12-2021

Comparative Genome Analysis of Fungal Antagonists Marinomonas ostreistagni 398 and M. spartinae 468.

Jessie L. Fields

Comparative Genome Analysis of Fungal Antagonists *Marinomonas ostreistagni* 398 and *M. spartinae* 468

by

Jessie Lynda Elizá-Beth Fields

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

December 2021

Approved by:

Dmitri Mavrodi, Ph.D., Thesis Advisor,
School of Biological, Environmental and Earth Sciences

Jacob Schaefer, Ph.D., Director,
School of Biological, Environmental and Earth Sciences

Sabine Heinhorst, Ph.D., Dean,
Honors College

ABSTRACT

Under certain conditions, the *Spartina alterniflora* and *Juncus roemerianus* grasses in marshes undergo progressive decline in an event known as Saltwater Marsh Dieback, which may be attributed to the presence of the plant pathogenic fungi *Fusarium*. The microbiomes of *S. alterniflora* and *J. roemerianus* from Deer Island, MS were characterized and *Fusarium* suppressing bacteria were identified. Among isolates capable of antagonizing *Fusarium* were *Marinomonas ostreistagni* 398 and *M. spartinae* 468. Despite the progress in understanding the diversity of *Marinomonas*, our ability to explain ecological, metabolic, and biochemical traits of marinomonads at the genomic sequence level remains limited. Analysis revealed that *Marinomonas* genomes form three distinct clades supported by the relatedness of orthologous genes. Heterogeneity is reflected in the core genome representing only 50-60% of any individual strain. Genes for the synthesis of siderophores and other secondary metabolites were identified. Clade- and strain-specific genomic regions contained mobile genetic elements. These results provide insights into the genomic diversity of *Marinomonas* by characterizing genes for the adaptation to hypersaline environments, pathways involved in the interaction with plants, and the production of antimicrobial compounds.

Keywords:

Genomic diversity, comparative analysis, *Marinomonas*, phylogeny, genome annotation

DEDICATION

Dedicated to my family who supported me through my research and to my mentor, Dr. Mavrodi, who pushed me to achieve goals that did not seem within my grasp and allowed me chances to learn and grow as a student.

ACKNOWLEDGMENTS

This study was made possible with funding received from the Drapeau Center for Undergraduate Research through the Eagle SPUR program and with contributions in sequencing from the Molecular and Genomics Core facility at the University of Mississippi Medical Center. The completion of this project would not have been made possible without the assistance received from the Honors College, Dr. Westley Follett, and Dr. Dmitri Mavrodi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

antiSMASH	Antibiotics & Secondary Metabolite Analysis Shell
ARTS	Antibiotic Resistant Target Seeker
BLAST	Basic Local Alignment Search Tool
BLASTn	Basic Local Alignment Search Tool Nucleotide
BLASTp	Basic Local Alignment Search Tool Protein
BPA	Bisphenol A
Contig	Contiguous Sequence
CDPS	Cyclodipeptide synthase
CDS	Coding DNA Sequence
COG	Clusters of Orthologous Groups
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic acid
EC	Enzyme Commission
FASTA	Fast-all
FastQC	Fast Quality Control
$FeSO_4$	Iron II Sulfate
GC	Guanfacine-Cytosine
GI	Genomic Island
GOLD	Genomes Online Database
HGT	Horizontal Gene Transfer
HMM	Hidden Markov Models

JGI-IMG/M	Joint Genome Institute - Integrated Microbial Genomes & Microbiomes
K_2HPO_4	Dipotassium Hydrogen Phosphate
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
MCL	Markov Cluster
NCBI	National Center for Biotechnology Information
NH_4Cl	Ammonium chloride
NRPS	Nonribosomal Peptide Synthase
ORF	Open Reading Frame
PAH	Polycyclic Aromatic Hydrocarbons
PATRIC	Pathosystems Resource Integration Center
PEC	Protein Enzyme Commission
PGFams	Global Protein Families
PHASTER	PHAge Search Tool Enhanced Release
PKS	Polyketide Synthase
RAST	Rapid Annotation using Subsystem Technology
RiPPS	Ribosomally Synthesized and Post-Translationally Modified Peptides
RNA	Ribonucleic Acid
RREFinder	RiPP Recognition Elements Finder
rRNA	Ribosomal RNA
Sec	General Secretory
SMD	Saltwater Marsh Dieback
SPI	Signal Peptidase I

SPII	Signal Peptidase II
SRA	Sequence Read Archive
T1PKS	Type 1 Polyketide Synthase
TMH	Transmembrane Helices
TAT	Tyrosine Aminotransferase
TMH	Transmembrane Helices
TMHMM	Transmembrane Helices: Hidden Markov Model
tRNA	Transfer RNA
UV	Ultraviolet
VFDB	Virulence Factor Database
v/v	Volume per Volume

CHAPTER I: INTRODUCTION

The first microbial genome sequences were generated 26 years ago when The Institute for Genomic Research (now J. Craig Venter Research Institute) published studies focused on genomes of human opportunistic pathogens *Haemophilus influenza* (Fleischmann et al., 1995) and *Mycoplasma genitalium* (Frazer et al., 1995). In the following years, rapid progress in the development of next-generation sequencing technologies resulted in the exponential growth in the number of prokaryotic genome sequences. For example, the latest release of the Genomes OnLine Database (GOLD) maintained by the Department of Energy Joint Genome Institute (JGI) contains 359,677 complete and draft genome assemblies representing 17,449 type strains of bacteria and archaea (Mukherjee et al., 2020). The rapid accumulation of these data prompted the development of comparative genomics, which employs diverse computational and functional approaches to study the evolution and function of individual genes and gene sets across taxa and entire microbial communities (Koonin et al., 2021).

One of the fundamental concepts of comparative genomics is gene orthology, which postulates that genes evolved from a common ancestor have similar roles in different organisms (Fitch, 1970). Sorting genes into COGs (clusters of orthologous genes) represents a robust approach for functional genome annotation and studying genome evolution (Trachana et al., 2011). A significant fraction of genes in any given bacterial genome is organized into co-transcribed operons, which helps to control and coordinate the expression of functionally related genes (Touchon et al., 2016). The computational prediction and analysis of operons can be used to identify genes involved in common cellular processes and predict functions of conserved hypothetical genes

through the “guilt by association” approach (Aravind, 2000). Finally, pangenome analyses are used to define “core” gene families shared among all members of a species and identify numerous moderately conserved and unique “accessory” genes (Golicz et al., 2020). Recent studies revealed that most free-living bacteria have open pangenomes with a compact conserved core and a large dynamic accessory genome, most of which evolved through horizontal gene transfer (HGT). The analysis of gene presence/absence patterns and copy number variation can reveal genetic sequences that correlate with certain traits or phenotypes.

Various comparative genomics approaches have been successfully used to characterize microbial genes involved in metabolic processes, regulatory mechanisms, pathogenic and symbiotic interactions, and environmental persistence (Delaux et al., 2017; Rouli et al., 2015; Sardi and Gasch, 2017). Global computational analysis of genomic islands from actinomycete genomes revealed thousands of biosynthetic gene clusters that potentially control the synthesis of numerous novel natural products (Cimermancic et al., 2014). This project employed comparative computational genomics to analyze genome sequences of strains *Marinomonas ostreistagni* 398 and *M. spartinae* 468 isolated from the rhizosphere of tidal marsh grasses *Spartina alterniflora* and *Juncus roemerianus* collected on Deer Island, MS (Mavrodi et al., 2018). Both strains exhibited antagonism towards phytopathogenic *Fusarium* fungi, which have been linked to the rapid decline of marsh grasses known as Saltwater Marsh Dieback (Elmer et al., 2013). This study aimed to characterize genomic features of strains 398 and 468 that allow these bacteria to produce antifungal metabolites and resist environmental stresses associated with the colonization of plants growing in brackish tidal marshes.

CHAPTER II: LITERATURE REVIEW

2.1 The Genus *Marinomonas*

Marinomonas are Gram-negative, moderately halophilic, motile, aerobic or facultative anaerobic Gammaproteobacteria. They belong to the *Oceanospirillaceae* family that, in addition to marinomonads, encompasses multiple genera of marine bacteria, such as *Marinobacter*, *Marinobacterium*, *Marinospirillum*, *Microbulbifer*, *Neptunomonas*, *Oceanobacter*, *Oceanospirillum* and *Pseudospirillum* (González and Whitman, 2006). Although the family is diverse and includes several distinct lineages, its monophyletic nature is supported by DNA-rRNA hybridization and 16S rRNA-based analyses. This group is also distantly related to *Shewanella*, *Alteromonas, Vibrio,* and *Photobacterium,* which are large genera that include numerous species ubiquitously distributed in marine habitats.

Early phenotypic characterization attempts classified *Marinomonas* isolates as different species of *Alteromonas* (Baumann et al., 1972). However, subsequent, more detailed morphological, physiological, and DNA-rRNA hybridization analyses revealed that these organisms form a group distinct from *Alteromonas* and members of the *Oceanospirillum* group (Bodwitch et al., 1984; Pot et al., 1989). In the early 1980s, 16S rRNA-based studies clearly differentiated marinomonads from other, closely related Gram-negative marine bacteria and resulted in the establishment of the *Marinomonas* genus (Van Landschoot and De Ley, 1983). Over the following two decades, numerous *Marinomonas* strains were isolated, characterized and assigned to at least 40 distinct species with validly published names (Euzéby, 1997).

Marinomonas spp. are chemoorganotrophs that thrive under elevated salt concentrations and are commonly found in seawater, marine animals, and aquatic plants (González and Whitman, 2006). These organisms can be cultured on defined and complex media. For example, Baumann et al. (1984) isolated *M. communis* and *M. vaga* from ocean water using Tris-buffered artificial seawater amended with NH_4Cl, K_2HPO_4, $FeSO_4$, and *m*-hydroxybenzoate as a carbon source. The marinomonads were also successfully recovered on a complex, defined culture medium containing a mixture of carbohydrates, sugar alcohols, amino acids, organic acids, glycerol, and synthetic seawater (Eilers et al., 2000). The type strain of *M. ostreistagni* was isolated from pearl oyster ponds by processing a sample of seawater through a polycarbonate filter, plating the filtered sample on nutrient agar amended with seawater, and incubating for 2-3 days at room temperature (Lau et al., 2006). In contrast, the type strain of *M. spartinae* was isolated by macerating and centrifuging surface-sterilized roots of *Spartina* plants (Lucena et al., 2016). The resultant apoplastic fluid was then plated on trypticase soy agar amended with sodium chloride and incubated at 28°C for 72 hours. The bacterial strains used in this study (i.e., *M. ostreistagni* 398 and *M. spartinae* 468) were isolated by dilution plating root washes from *S. alterniflora* and *J. roemerianus* on one-tenth strength Tryptic Soy Agar supplemented with 15% artificial saltwater (Mavrodi et al., 2018). The recovered isolates were purified by several passages on the same medium and stored at -80°C in Tryptic Soy broth-containing saltwater and 15% (v/v) glycerol.

2.2 Practical Importance of Marine Bacteria

Marinomonads, along with other members of the *Oceanospirillaceae*, play an important role in marine ecosystems by contributing to the cleanup of areas contaminated

with crude oil and other petroleum products. Many species of this group secrete bioemulsifiers and biosurfactants that reduce surface tension and facilitate the dispersion, emulsification, and biodegradation of hydrocarbons (Dang et al., 2016). There is a growing interest in the commercial use of such microbial biosurfactants because of their biodegradability and lower toxicity compared to the synthetic counterparts (Tripathi et al., 2018). *Marinomonas* spp. and other members of the *Oceanospirillaceae* were also identified as active degraders of polycyclic aromatic hydrocarbons (PAHs), which include many toxic, mutagenic, and persistent xenobiotics polluting marine environments (Dong et al., 2015).

In addition to thriving in seawater and sediments, many species of *Marinomonas* readily colonize marine macroorganisms. For example, *M. posidonica* and *M. mediterranea* constitute a significant proportion of the microbiome of the perennial Neptune grass (*Posidonia oceanica*) that grows in the Mediterranean Sea (Lucas-Elio et al., 2012). The genome analysis revealed that these microorganisms contribute to the global sulfur and carbon cycles by releasing dimethyl sulfide that affects the cloud formation and climate over the oceans. *Marinomonas spartinae* is a plant-beneficial microorganism associated with small (*Spartina maritima*) and smooth (*S. alterniflora*) cordgrass growing in tidal marshes (Lucena et al., 2016; Mavrodi et al., 2018). Other species of marinomonads are found in microbiomes of marine animals and frequently isolated from various corals, shellfish, and finfish (Kumari et al., 2014; Vignier et al., 2021; Villasante et al., 2020; Califano et al., 2017). Finally, *Marinomonas* spp. were studied for their physiological mechanisms of osmoadaptation that allow these moderately halophilic microorganisms to grow over a wide range of saline concentrations

(Vaidya et al., 2018). Several *Marinomonas* spp. have also been investigated for their ability to produce various enzymes and metabolites with potential clinical, environmental, and biotechnological applications. For example, the psychrophilic bacterium *M. primoryensis* produces unusual antifreeze proteins that modulate the ice nucleation rate and have potential applications in the areas of cryopreservation, materials science, biocatalysis, and climate control (Voets, 2017).

2.3 Goals of this Study

The genus *Marinomonas* encompasses a ubiquitous group of bacteria that thrive in diverse marine habitats and are investigated because of their environmental significance and biotechnological potential. Despite the progress in understanding the diversity of these organisms, our ability to explain ecological, metabolic, and biochemical traits of marinomonads at the genomic sequence level remains limited. The goal of this study was to address this gap by annotating and analyzing genome sequences of *M. ostreistagni* 398 and *M. spartinae* 468, which were isolated from the rhizosphere of *Spartina* collected in tidal marshes on Deer Island, Mississippi. This project aimed to provide new insights into the genomic diversity of the genus *Marinomonas* and identify genes that allow these bacteria to adapt to hypersaline environments, as well as pathways involved in the interaction with plants and production of antifungal compounds.

CHAPTER III: MATERIALS AND METHODS

3.1 Genome Sequencing, Assembly, and Annotation.

Bacterial strains *Marinomonas ostreistagni* 398 and *M. spartinae* 468 were cultured for 48 h at 27°C in Tryptic Soy broth (Difco) supplemented with 15% artificial saltwater SW30 (Dyall-Smith 2009), and high molecular weight samples of their DNA were prepared using the cetyltrimethylammonium bromide method (Ausubel et al., 2002). The DNA samples were shipped to the Molecular and Genomics Core facility of the University of Mississippi Medical Center where they were sequenced using a MiSeq instrument (Illumina, San Diego, CA). The resulting raw data in the form of paired reads were processed with Trim Galore v.0.6.1 (Krueger 2015) and assessed for quality with the FastQC toolkit (Andrews 2010). Once filtered, the quality reads were assembled in the Pathosystems Resource Integration Center (PATRIC) (Wattam et al., 2014) with Unicycler (Wick et al., 2017) and the resultant assemblies were annotated with the RASTtk (Rapid Annotation using Subsystem Technology) annotation engine (Brettin et al., 2015). During the annotation, the domain of bacteria was selected, with the taxonomy name of *Marinomonas*, and the genetic code was set to 11 (Archaea & most Bacteria). The genomes of strains 398 and 468 were deposited in GenBank under accession numbers JAEMUH000000000 and JAEMUI000000000, respectively. The raw reads were deposited in the Sequence Read Archive under BioProject PRJNA686781.

3.2 Phylogenetic Analysis

The taxonomic placements of strains *M. ostreistagni* 398 and *M. spartinae* 468 were determined by inferring a whole-genome maximum likelihood phylogeny using the Codon Tree method implemented in PATRIC (Wattam et al., 2014). The analysis

involved 39 draft and complete genomes of the *Marinomonas* group. The complete genome of *Neptunomonas phycophila* Scap09 was used as an outgroup. The parameters were set for 100 conserved shared genes with 0 maximum allowed deletions and duplications. The resultant bootstrapped phylogeny was exported in the Newick tree format and edited in Geneious Prime 2021 (Biomatters, Auckland, New Zealand).

3.3 Identification of Secreted and Membrane Proteins

The predicted proteomes of *M. ostreistagni* 398 and *M. spartinae* 468 were analyzed for the presence of signal peptides associated with Sec/SPI, Sec/SPII, and Tat/SPI secretion systems using the SignalP-5.0 server (Armenteros et al., 2019). The protein sequences were uploaded to SignalP-5.0 in the FASTA format and the Gram-negative organism group and long output format were selected before submitting the protein sequences for analysis. Genes associated with type III, type IV, and type VI protein secretion systems were identified with EffectiveDB (Eichinger et al., 2016). The proteomes were also screened for the presence of transmembrane helices (TMHs) using the TMHMM v. 2.0 server (Krogh et al., 2001). The analysis included the prediction of TMHs and the probability that the N-terminus of the protein is located on the cytoplasmic side of the cell membrane.

3.4 Secondary Metabolite Cluster Prediction

Secondary metabolite clusters were identified using the Antibiotics & Secondary Metabolite Analysis Shell (AntiSMASH) 5.0 genome mining pipeline (Blin et al., 2019). For this analysis, the bacterial version of AntiSMASH was used along with relaxed detection strictness and the extra features of KnownClusterBlast, SubClusterBlast, ActiveSiteFinder, and RREFinder being selected as well. Additional analyses were

conducted with the ARTS 2.0 pipeline (Mungan et al., 2020), where the phylogeny screen and resistance model screen options were selected to predict core genes, putative biosynthetic gene clusters, and antibiotic resistance factors.

3.5 Identification of Genomic Islands and Prophages

Prophage sequences were identified with PHASTER, which is an upgraded version of the PHAge Search Tool (PHAST) pipeline (Arndt et al., 2016). Each identified prophage element was then further analyzed for GC content, presence of an integrase gene, position within the genome, and the total number of ORFs encoding phage-like proteins and their predicted functions. Genomic islands within the genomes of *M. ostreistagni* 398 and *M. spartinae* 468 were detected with the IslandViewer4 server (Dhillon et al., 2015). This pipeline identifies horizontally transferred genome segments based on the codon usage-based prediction (the SIGI-HMM method) (Waack et al., 2006) or relying on the sequence composition biases and HMM profiles for mobility genes (the IslandPath-DIMOB approach) (Bertelli and Brinkman 2018). Both draft genomes were analyzed using a complete genome of *M. mediterranea* genome as a reference. Putative virulence factors were identified by comparison against curated databases, such as the Virulence Factor Database (VFDB) (Chen et al., 2012) and Victors (Sayers et al., 2019).

3.6 Pangenome Analyses and Orthologous Cluster Identification

The pangenome analysis of *M. ostreistagni* 398, *M. spartinae* 468, and selected reference strains was conducted with OrthoVenn2 (Xu et al., 2019). The parameters for mapping orthologous gene clusters included the E-value of 1e-15, the inflation value of 2, and the enabled annotation and protein similarity network functions. The partitioning of genes into core, non-core, and singleton parts of each strain's proteome were also done

by performing BLASTp (Altschul et al., 1990) comparisons with an E-value cutoff of 1e-06, identity of 40% and coverage of 60%. Finally, the functional annotation of the *M. ostreistagni* 398 and *M. spartinae* 468 genes was carried by KEGG Orthology (KO)-based mapping with BlastKOALA (Kanehisa et al., 2016). Circular genome maps were created with CGview (Stothard and Wishart, 2005).

CHAPTER IV: RESULTS

4.1 Genome Properties of *M. ostreistagni* 398 and *M. spartinae* 468

The sequencing of the *M. ostreistagni* 398 genome produced 1,902,445 × 2 high quality reads totaling 277,199,648 bp of data. The Unicycler assembly of these reads had 35 contigs with a N_{50} of 196,838 bp, N_{max} of 339,475 bp, and median coverage of 139× (Table 1). The genome of strain 398 is comprised of a 3,976,153-bp long chromosome with a 44.8% G + C content. The coding regions accounted for 91.6% of the genome and contained 64 RNA- and 3,730 protein-coding genes. Of these, 2,768 genes were assigned a function, with the remaining annotated as hypothetical or conserved hypothetical. A total of 1,030 genes encoded proteins with Enzyme Commission (EC) numbers.

For *M. spartinae* 468, the sequence run produced 1,313,234 × 2 high quality reads totaling 193,354,801 bp of data (Table 1). The final assembly had 58 contigs with a N_{50} of 158,408 bp, N_{max} of 406,278 bp, and median coverage of 80×. The draft genome of *M. spartinae* 468 was 4,842,952-bp long with a 44.4% G + C content. The coding regions accounted for 91.6% of the genome and contained 62 RNA- and 4,600 protein-coding genes. A total of 3,347 genes were assigned a function, and 1,220 genes encoded proteins with EC numbers. Also, 1,253 genes were annotated as conserved hypothetical or hypothetical. Finally, the sequenced genomes of *M. ostreistagni* 398 and *M. spartinae* 468 contained CRISPR arrays with 40 and 33 repeats, respectively.

Table 1. Genome features of *M. ostreistagni* 398 and *M. spartinae* 468.

Strain	Fold coverage (x)	Genome size (bp)	No. of contigs	N_{50} values (bp)	G+C (%)	Total no. of genes	No. of RNA genes	Total no. of CDSs[a]	No. of PECs[b]	No. of hypothetical proteins
398	139.27	3,976,153	35	196,838	44.84	3,878	64	3,730	1,030	962
468	79.63	4,842,952	58	158,408	44.39	4,728	62	4,600	1,220	1,253

[a] CDSs, coding sequences with proteins. [b] PECs, proteins with Enzyme Commission numbers.

A functional analysis of genes encoded by *M. ostreistagni* 398 and *M. spartinae* 468 was performed using BlastKOALA, an annotation engine that uses BLAST searches to assign query sequences to KEGG Orthology (KO) groups (Kanehisa et al., 2016). The KO assignment information is then used to describe gene functions, reconstruct metabolic pathways and their hierarchies, and infer high-level functions of the organism. The KO identifiers were successfully assigned to a total of 2,334 *M. ostreistagni* 398 proteins, 150 of which showed unclear function (Figure 1). In *M. spartinae* 468, a total of 2,661 proteins were assigned KO identifiers, with 155 showing unclear function. The predicted functions of these proteins could be organized in 38 different functional categories associated with metabolism, processing of genetic information, and signaling and cellular processes. A total of 70 complete pathway modules were identified and associated with metabolism of carbohydrates, energy, lipids, nucleotides, amino acids, glycan, cofactors and vitamins, xenobiotic biodegradation, and drug resistance.

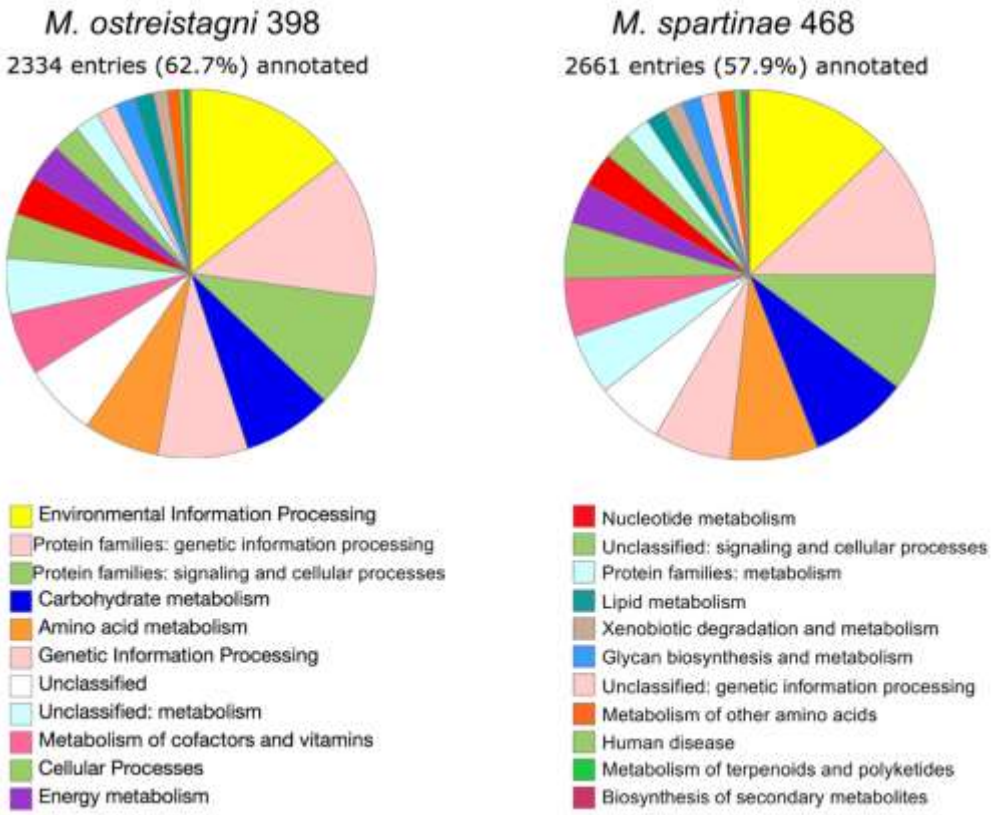

Figure 1. The KEGG Orthology classification of proteins encoded by the M. ostreistagni 398 and M. spartinae 468 genomes.

The analysis was conducted using KEGG's BlastKOALA server (Kanehisa et al., 2016).

The predicted proteomes of strains 398 and 468 were also screened for potential secreted and transmembrane proteins (Table 2). The SignalP-5.0 analysis revealed in each strain the presence of 400 proteins with signal peptides recognized by the general secretion (Sec) SecYEG pathway. In Gram-negative bacteria, the Sec machinery delivers unfolded proteins across or into the inner membrane (Dalbey et al 2012; Auclair et al., 2012). Approximately three quarters of the putative secreted proteins had signal peptides recognized by the membrane-bound signal peptidase I (SPI). After being transferred across the inner membrane, such proteins remain in the periplasm or are being

translocated through the outer membrane by another transport system (Paetzel et al., 2002).

Table 2. Secreted proteins of *M. ostreistagni* 398 and *M. spartinae* 468.

Strain	Total No. of Predicted Protein Sequences	No. of Signal Peptides (Sec/SPI)	No. of Lipoprotein Signal Peptides (Sec/SPII)	No. of TAT Signal Peptides (TAT/SPI)	No. of Trans-membrane Helices (TMHs)	No. of Unspecified Protein Sequences
398	3,730	329	107	14	908	3,280
468	4,600	316	120	16	1,080	4,148

Approximately a quarter of putative secreted proteins in each strain carried a type II signal peptide, which is first modified via attachment of a diacylglycerol membrane anchor (Okusa and Tokuda, 2011). The resultant lipoprotein is then cleaved by the signal peptidase II (SPII). In Gram-negative bacteria, most proteins processed by the SPII peptidase remain in the periplasm and stay anchored to the inner or outer membrane (Okusa and Tokuda, 2011). Also present were proteins carrying TAT/SPI signal peptides, which are recognized by the "twin arginine" TATABC translocon. Unlike the general secretion system, the Tat translocase moves across the inner membrane fully folded proteins, such as some toxins, iron–sulfur proteins, cytochromes, or other respiratory enzymes (Frain et al., 2019). The EffectiveDB searches revealed that although *M. ostreistagni* 398 and *M. spartinae* 468 lack type III and type IV protein secretion pathways, both strains carry well-conserved gene clusters that encode type VI protein secretion systems. Finally, *M. ostreistagni* 398 and *M. spartinae* 468 encoded, respectively, 908 and 1080 proteins with transmembrane helices. These structural

features act as a hydrophobic anchor that mediates the assembly and folding of membrane proteins (Walther, Ulrich 2014; Wong et al 2012).

4.2 Other Notable Genome Features

The antiSMASH and ARTS analyses revealed that the genomes of strain 398 and strain 468 contain multiple pathways for the synthesis of secondary metabolites. Specifically, in *M. ostreistagni* 398 the analysis identified putative genes encoding enzymes involved in the production of a beta-lactone, an arylpolyene, ectoine, and a bacteriocin, respectively (Table 3). Also identified was a cyclodipeptide synthase (CDPS) that belongs to a group of enzymes utilizing aminoacyl-tRNAs as substrates for the synthesis of antimicrobial cyclodipeptides (Canu et al., 2020).

Table 3. Summary of secondary metabolite gene clusters identified in the genomes of *M. ostreistagni* 398 and *M. spartinae* 468 by antiSMASH.

Cluster no.	Position (locus tag)	Cluster category	Most similar known cluster	Similar genes (%)	Putative biological activity of the product
M. ostreistagni 398					
1	133920-166585	Betalactone	None	0	Antibacterial
2	133531-177090	Arylpolyene	Aryl polyenes	38	Protection from oxidative stress
3	106080-116469	Ectoine	Ectoine	83	Osmotic stress tolerance
4	180133-191029	Bacteriocin	None	0	Antibacterial
5	36140-56910	CDPS	None	0	Antimicrobial

Table 3, continued

M. spartinae 468					
1	214378-224767	Ectoine	Ectoine	83	Osmotic stress tolerance
2	167252-211214	NRPS	None	0	Antimicrobial
3	3808-14707	Bacteriocin	None	0	Antibacterial
4	258117-273097	Siderophore	Vibrioferrin	63	Iron transport
5	122084-160927	NRPS/T1PKS	None	0	Antifungal
6	142811-153779	Bacteriocin	None	0	Antibacterial
7	63753-119951	NRPS	Cupriachelin siderophore	17	Iron transport

In addition to ectoine and bacteriocin clusters, *M. spartinae* 468 carried the genes for the biosynthesis of a siderophore and various polyketide metabolites, which involve enzymes with predicted non-ribosomal peptide synthetase (NRPS) and hybrid NRPS-type I polyketide synthase (NRPS/T1PKS) activities.

The genomes of *M. ostreistagni* 398 and *M. spartinae* 468 were also screened for the presence of prophages and genomic islands. The analysis of *M. ostreistagni* 398 with PHASTER (Arndt et al., 2016) identified four complete prophage regions showing partial similarity to viruses circulating in different Gram-positive and Gram-negative bacteria. These prophages included genes encoding structural components of phage capsids and tails, as well as viral proteases and integrases (Table 4). Compared to strain 398, the genome of *M. spartinae* 468 contained only three prophage regions that carried phage tail, plate, and protease genes.

Table 4. Summary of prophage regions predicted in the genomes of *M. ostreistagni* 398 and *M. spartinae* 468 by PHASTER.

Prophage no.	Position (locus tags)	Most similar known phage	Similar genes (%)	Notable features
M. ostreistagni 398				
1	104745-112214	*Synechococcus* phage ACG-2014f	25.0	Phage plate genes
2	202798-212079	*Bacillus* phage vB-BanS-Tsamsa	12.5	Phage plate genes, protease
3	50018-86305	*Enterobacter* phage Arya	26.5	Phage tail, capsid, and head genes, lysis genes and integrase
4	75563-97155	*Enterobacter* phage BP-4795	18.2	Phage head genes, integrase
M. spartinae 468				
1	60509-70174	*Bacillus* phage vB_BanS-Tsamsa	11.1	Phage plate genes, protease
2	87310-101565	*Escherichia* phage Mu	37.5	Phage tail genes
3	274938-283654	*Pseudomonas* phage OBP NC_01657	20	None

The IslandViewer4 (Dhillon et al., 2015) screen of strains 398 and 468 identified multiple genomic islands (GIs). In strain 398, the predicted GIs carried genes for the biosynthesis of thiamine, uptake and catabolism of urea, resistance to fosfomycin, DNA methylation, repair of UV radiation-induced DNA damage, assembly of fimbriae, diverse transport and regulation activities, and components of prophages and transposons (Table A.1). In *M. spartinae* 468, the GIs encoded assorted bacteriophage genes, DNA metabolism enzymes (including a lesion bypass DNA polymerase), heat shock chaperones, adhesins, and capsular synthesis enzymes (Table A.2). In both strains, the GI

regions also contained numerous genes encoding hypothetical and conserved hypothetical proteins.

4.3 Pan-Genome Analysis

To identify the relationship of *M. ostreistagni* 398 and *M. spartinae* 468 to other members of the genus, a whole-genome phylogenetic analysis that included a total of 41 strains representing at least 22 named *Marinomonas* species was performed. The maximum-likelihood phylogenic tree of *Marinomonas* showed three distinct and well-supported clades (Figure 2). The strains *M. ostreistagni* 468 and *M. spartinae* 398 used in this study belonged to clades I and II, respectively. Based on these results, we selected *M. posidonica* IVIA-Po-181, *M. primoryensis* AceL, *M. mediterranea* MMB-1, and *Marinomonas* sp. SBI8L to be used as typical representatives of the three clades in subsequent comparative genome analyses.

The OrthoVenn2 comparison of *M. ostreistagni* 398 and *M. spartinae* 468 with the selected reference strains revealed that the number of orthologous gene families per strain varied between 3,037 and 3,328 (Figure 3). The core genome was represented by 1,805 gene families, which constituted between 54.2% and 59.4% of the predicted proteome. The number of strain-specific gene families varied between 4 (IVIA-Po-181) and 95 (AceL). The pairs of strains that shared the highest proportion of gene families were *Marinomonas* sp. SBI8L and *M. mediterranea* MMB-1, *M. posidonica* IVIA-Po-181 and *M. mediterranea* MMB-1, and *M. spartinae* 468 and *M. primoryensis* AceL (Figure 3).

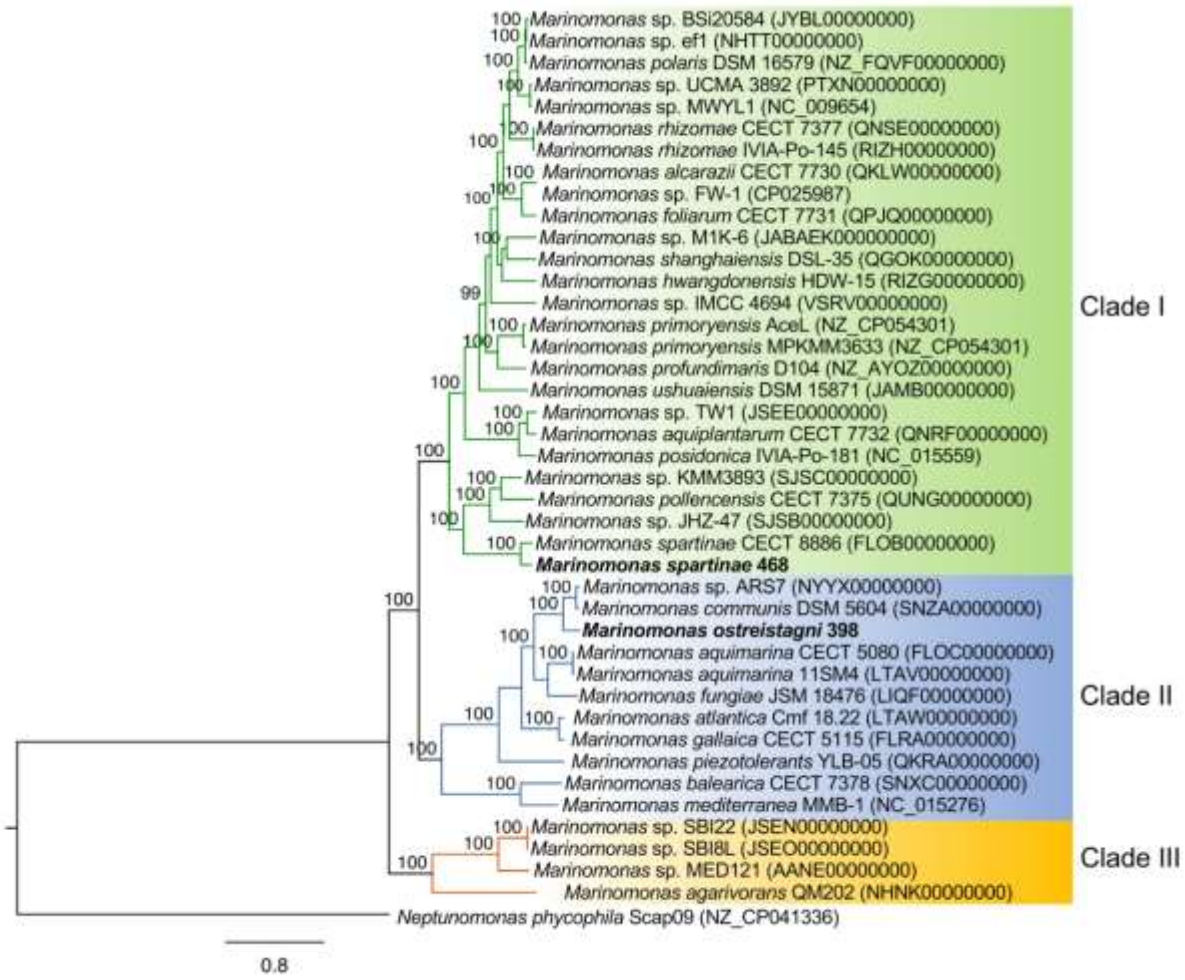

Figure 2. Whole-genome maximum likelihood phylogeny of 41 different strains showing three distinct Marinomonas clades.

The phylogeny was generated using the CodonTree method in PATRIC (Wattam et al., 2014) based on 100 shared single-copy genes identified by PGFams. Numbers above tree nodes are bootstrap values. *Neptunomonas phycophila* Scap09 was included in the analysis as an outgroup. The scale bar shows the number of nucleotide substitutions per site.

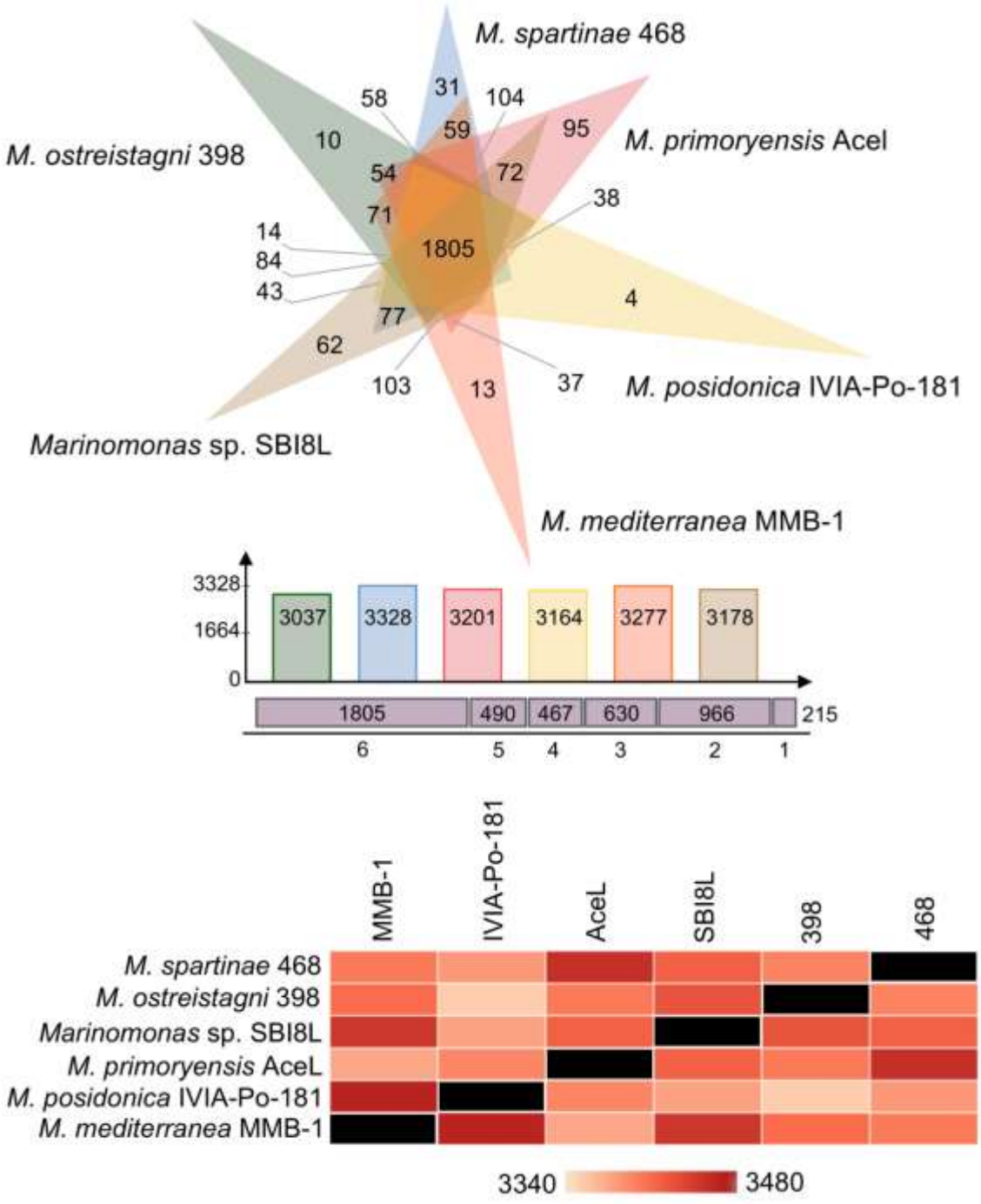

Figure 3. The distribution of orthologous gene families among M. ostreistagni 398, M. spartinae 468, and selected reference Marinomonas strains (top panel).

The Venn diagram and bar chart show the number of shared and unique orthologous gene families encoded by each analyzed genome. The number of the shared orthologs is also shown in the form of the pairwise heatmap (bottom panel). The calculations were conducted with OrthoVenn2 using an E-value of e-15 and MCL inflation factor of 2.0.

An in-depth analysis of *M. ostreistagni* 398 by multiway BLASTp revealed that its 3,730 predicted protein-coding genes were split unevenly between the core (2,315 genes or 62%), non-core (617 genes or 16.5%), and strain-specific (798 genes or 21.4%) parts of the genome. Similarly, the majority of 4,600 protein-coding genes of *M. spartinae* 468 were assigned to the core (2,315 genes or 50.3%), while the non-core and strain-specific parts of the genome were comprised, respectively, of 966 (21.0%) and 1,319 (28.7%) genes. The visual representations of the genomes of *M. ostreistagni* 398 and *M. spartinae* 468 show the genes that appear in both strains as well as those that appear in the reference genomes. We also identified lineage-specific genes that are shared by strains of clade I (*M. spartinae* 468, *M. posidonica* IVIA-Po-181, *M. primoryensis* AceL) or clade II (*M. ostreistagni* 398, *M. mediterranea* MMB-1). In *M. spartinae* 468, a total of 69 genes were identified as clade I-specific. A total of 79 *M. ostreistagni* 398 genes were identified as specific to clade II. The variable gene content was visualized on circular genome maps built with the help of CGview (Stothard and Wishart, 2005). The position of clade-strain and strain-specific genes with *M. ostreistagni* 398 and *M. spartinae* 468 genomes is visible as gaps in BLASTn hits against reference *Marinomonas* genomes (Figure 4).

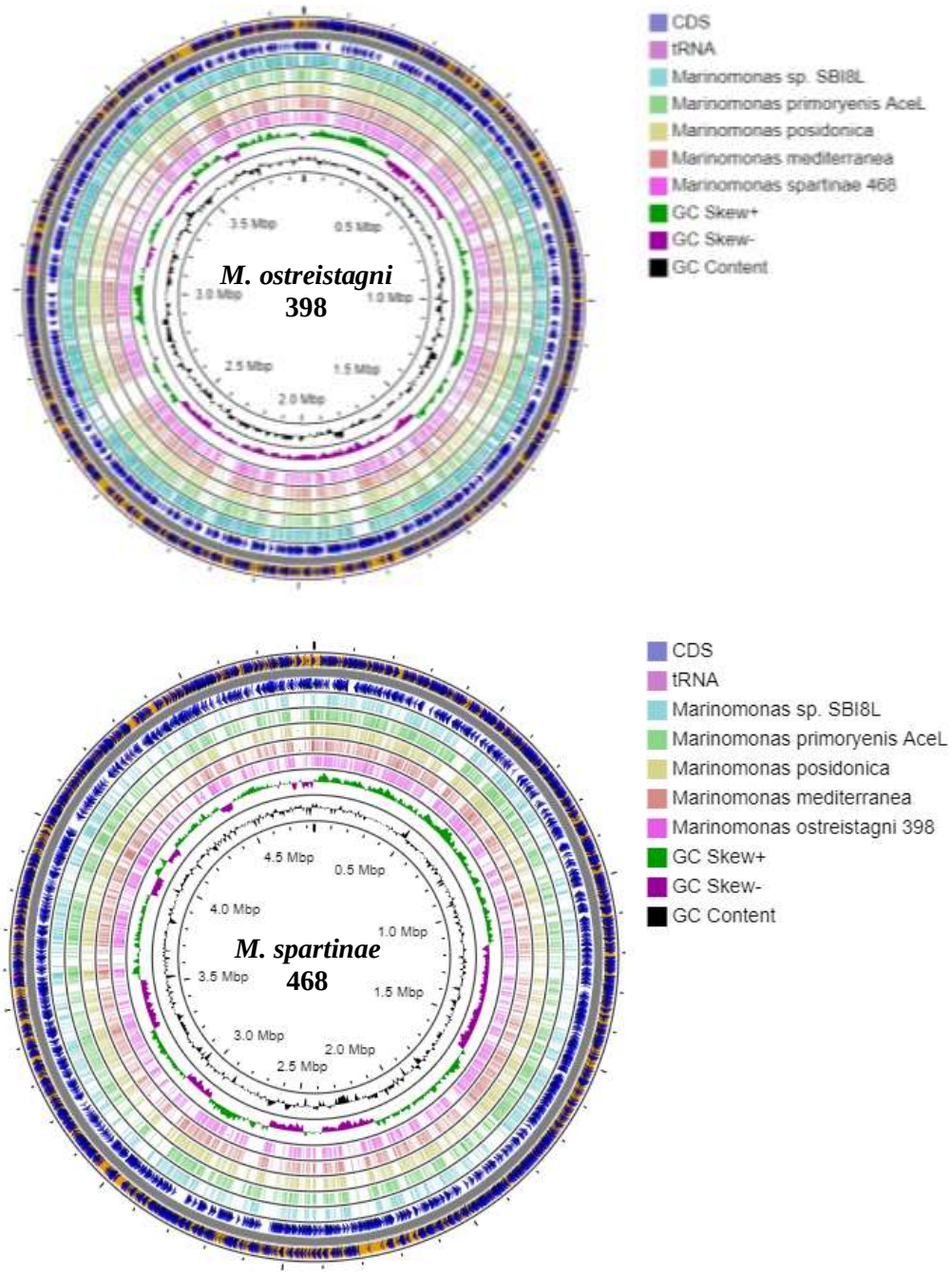

Figure 4. Circular representation of the M. ostreistagni 398 and M. spartinae 468 genomes and their comparison to representative genomes of the Marinomonas group.

Rings from inside to outside: (1) GC content, (2) GC skew, (3-7) BLASTn comparison with other *Marinomonas* genomes, (8, 9) CDSs predicted on the reverse and forward DNA strands. The inner scales designate coordinates in kilobase pairs. The analysis was performed with CGview (Stothard and Wishart, 2005).

CHAPTER V: DISCUSSION

This study provided new information about the biology of marinomonads based on a comparative genome analysis of bacterial strains *M. ostreistagni* 398 and *M. spartinae* 468, which antagonize the plant pathogenic fungus *Fusarium*. The genome of strain *M. ostreistagni* 398 is the first sequenced genome of this species. The phylogenetic analysis confirmed the taxonomic identity of both strains and placed them into two distinct clades within the *Marinomonas* genus. Although phenotypically similar, *M. ostreistagni* 398 and *M. spartinae* 468 differ substantially at the genetic level. The strain 468 possesses a much larger genome than strain 398 with a difference of about 0.87 megabase pairs. The two genomes also differ in the number of encoded proteins, as the genome of strain 398 contained just 3,730 CDSs compared to 4,600 CDSs in strain 468. In *M. ostreistagni* 398, the non-core and strain-specific genes collectively accounted for 38% of the gene content. The share of non-core and strain-specific genes was even higher in *M. spartinae* 468 and accounted for almost half of the gene content. The comparison of *M. ostreistagni* 398, *M. spartinae* 468 and selected reference genomes revealed that members of different clades share a core set of protein-coding genes performing housekeeping functions. This core constituted only 50.3 to 64% of individual strains' proteomes, further highlighting the importance of variable gene content in the biology of *Marinomonas*.

A substantial amount of the variable gene content in *M. ostreistagni* 398 and *M. spartinae* 468 was associated with genomic islands (GIs) and prophage regions. The GIs are clusters of genes in the genomes of bacteria that are unique to subsets of genomes and are of probable horizontal transfer origin, which is an evolutionary mechanism for

bacterial genomes that allows for the bacteria to quickly develop specific adaptations for survival (Edwards, Holt 2013; Bertelli et al. 2019). Examples of such adaptations include the ability to catabolize new carbon sources, resistance to heavy metals and antibiotics, synthesis of virulence factors, and symbiotic and stress response traits (Bertelli et al. 2019). Prophages play an important role in bacterial evolution by acting as vectors for the HGT-mediated acquisition of virulence factors in bacterial pathogens and of specialized ecological adaptations in nonpathogenic bacteria (Canchaya et al. 2003). In addition, prophages protect their bacterial host from the attack and destruction by other, closely related bacteriophages (Canchaya et al. 2003; Bertelli et al. 2019). The analysis of GIs, prophages, and other strain-specific genome segments in *M. ostreistagni* 398 and *M. spartinae* 468 revealed genes encoding numerous hypothetical proteins, membrane transporters, surface appendages, regulatory proteins, and various enzymes participating in catabolism, as well as DNA and energy metabolism. It is likely that these genes collectively contribute to the ability of these organisms to colonize new niches, compete with other microorganisms, and adapt to changing environmental conditions.

The functional analysis of proteins encoded by the genes in both *Marinomonas ostreistagni* 398 and *M. spartinae* 468 has revealed several predicted functions and applications of these proteins as defined by their KEGG orthology classifications. One of these functions is xenobiotic degradation and metabolism, which refers to the ability of an organism to degrade and/or metabolize non-native compounds in an environment. In general, halophilic bacteria have been known to be capable of degrading certain common xenobiotics that pollute the environment such as hydrocarbons, formaldehydes, and nitroaromatic compounds (Orellana et al, 2018). The degradation of xenobiotics such as

these by certain bacteria can remove harmful compounds from the environment and therefore can have significant applications in bioremediation. Specifically, within the *Marinomonas* genus, species such as *M. communis* and *M. mediterranea* have been reported to have xenobiotic degradation and bioremediation capabilities. The uptake and subsequent removal of arsenic from its environment has been reported in the *M. comminus* species while *M. mediterranea* has been shown to produce laccases, which have been reported to be capable of degrading a wide range of xenobiotics such as textile waste, antibiotics, bisphenol A (BPA), and polycyclic aromatic hydrocarbons (PAHs) (Takeuchi et al, 2007; Arregui et al 2019). Both of these species are closely related to *M. ostreistagni* 398, in that they exist within the same clade and have been shown to share some common non-core genes, so the ability of *M. ostreistagni* 398 to produce compounds and have bioremediation potential similar to these species is likely.

The genome analysis also revealed that *M. ostreistagni* 398 and *M. spartinae* 468 are capable of producing diverse secondary metabolites, which may help these bacteria to flourish in the rhizosphere of marsh grasses. Among these secondary metabolites is ectoine, an effective osmolyte that allows halophilic bacteria to balance the osmotic pressure across cellular membranes without compromising the structure of biomolecules (Yancey 2005). Both strains also secrete siderophores that help *Marinomonas* to effectively acquire iron, as well as bacteriocins that function to inhibit the growth of other bacteria, generally those that are closely related to the bacteria secreting the metabolite, and serve to reduce the number of bacteria competing for the same resources (Sharrar et al 2020). Yet another group of secondary metabolites includes aryl polyenes, some of which are known to function as oxidative stress protectants (Aleti et al 2018; Richter et al

2019). Additionally, *M. ostreistagni* 398 and *M. spartinae* 468 also carries non-ribosomal peptide synthetases and polyketide synthase, which are likely involved in the synthesis of antifungal metabolites responsible for the ability of these strains to inhibit the growth of the phytopathogenic *Fusarium*. The antifungal metabolites produced by *M. ostreistagni* 398 and *M. spartinae* 468 may also have clinical and biotechnological applications in the treatment of human diseases, such as mycotoxicosis and fusariosis, and the prevention of plant diseases caused by fungi such as those within the *Fusarium* genus (Batista et al, 2020).

Our further analysis of *Marinomonas* genomes indicated that, in addition to ectoine, siderophores, and NRPS- and PKS-derived compounds, these bacteria carry genes for the synthesis of ribosomally synthesized and post-translationally modified peptides (RiPPS), lassopeptides, beta-lactones, cyclodipeptides, thioamides, linaridins, and diverse saccharides (Figure 5). Collectively, these findings highlight *Marinomonas* spp. as a rich source of natural bioactive metabolites with potential antibacterial, antifungal, immunosuppressant, and analgesics properties.

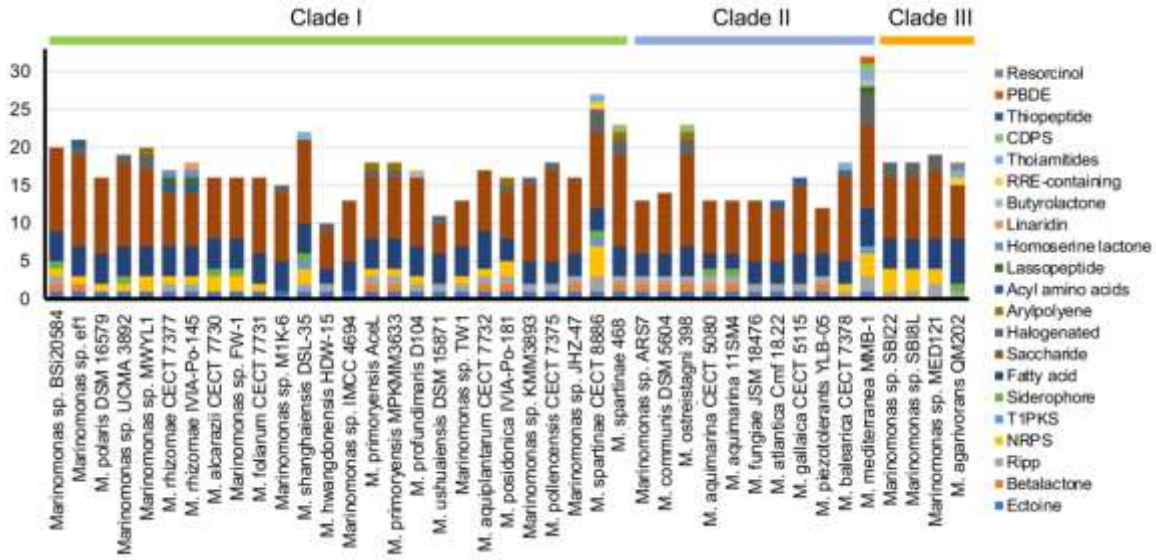

Figure 5. *Distribution and classification of antiSMASH hits for representative Marinomonas genomes.*

APPENDIX A: GENOMIC ISLANDS

Genomic island no.	Position (locus tag)	Prediction Method	# of hypothetical Proteins	Products other than hypothetical proteins
1	45203-49572	SIGI-HMM	1	Transcriptional regulator (AraC family), Pulcherriminic acid synthase, Uncharacterized MFS-type transporter
2	1040390-1045696	SIGI-HMM	0	Transcriptional repressor of ectoine biosynthetic genes, Deoxyribodipyrimidine photolyase, Transcriptional regulator (MerR family) associated with photolyase, Renalase, 1,2-dihydrooxidase, 1,6-dihydrooxidase, Uncharacterized membrane protein SO_4740, COG1683: Uncharacterized conserved protein / FIG143828: Hypothetical protein YbgA
3	1353454-1360136	SIGI-HMM	6	None
4	1580193-1585495	SIGI-HMM	2	Outer-membrane-phospholipid-binding lipoprotein MlaA, PQQ-dependent oxidoreductase, gdhB family
5	1877757-1884301	SIGI-HMM	3	Thiazole synthase, 2-iminoacetate synthase (ThiH)
6	2041945-2049397	SIGI-HMM	4	Chemotaxis regulator, DNA gyrase inhibitor YacG, Dephospho-CoA kinase, Type IV fimbrial assembly, ATPase PilB, Type IV pilin PilA, Two-component system sensor histidine kinase
7	2557169-2566388	SIGI-HMM	5	N-Acetylneuraminate cytidylyltransferase, Lipopolysaccharide biosynthesis protein, 2-keto-3-deoxy-D-arabino-heptulosonate-7- phosphate synthase I beta
8	2599382-2605403	SIGI-HMM	4	Iron-sulfur cluster regulator IscR, Transcriptional regulator (LysR family)
9	3987818-3992648	SIGI-HMM	4	Cysteine desulfurase (SufS), Sulfur acceptor protein (sulfur transfer pathway protein CsdE)
10	3995742-4010153	SIGI-HMM	4	Gfa-like protein, Fosfomycin resistance protein FosA, TPR domain protein (putative component of TonB system)

11	452502-485332	IslandPath-DIMOB	18	Putative glycosyltransferase, Transposase and inactivated derivatives, Teichoic acid export ATP-binding protein TagH
12	1069580-1082161	IslandPath-DIMOB	19	DNA translocase FtsK, C-5 cytosine-specific DNA methylase family protein, Phage DNA replication protein O
13	1091509-1108788	IslandPath-DIMOB	19	Phage head, portal protein B, Phage head head-tail preconnector protease C and scaffolding domain Nu3, Phage baseplate assembly protein GpJ and chaperone, Phage tail sheath protein FI, Phage-related protein
14	1346389-1361356	IslandPath-DIMOB	12	Transposase and inactivated derivatives
15	1667070-1675152	IslandPath-DIMOB	1	2-dehydropantoate 2-reductase, UPF0234 protein Yitk, methylated-DNA--protein-cysteine methyltransferase-related protein, Protein affecting phage T7 exclusion by the F plasmid, Heat shock protein 10 kDa family chaperone GroES, Heat shock protein 60 kDa family chaperone GroEL, Probable type IV pilus assembly FimV-related transmembrane protein
16	3734050-3762740	IslandPath-DIMOB	9	Urease accessory proteins (UreG, UreF, UreE, UreD), Urease subunits (alpha, beta, gamma), Urea ABC transporter ATPase proteins (UrtE, UrtD, UrtC, UrtB, UrtA), Permease of the drug/metabolite transporter (DMT) superfamily, DNA-binding transcriptional regulator (MocR family) / aminotransferase domain, transposase, Sugar fermentation stimulation protein SfsA, L-threonine 3-dehydrogenase, 2-amino-3-ketobutyrate coenzyme A ligase, Transcriptional regulator (LysR family), Putative 5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase oxidoreductase protein, ABC transporter permease proteins 1,2 (cluster 4), ABC transporter ATP-binding proteins 1,2 (cluster 4)

17	3986608-3999010	IslandPath-DIMOB	6	Cysteine desulfurase (SufS), Sulfur acceptor protein (sulfur transfer pathway protein CsdE), Phage integrase, Mobile element protein, Gfa-like protein

Table A.1 Summary of genomic islands predicted by IslandViewer 4 in the genome of *M. ostreistagni* 398.

Genomic island no.	Position (locus tags)	Prediction method	# of hypothetical proteins	Products other than hypothetical proteins
1	389163-395383	SIGI-HMM	4	P pilus assembly protein (porin PapC), Transcriptional regulator VpsT
2	1509150-1514663	SIGI-HMM	4	DNA-cytosine methyltransferase
3	1936866-1942963	SIGI-HMM	6	Transposase and inactivated derivatives
.4	2147186-2151697	SIGI-HMM	8	None
5	2780678-2786316	SIGI-HMM	3	Serine/threonine kinase, GTP cyclohydrolase II
6	2905677-2909931	SIGI-HMM	6	None
7	3427952-3432052	SIGI-HMM	2	Capsular polysaccharide synthesis enzyme CpsA, sugar transferase, Glycosyltransferase
8	4545402-4550376	SIGI-HMM	4	None
9	1576557-1583573	IslandPath-DIMOB	1	Phosphoribosylanthranilate isomerase, tRNA pseudouridine (38-40) synthase, Probable type IV pilus assembly FimV-related transmembrane protein, Heat shock protein 60 kDa family chaperone GroEL, Heat shock protein 10 kDa family chaperone GroES, Protein affecting phage T7 exclusion by the F plasmid, DNA base-flipping protein, UPF0234 protein Yitk
10	1839394-1846790	IslandPath-DIMOB	8	Phage tail sheath Mup39 L
11	2184425-2195884	IslandPath-DIMOB	57	Phage protein (ACLAME 1358), Phage protein, Phage head portal protein B, Phage head terminase DNA packaging protein A, Phage terminase (small subunit), Error-prone, lesion bypass DNA polymerase V (UmuC), DNA translocase FtsK
12	2994186-3010372	IslandPath-DIMOB	17	Transposase and inactivated derivatives, diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with

				PAS/PAC sensor(s), TagE protein, Transcription termination factor Rho, FOG: TPR repeat protein
13	3879100-3893968	IslandPath-DIMOB	25	Phage integrase, Prophage MuSo1 transcriptional regulator (Cro/CI family), Phage replication protein GpA (endonuclease), Caffeoyl-CoA O-methyltransferase
14	3908635-3921064	IslandPath-DIMOB	5*	FIG015547: peptidase (M16 family) / FIG015287: Zinc protease, MotA/TolQ/ExbB proton channel family protein, Tol biopolymer transport system (TolR protein), Phage integrase, Mobile element protein
15	4495875-4506333	IslandPath-DIMOB	1	Inner membrane protein translocase and chaperone YidC (long form), Ribonuclease P protein component, LSU ribosomal protein L34p, Chromosomal replication initiator protein DnaA, DNA polymerase III beta subunit, DNA recombination and repair protein RecF, DNA gyrase subunit B, Oxidoreductase (short-chain dehydrogenase/reductase family)
16	4849376-4870295	SIGI-HMM, IslandPath-DIMOB	10	Oxygen-insensitive NAD(P)H nitroreductase, Dihydropteridine reductase, DNA-binding protein, Predicted cell-wall-anchored protein SasA (LPXTG motif), T1SS secreted agglutinin RTX

Table A.2 Summary of genomic islands predicted by IslandViewer 4 in the genome of *M. spartinae* 468.

*Contains a bacteriophage hypothetical protein.

REFERENCES

Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. 2016. Identification of protein secretion systems in bacterial genomes. Sci. Rep. 6:23080. doi: 10.1038/srep23080.

Aleti G, Baker JL, Tang X, Alvarez R, Dinis M, Tran NC, Melnik A V., Zhong C, Ernst M, Dorrestein PC, et al. 2018. Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease. mBio 10(2): e00321-19. doi: 10.1128/mBio.00321-19.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215(3):403–410. Doi: 10.1016/S0022-2836(05)80360-2.

Amsaveni R, Sureshkumar M, Aravinth A, Mary JR, Vivekanandhan G. 2016. Production of non-ribosomal peptide synthetase (NRPS)- dependent siderophore by *Aeromonas* isolates. Iran. Biomed. J. 20(4):235. doi: 10.7508/ibj.2016.04.008.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. November 20 2020. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Aravind L. 2000. Guilt by association: contextual information in genome analysis. Genome Res. 10(8):1074–1077. doi: 10.1101/gr.10.8.1074.

Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37(4):420–423. doi: 10.1038/s41587-019-0036-z.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1):W16–W21. doi: 10.1093/nar/gkw387.

Auclair SM, Bhanu MK, Kendall DA. 2012. Signal peptidase I: cleaving the way to mature proteins. Protein Sci. 21(1):13–25. doi: 10.1002/pro.757.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. 2002. Short protocols in molecular biology, 5th ed. John Wiley and Sons, New York, NY.

Baumann L, Baumann P, Mandel M, Allen RD. 1972. Taxonomy of aerobic marine eubacteria. J. Bacteriol. 110(1):402–429.

Baumann P, Gauthier MJ, Baumann L. 1984. Genus *Alteromonas* Baumann, Baumann, Mandel and Allen 1972. pp.343–352. In: N. R. Krieg and J. G. Holt (Eds.) Bergey's Manual of Systematic Bacteriology. Williams and Wilkins, Baltimore, MD.

Bertelli C, Brinkman FSL. 2018. Improved genomic island predictions with IslandPath-DIMOB. Bioinformatics 34(13):2161–2167. doi: 10.1093/bioinformatics/bty095.

Bertelli C, Tilley KE, Brinkman FSL. 2019. Microbial genomic island discovery, visualization and analysis. Brief Bioinform. 20(5):1685–1698. doi: 10.1093/bib/bby042.

Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FSL. 2017. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45(W1):W30–W35. doi: 10.1093/nar/gkx343.

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47(W1):W81–W87. doi: 10.1093/nar/gkz310.

Bowditch RD, Baumann L, Baumann P. 1984. Description of *Oceanospirillum kriegii* sp. nov. and *O. jannaschii* sp. nov. and assignment of two species of *Alteromonas* to this genus as *O. commune* comb. nov. and *O. vagum* comb. nov. Curr. Microbiol. 10:221–230. doi: 10.1007/BF01627259.

Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5:8365. doi: 10.1038/srep08365.

Brown TA. 2017. Genomes 4. Garland Science, New York, NY. 520 p.

Califano G, Castanho S, Soares F, Ribeiro L, Cox CJ, Mata L, Costa R. 2017. Molecular taxonomic profiling of bacterial communities in a gilthead seabream (*Sparus aurata*) hatchery. Front Microbiol. 8:204. doi: 10.3389/fmicb.2017.00204.

Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. 67(2):238–276. doi:10.1128/mmbr.67.2.238-276.2003

Canu N, Moutiez M, Belin P, Gondry M. 2020. Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. Nat. Prod. Rep. 37(3):312–321. doi: 10.1039/c9np00036d.

Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. 40(D1):D641–D645. doi: 10.1093/nar/gkr989.

Chikindas ML, Weeks R, Drider D, Chistyakov VA, Dicks LM. 2018. Functions and emerging applications of bacteriocins. Curr. Opin. Biotechnol. 49:23–28. doi: 10.1016/j.copbio.2017.07.011.

Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158(2):412–421. doi: 10.1016/j.cell.2014.06.034.

Dang NP, Landfald B, Willassen NP. 2016. Biological surface-active compounds from marine bacteria. Environ. Technol. 37(9):1151–1158. doi: 10.1080/09593330.2015.1103784.

Dalbey RE, Wang P, van Dijl JM. 2012. Membrane proteases in the bacterial protein secretion and quality control pathway. Microbiol. Mol. Biol. Rev. 76(2):311–330. doi: 10.1128/mmbr.05019-11.

De Pascale G, Nazi I, Harrison PHM, Wright GD. 2011. β-Lactone natural products and derivatives inactivate homoserine transacetylase, a target for antimicrobial agents. J. Antibiot. (Tokyo). 64(7):483–487. doi: 10.1038/ja.2011.37.

Delaux PM. 2017. Comparative phylogenomics of symbiotic associations. New Phytol. 213(1): 89–94. doi: 10.1111/nph.14161.

Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, McArthur AG, Langille MGI, Brinkman FSL. 2015. IslandViewer 3: More flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res. 43(W1):W104–W108. doi: 10.1093/nar/gkv401.

Dong C, Bai X, Sheng H, Jiao L, Zhou H, Shao Z. 2015. Distribution of PAHs and the PAH-degrading bacteria in the deep-sea sediments of the high-latitude Arctic Ocean. Biogeosciences 12(7):2163–2177. doi: 10.5194/bg-12-2163-2015.

Dyall-Smith M. 2009. The Halohandbook: Protocols for Haloarchaeal Genetics. Available at: https://haloarchaea.com/halohandbook/

Edwards DJ, Holt KE. 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb. Inform. Exp. 3(1):2. doi: 10.1186/2042-5783-3-2.

Eichinger V, Nussbaumer T, Platzer A, Jehl MA, Arnold R, Rattei T. 2016. EffectiveDB--updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. Nucleic Acids Res. 44(D1):D669–D674. doi: 10.1093/nar/gkv1269.

Eilers H, Pernthaler J, Glöckner FO, Amann R. 2000. Culturability and in situ abundance of pelagic bacteria from the North Sea. Appl. Environ. Microbiol. 66(7):3044–3051. doi: 10.1128/AEM.66.7.3044-3051.2000.

Elmer WH, Useman S, Schneider RW, Marra RE, LaMondia JA, Mendelssohn IA, Jiménez-Gasco MM, Caruso FL. 2013. Sudden vegetation dieback in Atlantic and Gulf Coast salt marshes. Plant Dis. 97(4):436–445. doi: 10.1094/PDIS-09-12-0871-FE.

Euzéby JP. 1997. LPSN-List of Prokaryotic Names with Standing in Nomenclature. Accessed March 30, 2021. https://lpsn.dsmz.de/family/oceanospirillaceae

Fitch WM. 1970. Distinguishing homologous from analogous proteins. Syst. Biol. 19(2):99–106.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269(5223):496–512. doi: 10.1126/science.7542800.

Frain KM, Robinson C, van Dijl JM. 2019. Transport of folded proteins by the Tat system. Protein J. 38(4):377–388. doi: 10.1007/s10930-019-09859-y.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA, Venter JC. 1995. The minimal gene complement of *Mycoplasma genitalium*. Science 270(5235):397–403. doi: 10.1126/science.270.5235.397

Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. 2020. Pangenomics comes of age: from bacteria to plant and animal applications. Trends Genet. 36(2):132–145. doi: 10.1016/j.tig.2019.11.006

González JM, Whitman WB. 2006. *Oceanospirillum* and related genera. In: The Prokaryotes. Springer, New York, NY. p. 887–915.

Guimaraes LC, Florczak-Wyspianska J, de Jesus LB, Viana MVC, Silva A, Ramos RTJ, de Castro Soares S. 2015. Inside the pan-genome - methods and software overview. Curr. Genomics. 16(4):245–252. doi: 10.2174/1389202916666150423002311.

Hiller K, Grote A, Scheer M, Münch R, Jahn D. 2004. PrediSi: Prediction of signal peptides and their cleavage positions. Nucleic Acids Res. 32:W375–W379. doi:10.1093/nar/gkh378.

Johnstone TC, Nolan EM. 2015. Beyond iron: Non-classical biological functions of bacterial siderophores. Dalton Trans. 44(14):6320–6339. doi:10.1039/c4dt03559c.

Jousset A, Scheu S, Bonkowski M. 2008. Secondary metabolite production facilitates establishment of rhizobacteria by reducing both protozoan predation and the competitive effects of indigenous bacteria. Funct. Ecol. 22(4):714–719. doi: 10.1111/j.1365-2435.2008.01411.x.

Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428(4):726–731. doi: 10.1016/j.jmb.2015.11.006.

Koonin EV, Makarova KS, Wolf YI. 2021. Evolution of microbial genomics: conceptual shifts over a quarter century. Trends Microbiol. 29(7): 582–592. doi: 10.1016/j.tim.2021.01.005.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305(3):567–580. doi: 10.1006/jmbi.2000.4315.

Krueger F. 2015. Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Kumari P, Poddar A, Das SK. 2014. *Marinomonas fungiae* sp. nov., isolated from the coral *Fungia echinata* from the Andaman Sea. Int. J. Syst. Evol. Microbiol. 64(2):487–494. doi: 10.1099/ijs.0.054809-0.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178–2189. doi: 10.1101/gr.1224503.

Lau KWK, Ren J, Wai NLM, Lau SCL, Qian PY, Wong PK, Wu M. 2006. Marinomonas ostreistagni sp. nov., isolated from pearl-oyster culture pond in Sanya, Hainan Province, China. Int. J. Syst. Evol. Microbiol. 56(10):2271–2275. doi: 10.1099/ijs.0.64392-0.

Lu J, Salzberg SL. 2020. SkewIT: The Skew Index Test for large-scale GC Skew analysis of bacterial genomes. Rzhetsky A, editor. PLoS Comput. Biol. 16(12):e1008439. doi:10.1371/journal.pcbi.1008439.

Lucas-Elío P, Goodwin L, Woyke T, Pitluck S, Nolan M, Kyrpides NC, Detter JC, Copeland A, Lu M, Bruce D, Detter C, Tapia R, Han S, Land ML, Ivanova N, Mikhailova N, Johnston AW, Sanchez-Amat A. 2012. Complete genome sequence of *Marinomonas posidonica* type strain (IVIA-Po-181^{T}). Stand. Genomic Sci. 7(1):31–43. doi: 10.4056/sigs.2976373.

Lucena T, Mesa J, Rodriguez-Llorente ID, Pajuelo E, Caviedes MA, Ruvira MA, and Pujalte MJ. 2016. *Marinomonas spartinae sp. nov.*, a novel species with plant-

beneficial properties. Int. J. Syst. Evol. Microbiol. 66(4):1686–1691. doi: 10.1099/ijsem.0.000929.

Maffei B, Francetic O, Subtil A. 2017. Tracking proteins secreted by bacteria: What's in the toolbox? Front. Cell. Infect. Microbiol. 7:221. doi:10.3389/fcimb.2017.00221.

Mavrodi OV., Jung CM, Eberly JO, Hendry SV., Namjilsuren S, Biber PD, Indest KJ, Mavrodi DV. 2018. Rhizosphere microbial communities of *Spartina alterniflora* and *Juncus roemerianus* from restored and natural tidal marshes on Deer Island, Mississippi. Front. Microbiol. 9:3049. doi: 10.3389/fmicb.2018.03049.

Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39:W339. doi:10.1093/nar/gkr466.

Mishra AK, Choi J, Choi SJ, Baek KH. 2017. Cyclodipeptides: An overview of their biosynthesis and biological activity. Molecules. 22(10):1796. doi: 10.3390/molecules22101796.

Mungan MD, Alanjary M, Blin K, Weber T, Medema MH, Ziemert N. 2020 ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. Nucleic Acids Res. 48(W1): W546–W552. doi: 10.1093/nar/gkaa374.

Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Sundaramurthi JC, Lee J, Kandimalla M, Chen IM, Kyrpides NC, Reddy TBK. 2020. Genomes OnLine Database (GOLD) v.8: overview and updates. Nucl. Acids Res. 49(D1):D723–D733. doi: 10.1093/nar/gkaa983.

Okuda S, Tokuda H. 2011. Lipoprotein sorting in bacteria. Annu. Rev. Microbiol. 65:239–259. doi: 10.1146/annurev-micro-090110-102859.

Paetzel M, Karla A, Strynadka NC, Dalbey RE. 2002. Signal peptidases. Chem. Rev. 102(12):4549–4580. doi: 10.1021/cr010166y.

Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M. 2020. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. Int. J. Syst. Evol. Microbiol. 70(11):5607–5612. doi: 10.1099/ijsem.0.004332.

Peng C, Shi C, Cao X, Li Y, Liu F, Lu F. 2019. Factors influencing recombinant protein secretion efficiency in Gram-positive bacteria: signal peptide and beyond. Front. Bioeng. Biotechnol. 7:139. doi:10.3389/fbioe.2019.00139.

Pot B, Gillis M, Hoste B, van De Velde A, Bekaert F, Kersters K, De Ley J. 1989. Intra- and intergeneric relationships of the genus *Oceanospirillum*. Int. J. Syst. Bacteriol. 39(1):23–34. doi: 10.1099/00207713-39-1-23.

Richter AA, Mais CN, Czech L, Geyer K, Hoeppner A, Smits SHJ, Erb TJ, Bange G, Bremer E. 2019. Biosynthesis of the stress-protectant and chemical chaperon ectoine: biochemistry of the transaminase EctB. Front. Microbiol. 10:2811. doi: 10.3389/fmicb.2019.02811.

Roongsawang N, Washio K, Morikawa M. 2011. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int. J. Mol. Sci. 12(1):141–172. doi: 10.3390/ijms12010141.

Rouli L, Merhej V, Fournier P-E, Raoult D. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. New Microbes New Infect. 7:72–85. doi: 10.1016/j.nmni.2015.06.005.

Sardi M, Gasch AP. 2017. Incorporating comparative genomics into the design–test–learn cycle of microbial strain engineering. FEMS Yeast Res. 17(5):fox042. doi: 10.1093/femsyr/fox042.

Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y, Yang B, Zhang S, Fa Z, Zhao B, Xiang Z, Li Y, Zhao Z, Olszewski MA, Chen L, He Y. 2019. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. Nucleic Acid Res. 47(D1):D693–D700. doi: 10.1093/nar/gky999.

Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Banfield JF. 2020. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. MBio 11(3):1–17. doi:10.1128/MBIO.00416-20.

Sinha S, Nge CE, Leong CY, Ng V, Crasta S, Alfatah M, Goh F, Low KN, Zhang H, Arumugam P, et al. 2019. Genomics-driven discovery of a biosynthetic gene cluster required for the synthesis of BII-Rafflesfungin from the fungus *Phoma* sp. F3723. BMC Genomics. 20(1):374. doi: 10.1186/s12864-019-5762-6.

Sleator RD. 2013. A beginner's guide to phylogenetics. Microb. Ecol. 66(1):1–4. doi: 10.1007/s00248-013-0236-x.

Stothard P, Wishart DS. 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21(4):537–539. doi: 10.1093/bioinformatics/bti054.

Touchon M, Rocha EP. 2016. Coevolution of the organization and structure of prokaryotic genomes. Cold Spring Harb. Perspect. Biol. 8(1):a018168. doi: 10.1101/cshperspect.a018168.

Trachana K, Larsson TA, Powell S, Chen WH, Doerks T, Muller J, Bork P. 2011. Orthology prediction methods: a quality assessment using curated protein families. Bioessays 33(10):769–780. doi: 10.1002/bies.201100062.

Tripathi L, Irorere VU, Marchant R, Banat IM. 2018. Marine derived biosurfactants: a vast potential future resource. Biotechnol. Lett. 40(11):1441–1457. doi: 10.1007/s10529-018-2602-8.

Van Landschoot A, De Ley J. 1983. Intra-and intergeneric similarities of the rRNA cistrons of *Alteromonas*, *Marinomonas* (gen. nov.) and some other Gram-negative bacteria. Microbiology 129(10):3057–3074. doi: 10.1099/00221287-129-10-3057.

Vignier J, Laroche O, Rolton A, Wadsworth P, Kumanan K, Trochel B, Pochon X, King N. 2021. Dietary exposure of Pacific oyster (*Crassostrea gigas*) larvae to compromised microalgae results in impaired fitness and microbiome shift. Front Microbiol. 12:706214. doi: 10.3389/fmicb.2021.706214.

Villasante A, Catalán N, Rojas R, Lohrmann KB, Romero J. 2020. Microbiota of the digestive gland of red abalone (*Haliotis rufescens*) is affected by withering syndrome. Microorganisms 8(9):1411. doi: 10.3390/microorganisms8091411.

Voets IK. 2017. From ice-binding proteins to bio-inspired antifreeze materials. Soft Matter 13(28):4808–4823. doi: 10.1039/c6sm02867e.

Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics 7:142. doi: 10.1186/1471-2105-7-142.

Walther TH, Ulrich AS. 2014. Transmembrane helix assembly and the role of salt bridges. Curr. Opin. Struct. Biol. 27(1):63–68. doi: 10.1016/j.sbi.2014.05.003.

Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 42:D581–D591. doi: 10.1093/nar/gkt1099.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13(6):e1005595. doi: 10.1371/journal.pcbi.1005595.

Wong WC, Maurer-Stroh S, Schneider G, Eisenhaber F. 2012. Transmembrane helix: Simple or complex. Nucleic Acids Res. 40(W1):W370. doi: 10.1093/nar/gks379.

Wu Y, Zaiden N, Cao B. 2018. The core- and pan-genomic analyses of the genus *Comamonas*: from environmental adaptation to potential virulence. Front. Microbiol. 9:3096. doi: 10.3389/fmicb.2018.03096.

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, Wang Y. 2019. OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 47(W1): W52–W58. doi: 10.1093/nar/gkz333.

Yancey PH. 2005. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. J. Exp. Biol. 208(Pt 15):2819– 2830. doi: 10.1242/jeb.01730.

Zheng Y, Saitou A, Wang CM, Toyoda A, Minakuchi Y, Sekiguchi Y, Ueda K, Takano H, Sakai Y, Abe K, Yokota A, Shuhei Y. 2019. Genome features and secondary metabolites biosynthetic potential of the class *Ktedonobacteria*. Front. Microbiol. 10:893. doi: 10.3389/fmicb.2019.00893.

Zhu D, He J, Yang Z, Wang M, Jia R, Chen S, Liu M, Zhao X, Yang Q, Wu Y, Zhang S, Liu Y, Zhang L, Yu Y, Yu Y, Chen X, Cheng A. 2019. Comparative analysis reveals the genomic islands in *Pasteurella multocida* population genetics: on symbiosis and adaptability. BMC Genom. 20(1):63. doi: 10.1186/s12864-018-5366-6.