Genetic Variation in *Janthinobacterium lividum*;
An Examination of the Violacein Operon

By
Cerrise Weiblen
Department of Ecology and Evolutionary Biology, University of Colorado at Boulder

Defense Date March 9, 2016

Thesis Advisor:
Dr. Valerie McKenzie, Department of Ecology and Evolutionary Biology

Defense Committee:
Dr. Valerie McKenzie, Department of Ecology and Evolutionary Biology
Dr. Betsy Forrest, Department of Atmospheric and Oceanic Sciences
Dr. Pieter Johnson, Department of Ecology and Evolutionary Biology

Table of Contents

Index of Tables

Index of Figures

ABSTRACT

Probiotic therapeutics are revolutionizing the way we conceptualize and approach the management of pathogens in our environment, from hospitals and households to remote tropical rainforests. *Janthinobacterium lividum* is an excellent candidate for probiotic use because it produces violacein, a metabolite that destroys or inhibits many types of microorganisms, including bacteria, fungi, protozoans and nematodes, while being nonpathogenic to humans and other vertebrates. Researchers are exploring probiotic treatments based on the antibiotic metabolites produced by live strains of the bacterium in vitro, yet little is known about the broader phylogeny or regional variation of the organism in vivo. Here we examine the variability of the violacein gene coding region in *Janthinobacterium* species across geographic space and different environmental sources. Sixteen regional strains of the bacterium were included in this study, isolated from environmental sources, such as soil and water, and from the microbiome of living organisms, including amphibians. Genomic analysis focused on the five genes responsible for the production of violacein, along with the 16S rRNA gene, a commonly used bacterial marker gene. Phylogenetically, the violacein operon was more informative than the 16S rRNA gene for determining evolutionary relationships between strains. This operon may be the best choice for classifying violacein-producing bacteria at the species level, short of full genome assembly and analysis. This data set revealed no detectable correlation between environmental source and genotype (e.g. amphibian vs soil or water). A regional pattern was observed at the continental scale.

INTRODUCTION

Probiotic research aims to leverage the existing competitive relationships between microbial species, to undermine, weaken, or destroy pathogenic organisms. The fundamental concept behind probiotics is that pathogens do not evolve in a vacuum. They evolve within a community of other microorganisms, and are shaped by natural selection and environmental factors just like every other living organism (Bletz et al. 2013; Matz et al. 2004; Woodhams et al. 2014). Every pathogen we encounter has evolved in a community system with other organisms that are often well matched in terms of environmental adaptation, resource competition, and antagonistic interactions.

Some beneficial bacteria provide protective (probiotic) effects merely because they are so successful under certain conditions that they out-compete pathogenic organisms for resources (resource denial). We see this effect in yogurt: when milk is held at the correct temperature, beneficial bacteria are able to colonize the environment very rapidly and undermine the ability of other organisms to gain a foothold.

Another strategy used by bacteria to out-compete their neighbors is the production of substances that are inhibitory or toxic to other microorganisms. In cases where these bacterial metabolites are not strong enough to fully eliminate a pathogen, they can still be very effective as probiotics, because their overall impact is to weaken the pathogen and make it more susceptible to conventional treatments, medications, or to the natural immune system of the host. Probiotic effectiveness often depends on our ability to encourage the growth of a beneficial bacterium, in order for that organism to defeat a target pathogen, leveraging the naturally competitive relationships that have evolved between the microorganisms over billions of years (Matz et al. 2004; Ramsey et al. 2015).

In spite of our recent successes in the field of probiotics and microbiology, we still have far more questions than answers about the complex relationships between environmental variables, microscopic pathogens, and their natural competitors, which may prove to be our benefactors in the war against disease-causing microorganisms (Harris et al. 2009; Woodhams et al. 2014). One application for probiotic therapeutics under investigation is the mitigation of infectious disease in wildlife populations (Bletz et al. 2013; Harris et al 2009). This represents the frontier of ecosystem management, with the potential to reduce the impact of anthropogenic (caused by humans) pathogens on sensitive wildlife populations.

Some microorganisms, such as *Janthinobacterium lividum* stand out as very well suited to development as probiotics. *J. lividum* is a nonpathogenic bacterium, common in the microbiome of many vertebrate species, including human beings (Ramsey et al 2015). This benign organism produces a powerful microbiological weapon: the metabolite violacein, which is effective against other microorganisms, from bacteria to nematodes (Duran et al. 2012). Researchers in the field of probiotics have begun to enlist live cultures of *J. lividum* as foot soldiers in the ongoing war against many types of pathogenic organisms (Harris et al. 2009; Ramsey et al. 2015). Some of the most promising research has involved the use of *J. lividum* to protect amphibians against a fungal pathogen, *Batrachochytrium dendrobatidis,* (referred to as *Bd* hereafter) which causes the disease chytridiomycosis and is linked to amphibian declines worldwide (Bletz et al. 2013; Harris et al. 2009; Woodhams et al. 2014). *Janthinobacterium lividum* has proven effective against this insidious pathogen in vitro (in the lab) (Harris et al. 2009). Researchers are now attempting to develop methods for inoculating and protecting amphibians from *Bd* in vivo (in the field) (Bletz et al. 2013).

Towards a better understanding of the biology of *J. lividum* and the implications to probiotic research surrounding it, we employ a genomics-based approach to investigate genetic differences among strains. The first topic addressed in this study is an exploration of the violacein operon in *J. lividum* to determine which gene(s) are most informative for elucidating the phylogenetic relationships within the species and closely related taxa. Second, we address whether the strains that live symbiotically on amphibian skin are genetically distinct from the strains isolated from environmental substrates. Lastly, we examine whether genetic variability can be linked to the regional origin of *J. lividum* strains.

BACKGROUND

Janthinobacterium lividum is a ubiquitous bacterium which thrives in a wide range of environmental conditions, from the ballast tanks of North American freighters, to the mountain snows of Japan, and even the glacial soils of Antarctica (Garcia-Echauri et al. 2011; Segawa et al. 2005; Starliper et al. 2015). The purple pigment, violacein, is an important metabolite produced by the bacterium, and has been widely studied for its antibacterial and antifungal properties (Antonio and Creczynski-Pasa 2004; Brucker et al. 2008; Lu et al. 2009; Zhang and Enomoto 2011). Violacein has also led to the use of *J. lividum* for the production of dyes and pigments for a variety of industrial purposes (Lu et al. 2009, Pantanella et al. 2007).

Observations in the lab reveal that violacein production in *J. lividum* can be variable (Pantanella et al. 2007; Pers. Obs.; Schloss et al. 2010). For researchers developing probiotic or antibiotic treatments based on *J. lividum* or its metabolites, these types of differences between strains could be tremendously important. It stands to reason that natural selection in variable environments, especially extreme environments like the soil under a glacier, could lead to regional differences in geographically distant strains of the bacterium, including the organism's ability or propensity to produce violacein.

In *Janthinobacterium lividum*, violacein is a secondary metabolite, resulting from a pathway which requires the precursor molecule, l-Tryptophan (Hoshino et al. 2011). The genes responsible for the production of violacein have been identified and well-studied in *J. lividum* and several other bacteria, such as *Chromobacterium violaceum* (Antonio and Creczynski-Pasa 2004; Sanchez et al. 2006; Schloss et al. 2010). Methods have even been developed for mass-producing the metabolite in the lab, using genetically modified *Escherichia coli* (Hoshino 2011; Sanchez et al. 2006; Zhang and Enomoto 2011).

Other bacteria produce violacein, but are less suited to use as probiotics. For example, the bacterium *Chromobacterium violaceum* is hypothesized as the source of the violacein genes, which were acquired by organisms in the genus *Janthinobacterium* through horizontal gene transfer (sharing of genetic material between otherwise unrelated microorganisms) (Gillis and De Ley 2006). This similar-looking purple bacterium displays the same antibacterial and antifungal properties as *J. lividum,* but is a pathogen which is highly toxic to vertebrates, including humans, therefore it is unsuitable as a probiotic.

One study has examined the genetic variation within a metapopulation of *J. lividum* in a single geographic region (Saeger and Hale 1993). Saeger and Hale established that *J. lividum* does experience some variation, even across a relatively short geographic distance along a 3 km section of a single creek (1993). Considering the results of this narrower study, we expect greater variation across larger geographic space. Although research into the global population dynamics of *J. lividum* has not yet been published, detailed descriptions of the genes responsible for violacein production which have been published can be leveraged in the context of population dynamics, to determine whether significant regional variations in violacein production may occur.

We address an open question within the research community regarding the efficacy of different strains of *J. lividum* as probiotic agents for the treatment of a variety of pathogens by examining the regional differences between genomes of several strains of *J. lividum*, especially in the genes responsible for violacein production. If differences exist between regional strains, particularly in the amount, timing, or conditions required to produce violacein, researchers working with isolates from the field would benefit from such knowledge.

MATERIALS AND METHODS

In order to gather genomic data to compare different strains of *Janthinobacterium* across geographic regions and from different sources, we used: 1) live cultures that we isolated in the McKenzie lab at the University of Colorado, 2) live cultures isolated by collaborators from 4 different institutions, and 3) published sequences provided by other researchers (e.g., open access sources such as NCBI, the National Center for Biotechnology Information.) For the live cultures, we describe the methods used to obtain sequence information below. Ultimately we used two different gene regions for comparing strains, the 16S rRNA ribosomal gene, and the five gene operon that codes for the metabolite violacein.

In total, sixteen strains of *Janthinobacterium spp.* were examined in this study (Table 1). Five active strains (CCOS 423, DE0145, PAZ19G, PAZ21C, BTP022) and one inactive strain (VA0829) of *J. lividum* were processed in the McKenzie lab at CU for genomic sequencing. The sequences of 10 additional strains of *Janthinobacterium spp.* were obtained from the nucleotide database of the National Center for Biotechnology Information (NCBI Strain, Accession number: Marseille, NC_009659; CG3, NZ_APFF01000000; CG23-2, CYSS00000000; HH01, AMWD00000000; KBS0711, LBCO00000000; MTR, NZ_JRRH00000000; NBRC 102515, HG322949; PAMC 25724, NZ_AHHB00000000; RA13, NZ_JQNP01000001; RIT308, NZ_JFYR00000000).

Strain	Species	GenBank ID / NCBI reference	Origin	Region where collected	Color in culture	Vio genes
NBRC 102515 / DSM 9628	*J. agaricidamnosum*	HG322949	Mushroom (*Agaricus bisporus*)	United Kingdom	Buff	Yes >90%
BTP022	*J. lividum*	Valerie McKenzie	Boreal toad (*Anaxyrus boreas, Bufo boreas*) (wild)	Colorado, USA	Purple	Yes >90%
CCOS423	*J. lividum*	Doug Woodhams	Egg clutch of Midwife toad (*Alytes obstetricans*) (wild)	Basel, Switzerland	Purple	Yes >90%
DE0145	*J. lividum*	Molly Bletz	Fire salamander (*Salamandra salamandra*) (wild)	Sölling, Germany	Purple	Yes >90%
MTR	*J. lividum*	NZ_JRRH00000000	Soil (mud)	Cajon del Maipo, Chile	Purple	Yes >90%
PAMC 25724	*J. lividum*	NZ_AHHB00000000	Cryoconite (biofilm)	Innsbruck, Alps, Austria	Purple	None 00%
PAZ19G	*J. lividum*	Matthew Becker	Panamanian golden frog (*Atelopus zeteki*) (captive bred F1)	Maryland, USA Sora, Panama	Purple	Yes >90%
PAZ21C	*J. lividum*	Matthew Becker	Panamanian golden frog (*Atelopus zeteki*) (captive bred F1)	Maryland, USA Sora, Panama	Off-white	None 00%
RIT308	*J. lividum*	NZ_JFYR00000000	Shrub willow (Fabius cultivar: *Salix viminalis* x *S. miyabeana*) (endophyte)	New York, USA	Not recorded	Yes >90%
VA0829	*J. lividum*	Reid Harris	Four-toed salamander (*Hemidactylium scutatum*) (wild)	Virginia, USA	Purple	Yes >90%
CG23-2	*Janthinobacterium sp.*	CYSS00000000	Water (supraglacial stream)	Antarctic dry valleys, Antarctica	Purple	Yes >90%
CG3	*Janthinobacterium sp.*	NZ_APFF01000000	Water (supraglacial stream)	Cotton Glacier, Antarctica	Unpigmented	None

Strain	Species	GenBank ID / NCBI reference	Origin	Region where collected	Color in culture	Vio genes
HH01	*Janthinobacterium sp.*	AMWD00000000	Water (watering can)	Hamburg, Germany	Purple	Yes >90%
KBS0711	*Janthinobacterium sp.*	LBCO00000000	Soil (never-plowed)	Michigan, USA	Purple	Yes >90%
Marseille	*Janthinobacterium sp.*	NC_009659	Water (ultra-purified hemodialysis)	Marseille, France	Not recorded	No ~50%
RA13	*Janthinobacterium sp.*	NZ_JQNP01000001	Soil (lake sediment)	Washington, USA	Purple	Yes >90%

Table 1: Bacteria examined in this study include 16 strains of Janthinobacterium spp.

Field collected *Janthinobacterium lividum* from Colorado

We conducted field campaigns in May-July 2015 aimed at isolating a strain of *J. lividum* from the native Colorado boreal toad (*Anaxyrus boreas*). We were operating in accordance with a Colorado Parks and Wildlife Research Permit, an approved Institutional Animal Care and Use Committee (IACUC) protocol (#1505.04), and an approved Institutional Biosafety Committee (IBC) protocol (#BA15-EBIO-McK-01). We targeted populations in Chaffee County near the town of Buena Vista, Colorado. Live boreal toads were sampled from three different breeding sites consisting of relatively undisturbed alpine wetland systems, located above 2,900 meters elevation in the Rocky Mountains of Colorado. Toads were found in or near active beaver ponds within 50 m of lingering snow drifts. Ponds were surrounded by a heavy growth of macrophytes (water-loving plants), spreading into rough and rocky terrain with fast-moving streams and hillside stands of live aspen and evergreens interspersed with fallen aspen and pine logs.

We sampled toads by gently capturing them while wearing nitrile gloves, transporting them individually in single-use Ziploc bags to a dry location next to the wetland for sampling.

Each toad was carefully measured, rinsed with sterile water to remove transient microbes and soil (McKenzie et al. 2012), then swabbed with a sterile cotton-tipped swab and carried back to the exact location of their origin for release. All toads handled in this manner were seen to actively move into the undergrowth under their own power and with no apparent detrimental effects. Samples from dorsal and ventral skin surfaces were collected and immediately used to inoculate 100mm x 20mm culture plates filled with sterile R2A agar (Difco^{TM}; Becton, Dickinson and Company, Franklin Lakes, New Jersey). Culture plates were sealed with parafilm (Bemis North America, Neenah, Wisconsin), stored in a cooler with ice (ambient temp. ~4˚ C), and transported from the field site to the laboratory within 48 hours.

Bacterial culturing at CU

In the McKenzie lab at CU Boulder, I created subcultures daily from any bacterial colonies that grew in the field-collected culture plates using standard culturing methods on R2A media to increase the diversity of culturable isolates (Reasoner and Geldreich 1985). Purple colonies were identified on day four in one serial plate and one field-collected plate. One of these colonies was eventually isolated and identified as *Janthinobacterium lividum* by PCR analysis using the *J. lividum*-specific primers JlivF (5'-TACCACGAATTGCTGTGCCAGTTG-3') and JlivR (5'-ACACGCTCCAGGTATACGTCTTCA-3'), following the methods established by Harris et al. (2009). Both positive and negative control samples were included in the analysis to ensure accuracy. A positive control was used to confirm the analysis was effective. (Positive results means it worked.) A negative control was used to check for cross-contamination during processing. (A negative result means no cross-contamination occurred.)

This culture, from boreal toad number 22 was perpetuated and included in the study as *J. lividum* strain BTP022.

While the live cultures from the McKenzie lab and other labs were under my care, I developed procedures to prevent cross-contamination of the individual regional strains, including frequent glove changes, disinfection of surfaces and tools with ethanol (destroys live cells) and bleach (destroys DNA), separate and dedicated secondary storage containers for each strain, and the meticulous use of a Labconco, Purifier Logic+, Class II biosafety cabinet for all culturing and sampling activities (www.labconco.com). All active cultures of the bacterial isolates were perpetuated on R2A agar in 100mm x 20mm plates at 4º C and also stored in a standardized glycerol-based storage solution suggested by Jenifer Walke of Virginia Tech (pers. comm.) at -20° C and -70° C to be utilized in future research. All strains except the Colorado native strain were shipped as active culture plates in refrigerated shipping containers from their origins to the McKenzie Lab at the University of Colorado, Boulder, where they were immediately transferred to a refrigerator (4º C). The Virginia strain was obtained from an inactive culture plate, left over from previous research at the McKenzie Lab and stored at 4º C since 2012. The Colorado strain was isolated in-house during the summer of 2015, as described above.

DNA isolation and sequencing

DNA samples were collected by directly swabbing isolates growing on R2A culture media with sterile swabs. DNA extractions were performed in the McKenzie Lab, using the UltraClean Microbial DNA Isolation Kit from Mo Bio Laboratories Inc. (Carlsbad, CA) according to the manufacturer's instructions (www.mobio.com). Approximately 50 µL of DNA solution was prepared and quantified in the Schmidt Lab at CU Boulder using the Qubit Fluorometer (Thermofisher scientific, Waltham, MA) under the standard protocol (www.invitrogen.com/qubit). Quantified DNA samples at concentrations of at least 3,000 ng/mL were sent to the Genetic Sequencing and Analysis Facility of the Institute for Cellular and Molecular Biology, at the University of Texas, Austin, where they were sequenced using an Illumina NextSeq 500 instrument (Illumina, Inc., San Diego, CA) (www.icmb.utexas.edu/research/core-facilities/gsaf).

Alignments

The 16S rRNA genes of 16 regional strains and the violacein-producing gene regions (operons) of 14 strains of *Janthinobacterium spp.* were aligned with the Geneious 9.0.5 software package (Multiple Alignment, using Geneious Alignment algorithm with default settings. *Alignment type: Global alignment with free end gaps, Cost Matrix: 65% similarity (5.0/-4.0), Gap open penalty: 12, Gap extension penalty: 3, Refinement iterations: 2*) (www.geneious.com, Kearse et al. 2012a). The genes and operons were rarefied by trimming off overhanging bases at the beginning and/or end of the aligned sequences. The Geneious alignment included pairwise relational matrices for both gene regions of interest (Appendix 2).

Analyses

Overall, the analyses for this project involved the comparison of specific genes within the genome of several strains of the *J. lividum* bacterium and phylogenetic analysis to infer evolutionary relationships between the strains based on their genetic differences. After locating and downloading reliable reference sequences for the two gene regions we were interested in from a reputable source in the NCBI database, these reference sequences were used to assemble the DNA sequences of the gene regions for all of the strains we sampled in-house (6 strains). Ten additional strains of the *Janthinobacterium* genus, which contained the two gene regions were available from the NCBI nucleotide and genome databases. In cases where published genomes contained annotations for the 16S rRNA gene and / or the violacein operon, the annotations of the publisher were assumed to be accurate and used as-is. After extracting the genetic information for the genes we were interested in from the NCBI data, the individual genes of every sequence (lab and NCBI) were aligned and rarefied so they could be compared equitably. Finally, phylogenetic trees were created based on the relative similarity of the DNA sequences, to show the evolutionary relationships between the strains.

Preprocessing of the raw sequence data

Although the Geneious software package was used for assembling the genes, a significant amount of pre-processing was required in order to prepare the data for use in Geneious. To this end, I researched and wrote a “script” or series of instructions to be used in a command-line interface in order to properly prepare the raw sequence data. The development and rationale of the script is described in detail in Appendix 1. Generally, I downsampled the raw data from more than 9 million raw reads for each sequenced strain (6 strains) to a high-quality subset of 3 million reads per strain, in order to provide good depth of coverage (50) for an expected genome length of 6 million base pairs, while maximizing processing speed. (Coverage is conceptually defined in the Results section. Specific calculations used to determine coverage for this study are described in Appendix 1.) From this smaller dataset, I located and extracted only those high quality sequences that were relevant to the two specific genes we were interested in. This created an even smaller dataset, which was imported into Geneious 9.0.5 for reference-based assembly and alignment.

Violacein gene identification

In *J. lividum*, the five violacein genes occur together, in a single operon (Sanchez et al. 2006). I treated this contiguous sequence of more than 7,000 nucleotides as a single coding region for all violacein analyses.

The raw DNA sequence data for the six strains sampled and assembled in-house was processed using a software pipeline suggested by Jack Darcy of the Schmidt Lab, at the University of Colorado, Boulder (personal comm.). This pipeline involved the use of BLAST+ and the Bash command language interpreter in Linux Ubuntu version 14.04.3, plus the addition of FastX for certain formatting procedures (Bash 2010, Camacho et al. 2009, FastX 2016). A reference sequence of DNA for the *Janthinobacterium spp.* violacein-producing operon (genes vioA, vioB, vioC, vioD, and vioE) was downloaded from NCBI (Strain RA13, accession NZ_JQNP01000001). I chose this strain as a reference based on its level of completion (high-quality draft) and its origin from a reputable research group: The Joint Genome Institute. The violacein operon sequence from strain RA13 was used to identify similar sequences within the raw DNA sequence data files for each of the six in-house strains using BLAST+ (Camacho et al. 2009).

In addition to five out of the six strains of *J. lividum* sampled in-house, I was able to locate the violacein operon in seven out of ten strains of *Janthinobacterium spp*. held in the NCBI database with accompanying region of origin and source data, which made these strains suitable for inclusion in the study. The violacein reference sequence from strain RA13 was also used to identify matching gene regions within those genomes downloaded from NCBI, which lacked annotations, for consistency with the in-house procedures.

Unpigmented strains

In the lab, strain PAZ21C is morphologically similar in growth pattern on solid R2A media compared with the other (pigmented) strains, but does not exhibit purple coloration, so extensive analysis was needed to determine whether the genes are present in this strain, but non-functional, or whether they are completely lacking.

Full genome assemblies of *Janthinobacterium spp.* have shown the genes vioA-vioE are contiguous along the genome (e.g. strain RA13, accession JQNP01000001), therefore I hypothesized that the presence and location of portions of the sequence could be used to delineate a putative gene region, even if changes have made it nonfunctional as an operon. Raw sequences matching any of the genes could be aligned against the full assembly of the PAZ21C strain, in order to locate the region of the genome most likely to correspond with the violacein genes in other *J. lividum* strains. The sequence could then be extracted from the genome assembly and considered as the putative violacein gene-region for this strain, for purposes of this study. Per this hypothesis, strains showing no evidence of a violacein operon (Tables 1 and 2) were further investigated for each, individual gene, vioA through vioE, using the same BLAST+ search method employed for the full operon and the 16S rRNA gene.

16S rRNA gene identification

16S rRNA analysis was performed in order to place strains Marseille, PAZ21C, CG3 and PAMC 25724, into phylogenetic context with the other strains, since they do not contain violacein genes. I followed the same pre-processing methodology noted above, searching the raw sequences for any reads matching a reference sequence of 16S rRNA extracted from the same reference genome (RA13), for consistency.

Gene assemblies

Downsampled sequence data was imported into the Geneious software package for reference-based assembly (Kearse 2012a). The 16S rRNA and violacein operon sequences from strain RA13 were used as references to assemble the genes for all of the in-house strains and to identify genes in unannotated genomes downloaded from NCBI. In the published genomes used for this study, genes identified by annotations of the publisher were assumed to be accurate and used as-is. Default settings were used for the reference-based assembly, as recommended for the greatest accuracy by the manufacturer (*Mapper: Geneious, Sensitivity: Medium-low Sensitivity / Fast, Fine Tuning: Iterate up to 5 times, Do not trim, Map multiple best matches: Randomly*) (Kearse 2012b).

Phylogenetic inference and visualization

Phylogenetic trees, representing evolutionary relationships between the strains were created with two different software packages: FastTree and PhyML (Guindon et al. 2010, PhyML 2008, Price et al. 2009, 2010). Trees were generated in both programs with the manufacturers' recommended default settings and 1000 replicate bootstrap support where appropriate (PhyML). For consistency, FastTree was used to evaluate the violacein operon, although the program is optimized for 16S rRNA analysis (Price [date unknown]). Similarly, we analyzed the 16S rRNA gene with PhyML to provide consistency, not because it is particularly suited to 16S rRNA analysis. In this way, each gene was investigated with at least one tool well-suited to the specific type of gene region being analyzed.

Wide-view tree topologies were created using the bacterium *Chromobacterium violacein* as an outgroup, followed by finer resolution trees generated using only the *Janthinobacterium* strains being studied. Outgroups in the genera-specific trees were informed by the branch ordering of the wider scale phylogenetic analyses. *Chromobacterium violaceum* was chosen as the outgroup for these analyses because it is on the one hand distantly related to the *Janthinobacterium* genus and it contains all five genes of the violacein operon.

RESULTS

Data received from the sequencing facility were very favorable, with a minimum raw coverage of 116 (Table 2). The calculations used to determine coverage depth are described in Appendix 1: Pre-processing script development. Generally speaking, coverage depth is a calculation of the average number of overlapping reads per nucleotide base location. Higher coverage provides greater accuracy during genome assembly and analysis (Sims et al. 2014). Downsampling and filtering for relevant reads resulted in a minimum 16S rRNA coverage of 397 and a minimum violacein operon coverage of 26, disregarding strain PAZ21C as explained in the methods section above.

ID#	Strain	Raw Reads	*Raw Coverage	Violacein reads	**Violacein coverage	16S reads	***16S coverage
JD27	VA0829	9965099	166	2780	40	6106	407
JD28	DE0145	6981488	116	1912	27	6401	426
JD29	PAZ19G	9605463	160	2353	34	6082	405
JD30	PAZ21C	8174824	136	1	0.01	5951	397
JD31	BTP022	7667388	128	2927	42	6060	404
JD32	CCOS423	7984801	133	1846	26	5973	398

*Table 2: Raw sequence data for in-house strains. *Raw coverage was calculated assuming a genome of 6Mbp. **Violacein coverage calculated assuming an operon length of 7kbp. ***16S coverage calculated assuming a gene length of 1500bp.*

16S rRNA analysis places the strains PAZ21C, CG3 and PAMC 25724, which have no violacein operon into phylogenetic context with the other strains. Neither BLAST+ nor Geneious searches found any trace of the violacein operon in strain PAZ21C, processed in-house at the McKenzie Lab (Table 2) (Camacho et al. 2009, Kearse et al. 2012a). We found neither the operon as a whole, nor the five genes individually. Similar results occurred when strains PAMC 25724, NBRC 102515 and CG3, downloaded from NCBI, were analyzed. No portions of the violacein genes could be found.

Reads matching the 16S rRNA gene were found in the raw sequence data of the atypical strains, PAZ21C, NBRC 102515, PAMC 25724 and CG3: 1) confirming the raw sequence data and downloaded data was valid and I had made no detectable errors in processing, and 2) allowing these intriguing strains to be included in the overall analysis, even though they lack the operon we were most interested in for this study. The 16S rRNA alignment confirmed the newly isolated atypical PAZ21C strain is a member of the species *Janthinobacterium lividum* in spite of its lack of violacein producing genes, sharing a greater than 99.5% identity with five separate members of the *lividum* clade (genetic group) (Appendix 2: Matrix 1).

Another strain included in the 16S rRNA analysis but absent from the violacein analysis is the Marseille strain. Violacein analysis of this atypical strain reveals a segment which aligns with the violacein operon, but only at ~50% identity. This uncertainty is homogeneous across the entire operon. No segments within the aligned region are similar enough to the reference sequence to identify them as clearly analogous to functional gene regions of the violacein operon, therefore it was excluded from the violacein analyses.

The five-gene violacein operon showed variation across all strains examined (Appendix 1: Matrix 2). No two strains shared more than 98.655% identity within the violacein operon, as opposed to the 16S alignment in which the strains are highly similar. The greatest difference in the 16S rRNA gene between any two *Janthinobacterium* strains (of any species) in our study was only 4.061% (Appendix 2: Matrix 1). In fact, 16S rRNA analysis resulted in 100% shared identity between three pairs of strains: PAZ19G **|** MTR, CCOS 423 **|** VA0829, and DE0145 **|** RIT308 (Appendix 2: Matrix1).

Unlike the highly variable trees generated from the 16S rRNA gene, the topologies of paired phylogenetic trees created with PhyML and FastTree for the violacein operon were extremely similar, with a few differences in bootstrap support (a measure of the probability that a given topology is accurate) between the two methods, shown below in Figure 1.

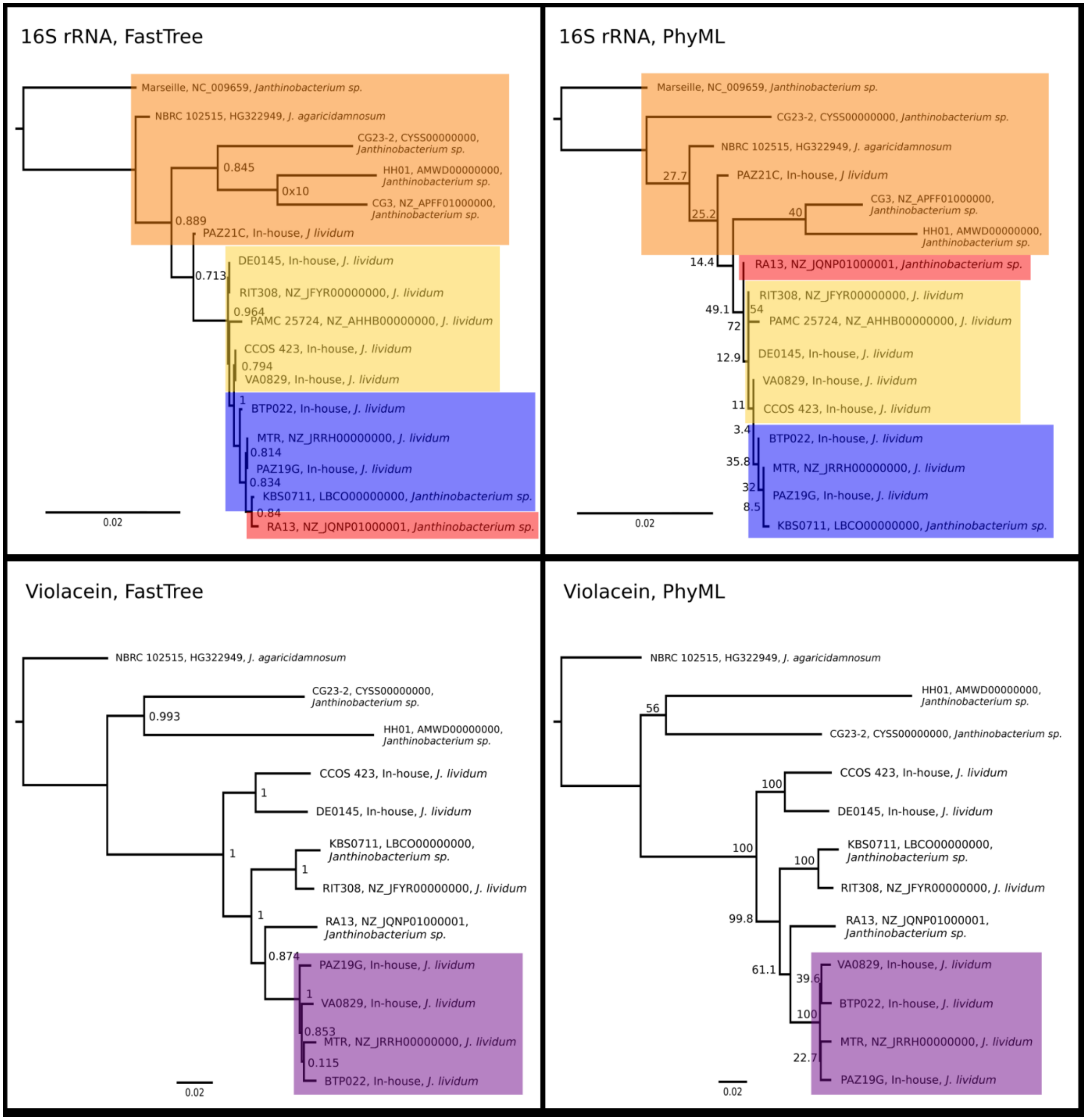
16S rRNA, FastTree
Marseille, NC_009659, Janthinobacterium sp.
NBRC 102515, HG322949, J. agaricidamnosum
CG23-2, CYSS00000000, Janthinobacterium sp.
0.845
HH01, AMWD00000000, Janthinobacterium sp.
0x10
CG3, NZ_APFF01000000, Janthinobacterium sp.
0.889
PAZ21C, In-house, J lividum
DE0145, In-house, J. lividum
0.713
RIT308, NZ_JFYR00000000, J. lividum
0.964
PAMC 25724, NZ_AHHB00000000, J. lividum
CCOS 423, In-house, J. lividum
0.794
VA0829, In-house, J. lividum
1
BTP022, In-house, J. lividum
MTR, NZ_JRRH00000000, J. lividum
0.814
PAZ19G, In-house, J. lividum
0.834
KBS0711, LBCO00000000, Janthinobacterium sp.
0.84
RA13, NZ_JQNP01000001, Janthinobacterium sp.
0.02
16S rRNA, PhyML
Marseille, NC_009659, Janthinobacterium sp.
CG23-2, CYSS00000000, Janthinobacterium sp.
NBRC 102515, HG322949, J. agaricidamnosum
27.7
PAZ21C, In-house, J lividum
CG3, NZ_APFF01000000, Janthinobacterium sp.
25.2
40
HH01, AMWD00000000, Janthinobacterium sp.
14.4
RA13, NZ_JQNP01000001, Janthinobacterium sp.
RIT308, NZ_JFYR00000000, J. lividum
49.1
54
72
PAMC 25724, NZ_AHHB00000000, J. lividum
12.9
DE0145, In-house, J. lividum
VA0829, In-house, J. lividum
11
CCOS 423, In-house, J. lividum
3.4
BTP022, In-house, J. lividum
35.8
MTR, NZ_JRRH00000000, J. lividum
32
PAZ19G, In-house, J. lividum
8.5
KBS0711, LBCO00000000, Janthinobacterium sp.
0.02
Violacein, FastTree
NBRC 102515, HG322949, J. agaricidamnosum
CG23-2, CYSS00000000, Janthinobacterium sp.
0.993
HH01, AMWD00000000, Janthinobacterium sp.
CCOS 423, In-house, J. lividum
1
DE0145, In-house, J. lividum
KBS0711, LBCO00000000, Janthinobacterium sp.
1
1
RIT308, NZ_JFYR00000000, J. lividum
1
RA13, NZ_JQNP01000001, Janthinobacterium sp.
0.874
PAZ19G, In-house, J. lividum
1
VA0829, In-house, J. lividum
0.853
MTR, NZ_JRRH00000000, J. lividum
0.115
BTP022, In-house, J. lividum
0.02
Violacein, PhyML
NBRC 102515, HG322949, J. agaricidamnosum
HH01, AMWD00000000, Janthinobacterium sp.
56
CG23-2, CYSS00000000, Janthinobacterium sp.
100
CCOS 423, In-house, J. lividum
DE0145, In-house, J. lividum
100
KBS0711, LBCO00000000, Janthinobacterium sp.
100
RIT308, NZ_JFYR00000000, J. lividum
99.8
RA13, NZ_JQNP01000001, Janthinobacterium sp.
61.1
VA0829, In-house, J. lividum
39.6
BTP022, In-house, J. lividum
100
MTR, NZ_JRRH00000000, J. lividum
22.7
PAZ19G, In-house, J. lividum
0.02

Figure 1: Evolutionary trees from the 16S rRNA gene compared with those from the violacein operon. Format: Strain, NCBI Accession #, Genus species. Color blocks denote areas of variation between two tree-building methods: FastTree and PhyML. Support statistics appropriate to each modeling program are shown: FastTree- local support values based on the Shimodaira-Hasegawa (SH) test, PhyML- 1,000 replicate bootstrap support.

To incorporate the unpigmented strains into the bigger picture, I combined the results from the 16S rRNA and violacein analyses (Figure 1) into a single informed hypothesis of the general topology of the evolutionary tree (cladogram) Figure 2.

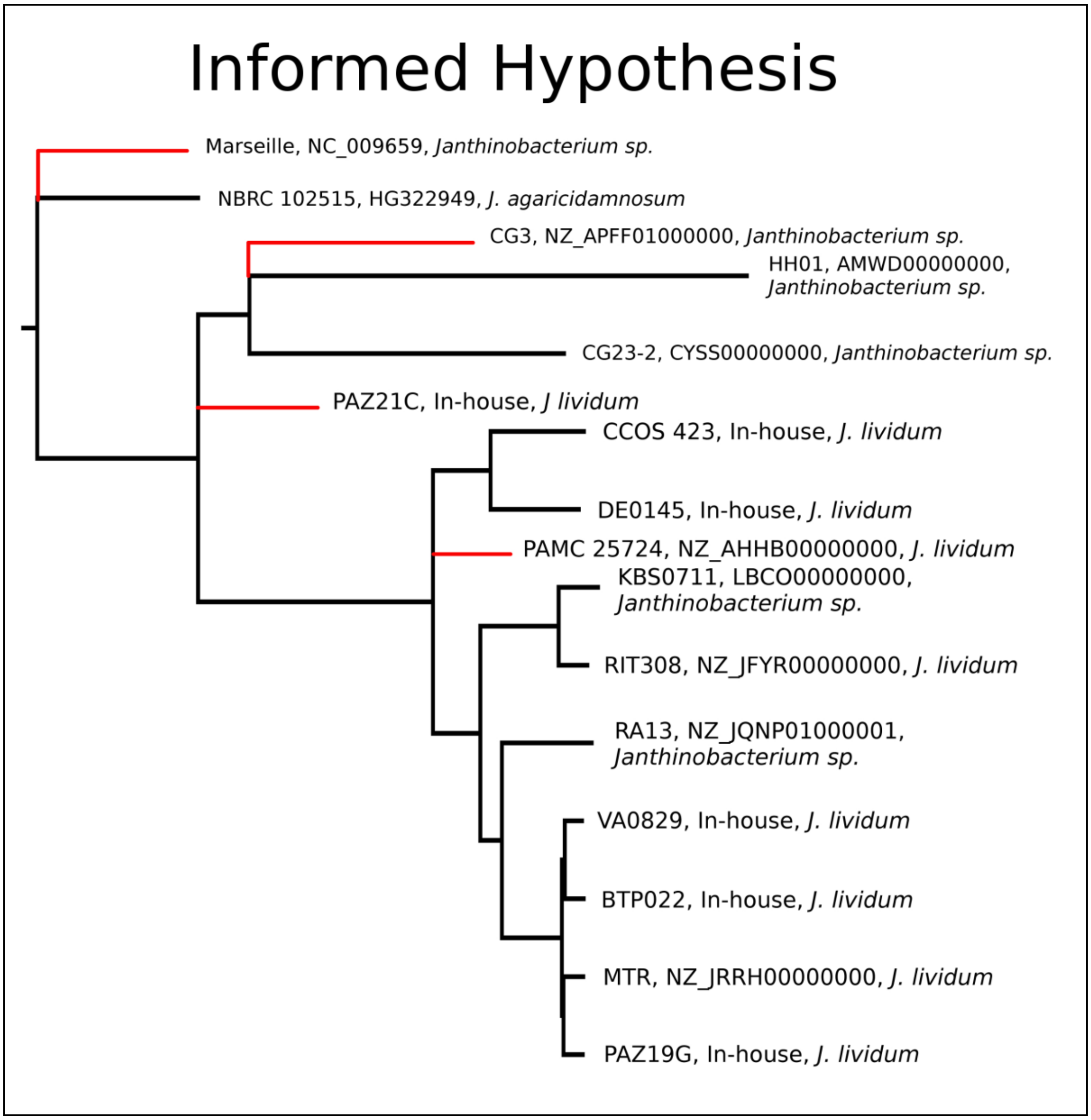
Informed Hypothesis
Marseille, NC_009659, *Janthinobacterium sp.*
NBRC 102515, HG322949, *J. agaricidamnosum*
CG3, NZ_APFF01000000, *Janthinobacterium sp.*
HH01, AMWD00000000, *Janthinobacterium sp.*
CG23-2, CYSS00000000, *Janthinobacterium sp.*
PAZ21C, In-house, *J lividum*
CCOS 423, In-house, *J. lividum*
DE0145, In-house, *J. lividum*
PAMC 25724, NZ_AHHB00000000, *J. lividum*
KBS0711, LBCO00000000, *Janthinobacterium sp.*
RIT308, NZ_JFYR00000000, *J. lividum*
RA13, NZ_JQNP01000001, *Janthinobacterium sp.*
VA0829, In-house, *J. lividum*
BTP022, In-house, *J. lividum*
MTR, NZ_JRRH00000000, *J. lividum*
PAZ19G, In-house, *J. lividum*

Figure 2: Combined results of 16S rRNA and violacein analyses. Formatted: Strain, NCBI Accession #, Genus species. Red branches denote unpigmented strains.

DISCUSSION

During our study investigating genetic variation in *J. lividum*, we observed some aspects of sequencing and analysis which may prove useful to systematists working in bacterial taxonomy and classification. We begin with an explanation of the ~7,000 base pair (7kbp) violacein operon as a more informative region than the ~1500bp 16S rRNA gene for determining evolutionary relationships between bacterial strains. This operon may be the best choice for classifying violacein-producing bacteria at the species level, short of full genome assembly and analysis. Controversy exists regarding the relevance and criteria of the species concept as it pertains to bacterial life forms, which we will not address here. We merely present the violacein operon as an aid in differentiating violacein-producing bacterial strains to a finer degree than is convenient with the standard 16S rRNA approach. We follow with a discussion of the results revealed from the robust tree we assembled combining the information from 16S and violacein phylogenies (evolutionary trees). While we found no clear and robust evidence for a correlation between genotype and environmental substrate or geographic region of origin, a hint of continental-scale grouping was observed.

Operon vs single gene for fine-scale analyses

Phylogenetic analysis of the 16S rRNA gene places some unpigmented strains at the base of the *lividum* clade (Figure 2), warranting further investigation and explanation. This placement of a group of unpigmented strains at the center of a genus known for its purple pigmentation is perplexing, until the life history of the *Janthinobacterium* genus is considered. The nearly complete lack of the violacein operon in three out of sixteen strains of mixed species sampled for our study is a reminder of the distinctive reproductive behavior of these bacteria. Members of the genus *Janthinobacterium* are known to exhibit intraclonal polymorphism, a characteristic driven by natural selection of asexual organisms evolving in highly variable environmental conditions (Gillis and De Ley 2006, Lincoln et al. 1999, Rainey et al. 1993). Under these conditions, the production of highly variable offspring increases the survivability of the species, therefore bacteria persisting in these environments encounter natural selection which favors the creation of genetically variable offspring, even within a single clonal lineage. A variety of mechanisms are employed by bacteria to introduce and perpetuate genetic variability within the context of clonal reproduction (binary fission) (Gillis and De Ley 2006). In *J. lividum* these mechanisms are more apparent than in other microorganisms, merely because they sometimes affect the violacein operon and the purple colored violacein it produces; a characteristic obvious to the naked eye. Many other intraclonal phenotypic changes also occur in these and other bacteria which are undetectable without specialized experimentation, and therefore go unnoticed (Gillis and De Ley 2006, Rainey et al. 1993).

It's ironic that the very qualities that make 16S rRNA appropriate and effective for distinguishing between microbial organisms at the genus level become a hindrance to identification at the species level. Ultra-conserved regions of the genome (regions which show very little change over long periods of evolutionary time) are effective for phylogenetic inference at a wide scale, when comparing highly divergent organisms, but they can only provide limited resolution when applied to closely-related organisms. The stereotypical result of narrow-focus genetic analyses from 16S rRNA genes is a flat tree, exhibiting large regions of polytomy, which indicate uncertainty or a lack of data necessary to differentiate between organisms. Even in cases where evolutionary relationships are detected by 16S rRNA analysis, bootstrap support for the relationships is very low (another indicator of uncertainty), as displayed in Figure 1, which shows the 16S phylogenies of the strains analyzed in this study. These results agree with previous indications that 16S rRNA identification of bacterial species cannot be relied upon as an indicator of specific phenotypic (or even genotypic) characteristics of any given strain (Janda and Abbot 2007).

A better choice for finer resolution phylogenetic analysis would be a relatively long sequence of conserved nucleotide bases which code for a trait subject to selective pressures. It just so happens that the violacein operon is exactly this sort of sequence. From a statistical perspective, a region of ~7,000 bp is more informative than the ~1500 bp 16S rRNA sequence, simply because it is longer. Added to the fundamental difference in mathematical robustness provided by a longer contiguous DNA sequence is the fact that the violacein operon is responsible for the production of a metabolite which provides the organism with a selective advantage in some environments (Mojib et al. 2013). Therefore the region is under selective pressure. This type of sequence would be inappropriate for differentiating between distantly related species, however for closely related organisms, it is bound to be highly informative.

Loss or damage to any single gene within the violacein operon results in loss of functionality (the operon no longer produces the metabolite) therefore functional codons are likely to be conserved within the gene regions of the operon (Sanchez et al. 2009). This operon is composed of a contiguous series of five conserved genes (vioA through vioE), connected by very short variable regions. This configuration is well suited to generating high-resolution phylogenetic trees and makes the violacein operon extremely informative; especially considering the relatively small investment of time and resources required for assembly and analysis, compared with whole genome assembly.

Figure 1 illustrates the high variability and uncertainty of phylogenetic inference based on the limited information from the 16S rRNA gene, and clearly shows the results became far less variable and more certain (higher bootstrap support) when the same two software programs were used to analyze the violacein operon. This is strong evidence that the operon is a more informative sequence than the 16S rRNA gene for genomic analysis at this resolution.

The greatest and most obvious weakness of utilizing the violacein operon to differentiate between strains of *Janthinobacterium* is that not all strains contain the operon. Consequently, it must be used in conjunction with other gene regions and the understanding that not all strains appear purple. Nonetheless it can be used to determine relationships that would otherwise be obscured in standard 16S analysis. Another important limitation of this study was the use of default settings in parts of the assembly, alignment, and phylogenetic inference. While these settings are recommended by their respective manufacturers for general research, each step through the processing could theoretically be optimized to produce an even clearer result.

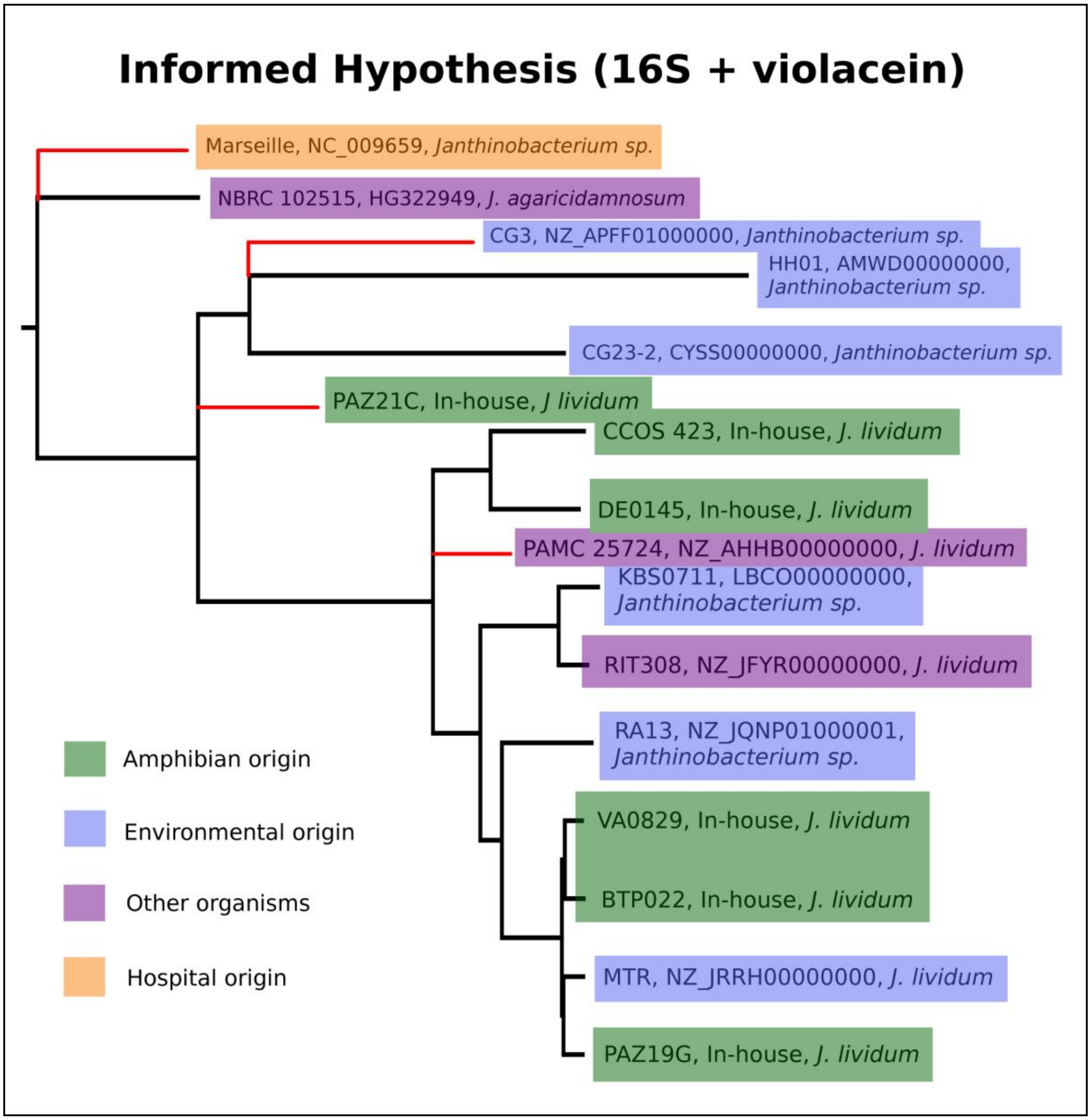
Informed Hypothesis (16S + violacein)
Marseille, NC_009659, Janthinobacterium sp.
NBRC 102515, HG322949, J. agaricidamnosum
CG3, NZ_APFF01000000, Janthinobacterium sp.
HH01, AMWD00000000, Janthinobacterium sp.
CG23-2, CYSS00000000, Janthinobacterium sp.
PAZ21C, In-house, J lividum
CCOS 423, In-house, J. lividum
DE0145, In-house, J. lividum
PAMC 25724, NZ_AHHB00000000, J. lividum
KBS0711, LBCO00000000, Janthinobacterium sp.
RIT308, NZ_JFYR00000000, J. lividum
RA13, NZ_JQNP01000001, Janthinobacterium sp.
VA0829, In-house, J. lividum
BTP022, In-house, J. lividum
MTR, NZ_JRRH00000000, J. lividum
PAZ19G, In-house, J. lividum
Amphibian origin
Environmental origin
Other organisms
Hospital origin

Figure 3: Correlation of substrate with strain. Format: Strain, NCBI Accession #, Genus species. Red branches denote unpigmented strains. Color blocks denote substrate.

Source variation

Speculation exists within the community of amphibian researchers regarding the bacteria which populate the amphibian microbiome. How might they differ from related taxa in the environment? Are they specially adapted to inhabit amphibian skin? Our research contains data which could be relevant to these questions.

This data set includes strains of *J. lividum* from both amphibians and soil samples. However, it should be noted that none of these were collected in conjunction. Environmental samples from naturally occurring soil or water examined in this study originated from six widespread locations, including from Antarctica, where amphibians do not occur, whereas the amphibian hosts were from six completely different locations. The geographically closest strains from the amphibian microbiome and a nearby environmental origin are DE0145 from an amphibian in Solling, Germany and HH01 from a water sample in Hamburg, Germany, 250 km away (97.4304% identity).

The geographic distance between the most closely located amphibian and environmental strains confounds the results to such a degree that any apparent trends are subject to a high degree of skepticism. Nonetheless, this data appears to support the hypothesis that amphibian and soil strains cannot be genetically differentiated and are therefore likely equivalent in terms of their niche in the microbiological ecosystem. Evidence for this hypothesis is presented in Figure 3, where the amphibian and environmental strains are spread widely across the phylogenetic analysis. If this is the case, it is possible that probiotic treatments for amphibians could be developed successfully from environmental strains of bacteria, making the development of these treatments much easier for researchers working in the field.

If a life-history aspect were linked to natural selection in these organisms, we would expect to see strains which share a particular life-history trait to be clustered together in a phylogenetic analysis. No clear evidence of clustering was observed in this tree (Fig. 3). In spite of these results, it should be noted that a better experiment could be conducted with methods specifically designed to answer this question. Researchers collecting microbiome samples from amphibians for live culturing should be encouraged to collect matching environmental samples, to create a database of paired cultures from amphibians and their surrounding environments, in order to answer these questions definitively.

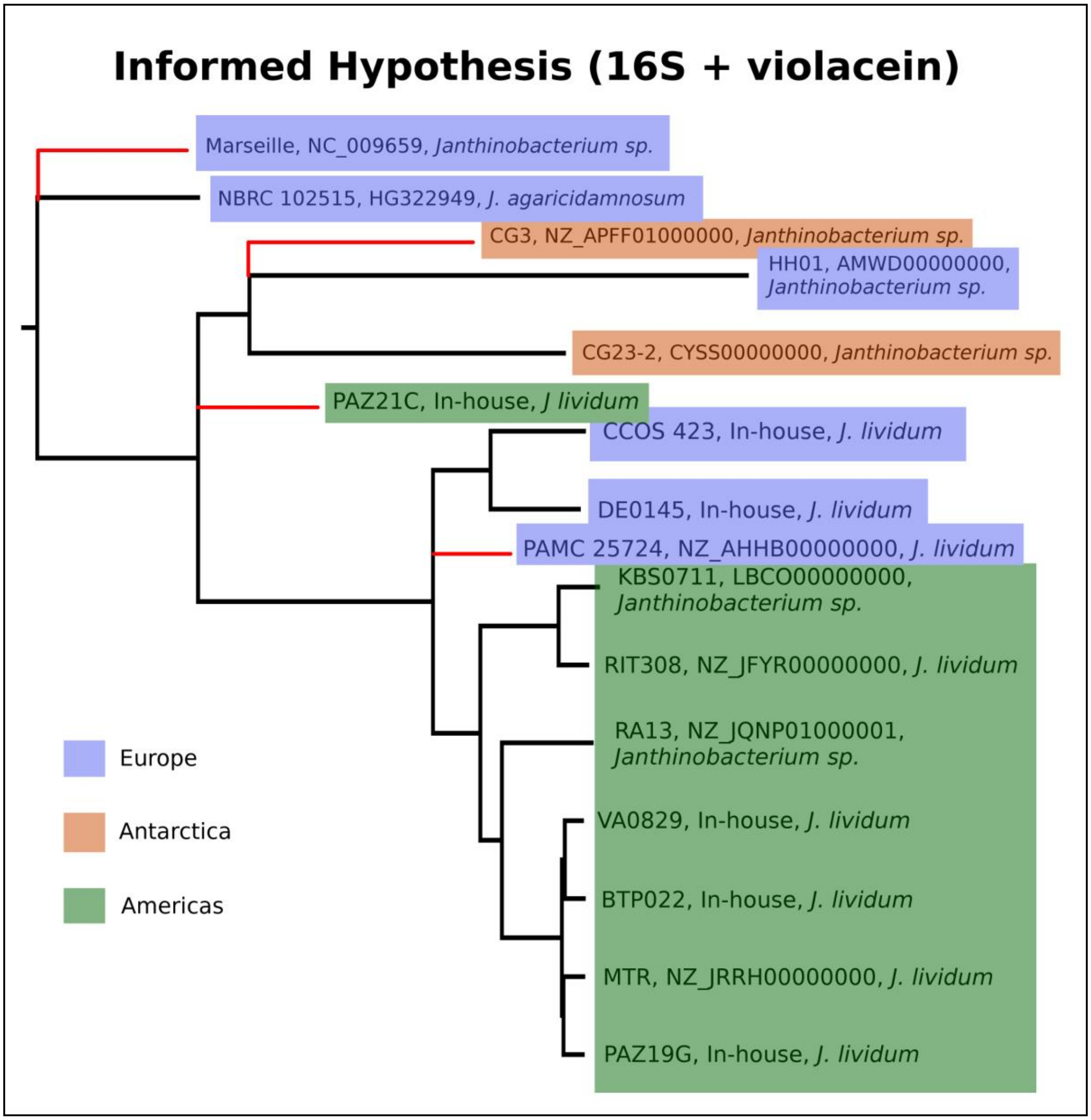
Informed Hypothesis (16S + violacein)
Marseille, NC_009659, *Janthinobacterium sp.*
NBRC 102515, HG322949, *J. agaricidamnosum*
CG3, NZ_APFF01000000, *Janthinobacterium sp.*
HH01, AMWD00000000, *Janthinobacterium sp.*
CG23-2, CYSS00000000, *Janthinobacterium sp.*
PAZ21C, In-house, *J lividum*
CCOS 423, In-house, *J. lividum*
DE0145, In-house, *J. lividum*
PAMC 25724, NZ_AHHB00000000, *J. lividum*
KBS0711, LBCO00000000, *Janthinobacterium sp.*
RIT308, NZ_JFYR00000000, *J. lividum*
RA13, NZ_JQNP01000001, *Janthinobacterium sp.*
VA0829, In-house, *J. lividum*
BTP022, In-house, *J. lividum*
MTR, NZ_JRRH00000000, *J. lividum*
PAZ19G, In-house, *J. lividum*
Europe
Antarctica
Americas

Figure 4: Correlation of region with strain. Format: Strain, NCBI Accession #, Genus species. Red branches denote unpigmented strains. Color blocks denote region of origin.

Regional variation

Analysis of the violacein operon reveals clear genetic differences between strains of *Janthinobacterium lividum* (Appendix 2: Matrix 2). Unfortunately, the number of strains available for violacein analysis was limited to only twelve. Although a clear association between genotype and regional location could not be identified in this study, it would not be reasonable to declare that none exists. The question remains as to whether some regional or environmental variable may be an indicator for violacein production in field-collected strains of *J. lividum.*

I expected to find a clear regional distinction between strains of *J. lividum,* however contrary to expectations, the data do not strongly support this hypothesis. Although the data hints at a general pattern of relatedness among strains from the Americas when compared to those from Europe, any definitive conclusion on the matter would require a much more exhaustive analysis from a larger dataset. That being said, the matrices (1 and 2) and the phylogenetic trees (Figures 3 and 4) reveal two general clusters of strains with a very high percent identity (% shared DNA sequence), which could indicate some regional variation at the continental scale. Examining Figure 4, which correlates the strains with region of origin, it appears that strains within the *lividum* clade: PAZ19G from Panama-Maryland, MTR from Chile, BTP022 from Colorado, VA0829 from Virginia and RA13 from Washington may be clustered into a related group based in the Americas, distinct from a European group that includes DE0145 from Solling Germany and CCOS 423 from Switzerland. *Janthinobacterium lividum s*trains KBS 0711 from Michigan, and RIT308 from New York may form a third group. More research will certainly be required to resolve the issue. The presence of regional variation within the *J. lividum* species would complicate the development of probiotic treatments based on the bacterium for the treatment of amphibian populations in situ (in their region of origin). Even if a single probiotic treatment could provide universal protection against a globally distributed pathogen like *Bd*, it could be considered ecologically irresponsible to apply a single probiotic treatment across widespread geographic regions. Careful consideration should be undertaken before the application of probiotics from one region to another.

Our investigation of regional differentiation is subject to some important confounding factors and uncertainty. This is especially true of strains PAZ19G and PAZ21C, described as Panama-Maryland in Table 1. Unfortunately, the source(s) of these two strains is somewhat obscured by the fact that they were isolated from the microbiome of captive-bred Panamanian Golden Frogs (*Atelopus zetekii*) which are housed in Maryland, USA. There is no way to conclusively determine whether the bacteria in the microbiome of these frogs originated from the homeland of their wild parents, Panama, or whether it originated from their surroundings in the US. One clue could be that two of the strains which share the lowest percent identity across the 16S rRNA gene (99.4087%) in the *lividum* clade are PAZ21C and PAZ19G (which shares 100% identity with MTR, a strain from Chile). Strain PAZ21C shares 99.6058% identity across the 16S rRNA gene with RIT308 from New York and DE0145 from Solling, Germany. This would seem to indicate that strain PAZ21C is more closely related to strains from the US and Europe than it is to strains from Central or South America. Considering that both PAZ19G and PAZ21C came from the same ostensible origin, this ambiguity of origin contributes uncertainty to the results regarding the hypothesis of regional variations between strains.

Another source of uncertainty in the study arises from the use of data downloaded from the public NCBI website. One good example is the published draft genome for strain PAMC 25724. While it has been described by the researchers who published it as being purple when cultured, no violacein operon or any fragmented violacein genes could be detected in the published genome (Kim et al. 2012). The data for this strain contains only 4 million base pairs, compared to an average of 6 million for the other strains. While this could be a legitimately simpler organism with a smaller genome, there is also the possibility that the data is incomplete. Another possibility is the violacein operon present in this strain of *J. livi*dum could be located outside of the nucleoid, on a plasmid, which was not included in the data published to NCBI. Or the strain could produce an alternative purple pigment, which is not violacein. The cause of the missing operon remains a mystery. Strain CG3, on the other hand, has a published genome of more than 6 million base pairs, but shows no sign of the violacein operon. Unlike PAMC 25724, the publishers of CG3 noted that the strain appears unpigmented in culture (Smith et al. 2013). This corroborative information provided by the publisher is helpful for reducing uncertainty, but not all publishers are so forthcoming with key information.

Future work with these published datasets would warrant contacting the publishers directly for any metadata they could share. This type of project is fundamentally a collaborative effort, so reaching out to other researchers could be of immeasurable benefit to the project and provide for more robust results overall.

CONCLUSIONS

We recommend analysis of the violacein operon for differentiating between violacein-producing organisms at the species level. Furthermore, researchers working with *Janthinobacterium lividum* should consider the strain. It has been well established that some strains produce violacein and others do not (Gillis and De Ley 2006, Lincoln et al. 1999, Rainey et al. 1993). The amount of violacein produced can vary between strains (Schloss et al. 2010). Schloss et al. also recorded that certain environmental conditions, such as temperature and light, can affect violacein production even within a single strain (2010). All of these variables could affect the success or failure of a given probiotic treatment. This is especially important for researchers interested in developing probiotic treatments based on live *J. lividum* strains. We have confirmed that 16S rRNA identification is not sufficient to determine whether a bacterium in the genus *Janthinobacterium* will produce violacein. If *J. lividum* is detected in environmental assays or by culture-independent methods, researchers cannot assume that it is a strain which produces violacein based on 16S analysis alone.

Additional research is warranted, to more fully describe the genera *Janthinobacterium.* Raw sequence data generated for this study can be used to assemble complete genomes for six new strains of *J. lividum,* tripling the number of published genomes assigned to this species. Further analysis should also include an examination of the genes associated with quorum sensing, the regulatory mechanism most associated with violacein production (Pantanella et al. 2006, Schloss et al. 2010). Possibly the most helpful future research would be a study designed to map genotypic variation in the genome of *J. lividum* with phenotypic characteristics including morphology and the details of fluctuating violacein production.

ACKNOWLEDGEMENTS

A very special thank you to Dr. Valerie McKenzie, who made this research possible and mentored me through the entire process. Without her confidence in my ability and her investment of time and money, this research would have been impossible. I'm also very thankful to the other members of the McKenzie lab who helped out in so many ways, even braving the rain at 3,000 m elevation during our trips into the field: Holly Archer, Jordan Kueneman, Christine Avena, Abby Kimball, & Lisette Arellano. I owe a huge debt of gratitude to members of the Schmidt lab, especially Jack Darcy, & Adam Solon for all of their help getting the DNA sequenced, strategizing the data processing pipeline, and bringing me up to speed on the appropriate software programs. Thank you very much to those who contributed bacterial strains and culturing expertise to facilitate this research: Doug Woodhams, Reid Harris, Molly Bletz, Matthew Becker, and Jenifer Walke, whose contributions were very important to the success of this study. Thank you to my committee members, especially Betsy Forrest who provided a valuable outside perspective during my defense talk. Many thanks are also owed to Barbara Demmig-Adams and Pieter Johnson who guided me through the Honors Thesis process and helped me survive my senior year. A final thank you to the Colorado Parks and Wildlife Boreal Toad Recovery Team for providing research permits and knowledge about active sites during the 2015 season. Funding to support this project was provided by research grants from the NSF and Templeton Foundation to Valerie J. McKenzie.

REFERENCES

Antônio RV, Creczynski-Pasa TB. 2004. Genetic analysis of violacein biosynthesis by *Chromobacterium violaceum*. Genetics and Molecular Research: GMR. 3(1):85–91.

Bash – GNU Bourne-Again SHell. 2010. London (UK): Canonical Group Limited; [cited 2016 Feb 20]. Available from http://manpages.ubuntu.com/manpages/lucid/man1 /bash.1.html

Bletz MC, Loudon AH, Becker MH, Bell SC, Woodhams DC, Minibiole KP, Harris RN. 2013. Mitigating amphibian chytridiomycosis with bioaugmentation: characteristics of effective probiotics and strategies for their selection and use. Ecology Letters. 16(6):807-820.

Brucker RM, Harris RN, Schwantes CR, Gallaher TN, Flaherty DC, Lam BA, Minbiole KPC. 2008. Amphibian chemical defense: antifungal metabolites of the microsymbiont *Janthinobacterium lividum* on the salamander *Plethodon cinereus*. Journal of Chemical Ecology. 34(11):1422–1429.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10(1):421.

Copeland WB, Bartley BA, Chandran D, Galdzicki M, Kim KH, Sleight SC, Maranas CD, Sauro HM. 2012. Computational tools for metabolic engineering. Metabolic Engineering. 14(3):270-280.

Duran M, Ponezi AN, Faljoni-Alario A, Teixeira MFS, Justo GZ, Duran N. 2012. Potential applications of violacein: a microbial pigment. Medicinal Chemistry Research. 21(7):1524-1532.

Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. PLoS ONE. 8(6):e66643.

FastX-Toolkit: FASTQ/A short-reads pre-processing tools. 2016. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory; [cited 2016 Feb 20]. Available from http://hannonlab.cshl.edu/fastx_toolkit/index.html

García-Echauri SA, Gidekel M, Gutiérrez-Moraga A, Santos L, De León-Rodríguez A. 2011. Isolation and phylogenetic classification of culturable psychrophilic prokaryotes from the Collins glacier in the Antarctica. Folia Microbiol. *56*(3):209–214.

Gillis M, De Ley J. 2006. The Genera *Chromobacterium* and *Janthinobacterium.* In Dworkin M, Falkow S Rosenberg E, Schleifer KH, Stackebrandt E, eds. The Prokaryotes. New York (NY): Springer New York. p. 737-746.

Glenn TC. 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources. 11(5):759-769.

Guindon S, Delsuc F, Dufayard J, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. David Posada ed. Bioinformatics for DNA Sequence Analysis, Springer Protocols. Methods in Molecular Biology:113-137.

Guindon S, Dufayard JF, Lefort V, Anisinova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate Maximum-Likelihood Phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology. 59(3): 307-321.

Harris RN, Brucker RM, Walke JB, Becker MH, Schwantes CR, Flaherty DC, Lam BA, Woodhams DC, Briggs CJ, Vredenburg VT, Minbiole KPC. 2009. Skin microbes on frogs prevent morbidity and mortality caused by a lethal skin fungus. The ISME Journal. 3(7):818–824.

Hoshino T. 2011. Violacein and related tryptophan metabolites produced by Chromobacterium violaceum: biosynthetic mechanism and pathway for construction of violacein core. Applied Microbiology and Biotechnology. 91(6):1463–1475.

Janda JM, Abbot SL. 2007. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. Journal of Clinical Microbiology. 45(9):2761-2764.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012a. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647-1649.

Kearse M, Sturrock S, Meintjes P. 2012b. The Geneious 6.0.3 Read Mapper [Internet]. Aukland (NZ): Biomatters Limited; [cited 2016 Feb 16]. Available from http://assets.geneious.com/documentation/geneious/GeneiousReadMapper.pdf

Kim SJ, Shin SC, Hong SG, Lee YM, Lee H, Lee J, Choi I, Park H. 2012. Genome sequence of *Janthinobacterium sp.* strain PAMC 25724, isolated from alpine glacier cryoconite. Journal of Bacteriology. 194(8):2096-2096.

Lincoln SP, Fermor TR, Tindall BJ. 1999. *Janthinobacterium agaricidamnosum* sp. nov., a soft rot pathogen of *Agaricus bisporus*. International Journal of Systematic and Evolutionary Microbiology. 49(4):1577–1589.

Lu Y, Wang L, Xue Y, Zhang C, Xing X-H, Lou K, Zhang Z, Li Y, Zhang G, Bi J, Su J. 2009. Production of violet pigment by a newly isolated psychrotrophic bacterium from a glacier in Xinjiang, China. Biochemical Engineering Journal. 43(2):135–141.

Massa S, Caruso M, Trovatelli F, and Tosques M. 1998. Comparison of plate count agar and R2A medium for enumeration of heterotrophic bacteria in natural mineral water. World Journal of Microbiology & Biotechnology. 14:727-730.

Matz C, Deines P, Boenigk J, Arndt H, Eberl L, Kjelleberg S, Jürgens K. 2004. Impact of violacein-producing bacteria on survival and feeding of bacterivorous nanoflagellates. Applied and Environmental Microbiology. 70(3):1593–1599.

McKenzie VJ, Bowers RM, Fierer N, Knight R, Lauber CL. 2012. Co-habiting amphibian species harbor unique skin bacterial communities in wild populations. The ISME Journal. 6(3):588-596.

Mojib N, Farhoomand A, Andersen DT, Bej AK. 2013. UV and cold tolerance of a pigment-producing Antarctic *Janthinobacterium sp.* Ant5-2. Extremophiles. 17(3):367-378.

Pantanella F, Berlutti F, Passariello C, Sarli S, Morea C, Schippa S. 2006. Violacein and biofilm production in *Janthinobacterium lividum*. Journal of Applied Microbiology. 102:992-999.

PhyML 3.0: new algorithms, methods and utilities. 2008. Montpellier (FR): Le Centre national de la recherché scientifique; [cited 2016 Feb 20]. Available from http://www.atgc-montpellier.fr/phyml/binaries.php

Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution. 26(7):1641–1650.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS One. 5(3):e9490.

Price MN. [date unknown]. FastTree. [Internet]. Berkeley (CA): Lawrence Berkeley National Lab; [cited 2016 Feb 18]. Available from http://www.microbesonline.org/fasttree/

Rainey PB, Moxon RE, Thompson IP. 1993. Intraclonal polymorphism in bacteria. In: Jones JG, editor. Advances in Microbial Ecology Vol. 13. Boston (MA): Springer US. p. 263-300.

Ramsey JP, Mercurio A, Holland JA, Harris RN, MinbioleKPC. 2015. The cutaneous bacterium *Janthinobacterium lividum* inhibits the growth of *Trichophyton rubrum* in vitro. International Journal of Dermatology. 54(2):156–159.

Reasoner DJ, Geldreich EE. 1985. A new medium for the enumeration and subculture of bacteria from potable water. Applied and Environmental Microbiology. 49(1):1–7.

Saeger JL, Hale AB. 1993. Genetic variation within a lotic population of *Janthinobacterium lividum.* Applied and Environmental Microbiology. 59(7):2214-2219.

Sánchez C, Braña AF, Méndez C, Salas JA. 2006. Reevaluation of the violacein biosynthetic pathway and its relationship to indolocarbazole biosynthesis. ChemBioChem. 7(8):1231–1240.

Schloss PD, Allen HK, Klimowicz AK, Mlot C, Gross JA, Savengsuksa S, McEllin J, Clardy J, Ruess RW, Handelsman J. 2010. Psychrotrophic strain of *Janthinobacterium lividum* from a cold Alaskan soil produces prodigiosin. DNA and Cell Biology. 29(9):533–541.

Segawa T, Miyamoto K, Ushida K, Agata K, Okada N, Kohshima S. 2005. Seasonal change in bacterial flora and biomass in mountain snow from the Tateyama mountains, Japan, analyzed by 16S rRNA gene sequencing and real-time PCR. Applied and Environmental Microbiology. 71(1):123–130.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nature Reviews Genetics. 15(2):121-132.

Smith H, Akiyama T, Foreman C, Franklin M, Woyke T, Teshima H, Davenport K, Daligault H, Erkkila T, Goodwin L, Gu W, Xu Y, Chain P. 2013. Draft genome sequence and description of *Janthinobacterium sp.* strain CG3, a psychrotolerant Antarctic supraglacial stream bacterium. Genome announcements. 1(6):e00960-13.

Starliper CE, Watten BJ, Iwanowicz DD, Green PA, Bassett NL, Adams CR. 2015. Efficacy of pH elevation as a bactericidal strategy for treating ballast water of freight carriers. Journal of Advanced Research. 6(3):501–509.

Woodhams DC, Brandt H, Baumgartner S, Kielgast J, Kupfer E, Tobler U, Davis LR, Schmidt BR, Bel C, Hodel S, Knight R, McKenzie V. 2014. Interacting symbionts and immunity in the amphibian skin mucosome predict disease risk and probiotic effectiveness. PloS ONE. 9(4):e96375.

Zhang X, Enomoto K. 2011. Characterization of a gene cluster and its putative promoter region for violacein biosynthesis in *Pseudoalteromonas sp*. *520P1*. Applied Microbiology and Biotechnology. 90(6):1963–1971.

Appendix 1:

Pre-processing script development

Although the Geneious software package was used for assembling and analyzing the genes, a significant amount of pre-processing was required in order to prepare the data for use in Geneious. To this end, I researched and wrote a “script” or series of instructions to be used in a command-line interface in order to properly prepare the raw sequence data.

First, I determined how many “reads” or 101 base pair sequences were produced by the Next-seq DNA sequencing process, by counting the number of lines present in the .fastq file supplied by the next-generation Illumina sequencer. The FastQ format includes four lines of information to describe each read. Therefore a FastQ file containing 36 million lines represents 9 million “reads” or “raw sequences”.

The *J. lividum* genome is usually expected to include approximately 5 to 6 million base pairs. Assuming an expected size of 6 million base pairs: 9 million, 101 base pair reads results in “coverage depth” of 909 million / 6 million = 151 bases per nucleotide location.

At this point, I employed “downsampling” for two purposes: to increase the overall quality of the data set and to facilitate faster processing speeds. Downsampling in this case consisted of extracting a subset of the highest-quality reads from the center of the data, because if and when anomalies occur in Next-gen sequencing data, they tend to occur at the beginning or end of the sequencing process (Eren et al. 2013; Glenn 2011). Extracting a set of 3 million reads still provided ample coverage of 303 million / 6 million = 50 reads per nucleotide base.

To further facilitate processing speeds, the downsampled sequences were examined using BLAST+, to locate reads with characteristics of the violacein genes under study. These specific sequences were then collected into a much smaller data file which could be easily and efficiently imported into the Geneious 9.0.5 software package for final reference-based assembly and analysis. The use of BLAST+ for this second step provided a second opportunity for quality filtering (e-value 10).

I began by converting the .fastq file type into .fasta format, using FastX. This step was necessary in order to create a file type required by the BLAST+ software. This .fasta file was then expanded into a BLAST+ formatted database. I located a *Janthinobacterium* genome in the NCBI nucleotide database which included an annotated sequence for the five gene operon responsible for violacein production (strain RA13). I extracted this 7,390 base pair sequence and exported it to a .fasta file-type compatible with BLAST+.

Finally, I used the reference sequence to search the downsampled database of raw sequences for any reads containing matching bases. After identifying the relevant reads, I created a new, and much smaller data set which only included the ~2,000 - 3,000 reads most likely to form a part of the five gene violacein operon. This resulted in coverage of 3,000 * 101 raw base pairs / 7,000 base pair length of operon = coverage of 44. This streamlined file was then imported into Geneious for reference-based assembly and alignment.

Annotated script

```
#Written by C. Weiblen. Use at your own risk. Back up your files first! :)
#For simplicity, put the reference sequence in the same location with the raw sequences.
#To run this code, three things are needed: path_to_file_location, Core_ID, and Reference_sequence.
#Generated filenames are designed to append stepwise descriptions to the end of the core ID.

#Core_ID - Search & replace the core ID 'JD29_R2'
#Reference_filename - Search & replace reference 'Violacein_Genes_LC000629.fasta'

#Specify path_to_file_location.
cd "/media/c/Storage/J-liv_Genomes/Sample_JD29"

#Count the lines in the raw sequence data file.
wc -l "JD29_R2.fastq" > "JD29_R2_linecount.txt"

#Downsample as needed, depending on line count. fastq contains four lines per sequence.
#fasta contains two lines per sequence.
#The following downsamples 9 million sequences of fastq formatted raw data to 3 million sequences.
#'tail' and 'head' determine how many lines are RETAINED from the source file.
cat "JD29_R2.fastq" | tail -n 24000000 | head -n 12000000 > "JD29_R2_3Mb.fastq"

#Convert .fastq to .fasta. Requires FastX.
#This is necessary because BLAST+ wants a .fasta file in order to build a database.
fastq_to_fasta -v -n -i "JD29_R2_3Mb.fastq" -o "JD29_R2_3Mb.fa"

#Format raw sequence .fasta files as BLAST db.
#Note this creates about a half-dozen files in the directory.
#Important: -max_file_sz is called here in order to override the default setting, which is only 1 Gig.
makeblastdb -in "JD29_R2_3Mb.fa" -parse_seqids -dbtype nucl -max_file_sz 1900000000B

#Compare the reference against raw sequences and output in tabular format.
#-evalue default is 10. Called here so it can be changed easily without having to look up the command.
#Important: -max_target_seqs MUST be called in order to override the default setting
#Default is only 500 - NOT enough for an assembly.
blastn -query Violacein_Genes_LC000629.fasta -db "JD29_R2_3Mb.fa" -out "JD29_R2_3Mb-violacein-
test.txt" -evalue 1e-10 -outfmt 6 -max_target_seqs 5000000

#Isolate record IDs of hits from column 2 of the blastn output format 6 (tabular).
cat "JD29_R2_3Mb-violacein-test.txt" | cut -f 2 > "JD29_R2_3Mb-violacein-test_hitslist.txt"

#Count the lines in the hitslist file, just to see what you've got.
wc -l "JD29_R2_3Mb-violacein-test-hitslist.txt" > "JD29_R2_3Mb-violacein-test-hitslist_linecount.txt"

#Use IDs of hits to extract the matching sequences from the downsampled raw sequence file.
blastdbcmd -db "JD29_R2_3Mb.fasta" -dbtype nucl -entry_batch "JD29_R2_3Mb-violacein-test_hitslist.txt"
-outfmt %f -out "JD29_R2_3Mb-violacein-test_hitsequences.fasta"
```

Appendix 2: Relationship matrices

	Marseille	HH01	CG3	CG23-2	NBRC 102515	PAZ21C	PAMC 25724	DE0145	CCOS423	KBS0711	RA13	RIT308	MTR	VA0829	PAZ19G	BTP022
Marseille	---	96.0656	95.9398	96.3303	97.3097	96.8504	96.6371	96.7192	96.6535	96.7192	96.7848	96.7182	96.6535	96.6535	96.6535	96.5879
HH01	60	---	97.7690	97.5049	97.4393	97.4393	97.2915	97.4393	97.3736	97.4393	97.5049	97.4393	97.3736	97.3736	97.3736	97.3079
CG3	62	34	---	97.1148	97.5082	98.0328	97.9836	98.0328	98.0984	98.0328	98.0984	98.0328	98.0984	98.0984	98.0984	98.0328
CG23-2	56	38	44	---	97.5066	97.5722	97.1949	97.4409	97.4409	97.5722	97.5066	97.4409	97.5722	97.4409	97.5722	97.5066
NBRC 102515	41	39	38	38	---	99.2116	99.1294	99.0145	98.9488	99.0145	99.0802	99.0145	98.9488	98.9488	98.9488	98.8830
PAZ21C	48	39	30	37	12	---	99.3922	99.6058	99.5401	99.4744	99.5401	99.6058	99.4087	99.5401	99.4087	99.4744
PAMC 25724	53	43	32	44	15	11	---	99.7208	99.7208	99.5894	99.6551	99.7208	99.5894	99.7208	99.5894	99.6551
DE0145	50	39	30	39	15	6	6	---	99.9343	99.8686	99.9343	100	99.8029	99.9343	99.8029	99.8686
CCOS 423	51	40	29	39	16	7	6	1	---	99.8029	99.8686	99.9343	99.8686	100	99.8686	99.9343
KBS0711	50	39	30	37	15	8	8	2	3	---	99.9343	99.8686	99.9343	99.8029	99.9343	99.8686
RA13	49	38	29	38	14	7	7	1	2	1	---	99.9343	99.8686	99.8686	99.8686	99.8029
RIT308	50	39	30	39	15	6	6	0	1	2	1	---	99.8029	99.9343	99.8029	99.8686
MTR	51	40	29	37	16	9	8	3	2	1	2	3	---	99.8686	100	99.9343
VA0829	51	40	29	39	16	7	6	1	0	3	2	1	2	---	99.8686	99.9343
PAZ19G	51	40	29	37	16	9	8	3	2	1	2	3	0	2	---	99.9343
BTP022	52	41	30	38	17	8	7	2	1	2	3	2	1	1	1	---

Table 3: Pairwise relationship matrix of strains based on the 16S rRNA gene. Top right: % identity shared by two strains. Lower left: total number of different bases between two strains. Highest and lowest values highlighted.

	HH01	NBRC 102515	CG23-2	CCOS 423	DE0145	KBS0711	RIT308	RA13	BTP022	PAZ19G	VA0829	MTR
HH01	---	80.7547	81.2525	77.4733	76.7151	80.5812	80.8121	80.9957	81.2033	81.2712	81.2081	81.2168
NBRC 102515	1428	---	83.1108	79.7303	78.5120	82.2768	82.4791	82.5946	82.5411	82.5000	82.4681	82.5605
CG23-2	1380	1252	---	79.8427	79.0373	82.8416	82.9776	83.0166	83.1963	83.2270	83.1575	83.0610
CCOS 423	1480	1343	1324	---	88.0081	87.9097	88.1245	87.9720	88.3060	88.3647	88.3986	88.2142
DE0145	1563	1455	1407	756	---	86.9275	87.0476	86.9246	87.5675	87.6294	87.5057	87.4699
KBS0711	1430	1314	1262	796	880	---	97.7527	93.3270	93.5680	93.4815	93.4586	93.4397
RIT308	1413	1299	1252	782	872	165	---	93.6659	93.8401	93.7398	93.7809	93.7670
RA13	1401	1292	1251	794	881	491	466	---	94.5098	94.6521	94.5998	94.5603
BTP022	1387	1297	1239	772	838	475	455	406	---	98.5466	98.6548	98.6513
PAZ19G	1385	1302	1240	772	835	484	465	398	112	---	98.4564	98.5262
VA0829	1382	1298	1237	765	841	485	461	402	104	121	---	98.5752
MTR	1380	1290	1244	774	841	481	457	399	101	112	112	---

Table 4: Pairwise relationship matrix of strains based on the violacein operon. Top right: % identity shared by two strains. Lower left: total number of different bases between two strains. Highest and lowest values highlighted.