Description of *Lycobacterium gen. nov.* and *Noelibacterium gen. nov.* and description of 17 novel species within these new genera

Presented to the faculty of Lycoming College in partial fulfillment of the requirements for Departmental Honors in Biology

By

Colin Allen

Lycoming College

6 May 2021

Approved by:

(Jeff Newman)

(Betty McCall)

(Michelle Briggs)

(Holly Bendorf)

Abstract

The names of different bacteria should indicate of their characteristics and genetic relationships. The present-day issue in taxonomy, however, is the presence of many broad, "catch all" groups that were established before genome sequencing. As a result, the significance of the names of these broad groups are diluted and have little meaning because many of the members have widely different characteristics. *Flavobacterium* is one such "catch all" genus that needs to be further explored to understand the relationships between members. From 2012 to 2020, 23 species of *Flavobacterium* were isolated from various creeks near Williamsport, PA. It was hypothesized that these organisms, *Flavobacterium* sp. CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* CRH, *F. sp.* GR2016-10, *F. sp.* CSZ, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* JLP, *F. sp.* MR2016-29, *F. sp.* MC2016-06, *F. sp.* ABG, *F. sp.* LC2016-12, *F. sp.* JRM, *F. sp.* KMS, *F. sp*. AJR, *F. sp.* ALJ2, *F. sp.* AED, and *F. sp.* HJJ were novel species. The genome of each organism was sequenced, annotated, and the organisms were compared using 16s rRNA gene sequence similarity, rpoB gene sequence similarity, estimated DNA-DNA hybridization (eDDH), average nucleotide identity (ANI), average amino acid identity (AAI), and the Genome Taxonomy Database (GTDB) Tree (Parks et al. 2018). Phenotypic analysis was also run to determine phenotypic differences amongst the novel species and their closest validly named relatives. These analyses led to the conclusion that 17 species were sufficiently divergent at the genomic level from their closest validly published relatives to be considered new species. The remaining 6 organisms were determined to be independent isolates of other various novel species listed here. Comparing AAI, all 17 novel species were <70% similar to *F. aquatile* LMG 4008^{T} meaning they should be classified into new genera based on a 70% threshold (Nicholson *et al.* 2020). Based on their estimated

placement on the GTDB tree, it is proposed that *F. sp.* CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* CRH, *F. sp.* GR2016-10, *F. sp.* CSZ, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* JLP, *F. sp.* MR2016-29, *F. sp.* MC2016-06, *F. sp.* ABG, *F. sp.* LC2016-12, *F. sp.* JRM, *F. sp.* KMS, *F. sp.* AJR, *F. sp.* ALJ2, all of their close relatives, and *F. johnsoniae* UW101^{T} should be transferred into the *Lycobacterium gen. nov.* and *F. sp.* AED, *F. sp.* HJJ and their close relatives, including *F. aquicola* DSM 100880^{T} should be transferred into the *Noelibacterium gen. nov*.

Background

Understanding how organisms are related is important. The names of these organisms should have meaning that reflect their characteristics. The name of a genus should create a clear image in one's mind of what that organism can do and not cause confusion as to what organism you are addressing. Therefore, groups must be sufficiently narrow to show the organisms will have similarities but be distinguishable from their neighboring groups. In the medical field, where there is a constant threat from the development of antibiotic-resistant strains of pathogens, knowing how organisms are related can help inform both sides of this issue: which organisms might be more prone to developing resistance, and which organisms are related to antibiotic producers that should be further studied to investigate their antibiotic potential. Especially in a world where emerging diseases are in the front of everyone's mind, it is important to strengthen a foundation upon which lifesaving research can continue to grow.

The present-day issues with bacterial taxonomy stem from the development of new, cost effective DNA sequencing technology, as well as a better understanding of the metrics that answer the question, "How different are they?". The traditional answer may be phenotypes, but what if the limited number of testable phenotypes are consistent across a specific phylum such as the *Bacteroidetes*? In these instances, where major testable phenotypes are similar or perceived to be the same, it becomes difficult to differentiate one species from another. For this reason, some genera have become "catch-all" or "default" groups, that is to say, novel species are classified into these groups due to having no defining characteristics to classify them elsewhere. Researchers are also not willing to spend time to check genus level comparisons, thinking that the genus placement is appropriate due to the similarity to the closest relatives. Only now, as more and more bacterial genomes are sequenced, are these "catch-all" groups revealed to be

polyphyletic and heterogenous, meaning they improperly include organisms that do not share a close common ancestor. This issue is exemplified in the genera of *Streptomyces, Bacillus,* and *Flavobacterium,* as *Streptomyces* and *Bacillus* each had over 200 species added to them between 2001 and 2018, while *Flavobacterium* added over 180 (Sant'Anna *et al*. 2020).

The *Flavobacterium* genus was originally named in 1923 to describe yellow-pigmented gram-negative, non-spore-forming rod-shaped bacteria that weakly produced acid from different carbohydrate sources (Bernardet *et al.* 1996). The type species for this genus is *Flavobacterium aquatile*, a species originally published as *Bacillus aquatilis* in 1889, highlighting the previously mentioned issues in taxonomy. The type species of a genus is the representative "standard" that all other organisms within the genus must be compared to. This idea of type also stands within species, as the type strain for a species is the representative to which all other strains must be compared. The *Flavobacterium* genus was restricted to those organisms that were non-motile, but this was only after it had acquired many poorly defined species, causing heterogeneity (Bernardet *et al.* 1996). This "Gordian Knot" is still being addressed today, and not only at the genus level, but on phylum levels of taxonomy as well. The *Bacteroidetes* still has issues because only one phenotypic difference delineates genera (Hahnke *et al*. 2016). This highlights the limitations of phenotype-based taxonomy, and other factors like horizonal gene transfer and convergent evolution that make using phenotypes to reflect evolutionary relationships difficult.

As genome sequences have become more readily available, increasing focus has been placed on different genome-based metrics. Methods like fatty acid and polar lipid analysis have been used to characterize bacteria, but like DNA-DNA hybridization (DDH) experiments, their low level of reproducibility has limited their use as a metric for comparison. DDH is a measurement of how well two strands of DNA from different organisms base pair with each

other, or hybridize. 16s rRNA gene sequence similarity has been used and recommended since the 1990s, though this measurement has low resolution at the species level due to high conservation and being a relatively short sequence. The 16s rRNA gene codes for the RNA found in the small subunit of the ribosome. The gene is highly conserved across different species due to its precise function, and too many mutations would result in dysfunctional ribosomes, eventually leading to cell death.

DDH was considered to be the gold standard for determining bacterial taxonomy but is now being replaced with new overall genome relatedness indices (OGRI) to address its low reproducibility and laborious experimental methods for physically measuring the parameter (Sant'anna *et al* 2020). Average Nucleotide Identity (ANI) has begun to replace DDH, and metrics like Average Amino Acid Identity (AAI) are also being used for genus level comparisons (Sant'Anna *et al* 2020). ANI measures genome similarity between two or more organisms by looking at each individual base in the genome and seeing if they are the same or different. Similarly, AAI does this but with amino acids, which are more likely to be conserved due to silent mutations in the genome having no impact on the resulting amino acid product after translation. ANI and AAI, while providing better resolution than 16s rRNA gene sequence similarity, result in one value for the comparison between genomes. These alignment-independent methods merely show the similarities without context, while alignment-based analyses, especially targeted at ubiquitous proteins, can provide much more detailed information on the relatedness of these organisms.

The Genome Taxonomy Database (GTDB) provides high resolution for comparisons across all levels of taxonomy with their method of using 120 concatenated, highly conserved protein sequences to construct a tree. The current version uses more than 30,000 dereplicated

genomes from the RefSeq database. The relative evolutionary distance used to determine relationships between organisms reflects both the level of similarity and the speed at which these organisms are evolving. Rigorous validation methods have shown the GTDB tree to be a robust and stable method, and their proposed taxonomy fixes many of the issues with polyphyletic branches of the current model (Parks *et al*. 2018). Though each of these methods have flaws and caveats of their own, using them together cover their weaknesses and provide a very strong portfolio of how bacteria are related.

For genera like *Flavobacterium*, however, the GTDB does not propose to split the genus. This leaves the genus very broad with over 250 widely varying species that are not sufficiently described by the species name. To address this problem, more presice gene level comparisons help determine what auxiliary genes are unique or shared within different groups, then leading to prediction of and testing for different phenotypes that could distinguish different clades. These distinguishing traits can then be used for groups like the *Flavobacterium* that are extremely broad. In this way, the genus can be divided into groups that share phylogenetically significant properties, each properly named so that the names depict the features of the organisms assigned to that phylogenetic group.

It is within this context, that we tested the hypothesis that *Flavobacterium* sp. CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* CRH, *F. sp.* GR2016-10, *F. sp.* CSZ, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* JLP, *F. sp.* MR2016-29, *F. sp.* MC2016-06, *F. sp.* ABG, *F. sp.* LC2016-12, *F. sp.* JRM, *F. sp.* KMS, *F. sp.* AJR, *F. sp.* ALJ2, *F. sp.* AED, and *F. sp.* HJJ, are all novel species. Organisms were recovered from frozen permanents and had their 16s rRNA gene PCR amplified and sequenced for identification. gDNA was extracted from those whose genomes were not yet sequenced and sent for

sequencing. The rpoB and 16s rRNA gene sequences were retrieved from the genome and used to create two phylogenetic trees. The genomes were then used to create overall genome relatedness indices (OGRI) tables to determine similarity to closest relatives. The novel organisms were characterized for phenotypic differences and compared to their closest relatives.

Methods

Isolation and identification of organisms

Flavobacterium sp. CNT (2020), *F. sp.* NLM (2015), *F. sp.* WLB (2017), *F. sp.* HTF (2017), *F. sp.* ANB (2019), *F. sp.* KMA (2013), *F. sp.* CRH (2015), *F. sp.* CSZ (2019), *F. sp.* KJJ, *F. sp.* JLP (2015), *F. sp.* ABG (2012), *F. sp.* JRM (2014), *F. sp.* KMS (2014), *F. sp*. AJR (2014), *F. sp.* ALJ2 (2020), *F. sp.* AED (2012), and *F. sp.* HJJ (2015) were isolated in January of their respective years from Loyalsock Creek, near Williamsport in north-central Pennsylvania. *F. sp.* LC2016-01, *F. sp.* LC2016-12, and *F. sp.* LC2016-23 were isolated the same creek in late March of 2016. *F. sp.* GR2016-10 was isolated from Graffius Run in Williamsport PA. *F. sp.* MR2016-29 was also isolated in late March of 2016 from Millers Run near Bruce Henry Park in Loyalsock Township, PA. *F. sp.* MC2016-06 was isolated in late March of 2016 from Mill Creek near Montoursville, PA. Samples were plated on nutrient agar (NA) media or R2A media and grown at room temperature. Well isolated colonies were selected, and pure cultures were obtained by streaking onto new plates, grown at 30°C for 2 days, then maintained in permanent stocks at -80°C with 20% glycerol. In August 2020, organisms were recovered by removing a bead from their respective stocks and rolling them on separate R2A plates. This was repeated in February 2021. The identified closest relatives were also recovered in the same fashion.

Many of the organisms already had their genomes sequenced previously, and *F. sp.* ANB, *F. sp.* CSZ, *F. sp.* HJJ, *F. sp.* JLP, *F. sp.* LC2016-12, and *F. sp.* MR2016-29 were without sequenced genomes. Genomic DNA (gDNA) of organisms identified as novel was isolated using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's instructions. Polymerase chain reaction (PCR) amplification of the 16s rRNA gene used 27f and 1492r primers (Weisburg *et al.*, 1991) and PCR product concentration was estimated using gel electrophoresis. Samples were sent to Genewiz for Sanger sequencing, and reads were assembled using CAP3 (Huang and Madan 1999). After assembly, 16s rRNA gene sequences for each organism were used in NCBI BLAST searches against the different databases and the top matches were recorded. The 16s rRNA gene sequences were also compared to previous sequences from past work to ensure the identity of the organisms.

Additionally, in preparation for depositing organisms to international culture collections, the identities of *F. sp.* CNT, *F. sp.* LC2016-01, *F. sp.* GR2016-10, *F. sp.* LC 2016-23, *F. sp.* MC2016-06, *F. sp.* KMA, and *F. sp.* ALJ2 needed to be confirmed. This process used the same PCR amplification of the 16s rRNA gene above. Instead of extracting the gDNA, however, cells were frozen and thawed twice to lyse cells and release the DNA to use as a template.

Genome Sequencing, Annotation, and Identification of Closest Relatives

After using the Qubit fluorometer to measure the gDNA concentration, samples for *F. sp.* ANB, *F. sp.* CSZ, *F. sp.* HJJ, *F. sp.* JLP, *F. sp.* LC2016-12, and *F. sp.* MR2016-29 were sent to the Microbial Genome Sequencing Center (MiGS) for sequencing using the Illumina NextSeq 550 platform. Reads were assembled with SPAdes in PATRIC (Davic *et al.* 2019), and

assemblies were uploaded to the RAST server for annotation (Aziz *et al*. 2008). Assembly is the process of taking the raw reads from the sequencing instrument and looking for areas of overlap. Annotation is the process of looking through an assembled genome and finding genes or potential transcripts and "annotating" them for future study. The assembled genomes were deposited to GenBank and the sequencing reads were uploaded to the Sequence Read Archive. These same steps were previously completed for all of the other novel species. After annotation, 16s rRNA gene sequences were isolated from the genome sequences using the SEED viewer (Overbeek *et al.* 2014) for every organism and used in NCBI BLAST searches against the nonredundant nucleotide collection and the 16s rRNA database for Refseq. The most similar results from the BLAST search were recorded for identifying the closest relatives. Sequences of these top hits were downloaded, aligned, and used in MEGA7 (Kumar *et al.* 2016) to create a 16s rRNA phylogenetic tree (Figure 4). In addition to the 16s rRNA gene sequence similarity, the RNA polymerase beta-subunit (rpoB) gene sequence was identified from the annotated genome and used in an NCBI BLAST search. Most similar sequences from the BLAST search were downloaded, aligned, and used to create a rpoB phylogenetic tree (Figure 5) in MEGA7(Kumar *et al.* 2016).

Overall Genome Relatedness Indices and GTDB Tree

After construction of the 16S rRNA gene sequence tree and rpoB tree, the organisms clustering with each of the novel species candidates were further compared using metrics such as Estimated DNA-DNA Hybridization (eDDH), Average Nucleotide Identity (ANI), and Average Amino Acid Identity (AAI). For eDDH calculations, the reference genome of the novel organism candidate and the closest relatives of the candidate were uploaded to the Genome-to-Genome

Distance Calculator from the Leibniz Institute (Meier-Kolthoff *et al*. 2013). Data outputs for each organism were saved and the values from formula two were reported. For ANI calculations**,** the novel species candidate and the closest relatives were used with the OrthoANI tool (Lee *et al.* 2016). For AAI calculations, the genomes of all organisms to be compared were uploaded to RAST. RAST's sequence-based comparison tool was then used with the novel species candidate as the reference and the closest relatives as the reference organisms. The results of this tool were exported and used in an AAI calculator available from http://lycofs01.lycoming.edu/~newman/AAI.html. The Genome Taxonomy Database (GTDB) (Parks et al. 2018) was used to place these organisms in the larger context of the genus *Flavobacterium*. The GTDB tree was edited in the NCBI Genome Workbench (NCBI Resource Coordinators 2018), using RefSeq accessions to locate organisms on the tree. The placement of the novel organisms on the tree was estimated using the closest relatives. Closest relatives were used to make Venn Diagrams and lists of unique and shared genes using RAST's sequence-based comparison tool (Aziz *et al.* 2008) and a tool developed by previous Newman lab students.

Phenotypic Characteristics

Using the cultures recovered from the Isolation and Identification of organisms step, single colonies were selected for each organism and plated onto new R2A plates. This process was repeated at least every two weeks to ensure organisms were freshly grown. For each organism, single colonies were used to test for the following phenotypes, or physical characteristics, and all methods and procedures were carried out as described in Reddy *et al.* (2007).

Resistance or susceptibility to different antibiotics was tested using the Kirby-Bauer disk diffusion assay. Organisms were grown overnight in liquid R2A medium, and 100 μL of inoculum was pipetted onto new R2A agar plates before being spread to cover the agar surface. Five-6 mm filter paper disks were used per plate with one of fifteen antibiotics added to each of the disks. These antibiotics were: amoxicillin, ampicillin, gentamycin, clavulanic acid, clavamox, clindamycin, chloramphenicol, chlortetracycline, erythromycin, kanamycin, nalidixic acid, penicillin, rifampicin, streptomycin, and tetracycline in amounts specified in Table 21. Plates were incubated for two days before zones of inhibition, areas of no growth around the disk, were measured.

Differential media, growth conditions, characteristics and biological activities were tested. Unless otherwise noted, all tests occurred with a two-day incubation period. Ability to grow in the presence of different salt concentrations was tested by plating each organism on R2A plates containing 1%, 2%, 3%, 4%, and 6% NaCl. These plates were compared to a normal R2A plate (0% NaCl) as a control. Incubation temperatures were tested by incubating inoculated R2A plates at 4°C, 20°C, 30°C, 37°C and 45°C, and the amount of growth was compared between all of them. Flexirubin pigments were tested for by dropping KOH onto thick areas of cell growth for each organism and seeing if a red color was observed. Glucose utilization and fermentation was tested using phenol red as a pH indicator for acid production and a Durham tube to capture gas. Cysteine desulfhydrase, indole production, motility, and Simmons Citrate were tested for using their agar slants while nitrate reduction and denitrification, and gelatinase activity were tested for using liquid media. For the gelatinase activity, organisms were incubated for five days as opposed to two. Presence of amylase, caseinase, esculin hydrolysis and growth in bile were

tested using solid media plates. Eosin methylene blue medium was also inoculated for each organism (Reddy *et al.* 2007).

Ninety-six different growth conditions were tested using Biolog Gen III plates, at least twice for each organism. Each organism was grown on the Biolog Universal growth medium + 5% sheep blood for two days. A small amount of inoculum was then collected and inoculated into IF-A broth to achieve a transmittance between 90% and 98% before the broth was pipetted into the ninety-six wells. Biolog GenIII plates were incubated for thirty-six hours, unless otherwise noted. Plates were read every fifteen minutes, and the results were exported from the Omnilog instrument after the final read was completed. After two plates were successfully read, the results were averaged (Supplemental Table1).

Results

Isolation and identification of organisms

Examples of recovery from frozen permanent stocks produced colonies shown in Figure 1. *F. sp.* LC2016-13 and *Flavobacterium reichenbachii* LMG 25512^{T} were not recovered due to issues with their frozen permanent stocks. All organisms had circular colonies except for *F. sp.* JLP, which had punctiform colonies. All colonies were convex and mucoid. *F. sp.* ANB (Figure 1A) was a darker, golden yellow, while all others were roughly the same yellow color. Contamination from one colony was well isolated on the *F. sp.* HJJ plate (Figure 1F). All plates were then restreaked for single colonies and used for the gDNA extraction. After extraction of gDNA and amplification of the 16s rRNA gene, samples were run on a gel (Figure 3 and 4) and concentrations of the extracted gDNA was measured using the Qubit Fluorometer. PCR product

amounts were estimated using the intensity of the bands after gel electrophoresis (Table 1). A small amount of PCR product was found in the negative control (-) well. PCR products of the twice frozen and thawed cells are shown in Figure 4. No PCR product was found in the negative control well. Sanger sequencing of the PCR products resulted in high quality reads, with traces of all read files having well-defined peaks and nucleotide calls (Figure 5). Reads were assembled into full contigs containing the 16s rRNA gene sequence for each organism, an example of which can be found in Figure 2. Between the 27f, 330f, and 1492r primers, full 16s rRNA genes sequences were assembled for all organisms.

Table 1. gDNA and PCR product concentrations

Flavobacterium sp.	[gDNA] (ng/µL)	[Estimated PCR product](ng/µL)
ANB	186	12
CSZ	340	25
HJJ	136	8
JLP	43.1	30
LC2016-12	400	10
MR2016-29	51.6	30

Table 2. Genome Assembly Details for novel *Flavobacterium* species

Flavobacterium sp.	ANB	CSZ	HJJ	JLP	LC2016-12	MR2016-29
Number of Contigs	48	74	92	28	38	28
Total length	5789321	6202203	5046692	45523663	5012773	4648293
Largest Contig	767906	531047	403808	1218236	1064498	1325413
N50	513529	228661	152496	441836	326950	427815
GC%	32.9	34.3	34.6	33.3	33.9	33.4
Average Coverage	83.2	91.1	170.8	171.8	113.7	168.6

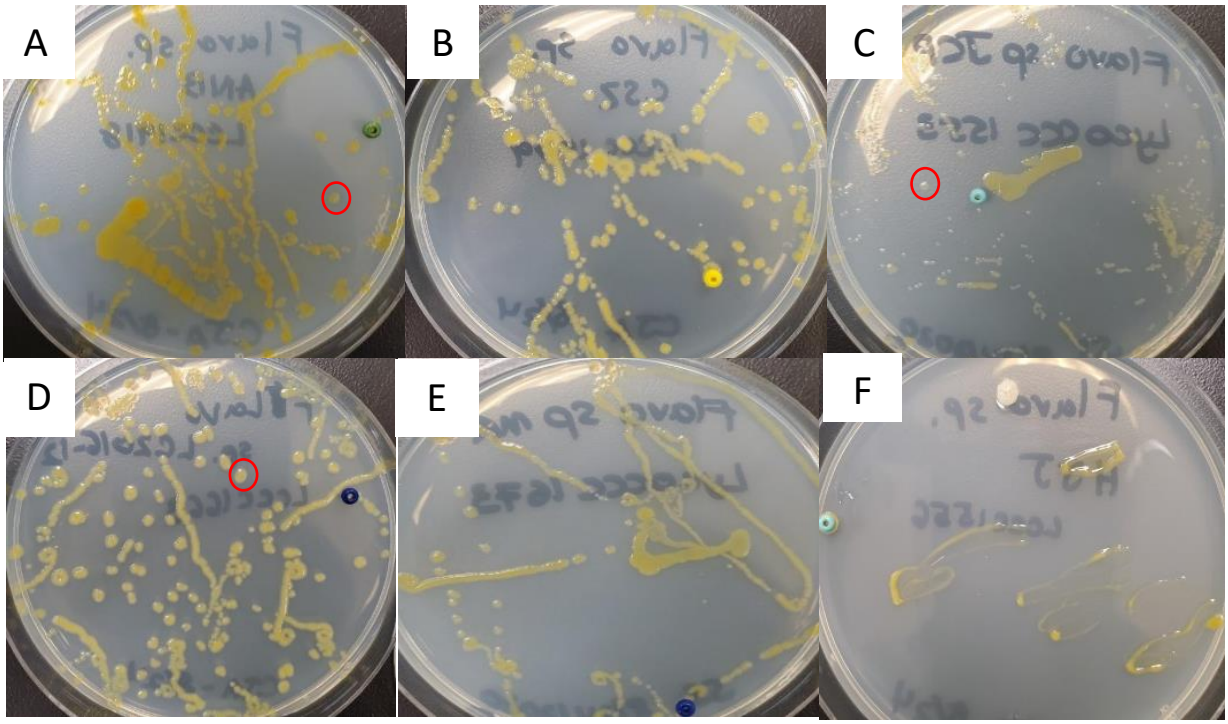

Figure 1. Photographs of recovery plate examples for: A) *Flavobacterium* sp. ANB; B) *F. sp.* CSZ; C) *F. sp.* HJJ; D) *F. sp.* JLP; E) *F. sp.* LC2016-12; and F) *F. sp.* MR2016-29. Well isolated colonies like those circled were selected and used for restreaking. A contaminant can be seen in F), but it is well isolated from the other yellow colonies, and there is no other evidence of contamination.

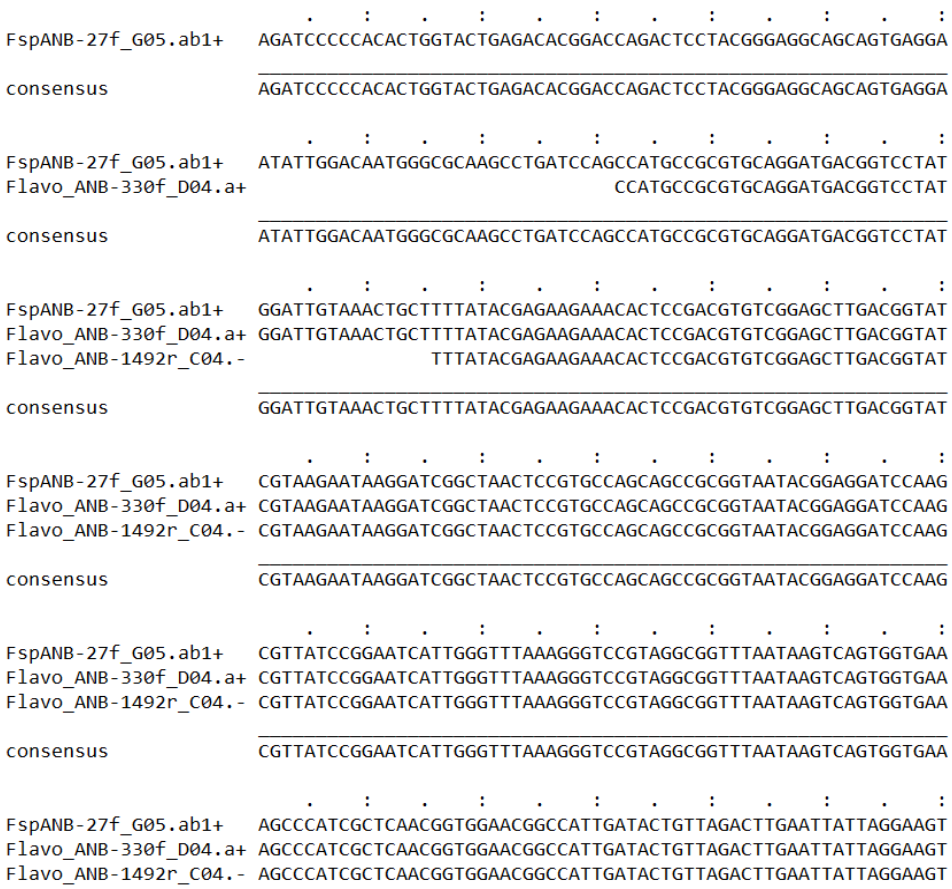

Figure 2. Example assembly of the 16s rRNA gene sequence of *F. sp. ANB*. The alternating dots serve as a shorthand for every five and every ten bases respectively. The consensus sequence shown under the line is the combination of the reads, which are listed to the left.

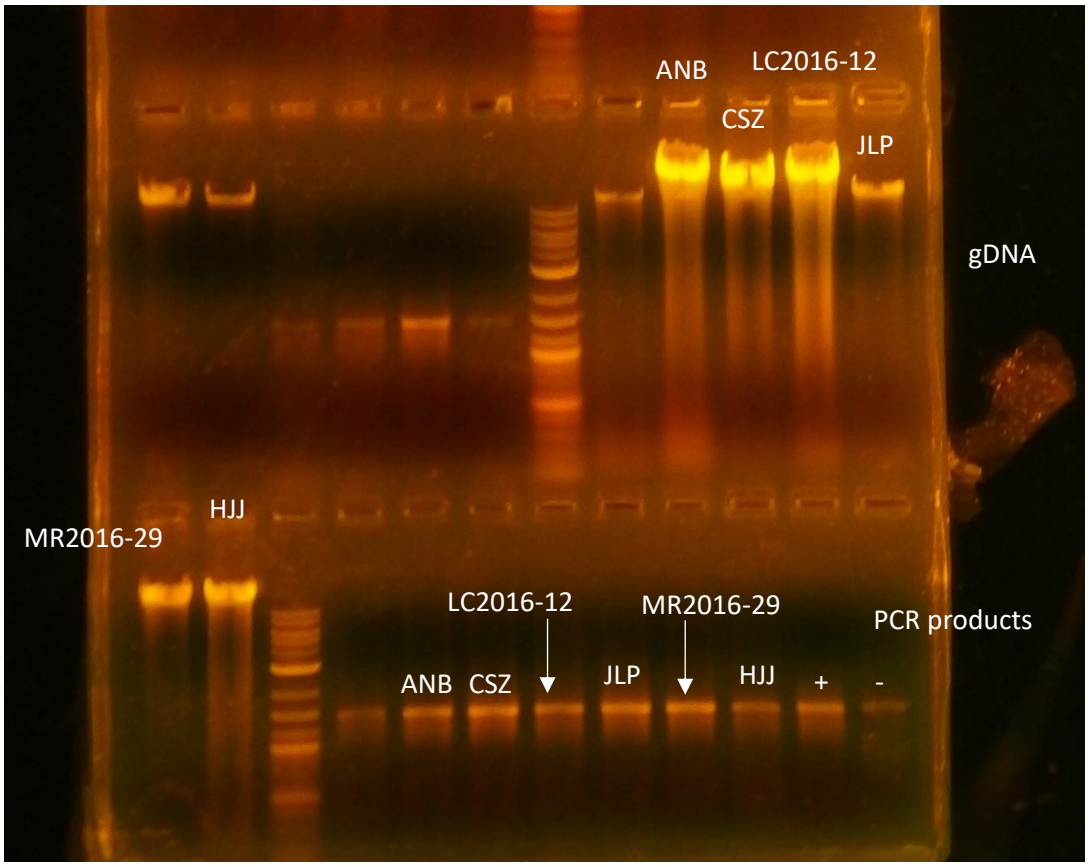

Figure 3. Gel electrophoresis of gDNA (top right and lower left) and 16s rDNA PCR products (lower right). (+) labels the positive control well and (–) labels the negative control well. All product wells (lower right) show a successful PCR.

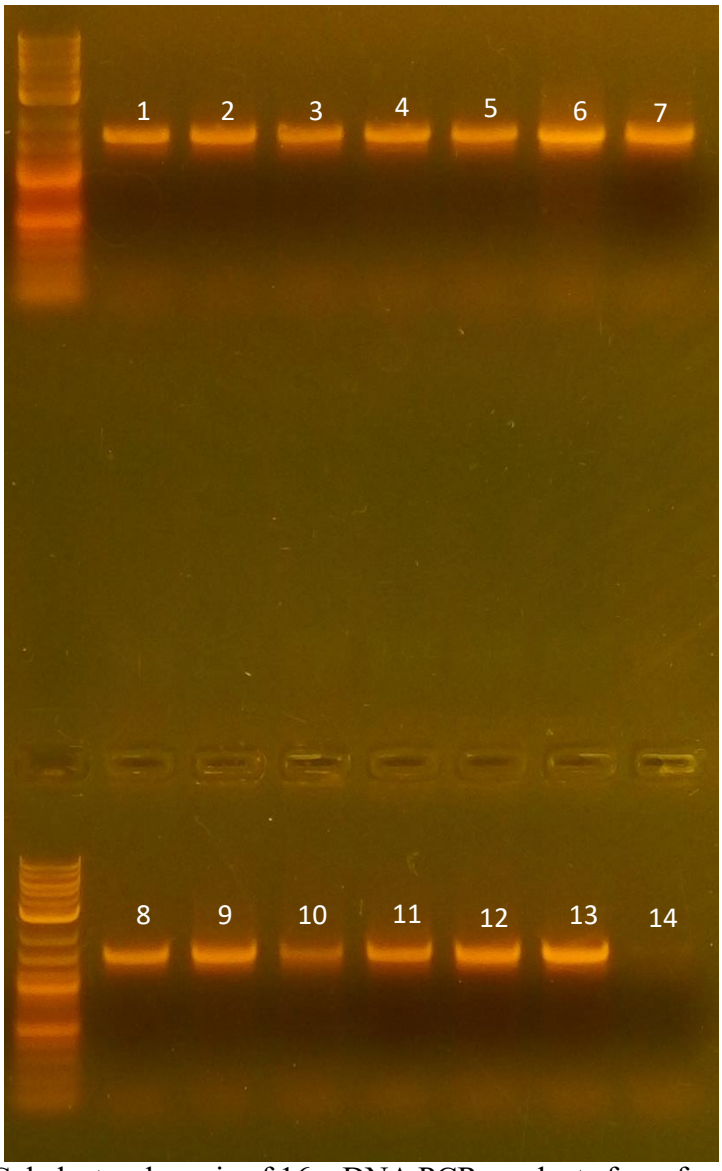

Figure 4. Gel electrophoresis of 16s rDNA PCR products from frozen/thawed cells. 1.) *F. sp.* CNT, 2.) *F. sp.* LC2016-01, 3.) *F. sp.* GR2016-10, 4.) *F. sp.* GR2016-10, 5.) *F. sp.* LC2016-23, 6.) *F. sp.* MC2016-06, 7.) *F. sp.* LC2016-12, 8.) *F. sp.* ANB, 9.) *F. sp.* KMA, 10.) *F. sp.* CSZ, 11.) *F. sp.* JLP, 12.) *F. sp.* ALJ2, 13.) *F. sp.* HJJ, 14.) negative control.

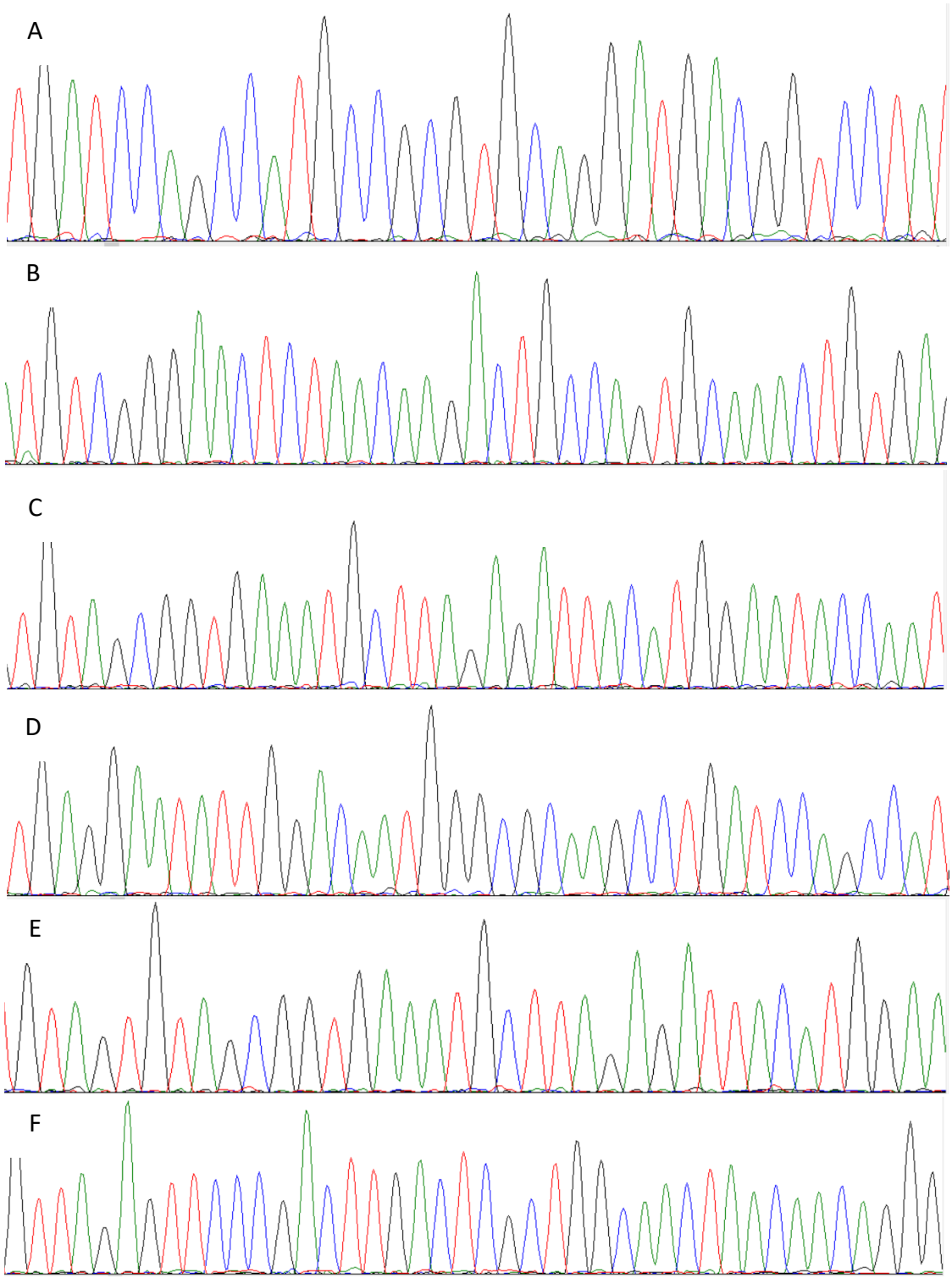

Figure 5. Examples of Sanger sequencing traces of the 16s rRNA gene: A) *Flavobacterium* sp. ANB; B) *F. sp.* CSZ; C) *F. sp.* HJJ; D) *F. sp.* JLP; E) *F. sp.* LC2016-12; and F) *F. sp.* MR2016-29. All sequences showed clean base calls.

16s rRNA gene sequences were used in BLAST (Zhang *et al.* 2000) searches against the non-redundant nucleotide collection and the 16s rRNA database using Refseq for each organism, all of which had the most similar sequences <99.5% similar. Most similar sequences were downloaded and used in MEGA 7 (Kumar *et al.* 2016) to create a neighbor-joining tree (Figure 6). The Neighbor-joining tree can be used to visualize similarities between different organisms. *F. sp.* CRH and *F. sp.* GR2016-10 clustered closely with *F. collinsii* 983-09 and *F.aquidurense* DSM18293^{T}. *F. sp.* KMA was also included in this cluster. *F. sp.* ANB and *F. sp.* LC2016-13 clustered extremely closely. Both, as well as *F. sp.* LC2016-12, clustered together and closely to *F. arucananum* LM-19-Fp. *F. sp.* NLM clustered closely to *F. sp.* LC2016-01 and *F. oncorhynchi* 631-08. *F. sp.* CSZ clustered closely to *F. pectinovorum* DSM 6368. *F. sp.* HJJ clustered with *F. gilvum* EM 1308. *F. sp.* CNT clustered on the same branch and closest to *F. ginsengisoli* DCY54. *F. sp.* JLP, *F. sp.* KJJ and *F. sp.* MR2016-29 clustered very close with each other and close to *F. psychrolimnae* LMG 22018^{T}. *F. sp.* ALJ2 clusted closest with *F. hydatis* NBRC 14958, while *F. sp.* LC2016-23 clustered on the same branch with *F. circumlabens* P5626^{T}. *F. sp.* MC2016-06 was isolated, but its closest clustering was to *F. glaciei* 0499. *F. sp.* JRM and AED were clustered together with its closest branching being to *F. sp.* MC2016-06.

gDNA was sent for sequencing at MiGS, resulting in reads that were then assembled using SPAdes in PATRIC (Davis *et al.* 2020). All genomes were assembled into 92 contigs or fewer with GC composition between 32% and 35% (Table 2). Genomes were uploaded to RAST (Aziz *et al.* 2008) for annotation, and the rpoB gene sequence from each organism was isolated and used in a BLAST (Zhang *et al.* 2000) search. The most similar sequences were downloaded and used in MEGA 7 (Kumar *et al.* 2016) to create a maximum-likelihood tree (Figure 7). The maximum-likelihood tree can be used to visualize similarities between different organisms using

this less conserved protein coding gene. *F. sp.* CNT clustered most closely with *F. ginsenosidimutans* THG01^{T}. *F. sp.* NLM was almost identical to *F. sp.* WLB, and clustered most closely to *F. plurextorum* CCUG60112^{T} and *F. sp.* HTF. *F. sp.* LC 2016-01 was close to *F. reichenbachii* LMG 25512^{T} and *F. hibernum* DSM 12661^{T}. *F. sp.* ANB clustered most with *F. sp.* LC2016-13, *F. sp.* KMA, and closely to *F. piscis* CCUG 60099^{T}. *F. sp.* CRH and *F. sp.* GR2016-10 were also close to *F. piscis* CCUG 60099^{T}. *F. sp.* CSZ clustered on this same branch, and most closely to *F. aquidurense* RC62 and *F. araucananum* DSM 24704. *F. sp.* LC 2016-23 was close to *F. cupreum* CCM 8825^{T}. *F. sp.* KJJ, *F. sp.* JLP, and *F. sp.* MR2016-29 clustered closely together and close to *F. hercynium* DSM 18292^{T}. *F. sp.* ABG and *F. sp.* MC2016-06 also clustered close to *F. hercynium* DSM 18292^{T}. *F. sp.* LC2016-12 clustered with *F. saccharophilum* DSM 1811. More distant from this group but still on the same branch, *F. sp.* JRM, *F. sp.* AJR, and *F. sp.* KMS all clustered extremely close to each other and close to *F. hydatis* DSM 2063T. *F. sp.* ALJ2 was also included in this grouping. *F. sp.* HJJ clustered on a completely different branch and most closely to *F. aquicola* DSM 100880. *F. sp.* AED clustered on this branch closest to *F. petrolei* Kopri-42^{T}.

Overall Genome Relatedness Indices and GTDB Tree

Organisms clustering closely to the novel species on the 16s rRNA phylogenetic tree (Figure 6) and the rpoB phylogenetic tree (Figure 7) were used for Genome wide measurements including eDDH, ANI, and AAI (Tables 3-19). When compared to their closest, validly named relative, all of the novel species had eDDH <70% and ANI <95.0% similar. Of the most similar ANI values to validly named type species, *F. sp.* CNT was most similar to *F. ginsenosidimutans* THG01^{T} (85.76%) (Table 3); *F. sp.* NLM was most similar to *F. johnsoniae* UW101^{T} (85.10%)

(Table 4); *F. sp.* HTF was most similar to *F. circumlabens* P5626^{T} (82.07%) (Table 5); *F. sp.* LC2016-01 was most similar to *F. resistens* DSM19382^{T} (87.32%) (Table 6); *F. sp.* ANB was most similar to *F. araucananum* DSM 24704 (83.36%) (Table 7); *F. sp.* KMA was most similar to *F. piscis* CCUG 60099^{T} (87.53%) (Table 8); *F. sp.* CRH was most similar to *F. aquidurense* DSM 18293^{T}(82.93%) (Table 9); *F. sp.* CSZ was most similar to *F. piscis* CCUG 60099 (89.21% ANI) (Table 10); *F. sp.* LC2016-23 was most similar to *F. cupreum* CCM 8825^{T} (90.15%) (Table 11); *F. sp.* KJJ was most similar to *F. saccharophilum* DSM 1811^{T} (83.54%) (Table 12); *F. sp.* MC2016-06 was most similar to *F. reichenbachii* DSM 21791^{T} (82.22%) (Table 13); *F. sp.* ABG was most similar to *F.* circumlabens P5626^{T} (82.65%) (Table 14); *F. sp.* LC2016-12 was most similar to *F. saccharophilum* DSM 1811 (90.91%) (Table 15). *F. sp.* JRM was most similar to *F. hydatis* ATCC29551^{T} (90.83%) (Table 16); *F. sp.* ALJ2 was most similar to *F. hydatis* ATCC29551^{T} (86.46%) (Table 17); *F. sp.* AED was most similar to *F. petrolei* Kopri-42^{T} (81.75%) (Table 18); and *F. sp.* HJJ was most similar to *F. aquicola* DSM 100880 (84.93%) (Table 19). Some novel species, however, were >95% similar in ANI and >70% similar in eDDH. When comparing ANI, *F. sp.* NLM and WLB were 99.78% similar (Table 4), *F. sp.* ANB and *F. sp.* LC2016-13 were 99.98% similar (Table 7), *F. sp.* CRH and *F. sp.* GR2016-10 were 97.89% similar (Table 9), and *F. sp.* KJJ was 98.27% similar to *F. sp.* JLP and 98.25% similar to *F. sp.* MR2016-29 (Table 12). Lastly, *F. sp.* JRM was 97.17% similar to *F. sp.* KMS and 97.07% similar to *F. sp.* AJR (Table 16).

The Genome Taxonomy Database Tree can be used to visualize how different organisms are related to each other in much higher resolution than a tree constructed from a single gene sequence. On the Genome Taxonomy Database Tree (Parks *et al*. 2018), *F. hercynium* DSM 18292^{T}, *F.* saccharophilum DSM 1811^{T}, *F. araucananum* DSM 24704^{T}, and *F.* piscis CCUG

60099^{T} clustered in the same branch as *F. johnsoniae* UW101^{T}. *F. aquicola* DSM 100880^{T} and *F. gilvum* EM1308^{T} clustered on a different branch from the *F. johnsoniae* group. The location of the novel species candidates was estimated based on their closest relative. Both groups were very far from the cluster containing *F. aquatile* LMG 4008^{T} (Supplemental Figure 1). A Venn diagram completed for *F. sp.* ANB (Figure 8) revealed 1921 genes were shared between all organisms, and *F. sp.* ANB had 961 unique genes. 1052 genes were shared in all organisms except for *F. aquatile* LMG 4008^{T}. The location of the novel species candidates was estimated based on their closest relatives. *F. sp.* CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* CRH, *F. sp.* GR2016-10, *F. sp.* CSZ, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* JLP, *F. sp.* MR2016-29, *F. sp.* MC2016-06, *F. sp.* ABG, *F. sp.* LC2016-12, *F. sp.* JRM, *F. sp.* KMS, *F. sp.* AJR, *F. sp.* ALJ2 were all estimated to be on the same branch as *F. johnsoniae* UW101^{T}. *F. sp.* AED and *F. sp.* HJJ were estimated to be on the same branch as *F. aquicola* DSM 100880^{T}.

Table 3. Overall Genome Relatedness Indices for *F. sp.* CNT

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. CNT	CNT	**	-	-	-	-
Flavobacterium ginsenosidimutans	THG 01^T	GCF_003254625	***	30.3	85.76	88.77
Flavobacterium resistens	DSM 19382^T	GCF_009674815	97.43%	26.3	82.95	84.81
Flavobacterium plurextorum	CCUG 60112^T	GCF_002217395	97.13%	25.7	82.76	83.71
Flavobacterium johnsoniae	UW101^T	GCF_000016645	97.21%	26.2	82.69	84.79
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	94.74%	19.6	73.53	67.46

**Accession not yet available from NCBI.

*** 16s rRNA sequence unavailable for proper comparison

Table 4. Overall Genome Relatedness Indices for *F. sp.* NLM

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. NLM	NLM	GCF_002210235	-	-	-	-
Flavobacterium sp. WLB	WLB	GCF_003070725	99.93%	98.8	99.78	99.76
Flavobacterium johnsoniae	UW101^T	GCF_000016645	97.22%	29.9	85.1	87.01
Flavobacterium plurextorum	CCUG 60112^T	GCF_002217395	96.59%	27.7	83.64	84.73
Flavobacterium resistens	DSM 19382^T	GCF_009674815	97.57%	27.1	83.16	85.06
Flavobacterium ginsenosidimutans	THG 01^T	GCF_003254625	97.34%	26.0	82.46	84.33
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	94.93%	20.1	73.23	67.47

Table 5. Overall Genome Relatedness Indices for *F. sp.* HTF

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. HTF	HTF	GCF_003105115	-	-	-	-
Flavobacterium circumlabens	P5626^T	GCF_004345565	97.70%	25.8	82.07	85.02
Flavobacterium reichenbachii	DSM 21791^T	GCF_002217435	97.87%	25.6	81.82	84.02
Flavobacterium cupreum	CCM 8825^T	GCF_003996965	97.90%	25.5	81.73	85.50
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.62%	24.5	80.92	82.71
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.46%	19.9	72.76	67.35

Table 6. Overall Genome Relatedness Indices for *F. sp.* LC2016-01

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. LC2016-01	LC2016-01	GCF_009711135	-	-	-	-
Flavobacterium resistens	DSM 19382^T	GCF_009674815	97.63%	32.7	87.32	91.22
Flavobacterium denitrificans	DSM 15936^T	GCF_000425445	96.34%	29.5	85.35	89.33
Flavobacterium plurextorum	CCUG 60112^T	GCF_002217395	97.47%	26.1	82.63	84.13
Flavobacterium johnsoniae	UW101^T	GCF_000016645	97.81%	26.0	82.5	84.65
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.46%	20.0	73.72	67.37

Table 7. Overall Genome Relatedness Indices for *F. sp.* ANB

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flacobacterium sp. ANB	ANB	GCF_015350765	-	-	-	-
Flavobacterium sp. LC2016-13	LC2016-13	GCF_009707955	100.00%	99.8	99.98	99.99
Flavobacterium frigidimaris	DSM 15937^T	GCF_002217275	98.24%	26.7	83.09	84.60
Flavobacterium saccharophilum	DSM 1811^T	GCF_900142735	98.33%	26.9	83.02	84.10
Flavobacterium aquidurense	DSM 18293^T	GCF_002217195	99.06%	26.9	82.97	84.50
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.45%	25.0	81.65	83.03
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.92%	19.8	73.54	67.82

Table 8. Overall Genome Relatedness Indices for *F. sp.* KMA

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. KMA	KMA	**	-	-	-	-
Flavobacterium piscis	CCUG 60099^T	GCF_001686925	99.52%	33.5	87.53	90.37
Flavobacterium aquidurense	DSM 18293^T	GCF_001404985	99.60%	30.2	85.84	83.77
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	98.49%	28.3	84.06	86.16
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.83%	24.8	81.18	83.18
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	96.19%	20.2	73.25	67.80

**Accession not yet available from NCBI.

Table 9. Overall Genome Relatedness Indices for *F. sp.* CRH

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. CRH	CRH	**	-	-	-	-
Flavobacterium sp. GR2016-10	GR2016-10	**	99.93%	80.9	97.89	98.29
Flavobacterium aquidurense	DSM 18293^T	GCF_001404985	98.59%	26.6	82.93	83.62
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	97.27%	26.5	82.93	83.97
Flavobacterium saccharophilum	DSM 1811^T	GCF_900142735	98.68%	26.8	82.93	84.22
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.29%	24.8	81.30	82.68
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.40%	19.8	73.25	67.82

**Accession not yet available from NCBI.

Table 10. Overall Genome Relatedness Indices for *F. sp.* CSZ

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. CSZ	CSZ	GCF_015350725	-	-	-	-
Flavobacterium piscis	CCUG 60099^T	GCF_001686925	99.52%	37.6	89.21	91.89
Flavobacterium aquidurense	DSM 18293^T	GCF_002217195	98.66%	31.4	85.91	83.98
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	97.40%	27.7	83.55	85.67
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.49%	24.4	80.81	82.82
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.77%	19.5	72.89	67.54

Table 11. Overall Genome Relatedness Indices for *F. sp.* LC2016-23

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. LC2016-23	LC2016-23	GCF_009674775	-	-	-	-
Flavobacterium cupreum	CCM 8825^T	GCF_003996965	98.62%	39.8	90.15	93.01
Flavobacterium circumlabens	P5626^T	GCF_004345565	99.54%	38.5	89.65	92.57
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	97.40%	24.4	81.74	83.07
Flavobacterium johnsoniae	UW101^T	GCF_000016645	95.76%	23.3	79.59	82.41
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.20%	19.1	72.35	68.02

Table 12. Overall Genome Relatedness Indices for *F. sp.* KJJ

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. KJJ	KJJ	GCF_000708595	-	-	-	-
Flavobacterium sp. JLP	JLP	GCF_015350755	99.87%	85	98.27	98.51
Flavobacterium sp. MR2016-29	MR2016-29	GCF_015350695	99.80%	85	98.25	98.27
Flavobacterium saccharophilum	DSM 1811^T	GCF_900142735	98.12%	27.1	83.54	86.11
Flavobacterium frigidimaris	DSM 15937^T	GCF_002217275	97.12%	26.9	83.36	84.97
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	96.72%	26.7	83.08	85.09
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.33%	25.0	81.65	83.82
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.33%	19.7	73.17	67.84

Table 13. Overall Genome Relatedness Indices for *F. sp.* MC2016-06

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. MC2016-06	MC2016-06	GCF_009711165	-	-	-	-
Flavobacterium reichenbachii	DSM 21791^T	GCF_002217435	97.34%	25.8	82.22	83.91
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	97.54%	25.6	81.87	83.06
Flavobacterium frigidimaris	DSM 15937^T	GCF_002217275	98.04%	25.4	81.59	83.13
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.55%	24.5	80.60	81.59
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	96.78%	19.5	73.40	67.72

Table 14. Overall Genome Relatedness Indices for *F. sp.* ABG

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. ABG	ABG	GCF_001027725	-	-	-	-
Flavobacterium circumlabens	P5626^T	GCF_004345565	***	26.6	82.65	84.65
Flavobacterium cupreum	CCM 8825^T	GCF_003996965	***	25.8	81.90	84.23
Flavobacterium hibernum	DSM 12611^T	GCF_000832125	***	24.8	80.81	82.10
Flavobacterium johnsoniae	UW101^T	GCF_000016645	***	23.5	79.24	80.58
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	***	19.9	72.74	67.49

*** 16s rRNA sequence unavailable for proper comparison

Table 15. Overall Genome Relatedness Indices for *F. sp.* LC2016-12

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. LC2016-12	LC2016-12	GCF_015351595	-	-	-	-
Flavobacterium saccharophilum	DSM 1811^T	GCF_900142735	98.54%	42.3	90.91	93.29
Flavobacterium frigidimaris	DSM 15937^T	GCF_002217275	98.03%	36.2	88.69	84.25
Flavobacterium piscis	CCUG 60099^T	GCF_001686925	98.36%	26.8	82.92	84.27
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.45%	24.5	81.15	82.74
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.66%	19.5	73.28	67.66

Table 16. Overall Genome Relatedness Indices for *F. sp.* JRM

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. JRM	JRM	GCF_000812985	-	-	-	-
Flavobacterium sp. KMS	KMS	GCF_000813005	98.94%	74.2	97.17	97.77
Flavobacterium sp. AJR	AJR	GCF_002156885	99.93%	73.5	97.07	97.39
Flavobacterium hydatis	ATCC 29551^T	GCF_000737695	96.99%	41.6	90.83	93.10
Flavobacterium saccharophilum	DSM 1811^T	GCF_900142735	97.85%	22.6	78.33	78.76
Flavobacterium hercynium	DSM 18292^T	GCF_002217285	97.07%	22.4	77.68	76.80
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.09%	21.8	77.32	77.13
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.59%	20.1	73.32	67.86

Table 17. Overall Genome Relatedness Indices for *F. sp.* ALJ2

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. ALJ2	ALJ2	GCF_015594725	-	-	-	-
Flavobacterium hydatis	ATCC 29551^T	GCF_000737695	97.54%	32.1	86.46	88.59
Flavobacterium saccharophilum	DSM 1811^T	GCF_900142735	98.47%	22.3	78.15	78.68
Flavobacterium hercynium	DSM 18292^T	GCF_002217285	97.87%	21.9	77.63	77.36
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.15%	21.5	77.30	77.18
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	96.37%	19.3	73.25	68.80

Table 18. Overall Genome Relatedness Indices for *F. sp.* AED

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. AED	AED	GCF_000812945	-	-	-	-
Flavobacterium petrolei	Kopri-42^T	GCF_003314435	96.40%	25.3	81.75	82.73
Flavobacterium granuli	DSM 17797^T	GCF_900129705	96.95%	22.2	78.44	78.82
Flavobacterium hydatis	ATCC 29551^T	GCF_000737695	95.83%	21.6	77.22	75.91
Flavobacterium johnsoniae	UW101^T	GCF_000016645	95.42%	20.9	75.48	73.64
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	95.00%	19.6	73.78	69.71

Table 19. Overall Genome Relatedness Indices for *F. sp.* HJJ

Genus species	Strain	Accession	16s rRNA	eDDH	ANI	AAI
Flavobacterium sp. HJJ	HJJ	GCF_015351475	-	-	-	-
Flavobacterium aquicola	DSM 100880^T	GCF_003385115	97.43%	30.7	84.93	86.19
Flavobacterium gilvum	EM1308^T	GCF_001761465	97.42%	24.5	80.14	81.80
Flavobacterium petrolei	Kopri-42^T	GCF_003314435	95.99%	21.1	75.69	74.55
Flavobacterium johnsoniae	UW101^T	GCF_000016645	96.75%	21.1	75.21	72.61
Flavobacterium aquatile	LMG 4008^T	GCF_000757385	94.80%	19.2	72.57	67.76

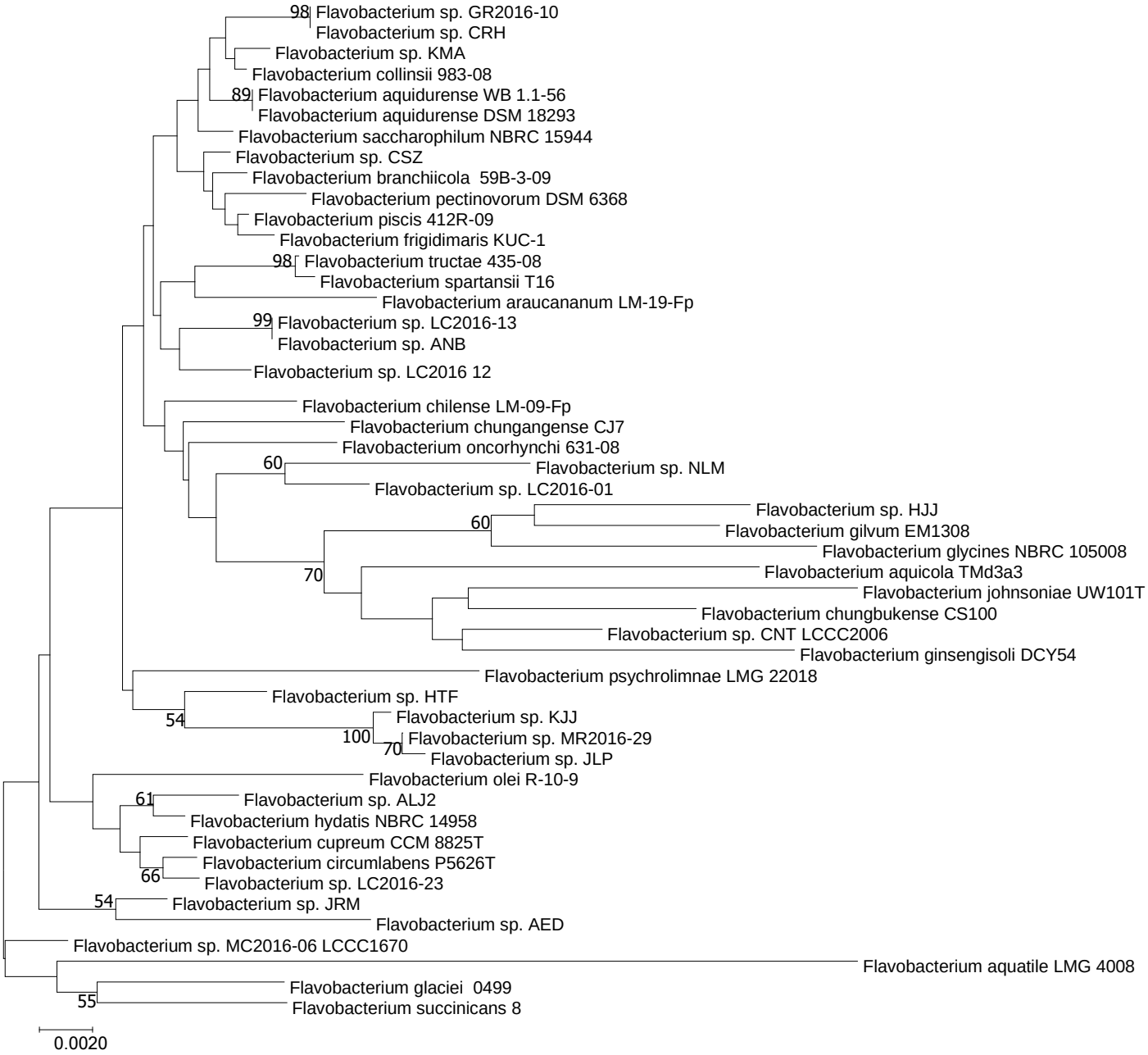

Figure 6. 16s rRNA Gene Similarity Tree. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 0.29924310 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura 1980) and are in the units of the number of base substitutions per site. The analysis involved 49 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1143 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al* 2016).

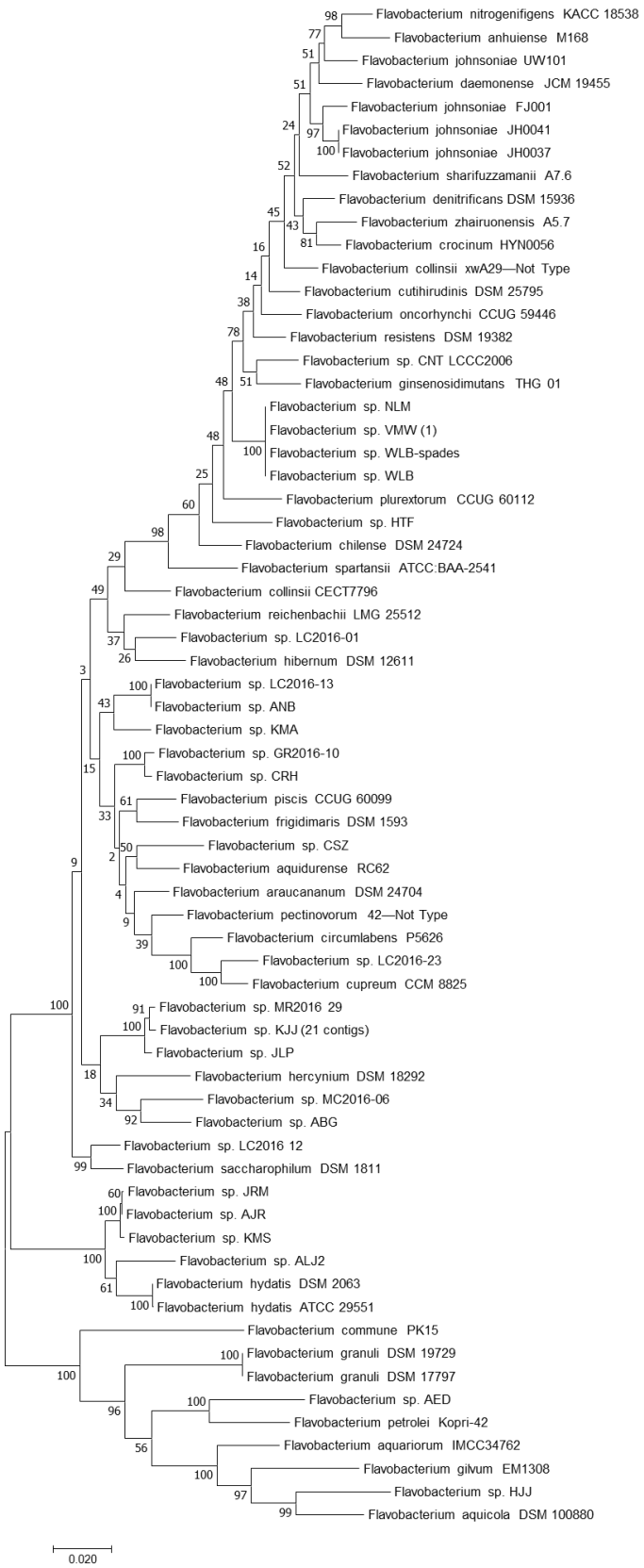

Figure 7. rpoB Sequence Similarity Tree

The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura 1980). The tree with the highest log likelihood (-34848.54) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. 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 Maximum Composite Likelihood (MCL) approach, 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. The analysis involved 66 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 3813 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar 2016).

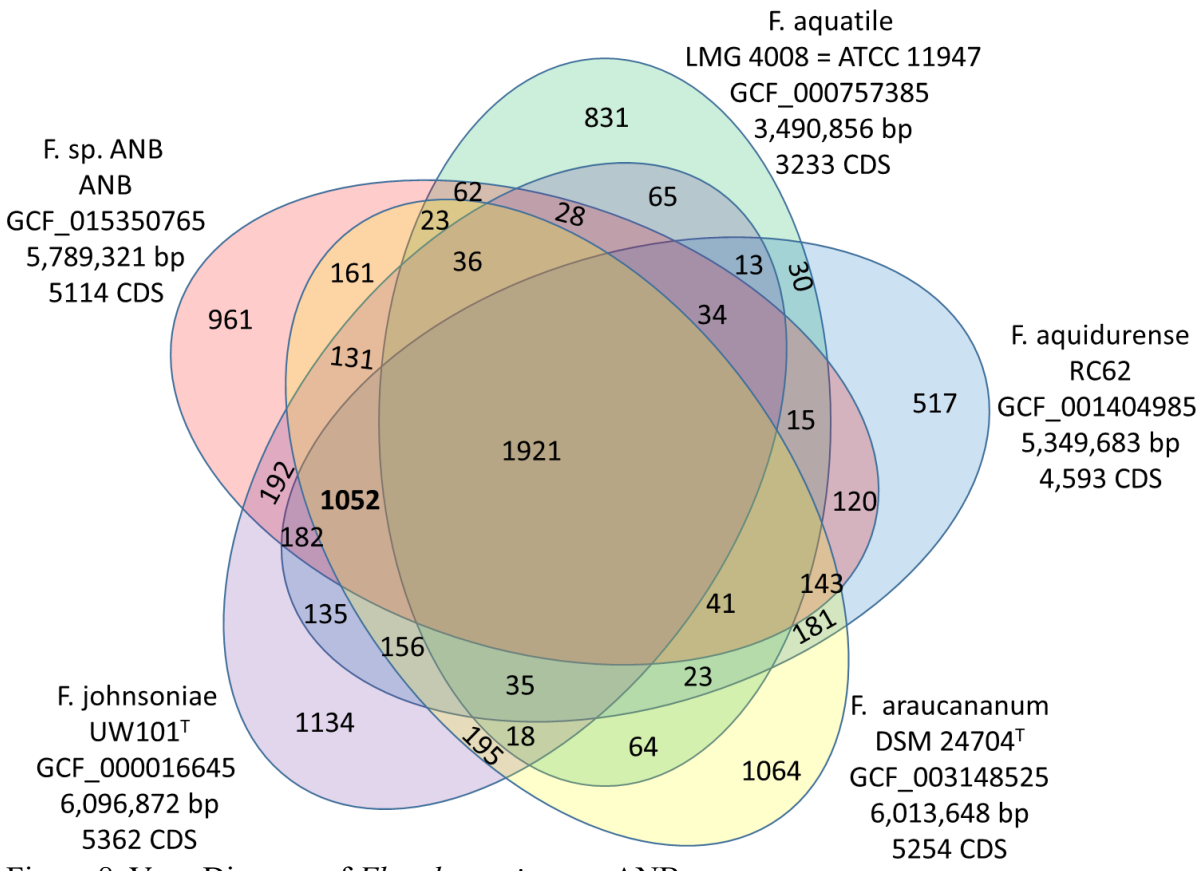

Figure 8. Venn Diagram of *Flavobacterium* sp. ANB.

Organisms chosen for this Venn Diagram were based on closest relatives, placement on the GTDB Tree, and the type representative of *Flavobacterium.* 1921 genes were shared between all organisms, 961 genes were unique to *F. sp.* ANB, and 1052 genes were shared between all organisms except for *F. aquatile* LMG 4088^{T}. These 1052 genes are indicative of the genes unique to the new genus.

Phenotypic Results

Phenotypic tests were run in parallel to the closest relatives of the novel species, but only the results of the novel organisms are described here. Results of the reference organisms are in Tables 20-23. None of the organisms were able to grow at 4°C, 37°C, 45°C, or at 3% NaCl, 4% NaCl, or 6% NaCl. All organisms but *F. sp.* KMS, *F. sp.* AED, and *F. sp.* HJJ were able to grow at 1% NaCl. All organisms could grow on R2A and Nutrient Agar plates, though *F. sp. HJJ* grew slowly. Weak growth occurred at 2% NaCl for *F. sp.* KMA, *F. sp.* CSZ, *F. sp.* MC2016-06, *F. sp.* LC2016-12, *F. sp.* JRM, and *F. sp.* AJR, while *F. sp.* ALJ2 grew well. All other organisms did not grow at 2% NaCl. All were positive for glucose utilization, but negative for fermentation. Only *F. sp.* KJJ and *F. sp.* HJJ did not exhibit amylase activity. *F. sp.* LC 2016-01, *F. sp.* KMA, *F. sp.* MC2016-06, and *F. sp.* HJJ showed weak caseinase activity, and *F. sp.* ALJ2, *F. sp.* AED, *F. sp.* KJJ, *F. sp.* MR2016-29, and *F. sp.* JLP were negative for caseinase. All were positive for esculin hydrolysis except for *F. sp.* ABG and weak activity in *F. sp.* AED and *F. sp.* HJJ. Only *F. sp.* KJJ displayed motility in the SIM agar. All organisms showed gelatinase activity except for *F. sp.* HJJ, and *F. sp.* ANB and *F. sp.* AED which were weak for gelatinase activity. Most organisms were able to grow in bile, with *F. sp.* GR2016-10, *F. sp.* CRH, and *F. sp.* MC2016-06 growing weakly and *F. sp.* AED and *F. sp.* HJJ not growing at all. On the EMB plates, *F. sp.* HJJ did not grow while *F. sp.* AED, MC2016-06 and *F. sp.* ANB grew weakly (Table 20). All biolog results can be found in Supplemental Table 1.

All organisms and reference organisms were tested for antibiotic resistance (Table 21). A zone of inhibition of 15 mm or larger was considered sensitive, while less than 15mm was considered resistant. *F. sp.* CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* GR2016-10, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* MR2016-29, *F. sp.*

JLP, *F. sp.* AJR, and *F. sp.* KMS, had zones of inhibition less than 15mm in radius for all antibiotics. Generally, however, all organisms had a zone of inhibition for tetracycline and nalidixic acid. Only *F. sp.* AED was affected by penicillin, while all others had no zone of inhibition. *F. sp.* AED was resistant to cephalexin, chloramphenicol, gentamycin, kanamycin, and penicillin, but sensitive to all other antibiotics. Ampicillin, cephalexin, chloramphenicol, and clavamox was resisted by all organisms except *F. sp.* AED. Kanamycin was resisted by all organisms and Sulfamethoxazole with Trimethoprim was resisted by all organisms except *F. sp.* AED.

Some organisms displayed gliding motility or spread out on the plate as they grew. This occurred on R2A media but not higher nutrient dense growth media like NA or TSBA. *F. sp.* NLM, *F. sp.* WLB, and *F. sp.* HTF all displayed gliding motility if grown at 30°C, but not at 22°C. Conversely, *F. sp.* CSZ, *F. sp*. JRM, *F. sp.* AJR, and *F. sp*. KMS, displayed gliding motility at 22°C but not at 30°C. *F. hydatis* DSM 2063^{T} and *F. hercynium* LMG 26768^{T} displayed a similar pattern of motility as *F. sp.* CSZ. Additionally, all organisms that were motile had an iridescent coloration to the leading edge of their growth. This iridescence was also observed on the colonies of *F. aquicola* DSM100880^{T}.

Table 20. Phenotypic Characteristics

	1	2	3	4	5	6	7	8	9	10
Gliding motility	-	+ at 30°C	+ at 30°C	-	-	+ at 30°C	+ at 30°C	-	+ at 20°C	-
Flexirubin Pigments	+	+	+	+	+	+	+	+	+	+
20°C	+	+	+	+	+	+	+	+	+	+
30°C	+	+	+	+	+	+	+	+	+	+
1% NaCl	+	+	+	+	+	+	+	+	+	+
2% NaCl	-	-	-	-	-	w	-	-	-	-
glucose utilization	+	+	+	+	+	+	+	+	+	+
amylase	+	+	+	+	+	+	+	+	+	+
caseinase	+	+	+	+	w	+	+	w	+	+
Esculin	+	+	+	+	+	+	+	+	+	+
nitrate reduction	-	+	+	w	-	+	+	+	+	-
denitrification	-	-	-	-	-	-	-	+	+	-
motility (SIM)	-	-	-	-	-	-	-	-	-	-
gelatinase	+	+	+	+	+	+	+	+	+	w
EMB	+	+	+	+	-	+	+	+	+	w
Bile	+	+	+	+	+	+	+	+	+	+

1. F. sp. CNT, 2. *F. sp.* NLM, 3. *F. sp.* WLB, 4. *F. resistens* DSM 19382^{T}, *5. Flavobacterium denitrificans* DSM 15936^{T}, *6. Flavobacterium johnsoniae* UW-101^{T}, *7. F. sp.* HTF, 8. *F. sp.* LC2016-01, 9. *Flavobacterium hibernum DSM 12611^{T}*, 10. *F. sp.* ANB. + Positive, -Negative, w Weak

Table 20 continued. Phenotypic Characteristics

	11	12	13	14	15	16	17	18	19	20
Gliding motility	-	-	-	-	+ at 20°C but not 30°C	-	-	-	-	-
Flexirubin Pigments	+	+	+	+	+	+	-	+	+	+
20°C	+	+	+	+	+	+	+	+	+	+
30°C	+	-	-	+	+	+	+	+	+	+
1% NaCl	+	+	+	+	+	+	+	+	+	+
2% NaCl	w	-	-	w	w	w	-	-	-	-
glucose utilization	+	+	+	+	+	+	+	+	+	+
amylase	+	+	+	+	+	w	+	-	+	+
caseinase	w	+	+	w	+	w	+	-	-	-
Esculin	+	+	+	+	+	+	+	+	+	+
nitrate reduction	w	+	-	-	-	w	w	w	w	-
denitrification	-	-	-	-	-	-	-	-	-	-
motility (SIM)	-	-	-	-	-	-	-	+	-	-
gelatinase	+	+	+	+	+	+	+	+	+	+
EMB	+	+	+	+	+	+	+	+dark	+	+dark
Bile	+	w	w	w	+	+	+	+	+	+

11. F. sp. KMA, *12. F. sp.* GR2016-10, 13. *F. sp.* CRH, 14. *Flavobacterium frigidimaris* DSM 15937^{T},15. *F. sp*. CSZ, 16. *Flavobacterium aquidurense* KCTC 22840^{T}, 17. *F. sp.* LC2016-23, 18. *F. sp.* KJJ, 19. *F. sp.* MC2016-29, 20. *F. sp.* JLP.+ Positive, -Negative, w Weak

Table 20 continued. Phenotypic Characteristics

	21	22	23	24	25	26	27	28	29	30
Gliding motility	+ at 20°C but not 30°C	-	-	-	-	+ at 20°C but not 30°C	+ at 20°C but not 30°C	+ at 20°C but not 30°C	+ at 20°C but not 30°C	-
Flexirubin Pigments	+	+	+	+	+	+	+	+	+	+
20°C	+	+	+	+	+	+	+	+	+	+
30°C	+	-	-	+	+	+	+	+	+	w
1% NaCl	+	+	+	+	+	+	+	-	+	+
2% NaCl	w	w	-	w	w	w	w	-	w	+
glucose utilization	+	+	+	+	+	+	+	+	+	+
amylase	+	+	+	+	+	+	+	+	+	+
caseinase	+	w	+	+	+	+	+	+	w	-
Esculin	+	+	-	+	-	+	+	+	+	+
nitrate reduction	w	+	w	-	-	+	+	w	+	+
denitrification	-	-	-	-	-	+	+	-	-	-
motility (SIM)	+	-	-	-	-	-	-	-	-	-
gelatinase	w	+	+	+	+	+	+	+	+	+
EMB	+	w	+	+	w	+dark	+dark	+dark	+	+
Bile	+	w	+	+	-	+	+	+	+	+

21. *Flavobacterium hercynium* LMG 26768^{T}, 22. *F. sp.* MC2016-06, 23. *F. sp.* ABG, 24. *F. sp.* LC2016-12, 25. *Flavobacterium saccharophilum* LMG 8384^{T}, 26. *F. sp.* JRM, 27. *F. sp.* AJR, 28. *F. sp.* KMS, 29. *Flavobacterium hydatis* DSM 2063^{T}, 30. *F. sp.* ALJ2. + Positive, -Negative, w Weak

Table 20 continued. Phenotypic Characteristics

	21	22	23	24	25	26	27	28	29	30
Gliding motility	+ at 20°C but not 30°C	-	-	-	-	+ at 20°C but not 30°C	+ at 20°C but not 30°C	+ at 20°C but not 30°C	+ at 20°C but not 30°C	-
Flexirubin Pigments	+	+	+	+	+	+	+	+	+	+
20°C	+	+	+	+	+	+	+	+	+	+
30°C	+	-	-	+	+	+	+	+	+	w
1% NaCl	+	+	+	+	+	+	+	-	+	+
2% NaCl	w	w	-	w	w	w	w	-	w	+
glucose utilization	+	+	+	+	+	+	+	+	+	+
amylase	+	+	+	+	+	+	+	+	+	+
caseinase	+	w	+	+	+	+	+	+	w	-
Esculin	+	+	-	+	-	+	+	+	+	+
nitrate reduction	w	+	w	-	-	+	+	w	+	+
denitrification	-	-	-	-	-	+	+	-	-	-
motility (SIM)	+	-	-	-	-	-	-	-	-	-
gelatinase	w	+	+	+	+	+	+	+	+	+
EMB	+	w	+	+	w	+dark	+dark	+dark	+	+
Bile	+	w	+	+	-	+	+	+	+	+

31. *Flavobacterium granuli* DSM 19729^{T}, 32. *F. sp.* AED, 33. *F. sp.* HJJ, 34. *Flavobacterium aquicola* DSM 100880^{T}, 35. *Flavobacterium aquatile* LMG 4008^{T}. + Positive, -Negative, w Weak

Table 21. Antibiotic Resistance

	1	2	3	4	5	6	7	8
Ampicillin (10µg)	R	R	R	R	R	R	R	R
Cephalexin (30 µg)	R	R	R	R	R	R	R	R
Chloramphenicol (30ug)	R	R	R	R	R	R	R	R
Ciprofloxacin (5 µg)	R	R	R	R	R	R	R	R
Clavamox (10µgClav+20µgAmox)	R	R	R	R	R	R	R	R
Clindamycin (2 µg)	R	R	R	R	R	R	R	R
Erythromycin (15µg)	R	R	R	R	R	R	R	R
Gentamycin (120µg)	R	R	R	R	R	R	R	R
Kanamycin (30µg)	R	R	R	R	R	R	R	R
Naladixic Acid (30µg)	R	R	R	R	R	S	R	R
Penicillin (10µg)	R	R	R	R	R	R	R	R
Rifampicin (5µg)	R	R	R	R	R	R	R	R
Streptomycin (10µg)	R	R	R	R	R	R	R	R
Sulfamethoxazole+ Trimethoprim	R	R	R	R	R	R	R	R
Tetracyclin (30µg)	R	R	R	R	R	S	R	R

1. F. sp. CNT, 2. *F. sp.* NLM, 3. *F. sp.* WLB, 4. *F. resistens* DSM 19382^{T}, *5. Flavobacterium denitrificans* DSM 15936^{T}, *6. Flavobacterium johnsoniae* UW-101^{T}, *7. F. sp.* HTF, 8. *F. sp.* LC2016-01, 9. *Flavobacterium hibernum DSM 12611^{T}*, 10. *F. sp.* ANB. R, Resistant, S, Sensitive, - Data not collected.

Table 21 Continued. Antibiotic Resistance

	9	10	11	12	13	14	15	16
Ampicillin (10µg)	R	R	R	R	R	R	-	R
Cephalexin (30 µg)	R	R	R	R	R	R	-	R
Chloramphenicol (30ug)	R	R	R	R	R	R	-	R
Ciprofloxacin (5 µg)	R	R	R	R	S	R	-	R
Clavamox (10µgClav+20µgAmox)	R	R	R	R	R	R	-	R
Clindamycin (2 µg)	R	R	R	R	S	R	-	R
Erythromycin (15µg)	R	R	R	R	R	R	-	R
Gentamycin (120µg)	R	R	R	R	R	R	-	R
Kanamycin (30µg)	R	R	R	R	R	R	-	R
Naladixic Acid (30µg)	R	R	R	R	S	S	-	R
Penicillin (10µg)	R	R	R	R	R	R	-	R
Rifampicin (5µg)	R	R	R	R	S	R	-	R
Streptomycin (10µg)	R	R	R	R	R	R	-	R
Sulfamethoxazole+ Trimethoprim	R	R	R	R	R	R	-	R
Tetracyclin (30µg)	R	R	R	R	S	S	-	R

11. F. sp. KMA, *12. F. sp.* GR2016-10, 13. *F. sp.* CRH, 14. *Flavobacterium frigidimaris* DSM 15937^{T},15. *F. sp*. CSZ, 16. *Flavobacterium aquidurense* KCTC 22840^{T}, 17. *F. sp.* LC2016-23, 18. *F. sp.* KJJ, 19. *F. sp.* MC2016-29, 20. *F. sp.* JLP. R, Resistant, S, Sensitive, - Data not collected.

Table 21 Continued. Antibiotic Resistance

	17	18	19	20	21	22	23	24
Ampicillin (10µg)	R	R	R	R	R	R	R	R
Cephalexin (30 µg)	R	R	R	R	R	R	R	R
Chloramphenicol (30ug)	R	R	R	R	R	R	R	R
Ciprofloxacin (5 µg)	R	R	R	R	R	S	S	R
Clavamox (10µgClav+20µgAmox)	R	R	R	R	R	R	R	R
Clindamycin (2 µg)	R	R	R	R	R	R	R	R
Erythromycin (15µg)	R	R	R	R	R	R	R	R
Gentamycin (120µg)	R	R	R	R	R	R	R	R
Kanamycin (30µg)	R	R	R	R	R	R	R	R
Naladixic Acid (30µg)	R	R	R	R	R	S	S	R
Penicillin (10µg)	R	R	R	R	R	R	R	R
Rifampicin (5µg)	R	R	R	R	R	R	R	R
Streptomycin (10µg)	R	R	R	R	R	R	R	R
Sulfamethoxazole+ Trimethoprim	R	R	R	R	R	R	R	R
Tetracyclin (30µg)	R	R	R	R	R	S	R	S

21. *Flavobacterium hercynium* LMG 26768^{T}, 22. *F. sp.* MC2016-06, 23. *F. sp.* ABG, 24. *F. sp.* LC2016-12, 25. *Flavobacterium saccharophilum* LMG 8384^{T}, 26. *F. sp.* JRM, 27. *F. sp.* AJR, 28. *F. sp.* KMS, 29. *Flavobacterium hydatis* DSM 2063^{T}, 30. *F. sp.* ALJ2. R, Resistant, S, Sensitive, - Data not collected.

Table 21 Continued. Antibiotic Resistance

	25	26	27	28	29	30	31	32	33	34	35
Ampicillin (10µg)	R	R	R	R	R	R	R	S	-	-	R
Cephalexin (30 µg)	R	R	R	R	R	R	R	R	-	-	R
Chloramphenicol (30ug)	R	R	R	R	R	R	R	R	-	-	R
Ciprofloxacin (5 µg)	R	R	R	R	R	R	R	S	-	-	S
Clavamox (10µgClav+20µgAmox)	R	R	R	R	R	R	R	S	-	-	R
Clindamycin (2 µg)	R	R	R	R	R	R	R	S	-	-	S
Erythromycin (15µg)	R	R	R	R	R	R	R	S	-	-	R
Gentamycin (120µg)	R	R	R	R	R	R	R	R	-	-	R
Kanamycin (30µg)	R	R	R	R	R	R	R	R	-	-	R
Naladixic Acid (30µg)	R	R	R	R	R	R	R	S	-	-	S
Penicillin (10µg)	R	R	R	R	R	R	R	R	-	-	R
Rifampicin (5µg)	R	R	R	R	R	R	R	S	-	-	S
Streptomycin (10µg)	R	R	R	R	R	R	R	S	-	-	S
Sulfamethoxazole+ Trimethoprim	R	R	R	R	R	R	R	S	-	-	R
Tetracyclin (30µg)	R	S	R	R	R	S	R	S	-	-	S

21. *Flavobacterium hercynium* LMG 26768^{T}, 22. *F. sp.* MC2016-06, 23. *F. sp.* ABG, 24. *F. sp.* LC2016-12, 25. *Flavobacterium saccharophilum* LMG 8384^{T}, 26. *F. sp.* JRM, 27. *F. sp.* AJR, 28. *F. sp.* KMS, 29. *Flavobacterium hydatis* DSM 2063^{T}, 30. *F. sp.* ALJ2, 31. *Flavobacterium granuli* DSM 19729^{T}, 32. *F. sp.* AED, 33. *F. sp.* HJJ, 34. *Flavobacterium aquicola* DSM 100880^{T}, 35. *Flavobacterium aquatile* LMG 4008^{T}. R, Resistant, S, Sensitive, - Data not collected.

Discussion

Recovery of organisms provided well isolated colonies for further isolation and use in gDNA extraction and 16s rRNA gene amplification. While one contaminant from the plating was observed on the *F. sp.* HJJ plate (Figure 1F), this colony was well isolated from the others, however, and restreaking did not result in a mixed culture. The gDNA extraction and PCR products resulted in concentrations acceptable for sequencing as specified by MiGS and Genewiz (Table 1). Analyzing these samples via gel electrophoresis showed high molecular weight bands for the gDNA extractions for all of the organisms. Based on intensity of the bands, *F. sp.* ANB, F. sp. CSZ, and *F. sp.* LC2016-12 had the highest gDNA concentrations (Figure 3), and this was confirmed in the Qubit measurements (Table 1). The 16s rDNA amplification using the 27f and 1492 primers showed bands of varying intensity and all of the same expected size of about 1500 base pairs. This is indicative that the PCR amplification worked as intended. The negative control well had some PCR product present. This was not cause for repeating the PCR, however, as the intensity of the negative control well was low, leading to an estimated DNA concentration of < 5ng/mL (Figure 3). A minimal amount of contamination of the primer, PCR reagent mix, or even aerosolization of DNA solution in the presence of a large excess of target DNA is not ideal. However, it was not a significant problem when just confirming the identity of the recovered organisms' DNA samples because the concentration of the contaminating DNA would be much lower that the target DNA sequence. The same observations can be made for the frozen/thawed cell PCR products (Figure 4). Very little product was seen in the negative control well, indicating a low level of contamination. The Sanger sequencing results of both samples showed traces with clean peaks and very little baseline noise (Figure 5). These Sanger Sequence derived 16s rRNA gene sequences were compared to the previous sequences for each organism, all resulting in

good matches. In summary, the initial recovery and identification of the organisms as the same novel species from previous years was successful.

The 16s rRNA gene sequence was derived from the genome sequences and used in NCBI BLAST searches against different databases. The results of these searches can be found on tables 3-19. The outsome of these searches resulted in 16s rRNA gene sequence similarities under 98.50%. Alternatively, the organisms were equally similar to multiple other named species. It is impossible for an organism to be two different species at the same time. Be it low similarity or equidistance between organisms, all novel species candidates were determined based on the 16S rRNA gene sequence to likely be novel. The 16s rRNA gene sequences were used to construct a phylogenetic tree (Figure 6) to visualize evolutionary relationships between the organisms. *Flavobacterium aquatile* LMG 4008^{T} was isolated alone on its own long branch, indicative that it is the least related to all other organisms on the tree. The sum of horizontal branch lengths between the novel species and their close relatives reflects their 16s rRNA gene similarity.

gDNA samples were sent for sequencing at MiGS, and the reads were then assembled with SPAdes on the PATRIC platform (Davis *et al* 2020). The assemblies were all less than 93 contigs and had a minimum average coverage of 83.2 (Table 3). With less than 100 total contigs, high N50 values, and high average coverage, these genome assemblies were considered high quality (Chun *et al.* 2018). After annotation with RAST (Aziz *et al.* 2008), rpoB gene sequences were used to look for other close relatives that were not revealed using the 16s rRNA gene. The rpoB gene experiences less evolutionary pressure than the 16s rRNA gene, allowing it to accumulate more mutations and differentiate between species more clearly than the 16s rRNA gene. Being a protein coding gene, however, these mutations cannot be loss of function, as then the organisms would not generate mRNA. The rpoB genes for each novel species was used in

BLAST searches, and the top hits for each organism was downloaded and used to generate a maximum likelihood tree (Figure 6). The clustering and distances on this tree, as well as the 16s rRNA phylogenetic tree, guided the organisms to be compared to in the OGRI tables.

OGRI tables included the 16s rRNA gene similarity, Estimated DNA-DNA (eDDH), Average Nucleotide Identity (ANI), and Average Amino Acid Identity (AAI). The long-established threshold for eDDH is 70% for species level comparisons. If two organisms have an eDDH less than 70%, they are considered different species (Wayne et al 1987). Similarly, the species threshold for ANI is 95%, meaning two organisms with an ANI less than 95% are different species (Kim et a. 2014). As seen in the OGRI Tables (Tables 3-19), all of the novel organisms have ANI and eDDH values that fall under these thresholds when compared to their validly named closest relatives, identifying the organisms as truly novel. Some of the novel organisms, however, have values above these thresholds when compared to each other. When comparing *F. sp.* JLP to *F. sp.* MR2016-29, the ANI value is above this threshold. Therefore, the initial hypothesis that they were novel species is further supported. *F. sp.* JLP and *F. sp.* MR2016-29 also exceed these threshold values when compared to *F. sp.* KJJ (Table 12). This indicates that all three of these unnamed organisms are the same species. This is highly interesting, as all were isolated in different years, and MR2016-29 was isolated from a different source entirely. Single strain species descriptions, that is, naming a new species that only has one strain, are one of the problems adding to the bloating of certain catch all groups. Using this same logic, it can be concluded that, counting *F. sp.* KJJ, there are five organisms that have multiple isolates; *F. sp.* NLM and *F. sp.* WLB are independent isolates of the same species (Table 4); *F. sp.* ANB and *F. sp.* LC2016-13 are independent isolates of the same species (Table 7), *F. sp.* CRH and *F. sp.* GR2016-10 are independent isolates of the same species (Table 9), the

aforementioned *F. sp.* KJJ, *F. sp.* JLP and *F. sp.* MR2016-29 are all independent isolates of the same species (Table 12) and *F. sp.* JRM, *F. sp.* KMS, and *F. sp.* AJR are all independent isolates of the same species (Table 16).

AAI is a measurement that is not yet widely used, and thresholds require more research before they can be established. Xu *et al.* described using AAI and evolutionary distance from a phylogenetic tree as a basis to split the *Erythrobacteraceae* into multiple genera, settling on a threshold of about 70% (2020). Nicholson *et al.* further refined the threshold, noting that AAI values for across different genera result in a bimodal distribution with a gap between 74% and 76%. Viewing this as a natural breaking point, organisms with an AAI of less than 70% are likely different genera, while organisms with an AAI of 75% or higher are in the same genus (Nicholson et al. 2020). For organisms in the "grey area" of between 70% and 75%, phylogenetic trees can help inform relationships and show in which groups different organisms may belong. *F. sp.* ANB (Table 4) was shown to be in the same genus as all of its close relatives, but when compared to *F. aquatile* LMG 4008^{T}, its AAI value was 67.82%. This result shows that *F. sp.* ANB should be classified in a different genus from *Flavobacterium*. This result is a trend with all of the novel species, uncovering two distinct genera from *Flavobacterium*, one containing *F. sp.* CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* CRH, *F. sp.* GR2016-10, *F. sp.* CSZ, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* JLP, *F. sp.* MR2016-29, *F. sp.* MC2016-06, *F. sp.* ABG, *F. sp.* LC2016-12, *F. sp.* JRM, *F. sp.* KMS, *F. sp.* AJR, *F. sp.* ALJ2, all of their close relatives, and *F. johnsoniae* UW101^{T}, and another containing *F. sp.* AED, *F. sp.* HJJ and its close relatives, including *F. aquicola* DSM 100880^{T}.

Though *F. sp* AED and *F. sp.* HJJ have AAI values above 70% compared to *F. johnsoniae* UW101^{T} (Table 6), they are in the grey area of between 70% and 75%. The Genome

Taxonomy Database tree can provide more context on these relationships. Though *F. sp.* HJJ and all of the novel organisms discussed here are not on the GTDB tree, their placement can be inferred using the placements of the closest relatives. Using the relative evolutionary distance of the organisms can provide the information required to determine if organisms should be in the same genus or not (Supplemental Figure 1). As seen on the tree, *F. sp.* AED is placed on a separate branch from *F. johnsoniae* UW101^{T}. The closest relatives to *F. sp.* HJJ cluster together, and it can be inferred that *F. sp.* HJJ is near this cluster. This branch is closer to *F. sp.* AED than it is to *F. johnsoniae* UW101^{T}, and following it reveals that *F. sp.* AED and *F. sp.* HJJ share a more recent common ancestor. With this context, it can be argued that these two organisms should be in the same genus as each other, but different from the genus containing *F. johnsoniae* UW101^{T}. It could also be argued that *F. sp.* AED is in a different genus than *F. sp.* HJJ because the AAI value of *F. sp.* AED compared to F. aquicola DSM 100880^{T} is 75.23%. This conclusion is also supported by the presence of flexirubin pigments in the *F. johnsoniae* UW101^{T} group and the *F. sp.* HJJ group that are absent in the *F. sp.* AED group (Table 20). However, as the AAI is also in the "grey area," it is unclear, and further work must be done to distinguish these two groups from each other and determine if they should truly be considered different genera.

Looking at the Venn Diagrams for these organisms can elucidate sets of genes that would be distinguishing characteristics for these groups. For example, *F. sp.* ANB's Venn Diagram revealed 1052 genes shared by its close relatives that were not found in *F. aquatile* LMG 4008^{T}. Given its distance on the GTDB tree and its AAI value compared to *F. aquatile* LMG 4008^{T}, the list of 1052 genes could be used to distinguish the genus of *F. sp.* ANB from *Flavobacterium*.

Conclusions

Based on these results, it can be concluded that *F. sp.* CNT, *F. sp.* NLM, *F. sp.* WLB, *F. sp.* HTF, *F. sp.* LC2016-01, *F. sp.* ANB, *F. sp.* KMA, *F. sp.* CRH, *F. sp.* GR2016-10, *F. sp.* CSZ, *F. sp.* LC2016-23, *F. sp.* KJJ, *F. sp.* JLP, *F. sp.* MR2016-29, *F. sp.* MC2016-06, *F. sp.* ABG, *F. sp.* LC2016-12, *F. sp.* JRM, *F. sp.* KMS, *F. sp*. AJR, *F. sp.* ALJ2 are different enough from the type species *F. aquatile* LMG 4008^{T} to be considered members of different genera, but they are similar enough to *F. johnsoniae* UW101^{T} to be members of the same genus. The name *Lycobacterium gen. nov* is proposed for this new genus. *F. sp.* AED and *F. sp.* HJJ are also sufficiently divergent from *F.* aquatile LMG 4008^{T} to be members of different genera but should be categorized in the same genus as *F. aquicola* DSM100880^{T}. The name *Noelibacterium gen. nov*. is proposed for this new genus. Additionally, *F. sp.* JLP and *F. sp.* MR2016-29 are independent isolates of same species, *F. sp.* NLM and *F. sp.* WLB are independent isolates of the same species, *F. sp.* CRH and *F. sp.* GR2016-10 are independent isolates of the same species, and *F. sp.* JRM, *F. sp.* KMS, and *F. sp.* AJR are all independent isolates of the same species.

Future work

To finish this project, Fatty Acid Methyl Ester analysis and Polar lipids must be completed for all of the organisms and their closest relatives. The following organisms must all be ordered and be fully characterized compared to their closest novel relative: *F. circumlabens* P5626^{T}, *F. cupreum* CCM8825^{T}, *F. ginsenosidimutans* THG 01^{T,} and *F. piscis* CCUG 60099^{T}. Due to problems with the recovery process, *F. sp.* LC2016-13 and *F. reichenbachii* DSM 21791^{T} must be fully characterized for phenotypes. All organisms need to be grown on plates with

different pH levels and all should have lists of unique and shared genes generated in order to better distinguish groups, especially for *F. sp.* AED and *F. sp*. HJJ. Due to contamination or lack of growth, *F. sp.* CSZ and *F. sp*. HJJ must be tested for resistance to different antibiotics. Many organisms also need to have two or more Biolog Gen III plates run to test for different phenotypes.

Acknowledgements

This work was supported by the Joanne and Arthur Haberberger Fellowship, awarded to Colin Allen at Lycoming College, Williamsport, PA.

References

Ali Z, Cousin S, Fruhling A, Brambilla E, schumann P, *et al*. Flavobacterium rivuli sp. nov., Flavobacterium subsaxonicum sp. nov., Flavobacterium swingsii sp. nov. and Flavobacterium reichenbachii sp. nov., isolated from a hard water rivulet. Int. J. System. Evol. Microbiol. 2009:59:2610-2617. DOI 10.1099/ijs.0.008771-0

Aslam Z, Im WT, Kim MK, and Lee ST. Flavobacterium granuli sp. nov., isolated from granules used in a wastewater treatment plant. Int. J. System. Evol. Microbiol. 2005:55:747-751. DOI 10.1099/ijs.0.63459-0

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, *et al.* The Rast Server: rapid annotation using subsystems technology. BMC Genomics 2008; 9:75.

Bernardet JF, Segers P, Vancanneyt M, Berthe F, Kersters L, *et al*. Cutting a Gordian Knot: emended classification and description of the genus *Flavobacterium*, emended description of the family *Flavobacteriaceae*, and proposal of *Flavobacterium hydatis* nom.nov (basonym, *Cytophaga aquatilis* Strohl and Tait 1978). Int. J. Syst. Bacteriol. 1996:46:128-148.

Chaudhary DK, Kim DU, Kim D, and Kim J. Flavobacterium petrolei sp. nov., a novel psychrophilic, diesel degrading bacterium isolated from oil-contaminated Arctic soil. Sci. Rep-UK. 2019:9:4134. https://doi.org/10.1038/s41598-019-40667-7

Cousin S, Pauker O, and Stackebrandt E. Flavobacterium aquidurense sp. nov. and Flavobacterium hercynium sp. nov., from a hard-water creek. Int. J. System. Evol. Microbiol. 2007:47:243-249. DOI 10.1099/ijs.0.64556-0

Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, *et al*. Proposed minimal standards for the use of genome data for taxonomy of prokaryotes. Int. J. System. Evol. Microbiol. 2018:68:461-466. Doi: 10.1099/ijsem.0.002516.

Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res. 2020 Jan 8;48(D1):D606-D612. doi: 10.1093/nar/gkz943.

Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 1985:39:783-791

Garcia-Lopez M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, et al. Analysis of 1,000 type-strain genomes improves taxonomic classification of Bacteroidetes. Front. Microbiol. 2019:10:2083. Doi: 10.3389/fmicb.2019.02083.

Hahnke RL, Meier-Kolthoff JP, Garcia-Lopez M, Mukherjee S, Huntemann M, *et al*. Genome-based taxonomic classification of *Bacteroidetes*. Front. Microbiol. 2016;7: article 3. Doi: 10.3389/fmicb.2016.02003.

Hatayama K, Ushida A, Kuno T. Flavobacterium aquicola sp. nov., isolated from river warer. Int. J. Syst. Evol. Microbiol. 2016:66:2789-2796. Doi:10.1099/ijsem.0.001057.

Horn MA, Ihssen J, Matthies C, Schramm A, Acker G, and Drake HL. Dechloromonas denitrificans sp. nov., Flavobacterium denitrificans sp. nov., Paenibacillus anaericanus sp. nov. and Paenibacillus terrae strain MH72, N2O-producing bacteria isolated from the gut of the earthworm Aporrectodea caliginosa. Int. J. Syst. Evol. Microbiol. 2005:55:1255-1265. DOI 10.1099/ijs.0.63484-0

Huang X. and Madan A. CAP3: A DNA sequence assembly program. Genome Res. 1999;9: 868-877.

Kämpfer P, Lodder N, Martin K, Avendaño-Herrera R. Flavobacterium chilense sp. nov. and Flavobacterium araucananum sp. nov., isolated from farmed salmonid fish. Int. J. Syst. Evol. Microbiol. 2012:62:1402-1408. Doi:10.1099/ijs.0.033431-0.

Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int K Syst Evol Microbiol 2014:64:346-351.

Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980:16:111-120.

Kralova S, Busse HJ, Svec P, Maslanova I, Stankova E, *et al.* Flavobacterium circumlabens sp. nov. and Flavobacterium cupreum sp.nov., two psychrotrophic species isolated from Antarctic environmental samples. Syst. Appl. Microbiol. 2019:42:291-301. https://doi.org/10.1016/j.syapm.2018.12.005

Kumar S, Strecher G. and Tamura K. MEGA7: Molecular Evolutionary Genetic Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 2016;33:1870-1874.

Lawson PA, Citron DM, Tyrrell KL, & Finegold SM. Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938. Anaerobe 2016; 40:95-9.

McCammon SA, Innes BH, Bowman JP, Franzmann PD, Dobson SJ, *et al*. Flavobacterium hibernum sp. Nov, a lactose-utilizing bacterium from a freshwater Antarctic lake. J. Syst. Evol. Microbiol. 1998:48:1405-1412

NCBI Resource Coordinators. Database resources of the National Center for Biothechnology Information. Nucleic Acids Res. 2018; 4:46 (Database issue): D8-D13. Doi: 10.1093/nar/gkx1095.

Nicholson AC, Gulvik CA, Whitney AM, Humrighouse BW, Bell ME, *et al.* Division of the genus Chryseobacterium: Observation of discontinuities in amino acid identity values, a possible consequence of major extinction events, guides transfer of nine species to the genus Epilithonimonas, eleven species to the genus Kaistella, and three species to the genus Halpernia gen. nov., with description of Kaistella daneshvariae sp. nov. and Epilithonimonas vandammei sp. nov. derived from clinical specimens. Int. J. Syst. Evol. Microbiol. 2020. https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.003935

Nogi Y, Soda K, and Oikawa T. Flavobacterium frigidimaris sp. nov., isolated from Antarctic seawater. Syst. Appl. Microbiol. 2005:28:310-315. DOI:10.1016/j.syapm.2005.01.001

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, *et al*. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 2018:36: 996–1004.

Reddy CA, Beveridge TJ, Breznak JA, Marzluf G, Schmidt TM et al. Methods for General and Molecular Microbiology, 3rd ed ASM Press; 2007.

Ryu SH, Park JH, Moon JC, Sung Y, Lee SS, and Jeon CO. Flavobacterium resistens sp. nov., isolated fromstream sediment. J. Syst. Evol. Microbiol. 2008:58:2266-2270. DOI 10.1099/ijs.0.65720-0

Saitou N. and Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 1987;4:406-425

Sasser M. Technical note # 101 microbial identification by Gas Chromatographic Analysis of Fatty Acid Methyl Esters (GC-FAME). 2009:1–6.

Sant'Anna FH, Bach E, Porto RZ, Guella F, Sant'Anna EH,Passaglia LMP. Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy. Crit. Rev. Microbiol. 2019:2:182-200. Doi: https://doi.org/10.1080/1040841X.2019.1569587.

Sant'Anna FH, Reiter KC, de Fatima Almeida P, Passaglia LM. Systematic review of descriptions of novel bacterial species: evaluation of the twenty-first century taxonomy through text mining. Int. J. Syst. Evol. Microbiol. 2020;70:2925-2936. Doi: 10.1099/ijsem.0.004070.

Shin SK, Ha Y, Cho YJ, Kwon S,Yong D, and Yi H. Flavobacterium gilvum sp. nov., isolated from stream water. Int J Syst Bacteriol. 2017:67:153-157. DOI 10.1099/ijsem.0.001594

Thompson CC, Amaral GR, Campeao M, Edwards RA, Polz ME et al. Microbial taxonomy in the post-genomic era: rebuilding from scratch?. Arch. Microbiol. 2015;197:359-370. Doi:10.1007/s00203-014-1071-2.

Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, et al. Report of the ad hoc committee on reconciliation of app. roaches to bacterial systematics. Int J Syst Bacteriol. 1987; 37:463–464.

Weisburg WG, Barns SM, Pelletier DA, and Lane DJ. 16S Ribosomal DNA Amplification for Phylogenetic Study. J. Bacteriol. 1991:173:697-703. DOI: 10.1128/jb.173.2.697-703.1991

Xu L, Sun C, Fang C, Oren A, and Xu XW. Genomic-based taxonomic classification of the family Erythrobacteracea. Int J Syst Bacteriol. 2020:70:4470-4495.

Yang JE, Kim SY, Im WT, and Yi TH. Flavobacterium ginsenosidimutans sp. nov., a bacterium with ginsenoside converting activity isolated from soil of a ginseng field. J. Syst. Evol. Microbiol. 2001:61:1408-1412. DOI 10.1099/ijs.0.025700-0

Zhang Z, Schwartz S, Wagner L, and Miller W. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 2000 :7:203-14.

Zamora L, Fernandez-Garayzabal JF, Sanchez-Porro C, Palacios MA, Moore ERB, *et al. Flavobacterium plurextorum sp. nov. Isolated from Farmed Rainbow Trout (Oncorhynchus mykiss.* PLOS ONE 2013:8:e67741.

Zamora L, Vela AI, Sanchez-Porro C, Palacios MA, Moore ERB, et al. Flavobacterium tructae sp. nov. and Flavobacterium piscis sp. nov., isolated from frarmed rainbow trout (Oncorhynchus mykiss). Int. J. Syst. Evol. Microbiol. 2014:64:392-399. Doi: 10.1099/ijs.0.056341-0