment of CsmA sequences** from Cab. thermophilum, Chl. aurantiacus, Chl. aggregans, Cba. tepidum, C. limicola, C. phaeobacteroides, Prosthecochloris aestuarii. The conserved histidine thought to ligate BChl a is indicated by the red star.

**Figure 4.17. Alignment of Cabther_17410, an uncharacterized protein, product of the gene located immediately upstream of csmA, with its homologs in Cfx. aurantiacus, Cfx. aggregans, Cba. tepidum, C. limicola, C. phaeobacteroides, Prosthecochloris aestuarii, C. vibrioforme, Pelodictyon. clathratiforme, C. ferroxidans.** Please note that C. phaeobacteroides, Cfx. aggregans and Cfx.
aurantiacus have two copies of this gene, which presumably arose by gene duplication.

**Figure 4.18. Absorption spectra of cells and chlorosomes of Candidatus Chloracidobacterium thermophilum.** A. Absorption spectra of whole cells of Cab. thermophilum (thick solid line), Cba. tepidum (dotted line) and Cfx. aurantiacus (thin solid line). B. Absorption spectra of isolated chlorosomes from Cab. thermophilum (thick solid line), Cba. tepidum (dotted line) and Cfx. aurantiacus (thin solid line).

**Figure 4.19. Fluorescence spectra chlorosomes of Candidatus Chloracidobacterium thermophilum under oxidizing and reducing conditions.** A. Fluorescence emission spectra of isolated chlorosomes from Cab. thermophilum under oxidizing (thick solid line) and reducing (thin solid line) conditions. B. Fluorescence emission spectra of isolated chlorosomes from Cab. thermophilum (thick solid line), Cba. tepidum (dotted line) and Cfx. aurantiacus (thin solid line). The excitation wavelength for these measurements was 440 nm. To produce oxic conditions, chlorosomes were diluted into air-saturated buffer, and to produce reducing conditions, 25 mM Na-dithionite was added to the chlorosomes which were incubated in the dark for ca. 2 hours.
Table 4.1. Masses of Bacteriochlorophyll c homologs

<table>
<thead>
<tr>
<th>Peak number</th>
<th>(M+H)</th>
<th>Macrocycle</th>
<th>[C8, C12 substituents]</th>
<th>Mass of alkyl fragment (Da)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>799</td>
<td>595</td>
<td>[Et, Et]</td>
<td>204</td>
<td>Farnesyl</td>
</tr>
<tr>
<td>2</td>
<td>813</td>
<td>609</td>
<td>[n-Pr, Et]</td>
<td>204</td>
<td>Farnesyl</td>
</tr>
<tr>
<td>3</td>
<td>827</td>
<td>ND</td>
<td>[i-But, Et]*</td>
<td>ND</td>
<td>Farnesyl*</td>
</tr>
<tr>
<td>4</td>
<td>853</td>
<td>ND</td>
<td>[Et, Me]*</td>
<td>ND</td>
<td>Geranylgeraniol*</td>
</tr>
<tr>
<td>5</td>
<td>867</td>
<td>595</td>
<td>[Et, Et]</td>
<td>272</td>
<td>Geranylgeraniol</td>
</tr>
<tr>
<td>6</td>
<td>881</td>
<td>609</td>
<td>[n-Pr, Et]</td>
<td>272</td>
<td>Geranylgeraniol</td>
</tr>
<tr>
<td>7</td>
<td>805</td>
<td>ND</td>
<td>[Et, Me]</td>
<td>224</td>
<td>Hexadecanol</td>
</tr>
<tr>
<td>8</td>
<td>819</td>
<td>ND</td>
<td>[i-But, Et]/[Et, Me]*</td>
<td>224/238</td>
<td>Hexadecanol/heptadecanol</td>
</tr>
<tr>
<td>9</td>
<td>817</td>
<td>581</td>
<td>581</td>
<td>236</td>
<td>Heptadecanol (1)</td>
</tr>
<tr>
<td>10</td>
<td>833</td>
<td>595</td>
<td>595</td>
<td>238</td>
<td>Heptadecanol</td>
</tr>
<tr>
<td>11</td>
<td>847</td>
<td>609</td>
<td>609</td>
<td>238</td>
<td>Heptadecanol</td>
</tr>
<tr>
<td>12</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>13</td>
<td>847</td>
<td>595</td>
<td>595</td>
<td>252</td>
<td>Octadecanol</td>
</tr>
<tr>
<td>14</td>
<td>861</td>
<td>609</td>
<td>609</td>
<td>252</td>
<td>Octadecanol</td>
</tr>
<tr>
<td>15</td>
<td>875</td>
<td>ND</td>
<td>[i-But, Et]*</td>
<td>ND</td>
<td>Octadecanol</td>
</tr>
</tbody>
</table>

*Assignments made based on full mass spectrum and adjacent homologs
Table 4.2. Mass of carotenoid and quinone species

<table>
<thead>
<tr>
<th>Compound</th>
<th>Mass (m/e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echinenone</td>
<td>551</td>
</tr>
<tr>
<td>Canthaxanthin</td>
<td>564</td>
</tr>
<tr>
<td><em>Cab. thermophilum</em> menaquinone-8(2H)</td>
<td>718</td>
</tr>
<tr>
<td><em>Escherichia coli</em> menaquinone-8</td>
<td>716</td>
</tr>
</tbody>
</table>
Table 4.3. Properties of proteins in chlorosomes of *Candidatus Chloracidobacterium thermophilum*

<table>
<thead>
<tr>
<th>Protein</th>
<th>Locus Tag</th>
<th>Mass (kDa)</th>
<th>Predicted Transmembrane domains*</th>
<th>Predicted pI*</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CsmA</td>
<td>17420</td>
<td>8.15</td>
<td>None</td>
<td>6.55</td>
<td>BChl -α binding base plate protein</td>
</tr>
<tr>
<td>CsmR</td>
<td>28310</td>
<td>9.82</td>
<td>None</td>
<td>6.72</td>
<td>Function unknown</td>
</tr>
<tr>
<td>CsmS</td>
<td>18160</td>
<td>13.07</td>
<td>2</td>
<td>8.89</td>
<td>Function unknown</td>
</tr>
<tr>
<td>CsmT</td>
<td>9770</td>
<td>13.48</td>
<td>2</td>
<td>4.72</td>
<td>Function unknown</td>
</tr>
<tr>
<td>CsmI</td>
<td>20870</td>
<td>29.34</td>
<td>1</td>
<td>7.67</td>
<td>Contain 2Fe-2S binding motif</td>
</tr>
</tbody>
</table>

*Predicted through the ExPASY Proteomics server (12)*
Figure 4.1
Figure 4.3
Figure 4.4
Figure 4.5
Figure 4.6
Figure 4.7

Absorbance at 667 nm

High light
(120 µE/ms)

Low light
(30 µE/ms)
Figure 4.8
Figure 4.9
Figure 4.10

A

B

C

Absorbance (270 nm)

Absorbance

Wavelength (nm)

Wavelength (nm)
Figure 4.11
Figure 4.13

Novel proteins (R,S,T)

A

J

Novel proteins (R,S,T)
Figure 4.14
Figure 4.15

CsmS

CsmT

CsmJ
Figure 4.16

* Probable Bchl a binding site
Figure 4.18

A

B
Figure 4.19

A

![Graph A]

B

![Graph B]
CHAPTER 5

Heterologous expression of carotenoid biosynthesis genes from Candidatus Chloracidobacterium thermophilum.

Publication:

ABSTRACT
Carotenoids are pigments characterized by a polyene backbone consisting of 9 to 13 conjugated double bonds. As isoprenoid derivatives, carotenoids are synthesized from isopentenyl pyrophosphate and dimethylallylpyrophosphate by either the mevalonate or the 1-deoxy-D-xylulose-5-phosphate (DOXP) pathways. Phytoene, the first committed precursor to all C40 carotenoids, is converted to lycopene by the product of a single gene (crtI) or by the cooperative activities of the products of three or four genes, crtP, crtQ and crtH (and sometimes mnrU/crtZ-Iso). Further modifications to the ψ ends of lycopene are partly responsible for generating the large variety of carotenoid species found in nature. These modifications include cyclizations, hydroxylations, desaturations, and glycosylations. Carotenoid biosynthesis in the chlorophototrophic acidobacterium Candidatus Chloracidobacterium thermophilum has been investigated. Three genes form this organism that are predicted to encode a lycopene cyclase, a ϕ–ring ketolase and a ψ–end hydroxylase, respectively, have been heterologously expressed in Escherichia coli, and the modifications they introduce to the ends of lycopene, γ-carotene and β-carotene investigated. Phylogenetic analyses of protein sequences derived from these genes show that they are most closely related to those found in the green chlorophototrophic bacteria and in Gram-positive soil microorganisms.
5.1. INTRODUCTION

Carotenoids are a family of isoprenoid molecules characterized by a backbone of a conjugated polyene chain with 9 to 13 double bonds and regularly spaced methyl substitutions (3). This extended conjugated system results in the ability to absorb light in the ca. 300-500 nm wavelength range. Carotenoids can be linear, monocyclic or dicyclic. In addition, the ends of carotenoids are usually modified with functional groups, including hydroxyl and keto groups, which may themselves be further modified by methylation, glycosylation, and/or acylation (3, 5). These modifications are responsible for the large diversity of carotenoids found in nature, with over 700 different types described to date (3).

Carotenoids are among the most widespread natural compounds found in nature and are synthesized by plants, algae, fungi, many microorganisms and aphids. Aphids are the first documented instance of an animal possessing the ability to synthesize carotenoids de novo (28). The roles of carotenoids are as widespread as their distribution, and range from enhancing the immune system and being the precursor to visual pigments in animals (36), to inducing phytohormones in plants (6), and to stabilizing cell membranes in some bacteria (18). In chlorophototrophic bacteria, carotenoids play roles in photoprotection and light harvesting (15). These roles are facilitated by the extended conjugated system found in these pigments. The relative low energy of the triplet state in carotenoids, with respect to the triplet state of chlorophylls or singlet state of oxygen, results in rapid quenching of these highly reactive species (1). Moreover, the relatively high energy of the excited state of carotenoids with respect to the chlorophylls effectively results in energy transfer
from carotenoids to chlorophylls (1, 9). As a result, carotenoids are found in most light-harvesting antenna systems (5).

Carotenoids are isoprenoid derivatives synthesized from the 5-carbon unit isopentenyl pyrophosphate (IPP) and its isomer derivative, dimethylallyl diphosphate (DMAPP). Two pathways in nature synthesize these precursors: the mevalonate (MVA) and 1-deoxy-D-xylulose-5-phosphate (DOXP) pathways (20). The DOXP pathway appears to be the more ancient of the two (2) and is found in chloroplasts, algae, *Cyanobacteria*, most eubacteria and some parasites (20). DOXP is derived from the condensation of a molecule of pyruvate with a molecule of glyceraldehyde 3-phosphate, with consequent decarboxylation and phosphorylation reactions to form IPP and DMPP (20). The MVA pathway is found in most eukaryotes, archaea, a few eubacteria and the cytosol and mitochondria of algae and plants (20). MVA is derived from the condensation of three acetyl-CoA molecules with consequent decarboxylation and phosphorylation steps to equally yield IPP and DMPP. The immediate precursor to carotenoids, geranylgeranyl pyrophosphate (GGPP) is generated by successive condensation reactions of IPP and DMAPP molecules generated by these two pathways.

Phytoene synthase, encoded by *crtB*, catalyzes the condensation of two GGPP molecules to form phytoene, the first intermediate in the biosynthesis of all C40 carotenoids, the most common type (16). Phytoene is a colorless, symmetrical molecule with three conjugated double bonds; it usually occurs as the *cis* isomer of the C15 and C15′ central carbon double bond.
Two alternative pathways result in the conversion of phytoene into lycopene, a linear carotenoid with 9 conjugated double bonds. In the first pathway, the product of a single gene, *crtI*, is responsible for either 3 or 4 consecutive desaturation reactions, depending on the nature of the enzyme (when only 3 desaturations take place the end product is neurosporene rather than lycopene) (25, 30). The second pathway, found in plants, *Cyanobacteria*, green sulfur bacteria (GSB) and Cab. thermophilum (17, 25, 31, chapter 3), involves three or, in some of these organisms, four enzymes coded by *crtP*, *crtQ* and *crtH* and the recently described *crtZ-ISO* (7) to carry out desaturations and isomerations. From lycopene, hundreds of different carotenoids are synthesized by modifying the ends of the molecule. Linear carotenoids, such as spirilloxanthin and spheroidenone found in the purple bacteria (34), are synthesized by adding hydroxyl groups, keto groups or both to the ψ end of lycopene. These functional groups can be further modified by methylation and glycosylation (25).

γ-carotene and β-carotene, the precursors to most monocyclic and dicyclic carotenoids, respectively, are synthesized from lycopene by lycopene cyclases. There are four known families of lycopene cyclases, and some organisms have more than one type, presumably to be able to differentially regulate the synthesis of mono and dicyclic carotenoids (18, 23-26). Like the ψ ends of lycopene, the rings of these carotenoids are often modified to generate diverse molecules, such as echinenone, with a keto group modifying a single ring, or astaxanthin, with both a keto and a hydroxyl group modifying each ring. In some carotenoids, such as chlorobactene and isorenieratene found in the green sulfur bacteria, a series of
desaturations result in aromatic rings (23). Monocyclic carotenoids may have substitutions at both ends of the molecule, both in the ring and in the linear ψ-end.

The chlorophototrophic acidobacterium Cab. thermophilum synthesizes the ketocarotenoids echinenone and canthaxanthin as its dominant carotenoids (17, chapter 5) and additionally produces myxol-like carotenoids. Although its antenna system resembles those found in the green bacteria, i.e., the chlorophototrophic \textit{Chlorobi} and the chlorosome-containing \textit{Chloroflexi}, Cab. thermophilum is unique among these organisms in the presence of ketocarotenoids in its chlorosomes. This chapter presents the results from heterologous expression in \textit{Escherichia coli} of three genes from Cab. thermophilum predicted to be involved in carotenoid biosynthesis. These genes are predicted to be responsible for introducing rings in lycopene and modifying intermediates through the addition of hydroxyl and keto groups (see figure 5.1).

\textbf{5.2. METHODS}

\textbf{5.2.1. Molecular manipulations and cloning.} Cab. thermophilum genomic DNA was isolated as previously described (Chapter 3). Bioinformatic analyses identified putative carotenoid biosynthesis genes (Chapter 3). BLASTP searches using amino acid sequences of carotenoid biosynthesis genes from the cyanobacterium \textit{Synechococcus} sp. PCC 7002 and the green sulfur bacterium \textit{Chlorobaculum tepidum} as queries identified additional open reading frames (ORFs) that might be involved in the synthesis of these compounds. Amplicons of these ORFs were generated using Phusion™ DNA polymerase (Finnzymes USA, Woburn, MA), following manufacturer instructions and ligated into either the NcoI-BamHI or the
NdeI-XhoI sites (see primer sequences in Table 5.1) in the vector pCOLA-Duet from Novagen (Madison, WI). Ligated plasmids were cloned into chemically competent DH10B *E. coli* strains, and transformants were selected by plating in LB medium supplemented with kanamycin (30-50 µg mL⁻¹). Positive transformants were identified by PCR amplification of the cloned gene, and one was selected for further verification by DNA sequencing. Plasmids from this transformant were isolated and used to transform chemically competent BL21(DE3) *E. coli* cells that contained either pACYC (10), pACYC and p16_CPL1 (24) or pBC01 (Julia Maresca) (see table 5.2 for plasmid properties).

5.2.2. *In vitro* assay of CrtO activity. CrtO activity was assayed in MOPS buffer, pH 8.0, containing cell extracts (1/5 total reaction volume) from an *E. coli* culture containing the pCOLA::crtO plasmid, and β-carotene (1/5 total reaction volume) that had been extracted from a culture of *E. coli* cells containing the pBC01 plasmid (see below for culture conditions). Control reactions containing no cell extract or β-carotene were also analyzed. The reaction mixtures were incubated in a shaker at 37 °C overnight, and the reactions were stopped by the extraction of carotenoids with diethyl ether.

5.2.3. Protein induction conditions. In initial protein expression experiments, *E. coli* was cultivated in M9 media at 28 °C and IPTG (1.0 mM) was added when cultures reached an OD₆₀₀ nm of ~0.6. In these experiments cells were harvested when the culture reached an optical density (OD₆₀₀ nm) of ~1.2. In general, however, the addition of IPTG was not required for protein expression, and cultures were routinely grown in 10-100 mL volumes in LB supplemented with the appropriate
antibiotics at 30 °C to 37 °C overnight from glycerol stocks. Arabinose (0.001% w/v) was added to the culture media for the expression of β-carotene biosynthesis genes when the plasmid pBC01 was used.

5.2.4. Pigment extraction and analysis. Pigments were extracted from *E. coli* cells by resuspending pellets in 7:2 acetone:methanol and incubating on ice in the dark for ca. 10 min. Following this incubation, whole cells and debris were pelleted by centrifugation, and the supernatant was filtered through a 0.2 µm syringe and injected into a 25 cm x 4.6 mm Discovery 5 µm C18 column (Supelco, Bellefonte, PA) as previously described (24).

5.2.5. Phylogenetic analyses. Phylogenetic trees were constructed using PHYLIP software (12) or ClustalW from protein alignments generated using ClustalW and manually modified with BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Protein sequences were obtained from NCBI or IMG. Phylogenetic trees were visualized with TreeView (29).

5.3. RESULTS

5.3.1. Expression of *crtYc-crtYd* in lycopene-producing *Escherichia coli*. *E. coli* cells harboring the pACLYC and pCOLA::crtYc-crtYd plasmids and grown in M9 media at 28 °C, and after IPTG induction produced γ- and β-carotene as shown in figures 5.2 and 5.3). Interestingly, different *E. coli* growth conditions and overexpression protocols generated different pigment patterns. When these cells were incubated overnight at 37 °C in LB broth without IPTG addition, no β-carotene was observed (figures 5.2 and 5.3). Under both conditions γ- and β-
carotene production was dependent on the presence of the \textit{crtYc-crtYd} genes. \textit{E. coli} cells that contained plasmids pACLYC and pCOLA-Duet with no insert produced lycopene but no other carotenoids (figure 5.2).

\textbf{5.3.2. Expression of \textit{crtC} in lycopene producing \textit{Escherichia coli}.} \textit{E. coli} cells containing the pACLYC and pCOLA::\textit{crtC} plasmids converted all of the lycopene to 1-hydroxylycopene and 1,1’-dihydroxylycopene as evidenced by HPLC analyses of carotenoids extracted from these cells (figure 5.4), and in-line absorption spectra of the eluting compounds (figure 5.5). Hydroxylation of lycopene was dependent on the presence of the \textit{crtC} gene from \textit{Cab. thermophilum}. \textit{E. coli} cells that contained plasmids pACLYC and pCOLA with no insert produced lycopene but no other carotenoids (figure 5.4).

\textbf{5.3.3. Expression of \textit{crtO} in \textit{\textgamma-}carotene producing \textit{Escherichia coli}.} \textit{E. coli} cells containing the pACLYC, p16_CPL1 and pCOLA::\textit{crtO} plasmids produced minor amounts of 4-keto-\textit{\textgamma-}carotene as evidenced by HPLC analyses of carotenoids extracted from these cells (figure 5.6) and absorption spectra of the eluting compounds (figure 5.7). Ketolation of \textit{\textgamma-}carotene was dependent on the presence of the \textit{crtO} gene from \textit{Cab. thermophilum}. As above, \textit{E. coli} cells that contained plasmids pACLYC, pCL16 and pCOLA-Duet produced \textit{\textgamma-}carotene but no other carotenoids (figure 5.6).

\textbf{5.3.4. \textit{In vitro} modification of \textit{\textbeta-}carotene by \textit{Candidatus Chloracidobacterium thermophilum} CrtO expressed in \textit{Escherichia coli}.} \textit{E. coli} cells containing the pBC01 and pCOLA::\textit{crtO} plasmids had slow growth rates and did not produce carotenoids under the conditions tested. Because of this, an \textit{in vitro} assay of CrtO
activity with β-carotene as substrate was performed in order to verify the ketolase activity of this protein. Carotenoids from this reaction were extracted and analyzed following overnight incubation at 37°C. With this assay, small amounts of both echinenone and canthaxanthin were detected (figures 5.8 and 5.9).

5.3.5. Expression of cruF in lycopene-producing Escherichia coli. E. coli cells containing the pACLYC and pCOLA::cruF plasmids synthesized lycopene and a more hydrophobic, unidentified compound that eluted 1 min after lycopene (figure 510). This compound had absorption maxima at 417, 439 and 466 nm. The synthesis of this compound was dependent on the presence of ORF 23570. E. coli cells that only contained the pACLYC, and pCOLA-Duet plasmids produced only lycopene.

5.3.6. Phylogeny of carotenoid-biosynthesis enzymes in Candidatus Chloracidobacterium thermophilum. Phylogenetic trees of the different families of carotenoid biosynthesis enzymes described in this chapter revealed that, phylogenetically, Cab. thermophilum proteins are most closely related to those of other green bacteria. For example, the lycopene cyclase CruA and the hydratase CrtC are both basal sequences in the Chlorobi clades for these respective enzymes (figures 5.13 and 5.15), whereas the ketolase CrtO is a basal sequence within the Chloroflexi clade (figure 5.14). In addition, an alignment of the amino acid sequence in the N-terminus of Cab. thermophilum CruA with those of various Cyanobacteria and Chlorobi (figure 5.12) show that, like the Chlorobi, the CruA sequence of Cab. thermophilum does not include the first 160 residues that are
thought to bind chlorophyll as a cofactor in CruA from *Cyanobacteria* (Wei Xiong, Yue-hui Zhu, and D. A. Bryant, personal communication).

Two of the protein families analyzed, the YcYd lycopene cyclase and the putative desaturase CrtD, do not occur in other green bacteria. In one of those cases, (CrtD), *Cab. thermophilum* sequences were associated with clades that contained predominately soil organisms, rather than with phototrophic groups like the *Cyanobacteria* or purple bacteria (figure 5.15). *Cab. thermophilum* is the first known chlorophototroph that possesses the heterodimeric lycopene cyclase YcYd, and this sequence was the earliest diverging sequence among all of the analyzed YcYd sequences. Interestingly, many members of this family of lycopene cyclases are halophilic organisms, including some halophilic *Archaea* (figure 5.13).

### 5.4 DISCUSSION

Heterologous expression of carotenoid biosynthesis genes in strains of *E. coli* that synthesize carotenoid intermediates has proven to be a direct approach to verify the potential functions of these genes (8, 11, 22, 32). In this study, this approach has been used to overexpress and analyze ORFs predicted to encode carotenoid biosynthesis genes in *Cab. thermophilum*. This organism synthesizes dicyclic ketocarotenoids as the major carotenoid species and linear myxol-like derivatives in minor amounts.

Four different families of enzymes are known to introduce rings at the ends of lycopene (24). The genome of *Cab. thermophilum* contains ORFs with strong sequence similarity to two of these families, the heterodimeric YcYd and CruA
families of lycopene cyclases. The data presented in this chapter clearly indicate that ORFs 06920 and 06910, coding for CrtYc and CrtYd homologs, have lycopene cyclase activity, and it is that is the function of these proteins in Cab. thermophilum. Despite its strong homology to other lycopene cyclases, attempts to express the CruA homolog encoded by ORF 12080 in the same lycopene-producing E. coli have thus far been unsuccessful. Likewise, expression of CruA-type lycopene cyclases from other organisms in lycopene-producing E. coli has yielded mixed results in the past. Although CruA from Cba. tepidum showed lycopene cyclase activity when expressed heterologously (24), the CruA enzyme from Synechococcus sp. PCC 7002 did not. Instead, its lycopene cyclase function was demonstrated through knockout mutants (18, 24). A possible explanation for this lack of activity of some CruA enzymes when heterologously expressed might be that E. coli is not providing a cofactor, such as chlorophyll, that is essential for this enzyme. As shown in figure 5.12, the amino acid sequence of this protein in Cab. thermophilum, however, does not contain what it is thought to be the putative cofactor binding site in the N-terminus of cyanobacterial CruA. Further studies on this enzyme, such as complementation of cruA mutants from other organisms, will aid in elucidating its activity and role in cells.

Because Cab. thermophilum synthesizes both monocyclic and dicyclic carotenoids, it is reasonable to speculate that each of the two lycopene cyclases present in this organism might be employed for one of these alternative branches of carotenoid biosynthesis. In this way the synthesis of different carotenoids could be regulated (23-26). Although overexpression conditions need to be optimized, the
heterologous complementation data showed that CrtY$_c$, CrtY$_d$ can synthesize both γ-carotene and β-carotene, especially when overexpressed at lower temperatures in the presence of IPTG. These results indicated that this enzyme in Cab. thermophilum is a lycopene cyclase capable of utilizing both lycopene and γ-carotene as substrates. Further analyses that involve expressing the genes encoding this heterodimeric cyclase in a γ-carotene background might aid in deducing whether lycopene or γ-carotene is the preferred substrate for this enzyme.

The genome of Cab. thermophilum contains the genes for two putative carotenoid hydratases, cruF and crtC. Both of these genes code for enzymes with similar 1,2-hydratase activity in different organisms. CruF participates in the biosynthesis of myxoxanthophyll and other similar hydroxylated carotenoids in *Cyanobacteria* and *Chloroflexi* (25), whereas CrtC has been found in purple bacteria and members of the *Chlorobi*, in which it participates in the biosynthesis of carotenoids such as spirilloxanthin and chlorobactene derivatives. Given their similar function, the presence of both of these genes in one organism was surprising. The data obtained indicate that ORF 23570, which has only weak homology to cruF in *Synechococcus* sp. PCC 7002, did not lead to the production of an enzyme that exhibited hydroxylase activity when expressed in a lycopene background. Instead, the absorbance maxima indicated that it probably generated a derivative of neurosporene, a precursor to lycopene generated by the *crtI* gene present in the pACLYC plasmid. Given these results we cannot conclude whether the product of this ORF is involved in carotenoid biosynthesis in Cab. thermophilum, because neurosporene is not produced in this organism. It could
modify a carotenoid other than lycopene, unknown to us at this time or, possibly, has a role in an altogether different pathway. ORF 05620 with homology to \textit{crtC}, however, exhibited the typical activity of known 1, 2-hydratases and produced 1-hydroxylycopene and 1,1′ dihydroxylycopene when expressed in a lycopene-producing strain of \textit{E. coli}.

The major carotenoids found in \textit{Cab. thermophilum} are the ketocarotenoids echinenone and canthaxanthin (chapter 4). ORF 10330 in the genome exhibits strong sequence similarity to the \textit{crtO} family of carotenoid ketolases, and the gene product was predicted to encode a φ–ring ketolase in this organism. This family of enzymes has been found in the \textit{Cyanobacteria}, purple bacteria and other nonphototrophic microorganisms such as \textit{Deinococcus} spp. (8, 13, 33, 35). In \textit{Cyanobacteria} these ketolases only seem to recognize β-carotene as substrate and yield the asymmetrical product, echinenone. Analyses on heterologously expressed CrtO from \textit{Rhodococcus erythropolis} strain AN12 and \textit{Deinococcus radiodurans} R1, however, have shown that in these organisms CrtO can add keto groups to both rings of β-carotene and generate the symmetrical product, canthaxanthin (35). Despite numerous attempts under various conditions, CrtO from \textit{Cab. thermophilum} did not show activity in β-carotene-producing \textit{E. coli}. However, the same construct in a γ-carotene-producing background led to small amounts of 4-keto-γ-carotene, supporting the hypothesis that this enzyme is indeed a carotene ketolase. An \textit{in vitro} reaction, in which extracts of an \textit{E. coli} strain that was expressing CrtO were incubated with β-carotene, resulted in small amounts of both echinenone and canthaxanthin, again indicating that this enzyme has ketolase
activity. The small amounts of γ-carotene and β-carotene ketolated by this enzyme are encouraging, but further optimization of the expression of CrtO are necessary. For example, it could be that most of the expressed protein is found in inclusion bodies and that an alternative expression system or conditions might yield more active enzyme. Alternatively, a crtO mutant strain from another organism (i. e., Rhodococcus or Deinococcus spp.) could be complemented with the gene from Cab. thermophilum.

An additional gene in Cab. thermophilum with strong homology to carotenoid biosynthesis genes in other organisms is ORF 20690 which codes for a putative desaturase, CrtD. Attempts to express this gene in E. coli producing either lycopene or γ-carotene, or in combination with CrtC in both of those backgrounds, have thus far failed to reveal any activity. Because the sequence similarity to known carotenoid desaturases is so strong (E value =1e^-86), the lack of detectable activity of this enzyme is most probably due to incompatibility with our expression system. Alternatively, it may also be that the correct substrate or expression conditions have not yet been identified. Complementation of crtD mutants of other organisms might be a more successful approach in studying the function of this enzyme.

The enzymes responsible for carotenoid biosynthesis in Cab. thermophilum have a diverse ancestry that includes the green bacteria, i. e., members of the Chlorobi and Chloroflexi, and Gram-positive organisms typically found in soils. None of the protein sequences analyzed here were phylogenetically closest to other chlorophototroph groups. In this aspect carotenoid biosynthesis differs from bacteriochlorophyll biosynthesis in this organism, for which some of the enzymes
involved in the early steps were related most closely to those found in the purple bacteria (4) whereas the enzymes responsible for the late steps were most closely related to those of the green bacteria (chapter 3).

In summary, we have presented here an initial analysis of carotenoid biosynthesis genes in Cab. thermophilum, the only known chlorophototrophic acidobacterium.
5.5. REFERENCES


FIGURE LEGENDS

Figure 5.1. Proposed carotenoid biosynthetic pathway in Candidatus Chloracidobacterium thermophilum.

Figure 5.2. HPLC analyses of carotenoids produced in lycopene-producing *Escherichia coli* strains. Elution profile of carotenoid extracts from *E. coli* cells containing A, the pACYC and pCOLA plasmids, and B and C, pACYC and pCOLA::crtY,Y,Y plasmids. Cells in panel C were grown at 28 °C and induced with IPTG. Cells in panel B were grown at 37 °C and no IPTG was added. Peaks 1, 2 and 3 identify lycopene, γ-carotene and β-carotene, respectively.

Figure 5.3. Absorption spectra of carotenoids identified in Figure 5.2.

Figure 5.4. HPLC analyses of carotenoids produced in lycopene-producing *Escherichia coli* strains. Elution profile of carotenoid extracts from *E. coli* cells containing A, the pACLYC and pCOLA plasmids, and B, pACLYC and pCOLA::crtC plasmids. Peaks 1, 2 and 3 identify lycopene, 1’-hydroxy-lycopene and 1, 1’-dihydroxy-lycopene, respectively.

Figure 5.5. Absorption spectra of carotenoids identified in Figure 5.4.

Figure 5.6. HPLC analyses of carotenoids in γ-carotene producing *Escherichia coli* strains. Elution profile of carotenoid extracts from *E. coli* cells containing A, the pACLYC, pACYC and pCOLA plasmids, and B, pACLYC, pACYC and pCOLA::crtO plasmids. Peaks 1, and 2 identify γ-carotene and 4-keto-γ-carotene, respectively.

Figure 5.7. Absorption spectra of carotenoids identified in Figure 5.6.

Figure 5.8. HPLC analyses of *in vitro* CrtO activity on β-carotene. Elution profile of carotenoid extracts from A, *E. coli* cells containing pBCO1 and pCOLA plasmids, and B, those same extracts incubated overnight at 37°C in MOPS buffer, pH 8, and with extracts from *E. coli* cells containing pCOLA::crtO.

Figure 5.9. Absorption spectra of carotenoids identified in Figure 5.8.

Figure 5.10. HPLC analyses carotenoids in γ-carotene producing *Escherichia coli* strains. Elution profile of carotenoid extracts from *E. coli* cells containing A, the pACLYC and pCOLA plasmids, and B, pACLYC and pCOLA::cruF plasmids. Peaks 1 and 2 identify lycopene and cis-lycopene, respectively. Peak 3 represents an unidentified species.

Figure 5.11. Absorption spectra of carotenoids identified in Figure 5.10.
Figure 5.12. Alignment of the N-terminal 400 amino acid residues of lycopene cyclases of the CruA family. Alignment was generated with ClustalW and edited with Bioedit. Note that the cyanobacterial sequences have an extra 140 residues at the N-termiuns that might be involved in binding Chl $a$ (W. Xiong, personal communication).

Figure 5.13. Phylogenetic tree of CrtY$_c$, CrtY$_d$ and CruA. Unrooted neighbor-joining tree of lycopene cyclases from the CrtY$_c$, CrtY$_d$ and CruA families.

Figure 5.14. Phylogenetic tree of CrtO. Unrooted neighbor-joining tree of CrtO carotene ketolases.

Figure 5.15. Phylogenetic tree of CrtC. Neighbor-joining tree of CrtC 1’,2’-carotene hydratases. CruF sequences from Cyanobacteria were included as root.

Figure 5.16. Phylogenetic tree of CrtD. Neighbor-joining tree of CrtD, methoxyneurosporene dehydrogenase. CrtI sequences were included as root.
### Table 5.1. Primers used in this study.

<table>
<thead>
<tr>
<th>Locus Tag</th>
<th>Predicted gene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>06920</td>
<td><em>crtYc</em></td>
<td><strong>CGC</strong>CATGGCGTACTTGCAGTTTCACCTTT**</td>
<td><strong>CCC</strong>GGATCCGCTGCGTCTCCCGTG**</td>
</tr>
<tr>
<td>06910</td>
<td><em>crtYd</em></td>
<td><strong>CGCC</strong>ATAGAAAACCAACTACCTGTTCC**</td>
<td><strong>CTCGAG</strong>TCATGCACGATGCCTCTCCG**</td>
</tr>
<tr>
<td>06910-06920</td>
<td><em>crtYc-crtYd</em></td>
<td><strong>CGCC</strong>ATGGCGTACTTGCAGTTTCACCTTT**</td>
<td><strong>CCC</strong>GGATCCGCTGCGTCTCCCGTG**</td>
</tr>
<tr>
<td>05620</td>
<td><em>crtC</em></td>
<td><strong>AATAA</strong>CATAGCGAATGATAGCCAGCCGCA**</td>
<td><strong>CTCGAG</strong>GACAAGCTCAGAAACCCAG**</td>
</tr>
<tr>
<td>20690</td>
<td><em>crtD</em></td>
<td><strong>CTGG</strong>CCATGGGTCCTTTAGCCGAAAACAC**</td>
<td><strong>GGATCC</strong>CAAGATGCCCGCTCCC**</td>
</tr>
<tr>
<td>10330</td>
<td><em>crtO</em></td>
<td><strong>CGCC</strong>ATAGTCCCGGAGCAACTCGCGGT**</td>
<td><strong>CTCGAG</strong>TAAGCAGCGAAGCCAGG**</td>
</tr>
<tr>
<td>23570</td>
<td><em>cruF</em></td>
<td><strong>GCGACTTT</strong>AATAGACCGATCTGCTGC**</td>
<td><strong>GGATCC</strong>GTTTCCACGACCAGACCTTC**</td>
</tr>
</tbody>
</table>
Table 5.2. Properties of plasmids used in this study.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Parent Vector</th>
<th>Origin of replication</th>
<th>Antibiotic casette</th>
<th>Carotenoid product</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC</td>
<td>pACYC184</td>
<td>p15A</td>
<td>Chloramphenicol</td>
<td>lycopene</td>
<td>Cunningham 1994</td>
</tr>
<tr>
<td>pCL1</td>
<td>pUC19</td>
<td>pMB1</td>
<td>Ampicillin</td>
<td>γ-carotene</td>
<td>Maresca 2007</td>
</tr>
<tr>
<td>pBC01</td>
<td>pBAD-TOPO</td>
<td>pMB1</td>
<td>Ampicillin</td>
<td>β-carotene</td>
<td>Martinez 2007*</td>
</tr>
<tr>
<td>pCOLA-Duet</td>
<td>pCOLA-Duet</td>
<td>ColA</td>
<td>Kanamycin</td>
<td>none</td>
<td>Novagen</td>
</tr>
<tr>
<td>pCOLA::crtYc-crtYd</td>
<td>pCOLA-Duet</td>
<td>ColA</td>
<td>Kanamycin</td>
<td>γ-carotene, β-carotene</td>
<td>this work</td>
</tr>
<tr>
<td>pCOLA::crtC</td>
<td>pCOLA-Duet</td>
<td>ColA</td>
<td>Kanamycin</td>
<td>hydroxylated lycopene</td>
<td>this work</td>
</tr>
<tr>
<td>pCOLA::crtO</td>
<td>pCOLA-Duet</td>
<td>ColA</td>
<td>Kanamycin</td>
<td>echinenone, canthaxanthin</td>
<td>this work</td>
</tr>
<tr>
<td>pCOLA::cruF</td>
<td>pCOLA-Duet</td>
<td>ColA</td>
<td>Kanamycin</td>
<td>unknown</td>
<td>this work</td>
</tr>
</tbody>
</table>

*pBC01 was constructed by Lee Miller under the supervision of Julia Maresca in Ed Delong’s laboratory at the Massachusetts Institute of Technology, Cambridge, MA. Genes from fosmid HF10-19P19 (Martinez 2007) were used to construct this plasmid.

^These products are obtained in combination with additional plasmids that synthesize lycopene, γ-carotene or β-carotene. See text for details.
Figure 5.1
Figure 5.2

A

B

C

Absorbance at 491 nm

Time (minutes)
Figure 5.3
Figure 5.4

Absorbance at 491 nm

A

B

Time (minutes)
Figure 5.5
Figure 5.6

Absorbance at 491 nm

A

B

1

2

Time (minutes)
Figure 5.7

Absorbance

Wavelength (nm)

350 400 450 500 550 600

1

462 492

2

468 486
Figure 5.8

Absorbance at 491 nm

A

B

Time (minutes)
Figure 5.9

![Absorbance graph with wavelengths 453, 478, and 465]
Figure 5.10

Absorbance at 491 nm

Time (minutes)
Figure 5.11

Absorbance

Wavelength (nm)
### Figure 5.12

<table>
<thead>
<tr>
<th>Cyananthaceae sp. PCC 7424</th>
<th>Micrasterias acruginosa PCC 7611</th>
<th>Mastac sp. 0700</th>
<th>Anabaena variabilis ATCC 29413</th>
<th>Synecococcus sp. PCC 7002</th>
<th>Synechocystis sp. PCC 6003</th>
<th>Synechocystis sp. JR-3-3ab</th>
<th>Gloeobacter violaceus PCC 7421</th>
<th>Chlorobium lutidum BGM 273</th>
<th>Chlorobium ferrooxidans BGM 13</th>
<th>Chlorobium phaeobacteroides BS</th>
<th>Chlorobium limicola BGM 245</th>
<th>Chlorobium tepidum TLS</th>
<th>Prosthecochloris aestuarii</th>
<th>Chlorarchaeon thalassium</th>
<th>Chlorarchaeon thermophil</th>
<th>Clastal Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyananthaceae sp. PCC 7424</td>
<td>Micrasterias acruginosa PCC 7611</td>
<td>Mastac sp. 0700</td>
<td>Anabaena variabilis ATCC 29413</td>
<td>Synecococcus sp. PCC 7002</td>
<td>Synechocystis sp. PCC 6003</td>
<td>Synechocystis sp. JR-3-3ab</td>
<td>Gloeobacter violaceus PCC 7421</td>
<td>Chlorobium lutidum BGM 273</td>
<td>Chlorobium ferrooxidans BGM 13</td>
<td>Chlorobium phaeobacteroides BS</td>
<td>Chlorobium limicola BGM 245</td>
<td>Chlorobium tepidum TLS</td>
<td>Prosthecochloris aestuarii</td>
<td>Chlorarchaeon thalassium</td>
<td>Chlorarchaeon thermophil</td>
<td>Clastal Consensus</td>
</tr>
<tr>
<td>Cyananthaceae sp. PCC 7424</td>
<td>Micrasterias acruginosa PCC 7611</td>
<td>Mastac sp. 0700</td>
<td>Anabaena variabilis ATCC 29413</td>
<td>Synecococcus sp. PCC 7002</td>
<td>Synechocystis sp. PCC 6003</td>
<td>Synechocystis sp. JR-3-3ab</td>
<td>Gloeobacter violaceus PCC 7421</td>
<td>Chlorobium lutidum BGM 273</td>
<td>Chlorobium ferrooxidans BGM 13</td>
<td>Chlorobium phaeobacteroides BS</td>
<td>Chlorobium limicola BGM 245</td>
<td>Chlorobium tepidum TLS</td>
<td>Prosthecochloris aestuarii</td>
<td>Chlorarchaeon thalassium</td>
<td>Chlorarchaeon thermophil</td>
<td>Clastal Consensus</td>
</tr>
<tr>
<td>Cyananthaceae sp. PCC 7424</td>
<td>Micrasterias acruginosa PCC 7611</td>
<td>Mastac sp. 0700</td>
<td>Anabaena variabilis ATCC 29413</td>
<td>Synecococcus sp. PCC 7002</td>
<td>Synechocystis sp. PCC 6003</td>
<td>Synechocystis sp. JR-3-3ab</td>
<td>Gloeobacter violaceus PCC 7421</td>
<td>Chlorobium lutidum BGM 273</td>
<td>Chlorobium ferrooxidans BGM 13</td>
<td>Chlorobium phaeobacteroides BS</td>
<td>Chlorobium limicola BGM 245</td>
<td>Chlorobium tepidum TLS</td>
<td>Prosthecochloris aestuarii</td>
<td>Chlorarchaeon thalassium</td>
<td>Chlorarchaeon thermophil</td>
<td>Clastal Consensus</td>
</tr>
</tbody>
</table>

214
Figure 5.13
Figure 5.14

![Dendrogram showing the relationships between various bacterial and cyanobacterial species.](image)

- **Cyanothece**
- **Lyngbia**
- **Nostoc azollae**
- **Nostoc punctiforme**
- **Synechocystis sp. PCC 6803**
- **Microcystis aeruginosa NIES-843**
- **Deinococcus radiodurans**
- **Deinococcus deserti**
- **Deinococcus geothermalis**
- **Rhodococcus jostii**
- **Rhodococcus erythropolis**
- **Rhodococcus opacus**
- **Cab. thermophilum**
- **Salinibacter ruber**
- **Roseiflexus castenholzii**
- **Roseiflexus sp. RS-1**
- **Cfx. aurantiacus**
- **Cfx. aggregans**
- **Herpetosiphon aurantiacus**
- **Nocardia farcinica**
- **Deinococcus geothermalis**
- **Deinococcus deserti**
- **Deinococcus radiodurans**
- **Rhodococcus erythropolis**
- **Nocardia farcinica**
- **Lyngbia sp. PCC 8106**
- **Nostoc punctiforme**
- **Anabaena variabilis**
- **Microcystis aeruginosa NIES-843**
- **Cfx. aurantiacus**
- **Cfx. aggregans**
- **Roseiflexus castenholzii**
Figure 5.15

Proteobacteria CrtC

- Bradyrhizobium sp. BTA1
- Allochlamatium vinosum
- Sarangium rubrum
- Rhodospirillum rubrum

Chlorobi CrtC

- Cab. thermophilum
- Chlorobium chlorochromatti
- Chlorobium parvum
- Chlorobium tepidum
- Prosthecochloris aestuarii
- Prosthecochloris vibrioformis
- Pelodictyon phaeodathraforme
- Chlorobium ferrooxidans

Cyanobacteria CruF

- Synechocystis sp. PCC 6803
- Synechococcus sp. PCC 7902
- Chlorobium phaeobacteroides
- Chlorobium limicola
- Thermosynechococcus elongatus
- Nostoc punctiforme PCC 73102
- Anabaena variabilis

0.1
Figure 5.16
CHAPTER 6. CONCLUDING REMARKS

This thesis has advanced our knowledge of bacterial chlorophototrophy at many different levels. First of all, the work presented here has challenged and expanded our understanding of chlorophototrophic physiological diversity in the bacterial world. For example, prior to the discovery of Candidatus Chloracidobacterium thermophilum, there were no groups of chlorophototrophs that were known to synthesize chlorosomes under oxic conditions (Chapter 2). The existence of these “aerobic chlorosomes” challenged the paradigm of chlorosomes as large and efficient antenna complexes that absorb and transfer light energy at low illumination intensity and under anoxic conditions. However, although Cab. thermophilum chlorosomes do exhibit some adaptations to highly oxic conditions, most notably the presence of ketocarotenoids, and other novel properties regarding protein composition and bacteriochlorophyll (BChl) arrangement (see below), the fluorescence emission spectra presented here show that energy transfer in these chlorosomes is redox-dependent and that they appear to function in a similar way to chlorosomes found in the Chlorobiales and Chloroflexi (Chapter 4). The pigment composition in these antennae further supports this premise, because the BChl c molecules identified structurally in this study are similar, in terms of methylations and esterifying alcohols, to those found in other green bacteria (Chapter 4).

The biochemical analyses of the pigment arrangement, proteins and lipids found in Cab. thermophilum chlorosomes presented in this work have additionally revealed unique properties that warrant further investigation (Chapter 4). For example, three previously uncharacterized proteins have been found in the
chlorosome envelope. Although genetic studies, such as the ones conducted on *Chlorobaculum tepidum* to elucidate the role of its chlorosome proteins, cannot yet be carried out with *Cab. thermophilum*, other approaches such as protein colocalization studies might provide clues about the possible roles of these peptides. Likewise, initial studies on the internal arrangement of the BChl molecules have revealed an alternative supramolecular organization of these pigments that clearly differs from that in chlorosomes of *Cba. tepidum* (Chapter 4). The data presented here seem to indicate that there are at least two different modes of BChl organization in chlorosomes, and further analyses of these chlorosomes will increase our understanding of this fascinating property of chlorosome BChls.

This work has also expanded and challenged our understanding of the microbial community in the chlorophototrophic mats of Octopus and Mushrooms Springs. The combined genomic and physiological studies describe a previously unknown member of this mat community, which is dependent on other members for at least fixed carbon, nitrogen and branched-chain amino acids and possibly reduced sulfur (Chapters 2 and 3). One can only speculate about the ecological contributions of *Cab. thermophilum* to the mat community. The data presented here suggest that it occupies a unique niche with low illumination and reduced oxygen conditions (Appendix C), and that it is a heterotroph that appears to consume the fermentation end products generated by the primary producers of the mat (chapters 2 and 3). It might play an essential role in preventing the accumulation of these products, although there are other heterotrophs in the mat with the capability to use these reduced carbon compounds (Chapter 1) that could also play a similar role. *In*
in situ analyses might provide additional clues to the interactions among the organisms found in these mats. For example, analyses of the stable isotope composition of organic molecules such as chlorophylls or lipids from Cab. thermophilum might reveal the nature of the carbon fixation pathway from the primary producer that provides carbon to Cab. thermophilum.

A functional survey of the distribution of Cab. thermophilum in other hot springs in Yellowstone National Park has revealed the existence of four subpopulations of these microorganisms that seem to be adapted to different temperatures (appendix D), a niche diversification also observed in other members of the mat (see Chapter 1).

In addition to revealing the nutritional requirements of Cab. thermophilum, the genome studies presented here predict this organism to have a highly branched electron transfer chain that might include both the cytochrome $bc_1$ complex and alternative complex III (Chapter 3). Biochemical isolation and enzymatic analyses of the electron donors and acceptors of these complexes, combined with gene-expression studies, might decipher the role of the different components of these electron transfer chains.

Other genome predictions included the nature of the carotenoid biosynthesis pathway in Cab. thermophilum (Chapter 3, Chapter 5). Heterologous expression of genes predicted to code for enzymes involved in carotenoid biosynthesis has provided initial evidence that confirms the biochemical identity of these enzymes (Chapter 5) and suggests possible biosynthetic pathways employed by this organism to synthesize carotenoids.
Lastly, this work has challenged and expanded our understanding of chlorophototrophic phylogenetic diversity in the bacterial world, bringing the number of bacterial phyla that contain chlorophototrophs to six. In doing so, it has advanced and expanded awareness of the physiological capabilities of the bacterial phylum Acidobacteria, capabilities that now include chlorophototrophy. This group of microorganisms is ubiquitous in soils worldwide but has proven to be difficult to grow in the laboratory with just a few representatives currently in culture (see Chapters 2 and 3). Despite these difficulties, the present work and those of others (see references 16 and 31 in Chapter 3) are beginning to define this group as a metabolic and phylogenetic diverse bacterial clade, similar in this respect to the Proteobacteria.
Figure A.1. HPLC analyses of carbon sources in spent media collected at various time points from a culture of Candidatus Chloracidobacterium thermophilum. Aliquots of a culture of Cab. thermophilum were collected every 24 hours, filter sterilized and analyzed by HPLC in a Shimadzu LC-20AB system equipped with a UV detector SPD-20A (210 nm) and a refractive index detector (RID-10A). Compounds were separated on a Supelcogel C610H column (Supelco), with HPLC grade 4 mM H₂SO₄ as the mobile phase. The flow rate was 0.5 ml min⁻¹ and the column was operated at 30 °C. The increase in BChl c absorbance throughout the incubation period, which signifies the growth of Cab. thermophilum, is superimposed in red. Succinate and lactate have identical elution times and could not be distinguished by this analysis. Note how the disappearance of succinate, lactate and glycolate from the medium correlates with growth of Cab. thermophilum.
Figure A.2. HPLC analyses of unknowns in spent media collected at various time points from a culture of *Candidatus Chloracidobacterium thermophilum*. Analyses of the same aliquots as those used for Figure 1 detected additional compounds eluting at different times than our known standards. Compounds are labeled according to their elution time in minutes. Compound eluting at 8 minutes is probably the buffer compound, HEPES, present in the culture medium. Please note how the compounds eluting at 8.8, 18.5 and 22.1 minutes are not present in the initial medium. Also note how the compound eluting at 22.1 minutes disappears proportionally to *Cab. thermophilum* exponential growth. Similar analyses carried out in for an axenic culture of *Anoxybacillus* sp. isolated from this enrichment show a similar appearance of peak 22.1 after 24 hours but, unlike in the culture of *Cab. thermophilum*, this peak does not disappear. Peak 18.1 was detected with the RID detector but not with the UV detector.
Figure B.1. Photodegradation of cell suspensions from representatives of all three phyla of green bacteria. Cab. thermophilum (orange triangles), Chloroflexus aurantiacus (yellow-green squares) and Chlorobaculum tepidum (blue-green diamonds) cells were resuspended in 10 mM Tris-HCl, pH 8 and exposed to light at 300 µmol photons m$^{-2}$ s$^{-1}$. Photodegradation of BChl c was measured over a period of 48 hours. Please note the sensitivity of Cfx. aurantiacus and Cba. thermophilum cells to high light intensities and the resistance of Cab. thermophilum to the same conditions. Please see Figure B.2 for images of these cell suspensions.
Figure B.2. Images of cell suspensions of representatives of all green bacteria before and after exposure to high light intensities. Cfx., *Cfx. aurantiacus*; Cab., *Cab. thermophilum*. Cba., *Cba. tepidum*. 

Cfx.  
↓  
Cab.  
↓  
Cba.
Figure B.3. Photodegradation Cab. thermophilum cell suspensions. Cab. thermophilum cells were grown under high (120 \text{µmol photons m}^{-2} \text{s}^{-1}) and low (30 \text{µmol photons m}^{-2} \text{s}^{-1}) light conditions and in the presence and absence of the lycopene cyclase inhibitor MPTA. Suspensions of these cells were exposed to a light intensity of 300 \text{µmol photons m}^{-2} \text{s}^{-1} and photodegradation of BChl \text{c} monitored. Cells incubated under high-light are shown in pink, cells incubated under low-light are shown in blue, cells incubated under high-light and in the presence of MPTA are shown in cyan, and cells incubated under low-light and in the presence of MPTA are shown in yellow. Notice how Cab. thermophilum cells grown in the presence of MPTA are sensitive to high-light exposure regardless of high or low-light incubation conditions.
Figure B.4. Photodegradation of chlorosome suspensions representatives from all three phyla of green bacteria. Cab. thermophilum (orange triangles), Chloroflexus aurantiacus (green squares) and Chlorobaculum tepidum (blue-gren diamonds) isolated chlorosomes were resuspended in 10mM Tris-HCl, pH 8.0 and exposed to light at 300 µmol photons m$^{-2}$ s$^{-1}$ intensity. Photodegradation of BChl c was measured over a period of ca. 10 hours. Please note the sensitivity of chlorosomes to light from representatives of all three different phyla, including Cab. thermophilum.
Figure C.1. *In situ* expression of the reaction center gene *pscA* from *Cab. thermophilum during a diel cycle*. RNA from mat samples from Octopus and Mushroom Springs were collected at 4 times during a diel cycle, converted into cDNA and sequenced by pyrosequencing. Bars representing the expression of the *Cab. thermophilum* gene *pscA* at these four time points are superimposed on a graph from Revsbeek and Ward (see reference 63 in Chapter 1) that shows the light intensities and oxygen partial pressures in these mats throughout a diel cycle. Please note how *Cab. thermophilum* is most actively transcribing at a time with low light intensities and almost anoxic conditions. At this time in the diel cycle, fermentations products such as glycolate and ethanol accumulate due to fermentation by *Synechococcus spp.* of glycogen accumulated during the day (see reference 4 in Chapter 1). Data collected by Zhenfeng Liu from Dr. Donald A. Bryant’s laboratory at the Pennsylvania State University and members of the laboratory of Dr. David Ward at Montana State University, Bozeman, MT.
APPENDIX D. Figure D.1. Neighbor-joining phylogenetic tree showing the relationships among various *C. thermophilum* pscA sequences. Tree was constructed using alignments of nucleotide sequences that had been obtained from IMG/M (http://img.jgi.doe.gov/cgi-bin/m/main.cgi) for the Bison Pool Q and *C. thermophilum* type M, PCR of *C. thermophilum* genomic DNA using pscA–specific primers (see Table 2.1 for primer sequence) and PCR (with same primers) of environmental DNA from phototrophic mats in Heart Lake, Imperial Geyser and White Creek Hot Spring all located in Yellowstone National Park and kindly provided by Dr. Eric Boyd from the laboratory of Dr. John W. Peters at Montana State University, Bozeman, MT and Dr. Scott Miller from University of Montana, Missoula, MT. DNA extracted from Chocolate Pots hot spring was kindly provided by Dr. Nicki Parenteau, but it failed to produce amplicons. Notice that four groups of pscA sequences are discernible, probably reflecting adaptation to temperature rather than geographic origin. When known, the temperature of the sample site was included next to the sample name in parenthesis. For additional metadata on White Creek samples please refer to reference 37 in Chapter 3. The tree was rooted with the pscA sequence from *Cba. tepidum*. 
AMAYA M. GARCIA COSTAS, *Curriculum vitae*

**EDUCATION**
- Ph.D. Biochemistry and Molecular Biology, 2010
  The Pennsylvania State University, University Park, PA
- M.S. Microbiology, 1999
  University of Illinois, Urbana-Champaign, IL
- B.S. Biology and Chemistry, 1996
  Rocky Mountain College, Billings, MT

**ADDITIONAL EDUCATION**
- Microbial Genomics and Metagenomics Workshop, August 2008
  Joint Genome Institute, Walnut Creek, CA
- Microbial Diversity course, Summer 2004 and Summer 2005
  Marine Biological Laboratory, Woods Hole, MA
- Exploring the Ecology of Yellowstone’s Microbes, July 2003
  Yellowstone Association, Anna Louise Reysenbach instructor
- Yellowstone Charismatic Microorganisms, July 2002
  Yellowstone Association, David Ward instructor
- Water Microbiology Course, Fall 1996
  Montana State University, Bozeman, MT
- Summer Undergraduate Research Experience, June-August 1995
  University of Oklahoma Medical Center, Oklahoma City, OK

**HONORS**
- Best Poster Award. International Symposium of Phototrophic Prokaryotes.
  Montreal, Canada. August 2009
- Invited speaker. American Society for Microbiology National Meeting.
  Philadelphia, PA. May 2009
- Graduate talk, 1st place winner. Montana Academy of Sciences. Montana Tech
  University, Butte, MT. March 2009
- Honorable Mention Recipient, 2008
  Robert D. Watkins Graduate Research Minority Fellowship Program,
  American Society for Microbiology
- Graduate talk, 1st place winner. ABASM, La Roche College. Pittsburgh, PA
  November 2007
- Deboer Research Fellowship, 1997-1999
  University of Illinois, Urbana-Champaign IL
- Dean’s Cup, given to the two seniors -male and female- with the highest
  G.P.A., May 1996. Rocky Mountain College, Billings, MT
- Spidel Scholar, a four-year tuition scholarship for Valedictorians from rural