ABSTRACT

Title of Dissertation:	BIOPROSPECTING MARINE ACTINOMYCETES FOR NOVEL ANTI-TUBERCULOSIS DRUGS
	Daniela Tizabi, Doctor of Philosophy, 2022
Dissertation directed by:	Dr. Russell T. Hill, Professor, Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science

Mycobacterium tuberculosis (*M. tb*), the causative agent of the infectious lung disease tuberculosis (TB), is estimated to infect approximately 1.7 billion people worldwide. This pathogen was responsible for more than 1.5 million deaths in 2020, and is likely to remain a global threat for many years to come due to the rising incidence of antibiotic resistance, as well as dramatic setbacks in treatment due to the ongoing COVID-19 pandemic. There is an urgent demand for novel therapeutics to treat the disease through unique mechanisms of action. In the search for these drugs, a novel collection of 101 marine actinomycetes previously isolated from a Caribbean giant barrel sponge *Xestospongia muta* was investigated for their ability to inhibit *M. tb* growth. Thirteen novel strains of *Micrococcus*, *Micromonospora*, *Brevibacterium*, and *Streptomyces* were identified as consistently producing extracts that inhibit *M. tb* in a dose-dependent manner. After sequencing the genomes of these strains, a

comparative analysis between three assembly algorithms (SPAdes, A5-miseq, Shovill) was performed to determine which program yielded the best assembly from Illumina MiSeq data for biosynthetic gene cluster (BGC) mining. Upon characterizing the biosynthetic potential of each strain, two isolates generating highly potent extracts – *Micrococcus* sp. strain R8502A1 and *Micromonospora* sp. strain R45601 – were selected for further analysis through a dual genomics and chemistry-enabled approach. No compounds with obvious anti-TB activity were detected in the genome of *Micrococcus* sp. strain R8502A1, suggesting production of an elusive and novel anti-TB compound through a cryptic pathway. A comprehensive examination of all BGC-associated domains was conducted to evaluate possible biosynthetic pathways linked to the anti-TB activity observed. The active component of the *Micrococcus* extract was further isolated with high performance liquid chromatography (HPLC) and is under investigation with liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR). In contrast, a BGC with 94% similarity to the selective and potent but poorly soluble anti-TB compound diazaquinomycin H/J was identified in the genome of *Micromonospora* sp. strain R45601, suggesting production of a chemical analog. LC-MS detected four peaks of interest, two of which are associated with mass-to-charge (*m/z*) values that do not correlate with any previously identified diazaquinomycin analogs. This analysis has identified at least two potentially novel anti-TB compounds, supporting continued investigation into sponge-associated marine actinomycetes for novel therapeutics.

BIOPROSPECTING MARINE ACTINOMYCETES FOR NOVEL ANTI-TUBERCULOSIS DRUGS

by

Daniela Tizabi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
 Professor Russell T. Hill, Chair
 Professor Allen Place
 Professor Yantao Li
 Research Chemist Mary Bedner
 Professor Jacques Ravel

Dedication

To my parents, who instilled in me the importance of education and have always supported my academic endeavors, and to Bas and Playa, for being my greatest role models and biggest cheerleaders. I love you all so much.

Acknowledgements

There are so many people who have helped me throughout my graduate career and I am deeply grateful for every one of you. Thank you to all of my family and friends, who have shown nothing but support in my pursuit of higher education. Mom, Dad, Reuel, Jonas, Lisa, Kelly, Katie, Madison, Emma, and Jillian - you all have been there since the beginning, and I cannot express how much your love and encouragement mean to me.

Although I wanted to pursue a career in marine biology ever since I first watched *Finding Nemo* in 4th grade (along with the rest of my classmates), it wasn't until high school that I really became serious about this. I am forever grateful that my 11th grade AP Biology teacher, Mrs. Pamela Leffler, invited me to attend the 2009 Howard Hughes Medical Institute's Holiday Lecture series "Exploring Biodiversity: The Search for New Medicines", where I was first introduced to (and immediately captivated by) the field of drug discovery. Additionally, I am very appreciate of Dr. Liza Merly, who taught a course on marine biomedicine my senior year of college, and introduced me to IMET and encouraged me to apply there for graduate studies.

Several fellowships supported my graduate studies over the past six years, each of which has introduced me to researchers who have had major impacts on my development as a scientist. I was very fortunate to receive the NIST-IMET Graduate Fellowship in Environmental Biotechnology as an incoming student. Dr. Trina Mouchahoir at the Institute for Bioscience and Biotechnology Research/ National Institute of Standards and Technology (NIST) was an incredible mentor to me during my summer as a NIST intern. The training she provided on how to analyze mass

independent way of thinking, have made me the scientist I am today. Your endless

supply of fascinating stories from life have also been much appreciated.

Many names remain unmentioned here, but you all have made my time at

IMET incredibly rewarding. Thank you all.

Table of Contents

List of Tables

List of Figures

xii

List of Abbreviations

A	Adenylation
ACP	Acyl carrier protein
ANI	Average nucleotide identity
ANIb	Average nucleotide identity based on BLAST+
antiSMASH	Antibiotics & Secondary Metabolite Analysis Shell
AT	Acyltransferase domain
AZT	Azidothymidine
BCG	Bacille Calmette-Guérin
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
BLASTN	Nucleotide Basic Local Alignment Search Tool
BLASTX	Translated nucleotide to protein Basic Local Alignment Search Tool
C	Condensation
CoA	Coenzyme A
COVID-19	Coronavirus disease 2019
CRISPR	Clustered regularly interspaced short palindromic repeats
crtB	Phytoene synthase
crtE	Geranylgeranyl diphosphate (GGPP) synthase
crtE2	Lycopene elongase
crtI	Phytoene desaturase or dehydrogenase
crtYe/crtYf	Epsilon(ε)-cyclase

crtYg and *crtYh*	Gamma(γ)-cyclase
crtX	Glycosyltransferase or glucosyltransferase
Cy	Heterocyclization
Cys	Cysteine
Da	Daltons
DH	Dehydratase
DMAPP	Dimethylallyl diphosphate
DMSO	Dimethyl sulfoxide
E	Epimerization
ER	Enoylreductase
ESI	Electrospray ionization
EtOAc	Ethyl acetate
XDR-TB	Extensively drug-resistant TB
F	Formylation
GGPP	Geranylgeranyl diphosphate
GNPS	Global Natural Products Social molecular networking
HGT	Horizontal gene transfer
HMA	High microbial abundance
HMMs	Hidden Markov models
HPLC	High-performance liquid chromatography
iChip	Isolation chips
ISP2	International *Streptomyces* project 2
KR	Ketoreductase

KS	Ketosynthase
LMA	Low microbial abundance
MBC	Minimum bactericidal concentration
MDR-TB	Multi-drug resistant TB
MeOH	Methanol
MEP	Methylerythritol 4-phosphate pathway
MIC	Minimum inhibitory concentration
MS	Mass spectrometry
MT	Methyltransferase
MVA	Mevalonate pathway
NAPAA	Non-alpha poly-amino acids like e-polylysin
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
nr/nt	Nucleotide collection
OSMAC	One strain many compounds
Ox	Oxidation
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PPT	Phosphopantetheine
PPTase	Phosphopantetheinyltransferase
PK	Polyketide

PKS	Polyketide synthase
QUAST	Quality Assessment Tool for Genome Assemblies
R2A	Reasoner's 2A agar
R2B	Reasoner's 2A broth
Re	Reductase
RiPPs	Ribosomally synthesized and post-translationally modified peptides
RRE	Rev response element
rRNA	Ribosomal RNA
RR-TB	Rifampicin-resistant TB
Ser	Serine
SMILES	Simplified Molecular Input Line Entry Specification
T	Thiolation
TB	Tuberculosis
Te	Thioesterase
WHO	World Health Organization

Statement of contribution

All of the Sanger sequencing and Illumina MiSeq whole genome sequencing was performed by Mrs. Sabeena Nazar from the BioAnalytical Services Laboratory. The Institute for Genome Sciences' Genomics Resource Center performed the PacBio sequencing, facilitated by Dr. Jacques Ravel. Dr. Tsvetan Bachvaroff provided extensive assistance in genome assembly and bioinformatics troubleshooting. LC-MS/MS analysis was conducted by Dr. Arpad Somogyi and Dr. Gong Wu of the Mass Spectrometry and Proteomics Facility at the Ohio State University, supported by NIH Award Number Grant P30 CA016058, and facilitated by Dr. Liva Harinantenaina Rakotondraibe. Fractionation with HPLC, HPLC-UV analysis and assistance with interpretation of results was also performed by Dr. Harinantenaina Rakotondraibe.

Chapter 1: Introduction

1.1 A brief history of tuberculosis

1.1.1 An ancient and peculiar record of tuberculosis

Tuberculosis (TB), the infectious lung disease caused by the pathogen *Mycobacterium tuberculosis* (*M. tb*), has a very long and complicated history. The first written record of the ailment dates back to 3,300-2,300 years ago, in India and China, respectively (Barberis et al., 2017; Brown, 1941; Cave & Demonstrator, 1939). However, skeletal deformities observed in excavated Egyptian mummies from 2400 BC as well as symbolism in ancient Egyptian art prove that TB has inflicted humans for much longer (Morse et al., 1964; Zimmerman, 1979). In the millennia since, TB has been the culprit of numerous epidemics, the first of which is believed to have hit the precolonial Americas over 1,500 years ago, and more recently ravaged Western Europe in the 18th century (Barberis et al., 2017; Daniel, 2000b; Morse, 1961; Salo et al., 1994). Before the term "tuberculosis" was adopted to describe the disease in the mid 1800s, numerous other names were given to various *M. tb* infections, including scrofula (*M. tb* infection in the lymph nodes), "King's Evil", phthisis, consumption, white plague, "the robber of youth", "the Captain of all these men of Death", and the graveyard cough (Barberis et al., 2017; Daniel, 2000a; Frith, 2014). It even receives mention in the bible under the name schachepheth (Daniel & Daniel, 1999). Just as numerous as the labels for this disease are the treatments. It was at different times believed that spa visits, fresh air, sea voyages, pulmonary collapse therapy, and more comically, milk, and the touch of royalty or visiting a royal tomb could cure infected individuals (Barberis et al., 2017; Daniel, 1997; Daniel, 2006; Pease, 1940). Although the contagious nature of the disease was first suggested in the

Ordinances of Manu in India in the 13th century BC (Blacklock, 1947), the bacterial origin of TB was not discovered until over 3,000 years later, when Dr. Robert Koch isolated the tubercle bacterium in 1882 (Barberis et al., 2017; Sakula, 1982).

This wasting disease has even made a peculiar impact in literature and popular culture. Poets of the 19th century began romanticizing the ailment, describing those suffering as appearing thin, of pale complexion, melancholy, delicate, and even angelic (Frith, 2014). Feminine beauty in Victorian England idealized a pallid and emaciated appearance, so much so that suffering in this manner became fashionable. As women grew thinner, they flaunted their newly slimmed waists with pointed corsets and skirts. In the following century, doctors began to fear that long skirts could sweep up germs responsible for TB and forewarn constrictive corsets, so hemlines rose a few inches to address this issue. Now with shoes more visible, women began to pay more attention to the style of their shoes as an essential part of their outfit. Men traded in their trendy moustaches and beards for a clean-shaven face for fear that the facial hair could transmit TB, among other infectious diseases. Even the modern concept of tanning has its roots in TB treatment of the early 1900s, when doctors began prescribing heliotherapy to build up immunity (Lo Grasso, 1928; Mullin, 2016). On a more morbid note, TB is also now believed to be the inspiration for the classical imagery of vampires (Byrne, 2011, as cited in Talairach-Vielmas, 2011). This is likely due to the aforementioned pale appearance of victims and their tendency to cough up blood as a symptom of infection.

1.1.2 Current status of and treatments for tuberculosis

Since Dr. Koch's discovery, it has been confirmed that TB spreads through the air, and when infected individuals expel bacteria into the air, such as by coughing or sneezing. Despite its ancient origins, this disease remains one of the top ten leading causes of death worldwide. The World Health Organization (WHO) has addressed this concern by developing a comprehensive "End TB Strategy" aimed at reducing TB incidence by 80%, TB deaths by 90%, and ruinous costs for all TB-affected households by 2030 (World Health Organization, 2020). It is currently estimated that 1.7 billion people are infected with TB (Houben & Dodd, 2016). Although most of these cases are latent (inactive and not contagious), 5-10% of those infected will likely develop the disease in their lifetime, and chances are even higher for individuals who are or become immunocompromised. An HIV-positive diagnosis is the highest risk factor, making the patient on average 18 times more likely to develop active TB. Additionally, undernutrition, diabetes, smoking and alcohol consumption can all raise the risk of TB reactivation. In 2020, approximately 9.9 million people became infected with TB, 1.5 million of whom died as a result of infection (Fig. 1.1). In the previous year, approximately 500,000 cases developed rifampicin-resistant infections (RR-TB), 78% of which further evolved into multi-drug resistant TB (MDR-TB) (Fig. 1.2). MDR-TB is defined as a TB infection that is resistant to both rifampicin and isoniazid. The most prominent areas of infection are Southeast Asia, Africa and the Western Pacific. Just eight countries account for nearly two-thirds of the global total – India (26%), Indonesia (8.5%), China (8.4%), the Philippines (6%), Pakistan (5.8%), Nigeria (4.6%), Bangladesh (3.6%) and South Africa (3.3%),

highlighting the need to focus efforts and resources in these areas (World Health Organization, 2020; World Health Organization, 2021a).

Currently, drug-susceptible TB cases require a lengthy six-month treatment with antibiotics to cure the infection. Isoniazid and rifamycin are the two first line drugs most commonly used to treat the disease, along with ethambutol and pyrazinamide. For RR-TB and MDR-TB infections, a more taxing treatment of nine to 20 months is required involving second-line injectables, albeit with only a 57% success rate for MDR-TB (World Health Organization, 2020). Recently, a new drug regimen was approved by the U.S. Food and Drug Administration (FDA) to treat extensively drug-resistant TB (XDR-TB) (U.S. Food and Drug Administration, 2019). As of January 2021, XDR-TB is defined as any *M. tb* strain causing infection that is resistant to all first-line drugs, any fluoroquinolone, as well as at least one of the most potent second-line drugs (levofloxacin, moxifloxacin, bedaquiline or linezolid) (World Health Organization, 2021b). Typical treatment for XDR-TB requires a combination of eight drugs for more than 18 months, but the success rate is only around 34% and treatment may result in deafness, nephrotoxicity or psychosis (World Health Organization, 2016; TB Alliance, 2019; World Health Organization, 2018). This new regimen (BPaL) involves a combination of three drugs – pretomanid, bedaquiline and delaminid. A recent clinical trial observed that nearly 90% of participants infected with XDR-TB recovered after following this treatment for six months. The most common side effects to this treatment include but are not limited to nerve damage, anemia, nausea, and vomiting (TB alliance, 2019). One singular vaccine developed from a live attenuated strain of *M. tb* known as bacille Calmette-Guérin (BCG) exists

to date, although it is only effective to prevent infection in infants and young children (Luca & Mihaescu, 2013). TB can infect various parts of the body, including the lymph nodes, nervous system, bones, joints, gastrointestinal tract and genitourinary tract, though it is most prevalent in the respiratory tract (Barberis et al., 2017; World Health Organization, 2013). The BCG vaccine only provides moderate protection against severe forms of TB, variable efficacy against pulmonary TB, and is universally used in 157 countries of the 180 for which data are available, with the exceptions including the USA, Canada, and nine other countries that have stopped universal BCG vaccination programs since the early 1980s (Zwerling et al., 2011). No vaccines are approved at present to treat adults; however, one vaccine candidate - M72/AS01E – has shown promising efficacy (50%) against *M. tb* infection in adults (Tait et al., 2019; World Health Organization, 2020). As of October 2021, 14 additional TB vaccine candidates are in various phases of clinical trials, including four candidates currently in phase III trials and suggested to protect against infection in adults (GamTBvac), adolescents (MIP), or both (VPM1002, MTBVAC) (Frick, 2021; King, 2021; World Health Organization, 2020).

Estimated TB incidence rates, 2020
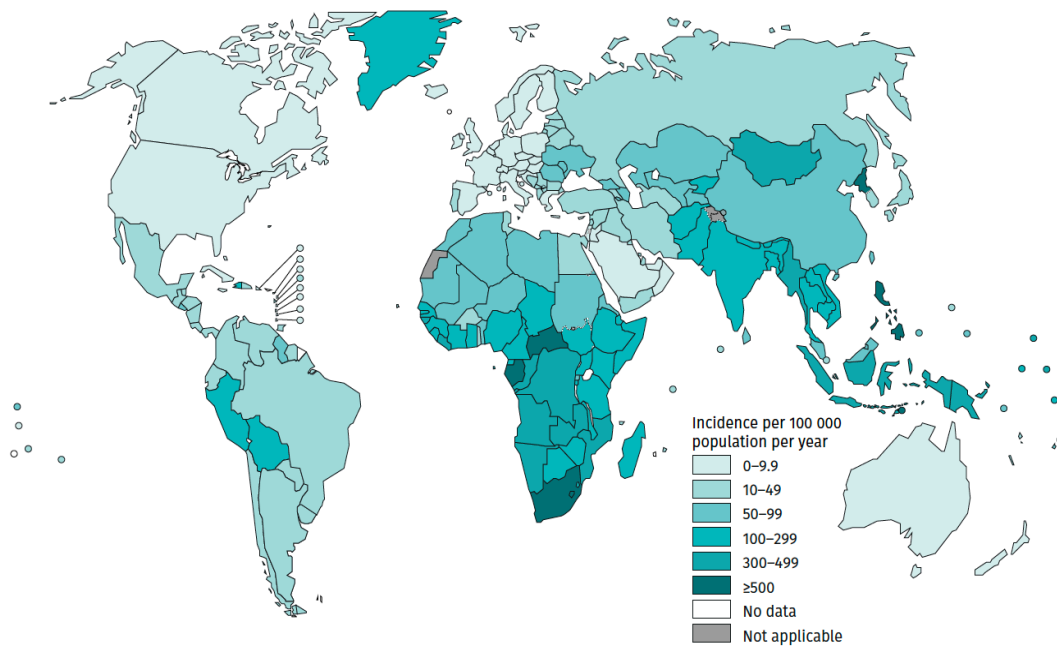
Figure 1.1. Most recent data on estimated TB incidence rates. Source: Global tuberculosis report 2021. World Health Organization. License: CC BY-NC-SA 3.0 IGO.

Percentage of new TB cases with MDR/RR-TB[a]
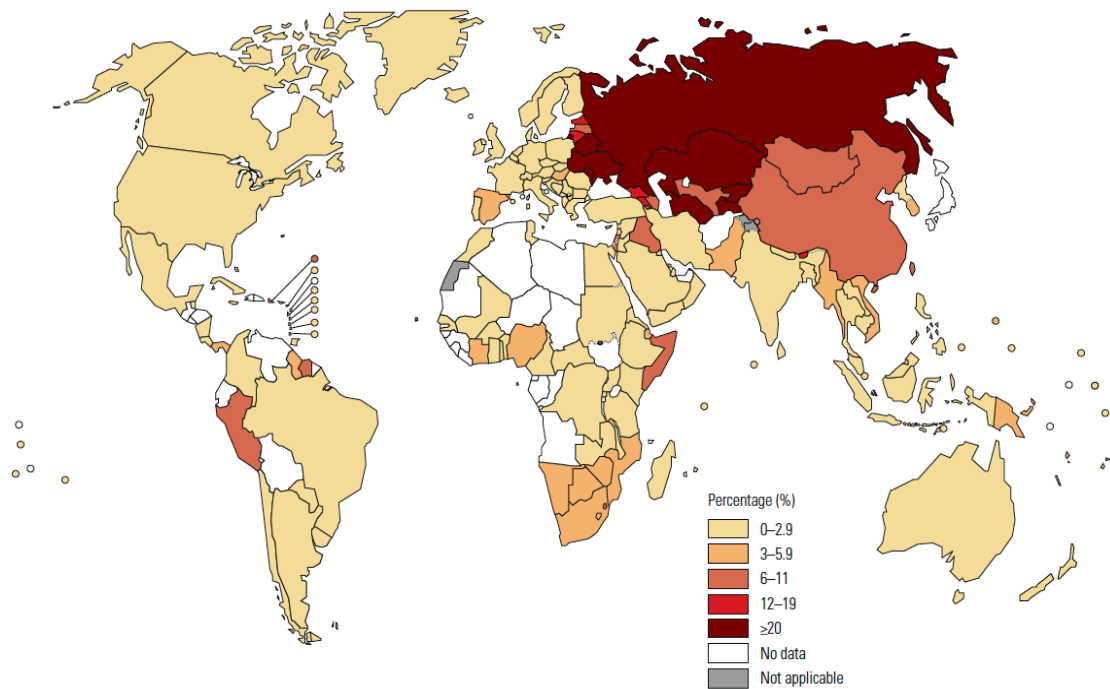
a Percentages are based on the most recent data point for countries with representative data from 2005 to 2020. Model-based estimates for countries without data are not shown. MDR-TB is a subset of RR-TB.

Figure 1.2. Most recent data on percentage of new TB cases documented with MDR/RR-TB. Source: Global tuberculosis report 2020. World Health Organization. License: CC BY-NC-SA 3.0 IGO.

1.1.3 Barriers to eradicating *Mycobacterium tuberculosis*

There are various reasons as to why TB remains such a prominent threat to public health. Regarding treatment compliancy, many patients find it difficult to maintain the long-term six month drug regimen, either due to inconvenience, adverse side effects, or the belief that they no longer require the drugs once they notice an improvement (Field, 2015). Furthermore, this is a disease of poverty, and access to treatment is most challenging for impoverished individuals, an already high-risk group. The newly approved treatment for XDR-TB appears very promising, but exorbitant costs for these drugs limit their range and prevent those who need them most from essential access. In addition, the coronavirus disease 2019 (COVID-19) pandemic has exacerbated concerns about achieving the WHO's 2020 milestones. By the end of 2019, the majority of high TB burden countries were not on track to reach their 2020 goals of reducing TB deaths by 35%, achieving a cumulative reduction of only 14% between 2015 and 2019. Emergency response forced health workers to redirect resources (diagnostic testing machines, staff, and budgets) towards COVID-19 efforts suddenly albeit temporarily. Reduced resources and mandatory quarantining led to severe underreporting in some of the countries with the highest TB burdens in 2020. Overall, an 18% global decline in reports and new diagnoses was observed between 2019 and 2020, and there is fear that the impacts of the pandemic on the fight against TB will linger for a several years to come. Despite increased social distancing, higher poverty, increased household exposure, and longer duration of infectiousness as a result of quarantining likely contributed to worsening outcomes. Health officials fear that disruption caused by COVID-19 could result in

over 1 million new cases of TB annually between 2020 and 2025 (World Health Organization, 2020, World Health Organization, 2021a). Furthermore, immunosuppression caused by COVID-19 increases the risk of reactivation in susceptible patients. At least 89 case studies of COVID-19/TB coinfection have already been documented and strongly suggest worse outcomes for these patients (Song et al., 2021).

Perhaps the most insidious barrier to ending TB is the rise in antibiotic resistance. Lapses in treatment, as described above, as well as natural evolution can result in resistance to treatment. The top two drugs used in treatment – isoniazid and rifampicin- were developed in 1952 and 1957, respectively (Daniel, 2006). Unfortunately, it did not take long for the first cases of drug-resistant *M. tb* to emerge, with resistance to isoniazid being reported almost immediately after its activity was discovered (Vilchèze & Jacobs, 2014). The fate of rifampicin was not much different. Introduced as a therapeutic one year after its discovery in 1957, the first case of resistance against rifampicin was observed only four year later, in 1962 (Lewis, 2013). The inevitability of pathogens developing resistance to antibiotics limits the efficacy of currently available treatments. It is critical that new therapeutic agents are developed that target *M. tb* through novel mechanisms of action to curtail the emergence of resistance.

1.2 The role of actinomycetes and their natural products in drug discovery
1.2.1 The golden age of drug discovery

In the search for novel anti-TB drugs, actinomycetes may serve as a promising source. Actinomycetes (of the Phylum Actinobacteria) are a diverse group of Gram-positive filamentous bacteria that are present throughout all environments – terrestrial, marine and aquatic (van der Meij et al., 2017). Interest in investigating these microbes for their pharmacological activity began in 1940, after the discovery of the first marketable antibiotic from a soil actinomycete. Waksman's graduate student H. Boyd Woodruff had actually isolated actinomycin four years prior (1940) and streptothricin two years prior (1942) to the discovery of streptomycin, but toxicity of both *Streptomyces* compounds precluded further development as clinical drugs (Waksman & Woodruff, 1940, 1942; Woodruff, 2014). The compound, which came to be known as streptomycin, was isolated from *Streptomyces griseus* by Dr. Waksman's laboratory[1] at Rutgers University, and investigated for its effectiveness against *M. tb* as well as various gram-negative pathogens (Schatz et al., 1944; Woodruff, 2014). The enduring importance of this discovery is highlighted by the recent naming of *S. griseus* as the State Microbe of New Jersey! (Bennett et al., 2021). Pharmaceutical companies quickly took notice of this potential and began rapid screening actinomycetes in a high throughput fashion (Katz & Baltz, 2016). In the following decades, drug discovery efforts experienced what is now fondly remembered as the "Golden Age of Drug Discovery". During this period, which lasted from the 1940s-1960s, tens of thousands of compounds were isolated from

---

[1] Despite Dr. Waksman receiving the Nobel Prize in Physiology or Medicine in 1952 for this discovery, the compound was actually discovered by his graduate student Albert Schatz. Unfortunately, rapid development of resistance to streptomycin and the eventual realization that it is only bacteriostatic and not bactericidal led to its eventual retirement as a monotherapy (Bogen, 1948; Lewis, 2013; Woodruff, 2014).

terrestrial actinomycete strains and screened for bioactivity against various pathogens/ailments in the pursuit of new and effective treatments (Bérdy, 2005; Kolter & van Wezel, 2016). Practically every important group of antibacterial compounds was discovered during this period, along with chemotherapeutic agents to treat cancer and viruses (Bérdy, 2005). This golden age did not last long, however; researchers began rediscovering more and more already-described compounds just a decade or two later. High rediscovery rates, low yields, increased costs of research, and increasingly risky drug leads led to pharmaceutical companies largely abandoning discovery efforts for more profitable pursuits, such as treatments for chronic illnesses (Bérdy, 2005, 2012; Fenical, 1999; Kolter & van Wezel, 2016; Lewis, 2013; Payne et al., 2007; Sekurova et al., 2019; Singh, 2014; van der Meij et al., 2017; Woon & Fisher, 2016). Eventually, rising rates of antibiotic resistance forced researchers to renew their bioprospecting efforts to develop new treatment options. The original discovery pipeline had long since dried up, and resuming the classical screening approach required a new source. Past sampling relied mainly on terrestrial environments due to ease of access (mainly soils) that were believed to have been exhausted as a source or novel actinomycetes at this point. By 2005, approximately 66% of all antibiotic discoveries could be attributed to actinomycetes, with the majority of compounds isolated originating from *Streptomyces* strains (Bérdy, 2005; Hodgson, 2000). Scientists began to advocate for the need to sample from remote environments to fully exploit the microbial biodiversity that has evolved in isolation (Bérdy, 2012; Kolter & van Wezel, 2016; Lam, 2006; Okami, 1982). Improved technologies and cultivation techniques and reduced costs over the years

11

finally enabled scientists to turn their attention towards a previously inaccessible source: the marine environment (Fenical & Jensen, 2006; Lindequist, 2016; Vogel, 2008).

1.2.2 Marine actinomycetes as a novel source of pharmaceuticals

Originally believed to be terrestrial strains that had simply found their way into the marine environment over time, it is now known that the ocean harbors unique actinomycetes, some of which are obligate marine species (Fenical & Jensen, 2006; Goodfellow & Williams, 1983; Maldonado et al., 2005). This observation does not preclude the possibility that they did originate from terrestrial strains, but clearly affirms that they have resided in the marine environment long enough to evolve phylotypes significantly distinct from their terrestrial counterparts (Ian et al., 2014; Penn et al., 2009; Penn & Jensen, 2012). This is understandable given the unique and often harsh conditions to which these microbes are exposed. Extreme temperatures, high pressure and high salinity are among the many factors to which marine bacteria have adapted to over time (Fenical, 1993; Lam, 2006). Often, marine actinomycetes are associated with marine invertebrates such as sponges as part of a symbiotic relationship. In exchange for shelter, concentrated access to nutrients, and the limiting element nitrogen, the microbial symbionts provide their immotile hosts with a chemical arsenal to defend against predators and compete for space and resources (Bérdy, 2005; Hentschel et al., 2012; van der Meij et al., 2017). Evidence that sponges actively select for symbionts and can transfer symbionts to their offspring further supports the notion that these microbes provide an advantage to their host

(Hentschel et al., 2002; Schmitt et al., 2007). These chemicals that they produce are secondary metabolites, meaning they are not required for growth, development or reproduction of the organism (Baltz, 2019). Therefore, they must confer a significant advantage for the microbe to retain the genes necessary for their expression and production. In the natural environment, where these excreted compounds are quickly diluted as they disperse, they likely serve as signaling molecules (Sengupta et al., 2013). When the bacteria are removed from the water and tested in the laboratory setting, the concentration of these secondary metabolites dramatically increases, and in these higher concentrations these compounds have been discovered to have diverse pharmaceutically-relevant activity, including antibacterial, antiviral, anticancer, antifungal and antiparasitic properties, among others (Abu-Salah, 1996; Bérdy, 2005; Nakagawa et al., 1981; Peraud, 2006; Ribeiro et al., 2020; Subramani & Aalbersberg, 2013).

Actinomycetes associated with marine sponges are of particular interest when searching for novel strains considering the sheer volume harbored inside the sessile host (Webster et al., 2010). Studies have shown that microbial symbionts can account for up to 40% of sponge biomass by volume (Vacelet, 1975). By 2014, 32 distinct bacterial phyla and at least 60 actinomycete genera had been isolated from marine sponges (Abdelmohsen et al., 2014; Schmitt et al., 2012), although more recent analyses have identified more than 40 and as many as 47 phyla, when including rare and candidate phyla (Reveillaud et al., 2014; Thomas et al., 2016).

Specifically considering the possibility that marine actinomycetes will synthesize antimycobacterial compounds, research suggests this likelihood is based

on more than just random chance due to sheer abundance of actinomycetes harbored within host tissue. In a recent study, 11 *Mycobacterium* species, together with an antimycobacterial *Salinispora* species, were isolated from the sponge *Amphimedon queenslandica* (Izumi et al., 2020). Several *Salinispora* species, including the strain isolated in the 2020 investigation, are capable of synthesizing rifamycins, a group of antibiotics that includes one of the top-line anti-TB drugs, rifampicin (Kim et al., 2006; Wilson et al., 2010). The authors of the study hypothesize that production of antimycobacterial compounds by marine actinomycetes may function in competition between the cohabiting sponge symbionts. Furthermore, several anti-TB compounds have already been isolated from marine sponges and associated actinomycetes. In addition to marine-sponge derived rifamycin-producing *Salinispora* strains (Kim et al., 2006), macfarlandins (anti-TB diterpenoids) were isolated from a Samoan *Chelonaplysilla* sponge (de Oliveira et al., 2020), and haliclonadiamines derivatives (antimycobacterial alkaloids) were isolated from an Okinawan *Haliclona* sponge species (Abdjul et al., 2018). It is important to note that only crude *Haliclona* and *Chelonaplysilla* extracts were tested for growth inhibition, and thus, the possibility cannot be ruled out that the isolated antimycobacterial compounds actually derive from associated microbes.

Approximately 10% of isolated bacterial strains tend to possess activity in preliminary screenings (Riyanti et al., 2020). Coupled with the fact that marine sponges also harbor rare actinomycete genera, it is highly likely that investigation of these strains will reveal novel compounds with unique activity and structural novelty (Abdelmohsen et al., 2014; Pye et al., 2017; Subramani & Aalbersberg, 2013). In fact,

a 2010 analysis found that approximately 71% of the molecular scaffolds described in the Dictionary of Marine Natural Products were exclusive to the marine environment (Kong et al., 2010). This claim is supported by the fact that numerous research endeavors have already identified bioactivity or molecular scaffolds indicative of biosynthetic potential from sponge-associated actinomycetes isolated from unique marine environments such as the Caribbean (Vicente et al., 2013), the South China Sea (Sun et al., 2015), Indonesian sponges (Riyanti et al., 2020), the Persian Gulf (Matroodi et al., 2020), deep sea habitats off the coast of Ireland (Borchert et al., 2016), in addition to sediment strains collected from the Northern coast of Portugal (Ribeiro et al., 2020), Valparaíso Bay, Chile (Claverías et al., 2015), and even at the deepest depths of the ocean - Challenger Deep (Pathom-aree et al., 2006). It should be noted that this is by no means an exhaustive list of successful sampling sites. Similar to observations in the terrestrial environment, the majority of compounds isolated from the marine environment still originate from *Streptomyces* strains, which remains to date the most abundant source of novel chemistry (Bérdy, 2012; Carroll et al., 2020; Patridge et al., 2016). In fact, as recently as 2018, > 69% of new marine microbe-derived natural products reported were isolated from *Streptomyces* strains (Carroll et al., 2020). Clearly, marine sponges are a tremendous reservoir for not-yet-cultured microbes, and prospects of discovering unique molecular scaffolds with novel mechanisms of action from these marine actinomycetes are favorable.

1.2.3 Current status of marine-derived pharmaceuticals

As of 2019, natural products or their semisynthetic derivatives accounted for over 50% of antibiotics (Newman & Cragg, 2020; Ribeiro et al., 2020). It was not until the 1950s that marine-derived natural products gained recognition. The first marine natural product was discovered from the marine sponge *Tectytethya crypta* in 1951 (Bergmann & Feeney, 1951). *T. crypta* produces two nucleosides – spongothymidine and spongouridine, that eventually became the basis for the first FDA-approved marine-derived anticancer and antiviral drugs, respectively, as well as azidothymidine (AZT), the first treatment for AIDS (Jimenez, 2014; Mayer, 2021). Pharmaceuticals have since been discovered from various marine organisms, including tunicates, mollusks, and fish (Lindequist, 2016; Mayer, 2021). There are currently 12 marine-derived drugs that have received approval by the FDA[2] **(Table 1.1)**. The majority of these drugs exhibit anti-tumor activity and are used to treat various cancers. Additionally, there are approximately 23 compounds in various phases of clinical trials at the moment (Mayer, 2021).

Due to the fact that actinomycetes are such a prominent source for antibiotics, it may be surprising at first glance that none of these FDA-approved drugs originate from their marine isolates. The field of marine drug discovery is still very young. Barely 18 years have passed since the first biosynthetic genes isolated from a marine invertebrate were confirmed to originate from a bacterial symbiont (Piel et al., 2004). Before this major discovery, it had long been suspected that microbial symbionts

---

[2] Vidarabine has since been discontinued as a drug, likely due to its narrow therapeutic window. It is included in Table 1.1 but not included in the tally of currently FDA-approved drugs of marine origin (Mayer et al., 2010).

were the true producers of many natural products being isolated from higher

organisms as evidenced by various studies finding bioactive compound sequestration

within symbiont cells (Bewley & Faulkner, 1998; Bewley et al., 1996; Unson et al.,

1994; Unson & Faulkner, 1993). Yet without confirming the origin of the

biosynthetic pathway, they could only be attributed to the hosts (Bewley & Faulkner,

1998; Piel et al., 2005). Only after it became apparent that compounds isolated from

sponges and animals of entirely different taxa retained very similar structures did

scientists begin to investigate a microbial link between the two (Newman & Hill,

2006; Perry et al., 1988; Piel et al., 2005; Sakemi et al., 1988; Thomas et al., 2010).

Today, many bioactive compounds originally isolated from higher organisms are

suspected to actually be produced by their symbionts (Lindequist, 2016). For

instance, halichondrin B, the compound isolated from the sea sponge *Halicondria*

okadai and the basis for the anticancer drug Halaven® has since been isolated from

various sponge species, strongly indicating a microbial producer, likely a

dinoflagellate (Lindequist, 2016; McCauley et al., 2020; Murakami et al., 1982;

Nakao & Fusetani, 2010). Dolastatin 10, the basis of the anti-cancer drug

brentuximab vedotin, is now known to originate from a cyanobacterium (Lindequist,

2016; Luesch et al., 2001). Subsequent studies have also since definitively assigned

production of anti-cancer ET-743 (Yondelis®) to microbial symbionts of the tunicate

Ecteinascidia turbinata (Lindequist, 2016; Rath et al., 2011). It is encouraging that

one compound, salinosporamide A, derived from a sediment-dwelling marine

actinomycete (*Salinispora tropica*), is among the drugs in phase III clinical trials.

Salinosporamide A appears to be a promising candidate and is currently being

investigated for its inhibitory activity against various cancers including multiple

myeloma and leukemia (Ghareeb et al., 2020; Lindequist, 2016). As recent as 2017,

an extensive review of antibacterial compounds isolated from marine bacteria

between 2010 and 2015 found that an overwhelming majority of these (69%) were

obtained from Actinobacteria (Schinke et al., 2017). With continued efforts to explore

novel marine environments and screen isolated actinomycetes for their diverse

bioactivity, we will unlock the true potential of marine actinomycetes in the fight

against global diseases and illnesses.

Table 1.1. Current FDA-approved marine-derived pharmaceuticals. Adapted
from Marine Pharmacology: Clinical Pipeline (Mayer, 2021) and Lindequist, 2016.
a-(Engene et al., 2015), b-(Rath et al., 2011), c-(Patel et al., 2021).

Drug Name (Trademark)	Year of FDA approval	Origin species	Compound description	Implication for Use
Cytarabine (Cytosar-U®)	1969	Sponge (*Tectytethya crypta*)	Synthetic analog of nucleoside	Anti-cancer – leukemia
Vidarabine (Vira-A®)	1976 *Since discontinued	sponge (*Tectytethya crypta*)	Synthetic analog of nucleoside	Anti-viral – herpes simplex and varicella zoster virus
Ziconotide (Prialt®)	2004	Cone snail (*Conus magus*)	Synthetic analog of omega-conotoxin	Analgesic – severe chronic pain
Omega-3-acid ethyl esters (Lovaza®)	2004	Fish oil *now cultured from microalgae	Poly-unsaturated fatty acid	Hypertriglyceridemia
Eribulin mesylate (Halaven®)	2010	Various Sponges/ dinoflagellate	Synthetic analog of halichondrin B	Anti-cancer – Metastatic breast cancer
Brentuximab vedotin (Adcetris®)	2011	Mollusk (*Dolabella auricularia*)/ cyanobacterium (*Caldora penicillata*)[a]	Synthetic analog of cytotoxic peptide dolastatin 10	Anti-cancer- Certain cases of Hodgkin lymphoma
Eicosapentaenoic acid ethyl ester (Vascepa®)	2012	Fish oil *now cultured from microalgae	Poly-unsaturated fatty acid	Hypertriglyceridemia

Table 1.1, continued

Omega-3-carboxylic acid (Epanova®)	2014	Fish oil *now cultured from microalgae	Poly-unsaturated fatty acid	Hypertriglyceridemia
Trabectedin (Yondelis®)	2015	Isolated from Tunicate (*Ecteinascidia turbinata*)/ Gammaproteobacteria (*Candidatus* Endoecteinascidia frumentensis)[b]	Alkaloid	Anti-neoplastic – soft-tissue sarcoma and ovarian cancer
Polatuzumab vedotin (Polivy™)	2019	Mollusk/ cyanobacterium	Antibody-drug conjugate	Anti-cancer – Various lymphomas including Non-Hodgkin lymphoma
Enfortumab vedotin-ejfv (PADCEV™)	2019	Mollusk/ cyanobacterium	Antibody-drug conjugate	Anti-cancer – Metastatic urothelial cancer
Lurbinectedin (Zepzelca™)	2020	Tunicate (*Ecteinascidia turbinata*)/ Gammaproteobacteria (*Candidatus* Endoecteinascidia frumentensis)	Synthetic alkaloid derivative of trabectedin[c]	Anti-cancer – Metastatic small cell lung cancer
Belantamab mafodotin-blmf (Blenrep™)	2020	Mollusk/ cyanobacterium	Antibody-drug conjugate	Anti-cancer – Relapsed/refractory multiple myeloma

1.3 Elucidating the genetic basis for natural products

1.3.1 Barriers to unlocking the full potential of microbial natural products

The classical screening approach for identifying natural products produced by actinomycetes begins by cultivating the isolates on agar plates to isolate individual strains. This method resulted in the immense number of compounds discovered during the "Golden Era" in the mid 20th century, but paradoxically also led to reduced interest as high rates of replication were encountered and discovery rates dramatically decreased. Even with the newfound interest in marine strains, there was concern that

19

the common cultivation techniques were not identifying the total community of actinomycetes present in a sample. Scientists had long since known that despite the high counts of viable cells observed under the microscope, the number of colonies observed to form on agar places was orders of magnitudes less (Jannasch & Jones, 1959; Winterberg, 1898). This phenomenon, known as the "great plate count anomaly", is observed in samples from all types of habitats – terrestrial, aquatic and marine (Staley & Konopka, 1985; ZoBell, 1946). Eventually it was determined that < 1% of seawater bacterial cells can be observed to form colonies on agar plates by using routine cultivation methods (Amann et al., 1995; Ferguson et al., 1984; Gram et al., 2010). Considering that microbial abundances are on average 10^6 per mL of seawater and 10^9 per mL in marine sediments, the probable amount of actinomycetes left unsampled is striking (Fenical & Jensen, 2006; Gram et al., 2010; Hentschel et al., 2006). Various innovative techniques have been employed to increase the number of viable microbial cells recovered during cultivation, such as altering incubation temperatures, media composition, carbon source, and more recently the use of diffusion chambers and isolation chips (iChip) (Button et al., 1993; Imhoff & Stohr, 2003; Kaeberlein et al., 2002; Nichols et al., 2010; Rappé et al., 2002). Research at the turn of the century provided evidence that many of these uncultivable bacteria grow very slowly and can be isolated in pure culture using extremely dilute, nutrient-poor media (Olson, Lord & Mccarthy, 2000; Watve et al., 2000). Despite commendable advances, there still remains an inevitable recovery bias, as no single media composition can equally support the growth of all microbes present in a given community (Foot & Taylor, 1949; Francisco et al., 1973; Koch, 1881; Roszak &

Colwell, 1987). Furthermore, it has been concluded that many of the bacteria present from aquatic and marine environments are isolated in a state known as viable but nonculturable, induced by a variety of stressors and lack of nutrients in their natural habitat (Xu et al., 1982).

Aside from the laborious efforts to culture this "hidden majority" of microbes, another equally frustrating phenomenon has hindered the drug discovery process. It is very common for an actinomycete strain to present an extract with interesting bioactivity that cannot be observed after repeated cultivation (Gross, 2009; Katz & Baltz, 2016). Again, various manipulations of growth conditions have been tested to promote production of the bioactive compound(s) within a given strain including medium composition (carbon, nitrogen, sulfur and phosphorus sources), salinity, temperature, pH, oxygen levels, addition of metal ions and trace elements, addition of a physical scaffold such as cotton or talc, as well as co-cultivation with other bacteria or fungi (Pan et al., 2019; Romano et al., 2018; Rutledge & Challis, 2015; Tomm et al., 2019). These strategies are part of what is known as the "one strain many compounds" (OSMAC) approach (Bode et al., 2002; Romano et al., 2018). The goal of the OSMAC approach is to identify the parameters that mimic the biotic and abiotic conditions of the natural environment to stimulate secondary metabolite production, yet even so the culture may fail to generate reproducible activity. To understand why, it is necessary to briefly review when and why actinomycetes produce these metabolites in the first place. Secondary metabolites are generated during the stationary phase of growth when nutrients and space become limited (Bibb, 2005; Chakraburtty & Bibb, 1997). Although not all actinomycetes undergo

sporulation, production of secondary metabolites typically coincides with the initiation of spores in strains that do (Adamidis et al., 1990; Chater et al., 2002; Goodfellow & Williams, 1983; van der Meij et al., 2017). During this time, the bacterium is undergoing a multitude of changes in gene expression to survive the unfavorable conditions. Therefore, the purpose of testing various cultivation conditions is to trigger expression of particular genes responsible for the biosynthesis of secondary metabolites that would otherwise remain "silent". When these purportedly simple cultivation manipulation techniques fail, more advanced and complicated genomic modification can be performed, either by editing the associated biosynthetic genes directly within the original producing strain (ex. ribosome engineering, biosynthetic gene cluster (BGC) refactoring using clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) to insert constitutive promoters, reported-guided mutant selection) or through heterologous expression in a more genetically tractable host (Mao et al., 2018; Ochi & Hosaka, 2013; Rutledge & Challis, 2015; Tomm et al., 2019). However, these experiments are much more complex and still require cultivation. An investigative method that offers direct access to these genes and curtails the need for extensive cultivation efforts would dramatically hasten the pace at which novel bioactive compounds can be identified and developed into pharmaceuticals.

1.3.2 Genomics as an alternative approach to drug discovery and the genetic patterns of natural products

Microbial natural products are defined as a group of highly diverse and specialized secondary metabolites of low molecular weight [< 3000 daltons (Da)] that exhibit a wide range of biological activities (Bérdy, 2005). They can be classified into various categories based on their chemical structure, the most common being polyketides (PKs), nonribosomal peptides (NRPs), ribosomally synthesized and post-translationally modified peptides (RiPPs), terpenoids and alkaloids (Hug et al., 2020; Medema et al., 2015). Directly studying the genes responsible for synthesizing these natural products enables scientists to assess the full biosynthetic potential of specific microbes despite the inability to culture the bacteria or lack of production of a particular product in the laboratory. This powerful approach is achievable by virtue of the fact that several classes of natural products are encoded by genes laid out in a recognizable pattern collectively known as a biosynthetic gene cluster (BGC) (Martín & Liras, 1989). The assembly-line nature of their synthesis facilitates their detection through genome-mining efforts (Van Lanen & Shen, 2006; Zerikly & Challis, 2009).

Both PKs and NRPs are synthesized by large, multimodular polypeptides, known respectively as polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs). While the molecular precursors differ between the two NPs, both mechanisms of synthesis rely on a collection of modules consisting of three core domains. These domains are responsible for the loading of molecular building blocks and elongation. In the case of PKs, malonyl-coenzyme A (CoA) and methylmalonyl-CoA serve as the main molecular precursors of the final product. Synthesis begins

with a 50 kDa acyltransferase domain (AT), which is responsible for selecting the carboxylic acid precursor and transferring it to the adjacent acyl carrier protein (ACP) domain (Donadio et al., 2007; Fischbach & Walsh, 2006). A phosphopantetheinyltransferase (PPTase) must first post-translationally modify a side chain serine (Ser) residue of the inactive apo-ACP domain by addition of a thiol-terminated phosphopantetheinyl arm from CoA to convert it into a functional holoenzyme (Lambalot et al., 1996). Once active, the 8-10 kDa ACP domain covalently tethers the acyl precursor unit through a thioester bond. Essentially, the ACP serves as a tethering site for reaction intermediates, which are ultimately linked together by 45 kDa ketosynthase (KS) domains. Oligomerization is carried out by the KS domain through decarboxylative condensation between the acyl thioester unit on the ACP domain of the same module and the reaction intermediate tethered to the ACP domain of the upstream (preceding) module. This template-directed elongation continues in an assembly-like manner for every module present in the BGC other than the first module (Donadio et al., 2007; Fischbach & Walsh, 2006). Although the first PKS module, known as the loading module, often lacks a KS domain, it has been reported to be present in some starting modules with the function of decarboxylating the initial activated monomer (Bisang et al., 1999; Fischbach & Walsh, 2006). The final product is eventually cleaved through hydrolysis or intramolecular macrocyclization by an additional 35 kDa thioesterase (Te) domain present in the final module of the PKS. Alternatively, a reductase (Re) domain can cleave the final PK by reducing the thioester to an aldehyde. Aside from the aforementioned core domains, accessory domains are located throughout modules for additional

processing, such as ß-ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains (Donadio et al., 2007; Fischbach & Walsh, 2006). In sum, the core sequence of a PKS consisting of n total modules is AT-ACP-(KS-AT-ACP)$_{n-2}$-KS-AT-ACP-Te, with accessory domains scattered in between.

Similar to the structure of PKs, the composition of NRPs is also dictated by the order of modules, known as the colinearity rule. For NRPs, over 500 different substrates including 20 proteinogenic amino acids, nonproteinogenic amino acids, and aryl acids can serve as monomers (Bloudoff & Schmeing, 2017; Fischbach & Walsh, 2006; Kleinkauf & von Döhren, 1990). Analogous to the AT domain of the PKS, NRPSs usually begin with an adenylation (A) domain. The ~500 amino acid (50 kDa) long A domains maintain 8-10 residues in their binding pocket that typically display specificity towards an amino acid substrate (Bloudoff & Schmeing, 2017; Donadio et al., 2007; Hur et al., 2012; Stachelhaus et al., 1999). Likewise, the amino acid sequence of the AT domain of PKSs has been determined to specify for malonyl-CoA or methylmalonyl-CoA (Haydock et al., 1995). This substrate specificity has been exploited by genome mining efforts to link the protein sequence of A/AT domains with putative amino acid/carboxylic acid residues present in the final NRP/PK (Challis, 2008; Challis et al., 2000; Stachelhaus et al., 1999). The amino acid is activated through adenylation and subsequently transfer to the ~80-100 amino acid (8-10 kDa) peptidyl carrier protein (PCP) domain, also known as the thiolation (T) domain. The adenylated monomer covalently attaches to the thiol terminus of the holo-T domain, activated by PPTase in the same manner as the ACP domain of PKSs. The ~450 amino acid (50 kDa) condensation (C) domain retains the same general

25

function as the KS domain, but catalyzes peptide bonds between adjacent aminoacyl

substrates (Bloudoff & Schmeing, 2017; Challis & Naismith, 2004; Donadio et al.,

2007; Hur et al., 2012). The C domain also exhibits substrate specificity, albeit to a

lesser extent than the A domain (Belshaw et al., 1999; Challis & Naismith, 2004).

The loading or initiation module of an NRPS often lacks a C domain, although in

some cases the C domain is retained and functions to N-acylate the initial amino acid

monomer (Fischbach & Walsh, 2006; Miao et al., 2005; Rausch et al., 2007). The

final peptide is released by the ~30 kDa Te domain through either hydrolysis or

cyclization. In some cases, the Te domain will catalyze oligomerization of the peptide

through nucleophilic attack, which must undergo further processing before final

release of the NRP (Bloudoff & Schmeing, 2017; Hoyer et al., 2007). As with PKSs,

additional tailoring enzymes are often present throughout the NRPS BGC, including

domains for epimerization (E), heterocyclization (Cy), oxidation (Ox), reduction

(Re), N- and C-methyltransferase (MT), formylation (F) and halogenation. Cy

domains can actually replace C domains, performing condensation and subsequent

heterocyclization of a Ser, cysteine (Cys) or threonine (Thr) residue to form a

thiazoline or oxazaline ring (Bloudoff & Schmeing, 2017; Challis & Naismith, 2004;

Donadio et al., 2007; Hur et al., 2012; Konz et al., 1997). Ultimately, the core

sequence of an NRPS consisting of n total modules is A-T-(C-A-T)$_{n-2}$-C-A-T-Te,

with the aforementioned accessory domains scattered among the modules.

The RiPP class of NPs has recently gained attention as another group that can be

efficiently mined within genomes using computational methods. RiPPs are ubiquitous

among all three domains of life and span a wide range of compounds with structural

diversity, but recent efforts to establish a universal nomenclature designate peptides up to 10 kDa with extensive post-translation/co-translational modifications as fitting within this class (Arnison et al., 2013). RiPPs begin as an unmodified precursor peptide ~20 -110 residues in length, with an N-terminus leader peptide and a C-terminal core peptide. The leader sequence is recognized by editing enzymes that heavily modify the core region before it is eventually cleaved, leaving behind the final mature core peptide. The leader sequence additionally is believed to play a role in export and immunity (Arnison et al., 2013). Although similar in peptide origin to NRPs and capable of undergoing identical modifications (ex. heterocyclization of Cys, Ser and Thr residues, macrocyclization, formylation), RiPPs retain significant advantages as potential drug leads as a result of their extensive modification (McIntosh et al., 2009). Thorough modifications generate a final product with restricted conformational flexibility, improved target recognition, reduced toxicity due to less off target effects, increased chemical stability, and increased resistance against degradation by exoproteases (Arnison et al., 2013; Cotter et al., 2013; McIntosh et al., 2009; Oman & Donk, 2010). In addition to these favorable qualities in a therapeutic agent, RiPPs are known to carry out multiple mechanisms of action simultaneously, often on highly conserved targets, thus further reducing the potential that a pathogen will develop resistance to the compound (Breukink et al., 1999; Cotter et al., 2013; Poorinmohammad et al., 2019). The fact that the leader peptide is singularly responsible for recruiting enzymes to modify the core peptide has enabled the core region to be hypervariable. Acquisition of new post-translational

modification enzymes has resulted in extensive evolution of RiPP structures and

functional diversity (Arnison et al., 2013; Oman & Donk, 2010).

Due to their similar coordinated mechanisms of synthesis, it is no surprise that

hybrid PKS-NRPS clusters and NRPS-RiPPs are commonly found in microbial

genomes (Baltz, 2016; Cimermancic et al., 2014; Donadio et al., 2007; Du et al.,

2000; Fischbach & Walsh, 2006; Medema & Fischbach, 2015; Udwary et al., 2007;

Zhang et al., 2015). Various algorithms (see Chapter 2 for a detailed review of these

programs) have been established to recognize common domains and clusters essential

for the synthesis of PKs, NRPs and RiPPs, as well as hybrid clusters, enabling rapid

examination of biosynthetic potential of actinomycetes. Despite the great

advancements in genome mining facilitated by template-directed elongation systems

of PKs and NRPs, significant caveats exist to the biosynthetic patterns described

above (Mootz et al., 2002; Shen, 2003; Wenzel & Müller, 2005). The cis form of the

KS-AT-ACP domains and C-A-T domains are in fact noted as type I PKSs and type

A NRPSs, respectively. Type I PKSs and type A linear NRPSs follow the modular

structure that is readily detected based on conserved motifs (De Crecy-Lagard et al.,

1995; Fischbach & Walsh, 2006; Mootz et al., 2002). However, type II PKSs are

nonmodular and distinguished by a trans organization of domains in which distinct

protein subunits iteratively catalyze addition of individual monomers to the growing

NP (Fischbach & Walsh, 2006; Sherman et al., 1989). Type C nonlinear NRPSs also

maintain a nonmodular structure in which domains may be used multiple times in

biosynthesis of a product as well as inserted out of the typical order, and therefore

modular arrangement has no correlation with the amino acid sequence of the final

NRP (Hur et al., 2012; Mootz et al., 2002). Type III PKSs and type B iterative NRPSs utilize all of their modules or domains multiple times for loading, oligomerizing subunits, and cyclization (Fischbach & Walsh, 2006; Hur et al., 2012; Mootz et al., 2002). More specifically, type III PKSs are homodimers that lack ACP domains and act directly on the malonyl-CoA starter units to carry out elongation (Fischbach & Walsh, 2006; Funa et al., 1999; Shen, 2003). The diversity in biosynthetic structure of PKSs extends even further, but that is beyond the scope of this discussion (see Shen 2003 for an extensive review).

Ultimately, these variations on the assembly-line structure underscore the barriers that still remain to studying and identifying unique, highly evolved pathways within a genome. Deviations from the classical BGC template cannot be as easily recognized by mining techniques, and novel pathways may go undetected as a result (Bachmann and Ravel, 2009; Baral, Akhgari and Metsä-Ketelä, 2018; Blin et al., 2019a; Challis, 2008; Wohlleben et al., 2016). It is therefore often necessary to employ chemical analysis to resolve the entire BGC. If structure determination can be achieved, chemistry can be cross referenced with genomic sequence to fully elucidate the biosynthetic pathway of the novel compound. Various techniques such as nuclear magnetic resonance (NMR), high-performance liquid chromatography (HPLC) and mass spectrometry (MS) are employed to isolate the bioactive compound and resolve its structure (Ebada et al., 2008; Matroodi et al., 2020; Motohashi et al., 2007; Peng, 2000; Riyanti et al., 2020; Sun et al., 2015). Identifying the functional groups present in the chemical structure by these methods can facilitate detection of corresponding BGC genes using bioinformatics.

1.3.3 Enhancing drug discovery efforts through genome mining

Provided the genomic DNA of a strain, we can now use an array of algorithms to detect putative BGCs in any bacterial isolate. The benefits of this newfound capability are several fold. First, *in silico* genome mining facilitates BGC analysis irrespective of a bacterium's amenability to laboratory culturing. This eliminates the need for extensive, time-consuming cultivation and dramatically reduces the resources necessary to carry out these studies. Even in cases where a strain can be cultured in the laboratory and consistently produce a particular compound, it is very possible for relevant bioactivity to go undetected by standard bioassays if production of the compound is too low (Rebets et al., 2014; Rutledge & Challis, 2015; Scherlach & Hertweck, 2021; Zhang et al., 2017). Bioinformatic analysis of genomic DNA enhances the likelihood that all bioactive secondary metabolites that a particular strain is capable of producing will be recognized, especially elucidating those linked to cryptic clusters. Secondly, quick detection of BGCs by computational methods enables rapid dereplication, avoiding the unfortunate but common occurrence of reisolating an already discovered compound (Baltz, 2005; Woodruff & McDaniel 1958). Quick assessment of the putative BGC repertoire of a genome enables prioritization of strains for further analysis (Ward & Allenby, 2018; Zhang et al., 2017). Aside from assessing the biosynthetic potential of newly isolated strains, former strains can also be reanalyzed to determine whether bioactivity was overlooked by classical screening approaches (Bentley et al., 2002; Cruz-Morales et al., 2016; Ikeda et al., 2003; Ōmura et al., 2001; Ward & Allenby, 2018).

Reassessment of strains previously investigated by bioactivity-guided screening has led to the discovery of various antibiotics such as the broad spectrum antibacterial agent ECO-0501 from *Amycolatopsis orientalis*, the potent antifungal agent ECO-02301 from *Streptomyces aizunensis* NRRL B-11277 and the broad-spectrum anti-oomycete and herbicidal agent phthoxazolin A from *Streptomyces avermitilis* (Banskota et al., 2006; McAlpine et al., 2005; Suroto et al., 2018).

As previously stated, genome mining is not sufficient as a stand-alone tool to discover novel compounds, especially if the corresponding pathways are highly evolved from the typical patterns that existing algorithms were developed to detect (Bachmann and Ravel, 2009; Baral, Akhgari and Metsä-Ketelä, 2018; Blin et al., 2019a; Challis, 2008; Medema and Fischbach, 2015; Wohlleben et al., 2016). However, provided current capabilities, this approach maximizes the chances of discovering novel compounds with relevant bioactivity in the most efficient way possible. Various mining techniques have been developed to address this shortcoming, such as EvoMining, which incorporates evolutionary principles into the detection of enzyme families repurposed from primary metabolism in non-standard biosynthetic pathways (Cruz-Morales et al., 2016). ClusterFinder can detect BGCs of unknown classes by utilizing a hidden Markov model-probabilistic algorithm to determine the probability that a Pfam domain belongs to a BGC by comparison of the frequency at which the Pfam domain appears in clusters implicated in secondary metabolism versus other areas of the genome. ClusterFinder relies on patterns of Pfam domain distribution in known BGCs and searches for similarly rich regions of Pfam domains instead of searching for specific domains. (Cimermancic et al., 2014;

Medema & Fischbach, 2015). Combining traditional genome mining algorithms with these more sophisticated "low-confidence/high-novelty" approaches (coined by Medema & Fischbach, 2015) will provide a more robust analysis of a bacterium's biosynthetic capabilities and promote efficient prioritization of strains likely to produce novel antibiotics.

1.3.4 Prevalence of biosynthetic gene clusters in actinomycetes

The field of genome mining for BGCs began in the early 2000s, after sequencing of the first genomes of two antibiotic producing actinomycete strains *Streptomyces coelicolor* and *Streptomyces avermitilis* was completed (Bentley et al., 2002; Ikeda et al., 2003). Genome annotation uncovered more than 20 putative BGCs encoded by each of these two strains, revealing a significant portion of natural products that had previously gone undetected by traditional analysis (Bentley et al., 2002; Ikeda et al., 2003; Ōmura et al., 2001). Since then, it has been estimated that actinomycetes reserve a significant portion of their genome (up to 10%) for the synthesis of secondary metabolites, although 90% of these predicted clusters are cryptic (Baltz, 2008, 2016). BGC detection is typically rare or altogether absent in genomes below 3 megabases (Mb), but a linear correlation between genome size and BGC count appears in genomes larger than 5 Mb (Donadio et al., 2007). BGCs are not evenly distributed among microbes, yet frequent detection in actinomycetes has encouraged focused efforts on this group of bacteria as a promising source of novel activity and structures (Baltz, 2008; Donadio et al., 2007). The collective group of actinomycetes are highly diverse, spanning species with sequenced genomes from less than 1 Mb (*Tropheryma whipplei*) to 13.1 Mb (*Nonomuraea* sp. ATCC 55076)

(Donadio et al., 2007; Nazari et al., 2017). It has been observed that actinomycete genomes > 8 Mb are capable of harboring over 30 putative BGCs (Baltz, 2016). A revisit of the genome of *Streptomyces bingchenggensis* BCW-1 with advanced bioinformatic analysis in 2016 by Baltz discovered that the 11.94 Mb genome codes a whopping 53 putative BGCs, most of which are attributed to NRPS, PKS or hybrid NRPS/PKS systems.

The pervasiveness of BGCs throughout actinomycete genomes excited scientists eager at the prospect of discovering novel biosynthetic pathways. Interestingly, detailed studies of distantly related bacteria detected similar or even identical BGCs in some cases (Baltz, 2005; Donadio et al., 2007; Fischbach et al., 2008; Humisto et al., 2018; Piel et al., 2004; Vior et al., 2018). It has been estimated that streptomycin, the first commercial antibiotic isolated from an actinomycete, is produced by 1% of soil actinomycetes (Baltz, 2005; Woodruff & McDaniel, 1958). This recurring observation suggested that these microbes can acquire new clusters through horizontal gene transfer (HGT), further expanding their potential repertoire (Donadio et al., 2007; Guerrero-Garzón et al., 2020; Lawrence & Roth, 1996; Ziemert et al., 2014). It has been suggested recently that HGT in *Streptomyces* does not occur as frequently as once believed, and therefore cannot account for the majority of genetic diversity observed within the genus. A comprehensive phylogenomic study of 122 *Streptomyces* genomes concluded that over a one million year time span, the genus acquires only ten genes through HGT, while it accumulates ~ 23,000 point mutations. However, of the genes that are acquired through HGT, the majority are genes related to secondary metabolism. Ninety-three percent of the BGCs analyzed in

33

the study obtained at least one gene through HGT within the last 50 million years, most composed of genes from multiple sources (Chase et al., 2020; McDonald & Currie, 2017). In addition to HGT, mediated by the exchange of genomic islands, transposons or plasmids between species, point mutations, module or gene duplications, deletions, recombination, domain rearrangement, and transposition are all methods by which biosynthetic pathways can evolve and diversify the existing set of natural products that actinomycetes are capable of producing (Baltz, 2008; Chase et al., 2020; Fischbach et al., 2008; Jenke-Kodama et al., 2006; Ziemert et al., 2014). By 2005, almost 10,000 antibiotics had been documented from Actinobacteria, although some estimates guess that they are capable of producing more than 100,000 antimicrobial compounds (Bérdy, 2005; Piel, 2011; Subramani & Aalbersberg, 2013; Watve et al., 2001). With rapid evolution evident, this number will only continue to rise. This ever-expanding collection of BGCs coupled with the high probability of finding unique molecular scaffolds in isolated or marine environments ensures that continued investigation into marine actinomycetes will lead to new clinical drugs.

1.4 Focus and objectives

This research hypothesizes that detailed investigation of an assemblage of novel sponge-derived marine actinomycetes will result in the discovery of compounds that inhibit *M. tb*. After preliminary screening of microbial extracts tested against *Mycobacterium* spp. identified extracts that consistently inhibited growth of *Mycobacterium* spp., further analysis involving both genomics and chemistry was performed to determine the compounds responsible for the inhibitory activity.

Chapter 2 focuses on assessing the biosynthetic potential of all actinomycete strains that were observed to inhibit *M. tb.* A comparative analysis testing three short-read assembly algorithms was performed to determine which method yields the most complete genome. The strengths and weaknesses of each assembly algorithm are evaluated, and the results of genome mining with several tools for each assembly are discussed. This analysis ultimately established a pipeline to rapidly assemble actinomycete genomes *de novo* and assess their biosynthetic capacity, which can then be used to prioritize strains for further analysis.

Chapter 3 investigates a *Micrococcus* isolate shown to have very potent anti-TB activity from disk-diffusion assays. Mining of genomes assembled with both short and long-read sequencing data did not identify any obvious BGCs encoding anti-TB compounds, so a comprehensive examination of all BGC-related domains with all genome mining tools employed was conducted to characterize the biosynthetic pathway that may be linked to the elusive compound with anti-TB activity. This analysis is speculative and demands further investigation with chemistry analytical techniques to identify the *Mycobacterium*-inhibiting metabolite. The results are highly promising in regards to the likelihood of identifying a novel bioactive compound.

Chapter 4 concentrates on one strain of *Micromonospora* identified through preliminary screening to inhibit *M. tb.* BGC analysis revealed a cluster with very high similarity to diazaquinomycin, a compound with selective and potent anti-TB activity. Chemical analysis was incorporated into this investigation to determine what

distinguishes this likely novel diazaquinomycin analog from those of known structures.

This research identifies thirteen actinomycete strains that inhibit *M. tb*, at least two of which produce novel compounds with anti-TB activity. These results support continued investigation into marine actinomycetes for pharmaceutically-relevant drugs.

Chapter 2: Comparative analysis of assembly algorithms to optimize biosynthetic gene cluster identification in novel actinomycete genomes

2.1 Abstract

Many marine sponges harbor dense communities of microbes that aid in the chemical defense of these nonmotile hosts. The metabolites that comprise this chemical arsenal have also been discovered to have pharmaceutically-relevant activities such as antibacterial, antiviral, antifungal and anticancer properties. Previous investigation of the Caribbean giant barrel sponge *Xestospongia muta* by Montalvo et al. (2005) revealed a microbial community dominated by novel Actinobacteria, a phylum well known for its production of antibiotic compounds. This novel assemblage was investigated for its ability to produce compounds that inhibit *M. tuberculosis* (*M. tb*) through a bioinformatics approach. Microbial extracts were tested for their ability to inhibit growth of *M. tb* and corresponding genomes of the 11 strains that showed anti-*M. tb* activity including *Micrococcus* (n=2), *Micromonospora* (n=4), *Streptomyces* (n=3) and *Brevibacterium* spp. (n=2) were sequenced by using Illumina MiSeq. Three assembly algorithms/pipelines (SPAdes, A5-miseq and Shovill) were compared for their ability to construct contigs with minimal gaps to maximize the probability of identifying complete biosynthetic gene clusters (BGCs) present in the genomes. Although A5-miseq and Shovill usually assembled raw reads into the fewest contigs, after necessary post-assembly filtering, SPAdes ultimately produced the most complete genomes with the fewest contigs in almost all cases. This study revealed the strengths and weaknesses of the different assemblers based on their ease of use and ability to be manipulated based on output format. It was concluded that none of the assembly methods handle contamination well and that high-quality DNA is ultimately necessary to produce a more complete

genome. Various BGCs of compounds with known anti-tuberculosis (TB) activity were identified in all *Micromonospora* and *Streptomyces* strains (genomes > 5 Mb), while no relevant BGCs were identified in *Micrococcus* or *Brevibacterium* strains (genomes < 5 Mb). The majority of the putative BGCs identified were located on contig edges, emphasizing the inability of short-read assemblers to resolve repeat regions and supporting the need for long-read sequencing to fully resolve BGCs.

2.2 Introduction

Marine sponges are found in all parts of the ocean, ranging from warm, shallow tropical waters to polar waters and even the deep ocean (Hentschel et al., 2006; Hooper & van Soest, 2002). Similarly, actinomycetes are found in a wide range of terrestrial, freshwater and marine environments (van der Meij et al., 2017). Marine sponges can harbor a huge amount and remarkable diversity of microbial symbionts both intracellularly and extracellularly within the mesohyl matrix (Hentschel et al., 2002; Vacelet, 1975; Vacelet & Donadey, 1977). As previously mentioned (Chapter 1), there is evidence that sponges actively select for the presence of species-specific symbiotic microbial communities (Hentschel et al., 2006; Lee et al., 2010; Montalvo & Hill, 2011; Taylor et al., 2004, 2007). Although a minimal core community has been identified to exist among sponges of various species and habitats, the majority of the microbial diversity present within a given sponge host is species-specific (sponge-specific or monophyletic), and generally distinct from the microbial diversity present in the surrounding seawater as evidenced by both culture-dependent and independent studies (Hentschel et al., 2002; Hill et al., 2006; Lee et al., 2010; Montalvo & Hill, 2011; Schmitt et al., 2012; Taylor et al., 2004, 2005, 2007; Webster & Hill, 2001; Wilkinson, 1978). Furthermore, studies assessing the intra- and interspecies variability of these microbial communities over distant geographic locations and temporal changes have found striking stability in at least a subpopulation of these communities (Montalvo & Hill, 2011; Schmitt et al., 2012; Taylor et al., 2004, 2005; Webster et al., 2004) . Depending on the density of symbiotic bacteria harbored by sponges, they can be classified as being either of low microbial abundance (LMA) or

high microbial abundance (HMA, aka 'bacteriosponge') (Hentschel et al., 2003; Reiswig, 1981; Vacelet & Donadey, 1977). LMA sponges have a similar bacterial density to that of the surrounding seawater (10^5-10^6 bacteria per gram of sponge wet weight), while HMA sponges can harbor a density of 10^8-10^{10} bacteria per gram of sponge wet weight (Hentschel et al., 2003; Reiswig, 1981). As these bacterial members constitute such a significant portion of their host by volume, it is not surprising that they play a critical role in nutrient cycling, of which actinomycetes contribute to nitrogen and phosphorus cycling as well as decomposition of organic material (Goodfellow & Williams, 1983; Hentschel et al., 2006; Sabarathnam et al., 2010; Weigel & Erwin, 2017; Zhang et al., 2019b).

The largest known sponge species, fittingly named the giant barrel sponge, *Xestospongia muta* (Schmidt, 1870) is an HMA sponge commonly found in coral reef communities in the Caribbean (Hentschel et al., 2006; McMurray et al., 2008). In its natural environment (ranging from Florida to Brazil), *X. muta* has been reported to reach up to 2 m in height and is believed to be long-lived, with some individuals estimated to be anywhere from 100 to 1,000 years old (Hechtel, 1983; Hentschel et al., 2006; Kerr & Kelly-Borges, 1994; McMurray et al., 2008; Montalvo et al., 2005; van Soest, 1980). Previous analysis by Montalvo et al. (2005) found Actinobacteria to dominate the microbial community composition of *X. muta* individuals isolated from the Florida Keys and identified a novel assemblage of actinobacteria isolates. Given that actinomycetes are such a prominent source of antibiotics, this novel assemblage was investigated for its potential to produce compounds with antimycobacterial activity with a primary focus on inhibition of *M. tb*. After screening microbial

extracts, a bioinformatics approach was undertaken to identify the biosynthetic gene clusters (BGCs) most probably associated with any compounds responsible for observed anti-tuberculosis (TB) activity, as well as to assess the full biosynthetic potential of interesting strains.

The GC-rich nature of actinomycete genomes makes sequencing and assembly very challenging (Rajwani et al., 2021). To optimize the success of this genome mining strategy, three assembly pipelines were compared for their particular ability to efficiently assemble reads and minimize gaps in actinomycete genomes. Based on successful strategies described in the literature, the following pipelines were evaluated: SPAdes (Bankevich et al., 2012), Shovill (Seemann, 2020) and A5-miSeq (Coil et al., 2015; Tritt et al., 2012) (Bellassi et al., 2020; Blackwell et al., 2021; Durrell et al., 2017; Egidi et al., 2017; Kincheloe et al., 2017; Klein et al., 2016; Koenigsaecker et al., 2016; Schorn et al., 2016; Soldatou et al., 2021; Tarlachkov et al., 2021). All three pipelines assemble raw Illumina data with slightly different processing steps. SPAdes involves four major stages: assembly graph construction, k-mer adjustment, paired assembly graph construction, and contig construction (Bankevich et al., 2012). The Shovill pipeline utilizes SPAdes genome assembler to assemble reads but involves modified pre- and post-assembly steps, and is optimized for smaller genomes (Seemann, 2020). The A5-miSeq workflow consists of five steps: read cleaning, contig assembly, crude scaffolding, mis-assembly correction, and final scaffolding. Similar to SPAdes, all steps are automated, but what makes this pipeline unique is the fact that all parameters are fixed, without the option to adjust (Coil et al., 2015). The following programs were employed post-assembly to detect

putative BGCs: antiSMASH (antibiotics & Secondary Metabolite Analysis Shell)

(Blin et al., 2019b) and NP.Searcher (Li et al., 2009). The most comprehensive of

these programs, antiSMASH, can detect biosynthetic genes associated with more than

20 natural product classes including polyketides (PKs), nonribosomal peptides

(NRPs), terpenes, and bacteriocins (Medema et al., 2011; Medema & Fischbach,

2015). The antiSMASH program uses Hidden Markov Models (HMMs) via HMMer3

to detect possible clusters through alignment of the translated nucleotide sequence

with proteins or protein domains determined to be exclusively present in particular

BGCs and maintains an extensive database of known BGCs to facilitate comparative

cluster analysis. NP.Searcher utilizes Basic Local Alignment Search Tool (BLAST)

to rapidly scan genomes and determine substrate specificities of adenylation (A) and

acyltransferase (AT) domains in nonribosomal peptide synthetases (NRPSs) and

polyketide synthases (PKSs), respectively. This program also conveniently provides

the Simplified Molecular Input Line Entry Specification (SMILES) output of

expected products and links results with programs that predict the associated 2D and

3D chemical structures of the natural product (Li et al., 2009).

2.3 Materials and methods

2.3.1 Cultivation of actinomycetes and *Mycobacterium* strains

Strains were previously isolated from *X. muta* samples collected at a depth of

20 m collected by SCUBA at Conch Reef, Key Large, Florida in July 2001 and June

2004 (latitude 24°56.82' N, longitude 80°27.40' W). Strains were isolated and stored

as described by Montalvo et al. (2005). Cryovials of isolates were plated out on both

International *Streptomyces* Project 2 (ISP2) agar (BD-Difco™, Franklin Lakes, NJ, USA) supplemented with 20% salt (granular sodium chloride - J.T. Baker, Phillipsburg, NJ, USA) and Reasoner's 2A Agar (R2A) (BD - Difco™, Franklin Lakes, NJ, USA). ISP2 and R2A were chosen for culturing because they were frequently used by Montalvo (2011) to isolate the actinomycete strains in the original study. Both media are commonly employed in the literature to isolate actinomycetes (Liu et al., 2019; Rangseekaew & Pathom-aree, 2019).

Plates were incubated at 30°C until growth of individual colonies could be observed. Two individual colonies per plate were then transferred to the corresponding medium from which they were initially isolated – 100 mL of ISP2 or Reasoner's 2A broth (R2B) (EZ-Media - Microbiology International, Frederick, MD, USA) – in 250 mL baffled flasks to provide sufficient aeration, and incubated at 30°C with shaking at 150 rpm for a minimum of two weeks, until cultures appeared dense. ISP2 liquid medium was prepared using yeast extract (BD-Bacto™, Franklin Lakes, NJ, USA), dextrose (Fisher Scientific, Hampton, NH, USA), malt extract (MP Bio, Santa Ana, CA, USA) and 20% salt (granular sodium chloride - J.T. Baker, Phillipsburg, NJ, USA).

Mycobacterium tuberculosis H37Ra (avirulent), *Mycobacterium marinum* ATCC 927 and *Mycobacterium smegmatis* MC^2 155 were all plated from cryovials onto Middlebrook M7H10 agar (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% OADC [oleic acid (Sigma-Aldrich, St. Louis, MO, USA), bovine serum albumin fraction V (Roche – Sigma-Aldrich, St. Louis, MO, USA), dextrose (Fisher Scientific, Hampton, NH, USA), catalase (Sigma, St. Louis, MO, USA), sodium

44

chloride (enzyme grade - Fisher Scientific, Hampton, NH, USA)] and subsequently cultured in Middlebrook M7H9 liquid medium (BD-Difco™, Franklin Lakes, NJ, USA) supplemented with 10% ADC (bovine serum albumin fraction V, dextrose, catalase, sodium chloride) and 250 µL Tween 80 (Amresco, Solon, OH, USA). *M. marinum* ATCC 927 and *M. smegmatis* MC^2 155 were incubated at 30°C with shaking at 150 rpm and *M. tuberculosis* H37Ra was incubated at 37°C with shaking at 150 rpm.

2.3.2 Preparation of extracts and antimycobacterial activity assay

After dense growth was evident in actinobacterial cultures, the cultures were extracted with high-performance liquid chromatography (HPLC) Plus grade ethyl acetate (EtOAc) (Sigma-Aldrich, St. Louis, MO, USA). A 1:1 volume of EtOAc was added to each culture and incubated overnight at 30°C with shaking at 150 rpm. The organic phases were dried by rotary evaporation, and the final marcs were dried in gas chromatography (GC) vials using a Savant SpeedVac® PLUS SC210A. The aqueous phases were discarded.

Extracts were dissolved in dimethyl sulfoxide (DMSO) at a concentration of 25 µg/10 µL and 250 µg/10 µL and applied to 6 mm Whatman filter discs to establish a dose response. Discs were applied to plates inoculated with *M. tuberculosis* H37Ra, *M. marinum* ATCC 927 and *M. smegmatis* MC^2 155 at exponential phase. A disc inoculated with 10 µL of DMSO was used as a negative control. All extracts were tested at both concentrations in duplicates. Plates of *M. marinum* ATCC 927 and *M. smegmatis* MC^2 155 were incubated for several days at 30°C and *M. tuberculosis*

H37Ra at 37°C until dense lawn growth and inhibition zones could be observed.

Resulting zones of inhibition were measured by using an illuminated colony counter.

2.3.3 Genomic DNA extraction, identification and whole genome sequencing

Strains were assigned taxonomic classifications after initial isolation, which

were confirmed at the time of this study on the basis of partial 16S ribosomal RNA

(rRNA) gene sequence analysis. DNA was extracted using the UltraClean® Microbial

DNA Isolation Kit (MO Bio Laboratories Inc., Carlsbad, CA, USA). Genomic DNA

was quantified using a Nanodrop 2000 Spectrophotometer (Thermo Scientific,

Waltham, MA, USA). Degenerate primers 27F 5'-AGAGTTTGATCMTGGCTCAG-

3' and 1492R 5'-CGGTTACCTTGTTACGACTT-3' were used to amplify 16S rRNA

gene fragments, and polymerase chain reaction (PCR) amplification was performed

on a PTC-200 Peltier Thermal Cycler (MJ Research, St. Bruno, QC, CA). The PCR

reaction mix consisted of 12.5 µL JumpStart™ REDTaq® ReadyMix™ Reaction

Mix (Sigma-Aldrich, St. Louis, MO, USA), at least 15 ng of DNA template, 1 µL

each of primers 27F and 1492R (10 µM stock) and sterile deionized water for a final

volume of 25 µL. The PCR was programmed to the following protocol: 31 cycles of

denaturation at 95°C for 1 min 30 sec, annealing at 55°C for 1 min 30 sec, and

elongation at 72°C for 1 min 30 sec, followed by a final extension step at 72°C for 7

min. PCR products were separated on a 2% agarose gel to confirm amplification and

purified with ExoSAP-IT™ or ExoSAP-IT™ Express PCR Product Cleanup Reagent

(ThermoFisher Scientific, Waltham, MA, USA). Forward and reverse Sanger

sequences were trimmed and assembled using CLC Main Workbench 7, and the

resulting consensus sequence was compared against the National Center for

Biotechnology Information (NCBI) database with Nucleotide BLAST (BLASTN) to

identify the strain. Sequence errors were corrected manually by visual inspection of

chromatograms.

For genomic sequencing, DNA was sequenced on the MiSeq sequencer

(Illumina) using the MiSeq version 2.4.0.4 Reagent Kit. The Nextera DNA Flex

Library Prep Kit (100 ng DNA) was used to prepare the sequencing libraries for

Brevibacterium sp. strain XM4083and *Micromonospora* sp. strain XM-20-01 (300

cycles each), while the Nextera XT Library Prep Kit (1 ng DNA) was used to prepare

sequencing libraries for the remaining nine strains (2 x 250 bp paired-end reads for a

total of 500 cycles).

2.3.4 Genome assembly, annotation and biosynthetic gene cluster analysis

Assembly was performed using three *de novo* methods: 1) reads were trimmed

using Trimmomatic version 0.30 (Bolger et al., 2014) and assembled using SPAdes

version 3.14.1 (Bankevich et al., 2012); 2) reads were assembled using A5-miseq

version 20160825 assembly pipeline (Coil et al., 2015), or 3) reads were assembled

using Shovill version 1.1.0 (uses SPAdes version 3.14.1), which includes an optional

step to trim adaptors (Trimmomatic version 0.39). Initial assembly statistics were

evaluated with the Quality Assessment Tool for Genome Assemblies (QUAST)

(Gurevich et al., 2013). Contigs were then filtered primarily based on coverage,

followed by match identity after comparison with the Nucleotide collection (nr/nt)

BLAST database. If a contig did not return a BLASTN (Zhang et al., 2000) hit, the

translated nucleotide to protein BLAST (BLASTX) function (Altschul et al., 1997) was performed and the contig was retained if it returned a hit to a protein identified from the expected genus with substantial query coverage and percent identity. Resulting contigs were also aligned to contigs of other genomes sequenced in the same Illumina MiSeq run to identify and remove spillover reads or cross contamination. Manual filtering was performed to remove any additional contigs with dubious coverage (determined cut-off value varied per assembly). PATRIC version 3.5.41 was used to perform genome annotation and to calculate post-filtering statistics (Brettin et al., 2015; Davis et al., 2020). The final assemblies were validated by evaluating contamination and completeness values, calculated using CheckM version 1.0.18 app through KBase (kbase.us) (Parks et al., 2015). Scaffolding was performed with MeDuSa version 1.6 (Bosi et al., 2015). Genomes were scaffolded by comparison to all available complete or nearly complete genome sequences in the NCBI BLASTN database that aligned to the trimmed 16S rRNA gene sequence of the particular strain. If the trimmed 16S rRNA gene sequence did not return any hits to complete genome sequences, the trimmed forward or reverse read was analyzed by BLASTN, and complete genome sequences were selected from the resulting list for scaffolding. Default parameters were used for all software packages. However, for Trimmomatic, a slightly modified script was used that was more sensitive for adapters and also included a sliding window of four bases to scan the reads and remove bases when the average quality per base was below 15. Codes run for each software package are provided in the appendix. BGCs were identified using the following algorithms: antiSMASH version 5.0 (Blin et al., 2019b) in relaxed mode

and NP.Searcher (Li et al., 2009). For a schematic overview of this analysis, see

Figure 2.1.
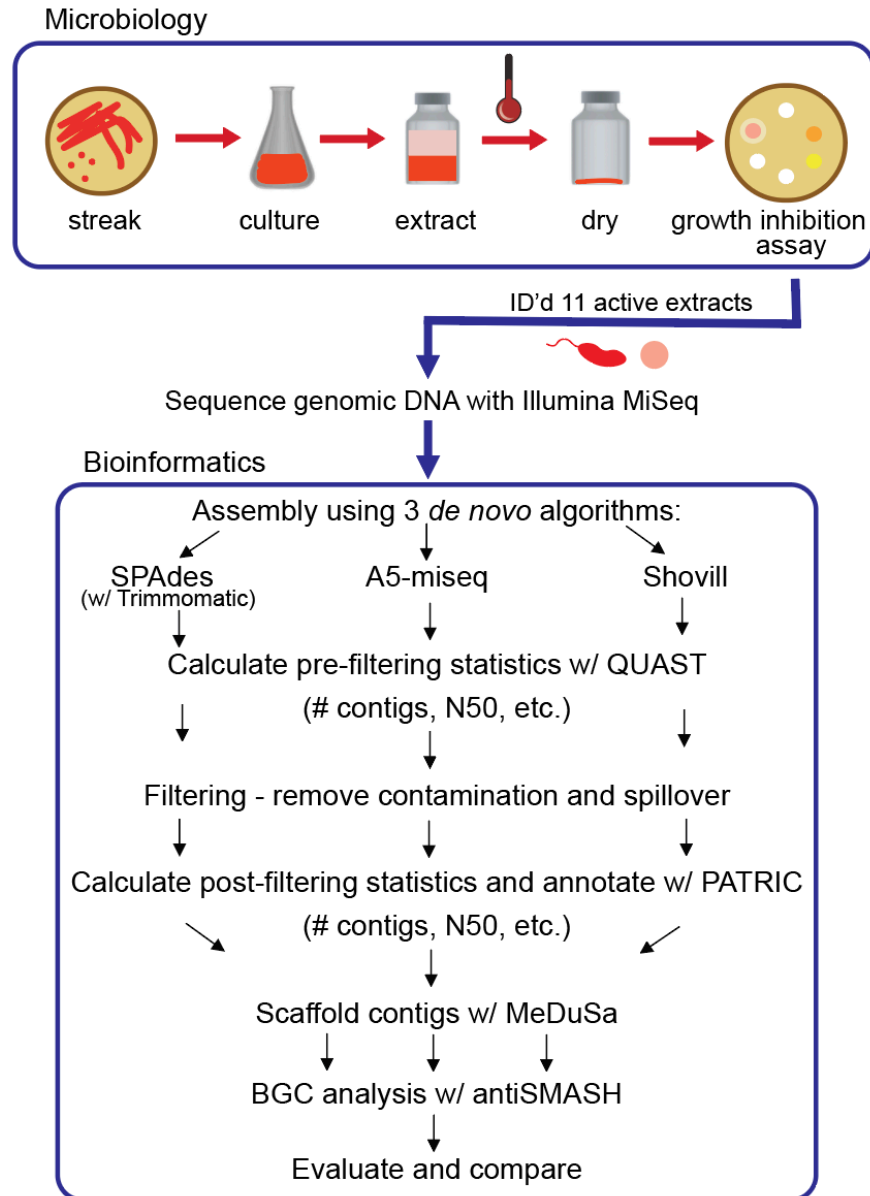
Figure 2.1. Schematic overview of research pipeline. Microbiology procedures and experiments to identify strains producing extracts with inhibitory activity against *M. tb* and comparative bioinformatic analysis of three *de novo* assembly pipelines with evaluation parameters to optimize assembly completeness for genome mining. Note: Some symbols sourced from Integration and Application Network (ian.umces.edu/media-library).

2.3.5 Genome comparison

In addition to 16S rRNA gene sequence analysis from Sanger sequencing data, the housekeeping genes *recA* and *gyrB* were identified from PATRIC annotation of assembled Illumina MiSeq data and aligned with CLC Main Workbench 7. These genes were chosen based on studies supporting their use as supplemental (or even better than 16S rRNA in some cases) phylogenetic markers for the classification of related bacterial strains (Liu et al., 2012; Rossi et al., 2006; Zhang et al., 2019a). Pairwise genome comparison using all three assemblies (SPAdes, A5-miseq, Shovill) per genome was performed by calculating average nucleotide identity (ANI) based on BLAST and MUMmer with JSpecies Web Server (JSpeciesWS) (Goris et al., 2007; Kurtz et al., 2004; Richter et al., 2016). Genome dot plots were created using D-Genies web application aligned with Minimap2 version 2.24 (Cabanettes & Klopp, 2018).

2.4 Results

2.4.1 Small-scale fermentation and anti-TB activity

From the original collection of 101 novel actinomycetes previously isolated from *X. muta*, 58 strains were recovered from storage and grew on either on ISP2 or R2A medium. Of these, 13 strains were found to produce extracts that consistently inhibit the growth of *M. tuberculosis* H37Ra (Table 2.1). Despite the fact that the 16S rRNA gene sequences of several strains returned BLASTN hits with 100% identity to other sequences in the database, these strains will be described as novel throughout this study, since the complete 16S rRNA gene sequences were not obtained. Previous

studies have also shown that actinomycete strains with very similar 16S rRNA gene

sequences can produce different bioactive compounds (Antony-Babu et al., 2017).

The extracts of four of these isolates were observed to have broad activity as they

were shown to consistently inhibit the growth of *M. tuberculosis* H37Ra, *M.*

smegmatis MC^2 155 and *M. marinum* ATCC 927. *M. smegmatis* MC^2 155 and *M.*

marinum ATCC 927 were used as preliminary indicators for anti-TB activity, as they

are both closely related to *M. tuberculosis* and replicate much more rapidly.

Additionally, *M. marinum* ATCC 927 is a known pathogen that rarely infects humans

but causes a "tuberculosis-like illness in fish" (Akram & Aboobacker, 2021). Two

strains (*Micrococcus* sp. strain R8502A1 and *Micromonospora* sp. strain R45601) are

omitted from this discussion, as they are investigated in more detail in Chapters 3 and

4, respectively. Cultures that were incubated for a minimum of two weeks and as long

as four months (the maximum incubation period tested) retained inhibitory activity. In

every case, extracts tested at 250 µg/10 µL DMSO were shown to produce a zone of

inhibition greater than when tested at 25 µg/10 µL DMSO, confirming a dose-

dependent response.

Table 2.1. Strain identification of active extracts, nearest well-identified BLASTN hit and observed bioactivity.

Isolate ID	Nearest BLASTN Hit (NCBI Accession no.)	% ID	*M. tb* H37Ra	*M. smegmatis* MC^2 155	*M. marinum* ATCC 927
Micrococcus sp. strain XM4230A	*Micrococcus luteus* strain OsEp_A&N_15A1 (MT367834.1)	99.93%	+	+	+
Micrococcus sp. strain XM4230B	*Micrococcus luteus* strain OsEp_A&N_15A1 (MT367834.1)	99.93%	+	+	+
Micromonospora sp. strain R42003	*Micromonospora* sp. 201808 (EU437803.1)	100%	+	-	-
Micromonospora sp. strain R42004	*Micromonospora* sp. 201807 (EU437802.1)	99.93%	+	-	-
Micromonospora sp. strain R42106	*Micromonospora chalcea* strain IMB16-203 (MG190678.1)	100%	+	-	-
Micromonospora sp. strain XM-20-01	*Micromonospora chalcea* strain IMB16-203 (MG190678.1)	100%	+	+	+
Brevibacterium sp. strain R8603A2	*Brevibacterium* sp. CS2 (CP040020.1)	100%	+	-	-
Brevibacterium sp. strain XM4083	*Brevibacterium* sp. strain AKR2 (MN932133.1)	99.57%	+	+	+
Streptomyces sp. strain XM4011	*Streptomyces* sp. strain BI87 (KU058407.1)	99.86%	+	-	-
Streptomyces sp. strain XM83C	*Streptomyces thermocoprophilus* strain NBRC 100771 (NR_112594.1)	99.56%	+	-	-
Streptomyces sp. strain XM4193	*Streptomyces* sp. P38-E01 (MW144955.1)	99.78%	+	-	-

Note: *Streptomyces* spp. strains XM83C and XM4011 were cultured in R2A. All other strains were cultured in ISP2.

2.4.2 Genome assembly pipeline comparison

Genome assembly for each strain was performed using paired-end reads with three separate assembly pipelines: (1) Trimmomatic and SPAdes, (2) Shovill, and (3) A5-miseq. In each case, default parameters were used in assembly. As Trimmomatic is merely a trimming function performed pre-assembly, the first pipeline will be hereafter referred to as SPAdes-assembled for simplicity. It is important to note that there is a slight difference in the way QUAST and PATRIC calculate assembly statistics. The pre-filtering assembly statistics (generated by QUAST) only consider contigs of at least 500 bp when calculating GC content and N50 values, while PATRIC generates these two statistics for the final post-filtered assembly by considering all contigs, irrespective of length. Furthermore, pre- and post-filtering sequence coverage values actually refer to average k-mer coverage for SPAdes and Shovill assemblies. Although Shovill is based on SPAdes, this pipeline varies in that it estimates coverage depth by calculating the ratio of total reads over genome size and automatically downsamples fastq files to a depth of 150x. In addition, Shovill removes any contigs below a coverage of 2x by default. For A5-miseq assemblies, pre-filtering coverage values are calculated from the average coverage of reads included in the final assembly after quality control and error correction. Because A5-miseq does not provide individual coverage values for assembled contigs, it is not possible to calculate a coverage depth of the assembly post-filtering. These inherent differences in the three pipelines result in highly variable coverage values between assemblies in some cases.

Micrococcus

Prefiltering. The two *Micrococcus* strains' assemblies were the smallest genomes analyzed in this set, with *Micrococcus* sp. strain XM4230A (Table 2.2) and *Micrococcus* sp. strain XM4230B (Table 2.3) having genomes of approx. 2.5 to 3.1 Mb, regardless of assembly method employed (Fig. 2.2). Shovill assembled raw data into the fewest initial contigs for strain XM4230A (63) while A5-miseq assembled raw data into the fewest initial contigs for strain XM4230B (121) (Fig. 2.3). For strain XM4230A, the genome was determined to be between 2.6 and 3.1 Mb and have a GC content of 72.7%. For strain XM4230B, the genome size was much more consistent, with all assembly methods yielding a genome of 2.5-2.6 Mb with a GC content of 72.6-72.8%. Shovill yielded the highest N50 value for strain XM4230A (287,979) while SPAdes generated contigs with the highest N50 value for strain XM4230B (76,897) (Fig. 2.4). Raw read files for strain XM4230A were substantially smaller (~300 Mb each) compared to those of all other genomes analyzed (usually ~1 Gb each).

Post-filtering. Substantial removal of contaminant *Micromonospora* contigs was necessary for strain XM4230A when assembled with SPAdes (95% of total contigs (985/1,038) removed) and A5-miseq (73% of total contigs (116/158) removed). Ultimately, A5-miseq generated the best assembly for strain XM4230A and yielded the fewest contigs (42) (Fig. 2.3), although the Shovill assembly retained a slightly higher N50 value (287,979 v. 222,030) (Fig. 2.4). The final GC content of strain XM4230A was 72.8% in every assembly. After removal of contaminant contigs [44%

of total contigs (63/143)] and filtering out low coverage contigs, SPAdes yielded the best assembly for strain XM4230B with fewest contigs (80) and a final GC content of 72.8% (Fig. 2.3). The N50 values did not change for any genomes of strain XM4230B post-filtering.

Scaffolding. Based on BLASTN hits aligning to the trimmed and error-corrected 16S rRNA gene sequences for the *Micrococcus* isolates, strain XM4230A was compared to two complete reference genomes, and strain XM4230B was compared to three complete reference genomes for scaffolding. Strain XM4230A assembled into seven to 12 scaffolds, with A5-assembled data yielding the best output. Strain XM4230B assembled into eight to 12 scaffolds, with A5-miseq-assembled contigs yielding the fewest scaffolds.

Table 2.2. Comparative assembly statistics for *Micrococcus* sp. strain XM4230A.

Micrococcus sp. strain XM4230A	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	3,062,987	2,650,899	2,580,442
Contigs	1,038	158	63
GC	72.66%	72.69%	72.73%
N50	222,067	222,030	287,979
Coverage	~5x	~107x	~75x
Paired Reads (no. clusters)	784,051	784,051	784,051
Post-filtering:			
Total Length	2,570,984	2,573,118	2,574,929
Contigs	53	42	62
GC	72.78%	72.78%	72.76%
N50	253,608	222,030	287,979
Genes	2,465	2,477	2,464
Protein Coding Seqs	2,415	2,427	2,411
Coverage	~63x	n/a	~71x
Structural RNA (tRNA/rRNA)	48/2	48/2	48/5
Completeness	98.65	98.65	98.65
Contamination	1.61	1.61	1.38
Scaffolding:	**MeDuSa**		
Scaffolds	12	7	12
Length (includes Ns)	2,573,184	2,574,518	2,577,529
N50	923,726	741,075	746,446
No. of Genomes Compared To	2	2	2

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.

Table 2.3. Comparative assembly statistics for *Micrococcus* sp. strain XM4230B.

Micrococcus sp. strain XM4230B	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	2,573,696	2,557,537	2,540,703
Contigs	143	121	195
GC	72.61%	72.65%	72.84%
N50	76,897	45,729	26,846
Coverage	~138x	~477x	~40x
Paired Reads (no. clusters)	3,924,599	3,924,599	3,924,599
Post-filtering:			
Total Length	2,535,083	2,540,778	2,540,703
Contigs	80	98	195
GC	72.84%	72.84%	72.81%
N50	76,897	45,729	26,846
Genes	2,432	2,463	2,455
Protein Coding Seqs	2,382	2,413	2,401
Coverage	~215x	n/a	~40x
Structural RNA (tRNA/rRNA)	48/2	48/2	48/6
Completeness	98.65	98.65	98.54
Contamination	1.38	1.38	1.61
Scaffolding:	**MeDuSa**		
Scaffolds	10	8	12
Length (includes Ns)	2,538,583	2,545,278	2,549,403
N50	2,517,556	850,539	235,7261
No. of Genomes Compared To	3	3	3

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.

Micromonospora

Prefiltering. Three of the four *Micromonospora* strains had very similar assemblies, with each genome assembling to approx. 6.7 Mb, irrespective of assembly method used (Fig. 2.2). In the case of *Micromonospora* sp. strain XM-20-01 (Table 2.4), significant fungal contamination was evidenced by the concerningly large total genome length initially assembled by SPAdes (33.0 Mb) and A5-miseq (27.0 Mb) and the reduced GC content (< 60%) compared to other actinomycetes (Fig. 2.2). This

suspicion was confirmed by BLAST hit comparison of contigs against the NCBI nt database. For three of the four strains [*Micromonospora* spp. strain XM-20-01, strain R42003 (Table 2.5) and strain R42004 (Table 2.6)], Shovill assembled the genomes into the fewest contigs pre-filtering (Fig. 2.3). For *Micromonospora* sp. strain R42106 (Table 2.7), SPAdes yielded the fewest contigs before filtering (Fig. 2.3). For every strain except XM-20-01, all three assemblers yielded pre-filtered genomes with a GC content of 72.91 or 72.92%. N50 values were consistently highest for SPAdes assemblies, except for strain XM-20-01, for which Shovill provided the largest N50 value (110,598) (Fig. 2.4).

Post-filtering. After comparing assembled contigs against the BLAST nt database, filtering for coverage and spillover contamination, SPAdes assemblies consistently yielded the fewest contigs (171-291), except for strain XM-20-01, for which Shovill yielded almost 10% fewer contigs (154) than it did with SPAdes or A5-miseq (Fig. 2.3). Every assembly method for every *Micromonospora* strain produced a final genome size of 6.7 Mb with a final GC content of 72.9% (Fig. 2.2). After removal of all contaminant contigs, the genome of strain XM-20-01 was 6.7 Mb with a GC content of approx. 72.9%, in line with the other three strains (Fig. 2.2). The N50 values did not change post-filtering for strains R42003, R42004, and R42106, since the contigs removed were small and had little impact on the overall assembly length (Fig. 2.4). However, after removing all contaminant contigs from the XM-20-01 genome, the N50 value rose from less than 2,000 to over 90,000 bases for SPAdes and A5-miseq assemblies (Fig. 2.4).

Scaffolding. Based on BLASTN identity of the trimmed reverse partial 16S rRNA gene sequences, *Micromonospora* sp. strain XM-20-01 contigs were scaffolded by comparison to two other complete genome sequences, resulting in 23 to 30 final scaffolds, with the fewest resulting when using the final SPAdes-assembled contigs. Strain R42003 ultimately assembled into six to eight contigs when compared to five reference genomes, with the SPAdes-assembled data yielding the fewest contigs again. In the case of strains R42004 and R42106, A5-miseq-assembled contigs yielded the fewest scaffolds with MeDuSa when each compared against three complete reference genomes, with the respective ranges being 13 to 21 scaffolds for strain R42004 and four to 15 scaffolds for strain R42016.

Table 2.4. Comparative assembly statistics for *Micromonospora* sp. strain XM-20-01.

Micromonospora sp. strain XM-20-01	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	33,004,014	27,015,965 (Ns)	7,446,207
Contigs	31,953	22,545	1,020
GC	58.42%	58.97%	71.00%
N50	1,606	1,539	110,598
Coverage	~2x	~16x	~6x
Paired Reads (no. clusters)	1,565,813	1,565,813	1,565,813
Post-filtering:			
Total Length	6,720,446	6,740,150 (Ns)	6,726,690
Contigs	171	172	154
GC	72.90%	72.86%	72.90%
N50	99,581	90,238	112,020
Genes	6,416	6,440	6,388
Protein Coding Seqs	6,362	6,383	6,330
Coverage	~31x	n/a	~27x
Structural RNA (tRNA/rRNA)	51/3	51/6	52/6
Completeness	100	100	100
Contamination	1.84	2.11	1.84
Scaffolding:	**MeDuSa**		
Scaffolds	23	27	30
Length (includes Ns)	6,727,646	6,747,850	6,732,690
N50	1,498,014	4,081,964	3,706,571
No. of Genomes Compared To	2	2	2

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted. A5-miseq added strings of Ns into the assembly, which are included in the total length pre- and post-filtering.

Table 2.5. Comparative assembly statistics for *Micromonospora* sp. strain R42003.

Micromonospora sp. strain R42003	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	6,715,865	6,713,975 (Ns)	6,711,764
Contigs	289	296	268
GC	72.92%	72.92%	72.92%
N50	46,788	41,179	43,212
Coverage	~22x	~75x	~26x
Paired Reads (no. clusters)	1,927,482	1,927,482	1,927,482
Post-filtering:			
Total Length	6,704,703	6,713,975 (Ns)	6,711,509
Contigs	262	296	267
GC	72.92%	72.92%	72.92%
N50	46,788	41,179	43,212
Genes	6,377	6,408	6,369
Protein Coding Seqs	6,321	6,349	6,312
Coverage	~19x	~75x	~26x
Structural RNA (tRNA/rRNA)	51/5	51/8	51/6
Completeness	100	100	100
Contamination	1.84	1.84	1.84
Scaffolding:	**MeDuSa**		
Scaffolds	6	8	7
Length (includes Ns)	6,717,903	6,727,275	6,724,909
N50	4,376,958	6,107,504	6,719,948
No. of Genomes Compared To	5	5	5

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted. A5-miseq added strings of Ns into the assembly, which are included in the total length pre- and post-filtering.

Table 2.6. Comparative assembly statistics for *Micromonospora* sp. strain R42004.

Micromonospora sp. strain R42004	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	6,727,045	6,716,786	6,726,597
Contigs	311	313	297
GC	72.92%	72.91%	72.91%
N50	41,796	39,772	40,911
Coverage	~21x	~83x	~25x
Paired Reads (no. clusters)	2,256,824	2,256,824	2,256,824
Post-filtering:			
Total Length	6,718,855	6,716,786	6,726,597
Contigs	291	313	297
GC	72.91%	72.91%	72.91%
N50	41,443	39,772	40,911
Genes	6,424	6,428	6,407
Protein Coding Seqs	6,368	6,369	6,350
Coverage	~19x	~83x	~25x
Structural RNA (tRNA/rRNA)	51/5	51/8	51/6
Completeness	100	100	100
Contamination	1.84	1.84	1.84
Scaffolding:	**MeDuSa**		
Scaffolds	21	13	15
Length (includes Ns)	6,732,655	6,731,986	6,740,597
N50	3,655,554	1,604,887	2,262,276
No. of Genomes Compared To	3	3	3

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.

Table 2.7. Comparative assembly statistics for *Micromonospora* sp. strain R42106.

Micromonospora sp. strain R42106	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	6,721,542	6,726,769	6,723,783
Contigs	194	222	267
GC	72.91%	72.92%	72.91%
N50	71,018	53,174	45,057
Coverage	~47x	~154x	~23x
Paired Reads (no. clusters)	4,052,164	4,052,164	4,052,164
Post-filtering:			
Total Length	6,715,842	6,726,769	6,723,783
Contigs	182	222	267
GC	72.91%	72.92%	72.91%
N50	71,018	53,174	45,057
Genes	6,346	6,369	6,384
Protein Coding Seqs	6,289	6,309	6,326
Coverage	~40x	~154x	~23x
Structural RNA (tRNA/rRNA)	51/6	51/9	51/7
Completeness	100	100	100
Contamination	1.84	2.37	1.84
Scaffolding:	**MeDuSa**		
Scaffolds	6	4	15
Length (includes Ns)	6,725,042	6,737,269	6,736,283
N50	6,176,325	6,734,706	6,265,841
No. of Genomes Compared To	3	3	3

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.

Brevibacterium

Prefiltering. All three software packages assembled strain XM4083 (Table 2.8) into a genome of approximately 4.0 Mb with a GC content of 68.02%, and strain R8603A2 (Table 2.9) into a genome of approx. 3.3 Mb with a GC content of 70.4% (Fig. 2.2). A5-miseq yielded by far the fewest contigs for strain XM4083 prefiltering (16), and Shovill yielded significantly fewer contigs for strain R8603A2 (66) than A5-miseq (79) or SPAdes (220) (Fig. 2.3). N50 values for strain XM4083 were highly variable

depending on the assembly method employed, and A5-miseq yielded the largest N50

value (801,351) (Fig. 2.4). For strain R8603A2, N50 values were much more

consistent regardless of assembly method, and Shovill yielded the largest N50 value

(267,419) (Fig. 2.4).

Post-filtering. Regardless of assembly method employed, the final genome size of

strain XM4083 was approximately 4.0 Mb with a GC content of 68.0%, and the final

genome size of strain R8603A2 was approximately 3.3 Mb with a GC content of

70.5% (Fig. 2.2). Significant removal of contigs was necessary for the SPAdes

assemblies of both strains (83% of total contigs (84/101) for strain XM4083 and 71%

of total contigs (157/220) for strain R8603A2) (Fig. 2.3). Post-filtering, A5-miseq

still yielded the fewest contigs for strain XM4083 (16) although SPAdes was not very

far off, yielding 17 (Fig. 2.3). The final contig count for strain R8603A2 was very

similar for all three assembly methods, and SPAdes now produced the fewest contigs

(63). N50 values did not change from the original assemblies (Fig. 2.4).

Scaffolding. Trimmed and error-corrected 16S rRNA gene sequences were compared

against the NCBI BLAST nr/nt database to determine closely-related complete

genomes to use as references for scaffolding both *Brevibacterium* isolates.

Brevibacterium sp. strain XM4083 was scaffolded by comparison to six

Brevibacterium reference genomes, and yielded six to 12 final scaffolds, with the

SPAdes assembly producing the fewest final scaffolds. Strain R8603A2 was

compared to two complete reference genomes and yielded 12 to 14 final scaffolds.

Both SPAdes- and A5-miseq-assembled contigs yielded 12 final scaffolds.

Table 2.8. Comparative assembly statistics for *Brevibacterium* sp. strain XM4083.

Brevibacterium sp. strain XM4083	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	4,052,273	4,032,593	4,034,963
Contigs	101	16	31
GC	68.02%	68.02%	68.02%
N50	761,026	801,351	489,145
Coverage	~58x	~300x	~46x
Paired Reads (no. clusters)	4,324,102	4,324,102	4,324,102
Post-filtering:			
Total Length	4,030,327	4,032,593	4,033,869
Contigs	17	16	26
GC	68.02%	68.01%	68.01%
N50	761,026	801,351	489,145
Genes	3,788	3,793	3,809
Protein Coding Seqs	3,738	3,741	3,756
Coverage	~184x	~300x	~50x
Structural RNA (tRNA/rRNA)	47/3	47/5	47/6
Completeness	100	100	100
Contamination	0.19	0.19	0.19
Scaffolding:	**MeDuSa**		
Scaffolds	6	7	12
Length (includes Ns)	4,030,727	4,033,093	4,034,669
N50	3,998,910	2,245,819	1,177,182
No. of Genomes Compared To	6	6	6

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.

Table 2.9. Comparative assembly statistics for *Brevibacterium* sp. strain R8603A2.

Brevibacterium sp. strain R8603A2	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	3,340,005	3,289,503	3,267,070
Contigs	220	79	66
GC	70.43%	70.38%	70.44%
N50	248,801	217,368	267,419
Coverage	~90x	~342x	~109x
Paired Reads (no. clusters)	3,358,346	3,358,346	3,358,346
Post-filtering:			
Total Length	3,255,145	3,270,872	3,261,557
Contigs	63	64	65
GC	70.48%	70.45%	70.47%
N50	248,801	217,368	267,419
Genes	3,146	3,185	3,155
Protein Coding Seqs	3,098	3,133	3,105
Coverage	~184x	n/a	~89x
Structural RNA (tRNA/rRNA)	45/3	46/6	46/4
Completeness	100	100	100
Contamination	0.0	0.0	0.0
Scaffolding:	**MeDuSa**		
Scaffolds	12	12	14
Length (includes Ns)	3,257,245	3,273,372	3,263,757
N50	709,815	604,567	902,778
No. of Genomes Compared To	2	2	2

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.

Streptomyces

Prefiltering. The genome size of strain XM4011 (Table 2.10) varied between
approximately 5.9 and 7.5 Mb depending on assembler, although GC content was
more consistent, ranging between 72.5 and 72.8% (Fig. 2.2). This discrepancy in
genome size despite consistency in GC content was indicative of contamination with
another actinomycete strain. In fact, a large amount of *Micromonospora* reads was
found to contaminate the genome of strain XM4011. The genome sizes and GC

contents of strains XM83C (Table 2.11) (approx. 6.8 Mb and 72.23% GC) and

XM4193 (Table 2.12) (approx. 6.1 Mb and 72.0% GC) were much more consistent

irrespective of software package. Despite these relatively similar genome sizes, contig

count varied widely for each genome depending on assembly method (Fig. 2.3).

Shovill produced the fewest contigs for strain XM4011 (749), and A5-miseq

produced the fewest contigs for strain XM83C (805) and strain XM4193 (55). For all

three *Streptomyces* genomes, SPAdes consistently yielded the highest N50 value (Fig.

2.4).

Post-filtering. Significant contaminant contig removal was necessary for the SPAdes

[87% of total contigs (2,842/3,265)] and A5-miseq [56% of total contigs (712/1,268)]

assemblies of strain XM4011. The final genome size of strain XM4011 was ~ 5.9 Mb

with a GC content of approximately 73% for all assemblers (Fig. 2.2). All three

software packages assembled strain XM83C into a genome of ~ 6.8 Mb with a final

GC content of 72.2% (Fig. 2.2). Strain XM4193 assembled into a final genome of

approximately 6.1 Mb with a GC content of 72.0% with all three methods (Fig. 2.2).

Post-filtering of the SPAdes assembly for strain XM4193 required significant

removal of contigs [77% of total contigs (83/108)] (Fig. 2.3). Ultimately, SPAdes

yielded the fewest contigs as well as the highest N50 values in the final genomes for

all three *Streptomyces* strains (Fig. 2.3, 2.4).

Scaffolding. Streptomyces sp. strain XM4011 was scaffolded by comparison against

the only available closely-related reference genome (*Streptomyces harbinensis* strain

NA02264), found by aligning the trimmed forward 16S rRNA gene fragment of strain XM4011 against the NCBI BLAST nr/nt database. The final genome consisted of seven to 27 scaffolds depending on which assembler data was used, with SPAdes contigs resulting in the fewest scaffolds. Strain XM83C was compared against six reference genomes and yielded 187 to 202 scaffolds, with A5-miseq producing the best final assembly. Unfortunately, Shovill-assembled data for strain XM83C could not be scaffolded with MeDuSa. The exact reason for this failure to scaffold is unknown, as multiple rounds were tested with various sets of reference genomes (one to six), although it should be mentioned that this contig set had the highest number of contigs post-filtering (943), the most for any of the data sets analyzed in this study. Strain XM4193 was compared against five reference genomes and yielded 11 to 20 scaffolds, with SPAdes-assembled data resulting in the fewest scaffolds.

Table 2.10. Comparative assembly statistics for *Streptomyces* sp. strain XM4011.

Streptomyces sp. strain XM4011	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	7,581,305	6,428,810 (Ns)	5,952,194
Contigs	3,265	1,268	749
GC	72.56%	72.71%	72.88%
N50	22,794	18,944	14,732
Coverage	~6x	~180x	~15x
Paired Reads (no. clusters)	4,639,399	4,639,399	4,639,399
Post-filtering:			
Total Length	5,950,109	5,957,816 (Ns)	5,951,019
Contigs	423	556	746
GC	72.95%	72.96%	72.86%
N50	26,659	21,228	14,732
Genes	5,679	5,742	5,805
Protein Coding Seqs	5,620	5,681	5,742
Coverage	~36x	n/a	~15x
Structural RNA (tRNA/rRNA)	55/4	55/6	55/8
Completeness	95.82	95.46	95.46
Contamination	0.7	0.7	1.18
Scaffolding:	**MeDuSa**		
Scaffolds	7	18	27
Length (includes Ns)	5,970,909	5,985,216	5,987,119
N50	5,944,569	5,920,389	5,818,621
No. of Genomes Compared To	1	1	1

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted. A5-miseq added strings of Ns into the assembly, which are included in the total length pre- and post-filtering.

Table 2.11. Comparative assembly statistics for *Streptomyces* sp. strain XM83C.

Streptomyces sp. strain XM83C	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	6,838,614	6,830,359 (Ns)	6,801,453
Contigs	822	805	946
GC	72.23%	72.23%	72.23%
N50	18,780	15,938	14,112
Coverage	~38x	~130x	~29x
Paired Reads (no. clusters)	3,348,454	3,348,454	3,348,454
Post-filtering:			
Total Length	6,797,053	6,827,462 (Ns)	6,800,452
Contigs	729	802	943
GC	72.23%	72.22%	72.21%
N50	18,780	15,938	14,066
Genes	6,616	6,731	6,698
Protein Coding Seqs	6,541	6,650	6,625
Coverage	~39x	n/a	~29x
Structural RNA (tRNA/rRNA)	70/5	70/11	68/5
Completeness	100	99.86	99.43
Contamination	0.79	0.86	1.07
Scaffolding:	**MeDuSa**		
Scaffolds	202	187	
Length (includes Ns)	6,822,453	6,856,962	FAILED
N50	1,731,190	1,043,795	
No. of Genomes Compared To	6	6	

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted. A5-miseq added one string of Ns into the assembly, which is included in the total length pre- and post-filtering.

Table 2.12. Comparative assembly statistics for *Streptomyces* sp. strain XM4193.

Streptomyces sp. strain XM4193	Assembler		
Pre-filtering:	**SPAdes**	**A5-miseq**	**Shovill**
Total Length	6,095,217	6,072,697	6,057,528
Contigs	108	55	74
GC	71.97%	71.96%	72.00%
N50	524,754	355,710	197,671
Coverage	~256x	~636x	~114x
Paired Reads (no. clusters)	11,457,110	11,457,110	11,457,110
Post-filtering:			
Total Length	6,050,740	6,054,796	6,050,682
Contigs	25	40	70
GC	72.03%	72.02%	72.02%
N50	524,754	355,710	197,671
Genes	5,359	5,370	5,378
Protein Coding Seqs	5,283	5,296	5,301
Coverage	~312x	n/a	~94x
Structural RNA (tRNA/rRNA)	66/10	66/8	66/11
Completeness	97.15	97.15	97.15
Contamination	0.18	0.18	0.18
Scaffolding:	MeDuSa		
Scaffolds	11	12	20
Length (includes Ns)	6,051,440	6,056,096	6,053,382
N50	5,910,061	5,897,326	4,138,251
No. of Genomes Compared To	5	5	5

Assemblies yielding the fewest contigs pre- and post-filtering are highlighted.
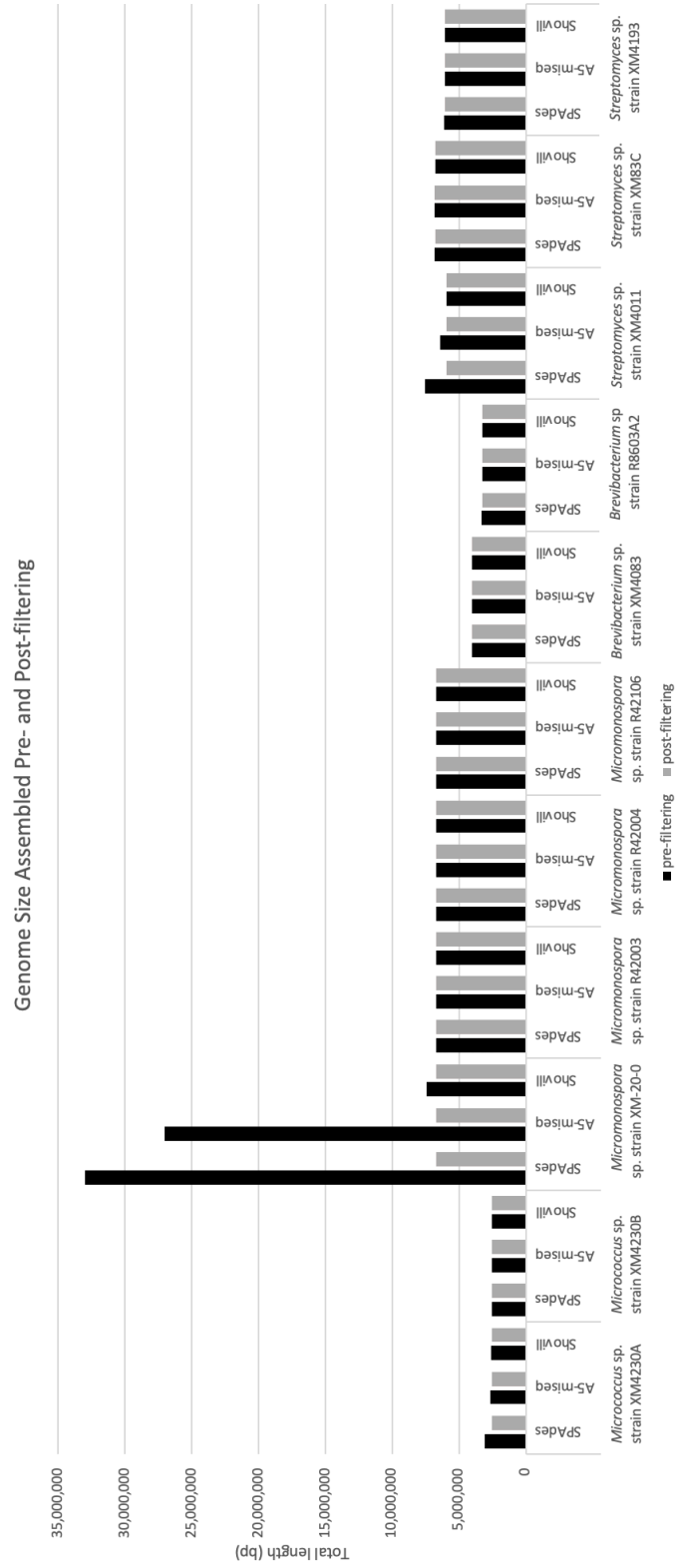
Figure 2.2. Total genome size calculated pre- and post-filtering for contamination and spillover. Filtering procedures did not significantly alter genome size, except for in the case of *Micromonospora* sp. strain XM-20-01, which was heavily contaminated with fungal contigs.

No. of Contigs Assembled Pre- and Post-filtering
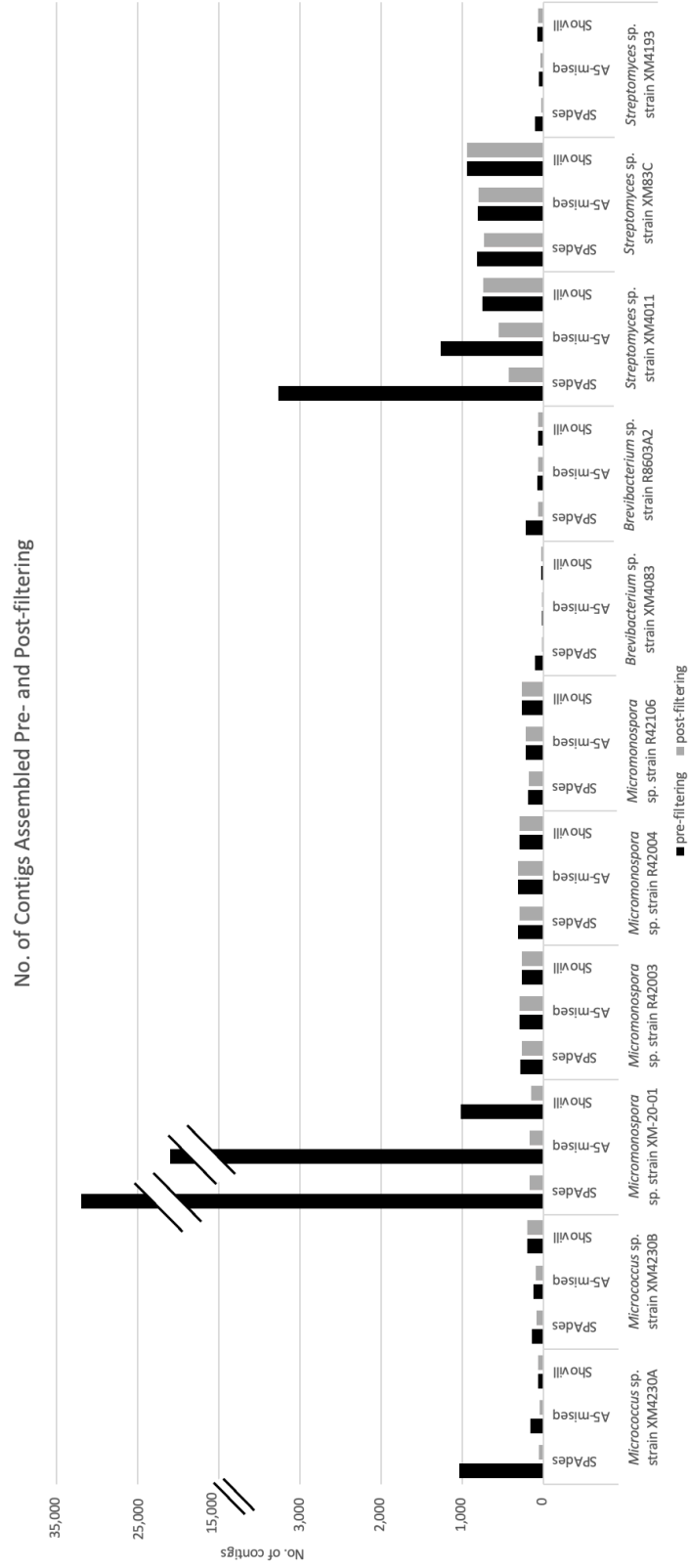
Figure 2.3. Total number of contigs assembled pre- and post-filtering for contamination and spillover. Removal of contigs post-filtering varied per genome based on assembly method, with some assemblies requiring much more contig removal than others. Significant contig removal was necessary for *Micromonospora* sp. strain XM-20-01 (heavily contaminated with fungal contigs) and also for *Streptomyces* sp. strain XM4011, albeit to a lesser extent.

73
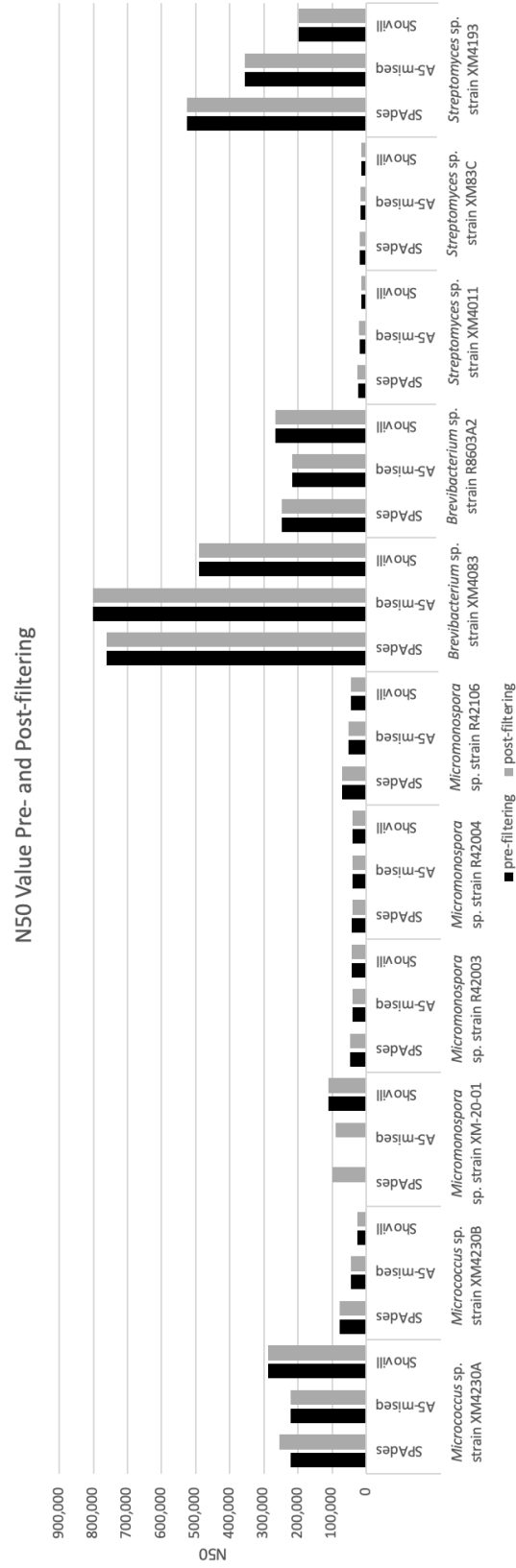
Figure 2.4. N50 values calculated for genome assemblies pre- and post-filtering for contamination and spillover. Values varied for certain genomes by assembler employed but were not affected by filtering procedures. The exception is *Micromonospora* sp. strain XM-20-01, which was heavily contaminated with fungal contigs (the pre-filtering N50 value for XM-20-01 assembled with SPAdes is 1,606 and assembled with A5-miseq is 22,545).

2.4.3 Biosynthetic gene cluster identification

Genomes assembled with all three software packages were analyzed with antiSMASH to identify potential BGCs encoded (Table 2.13). Only BGCs with at least 40% similarity to a known cluster were considered valid, although several clusters with less than 40% similarity are still mentioned if they are identified in other assemblies of the same genome with greater similarity. The majority of BGCs identified were classified as type I PKS, type III PKS, NRPS, terpene, or were characterized as "other", although many type I PKS and type III PKS were associated with hybrid NRPS clusters. There was no difference in putative BGCs detected between the three assembly methods employed for *Micrococcus* sp. strain XM4230A, *Micrococcus* sp. strain XM4230B, *Brevibacterium* sp. strain XM4083 and *Brevibacterium* sp. strain R8603A2. Only one putative BGC encoding a carotenoid was identified for both *Micrococcus* strains and *Brevibacterium* sp. strain R8603A2. No BGCs with known anti-TB activity were detected in any assemblies of any strains of *Micrococcus* and *Brevibacterium* spp. Slight variations were detected among the results for all *Streptomyces* and *Micromonospora* strains depending on assembly method employed, although *Micromonospora* sp. strain XM-20-01 uniquely had the most variability in putative BGCs identified between different assemblies. All *Streptomyces* strains had the largest number of putative BGCs identified by antiSMASH (four to nine depending on the assembly), with *Streptomyces* sp. strain XM4193 consistently identifying the most putative clusters, as well as clusters with 100% similarity to known clusters. For all *Micrococcus* and *Brevibacterium* sp. analyzed, no BGCs with known anti-TB activity were detected. For all *Streptomyces*

and *Micromonospora* strains, at least one compound with known anti-TB activity was detected.

Ectoine, a stress protective osmolyte, was identified in all assemblies of all *Streptomyces* strains (100% similarity) and of *Brevibacterium* sp. strain XM4083 (75% similarity), and likely aids in the adaptation of these bacteria to the high salinity marine environment (Richter et al., 2019). A putative BGC with 32-42% similarity to ECO-02301(an anti-fungal polyketide first isolated from a *Streptomyces*) was identified in every assembly of every *Micromonospora* genome except for the Shovill assembly of strain XM-20-01 (McAlpine et al., 2005). In several cases, the two putative ECO-02301 clusters were identified in the same assembly. Putative clusters with 28-71% similarity to alkyl-O-dihydrogeranyl-methoxyhydroquinone (a type III polyketide first discovered from a rare actinomycete *Actinoplanes missouriensis*) were identified in all assemblies of *Micromonospora* spp. strains XM-20-01, R42003, and R42106 (Awakawa et al., 2011). A type I polyketide BGC with 53% similarity to griseochelin (an antibiotic first isolated from *Streptomyces griseus* and active against Gram-positive bacteria, though no mentions of activity against *M. tb* in the literature) was identified only in the Shovill-assembled genome of *Micromonospora* sp. strain XM-20-01 (Gräfe et al., 1984). Within this same assembly, a type I PKS/ NRPS hybrid BGC was identified with 73% similarity to the anti-cancer macrocyclic depsipeptide rakicidin A/B (Tsakos et al., 2016). All assemblies of *Streptomyces* sp. strain XM4011 contained two putative NRPS clusters (one with 57% similarity and the other with less than 40% similarity) to coelibactin, a zincophore NRP known to play a role in sporulation and regulation of antibiotic biosynthesis (Hesketh et al.,

2009; Kallifidas et al., 2010; Zhao et al., 2012). Also exclusive to the assembled

genomes of *Streptomyces* sp. strain XM4011 is a putative terpene cluster with 100%

similarity to geosmin, the compound known for the distinct earthy odor associated

with fresh soil (Gerber, 1979). A putative BGC with 40% similarity to the pigment

melanin was identified in both the SPAdes and A5-miseq assemblies of *Streptomyces*

sp. strain XM4011 but was absent from the Shovill assembly. Two clusters with 57%

and 60% similarity to melanin were also consistently identified in all assemblies for

Streptomyces sp. strain XM83C, as well as a putative type II PKS cluster with 66%

similarity to a spore pigment. A BGC with 61% similarity to hopene, a terpene

involved in stabilizing the cell membrane and regulating fluidity, was identified in

both the SPAdes and Shovill assemblies of *Streptomyces* sp. strain XM83C, but two

separate putative clusters with 30% and 38% similarity to hopene were detected in the

A5-miseq assembly (Kannenberg & Poralla, 1999). The complete BGC encoding

gamma-butyrolactone, a quorum sensing signaling molecule with roles including

nutrient utilization and activation of antibiotic production, was also identified in all

assemblies of *Streptomyces* sp. strain XM83C (Du et al., 2011). Lastly, for

Streptomyces sp. strain XM83C, the complete BGC encoding the terpene antibiotic

albaflavenone (with anti-*Bacillus* activity reported) was identified, but only in the

SPAdes and A5-miseq assemblies (Gürtler et al., 1994). All assemblies of

Streptomyces sp. strain XM4193 uniquely contained the complete sequence for three

BGCs: isorenieratene (carotenoid), alkylresorcinol (type III polyketide conferring

rigidity to cell membrane) and staurosporine (alkaloid). Staurosporine is reported to

have a wide range of activities including anti-fungal activity and anti-cancer activity

based on its potent inhibition of protein kinases (Belmokhtar et al., 2001; Funabashi et al., 2008; Tamaoki et al., 1986). A hybrid cluster with type I PKS and NRPS-like properties was identified only in *Streptomyces* sp. strain XM4193 assemblies with 85-90% similarity to candicidin, a fungicidal polyketide (Waksman et al., 1965). The same assemblies also all contained a putative NRPS cluster with 41% similarity to the siderophore streptobactin and a putative lassopeptide/terpene cluster with 40% similarity to the ribosomally synthesized and post-translationally modified peptide (RiPP) keywimysin. Both the SPAdes and Shovill-assembled genomes of *Streptomyces* sp. strain XM4193 also retained a cluster with 95% similarity to the anthelminthic NRP WS9326, while the A5-miseq assembly contained two putative clusters with 42% and 57% similarity to WS9326 (Yu et al., 2012). None of the BGCs described above are known to exhibit anti-TB activity based on the available literature.

A total of six BGCs corresponding to PKS (two), NRPS (two), or other (one hybrid, one "other") BGC classes with known anti-TB activity were identified in this study. Interestingly, all assemblies of *Micromonospora* sp. strain R42004 and *Streptomyces* sp. strain XM4193, but only the SPAdes-assembled genome of *Micromonospora* sp. strain R42003, contained a cluster with 100% similarity to the siderophore desferrioxamine E. Studies have shown this siderophore to inhibit biofilm formation of *Mycobacterium* spp. and support its use as adjunctive therapy with current anti-TB drugs (Cahill et al., 2021; Ishida et al., 2011). Additionally, all assemblies of *Streptomyces* sp. strain XM83C contained a putative BGC with 66% similarity to desferrioxamin B [sic] /desferrioxamine E (referred to as

desferrioxamine B from here on). Desferrioxamine B has been shown to inhibit drug-resistant *M. tb* strains in vitro (Gokarn & Pal, 2017). Only in the A5-miseq assembly of *Micromonospora* sp. strain R42003 was a type I PKS cluster with similarity (44%) to oligomycin, a non-selective ATP synthase inhibitor, identified. Although oligomycin can block ATP synthase in *M. tb*, its non-selective nature means it also interferes with enzyme activity in mitochondria, prohibiting its use as an effective anti-TB therapy (Haagsma et al., 2009). All assemblies of *Streptomyces* sp. strain XM4011 identified a putative NRPS cluster with 40% similarity to valinomycin, a compound with a wide range of activities including antifungal, antiviral and antitumor, but first discovered to have antibacterial activity against *M. tb* (Brockmann & Schmidt-Kastner, 1955; Huang et al., 2021). A type I PKS cluster with 57% similarity to macrolide antibiotic methymycin was identified only in the SPAdes assembly of *Micromonospora* sp. strain XM-20-01 (Kittendorf & Sherman, 2009). Methymycin displays activity against Gram-positive pathogens and has been observed to inhibit *M. tb* in vitro (Dutcher et al., 1956). Present in the SPAdes and A5-miseq assemblies of *Streptomyces* sp. strain XM4011 but absent from the Shovill assembly was a putative NRPS with 52% similarity to ecumicin, a cyclic tridecapeptide with inhibitory activity against *M. tb,* including resistant strains (Gao et al., 2015). Strangely, all assemblies of all *Micromonospora* strains investigated except for the Shovill-assembled genome of strain XM-20-01 are reported to contain a hybrid cluster with 94% similarity to diazaquinomycin H/J, a highly potent and selective anti-TB metabolite (Mullowney et al., 2015). Unique to the Shovill assembly of strain XM-20-01 was a type III PKS cluster with 70% similarity to

diazepinomicin, a terpene alkaloid with antitumor, antioxidant and anti-protease activity that shares several genes with the diazaquinomycin cluster (Abdelmohsen et al., 2012; Braesel et al., 2019; Campas, 2009). It is possible that these differences in the Shovill assemblies of strain XM-20-01 are linked to this particular assembler's capacity to deal with the high fungal contamination present in this DNA sample.

Inter-assembly BGC analysis with NP.Searcher also yielded very similar cluster detections for all actinomycete genomes analyzed (Table A.2.1).

Table 2.13. Putative BGCs identified by antiSMASH for actinomycete strains based on assembly method employed.

	SPAdes	AS-miseq	Showill
Micrococcus sp. strain XM4230A	carotenoid (66%)	carotenoid (66%)	carotenoid (66%)
Micrococcus sp. strain XM4230B	carotenoid (66%)	carotenoid (66%)	carotenoid (66%)
Micromonospora sp. strain XM-20-01	diazaquinomycin H/J (94%) ECO-02301 (42%• and 32%••) alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%)• methymycin (57%)•	diazaquinomycin H/J (94%) ECO-02301 (42%) • alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%)	alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%) diazepinomicin (70%)• rakicidin A/B (73%) griseochelin (53%)•
Micromonospora sp. strain R42003	diazaquinomycin H/J (94%) ECO-02301 (42%• and 35%••) alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%)• desferrioxamine E (100%)	diazaquinomycin H/J (94%) ECO-02301 (42%)• alkyl-O-dihydrogeranyl-methoxyhydroquinones (28%)• oligomycin (44%)•	diazaquinomycin H/J (94%) ECO-02301 (35%)• alkyl-O-dihydrogeranyl-methoxyhydroquinones (28%)•
Micromonospora sp. strain R42004	diazaquinomycin H/J (94%) ECO-02301 (32%)• desferrioxamine E (100%)	diazaquinomycin H/J (94%) ECO-02301 (42%)• desferrioxamine E (100%)	diazaquinomycin H/J (94%) ECO-02301 (42%• and 32%•) desferrioxamine E (100%)
Micromonospora sp. strain R42106	diazaquinomycin H/J (94%) ECO-02301 (42% and 32%••) alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%)•	diazaquinomycin H/J (94%) ECO-02301 (42%) • alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%)•	diazaquinomycin H/J (94%) ECO-02301 (42%• and 32%•) alkyl-O-dihydrogeranyl-methoxyhydroquinones (57%)•
Brevibacterium sp. strain XM4083	carotenoid (57%) ectoine (75%)	carotenoid (57%) ectoine (75%)	carotenoid (57%) ectoine (75%)
Brevibacterium sp. strain R8603A2	carotenoid (50%)	carotenoid (50%)	carotenoid (50%)
Streptomyces sp. strain XM4011	ectoine (100%) valinomycin (40%)• geosmin (100%)• coelibactin (54%• and 27%•) melanin (40%)• ecumicin (52%)•	ectoine (100%) valinomycin (40%)• geosmin (100%)• coelibactin (54%• and 27%•) melanin (40%)• ecumicin (52%)•	ectoine (100%) valinomycin (40%)• geosmin (100%)• coelibactin (54%• and 18%•)
Streptomyces sp. strain XM83C	desferrioxamine B/ E (66%)• hopene (61%)• ectoine (100%)• spore pigment (66%)• melanin (57%• and 60%•) gamma-butyrolactone (100%)• albaflavenone (100%)	desferrioxamine B/ E (66%)• hopene (30%• and 38%•) ectoine (100%)• spore pigment (66%)• melanin (57%• and 60%•) gamma-butyrolactone (100%)• albaflavenone (100%)	desferrioxamine B/ E (66%)• hopene (61%)• ectoine (100%)• spore pigment (66%)• melanin (57%• and 60%•) gamma-butyrolactone (100%)•
Streptomyces sp. strain XM4193	alkylresorcinol (100%) isorenieratene (100%) ectoine (100%) candicidin (85%)• staurosporine (100%) streptobactin (41%) desferrioxamine E (100%)• keywimysin (40%) WS9326 (95%)	alkylresorcinol (100%)• isorenieratene (100%) ectoine (100%) candicidin (85%)• staurosporine (100%) streptobactin (41%)• desferrioxamine E (100%)• keywimysin (40%) WS9326 (57%• and 42%•)	alkylresorcinol (100%)• isorenieratene (100%) ectoine (100%) candicidin (90%) staurosporine (100%)• streptobactin (41%) desferrioxamine E (100%)• keywimysin (40%) WS9326 (95%)

Note: If a cluster is identified for a strain with all three assembly methods but does not meet the cutoff threshold of 40% similarity for a particular assembly method, the cluster is labeled in light grey. BGCS located on a contig edge are identified with the symbol "•".

2.4.4 Resolving highly similar isolates

The highly similar assembly statistics calculated for both *Micrococcus* isolates as well as the high similarity observed among all *Micromonospora* strains warranted further consideration as to whether these strains were in fact identical. It is important to note that although the actinomycetes in the original collection were derived from multiple *X. muta* samples, both *Micrococcus* strains were isolated from the same sponge sample, and all four *Micromonospora* strains were isolated from the same sponge sample. Both *Micrococcus* spp. strains XM4230A and XM4230B assembled to genomes of approximately 2.5 Mb with a GC content of ~ 72.8% and had identical partial 16S rRNA gene sequences (1353 bp). All four *Micromonospora* genomes were approximately 6.7 Mb with a GC content of 72.9%. Furthermore, the partial 16S rRNA genes (1355 bp) sequenced with Sanger are identical for all four *Micromonospora* isolates. Therefore, further analysis was required. As an additional measure of identity, the sequences of the housekeeping genes *recA* and *gyrB* as annotated by PATRIC were compared among all isolates. *Micrococcus* spp. XM4230A and XM4230B had identical sequences for both *recA* and *gyrB*. All four *Micromonospora* spp. strains (XM-20-01, R42003, R42004, R42106) had identical sequences as well for both housekeeping genes. ANI values were calculated in pairwise-fashion for all *Micromonospora* strains. Irrespective of which final genome assembly was used for comparison, all *Micromonospora* strain comparisons had ANI values > 99.9%, well exceeding the species delineation threshold of 95% (Table 2.14). Similarly, *Micrococcus* sp. strain XM4230A had an ANI value > 99.9% compared to strain XM4230B (Table 2.15). An ANI value of 95% ± 0.5%

82

corresponds to the DNA-DNA hybridization (DDH) species cutoff value of 70%

(Goris et al., 2007). To avoid redundancy, only the SPAdes-assembled genomes

comparisons are presented below.

Table 2.14. Pairwise ANI comparison of SPAdes-assembled *Micrococcus* genomes based on BLAST+ (ANIb).

Strain	XM4230A	XM4230B
XM4230A	*	99.97 (94.46)
XM4230B	99.99 (96.33)	*

Key: ANIb values [aligned nucleotides] (%).

Table 2.15. Pairwise ANI comparison of SPAdes-assembled *Micromonospora* genomes based on BLAST+ (ANIb).

Strain	XM-20-01	R42003	R42004	R42106
XM-20-01	*	99.99 (93.81)	99.99 (93.79)	99.99 (94.09)
R42003	99.98 (94.78)	*	99.99 (94.99)	100 (95.11)
R42004	99.97 (94.65)	99.99 (94.80)	*	99.99 (94.95)
R42106	99.97 (94.60)	99.99 (94.56)	99.98 (94.56)	*

Key: ANIb values [aligned nucleotides] (%).

As a final proxy measure of identity and similarity, genome dot plots were

performed to visualize the alignment of whole genomes of *Micrococcus* strains in

pairwise fashion (Fig. 2.5). Similar results were observed for comparisons using all

three versions of assembled genomes (only genomes assembled with the same

software package were compared to each other for consistency), so only SPAdes-

assembled genomes are presented to avoid redundancy. The same plots are presented

for all *Micromonospora* strains as well (Fig. 2.6-2.11).

Figure 2.5. Genome dot plot comparing *Micrococcus* spp. strains XM4230A and XM4230B.
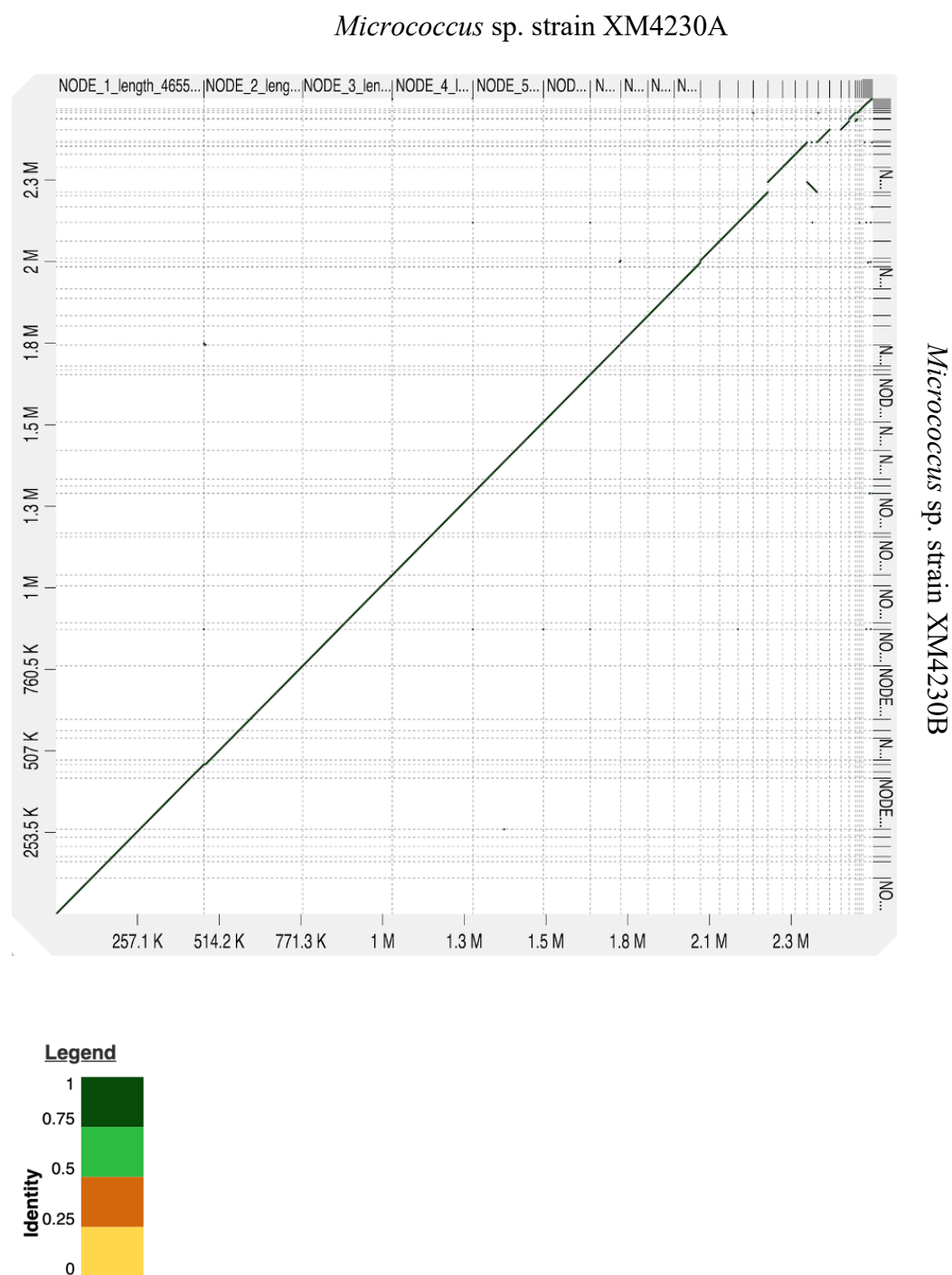
Figure 2.6. Genome dot plot comparing *Micromonospora* spp. strains R42106 and XM-20-01.

Micromonospora sp. strain R42106
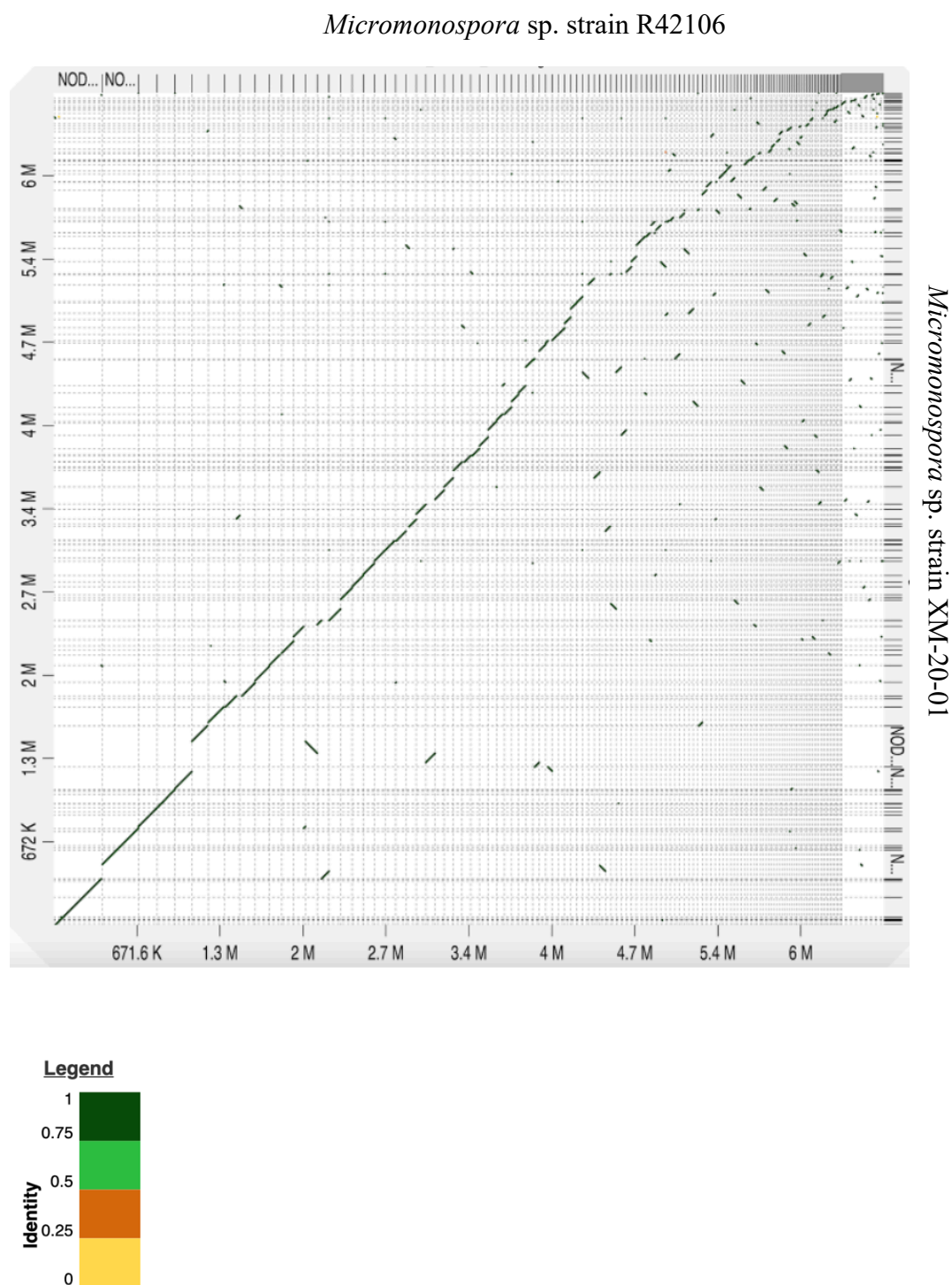
Figure 2.7. Genome dot plot comparing *Micromonospora* spp. strains R42004 and XM-20-01.

Micromonospora sp. strain R42004

Micromonospora sp. strain XM-20-01
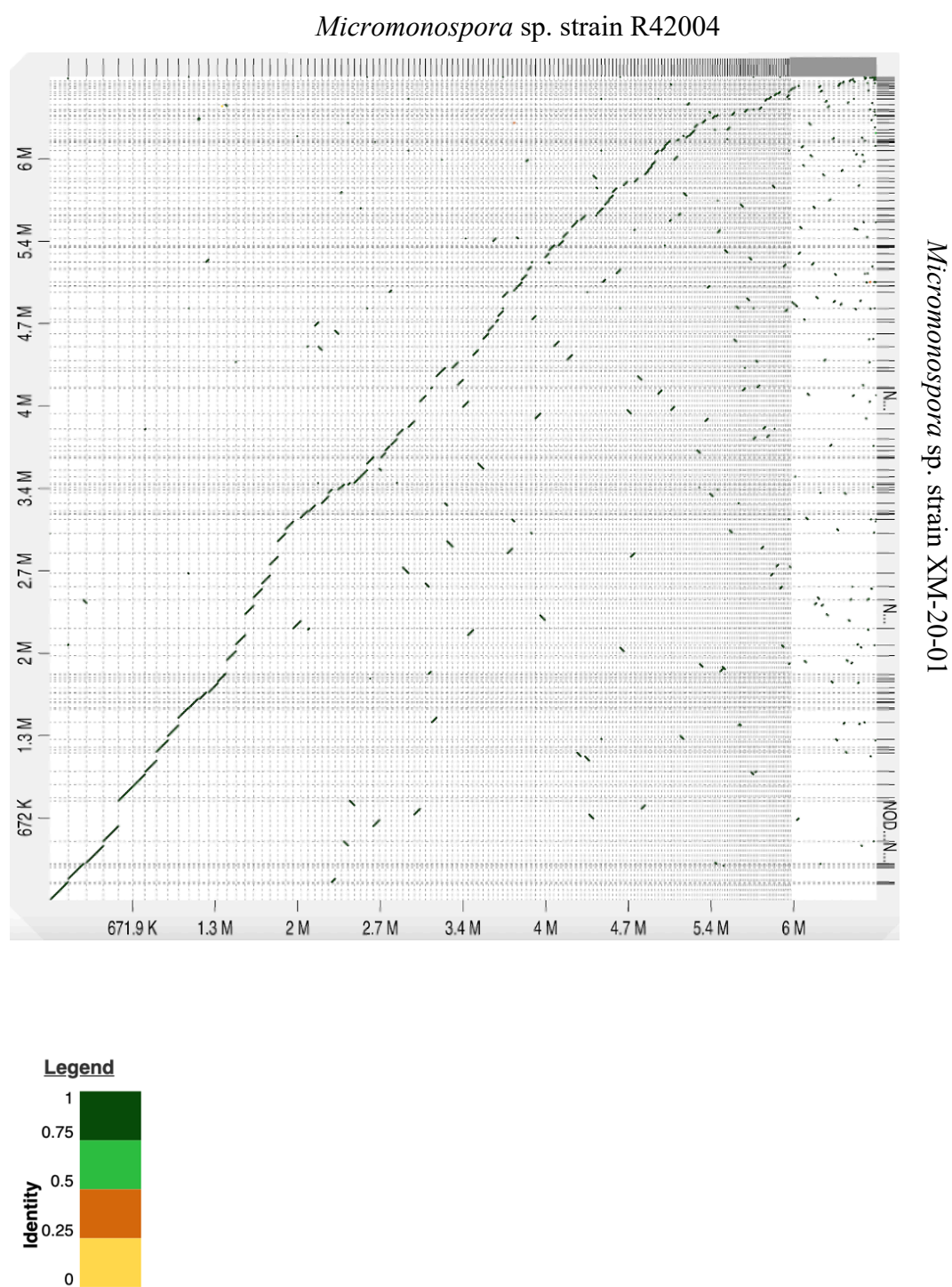
Figure 2.8. Genome dot plot comparing *Micromonospora* spp. strains R42004 and R42106.

Micromonospora sp. strain R42004

Micromonospora sp. strain R42016
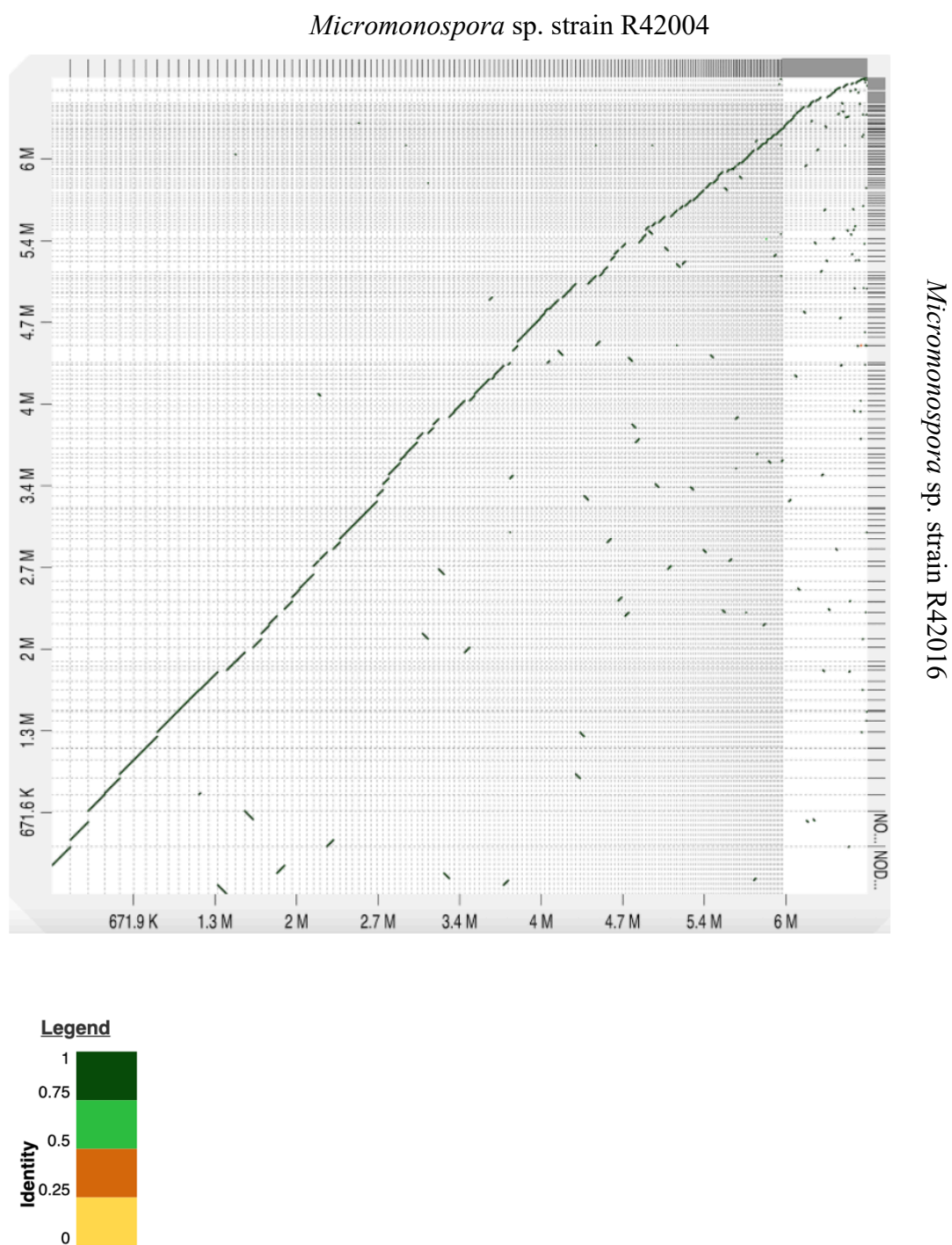
Figure 2.9. Genome dot plot comparing *Micromonospora* spp. strains R42003 and R42106.

Micromonospora sp. strain R42003

Micromonospora sp. strain R42016
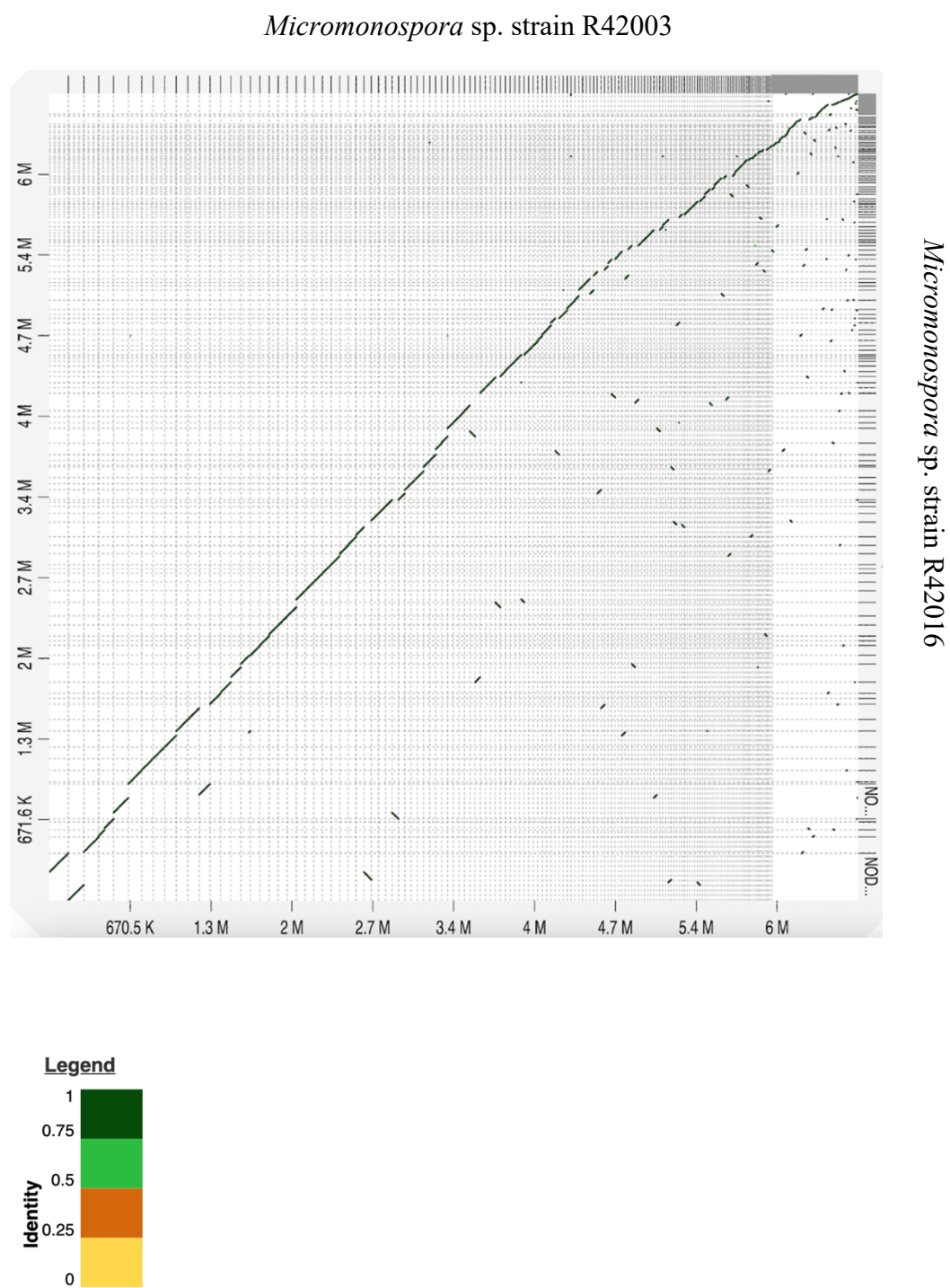
Figure 2.10. Genome dot plot comparing *Micromonospora* spp. strains R42003 and XM-20-01.

Micromonospora sp. strain R42003
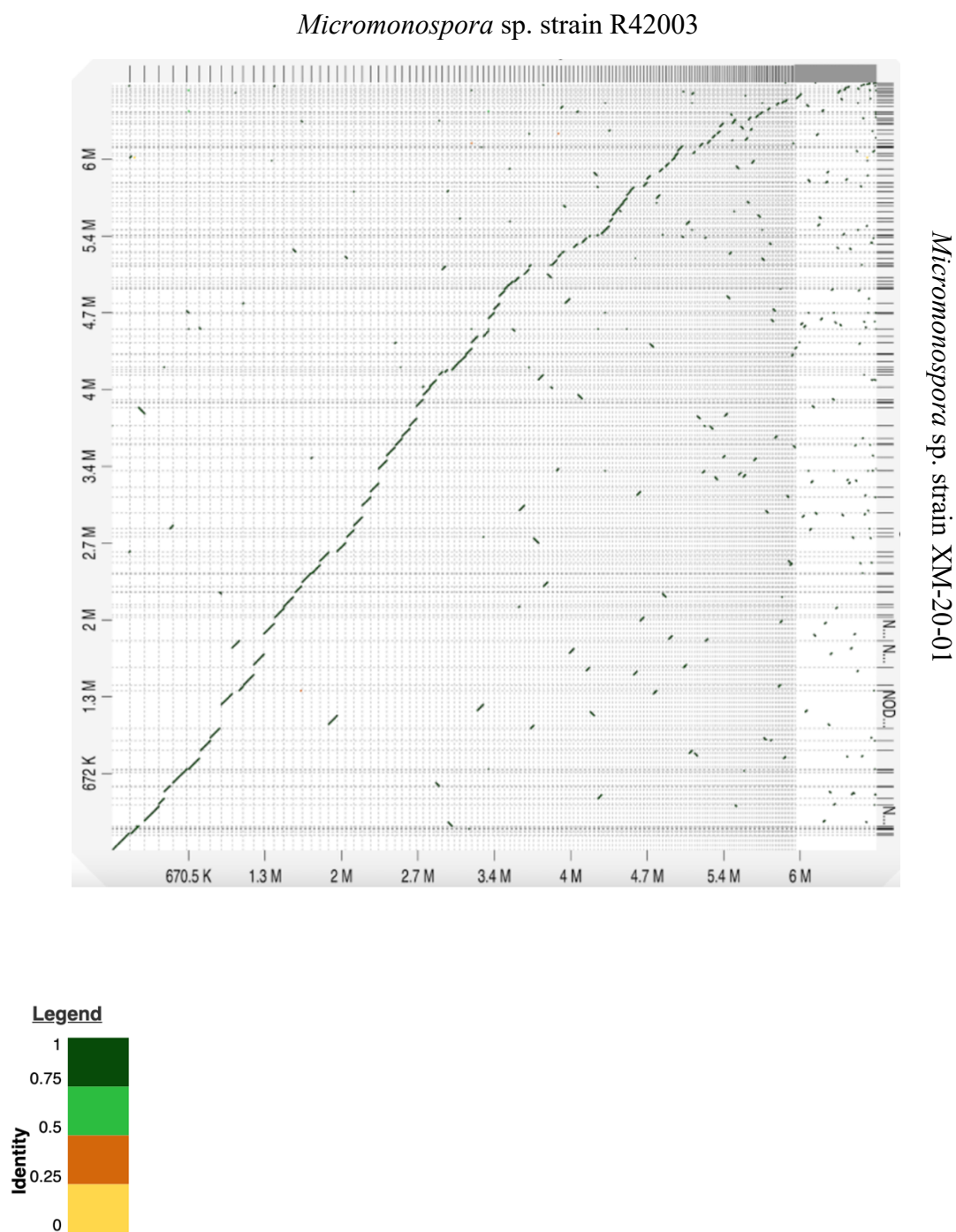
Figure 2.11. Genome dot plot comparing *Micromonospora* spp. strains R42003 and R42004.

Micromonospora sp. strain R42003
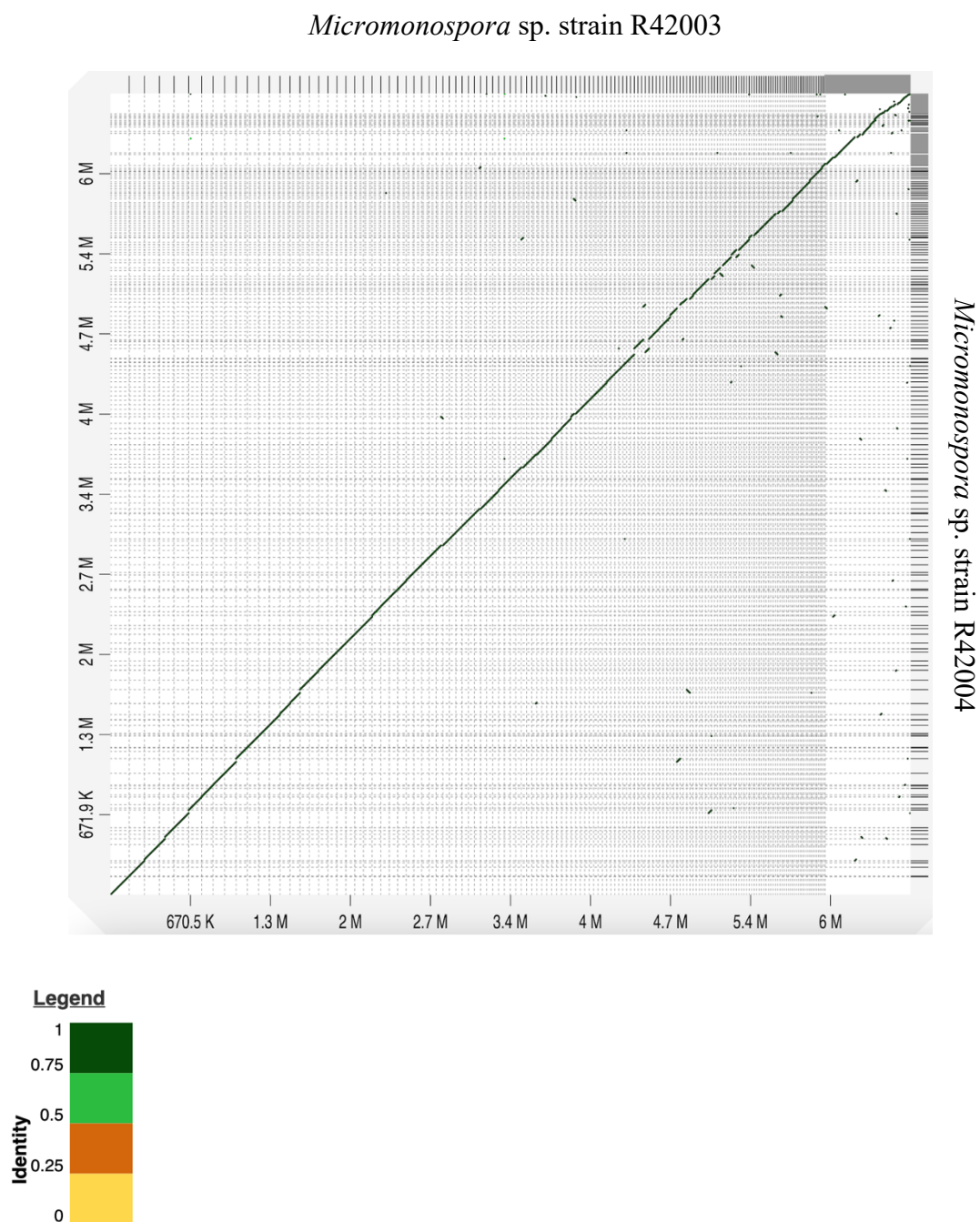
2.5 Discussion

The primary objective of this study was to determine the most efficient method of genome assembly with 250 base paired-end reads that could be applied to the challenging GC-rich actinomycetes with a wide range of genome sizes from various genera. Previous comparative studies on genome assemblers consistently identified SPAdes as generally producing some of the best assemblies of bacterial genomes, and thus this assembler was used as the focal point in this study, to be compared to more recently developed assembly algorithms (Acuña-Amador et al., 2018; Magoc et al., 2013). Although all three methods are fairly similar, they employ slightly variable steps that affect their final output. SPAdes, and thus Shovill, assemble contigs using multi-sized de Bruijn graphs, while A5-miseq contig assembly is performed with the more recently developed IDBA-UD algorithm (Coil et al., 2015; Peng et al., 2012). The de Bruijn graph approach still serves as the base of the IDBA-UD algorithm, although it employs a different method for error correction of k-mers based on coverage depth (Bankevich et al., 2012; Peng et al., 2012). SPAdes uses default k-mer sizes of 21, 33, 55, 77, 99 and 127, while Shovill sets the default k-mer sizes for assembly with SPAdes to 31, 55, 79, 103, and 127. A5-miseq and Shovill both work exclusively for paired end Illumina data, while SPAdes has vastly more capabilities, including the ability to support unpaired reads, as well as hybrid assemblies with long read sequencing data. When using SPAdes, users must also be sure to perform an initial trimming step with another software package, such as Trimmomatic, to remove adapters before assembling raw reads. Shovill addresses this issue by incorporating Trimmomatic into its pipeline, albeit

with predetermined settings that cannot be manually edited. This makes it very simple to use for the coding novice, but is less desirable for any cases where it is would be advantageous to modify the script. A5-miseq also prescreens raw reads for adapters with Trimmomatic before assembly, and features the option to provide an alternative adapter file if necessary. The developers of the original A5/A5-miseq pipelines assert that the main advantage of their software is the ability to produce quality genomes without any prior knowledge of the genome under assembly or parameter tuning, making this pipeline an enticing option for those with a limited bioinformatics background (Tritt et al., 2012).

Considering the initial assembly before filtering, either A5-miseq or Shovill always provided the fewest contigs per genome, except for *Micromonospora* sp. strain R42106. In four of the 11 genomes analyzed in this study, A5-miseq provided the fewest contigs pre-filtering, including for one *Micrococcus* genome and at least one representative for both *Streptomyces* and *Brevibacterium.* Shovill also yielded the fewest contigs for six of the 11 genomes analyzed, including for one *Micrococcus* genome, one *Streptomyces* genome, one *Brevibacterium* genome, and all but one *Micromonospora* genome. This is due to the final steps in the Shovill pipeline, in which minor assembly errors are corrected and any contigs deemed too short, with insufficient coverage (< 2x) or homopolymers, are removed. In many cases, no post-filtering was required of Shovill-assembled contigs. Only in one case was substantial filtering of contigs required post-assembly with Shovill. For *Micromonospora* sp. strain XM-20-01 data, which was heavily contaminated with fungal DNA, 85% of contigs were removed. A5-miseq also removed regions of mis-assembly, albeit most

genomes assembled with this software still required contig filtering post-assembly. Interestingly, no post-processing was necessary for the good quality *Micromonospora* data (all but strain XM-20-01). On the other hand, SPAdes retained low coverage contigs in its output file, which must later be removed. Because the SPAdes *contigs.fasta* output file labels each "node" with a k-mer coverage value, identification of poor coverage contigs and subsequent removal is fairly straightforward. Unfortunately, A5-miseq does not provide a coverage value for individual contigs, and it is therefore more difficult to discern short but legitimate contigs from contaminants or erroneous sequences. In this case, judgement calls on filtering are dependent on match identity by comparison to the NCBI BLAST database. However, further complicating this process is the unique feature of the A5-miseq assembly in which ambiguous nucleotide codes are included in the output contigs file, which results in underestimates in alignment scores with database entries. This did not significantly disrupt post-filtering for most genomes with good quality data, but in the case of *Streptomyces* sp. strain XM4011, lack of coverage values and ambiguous nucleotide codes severely increased the time required to complete filtering and removal. Thus, one should ensure that the DNA sequenced is high quality and void of contamination prior to sequencing and assembly. A similar conclusion was drawn by previous studies in which assembly algorithm performance was compared on various data sets and data quality was found to has a greater impact than the particular assembler on the final assembly (Salzberg et al., 2012).

For every genome assembled in this study, SPAdes-assembled contigs required the most filtering post-assembly. However, the ability to modify pipeline

options and the fact that this assembler maintains a coverage depth for each contig

assembled makes these data much easier to manipulate and correct. Although A5-

miseq normally requires minimal post-filtering, the inability to identify coverage

depth of individual contigs makes filtering post-assembly more complicated and

uncertain. Furthermore, A5-miseq is the only assembler among the three tested that

adds strings of Ns into the assembly during scaffolding. These ambiguous sequences

inflate the total genome size and, similar to the ambiguous nucleotide codes, they

make it more difficult to accurately align assembled contigs with sequences in the

BLAST database. Shovill provides individual coverage depth values for each contig

assembled, enabling easier manipulation of post-assembly results. Despite the fact

that no post-assembly filtering was normally needed for Shovill-assembled genomes,

this software package rarely produced the best assembly in terms of final contig count

and N50 value. It is interesting to note that for the "good quality" *Micromonospora*

data (all but strain XM-20-01), A5-miseq assemblies required absolutely no post-

assembly filtering, and Shovill assemblies required minimal to no filtering (only one

contig was removed from the Shovill assembly for genome *Micromonospora* sp.

strain R42003). However, based on the results, no correlations between genome

assembler performance and bacterial genome size/actinomycete genus were observed.

The most consistent observation was that in eight of the 11 genomes analyzed,

SPAdes ultimately produced the best genome assembly when evaluated with contigs

and N50 value as metrics. Only in the case of *Brevibacterium* sp. strain R8603A2 did

SPAdes yield the fewest final contigs (best assembly) while Shovill-assembled data

retained the slightly higher N50 value. Ultimately, this highlights the importance of

the manual post-filtering required for SPAdes-assembled data, the main step distinguishing SPAdes from Shovill. Previous studies also advocate the importance of additional manual filtering during assembly as opposed to simply autoassembling (Smits, 2019).

In theory, both A5-miseq and Shovill require no additional processing of raw data pre- or post-assembly, but as evident from this study, that is not always the case. Even if very high quality pure DNA is extracted and used for sequencing, contamination in not uncommon from the sequencing process itself. For instance, in several cases, contigs aligning to the blue crab *Callinectes sapidus* genome and barley *Hordeum vulgare*, both organisms known to be sequenced by the same sequencing laboratory, were identified among contigs assembled by A5-miseq that had to subsequently be removed from the final assembly. Contaminant identification was not consistent among assemblers. In some instances, what appeared to be spillover contigs from other actinomycetes sequenced in the same Illumina MiSeq run were detected. These spillover contigs were most easily detected through SPAdes-generated assemblies, often marked by short contig length (usually 1000 bp or less) and low coverage (< 2x). This contamination became more difficult to detect when trying to determine spillover contigs among genomes of the same genus. Multiple *Micromonospora* strains were sequenced in the same run, so to determine which *Micromonospora* isolate small contigs likely belonged to, they were aligned to all contigs assembled from all *Micromonospora* genomes sequenced at the same time. If a contig aligned to the assembly of another *Micromonospora* strain with high percent similarity (at least 90%) and to a node with significantly higher coverage, it was

95

considered a spillover contig and removed. This decision was justified based on the fact that this process ensured that the ambiguous sequence would be retained in at least one other *Micromonospora* assembly, guaranteeing that it would not be overlooked during BGC analysis. Spillover contigs were virtually absent from Shovill assemblies, due to this assembler's aforementioned post-processing filtering step.

One major caveat of the conclusions in this study is that no complete reference genomes were available for any of these strains, as they are all novel isolates. Without a reference genome, it is impossible to fully assess the assembly accuracy of any of the software packages tested. Previous comparative analyses have used reference genomes of closely related species to evaluate assembly correctness, but they acknowledge that true differences existing between the sequenced genome and reference may be considered errors by this method (Magoc et al., 2013). Therefore, to avoid unnecessary complications, assemblies were not compared to any closely-related genomes. Mis-joins including relocations, translocations, inversions, as well as indels and unnecessarily duplicated or compressed repeats could not be identified as a result. Of course, long-read technology remains the superior method for complete and accurate genome assembly.

Despite the deeper coverage provided with sequencing short reads, it is not possible to resolve repeat regions (often longer than the maximum read length) in a fragmented final assembly, as evidenced by the data presented in this study. In bacteria, an estimated 5-10% of the genome consists of genomic repeats (Hofnung & Shapiro, 1999; Parkhill et al., 2000; Shapiro & von Sternberg, 2005). Further assembly of contigs into scaffolds was attempted using the web interface for

MeDuSa. In every case, MeDuSa was able to use genomes of closely related strains

to join contigs. It should be noted that in repeated scaffolding attempts on a particular

assembly with comparison to the same set of reference genomes, the final results

varied slightly. When a gap was determined to be present between two contigs, a

string of 100 Ns was inserted between them. Therefore, repeated scaffolding on the

same genome resulted in slightly different scaffold counts every time, with a genome

size varying by $100n$ bases, where n is difference in amount of N strings inserted.

This scaffolding technique is advantageous based on the fact that it enables better

understanding of how the contigs are linked together, but still leaves an unknown

regarding accurate genome size. The gap regions are often flanked by repeat

sequences, confirming the universally poor performance of assemblers in

reconstructing repeat regions from short reads.

     The frequency with which repeat regions were observed to flank assembly

breaks was reflected in the tendency of the BGCs identified by antiSMASH to be

located on contig edges. Overall, no major differences were observed in the BGCs

identified for a particular genome between the different assembly methods. Still, in 22

of the 33 assemblies analyzed, at least 50% of all the BGCs identified (including

those below the 40% similarity cutoff threshold) fell on contig edges. In the Shovill

assembly of *Streptomyces* sp. strain XM83C, 100% of the 33 putative BGCs

identified were located on contig edges. For every assembly of every

Micromonospora strain analyzed, at least 50% of the BGCs identified were located

on contig edges, including the BGCs for ECO-02301, alkyl-O-dihydrogeranyl-

methoxyhydroquinone, oligomycin, methymicin, diazepinomicin, and griseochelin.

This explains why in some cases a BGC was identified above the 40% cutoff threshold in one assembly but below the cutoff in another (ex. alkyl-O-dihydrogeranyl-methoxyhydroquinone and ECO-02301). It is likely that *Micromonospora* sp. strain XM-20-01 also contains the diazaquinomycin H/J cluster, as opposed to diazepinomicin, since the latter BGC falls on a contig edge and they share genes. Due to the high percent similarity with the diazaquinomycin H/J cluster, it is very likely that a chemical analogue is responsible for the anti-TB activity observed for all *Micromonospora* extracts. Despite the fact that antiSMASH did not detect 6% of the cluster, the possibility cannot be ruled out that the BGC domains are arranged differently in these genomes so that the entire cluster does not fall on one single contig (modular arrangement), and that these strains do in fact contain the entire BGC for diazaquinomycin H/J. This compound/BGC will be discussed in greater detail in Chapter 4. Likewise, it is possible that the *Streptomyces* sp. strain XM4011 genome contains the BGCs with greater % similarity than reported to the known anti-TB compounds valinomycin (40%) and ecumicin (52%), both of which are located on contig edges and thus possibly unresolved. Desferrioxamine B putatively identified by antiSMASH with 66% similarity is also on a contig edge and may in fact be the true compound effecting growth inhibition of *M. tb*. The complete BGC for desferrioxamine E was identified in every genome assembly of *Streptomyces* sp. strain XM4193 and is the only compound with known anti-TB activity identified for this strain. However it cannot be ruled out that this strain produces another novel compound with anti-TB activity that was not detected by antiSMASH. For *Brevibacterium* sp. strain XM4083, no BGCs were reported to be

located on contig edges in the A5-miseq or Shovill assemblies, and only one putative

BGC fell on a contig edge in the SPAdes assembly, although for a cluster that did not

meet the cutoff threshold (no % similarity was provided). Since antiSMASH did not

identify any BGCs in the genomes of these strains nor in any of the other

Micrococcus or *Brevibacterium* strains analyzed related to BGCs previously reported

to encode compounds that inhibit *M. tb*, it is likely that these strains are producing

novel compounds with anti-mycobacterial activity (or at least compounds not in the

MIBiG database). Further genomic analysis with long-read sequencing technology for

all of these genomes with unresolved BGCs, as well as chemical analysis of extracts

would provide much more critical information and is necessary to determine exactly

what compounds these strains are producing that could inhibit *M. tb*.

Based on the various methods used to compare the genome of *Micrococcus*

sp. strain XM4230A with strain XM4230B, as well as all four *Micromonospora*

genomes, it is highly likely that these strains are in fact identical and comprise one

unique *Micrococcus* and one unique *Micromonospora* strain. It may raise concern

that Illumina MiSeq data that can be somewhat error-prone was used to obtain

sequences for the *recA* and *gyrB* genes instead of the more accurate Sanger

sequencing technology. Nevertheless, the fact that all four strains had completely

identical sequence for a span of 1050 base pairs of *recA* (1065 base pairs in the case

of *Micrococcus*) and 1947 base pairs of *gyrB* (2151 base pairs in the case of

Micrococcus) is so unlikely to occur otherwise that further analysis with Sanger was

deemed unnecessary. Without a doubt we can be certain that *Micrococcus* sp. strain

XM4230A and strain XM4230B belong to the same species, and that all four

Micromonospora strains investigated also belong to the same species. Hesitation to declare the *Micrococcus* strains as identical comes from the fact that strain XM4230A consistently assembled to a genome of 2.57 Mb while strain XM4230B consistently assembled to a genome of 2.54 Mb. Similar to the results reported by Antony-Babu et al. (2017), the antiSMASH results were not identical among the *Micromonospora* assemblies, possible indicating that they retain different metabolic profiles. All strains were treated as unique isolates based on their original isolation as separate colonies on initial isolation plates, although it is possible that this analysis is truly dealing with one unique *Micrococcus* and one unique *Micromonospora* strain. Ultimately, assembling the genomes of what are now known to be highly similar strains with these various assemblers served as a sort of control for validating algorithm correctness. The highly similar results obtained for assembly statistics and BGC identification confirm that each assembly method employed was fairly precise, but point out slight differences in contig assembly, as reflected by antiSMASH results. Additionally, this repetitive comparison highlighted the issues that can arise with particular assemblers when dealing with poor quality/contaminated data, as only the Shovill assembly of the highly contaminated *Micromonospora* sp. strain XM-20-01 genome yielded vastly different putative BGCs. This perturbance is also reflected in the genome dot plots, of which comparisons to *Micromonospora* sp. strain XM-20-01 produce the least linear correlations.

This is the first study to compare the effectiveness of various short-read *de novo* bacterial genome assemblers specifically for actinomycete strains with exceedingly high GC content. Although a side-by-side comparison of SPAdes and

100

A5-miseq, among other assemblers, was performed by Acuña-Amador et al. in 2018, no studies specifically assess their performance in assembling genomes of high-GC content bacteria. Past studies have observed that regions of high GC-bias, (either GC-rich or GC-poor) tend to have low coverage of reads, which in turn contributes to assembly breaks and reduces assembly completeness (Browne et al., 2020; Chen et al., 2013). This phenomenon was observed for all assemblies analyzed in the previous study. Although none of the algorithms investigated in this present study were compared in the past analysis, all short-read assemblers suffer from this same weakness, as evidenced by the high levels of fragmentation observed in all 11 genomes. A recently developed assembly algorithm (dnaasm) claims to properly assemble regions of tandem repeats and maintain the ability to restore repetitive regions of the genome covered by only a single read (Kuśmirek & Nowak, 2018). The dnaasm algorithm uses the relative frequency of reads to reconstruct tandem repeats. Unfortunately, the varied read coverage characteristic of genomes with high GC content means that this method would still likely be insufficient to completely resolve assembly breaks in actinomycete genomes. To better understand exactly how different assemblers handle sequences of high GC content and how this affects assembly breakage, future research should investigate the GC content distribution (as opposed to the average value) in these actinomycete genomes. Nevertheless, when only short-read sequencing data is available for genomes with significant GC bias, employing SPAdes with a pre-assembly trimming step and post-assembly manual filtering ultimately yielded the most sufficient assemblies for BGC analysis.

Chapter 3: Investigation into the antimycobacterial activity of a novel marine *Micrococcus* sp.

3.1 Abstract

The Caribbean giant barrel sponge *Xestospongia muta* is associated with a diverse bacterial community that includes members of the Actinobacteria, a group with an excellent track-record of producing bioactive compounds of pharmaceutical importance. A novel collection of actinomycetes previously isolated from *X. muta* were investigated for the production of secondary metabolites that inhibit *Mycobacterium* spp., with a particular emphasis on activity against *Mycobacterium tuberculosis* (*M. tb*). Bioactivity screening identified a strain of the more rarely isolated *Micrococcus* spp. that consistently inhibits growth of *M. tb*, often with great potency. Given the fact that *Micrococcus* spp. tend to have small genomes (< 3 Mb), a size at which biosynthetic gene clusters (BGCs) are rare or absent altogether (Donadio et al., 2007), a genome mining approach was taken to investigate this intriguing activity. Genomes assembled from both short-read sequencing data with Illumina MiSeq and long-read sequencing data with PacBio were mined with various genome mining tools (antiSMASH, BAGEL4, NP.Searcher, and NaPDoS). Although both assemblies yielded similar results, no putative clusters were identified pertaining to compounds with obvious anti-tuberculosis (TB) activity. These results strongly suggest that a novel compound is responsible for the growth inhibition observed. A comprehensive examination of all relevant BGC domains detected in the *Micrococcus* sp. strain R8502A1 genome was performed in an attempt to deduce what type of compound this strain produces that is capable of inhibiting *M. tb*. Ultimately, two key findings were deduced from this analysis: detection of the complete BGC encoding the carotenoid sarcinaxanthin as well as a potentially novel hybrid polyketide

synthase (PKS) - nonribosomal peptide synthetase (NRPS) cluster that may be implicated in the anti-TB activity observed.

3.2 Introduction

The giant barrel sponge *Xestospongia muta* is highly prevalent throughout Caribbean coral reefs and is associated with an immensely diverse bacterial community (McMurray et al., 2008; Montalvo et al., 2005, 2014; Montalvo & Hill, 2011). This sponge was characterized in the late 1800s (Schmidt, 1870), but studies investigating its associated bacterial community are entirely absent from the literature before the turn of the 21st century (Hentschel et al., 2006; Montalvo et al., 2014). In fact, the first discovery of microbial symbionts in any *Xestospongia* species dates to 1985 (Rützler, 1985). Analysis of *X. muta* individuals by Hill and colleagues revealed a novel assemblage of Actinobacteria isolates comprising 17 different genera (Montalvo, 2011). Identified genera ranged from very common *Streptomyces* to rarer *Brevibacterium* and *Nocardiopsis.* Because Actinobacteria are known to be prolific producers of compounds with pharmaceutically-relevant activity, these novel strains were investigated for their capacity to produce bioactive compounds that inhibit *Mycobacterium*, with a particular emphasis on *M. tuberculosis* (*M. tb*). Preliminary investigation identified a strain of a rarely isolated *Micrococcus* spp. that consistently generates extracts effecting potent growth inhibition of *M. tb.* Interestingly, the first study to ever discern the true origin of a sponge-ascribed secondary metabolite discovered a *Micrococcus* sp. to be the true producer, although later studies

questioned the legitimacy of attributing the compound to sponges in the first place (Faulkner et al., 2000; Hill, 2014; Stierle et al., 1988).

Micrococcus is a genus of non-spore forming actinomycetes (family *Micrococcaceae*) ubiquitous throughout terrestrial, aquatic and marine environments (Nuñez, 2014). The genus was first described in the late 1800s, and the type strain *Micrococcus luteus* was originally isolated by Alexander Fleming as *M. lysodeikticus* in 1922 (Cohn, 1872; Fleming & Allison, 1922; Wieser et al., 2002). It is commonly found as a resident of human skin microbiota as well as in the microbiome of dairy products such as raw milk and cheese and has even been isolated from amber (Bhowmik & Marth, 1990; Chiller et al., 2001; Davis, 1996; Greenblatt et al., 2004; Lakshminarasim & Iya, 1955; Nuñez, 2014). In the marine environment, these bacteria have been isolated from sediments as well as marine invertebrates such as sponges and corals (Montalvo et al., 2005; Wang et al., 2021; Wilson et al., 2012). *Micrococcus* are generally considered to be nonpathogenic, although some species have been the culprit of several infections and therefore can act opportunistically (Albertson et al., 1978; Fosse et al., 1985; Nuñez, 2014). Isolates are often vividly pigmented, with yellow, orange, green, pink, red and white colonies having been reported (Jagannadham et al., 1991; Kocur, 1986). Despite considerable investigation into the pigments produced by *Micrococcus* spp., it was noted three quarters of a century ago, and again almost a decade ago, that there is a severe lack of research into their biosynthetic potential (Palomo et al., 2013; Su, 1948). The first *Micrococcus* genome was sequenced in 2009 (Young et al., 2010), and it was revealed to have an exceedingly small genome of only 2.5 Mb. Extensive genomic analysis has concluded

that genomes less than 3 Mb rarely contain biosynthetic gene clusters (Donadio et al., 2007), making the potent anti-TB activity observed by the *Micrococcus* extract all the more puzzling. A genomics-enabled approach was carried out in an attempt to identify the biosynthetic pathway(s) linked to the compound responsible for this activity.

3.3 Materials and methods

3.3.1 Cultivation of *Micrococcus* and *Mycobacterium* strains

Micrococcus sp. strain R8502A1 was previously isolated from *X. muta* and stored as described by Montalvo et al. (2005). A cryovial of the isolate was plated out on both International *Streptomyces* Project 2 (ISP2) agar (BD-Difco™, Franklin Lakes, NJ, USA) supplemented with 20% salt (granular sodium chloride - J.T. Baker, Phillipsburg, NJ, USA) and Reasoner's 2A Agar (R2A) (BD - Difco™, Franklin Lakes, NJ, USA). Plates were incubated at 30°C until growth of individual colonies could be observed. Two individual colonies per plate were then transferred to 100 mL of ISP2 or Reasoner's 2A Broth (R2B) (EZ-Media - Microbiology International, Frederick, MD, USA) in 250 mL baffled flasks to provide sufficient aeration, and incubated at 30°C with shaking at 150 rpm for a minimum of two weeks, until cultures appeared dense. ISP2 cultures (100 mL) were incubated for up to six months before preparing extracts to assess stability of compound production and potency. To ensure homogeny of all cultures during the six month culturing experiment, 1.2 L of ISP2 was inoculated with 10 mL of a starter culture from a 100 mL flask and incubated for several days until sufficient growth was evident. The volume of this

flask was then divided equally among twelve 250 mL baffled flasks under sterile conditions to generate 12 identical 100 mL cultures. Extracts were prepared from two cultures each month in duplicate by selecting two flasks at random. Larger volumes of 500 mL were cultured for up to six months to generate more extract mass. ISP2 liquid medium was prepared as previously described in section 2.3.1.

Mycobacterium tuberculosis H37Ra (avirulent), *Mycobacterium marinum* ATCC 927 and *Mycobacterium smegmatis* MC^2 155 were all plated from cryovials as previously described in section 2.3.1. *M. marinum* ATCC 927 and *M. smegmatis* MC^2 155 were incubated at 30°C with shaking at 150 rpm, and *M. tuberculosis* H37Ra was incubated at 37°C with shaking at 150 rpm.

3.3.2 Preparation of extracts and antimycobacterial activity assay

See section 2.3.2 for a description of how organic extracts were prepared and tested against *Mycobacteria*. Ectoine (> 95.0%, HPLC) (Sigma-Aldrich, St. Louis, MO, USA) standard was dissolved in MilliQ water to concentrations of 5, 25, 50, 100, and 250 µg/10 µL water and applied to 6 mm Whatman filter discs.

3.3.3 Genomic DNA extraction, identification and whole genome sequencing

Micrococcus sp. strain R8502A1 was assigned taxonomic classification as previously described in section 2.3.3. Sequence errors were corrected manually by visual inspection of chromatograms. For short-read genomic sequencing, DNA was sequenced by using a MiSeq sequencer (Illumina) using the MiSeq version 2.4.0.4

Reagent Kit. The Nextera XT Library Prep Kit (1 ng DNA) was used to prepare the

sequencing library (2 x 250 bp paired end reads for a total of 500 cycles).

For long-read sequencing, the Nanobind CBB Big DNA Kit (Circulomics -

Pacific Biosciences of California, Inc., Menlo Park, CA, USA) was used to extract

high molecular weight genomic DNA from *Micrococcus* sp. strain R8502A1. The

sequencing library was prepared using the SMRTbell™ Express Template Prep Kit

2.0 (Pacific Biosciences of California, Inc., Menlo Park, CA, USA), size selected (5-

20 kbp) on a BluePippin™ platform (Sage Science, Inc., Beverly, MA, USA) and

sequenced using a PacBio Sequel II System (Pacific Biosciences of California, Inc.,

Menlo Park, CA, USA) with a SMRT Cell 8M. Sequencing generated 3,007,805 raw

reads with an average length of 10,765 bp (stdev: 4,687 bp, median: 10,950 bp).

These reads reduced down to 220,106 CCS reads with an average length of 12,322 bp

(stdev: 2,701 bp, median: 11,743 bp).

3.3.4 Genome assembly, annotation and biosynthetic gene cluster analysis

Assembly of Illumina MiSeq reads was performed *de novo* using

Trimmomatic version 0.30 (Bolger et al., 2014) and assembled using SPAdes version

3.14.1 (Bankevich et al., 2012). PacBio reads were assembled *de novo* with Canu

version 2.1.1. Initial assembly statistics were evaluated with the Quality Assessment

Tool for Genome Assemblies (QUAST) (Gurevich et al., 2013). SPAdes-assembled

contigs were filtered as previously described in section 2.3.4. Genome annotation for

both assemblies was carried out using PATRIC version 3.5.41 (Brettin et al., 2015;

Davis et al., 2020). The nucleotide sequence for the chromosomal replication initiator

protein DnaA was extracted from the PATRIC annotation and used to properly orient

the complete PacBio assembly using Circlator version 1.5.5 (Hunt et al., 2015). The

final assemblies were validated by evaluating contamination and completeness

values, calculated using CheckM version 1.0.18 app through KBase (kbase.us) (Parks

et al., 2015). Scaffolding for the SPAdes assembly was performed with MeDuSa

version 1.6 (Bosi et al., 2015). This genome was scaffolded as previously described in

section 2.3.4. BGCs were identified using the following algorithms: antiSMASH

(antibiotics & Secondary Metabolite Analysis Shell) version 5.0 (Blin et al., 2019b)

in relaxed mode, BAGEL4 (van Heel et al., 2018), NaPDoS (Ziemert et al., 2012) and

NP.Searcher (Li et al., 2009). Alignment of the SPAdes and Canu genome assemblies

to each other was visualized using progressive Mauve aligner from Mauve version

snapshot 2015_02_25 (Darling et al., 2004).

3.3.5 Phylogenetic analysis

Partial 16S rRNA gene sequences of *Micrococcus* strains from the literature

were extracted from GenBank and aligned with CLC Main Workbench 7. Sequences

deemed too short or of poor quality were removed from the analysis. After aligning

strains from the literature with the PCR-amplified partial 16S rRNA gene sequence of

Micrococcus sp. strain R8502A1, fragments with a minimum length of 1311 bp were

used to construct a phylogenetic tree using MegaX version 10.2.6 (Stecher et al.,

2020).

3.4 Results

3.4.1 Characterization of strain and extract

Colonies of *Micrococcus* sp. strain R8502A1 are bright yellow with two distinct morphologies observed – either a fast-growing phenotype consisting of tiny colonies, or a much slower-growing phenotype displaying larger colonies that eventually form pimple-like aggregates (Fig. 3.1). Identity of both colony types was confirmed by using 16S rRNA gene sequencing. Strain R8502A1 grew well on both ISP2 + 20% NaCl and ISP2 without added salt. ISP2 control discs inoculated with 250 µg of extract displayed a transient clearance zone followed several days later by *M. tb* growth surrounding the disc. ISP2 is a rich medium, and it is possible medium components may be inhibiting *M. tb* H37Ra. However, this seems very unlikely considering that not all actinomycete strains cultured in ISP2 (tested along with *Micrococcus* sp. strain R8502A1) displayed a similar growth inhibition. To ensure that this medium-derived compound was not the same compound inhibiting *M. tb* H37Ra growth in the *Micrococcus* extract, strain R8502A1 was also cultured in the low-nutrient R2B. The R2B-cultured *Micrococcus* extract tested at 200 µg/disc was observed to inhibit *M. tb* H37Ra, while no inhibition was observed from the R2B control disc (tested at 250 µg/disc). This independently confirmed in a different culture medium that strain R8502A1 produces a compound with anti-TB activity. Growth-inhibiting activity of *M. tb* H37Ra by one of the most potent *Micrococcus* sp. strain R8502A1 extracts prepared in this study is displayed in Figure 3.2.

Cultures appear as an opaque yellow/orange homogenous solution and yield dark orange/brown extracts (Fig 3.1). Cultures incubated for two weeks and for up to

six months displayed inhibition against *M. tb* H37Ra with varying strength, but no correlation between incubation time and potency of extract was observed (Tables 3.1 and 3.2), nor any correlation between mass yield of extract and potency (data not shown).

Figure 3.1. Observed colony morphologies (A) of *Micrococcus* sp. strain R8502A1 and extract (B).
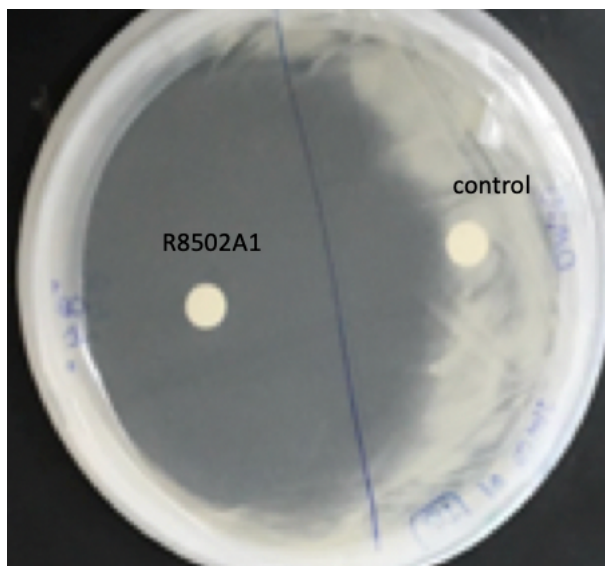
Figure 3.2. Growth-inhibition of *M. tb* H37Ra by a potent *Micrococcus* sp. strain R8502A1 extract. Note: 10 µL of crude extract (dissolved in 100 µL) tested, concentration unknown.

Table 3.1. Observed inhibition of *Micrococcus* sp. strain R8502A1 extracts against *M. tb* H37Ra over a two-month time course experiment.

Age of Extract (weeks)	25 µg/disc - Zone of Inhibition (mm)	250 µg/disc - Zone of Inhibition (mm)
1	0	9.25
2	0	11
4	UNK	10
8	UNK	8.75

Note: Zone values are average measurements taken from discs prepared from two separate cultures tested in duplicate. All concentrations were tested in the same experiment. Discs of four and eight-week old extracts containing 25 µg of extract could not be resolved initially due to overlapping zones and insufficient extract mass prohibited retesting.

Table 3.2. Observed inhibition of *Micrococcus* sp. strain R8502A1 extracts against *M. tb* H37Ra over a six-month time course experiment.

Age of Extract (months)	25 µg/disc - Zone of Inhibition (mm)	250µg/disc - Zone of Inhibition (mm)
1	0	<1
2	0	6
3	1	6.38
4	0	6.38
5	0	3.13
6	0	NA

Note: Zone values are averages measurements taken from discs prepared from two separate cultures tested in duplicate. Different concentrations were tested throughout several experiments. Discs of six-month old extracts containing 250 µg of extract were contaminated and removed from analysis.

3.4.2 Genome assembly and biosynthetic gene cluster analysis

Partial 16S rRNA gene sequence (1354 bp) returned a BLASTN hit with 100% identity to multiple *Micrococcus* strains, with the top hit being *Micrococcus luteus* strain KUK-2-23 (AN: MN826453.1). Sanger sequencing consistently produced a chromatogram with a double peak around base 340. This C/A transversion was validated by the presence of two 16S rRNA gene copies annotated in the Canu assembled genome. Only one 16S rRNA gene copy (with a C at base 380 of 1529) was identified in the SPAdes assembly (Fig. 3.3). Significant *Bacillus* contamination

was present in the Illumina MiSeq assembly and was manually filtered out. Circular

views of both Illumina MiSeq and PacBio-sequenced genomes are presented in

Figure 3.4, and assembly statistics for both are compared in Table 3.3. Scaffolding of

the SPAdes assembly performed by MeDuSa by comparison to two reference

genomes of closely related strains (*M. luteus* strain anQ-m1 and *M. luteus* strain

SGAir0127) yielded 13 scaffolds.
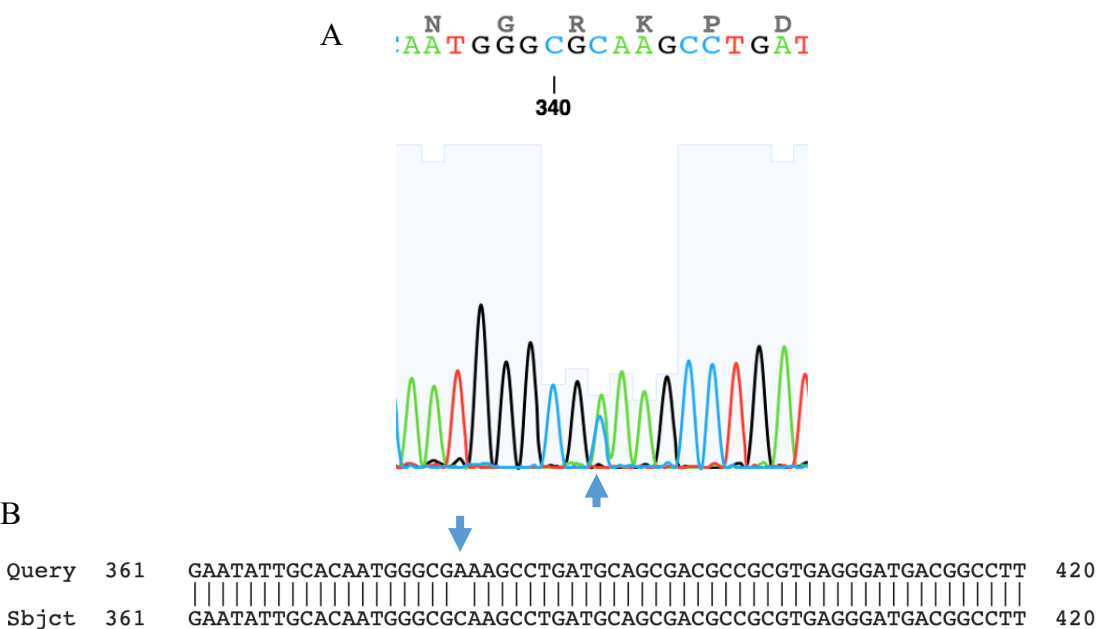
Figure 3.3. Confirming two copies of 16S rRNA gene in the *Micrococcus* sp. strain R8502A1 genome. Chromatogram of partial 16S rRNA gene sequence of *Micrococcus* sp. strain R8502A1 sequenced by Sanger method with double peak (A) and partial alignment of 16S rRNA gene sequences as annotated by PATRIC (B). Both query and subject sequences are derived from PacBio sequencing data assembled with Canu.

113

PacBio

Illumina MiSeq

Key:
- Position Label (Mbp)
- Contigs/Chromosomes
- CDS – FWD
- CDS – REV
- Non-CDS Features
- AMR Genes
- VF Genes
- Transporters
- Drug Targets
- GC Content
- GC Skew
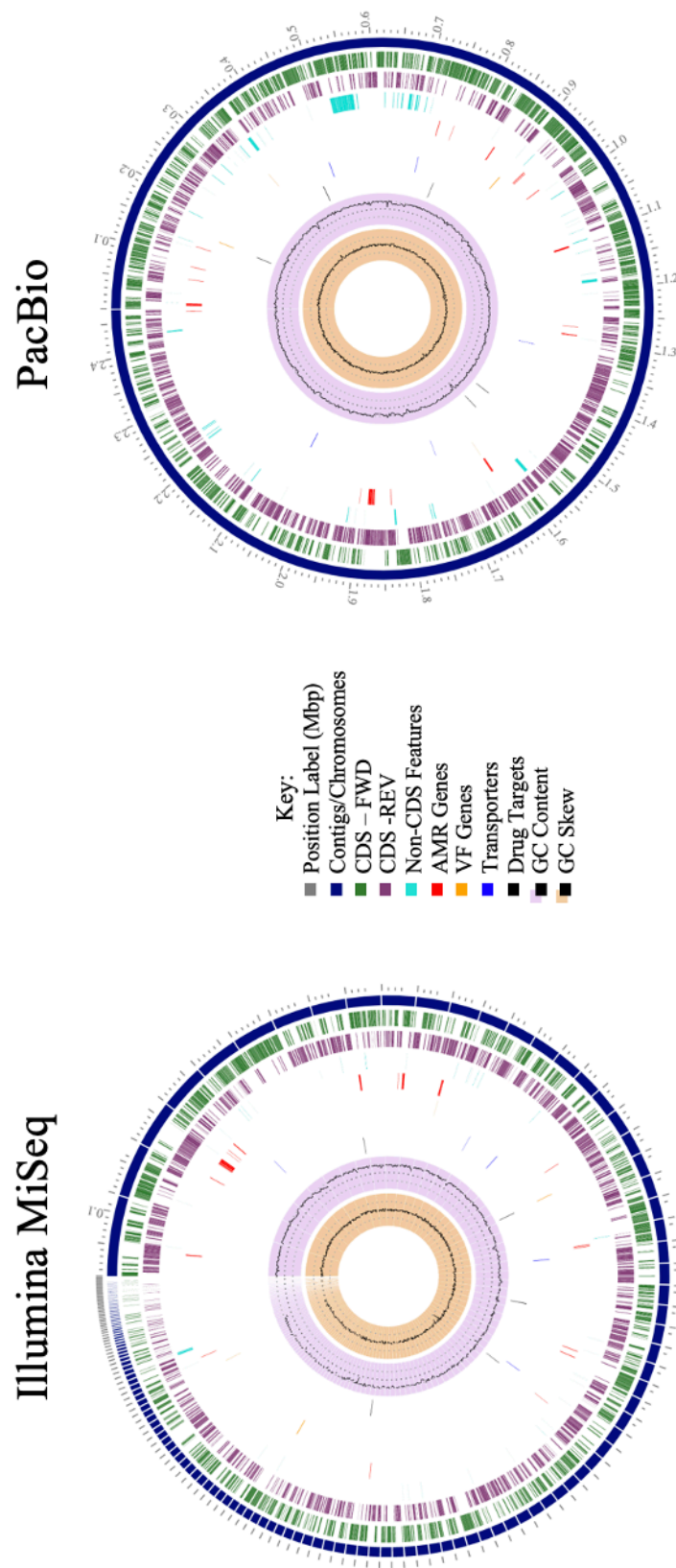
Figure 3.4. Circular view of assembled genomes of *Micrococcus* sp. strain R8502A1 sequenced with Illumina MiSeq and PacBio. Note: Key adapted from PATRIC.

Table 3.3. Assembly statistics of the *Micrococcus* sp. strain R8502A1 genomes sequenced with Illumina MiSeq and PacBio.

SPAdes	Assembly Statistics	Canu
2,422,910	Total Length	2,470,255
137	Contigs	1
73.00%	GC	72.95%
32,728	N50	2,470,255
2,353	Genes	2,333
2,302	Protein Coding Seqs	2,280
~122x	Coverage	800x
49/2	Structural RNA (tRNA/rRNA)	49/4
98.93	Completeness	98.93
0.69	Contamination	1.15

All genome mining algorithms employed (antiSMASH, BAGEL4, NaPDoS, and NP.Searcher) yielded identical results for the SPAdes and Canu assemblies. BAGEL4 did not identify any areas of interest pertaining to bacteriocins or ribosomally synthesized and post-translationally modified peptides (RiPPs). NP.Searcher detected one modular NRPS with the predicted amino acid sequence phenylalanine - glycine - proline (Phe Gly Pro) terminating with a thioesterase (Te) domain (on contig 17 in the SPAdes assembly), and one non-mevalonate terpenoid mep (methylerythritol 4-phosphate pathway) gene (contig/location unknown). The sequence of this non-mevalonate terpenoid mep gene varied slightly between assemblies. In the SPAdes assembly, a cluster was detected with the sequence "NRP NRP Mal Tyr", while in the Canu assembly, the detected sequence was "Phe Gly Pro Tyr Mal NRP". NaPDoS identified two ketosynthase (KS) domains – one modular domain class (on contig 1 in the SPAdes assembly) with 34% identity to the

macrolide antibiotic chlorothricin and one fatty acid synthesis domain (on contig 31

in the SPAdes assembly) with 54% identity to a fatty acid from *Streptomyces*. The

detected KS domain with similarity to chlorothricin is labeled as a KS-beta domain,

also known as a chain length factor, which determines the number of iterative

condensation steps. It should be noted that PATRIC annotates an acyl carrier protein

domain on contig 31, but no KS domain. No condensation (C) domains were

identified by NaPDoS. Interestingly, no regions of interest were identified in nodes 1

or 31 by antiSMASH (Table 3.4). Instead, antiSMASH identified five putative

clusters, with only one cluster maintaining significant similarity (> 40%) to a known

cluster in the MIBiG database (Table 3.4). The organization of these clusters is

presented for the PacBio assembly (Fig. 3.5).

Table 3.4. Putative BGCs in the *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH.

Putative Cluster Type	Most Similar Known Cluster	Location in Canu Assembly	Location in SPAdes Assembly
ectoine	NA	255,947 - 266,318	contig 39 (1 - 5,760) *
RRE-containing	NA	578,485 – 598,817	contig 11 (23,864 - 44,196)
betalactone	microansamycin (7%)	1,516,511 – 1,543,473	contig 8 (4,308 - 31,270)
NAPAA	stenothricin (31%)	1,572,530 – 1,606,537	contig 28 (1 - 28,078) *
terpene	carotenoid (66%)	2,260,208 – 2,281,104	contig 44 (52 - 19,965) *

Note: Putative clusters located on a contig edge are denoted with a symbol (*). Key: RRE = Rev response element, NAPAA = non-alpha poly-amino acids like e-polylysin.

Figure 3.5. Putative BGCs in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH.

Despite the low similarity, it is important to note that microansamycin is a macrolactam antibiotic with activity against mainly Gram-positive and to a lesser extent Gram-negative bacteria, and falls in the same category of antibiotics as rifamycin, one of the top-line anti-TB drugs (Hifnawy et al., 2020). Although there is no evidence in the literature that stenothricin inhibits *M. tb*, this compound is also active against Gram-positive and Gram-negative bacteria and inhibits bacterial cell wall synthesis (Liu et al., 2014). Thus, further analysis of these putative BGCs was carried out to assess the likelihood that these clusters synthesize the compound(s) responsible for the inhibition observed in bioassays. Particular interest was paid to the betalactone (Fig. 3.6; Table 3.5) and "non-alpha poly-amino acids like e-polylysin" (NAPAA) clusters (Fig. 3.7; Table 3.6), as they are closely located to each other in the genome (29,057 bp apart). Clusters were annotated and searched for common NRPS and PKS core domains as well as tailoring domains. Comparisons between the putative BGCs detected by antiSMASH and similar known clusters were also performed to highlight differences between the two (Fig. 3.7, 3.9). Gene annotation varied slightly between the SPAdes and Canu assemblies but was negligible. Only the Canu assembly is analyzed moving forward, as it provided a complete assembly.

Additionally, although it in unclear how similar the putative ectoine cluster is to ectoine, HPLC-grade ectoine was tested against *M. tb* by disk-diffusion assay at

concentrations ranging from 5 – 250 µg/disc. No inhibition was observed at any

concentration tested, confirming that this compound does not possess anti-TB

activity.

Table 3.5. Annotated genes of putative betalactone BGC with similarity to microansamycin in the Canu-assembled _Micrococcus_ sp. strain R8502A1 genome identified by antiSMASH.

Gene	Putative Function	Location
A	Unknown	1,516,535 - 1,516,921
B	ImpB/MucB/SamB family	1,517,055 - 1,518,350
C	Polyprenyl synthetase	1,518,406 - 1,519,521
D	Rv2175c C-terminal domain of unknown function	1,519,566 - 1,520,060
E	LysM domain	1,520,161 - 1,521,348
F	Serine/Threonine protein kinase	1,521,404 - 1,523,341
G	Class-II DAHP synthetase family	1,523,366 - 1,524,778
H	Acyltransferase*	1,524,850 - 1,525,557
I	Abhydrolase_6	1,525,626 - 1,526,396
J	AMP-dependent synthetase and ligase	1,526,511 - 1,528,346
K	Pyridine nucleotide-disulfide oxidoreductase*	1,528,409 - 1,529,947
L	Isopropylmalate synthase or Biotin carboxylase	1,529,955 - 1,533,473
M	MerR HTH family regulatory protein	1,533,632 - 1,534,195
N	Unknown	1,534,408 - 1,534,923
O	MerR HTH family regulatory protein	1,534,916 - 1,535,659
P	FHA domain	1,535,672 - 1,536,097
Q	Glycine cleavage H-protein	1,536,203 - 1,536,589
R	Major facilitator transporter	1,536,675 - 1,538,210
S	Phosphoribosyl transferase domain	1,538,311 - 1,538,832
T	Unknown	1,539,446 - 1,539,925
U	Ribosome biogenesis GTP-binding protein YsxC	1,540,030 - 1,541,568
V	Cytidylate kinase	1,541,565 - 1,542,296
W	Prephenate dehydrogenase	1,542,298 - 1,543,458

Note: If no function was annotated by antiSMASH, then the highest PFAM hit was used to determine the putative gene function. Multiple gene functions are provided if applicable based on varying annotation and PFAM hits. Common NRPS or PKS core and tailoring domains are denoted by a symbol (*).
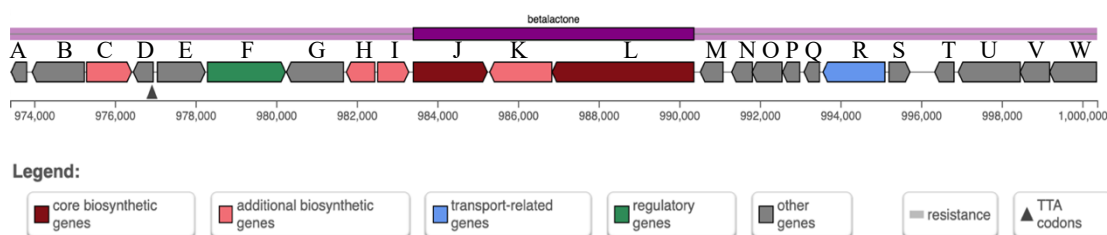
Figure 3.6. Putative BGC of a betalactone with similarity to microansamycin in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH. Note: Legend as provided by antiSMASH.
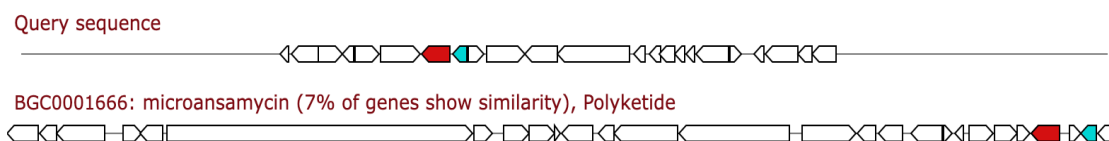
Figure 3.7. Shared genes between putative betalactone BGC in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH and microansamycin. Only two genes are shared between the putative betalactone BGC of *Micrococcus* sp. strain R8502A1 and the known microansamycin cluster. The gene highlighted in red codes for an aDAHP synthase and the gene highlighted in blue codes for a putative phosphate acyltransferase.
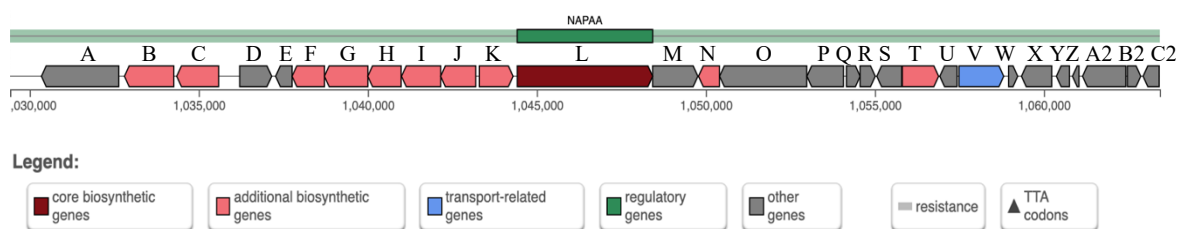
Figure 3.8. Putative BGC of a NAPAA with similarity to stenothricin in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH. Note: Legend as provided by antiSMASH.

Table 3.6. Annotated genes of putative NAPAA BGC with similarity to stenothricin in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH.

Gene	Putative Function	Location
A	AAA domain	1,573,455 - 1,575,743
B	Argininosuccinate lyase/adenylosuccinate lyase	1,575,911 - 1,577,371
C	Argininosuccinate synthase	1,577,457 - 1,578,698
D	Transglycosylase-like domain or LysM domain	1,579,320 - 1,580,267
E	Arginine repressor domain	1,580,382 - 1,580,870
F	Ornithine carbamoyltransferase	1,580,870 - 1,581,829
G	Aminotransferase class-III*	1,581,826 - 1,583,112
H	Acetylglutamate kinase	1,583,109 - 1,584,095
I	Bifunctional ornithine	1,584,107 - 1,585,267
J	N-acetyl-gamma-glutamyl-phosphate reductase or Semialdehyde dehydrogenase, dimerisation domain*	1,585,264 - 1,586,304
K	Crotonyl-CoA reductase / alcohol dehydrogenase*	1,586,407 - 1,587,408
L[a]	AMP-dependent synthetase and ligase or Phosphopantetheine attachment site*	1,587,530 - 1,591,537
M	Peptidase M1 domain	1,591,534 - 1,592,874
N	Phosphopantetheinyl transferase*	1,592,893 - 1,593,513
O	Putative tRNA binding domain	1,593,516 - 1,596,095
P	Aminoacyl tRNA synthetase class II	1,596,099 - 1,597,181
Q	CHY zinc finger	1,597,268 - 1,597,675
R	Unknown	1,597,672 - 1,598,121
S	Pyroglutamyl peptidase	1,598,167 - 1,598,829
T	Radical SAM superfamily	1,598,922 - 1,599,995
U	NUDIX domain	1,600,030 - 1,600,542
V	Major facilitator transporter	1,600,599 - 1,601,921
W	Transglycosylase associated protein	1,602,064 - 1,602,336
X	SpoU rRNA Methylase family)	1,602,446 - 1,603,342
Y	Ribosomal protein L20	1,603,479 - 1,603,865
Z	Ribosomal protein L35	1,603,954 - 1,604,148
A2	Translation initiation factor IF-3 domain	1,604,253 - 1,605,533
B2	Unknown	1,605,640 - 1,605,975
C2	Unknown	1,606,043 - 1,606,519

Note: If no function was annotated by antiSMASH, then the highest PFAM hit was used to determine the putative gene function. Multiple gene functions are provided if applicable based on varying annotation and PFAM hits. Common NRPS or PKS core and tailoring domains are denoted by a symbol (*). [a]Detailed domain annotation reveals that gene "L" contains an adenylation and peptidyl carrier protein domain.

BGC0000431: stenothricin (31% of genes show similarity), NRP:Cyclic depsipeptide

Figure 3.9. Shared genes between putative NAPAA BGC in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH and stenothricin. Seven genes are shared between the putative NAPAA BGC of *Micrococcus* sp. strain R8502A1 and the known stenothricin cluster. In order of appearance from left to right, the highlighted genes code for the following functions: argininosuccinate lyase, argininosuccinate synthase, arginine repressor, N2-acetyl-L-ornithine:2-oxoglutarate aminotransferase, N-acetylglutamate kinase, N2-acetyl-L-ornithine:L-glutamate N-acetyltransferase, and N-acetyl-gamma-glutamylphosphate reductase.

Given the relatively high similarity of the putative terpene cluster detected by antiSMASH with a carotenoid, annotation was performed to determine which carotenoids *Micrococcus* sp. strain R8502A1 is capable of synthesizing (Fig. 3.10; Table 3.7). To provide more clarity as to possible functions of "unknown" genes as labeled by antiSMASH, annotations of the most similar known gene cluster (*M. luteus* strain DE0566) as identified by ClusterBlast are provided (Table 3.7). Clearly the strain is capable of producing at least one carotenoid, evidenced by the characteristic yellow colonies. Based on the annotations and PFAM hits from antiSMASH, it was determined that this cluster encodes the necessary genes to synthesize a pro-vitamin A carotenoid (gamma (γ)-cyclase converts lycopene to beta-carotene). When searching for carotenoid genes in the PATRIC annotation output, the same region noted by antiSMASH has a slightly modified gene assembly. Unlike antiSMASH, PATRIC detected the presence of a C_{50} ε-cyclase (*crtYe/crtYf*) as opposed to a γ-cyclase (*crtYg* and *crtYh*). Furthermore, not identified by antiSMASH but annotated by PATRIC is geranylgeranyl diphosphate (GGPP) synthase (C_{20}

precursor) (*crtE*), the first essential gene in carotenoid synthesis necessary to link three isopentenyl diphosphates (C_5 monomer) to dimethylallyl diphosphate (DMAPP, C_5 monomer) (Rodriguez-Concepcion et al., 2018).

PATRIC annotation reveals a cluster containing all the genes necessary for the synthesis of the rare C_{50} carotenoid sarcinaxanthin (Netzer et al., 2010). Sarcinaxanthin is not in the MIBiG database, but is known to be synthesized by *Micrococcus* spp. (Netzer et al., 2010; Osawa et al., 2010). The presence of a putative glycosyltransferase (*crtX*) clustered with other carotenoid-associated genes by PATRIC signals that *Micrococcus* sp. strain R8502A1 likely synthesizes γ-cyclic sarcinaxanthin. This glycosyltransferase gene may be identical to the C_{50} carotenoid glucosyltransferase (*crtX*) annotated by antiSMASH. In order, the γ-cyclic sarcinaxanthin BGC is comprised of the following genes: *crtE, crtB* (phytoene synthase)*, crtI* (phytoene desaturase or dehydrogenase)*, crtE2* (lycopene elongase)*, crtYg, crtYh, crtX*. Further analysis is required to determine whether *Micrococcus* sp. strain R8502A1 is capable of synthesizing any other carotenoids, as suggested by the presence of another GGPP synthase within the mentioned cluster (2,267,274 – 2,266,918 bp on antisense strand, just after the C_{50} carotenoid ε-cyclase), an additional GGPP synthase located approximately 747,000 bp upstream of the putative sarcinaxanthin cluster, a GGPP reductase (approximately 341,000 bp upstream), another *crtI* (approximately 553,000 bp upstream) and nearly a dozen glycosyltransferases scattered throughout the genome upstream.

Figure 3.10. Putative BGC of a terpene with similarity to a carotenoid in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH. Note: Legend as provided by antiSMASH.

Table 3.7. Annotated genes of putative terpene BGC with similarity to a carotenoid in the Canu-assembled *Micrococcus* sp. strain R8502A1 genome identified by antiSMASH.

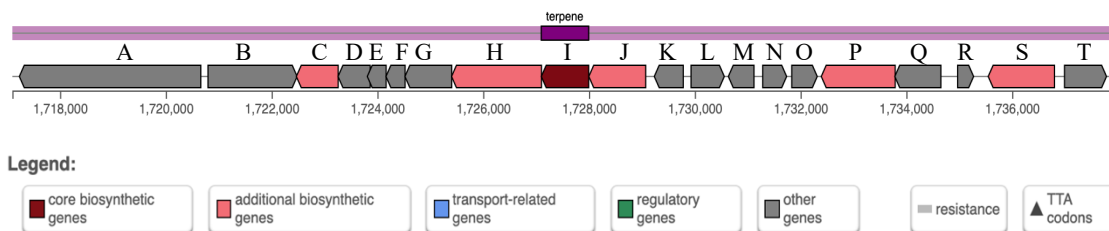| Gene	Putative Function	Closest hit: *M. luteus* strain DE0566 annotation (% ID/ % coverage)
A	SWIM zinc finger or SNF2 family N-terminal domain or helicase conserved C-terminal domain	DEAD/DEAH box helicase family protein (99/100)
B	NmrA family protein	SDR family oxidoreductase (100/100)
C	Glycosyl transferase/polyprenol-monophosphomannose synthase ppm 1	C_{50} carotenoid glucosyltransferase CrtX (99/100)
D	Unknown	Hypothetical protein (99/91)
E	Unknown	C_{50} carotenoid gamma cyclase subunit beta CrtYh (97/99)
F	Unknown	C_{50} carotenoid gamma cyclase subunit alpha CrtYg (98/100)
G	UbiA prenyltransferase family	Prenyltransferase (99/100)
H	Dehydrogenase or Flavin containing amine oxidoreductase	Phytoene desaturase (99/100)
I	Phytoene synthase	Squalene/phytoene synthase family protein (99/100)
J	Polyprenyl synthetase	Polyprenyl synthase family protein (100/100)
K	Unknown	Hypothetical protein (100/100)
L	NUDIX domain	Isopentenyl-diphosphate Delta-isomerase (95/100)
M	Thioredoxin	Thioredoxin (86/91)
N	Dehydroquinase class II	Type II 3-dehydroquinate dehydratase (99/100)
O	Pyridoxamine 5'-phosphate oxidase like	Pyridoxamine 5'-phosphate oxidase family protein (99/100)
P	PF04055 (Radical SAM superfamily)	23S rRNA (adenine(2503)-C(2))-methyltransferase RlmN (100/100)
Q	ER-bound oxygenase mpaB/B'/Rubber oxygenase	DUF2236 domain-containing protein (98/100)
R	Unknown	Hypothetical protein (100/100)
S	Isochorismate synthase	Chorismate-binding protein (99/99)
T	Low molecular weight phosphotyrosine protein phosphatase	Protein-tyrosine-phosphatase (99/100)

Note: If no function was annotated by antiSMASH, then the highest PFAM hit was used to determine the putative gene function. Multiple gene functions are provided if applicable based on varying annotation and PFAM hits. Annotations for a similar cluster found in another *Micrococcus* genome as determined by ClusterBlast are provided.

To see the overall arrangement of commonly detectable NRPS or PKS domains throughout the entire genome by all genome mining tools employed as well as PATRIC annotations, the SPAdes assembly was aligned to the Canu assembly. As shown in Figure 3.11, NRPS or PKS domains were detected in six regions of the *Micrococcus* sp. strain R8502A1 genome. From left to right, contig 11 (as annotated in the SPAdes assembly) encodes a putative acyltransferase (AT) family gene (belonging to the putative RRE-containing cluster as annotated by antiSMASH), contig 1 encodes a KS domain, contig 31 encodes a KS domain (although this corresponds to a fatty acid), as well as an acyl carrier protein domain (identified by PATRIC), contig 8 encodes an AT and an oxidoreductase domain (identified by antiSMASH), contig 28 encodes two putative reductase (Re) or dimerization domains, an aminotransferase, a phosphopantetheine (PPT) attachment site and a phosphopantetheinyl transferase, as well as an adenylation (A) domain and peptidyl carrier protein domain (PCP) (identified by antiSMASH), and contig 17 encodes a modular NRPS with a Te domain (identified by NP.Searcher). Contig 44 (identified as containing the putative terpene cluster by antiSMASH in the SPAdes assembly) is also labeled to point out its close proximity to contig 17. The detection of a tyrosine phosphatase by antiSMASH in contig 44 likely means that this cluster is associated with the non-mevalonate terenoid mep gene identified by NP.Searcher (sequence either "NRP NRP Mal **Tyr**" or "NRP Phe Gly Pro **Tyr** Mal"). Considering the Canu assembly, approximately 156,000 bp separate contig 11 from contig 1, and approx. 79,000 bp separate contig 1 from contig 31. Contig 31 is separated from contig 8 by 509,000 bp, contigs 8 and 28 are linked directly, approximately 583,000 bases

separate contig 28 from contig 17, and approx. 43,000 bp separate contig 17 from

contig 44. This entire genomic region spanning contig 11 to contig 44 covers over

half the total genome size, with more than 1.7 million bp between them.
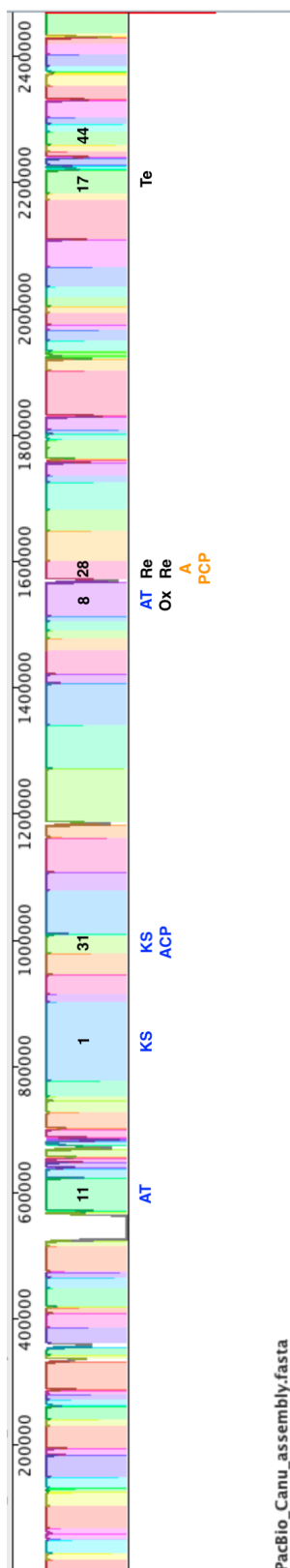
Figure 3.11. Alignment of PacBio (Canu-assembled) genome with Illumina (SPAdes-assembled) genome (not shown) with Mauve version snapshot_2015_02_25. Regions of PacBio assembly pertaining to contigs (as labeled in SPAdes assembly) with standard NRPS or PKS domains are labeled. Contig 44 (putative terpene cluster) is labeled to point out proximity to contig 17. PKS-specific domains are labeled in blue and NRPS-specific domains are labeled in orange. Domains associated with both PKSs and NRPSs, including tailoring domains, are labeled in black. Key: AT (acyltransferase domain), KS (ketosynthase domain), ACP (acyl carrier protein domain), Ox (oxidoreductase domain), Re (reductase domain), A (adenylation domain), PCP (peptidyl carrier protein domain), Te (thioesterase domain).

127

3.4.3 Phylogenetic analysis

To better understand the biosynthetic potential of *Micrococcus* sp. strain

R8502A1 as it relates to the activity of *Micrococcus* strains previously investigated, a

phylogenetic tree was constructed from a nearly exhaustive literature search of

terrestrial and marine-derived strains (Fig. 3.12). Unfortunately, the majority of

Micrococcus strains discovered to produce bioactive compounds, including

antibacterial pigments [*Micrococcus luteus* strain MKVKUD 2013 (AN:

KF532949.1) and *Micrococcus* sp. MP76 (AN: KT804695.1)], anticancer compounds

[*Micrococcus* sp. OUS9 (AN: MN108086.1)] and antibacterial metabolites

[*Micrococcus* sp. strain SB58 (AN: AF218240.1)] had to be excluded from this study

either due to insufficient sequence length for analysis or poor quality. In fact, of the

21 studies that identified bioactivity in *Micrococcus* strains, only seven papers

provided sequence data, from which only six strains were included in the

phylogenetic analysis. *Micrococcus yunnanensis* F-256,446 sequence data was

requested directly from the authors. *M. yunnanensis* F-256,446 is the only strain

included in the phylogenetic tree for which a bioactive compound was isolated. This

isolate is also sponge-derived and produces kocurin, a thiopeptide antibiotic (RiPP)

with anti-MRSA activity (Palomo et al., 2013). A minimal sequence length of 1311

bp for alignment was determined as it covered the region of the 16S rRNA gene that

distinguishes *Micrococcus* sp. strain R8502A1 (by a single nucleotide insertion) from

closely related *Micrococcus* spp. strains XM4230A and XM4230B (as discussed in

Chapter 2). As mentioned in Chapter 2, both strains XM4230A and XM4230B were

isolated from the same *X. muta* sample, while strain R8502A1was isolated from a

separate sponge sample.

MT135569.1 *Micrococcus* sp. strain KRD128 ▲
KF550912.1 *Micrococcus* luteus strain BAA2 ●
MT136106.1 *Micrococcus* sp. strain KRD026 ▲ ☆
MT136242.1 *Micrococcus* sp. strain KRD012 ▲
60
67
MZ156964.1 *Micrococcus luteus* strain MT1691313 ▲
NR 075062.2 *Micrococcus luteus* strain NCTC 2665 ●
AJ409096.1 *Micrococcus luteus* strain Ballarat ●
KF524364.1 *Micrococcus aloeverae* strain AE-6 ●
JX094178.1 *Micrococcus yunnanensis* strain N1-7 ●
KJ843153.1 *Micrococcus luteus* strain VL-80 ●
KU991822.1 *Micrococcus yunnanensis* strain rsk5 ● ☆
AJ312751.1 *Micrococcus luteus* strain 118 ●
Micrococcus yunnanensis F-256,446 ▲ ☆
52
MT135519.1 *Micrococcus* sp. strain KRD153 ▲ ☆
FR848405.1 *Micrococcus luteus* strain 0310ARD7G 6 ●
50
HE613658.1 *Micrococcus luteus* strain TUB6 ● ☆
JN709478.1 *Micrococcus luteus* strain ODB36 ●
100
strain XM4230A ▲ ☆
strain R8502A1 ▲ ☆
82
strain XM4230B ▲ ☆
MT136510.1 *Micrococcus* sp. strain KRD096 ▲ ☆
AJ409095.1 *Micrococcus luteus* strain D7 ●
NR 116578.1 *Micrococcus yunnanensis* strain YIM 65004 ●
MT135795.1 *Micrococcus* sp. strain KRD077 ▲ ☆
AB247644.1 *Micrococcus thailandicus* ●
EU379020.1 *Micrococcus* sp. LJY5 ●
KT719420.1 *Micrococcus yunnanensis* strain MER TA 11 ●
62
KR047779.1 *Micrococcus luteus* isolate IC C3.2 ▲
53
AJ005932.1 *Micrococcus psychrophilum* ●
AY702663.1 *Micrococcus alkanovora* ▲
EU005372.1 *Micrococcus endophyticus* strain YIM 56238 ●
80
67
EU169174.1 *Micrococcus indicus* strain B7 ●
68
NR 026200.1 *Micrococcus lylae* strain DSM 20315 ●
FR832424.1 *Micrococcus cohnii* strain WS4601T ●
100
DQ416794.1 *Micrococcus* sp. G2DM-40 ●
DQ923132.1 *Micrococcus xinjiangensis* strain YIM 90729 ●
52
FJ999947.1 *Micrococcus luteus* strain EQH18 ●
FJ423763.1 *Micrococcus terreus* strain V3M1 ●
99
68
DQ660308.1 *Micrococcus chenggongense* strain AS1.3025 ●
KT958874.1 *Streptomyces lividans* strain S19 ●

Legend:
● terrestrial
▲ marine
☆ bioactivity detected

0.010

Figure 3.12. Maximum-likelihood phylogenetic tree based on partial 16S rRNA gene sequences of *Micrococcus* sp. strain R8502A1, *Micrococcus* spp. strains XM4230A and XM4230B (from Chapter 2), and related strains from literature. Sequences from this study and Chapter 2 are in bold. Adapted from MegaX: The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model (Tamura & Nei, 1993). The tree with the highest log likelihood (-3406.79) is shown. Bootstrap values are calculated from 1000 sampling replicates. The percentage of trees in which the associated taxa clustered together is shown next to the branches (only values > 50% shown). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 40 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 1325 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018; Stecher et al., 2020). *Streptomyces lividans* strain S19 (AN: KT958874.1) was used as the outgroup.

3.5 Discussion

Several studies have shown the great diversity of "rare" (non-streptomycete) actinomycetes, as well as the ability of these diverse actinomycetes to synthesize complex metabolites, often with low toxicity (Bérdy, 2005; Kurtböke, 2012; Schorn et al., 2016). In fact, only approximately 10% of biosynthetic compounds produced by rare actinomycetes are also found in *Streptomyces* (Bérdy, 2005). Yet, as recently as 2012, literary reviews documenting the biosynthetic potential of rare actinomycete species excluded *Micrococcus* spp., indicating that this genus has historically been overlooked in drug discovery efforts (Bérdy, 2005; Kurtböke, 2012). Some investigations into the bioactivity of *Micrococcus* strains have been conducted, though they are far and few between. The antibacterial activity of *Micrococcus* strains has been known since 1889, when the first photographic record of antibiosis was prepared with *M. anthracotoxicus* (Doehle, 1889, as cited in Florey, 1946; Su, 1948). Several studies between in 1902 and 1948 identified antibacterial activity of

130

Micrococcus strains as well (Dujardin-Beaumetz, 1934; Hutchinson et al., 1943; Löde, 1902; Su, 1948). Although the original work is absent from data records, Su (1948) notes that Löde isolated the active substance from a *Micrococcus* strain in 1902 for animal studies that ultimately failed to demonstrate any therapeutic results. This likely makes this unidentified compound the first bioactive metabolite to ever be isolated from a *Micrococcus*.

Aside from characterization studies, the majority of the literature focuses on the pigments that *Micrococcus* spp. produce, several times documenting their antioxidant, antibiotic, antifungal and anticancer properties (Cooney et al., 1966; Jagannadham et al., 1991; Karbalaei-Heidari et al., 2020; Majeed, 2017; Mohana et al., 2013; Netzer et al., 2010; Nisha et al., 2020; Rostami et al., 2016; Sobin & Stahly, 1942; Surekha et al., 2016; Umadevi & Krishnaveni, 2013; Ungers & Cooney, 1968; Ushasri & Gods Will Shalomi, 2015). Several studies have tested crude *Micrococcus* extracts for various pharmaceutically-relevant activity, but fail to isolate the active component (Akbar et al., 2014; Anteneh et al., 2021; Cita et al., 2017; Li et al., 2012; Sharma et al., 2012; Ushasri & Gods Will Shalomi, 2015). In fact, a literature search reveals only nine papers describing isolated bioactive compounds (including pigments) from *Micrococcus* spp. (Biskupiak et al., 1988; Bultel-Poncé et al., 1999; Eltamany et al., 2014; Mohana et al., 2013; Nisha et al., 2020; Palomo et al., 2013; Shanthi Kumari et al., 2020; Stierle et al., 1988, Su, 1948). Surprisingly, six of these papers pertain to marine isolates, four of which are derived from sponges. Micrococcin, the thiopeptide (RiPP) identified by Su in 1948, is active against Gram-positive bacteria including *M. tb*.

The serious lack of information regarding *Micrococcus*-derived compounds in the literature is at once both frustrating and intriguing. Without the ability to identify any known compounds in the genome of *Micrococcus* sp. strain R8502A1, genome mining efforts can only go so far before chemical analysis is required. On the other hand, this dearth of knowledge signifies that this novel strain is a promising source for drug discovery. The genome of *Micrococcus* sp. strain R8502A1 is approximately 2.47 Mb, a size at which PKSs and NRPSs are rarely, if ever, detected (Donadio et al., 2007). Yet repeated disk diffusion assays consistently showed very potent growth inhibition against various *Mycobacterium* spp. including *M. tb* in this study. There are several possible reasons for this discrepancy in activity observed versus BGCs identified. It is possible that the compound inhibiting *M. tb* is a significantly modified version of a known anti-TB compound, to the point that it evades detection with genome mining algorithms. It is also possible that the compound is neither an NRPS nor a PKS, but a more elusive compound of a group with more scaffold versatility such as an alkaloid, quinone or xanthone that is not as readily detectable by typical genome mining algorithms (Schorn et al., 2016).

The putative BGC detected by antiSMASH with the closest similarity to a known compound is a terpene cluster that likely synthesizes a carotenoid. Although mainly studied for their antioxidant and anticancer properties, there are at least three known carotenoids with anti-TB activity. One compound, fucoxanthin, is a marine-derived orange xanthophyll produced by both brown seaweed and diatoms rather than a bacterium (Peng et al., 2011; Šudomová et al., 2019). Fucoxanthin is highly abundant in the marine environment and is estimated to contribute more than 10% of

total carotenoid production (Liaaen-Jensen et al., 1978; Viera et al., 2018). The other two carotenoids, flexirubin [a yellow-orange pigment isolated from *Flavobacterium* sp. Ant342 (F-YOP)] and violacein [a violet purple pigment isolated from *Janthinobacterium* sp. Ant5-2 (J-PVP)], originate from a freshwater lake in Antarctica (Mojib et al., 2010). Violacein, isolated from bacteria including *Chromobacterium violaceum,* has shown antimycobacterial activity (de Souza et al., 1999; Durán & Menck, 2001; Mojib et al., 2010). Fucoxanthin targets *M. tb* by interfering with cell wall biosynthesis (Šudomová et al., 2019). The mechanism of action of flexirubin and violacein remain undocumented, although one study notes that the antibacterial mechanism of violacein against methicillin-resistant *Staphylococcus aureus* (MRSA) differs from that of other common antibiotics (Choi et al., 2015, 2021). Furthermore, studies show that fatty acid-carotenoid complexes isolated from the microalga *Chlorella vulgaris*, composed of oleic acid or lineolic acid and various carotenoids including canthaxanthin, neoxanthin, cryptoxanthin and echinenone, can act as potent therapeutic agents against multi-drug resistant strains of *M. tb* (Kumar et al., 2020). Essential genes for the biosynthesis of fucoxanthin were not detected in the *Micrococcus* sp. strain R8502A1 genome by any genome mining tools nor by manual investigation of the PATRIC annotation output. The BGCs for violacein and flexirubin are in the MIBiG database, and neither were detected by antiSMASH. There are no studies to date investigating the antibacterial properties of sarcinaxanthin, the only complete carotenoid that this study concluded that *Micrococcus* sp. strain R8502A1 is capable of synthesizing (at least in theory).

The close proximity of the putative betalactone and NAPAA clusters to each other and their relatively low similarity to known compounds suggests that the domains identified between both clusters may in fact comprise a novel hybrid cluster. It is possible that the AT and tailoring oxidoreductase domains identified in the putative betalactone cluster work with the Re or dimerisation domains and phosphopantetheinyl transferase domain detected in the putative NAPAA cluster by antiSMASH to synthesize a novel NRPS-PKS hybrid compound. Further support for at least a partial NRP is the detection of one modular NRPS with a Te domain on contig 17 by NP.Searcher. Support for at least a partial PK is provided by the detection of two KS domains (nodes 1 and 31) by NaPDoS, which were not detected by antiSMASH but may also be involved in synthesizing this same compound. Further complicating detection is the possibility that the cluster is nonmodular and contains domains acting in trans (associated with type II PKSs and type C nonlinear NRPSs). In this case, genes located in distant parts of the genome may be working together to synthesize a compound that evades detection by current genome mining algorithms. From all the genome mining tools employed in this study and PATRIC annotations, a total of one PPT attachment site and PPTase, two AT domains, one acyl carrier protein (ACP) domain, one PCP domain, two KS domains, one Te domain, two Re domains (or one Re and one dimerisation domain), and one aminotransferase domain were detected throughout the entire *Micrococcus* sp. strain R8502A1 genome. The distance between individual domains, coupled with the detection of a putative KS-beta domain (chain length factor) by NaPDoS strongly suggests that this hybrid cluster is not only trans, but also iterative. The likelihood of

134

an iterative cluster is further supported by the sparse BGC domains detected. Further analysis is necessary to determine exactly how many BGCs *Micrococcus* sp. strain R8502A1 is capable of synthesizing, as well as which domains act together to synthesize a single compound. Previous investigations have identified a PK (ECO-02301) that requires more than 120 domains for synthesis and an NRP (syringopeptin) that requires 68 domains for synthesis and spans approximately 73.8 kb of DNA (Fischbach & Walsh, 2006; McAlpine et al., 2005; Scholz-Schroeder et al., 2003). Such a large BGC is unlikely to be the case here considering the small size of the *Micrococcus* genome and the fact that more than 90% of actinomycetes devote less than 7.5% of their genome to secondary metabolite production (Challis and Hopwood, 2003; Cimermancic et al., 2014; Nazari et al., 2017).

Despite the current limited literature on known *Micrococcus* strains and the compounds they produce, the phylogenetic analysis is interesting in regard to the relationship (or lack thereof) between bioactivity and taxonomy. It should be noted that all of the aforementioned excluded strains (*M. luteus* strain MKVKUD 2013, *Micrococcus* sp. MP76, *Micrococcus* sp. OUS9 and *Micrococcus* sp. strain SB58) are of marine origin, with strain SB58 being isolated from a marine sponge (Hentschel et al., 2001; Karbalaei-Heidari et al., 2020; Shanthi Kumari et al., 2020; Umadevi & Krishnaveni, 2013). Exclusion of these marine strains from phylogenetic analysis most certainly precludes a comprehensive understanding of the chemotaxonomic relationship among *Micrococcus* isolates. All novel strains previously isolated from *X. muta* by Montalvo et al. (2005) (strains R8502A1, XM4230A, and XM4230B) cluster together, which is expected due to the fact that they were isolated from the

same region. *Micrococcus yunnanensis* F-256,446 was also isolated from a marine sponge from the Florida Keys and clusters with the novel isolates, along with terrestrial strains isolated from a medicinal plant from Saudi Arabia (either *Coleus forskohlii* or *Plectranthus tenuiflorus*) (*M. luteus* strain TUB6), Indian soil (*M. luteus* strain ODB36), the air in a subterranean Spanish show cave (*M. luteus* strain 0310ARD7G_6), and several strains isolated from Arctic marine sediment (*Micrococcus* spp. strain KRD153 and KRD096) (El-Deeb et al., 2012; Fernandez-Cortes et al., 2011; Soldatou et al., 2021). *Micrococcus* sp. strain KRD153 was shown to produce an extract with antibacterial activity against *Staphylococcus aureus, Escherichia coli* and *Pseudomonas aeruginosa*. *Micrococcus* sp. strain KRD093 is active against *E. coli* (Soldatou et al., 2021). *Micrococcus luteus* strain TUB6 generates a crude extract with antimicrobial activity against the human pathogen *Proteus mirabilis* (El-Deeb et al., 2012). Within a larger clustering is *M. yunnanensis* strain rsk5, isolated from a plant from India (*Catharanthus roseus*) and shown to produce a crude extract with broad-spectrum antibacterial activity, along with three more marine strains isolated from Polar regions (Ranjan & Jadeja, 2017; Soldatou et al., 2021). The extract of *Micrococcus* sp. strain KRD026 was shown to inhibit *S. aureus* and *E. coli*, while no antibacterial activity was observed from the extracts of strains KRD128 or KRD012 (Soldatou et al., 2021). The study analyzing *M. yunnanensis* F-256,446 as well as two other *Kocuria* strains for their production of the anti-MRSA thiopeptide kocurin noted that although all three microbes were isolated from the same region and found to produce the same compound, they all retain different metabolic gene amplification patterns (Palomo et al., 2013). Coupled

136

with the fact that actinomycetes of distantly related genera isolated from Antarctica

have also been found to produce kocurin, this suggests that geographic proximity

does not necessarily correlate with chemosimilarity (Palomo et al., 2013). All six

strains included in the phylogenetic analysis (both terrestrial and marine isolates) that

have been identified as having antibacterial properties cluster together. This may be

merely coincidental, and we cannot make any conclusions based on this observation

until activity is confirmed or excluded from all other strains included in the analysis.

Again, this highlights the lack of data regarding the biosynthetic potential of

Micrococcus strains and emphasizes the need for continued research in this area. One

final and essential note on the phylogenetic analysis performed is that the results are

entirely dependent on the accuracy of the sequence data provided. Without access to

chromatograms of the sequenced 16S rRNA gene fragments, it is difficult to ascertain

the correctness of the published sequences. Several sequences were removed during

this analysis based on highly unlikely base diversions, yet we cannot be certain that

the provided strains are 100% accurate. In any case, the analysis performed here

provides an insightful, general look into the taxonomic relationship of *Micrococcus*

strains isolated from widely disparate environments.

To ensure a thorough probe of BGCs in *Micrococcus* sp. strain R8502A1,

long-read sequencing was performed with PacBio. This method enabled assembly of

a complete genome void of gaps, a feat impossible to achieve with short-read

sequencing data of genomes with long repetitive regions and high GC content, both

characteristics of actinomycete genomes. This is evidenced by the fact that

antiSMASH analysis of the SPAdes assembly identified three putative clusters on

contig edges, all of which were completely resolved within the Canu assembly. Interestingly, genome mining of the Canu assembly did not reveal any more information than the SPAdes assembly, making identification of the anti-TB compound all the more puzzling. Despite the numerous genome mining tools employed to search for putative BGCs, the compound responsible for the anti-TB activity observed remains unknown. The detection of a possibly novel hybrid PKS-NRPS cluster in this study provides insight regarding the bioactivity of this *Micrococcus* strain yet it highlights the limitations of this approach to drug discovery. Currently, there are no *Micrococcus*-derived BGCs in the MIBiG database nor in the BAGEL4 database. In addition, to date, there are no hybrid PKS-NRPS clusters described from *Micrococcus* species. Evidently, genome mining is insufficient as a stand-alone tool, and this shortcoming is exacerbated by the paucity of research into the biosynthetic potential of *Micrococcus* spp.

Drug discovery is an arduous undertaking that demands collaboration between microbiologists, chemists and bioinformaticians, among others. Biology informs chemistry and vice versa in a continuous cycle to further isolate the bioactive compound and confirm activity. After the first round of screening for growth inhibition, bioassay-guided fractionation (with 40%, 70% and 100% methanol) was performed on *Micrococcus* sp. strain R8502A1 extracts. Both fractions "70" and "100" were shown to retain inhibitory activity when tested against *M. tb* H37Ra. Chemical analysis is now required to isolate the compound and elucidate its structure. Furthermore, to characterize the biosynthetic pathway of the compound responsible for inhibiting growth of *M. tb*, stable isotope feeding studies should be conducted.

These experiments can elucidate the compound corresponding to the modular NRPS with the amino acid sequence Phe Gly Pro predicted by NP.Searcher (Ikeda et al., 1992; Klitgaard et al., 2015; McCaughey et al., 2022). This method can also be used to determine whether the non-mevalonate terpenoid detected by NP.Searcher (containing a tyrosine) is responsible for the antimycobacterial activity observed. Nevertheless, this method must be coupled with analytical chemistry techniques, such as HPLC, mass spectrometry (MS), and nuclear magnetic resonance (NMR).

Chapter 4: Characterization of potential chemical analogues of the potent anti-TB diazaquinomycin from a marine *Micromonospora* sp.

4.1 Abstract

Rare actinomycetes are increasingly becoming the focus of drug discovery efforts, due to their high potential for chemical diversity and the fact that they have been historically underexploited. Genomes of the rare actinomycete genus *Micromonospora* cumulatively encode over 2,300 biosynthetic gene clusters (BGCs) linked to more than 1,000 distinct BGC-families, of which more than 98.5% show significant divergence from known BGCs (Hifnawy et al., 2020). Previous analysis has identified a freshwater *Micromonospora* strain to produce potent and selective anti-tuberculosis (TB) analogs of the diazaquinomycin compound class (Mullowney et al., 2015), but poor solubility of the drug prohibits further development into an effective anti-TB therapeutic (Braesel et al., 2019). Investigation of a novel assemblage of marine sponge-derived actinomycetes screened for production of compounds that inhibit *Mycobacterium tuberculosis* (*M. tb*) identified a novel strain of *Micromonospora* with potent inhibitory activity. Genome mining revealed the presence of a cluster with 94% similarity to the biosynthetic pathway for diazaquinomycin H/J. Chemical analysis of microbial extracts by liquid chromatography-mass spectrometry (LC-MS) resulted in the detection of four ions that may correlate to diazaquinomycin structures, two of which correspond to chemical formulas not associated with any previously identified analogs. In the hopes of discovering analogs with improved solubility for drug development, further analysis was performed to determine the uniqueness of these potentially novel diazaquinomycins. Additional chemical and spectroscopic analyses are required to confirm the structures of these detected molecular ions.

4.2 Introduction

In the search for novel bioactive compounds of pharmaceutical relevance, research is gradually shifting from commonly studied *Streptomyces* species to the less frequently cultured actinomycete genera. *Micromonospora* is a rare genus of slow-growing, spore-forming actinomycetes belonging to the *Micromonosporaceae* family. First described in 1923, this family is now known to encompass 29 genera and at least 126 validly published *Micromonospora* species and subspecies (lpsn.dsmz.de, accessed Feb 2022). Notably, this family includes the first seawater-obligate actinomycete isolate to ever be described (*Salinospora* sp.) (Inahashi et al., 2010; Maldonado et al. 2005; Mincer et al., 2002; Ørskov, 1923). *Micromonospora* have the ability to degrade various complex materials including latex, cellulose and chitin, suggesting an essential role in organic matter decomposition in the environment (de Menezes et al., 2008; Erikson, 1941; Jendrossek et al., 1997; Jensen, 1930). A 2013 review of marine actinomycetes screened for activity over a five-year period reported 80 new rare *Actinomycete* species, of which *Micromonosporaceae* (family) isolates dominated as sources of novel chemical diversity (Subramani & Aalbersberg, 2013). Furthermore, a more recent analysis taking an extensive look at the biosynthetic potential of *Micromonospora* strains found that species within the genus are capable of producing more than 2,300 biosynthetic gene clusters (BGCs) that can be grouped into 1,033 BGC-families. Less than 1.5% of these BGC-families show identity or high similarity to known BGCs, supporting continued investigation into this genus as a source of novel secondary metabolites (Hifnawy et al., 2020). Notable antibiotics discovered from *Micromonospora* strains include the antibacterial gentamicin and

antitumor enediynes (Lee et al., 1989; Weinstein et al., 1963). More recently investigated marine isolates have also been discovered to produce novel compounds with therapeutic potential including the potent antitumor rakicidins as well as the antibacterial tetrocarcin Q (Chen et al., 2018; Gong et al., 2018; Sun et al., 2019). For an up-to-date and in-depth review of compounds known to be produced by marine *Micromonospora*, see Qi et al. 2020.

Although ubiquitous throughout terrestrial, freshwater and marine ecosystems, *Micromonospora* are highly prevalent in freshwater habitats, particularly lake sediments (Cross, 1981; Erikson, 1941; de Menezes et al., 2012). Research into a *Micromonospora* strain isolated from Lake Michigan sediment (*Micromonospora* sp. B026) led to the discovery of two compounds with potent inhibitory activity against *Mycobacterium tuberculosis* (*M. tb*) H37Rv: diazaquinomycin H and J (Mullowney et al., 2015). Despite their potent and selective anti-tuberculosis (TB) activity, poor water solubility has significantly hindered their development into an effective drug against *M. tb* (Braesel et al., 2019)*.* This hydrophobicity is in part due to the various alkyl side chains associated with diazaquinomycins (Braesel et al., 2019; Rinehart & Renfroe, 1961).

This work investigated a novel collection of marine sponge-derived actinomycetes and identified a strain of *Micromonospora* (*Micromonospora* sp. strain R45601) that inhibits *M. tb* H37Ra, and BGC analysis reveals a putative cluster with 94% similarity to diazaquinomycin H/J. Given this highly similar but not identical cluster, in addition to the fact that these strains inhabit totally different environments, it is very likely that this sponge symbiont produces a novel diazaquinomycin analog.

A dual genomics and chemistry-enabled approach was undertaken to assess the uniqueness of this compound. Further analysis of the extract of *Micromonospora* sp. strain R45601 was performed to determine the structure of this putative diazaquinomycin analog with particular attention to functional groups present.

4.3 Materials and methods

4.3.1 Cultivation of *Micromonospora* and *Mycobacterium* strains

Micromonospora sp. strain R45601 was previously isolated from *Xestospongia muta* and stored as described by Montalvo et al. (2005). A cryovial of the isolate was plated out on both International *Streptomyces* Project 2 (ISP2) agar (BD-Difco™, Franklin Lakes, NJ, USA) supplemented with 20% salt (granular sodium chloride - J.T. Baker, Phillipsburg, NJ, USA) and Reasoner's 2A Agar (R2A) (BD - Difco™, Franklin Lakes, NJ, USA). Plates were incubated at 30°C until growth of individual colonies could be observed. Two individual colonies per plate were then transferred to 100 mL of ISP2 or Reasoner's 2A Broth (R2B) (EZ-Media - Microbiology International, Frederick, MD, USA) in 250 mL baffled flasks to provide sufficient aeration, and incubated at 30°C with shaking at 150 rpm for a minimum of two weeks, until cultures appeared dense. ISP2 cultures (100 mL) were incubated for up to six months before preparing extracts to assess stability of compound production and potency. Larger volumes of 500 mL were cultured for up to six months to generate more extract mass. ISP2 liquid medium was prepared as previously described in section 2.3.1.

Mycobacterium tuberculosis H37Ra (avirulent), *Mycobacterium marinum*
ATCC 927 and *Mycobacterium smegmatis* MC^2 155 were all plated from cryovials as
previously described in section 2.3.1. *M. marinum* ATCC 927 and *M. smegmatis* MC^2
155 were incubated at 30°C with shaking at 150 rpm and *M. tuberculosis* H37Ra was
incubated at 37°C with shaking at 150 rpm.

4.3.2 Preparation of extracts and antimycobacterial activity assay

See section 2.3.2 for a description of how organic extracts were prepared and
tested against *Mycobacteria*.

4.3.3 Genomic DNA extraction, identification and whole genome sequencing

Micromonospora sp. strain R45601 was assigned taxonomic classification as
previously described in section 2.3.3. Sequence errors were corrected manually by
visual inspection of chromatograms.

Genomic DNA was sequenced by using a MiSeq sequencer (Illumina) using
the MiSeq version 2.4.0.4 reagent kit. The Nextera XT Library Prep Kit (1 ng DNA)
was used to prepare the sequencing library (2 x 250 bp paired end reads for a total of
500 cycles).

4.3.4 Genome assembly, annotation and biosynthetic gene cluster analysis

Assembly of Illumina MiSeq reads was performed *de novo* using
Trimmomatic version 0.30 (Bolger et al., 2014) and assembled using SPAdes version
3.14.1 (Bankevich et al., 2012). Initial assembly statistics were evaluated with the

Quality Assessment Tool for Genome Assemblies (QUAST) (Gurevich et al., 2013).

Contigs were filtered as previously described in section 2.3.4. Genome annotation

was carried out using PATRIC version 3.5.41 (Brettin et al., 2015; Davis et al., 2020).

The final assemblies were validated by evaluating contamination and completeness

values, calculated using CheckM version 1.0.18 app through KBase (kbase.us) (Parks

et al., 2015). Scaffolding of the genome was performed as previously described in

section 2.3.4. BGCs were identified using antiSMASH (antibiotics & Secondary

Metabolite Analysis Shell) version 5.0 in relaxed mode (Blin et al., 2019b).

4.3.5 HPLC/LC-MS analyses of extracts

The extract generated from *Micromonospora* sp. strain R45601 was dissolved

in 100 µL methanol (MeOH) for LC-MS analysis. A 10 µL sample was diluted to

50% MeOH and 5 µL was injected for LC-MS analysis. HPLC analysis was

performed on an UltiMate™ 3000 Cap-LC series system (Dionex, Sunnyvale, CA,

USA) equipped with an XSelect HSS T3 RP-18 column (1.0 x 100 mm, 3.5 µm

particle size, Waters) at 40°C coupled with a Q Exactive Plus mass spectrometer

(Thermo Scientific, Waltham, MA, USA). The mobile phases consisted of 0.1%

formic acid and water (A) and 0.1% formic acid and acetonitrile (B). The flow rate

was 0.2 mL/min, and the gradient program was as follows: 5% B (0 min), 5-95% B

(0–20 min).

Ionization was achieved using electrospray ionization (ESI) in positive ion

mode (mass range: *m/z* 100 – 1,500). Mass spectrometry was performed with an ESI

source in the positive-ionization mode with a sheath gas of 15 arbitrary units (arb)

and auxillary gas of 10 arb. The ion spray voltage was set at 5,000 V, and the source

temperature was 320°C.

LC/MS/MS data were analyzed by Progenesis QI (Waters, Milford, MA). The

p value of the ANOVA test was calculated to determine significance of peak

identification, and a p value < 0.05 was considered significant. MS/MS fragment data

was also used to search the National Institute of Standards and Technology (NIST)

database.

For HPLC-UV data, 90 µL of sample was injected into a Primaide system

(Hitachi, Chiyoda City, Tokyo, Japan) equipped with a Primaide 1430 Diode Array

Detector, Primaide 1210 Autosampler, Primaide 1110 pump with a degasser, and an

octadecyl-silica (C18) column (4.6x240mm, 5 µm particle size, Ultrasphere®). The

mobile phase consisted of 1% formic acid and water (A) and 1% formic acid and

acetonitrile (B) with a flow rate of 1 mL/min and a stepwise gradient as follows: 25-

35% B (5 min), 35-40% B (5 min), 40-60% B (5 min), 60-70% B (4 min), 70-80% B

(4 min), 80-100% B (6 min), and a re-equilibration step at 100% B (5 min).

4.4 Results

4.4.1 Characterization of strain and extract

Micromonospora sp. strain R45601 colonies on ISP2 agar grow initially with

an orange coloring that changes from brown to black, with an increasingly oily-

looking appearance as the colonies age (Fig. 4.1). The colonies have a fungal-like

appearance due to their production of aerial and substrate mycelia (Fig. 4.1)

(Maldonado et al., 2015). Pigment is diffusible as evidenced by darkening of the agar

or liquid medium over time. In liquid medium, cultures grow in clumps that appear orange/pink and eventually turn black (Fig. 4.1). Strain R45601 grew well on both ISP2 + 20% NaCl and ISP2 without added salt. ISP2 control discs inoculated with 250 μg displayed a strange clearance zone followed several days later by *M. tb* growth surrounding the disc. ISP2 is a rich medium, and the possibility that medium components may be inhibiting *M. tb* H37Ra has to be considered. This seems very unlikely because not all actinomycete strains cultured in ISP2 (tested along with *Micromonospora* sp. strain R45601) displayed growth inhibition. Nonetheless, to ensure that this medium-derived compound was not the same compound inhibiting *M. tb* H37Ra growth in the *Micromonospora* extract, strain R45601 was also cultured in the low-nutrient R2B. The R2B-cultured *Micromonospora* extract tested at a concentration of 200 μg/disc was observed to inhibit *M. tb* H37Ra, while no inhibition was observed from the R2B control disc (tested at 250 μg/disc). This confirmed that strain R45601 produces a compound with anti-TB activity.

Cultures incubated for two weeks and up to six months display inhibition against *M. tb* H37Ra with varying strength, however, no correlation between incubation time and potency of extract was observed, nor any correlation between mass yield of extract and potency (data not shown). Growth-inhibiting activity of *M. tb* H37Ra by two different concentrations of a potent *Micromonospora* sp. strain R45601 extract prepared in this study is displayed in Figure 4.2.
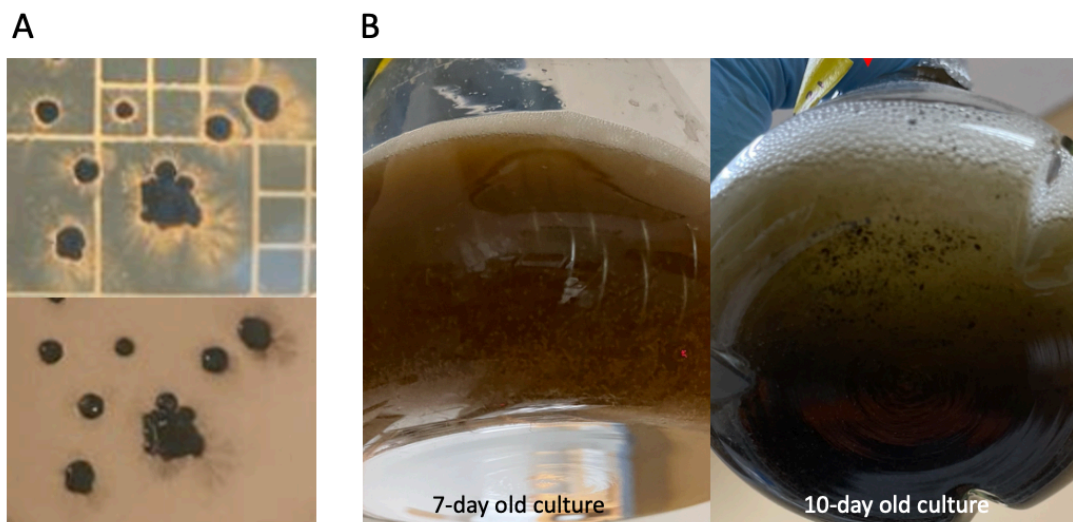
Figure 4.1. Observed colony morphology of *Micromonospora* sp. strain R45601 on ISP2 agar (A) and in liquid ISP2 medium (B). On agar, colonies appear oily and present substrate and aerial mycelia. Colonies were photographed by using an illuminated colony counter as well as in natural light. In liquid, cultures form tiny "ball-like" aggregates that darken from orange to black over time.
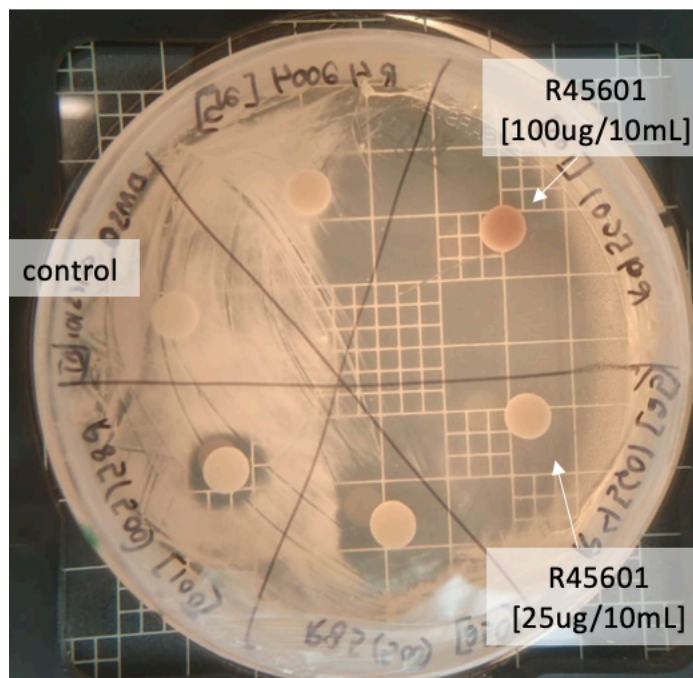
Figure 4.2. Growth-inhibition of *M. tb* by different concentrations of a potent *Micromonospora* sp. strain R45601 extract.

149

4.4.2 Genome assembly and biosynthetic gene cluster analysis

Partial 16S rRNA gene sequence (1375 bp) returned a BLAST hit with 100%
identity to multiple *Micromonospora* strains, with the top hit being *Micromonospora*
chalcea strain IMB16-203 (AN: MG190678.1). The draft genome consists of 207
contigs that assemble to 6.7 Mb with a GC content of 72.92%, consistent with the
high GC bias associated with actinomycetes. A circular view of the Illumina MiSeq-
sequenced genome is presented in Figure 4.3, and assembly statistics are presented in
Table 4.1. Scaffolding of the SPAdes assembly by MeDuSa by comparison to the
complete genome of *Micromonospora* sp. WMMA2032 yielded 14 scaffolds.
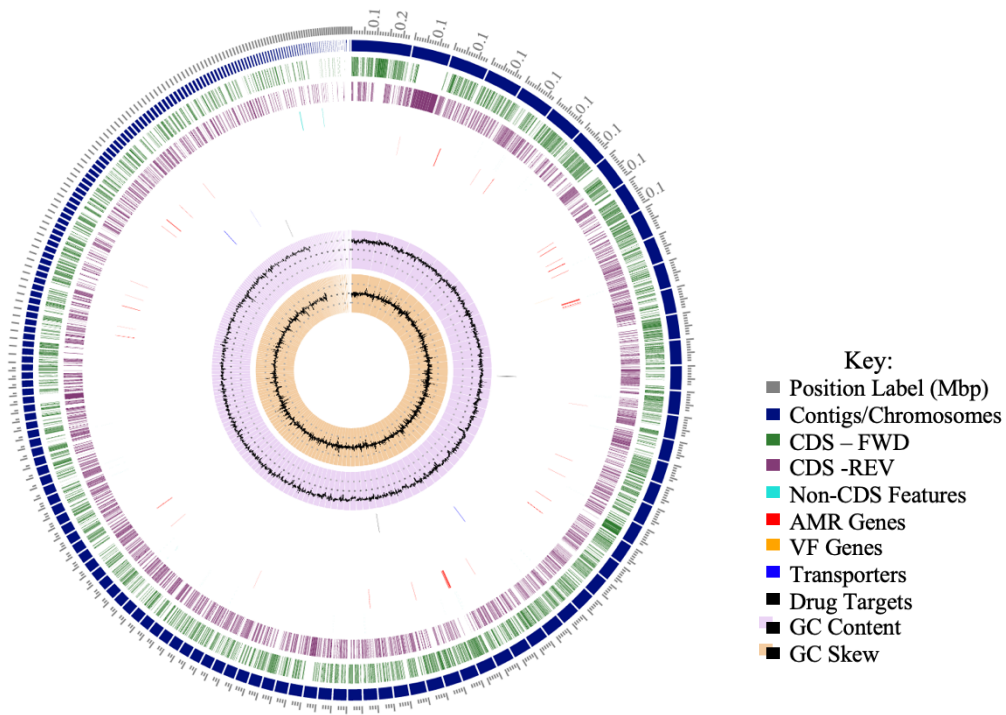
**Figure 4.3. Circular view of assembled genome of *Micromonospora* sp. strain
R45601.** Note: Key adapted from PATRIC.

Table 4.1. Assembly statistics of the *Micromonospora* sp. strain R45601 genome.

Assembly Statistics	
Total Length	6,712,873
Contigs	207
GC	72.92%
N50	64,854
Genes	6,357
Protein Coding Seqs	6,301
Coverage	30x
Structural RNA (tRNA/rRNA)	51/5
Completeness	100
Contamination	1.84

A total of 25 putative BGCs were identified by antiSMASH, although only

three clusters had considerable similarity (> 40%) to known clusters in the MIBiG

database (Table 4.2). Of these, the putative T1/T3 polyketide synthase

(PKS)/nonribosomal peptide synthetase (NRPS) cluster with similarity to

diazaquinomycin H/J stood out for its very high identity (94%). The full annotation of

this entire putative cluster consisting of 74 genes and spanning 117,500 bp is

provided in Figure 4.4 and Table 4.3. Furthermore, a direct comparison of the

putative cluster with the diazaquinomycin H/J cluster is provided to discern what

distinguishes this cluster from the known diazaquinomycin biosynthetic pathway

(Fig. 4.5). Almost half of the entire length of the putative cluster (> 117,500 bases)

uniquely consists of mixed PKS and NRPS-associated genes (> 62,000 bases), none

of which are not present in the known diazaquinomycin H/J cluster.

Table 4.2. Putative BGCs in the *Micromonospora* sp. strain R45601 genome identified by antiSMASH.

Location in Genome	Putative Cluster Type	Most Similar Known Cluster (% similarity)
contig 1 (90,094 – 207,644)	T1PKS, NRPS-like, NRPS, T3PKS, other	Diazaquinomycin H/J (94%)
contig 2 (19,587 – 149,555)*	T1PKS	ECO-02301 (42%)
contig 3 (12,983 – 58,754)	T1PKS	Calicheamicin (22%)
contig 3 (74,516 – 108,064)	LAP	NA
contig 6 (54,436 – 91,030)	Thiopeptide, thioamitides, LAP	NA
contig 8 (39,213 – 116,253)	Thioamide-NRP, Siderophore	Azicemicin B (13%)
contig 9 (36,083 – 75,143)	Oligosaccharide, terpene	Lobosamide A/B/C (13%)
contig 11 (70,196 – 100,899)*	T3PKS	Alkyl-O-dihydrogeranyl-methoxyhydroquinones (71%)
contig 13 (6,038 – 30,570)	Terpene, RiPP-like	Lymphostin/neolymphostinol B/lymphostinol/neolymphostin B (30%)
contig 14 (3 – 62,058)	NRPS-like, NRPS, T1PKS, PKS-like	Naphthyridinomycin (14%)
contig 16 (29,823 – 50,863)	Terpene	NA
contig 22 (60,033 – 81,937)	RRE-containing	NA
contig 23 (14,877 – 80,116)*	NRPS, T1PKS, NRPS-like	Rifamycin (35%)
contig 32 (39,210 – 53,944)	NAGGN	NA
contig 40 (30,793 – 51,719)	Terpene	Isorenieratene (25%)
contig 44 (3,319 – 51,746)*	NRPS, T1PKS, NRPS-like	Crochelin A (7%)
contig 46 (1 – 13,770)*	Lanthipeptide class I	NA
contig 66 (1 – 37,575)*	T2PKS	Pradimicin A (25%)
contig 79 (10,636 – 28,979)*	Terpene	Phosphonoglycans (3%)
contig 83 (1 – 25,839)*	T1PKS	Cremimycin (20%)
contig 101 (1 – 18,074)*	RRE-containing	Actagardine (6%)
contig 103 (4,942 – 19,327)*	Terpene	Nocathiacin (4%)
contig 143 (1 – 10,858)*	Lanthipeptide class III	NA
contig 171 (1 - 4,249)*	T1PKS	NA
contig 178 (1 – 3,305)*	NRPS-like	NA

Note: Putative clusters located on a contig edge are denoted with a symbol (*). Key: T1PKS = type I PKS, NRPS = non-ribosomal peptide synthetase, LAP = linear azol(in)e-containing peptides, T3PKS = type III PKS, RiPP = ribosomally synthesized and post-translationally modified peptide, PKS = polyketide synthase, RRE = Rev response element, NAGGN = N-acetylglutaminylglutamine amide, T2PKS = type II PKS.
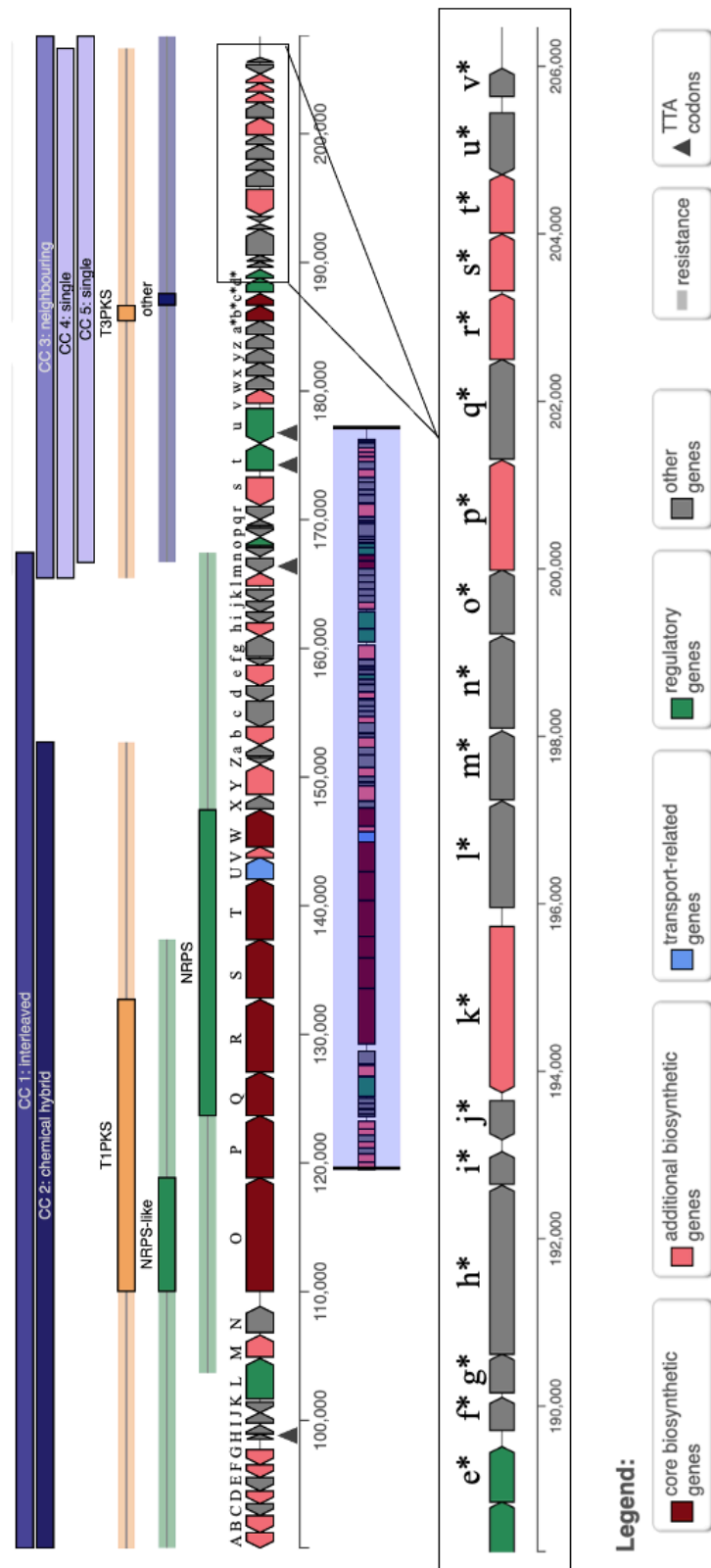
Figure 4.4. Putative BGC of a hybrid NRPS/PKS with similarity to diazaquinomycin H/J in the *Micromonospora* **sp. strain R45601 genome identified by antiSMASH.** Note: Legend as provided by antiSMASH.

Table 4.3. Annotated genes of putative hybrid NRPS/PKS BGC with similarity to diazaquinomycin H/J in the *Micromonospora* sp. strain R45601 genome identified by antiSMASH.

Gene	Putative Function	Gene	Putative Function
A	Acyltransferase 3	l	3-oxoacyl-(acyl carrier protein) synthase III
B	Glycosyltransferase, MGT family	m	DDE superfamily endonuclease
C	Diacylglycerol kinase catalytic domain	n	NmrA family protein
D	NAD-dependent epimerase/dehydratase	o	TetR family transcriptional regulator
E	AIR synthase related protein	p	NADPH-dependent FMN reductase
F	ubiA prenyltransferase	q	Domain of unknown function (DUF4440)
G	8-amino-7-oxononanoate synthase OR Aminotransferase class I and II	r	Nitronate monooxygenase
H	Glyoxalase-like domain	s	MMPL domain-containing transport protein
I	Peptidase family M23	t	LuxR family transcriptional regulator
J	Unknown	u	LuxR family transcriptional regulator
K	S-adenosyl methyltransferase	v	Polyprenyl synthetase
L	Transcriptional regulator, SARP family	w	GHMP kinases
M	Monooxygenase FAD-binding	x	GHMP kinases N terminal domain OR Mevalonate 5-diphosphate decarboxylase C-terminal domain
N	Unknown	y	GHMP kinases
O	T1PKS: hyb_KS, T1PKS_AT, NRPS-like:AMP-binding, NRPS-like:PP-binding, Beta-ketoacyl synthase	z	FMN-dependent dehydrogenase
P	T1PKS: hyb_KS, T1PKS_AT, NRPS-like:AMP-binding, NRPS-like:PP-binding, Beta-ketoacyl synthase	a*	Hydroxymethylglutaryl-coenzyme A reductase
Q	NRPS: Condensation, NRPS:AMP-binding, PP-binding, polyketide synthase associated protein papA3	b*	T3PKS: Chal_sti_synt_N OR hydroxymethylglutaryl-CoA synthase
R	T1PKS: hyb_KS, T1PKS_AT, NRPS-like:AMP-binding, NRPS-like:PP-binding, Beta-ketoacyl synthase	c*	Aromatic prenyltransferase Orf2
S	NRPS: Condensation, NRPS:AMP-binding, PP-binding, AMP-dependent synthetase and ligase	d*	Sensor histidine kinase
T	NRPS: Condensation, NRPS:AMP-binding, PP-binding, AMP-dependent synthetase and ligase	e*	LuxR family DNA-binding response regulator
U	ABC transporter ATP-binding protein	f*	Unknown
V	Abhydrolase_6 OR thioesterase	g*	Unknown

Table 4.3, continued

W	NRPS: Condensation, NRPS:AMP-binding, AMP-dependent synthetase and ligase	h*	Multicopper oxidase
X	Taurine catabolism dioxygenase TauD, TfdA family	i*	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily
Y	MMPL domain-containing transport protein	j*	Unknown
Z	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily	k*	Isochorismate synthase
a	Peptidase family M23	l*	Amino acid kinase family, ACT domain
b	Argininosuccinate lyase/adenylosuccinate lyase	m*	DeoC/LacD family aldolase
c	FAD-NAD(P)-binding	n*	3-dehydroquinate synthase II
d	FAD binding domain, hypothetical protein	o*	Amidohydrolase
e	AMP-dependent synthetase and ligase	p*	AMP-binding
f	Polyketide cyclase / dehydrase and lipid transport	q*	FAD binding domain, hypothetical protein
g	GGDEF-like domain OR PucR C-terminal helix-turn-helix domain	r*	Short-chain dehydrogenase/reductase SDR
h	Crotonyl-CoA reductase / alcohol dehydrogenase	s*	Isochorismatase
i	N-acetyltransferase	t*	Short-chain dehydrogenase/reductase SDR
j	N-acetyltransferase	u*	Unknown
k	Amidohydrolase	v*	Protein of unknown function, DUF485
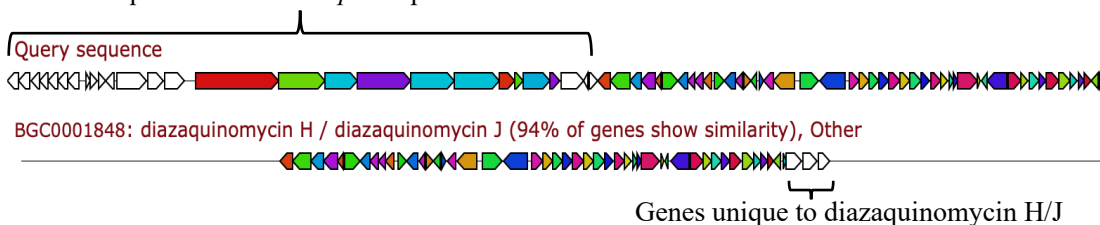
Figure 4.5. Shared genes between the putative diazaquinomycin BGC in the *Micromonospora* sp. strain R45601 genome identified by antiSMASH and known diazaquinomycin H/J cluster. Genes A – a (as labeled in Fig. 4.4/Table 4.3) are unique to the *Micromonospora* sp. strain R45601 BGC, and genes b – v* are shared with the known diazaquinomycin H/J BGC. Three genes are unique to diazaquinomycin H/J (from left to right): SSS sodium solute transporter superfamily, hypothetical protein, 2-keto-3-deoxy-D-arabino-heptulosonate-7- phosphate synthase II.

4.4.3 HPLC/LC-MS analyses of extracts

LC-MS data generated from a *Micromonospora* sp. strain R45601 extract were scanned for *m/z* values corresponding to known and hypothetical diazaquinomycin-like structures. These hypothetical structures and their chemical formulas were calculated based from existing diazaquinomycin analogs A-H and J (Fig. 4.6), with the addition of theoretical alkyl group side chains. From this analysis, four hypothetical compounds were proposed (Table 4.4). The presence of diazaquinomycins was strongly supported by HPLC-UV data (Figure 4.8). Putative diazaquinomycins eluted with a retention time of approximately 6 min.
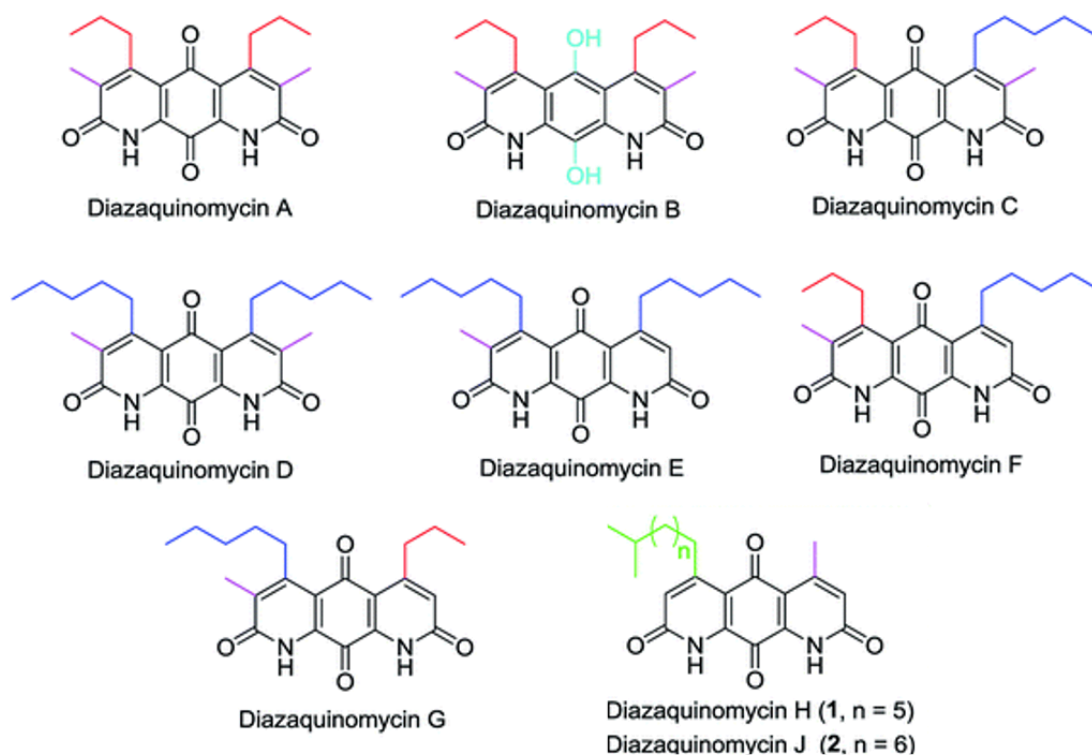
Figure 4.6. Structures of naturally occurring diazaquinomycin analogs.
Reproduced with slight adaptation from Prior & Sun (2019) with permission from the Royal Society of Chemistry.

Table 4.4. Hypothetical diazaquinomycin analogs suggested by LC-MS data.

Detected m/z	Hypothetical m/z	Likely chemical formula	Corresponding likely to:
321.0870	321.0846	$C_{16}H_{14}N_2O_4$	$[M+Na]^+$
369.1809	369.1809	$C_{21}H_{24}N_2O_4$	$[M+H]^+$
397.2133	397.2122	$C_{23}H_{28}N_2O_4$	$[M+H]^+$
489.2715	489.2723	$C_{28}H_{38}N_2O_4$	$[M+Na]^+$

Note: M = molecular mass.

The detected m/z of 369.1809 with a chemical formula $C_{21}H_{24}N_2O_4$ can be attributed to diazaquinomycin F or G, or a novel structure, due to the presence of peaks at: m/z 271.0708 corresponding to a loss of C_7H_{14} and m/z 285.0867 corresponding to a loss of C_6H_{12} (Figure 4.7). The observed loss of $(CH_2)_n$ groups (where n = number of CH_2 moieties lost), as opposed to loss of CH_3 and close proximity of the side chains to a carbonyl group on the core ring structure of diazaquinomycin suggests a McLafferty rearrangement or a similar type of rearrangement.

Similarly, the detected m/z of 397.2133 with a chemical formula $C_{23}H_{28}N_2O_4$ may be attributed to diazaquinomycin E or J, or a novel structure, due to presence of peaks at: m/z 271.0709 corresponding to a loss of C_9H_{18}, and m/z 285.0875 corresponding to a loss of C_8H_{16} (Figure 4.7). Again, loss of $(CH_2)_n$ groups and close proximity of the side chains to a carbonyl group suggests a McLafferty rearrangement or a similar type of rearrangement.

The molecular ion peaks at m/z values 321.0870 (attributed to a chemical formula $C_{16}H_{14}N_2O_4$) and 489.2715 (attributed to a chemical formula $C_{28}H_{38}N_2O_4$) do

not correspond to any known diazaquinomycin analogs and therefore may indicate

novel diazaquinomycin structures produced by *Micromonospora* sp. strain R45601.

A. MS of 321.0870 [M+Na]$^+$ (product ion scan)
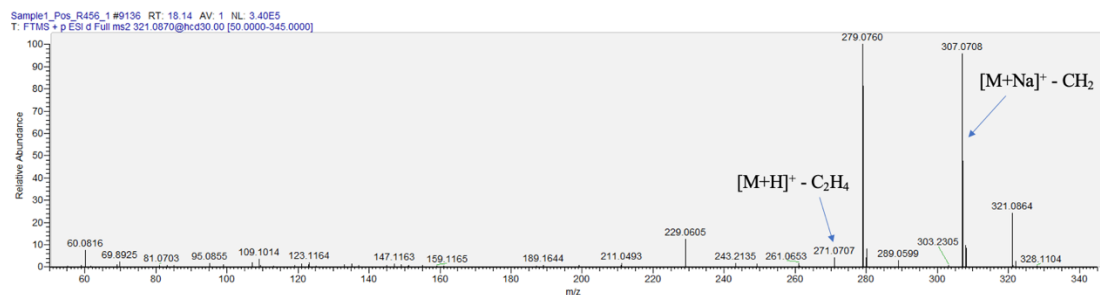
B. MS of 369.1809 [M+H]$^+$ (product ion scan)
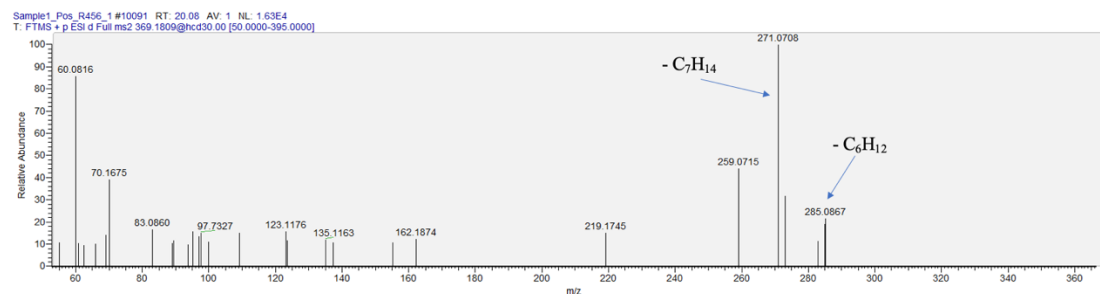
Figure 4.7. Partial MS/MS spectra of selected ions with possible fragmentation patterns observed. Fragmentation based on loss of C_nH_{2n} group (n = number of CH_2 moieties lost).

Figure 4.7, continued

C. MS of 397.2122 [M+H]$^+$ (product ion scan)
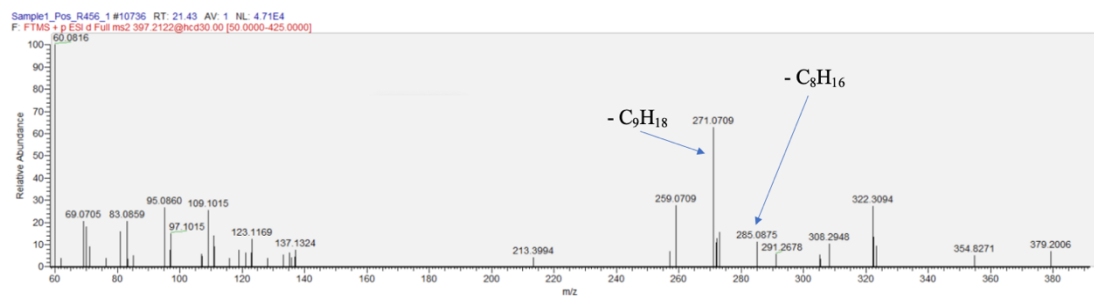
D. MS of 489.2707 [M+Na]$^+$ (product ion scan)
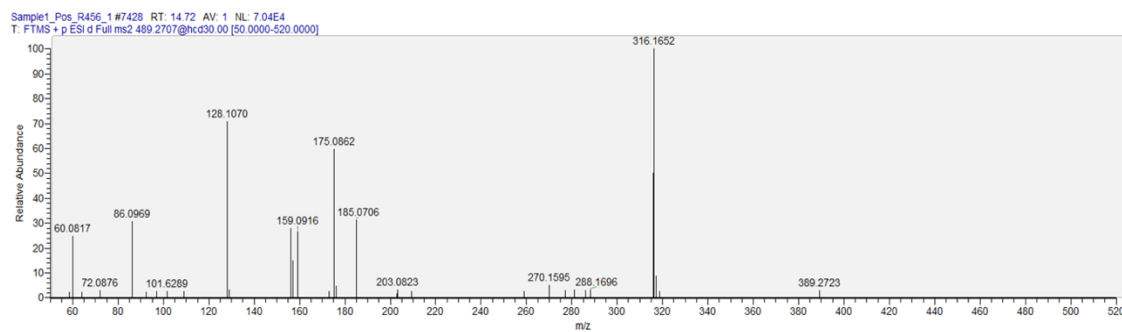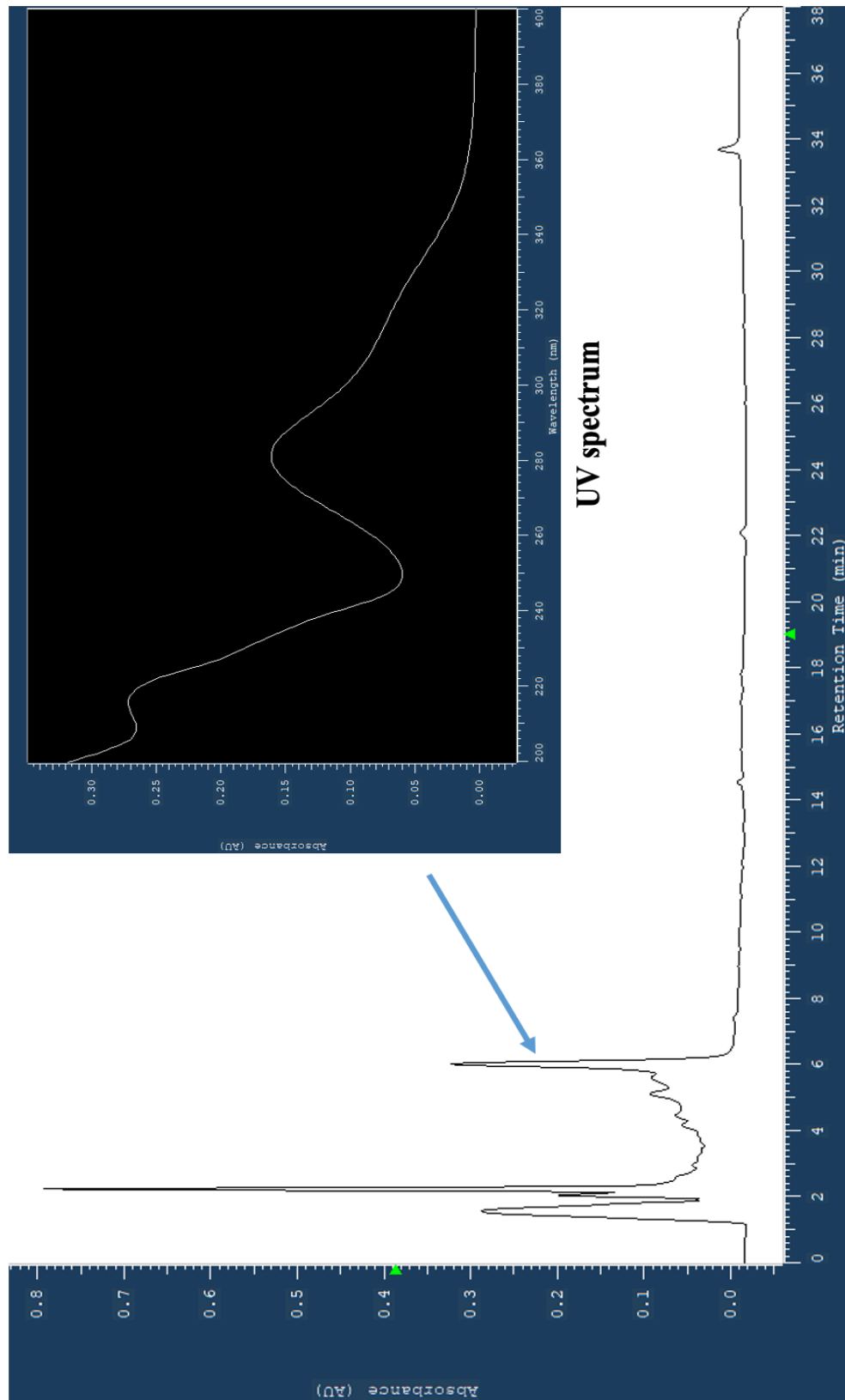
Figure 4.8. HPLC chromatogram of a *Micromonospora* sp. strain R45601 extract overlaid with the UV spectrum of the selected peak. The UV spectrum was consistent with UV spectra of diazaquinomycin analogs from the literature (Braesel et al., 2019).

4.5 Discussion

There is no argument that the genus *Micromonospora* is a prolific source of bioactive compounds. Previous reviews dating back to the 1980s document the extensive set of antibiotics that they are known to produce (in addition to compounds with other pharmaceutically-relevant bioactivities), including aminoglycosides, macrolides, ansamycins, everninomicins, actinomycins, alkaloids, and quinones (Boumehira et al., 2016; Hifnawy et al., 2020; Qi et al., 2020; Sun et al., 2019; Wagman & Weinstein, 1980). Genomic-based studies investigating the biosynthetic potential of *Micromonospora* identify strains with genomes encoding an average of 15 BGCs, with some harboring more than 28 putative BGCs (Hifnawy et al., 2020). Despite the fact that this compound class was first isolated from a *Micromonospora* strain in 2015 (Mullowney et al., 2015), the more recent reviews have not included diazaquinomycin in their discussions (Hifnawy et al., 2020; Qi et al., 2020; Sun et al., 2019).

Diazaquinomycin A, originally reported as OM-704 A, is a diaza-anthracene antibiotic first isolated from *Streptomyces* sp. OM-704 in 1982 (Ōmura et al., 1982). Diazaquinomycin B was isolated the following year, and both compounds were found to exhibit antibacterial activity against Gram-positive bacteria including *Staphylococcus aureus* and *Streptococcus faecium* (Ōmura et al., 1982, 1983). Since then, a total of nine diazaquinomycin analogs (A-H and J) have been discovered and reported to maintain the same antibacterial activity (Braesel et al., 2019; Maskey et al., 2005; Mullowney et al., 2014, 2015). In addition to antibacterial activity, several diazaquinomycin analogs are now known to be cytotoxic against various cancer cells

including drug resistant ovarian cancer and carcinoma. The mechanism of action of diazaquinomycins remains unknown. It was once believed to target mammalian and bacterial thymidylate synthase, which plays a critical role in DNA replication and repair (Braesel et al. 2019; Mullowney et al. 2015; Murata et al., 1985), but this hypothesis was discredited in experiments with *M. tb* performed by Mullowney and colleagues in 2015. Furthermore, despite extensive analysis into potential mechanisms of action, not a single molecular target of diazaquinomycins has been determined (Mullowney, 2016).

Diazaquinomycins H and J were the first diazaquinomycins isolated from a *Micromonospora* sp. (Mullowney et al., 2015). These two analogs, which were isolated from a freshwater *Micromonospora* sp. strain B026, also provided the first indication of anti-TB activity of this compound class (Mullowney et al., 2015). Diazaquinomycins are not only potent against but also selective toward *M. tb*, suggesting great promise as a novel therapeutic. Unfortunately, as previously mentioned, further development into an effective anti-TB drug has been significantly hampered by the poor solubility of these compounds in both organic and aqueous solvents (Mullowney et al., 2014; Tsuzuki et al., 1989). Efforts to improve solubility have only focused on diazaquinomycin A, with some success in increasing solubility in chloroform and in water by substituting short-chain ester and ether derivatives for C-3 and C-6 methyl groups (Mullowney et al., 2014; Tsuzuki et al., 1989).

The high potency and high hydrophobicity of diazaquinomycins H and J are not unique to this compound class. In fact, many marine-derived metabolites exhibit higher potency compared to their terrestrial counterparts, believed to be due to their

162

higher dilution in the marine environment (Newman et al., 2009; Schneider, 2021). Marine microbes must generate compounds with higher potency to be effective in such an environment. On the other hand, the marine environment is also responsible for the reduced hydrophilicity of compounds. Oxygen and nitrogen atom content are the main predictors of hydrophilicity in natural products (Kong et al., 2010). The hypoxic nature of the marine environment in comparison to terrestrial habitats means less oxygen is available for organisms to incorporate into metabolites, resulting in more frequent reports of hydrophobic properties from marine natural products (Kong et al., 2010; Schneider, 2021).

Given the very high percent similarity of the putative diazaquinomycin cluster (94%) to diazaquinomycin H or J, LC-MS data was analyzed in an attempt to further characterize the diazaquinomycin analog produced by *Micromonospora* sp. strain R45601. To do so, analysis focused on identifying peaks that retain the core diazaquinomycin structure containing four oxygen atoms, two nitrogen atoms, and a combination of carbon and hydrogen atoms with a minimum of 12 carbon atoms. Based on the structures of known analogs A-H and J, *m/z* values of ions (M^+, $[M+H]^+$, $[M+Na]^+$) were calculated that could theoretically be detected in positive ion mode. Additionally, *m/z* values of the above three ions were calculated for theoretical but unknown diazaquinomycin-like structures by incremental addition or subtraction of C_nH_{2n} groups. Of the four peaks identified that may pertain to diazaquinomycin ions, two correspond to chemical formulas of known diazaquinomycin structures ($C_{21}H_{24}N_2O_4$ - diazaquinomycin F and G, $C_{23}H_{28}N_2O_4$ – diazaquinomycin E and J). However, this does not necessarily mean that these ions correspond to compounds

with identical structures of known diazaquinomycins. It is possible that the ions

maintain the same chemical formula but a different distribution of side chains. The

fact that already multiple known diazaquinomycin analogs (E and J, F and G) have

the same chemical formulas further supports this possibility.

The smallest diazaquinomycin analog (diazaquinomycin A) contains 20

carbon atoms in its structure, while the largest (diazaquinomycin D) contains 24

carbon atoms (Fig. 4.6). The two potentially novel diazaquinomycin compounds

identified fall outside of this range. It should be noted that the same study that

characterized diazaquinomycins E-G (and reisolated analog A) observed an additional

10 diazaquinomycin analogs by UV spectra and MS/MS fragmentation patterns, with

positive ion values ranging from m/z 355.1 $[M+H]^+$ to 425.2 $[M+H]^+$ (Mullowney et

al., 2014). The larger m/z detected in the present study (489.2723) possibly pertains to

the chemical formula $C_{28}H_{38}N_2O_4$. If indeed this is confirmed as a novel

diazaquinomycin, this would be the largest analog characterized to date.

Unfortunately, it is highly unlikely that this larger compound would have improved

solubility compared to the known analogs, as solubility is intrinsically related to

particle size (Sareen et al., 2012). Side chains composed of smaller functional groups

would increase the surface area of the compound and promote solubility (Sareen et

al., 2012). In this regard, the fourth m/z detected at 321.0846, pertaining to the

chemical formula $C_{16}H_{14}N_2O_4$, may be the most promising candidate. With only four

carbon atoms to incorporate into side chains, the largest side chain that the theoretical

diazaquinomycin structure could maintain is a butyl group. However, based on the

presence of at least one methyl group in all known diazaquinomycin analogs, it is

164

more likely that this analog would contain at least one methyl group and some

combination of methyl and ethyl groups or one additional propyl/isopropyl group. If

truly a novel diazaquinomycin, this would be the smallest analog characterized to

date. Possible structures with the chemical formula $C_{16}H_{14}N_2O_4$ are presented in

Figure 4.9, with the assumption that at least one side chain is a methyl group.
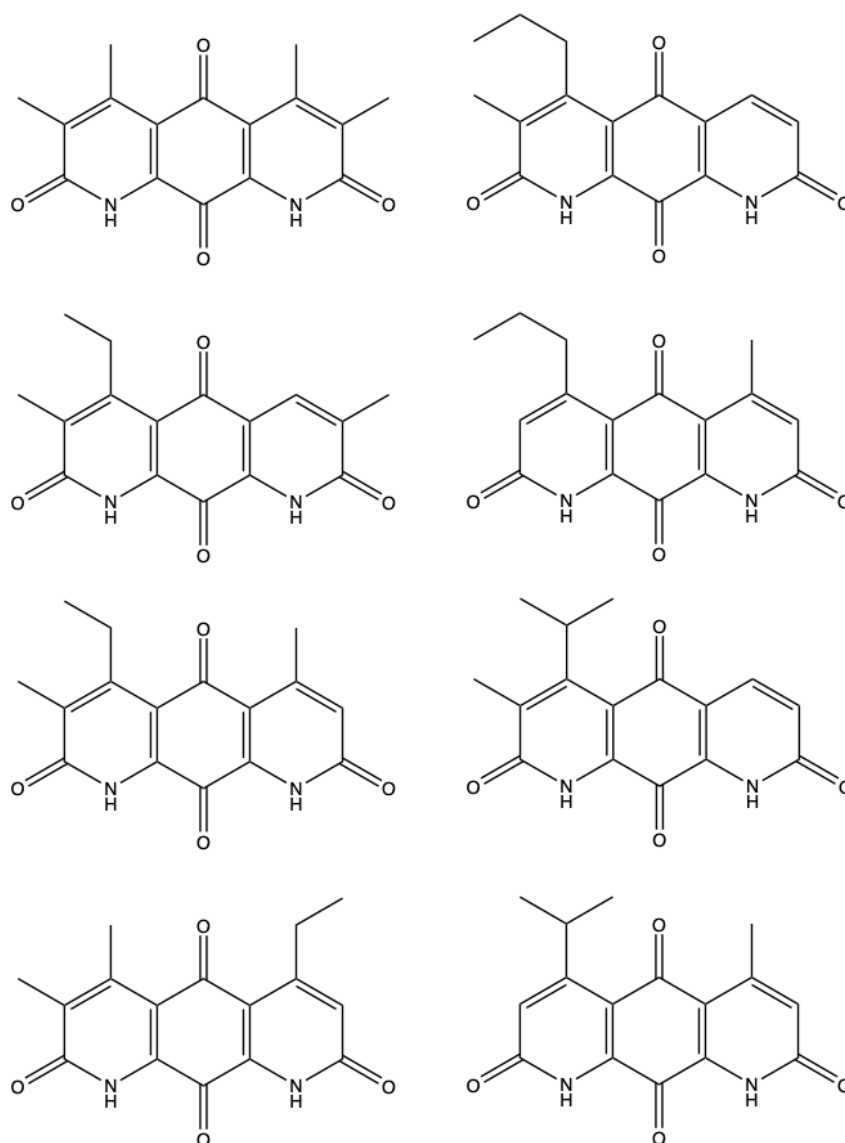
Figure 4.9. Possible diazaquinomycin-like structures in the *Micromonospora* sp. strain R45601 extract with the chemical formula C$_{16}$H$_{14}$N$_{2}$O$_{4}$. Enantiomers are excluded for simplicity.

In the study performed by Tsuzuki and colleagues to improve solubility of

diazaquinomycin A, the highest solubility was achieved by substituting short-chain

ester and ether derivatives for C-3 and C-6 methyl groups (Tsuzuki et al., 1989) (Fig.

4.10). Provided that this novel diazaquinomycin structure is synthesized by

Micromonospora sp. strain R45601, similar experiments should be performed to

substitute the same derivates at C-3 and C-6 to understand how the shorter C-4 and C-

5 alkyl groups in this structure ($C_{16}H_{14}N_2O_4$) affect solubility. Nevertheless, the next

steps moving forward must first focus on structure elucidation of these possibly novel

diazaquinomycin analogs through a combination of LC-UV-HRMS to confirm that

they are in fact present.
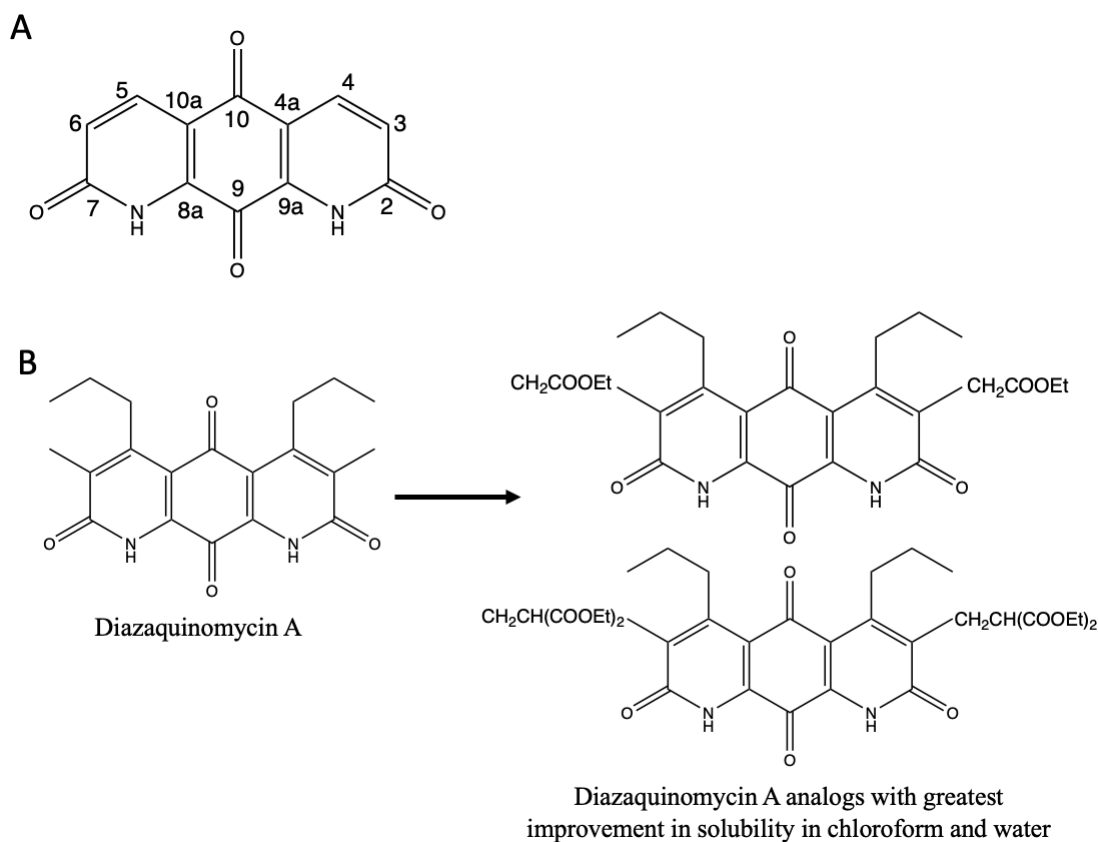
Figure 4.10. Naming convention of the diazaquinomycin core structure (A) and synthesized diazaquinomycin A analogs with improved solubility as described by Tsuzuki et al. (1989) (B). Naming convention adapted from scheme as detailed by Mullowney et al. (2014).

As discussed in Chapters 2 and 3, effective genome mining depends on the quality of the sequenced genome. To overcome the hurdles of assembling sequences with high GC bias and repetitive regions characteristic of actinomycete genomes, long-read sequencing is necessary. Half of all of the putative BGCs identified by antiSMASH (13/25) in the genome of *Micromonospora* sp. strain R45601 are located on contig edges. Long-read sequencing would enable complete resolution of these clusters and possibly elucidate other BGCs encoded by *Micromonospora* sp. strain R45601. The presence of PKS and NRPS-associated genes spanning a large portion of the putative cluster may be resolved with long-read sequencing. If these genes are legitimately associated with the diazaquinomycin cluster, this would suggest a hybrid BGC synthesizing a very large compound. It is more likely that these genes actually comprise a separate cluster and are not associated with the diazaquinomycin BGC. Perhaps these clusters were merged by the antiSMASH mining algorithm due to their close proximity. Mining with antiSMASH should be repeated for a *Micromonospora* sp. strain R45601 genome sequenced by PacBio to confirm this hypothesis.

The possibility that this strain produces multiple compounds with anti-TB activity should not be discounted. For the purposes of this study, short-read sequencing was sufficient to detect a BGC likely associated with a novel analog of a compound with known anti-TB activity. If the putative diazaquinomycin cluster was located on a contig edge, there would be the concern that the high percent similarity, as opposed to 100% identity, is due to incomplete resolution of the BGC. However, the putative cluster starts 90 kb into contig 1 and terminates almost 30 kb before the end of the same contig, decreasing the likelihood that this is the case. Nevertheless,

this study once again highlights the limitations of genome mining and bioinformatics in the drug discovery process. Although discerning the biosynthetic pathway can significantly help guide chemists in their analysis, analytical chemistry tools are essential to discovering novel compounds, as they enable isolation of the compound of interest and structure elucidation. In this present study, genomics has already informed chemistry and provided a core structure of interest. Additional chemistry will ultimately inform efforts to elucidate the full biosynthetic pathway of this putatively novel diazaquinomycin analog.

Chapter 5: Conclusions and future directions

5.1 Conclusions

It is estimated that given a collection of bacterial strains under investigation for antibacterial activity, preliminary screening will identify approximately 10% of these as displaying activity (Riyanti et al., 2020). In this study investigating a novel collection of 101 actinomycetes, 58 were recovered through culturing efforts, of which 13 strains were detected to inhibit *Mycobacterium* in preliminary screening. This preliminary hit rate of 22% strongly supports continued research into novel marine actinomycetes for their biosynthetic potential. Even if the highly similar *Micrococcus* and *Micromonospora* isolates investigated in Chapter 2 are indeed identical, the preliminary hit rate (15%) still falls well within the expected range. Nevertheless, this initial hit rate has no correlation with the success rate of developing a new drug. In fact, for every 5,000 compounds screened, only one will be developed into a marketable drug (Kraljevic et al., 2004). Furthermore, the journey to drug development is long and costly, with most clinical trials spanning over nine years and costing an average of $1 billion, with many projects costing much more (Brown et al., 2021; DiMasi et al., 2016; Wouters et al., 2020). It should be emphasized that this does not include the costs entailed by preclinical trials/ animal studies, which can account for one third of the total developmental costs, nor the highly variable costs to originally identify and isolate the compound in the laboratory (Paul et al., 2010). Therefore, it is essential to prioritize strains that appear most promising early on. These high average costs also mean that successful development of a new drug almost always depends on major investments by a large pharmaceutical company (Woon & Fisher, 2016). However, many pharmaceutical companies are no longer

exploring natural products for drug discovery. This leaves an important niche for academic groups and small biotechnology companies in the initial, very high-risk stages for drug discovery (Frearson & Wyatt, 2010; Mossialos et al., 2010; Sekurova et al., 2019; Woon & Fisher, 2016). The increased flexibility in research available in an academic setting, as well as the highly collaborative environment between experts across various disciplines, fosters innovation and enables more high-risk investigations (Frearson & Wyatt, 2010). For academic groups at least, even if a new drug is not found, many important scientific insights can be made during the initial steps of examining natural sources for bioactivity. These results are often published and made available for the benefit of other researchers.

In this research, to quickly prioritize actinomycete strains for further investigation, a genomics-enabled approach was first performed to evaluate the biosynthetic potential of each strain. Through genome mining, it was possible to rule out strains that likely cause growth inhibition by synthesizing known anti-tuberculosis (TB) compounds. Of course, there is always the possibility that strains are erroneously rejected for further analysis. A strain may produce multiple compounds with anti-TB activity, some of which are known and others that are novel. Additionally, a strain may produce a compound for which the biosynthetic gene cluster (BGC) is cryptic and evades detection by standard genome mining methods. The comparative assembly analysis described in Chapter 2 was performed to minimize this possibility and to maximize resolution of all BGCs detected. The most complete assembly of each genome (the best possible with short-read sequencing) was necessary for effective BGC analysis with the various genome mining tools

employed. Nevertheless, as evidenced by the lack of relevant BGCs detected in any of the *Micrococcus* and *Brevibacterium* genomes analyzed in this study, several of the novel isolates still retain cryptic or highly evolved clusters. Another possibility is that a strain was erroneously discarded from genomic analysis as a result of inactivity. Prioritizing strains based on the initial screening step presumes that there are no silent BGCs in the genome. As discussed in Chapter 1, it is very uncommon for bacteria to express all of the BGCs in their repertoire at once, especially in a laboratory setting. Because this study detected several strains consistently producing extracts with inhibitory activity, these isolates were assigned priority, and revisiting the aforementioned scenarios was unnecessary.

In this study, two strains stand out for entirely different reasons. The anti-TB activity of *Micrococcus* sp. strain R8502A1 is very intriguing due its very small genome size that is seemingly void of relevant BGCs, which contrasts with the potency of inhibition by the corresponding extract. The potential for drug discovery with this strain is at once both appealing and daunting, as it is entirely *de novo*. Without an idea of what type of compound this strain may be producing, extensive research is still needed, including data from additional experiments of bioassay-guided fractionation, liquid chromatography-mass spectrometry (LC-MS), and nuclear magnetic resonance (NMR) analysis. Continuous separation of the bioactive fraction and confirmation of activity retention by disk-diffusion assay will ultimately isolate the compound responsible for inhibiting growth of *Mycobacterium* spp. Characterizing this novel antibiotic from strain R48502A1 would be a significant addition to the currently limited knowledge of *Micrococcus*-derived compounds and

would support continued investigation into this genus for its biosynthetic potential, which has been historically overlooked due to the lack of detected BGCs in the relatively small genomes. A comprehensive analysis of all identifiable BGC-associated domains shows that this strain harbors at least one full set of core domains required for a complete module. These include a mix of PKS- and NRPS-associated domains, suggesting a hybrid BGC. Moreover, because these domains are scattered throughout the genome, it is likely that if a PKS or NRPS is indeed present, it maintains a nonmodular/nonlinear arrangement. We cannot rule out the probability that a natural product of an entirely novel class, thus far unidentified, is responsible for the anti-TB activity. These results highlight the limits of current genome mining approaches and encourage efforts to better understand the biosynthetic mechanisms of more elusive secondary metabolites, as occurred with terpenes not too long ago (Soldatou et al., 2021; Yamada et al., 2015).

The anti-TB activity observed by *Micromonospora* sp. strain R45601 is less of a mystery, but nonetheless promising. The genome mining results combined with preliminary LC-MS investigation convincingly point to a diazaquinomycin being responsible for the observed growth inhibition. As discussed in depth in Chapter 4, the goal of isolating a novel diazaquinomycin analog is to find a structure with improved solubility for drug development. The potency and selectivity of this compound class against *Mycobacterium tuberculosis* (*M. tb*) has already been confirmed, but solubility issues prohibit formulation development for preclinical studies. Of the four peaks identified to possibly pertain to diazaquinomycin ions, two are likely novel, as the corresponding chemical formulas of the *m/z* values detected do

not match that of any known diazaquinomycin analog (A-H or J). Both calculated

chemical formulas fall outside the size range of known analogs, on opposite ends of

the spectrum. If they are indeed diazaquinomycins, the peak pertaining to the

chemical formula $C_{16}H_{14}N_2O_4$ would be the smallest analog known to date, and the

peak pertaining to the chemical formula $C_{28}H_{38}N_2O_4$ would be the largest known to

date.

The rising threat of antibiotic resistance demands therapeutics with novel

mechanisms of action to treat infectious diseases like *M. tb.* Resistance has already

been detected against almost every antibiotic currently on the market, and there is

great concern that the rate of drug discovery by the classical screening approach

cannot keep up with the pace at which pathogens develop resistance (Lewis, 2013;

Tomm et al., 2019; Ventola, 2015). Alarmingly, resistance to some antibiotics was

detected in the same year that it was introduced to the market, such as in the case of

levofloxacin (Ventola, 2015). Alternative approaches to drug discovery, such as

combinatorial chemistry and rational drug design, gained prominence in the 1990s in

an attempt to focus compound screening efforts, as sampling studies increasingly

resulted in rediscoveries (Lewis, 2013; Wright, 2017). Combinatorial chemistry has

enabled scientists to enhance the activity of countless natural products, but

unfortunately, these approaches are largely unsuccessful in producing new drugs *de*

novo (Lewis, 2013; Paterson & Anderson, 2005; Tomm et al., 2019). In fact, only

three combinatorial compounds have been approved as drugs since efforts began in

the late 1990s, all of which are implicated in various cancer treatments (Newman &

Cragg, 2020). Natural products, especially those marine-derived, offer a great

advantage over synthetically-derived compounds, as they often possess highly

complex and specialized structures with unique chemical moieties (Blunt et al., 2018;

Henkel et al., 1999; Montaser & Luesch, 2011; Tiwari & Gupta, 2012; Tomm et al.,

2019). Compared to synthetically-designed products, natural compounds retain higher

specificity and potency (Paterson & Anderson, 2005). This is a consequence of the

many years over which they have been able to evolve highly optimized and targeted

structures (O'Shea & Moser, 2008; Paterson & Anderson, 2005). High rediscovery

rates from terrestrial sampling sources indicate that unique and underexplored

environments should be pursued to maximize the chances of discovering novel,

highly evolved natural products. Numerous studies have identified unique metabolites

from rare actinomycetes, and some have found geographic location to have a

dramatic effect on the specialized metabolism of isolates of the same genus sampled

from distinct regions (Parra & Duncan, 2019). With at least 71% of known marine

scaffolds being used exclusively by marine organisms, and 53% of marine scaffolds

detected from only one source thus far (Kong et al., 2010), the probability of

discovering novel chemistry from the marine environment is all but guaranteed. It is

for these reasons that marine actinomycetes, and particularly those of rare genera,

were investigated in this study for their ability to produce novel compounds with anti-

TB activity. Several recent studies have taken a similar approach to drug discovery,

albeit with strains of the more commonly isolated *Streptomyces*, incorporating

antimicrobial assays, genome mining, and chemical analysis (Guerrero-Garzón et al.,

2020; Shi et al., 2022). With continued sampling of underexplored marine

environments for novel actinomycete strains and effective prioritizing strategies, we are sure to discover novel therapeutics.

5.2 Future directions

As previously discussed, actinomycetes typically do not produce all the secondary metabolites that they are capable of synthesizing in the laboratory, and extensive culturing modifications ("one strain many compound" (OSMAC) approach) are often necessary to induce expression of these silent BGCs. It is very possible that some of the strains investigated in this study were erroneously labeled as "inactive" in preliminary screenings but do in fact produce compounds that inhibit *Mycobacterium* that were either not synthesized or produced in concentrations too low to permit detection. This will always be a concern in preliminary screenings, so priority must be assigned to strains early due to resource limitations and time constraints. Thus, only strains displaying inhibitory activity consistently in preliminary screens were selected for further investigation. Once strains of interest are identified, we can then return to the OSMAC approach to alter culturing conditions in an effort to maximize extract yield.

Substantial crude extract is required to purify as much of the compound of interest as possible, which must then be confirmed to retain anti-TB activity by disk-diffusion assay. Isolation of the purified bioactive compound will also enable calculation of critical activity characteristics, such as the minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) of the natural

product. Comparison of the MIC and MBC of the novel compound to known antibiotics will inform whether this drug merits further development.

An interesting discovery made in Chapter 2 was the very high similarity/identity of several strains to one another. Although it is possible that these strains were reisolates of a single strain from different sponges or different samples taken from the same sponge during the original isolation and characterization studies and are indeed identical, confirming this will require performing a DNA-DNA hybridization study. It should be noted that *Micromonospora* sp. strain R45601 shares 100% identity of partial 16S rRNA gene sequence with the other four *Micromonospora* strains discussed in Chapter 2, although it was isolated from a separate specimen of *Xestospongia muta*. All five *Micromonospora* strains were detected to encode a BGC with 94% similarity to diazaquinomycin H/J, but further analysis was carried out only for the cluster encoded by strain R45601. A recent study investigating culturable bacteria from multiple sponge samples of the same species (collected from the same region) finds an inconsistent correlation between secondary metabolite production below the level of species of bacterial isolate (Clark et al., 2022). The results of that study support analyzing the biosynthetic potential of these strains individually. It is possible that these strains have evolved slightly different BGCs that encode distinct diazaquinomycin analogs. A similar case can be made for the three *Micrococcus* strains studied here. *Micrococcus* spp. strain XM4230A and strain XM4230B, which were isolated from the same sponge sample, have identical partial 16S rRNA gene sequences, and only one base change distinguishes both strains from strain R8502A1, which was isolated from a separate *X. muta* specimen.

Genome mining did not identify any relevant BGCs in any of the *Micrococcus* strains investigated, suggesting that they all encode cryptic clusters. In addition to isolating and characterizing the bioactive compound from *Micrococcus* sp. strain R8502A1, further work should focus on investigating the activity of strains XM4230A and XM4230B. The fact that these strains were not all isolated from the same sponge sample increases the likelihood that they encode distinct metabolites.

In the future with more resources available, every novel actinomycete strain in the Hill Laboratory collection of several hundred strains could be examined. Beginning with the OSMAC approach, it would be important to culture every strain in a variety of media to maximize the probability of extracting all of the bioactive compounds that the strain is capable of synthesizing. Moreover, continuous culturing of the strains in large volumes would provide sufficient sample to conduct more extensive screening assays. The objective of this study was to screen extracts for antibacterial activity, specifically against *M. tb.* However, it is highly likely that these strains produce secondary metabolites with a diverse array of bioactivities. With unlimited sample, extracts could be tested for multiple activities at once, including against various bacteria, fungi, viruses and cancer cell lines.

Even with extensive modification of culturing parameters, it is still possible that cultured strains maintain silent BGCs. Genomic analysis was performed to circumvent this obstacle and efficiently obtain an estimate of each strain's biosynthetic capacity. Previous studies comparing short read assembly algorithms conclude that repeat regions are consistently the least well-reconstructed annotation feature, regardless of assembly method employed (Earl et al., 2011). Future genomic

studies could employ additional short read bacterial genome assembly programs that

have been successful in previous analyses (Magoc et al., 2013) although not tested

here, such as MaSuRCA (Zimin et al., 2013) and ABySS (Simpson et al., 2009).

However, any short read assembly algorithm will suffer from the aforementioned

deficiency. PacBio sequencing was only performed on one strain in this study due to

cost restraints, but to improve the efficacy of genome mining, long-read sequencing

data must be used to assemble each genome *de novo*. Assembly of a singular contig

ensures that all BGCs are fully resolved and that none are located on contig edges (an

issue that often obstructs proper identification). In the future, additional genome

mining tools could be utilized to provide a more robust assessment of BGC capacity,

including PRISM (Skinnider et al., 2017), EvoMining (Cruz-Morales et al., 2016),

and NRPSpredictor2 (Röttig et al., 2011). Nevertheless, regardless of sequencing

method employed, genome mining is insufficient as a stand-alone tool for discovery.

In addition to completeness of genome assembly, detection of putative BGCs depends

on the strength of mining algorithms as well as the database of known compounds

available at the time of analysis. A cluster with significant domain modifications

could very well generate novel chemistry, but will evade detection by current mining

programs. Additionally, natural products that do not follow typical enzymatic

assembly-line synthesis are not recognized by these algorithms. Thus, chemical

analysis will always remain an essential requirement for drug discovery.

Extensive analysis is still required to determine the novel anti-TB compound

produced by *Micrococcus* sp. strain R8502A1. After performing LC-MS/MS on an

extract of *Micrococcus* sp. strain R8502A1, this data can be cross-referenced with

Global Natural Products Social molecular networking (GNPS) spectra for dereplication purposes (Wang et al., 2016). Additionally, the NPLinker algorithm can be used to link GNPS spectra with BGCs in the MIBiG database through a feature-based approach (Hjörleifsson Eldjárn et al., 2021). Rosetta, a recently developed and innovative scoring tool through NPLinker, offers extensive filtering capabilities that have successfully linked spectra of several compounds, including ectoine, with their BGCs (Soldatou et al., 2021). Ectoine was already ruled out by laboratory testing as a potential inhibitor of *M. tb,* but this tool could be applied to rapidly confirm presence or absence of ectoine in the active fractions "70" and "100" of *Micrococcus* sp. strain R8502A1 extract and further aid in dereplication efforts. The same tools can be applied to extracts of *Micromonospora* sp. strain R45601. LC-MS/MS analysis of a *Micromonospora* sp. strain R45601 extract has thus far identified four peaks possibly corresponding to diazaquinomycin analogs. MS/MS fragmentation patterns of known diazaquinomycin analogs E, F, G, and J could be compared to the MS/MS spectra obtained from this study to determine whether the two detected *m/z* values of 369.1808889 and 397.213354 correspond to known diazaquinomycin analogs or novel structures. Since solubility correlates negatively with particle size, future research should also focus on confirming the structure of the detected ion with an *m/z* value of 321.0846, as this compound would most likely possess improved solubility compared to all known analogs. Additional analysis with NMR and ultraviolet-high-resolution mass spectrometry (UV-HRMS) is required for structure elucidation.

As of 2019, there were at least 176 marine derived compounds with anti-TB activity (Hou et al., 2019). These natural products belong to a range of biosynthetic

classes including alkaloids, terpenoids, peptides, quinones, sterols, pyrones and lactones, as well as other compound types, and were isolated from a diverse array of organisms such as sponges, fungi and sediment-derived bacteria (Hou et al., 2019). As described above, BGC or spectra information corresponding to these compounds, if available, can be compared to the genome mining data and LC-MS/MS spectra obtained for *Micrococcus* sp. strain R8502A1 and *Micrococcus* sp. strain R45601 for dereplication. Furthermore, this information can be used to prioritize any other strains under investigation for anti-TB activity in the future.

Appendices

Appendix Chapter 2

Scripts Run for assembly algorithms as described in Chapter 2

For Trimmomatic:

```
nice nohup java –jar /usr/local/bin/Trimmomatic–
0.30/trimmomatic.jar PE –threads 30 –phred33
~/genomeX_L001_R1_001.fastq.gz ~/genomeX_L001_R2_001.fastq.gz
TRIM_genomeX_pair_R1.fastq TRIM_genomeX_unpair_R1.fastq
TRIM_genomeX_pair_R2.fastq TRIM_genomeX_unpair_R2.fastq
ILLUMINACLIP:TruSeq3–PE.fa:2:30:10:2:keepBothReads LEADING:3
TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 &
```

For SPAdes:

```
nice –n 16 nohup ~/SPAdes–3.14.1–Linux/bin/spades.py ––pe1–1
/data1/dtizabi/genomeX_forward_reads.fastq ––pe1–2
/data1/dtizabi/genomeX_reverse_reads.fastq –o
spades_genomeX_output &
```

To check quality of assembly PRE-filtering:

```
quast.py
/data1/dtizabi/spades_genomeX_output/genomeX_contigs.fasta –o
quast_genomeX
```

To BLAST and remove all non-actinomycete contigs:

```
nohup blastn –query genomeX_contigs.fasta –db
/data1/blastdb/nt –evalue 1e–10 –outfmt '6 std stitle' –
max_target_seqs 1 –max_hsps 1 > genomeX_hsp_Blast &
```

Manually removed all non-actinomycete contigs and renamed file to "genomeX_actino_contigs.fasta"

To create database of contigs of other genomes sequenced during same Illumina MiSeq run and determine coverage cutoff and spillover:

```
perl –pi –e "s/^>/>genomeX_/g"
ForBLAST_genomeX_actino_contigs.fasta
```

-combined 7 genomes at a time to create a larger subject database for BLAST comparisons:

```
cat ForBLAST_genomeA_actino_contigs.fasta
ForBLAST_genomeB_actino_contigs.fasta
ForBLAST_genomeC_actino_contigs.fasta
ForBLAST_genomeD_actino_contigs.fasta
ForBLAST_genomeE_actino_contigs.fasta
ForBLAST_genomeF_actino_contigs.fasta
ForBLAST_genomeG_actino_contigs.fasta > BLASTvgenomeX.fasta

nohup blastn –query genomeX_actino_contigs.fasta –subject
BLASTvgenomeX.fasta –evalue 1e–10 –outfmt '6 std qlen slen' –
max_target_seqs 1 –max_hsps 1 > comparegenomeX_to_7others &
```

To reformat and filter:

```
perl –pi –e 's/_/ /g' genomeX_actino_contigs.fasta

perl –pi –e 's/([ACTG])\n/\1/g' genomeX_actino_contigs.fasta

perl –pi –e 's/([ACTG])>/\1\n>/g' genomeX_actino_contigs.fasta

perl –pi –e 's/([0–9])\n/\1 /g' genomeX_actino_contigs.fasta

awk '$6>VALUE {print}' genomeX_actino_contigs.fasta >
genomeX_final_contigs.fasta
```

where VALUE is the cutoff coverage value determined individually for each genome

```
perl –pi –e 's/([0–9]) ([ACTG])/\1\n\2/g'
genomeX_final_contigs.fasta

perl –pi –e 's/ /_/g' genomeX_final_contigs.fasta
```

Finally, manually deleted any nodes that were not removed based on coverage and that did not belong in final assembly

For A5:
```
nohup a5_miseq_linux_20160825/bin/a5_pipeline.pl
genomeX_forward_reads.fastq genomeX_reverse_reads.fastq
A5_genomeX &

nohup blastn –query A5_genomeX.contigs.fasta –db
/data1/blastdb/nt –evalue 1e–10 –outfmt '6 std stitle' –
max_target_seqs 1 –max_hsps 1 > A5_genomeX_Blast &
```

For Shovill:

184

```
nohup shovill --ram 48 --trim --outdir Shovill_genomeX --R1
genomeX_L001_R1_001.fastq --R2 genomeX_L001_R2_001.fastq &
```

***changed ram setting (default 16) to 48

```
nohup blastn -query ~/Shovill_genomeX/contigs.fa -db
/data1/blastdb/nt -evalue 1e-10 -outfmt '6 std stitle' -
max_target_seqs 1 -max_hsps 1 > BLAST_Shovill_genomeX &
```

Table A.2.1. NP.Searcher results for actinomycete genomes based on assembly method.

Strain	SPAdes	A5-miseq	Shovill
Micrococcus sp. strain XM4230A	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes
Micrococcus sp. strain XM4230B	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes
Micromonospora sp. strain XM-20-01	4 modular PKSs 4 mixed modular NRPS/PKSs 3 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	2 modular PKSs 6 mixed modular NRPS/PKSs 1 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	5 modular PKSs 4 mixed modular NRPS/PKSs 2 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes
Micromonospora sp. strain R42003	2 modular PKSs 4 mixed modular NRPS/PKSs 1 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	2 modular PKSs 5 mixed modular NRPS/PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	2 modular PKSs 5 mixed modular NRPS/PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes
Micromonospora sp. strain R42004	2 modular PKSs 5 mixed modular NRPS/PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	3 modular PKSs 5 mixed modular NRPS/PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	2 modular PKSs 6 mixed modular NRPS/PKSs 1 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes
Micromonospora sp. strain R42106	5 mixed modular NRPS/PKSs 2 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	5 mixed modular NRPS/PKSs 3 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes	5 mixed modular NRPS/PKSs 1 trans AT PKSs 1 mevalonate terpenoid mva genes 1 non-mevalonate terpenoid mep genes
Brevibacterium sp. strain XM4083	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes
Brevibacterium sp. strain R8603A2	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes
Streptomyces sp. strain XM4011	4 modular NRPSs 1 mixed modular NRPS/PKSs 2 non-mevalonate terpenoid mep genes	3 modular NRPSs 1 mixed modular NRPS/PKSs 2 non-mevalonate terpenoid mep genes	3 modular NRPSs 1 modular PKSs 1 trans AT PKSs 2 non-mevalonate terpenoid mep genes
Streptomyces sp. strain XM4193	5 modular NRPSs 2 modular PKSs 1 non-mevalonate terpenoid mep genes	5 modular NRPSs 2 modular PKSs 2 mixed modular NRPS/PKSs	5 modular NRPSs 1 modular PKSs 1 non-mevalonate terpenoid mep genes

		1 non-mevalonate terpenoid mep genes	
Streptomyces **sp. strain XM83C**	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes	1 non-mevalonate terpenoid mep genes

Key: PKS = polyketide synthetase, NRPS = nonribosomal peptide synthetase, AT = acyltransferase domain, mva = mevalonate pathway, mep = methylerythritol 4-phosphate pathway.

Bibliography

Abdelmohsen, U. R., Bayer, K., & Hentschel, U. (2014). Diversity, abundance and natural products of marine sponge-associated actinomycetes. *Natural Product Reports*, *31*(3), 381-399. https://doi.org/10.1039/c3np70111e

Abdelmohsen, U. R., Szesny, M., Othman, E. M., Schirmeister, T., Grond, S., Stopper, H., & Hentschel, U. (2012). Antioxidant and anti-protease activities of diazepinomicin from the sponge-associated *Micromonospora* strain RV115. *Marine Drugs*, *10*(10), 2208-2221. https://doi.org/10.3390/MD10102208

Abdjul, D. B., Yagi, A., Yamazaki, H., Kirikoshi, R., Takahashi, O., Namikoshi, M., & Uchida, R. (2018). Anti-mycobacterial haliclonadiamine alkaloids from the Okinawan marine sponge *Haliclona* sp. collected at Iriomote Island. *Phytochemistry Letters*, *26*, 130–133. https://doi.org/10.1016/j.phytol.2018.05.028

Abu-Salah, K. M. (1996). Amphotericin B: An update. *British Journal of Biomedical Science*, *53*(2),122-133. https://pubmed.ncbi.nlm.nih.gov/8757689/

Acuña-Amador, L., Primot, A., Cadieu, E., Roulet, A., & Barloy-Hubler, F. (2018). Genomic repeats, misassembly and reannotation: A case study with long-read resequencing of *Porphyromonas gingivalis* reference strains. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-017-4429-4

Adamidis, T., Riggle, P., & Champness, W. (1990). Mutations in a new *Streptomyces coelicolor* locus which globally block antibiotics biosynthesis but not sporulation. *Journal of Bacteriology*, *172*(6), 2962–2969. https://doi.org/10.1128/jb.172.6.2962-2969.1990

Akbar, A., Sitara, U., Ali, I., Muhammad, N., & Khan, S. A. (2014). Isolation and characterization of biotechnologically potent *Micrococcus luteus* strain from environment. *Pakistan Journal of Zoology*, *46*(4), 967–973.

Akram, S., & Aboobacker, S. (2021). *Mycobacterium Marinum*. StatPearls Publishing. https://doi.org/10.1016/S1294-5501(05)80179-7

Albertson, D., Natsios, G. A., & Gleckman, R. (1978). Septic shock with *Micrococcus luteus*. *Archives of Internal Medicine*, *138*(3), 487–488. https://doi.org/10.1001/ARCHINTE.1978.03630270093032

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/NAR/25.17.3389

Amann, R. I., Ludwig, W., & Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), 143–169. https://doi.org/10.1128/MR.59.1.143-169.1995

Anteneh, Y. S., Yang, Q., Brown, M. H., & Franco, C. M. M. (2021). Antimicrobial activities of marine sponge-associated bacteria. *Microorganisms*, *9*(1), 1–19. https://doi.org/10.3390/MICROORGANISMS9010171

Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., & Suzuki, M. T. (2017). Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Scientific Reports, 7*(1), 1–8. https://doi.org/10.1038/s41598-017-11363-1

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., … van der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Natural Product Reports*, *30*(1), 108. https://doi.org/10.1039/C2NP20085F

Awakawa, T., Fujita, N., Hayakawa, M., Ohnishi, Y., & Horinouchi, S. (2011). Characterization of the biosynthesis gene cluster for alkyl-O-dihydrogeranyl-methoxyhydroquinones in *Actinoplanes missouriensis*. *Chembiochem : A European Journal of Chemical Biology, 12*(3), 439–448. https://doi.org/10.1002/CBIC.201000628

Bachmann, B. O., & Ravel, J. (2009). Chapter 8 methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods in Enzymology*, *458*(A), 181-217. https://doi.org/10.1016/S0076-6879(09)04808-3

Baltz, R. H. (2005). Antibiotic discovery from actinomycetes: Will a renaissance follow the decline and fall? *SIM News*, *55*, 186-196.

Baltz, R. H. (2008). Renaissance in antibacterial discovery from actinomycetes. *Current Opinion in Pharmacology, 8*(5), 557-563. https://doi.org/10.1016/j.coph.2008.04.008

Baltz, R. H. (2016). Gifted microbes for genome mining and natural product discovery. *Journal of Industrial Microbiology and Biotechnology*, *44*(4–5), 573–588. https://doi.org/10.1007/s10295-016-1815-x

Baltz, R. H. (2019). Natural product drug discovery in the genomic era: Realities, conjectures, misconceptions, and opportunities. *Journal of Industrial*

Microbiology and Biotechnology, *46*(3–4), 281–299.
https://doi.org/10.1007/s10295-018-2115-4

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455. https://doi.org/10.1089/CMB.2012.0021

Banskota, A. H., McAlpine, J. B., Sørensen, D., Ibrahim, A., Aouidate, M., Piraee, M., Alarco, A. M., Farnet, C. M., & Zazopoulos, E. (2006). Genomic analyses lead to novel secondary metabolites. Part 3. ECO-0501, a novel antibacterial of a new class. *The Journal of Antibiotics*, *59*(9), 533–542. https://doi.org/10.1038/JA.2006.74

Baral, B., Akhgari, A., & Metsä-Ketelä, M. (2018). Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synthetic and Systems Biotechnology*, *3*(3), 163-178. https://doi.org/10.1016/j.synbio.2018.09.001

Barberis, I., Bragazzi, N. L., Galluzzo, L., & Martini, M. (2017). The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus. *Journal* of *Preventive Medicine* and *Hygiene, 58*(1), E9-E12.

Bellassi, P., Cappa, F., Fontana, A., & Morelli, L. (2020). Phenotypic and genotypic investigation of two representative strains of *Microbacterium* species isolated from micro-filtered milk: Growth capacity and spoilage-potential assessment. *Frontiers in Microbiology*, *11*, 554178. https://doi.org/10.3389/FMICB.2020.554178/FULL

Belmokhtar, C. A., Hillion, J., & Ségal-Bendirdjian, E. (2001). Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms. *Oncogene*, *20*(26), 3354–3362. https://doi.org/10.1038/SJ.ONC.1204436

Belshaw, P. J., Walsh, C. T., & Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, *284*(5413), 486–489. https://doi.org/10.1126/science.284.5413.486

Bennett, J. W., Häggblom, M. M., & Sullivan R. F. (2021). The microbiology of sheer fun: A memorial to Douglas E. Eveleigh (1933-2019). *SIMB News*, *71*(2), 44-59.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A.,

Goble, A., Hidalgo, J., … Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, *417*(6885), 141–147. https://doi.org/10.1038/417141a

Bérdy, J. (2005). Bioactive microbial metabolites. *The Journal of Antibiotics, 58*(1), 1–26. https://doi.org/10.1038/ja.2005.1

Bérdy, J. (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. *The Journal of Antibiotics*, *65*(8), 385–395. https://doi.org/10.1038/ja.2012.27

Bergmann, W., & Feeney, R. J. (1951). Contributions to the study of marine products. XXXII. The nucleosides of sponges. I. *Journal of Organic Chemistry*, *16*(6), 981–987. https://doi.org/10.1021/JO01146A023/ASSET/JO01146A023.FP.PNG_V03

Bewley, C. A., & Faulkner, D. J. (1998). Lithistid sponges: Star performers or hosts to the stars. *Angewandte Chemie International Editino England.* https://doi.org/10.1002/(SICI)1521-3773(19980904)37:16<2162::AID-ANIE2162>3.0.CO;2-2

Bewley, C. A., Holland, N. D., & Faulkner, D. J. (1996). Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. *Experientia*, *52*(7), 716–722. https://doi.org/10.1007/BF01925581

Bhowmik, T., & Marth, E. H. (1990). Rote of *Micrococcus* and *Pediococcus* species in cheese ripening: A review. *Journal of Dairy Science*, *73*(4), 859–866. https://doi.org/10.3168/JDS.S0022-0302(90)78740-1

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology, 8*(2), 208-215. https://doi.org/10.1016/j.mib.2005.02.016

Bisang, C., Long, P. F., Cortés, J., Westcott, J., Crosby, J., Matharu, A. L., Cox, R. J., Simpson, T. J., Staunton, J., & Leadlay, P. F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, *401*(6752), 502–505. https://doi.org/10.1038/46829

Biskupiak, J. E., Meyers, E., Gillum, A. M., Dean, L., Trejo, W. H., & Kirsch, D. R. (1988). Neoberninamycin, a new antibiotic produced by *Micrococcus luteus*. *The Journal of Antibiotics*, *41*(5), 684-687. doi: 10.7164/antibiotics.41.684

Blacklock, J. W. S. (1947). The epidemiology of tuberculosis. *British Medical Journal*, *1*(4507), 707-712.

Blackwell, G. A., Hunt, M., Malone, K. M., Lima, L., Horesh, G., Alako, B. T. F.,

Thomson, N. R., & Iqbal, Z. (2021). Exploring bacterial diversity via a curated and searchable snapshot of archived DNA sequences. *PLoS Biology*, *19*(11). https://doi.org/10.1371/journal.pbio.3001421

Blin, K., Kim, H. U., Medema, M. H., & Weber, T. (2019a). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Briefings in Bioinformatics*, *20*(4), 1103–1113. https://doi.org/10.1093/bib/bbx146

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019b). AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, *47*(W1), W81–W87. https://doi.org/10.1093/nar/gkz310

Bloudoff, K., & Schmeing, T. M. (2017). Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: Discovery, dissection and diversity. *himica et Biophysica Acta - Proteins and Proteomics*, *1865*(11), 1587–1604. https://doi.org/10.1016/j.bbapap.2017.05.010

Blunt, J. W., Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., & Prinsep, M. R. (2018). Marine natural products. *Natural Product Reports*, *35*(1), 8–53. https://doi.org/10.1039/C7NP00052A

Bode, H. B., Bethe, B., Höfs, R., & Zeeck, A. (2002). Big effects from small changes: Possible ways to explore nature's chemical diversity. *ChemBioChem*, *3*(7), 619-627. doi: 10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9. PMID: 12324995.

Bogen, E. (1948). Streptomycin treatment for tuberculosis. *Journal of the* National *Medical Association*, *40*(1), 32.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Borchert, E., Jackson, S. A., O'Gara, F., & Dobson, A. D. W. (2016). Diversity of natural product biosynthetic genes in the microbiome of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa*, and *Stelletta normani*. *Frontiers in Microbiology, 7*, 1027. https://doi.org/10.3389/FMICB.2016.01027

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M. F., Lió, P., Crescenzi, P., Fani, R., & Fondi, M. (2015). MeDuSa: A multi-draft based scaffolder. *Bioinformatics*, *31*(15), 2443–2451. https://doi.org/10.1093/BIOINFORMATICS/BTV171

Boumehira, A. Z., El-Enshasy, A., Hacène, H., Elsayed, A., Aziz, R., & Park, E. Y.

192

(2016). Recent progress on the development of antibiotics from the genus *Micromonospora*. *Biotechnology and Bioprocess Engineering*, *21*, 199–223. https://doi.org/10.1007/s12257-015-0574-2

Braesel, J., Lee, J. H., Arnould, B., Murphy, B. T., & Eustáquio, A. S. (2019). Diazaquinomycin biosynthetic gene clusters from marine and freshwater actinomycetes. *Journal of Natural Products*, *82*(4), 937–946. https://doi.org/10.1021/acs.jnatprod.8b01028

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports, 5*, 8365. https://doi.org/10.1038/SREP08365

Breukink, E., Wiedemann, I., Van Kraaij, C., Kuipers, O. P., Sahl, H. G., & de Kruijff, B. (1999). Use of the cell wail precursor lipid II by a pore-forming peptide antibiotic. *Science*, *286*(5448), 2361–2364. https://doi.org/10.1126/science.286.5448.2361

Brockmann, H., & Schmidt-Kastner, G. (1955). Valinomycin I, XXVII. Mitteil. über Antibiotica aus Actinomyceten. *Chemische Berichte*, *88*(1), 57–61. https://doi.org/10.1002/CBER.19550880111

Brown, L. 1941. The story of clinical pulmonary tuberculosis (1st ed.). Williams & Wilkins.

Brown, D. G., Wobst, H. J., Kapoor, A., Kenna, L. A., & Southall, N. (2021). Clinical development times for innovative drugs. *Nature Reviews. Drug Discovery*. https://doi.org/10.1038/D41573-021-00190-9

Browne, P. D., Nielsen, T. K., Kot, W., Aggerholm, A., Gilbert, M. T. P., Puetz, L., Rasmussen, M., Zervas, A., & Hansen, L. H. (2020). GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. *GigaScience*, *9*(2). https://doi.org/10.1093/gigascience/giaa008

Bultel-Poncé, V., Berge, J. P., Debitus, C., Nicolas, J. L., & Guyot, M. (1999). Metabolites from the sponge-associated bacterium *Pseudomonas* species. . *Marine Biotechnology*, *1*(4), 384–390. https://doi.org/10.1007/PL00011792

Button, D. K., Schut, F., Quang, P., Martin, R., & Robertson, B. R. (1993). Viability and isolation of marine bacteria by dilution culture: Theory, procedures, and initial results. *Applied and Environmental Microbiology*, *59*(3), 881. https://doi.org/10.1128/aem.59.3.881-891.1993

Byrne, K. (2011). *Tuberculosis and the Victorian literary imagination*. Cambridge University Press.

Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6, e4958. https://doi.org/10.7717/peerj.4958

Cahill, C., O'connell, F., Gogan, K. M., Cox, D. J., Basdeo, S. A., O'Sullivan, J., Gordon, S. V., Keane, J., & Phelan, J. J. (2021). The iron chelator desferrioxamine increases the efficacy of dedaquiline in primary human macrophages infected with BCG. *International Journal of Molecular Sciences*, *22*(6), 1–15. https://doi.org/10.3390/IJMS22062938

Campas, C. (2009). Diazepinomicin. *Drugs of the Future*, *34*(5), 349–351. doi: 10.1358/dof.2009.034.05.1370797.

Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., & Prinsep, M. R. (2020). Marine natural products. *Natural Product Reports*, *37*(2), 175–223. https://doi.org/10.1039/C9NP00069K

Cave, A. J. E. & Demonstrator, A. (1939). The evidence for the incidence of tuberculosis in ancient Egypt. *British Journal* of *Tuberculosis*, *33*, 142–152.

Chakraburtty, R., & Bibb, M. (1997). The ppGpp synthetase gene (relA) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation. *Journal of Bacteriology, 179*(18), 5854. https://doi.org/10.1128/JB.179.18.5854-5861.1997

Challis, G. L. (2008). Genome mining for novel natural product discovery. *Journal of Medicinal Chemistry, 51*(9), 2618-2628. https://doi.org/10.1021/jm700948z

Challis, G. L., & Hopwood, D. A. (2003). Colloquium paper: Chemical communication in a post-genomic world: Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proceedings of the National Academy of Sciences of the United States of America, 100* (Suppl 2), 14555-14561. https://doi.org/10.1073/PNAS.1934677100

Challis, G. L., & Naismith, J. H. (2004). Structural aspects of non-ribosomal peptide biosynthesis. *Current Opinion in Structural Biology, 4*(6), 748-756. https://doi.org/10.1016/j.sbi.2004.10.005

Challis, G. L., Ravel, J., & Townsend, C. A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry and Biology*, *7*(3), 211–224. https://doi.org/10.1016/S1074-5521(00)00091-0

Chase, A. B., Sweeney, D., Muskat, M. N., Guillén-Matus, D., & Jensen, P. R. (2020). Complex evolutionary dynamics govern the diversity and distribution of biosynthetic gene clusters and their encoded specialized metabolites. *BioRxiv*. https://doi.org/10.1101/2020.12.19.423547

Chater, K. F., Bucca, G., Dyson, P., Fowler, K., Gust, B., Herron, P., Hesketh, A., Hotchkiss, G., Kieser, T., Mersinias, V., & Smith, C. P. (2002). *Streptomyces coelicolor* A3(2): from genome sequence to function. *Methods in Microbiology*, *33*, 321–336. https://doi.org/10.1016/S0580-9517(02)33018-6

Chen, Y. C., Liu, T., Yu, C. H., Chiang, T. Y., & Hwang, C. C. (2013). Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS ONE*, *8*(4), 62856. https://doi.org/10.1371/JOURNAL.PONE.0062856

Chen, L., Zhao, W., Jiang, H. L., Zhou, J., Chen, X. M., Lian, Y. Y., Jiang, H., & Lin, F. (2018). Rakicidins G - I, cyclic depsipeptides from marine *Micromonospora* chalcea FIM 02-523. *Tetrahedron*, *74*(30), 4151–4154. https://doi.org/10.1016/j.tet.2018.06.039

Chiller, K., Selkin, B. A., & Murakawa, G. J. (2001). Skin microflora and bacterial infections of the skin. *The Journal of Investigative Dermatology. Symposium Proceedings*, *6*(3), 170–174. https://doi.org/10.1046/J.0022-202X.2001.00043.X

Choi, S. Y., Kim, S., Lyuck, S., Kim, S. B., & Mitchell, R. J. (2015). High-level production of violacein by the newly isolated *Duganella violaceinigra* str. NI28 and its impact on *Staphylococcus aureus*. *Scientific Reports*, *5*, 15598. https://doi.org/10.1038/SREP15598

Choi, S. Y., Lim, S., Yoon, K. hye, Lee, J. I., & Mitchell, R. J. (2021). Biotechnological activities and applications of bacterial pigments violacein and prodigiosin. *Journal of Biological Engineering*, *15*(1), 1–16. https://doi.org/10.1186/S13036-021-00262-9

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., & Fischbach, M. A. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, *158*(2), 412. https://doi.org/10.1016/J.CELL.2014.06.034

Cita, Y. P., Suhermanto, A., Radjasa, O. K., & Sudharmono, P. (2017). Antibacterial activity of marine bacteria isolated from sponge *Xestospongia testudinaria* from Sorong, Papua. *Asian Pacific Journal of Tropical Biomedicine*, *7*(5), 450–454. https://doi.org/10.1016/J.APJTB.2017.01.024

Clark, C. M., Hernandez, A., Mullowney, M. W., Fitz-Henley, J., Li, E.,

Romanowski, S. B., Pronzato, R., Manconi, R., Sanchez, L. M., & Murphy, B. T. (2022). Relationship between bacterial phylotype and specialized metabolite production in the culturable microbiome of two freshwater sponges. *ISME Communications*, *2*(1). https://doi.org/10.1038/s43705-022-00105-8

Claverías, F. P., Undabarrena, A. N., González, M., Seeger, M., & Cámara, B. P. (2015). Culturable diversity and antimicrobial activity of Actinobacteria from marine sediments in Valparaíso bay, Chile. *Frontiers in Microbiology, 6*, 737. https://doi.org/10.3389/FMICB.2015.00737

Cohn, F. (1872). Untersuchungen über Bakterien. *Beiträge zur Biologie der* Pflanzen*, 1*, 127-244.

Coil, D., Jospin, G., & Darling, A. E. (2015). A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, *31*(4), 587–589. https://doi.org/10.1093/bioinformatics/btu661

Cooney, J. J., Marks, H. W., & Smith, A. M. (1966). Isolation and identification of canthaxanthin from *Micrococcus roseus*. *Journal of Bacteriology*, *92*(2), 342–345. https://doi.org/10.1128/JB.92.2.342-345.1966

Cotter, P. D., Ross, R. P., & Hill, C. (2013). Bacteriocins - a viable alternative to antibiotics? *Nature Reviews Microbiology*, *11*(2), 95-105. https://doi.org/10.1038/nrmicro2937

Cross, T. (1981). Aquatic actinomycetes: A critical survey of the occurrence, growth and role of actinomycetes in aquatic habitats. *Journal of Applied Bacteriology*, *50*(3), 397–423. doi:10.1111/j.1365-2672.1981.tb04245.x

Cruz-Morales, P., Kopp, J. F., Martínez-Guerrero, C., Yáñez-Guerra, L. A., Selem-Mojica, N., Ramos-Aboites, H., Feldmann, J., & Barona-Gómez, F. (2016). Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. *Genome Biology and Evolution*, *8*(6), 1906–1916. https://doi.org/10.1093/gbe/evw125

Daniel, T. M. (1997). *Captain of death: The story of tuberculosis*. University of Rochester Press.

Daniel, T. M. (2000a). *Pioneers in medicine and their impact on tuberculosis*. University of Rochester Press.

Daniel, T. M. (2000b). The origins and precolonial epidemiology of tuberculosis in the Americas: Can we figure them out? *International Journal of Tuberculosis and Lung Disease*, *4*(5), 395-400.

Daniel, T. M. (2006). The history of tuberculosis. *Respiratory Medicine*, *100*(11),

1862–1870. https://doi.org/10.1016/J.RMED.2006.08.006

Daniel, V. S., & Daniel, T. M. (1999). Old Testament biblical references to tuberculosis. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 29*(6), 1557–1558. https://doi.org/10.1086/313562

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–1403. https://doi.org/10.1101/GR.2289704

Davis, C. P. (1996). Normal flora. *Definitions*. https://doi.org/10.32388/sfi72p

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., MacHi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., … Stevens, R. (2020). The PATRIC bioinformatics resource center: Expanding data and analysis capabilities. *Nucleic Acids Research*, *48*(D1), D606–D612. https://doi.org/10.1093/NAR/GKZ943

De Crecy-Lagard, V., Marliere, P., & Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: Identification of the functional domains catalysing peptide elongation and epimerisation. *Comptes Rendus de l'Academie Des Sciences - Serie III*, *318*(9), 927–936. https://pubmed.ncbi.nlm.nih.gov/8521076/

de Menezes, A. B., Lockhart, R. J., Cox, M. J., Allison, H. E., & McCarthy, A. J. (2008). Cellulose degradation by Micromonosporas recovered from freshwater lakes and classification of these actinomycetes by DNA gyrase B gene sequencing. *Applied and Environmental Microbiology*, *74*(22), 7080–7084. https://doi.org/10.1128/AEM.01092-08

de Menezes, A. B., McDonald, J. E., Allison, H. E., & McCarthy, A. J. (2012). Importance of *Micromonospora* spp. as colonizers of cellulose in freshwater lakes as demonstrated by quantitative reverse transcriptase PCR of 16S rRNA. *Applied and environmental microbiology*, *78*(9), 3495–3499. https://doi.org/10.1128/AEM.07314-11

de Oliveira, J. A. M., Williams, D. E., Bonnett, S., Johnson, J., Parish, T., & Andersen, R. J. (2020). Diterpenoids isolated from the Samoan marine sponge *Chelonaplysilla* sp. inhibit *Mycobacterium tuberculosis* growth. *Journal of Antibiotics*, *73*(8), 568–573. https://doi.org/10.1038/s41429-020-0315-4

de Souza, A., Aily, D., Sato, D., & Durán, N. (1999). Atividade da violaceína in vitro sobre o *Mycobacterium turbeculosis* H37RA. *Revista Do Instituto Adolfo Lutz*, *58*(1), 59–62. https://doi.org/10.53393/RIAL.1999.V58.36676

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, *47*, 20–33. https://doi.org/10.1016/J.JHEALECO.2016.01.012

Doehle, P. (1889). *Beobachtunger über einen Antagonisten des Milzbrandes*. Schmidt & Klaunig.

Donadio, S., Monciardini, P., & Sosio, M. (2007). Polyketide synthases and nonribosomal peptide synthetases: The emerging view from bacterial genomics. *Natural Product Reports, 24*(5), 1073-1079. https://doi.org/10.1039/b514050c

Du, L., Sánchez, C., Chen, M., Edwards, D. J., & Shen, B. (2000). The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chemistry and Biology*, *7*(8), 623–642. https://doi.org/10.1016/S1074-5521(00)00011-9

Du, Y. L., Shen, X. L., Yu, P., Bai, L. Q., & Li, Y. Q. (2011). Gamma-butyrolactone regulatory system of *Streptomyces chattanoogensis* links nutrient utilization, metabolism, and development. *Applied and Environmental Microbiology, 77*(23), 8415. https://doi.org/10.1128/AEM.05898-11

Dujardin-Beaumetz, E. *(*1934). *Comptes Rendus des Seances de la Societe de Biologie et de Ses Filiales, 117,* 1178.

Durán, N., & Menck, C. F. M. (2001). *Chromobacterium violaceum*: A review of pharmacological and industiral perspectives. *Critical Reviews in Microbiology, 27*(3), 201-222. https://doi.org/10.1080/20014091096747

Durrell, K., Prins, A., & Le Roes-Hill, M. (2017). Draft genome sequence of *Gordonia lacunae* BS2T. *Genome Announcements*, *5*(40), 959–976. https://doi.org/10.1128/GENOMEA.00959-17

Dutcher, J. D., Richard, D., Heuser, L. J., Pagano, J. F., & Perlman, D. (1956). *Methymycin* (United States US2916483A). Olin Corp. https://patents.google.com/patent/US2916483A/en

Earl, D., Bradnam, K., St. John, J., Darling, A., Lin, D., Fass, J., Yu, H. O. K., Buffalo, V., Zerbino, D. R., Diekhans, M., Nguyen, N., Ariyaratne, P. N., Sung, W. K., Ning, Z., Haimel, M., Simpson, J. T., Fonseca, N. A., Birol, I., Docking, T. R., … Paten, B. (2011). Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Research*, *21*(12), 2224. https://doi.org/10.1101/GR.126599.111

Ebada, S. S., Edrada, R. A., Lin, W., & Proksch, P. (2008). Methods for isolation,

purification and structural elucidation of bioactive secondary metabolites from marine invertebrates. *Nature Protocols*, *3*(12), 1820–1831. https://doi.org/10.1038/nprot.2008.182

Egidi, E., Wood, J. L., Fox, E. M., Liu, W., & Franks, A. E. (2017). Draft genome sequence of *Leifsonia* sp. strain NCR5, a rhizobacterium isolated from cadmium-contaminated Soil. *Genome Announcements, 5*(23). https://doi.org/10.1128/GENOMEA.00520-17

El-Deeb, B., Fayez, K., & Gherbawy, Y. (2012). Isolation and characterization of endophytic bacteria from *Plectranthus tenuiflorus* medicinal plant in Saudi Arabia desert and their antimicrobial activities. *Journal of Plant Interactions, 8*(1), 56-64. https://doi.org/10.1080/17429145.2012.680077

Eltamany, E. E., Abdelmohsen, U. R., Ibrahim, A. K., Hassanean, H. A., Hentschel, U., & Ahmed, S. A. (2014). New antibacterial xanthone from the marine sponge-derived *Micrococcus* sp. EG45. *Bioorganic & Medicinal Chemistry Letters*, *24*(21), 4939–4942. https://doi.org/10.1016/J.BMCL.2014.09.040

Engene, N., Tronholm, A., Salvador-Reyes, L. A., Luesch, H., & Paul, V. J. (2015). *Caldora penicillata* gen. nov., comb. nov. (Cyanobacteria), a pantropical marine species with biomedical relevance. *Journal of Phycology*, *51*(4), 670. https://doi.org/10.1111/JPY.12309

Erikson D. (1941). Studies on some lake-mud strains of *Micromonospora*. *Journal of bacteriology*, *41*(3), 277–300. https://doi.org/10.1128/jb.41.3.277-300.1941

Faulkner, D. J., Harper, M. K., Haygood, M. G., Salomon, C. E., & Schmidt, E. W. (2000). Symbiotic bacteria in sponges: Sources of bioactive substances. In N. Fusetani (Ed.), *Drugs from the sea* (pp.107–119). Karger. https://doi.org/10.1159/000062486

Fenical, W. (1993). Chemical studies of marine bacteria: Developing a new resource. *Chemical Reviews*, *93*(5), 1673–1683. https://doi.org/10.1021/CR00021A001/ASSET/CR00021A001.FP.PNG_V03

Fenical, W., Baden, D., Burg, M., de Goyet, C. V., Grimes, J. D., Katz, M., Marcus, N. H., Pomponi, S., Rhines, P., Tester, P., & Vena, J. (1999). Marine derived pharmaceuticals and related bioactive compounds. In W. Fenical (Ed.), *From monsoons to microbes: Understanding the ocean's role in human health* (pp. 71-86). National Academies Press.

Fenical, W., & Jensen, P. R. (2006). Developing a new resource for drug discovery: Marine actinomycete bacteria. *Nature Chemical Biology, 2*(12), 666–673. https://doi.org/10.1038/nchembio841

Ferguson, R. L., Buckley, E. N., & Palumbo, A. V. (1984). Response of marine bacterioplankton to differential filtration and confinement. *Applied and Environmental Microbiology*, *47*(1), 49–55. https://doi.org/10.1128/AEM.47.1.49-55.1984

Fernandez-Cortes, A., Cuezva, S., Sanchez-Moral, S., Cañaveras, J. C., Porca, E., Jurado, V., Martin-Sanchez, P. M., & Saiz-Jimenez, C. (2011). Detection of human-induced environmental disturbances in a show cave. *Environmental Science and Pollution Research*, *18*(6), 1037–1045. https://doi.org/10.1007/S11356-011-0513-5/TABLES/2

Field, S. K. (2015). Bedaquiline for the treatment of multidrug-resistant tuberculosis: great promise or disappointment? *Therapeutic Advances in Chronic Disease*, *6*(4), 170–184. https://doi.org/10.1177/2040622315582325

Fischbach, M. A., & Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chemical Reviews, 106*(8), 3468-3496. https://doi.org/10.1021/cr0503097

Fischbach, M. A., Walsh, C. T., & Clardy, J. (2008). The evolution of gene collectives: How natural selection drives chemical innovation. *Proceedings of the National Academy of Sciences*, *105*(12), 4601–4608. https://doi.org/10.1073/PNAS.0709132105

Fleming, A., & Allison, V. D. (1922). Observations on a bacteriolytic substance ("lysozyme") found in secretions and tissues. *British Journal of Experimental Pathology, 3*(5), 252-260. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2047739/

Florey, H. W. (1946). The use of micro-organisms for therapeutic purposes. *The Yale Journal of Biology and Medicine*, *19*(1), 101–117.

Foot, C. H., & Taylor, C. B. (1949). The influence of the composition of the medium on the growth of bacteria from water. *Proceedings of the Society for Applied Bacteriology*, *1*, 11–13.

Fosse, T., Toga, B., Peloux, Y., Granthil, C., Bertrando, J., & Sethian, M. (1985). Meningitis due to *Micrococcus luteus*. *Infection*, *13*(6), 280–281. https://doi.org/10.1007/BF01645439

Francisco, D. E., Mah, R. A., & Rabin, A. C. (1973). Acridine orange-epifluorescence technique for counting bacteria in natural waters. *Transactions of the American Microscopical Society*, *92*(3), 416–421. https://doi.org/10.2307/3225245

Frearson, J., & Wyatt, P. (2010). Drug discovery in academia- the third way?. *Expert opinion on drug discovery*, *5*(10), 909–919. https://doi.org/10.1517/17460441.2010.506508

Frick, M. (2021). *Pipeline report 2021. Tuberculosis vaccines*. Treatment Action Group. https://www.treatmentactiongroup.org/wp-content/uploads/2021/10/2021_pipeline_TB_vaccines_final.pdf

Frith, J. (2014). History of tuberculosis Part 1 – Pthisis, consumption and the White Plague. *Journal of Military and Veterans' Health*, *22*(2).

Funa, N., Ohnishi, Y., Fujli, I., Shibuya, M., Ebizuka, Y., & Horinouchi, S. (1999). A new pathway for polyketide synthesis in microorganisms. *Nature*, *400*(6747), 897–899. https://doi.org/10.1038/23748

Funabashi, M., Funa, N., & Horinouchi, S. (2008). Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *The Journal of Biological Chemistry*, *283*(20), 13983–13991. https://doi.org/10.1074/JBC.M710461200

Gao, W., Kim, J. Y., Anderson, J. R., Akopian, T., Hong, S., Jin, Y. Y., Kandror, O., Kim, J. W., Lee, I. A., Lee, S. Y., McAlpine, J. B., Mulugeta, S., Sunoqrot, S., Wang, Y., Yang, S. H., Yoon, T. M., Goldberg, A. L., Pauli, G. F., Suh, J. W., … Cho, S. (2015). The cyclic peptide ecumicin targeting CLpC1 is active against *Mycobacterium tuberculosis* in vivo. *Antimicrobial Agents and Chemotherapy*, *59*(2), 880–889. https://doi.org/10.1128/AAC.04054-14

Gerber, N. N. (1979). Volatile substances from actinomycetes: Their role in the odor pollution of water. *CRC Critical Reviews in Microbiology*, *7*(3), 191–214. https://doi.org/10.3109/10408417909082014

Ghareeb, M. A., Tammam, M. A., El-Demerdash, A., & Atanasov, A. G. (2020). Insights about clinically approved and preclinically investigated marine natural products. *Current Research in Biotechnology*, *2*, 88–102. https://doi.org/10.1016/J.CRBIOT.2020.09.001

Gokarn, K., & Pal, R. B. (2017). Preliminary evaluation of anti-tuberculosis potential of siderophores against drug-resistant *Mycobacterium tuberculosis* by mycobacteria growth indicator tube-drug sensitivity test. *BMC Complementary and Alternative Medicine, 17*(1). https://doi.org/10.1186/S12906-017-1665-8

Gong, T., Zhen, X., Li, X. L., Chen, J. J., Chen, T. J., Yang, J. L., & Zhu, P. (2018). Tetrocarcin Q, a new spirotetronate with a unique glycosyl group from a marine-derived actinomycete *Micromonospora carbonacea* LS276. *Marine Drugs*, *16*(2). https://doi.org/10.3390/MD16020074

Goodfellow, M., & Williams, S. T. (1983). Ecology of actinomycetes. *Annual Review of Microbiology, 37*, 189-216. https://doi.org/10.1146/annurev.mi.37.100183.001201

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology 57*(1), 81–91. https://doi.org/10.1099/IJS.0.64483-0/CITE/REFWORKS

Gräfe, U., Schade, W., Roth, M., Radics, L., Incze, M., & Ujszászy, K. (1984). Griseochelin, a novel carboxylic acid antibiotic from *Streptomyces griseus*. *The Journal of Antibiotics*, *37*(8), 836–846. https://doi.org/10.7164/ANTIBIOTICS.37.836

Gram, L., Melchiorsen, J., & Bruhn, J. B. (2010). Antibacterial activity of marine culturable bacteria collected from a global sampling of ocean surface waters and surface swabs of marine organisms. *Marine Biotechnology*, *12*(4), 439–451. https://doi.org/10.1007/s10126-009-9233-y

Greenblatt, C. L., Baum, J., Klein, B. Y., Nachshon, S., Koltunov, V., & Cano, R. J. (2004). *Micrococcus luteus* - survival in amber. *Microbial Ecology*, *48*(1), 120–127. https://doi.org/10.1007/S00248-003-2016-5/FIGURES/6

Gross, H. (2009). Genomic mining - a concept for the discovery of new bioactive natural products. *Current Opinion* in *Drug Discovery & Development*, *12*, 207-219.

Guerrero-Garzón, J. F., Zehl, M., Schneider, O., Rückert, C., Busche, T., Kalinowski, J., Bredholt, H., & Zotchev, S. B. (2020). *Streptomyces* spp. from the marine sponge *Antho dichotoma*: Analyses of secondary metabolite biosynthesis gene clusters and some of their products. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00437

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/BIOINFORMATICS/BTT086

Gürtler, H., Pedersen, R., Anthoni, U., Christophersen, C., Nielsen, P. H., Wellington, E. M. H., Pedersen, C., & Bock, K. (1994). Albaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a streptomycete with a new rope morphology. *The Journal of Antibiotics*, *47*(4), 434–439. https://doi.org/10.7164/antibiotics.47.434

Haagsma, A. C., Abdillahi-Ibrahim, R., Wagner, M. J., Krab, K., Vergauwen, K., Guillemont, J., Andries, K., Lill, H., Koul, A., & Bald, D. (2009). Selectivity of

TMC207 towards mycobacterial ATP synthase compared with that towards the eukaryotic homologue. *Antimicrobial Agents and Chemotherapy*, *53*(3), 1290. https://doi.org/10.1128/AAC.01393-08

Haydock, S. F., Aparicio, J. F., Molnár, I., Schwecke, T., Khaw, L. E., König, A., Marsden, A. F. A., Galloway, I. S., Staunton, J., & Leadlay, P. F. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Letters*, *374*(2), 246–248. https://doi.org/10.1016/0014-5793(95)01119-Y

Hechtel, G. J. (1983). New species of marine Demospongiae from Brazil. *Iheringia. Série Zoologia*, *63*, 59-89.

Henkel, T., Brunne, R. M., Müller, H., & Reichel, F. (1999). Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angewandte Chemie (International Edition in English)*, *38*(5), 643–647. https://doi.org/10.1002/(sici)1521-3773(19990301)38:5<643::aid-anie643>3.0.co;2-g

Hentschel, U., Fieseler, L., Wehrl, M., Gernert, C., Steinert, M., Hacker, J., & Horn, M. (2003). Microbial diversity of marine sponges. *Progress in Molecular and Subcellular Biology*, *37*, 59-88. https://doi.org/10.1007/978-3-642-55519-0_3

Hentschel, U., Hopke, J., Horn, M., Friedrich, A. B., Wagner, M., Hacker, J., & Moore, B. S. (2002). Molecular evidence for a uniform microbial community in sponges from different oceans. *Applied and Environmental Microbiology*, *68*(9), 4431–4440. https://doi.org/10.1128/AEM.68.9.4431-4440.2002

Hentschel, U., Piel, J., Degnan, S. M., & Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology*, *10*(9), 641-654. https://doi.org/10.1038/nrmicro2839

Hentschel, U., Schmid, M., Wagner, M., Fieseler, L., Gernert, C., & Hacker, J. (2001). Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges *Aplysina aerophoba* and *Aplysina cavernicola*. *FEMS Microbiol Ecol*, *35*(3), 305–312. https://doi.org/10.1111/J.1574-6941.2001.TB00816.X

Hentschel, U., Usher, K. M., & Taylor, M. W. (2006). Marine sponges as microbial fermenters. *FEMS Microbiology Ecology*, *55*(2), 167–177. https://doi.org/10.1111/j.1574-6941.2005.00046.x

Hesketh, A., Kock, H., Mootien, S., & Bibb, M. (2009). The role of absC, a novel regulatory gene for secondary metabolism, in zinc-dependent antibiotic production in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, *74*(6),

1427–1444. https://doi.org/10.1111/J.1365-2958.2009.06941.X

Hifnawy, M. S., Fouda, M. M., Sayed, A. M., Mohammed, R., Hassan, H. M., AbouZid, S. F., Rateb, M. E., Keller, A., Adamek, M., Ziemert, N., & Abdelmohsen, U. R. (2020). The genus *Micromonospora* as a model microorganism for bioactive natural product discovery. *RSC Advances*, *10*(35), 20939–20959. https://doi.org/10.1039/D0RA04025H

Hill, M., Hill, A., Lopez, N., & Harriott, O. (2006). Sponge-specific bacterial symbionts in the Caribbean sponge, *Chondrilla nucula* (Demospongiae, Chondrosida). *Marine Biology*, *148*(6), 1221–1230. https://doi.org/10.1007/s00227-005-0164-5

Hill, R. T. (2014). Microbes from marine sponges: A treasure trove of biodiversity for natural products discovery. In A. T. Bull (Ed.), *Microbial diversity and bioprospecting* (pp. 177–190). ASM Press. https://doi.org/10.1128/9781555817770.CH18

Hjörleifsson Eldjárn, G., Ramsay, A., van der Hooft, J. J. J., Duncan, K. R., Soldatou, S., Rousu, J., Daly, R., Wandy, J., & Rogers, S. (2021). Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions. *PLoS Computational Biology*, *17*(5), e1008920. https://doi.org/10.1371/journal.pcbi.1008920

Hodgson, D. A. (2000). Primary metabolism and its control in *Streptomycetes*: A most unusual group of bacteria. *Advances in Microbial Physiology*, *42*, 47-238. https://doi.org/10.1016/s0065-2911(00)42003-5

Hofnung, M., & Shapiro, J. A. (1999). Introduction. *Research in Microbiology, 150*(9-10), 577–578. doi:10.1016/s0923-2508(99)00133-3

Hooper, J. N. A., & van Soest, R. W. M. (2002). Systema porifera. A guide to the classification of sponges. In J. N. A. Hooper & R. W. M. van Soest (Eds.), *Systema porifera: A guide to the classification of sponges* (pp. 1-7). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0747-5_1

Hou, X. M., Wang, C. Y., Gerwick, W. H., & Shao, C. L. (2019). Marine natural products as potential anti-tubercular agents. *European Journal of Medicinal Chemistry*, *165*, 273–292. https://doi.org/10.1016/J.EJMECH.2019.01.026

Houben, R. M. G. J., & Dodd, P. J. (2016). The global burden of latent tuberculosis infection: A re-estimation using mathematical modelling. *PLoS Medicine*, *13*(10), e1002152. https://doi.org/10.1371/journal.pmed.1002152

Hoyer, K. M., Mahlert, C., & Marahiel, M. A. (2007). The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization. *Chemistry and Biology*,

14(1), 13–22. https://doi.org/10.1016/j.chembiol.2006.10.011

Huang, S., Liu, Y., Liu, W., Neubauer, P., & Li, J. (2021). The nonribosomal peptide valinomycin: From discovery to bioactivity and biosynthesis. *Microorganisms*, *9*(4). https://doi.org/10.3390/MICROORGANISMS9040780/S1

Hug, J. J., Krug, D., & Müller, R. (2020). Bacteria as genetically programmable producers of bioactive natural products. *Nature Reviews Chemistry*, *4*(4), 172–193. https://doi.org/10.1038/s41570-020-0176-1

Humisto, A., Jokela, J., Liu, L., Wahlsten, M., Wang, H., Permi, P., Machado, J. P., Antunes, A., Fewer, D. P., & Sivonen, K. (2018). The swinholide biosynthesis gene cluster from a terrestrial cyanobacterium, *Nostoc* sp. strain UHCC 0450. *Applied and Environmental Microbiology*, *84*(3). https://doi.org/10.1128/AEM.02321-17

Hunt, M., Silva, N. De, Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, *16*(1), 1–10. https://doi.org/10.1186/S13059-015-0849-0/FIGURES/3

Hur, G. H., Vickery, C. R., & Burkart, M. D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural Product Reports*, *29*(10), 1074-1098. https://doi.org/10.1039/c2np20025b

Hutchinson, D., Weaver, R. H., & Scherage, M. (1943). The incidence and significance of microorganisms antagonistic to *E. coli* in water. *Journal of Bacteriology*, *45*, 29-34.

Ian, E., Malko, D. B., Sekurova, O. N., Bredholt, H., Rückert, C., Borisova, M. E., Albersmeier, A., Kalinowski, J., Gelfand, M. S., & Zotchev, S. B. (2014). Genomics of sponge-associated *Streptomyces* spp. closely related to *Streptomyces albus* J1074: Insights into marine adaptation and secondary metabolite biosynthesis potential. *PLoS ONE*, *9*(5). https://doi.org/10.1371/JOURNAL.PONE.0096719

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., & Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnology*, *21*(5), 526–531. https://doi.org/10.1038/NBT820

Ikeda, Y., Naganawa, H., Kondo, S., & Takeuchi, T. (1992). Biosynthesis of bellenamine by *Streptomyces nashvillensis* using stable isotope labeled compounds. *The Journal of Antibiotics*, *45*(12), 1919–1924. https://doi.org/10.7164/ANTIBIOTICS.45.1919

Imhoff, J. F. & Stöhr, R. (2003). Sponge-associated bacteria: General overview and special aspects of bacteria associated with *Halichondria panicea*. *Progress in Molecular and Subcellular Biology, 37*, 35-57. doi: 10.1007/978-3-642-55519-0_2

Inahashi, Y., Matsumoto, A., Danbara, H., Ōmura, S., & Takahashi, Y. (2010). *Phytohabitans suffuscus* gen. nov., sp. nov., an actinomycete of the family Micromonosporaceae isolated from plant roots. *International Journal of Systemic and Evolutionary Microbiology*, *60*(11), 2652-2658. doi: 10.1099/ijs.0.016477-0.

Ishida, S., Arai, M., Niikawa, H., & Kobayashi, M. (2011). Inhibitory effect of cyclic trihydroxamate siderophore, desferrioxamine E, on the biofilm formation of *Mycobacterium* species. *Biological & Pharmaceutical Bulletin*, *34*(6), 917–920. https://doi.org/10.1248/BPB.34.917

Izumi, H., Gauthier, M. E. A., Degnan, B. M., Ng, Y. K., Hewavitharana, A. K., Shaw, P. N., & Fuerst, J. A. (2010). Diversity of *Mycobacterium* species from marine sponges and their sensitivity to antagonism by sponge-derived rifamycin-synthesizing actinobacterium in the genus *Salinispora*. *FEMS Microbiology Letters*, *313*(1), 33–40. doi: 10.1111/j.1574-6968.2010.02118.x.

Jagannadham, M. V., Rao, V. J., & Shivaji, S. (1991). The major carotenoid pigment of a psychrotrophic *Micrococcus roseus* strain: Purification, structure, and interaction with synthetic membranes. *Journal of Bacteriology*, *173*(24), 7911. https://doi.org/10.1128/JB.173.24.7911-7917.1991

Jannasch, H. W., & Jones, G. E. (1959) Bacterial populations in sea water as determined by different methods of enumeration 1. *Limnology and Oceanography*, *4*, 128–139. doi:10.4319/lo.1959.4.2.0128.

Jendrossek, D., Tomasi, G., & Kroppenstedt, R. M. (1997). Bacterial degradation of natural rubber: a privilege of actinomycetes? *FEMS Microbiology Letters*, *150*(2), 179-88. doi: 10.1016/s0378-1097(97)00072-4.

Jenke-Kodama, H., Börner, T., & Dittmann, E. (2006). Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLOS Computational Biology*, *2*(10), e132. https://doi.org/10.1371/JOURNAL.PCBI.0020132

Jensen, H. L. (1930) The genus *Micromonospora* Ørskov -a little known group of soil microorganisms. *Proceedings of the Linnean Society of New South Wales*, *55*, 231-248.

Jimenez, P. C. (2014). The first, the next, and the cinematographed versions of AZT. *Vitae*, *21*(2), 79–80.

Kaeberlein, T., Lewis, K., & Epstein, S. S. (2002). Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science*, *296*(5570), 1127–1129. https://doi.org/10.1126/science.1070633

Kallifidas, D., Pascoe, B., Owen, G. A., Strain-Damerell, C. M., Hong, H. J., & Paget, M. S. B. (2010). The zinc-responsive regulator zur controls expression of the coelibactin gene cluster in *Streptomyces coelicolor*. *Journal of Bacteriology*, *192*(2), 608. https://doi.org/10.1128/JB.01022-09

Kannenberg, E. L., & Poralla, K. (1999). Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften*, *86*, 168-176.

Karbalaei-Heidari, H. R., Partovifar, M., & Memarpoor-Yazdi, M. (2020). Evaluation of the bioactive potential of secondary metabolites produced by a new marine *Micrococcus* species isolated from the Persian Gulf. *Avicenna Journal of Medical Biotechnology*, *12*(1), 61-65.

Katz, L., & Baltz, R. H. (2016). Natural product discovery: Past, present, and future. *Journal of Industrial Microbiology and Biotechnology*, *43*(2-3), 155–176. https://doi.org/10.1007/s10295-015-1723-5

Kerr, R. G. & Kelly-Borges, M. (1994). Biochemical and morphological heterogeneity in the Caribbean sponge *Xestospongia m*uta (Petrosida: Petrosiidae). In R. W. M. van Soest, T. M. G. van Kempen, & J. C. Braekman (Eds.), *Sponges in time and space* (pp. 65-73). Balkema.

Kim, T. K., Hewavitharana, A. K., Shaw, P. N., & Fuerst, J. A. (2006). Discovery of a new source of rifamycin antibiotics in marine sponge actinobacteria by phylogenetic prediction. *Applied and Environmental Microbiology*, *72*(3), 2118–2125. https://doi.org/10.1128/AEM.72.3.2118-2125.2006

Kincheloe, G. N., Eisen, J. A., & Coil, D. A. (2017). Draft genome sequence of *Arthrobacter* sp. strain UCD-GKA (phylum Actinobacteria). *Genome Announcements*, *5*(6). https://doi.org/10.1128/GENOMEA.01599-16

King, A. (2021, July 1) *Tuberculosis: The forgotten pandemic*. The Scientist. https://www.the-scientist.com/features/tuberculosis-the-forgotten-pandemic-68894?utm_campaign=TS_DAILY_NEWSLETTER_2021&utm_medium=email&_hsmi=162765238&_hsenc=p2ANqtz--RAnlRYt51e0dDraa3yfwXyqd3NWMxgxLJ2-C7VyTQVjxQ8docID21kIBk38OhNdQHhZbEubzSa_gEvRrGZiJP79AIJA&utm_content=162765238&utm_source=hs_email

Kittendorf, J. D., & Sherman, D. H. (2009). The methymycin/pikromycin pathway: A model for metabolic diversity in natural product biosynthesis. *Bioorganic &*

Medicinal Chemistry, *17*(6), 2137–2146.
https://doi.org/10.1016/J.BMC.2008.10.082

Klein, B. A., Lemon, K. P., Faller, L. L., Jospin, G., Eisen, J. A., & Coil, D. A. (2016). Draft genome sequence of *Curtobacterium* sp. strain UCD-KPL2560 (phylum Actinobacteria). *Genome Announcements*, *4*(5). https://doi.org/10.1128/GENOMEA.01040-16

Kleinkauf, H., & von Döhren, H. (1990). Nonribosomal biosynthesis of peptide antibiotics. *European Journal of Biochemistry*, *92*(1), 1–15. https://doi.org/10.1111/j.1432-1033.1990.tb19188.x

Klitgaard, A., Nielsen, J. B., Frandsen, R. J. N., Andersen, M. R., & Nielsen, K. F. (2015). Combining stable isotope labeling and molecular networking for biosynthetic pathway characterization. *Analytical Chemistry*, *87*(13), 6520–6526. https://doi.org/10.1021/ACS.ANALCHEM.5B01934/SUPPL_FILE/AC5B01934_SI_001.PDF

Koch, R. (1881) *Zur Untersuchung von pathogenen Organismen*. Norddeutschen Buchdruckerei und Verlagsanstalt.

Kocur, M. (1986). Genus I. Micrococcus Cohn 1872, 151[AL]. In N. R. Krieg & J. G. Holt (Eds.), Bergey's manual of systematic bacteriology (vol. 2, pp. 1004-1008). The Williams & Wilkins Co.

Koenigsaecker, T. M., Eisen, J. A., & Coil, D. A. (2016). Draft genome sequence of *Gordonia* sp. strain UCD-TK1 (phylum Actinobacteria). *Genome Announcements*, *4*(5), 1121–1137. https://doi.org/10.1128/GENOMEA.01121-16

Kolter, R., & van Wezel, G. P. (2016). Goodbye to brute force in antibiotic discovery? *Nature Microbiology*, *1*(2), 1–2. https://doi.org/10.1038/nmicrobiol.2015.20

Kong, D. X., Jiang, Y. Y., & Zhang, H. Y. (2010). Marine natural products as sources of novel scaffolds: Achievement and concern. *Drug Discovery Today*, *15*(21-22), 884–886. https://doi.org/10.1016/j.drudis.2010.09.002

Konz, D., Klens, A., Schörgendorfer, K., & Marahiel, M. A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: Molecular characterization of three multi-modular peptide synthetases. *Chemistry and Biology*, *4*(12), 927–937. https://doi.org/10.1016/S1074-5521(97)90301-X

Kraljevic, S., Stambrook, P. J., & Pavelic, K. (2004). Accelerating drug discovery. *EMBO Reports*, *5*(9), 837. https://doi.org/10.1038/SJ.EMBOR.7400236

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/MOLBEV/MSY096

Kumar, T. S., Josephine, A., Sreelatha, T., Azger Dusthackeer, V. N., Mahizhaveni, B., Dharani, G., Kirubagaran, R., & Raja Kumar, S. (2020). Fatty acids-carotenoid complex: An effective anti-TB agent from the chlorella growth factor-extracted spent biomass of *Chlorella vulgaris*. *Journal of Ethnopharmacology*, *249*, 112392. https://doi.org/10.1016/J.JEP.2019.112392

Kurtböke, D. İ. (2012). Biodiscovery from rare actinomycetes: An eco-taxonomical perspective. *Applied Microbiology and Biotechnology*, *93*, 1843–1852. https://doi.org/10.1007/s00253-012-3898-2

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), 1–9. https://doi.org/10.1186/GB-2004-5-2-R12/FIGURES/3

Kuśmirek, W., & Nowak, R. (2018). De novo assembly of bacterial genomes with repetitive DNA regions by dnaasm application. *BMC Bioinformatics*, *19*(1), 1–10. https://doi.org/10.1186/S12859-018-2281-4/TABLES/6

Lakshminarasim, V. & Iya, K. K. (1955). Studies on the micrococci in milk. Part I. Incidence and distribution. *Indian Journal* of *Dairy Science*, *8*, 67**.**

Lam, K. S. (2006). Discovery of novel metabolites from marine actinomycetes. *Current Opinion in Microbiology 9*(3), 245–251. https://doi.org/10.1016/J.MIB.2006.03.004

Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., & Walsh, C. T. (1996). A new enzyme superfamily – the phosphopantetheinyl transferases. *Chemistry and Biology*, *3*(11), 923–936. https://doi.org/10.1016/S1074-5521(96)90181-7

Lawrence, J. G., & Roth, J. R. (1996). Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics*, *143*(4), 1843-1860. https://academic.oup.com/genetics/article/143/4/1843/6016793

Lee, M. D., Manning, J. K., Williams, D. R., Kuck, N. A., Testa, R. T., & Borders, D. B. (1989). Calicheamicins, a novel family of antitumor antibiotics. 3. Isolation, purification and characterization of calicheamicins beta 1Br, gamma 1Br, alpha 2I, alpha 3I, beta 1I, gamma 1I and delta 1I. *Journal of Antibiotics, 42*(7), 1070-1087. doi: 10.7164/antibiotics.42.1070. PMID: 2753814

Lee, O. O., Wang, Y., Yang, J., Lafi, F. F., Al-Suwailem, A., & Qian, P.-Y. (2010). Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *The ISME Journal*, *5*(4), 650–664. https://doi.org/10.1038/ismej.2010.165

Lewis, K. (2013). Platforms for antibiotic discovery. *Nature Reviews Drug Discovery*, *12*(5), 371–387. https://doi.org/10.1038/nrd3975

Li, J., Zhao, G. Z., Huang, H. Y., Qin, S., Zhu, W. Y., Zhao, L. X., Xu, L. H., Zhang, S., Li, W. J., & Strobel, G. (2012). Isolation and characterization of culturable endophytic Actinobacteria associated with *Artemisia annua* L. . *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *101*(3), 515–527. https://doi.org/10.1007/S10482-011-9661-3/TABLES/4

Li, M. H. T., Ung, P. M. U., Zajkowski, J., Garneau-Tsodikova, S., & Sherman, D. H. (2009). Automated genome mining for natural products. *BMC Bioinformatics*, *10*, 185. https://doi.org/10.1186/1471-2105-10-185

Liaaen-Jensen, S. (1978). Marine carotenoids. In P. J. Scheuer (Ed.), *Marine natural products. Chemical and biological perspectives* (vol. 2., pp. 1-73). Academic Press.

Lindequist, U. (2016). Marine-derived pharmaceuticals - challenges and opportunities. *Biomolecules & Therapeutics*, *24*(6), 561–571. https://doi.org/10.4062/biomolther.2016.181

Liu, W., Li, L., Khan, M. A., & Zhu, F. (2012). Popular molecular markers in bacteria. *Molecular Genetics, Microbiology and Virology*, *27*(3), 103–107. https://doi.org/10.3103/S0891416812030056

Liu, W. T., Lamsa, A., Wong, W. R., Boudreau, P. D., Kersten, R., Peng, Y., Moree, W. J., Duggan, B. M., Moore, B. S., Gerwick, W. H., Linington, R. G., Pogliano, K., & Dorrestein, P. C. (2014). MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *Journal of Antibiotics*, *67*(1), 99. https://doi.org/10.1038/JA.2013.99

Liu, T., Wu, S., Zhang, R., Wang, D., Chen, J., & Zhao, J. (2019). Diversity and antimicrobial potential of Actinobacteria isolated from diverse marine sponges along the Beibu Gulf of the South China Sea. *FEMS Microbiology Ecology*, *95*(7). https://doi.org/10.1093/FEMSEC/FIZ089

Lo Grasso, H. (1928). The treatment of tuberculosis by heliotherapy. *Radiology, 11*(3). https://doi.org/10.1148/11.3.217

Löde, A. (1902). *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene (1 Orig)*, *33*, 196.

Luca, S., & Mihaescu, T. (2013). History of BCG vaccine. *Maedica - A Journal of Clinical Medicine*, *8*(1), 53–58.

Luesch, H., Moore, R. E., Paul, V. J., Mooberry, S. L., & Corbett, T. H. (2001). Isolation of dolastatin 10 from the marine cyanobacterium *Symploca* species VP642 and total stereochemistry and biological evaluation of its analogue symplostatin 1. *Journal of Natural Products*, *64*(7), 907–910. https://doi.org/10.1021/NP010049Y

Magoc, T., Pabinger, S., Canzar, S., Liu, X., Su, Q., Puiu, D., Tallon, L. J., & Salzberg, S. L. (2013). GAGE-B: An evaluation of genome assemblers for bacterial organisms. *Bioinformatics*, *29*(14), 1718. https://doi.org/10.1093/BIOINFORMATICS/BTT273

Majeed, H. Z. (2017). Antimicrobial activity of *Micrococcus luteus* cartenoid pigment. *Al-Mustansiriyah Journal of Science*, *28*(1), 64. https://doi.org/10.23851/MJS.V28I1.314

Maldonado, L. A., Fenical, W., Jensen, P. R., Kauffman, C. A., Mincer, T. J., Ward, A. C., Bull, A. T., & Goodfellow, M. (2005). *Salinispora arenicola* gen. nov., sp. nov. and *Salinispora tropica* sp. nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. *International Journal of Systematic and Evolutionary Microbiology*, *55*(Pt 5), 1759–1766. https://doi.org/10.1099/IJS.0.63625-0

Maskey, R. P., Grün-Wollny, I., & Laatsch, H. (2005). Isolation and structure elucidation of diazaquinomycin C from a terrestrial *Streptomyces* sp. and confirmation of the akashin structure. *Natural Product Research*, *19*(2), 137–142. https://doi.org/10.1080/14786410410001704741

Mao, D., Okada, B. K., Wu, Y., Xu, F., & Seyedsayamdost, M. R. (2018). Recent advances in activating silent biosynthetic gene clusters in bacteria. *Current Opinion in Microbiology*, *45*, 156–163. https://doi.org/10.1016/j.mib.2018.05.001

Martín, J. F., & Liras, P. (1989). Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annual Review of Microbiology*, *43*, 173-206. https://doi.org/10.1146/annurev.mi.43.100189.001133

Matroodi, S., Siitonen, V., Baral, B., Yamada, K., Akhgari, A., & Metsä-Ketelä, M. (2020). Genotyping-guided discovery of persiamycin A from sponge-associated halophilic *Streptomonospora* sp. PA3. *Frontiers in Microbiology*, *11*, 1237.

https://doi.org/10.3389/fmicb.2020.01237

Mayer, A. M. S. (2021, October). *Marine pharmacology: Clinical pipeline*. Midwestern University. https://www.midwestern.edu/departments/marinepharmacology/clinical-pipeline.

Mayer, A.M., Glaser, K. B., Cuevas, C., Jacobs, R. S., Kem, W., Little, R. D., McIntosh, J. M., Newman, D. J., Potts, B. C., & Shuster, D. E. (2010). The odyssey of marine pharmaceuticals: a current pipeline perspective. Trends in Pharmacological Sciences, 31(6):255-65. doi: 10.1016/j.tips.2010.02.005.

McAlpine, J. B., Bachmann, B. O., Piraee, M., Tremblay, S., Alarco, A. M., Zazopoulos, E., & Farnet, C. M. (2005). Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *Journal of Natural Products*, *68*(4), 493–496. https://doi.org/10.1021/NP0401664/SUPPL_FILE/NP0401664_S.PDF

McCaughey, C. S., van Santen, J. A., van der Hooft, J. J. J., Medema, M. H., & Linington, R. G. (2022). An isotopic labeling approach linking natural products with biosynthetic gene clusters. *Nature Chemical Biology*, *18*(3), 295–304. https://doi.org/10.1038/S41589-021-00949-6

McCauley, E. P., Piña, I. C., Thompson, A. D., Bashir, K., Weinberg, M., Kurz, S. L., & Crews, P. (2020). Highlights of marine natural products having parallel scaffolds found from marine-derived bacteria, sponges, and tunicates. *The Journal of Antibiotics, 73*(8), 504–525. https://doi.org/10.1038/s41429-020-0330-5

McDonald, B. R., & Currie, C. R. (2017). Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *MBio*, *8*(3). https://doi.org/10.1128/mBio.00644-17

McIntosh, J. A., Donia, M. S., & Schmidt, E. W. (2009). Ribosomal peptide natural products: Bridging the ribosomal and nonribosomal worlds. *Natural Product Reports*, *26*(4), 537-559. https://doi.org/10.1039/b714132g

McMurray, S. E., Blum, J. E., & Pawlik, J. R. (2008). Redwood of the reef: Growth and age of the giant barrel sponge *Xestospongia muta* in the Florida Keys. *Marine Biology*, *155*(2), 159–171. https://doi.org/10.1007/S00227-008-1014-Z

Medema, M. H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., 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 Research*, *39*(Web Server issue). https://doi.org/10.1093/NAR/GKR466

Medema, M. H., & Fischbach, M. A. (2015). Computational approaches to natural product discovery. *Nature Chemical Biology*, *11*(9), 639. https://doi.org/10.1038/NCHEMBIO.1884

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., … Glöckner, F. O. (2015). Minimum information about a biosynthetic gene cluster. *Nature Chemical Biology*, *11*(9), 625–631. https://doi.org/10.1038/nchembio.1890

Miao, V., Coëffet-LeGal, M.-F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C. J., Wrigley, S. K., & Baltz, R. H. (2005). Daptomycin biosynthesis in *Streptomyces roseosporus*: Cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology*, *151*(5), 1507–1523. https://doi.org/10.1099/MIC.0.27757-0

Mincer, T. J., Jensen, P. R., Kauffman, C. A., & Fenical, W. (2002). Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Applied and Environmental Microbiology*, *68*(10), 5005–5011. https://doi.org/10.1128/AEM.68.10.5005-5011.2002

Mohana, D. C., Thippeswamy, S., & Abhishek, R. U. (2013). Antioxidant, antibacterial, and ultraviolet-protective properties of carotenoids isolated from *Micrococcus* spp. *Radiation Protection and Environment*, *36*(4), 168. https://doi.org/10.4103/0972-0464.142394

Mojib, N., Philpott, R., Huang, J. P., Niederweis, M., & Bej, A. K. (2010). Antimycobacterial activity in vitro of pigments isolated from Antarctic bacteria. *Antonie Van Leeuwenhoek*, *98*(4), 531-540. https://doi.org/10.1007/s10482-010-9470-0

Montalvo, N. F. (2011). *The bacterial communities associated with two marine sponges of the genus Xestospongia muta.* [Doctoral dissertation, University of Maryland Baltimore]. http://hdl.handle.net/10713/788

Montalvo, N. F., Davis, J., Vicente, J., Pittiglio, R., Ravel, J., & Hill, R. T. (2014). Integration of culture-based and molecular analysis of a complex sponge-associated bacterial community. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0090517

Montalvo, N. F., & Hill, R. T. (2011). Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts. *Applied and Environmental Microbiology*, *77*(20), 7207–7216. https://doi.org/10.1128/AEM.05285-11

Montalvo, N. F., Mohamed, N. M., Enticknap, J. J., & Hill, R. T. (2005). Novel Actinobacteria from marine sponges. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *87*(1), 29–36. https://doi.org/10.1007/s10482-004-6536-x

Montaser, R., & Luesch, H. (2011). Marine natural products: A new wave of drugs? *Future Medicinal Chemistry*, *3*(12), 1475–1489. https://doi.org/10.4155/FMC.11.118

Mootz, H. D., Schwarzer, D., & Marahiel, M. A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem*, *3*(6), 490–504.. https://doi.org/10.1002/1439-7633(20020603)3:6<490::AID-CBIC490>3.0.CO;2-N

Morse, D. (1961). Prehistoric tuberculosis in America. *The American Review of Respiratory Disease*, *83*, 489–504. https://doi.org/10.1164/ARRD.1961.83.4.489

Morse, D., Brothwell, D. R., & Ucko, P. J. (1964). Tuberculosis in ancient Egypt. *The American Review of Respiratory Disease*, 90, 524-541.

Mossialos, E., Morel, C. M., Edwards, S. E., Berensen, J., Gemmill, M., & Brogen, D. (2010). Policies and incentives for promoting innovation in antibiotic research. *European Observatory on Health Systems and Policies*, 1–197. http://www.euro.who.int/en/home/projects/observatory/publications

Motohashi, K., Ueno, R., Sue, M., Furihata, K., Matsumoto, T., Dairi, T., Omura, S., & Seto, H. (2007). Studies on terpenoids produced by actinomycetes: Oxaloterpins A, B, C, D, and E, diterpenes from *Streptomyces* sp. KO-3988. *Journal of Natural Products*, *70*(11), 1712–1717. https://doi.org/10.1021/NP070326M

Mullin, E. (2016, May 10*). How tuberculosis shaped victorian fashion.* Smithsonian Magazine. https://www.smithsonianmag.com/science-nature/how-tuberculosis-shaped-victorian-fashion-180959029/

Mullowney, M. W. (2016). Antibiotics from aquatic-derived actinomycete bacteria that inhibit *M. tuberculosis*. [Doctoral dissertation, University of Illinois at Chicago]. https://hdl.handle.net/10027/21542

Mullowney, M. W., Hwang, C. H., Newsome, A. G., Wei, X., Tanouye, U., Wan, B., Carlson, S., Barranis, N. J., Ó Hainmhire, E., Chen, W. L., Krishnamoorthy, K., White, J., Blair, R., Lee, H., Burdette, J. E., Rathod, P. K., Parish, T., Cho, S., Franzblau, S. G., & Murphy, B. T. (2015). Diaza-anthracene antibiotics from a freshwater-derived actinomycete with selective antibacterial activity toward *Mycobacterium tuberculosis*. *ACS Infectious Diseases*, *1*(4), 168–174. https://doi.org/10.1021/ACSINFECDIS.5B00005

214

Mullowney, M. W., Shaikh, A., Wei, X., Tanouye, U., Santarsiero, B. D., Burdette, J. E., & Murphy, B. T. (2014). Diazaquinomycins E-G, novel diaza-anthracene analogs from a marine-derived *Streptomyces* sp. *Marine Drugs*, *12*(6), 3574–3586. https://doi.org/10.3390/md12063574

Murakami, Y., Oshima, Y., & Yasumoto, T. (1982). Identification of okadaic acid as a toxic component of a marine dinoflagellate *Prorocentrum lima*. *Bulletin* of the Japanese Society of Scientific Fisheries, *48*(1), 69–72. https://doi.org/10.2331/suisan.48.69

Murata, M., Miyasaka, T., Tanaka, H., & Omura, S. J. (1985). *The Journal of Antibiotics*, *38*(8), 1025–1033.

Nakagawa, A., Iwai, Y., Hashimoto, H., Miyazaki, N., ŌIwa, R., Takahashi, Y., Hirano, A., Shibukawa, N., Kojima, Y., & ŌMura, S. (1981). Virantmycin, a new antiviral antibiotic produced by a strain of *Streptomyces*. *Journal of Antibiotics*, *34*(11), 1408–1415. https://doi.org/10.7164/antibiotics.34.1408

Nakao, Y., & Fusetani, N. (2010). Marine invertebrates: Sponges. In L. Mander & H. Liu (Eds.), *Comprehensive Natural Products II: Chemistry and Biology* (vol. 2, pp. 327-362). Elsevier Ltd.

Nazari, B., Forneris, C. C., Gibson, M. I., Moon, K., Schramma, K. R., & Seyedsayamdost, M. R. (2017). *Nonomuraea* sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *MedChemComm*, *8*(4), 780–788. https://doi.org/10.1039/c6md00637j

Netzer, R., Stafsnes, M. H., Andreassen, T., Goksøyr, A., Bruheim, P., & Brautaset, T. (2010). Biosynthetic pathway for γ-cyclic sarcinaxanthin in *Micrococcus luteus*: Heterologous expression and evidence for diverse and multiple catalytic functions of C50 carotenoid cyclases. *Journal of Bacteriology*, *192*(21), 5688–5699. https://doi.org/10.1128/JB.00724-10

Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, *83*(3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

Newman, D. J., Cragg, G. M., & Battershill, C. N. (2009). *Therapeutic agents from the sea: Biodiversity, chemo-evolutionary insight and advances to the end of Darwin's 200th year*. *Diving and Hyperbaric Medicine*, 39(4), 216-225.

Newman, D. J., & Hill, R. T. (2006). New drugs from marine microbes: The tide is turning. *Journal of Industrial Microbiology and Biotechnology*, *33*(7), 539–544. https://doi.org/10.1007/S10295-006-0115-2

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., & Epstein, S. S. (2010). Use of ichip for high-throughput *in situ* cultivation of "uncultivable" microbial species. *Applied and Environmental Microbiology*, *76*(8), 2445. https://doi.org/10.1128/AEM.01754-09

Nisha, P., John, N., Mamatha, C., & Thomas, M. (2020). Characterization of bioactive compound produced by microfouling Actinobacteria (*Micrococcus luteus*) isolated from the ship hull in Arabian Sea, Cochin. Kerala. *Materials Today: Proceedings*, 25, 257–264. https://doi.org/10.1016/j.matpr.2020.01.362

Nuñez, M. (2014). *Encyclopedia of Food Microbiology* (2nd ed., pp. 627-633). Elsevier.

O'Shea, R., & Moser, H. E. (2008). Physicochemical properties of antibacterial compounds: Implications for drug discovery. *Journal of Medicinal Chemistry*, *51*(10), 2871–2878. https://doi.org/10.1021/jm700967e

Ochi, K., & Hosaka, T. (2013). New strategies for drug discovery: Activation of silent or weakly expressed microbial gene clusters. *Applied Microbiology and Biotechnology*, *97*(1), 87–98. https://doi.org/10.1007/s00253-012-4551-9

Okami, Y. (1982). Potential use of marine microorganisms for antibiotics and enzyme production. *Pure and Applied Chemistry*, *54*(10), 1951–1962. https://doi.org/10.1351/pac198254101951

Olson, J. B., Lord, C. C., & Mccarthy, P. J. (2000). Improved recoverability of microbial colonies from marine sponge samples. *Microbial Ecology*, *40*(2), 139-147. https://doi.org/10.1007/s002480000058

Oman, T. J., & van der Donk, W. A. (2010). Follow the leader: The use of leader peptides to guide natural product biosynthesis. *Nature Chemical Biology*, *6*(1), 9. https://doi.org/10.1038/NCHEMBIO.286

Ōmura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., & Hattori, M. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proceedings of the National Academy of Sciences*, *98*(21), 12215–12220. https://doi.org/10.1073/PNAS.211433198

Omura, S., Nakagawa, A., Aoyama, H., Hinotozawa, K., & Sano, H. (1983). The structures of diazaquinomycins A and B, new antibiotic metabolites. *Tetrahedron Letters*, *24*, 3643–3646. doi: 10.1016/S0040-4039(00)88190-3

Ørskov, J. (1923). Investigations into the morphology of the ray fungi. Levin & Munksgaard.

Osawa, A., Ishii, Y., Sasamura, N., Morita, M., Kasai, H., Maoka, T., & Shindo, K. (2010). Characterization and antioxidative activities of rare C 50 carotenoids-sarcinaxanthin, sarcinaxanthin monoglucoside, and sarcinaxanthin diglucoside-obtained from *Micrococcus yunnanensis*. *Journal of Oleo Science*, *59*(12). http://www.jstage.jst.go.jp/browse/jos/

Palomo, S., González, I., De La Cruz, M., Martín, J., Tormo, J. R., Anderson, M., Hill, R. T., Vicente, F., Reyes, F., & Genilloud, O. (2013). Sponge-derived *Kocuria* and *Micrococcus* spp. as sources of the new thiazolyl peptide antibiotic kocurin. *Marine Drugs*, *11*(4), 1071. https://doi.org/10.3390/MD11041071

Pan, R., Bai, X., Chen, J., Zhang, H., & Wang, H. (2019). Exploring structural diversity of microbe secondary metabolites using OSMAC strategy: A literature review. *Frontiers in Microbiology*, *10*, 294. https://doi.org/10.3389/fmicb.2019.00294

Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., Davies, R. M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Moule, S., … Barrell, B. G. (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature*, *404*(6777), 502–506. https://doi.org/10.1038/35006655

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

Parra, J., & Duncan, K. R. (2019). Assessing metabolite biogeography of *Micrococcus* spp. and *Pseudonocardia* spp. isolated from marine environments. *Access Microbiology*, *1*(1A), 587. https://doi.org/10.1099/ACMI.AC2019.PO0359

Patel, S., Petty, W. J., & Sands, J. M. (2021). An overview of lurbinectedin as a new second-line treatment option for small cell lung cancer. *Therapeutic Advances in Medical Oncology*, *13*. https://doi.org/10.1177/17588359211020529

Paterson, I., & Anderson, E. A. (2005). Chemistry. The renaissance of natural products as drug candidates. *Science (New York, N.Y.)*, *310*(5747), 451–453. https://doi.org/10.1126/SCIENCE.1116364

Pathom-aree, W., Stach, J. E. M., Ward, A. C., Horikoshi, K., Bull, A. T., & Goodfellow, M. (2006). Diversity of actinomycetes isolated from Challenger

Deep sediment (10,898 m) from the Mariana Trench. *Extremophiles*, *10*(3), 181–189. https://doi.org/10.1007/s00792-005-0482-z

Patridge, E., Gareiss, P., Kinch, M. S., & Hoyer, D. (2016). An analysis of FDA-approved drugs: Natural products and their derivatives. *Drug Discovery Today*, *21*(2), 204–207. https://doi.org/10.1016/J.DRUDIS.2015.01.009

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, *9*(3), 203–214. https://doi.org/10.1038/NRD3078

Payne, D., Gwynn, M., Holmes, D., & Pompliano, D. (2007). Drugs for bad bugs: Confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery*, *6*(1), 29–40. https://doi.org/10.1038/NRD2201

Pease, A. S. (1940). Some remarks on the diagnosis and treatment of tuberculosis in antiquity. *Isis*, *31*(2), 380–393. https://doi.org/10.1086/347595

Peng, J., Yuan, J. P., Wu, C. F., & Wang, J. H. (2011). Fucoxanthin, a marine carotenoid present in brown seaweeds and diatoms: metabolism and bioactivities relevant to human health. *Marine Drugs*, *9*(10), 1806. https://doi.org/10.3390/MD9101806

Peng, S. X. (2000). Hyphenated HPLC-NMR and its applications in drug discovery. *Biomedical Chromatography*, *14*(6), 430–441. https://doi.org/10.1002/1099-0801(200010)14:6<430::AID-BMC32>3.0.CO;2-P

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), 1420–1428. https://doi.org/10.1093/BIOINFORMATICS/BTS174

Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., Foster, B., Lapidus, A., Podell, S., Allen, E. E., Moore, B. S., & Jensen, P. R. (2009). Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *The ISME Journal*, *3*(10), 1193. https://doi.org/10.1038/ISMEJ.2009.58

Penn, K., & Jensen, P. R. (2012). Comparative genomics reveals evidence of marine adaptation in *Salinispora* species. *BMC Genomics 2012 13:1*, *13*(1), 1–12. https://doi.org/10.1186/1471-2164-13-86

Peraud, O. (2006). *Isolation and characterization of a sponge-associated actinomycete that produces manzamines*. [Doctoral dissertation, University of Maryland, College Park]. ProQuest Dissertations Publishing.

Perry, N. B., Blunt, J. W., Munro, M. H., & Pannell, L. K. (1988). Mycalamide A, an antiviral compound from a New Zealand sponge of the genus *Mycale*. *Journal of the American Chemical Society*, *110*(14), 4850–4851. https://doi.org/10.1021/ja00222a067

Piel, J. (2011). Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annual Review of Microbiology*, *65*, 431–453. https://doi.org/10.1146/ANNUREV-MICRO-090110-102805

Piel, J., Butzke, D., Fusetani, N., Hui, D., Platzer, M., Wen, G., & Matsunaga, S. (2005). Exploring the chemistry of uncultivated bacterial symbionts: Antitumor polyketides of the pederin family. *Journal of Natural Products*, *68*(3), 472–479. https://doi.org/10.1021/np049612d

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., & Matsunaga, S. (2004). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proceedings of the National Academy of Sciences*, *101*(46), 16222–16227. https://doi.org/10.1073/PNAS.0405976101

Poorinmohammad, N., Bagheban-Shemirani, R., & Hamedi, J. (2019). Genome mining for ribosomally synthesised and post-translationally modified peptides (RiPPs) reveals undiscovered bioactive potentials of Actinobacteria. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *112*(10), 1477–1499. https://doi.org/10.1007/s10482-019-01276-6

Prior, A. M., & Sun, D. (2019). Total synthesis of diazaquinomycins H and J using double Knorr cyclization in the presence of triisopropylsilane. *RSC Advances*, *9*(4), 1759–1771. https://doi.org/10.1039/C8RA09792E

Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H., & Linington, R. G. (2017). Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(22), 5601. https://doi.org/10.1073/PNAS.1614680114

Qi, S., Gui, M., Li, H., Yu, C., Li, H., Zeng, Z., & Sun, P. (2020). Secondary metabolites from marine *Micromonospora*: Chemistry and bioactivities. *Chemistry and Biodiversity*, *17*(4). https://doi.org/10.1002/cbdv.202000024

Rajwani, R., Ohlemacher, S. I., Zhao, G., Liu, H.-B., & Bewley, C. A. (2021). Genome-guided discovery of natural products through multiplexed low-coverage whole-genome sequencing of soil actinomycetes on Oxford Nanopore Flongle. *MSystems*, *6*(6). https://doi.org/10.1128/msystems.01020-21

Rangseekaew, P., & Pathom-aree, W. (2019). Cave actinobacteria as producers of bioactive metabolites. *Frontiers in Microbiology 10*, 387.

https://doi.org/10.3389/fmicb.2019.00387

Ranjan, R., & Jadeja, V. (2017). Isolation, characterization and chromatography based purification of antibacterial compound isolated from rare endophytic actinomycetes *Micrococcus yunnanensis*. *Journal of Pharmaceutical Analysis*, *7*(5), 343–347. https://doi.org/10.1016/J.JPHA.2017.05.001

Rappé, M. S., Connon, S. A., Vergin, K. L., & Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature*, *418*(6898), 630–633. https://doi.org/10.1038/nature00917

Rath, C. M., Janto, B., Earl, J., Ahmed, A., Hu, F. Z., Hiller, L., Dahlgren, M., Kreft, R., Yu, F., Wolff, J. J., Kweon, H. K., Christiansen, M. A., Håkansson, K., Williams, R. M., Ehrlich, G. D., & Sherman, D. H. (2011). Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chemical Biology*, *6*(11), 1244. https://doi.org/10.1021/CB200244T

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., & Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, *7*(1), 1–15. https://doi.org/10.1186/1471-2148-7-78

Rebets, Y., Brötz, E., Tokovenko, B., & Luzhetskyy, A. (2014). Actinomycetes biosynthetic potential: How to bridge in silico and in vivo? *Journal of Industrial Microbiology & Biotechnology*, *41*(2), 387–402. https://doi.org/10.1007/S10295-013-1352-9

Reiswig, H. M. (1981). Partial carbon and energy budgets of the bacteriosponge *Verohgia fistularis* (Porifera: Demospongiae) in Barbados. *Marine Ecology*, *2*(4), 273–293. https://doi.org/10.1111/J.1439-0485.1981.TB00271.X

Reveillaud, J., Maignien, L., Eren, M. A., Huber, J. A., Apprill, A., Sogin, M. L., & Vanreusel, A. (2014). Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal*, *8*(6), 1198–1209. https://doi.org/10.1038/ISMEJ.2013.227

Ribeiro, I., Girão, M., Alexandrino, D. A. M., Ribeiro, T., Santos, C., Pereira, F., Mucha, A. P., Urbatzka, R., Leão, P. N., & Carvalho, M. F. (2020). Diversity and bioactive potential of Actinobacteria isolated from a coastal marine sediment in northern Portugal. *Microorganisms*, *8*(11), 1691. https://doi.org/10.3390/microorganisms8111691

Richter, A. A., Mais, C. N., Czech, L., Geyer, K., Hoeppner, A., Smits, S. H. J., Erb, T. J., Bange, G., & Bremer, E. (2019). Biosynthesis of the stress-protectant and chemical chaperon ectoine: Biochemistry of the transaminase EctB. *Frontiers in*

Microbiology, *10*, 2811. https://doi.org/10.3389/FMICB.2019.02811/BIBTEX

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., & Peplies, J. (2016). JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*, *32*(6), 929–931. https://doi.org/10.1093/BIOINFORMATICS/BTV681

Rinehart, K. L., & Renfroe, H. B. (1961). The structure of nybomycin. *Journal of the American Chemical Society*, *83*(17), 3729–3731. https://doi.org/10.1021/JA01478A049/ASSET/JA01478A049.FP.PNG_V03

Riyanti, Balansa, W., Liu, Y., Sharma, A., Mihajlovic, S., Hartwig, C., Leis, B., Rieuwpassa, F. J., Ijong, F. G., Wägele, H., König, G. M., & Schäberle, T. F. (2020). Selection of sponge-associated bacteria with high potential for the production of antibacterial compounds. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-76256-2

Rodriguez-Concepcion, M., Avalos, J., Bonet, M. L., Boronat, A., Gomez-Gomez, L., Hornero-Mendez, D., Limon, M. C., Meléndez-Martínez, A. J., Olmedilla-Alonso, B., Palou, A., Ribot, J., Rodrigo, M. J., Zacarias, L., & Zhu, C. (2018). A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health. *Progress in Lipid Research*, *70*, 62–93. https://doi.org/10.1016/J.PLIPRES.2018.04.004

Romano, S., Jackson, S. A., Patry, S., & Dobson, A. D. W. (2018). Extending the "one strain many compounds" (OSMAC) principle to marine microorganisms. *Marine Drugs*, *16*(7). https://doi.org/10.3390/MD16070244

Rossi, F., Dellaglio, F., & Torriani, S. (2006). Evaluation of recA gene as a phylogenetic marker in the classification of dairy propionibacteria. *Systematic and Applied Microbiology*, *29*(6), 463–469. https://doi.org/10.1016/J.SYAPM.2006.01.001

Rostami, H., Hamedi, H., & Yolmeh, M. (2016). Some biological activities of pigments extracted from *Micrococcus roseus* (PTCC 1411) and *Rhodotorula glutinis* (PTCC 5257). *International Journal of Immunopathology and Pharmacology*, *29*(4), 684–695. https://doi.org/10.1177/0394632016673846

Roszak, D. B., & Colwell, R. R. (1987). Survival strategies of bacteria in the natural environment. *Microbiological Reviews*, *51*(3), 365–379. https://doi.org/10.1128/mr.51.3.365-379.1987

Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., & Kohlbacher, O. (2011). NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, *39*(Web Server issue), W362. https://doi.org/10.1093/NAR/GKR323

Rutledge, P. J., & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*, *13*(8), 509–523. Nat Rev Microbiol. https://doi.org/10.1038/nrmicro3496

Rützler, K. (1985). Associations between Caribbean sponges and photosynthetic organisms. In K. Rützler (Ed.), *New perspectives in sponge biology* (pp. 455-466). Smithsonian Institution Press.

Sabarathnam, B., Manilal, A., Sujith, S., Kiran, G. S., Selvin, J., Thomas, A., & Ravji, R. (2010). Role of sponge associated actinomycetes in the marine phosphorus biogeochemical cycles. *American-Eurasian Journal of Agricultural & Environmental Sciences*, *8*(3), 253–256. http://www.idosi.org/aejaes/jaes8(3)/4.pdf

Sakemi, S., Ichiba, T., Kohmoto, S., Saucy, G., & Higa, T. (1988). Isolation and structure elucidation of onnamide A, a new bioactive metabolite of a marine sponge, *Theonella* sp. *Journal of the American Chemical Society*, *110*(14), 4851–4853. https://doi.org/10.1021/JA00222A068/SUPPL_FILE/JA00222A068_SI_001.PDF

Sakula, A. (1982). Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax*, *37*(4), 246–251. https://doi.org/10.1136/THX.37.4.246

Salo, W. L., Aufderheide, A. C., Buikstra, J., & Holcomb, T. A. (1994). Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(6), 2091. https://doi.org/10.1073/PNAS.91.6.2091

Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marcxais, G., Pop, M., & Yorke, J. A. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, *22*(3), 557. https://doi.org/10.1101/GR.131383.111

Sareen, S., Joseph, L., & Mathew, G. (2012). Improvement in solubility of poor water-soluble drugs by solid dispersion. *International Journal of Pharmaceutical Investigation*, *2*(1), 12. https://doi.org/10.4103/2230-973x.96921

Schatz, A., Bugie, E., Waksman, S. A. (1944). Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine*, *55*(1), 66-69. doi:10.3181/00379727-55-14461

Scherlach, K., & Hertweck, C. (2021). Mining and unearthing hidden biosynthetic potential. *Nature Communications*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-24133-5

Schinke, C., Martins, T., Queiroz, S. C. N., Melo, I. S., & Reyes, F. G. R. (2017). Antibacterial compounds from marine bacteria, 2010-2015. *Journal of Natural Products*, *80*(4), 1215–1228. https://doi.org/10.1021/acs.jnatprod.6b00235

Schmidt, O. (1870). *Grundzüge einer Spongien-Fauna des atlantischen Gebietes* (pp. 44-45). Wilhelm Engelmann.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., Perez, T., Rodrigo, A., Schupp, P. J., Vacelet, J., Webster, N., Hentschel, U., & Taylor, M. W. (2012). Assessing the complex sponge microbiota: Core, variable and species-specific bacterial communities in marine sponges. *ISME Journal*, *6*(3), 564–576. https://doi.org/10.1038/ismej.2011.116

Schmitt, S., Weisz, J. B., Lindquist, N., & Hentschel, U. (2007). Vertical transmission of a phylogenetically complex microbial consortium in the viviparous sponge *Ircinia felix*. *Applied and Environmental Microbiology*, *73*(7), 2067. https://doi.org/10.1128/AEM.01944-06

Schneider, Y. K. (2021). Bacterial natural product drug discovery for new antibiotics: Strategies for tackling the problem of antibiotic resistance by efficient bioprospecting. *Antibiotics*, *10*(7), 842. https://doi.org/10.3390/antibiotics10070842

Scholz-Schroeder, B. K., Soule, J. D., & Gross, D. C. (2003). The sypA, sypB, and sypC synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. syringae B301D. *MPMI*, *16*(4), 271–280.

Schorn, M. A., Alanjary, M. M., Aguinaldo, K., Korobeynikov, A., Podell, S., Patin, N., Lincecum, T., Jensen, P. R., Ziemert, N., & Moore, B. S. (2016). Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology (Reading, England)*, *162*(12), 2075. https://doi.org/10.1099/MIC.0.000386

Seemann, T. (2020, March 13). *Shovill*. GitHub. https://github.com/tseemann/shovill

Sekurova, O. N., Schneider, O., & Zotchev, S. B. (2019). Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microbial Biotechnology*, *12*(5), 828–844. https://doi.org/10.1111/1751-7915.13398

Sengupta, S., Chattopadhyay, M. K., & Grossart, H.-P. (2013). The multifaceted roles

of antibiotics and antibiotic resistance in nature. *Frontiers in Microbiology*, *4*, 47. https://doi.org/10.3389/FMICB.2013.00047

Shanthi Kumari, K., Shivakrishna, P., & Ganduri, V. S. R. (2020). Wound healing activities of the bioactive compounds from *Micrococcus* sp. OUS9 isolated from marine water. *Saudi Journal of Biological Sciences*, *27*(9), 2398–2402. https://doi.org/10.1016/J.SJBS.2020.05.007

Shapiro, J. A., & von Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biological Reviews of the Cambridge Philosophical Society*, *80*(2), 227–250. https://doi.org/10.1017/S1464793104006657

Sharma, S. C. D., Shovon, M. S., Jahan, M. G. S., Asaduzzaman, A. K. M., Khatun, B., Yeasmin, T., & Roy, N. (2012). Antibiotic sensitivity and antibacterial activity of *Micrococcus* sp SCS1. *Research & Reviews in BioSciences*, 6(10), 304–310.

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, *7*(2), 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6

Sherman, D. H., Malpartida, F., Bibb, M. J., Kieser, H. M., Bibb, M. J., & Hopwood, D. A. (1989). Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tü22. *The EMBO Journal*, *8*(9), 2717. https://doi.org/10.1002/j.1460-2075.1989.tb08413.x

Shi, S., Cui, L., Zhang, K., Zeng, Q., Li, Q., Ma, L., Long, L., & Tian, X. (2022). *Streptomyces marincola* sp. nov., a novel marine actinomycete, and its biosynthetic potential of bioactive natural products. *Frontiers in microbiology*, *13*, 860308. https://doi.org/10.3389/fmicb.2022.860308

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–1123. https://doi.org/10.1101/gr.089532.108

Singh, S. B. (2014). Confronting the challenges of discovery of novel antibacterial agents. *Bioorganic & Medicinal Chemistry Letters*, *24*(16), 3683–3689. https://doi.org/10.1016/J.BMCL.2014.06.053

Skinnider, M. A., Merwin, N. J., Johnston, C. W., & Magarvey, N. A. (2017). PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research*, *45*, 49–54. https://doi.org/10.1093/nar/gkx320

Smits, T. H. M. (2019). The importance of genome sequence quality to microbial comparative genomics. *BMC Genomics*, *20*(1). https://doi.org/10.1186/S12864-

019-6014-5

Sobin, B., & Stahly, G. L. (1942). The isolation and absorption spectrum maxima of bacterial carotenoid pigments. *Journal of Bacteriology*, *44*(3), 265. https://doi.org/10.1128/jb.44.3.265-276.1942

Soldatou, S., Eldjárn, G. H., Ramsay, A., van der Hooft, J. J. J., Hughes, A. H., Rogers, S., & Duncan, K. R. (2021). Comparative metabologenomics analysis of polar actinomycetes. *Marine Drugs*, *19*(2), 103. https://doi.org/10.3390/MD19020103

Song, W. M., Zhao, J. Y., Zhang, Q. Y., Liu, S. Q., Zhu, X. H., An, Q. Q., Xu, T. T., Li, S. J., Liu, J. Y., Tao, N. N., Liu, Y., Li, Y. F., & Li, H. C. (2021). COVID-19 and tuberculosis coinfection: An overview of case reports/case series and meta-analysis. *Frontiers in Medicine*, *8*, 657006. https://doi.org/10.3389/fmed.2021.657006

Stachelhaus, T., Mootz, H. D., & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry and Biology*, *6*(8), 493–505. https://doi.org/10.1016/S1074-5521(99)80082-9

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology*, *39*, 321–346. https://doi.org/10.1146/annurev.mi.39.100185.001541

Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution*, *37*(4), 1237–1239. https://doi.org/10.1093/MOLBEV/MSZ312

Stierle, A. C., Cardellina, J. H., & Singleton, F. L. (1988). A marine *Micrococcus* produces metabolites ascribed to the sponge *Tedania ignis*. *Experientia*, *44*(11–12), 1021. https://doi.org/10.1007/BF01939910

Su, T.L. (1948). Micrococcin, an antibacterial substance formed by a strain of *Micrococcus*. *British Journal of Experimental Pathology*, 29(5), 473-481.

Subramani, R., & Aalbersberg, W. (2013). Culturable rare actinomycetes: Diversity, isolation and marine natural product discovery. *Applied Microbiology and Biotechnology*, *97*(21), 9291–9321. https://doi.org/10.1007/s00253-013-5229-7

Šudomová, M., Shariati, M. A., Echeverría, J., Berindan-Neagoe, I., Nabavi, S. M., & Hassan, S. T. S. (2019). A microbiological, toxicological, and biochemical study of the effects of fucoxanthin, a marine carotenoid, on *Mycobacterium tuberculosis* and the enzymes implicated in its cell wall: A link between mycobacterial infection and autoimmune diseases. *Marine Drugs*, *17*(11), 641.

https://doi.org/10.3390/MD17110641

Sun, W., Wu, W., Liu, X., Zaleta-Pinet, D. A., & Clark, B. R. (2019). Bioactive compounds isolated from marine-derived microbes in China: 2009–2018. *Marine Drugs*, *17*(6). https://doi.org/10.3390/MD17060339

Sun, W., Zhang, F., He, L., Karthik, L., & Li, Z. (2015). Actinomycetes from the South China Sea sponges: Isolation, diversity, and potential for aromatic polyketides discovery. *Frontiers in Microbiology*, *6*, 1048. https://doi.org/10.3389/fmicb.2015.01048

Surekha, P. Y., P, D., MK, S. J., S, P., & Benjamin, S. (2016). *Micrococcus luteus* strain BAA2, a novel isolate produces carotenoid pigment. *Electronic Journal of Biology*, *12*(1), 83–89. https://ejbio.imedpub.com/micrococcus-luteus-strain-baa2-a-novel-isolate-produces-carotenoid-pigment.php?aid=8403

Suroto, D. A., Kitani, S., Arai, M., Ikeda, H., & Nihira, T. (2018). Characterization of the biosynthetic gene cluster for cryptic phthoxazolin A in *Streptomyces avermitilis*. *PLoS ONE*, *13*(1). https://doi.org/10.1371/JOURNAL.PONE.0190973

Tait, D. R., Hatherill, M., Van Der Meeren, O., Ginsberg, A. M., Van Brakel, E., Salaun, B., Scriba, T. J., Akite, E. J., Ayles, H. M., Bollaerts, A., Demoitié, M. A., Diacon, A., Evans, T. G., Gillard, P., Hellström, E., Innes, J C., Lempicki, M., Malahleha, M., Martinson, N., Mesia Vela, D., Muyoyeta, M., Nduba, V., Pascal, T. G., Tameris, M., Thienemann, F., Wilkinson, R. J., & Roman, F. (2019). Final analysis of a trial of M72/AS01$_E$ vaccine to prevent tuberculosis. *The New England Journal of Medicine*, 381(25), 2429-2439. doi: 10.1056/NEJMoa1909953

Talairach-Vielmas, L. (2011). *Katherine Byrne, Tuberculosis and the Victorian literary imagination*. Cambridge University Press. https://doi.org/10.4000/MIRANDA.5085

Tamaoki, T., Nomoto, H., Takahashi, I., Kato, Y., Morimoto, M., & Tomita, F. (1986). Staurosporine, a potent inhibitor of phospholipid/Ca++dependent protein kinase. *Biochemical and Biophysical Research Communications*, *135*(2), 397–402. https://doi.org/10.1016/0006-291X(86)90008-2

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040023

Tarlachkov, S. V., Starodumova, I. P., Dorofeeva, L. V., Prisyazhnaya, N. V., Roubtsova, T. V., Chizhov, V. N., Nadler, S. A., Subbotin, S. A., & Evtushenko,

L. I. (2021). Draft genome sequences of 28 Actinobacteria of the family Microbacteriaceae associated with nematode-infected plants. *Microbiology Resource Announcements*, *10*(9). https://doi.org/10.1128/MRA.01400-20

Taylor, M. W., Radax, R., Steger, D., & Wagner, M. (2007). Sponge-associated microorganisms: Evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*, *71*(2), 295–347. https://doi.org/10.1128/mmbr.00040-06

Taylor, M. W., Schupp, P. J., Dahllöf, I., Kjelleberg, S., & Steinberg, P. D. (2004). Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environmental Microbiology*, *6*(2), 121–130. https://doi.org/10.1046/j.1462-2920.2003.00545.x

Taylor, M. W., Schupp, P. J., De Nys, R., Kjelleberg, S., & Steinberg, P. D. (2005). Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*. *Environmental Microbiology*, *7*(3), 419–433. https://doi.org/10.1111/j.1462-2920.2004.00711.x

TB Alliance. (2019, June 6). *Pretomanid and BPaL regimen for treatment of highly resistant tuberculosis*. [PowerPoint slides]. Antimicrobial Drugs Advisory Committee. https://www.fda.gov/media/128001/download

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J. R., Easson, C., Astudillo-García, C., Olson, J. B., Erwin, P. M., López-Legentil, S., Luter, H., Chaves-Fonnegra, A., Costa, R., Schupp, P. J., Steindler, L., Erpenbeck, D., Gilbert, J., Knight, R., Ackermann, G., Victor Lopez, J., … Webster, N. S. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, *7*, 1-12. https://doi.org/10.1038/ncomms11870

Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., Heidelberg, K. B., Egan, S., Steinberg, P. D., & Kjelleberg, S. (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME Journal*, *4*(12), 1557–1567. https://doi.org/10.1038/ismej.2010.74

Tiwari, K., & Gupta, R. K. (2012). Rare actinomycetes: A potential storehouse for novel antibiotics. *Critical Reviews in Biotechnology*, *32*(2), 108–132. https://doi.org/10.3109/07388551.2011.562482

Tomm, H. A., Ucciferri, L., & Ross, A. C. (2019). Advances in microbial culturing conditions to activate silent biosynthetic gene clusters for novel metabolite production. *Journal of Industrial Microbiology & Biotechnology*, *46*(9), 1381–1400. https://doi.org/10.1007/S10295-019-02198-Y

Tritt, A., Eisen, J. A., Facciotti, M. T., & Darling, A. E. (2012). An integrated

pipeline for de novo assembly of microbial genomes. *PLoS ONE, 7*(9). https://doi.org/10.1371/JOURNAL.PONE.0042304

Tsakos, M., Clement, L. L., Schaffert, E. S., Olsen, F. N., Rupiani, S., Djurhuus, R., Yu, W., Jacobsen, K. M., Villadsen, N. L., & Poulsen, T. B. (2016). Total synthesis and biological evaluation of rakicidin A and discovery of a simplified bioactive analogue. *Angewandte Chemie International Edition*, *55*(3), 1030–1035. https://doi.org/10.1002/ANIE.201509926

Tsuzuki, K., Yokozuka, T., Murata, M., Tanaka, H., & Ōmura, S. (1989). Synthesis and biological activity of analogues of diazaquinomycin A, a new thymidylate synthase inhibitor. *The Journal of Antibiotics*, *42*(5), 727–737. https://doi.org/10.7164/antibiotics.42.727

U.S. Food and Drug Administration. (2019, August 14). *FDA approves new drug for treatment-resistant forms of tuberculosis that affects the lungs* [Press release.] https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-treatment-resistant-forms-tuberculosis-affects-lungs

Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R., & Moore, B. S. (2007). Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(25), 10376–10381. https://doi.org/10.1073/pnas.0700962104

Umadevi, K., & Krishnaveni, M. (2013). Antibacterial activity of pigment produced from *Micrococcus luteus* KF532949. *International Journal of Chemical and Analytical Science*, *4*(3), 149–152. https://doi.org/10.1016/J.IJCAS.2013.08.008

Ungers, G. E., & Cooney, J. J. (1968). Isolation and characterization of carotenoid pigments of *Micrococcus roseus*. *Journal of Bacteriology*, *96*(1), 234–241. https://doi.org/10.1128/JB.96.1.234-241.1968

Unson, M. D., & Faulkner, D. J. (1993). Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia*, *49*(4), 349–353. https://doi.org/10.1007/BF01923420

Unson, M. D., Holland, N. D., & Faulkner, D. J. (1994). A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Marine Biology*, *119*(1), 1–11. https://doi.org/10.1007/BF00350100

Ushasri, R., & Gods Will Shalomi, C. (2015). A study on in vitro anti breastcancer activity of crude ethanol and acetone pigment extracts of *Micrococcus luteus* by MTT assay and analysis of pigment by thin layer chromatography. *International Journal of Pharmacy and Biological Science*, *5*(1), 59–65.

www.ijpbs.comorwww.ijpbsonline.com

Vacelet, J. (1975). Étude en microscopie électronique de l'association entre bactéries et spongiaires du genre *Verongia* (Dictyoceratida). *Journal de Microscopie et de Biologie Cellulaire*, 23, 271–288.

Vacelet, J., & Donadey, C. (1977). Electron microscope study of the association between some sponges and bacteria. *Journal of Experimental Marine Biology and Ecology*, *30*(3), 301–314. https://doi.org/10.1016/0022-0981(77)90038-7

van der Meij, A., Worsley, S. F., Hutchings, M. I., & van Wezel, G. P. (2017). Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiology Reviews*, *41*(3), 392-416. https://doi.org/10.1093/femsre/fux005

van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J., & Kuipers, O. P. (2018). BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, *46*(W1), W278–W281. https://doi.org/10.1093/nar/gky383

Van Lanen, S. G., & Shen, B. (2006). Microbial genomics for the improvement of natural product discovery. *Current Opinion in Microbiology*, *9*(3), 252–260. https://doi.org/10.1016/j.mib.2006.04.002

van Soest, R. W. M. (1980). Marine sponges from Curaçao and other Caribbean localities Part II. Haplosclerida. *Studies on the Fauna of Curaçao and Other Caribbean Islands*, *62*(1), 1–173.

Ventola, C. (2015). The antibiotic resistance crisis: Causes and threats. *P & T Journal*, *40*(4), 277–283. https://doi.org/Article

Vicente, J., Stewart, A., Song, B., Hill, R. T., & Wright, J. L. (2013). Biodiversity of actinomycetes associated with Caribbean sponges and their potential for natural product discovery. *Marine Biotechnology*, *15*(4), 413–424. https://doi.org/10.1007/s10126-013-9493-4

Viera, I., Pérez-Gálvez, A., & Roca, M. (2018). Bioaccessibility of marine carotenoids. *Marine Drugs*, *16*(10). https://doi.org/10.3390/MD16100397

Vilchèze, C., & Jacobs, W. J. R. (2014). Resistance to isoniazid and ethionamide in *Mycobacterium tuberculosis* : Genes, mutations, and causalities. *Microbiology Spectrum*, *2*(4), MGM2 -0014-2013. https://doi.org/10.1128/microbiolspec.mgm2-0014-2013

Vior, N. M., Lacret, R., Chandra, G., Dorai-Raj, S., Trick, M., & Truman, A. W. (2018). Discovery and biosynthesis of the antibiotic bicyclomycin in distantly related bacterial classes. *Applied and Environmental Microbiology*, *84*(9).

https://doi.org/10.1128/AEM.02828-17

Vogel, G. (2008). The inner lives of sponges. *Science*, *320*(5879), 1028–1030. https://doi.org/10.1126/SCIENCE.320.5879.1028/ASSET/F9C1948C-37F7-4FA1-B04E-FA45FE3845C2/ASSETS/SCIENCE.320.5879.1028.FP.PNG

Wagman, G. H., & Weinstein, M. J. (1980). Antibiotic from *Micromonospora*. *Annual Review of Microbiology*, *34*, 537–557. https://doi.org/10.1146/ANNUREV.MI.34.100180.002541

Waksman, S. A., Lechevalier, H. A., & Schaffner, C. P. (1965). Candicidin and other polyenic antifungal antibiotics. *Bulletin of the World Health Organization*, *33*(2), 219–226. https://pubmed.ncbi.nlm.nih.gov/5320588/

Waksman, Selman A., & Woodruff, H. B. (1940). The soil as a source of microorganisms antagonistic to disease-producing bacteria. *Journal of Bacteriology*, *40*(4), 581. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC374661/

Waksman, Selman A., & Woodruff, H. B. (1942). Streptothricin, a new selective bacteriostatic and bactericidal agent, particularly active against gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine*, *49*(2), 207–210. https://doi.org/10.3181/00379727-49-13515

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., … Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with GNPS. *Nature Biotechnology*, *34*(8), 828. https://doi.org/10.1038/NBT.3597

Wang, Z., Feng, J., Li, X., & Zhang, J. (2021). Whole-genome sequencing of *Micrococcus luteus* MT1691313, isolated from the Mariana Trench. *Microbiology Resource Announcements*, *10*(23). https://doi.org/10.1128/MRA.00369-21

Ward, A. C., & Allenby, N. E. (2018). Genome mining for the search and discovery of bioactive compounds: the *Streptomyces* paradigm. *FEMS Microbiology Letters*, *365*(24). https://doi.org/10.1093/FEMSLE/FNY240

Watve, M. G., Shejval, V., Sonawane, C., Rahalkar, M., Matapurkar, M., Shouche, Y., Patole, M., Phadnis, N., Champhekar, A., Damle, K., Karandikar, S., Kshirsagar, V., Jog, M. (2000). The 'K' selected oligophilic bacteria: A key to uncultured diversity? *Current Science*, 78(12).

Watve, M. G., Tickoo, R., Jog, M. M., & Bhole, B. D. (2001). How many antibiotics

are produced by the genus *Streptomyces*? *Archives of Microbiology*, *176*(5), 386–390. https://doi.org/10.1007/s002030100345

Webster, N. S., & Hill, R. T. (2001). The culturable microbial community of the Great Barrier Reef sponge *Rhopaloeides odorabile* is dominated by an α-Proteobacterium. *Marine Biology*, *138*(4), 843–851. https://doi.org/10.1007/S002270000503

Webster, N. S., Negri, A. P., Munro, M. M. H. G., & Battershill, C. N. (2004). Diverse microbial communities inhabit Antarctic sponges. *Environmental Microbiology*, *6*(3), 288–300. https://doi.org/10.1111/j.1462-2920.2004.00570.x

Webster, N. S., Taylor, M. W., Behnam, F., Lücker, S., Rattei, T., Whalan, S., Horn, M., & Wagner, M. (2010). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology*, *12*(8), 2070–2082. https://doi.org/10.1111/j.1462-2920.2009.02065.x

Weigel, B. L., & Erwin, P. M. (2017). Effects of reciprocal transplantation on the microbiome and putative nitrogen cycling functions of the intertidal sponge, *Hymeniacidon heliophila*. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/srep43247

Weinstein, M. J., Luedemann G. M., Oden, E. M., Wagman, G. H., Rosselet, J. P., Marquez, J. A., Coniglio, C. T., Charney, W,, Herzog, H. L., & Black, J. (1963). Gentamicin, a new antibiotic complex from *Micromonospora*. *Journal of Medicinal Chemistry*, 6(4), 463-464. doi: 10.1021/jm00340a034

Wenzel, S. C., & Müller, R. (2005). Formation of novel secondary metabolites by bacterial multimodular assembly lines: Deviations from textbook biosynthetic logic. *Current Opinion in Chemical Biology*, *9*(5), 447–458. https://doi.org/10.1016/J.CBPA.2005.08.001

Wieser, M., Denner, E. B. M., Kämpfer, P., Schumann, P., Tindall, B., Steiner, U., Vybiral, D., Lubitz, W., Maszenan, A. M., Patel, B. K. C., Seviour, R. J., Radax, C., & Busse, H. J. (2002). Emended descriptions of the genus *Micrococcus*, *Micrococcus luteus* (Cohn 1872) and *Micrococcus lylae* (Kloos et al. 1974). *International Journal of Systematic and Evolutionary Microbiology*, *52*(2), 629–637. https://doi.org/10.1099/00207713-52-2-629/CITE/REFWORKS

Wilkinson, C. R. (1978). Microbial associations in sponges. II. Numerical analysis of sponge and water bacterial populations. *Marine Biology 1978 49:2*, *49*(2), 169–176. https://doi.org/10.1007/BF00387116

Wilson, B., Aeby, G. S., Work, T. M., & Bourne, D. G. (2012). Bacterial communities associated with healthy and *Acropora* white syndrome-affected

corals from American Samoa. *FEMS Microbiology Ecology*, *80*(2), 509–520. https://doi.org/10.1111/J.1574-6941.2012.01319.X

Wilson, M. C., Gulder, T. A. M., Mahmud, T., & Moore, B. S. (2010). Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the *sare1259*-encoded cytochrome P450. *Journal of the American Chemical Society*, *132*(36), 12757–12765. https://doi.org/10.1021/ja105891a

Winterberg, H. (1898). Zur Methodik der Bakterienzählung. *Zeitschrift Für Hygiene Und Infectionskrankheiten*, *29*(1), 75–93. https://doi.org/10.1007/BF02217377

Wohlleben, W., Mast, Y., Stegmann, E., & Ziemert, N. (2016). Antibiotic drug discovery. *Microbial Biotechnology*, *9*(5), 541–548. https://doi.org/10.1111/1751-7915.12388

Woodruff, H. B. (2014). Selman A. Waksman, winner of the 1952 Nobel Prize for physiology or medicine. *Applied and Environmental Microbiology*, *80*(1), 2. https://doi.org/10.1128/AEM.01143-13

Woodruff, H. B. & McDaniel, L. E. (1958). The antibiotic approach. In S. T. Cohen & R. Rowatt (Eds.), *The strategy of chemistry* (pp. 29-48). Cambridge University Press.

Woon, S.-A., & Fisher, D. (2016). Antimicrobial agents – optimising the ecological balance. *BMC Medicine*, *14*(1), 1–9. https://doi.org/10.1186/S12916-016-0661-Z

World Health Organization. (2013). *Definitions and reporting framework for tuberculosis - 2013 revision: updated December 2014 and January 2020.* World Health Organization. https://apps.who.int/iris/handle/10665/79199

World Health Organization. (2016). *WHO treatment guidelines for drug-resistant tuberculosis, 2016 update. October 2016 revision.* https://apps.who.int/iris/bitstream/handle/10665/250125/9789241549639-eng.pdf

World Health Organization. (2018). *Global tuberculosis report 2018*. World Health Organization. https://apps.who.int/iris/handle/10665/274453

World Health Organization. (2020). *Global tuberculosis report 2020*. World Health Organization. https://apps.who.int/iris/handle/10665/336069

World Health Organization. (2021a). *Global tuberculosis report 2021*. World Health Organization. https://apps.who.int/iris/handle/10665/346387

World Health Organization. (2021b, January 27). *WHO announces updated definitions of extensively drug-resistant tuberculosis* [Press release].

https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis

Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA*, *323*(9), 844–853. https://doi.org/10.1001/JAMA.2020.1166

Wright, G. D. (2017). Opportunities for natural products in 21st century antibiotic discovery. *Natural Product Reports*, *34*(7), 694–701. https://doi.org/10.1039/C7NP00019G

Xu, H. S., Roberts, N., Singleton, F. L., Attwell, R. W., Grimes, D. J., & Colwell, R. R. (1982). Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microbial Ecology*, *8*(4), 313–323. https://doi.org/10.1007/BF02010671

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., & Ikeda, H. (2015). Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(3), 857–862. https://doi.org/10.1073/PNAS.1422108112/-/DCSUPPLEMENTAL

Young, M., Artsatbanov, V., Beller, H. R., Chandra, G., Chater, K. F., Dover, L. G., Goh, E.-B., Kahan, T., Kaprelyants, A. S., Kyrpides, N., Lapidus, A., Lowry, S. R., Lykidis, A., Mahillon, J., Markowitz, V., Mavromatis, K., Mukamolova, G. V, Oren, A., Rokem, J. S., … Greenblatt, C. L. (2010). Genome sequence of the Fleming strain of *Micrococcus luteus*, a simple free-living actinobacterium. *Journal of Bacteriology*, *192*(3), 841–860. https://doi.org/10.1128/JB.01254-09

Yu, Z., Vodanovic-Jankovic, S., Kron, M., Shen, B., & Lett, O. (2012). New WS9326A congeners from *Streptomyces* sp. 9078 inhibiting *Brugia malayi* asparaginyl-tRNA synthetase. *Organic Letters*, *14*(18), 4946–4949. https://doi.org/10.1021/ol302298k

Zerikly, M., & Challis, G. L. (2009). Strategies for the discovery of new natural products by genome mining. *ChemBioChem*, *10*(4), 625–633. https://doi.org/10.1002/cbic.200800389

Zhang, C., Li, X., Yin, L., Liu, C., Zou, H., Wu, Z., & Zhang, Z. (2019a). Analysis of the complete genome sequence of *Brevibacterium frigoritolerans* ZB201705 isolated from drought- and salt-stressed rhizosphere soil of maize. *Annals of Microbiology*, *69*(13), 1489–1496. https://doi.org/10.1007/s13213-019-01532-0

Zhang, F., Jonas, L., Lin, H., & Hill, R. T. (2019b). Microbially mediated nutrient cycles in marine sponges. *FEMS Microbiology Ecology*, *95*(11). https://doi.org/10.1093/femsec/fiz155

Zhang, M. M., Qiao, Y., Ang, E. L., & Zhao, H. (2017). Using natural products for drug discovery: The impact of the genomics era. *Expert Opinion on Drug Discovery*, *12*(5), 475. https://doi.org/10.1080/17460441.2017.1303478

Zhang, Q., Doroghazi, J. R., Zhao, X., Walker, M. C., & van der Donk, W. A. (2015). Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in Actinobacteria. *Applied and Environmental Microbiology*, *81*(13), 4339–4350. https://doi.org/10.1128/AEM.00635-15

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7*(1-2), 203–214. Mary Ann Liebert, Inc. https://doi.org/10.1089/10665270050081478

Zhao, B., Moody, S. C., Hider, R. C., Lei, L., Kelly, S. L., Waterman, M. R., & Lamb, D. C. (2012). Structural analysis of cytochrome P450 105N1 involved in the biosynthesis of the zincophore, coelibactin. *International Journal of Molecular Sciences*, *13*(7), 8500. https://doi.org/10.3390/IJMS13078500

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguiñaga, N., Chavarria, K. L., & Jensen, P. R. (2014). Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proceedings of the National Academy of Sciences*, *111*(12), E1130–E1139. https://doi.org/10.1073/PNAS.1324161111

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., & Jensen, P. R. (2012). The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE*, *7*(3), 34064. https://doi.org/10.1371/JOURNAL.PONE.0034064

Zimin, A. V, Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Zimmerman M. R. (1979). Pulmonary and osseous tuberculosis in an Egyptian mummy. *Bulletin of the New York Academy of Medicine*, *55*(6), 604–608.

ZoBell, C. E. (1946). *Marine microbiology, a monograph on hydrobacteriology*. Chronica Botanica Company. https://doi.org/10.5962/bhl.title.10079

Zwerling, A., Behr, M. A., Verma, A., Brewer, T. F., Menzies, D., & Pai, M. (2011). The BCG world atlas: A database of global BCG vaccination policies and practices. *PLoS Medicine*, *8*(3), e1001012. https://doi.org/10.1371/JOURNAL.PMED.1001012