Genomic Analysis and Metabolic Potential of Myxobacteria

Hanna Albataineh

Follow this and additional works at: https://egrove.olemiss.edu/etd

Recommended Citation
https://egrove.olemiss.edu/etd/2349

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.
GENOMIC ANALYSIS AND METABOLIC POTENTIAL OF MYXOBACTERIA

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of BioMolecular Sciences
Division of Pharmacognosy
The University of Mississippi

By

HANAN ALBATAINEH

August 2022
ABSTRACT

Myxobacteria are known for their large genomes that bestow them with a distinct lifestyle that is characterized by fruiting body formation, gliding motility, social interactions, predation, and secondary metabolite production. The powerful combination of whole-genome sequencing and bioinformatics-driven analysis of sequence data has delivered unexpected insights into myxobacterial microbial communities. We applied these two approaches to investigate the chemical space and biosynthetic potential of myxobacteria. Mining different myxobacterial genomes for biosynthetic gene clusters led us to the discovery of a cryptic acyl-homoserine lactone (AHL) synthase encoded in the genome of the myxobacterium Archangium gephyra. Bioinformatic analysis of this AHL synthase using antiSMASH, blastp, MEGA7 HMMSEARCH, and EFI-EST showed that it is highly homologous to other functional bacterial AHL synthases with no cognate AHL receptor encoded in the genome. Heterologous expression of this AHL synthase in Escherichia coli BL21 was performed and it was found to produce detectable quantities of three AHL signals. This is the first example of a functional orphaned LuxI-type AHL synthase reported. This result provides a unique perspective on interspecies cross-talk within polymicrobial communities. We also utilized long-read genomes and modern comparative genomics to investigate the taxonomic rank and biosynthetic potential of four environmental myxobacteria isolated from North American soils and two myxobacteria deposited at the ATCC. Average nucleotide identity and digital DNA–DNA hybridization scores from comparative genomics suggest that the reclassification of Archangium primigenium ATCC 29037 and Chondrococcus macrosorus ATCC 29039 to be potentially novel members of the genera Melittangium and
Corallococcus, respectively, and the four isolated myxobacteria to be species from Corallococcus, Pyxidicoccus, and Myxococcus, respectively. In addition, further analysis to assess the biosynthetic potential of these bacteria using high antiSMASH and BiG-SCAPE platforms suggests genus-level conservation of biosynthetic pathways which support our preliminary taxonomic assignments. This result highlights the significance of applying modern genomics to revise myxobacterial taxonomy and improve our understanding of the genetic basis of the social activities and specialized metabolism of myxobacteria.
DEDICATION

To my parents, Nawal and Abdelfattah

My siblings, Haneen, Mohammad, and Ahmad

And all of the people who did not give up on me to finish my degree

Without whom none of my success would be possible
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHL</td>
<td>Acylhomoserine Lactone</td>
</tr>
<tr>
<td>ANI</td>
<td>Average Nucleotide Identity</td>
</tr>
<tr>
<td>antiSMASH</td>
<td>Antibiotics &amp; Secondary Metabolite Analysis Shell</td>
</tr>
<tr>
<td>ARTS</td>
<td>Antibiotic Resistant Target Seeker</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>BCCT</td>
<td>Betaine/Carnitine/Choline Transporter</td>
</tr>
<tr>
<td>BGCs</td>
<td>Biosynthetic Gene Clusters</td>
</tr>
<tr>
<td>BiG-SCAPE</td>
<td>Biosynthetic Gene Similarity Clustering and Prospecting Engine</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-Induced Dissociation</td>
</tr>
<tr>
<td>dDDH</td>
<td>Digital DNA-DNA Hybridization</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DSM</td>
<td>Deutsche Sammlung von Mikroorganismen und Zellkulturen</td>
</tr>
<tr>
<td>EIC</td>
<td>Extracted Ion Chromatograph</td>
</tr>
<tr>
<td>GBDP</td>
<td>Genome BLAST Distance Phylogeny approach</td>
</tr>
<tr>
<td>GNPS</td>
<td>Global Natural Products Social Molecular Networking</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance Liquid Chromatography</td>
</tr>
</tbody>
</table>
LB  Luria-Bertani
LC-MS  Liquid Chromatography with Mass Spectrometry
LC-MS/MS  Liquid Chromatography with Tandem Mass Spectrometry
MIBiG  Minimum Information about a Biosynthetic Gene cluster
MMC  Marine Myxobacteria Cluster
NAP  Network Annotation Propagation
NGS  Next-generation Sequencing
NRPS  Nonribosomal Peptide Synthases
OD  Optical Density
PCR  Polymerase Chain Reaction
PKS  Polyketide Synthases
QS  Quorum Signals
RiPPs  Ribosomally synthesized and Post-translationally modified Peptides
rRNA  Ribosomal Ribonucleic Acid
SEM  Scanning Electron Microscopy
SMRT  Single-Molecule Real-Time
SSN  Sequence Similarity Network
TYGS  Type (Strain) Genome Server
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VY/2</td>
<td>Half-strength Yeast cell Vitamin B12 medium</td>
</tr>
<tr>
<td>WoRMS</td>
<td>World Register of Marine Species</td>
</tr>
<tr>
<td>XCMS-MRM</td>
<td>Various forms (X) of mass spectrometry-Multiple Reaction Monitoring</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

This work would not have been possible without the constant sincerity, encouragement, and patience of my first advisor Prof. D. Cole Stevens. Through his guidance and encouragement, I have grown substantially as a scientist and been exposed to a breadth of fascinating research experiences. I bestow special thanks to lab members past and present, especially to Barbara, Andrew, Maya, Kayleigh, and Shailaja for a cherished time spent together in the lab, the stimulating discussions, and all the fun we have had in the last years. It has been my privilege to be a part of the Stevens lab and I am forever indebted for their support in both life and science.

I would like to express my deep gratitude to my committee members Prof. Dale G. Nagle, Prof. Joshua Sharp, and Prof. Vitor H. Pomin for their efforts and contributions to this work, and their support in facilitating my learning to be a better scientist.

I am thankful for the funding sources that made my Ph.D. work possible. I was funded by Jordanian American Commission for Educational Exchange (JACEE) for four years and later on by the Glycoscience Center of Research Excellence at the University of Mississippi.

I am indebted also to several professors, staff, and students at the Department of BioMolecular Sciences for their continuous encouragement and suggestions. I would also like to acknowledge Dr. Sandeep Misra and Mr. Anter Shami for all their help with mass spectrometric analysis. My appreciation goes to the administrative staff, Ms. Sherrie Gussow, Ms. Danielle Noonan, and Ms. Candace G. Lowstuter, for their assistance with departmental-related issues. Thanks also to all of the students I met throughout my study, mainly Manal Alhusban, Abidah
Parveen, Hoda Ahmed, and Alaa Qareya, your friendship, and encouragement have been very helpful during my study. I also thank Mohammed Ahmed for his advice and support.

Also, I would also like to extend my thanks to my Professors, coworkers, staff, and fellow students at the Jordan University of Science and Technology and the Yarmouk University for their endless enthusiasm for research and science.

I acknowledge my second family in the US, the Harris family, Mr. Kim, Mrs. Pam, Miss Linda, Aaron, Emmy, and Blake for their continual love and support. They have made my study and life in the US a wonderful time. I would also like to thank my cat Fluffy for all the entertainment and emotional support.

Also, I acknowledge my cousins, Muna and Qusai, for encouraging me to apply to the Fulbright Program. And finally, thanks to my parents and siblings who endured this long process with me, always offering support and love. Without their tremendous understanding and encouragement over the past few years, it would be impossible for me to complete my study.
TABLE OF CONTENTS

ABSTRACT...............................................................................................................................ii

DEDICATION...........................................................................................................................iv

LIST OF ABBREVIATIONS........................................................................................................v

ACKNOWLEDGEMENT.............................................................................................................viii

LIST OF TABLES....................................................................................................................x

LIST OF FIGURES ..................................................................................................................xv

CHAPTER 1. INTRODUCTION ................................................................................................1

1.1 Myxobacteria.....................................................................................................................1

1.2 Myxobacteria Biosynthetic Capacity...............................................................................2

1.3 Myxobacteria Genomics ..............................................................................................3

1.4 Classification of Myxobacteria......................................................................................4

1.5 Outline of this work.......................................................................................................6

CHAPTER 2. MARINE MYXOBACTERIA: A FEW GOOD HALOPHILES...............................11

2.1 Abstract ..........................................................................................................................12

2.2 Introduction.....................................................................................................................12

2.3 Characteristics Unique to Marine Myxobacteria..........................................................13

2.4 Haliangium ochraceum................................................................................................17

2.5 Enhygromyxa salina.......................................................................................................19

2.6 Plesiocystis pacifica.......................................................................................................20

2.7 Haliangium tepidum.......................................................................................................21
CHAPTER 3. IDENTIFICATION OF A SOLO ACYLHOMOSERINE LACTONE SYNTHASE FROM THE MYXOBACTERIUM ARCHANGIUM GEPHYRA

3.1 Abstract .......................................................... 25
3.2 Introduction.......................................................... 25
3.3 Materials and Methods.............................................. 27
  3.3.1 Cultivation of Archangium gephyra.......................... 27
  3.3.2 Bioinformatic Assessment of AgpI and VitI.................. 27
  3.3.3 Autoinducer Binding Site Search............................. 28
  3.3.4 Heterologous Expression of AgpI and VitI in E. coli........ 28
  3.3.5 Metabolite Extraction and Analysis.......................... 29
  3.3.6 Exogenous AHL Exposure Experiments...................... 30
  3.3.7 GNPS Dataset................................................... 30
3.4 Results................................................................... 30
  3.4.1 AgpI is Highly Homologous to Functional AHL Synthases..... 30
  3.4.2 Absence of a Cognate AHL Receptor in the Genome of A. gephyra ... 34
  3.4.3 A. gephyra does not Produce AHLs During Axenic Cultivation...... 36
  3.4.4 Exogenous AHLs do not Activate AgpI.......................... 36
  3.4.5 Heterologous Expression of AgpI Confirms Functional Production of AHLs ... 37
3.5 Discussion.................................................................. 39
CHAPTER 4. ASSESSMENT OF EVOLUTIONARY RELATIONSHIPS FOR PRIORITIZATION OF MYXOBACTERIA FOR NATURAL PRODUCT DISCOVERY

4.1 Abstract ........................................................................................................44
4.2 Introduction....................................................................................................44
4.3 Materials and Methods................................................................................48
  4.3.1 Bacterial Strains and Growth Conditions...............................................48
  4.3.2 Isolation of Environmental Myxobacteria...............................................48
  4.3.3 Genomic DNA Isolation, Sequencing, Assembly, and Annotation...........49
  4.3.4 Comparative Genomic Studies.................................................................50
  4.3.5 BIG-SCAPE Analysis...............................................................................50
4.4 Results...........................................................................................................51
  4.4.1 Comparative Genomics and Taxonomic Assessment of Archangium Primigenium, Chondrococcus Macrosporus, and Environmental Isolates.......................51
  4.4.2 Biosynthetic Potential and Genus Level Correlations...............................54
4.5 Discussion......................................................................................................60

CHAPTER 5. RECLASSIFICATION OF ARCHANGIUM PRIMIGENIUM AS MELITTANGIUM PRIMIGENIUM COMB. NOV. AND ASSESSMENT OF ITS BIOSYNTHETIC POTENTIAL..................................................................................62

5.1 Abstract ........................................................................................................62
5.2 Introduction....................................................................................................62
5.3 Materials and Methods................................................................................63
5.3.1 Bacterial Strains and Growth Conditions.................................................................63

5.3.2 Genomic DNA Isolation, Sequencing, Assembly, and Annotation.........................64

5.3.3 Polyphasic Taxonomy of Archangium primigenium ATCC 29037..........................64

5.3.3.1 Phenotypic Analysis..........................................................................................64

5.3.3.1.a Microscopy and morphological examination.................................................64

5.3.3.1.b Growth response to pH, temperature, and antibiotics....................................64

5.3.3.1.c Enzymatic tests............................................................................................65

5.3.3.2 Genotypic Analysis.........................................................................................65

5.3.4 Metabolic Profiling of Archangium primigenium ATCC 29037............................66

5.3.4.1 Seed cultures and fermentation for secondary metabolites extraction..................66

5.3.4.2 Extraction and Fractionation...........................................................................67

5.3.4.3 Metabolic analysis..........................................................................................67

5.3.4.4 GNPS Analysis..............................................................................................67

5.3.4.5 Antibacterial Activity ....................................................................................68

5.4 Results and Discussion...............................................................................................68

5.4.1 Phenotypic Analysis.............................................................................................68

5.4.2 Genotypic analysis..............................................................................................70

5.4.3 Estimation of biosynthetic potential of Archangium. primigenium.......................74

5.4.4 Secondary Metabolites Extraction .......................................................................75

5.5 Conclusion and Recommendations............................................................................76

LIST OF REFERENCES.................................................................................................78
LIST OF APPENDICES .................................................................................................................. 98

Appendix A. Sequence data for codon optimized AgpI construct purchased from Genscript and Sequence data for codon optimized VitI construct purchased from Genscript................................. 99

Appendix B. Locations of soil samples used for environmental strains isolation...................... 100

Appendix C. A list of all myxobacteria and their accession numbers used in BiG-SCAPE analysis.....................................................................................................................................................101

Appendix D. Minimum evolution tree from the whole genomes of Corallococcus

macrosorus ATCC 29039, NCCRE002, NCSPR001 and different members of genus Corallococcus using the GBDP approach........................................................................................................ 102

Appendix E. Minimum evolution tree from the whole genomes of SCHIC003, SCPEA002, and different members of Myxococcus and Pyxidicoccus using the GBDP approach.................................................................................... 103

Appendix F. API 20 NE and API Zym results for A. primigenium ATCC 29037.......................... 104

Appendix G. Dotplots representing genomic alignments of Archangium primigenium ATCC® 29037 with fully sequenced myxobacteria using YASS algorithm. Forward matches are shown in green, while reverse matches are shown in red........ 105

Appendix H. Molecular network of fractions from RP-SPE column as visualized with

Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional with similarity to MS/MS from GNPS libraries................................................................. 106

Appendix I. Molecular network of fractions from RP-SPE column as visualized with
Cytoscape. Each node is represented with pie-chart that represent the fractions in which that MS/MS appears.

Appendix J. Antibacterial activities of SPE and Sephadex LH-20 column chromatography fractions against different bacteria.

Appendix K. Molecular network of fractions from Sephadex LH-20 column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appear.

Appendix L. Molecular network of fraction 3 from Sephadex LH-20 column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appear.

Appendix M. Molecular network of fraction 4 from Sephadex LH-20 column as visualized Cytoscape showing the peptide rich fractions. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is
represented with pie-chart that represent the fractions in which that MS/MS appear.

Appendix N. Molecular network of fraction 4 from Sephadex LH-20 column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appear.
LIST OF TABLES

Table 1.1  The order of Myxococcales with suborders and families……………………………7
Table 1.2  The proposed phylum of Myxococcota (myxobacteria) with orders and families (as of
March 2021)……………………………………………………………………………………………..9
Table 2.1  Putative osmolyte synthases and transporters from sequenced marine myxobacteria.
BCCT—betaine/carnitine/choline transporter………………………………………………………14
Table 4.1  Comparison of biosynthetic gene clusters (BGCs) located on contig edges from
previously sequenced members of the genus Corallococcus. Included BGC data
provided by whole genome analysis using antiSMASH version 5……………………………46
Table 4.2  Genome properties and general features of myxobacteria under investigation in this
study…………………………………………………………………………………………………………53
Table 4.3  Overview of BiG-SCAPE BGC sequence similarity networks of the six strains under
investigation in this study……………………………………………………………………………60
Table 5.1  Biochemical Activities of A. primigenium ATCC 29037 as tested by API 20 NE and
API Zym……………………………………………………………………………………………71
Table 5.2  16S rRNA identity, ANI, and dDDH values for pairwise comparisons between A.
primigenium with the most similar fully sequenced myxobacteria………………………………72
Table 5.3  YASS analysis of nine sequenced myxobacteria with A. primigenium………………….74
LIST OF FIGURES

Figure 1.1 Chemical structures of epothilone B and ixabepilone (Ixempra®). In epothilone B, an oxygen atom is in the macrolide ring; in ixabepilone, that oxygen atom is replaced with a nitrogen atom. .................................................................2

Figure 1.2 Workflow of genome-based classification at the species level. .........................8

Figure 2.1 (A) Phylogenetic tree of cultivable Myxococcales with the discussed myxobacteria bolded; (B) Phylogenetic tree of myxobacteria, that includes the MMC and other myxobacteria. Using MEGA7, 16s rRNA sequences were aligned with ClustalW using general settings, and a phylogenetic tree was generated using the neighbor-joining method. ..................................................................................16

Figure 2.2 Haliangium ochraceum secondary metabolites. Of note, cis isomers about the epoxide moiety have also been reported for all of the haliangicins. .................................18

Figure 2.3 Enhygromyxa salina secondary metabolites and hydroxyectoine. .........................19

Figure 2.4 Secondary metabolites from potential marine myxobacteria. .........................22

Figure 3.1 (A) Cluster 32 from A. gephyra deposited in the antiSMASH database which includes the putative AHL synthase, agpl. (B) Genomic context for vitI from Vitiosangium sp. All annotated features within NCBI are labeled and all hypothetical features are in grey. Percentage GC content for each gene within the cluster provided for comparison and depicted in parentheses. ........................................................................31

Figure 3.2 (A) Alignment of LuxI synthases including Agpl and VitI with conserved residues boxed in black. (B) Alignment of myxobacterial features homologous to Agpl with
asterisks (*) indicating conserved residues associated with the LasI autoinducer
domain (COG3916) and red lines (-) indicating residues from the LasI domain that
are not conserved. (C) Minimum Evolution tree including AgpI and VitI rendered in
MEGA7 using ClustalW aligned with AHL synthases experimentally confirmed to
produce AHLs (68). Branch lengths ≤ 0.2 not depicted……………………………………32

Figure 3.3 Sequence similarity network rendered by EFI-EST analysis of AgpI amino acid
sequence data with AgpI (red diamond) and VitI (blue diamond) indicated. LuxI-type
AHL synthases are represented as grey nodes (a total of 1,001 nodes clustered in 2
groups) with grey edged connecting the nodes (a total of 124,346 edges). To reduce
complexity all nodes with ≥90% sequence similarity are represented as an individual
aggregate node. For Cytoscape visualization, Attribute Circle Layout was applied….33

Figure 3.4 Genomic region of Variosangium. sp. GDMCC 1.1324 including the putative LuxR-
type receptor (WP_108076247.1) and surrounding genes. All annotated features within
NCBI are labeled and all hypothetical features are in grey……………………………………..35

Figure 3.5 Molecular family from the molecular network of LC–MS/MS datasets from extracts of
heterologous E. coli expressing AgpI rendered by GNPS (50). Detected m/z values
from raw data positioned over each node with node diameter depicting associated
intensities for each AHL……………………………………………………………………………38

Figure 3.6 Extracted-ion chromatographs 228-229 m/z (Top) and 242.1-242.2 m/z (Bottom)
depicting presence of C8-AHL and C9-AHL, respectively, in extracts from E. coli
K207-3 heterologously expressing AgpI (blue) and VitI (red) and absence of C8-AHL
and C9-AHL in extracts from *E. coli* K207-3 + pET28b (green) and *E. coli* K207-3 with no plasmid (pink)………………………………………………………………………………………………..39

Figure 3.7 MS/MS fragmentation spectra with diagnostic fragments indicated for each AHL detected in extracts from heterologous *E. coli* expressing AgpI…………………………………40

Figure 4.1 Myxobacterial fruiting bodies from strains NCSPR001, NCCRE002, SCHIC003, SCPEA002, and the strains *C. macrosporus* ATCC 29039 and *A. primigenium* ATCC 29037…………………………………………………………………………………………………47

Figure 4.2 Minimum evolution tree from the whole genomes of different myxobacteria including the six strains under investigation in this study using the GBDP approach. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 100.0%. Branch pseudo-bootstrap less than 50% are not shown. The numbers below branches are branch lengths scaled in terms of GBDP distance formula d5. The tree was rooted at the midpoint………………………………………………………………………………………………52

Figure 4.3 Minimum evolution tree from the whole genomes of *A. primigenium* and different members of the family *Archangiaceae* using the GBDP approach. To the right of the tree a heat map shows dDDH and ANI values from pairwise comparisons of *A. primigenium* with different myxobacteria. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 100.0%. The numbers below branches are branch lengths scaled in terms of GBDP distance formula d5. Branch lengths less than 0.05 are not
shown. The tree was rooted at the midpoint. *M. xanthus* DK 1622 is included in the GBDP tree as an outgroup.

Figure 4.4 dDDH and ANI from pairwise whole-genome comparisons of *C. macrosporus* ATCC 29039, NCCRE002, and NCSPR001 with sequenced members of the genus *Corallococcus* (Top), and SCPEA002 and SCHIC003 with sequenced members of the genera *Myxococcus* and *Pyxidicoccus* (Bottom). Double-bordered box represents species that would be designated as the same using the designated method (< 95% for ANI and 70% for dDDH). Single-bordered represents the highest ANI/dDDH value for each myxobacterium under investigation.

Figure 4.5 Common specialized metabolites from myxobacteria associated with characterized BGCs present in the six investigated strains of myxobacteria. * Myxochelin BGC not present in SCHIC003 genome data. # Myxoprincomide BGC not present in *A. primigenium* genome data.

Figure 4.6 (A) Specialized metabolites produced by members of the genera *Corallococcus*, *Melittangium*, and *Pyxidicoccus* with no associated BGCs observed in any of the six investigated myxobacterial strains. (B) Comparisons of the rhizopodin, phenalamide A2, and fulvuthiacene BGCs from SCHIC003 genome data and the characterized pathways from *M. stipitatus* and *M. fulvus*. All SCHIC003 gene products, excluding RizC, had coverages ≥99% with the indicated homolog. * RizC is located on a contig edge and is incomplete in SCHIC003 genome data.

Figure 4.7 BiG-SCAPE BGC sequence similarity networks (c = 0.3) as visualized with
Cytoscape. The network is generated from *A. primigenium*, *C. macrosporus*, NCCRE002, NCSPR001, SCHIC003, SCPEA002, and all myxobacteria belonging to the Cystobacterineae suborder with genomes deposited in NCBI. Each node represents one BGC identified by AntiSMASH 5.0, where the colors and shapes of the nodes represent different genera and AntiSMASH-predicted classes, respectively. Nodes representing BGCs from newly sequenced myxobacteria included in this study are enlarged. BGCs included as singletons in the original BiG-SCAPE analysis were removed.

Figure 5.1 Colony and cellular morphology of *A. primigenium* ATCC 29037. Left: Colony morphology of *A. primigenium* swarm with; Right: Cellular morphology as visualized with scanning electron microscopy (SEM).

Figure 5.2 Minimum evolution tree of *A. primigenium* and different members of the family *Archangiaceae* (A) based on 16S rDNA gene sequence (B) using the GBDP approach. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 100.0%. The tree was rooted at the midpoint. *Mycococcus xanthus* DK 1622 is included in the trees as an outgroup.

Figure 5.3 A) antiSMASH v6.0 predicted biosynthetic gene clusters form the genome of *A. primigenium*. B) A representation of the location of different BGCs on *A. primigenium* genome.
CHAPTER 1. INTRODUCTION

1.1 Myxobacteria

Myxobacteria are aerobic, Gram-negative bacteria with rod-shaped cells typically 0.5 microns in width by 3–8 microns in length.\textsuperscript{1–5} They are mostly known to be ubiquitous soil-dwelling bacteria.\textsuperscript{1–8} However, different studies have shown that they can survive moderate to extreme environments such as freshwater and saltwater, slightly acidic and basic soils, and anaerobic and hypothermic habitats.\textsuperscript{9–14} Importantly, they also display social behaviors which are unique among bacteria and allow them to live in large swarms.\textsuperscript{1–5}

Most myxobacteria swarm and collectively hunt other microorganisms, including bacteria, fungi, and protozoa.\textsuperscript{5–8} Predation is achieved by the secretion of antimicrobial metabolites and digestive enzymes that kill and lyse their prey microorganisms.\textsuperscript{15–17} Under starvation conditions, the rod-shaped vegetative cells coordinate their motility to form multicellular aggregates known as fruiting bodies.\textsuperscript{1–3} Fruiting bodies have various shapes depending on species and are usually visible to the human eye. Within these bodies, the cells start to develop into dormant spherical immotile cells with thick cell walls known as myxospores which can survive harsh conditions and germinate when nutrients become available once again.\textsuperscript{18–20}

Myxobacteria glide in swarms over solid surfaces using two distinct motility systems: A (adventurous) motility and S (social) motility.\textsuperscript{1–5} A-motility appears on hard surfaces where individual cells at the swarm edges glide away to explore the surroundings. This is accomplished by the translocation of protein cytoskeletons in a helical loop against surfaces hydrated by self-secreted polysaccharide slime.\textsuperscript{21} On the other hand, S-motility is an effective, collective effort of
swarming that enables cells in groups to act as a coordinated swarm during hunting and predation. This type of motility is achieved by the extension and retraction of type-IV pili at the leading pole of the cell.\textsuperscript{22–24} It is to be noted that the term "swarming" is used in its general sense to describe a process "in which motile organisms actively spread on the surface of a suitably moist solid medium".\textsuperscript{20}

1.2 Myxobacteria Biosynthetic Capacity

Different secondary metabolites with bioactivities of pharmaceutical interest have been isolated from different strains of myxobacteria.\textsuperscript{25–30} One of the most celebrated compounds of myxobacterial origin is epothilone B which was isolated from \textit{Sorangium cellulosum} and shown to have cytotoxic activity.\textsuperscript{31} A semisynthetic analog of epothilone B, ixabepilone (Ixempra\textsuperscript{®}), is an FDA-approved treatment for breast cancer that has a myxobacterial origin (Figure 1.1).\textsuperscript{32} As myxobacteria are slow-growing and generally difficult to cultivate under laboratory conditions and more myxobacterial genomes are getting sequenced, most recent myxobacterial metabolites discovery efforts are being conducted by applying genome-mining approaches.\textsuperscript{26,33–38}

![Figure 1.1 Chemical structures of epothilone B and ixabepilone (Ixempra\textsuperscript{®}). In epothilone B, an oxygen atom is in the macrolide ring; in ixabepilone, that oxygen atom is replaced with a nitrogen atom.](image)

Many studies have shown that myxobacteria are one of the most gifted microbes with large genomes having multiple biosynthetic gene clusters (BGCs) that encode multiple secondary
metabolites. Medema and coworkers define a BGC as “a physically clustered group of two or more genes in a particular genome that together encode a biosynthetic pathway for the production of a specialized metabolite (including its chemical variants)”.

Different bioinformatics tools and platforms can be applied to identify and analyze secondary metabolite BGCs from myxobacteria. antiSMASH (antibiotics & Secondary Metabolite Analysis Shell) is the most versatile software tool to predict and identify BGCs from genome sequences. antiSMASH predicts BGC types through the implementation of a sequence alignment-based profile in a Hidden Markov Model (HMM) of genes that are known to be specific for certain types of BGCs. According to the antiSMASH webserver, the current version can identify 70 different BGCs. BiGSCAPE (Biosynthetic Gene Similarity Clustering and Prospecting Engine) platform is used to determine the gene cluster families (GCFs) of different BGCs and then generate a sequence similarity network to analyze the relationship between these BGCs.

1.3 Myxobacteria Genomics

Myxobacteria are known to possess one of the largest genomes amongst bacteria (9-13 Mbp) with a DNA of high GC content (percentage of G–C base pairs in the genome) of 66-72%. The first myxobacterial genome to be sequenced and described in the literature in 2006 is of Myxococcus xanthus DK1622. This was followed by the sequencing of other myxobacteria including Anaeromyxobacter dehalogenans 2CP-C in 2006, and Plesiocystis pacifica SIR-1, Sorangium cellulosum So ce56, and Haliangium ochraceum SMP-2 in 2007.

As new sequencing technologies have become more accessible and affordable, more myxobacteria have been sequenced over the years. Illumina technology is a short-read DNA,
inexpensive sequencing technique that results in high-quality draft genomes. As most secondary metabolites BGCs in bacteria are large, Illumina sequencing is not efficient for genome-mining drug discovery. Recently, a long-read DNA sequencing method developed by Pacific Biosciences (PacBio) has been developed to obtain high-quality sequences of very large genomes from myxobacteria and other bacteria.

Since 2006 there has been an exponential increase in publicly-available myxobacterial genome sequences with a 7-fold increase from 2016 to 2020. This increase is the second-highest among all bacterial strains (the highest is Micromonospora with a 7.67-fold increase from 2016 to 2020). This can be attributed to the fact that myxobacterial genome mining led to the realization that myxobacteria are one of the most gifted microbes for drug discovery. However, up until July 2020, only 11.5% of publicly available myxobacterial genomes were complete. High-quality complete genomes enable the full assembly of large BGCs. Thus, it is important to have finished quality myxobacterial genomes to identify novel secondary metabolite BGCs.

1.4 Classification of Myxobacteria

Due to their fruiting bodies, myxobacteria were considered to be fungi when they were first isolated and studied in 1809. Eighty-three years later, botanist Roland Thaxter was the first to recognize them as a distinct group of bacteria and named them “Myxobacteria”. The name comes from the Greek word myxo, meaning slime or mucus as myxobacterial swarms are covered with a slimy self-secreted extracellular matrix (polysaccharides).

While their distinguished morphological characteristics were the basis for their isolation and classification, the current classification of myxobacteria is mainly based on a broad polyphasic comparison of the novel bacteria with pre-existing taxa and/or Type strain. Type strain is
defined as the strain that was studied when the species was described for the first time in the literature.\textsuperscript{56,57} This comparison is in consideration of cell morphology, fruiting body morphology, colony morphology, 16S rRNA conservation, DNA–DNA hybridization (DDH), and biochemical/physiological assays.\textsuperscript{54,58–65} This places the order of Myxococcales (Myxobacteria) within the phylum of Proteobacteria and divides it into three suborders: Cystobacterineae, Nannocystineae, and Sorangiineae (Table 1.1).\textsuperscript{1,52,66}

16S rRNA gene comparison and DNA–DNA hybridization (DDH) have been considered the ‘the gold standard’ in prokaryote taxonomy.\textsuperscript{65,67,68} The 16S rRNA gene can be amplified and sequenced to compare different bacteria.\textsuperscript{11} The general principle of experimental DNA-DNA is to generate hybrid sequences of double-strand DNA that resulted from the shearing, dissociation, and subsequent annealing of double-stranded DNA from the two organisms under investigation. The melting point of the hybridized sequences reflects the high degree of similarity between the two genomes.\textsuperscript{62} With the advent of whole-genome sequencing and bioinformatics, modern tools are increasingly being applied to support the classification of bacteria. Recently, overall genome-related index (OGRI) bioinformatics tools have been applied for taxonomic resolution of closely related species. Among the OGRIs, average nucleotide identity (ANI) and digital DDH (dDDH) have been the most widely used.\textsuperscript{62,65,68,69} The ANI (average nucleotide identity) value is calculated computationally from pairwise comparisons of all sequences shared between two genomes, not just the 16S rRNA gene.\textsuperscript{68} dDDH is calculated using Genome-to-Genome Distance Calculator (GGDC).\textsuperscript{68} Another bioinformatics approach is the phylogenomic treeing which can be achieved by the analysis of multiple genes, instead of a single gene such as 16S (phylogenetic treeing).

The general procedure for genome-based bacterial classification starts with complete 16S rRNA gene sequencing and comparison. If two species show 98.7\% or higher 16S similarity, then
ANI and dDDH can be calculated. The two species are distinct if ANI and dDDH values are < 95–96 and < 70%, respectively (Figure 1.2). Some researchers suggest ANI > 70–80 for two species to be considered of the same genera.

The ongoing isolation of new myxobacteria and the reclassification of some known myxobacteria reveal the importance of updating the current taxonomy status of myxobacteria. Recently, a genomic analysis of 16248 genomes of Proteobacteria has suggested the reclassification of an entire Phylum of Proteobacteria. The study proposes that the members of myxobacteria, currently classified as an order named (Myxococcales) within the class Deltaproteobacteria within the phylum Proteobacteria, should be reclassified as an independent phylum Myxococcota that is further divided into four orders (Myxococcales, Polyangiales, Nannocystales, Haliangiales) comprising seven genera and 19 families (Table 1.2).

1.5 Outline of this work

Whole-genome sequencing of bacteria has become a more accessible, affordable tool in modern microbiology. The powerful combination of whole-genome sequencing and bioinformatics-driven analysis of sequence data has delivered unexpected insights into microbial communities. Myxobacteria are a unique group of Proteobacteria that are best known for their large genomes, well-coordinated social behavior, and ability to predate other microbes. Genome sequencing of myxobacteria has demonstrated that they are a “gifted” taxon as their genomes house multiple biosynthetic gene clusters (BGCs) to produce specialized metabolites with drug-like properties. Here, we present five chapters where we utilized genomics and bioinformatics to further study different species of myxobacteria.
Table 1.1 The order of Myxococcales with suborders and families.¹

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Class-Order</th>
<th>Sub-Order</th>
<th>Family</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>Cystobacterineae</td>
<td>Myxococcaceae</td>
<td>Myxococcus</td>
</tr>
<tr>
<td></td>
<td>- Myxococcales</td>
<td></td>
<td></td>
<td>Corallococcus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pyxidicoccus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Angiococcus</td>
</tr>
<tr>
<td></td>
<td>Archangiaceae</td>
<td></td>
<td></td>
<td>Cystobacter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Archangium</td>
<td></td>
<td>Stigmatella</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vitosangium</td>
<td></td>
<td>Melittangium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hyalangium</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vulgatibacteraceae</td>
<td>Vulgatibacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anaeromyxobacteraceae</td>
<td>Anaeromyxobacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorangiineae</td>
<td>Polyangiaceae</td>
<td></td>
<td>Chondromyces</td>
<td>Sorangium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sorangium</td>
<td>Minicystis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phaseliangium</td>
<td>Phaselicystis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pajaroellobacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Labilitrichaceae</td>
<td>Labilitrix</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sandaracinaceae</td>
<td>Sandaracinus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nannocystineae</td>
<td>Nannocystaceae</td>
<td>Nannocystis</td>
<td></td>
<td>Plesiocystis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plesiocystis</td>
<td>Enhygromyxa</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Enhygromyxa</td>
<td>Pseudoiygromyxa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haliangiaceae</td>
<td>Haliangium</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2 explores marine myxobacteria, their unique features, and their potential as a source of novel secondary metabolites. This chapter in full, is a reprint of the material as it appears in: Albataineh, H.; Stevens, D. Marine Myxobacteria: A Few Good Halophiles. *Marine Drugs* 2018, 16 (6), 209.

Chapter 3 describes the bioinformatic analysis and functional assessment of the first characterized solo acylhomoserine lactone synthase (AHL) in myxobacteria. The results suggest that Agpl of *Archangium gephyra* (DSM 2261) is a functional, orphaned AHL synthase. While the regulatory mechanism and its utility to the predatory myxobacterium remain unknown, the result provides a unique perspective and supports the continued investigation of small molecule interactions that contribute to microbial community structures. This chapter in full, is a reprint of the material as it appears in: Albataineh, H.; Duke, M.; Misra, S. K.; Sharp, J. S.; Stevens, D. C. Identification of a Solo Acylhomoserine Lactone Synthase from the Myxobacterium *Archangium gephyra*. *Scientific Reports* 2021, 11.
Table 1.2 The proposed phylum of *Myxococcota* (myxobacteria) with orders and families (as of March 2021).  

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myxococcota</td>
<td>Myxococcia</td>
<td>Myxococcales</td>
<td>Myxococcaceae</td>
<td>Myxococcus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Corallococcus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cystobacter</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Archanigium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stigmatella</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hyalangium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Melittangium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pyxidicoccus</td>
</tr>
<tr>
<td></td>
<td>Vulgatibacteraceae</td>
<td>Vulgatibacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anaeromyxobacteraceae</td>
<td>Anaeromyxobacter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyangia</td>
<td>Polyangiales</td>
<td>Polyangiaceae</td>
<td>Chondromyces</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sorangium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Minicystis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Labilitrix</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pajaroollobacter</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Phaselicystis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sandaracinaceae</td>
<td>Sandaracinus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nannocystales</td>
<td>Nannocystaceae</td>
<td></td>
<td>Nannocystis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plesiocystis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Enhygromyxa</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pseudohygromyxa</td>
<td></td>
</tr>
<tr>
<td>Haliangiales</td>
<td>Haliangiaceae</td>
<td></td>
<td>Haliangium</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4 aims to utilize long-read genomes and modern comparative genomics of two previously classified mycobacteria and four newly isolated myxobacteria for the assessment of evolutionary relationships of myxobacteria and the determination of biosynthetic potential for prioritization of natural product discovery. This chapter in full, is a reprint of the material as it appears in: Ahearne, A.; Albataineh, H.; Dowd, S. E.; Stevens, D. C. Assessment of Evolutionary Relationships for Prioritization of Myxobacteria for Natural Product Discovery. *Microorganisms* 2021, 9 (7), 1376.

Chapter 5 deals with the characterization and description of *A. primigenium* ATCC 29037 by applying a polyphasic approach, including substrate utilization, sugar assimilation assays, and antibiotic resistance assays. Utilizing mass spectrometry data, we assess the biosynthetic potential of this strain, and we describe our attempt to isolate pharmacologically active metabolites.
CHAPTER 2. MARINE MYXOBACTERIA: A FEW GOOD HALOPHILES

This chapter in full, is a reprint of the material as it appears in: Albataineh, H.; Stevens, D. *Marine Drugs* 2018, 16 (6), 209. This chapter has been formatted to follow the University of Mississippi’s guidelines for dissertation preparation.

H.A. and D.C.S. compared marine myxobacterial genomes via BLAST and wrote the manuscript.

DOI: 10.3390/md16060209
2.1 Abstract

Currently considered an excellent candidate source of novel chemical diversity, the existence of marine myxobacteria was in question less than 20 years ago. This review aims to serve as a roll call for marine myxobacteria and to summarize their unique features when compared to better-known terrestrial myxobacteria. Characteristics for discrimination between obligate halophilic, marine myxobacteria and halotolerant, terrestrial myxobacteria are discussed. The review concludes by highlighting the need for continued discovery and exploration of marine myxobacteria as producers of novel natural products.

2.2 Introduction

In a myxobacterial ecology review published in 1999, Hans Reichenbach asked, “Are there marine myxobacteria?”. Reichenbach’s succeeding paragraph provides insight into the uncertainty surrounding marine myxobacteria prior to the routine practice of molecular taxonomy.9 Two of the major factors contributing to this obscurity were the incorrect assignment of marine Bacteroidetes as lower order myxobacteria or ‘Myxobacteria imperfecta’ and the ubiquitous distribution and reported isolations of myxobacteria, with varying halotolerances among interfacial environments, such as sediments from beaches and shores.9,73,74 While the 16s rRNA sequence analysis has mostly addressed the former, determining that what were considered lower-order marine myxobacteria, were instead Bacteroidetes from the genera Cytophaga and Flavobacterium, the latter continues to obfuscate the distinction between marine and terrestrial myxobacteria.10,75,76 Literature concerning the halotolerant myxobacterium Myxococcus fulvus HW-1 (ATCC BAA-855) typifies this issue.14,67 While Myxococcus fulvus HW-1 has been reported to tolerate a salinity as high as 3% and is listed in the World Register of Marine Species
(WoRMS), the strain displays attenuated morphologies and social behaviors that are typical of myxobacteria, such as a fruiting body formation on agar media with low concentrations of seawater or salts. This review aims to clarify such confusion by highlighting the unique characteristics of halophilic myxobacteria when compared to better-known halotolerant soil myxobacteria, and to encourage further discovery and investigation of marine myxobacteria as a source of structurally unique secondary metabolites.

2.3 Characteristics Unique to Marine Myxobacteria

The recent retrospective analysis of natural product discovery trends reported by Pye et al. concluded that marine organisms are, at least upon discovery, productive sources of novel chemical diversity. Considering this observation, combined with the abundance of biologically active myxobacterial metabolites, we anticipate that an investigation of marine myxobacteria, as producers of secondary metabolites with unique molecular scaffolds and activities, will become a priority for drug discovery effort. While limited by the scarcity of cultivable marine myxobacteria, the following characteristics are meant to differentiate known terrestrial, halotolerant myxobacteria from marine, halophilic myxobacteria. These characteristics are centered on halophilicity, multicellular behaviors that have been observed at saline cultivation conditions, phylogeny, and ecology. Halophilic bacteria rely on the cellular accumulation of organic osmolytes to prevent dehydration in osmotic environments, such as seawater. Recently, differing strategies for osmolyte accumulation were reported for Enhygromyxa salina SWB007 and Pleiocystis pacifica SIR-1\(^T\). While \(P.\ pacifica\) accumulates amino acids to offset osmotic stress, \(E.\ salina\) produces the ubiquitous osmolytes betaine and hydroxyectoine.
Table 2.1 Putative osmolyte synthases and transporters from sequenced marine myxobacteria. BCCT—betaine/carnitine/choline transporter.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Gene</th>
<th>NCBI Reference Sequence</th>
<th>Length (aa)</th>
<th>Highest Homology</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>H. ochraceum</em></td>
<td>ectC</td>
<td>WP_012827762.1</td>
<td>142</td>
<td>ectoine synthase (Hydrogenophaga crassostreae)</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>betT</td>
<td>WP_012829907.1</td>
<td>538</td>
<td>BCCT family transporter (Desulfovermiculus halophilus)</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td>sodium/proline symporter</td>
<td>WP_012825704.1</td>
<td>598</td>
<td>hypothetical protein (Hymenobacter terrenus)</td>
<td>49%</td>
</tr>
<tr>
<td><em>E. salina</em></td>
<td>ectC</td>
<td>WP_106088059.1</td>
<td>126</td>
<td>ectoine synthase (Blastopirellula marina)</td>
<td>63%</td>
</tr>
<tr>
<td></td>
<td>ectD</td>
<td>AMH38938.1</td>
<td>298</td>
<td>ectoine hydroxylase (Blastopirellula marina)</td>
<td>59%</td>
</tr>
<tr>
<td></td>
<td>betT</td>
<td>AMH38943.1</td>
<td>492</td>
<td>BCCT family transporter (Spingomonas sp. Leaf30)</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td>sodium/proline symporter</td>
<td>WP_106088061.1</td>
<td>481</td>
<td>sodium/proline symporter (Rubinisphaera brasiliensis)</td>
<td>59%</td>
</tr>
<tr>
<td></td>
<td>sodium/glutamate symporter</td>
<td>KIG18073.1</td>
<td>469</td>
<td>hypothetical protein (P. pacifica)</td>
<td>64%</td>
</tr>
<tr>
<td><em>P. pacifica</em></td>
<td>betT (BCCT family transporter)</td>
<td>EDM75025.1</td>
<td>512</td>
<td>BCCT family transporter (Spingomonas sp. Leaf10)</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>sodium/proline symporter</td>
<td>WP_006976305.1</td>
<td>484</td>
<td>sodium/proline symporter (E. salina)</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td>sodium/glutamate symporter</td>
<td>WP_006969752.1</td>
<td>478</td>
<td>sodium/glutamate symporter (E. salina)</td>
<td>64%</td>
</tr>
</tbody>
</table>

Interestingly, all of the sequenced myxobacterial halophiles possess a betaine/carnitine/choline transporter (BCCT) or BetT homolog, while the sequenced terrestrial myxobacteria do not, regardless of the reported halotolerances (Table 2.1). The gene loci that encode solute biosynthetic pathways, such as the confirmed ectoine/hydroxyectoine cluster from *E. salina* SWB007, cannot currently be considered as critical aspects of marine myxobacteria, as *P. pacifica* SIR-1T has no such pathway and instead relies on the accumulation of glutamate and glycine to prevent osmotic stress.79
Excluding *Myxococcus fulvus* HW-1, the only cultivable myxobacteria found to be obligate halophiles have been associated with the ‘marine’ moniker. While among the order of the *Myxococcales* obligate halophilicity is wholly unique to marine myxobacteria, all myxobacteria participate in unique multicellular behaviors, such as fruiting body formation, social swarming, and organized predation.\(^2,80,81\) Salinity impedes fruiting body formation of the vast majority of terrestrial and halotolerant myxobacteria.\(^14\) Instead, halotolerant myxobacteria adopt a less complicated, unicellular growth strategy that is independent of cell density, when grown at salinities reflective of seawater conditions.\(^14\) While this strategy provides the resilience required for halotolerant myxobacteria to survive in sandy beaches and shoreline soils, it impedes their ability to thrive in seawater. An outlier, halotolerant *Pseudenygromyxa salsuginis* SYR-2\(^T\), designated a ‘brackish water myxobacterium’, displays fruiting bodies similar to *E. salina* SHK-1\(^T\) and *P. pacifica* SIR-1\(^T\) at various levels of salinity.\(^82–84\) Synchronized motility, referred to as swarming or gliding, is a defining feature of myxobacteria, has been observed for all marine myxobacteria as well as halotolerant myxobacteria, such as *Myxococcus fulvus* HW-1, when grown at varying salinities.\(^29,81–85\) Halophilic myxobacteria swarm in radial patterns with varying amounts of etching occurring on the surface of agar media.\(^29,82–85\)

Typical of terrestrial myxobacteria, marine myxobacteria do not form radial veins or distinct waves when swarming. Instead, marine myxobacteria aggregate at peripheral bands along the outer circumference of swarms.\(^29,82–85\) Unlike the predation efforts of the model myxobacterium *Myxococcus xanthus*, organized predation strategies of marine myxobacteria have yet to be explored in detail. While all halophilic myxobacteria discovered to date can lyse a variety of Gram-negative bacteria, only *H. ochraceum*, *H. tepidum*, and *Enhygromyxa niigatensis* are able to lyse *Saccharomyces cerevisiae* cells.\(^29,82–85\) The suborder *Nannocystineae* exclusively consists
of halotolerant and halophilic myxobacteria, including all of the discovered cultivable halophilic myxobacteria, with the only potential outlier being the lesser studied, terrestrial *Kofleria flava* (Figure 2.1 A).

Yet again exceptional, *Myxococcus fulvus* HW-1 is the only purported marine myxobacterium to instead belong to the suborder *Cystobacterineae* (Figure 2.1 A). However, Brinkhoff et al. recently reported a distinct cluster exclusively of marine myxobacteria (Figure 2.1 B).\(^{12}\) Aptly designated as the marine myxobacteria cluster (MMC), associated marine myxobacteria were observed primarily from sediment samples and water column samples near the sediment surface worldwide, at salinities ranging from brackish to marine.\(^{12}\) Interestingly, the myxobacteria distinctly clustered between the MMC and the defined suborders *Nannocystineae*, *Sorangiineae*, and *Cystobacterineae* were from a variety of diverse habitats including volcanic sediment, a wastewater treatment plant, a glacier, a uranium mining waste pile, a microbial biofilm,
a hypersaline microbial mat, and various marine sediments. The prevalence of halotolerant and halophilic myxobacteria within the suborder *Nannocystineae*, the observation of clustered myxobacteria from fluid habitats, and the exclusivity of the MMC provide a unique phylogenetic landscape within the order *Myxococcales*, where physiological adaptability to environmental volatility seems apparent. The observation of phylogeographic separation of marine and terrestrial myxobacteria, reported by Jiang et al., supports this conspicuous delineation. While only recently observed, the MMC and the assumed capacity of the myxobacteria within to produce secondary metabolites exemplifies the need for continued efforts focused on the isolation and cultivation of marine myxobacteria.

2.4 *Haliangium ochraceum*

Originally isolated from a seaweed sample collected from a beach on the Miura Peninsula of Japan and reported as *Nannocystis* sp. strain SMP-2 in 1998 by Iizuka et al., *H. ochraceum*, along with what would become *Plesiocystis pacifica*, was one of the first halophilic marine myxobacteria, which were confirmed by 16s rRNA sequencing to be a member of the myxobacterial suborder *Nannocystineae*. The genome for type strain *H. ochraceum* SMP-2T (DSM 14365T) was published in 2010 (NC_013440.1). Originally cultivated at 30–34 °C on modified VY/2 agar media (Baker’s yeast 5 g L⁻¹, cyanocobalamin 0.5 mg L⁻¹, agar 15 g L⁻¹) supplemented with sea water, *H. ochraceum* was observed to grow at NaCl concentrations of 0.2–5% (w/v), with optimal growth between 2–3%. Fudou et al. reported the first discovery of a secondary metabolite from marine myxobacteria, haliangicin, a polyketide with antifungal activity isolated from *Haliangium luteum* in 2001. Fudou et al. later reclassified *Haliangium luteum* as *H. ochraceum*. *H. ochraceum* has since been reported to produce a variety of haliagcin congeners, as well as the hybrid polyketide-nonribosomal peptide haliamide (Figure 2.2), via a
type-I polyketide synthase pathway and a hybrid polyketide-nonribosomal biosynthetic pathway respectively.\textsuperscript{39,88–90} The features of \textit{H. ochraceum}, denoted as differentiated from soil myxobacteria, include obligate halophilicity, palmitic acid as a principle fatty acid, and presence of anteiso-branched fatty acids.\textsuperscript{85} Fruiting body formation has been reported from both solid and liquid cultures of \textit{H. ochraceum}, regardless of salinity.\textsuperscript{85} \textit{H. ochraceum} swarms form slightly sunken radial bands within agar, generating a tough slime film.\textsuperscript{85} Associated with the \textit{Haliangiaceae} clade of the suborder \textit{Nannocystinea}, \textit{H. ochraceum} shows a higher 16S rRNA sequence similarity to terrestrial myxobacteria than other halophilic marine myxobacteria.\textsuperscript{87,88} As previously mentioned, \textit{H. ochraceum} is capable of lysing both Gram-negative bacteria, specifically \textit{Escherichia coli} and \textit{Micrococcus luteus}, but also \textit{S. cerevisiae}.\textsuperscript{85}

![Diagram of Haliangium ochraceum secondary metabolites](image)

*Figure 2.2 Haliangium ochraceum secondary metabolites\textsuperscript{87–89,91}. Of note, \textit{cis} isomers about the epoxide moiety have also been reported for all the haliangicins.\textsuperscript{88}*
2.5 *Enhygromyxa salina*

Also discovered by Izuka et al., *E. salina* was initially isolated from a lagoon shore on the north coast of Hokkaido, Japan.\(^8\) Three strains of *E. salina* have been sequenced, including the type strain *E. salina* SMK-1\(^T\) (DSM 15217\(^T\)) and strains SWB005 and SWB007.\(^{29,92}\) Belonging to the *Plesiocystis/Enhygromyxa* clade of the suborder *Nannocystineae*, numerous unique strains of *E. salina* have been reported.\(^{11,29,83}\) *E. salina* produces fruiting bodies varying from white to orange when grown on VY/2 supplemented with salt water.\(^8\) An obligate halophile, *E. salina* tolerates NaCl concentrations of 0.1–4.0% (w/v) with an optimal range of 1.0–2.0% NaCl.\(^8\) However, the numerous strains of *E. salina* all demonstrate varying ranges of salt tolerance with minimum concentrations as high as 1% NaCl (w/v) and maximum concentrations of 7% NaCl.\(^{29,83}\) *E. salina* swarms form circular patterns, leaving deeply etched craters within agar surfaces.\(^93\) While able to survive on media with yeast as the sole nitrogen source, *E. salina* is only capable of lysing Gram-negative bacteria and is unable to lyse *S. cerevisiae* cells.\(^8\) Numerous secondary metabolites have been discovered from a variety of *E. salina* strains (Figure 2.3); the activities and biosynthetic assembly of these metabolites have been well reviewed elsewhere.\(^{25,28,94,95}\)

![Figure 2.3 Enhygromyxa salina secondary metabolites and hydroxyectoine](image-url)
Of note, comparative antiSMASH analysis of the three sequenced strains of *E. salina* suggests strain SWB005 to be the only sequenced strain of *E. salina* without a predicted trans-AT polyketide synthase, and strain SWB007 to be the only strain with an identified thiopeptide biosynthetic pathway. Interestingly, only *E. salina* seems to produce the osmolyte hydroxyectoine, as no other EctD homologue is apparent in the genomes of other sequenced marine myxobacteria (Table 2.1).

2.6 *Plesiocystis pacifica*

Discovered, yet again, by Iizuka et al., *P. pacifica*, originally identified as *Nannocystis* sp. SHI-1, was isolated from a beach on Iriomote-jima Island, Japan in 1997. There are currently two reported strains, the type strain *P. pacifica* SIR-1\(^T\) (DSM 14875\(^T\)) and SHI-1 (DSM 14876). Both strains produce pinkish-orange to brownish-orange fruiting bodies when grown on VY/2 that is supplemented with salt water, and require NaCl concentrations of 1% (w/v) for growth with optimum salinities of 2.0–3.0%. *P. pacifica* forms radial bands at the perimeter of its swarms, leaving cloudy etches in agar surfaces. Similar to the other members of the *Plesiocystis/Enhygromyxa* clade of the suborder *Nannocystineae*, *P. pacifica* lyases Gram-negative bacteria and is unable to lyse *S. cerevisiae*. While no secondary metabolites have been reported from either strain, an antiSMASH analysis on the draft genome of *P. pacifica* SIR-1\(^T\) (GCA_000170895.1) revealed numerous secondary metabolite biosynthetic pathways. This analysis portrays *P. pacifica* as an excellent candidate for future natural product discovery efforts. Instead, the haloalkane dehalogenase DppA from *P. pacifica* SIR-1\(^T\) has garnered interest as a potential biocatalyst for bioremediation of aromatic pollutants. DppA shows unique specificities towards brominated \(\alpha,\beta\)-haloalkanes, with no activity observed towards the
chlorinated substrates. Only briefly referenced in a PCR survey of polyketide synthase genes from various myxobacteria, genomic DNA from *Plesiocystis* sp. strain SIS-2 was found to contain several polyketide synthases. However, whether *Plesiocystis* sp. strain SIS-2 is a third strain of *P. pacifica* or a unique member of the genera *Plesiocystis* remains unclear.

2.7 *Haliangium tepidum*

The lesser investigated of the marine members of the *Haliangiaceae* clade of the suborder *Nannocystineae*, *Haliangium tepidum* SMP-10T (DSM 14436T) was first reported by Fudou et al. Found to be an obligate halophile, *H. tepidum* grows and produces fruiting bodies at salinities ranging from 0.5–6.0% NaCl (w/v). *H. tepidum* is able to lyse both Gram-negative bacteria and *S. cereviseae*, and swarms in radial patterns leaving slime sheets slightly etched into agar surfaces. As the species designation suggests, *H. tepidum* grows at moderately warm temperatures when compared to other marine myxobacteria with an optimal temperature range of 37–40 °C. While no genome sequence data or discovered natural products have been reported, a PCR survey of myxobacterial genomic DNA found an abundance of polyketide synthases within the genome of *H. tepidum*.

2.8 Potential Marine Myxobacteria

While the six previously discussed marine myxobacteria are the best characterized to date, there have been several recently reported potential marine myxobacteria. Tomura et al. discovered three new myxobacterial natural products that were produced by *Enhygromyxa niigataensis* or *Enhygromyxa* sp. SNB-1 (Figure 2.4). With a 97% similarity to the 16S rRNA sequence of *E. salina* SWB004, *E. niigataensis* SNB-1 was determined to be a new species within the genera
*Enhygromyxa*. Although the phylogenetic position of *E. niigataensis* would suggest that it is indeed a marine myxobacterium, the halotolerance levels and morphological features of the strain are currently unreported.\(^{101}\) Originally isolated from a marsh bank in Shikoku, Japan, *Pseudenygromyxa salsuginis* SYR-2\(^T\) (DSM 21377\(^T\)) develops fruiting bodies at NaCl concentrations of up to 2.5% (w/v), and forms slightly sunken, radial swarms.\(^{82}\) However, *P. salsuginis* is not an obligate halophile.\(^{82}\) Optimal growth conditions for *P. salsuginis* were determined to be somewhat saline between 0.2–1.0% NaCl.\(^{82}\) The species identifier for *P. salsuginis*, translated as “of brackish water”, aptly summarizes this observation. The obligate halophile *Paraliomyxa miuraensis* SMH-27-4 produces a variety of halogenated hybrid polyketide-nonribosomal peptide metabolites (Figure 2.4).\(^{102}\) Although its halophilic nature strongly suggests the strain to be a marine myxobacterium, the morphological features for *P. miuraensis* at saline cultivation conditions have not been reported.

Figure 2.4. Secondary metabolites from potential marine myxobacteria \(^{101–103}\)
2.9 Conclusions

Ubiquitous to marine environments worldwide, cultivable marine myxobacteria remain a relatively underexplored resource. Ideally, this roll call for known marine myxobacteria and corresponding descriptions will provide clarity to the early literature surrounding halophilic and halotolerant myxobacteria, as well as encourage the continued discovery of new marine myxobacteria. The recent expansion of the order Myxococcales, with the addition of myxobacteria associated with the MMC, suggests that the vast majority of marine myxobacteria have yet to be discovered. Although the limited number seems to suggest a scarcity, it should be recognized that the covered marine myxobacteria have been discovered thanks to the tremendous efforts of only a few research groups. This dearth of cultivable marine myxobacteria has not, however, limited the chemical diversity of their cognate reported natural products. To date, natural product classes discovered from marine and potential marine myxobacteria include polyketides, hybrid polyketide-nonribosomal peptides, degraded sterols, diterpenes, cyclic depsipeptides, and \( \gamma \)-alkylidenebutenolides. The capability to produce natural products with novel chemical scaffolds, such as salimabromide, will ensure the continued investigation of marine myxobacteria as a resource for the discovery of new therapeutics.
CHAPTER 3. IDENTIFICATION OF A SOLO ACYLHOMOSERINE LACTONE SYNTHASE FROM THE MYXOBACTERIUM ARCHANGIUM GEPHYRA

This chapter in full, is a reprint of the material as it appears in: Albataineh, H.; Duke, M.; Misra, S. K.; Sharp, J. S.; Stevens, D. C. Scientific Reports 2021, 11 (1). The dissertation author was the primary investigator of this paper, with the supervision of Prof. D. Cole Stevens. This chapter has been formatted to follow the University of Mississippi’s guidelines for dissertation preparation.

H.A. and M.D. conducted the extraction and heterologous expression experiments, S.K.M. conducted the MS experiments; H.A. and D.C.S. analyzed the data and wrote the manuscript; H.A., M.D., S.K.M., J.S.S., and D.C.S. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

DOI: 10.1038/s41598-021-82480-1
3.1 Abstract

Considered a key taxon in soil and marine microbial communities, myxobacteria exist as coordinated swarms that utilize a combination of lytic enzymes and specialized metabolites to facilitate predation of microbes. This capacity to produce specialized metabolites and the associated abundance of biosynthetic pathways contained within their genomes have motivated continued drug discovery efforts from myxobacteria. Of all myxobacterial biosynthetic gene clusters deposited in the antiSMASH database, only one putative acylhomoserine lactone (AHL) synthase, *agpI*, was observed, in genome data from *Archangium gephryra*. Without an AHL receptor also apparent in the genome of *A. gephryra*, we sought to determine if AgpI was an uncommon example of an orphaned AHL synthase. Herein we report the bioinformatic assessment of AgpI and discovery of a second AHL synthase from *Vitosangium* sp. During axenic cultivation conditions, no detectible AHL metabolites were observed in *A. gephryra* extracts. However, heterologous expression of each synthase in *Escherichia coli* provided detectible quantities of 3 AHL signals including 2 known AHLs, C8-AHL and C9-AHL. These results suggest that *A. gephryra* AHL production is dormant during axenic cultivation. The functional, orphaned AHL synthase, AgpI, is unique to *A. gephryra*, and its utility to the predatory myxobacterium remains unknown.

3.2 Introduction

Ubiquitous throughout soils and marine sediments, myxobacteria utilize cooperative features to facilitate uniquely social lifestyles and exhibit organized predation of microbial prey. Often attributed to their predatory capabilities, an extraordinary number of biologically active specialized metabolites have been discovered from myxobacteria. Interest in this
chemical space and the therapeutic potential associated with each elucidated natural product has motivated significant efforts towards continued discovery. Our recent survey of the unexplored, biosynthetic gene clusters from myxobacteria included in the antiSMASH database determined that the potential for such discovery from cultivable myxobacteria remains high.\textsuperscript{30,96} An oddity reported by this survey was the presence of a solo acylhomoserine lactone (AHL) synthase within the genome of the myxobacterium \textit{Archangium gephya}.\textsuperscript{30,106,107} As obligate cooperators numerous signaling systems have been associated with the coordination of myxobacterial motility and predation including A-signal, a quorum-like signal. However, no myxobacteria have been observed to produce AHL quorum signals.

Acylhomoserine lactone quorum signaling (QS) systems are abundant throughout Proteobacteria at-large.\textsuperscript{108} Considered autoinducers, AHLs bind to LuxR-type receptors which in turn induce expression of LuxI-type AHL synthases. While a recent assessment of LuxR receptors included within or nearby specialized metabolite biosynthetic gene clusters (BGCs) reported the presence of a putative LuxR receptor from the marine myxobacterium \textit{Haliangium ochraceum} DSM 14365, no AHL quorum signals or functional LuxI-type AHL synthases have been reported from myxobacteria.\textsuperscript{109} Additionally, the presence of putative LuxR receptors within numerous members of the genera \textit{Myxococcus} and \textit{Corallococcus} has been reported in a recent survey of myxobacterial signaling proteins.\textsuperscript{110} Intriguingly, the model myxobacterium \textit{Myxococcus xanthus} demonstrates enhanced predatory features when exposed to a variety of exogenous AHLs despite having no obvious LuxR receptor within its genome.\textsuperscript{111} This phenomenon, often referred to as “eavesdropping,” has become a generally accepted cornerstone in hypotheses surrounding interspecies cross talk within polymicrobial communities, and the presence of solo or orphan LuxR receptors from species that do not produce AHL signals supports such communication.\textsuperscript{108,111–114}
Putative solo-LuxR transcription factors with no accompanying LuxI synthases account for the majority of annotated LuxR proteins. However, as with *Myxococcus xanthus*, there are no LuxR receptors apparent in the genome of *A. gephyra*. This suggests that the observed LuxI-type synthase, AgpI, from *A. gephyra* is a solo-LuxI synthase. Considering the abundance of AHL QS systems throughout Proteobacteria other than myxobacteria, the uniqueness of this AHL synthase from *A. gephyra*, and the generalist diet of predatory myxobacteria that includes large swaths of AHL signaling proteobacteria, supports the assumption that *agpI* might have been acquired horizontally\(^7,15,115\). Conversely, the benefit AHL production might provide a predatory myxobacterium remains non-obvious. Herein we report bioinformatic analysis, functional assessment, and heterologous expression of the myxobacterial AHL synthase AgpI.

3.3 Materials and Methods

3.3.1 Cultivation of *A. gephyra*

*Archangium gephyra* (DSM 2261) was initially obtained from German Collection of Microorganisms in Braunschweig was grown at 30 °C on VY/2 agar (5 g/L baker’s yeast, 1.36 g/L CaCl\(_2\), 0.5 mg/L vitamin B12, 15 g/L agar, pH 7.2).

3.3.2 Bioinformatic Assessment of AgpI and VitI

The amino acid sequence for AgpI (WP_047862734.1) was submitted for blastp analysis against the non-redundant protein sequences database. The amino acid sequences for AgpI (WP_047862734.1) and VitI (WP_108069305.1) were submitted to EFI-EST analysis (https://efi.igb.illinois.edu/efi-est/) to construct a sequence similarity network (SSN) of LuxI-type AHL synthases (PF00765) using the default settings. Results from EFI-EST analysis were
visualized using Cytoscape (3.8.2). Alignments from ClustalW, minimum evolution, and maximum likelihood phylogenetic trees were rendered using MEGA 7.86

3.3.3 Autoinducer Binding Site Search

All 3014 domains annotated as autoinducer binding domains (PF03472) deposited in Pfam were subjected to blastp analysis against the *A. gephyra* genome (NZ_CP011509.1). For HMMSEARCH analysis, the raw HHM for autoinducer binding domains was downloaded from Pfam (PF03472) and utilized as input for profile-HMM vs protein sequence database via HMMSEARCH with the taxonomy restrictions set to limit analysis to *A. gephyra* or *Vitiosangium* sp. The amino acid sequence for AinR (AAW85531.1) and LuxN (BAF43687.1) was submitted for blastp analysis against the *A. gephyra* (NZ_CP011509) and *Vitiosangium* sp. (NZ_PZOX00000000.1) genomes.

3.3.4 Heterologous Expression of AgpI and VitI in *E. coli*

Constructs of AgpI and VitI codon optimized for expression in *E. coli* situated in pET28b were purchased from Genscript (Piscataway, NJ). Sequence data for these constructs are provided in the Appendix A. The heterologous host *E. coli* K207-3 was transformed with each plasmid individually by electroporation (BTX Gemini Sc2, Harvard apparatus) to provide an *E. coli* strain capable of expressing AgpI and an *E. coli* strain capable of expressing VitI. In addition, an *E. coli* negative control was generated by transforming *E. coli* K207-3 with an empty pET28b vector (no AHL synthase construct). The transformed *E. coli* strains were grown at 37 °C in LB broth supplemented with 50 μg/mL kanamycin, induced with IPTG (final concentration of 1.0 mM) at OD600 = 0.6, and grown overnight at 14 °C with shaking to facilitate heterologous protein
expression. *E. coli* K207-3 with no included plasmid was also included as a negative control and was grown without the addition of kanamycin but otherwise under the same conditions.

3.3.5 Metabolite Extraction and Analysis

After 21 days of cultivation, *A. gephyra* plates were manually diced and extracted with excess EtOAc. Pooled EtOAc was filtered and dried in vacuo to provide crude extracts for LC–MS/MS analysis. Extracts from controls and heterologous strains of *E. coli* were generated by Amberlite XAD-16 absorber resin (Alfa Aesar) facilitated extraction of clarified culture broths following cell lysis. Resins were removed by filtration and were eluted with MeOH to provide extracts for LC–MS/MS analysis. Extraction for all strains were performed in triplicate.

LC–MS/MS analysis of the extracted samples was performed on an Orbitrap Fusion instrument (Thermo Scientific, San Jose, CA) controlled with Xcalibur version 2.0.7 and coupled to a Dionex Ultimate 3000 nanoUHPLC system. Samples were loaded onto a PepMap 100 C18 column (0.3 mm × 150 mm, 2 μm, Thermo Fisher Scientific). Separation of the samples was performed using mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile) at a rate of 6 μL/min. The samples were eluted with a gradient consisting of 5 to 60% solvent B over 15 min, ramped to 95% B over 2 min, held for 3 min, and then returned to 5% B over 3 min and held for 8 min. All data were acquired in positive ion mode. Collision-induced dissociation (CID) was used to fragment molecules, with an isolation width of 3 m/z units. The spray voltage was set to 3600 V, and the temperature of the heated capillary was set to 300 °C. In CID mode, full MS scans were acquired from m/z 150 to 1200 followed by eight subsequent MS2 scans on the top eight most abundant peaks. The orbitrap resolution for both the MS1 and MS2 scans was 120,000. The expected mass accuracy was < 3 ppm. All extracts from heterologous
expression experiments for AgpI and VitI performed in triplicate were analyzed with AHL production confirmed in each as described.

3.3.6 Exogenous AHL Exposure Experiments

Stock solutions (10 mM) of N-hexanoyl-L-homoserine lactone-d₃ (C₆-AHL-d₃) and N-butyryl-L-homoserine lactone-d₅ (C₄-AHL-d₅) (Cayman Chemical) were prepared in DMSO. The required volumes of these stock solutions were filter sterilized and added to 14 days growing A. gephyra plates to give a final concentration of 30 μM. After 7 days of exogenous AHL exposure, A. gephyra plates were manually diced, extracted with excess EtOAc, and submitted to LC–MS/MS analysis as previously described.

3.3.7 GNPS Dataset

Generated data were converted to .mzXML files using MS-Convert and mass spectrometry molecular networks were generated using the GNPS platform (http://gnps.ucsd.edu)52. LC–MS/MS data for this analysis were also deposited in the MassIVE Public GNPS data set (MSV000084574).

3.4 Results

3.4.1 AgpI is Highly Homologous to Functional AHL Synthases

Located in the 20.6 kb BGC referenced as cluster 32 from A. gephyra (NZ_CP011509) in the antiSMASH database (5.1.1) the 210 aa gene product, AgpI (WP_047862734.1), is annotated as a putative autoinducer synthesis protein homologous to the GNAT family N-acetyltransferase, LuxI class of AHL synthases (Figure 3.1.A). None of the other annotated features neighboring
agpl are obviously associated with AHL quorum signaling systems. Assessment of Agpl (WP_047862734.1) with highly homologous LuxI synthases using blastp against the non-redundant protein sequences database provided two additional putative AHL synthases within the genome of another myxobacterium, Vitiosangium sp. GDMCC 1.1324. These included a GNAT family N-acetyltransferase deemed VitI (WP_108069305.1) with 68.12% identity when comparing amino acid sequence data with Agpl (Figure 3.1.B) and an annotated autoinducer synthase protein (WP_158502406.1) with 69.52% identity with Agpl amino acid sequence. The absence of genome data for Vitiosangium sp. in version 4.2.1 of the antiSMASH database explains the omission of this putative AHL synthase from our previous survey of myxobacterial biosynthetic space. The next highest scoring sequence from this analysis, a GNAT family N-acetyltransferase (WP_055459978.1) from Chelatococcus sambhunathii has 96% coverage and 56.44% identity with Agpl.117,118

Figure 3.1 (A) Cluster 32 from A. gephrya deposited in the antiSMASH database which includes the putative AHL synthase, agpl. (B) Genomic context for vitl from Vitiosangium sp. All annotated features within NCBI are labeled and all hypothetical features are in grey. Percentage GC content for each gene within the cluster provided for comparison and depicted in parentheses.
When restricting the blastp search to only provide results from myxobacteria (taxid: 29), only 1 other GNAT N-acetyltransferase (WP_169850287.1) from Corallococcus exiguus was found to have >45% identity with AgpI albeit at 50% coverage. Interestingly, an alignment of these four putative LuxI-type AHL synthases from myxobacteria revealed that only AgpI and VitI possessed all eight of the highly conserved residues associated with the LasI autoinducer domain (COG3916) (Figure 3.2 B). Alignment and phylogenetic analysis of AgpI and VitI against an assortment of 17 LuxI-type synthases experimentally validated to produce AHL QS molecules, suggests common ancestry with the AHL synthases LuxI, LasI, and TraI from Aliivibrio fischeri, Pseudomonas aeruginosa, and Rhizobium radiobacter, respectively (Figure 3.2 A, C).
Utilizing the genomic enzymology web tool EFI-EST developed by the Enzyme Function Initiative (EFI) to construct a sequence similarity network (SSN) that included 1001 LuxI-type AHL synthases (PF00765) as nodes clustered in different groups and 124,346 edges, both AgpI and VitI are included in the central cluster family that contains the vast majority of homologous LuxI-type AHL synthases (Figure 3.3)\textsuperscript{125}.

Figure 3.3 Sequence similarity network rendered by EFI-EST analysis of AgpI amino acid sequence data with AgpI (red diamond) and VitI (blue diamond) indicated. LuxI-type AHL synthases are represented as grey nodes (a total of 1,001 nodes clustered in 2 groups) with grey edged connecting the nodes (a total of 124,346 edges). To reduce complexity all nodes with \( \geq 90\% \) sequence similarity are represented as an individual aggregate node. For Cytoscape visualization, Attribute Circle Layout was applied.

Interestingly, although AgpI and VitI are highly homologous and both include all of the conserved residues associated with LuxI synthases, they share no neighboring similarities when comparing genomic context and surrounding features (Figure 3.1). Considering the typical GC-rich genomes of myxobacteria, we sought to compare the GC content of the genes surrounding AgpI and VitI to determine if either were obviously less GC-rich which would support horizontal acquisition. The genomes of both \textit{A. gephyra} and \textit{Vitiosangium} sp are 69.4\% and 68.3\% GC, respectively. The GC content of AgpI (57\%) and VitI (56\%) are indeed lower than the genome GC content of each myxobacterium as well as the averages of the genes surrounding them (Figure 3.1).
Ultimately these analyses did not provide a clear candidate LuxI-containing gene cluster from which AgpI was acquired; these results do not preclude the possibility that AgpI was indeed acquired as a genetic insertion with various additional features including an associated LuxR-type receptor that have since been lost due to genome reduction. Overall, we suggest that the shared ancestry observed from phylogenetic analysis of AgpI and VitI with known LuxI synthases and highly conserved active site residues suggest both AgpI and VitI are indeed LuxI synthases as originally predicted by antiSMASH.

3.4.2 Absence of a Cognate AHL Receptor in the Genome of *A. gephyra*

While no obvious AHL-binding LuxR homolog was identified in the chromosome of *A. gephyra*, we sought to determine the presence of any potential AHL-binding domain using the conserved sequence for autoinducer binding domains (PF03472). Utilizing the blastp suite at NCBI, we assessed all 3014 domains within the Pfam database classified as autoinducer binding domains for homology against the deposited genome of *A. gephyra*. No features within the proteome of *A. gephyra* were sufficiently homologous to be considered to include an autoinducer binding domain. We next queried the Hidden Markov Model (HMM) associated with autoinducer binding domains deposited in Pfam against the proteome of *A. gephyra* using HMMSEARCH. The most significant hit (E-value 0.0015) a PAS domain S-box-containing protein also annotated as a GAF-domain-containing protein (WP_053066299.1) does not include significant sequence homology with LuxR-type, AHL receptors.

Utilizing blastp, the genome of *A. gephyra* was also assessed for features homologous to the alternative AHL receptors AinR (AAW85531.1) and LuxN (BAF43687.1), which do not include the conserved autoinducer binding domain associated with LuxR. No homologues with
significant homology (>30% identity) for either alternative receptor was observed. The only resulting features were an annotated response regulator (WP_047859337.1; 43% coverage and 27% identity with AinR) and an annotated MASE1 domain-containing protein (WP_047860847.1; 60% coverage and 20.04% identity with LuxN). Although the annotated MASE1 domain-containing protein is predicted to include an N-terminal transmembrane region, considered to be the recognition site for AHL quorum signals in LuxN, the only homology was associated with the C-terminal response regulatory receiver domain (PF00072.24) of LuxN.

Similar analysis of *Vitiosangium* sp. GDMCC 1.1324 provided a highly homologous LuxR-type receptor (WP_108076247.1). While the AHL receptor identified in the genome of *Vitiosangium* sp. is not clustered near vitI as is typical of LuxI-LuxR type synthase-receptor pairs, we cannot assume both are unpaired orphans and instead consider VitI might not be a truly solo AHL synthase (Figure 3.4). Interestingly, the solo LuxR from *Escherichia coli* SdiA (PRK10188) was the highest scoring domain hit provided by blastp analysis of the LuxR from *Vitiosangium* sp. From these data we determined AgpI to be an orphaned AHL synthase without any cognate LuxR, AinR, or LuxN receptor present in the genome of *A. gephyra*, and despite the unclustered nature of the LuxR homologue identified in the genome of *Vitiosangium* sp., VitI and the annotated autoinducer synthase (WP_158502406.1) cannot be considered solo LuxI synthases without further investigation.

Figure 3.4 Genomic region of *Vitiosangium* sp. GDMCC 1.1324 including the putative LuxR-type receptor (WP_108076247.1) and surrounding genes. All annotated features within NCBI are labeled and all hypothetical features are in grey.
3.4.3 *A. gephyra* does not Produce AHLs During Axenic Cultivation

Cultivation of *A. gephyra* on VY/2 agar plates at 30 °C for 21 days provided fully developed, wispy myxobacterial swarms encompassing the entirety of the plate surface. Homogenized agar and cellular contents were extracted using traditional organic phase techniques to provide extracts for LC–MS/MS analysis. The resulting datasets from LC–MS/MS analysis of *A. gephyra* extracts were analyzed against datasets generated from analytical standards for a variety of AHLs including C9-AHL, C8-AHL, and C11-AHL to determine the presence of any produced AHL-like metabolites. Data from resulting mass spectra were scrutinized using the Global Natural Products Social Molecular Networking (GNPS) platform to generate molecular networks depicting similarities in detected metabolite scaffolds inferred from ionized fragment commonalities. No metabolites that included the diagnostic AHL fragments at 102.0547 m/z and 74.0599 m/z associated with the core homoserine lactone moiety were detected in extracts from *A. gephyra*. This data supports any one of the following conclusions: *A. gephyra* does not produce AHL-like metabolites when grown axenically but may be active under other growth conditions; metabolites produced by AgpI do not possess structural similarity with typical AHL metabolites; or AgpI is simply nonfunctional. Silent or dormant BGCs are commonly observed during natural product discovery efforts, and various strategies to activate silent BGCs have been developed including addition of exogenous chemical elicitors and heterologous expression of silent BGCs in an alternative host.

3.4.4 Exogenous AHLs do not Activate AgpI

Considering the typical autoinduction of LuxI synthases, we sought to determine if exogenous AHL metabolites might induce AgpI and provide observable AHL-like metabolites from *A. gephyra*. These experiments were conducted despite the absence of a feature that includes
the conserved autoinducer domain encoded in the genome of *A. gephyra* to determine if a regulatory element with a non-typical AHL binding domain might induce expression of AgpI. Experiments introducing the deuterated AHLs *N*-hexanoyl-**L**-homoserine lactone-d₃ (C₆-AHL-d₃) and *N*-butyryl-**L**-homoserine lactone-d₅ (C₄-AHL-d₅) to *A. gephyra* plates at 30 μM after two weeks of growth at 30 °C were conducted to determine if exogenous AHLs induce AgpI activity. Deuterium-labeled analogs of C₆-AHL and C₄-AHL were utilized to provide the ability to decouple exogenous signals from structurally similar AHLs potentially induced by exogenous AHL introduction. Using LC–MS/MS and molecular networking as previously described, no metabolites possessing the core homoserine lactone moiety were detected in the deuterated AHL-exposed extracts from *A. gephyra* suggesting that AgpI activity is not induced by exogenous AHLs. Ultimately, these experiments confirm that the exogenous AHL signals C₆-AHL and C₄-AHL do not induce biosynthesis of AHL metabolites from *A. gephyra*.

3.4.5 Heterologous Expression of AgpI Confirms Functional Production of AHLs

To explore the functionality of both AgpI and VitI and assumed biosynthesis of AHL-like metabolites, IPTG-inducible codon-optimized constructs of *agpI* and *vitI* included in replicating pET-28b(+) plasmids suitable for expression in Escherichia coli BL21 were purchased. Heterologous expression of AgpI and VitI, subsequent extraction, LC–MS/MS analysis, and evaluation of molecular networks rendered by GNPS as previously described, provided a cluster family including 2 of 3 total nodes identified as C₈-AHL (228.159 m/z) and C₉-AHL (242.174 m/z) from internal GNPS public datasets as well as a third AHL metabolite detected at 226.144 m/z (Figure 3.5). This cluster family was identical in both heterologous expression experiments suggesting that AgpI and VitI produce the same 3 AHL metabolites when heterologously expressed in *E. coli* with similar detected intensities for each AHL. Both C₈-AHL and C₉-AHL
were confirmed to be present in AgpI and VitI extracts using analytical standards. Based on associated intensities, C8-AHL was the most abundant and the metabolite detected at 226.144 m/z was the least abundant AHL.

Figure 3.5 Molecular family from the molecular network of LC–MS/MS datasets from extracts of heterologous E. coli expressing AgpI rendered by GNPS. Detected m/z values from raw data positioned over each node with node diameter depicting associated intensities for each AHL.

No AHL-like entities were detected in control extracts from E. coli containing no plasmid and E. coli containing an empty pET28b expression plasmid (Figures 3.6, Figure 3.7). From the mass difference between C8-AHL and the unknown AHL detected at 226.144 m/z (2.015 Da measured vs. 2.01565 theoretical), as well as shared fragmentation patterns, we suggest the metabolite detected at 226.144 m/z to likely be an unsaturated analog of C8-AHL. From these experiments we determined that both AgpI and VitI are functional AHL synthases capable of producing the previously characterized AHLs C8-AHL and C9-AHL. These results suggest A. geophyra could produce AHLs and likely requires environmental cues or specific nutrients not present during our axenic cultivation conditions.
Figure 3.6 Extracted-ion chromatographs 228-229 m/z (Top) and 242.1-242.2 m/z (Bottom) depicting presence of C₈-AHL and C₉-AHL, respectively, in extracts from *E. coli* K207-3 heterologously expressing AgpI (blue) and VitI (red) and absence of C8-AHL and C9-AHL in extracts from *E. coli* K207-3 + pET28b (green) and *E. coli* K207-3 with no plasmid (pink).

3.5 Discussion

Ultimately, we conclude that the myxobacteria *A. gephyra* and *Vitiosangium* sp. encode functional AHL syntheses that produce the AHL signals C₈-AHL and C₉-AHL when heterologously expressed in *E. coli*. Considering the strong precedent for heterologous expression of AHL syntheses in *E. coli* to determine produced AHL metabolites, this suggests that both *A. gephyra* and *Vitiosangium* sp. can produce one or all of the observed AHL signals and that AgpI is merely silent or cryptic during axenic cultivation of *A. gephyra*. However, provided the
subtle differences in LuxI synthase homologies and resulting chemical diversity of produced 
AHLs\textsuperscript{124}, we should also consider that these synthases could instead utilize an acyl-ACP or CoA 
thioester precursor not available to the heterologous \textit{E. coli} host, and we are actively exploring 
cultivation conditions that might induce native AHL production from \textit{A. gephyra}\.\textsuperscript{135,139}

![Diagram](image)

Figure 3.7 MS/MS fragmentation spectra with diagnostic fragments indicated for each AHL detected in 
extracts from heterologous \textit{E. coli} expressing AgpI.

While numerous bacteria have been observed to possess orphaned LuxR-type AHL 
receptors, a functional solo LuxI synthase without any cognate LuxR receptor also present in the 
genome has yet to be reported.\textsuperscript{113,140,141} Although a functional orphaned LuxI-type synthase
capable of producing AHLs has been reported from the sponge symbiont *Ruegeria* sp. KLH11, the strain also harbors two pairs of clustered LuxI/LuxR homologues.\textsuperscript{142} We suggest that production of AHL quorum signals by myxobacteria would support the theoretical benefits of interspecies cross talk similar to functional, solo LuxR receptors.\textsuperscript{142–144} We also propose that the more typical abundance of orphan LuxR receptors compared to the seemingly exceptional solo LuxI synthase reported here might correlate with the rarity of cooperative generalist predators.

The absence of any AHL metabolites during axenic cultivation of *A. gephyra* suggests an unknown regulatory mechanism independent from a LuxR receptor might be involved. However, previously reported eavesdropping by *Myxococcus xanthus* and response to exogenous AHLs despite the absence of any AHL receptor with homology to LuxR suggests myxobacteria may possess an undiscovered, alternative means of AHL detection and response.\textsuperscript{111} An alternative explanation for the absence of AHL metabolites in the extracts of *A. gephyra* would be that AgpI is merely a non-functional feature that has been acquired but is not utilized. Although this seems unlikely due to the presence of additional LuxI-type synthases in the genome of *Vitiosangium* sp. and the understanding that specialized metabolite biosynthetic genes are often silent or unexpressed during axenic cultivation,\textsuperscript{30,135,136} this explanation should be considered until either myxobacterium is observed to produce AHLs metabolites. While the benefit afforded predatory myxobacteria remains unclear, production of AHL signals known to regulate QS-associated physiological functions such as biofilm formation, specialized metabolism, and motility offers some insight.\textsuperscript{108} Predatory disruption of any one of these functions would likely benefit the fitness of *A. gephyra* by improving predation of quorum signaling prey. For example, *Pseudomonas putida* biofilm formation is negatively regulated by the presence of AHLs,\textsuperscript{119,145} and biofilm formation is commonly associated with predator avoidance.\textsuperscript{146,147} Myxobacterial production of
AHLs would therefore inhibit biofilm formation of *P. putida* which would benefit predation. However, without direct evidence of myxobacterial AHL production any interplay between AHL biosynthesis and predator–prey interactions remain hypothetical. Overall, we consider the presence of two functional AHL synthases within the genomes of two predatory myxobacteria provides a unique perspective and supports the continued investigation of small molecule interactions that contribute to microbial community structures.
CHAPTER 4. ASSESSMENT OF EVOLUTIONARY RELATIONSHIPS FOR PRIORITIZATION OF MYXOBACTERIA FOR NATURAL PRODUCT DISCOVERY

This chapter in full, is a reprint of the material as it appears in: Ahearne, A.; Albataineh, H.; Dowd, S. E.; Stevens, D. C. Assessment of Evolutionary Relationships for Prioritization of Myxobacteria for Natural Product Discovery. *Microorganisms* **2021**, *9* (7), 1376.

The dissertation author was the primary co-investigator of this paper, with the supervision of Prof. D. Cole Stevens. This chapter has been formatted to follow the University of Mississippi’s guidelines for dissertation preparation.

A.A conducted the isolation of environmental Myxobacteria, A.A and H.A conducted the genomic DNA isolation and comparative genomic studies, H.A. conducted the BIG-SCAPE analysis, S.E.D. conducted the genomic sequencing and assembly; A.A, H.A., and D.C.S. analyzed the data and wrote the manuscript; A.A, H.A., S.E.D., and D.C.S. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

DOI: 10.3390/microorganisms9071376
4.1 Abstract

Discoveries of novel myxobacteria have started to unveil the potentially vast phylogenetic diversity within the family Myxococcaceae and have brought about an updated approach to myxobacterial classification. While traditional approaches focused on morphology, 16S gene sequences, and biochemistry, modern methods including comparative genomics have provided a more thorough assessment of myxobacterial taxonomy. Herein, we utilize long-read genome sequencing for two myxobacteria previously classified as *Archangium primigenium* and *Chondrococcus macrosporus*, as well as four environmental myxobacteria newly isolated for this study. Average nucleotide identity and digital DNA–DNA hybridization scores from comparative genomics suggest previously classified as *Archangium primigenium* to instead be a novel member of the genus *Melittangium*, *Corallococcus macrosporus* to be a potentially novel member of the genus *Corallococcus* with high similarity to *Corallococcus exercitus*, and the four isolated myxobacteria to include another novel *Corallococcus* species, a novel *Pyxidicoccus* species, a strain of *Corallococcus exiguis*, and a potentially novel *Myxococcus* species with high similarity to *Myxococcus stipitatus*. We assess the biosynthetic potential of each sequenced myxobacterium and suggest that genus-level conservation of biosynthetic pathways support our preliminary taxonomic assignment. Altogether, we suggest that long-read genome sequencing benefits the classification of myxobacteria and improves determination of biosynthetic potential for prioritization of natural product discovery.

4.2 Introduction

Over the last decade, 34 novel species of myxobacteria have been described including representatives from 10 newly described genera within the order *Myxococcales*.54,82,116,148–158
Prevalent in soils and marine sediments, predatory and cellulolytic myxobacteria contribute to nutrient cycling within microbial food webs. Perhaps most-studied for their cooperative lifestyles, myxobacteria have been an excellent resource for investigations concerning developmental multicellularity and two-component signaling, swarming motilites and predatory features, and the discovery of biologically active metabolites.\textsuperscript{1,8,16,26,27} Each of these areas of interest have benefited from the increased utility and accessibility of next-generation sequencing (NGS) technologies. The driving force behind the recent surge in efforts to discover novel species of myxobacteria can also be attributed to advances in sequencing technologies. Genome sequencing of myxobacteria has demonstrated that they possess large genomes replete with biosynthetic gene clusters, and myxobacteria have recently been deemed a “gifted” taxon to produce specialized metabolites with drug-like properties.\textsuperscript{25–27,38,40} These efforts, combined with a thorough metabolic survey of over 2000 strains within the order \textit{Myxococcales}, concluded that the odds of novel metabolite discovery increase when exploring novel genera of myxobacteria.\textsuperscript{33} Motivated by these observations, we sought to isolate novel myxobacteria from lesser studied North American soils.

Recently, comparative genomic analyses have been utilized to provide efficient preliminary classification of novel myxobacteria, and we considered that such an approach would expedite prioritization of strains for future metabolic studies.\textsuperscript{11,54,65,150,158,159} While traditional myxobacterial classification efforts relied on morphology, biochemistry, and the conservation of 16S gene sequences, updated methods including genome-based taxonomy have provided excellent preliminary taxonomic classification of myxobacterial isolates.\textsuperscript{55} Considering that genome sequencing would also afford the biosynthetic potential of any isolated myxobacteria, we sought to employ long-read sequencing to generate high-quality draft genomes hoping to avoid fragmented, partial biosynthetic pathways. For example, of the 11 currently sequenced
myxobacteria from the genus *Corallococcus*, 68% of the 621 total putative biosynthetic gene clusters (BGCs) predicted by the analysis platform AntiSMASH are positioned on a contig edge and are potentially incomplete (Table 4.1). In fact, the only two *Corallococcus* genomes sequenced with long-read techniques (*Corallococcus coralloides* DSM 2259<sup>T</sup> and *Corallococcus coralloides* strain B035) each included 34 predicted BGCs with none located on a contig edge.160,161 Ideally, larger contigs generated from long-read sequencing might benefit the comparative genomic analyses and provide a more complete assessment of biosynthetic potential.

Table 4.1. Comparison of biosynthetic gene clusters (BGCs) located on contig edges from previously sequenced members of the genus *Corallococcus*. Included BGC data provided by whole genome analysis using antiSMASH version 5.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Accession #</th>
<th>Total BGCs</th>
<th>BGCs on contig edge</th>
<th>Average % of BGCs on contig edge</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. coralloides</em> DSM 2259&lt;sup&gt;T&lt;/sup&gt;</td>
<td>NC_017030.1</td>
<td>34</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td><em>C. coralloides</em> B035</td>
<td>NZ_CP034669.1</td>
<td>34</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td><em>C. exiguus</em> DSM 14696&lt;sup&gt;T&lt;/sup&gt;</td>
<td>JAAAPK010000001.1</td>
<td>44</td>
<td>13</td>
<td>29.55</td>
</tr>
<tr>
<td><em>C. interemptor</em> AB047A&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAWM01000001.1</td>
<td>45</td>
<td>34</td>
<td>75.56</td>
</tr>
<tr>
<td><em>C. praedator</em> CA031B&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAWI01000001.1</td>
<td>70</td>
<td>66</td>
<td>94.29</td>
</tr>
<tr>
<td><em>C. aberystwythensis</em> AB050A&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAWK01000001.1</td>
<td>46</td>
<td>41</td>
<td>89.13</td>
</tr>
<tr>
<td><em>C. sicarius</em> CA040B&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAWG01000001.1</td>
<td>68</td>
<td>61</td>
<td>89.71</td>
</tr>
<tr>
<td><em>C. carmarthensis</em> CA043D&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAWE01000001.1</td>
<td>55</td>
<td>46</td>
<td>83.64</td>
</tr>
<tr>
<td><em>C. exercitus</em> AB043B</td>
<td>JABFJV010000005.1</td>
<td>59</td>
<td>52</td>
<td>88.14</td>
</tr>
<tr>
<td><em>C. terminator</em> CA054A&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAVZ01000001.1</td>
<td>71</td>
<td>68</td>
<td>95.77</td>
</tr>
<tr>
<td><em>C. llansteffanensis</em> CA051B&lt;sup&gt;T&lt;/sup&gt;</td>
<td>RAWB01000001.1</td>
<td>95</td>
<td>95</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>621</td>
<td>476</td>
<td>67.80%</td>
</tr>
</tbody>
</table>

In addition to four environmental isolates of putative myxobacteria included in this study, we acquired two previously characterized myxobacteria from the American Type Culture Collection (ATCC): *Archangium primigenium* ATCC 29,037 and *Chondrococcus macrosorus* ATCC 29039. Previously miscategorized as *Polyangium primigenium*, the original morphological descriptions for *Archangium primigenium* were remarkably apt for the strain acquired from the ATCC and cultivated in our lab, including obvious fruiting body formation and carotene-like pigmentation (Figure 4.1).162–164 The original description of *Archangium primigenium* fruiting
bodies initially piqued our interest in the strain as members of the genus *Archangium* typically do not or very rarely form defined fruiting bodies when cultivated with standard laboratory conditions.\textsuperscript{66,165} *Archangium* species have previously been referred to as “degenerate forms” of myxobacteria due to diminished fruiting bodies with no sporangioles or absent fruiting body formation.\textsuperscript{165}

Comparatively, little historical data is available for *Corallococcus macrosporus* ATCC 29039. The strain was deposited at the ATCC by distinguished taxonomist Professor V. B. D. Skerman and was subsequently included in a methodology study focused on isolating myxobacteria from soils.\textsuperscript{166–168} The decision to change the genus *Chondrococcus* to instead be *Corallococcus* has been validated with many novel *Corallococcus* species being described afterwards.\textsuperscript{150,169,170} However, we were curious to determine the status of *Corallococcus macrosporus* ATCC 29039. Considering the proposed reassignment of *Corallococcus macrosporus* DSM 14697\textsuperscript{T} to the genus *Myxococcus*, it was unclear if *Corallococcus macrosporus* ATCC 29,039 should also be reassigned. Both characterized using traditional approaches that heavily relied on morphology, we sought to determine how genomic comparisons might impact the current taxonomic assignments of these available myxobacteria.

Figure 4.1. Myxobacterial fruiting bodies from strains NCSPR001, NCCRE002, SCHIC003, SCPEA002, and the strains *C. macrosporus* ATCC 29039 and *A. primigenium* ATCC 29037.
4.3 Materials and Methods

4.3.1 Bacterial Strains and Growth Conditions

*Archangium primigenium* and *Corallococcus macrosorus* were procured from the ATCC as strain numbers ATCC 29037 and ATCC 29039, respectively. The remaining strains were isolated from soil as described later. All strains were cultured either on VY/2 or VY/4 agar plates (5 or 2.5 g/L baker’s yeast, 1.5 g/L CaCl$_2$·2H$_2$O, 0.5 mg/L vitamin B12, 15 g/L agar, pH 7.2). Swarming and fruiting bodies on agar plates were observed under a Zeiss discovery V12 stereo microscope and photographed using a Zeiss axicam105.

4.3.2 Isolation of Environmental Myxobacteria

Soil samples, collected in Asheville, NC and Tryon, SC, were taken from the base of trees and dried in open air before storage. Detailed location data are provided in the Appendix B. Myxobacteria were isolated using a slightly modified Coli-spot method.\textsuperscript{171} A 1 mg/mL solution of cycloheximide/nystatin was used to wet the soil sample to a paste-like consistency before inoculation onto an *Escherichia coli* baited WAT agar plate (1 g/L CaCl$_2$·2H$_2$O, 15 g/L agar, 20 mM HEPES).

To prepare the baiting plate, a lawn of *E. coli* was grown overnight on tryptone soya broth (TSB) with agar (1.5%), and the cells were scraped and suspended in 2 mL of sterile deionized water. Four hundred μL of the *E. coli* suspension was spread over the surface of a WAT agar plate to create a bait circle of approximately two inches in diameter and let dry. Once the *E. coli* was dried, a pea sized amount of soil paste was placed at the center of the bait circle. Plates were incubated at 25 °C for up to a month, and degradation of the *E. coli* was monitored over time. Visible degrading swarms were seen after a few days, and swarm edges or fruiting bodies were...
passaged onto VY/4 media for purification. Purification was accomplished by repeated swarm edge transfer.

4.3.3 Genomic DNA Isolation, Sequencing, Assembly, and Annotation

Genomic DNA for NGS was obtained from actively growing bacteria on VY/2 or VY/4 plates using NucleoBond high molecular weight DNA kit (Macherey-Nagel, Bethlehem, PA, USA). The quantity and quality of the extraction were checked by Nanodrop (Thermo Scientific NanoDrop One) and followed by Qubit quantification using Qubit®® dsDNA HS Assay Kit (ThermoFisher Scientific, Suwanee, GA, USA).

Sequencing for all samples was performed on a Pacific Biosciences single-molecule real-time (SMRT) sequencing platform at the MR DNA facility (Shallowater, TX, USA). The SMRTbell libraries for the sample were prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) following the manufacturer’s user guide. Following library preparation, the final concentration of each library was measured using the Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific, Suwanee, GA, USA), and the average library sizes were determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Each library pool was then sequenced using the 10-h movie time on the PacBio Sequel (Pacific Biosciences, Menlo Park, CA, USA). De Novo Assembly of each genome was accomplished using the PacBio SMRT Analysis Hierarchical Genome Assembly Process (HGAP). Sequencing data have been deposited in NCBI under the accession numbers JADWY100000000.1, JAFIMU00000000, JAFIMS000000000, JAFIMT000000000, CP071090, and CP071091 for strains Archangium primigenium, Corallococcus macrosorus, NCSPR001, NCCRE002, SCPEA002, and SCHIC003, respectively.
4.3.4 Comparative Genomic Studies

The genome sequence data were uploaded to the Type (Strain) Genome Server (TYGS), a free bioinformatics platform available under https://tygs.dsmz.de (accessed 10 January 2021), for a whole genome-based taxonomic analysis. TYGS was used to calculate the dDDH values and construct minimum evolution trees using the Genome BLAST Distance Phylogeny approach (GBDP).\textsuperscript{172,173} GBDP trees were visualized using MEGA-X.\textsuperscript{86} The average nucleotide identity (ANI) was calculated using the ANI/AAI-Matrix calculator.\textsuperscript{174}

4.3.5 BIG-SCAPE Analysis

Genome data for all myxobacteria belonging to the \textit{Cystobacterineae} suborder were downloaded from the NCBI database. A list of all myxobacteria used in this analysis are listed in Appendix C. These genomes in addition to genomes of \textit{Archangium primigenium}, \textit{Corallococcus macrosporus}, and the environmental isolates were analyzed by the AntiSMASH platform (version 5 available at https://docs.antismash.secondarymetabolites.org; accessed 1 February 2021) to assess specialized metabolite gene clusters using the “relaxed” strictness setting.\textsuperscript{44} A total of 1826 predicted BGCs (.gbk files) were then processed locally using the BiG-SCAPE program (version 20181005, available at https://git.wageningenur.nl/medema-group/BiG-SCAPE; accessed 1 February 2021), with the MiBIG database (version 2.0 available at https://mibig.secondarymetabolites.org; accessed 1 February 2021) as reference.\textsuperscript{46,175} BiG-SCAPE analysis was supplemented with Pfam database version 33.1.\textsuperscript{176} The singleton parameter in BiG-SCAPE was selected to ensure that BGCs with distances lower than the default cutoff distance of 0.3 were included in the corresponding output data. The hybrids-off parameter was selected to prevent hybrid BGC redundancy.
Generated network files separated by BiG-SCAPE class were combined for visualization using Cytoscape version 3.8.2 (http://www.cytoscape.org; accessed 1 February 2021). Annotations associated with each BGC were included in Cytoscape networks by importing curated tables generated by BiG-SCAPE.

4.4 Results
4.4.1 Comparative Genomics and Taxonomic Assessment of *Archangium Primigenium*, *Chondrococcus Macrosorus*, and Environmental Isolates

Genome sequencing provided high quality draft genomes for each of the six investigated myxobacteria, as indicated by the summary of general features in Table 4.2. The total genome sizes ranged from ~9.5–13 Mb, and the %GC content varied around ~69–71%. Of the six genomes, both environmental strains SCHIC003 and SCPEA002 were assembled on a single contig. Overall, the assemblies for each genome provided much lower total contig counts (1–17) than recently sequenced myxobacterial genomes. Interestingly, a minimum evolution of phylogenetic trees generated from the whole genome sequence data clustered *Archangium primigenium* with *Melittangium boletus* DSM 14713\textsuperscript{T} and not with the three currently sequenced strains from the genus *Archangium* (Figure 4.2, Figure 4.3). Accordingly, ANI and dDDH values supported the placement of *Archangium primigenium* in the genus *Melittangium* (Figure 4.3) as a novel species with both values well below the established cutoffs for classification of distinct species (< 95% ANI; < 70% dDDH). These data suggest *Archangium primigenium* is currently misclassified as a member of the genus *Archangium* and should instead be placed in the genus *Melittangium*. 
Figure 4.2. Minimum evolution tree from the whole genomes of different myxobacteria including the six strains under investigation in this study using the GBDP approach. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 100.0%. Branch pseudo-bootstraps less than 50% are not shown. The numbers below branches are branch lengths scaled in terms of GBDP distance formula d5. The tree was rooted at the midpoint.
Table 4.2 Genome properties and general features of myxobacteria under investigation in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (bp)</th>
<th>CDS</th>
<th>GC%</th>
<th>N50</th>
<th>L50</th>
<th>Contigs</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. primigenium</td>
<td>9,491,554</td>
<td>7,873</td>
<td>70.7%</td>
<td>9,468,833</td>
<td>1</td>
<td>3</td>
<td>441x</td>
</tr>
<tr>
<td>C. macrosporus</td>
<td>9,811,739</td>
<td>7,977</td>
<td>70.4%</td>
<td>1,094,727</td>
<td>2</td>
<td>17</td>
<td>300x</td>
</tr>
<tr>
<td>NCSPR001</td>
<td>9,785,177</td>
<td>8,033</td>
<td>70.1%</td>
<td>9,343,940</td>
<td>1</td>
<td>3</td>
<td>312x</td>
</tr>
<tr>
<td>NCCRE002</td>
<td>10,538,407</td>
<td>8,589</td>
<td>69.7%</td>
<td>3,024,381</td>
<td>2</td>
<td>8</td>
<td>479x</td>
</tr>
<tr>
<td>SCPEA002</td>
<td>13,211,253</td>
<td>10,588</td>
<td>69.6%</td>
<td>N/A</td>
<td>1</td>
<td>1</td>
<td>144x</td>
</tr>
<tr>
<td>SCHIC003</td>
<td>10,367,529</td>
<td>8,339</td>
<td>68.6%</td>
<td>N/A</td>
<td>1</td>
<td>1</td>
<td>301x</td>
</tr>
</tbody>
</table>

Figure 4.3 Minimum evolution tree from the whole genomes of A. primigenium and different members of the family Archangiaceae using the GBDP approach. To the right of the tree a heat map shows dDDH and ANI values from pairwise comparisons of A. primigenium with different myxobacteria. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 100.0 %. The numbers below branches are branch lengths scaled in terms of GBDP distance formula d5. branch lengths less than 0.05 are not shown. The tree was rooted at the midpoint. Myxococcus xanthus DK 1622 is included in the GBDP tree as an outgroup.

The calculated ANI and dDDH values for the sequenced Corallococcus marcosporus strain acquired from the ATCC support the original assignment to the genus Chondrococcus, now Corallococcus.68,170 As opposed to the recently reclassified Myxococcus macrosporus DSM 14697T, previously Corallococcus macrosporus, the minimum evolution phylogenetic tree suggested Corallococcus macrosporus ATCC 29039 to be a member or the genus Corallococcus.
most similar to Corallococcus exercitus DSM 108849T (Figure 4.2, Appendix D).\textsuperscript{170} The isolated strains NCCRE002 and NCSPR001 were also determined to be members of the genus Corallococcus (Figure 4.2, Appendix D). Comparative genome analyses implied that strain NCCRE002 is an isolate of Corallococcus exigus DSM 14696\textsuperscript{T}. However, the ANI and GBDP trees suggested that strain NCSPR001 is a novel member of the genus Corallococcus most similar to Corallococcus coralloides DSM 2259\textsuperscript{T} (Figure 4.4 Top, Appendix D).

The isolated SCHIC003 and SCPEA002 strains were initially determined to be members of the genus Myxococcus. However, inclusion of sequenced representatives from the genus Pyxidicoccus (considered to be synonymous with Myxococcus) in our comparative analysis grouped strain SCPEA002 within the Pyxidicoccus clade (Figure 4.2, Appendix E).\textsuperscript{158} Most similar to Pyxidicoccus caerfyrrddinensis CA032AT, dDDH and ANI analysis suggested the SCPEA002 strain to be a novel member of the genus Pyxidicoccus (Figure 4.4). Similarly, comparative genome analysis determined that strain SCHIC003 is likely be a novel member of the genus Myxococcus, albeit highly similar to Myxococcus stipitatus DSM 14675\textsuperscript{T} with ANI and dDDH values just below the cutoffs for species differentiation (Figure 4.2, Figure 4.4).\textsuperscript{58,70,159}

4.4.2 Biosynthetic Potential and Genus Level Correlations

Analysis of our draft genomes using the biosynthetic pathway prediction platform AntiSMASH revealed a range of 29–42 total predicted BGCs with Corallococcus macrosporus including the highest total of BGCs. However, the draft genome for Corallococcus macrosporus also included the highest total of four partial BGCs positioned on the edges of contigs. No BGCs occurring on contig edges were observed from Archangium primigenium, NCSPR001, or SCPEA002. All of the sequenced strains included highly similar (≥75% similarity score)
biosynthetic pathways for the signaling terpene geosmin,\textsuperscript{178} the signaling lipids VEPE/AEPE/TG-1,\textsuperscript{179,180} and carotenoids (Figure 4.5).\textsuperscript{181–184}

Excluding SCHIC003, each genome included a BGC highly homologous to the pathway associated with the myxobacterial siderophore myxochelin.\textsuperscript{185,186} Pathways somewhat similar
(similarity scores of 66%) to the myxoprincomide-c506 BGC were observed in every genome except the *Archangium primigenium* genome.\textsuperscript{187} Clusters with \textgeq75\% similarity to pathways from *Myxococcus stipitatus* DSM 14675\textsuperscript{T} associated with the metabolites rhizopodin\textsuperscript{188,189} and phenalamide A2\textsuperscript{190} were observed in the SCHIC003 draft genome as well as clusters also present in the *Myxococcus stipitatus* DSM 14675\textsuperscript{T} genome deposited in the AntiSMASH database, including the dkxanthene,\textsuperscript{191} fulvuthiacene,\textsuperscript{192} and violacein BGCs (Figure 4.6).\textsuperscript{193–195}

![Common specialized metabolites from myxobacteria associated with characterized BGCs present in the six investigated strains of myxobacteria. * Myxochelin BGC not present in SCHIC003 genome data. # Myxoprincomide BGC not present in *A. primigenium* genome data.](image-url)
Considering previously characterized BGCs from each genus associated with the six investigated myxobacteria, the corallopyronin BGC from *Corallococcus coralloides* B035 was absent from all three of the putative *Corallococcus* strains, the melithiazol BGC from *Melittangium lichenicola* Me I46 was not present in *Archangium primigenium*, and neither the disciformycin/gulmirecin BGC or the pyxidicycline BGC from *Pyxidicoccus fallax* were present in SCPEA002.

Figure 4.6 (A) Specialized metabolites produced by members of the genera *Corallococcus*, *Melittangium*, and *Pyxidicoccus* with no associated BGCs observed in any of the six investigated myxobacterial strains. (B) Comparisons of the rhizopodin, phenalamide A2, and fulvuthiacene BGCs from SCHIC003 genome data and the characterized pathways from *M. stipitatus* and *M. fulvus*. All SCHIC003 gene products, excluding RizC, had coverages ≥99% with the indicated homolog. * RizC is located on a contig edge and is incomplete in SCHIC003 genome data.
Utilizing the BiG-SCAPE platform to render BGC sequence similarity networks, we sought to determine the extent of homology between BGCs from our six sequenced myxobacteria and BGCs from all currently sequenced members of the suborder *Cystobacterineae*. The resulting sequence similarity network included 1080 BGCs connected by 3046 edges (not including self-looped nodes/singletons) and depicted genus-level homologies across all BGCs from the newly sequenced myxobacteria corroborating our suggested taxonomic assignments (Figure 4.7 and Table 4.3). For example, BGCs from the three newly sequenced samples *Corallococcus macrosorus*, NCSPR001, and NCCRE002 were almost exclusively clustered with BGCs from members of the genus *Corallococcus*, and BGCs from SCHIC003 and SCPEA002 samples clustered with the genera *Myxococcus* and *Pyxicidoccus* (Figure 4.7). However, SCPEA002 BGCs do not cluster as frequently with *Pyxicidoccus* BGCs as they do *Myxococcus* BGCs, and the majority (76.5%) were not clustered with any BGC within the network (Table 4.3). This is likely due to the highly fragmented nature of available *Pyxicidoccus* genomes resulting in many incomplete or partial BGCs. Therefore, few *Pyxicidoccus* pathways appear in the similarity network, and the percentage of unique pathways associated with SCPEA002 is likely overestimated. Regardless, the limited number of SCPEA002 BGCs clustered with BGCs from *Myxococcus/Pyxicidoccus* genomes indicates a potential to discover novel metabolites despite placement in the highly scrutinized clade. The only clustered groups with numerous edges formed between BGCs from the genera *Myxococcus* and *Corallococcus* included characterized biosynthetic pathways for ubiquitous signaling lipids VEPE/AEPE/TG-1, carotenoids, and the siderophore myxochelin as well as two uncharacterized BGCs predicted to produce ribosomally synthesized and post-translationally modified peptides (RiPPs).
Figure 4.7 BiG-SCAPE BGC sequence similarity networks (c = 0.3) as visualized with Cytoscape. The network is generated from *A. primigenium*, *C. macrosporus*, NCCRE002, NCSPR001, SCHIC003, SCPEA002, and all myxobacteria belonging to the Cystobacterineae suborder with genomes deposited in NCBI. Each node represents one BGC identified by AntiSMASH 5.0, where the colors and shapes of the nodes represent different genera and AntiSMASH-predicted classes, respectively. Nodes representing BGCs from newly sequenced myxobacteria included in this study are enlarged. BGCs included as singletons in the original BiG-SCAPE analysis were removed.
Interestingly, a total of twenty-three *Archangium primigenium* BGCs (out of thirty-two BGCs) appear as singletons in the network with no homology to any of the included BGCs from *Cystobacterineae*. In fact, aside from the VEPE/AEPE/TG-1 cluster and a terpene cluster that included members of the genera *Archangium* and *Cystobacter*, all remaining BGCs from *Archangium primigenium* had connecting edges to BGCs from *Melittangium boletus* DSM 14713T. Out of twenty-one edges formed by *Archangium primigenium* in the network, four edges were formed with four species of *Corallococcus* (a total of eleven *Corallococcus* species in the network), four edges were formed with all species of *Cystobacter* (three species in the network), six edges were formed with all species of *Archangium* (three species in the network), and seven edges were formed with the only *Melittangium* species in the network, *M. boletus* DSM 14713T. These data corroborate our preliminary taxonomic assignments and suggest that the prioritization of *A. primigenium* for subsequent discovery efforts is most likely to yield novel metabolites.

Table 4.3 Overview of BiG-SCAPE BGC sequence similarity networks of the six strains under investigation in this study.

<table>
<thead>
<tr>
<th>Myxobacteria</th>
<th># of Total BGCs</th>
<th># and % of Singletons</th>
<th># of Edges Formed with other BGCs</th>
<th># of BGCs with 1 or 2 Edges</th>
<th># of BGCs with 3 or More Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. primigenium</em> ATCC 29037</td>
<td>32</td>
<td>24 (75%)</td>
<td>21</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td><em>C. macrosporus</em> ATCC 29039 *</td>
<td>42</td>
<td>9 (21.4%)</td>
<td>228</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>NCSPRO001</td>
<td>32</td>
<td>1 (3.1%)</td>
<td>248</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>NCCRE002 *</td>
<td>36</td>
<td>3 (16.7%)</td>
<td>231</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>SCPEA002</td>
<td>34</td>
<td>26 (76.5%)</td>
<td>62</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>SCHIC003</td>
<td>29</td>
<td>8 (27.6%)</td>
<td>85</td>
<td>13</td>
<td>8</td>
</tr>
</tbody>
</table>

4.5 Discussion:

As novel myxobacteria continue to be isolated and explored for natural product discovery, efficient approaches for approximate taxonomic placement will assist the prioritization of lesser studied genera. Utilizing long-read genome sequencing and comparative genomic analyses, we
determine preliminary taxonomic placement for four myxobacteria isolated from North American soils and two myxobacteria deposited at the ATCC. This approach indicated that previously classified *Archangium primigenium* ATCC 29037 is instead a novel member of the genus *Melittangium*, and that three of our four environmental isolates included potentially novel members of the genera *Corallococcus*, *Myxococcus*, and *Pyxidicoccus*. Previously classified *Chondorococcus macrosorus* ATCC 29039 was also determined to be a potentially novel member of the genus *Corallococcus*, with high similarity to *Corallococcus exercitus* DSM 108849T and phylogenetically distinct from *Myxococcus macrosorus* DSM 14697T previously assigned to the genus *Corallococcus*.

Subsequent bioinformatic analysis of biosynthetic pathways included in the newly sequenced genomes corroborated our preliminary taxonomic placements for each sample. Ultimately, this process identified *Archangium primigenium* to be a member of the lesser studied genus *Melittangium* and indicated that it should be prioritized for continued natural product discovery efforts. Of the environmental isolates, BGCs from SCPEA002 were determined to include the least amount of overlap with BGCs from other *Myxococcus/Pyxidicoccus* species. While environmental isolates SCHIC002 and NCSPR001 were also identified as novel members of the genera *Myxococcus* and *Corallococcus*, respectively, the apparent overlap in BGCs from thoroughly explored myxobacteria determined from sequence similarity network analysis suggests a limited potential for discovery of novel specialized metabolites. Overall, comparative genomic techniques including the assessment of biosynthetic potential enabled a phylogenetic approximation and suggested prioritization of *Archangium primigenium* for natural product discovery efforts from a sample set of six newly sequenced myxobacteria.
CHAPTER 5. RECLASSIFICATION OF *ARCHANGIUM PRIMIGENIUM* AS *MELITTANGIUM PRIMIGENIUM* COMB. NOV. AND ASSESSMENT OF ITS BIOSYNTHETIC POTENTIAL

5.1 Abstract

Myxobacteria are a unique group of Gram-negative bacteria best known for their well-coordinated social behaviors and the production of a plethora of specialized metabolites. Different myxobacteria need to be sequenced and classified to investigate the genetic basis of these activities and fully access novel specialized metabolites from myxobacteria. Here, we report the complete genome, description, and metabolic capacity of the myxobacterium *Archangium primigenium* ATCC 29037. *Archangium primigenium* ATCC 29037 displays the general myxobacterial features like rod-shaped vegetative cells, swarming on solid surfaces, and fruiting body formation. DNA–DNA hybridization and average nucleotide identity show relatedness between the *Archangium primigenium* ATCC 29037 and the closest recognized species, *Melittangium boletus* DSM 14713. Based on our analysis, we suggest reclassifying this bacterium to *Melittangium primigenium* and investigating its biosynthetic capacity.

5.2 Introduction

Previously classified as *Polyangium primigenium* Quehl 1906, *Archangium primigenium* (Quehl 1906) Jahn 1924 was investigated a few times in the literature. One of the earliest mentions of *Archangium primigenium* was by Elbein when he isolated six myxobacteria, including
Archangium primigenium, from the arid soil of Tuscon, Arizona, during his MSc training in 1956. Elbein isolated different myxobacteria and studied their association with organic matter, oxygen, and other organisms. In 1957, Peterson isolated and described the morphology of Archangium primigenium and other myxobacteria from different tree bark samples across Missouri for his Ph.D. work. This was the first mention of the isolation of myxobacteria from tree bark. In 1967, McDonald published two article series discussing the morphology and growth conditions of two species belonging to the Genus Archangium: Archangium gephyra and Archangium primigenium.

In 1972, Archangium primigenium MQ19 was isolated from soil and bark in southeast Queensland, Australia, by Skerman and McNeil who later deposited it in ATCC as Archangium primigenium (Quehl) Jahn 29037™.

Brockman and Todd described the fruiting bodies of different myxobacteria, including Archangium primigenium. They described the fruiting body of Archangium primigenium as ‘an irregular, swollen surface’. This description is confirmed by our observation of fruiting bodies formation while growing the strain in our lab.

As most of the work done on Archangium primigenium is on its morphology, we decided to apply modern genomics to study this myxobacterium further. The present study aims to justify and establish a new taxonomic position for Archangium primigenium ATCC 29037™ as Melittangium primigenium. Thus, we used divergent approaches based on a genotypic, phenotypic, and biochemistry analysis.

5.3 Materials and Methods
5.3.1. Bacterial Strains and Growth Conditions
*Archangium primigenium* ATCC 29037 was procured from the American Type Culture Collection (ATCC). It was cultured on VY/4 agar plates (2.5 g/L baker’s yeast, 1.5 g/L CaCl₂·2H₂O, 0.5 mg/L vitamin B₁₂, 15 g/L agar, pH 7.2)

5.3.2. Genomic DNA Isolation, Sequencing, Assembly, and Annotation

Genomic DNA of *Archangium primigenium* ATCC 29037 was obtained from actively growing bacteria using NucleoBond high molecular weight DNA kit (Macherey-Nagel, Bethlehem, PA, USA). The quantity and quality of the extraction were checked by gel electrophoresis along with Nanodrop method and followed by Qubit quantification using Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific). Sequencing was performed on a Pacific Biosciences single-molecule real-time (SMRT) sequencing platform at the MR DNA facility (Shallowater, TX, USA). Sequencing data have been deposited in NCBI under the accession number JADWYI000000000.1. Genome annotation was done the NCBI Prokaryotic Genome Annotation Pipeline when the genome was deposited in NCBI.

5.3.3 Polyphasic Taxonomy of *Archangium primigenium* ATCC 29037

5.3.3.1. Phenotypic Analysis

5.3.3.1.a Microscopy and morphological examination

Swarming and fruiting bodies on agar plates were observed under a Zeiss discovery V12 stereo microscope and photographed using a Zeiss axiocam105. Vegetative cells were also observed through field emission scanning electron microscopy (JEOL JSM-7200 FLV FESEM). Microscopy was performed in the Glycoscience Center of Research Excellence at the University of Mississippi.

5.3.3.3.1.b Growth response to pH, temperature, and antibiotics
Growth characteristics of *Archangium primigenium* ATCC 29037 were assessed at various temperatures at pH 7.5 and 0% NaCl, pH values at 30 °C and 0% NaCl, and salinities at pH 7.5 and 30 °C on VY/4 agar. Growth at different conditions was tested with 100 µL (OD<sub>500</sub> = 0.5) of liquid culture inoculated in the VY4 agar plates. The swarm diameter was measured after five days. Antibiotic resistance against eight antibiotics was tested on VY/4 agar at 30 °C for five days. Filter-sterilized antibiotics stock solutions were added to pre-cooled VY/4 agar. The final concentrations were set to 100 µg/mL for ampicillin, 30 µg/mL for chloramphenicol, 10 µg/mL for tetracycline hydrochloride, 40 µg/mL for kanamycin, 150 µg/mL for hygromycin, 50 µg/mL for gentamicin, 50 µg/mL for polymyxin B sulphate, and 60 µg/mL for spectinomycin. For the negative resistance blank, the test strain was grown on VY/4 agar without antibiotics and incubated for seven days at 30 °C.

5.3.3.1.c Enzymatic tests

The biochemical properties of *Archangium primigenium* ATCC 29037 were characterized using the API ZYM and API 20E kits (bioMérieux) according to the manufacturer’s instructions. Both kits are semi-quantitative micro methods consisting of a strip containing 20 microwells (capsules) containing the synthetic substrates as non-woven fibers. These capsules were inoculated with drops of a dense suspension of *Archangium primigenium* (OD<sub>500</sub> = 0.5). After 24 hrs incubation at 30 °C, the products produced in each capsule were detected by adding reagents that changed color in the event of a positive detection.

5.3.3.2. Genotypic Analysis

The bioinformatics tools we used for DNA analysis were Type (Strain) Genome Server (TYGS), Molecular Evolutionary Genetics Analysis across Computing Platforms (MEGA-X), ANI calculator, Yet Another Similarity Searcher (YASS), Antibiotics Resistance Target seeker.
(ARTS), and antiSMASH 6.0. TYGS is a free bioinformatics platform for whole genome-based taxonomic analysis (available at https://tygs.dsmz.de). It was used to calculate the dDDH values and construct minimum evolution trees using the Genome BLAST Distance Phylogeny approach (GBDP).\textsuperscript{173} GBDP trees were visualized using MEGA-X.\textsuperscript{86} ANI calculator is a free platform to calculate ANI scores between two genomic sequences (Available at http://enve-omics.ce.gatech.edu/).\textsuperscript{174}

YASS was used to align the myxobacteria genomes, analyze genomic similarity, and visualize the alignments (available at https://bioinfo.lifl.fr/yass/index.php).\textsuperscript{205} antiSMASH 6.0 is an online platform to identify BGCs in bacterial genomes based on hidden Markov models of genes. It was used for BGCs prediction using the “relaxed” strictness setting (available at https://docs.antismash.secondarymetabolites.org).\textsuperscript{44}

5.3.4 Metabolic Profiling of \textit{Archangium primigenium} ATCC 29037

5.3.4.1 Seed cultures and fermentation for secondary metabolites extraction

\textit{Archangium primigenium} ATCC 29037 was cultivated in a 50 mL pre-culture using a 250 mL shake flask containing 50 mL of optimized CYH medium (0.2% Bacto yeast extract, 0.3% Bacto casitone, 0.2% glucose, 0.1% CaCl$_2$.2H$_2$O, 0.1% MgSO$_4$.7H$_2$O, 50 mMHEPES, pH adjusted to 7.5 with KOH before autoclaving, after autoclaving Fe-EDTA (8 mg/L) and Vit B12 (5 mg/L) were added) for 8 days at 30 °C and 110rpm. This 50 mL pre-culture was used to inculcate 500 mL CYH media in 2 L shake flasks (10 flasks, a total of 5 L). After 10 days of incubation at 30°C and 110rpm, the culture was complemented with 40 mL of a 50% v/v sterilized XAD-16 resin and shaken for 4 days. The total fermentation period is 14 days.
5.3.4.2 Extraction and Fractionation

At the end of fermentation, wet cell mass and adsorbent resin XAD-16 were harvested by centrifugation at room temperature (20 min, 5000rpm). The supernatant was discarded, and the pellet was washed with desalted water and then lyophilized. The pellet was extracted with two portions of acetone followed by four portions of methanol (500 mL each for 30 minutes). The combined organic fractions were evaporated under reduced pressure, and the resulting residue was either dissolved in acetonitrile or water. One part of the crude extract was fractionated using reverse-phase C18 solid-phase extraction (SPE) using acetonitrile: water gradient to yield seven fractions. Another part of the crude extract was fractionated using Sephadex LH-20 column chromatography to yield 16 fractions. Fractions 3 and 4 of the Sephadex LH-20 column were further fractioned by another Sephadex LH-20 column.

5.3.4.3 Metabolic analysis

Fractions were sent for HPLC-MS/MS measurements. This analysis was performed on an Orbitrap Fusion instrument (Thermo Scientific). Samples were loaded onto a PepMap 100 C18 column, and separation of the samples was performed using mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile) at a rate of 6 µL/min. All data were acquired in positive ion mode. MS analysis was performed in the Glycoscience Center of Research Excellence at the University of Mississippi.

5.3.4.4 GNPS Analysis

MS raw data were converted to mzXML files using MSConvert. Mass spectrometry molecular networks were generated using the Global Natural Products Social Molecular Networking (GNPS) platform. Different GNPS in silico tools were used for the analysis, including DEREPLICATOR® (in silico database search tool for the annotation of metabolites in MS/MS data.
using in silico fragmentation tree), DEREPLICATOR (in silico database search tool for the annotation of known peptidic natural products in MS/MS data using in silico fragmentation tree), NAP (Network Annotation Propagation: in silico tool to improves the accuracy of in silico predictions), MolNetEnhancer (in silico tool to integrate the results of other metabolome annotation tools). Molecular networks were rendered, visualized, and annotated using Cytoscape version 3.9.1.

5.3.4.5 Antibacterial Activity

These fractions were tested for antibacterial activity against *Bacillus subtilis* ATCC 6051, and *Pseudomonas aeruginosa* ATCC 55953 using 96-well plate assayes at a final fraction concentration of 1 mg/mL. Water and DMSO were used as a negative control and tetracycline hydrochloride was used as a positive control.

5.4 Results and Discussion

5.4.1 Phenotypic Analysis

*Archangium primigenium* ATCC 29037 has bright reddish-orange, rod-shaped cells. Its swarm colony is soft, thin, and surrounded by a clear halo due to the lysis of yeast cells as often described for most predatory myxobacteria. *Archangium primigenium* shows the typical gliding and circular swarming activity on solid agar surfaces. Under the microscope, vegetative cells were reddish-orange, rod-shaped with slight tapering ends, measuring 0.35 - 0.45 µm × 4.00 - 4.70 µm (Figure 5.1).

The growth response of the strain to different temperatures was tested at 20, 25, 30, 35, 37, and 40 °C (at pH 7.5). Growth is possible between 20 °C and 35 °C, with optimum growth temperature is at 30 °C, with no growth at 37 °C or above. The pH dependence of *Archangium*
*primigenium* strains growth was also tested (at 30 °C). The bacterium tolerates a wide pH range, from pH 6 - 10, but maximum growth was seen at pH 7 - 7.5, and there was no evidence of growth below pH 5.5. *Archangium primigenium* shows weak growth at 1% NaCl and no growth at 2%, 3%, 4%, and 5% NaCl. Antibiotic resistance testing shows that *Archangium primigenium* is sensitive to tetracycline hydrochloride, ampicillin, kanamycin, and chloramphenicol, and resistant to gentamicin and polymyxin B sulphate. ARTS demonstrated the presence of different genes encoding multidrug resistance efflux pumps, beta-lactamase, and aminoglycoside acetyltransferase within the genome of this strain.

Figure 5.1 Colony and cellular morphology of *A. primigenium* ATCC 29037. Left: Colony morphology of *A. primigenium* swarm with; Right. Cellular morphology as visualized with Scanning Electron Microscopy (SEM).

API 20 NE and API Zym are general biochemical (i.e. enzymological) evaluation methods for assessing microbial functional diversity.²⁰⁷,²⁰⁸ The API system has successfully been used to produce an enzymatic profile useful for the taxonomic classification of organisms including myxobacteria.²⁰⁹ The biochemical activities of these strains were tested using API 20 NE and API Zym kits and the results are summarized in Table 5.1 and Appendix F.
5.4.2 Genotypic analysis

The genome sequence of *Archangium primigenium* ATCC® 29037 is approximately 9.49 Mb in size, with a G+C content of 70.7%, and contains a total of 7,691 protein-coding sequences (CDSs), 12 rRNA (16S-23S-5S), and 63 tRNA genes, according to the annotation NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

Full sequence 16S rRNA gene analysis shows that *Archangium primigenium* ATCC 29037 is closer to *Melittangium boletus* DSM 14713T than *Archangium gephyra* DSM 2261T. The full 16S gene sequence of *Archangium primigenium* ATCC 29037 has a homology of 98.89% with the type strain *Melittangium boletus* DSM 14713T, 98.7% with *Corallococcus fuscus* DSM 52655, and 97.72% with the type strain *Archangium gephyra* DSM 2261T (Table 5.2). According to the 16S similarity cut-off, if the 16S similarity is below 98.7%, the strain can be categorized as a new species. This means that *Archangium primigenium* ATCC 29037 could be a sub-species of *Melittangium boletus* DSM 14713T based on the 16S homology.

The phylogenetic tree reconstructed by the maximum-likelihood method from the whole genomes of different myxobacteria using the GBDP approach shows that the three *Archangium* strains, with full/partial genomes deposited in NCBI (*Archangium sp*. Cb G35, *Archangium gephyra* DSM 2261T, and *Archangium violaceum* Cb vi76), formed a separate cluster without *Archangium primigenium* ATCC 29037. *Archangium primigenium* ATCC 29037 clustered with *Melittangium boletus* DSM 14713T within the tree. Similar trend can be seen in phylogenetic trees based on 16S rDNA. The phylogenetic trees based on 16S rDNA gene sequence and whole-genome sequencing are in Figure 5.2.A and B.
Table 5.1 Biochemical Activities of *A. primigenium* ATCC 29037 as tested by API 20 NE and API Zym

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Activity</th>
<th><em>A. primigenium</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Potassium nitrate</td>
<td>Nitrate reduction</td>
<td>-</td>
</tr>
<tr>
<td>L-Tryptophane</td>
<td>Indole production</td>
<td>-</td>
</tr>
<tr>
<td>L-Arginine</td>
<td>Arginine dihydrolase</td>
<td>-</td>
</tr>
<tr>
<td>Urea</td>
<td>Urease</td>
<td>-</td>
</tr>
<tr>
<td>Esculin ferric citrate</td>
<td>Esculin hydrolysis</td>
<td>++++</td>
</tr>
<tr>
<td>Gelatin (bovine origin)</td>
<td>Gelatin hydrolysis</td>
<td>++++</td>
</tr>
<tr>
<td>4-Nitrophenyl-β-D-galactopyranoside</td>
<td>p-Nitrophenyl-β-D-galactopyranoside</td>
<td>++++</td>
</tr>
<tr>
<td>Phenylacetic acid</td>
<td>Phenyl acetate assimilation</td>
<td>-</td>
</tr>
<tr>
<td>2-Naphthyl phosphate</td>
<td>Alkaline dephosphorylation</td>
<td>++++</td>
</tr>
<tr>
<td>2-Naphthyl butyrate</td>
<td>Esterase (C₄)</td>
<td>+++</td>
</tr>
<tr>
<td>2-Naphthyl caprylate</td>
<td>Esterase Lipase (C₈)</td>
<td>+++</td>
</tr>
<tr>
<td>2-Naphthyl myristate</td>
<td>Lipase (C₁₄)</td>
<td>+++</td>
</tr>
<tr>
<td>L-Leucyl-2-naphthylamide</td>
<td>Leucine arylamidase</td>
<td>++++</td>
</tr>
<tr>
<td>L-Valyl-2-naphthylamide</td>
<td>Valine arylamidase</td>
<td>++++</td>
</tr>
<tr>
<td>L-Cystyl-2-naphthylamide</td>
<td>Cystine arylamidase</td>
<td>+++</td>
</tr>
<tr>
<td>N-Benzoyl-DL-arginine-2-naphthylamide</td>
<td>Trypsin</td>
<td>+++</td>
</tr>
<tr>
<td>N-Glutaryl-phenylalanine-2-naphthylamide</td>
<td>D-chymotrypsin</td>
<td>-</td>
</tr>
<tr>
<td>2-Naphthyl phosphate</td>
<td>Acid phosphatase</td>
<td>++++</td>
</tr>
<tr>
<td>Naphthol-AS-BI-phosphate</td>
<td>Naphthol-AS-BI-phosphohydrolase</td>
<td>++++</td>
</tr>
<tr>
<td>6-Br-2-Naphthyl-β-D-galactopyranoside</td>
<td>D-galactosidase</td>
<td>-</td>
</tr>
<tr>
<td>2-Naphthyl-β-D-galactopyranoside</td>
<td>β-galactosidase</td>
<td>+</td>
</tr>
<tr>
<td>Naphthol-AS-BI-β-D-glucuronide</td>
<td>β-glucuronidase</td>
<td>-</td>
</tr>
<tr>
<td>2-Naphthyl-β-D-glucopyranoside</td>
<td>D-glucosidase</td>
<td>+</td>
</tr>
<tr>
<td>6-Br-2-Naphthyl-β-D-glucopyranoside</td>
<td>β-glucosidase</td>
<td>++</td>
</tr>
<tr>
<td>1-Naphthyl-N-acetyl-β-D-glucosaminide</td>
<td>N-acetyl-β-D-glucosaminidase</td>
<td>++++</td>
</tr>
<tr>
<td>6-Br-2-Naphthyl-β-D-mannopyranoside</td>
<td>D-mannosidase</td>
<td>-</td>
</tr>
<tr>
<td>2-Naphthyl-β-D-fucopyranoside</td>
<td>D-fucosidase</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 5.2 16S rRNA identity, ANI, and dDDH values for pairwise comparisons between *A. primigenium* with the most similar fully sequenced myxobacteria.

<table>
<thead>
<tr>
<th>species</th>
<th>16s rRNA</th>
<th>dDDH</th>
<th>ANI</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. boletus</em> DSM 14713&lt;sup&gt;T&lt;/sup&gt;</td>
<td>98.89%</td>
<td>29.5</td>
<td>86.1%</td>
</tr>
<tr>
<td><em>C. fuscus</em> DSM 52655</td>
<td>98.7%</td>
<td>24.5</td>
<td>83.29%</td>
</tr>
<tr>
<td><em>A. gephyra</em> DSM 2261&lt;sup&gt;T&lt;/sup&gt;</td>
<td>97.72%</td>
<td>23.2</td>
<td>81.41%</td>
</tr>
<tr>
<td><em>S. aurantiaca</em> DW43-1</td>
<td>96.06%</td>
<td>20</td>
<td>78.9%</td>
</tr>
<tr>
<td><em>M. macrosporus</em> DSM 14697&lt;sup&gt;T&lt;/sup&gt;</td>
<td>96.63%</td>
<td>19.8</td>
<td>78.85%</td>
</tr>
</tbody>
</table>

Figure 5.2 Minimum evolution tree of *A. primigenium* and different members of the family Archangiaceae (A) based on 16S rDNA gene sequence (B) using the GBDP approach. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 100.0%. The tree was rooted at the midpoint. *Myxococcus xanthus* DK 1622 is included in the trees as an outgroup.

A genomic comparison between a strain and a Type strain is the way to obtain a robust conclusion about a novel species. Novelty can be assumed when ANI < 95–96% or DDH of 70%.<sup>68</sup> According to our analysis, the ANI value of strain *Archangium primigenium* against *Melittangium boletus* DSM 14713<sup>T</sup> and *Archangium gephyra* DSM 2261<sup>T</sup> are 86.1% and 81.41%, respectively.
These values show that strains *Archangium primigenium* and *Melittangium boletus* are not identical, and the former could be a strain that belongs to the genus *Melittangium*, not *Archangium*. Genome Blast Distance Phylogeny (GBDP) analysis produced a DDH estimation for *Archangium primigenium* compared to *Melittangium boletus* of 29.5 (Table 5.2, Figure 5.2.B. According to this result, the two strains are distinct species.

We also performed analyses with other bioinformatics tools. The aim is to analyze the differences and similarities between the genomic features of strains *Archangium primigenium*, *Melittangium boletus* DSM 14713<sup>T</sup>, and *Archangium gephyra* DSM 2261<sup>T</sup>. We compared the genome of *Archangium primigenium* with *Melittangium boletus*, *Archangium gephyra*, and other completely sequenced *Myxococcales* genomes by aligning their DNA sequences using the YASS platform. This is a genome alignment pipeline that can rapidly align complete or draft genomes. YASS generates dotplots from genomic alignment for easier visualization of these alignments. Dotplots compare two sequences by organizing the one genome on each axis (x-axis, y-axis) of the plot. A dot is drawn in a location when the residues of both sequences match at the same location. Both forward and reverse complements were performed in these dotplots. Forward matches are shown in green, while reverse matches are shown in red. Our dotplots were generated at an E-value threshold of 1.0E<sup>-40</sup> and an X-drop threshold of 80. The number of alignments used to draw these dotplots is listed in Table (5.3). The genome of *Melittangium boletus* DSM 14713 shows a unique pattern of collapse when compared with the genome of *Archangium primigenium*. The rest of the myxobacteria show different patterns of insertions when comparing their genomes to the genome of *Archangium primigenium*. The generated dotplots can be found in Appendix G. YASS results show that the genome of the Myxobacterium *Archangium primigenium* aligns well and with a higher average identity of alignments with *Melittangium boletus* (87.81%) and
Corallococcus fuscus (86.13%) than Archangium gephyra (85.26%). The result of the YASS analysis of A is summarized in Table (5.3)

Table 5.3 YASS analysis of nine sequenced myxobacteria with *A. primigenium*

<table>
<thead>
<tr>
<th>Myxobacteria</th>
<th>% Aligned Bases</th>
<th>Number of Alignments</th>
<th>Total Length of Alignments</th>
<th>Avg Length of Alignments</th>
<th>% Avg Identity of Alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. boletus</em></td>
<td>64.47%</td>
<td>2047</td>
<td>6067966</td>
<td>2964.32</td>
<td>87.81%</td>
</tr>
<tr>
<td><em>C. fuscus</em></td>
<td>45.49%</td>
<td>2505</td>
<td>4282700</td>
<td>1709.66</td>
<td>86.13%</td>
</tr>
<tr>
<td><em>A. gephyra</em></td>
<td>36.72%</td>
<td>2288</td>
<td>3442686</td>
<td>1504.67</td>
<td>85.26%</td>
</tr>
<tr>
<td><em>M. macrosporus</em></td>
<td>15.16%</td>
<td>1364</td>
<td>1417458</td>
<td>1039.19</td>
<td>83.94%</td>
</tr>
<tr>
<td><em>C. coralloides</em></td>
<td>15.61%</td>
<td>1481</td>
<td>1461148</td>
<td>986.60</td>
<td>83.92%</td>
</tr>
<tr>
<td><em>M. fulvus</em></td>
<td>14.76%</td>
<td>1385</td>
<td>1384175</td>
<td>999.40</td>
<td>83.92%</td>
</tr>
<tr>
<td><em>M. stipitatus</em></td>
<td>12.54%</td>
<td>1180</td>
<td>1168526</td>
<td>990.28</td>
<td>83.91%</td>
</tr>
<tr>
<td><em>M. xanthus</em></td>
<td>13.10%</td>
<td>1198</td>
<td>1225205</td>
<td>1022.71</td>
<td>83.86%</td>
</tr>
<tr>
<td><em>M. hansupus</em></td>
<td>13.87%</td>
<td>1301</td>
<td>1300928</td>
<td>999.94</td>
<td>83.81%</td>
</tr>
<tr>
<td><em>S. aurantiaca</em></td>
<td>16.73%</td>
<td>1413</td>
<td>1559511</td>
<td>1103.69</td>
<td>83.80%</td>
</tr>
</tbody>
</table>

5.4.3 Estimation of biosynthetic potential of *Archangium primigenium*

The genome sequence of *Archangium primigenium* was screened for natural products biosynthetic gene clusters with the antiSMASH, version 5.2.0, tool. 32 unique secondary metabolite biosynthetic pathways were identified, mainly pathways for terpenes, ribosomally synthesized and post-translationally modified peptides (RiPPs), nonribosomal peptides (NRPs), polyketides (PKs), and hybrid nonribosomal peptide-polyketides (Figure 5.3). These pathways were sequenced in full as we have very high quality one long read contig; no BGC is located on contig edges. Half of BGCs are similar to known BGCs from the MIBiG (Minimum Information about a Biosynthetic Gene cluster) repository (only 7 BGCs with similarity >50%) suggesting high
potential for encoding novel metabolites. Similar observations were made in chapter four, in our BiG-SCAPE analysis (Figure 4.7). Some of these BGCs are similar to other known BGCs with characterized products: geosmin (a terpene molecule found in almost all myxobacterial extracts), 1-heptadecene, aurafuron (antibiotic polyketide), and carotenoid. Generally, *Archangium primigenium* devotes 12.87% of the genome to secondary metabolism.

![Figure 5.3 A) antiSMASH v6.0 predicted biosynthetic gene clusters form the genome of *A. primigenium*. B) A representation of the location of different BGCs on *A. primigenium* genome](image)

5.4.4 Secondary Metabolites Extraction

A figure of the corresponding Cytoscape-rendered molecular network of SPE fractions is provided (Appendix H). The fractions are inactive against the tested organisms (Appendix I). As most of the BGCs in this genome encode RiPPs and NRPs secondary metabolites, we decided to follow a GNPS-guided fragmentation to follow the fractions with the highest peptide content based on the MS/MS fragmentation pattern of peptides. This can be done using different in-silico tools available in GNPS. These include DEREPLICATOR+ (in-silico database search tool for the annotation of metabolites in MS/MS data using in silico fragmentation tree), DEREPLICATOR
From this analysis, we can see that Fraction 3 and Fraction 4 are prioritized for peptide isolation (Appendices J, K, L).

5.5 Conclusion and Recommendations

To fully access novel specialized metabolites from myxobacteria, it is essential to understand the relationships between myxobacterial evolution, taxonomy, and genomic variation. This study aims to reclassify *Archangium primigenium* ATCC 29037 to *Melittangium primigenium* ATCC 29037\(^T\) and to report its complete genome, description, and metabolic capacity. In all pairwise comparisons we performed (ANI, dDDH, YASS, 16S RNA), *Archangium primigenium* ATCC 29037 shows higher similarity to *Melittangium boletus* DSM 14713\(^T\) than *Archangium gephyra* DSM 2261\(^T\). As *Archangium primigenium* ATCC 29037 shares 16S similarity > 98.7%, ANI < 95–96, dDDH < 70% with *Melittangium boletus* DSM 14713\(^T\), we propose *Archangium primigenium* to be a distinct strain of the genus *Melittangium* (Figure 1.2).

**Description of *Melittangium primigenium* comb. nov.**

*Melittangium primigenium* (Gr. n. *melitta*, bee; Gr. neut. n. *angeion*, vessel (Latin transliteration angium); N.L. neut. n. *Melittangium*, a vessel resembling a honeycomb, pri.mi’ge.ni.um. L. gen. n. from primus (first) + genus (birth, origin) + ium. Vegetative cells are reddish-orange, rod-shaped with slight tapering ends, 0.7 µm wide and 5.6 µm long. Fruiting bodies are spherical to irregularly shaped, cushion-like, and reddish orange in color. Growth is possible between 20 °C and 35 °C, with optimum growth conditions
for strain 30 °C in a pH range of 7.0 – 7.5. They do not utilize monosaccharides or disaccharides well. Hydrolyses chitin, gelatin, and esculin, but not starch, cellulose, or agar.

Finding new antimicrobials from bacteria is a challenging process. Myxobacteria are slow-growing bacteria which means more time to collect the crude extract and high possibility of contamination. In our case, antiSMASH and Big-Scape analysis shows the high potential of this strain. However, screening for antibacterial bioactivity showed that all fractions of all different chromatographic systems we tested (RP-SPE, and Sephadex LH-20) are inactive against the tested bacteria. It is possible that more bacteria needed to be tested. Additionally, other microorganisms can be taken into consideration (e.g., Fungi). Another possibility is that these BGCs are silent or produce undetectable amounts of the products in the tested growth conditions.

Bacterial growth and secondary metabolites production depend heavily upon environmental conditions. When bacteria grow in media with different parameters (such as temperature, pH and nutrients) from their original environment, their production of secondary metabolites might be altered or stopped. Optimizing growth conditions (different growth media or co-culture with other organisms) or genetic manipulation (heterologous expression in a more suitable host or promoter insertion to overexpress the BGCs) can improve the process of isolation of active metabolites. In this study we were able to identify fractions rich with peptides. Extra fractionation is necessary to get pure peptides.
LIST OF REFERENCE


(54) Mohr, K. I.; Wolf, C.; Nübel, U.; Szafrańska, A. K.; Steglich, M.; Hennessen, F.; Gemperlein, K.; Kämpfer, P.; Martin, K.; Müller, R.; Wink, J. A Polyphasic Approach Leads to Seven New Species of the Cellulose- Decomposing Genus *Sorangium*, *Sorangium ambruticinum* sp. nov., *Sorangium arenae* sp. nov., *Sorangium bulgaricum* sp. nov., *Sorangium dawidii* sp. nov., *Sorangium kenyense* sp. nov., *Sorangium orientale* sp. nov. and *Sorangium reichenbachii* sp. nov. *International Journal of Systematic and


(64) Das, S.; Dash, H. R.; Mangwani, N.; Chakraborty, J.; Kumari, S. Understanding Molecular Identification and Polyphasic Taxonomic Approaches for Genetic Relatedness


(156) Yamamoto, E.; Muramatsu, H.; Nagai, K. *Vulgatibacter incomptus* gen. nov., sp. nov. and *Labilithrix luteola* gen. nov., sp. nov., Two Myxobacteria Isolated from Soil in Yakushima


(193) Hoshino, T. Violacein and Related Tryptophan Metabolites Produced by *Chromobacterium violaceum*: Biosynthetic Mechanism and Pathway for Construction of


(201) Elbein, A. D. Studies of Some of the Myxobacteria Isolated from Soils of the Tucson Area, The University of Arizona, 1956.


LIST OF APPENDICES
Appendix A.

Sequence data for codon optimized AgpI construct purchased from Genscript

CATATGTTCCAGGTCACATTCATCCACGCCGCTAGAACGTCATGACGGGACTGGACATCGAACAATCC
CTTTCGTATCCGCTACGATATTTGATGGTCGAAATGGCTGATGACGCTGATGACGCTGATGACGCTG
GTTGCGGGGTGAGGCCTCGCTGCCGACCCGGAGCCTCGCTGCCGACCCGGAGCCTCGCTGCCGACCCGG
GAACATCGTGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTG
TGCTGAAAGCCACGGTCATCAGGCGGGATGACGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCT
GCCAAAACATCCGTAAAATAGCGCGCGCGTTGTCAGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCT
TGCAAGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCT
TTAGCATTCGCTTTAACCGTGAACTTC

Sequence data for codon optimized VitI construct purchased from Genscript

CATATGTTCAAGATCCACGTCATCCACGCCGCTAGAACGTCATGACGGGACTGGACATCGAACAATCC
CTTTCGTATCCGCTACGATATTTGATGGTCGAAATGGCTGATGACGCTGATGACGCTGATGACGCTG
GTTGCGGGGTGAGGCCTCGCTGCCGACCCGGAGCCTCGCTGCCGACCCGGAGCCTCGCTGCCGACCCGG
GAACATCGTGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTG
TGCTGAAAGCCACGGTCATCAGGCGGGATGACGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCT
GCCAAAACATCCGTAAAATAGCGCGCGCGTTGTCAGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCT
TGCAAGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCTGCTGGCAGCTGCGCT
TTAGCATTCGCTTTAACCGTGAACTTC
Appendix B.

Locations of soil samples used for environmental strains isolation

<table>
<thead>
<tr>
<th>Myxobacteria</th>
<th>City</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCSPR001 &amp; NCCRE002</td>
<td>Asheville, NC, USA</td>
<td>35°37'30.3&quot;N 82°32'53.2&quot;W</td>
</tr>
<tr>
<td>SCHIC003 &amp; SCPEA002</td>
<td>Landrum, SC, USA</td>
<td>35°08'41.3&quot;N 82°09'48.2&quot;W</td>
</tr>
</tbody>
</table>
Appendix C.

A list of all myxobacteria and their accession numbers used in BiG-SCAPE analysis.

1. Aggregicoccus sp.17bor-14 (NZ_VJZZ00000000.1)
2. Anaeromyxobacter dehalogenans 2CP-C (NC_007760.1)
3. Anaeromyxobacter sp Fw109-5 (NC_009675.1)
4. Anaeromyxobacter sp. K (NC_011145.1)
5. Anaeromyxobacter dehalogenans 2CP-1 DSM 21875\(^T\) (NC_011891.1)
6. Archangium sp.Cb G35 (NZ_MPOI00000000.1)
7. Archangium gephyra DSM 2261\(^T\) (NZ_CP011509.1)
8. Archangium violaceum Cb vi76 (NZ_JPMI00000000.1)
9. Corallococcus exiguus DSM 14696\(^T\) (NZ_JAAAPK00000000.1)
10. Corallococcus exercitus AB043B (NZ_JABFJV00000000.1)
11. Corallococcus carmarthensis CA046B (JABFJX00000000.1)
12. Corallococcus macrosporus HW-1 (NC_015711.1)
13. Corallococcus coralloides DSM 2259\(^T\) (NC_017030.1)
14. Corallococcus coralloides B035 (NZ_CP034669.1)
15. Corallococcus terminatus CA054A DSM 108848\(^T\) (RAVZ00000000.1)
16. Corallococcus lansteffanensis CA051B DSM 108844\(^T\) (RAWB00000000.1)
17. Corallococcus sicarius CA040B DSM 108850\(^T\) (RAWG00000000.1)
18. Corallococcus aberystwythensis AB050A DSM 108846\(^T\) (RAWK00000000.1)
19. Corallococcus intemperatum AB047A DSM 108843\(^T\) (NZ_RAWM00000000.1)
20. Cystobacter fuscus DSM 2262\(^T\) (NZ_ANAH00000000.2)
21. Cystobacter fuscus DSM 52655 (NZ_CP022098.1)
22. Cystobacter ferruginues Cbfe23 (NZ_MPIN00000000.1)
23. Hyaangium minutum DSM 14724\(^T\) (NZ_JMCB00000000.1)
24. Melittangium boletus DSM 14713\(^T\) (NZ_CP022163.1)
25. Myxococcus fulvus 124B02 (CP006003.1)
26. Myxococcus fulvus DSM 16525\(^T\) (NZ_FOIB00000000.1)
27. Myxococcus hansupus mixupus (NZ_CP012109.1)
28. Myxococcus macrosporus DSM 14697\(^T\) (NZ_CP022203.1)
29. Myxococcus stipitatus DSM 14675\(^T\) (NC_020126.1)
30. Myxococcus xanthus DK1622 (NC_008095.1)
31. Myxococcus xanthus GH3.5.6c2 (NZ_CP017169.1)
32. Myxococcus xanthus GH5.1.9c20 (NZ_CP017170.1)
33. Myxococcus xanthus KF3.2.8c11 (NZ_CP017171.1)
34. Myxococcus xanthus KF4.3.9c1 (NZ_CP017172.1)
35. Myxococcus xanthus MC3.3.5c16 (NZ_CP017173.1)
36. Myxococcus xanthus MC3.5.9c15 (NZ_CP017174.1)
37. Myxococcus xanthus DK1622 (NZ_CP065375.1)
38. Myxococcus virescens NBRC 100334 (NZ_JBVY00000000.1)
39. Myxococcus virescens DSM 2260\(^T\) (NZ_FAJ00000000.1)
40. Ptyxadicoccus fallax DSM 14698\(^T\) (JABBJ00000000.1)
41. Stigmatella aurantiaca DSM 17044\(^T\) (NZ_FOAP00000000.1)
42. Stigmatella aurantiaca DW43-1 (NC_014623.1)
43. Stigmatella aurantiaca DW43-1 (NC_014623.1)
44. Vitosangium sp. GDMCC1.1324 (NZ_PZOX00000000.1)
45. Vulgatiibacter inquisitus DSM 27710 (NZ_CP012332.1)
Appendix D

Minimum evolution tree from the whole genomes of *Corallococcus macrosporus* ATCC 29039, NCCRE002, NCSPR001 and different members of genus *Corallococcus* using the GBDP approach. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 100.0 %. Branch pseudo-bootstrap less than 50% are not shown. The numbers below branches are branch lengths scaled in terms of GBDP distance formula d5. The tree was rooted at the midpoint.
Appendix E

Minimum evolution tree from the whole genomes of SCHIC003, SCPEA002, and different members of *Myxococcus* and *Pyxidicoccus* using the GBDP approach. The numbers in bold above the branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 100.0 %. Branch pseudo-bootstrap less than 50% are not shown. The numbers below branches are branch lengths scaled in terms of GBDP distance formula $d_5$. The tree was rooted at the midpoint.
Appendix F

API 20 NE and API Zym results for *A. primigenium* ATCC 29037
Appendix G

Dotplots representing genomic alignments of *Archangium primigenium* ATCC 29037 with fully sequenced myxobacteria using YASS algorithm. Forward matches are shown in green, while reverse matches are shown in red.

*Melittangium boletus* DSM 14713

*Cystobacter fuscus* DSM 52655

*Stigmatella aurantiaca* DW4/3-1

*Archangium gephyra* DSM 2261

*Sorangium cellulosum* So ce 56

*Myxococcus stipitatus* DSM 14675
Appendix H

Molecular network of fractions from RP-SPE column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional with similarity to MS/MS from GNPS libraries.
Appendix I:
Molecular network of fractions from RP-SPE column as visualized with Cytoscape. Each node is represented with pie-chart that represent the fractions in which that MS/MS appears.

Fraction number:
Appendix J

Antibacterial activities of SPE and Sephadex LH-20 column chromatography fractions against different bacteria
Appendix K

Molecular network of fractions from Sephadex LH-20 column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right) Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appears.

**Color code:**
- Red: Peptide and Amino acids
- Blue: Terpenes
- Green: Flavonoids
- Lime: Phenolic compounds
- Purple: Macrolides
- Orange: Steroids
- Black: No Match

**Color Code:**
- Green: Fraction 3
- Yellow: Fraction 4
- Purple: Fraction 5
- Red: Fraction 6
- Blue: Fraction 7
- Orange: Fraction 8
Appendix L

Molecular network of fraction 3 from Sephadex LH-20 column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appear.

**Color code:**
- Red: Peptide and Amino acids
- Blue: Terpenes
- Green: Flavonoids,
  Lime: Phenolic compounds,
- Purple: Macrolides
- Orange: Steroids
- Black: No Match

**Color Code:**
- Green: Fraction 3-1 and 3-2
- Yellow: Fraction 3-3 and 3-4
- Purple: Fraction 3-5 and 3-6
- Red: Fraction 3-7, 3-8, and 3-9
Appendix M

Molecular network of fraction 3 from Sephadex LH-20 column as visualized with Cytoscape showing the peptide rich fractions. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appear.

**Color code:**
- Red: Peptide and Amino acids
- Blue: Terpenes
- Green: Flavonoids,
- Lime: Phenolic compounds,
- Purple: Macrolides
- Orange: Steroids
- Black: No Match

**Color Code:**
- Green: Fraction 3-1 and 3-2
- Yellow: Fraction 3-3 and 3-4
- Purple: Fraction 3-5 and 3-6
- Red: Fraction 3-7, 3-8, and 3-9
Appendix N

Molecular network of fraction 4 from Sephadex LH-20 column as visualized with Cytoscape. Each node represents one MS/MS spectrum, where the colors of the nodes represent different chemical classes. Size of the nodes is proportional to the intensity of MS. Left: Molecular network by chemical class. Right: Molecular network with each node is represented with pie-chart that represent the fractions in which that MS/MS appear.

**Color code:**
- Red: Peptide and Amino acids
- Blue: Terpenes
- Green: Flavonoids
- Lime: Phenolic compounds
- Purple: Macrolides
- Orange: Steroids
- Black: No Match

**Color Code:**
- Green: Fraction 4-1 and 4-2
- Yellow: Fraction 4-3 and 4-4
- Purple: Fraction 4-5 and 4-6
- Red: Fraction 4-7 and 4-8
- Blue: Fractions 5-3 and 5-4
VITA

EDUCATION & RESEARCH EXPERIENCE

2009 Bachelor of Science in Pharmacy, Faculty of Pharmacy, Jordan University of Science and Technology, Jordan

2009-2010 Teaching Assistant, Faculty of Pharmacy, Jordan University of Science and Technology, Jordan

2010-2013 Graduate Teaching and Research Assistant, Faculty of Pharmacy, Jordan University of Science and Technology, Jordan

2014 Master of Science in Pharmacognosy, Faculty of Pharmacy, Jordan University of Science and Technology, Jordan

2014-2015 Lecturer, Department of Pharmacy, Faculty of Pharmacy, Al-Zaytoonah University of Jordan, Jordan

2015-2017 Lecturer, Faculty of Pharmacy, Yarmouk University, Jordan

2022 Doctor of Philosophy in Pharmaceutical Sciences, Department of BioMolecular Sciences, School of Pharmacy, University of Mississippi, University, MS, USA

FUNDING & SELECTED AWARDS

2021 Outstanding Graduate Student in Pharmacognosy Award. University of Mississippi, Department of Biomolecular Sciences

2017-2021 Fulbright Pre-Doctoral Scholarship/Ph.D.: Amideast-Jordan

PUBLICATIONS

*Authors contributed equally


PRESENTATIONS
• 2022 American Society of Pharmacognosy Annual Meeting, Charleston, SC. Poster (July, 2022)
• The Annual GSC Research Symposium, University of Mississippi, MS. Virtual Conference: Poster (April, 2021)
• World Microbe Forum by the American Society for Microbiology (ASM) and the Federation of European Microbiological Societies (FEMS). Virtual Conference: Poster (June, 2021)
• Drug Discovery and Development Colloquium (DDDC 2021) by the University of Arkansas for Medical Sciences and the University of Mississippi. Virtual Conference: Poster (July, 2021)
• Vanguards of Natural Product Research 2021 by the American Society of Pharmacognosy Virtual Conference: Poster (July, 2021)
• The 2021 SACNAS National Diversity in STEM (NDiSTEM) Digital Conference: Presentation (October, 2021)
• GlyCORE Symposium, University of Mississippi, MS. In-person Conference: Poster (October 2021)
• The School of Pharmacy Annual Poster Session. University of Mississippi, MS. In-person Conference: Poster (October, 2021)