Detecting Flavobacterial Fish Pathogens in the Environment Using High-Throughput Community Analysis

Todd Testerman[1], Lidia Beka[1*], Emily Ann McClure[1†], Stephen R. Reichley[2‡], Stacy King[2], Timothy J. Welch[3], and Joerg Graf[1#]

[1]University of Connecticut, Department of Molecular and Cell Biology, Storrs, Connecticut, 06269, USA
[2]Riverence Provisions, Buhl, Idaho, USA
[3]National Center for Cool and Cold Water Aquaculture, Agricultural Research Service/U.S. Department of Agriculture, Kearneysville, West Virginia, USA.

Keywords: aquaculture, pathogen, detection, metagenomics

#Corresponding author: Joerg Graf, University of Connecticut, Department of Molecular and Cell Biology, Storrs, Connecticut 06269, USA, joerg.graf@uconn.edu, Phone: 860-486-9284

*Current address: Independent Researcher, Glastonbury, CT, USA
†Current address: Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA
‡Current address: College of Veterinary Medicine, Mississippi State University, Starkville, MS, USA

Abstract

Diseases caused by the fish pathogens *Flavobacterium columnare* and *Flavobacterium psychrophilum* are major contributors of preventable losses in the aquaculture industry. The persistent and difficult to control infections caused by these bacteria make timely intervention and prophylactic elimination of pathogen reservoirs important measures to combat these disease-causing agents. In the present study, we present two independent assays for detecting these pathogens in a range of environmental samples. Natural water samples were inoculated with *F. columnare* and *F. psychrophilum* cells, and pathogen levels were detected using Illumina MiSeq sequencing and droplet digital PCR. Both detection methods accurately identified pathogen-positive samples and showed good agreement in quantifying each pathogen. Additionally, the real-world application of these approaches was demonstrated using environmental samples collected at a rainbow trout aquaculture facility. These results show that both methods can serve as useful tools for surveillance efforts in aquaculture facilities, where the early detection of these flavobacterial pathogens may direct preventative measures to reduce disease occurrence.

Importance

Early detection of a deadly disease outbreak in a population can be the difference between mass fatality or mitigated effects. In the present study, we evaluated and compared two techniques for detecting economically impactful aquaculture pathogens. We demonstrate that one of these techniques, 16S rRNA gene sequencing using Illumina MiSeq technology, provides the ability to accurately detect two fish pathogens, *F. columnare* and *F. psychrophilum*, while simultaneously profiling the native microbial

community. The second technique, droplet digital PCR, is commonly used for pathogen detection, and the results obtained using the assays we designed with this method served to validate those obtained using the MiSeq method. These two methods offer distinct advantages. The MiSeq method pairs pathogen detection and microbial community profiling to answer immediate and long-term fish health concerns, while droplet digital PCR method provides fast and highly sensitive detection that is useful for surveillance and rapid clinical responses.

Introduction

Improving detection methods for bacterial pathogens is a top priority in the global aquaculture industry, as delayed detection and identification of a pathogen can lead to devastating disease outbreaks and high production losses (1). *Flavobacterium columnare* and *Flavobacterium psychrophilum*, the causative agents of columnaris and bacterial cold-water disease (BCWD), respectively, are important threats to the fish farming industry (2). Once introduced into a fish-rearing facility, these pathogens have the potential to rapidly spread and establish biofilms, allowing them to escape biocidal treatments (3, 4). Due to their speed of transmission and long-term survival on surfaces, *F. columnare* and *F. psychrophilum* are immensely difficult to control in fish farms (5, 6) and thus have contributed to major economic losses (7).

Although considerable efforts have been made to control infections caused by flavobacteria over the last fifty years, rainbow trout production losses caused by these pathogens remain high. Studies point to the ability of *Flavobacterium* spp. to survive in a variety of environments, including aquatic and fish-host associated biomes (5, 6).

Biofilm formation and the rapid acquisition of antibiotic resistance genes reduce the efficacy of antibiotic therapies for trout. Antibiotics are often ineffective at eliminating pathogens in rearing-facilities and increase the potential for the development of drug-resistant *Flavobacterium* species (8, 9). Genome rearrangement and mutation can also lead to antigen variation in the fish host, reducing the efficacy of new vaccines, potentially explaining why studies report a low reproducibility of results (9–11). The potential of *Flavobacterium* pathogens to linger in fish farms necessitates improved detection methods that allow for early detection and provide insights into potential reservoirs in aquaculture facilities.

Current *F. columnare* and *F. psychrophilum* identification methods range widely depending on the need and capabilities of the laboratory performing the identification (12) and include classical biochemical (13), antigen-based, PCR-based, and DNA array-based techniques (14). The detection of *F. columnare* and *F. psychrophilum* in water and surface samples is further complicated by the presence of multitudinous non-pathogenic *Flavobacterium* species. These environments also commonly contain PCR inhibitors that can reduce the reliability of PCR-based assays (15). Furthermore, *F. psychrophilum* can also be challenging to identify through culture-based approaches because of its slow growth rate and overgrowth by other bacteria (16). These growth characteristics of flavobacteria present a challenge for timely identification, and alternative identification methods are needed.

At aquaculture facilities, similar to natural settings, many areas can harbor pathogens and warrant careful sampling for pathogen detection. For example, after entering a facility with the incoming water, bacterial pathogens can colonize and form

biofilms on various surfaces, including the fish raceway walls, containment screens, or baffles (structures commonly used to direct the flow of water downward within raceways). Bacteria in biofilms are more resistant to common biocidal treatment methods, which is a serious problem (17). If the pathogens attach to the surfaces of water pipes or other system components, they can escape disinfection or cleaning attempts and potentially serve as sources of disease recurrence. Many of these issues are of particular concern for flow-through facilities that use untreated spring, ground, or river water. Thus, this specific type of farm could benefit from surveillance at the water inflow. However, even with strategic and frequent sampling, the success of pathogen surveillance and outbreak prevention will be contingent on rapid, sensitive, and specific detection techniques.

Next-generation sequencing (NGS) has been employed in source-tracking pathogens and has the advantage of being able to detect a wide range of pathogens at once, and it is increasingly becoming an affordable tool for use in molecular epidemiology (18, 19). While diagnostic metagenomics is still an unorthodox method in fish disease diagnostics, recent studies on viral detection and persistence in aquatic habitats (20, 21) have demonstrated the potential advantages of this technique. Amplicon deep-sequencing provides an opportunity to detect the presence of lingering etiological agents (20) and may aid in preventing future outbreaks in aquaculture settings.

Marker-gene surveys in particular are common and canonically performed by sequencing a variable region of the 16S rRNA gene, which is ubiquitous and allows for the discernment of bacterial taxa (22). Various approaches are available to analyze 16S

rRNA gene sequence data. While traditional approaches use a sequence clustering technique, a more recent approach identifies amplicon sequence variants (ASVs), where the amplified sequence reads are error-corrected, and identical sequences are placed into an ASV. This approach improves pathogen detection sensitivity and is implemented by packages such as QIIME 2 (23) and DADA2 (24). Each ASV is classified by comparing it to a reference dataset. Although the resolution obtained by sequencing a short region of the 16S rRNA gene is not sufficient to identify most taxa at the species level, different variable regions provide greater resolution for some taxa.

Quantitative PCR (qPCR) has long been considered a gold standard for the detection and quantification of disease-causing agents. More recently, droplet digital PCR (ddPCR) has emerged as a promising advancement on qPCR. Due to its partitioning technology and robustness against inhibitory particles, ddPCR has been shown to be an efficient and sensitive approach for detecting low-copy DNA targets and has even begun to replace traditional qPCR in specific applications (25, 26). Additionally, ddPCR results report absolute counts in a sample rather than relative frequencies as are reported with NGS-based methods (27).

In the present study we developed two new assays to detect two pathogenic *Flavobacterium* species, *F. columnare* and *F. psychrophilum*, in water and surface samples by (i) Illumina MiSeq sequencing and (ii) ddPCR. These methods will be useful detection tools for aquaculture facilities. We tested the utility of Illumina MiSeq high-throughput sequencing of the 16S rRNA V4 region to detect these pathogens in inoculated water samples collected from a natural source. Subsequently, these results were independently validated using a ddPCR assay targeting a different housekeeping

gene, ATP synthase subunit alpha (*atpA*), which is commonly used as a genetic marker in flavobacterial typing(9). In addition, water and surface samples collected from an aquaculture facility were also tested using both methods, and the results were compared.

Results

Differentiation of *Flavobacteria* to the species level using the V4 region of the 16S rRNA gene.

A conventional approach for studying microbial communities is to perform high-throughput sequencing of a variable region of the 16S rRNA gene. However, accurate characterization of a polymicrobial sample and detection of specific pathogenic species require that the genetic diversity of the variable region permits species-level classification. Thus, to determine whether the V4 region of the 16S rRNA gene would suffice for species-level identification of fish pathogens of interest, we downloaded the 16S rRNA gene sequences of all 143 available type strains of *Flavobacterium* species from the RDP database to serve as an up-to-date reference gene bank. We trimmed the sequences to the 253 base pairs of the widely used V4 region and constructed a phylogeny. The resulting phylogeny (Figure 1) and the distance matrix (data not shown) identified the closest neighbors of *F. columnare* and *F. psychrophilum* as *F. verecundum* and *F. limicola*, which differed by 4 and 6 base pairs, respectively. Using this approach, the V4 region was observed to provide sufficient resolution to identify *F. columnare* and *F. psychrophilum* to the species level.

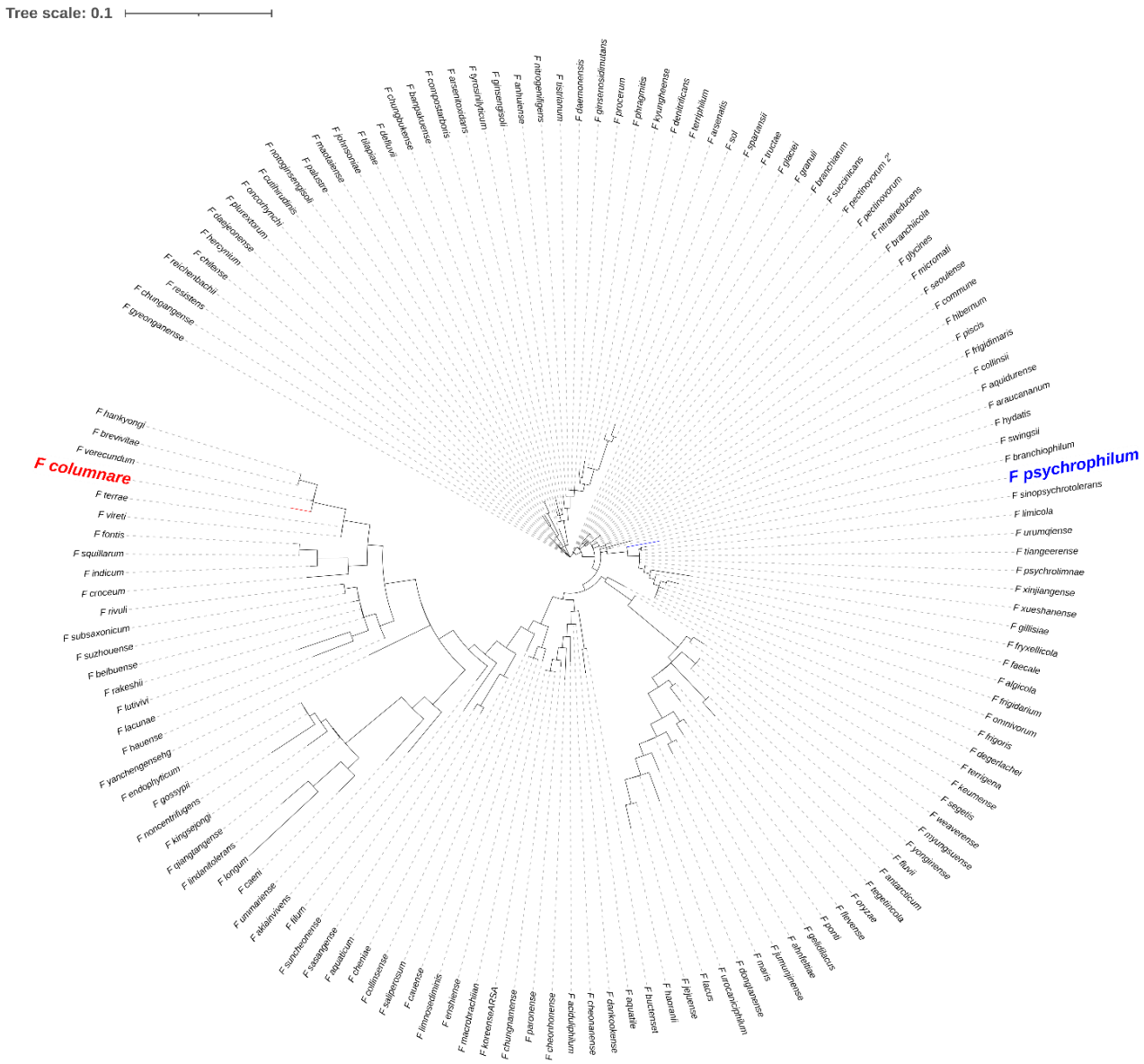


Figure 1: 16S rRNA Gene Cladogram for the Genus *Flavobacterium.* Sequences of the 253 base pair V4 region of the 16S rRNA gene from *Flavobacterium* type strains were aligned, and a phylogenetic tree was constructed using Fasttree. *Flavobacterium columnare* (red) is 4 base pairs away from *F. verecundum*, and *F. psychrophilum* (blue) is 6 base pairs away from *F. limicola*.

MiSeq: Detection of spiked *F. columnare* and *F. psychrophilum* in natural water samples

The ability to detect *F. columnare* and *F. psychrophilum* on an Illumina MiSeq was tested by inoculating and sequencing natural water samples with increasing amounts of these pathogens in three distinct dilution series (*F. columnare* alone, *F. psychrophilum* alone, or their combination). Each sample set was spiked with 10-fold dilutions of pathogen(s) (alone or in combination) over a five order of magnitude range. The relative levels of *F. columnare* and *F. psychrophilum* were normalized using a hemocytometer; however, accurate viable counts were difficult to obtain due to the clumping nature of *Flavobacterium* spp. Two uninoculated samples were used as controls to determine whether native *F. columnare* and/or *F. psychrophilum* could be detected and assess potential filtration system carryover between sample sets. The number of classified *F. columnare* and *F. psychrophilum* was determined using QIIME 2 with a DADA2 plugin for detecting ASVs.

F. columnare was detected in all inoculated water samples excluding the lowest-inoculum sample (Figure 2) for both the *F. columnare*-only and co-inoculated sample series. *F. psychrophilum* was detected in all inoculated samples in both dilution series. No sequencing reads were assigned to either *F. columnare* or *F. psychrophilum* in the uninoculated controls. The average read depth of the 17 inoculated samples was 43,853 reads with a standard error of the mean (SEM) of 3,636 reads. A total of 83,143 reads in the entire data set were assigned to *F. columnare* and a total of 38,450 reads were assigned to *F. psychrophilum. F. columnare* was not detected in any samples where it was not expected (samples where *F. columnare* was not added). Only four *F.*

psychrophilum reads were recovered from samples where it was not expected (the highest inoculum *F. columnare* sample).

ddPCR: Detection of spiked *F. columnare* and *F. psychrophilum* in natural water samples

A ddPCR-based approach was developed to independently quantify *F. columnare* and *F. psychrophilum* cells using a single copy housekeeping gene (*atpA*) that has been used previously for species-level discernment in a multi-locus sequence typing scheme for flavobacteria (9). These assays were optimized, validated, and tested for specificity as described in the methods section.

The *atpA*-ddPCR assay was used to determine the number of *F. columnare* and *F. psychrophilum* cells present in the same inoculated water samples evaluated in the Illumina MiSeq detection assay (Figure 2). As expected, the concentrations of both pathogens in each sample series mirrored the 10-fold dilutions over five orders of magnitude using the ddPCR-based assay (Figure 2). We detected ~1,000 copies of *atpA* per liter of filtered water in the most dilute samples. Since the *F. columnare* and *F. psychrophilum* genomes each harbor one copy of *atpA*, the observed copy number corresponds to 20 genome equivalents/µL of extracted DNA added to the ddPCR reaction mix (total DNA extraction volume was 50 µL).

The ddPCR analysis of samples that were not inoculated allowed us to determine the background levels of *F. columnare* (280 cells/L) and *F. psychrophilum* (12,800 cells/L) in the natural water samples. These counts are much higher than those obtained from the MiSeq experiment, where only a single sample, the highest inoculum *F. columnare* sample, had 3,750 cells/L *F. psychrophilum* detected (four MiSeq reads).

This discrepancy could be due to *F. columnare* or *F. psychrophilum* cells being present in the natural water source at levels below the detection limit of the MiSeq method, highlighting the superior sensitivity of ddPCR. Additionally, contamination may have occurred during sample processing or as a result of carry-over from previous filtration operations at levels the MiSeq approach could not easily detect. Lastly, unknown non-target organisms may have cross-reacted with the *F. psychrophilum* assay. To address this discrepancy, the average background level for each method from all non-spiked samples was subtracted from the counts shown in Figure 2.

For ddPCR, a no template control (NTC) as well as two positive controls containing one pathogen and one positive control containing both pathogens were assayed in parallel. The NTC was negative for both pathogens, and only the added target pathogen(s) were detected in the positive controls, indicating that the occurrence of contamination during ddPCR sample preparation is unlikely to explain the background levels of pathogens detected by ddPCR. Droplet counts per reaction ranged from 13,528 to 20,005 droplets with a median droplet count of 18,212. Overall, these data indicate that *F. columnare* and *F. psychrophilum* can be quantified over a wide range of concentrations in natural water samples using ddPCR, but proper controls must be used to ensure accuracy.

Comparison of MiSeq and ddPCR Results

We were interested in comparing the MiSeq and ddPCR data more directly. As 16S rRNA gene MiSeq data are relative and ddPCR data are absolute, it was necessary to convert absolute numbers while accounting for multiple copies of the 16S rRNA gene

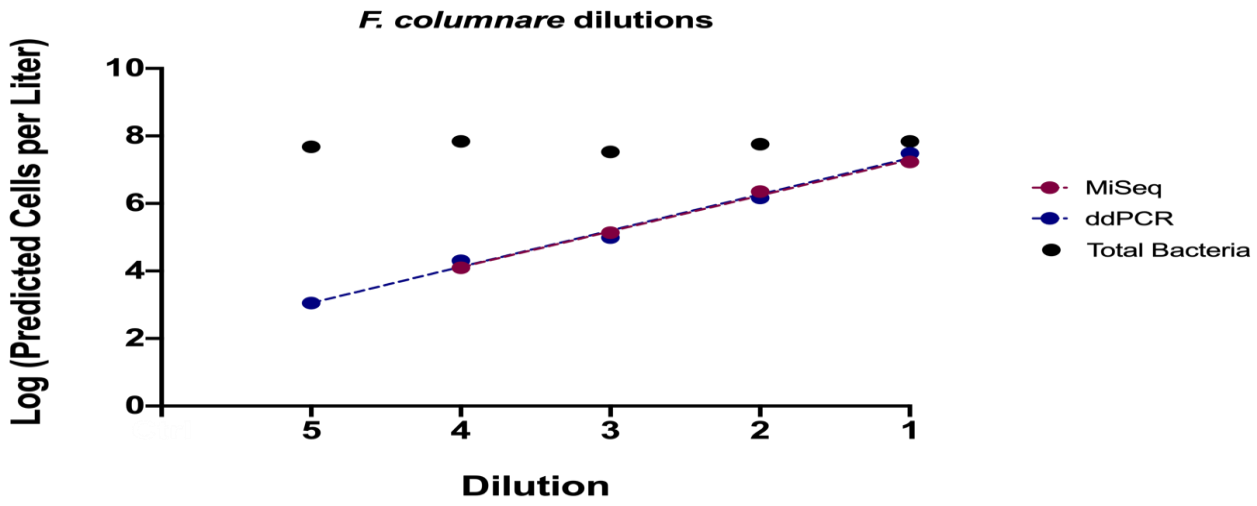
246 in individual bacterial cells. Using a universal 16S qPCR assay, we determined the 16S
247 rRNA gene copy number in each sample (mean = 1.36×10^8 copies, SEM = 1.52×10^7
248 copies). Additionally, using an estimate of the 16S rRNA gene copy number based on
249 the taxonomic composition of these samples (28), we converted this value to the
250 estimated total number of bacterial cells per sample (mean = 3.6×10^7 cells/mL, SEM =
251 3.84×10^6 cells/mL; Figure 2, Supplementary Figure 2).

252 Subsequently, using the absolute concentration data from both the MiSeq and
253 ddPCR assays, we compared the *F. columnare* and *F. psychrophilum* counts in the
254 spiked samples (Supplementary Table 1). Figure 2 displays the line of best fit (dotted
255 lines) produced from a linear regression analysis using log transformed count values.
256 Overall, the results were highly similar between the MiSeq- and ddPCR-generated data
257 and the single and co-inoculation samples. However, for *F. columnare*, the limit of
258 detection differed. In the samples that received the fewest cells, *F. columnare* was not
259 detected using the MiSeq method, whereas ~1,000 *F. columnare* genome copies were
260 detected with the ddPCR assay (Figure 2a). This difference in detection levels is likely
261 due to the MiSeq producing 10-100 thousand reads and having ~100,000,000 16S
262 rRNA gene copies in the sample, which may lead to rare sequences being drowned out
263 by more dominant species. Overall, there appeared to be excellent qualitative
264 agreement between the two methods.

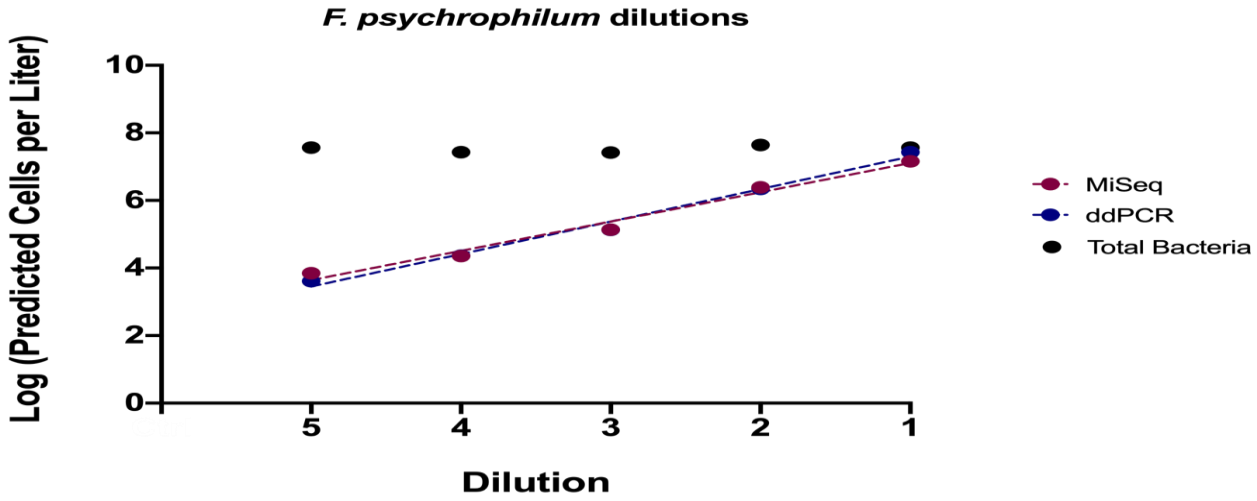
265 On a quantitative level, the estimated cell count values (Figure 2) obtained were
266 highly similar between the two methods. For *F. psychrophilum*, the ddPCR approach
267 yielded higher counts than the MiSeq method for 6 of 10 samples, with an average
268 coefficient of variance (CoV) of 22%. A paired Wilcoxon signed-rank test did not support

a significant difference between the methods (p = 0.375). The *F. columnare* counts obtained by ddPCR were higher than those obtained with the MiSeq approach for 6 out 8 samples (only considering samples where signal was detected for both methods), with an average CoV of 25.5%, a difference that was not significant using a paired Wilcoxon signed-rank test (p = 0.25).

A



B



C

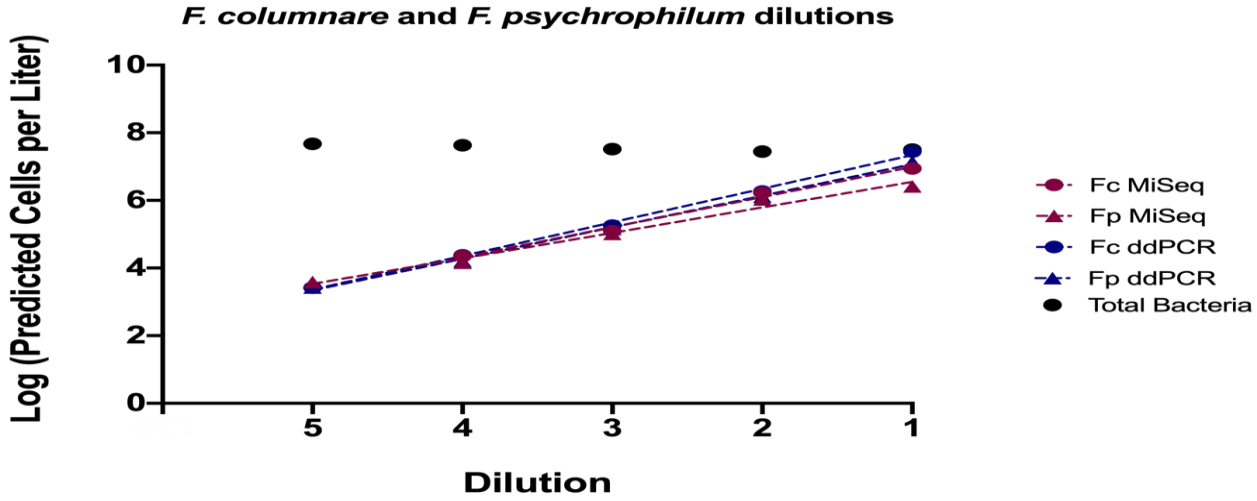


Figure 2: Comparison of MiSeq and ddPCR Assay Results. Estimated cell counts of *F. columnare* (red) and *F. psychrophilum* (blue) are shown for each of the five 10-fold dilutions assayed. Counts are shown on the y-axis as log values, while dilutions are shown on the x-axis and increase from the lowest concentration sample (5) to the highest concentration sample (1). (A) Samples inoculated with *F. columnare* alone, (B) *F. psychrophilum* alone, or (C) *F. columnare*/*F. psychrophilum* combined. The co-inoculated samples (C) show *F. columnare* counts with circles and *F. psychrophilum* counts with triangles. Black circles indicate the total estimated bacterial cell count within the sample calculated using qPCR.

Pathogen Detection in Environmental Samples

We used both the MiSeq and ddPCR methods to evaluate water samples and surface swabs we collected as part of a microbial community survey of an aquaculture facility. Twelve samples were collected from a commercial, spring-fed, rainbow trout farm that had previously experienced periods of BCWD and columnaris disease outbreaks. Table 1 shows the results of the comparison between the two assays for *F. columnare*. In the 6 samples that were positive using the MiSeq method, the ddPCR method also yielded a positive result, while 4 of 6 samples that were negative for *F. columnare* using the MiSeq method yielded extremely low positive results for the ddPCR method. As observed in the previous experiment with spiked samples of lake water, these data also suggest that the MiSeq-based approach of sequencing the V4 region of the 16S rRNA gene is less sensitive than an *atpA*-based ddPCR assay. Alternatively, these results may also suggest that qPCR-based methods are more sensitive to contamination. As

none of the aquaculture samples contained *F. psychrophilum* as detected by either method, we could not perform this comparison.

Table 1: MiSeq and ddPCR *F. columnare* Detection Results. A positive call using the MiSeq required one or more *F. columnare* reads when normalized to 10,000 reads. A positive call for ddPCR required 40 or more *F. columnare* copies to be detected per 50 µL DNA extract. (+) indicates that less than 2 copies were detected per DNA extract with the ddPCR method or less than 1 read per 10,000 detected with the MiSeq method.

	MiSeq (16S-V4)	**ddPCR (*atpA*)**
Natural Source Water	(+)	(+)
Global Hatchery Outflow Water	+	+
Individual Hatchery Pen Outflow Water	+	+
Surface Swab from Tailscreen	+	+
Adult Raceway Outflow Water 1	+	+
Adult Raceway Outflow Water 2	+	+
Surface Swab from Hatchery Baffle 1	-	-
Surface Swab from Hatchery Baffle 2	-	(+)
Surface Swab from Hatchery Baffle 3	-	(+)
Surface Swab from Hatchery Baffle 4	-	(+)
Gill Swab from Adult Fish	-	(+)
Global Hatchery Inflow Water	-	-

Microbial Community Analysis of Natural Water Samples

While not the focus of the present study, a high-throughput and multiplexable 16S rRNA gene survey provides useful microbial community composition data as an added benefit. This advantage is particularly notable for the V4 region of the 16S rRNA gene, which serves as an ideal molecular marker for identifying the widest diversity of bacteria (29). Therefore, to further demonstrate the advantages of this approach, we sequenced and analyzed 17 biological replicates of the pelagic community of a freshwater lake (Figure 3, Supplementary Figure 3). As the water was prefiltered

316 through a 2 µm pore sized prefilter, we considered the cells captured on the 0.2 µm
317 Sterivex filter to be planktonic microorganisms.

318 At the order level, Betaproteobacteriales dominated the community, accounting
319 for 51.8% of all reads, with constituents present at or above 1% abundance including
320 Flavobacteriales (20.7%), Cytophagales (5.7%), Frankiales (4.9%), Chitinophagales
321 (4.0%), Sphingobacteriales (3.4%), and chloroplast-associated reads (1.2%) (Figure
322 3A). A visualization of this community with the spiked *Flavobacterium* spp. removed is
323 presented in Supplementary Figure 3A.

324 At the genus level, 20.0% of reads belonged to the genus *Flavobacterium*, which
325 included the spiked *Flavobacterium* spp. In addition, 14.0, 8.7, 5.5% of the total reads
326 belonged to the genera *Rhodoferax*, *Polaromonas*, and *Polynucleobacter*, respectively,
327 all of which are members of the family Burkholderiaceae. An additional 14.6% of reads
328 were assigned to the family Burkholderiaceae but could not be confidently assigned to
329 any specific genus. Other genera present at or above 1% included *Pseudarcicella*
330 (5.6%), *Methylotenera* (4.7%), *Sediminibacterium* (3.7%), *Limnohabitans* (2.1%), and
331 *Pedobacter* (1.1%) (Figure 3B). A visualization of this community with the spiked
332 *Flavobacterium* spp. removed is presented in Supplementary Figure 3B.
333

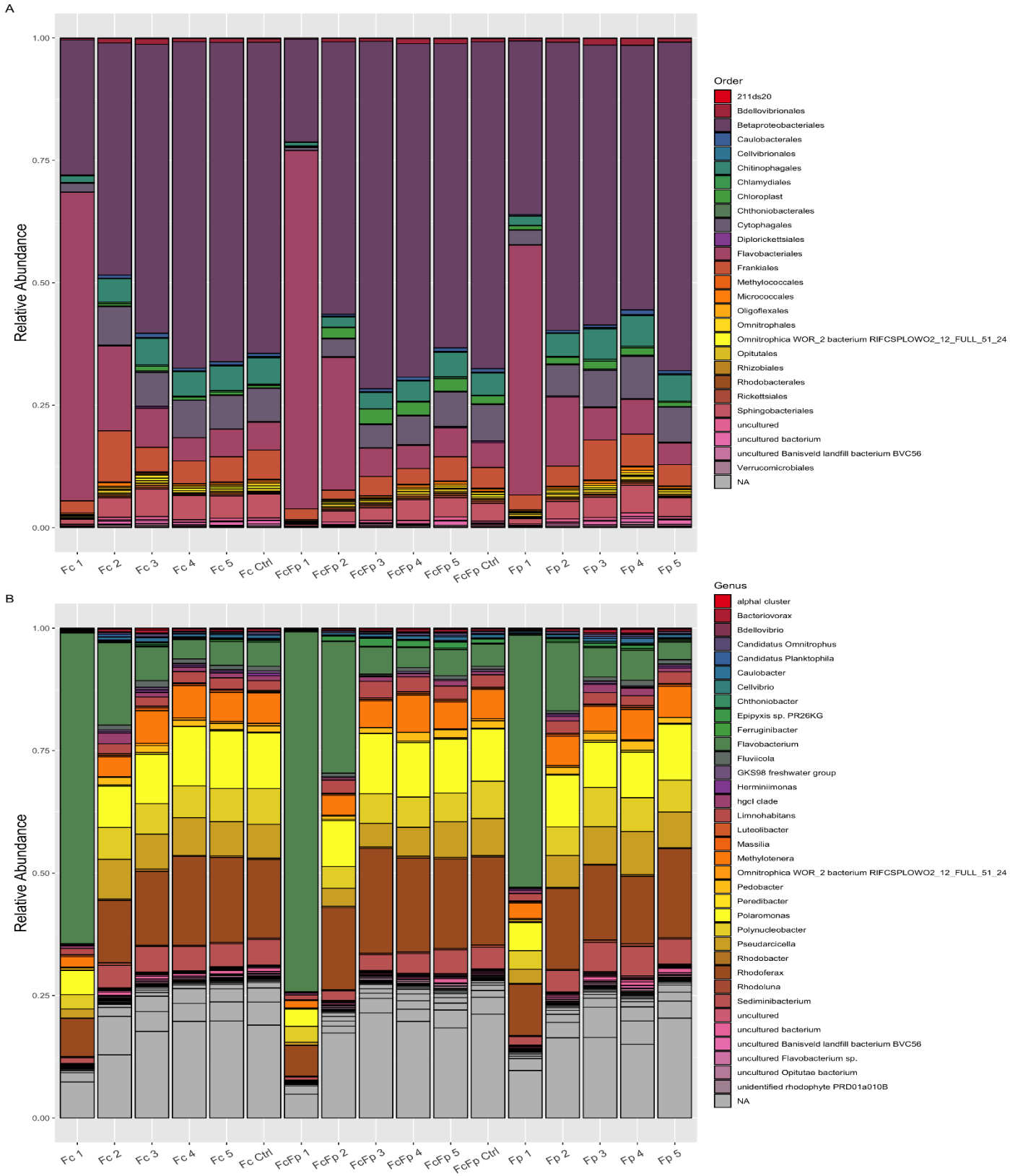


Figure 3: Order and Genus Level Taxonomic Barplots of Water Samples. (A) Order level taxonomic breakdown for each sample. (B) Genus level taxonomic breakdown for each sample. Labels on the x-axis correspond to the dilution series. Barplots are presented as relative abundances. *Flavobacterium* inoculum level decreases from left to right for each series. Fc, Fp and FcFp indicate the addition of *F. columnare* alone, *F. psychrophilum* alone or their combination, respectively. NA – no assignment could be made at this taxonomic level.

Discussion

In the present study, we describe the use of Illumina MiSeq sequencing and ddPCR for the detection of two economically impactful fish pathogens, *F. columnare* and *F. psychrophilum.* We provide direct evidence that MiSeq sequencing of the V4 region of the 16S rRNA gene allows species-level identification of *F. columnare* and *F. psychrophilum*. In addition, we describe a ddPCR assay to detect and quantify these two important fish pathogens, which were detected at similar levels independent of the method used.

Overall, there was good qualitative agreement between the Illumina MiSeq and ddPCR detection of *F. columnare* and *F. psychrophilum.* Statistical testing did not show a significant difference between the quantified levels of *F. psychrophilum* and *F. columnare* using the MiSeq and ddPCR approaches. However, one important difference between the approaches is the limit of detection. Based on the ddPCR quantification data, a detection limit of ~10^3 *F. columnare* cells/L was observed for the MiSeq method, while the ddPCR method could detect 10^2 cells/L. The limit of detection of the MiSeq can be lowered by sequencing to a greater depth, but this would also drive up the

358 overall sequencing costs. A non-targeted NGS-based detection approach requires that
359 the organism of interest be present at a high enough concentration to yield a signal in a
360 complex DNA background (30), which is affected by the sequencing depth. A targeted
361 PCR approach will always outperform a non-specific approach in regard to sensitivity.
362 From a clinical perspective, this difference in the detection limit may be inconsequential,
363 as LD_{50} measurements of *F. columnare* and *F. psychrophilum* are on the scale of 10^8-
364 10^{10} cells/L for immersion or bath-based infection of rainbow trout (31, 32), well within
365 the detectable range for NGS. Furthermore, due to the slow growth rates of *F.*
366 *columnare* and *F. psychrophilum*, remediation begun at concentrations of 10^3 cells/L will
367 likely be effective in keeping concentrations well below the LD_{50}
368 Compared to *F. columnare*, determining the limit of detection for *F.*
369 *psychrophilum* was more challenging due to an apparent background level of *F.*
370 *psychrophilum* in the water samples. This may have been caused by the true presence
371 of *F. psychrophilum*, cross-reaction between an off target organism and the primers and
372 probe, or potential carry-over within the filtration system. However, if contamination
373 occurred or *F. psychrophilum* was present in the water, it was not detected by the
374 MiSeq assay, suggesting a cross-reaction with an unknown organism was the likely
375 cause. While the specificity of the primers and probe were bioinformatically verified, the
376 potential for uncharacterized species presents a risk for a false positive signal arising.
377 We observed a baseline of ~10^4 *F. psychrophilum* cells in the ddPCR analysis of
378 negative samples. This cross-reactivity was accounted for using a baseline subtraction
379 step for the ddPCR data. By adding a known quantity of target DNA to replicate

samples, the presence of cross-reacting DNA can be determined and accounted for in the data analysis.

The benefit of the discovery power added by characterizing the microbial community provided by NGS cannot be overstated. Using this approach, associations between various bacterial taxa and the pathogen of interest can be elucidated, whereas a targeted approach would miss potential hypothesis-generating observations. The potential of other pathogenic bacteria being identified using NGS would also increase the utility of this approach and should be assessed in the future. Compared to ddPCR, a false positive signal arising from NGS is unlikely but possible in the case of primer switching or contamination from wells on the same PCR plate. However, the increased information density produced by this approach increases the odds of correctly parsing out a misidentified sequence. Additionally, it is important to note that because an abundance-filtering step is often performed during pre-processing with NGS datasets (33), low levels of false positive or contaminant reads would be removed as a normal portion of the pipeline. In addition, the inclusion of reagent, negative, and positive controls ensures that contaminants can be more readily detected and removed.

The results obtained for the fish farm samples displayed good agreement between the two methods in detecting the presence/absence of *F. columnare* (Table 1). This result is important, as it was possible to detect this pathogen in both the water and swab samples with the MiSeq. However, the two methods showed disagreement for four surface swab samples in which extremely low levels of *F. columnare* were detected using the ddPCR method but not the MiSeq method. This result may well demonstrate the superior sensitivity of ddPCR in detecting extremely low levels of target sequences.

403 However, this finding may also indicate minor contamination or a PCR artifact leading to
404 a false positive signal. An additional benefit to ddPCR methods is the ability to rapidly
405 process low numbers of samples without increasing the price per sample, while many
406 samples must be assayed at once to decrease the financial input when using the
407 MiSeq.

408 Although previous studies have demonstrated the potential of NGS in pathogen
409 detection, these investigations typically used shotgun metagenomic sequencing (34, 35)
410 or targeted approaches with genus-specific primers (36). While shotgun approaches
411 can generally provide better resolution, they are significantly more expensive (37),
412 suffer from background or host DNA interfering with detection (34), and require a more
413 complex bioinformatic approach. Targeted NGS methods provide useful information on
414 strain variation within the particular genus of interest but lose the survey capabilities
415 borne from an unbiased approach. In the present study, we demonstrate that specific
416 pathogens with distinct hypervariable regions in the 16S rRNA gene can be reliably and
417 semi-quantitatively detected using a non-targeted NGS method. While not solely
418 intended for use in diagnostic settings, we posit that an NGS-based approach to detect
419 *F. columnare* and *F. psychrophilum* is reliable, has similar processing time compared to
420 culture-based methods (38), and provides additional discovery power that can justify
421 increased cost for a research group.

422

423

424

Methods

Phylogeny of *Flavobacterium* type strain 16S rRNA gene sequences

The 16S rRNA gene sequences were downloaded from the Ribosomal Database Project (RDP) database (39) in February 2021. The 253 bp V4 region of the 16S rRNA gene was extracted and aligned using Geneious (version 10.0.6; Biomatters, Auckland, NZ) using a Geneious alignment algorithm. A phylogeny was then constructed in Geneious using Fast Tree (version 2.1.5) (40, 41) with a generalized time reversible model. A Newick file was then exported from Geneious into iTOL (version 6) (42) for additional annotation and export.

Flavobacterium cells and growth conditions

The *F. columnare* ATCC 23463 and *F. psychrophilum* ATCC 49418 type strains were obtained from ATCC (Manassas, VA, USA). Strains were grown on Anacker and Ordall (A&O) agar (43) to reduce clumping (44) and incubated at 22°C for ~36 h. For liquid cultures, a single colony was grown in A&O broth for 36 h with shaking at 22°C for 36 h.

Sample filtration

Preparation of F. columnare- and F. psychrophilum-spiked water samples

Lake water was collected to ensure that samples would contain a natural freshwater microbial community and have sufficient biomass for downstream processing. Nineteen liters of water were collected on March 29, 2019 from the Mansfield Hollow Lake in Mansfield, CT, USA (41.768568°, -72.174603°). The lake water was prefiltered before use as described in the next section. Then, the cultured bacteria were enumerated with

448 a hemocytometer and added to 1 L aliquots of prefiltered lake water to generate a 10-
449 fold dilution series. To ensure proper mixing, bacterial aliquots were added to 50 mL of
450 sterile water prior to being added to 1 L of the prefiltered lake water.
451
452 *Water filtration*
453 A previously described filtration system (45) was employed in the present study with the
454 following modifications. First, larger particles capable of clogging the system were
455 removed by pumping the lake water sample through a 2 µm pore size prefilter (Sigma
456 Aldrich, St. Louis, MO, USA, Cat # AP2502500) and collected in a sterile vacuum flask.
457 After inoculating the filtrate, the total sample was aseptically mixed and filtered using a
458 0.2 µm pore size Sterivex filter unit (Fisher Scientific, Waltham, MA, USA, Cat #
459 SVGPL10RC).
460
461 One liter of molecular grade water (Fisher Scientific, Cat # BP2819-1) was filtered
462 before the experimental samples and analyzed as a system control to ensure the
463 filtration system was sterile. The samples were filtered in the following order:
464 uninoculated lake water; increasing concentrations of *F. columnare* cells; increasing
465 concentrations of *F. psychrophilum*; uninoculated lake water; and increasing
466 concentrations of 1:1 mixtures of *F. columnare* and *F. psychrophilum*. The filtration
467 system was rinsed with 70% ethanol and molecular grade water (Fisher Scientific, Cat #
468 BP2819-1) between each filtration group before the next sample set. After filtering, the
469 Sterivex filter units were immediately frozen at -20°C.
470

471 DNA isolation

472 DNA was extracted from Sterivex filters using a Qiagen (formerly MoBio) Sterivex
473 PowerWater DNA Isolation Kit (Qiagen, Hilden, Germany) following the manufacturer's
474 protocol. DNA was eluted in 50 µL of buffer EB and quantified with a Qubit HS dsDNA
475 Assay Kit (Life Technologies, Carlsbad, CA, USA), with dsDNA concentrations ranging
476 from 4.1 to 11.8 ng/µL.

477

478 16S rRNA gene amplicon sequencing on an Illumina MiSeq

479 *PCR amplification*

480 The V4 hypervariable region of the 16S rRNA gene was amplified in each sample using
481 previously validated primers that contained dual-end adapters for indexing. PCR
482 reactions were set-up in a total volume of 83.4 µL. Reactions contained 3 µL of BSA
483 (New England Biosciences (NEB), Ipswich, MA, USA), 41.7 µL of Phusion 2× Master
484 Mix (NEB), 2.5 µL of 10 µM primer mix (515F forward and 806R reverse rRNA gene V4
485 primers with Illumina MiSeq adaptors), and 30 ng of sample DNA. Each sample was
486 assayed in triplicate using a Bio-Rad C1000 Touch Thermocycler (Bio-Rad
487 Laboratories Inc., Hercules, CA, USA) with the following parameters: an initial
488 denaturation for 3 min at 94°C followed by 30 cycles of 45 sec at 94°C, 60 sec at
489 50°C, and 90 sec at 72°C, with a final elongation step of 10 min at 72°C.

490

491 *Library preparation and MiSeq sequencing*

492 Triplicate reactions were pooled and assayed on a QIAxcel (Qiagen) to verify product
493 size and determine DNA concentrations. Samples were then pooled and cleaned using

a GeneRead Size Selection Kit (Qiagen, Cat # 180514) before being diluted to approximately 4 nM for loading onto an Illumina MiSeq (Illumina, San Diego, CA, USA). An Illumina 500 cycle MiSeq Reagent Kit v2 (Illumina, Cat # MS-102-2003) was used for paired-end sequencing of all libraries.

Bioinformatic processing

QIIME 2

16S rRNA V4 gene sequence data obtained from an Illumina MiSeq were demultiplexed through BaseSpace (basespace.illumina.com). Paired-end data was then imported into QIIME 2 (version 2019.4; https://docs.qiime2.org/2019.4/), and reads were filtered, denoised, and dereplicated using the DADA2 denoise-paired plugin (24)Sequences were taxonomically classified and compiled using the feature-classifier (46)and taxa (https://github.com/qiime2/q2-taxa) plugins. Taxonomy was assigned using a pre-trained Naïve Bayes classifier based on the Silva 132 99% OTU database (https://www.arb-silva.de/documentation/release-132/), with reads trimmed to only include the region bound by the 515F/806R primer pair. Read counts for taxa of interest were obtained from the csv download function within the barplot visualization provided by QIIME 2. Default QIIME 2 processing parameters were used throughout the workflow except where otherwise noted. All commands and related parameters can be viewed at https://github.com/joerggraflab.

ddPCR

Primers and probes were designed using Geneious (version 10.0.6; Biomatters) and are listed in Supplementary Table 2. A 200-bp region of the *atpA* gene was selected for ddPCR analysis using primers and probes unique to *F. columnare* and *F. psychrophilum.* Although a range of primer and probe concentrations were examined for optimum positive/negative droplet separation, the Bio-Rad recommendations of 900 nM primer and 250 nM probe were ultimately used to generate the data shown in the present study. The assay was also optimized using an annealing temperature gradient of 54-59°C. Optimal positive/negative droplet separation occurred at 58°C, with no improvement noted at increased temperatures and a slight decrease in the signal:noise ratio observed at 59°C. An annealing/extension temperature of 58°C was selected for use in all subsequent experiments. *Hae*III was used for restriction digestion of template DNA to decrease droplet rain and, per Bio-Rad recommendations, was included within the reaction mixture after being bioinformatically and experimentally shown not to cut within the target region.

Primer specificity was tested on 21 strains of bacteria including *Flavobacterium* spp. (including the ATCC type strains used in the present study) and other closely related genera. Probe specificity was tested on 9 *F. columnare* strains representing the 4 groups of this species (9). The *F. columnare* assay was shown to be specific for group 1 *F. columnare* (which includes the type strain used in the present study). The *F. psychrophilum* assay was specific for all tested *F. psychrophilum* strains. Neither assay reacted with non-target *Flavobacterium* spp., which included no cross-reactivity between the *F. columnare* and *F. psychrophilum* assays.

538 Dilution series of *F. columnare* and *F. psychrophilum* in water were used to determine
539 the limit of detection for this method, which was calculated as 1×10^{-5} ng of genomic
540 template per reaction with a limit of quantitation of 1×10^{-4} ng of genomic template per
541 reaction. Primer specificity was tested by performing ddPCR in the presence of an
542 excess of eukaryotic (fish) or prokaryotic [ZymoBIOMICS Microbial Community DNA
543 Standard (Zymo Research, Irvine, CA, USA)] DNA. At excess DNA concentrations of up
544 to 12 ng/reaction (prokaryotic) or 19 ng/reaction (eukaryotic) (~$1\text{-}2 \times 10^5$ times more
545 than genomic template), the limit of quantitation was maintained at 1×10^{-4} ng genomic
546 template per reaction.
547 The optimum reaction contained 1× ddPCR SuperMix for Probes (no dUTP) (Bio-Rad),
548 900 nM forward primer, 900 nM reverse primer, 250 nM each probe, 10 U of *Hae*III
549 (NEB), and >0.1 pg of genomic template DNA. The optimum reaction conditions were
550 as follows:10 min of activation at 95°C activation followed by 40 cycles of (30 sec of
551 denaturation at 94°C and 1 min of annealing/extension at 58°C, with a final incubation at
552 10 min 98°C for deactivation. Reactions and droplets were prepared according to
553 recommendations outlined by Bio-Rad. Droplets were read and analyzed using a
554 QX200 ddPCR system (Bio-Rad) with the QuantaSoft software package (version
555 1.7.4.0917).
556
557 One-microlter aliquots of DNA extracted from the spiked natural water samples were
558 used to determine total numbers of pathogenic flavobacteria using ddPCR. DNA from
559 samples with high *Flavobacterium* loads were diluted 1:10 to prevent droplet saturation.
560 All reactions were simultaneously assayed in duplex to detect both *F. columnare* and *F.*

psychrophilum. For comparison to the MiSeq data, read counts were adjusted by subtracting the average read count from the non-spiked samples (7 samples for *F. columnare* and *F. psychrophilum* each).

qPCR

A qPCR assay for 16S rRNA gene quantification was utilized for the present study as previously described (47) (Supplementary Table 2). This assay targets the region of the 16S rRNA gene bound by the 27F and 519R eubacterial primers (48, 49).

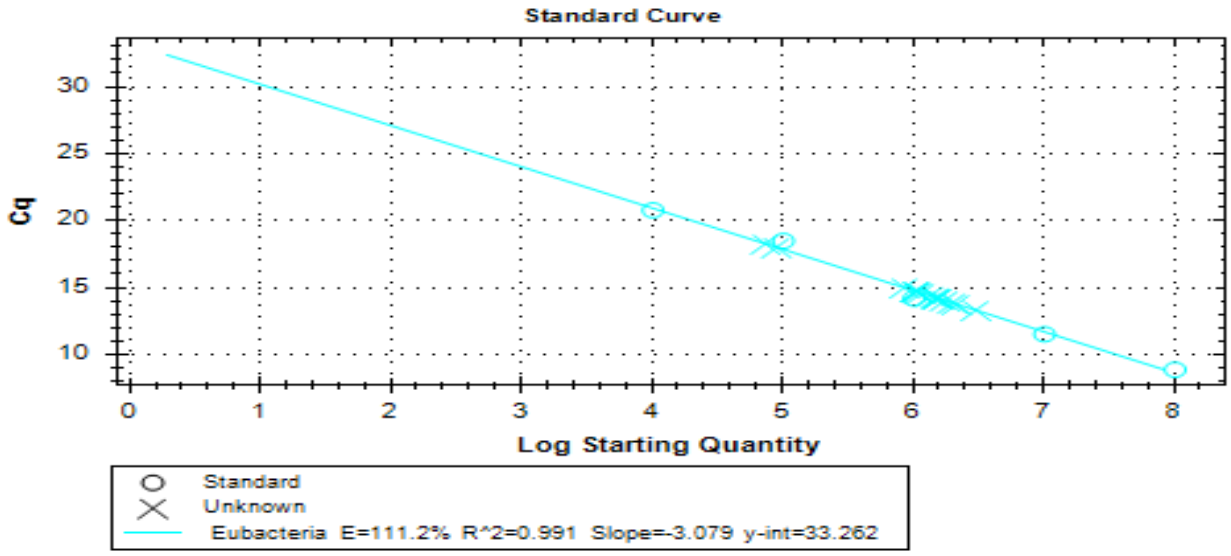
qPCR was performed using an EvaGreen (dye-based) assay on a CFX96 Touch Real-Time Thermocycler (Bio-Rad). Reactions were assembled using 2× SsoFast EvaGreen Master Mix (Bio-Rad), 27F and 519R eubacterial 16S primers at 300 nM final concentration, 5 µL of sample or standard DNA, and PCR-grade water to final volume of 20 µL per reaction. Thermocycling conditions were as follows: 98°C for 3 min followed by 40 cycles of 98°C for 10 sec, 55°C for 30 sec with a plate read step directly following each cycle. Data analysis was performed using Bio-Rad CFX Manager (version 3.1). Single threshold quantification cycle (Cq) value determinations were made with a baseline subtracted curve fit. Baseline and threshold values were automatically determined by CFX Manager.

Standards were generated from extracted *F. psychrophilum* genomic DNA. The genome size of the *F. psychrophilum* type strain (2.72 Mb) was retrieved from NCBI and used to calculate the estimated number of copies of the 16S rRNA gene per nanogram of

584 genomic DNA taking into account the gene copy number (6 copies). The standard curve
585 had an efficiency of 111.2% and an R^2 value of 0.991 with a slope of -3.079 and a y-
586 intercept at 33.262 (Supplementary Figure 1).
587
588 All samples (diluted 1:10) fell within the range of the standard curve except two
589 samples; the negative control *F. columnare* sample and *F. psychrophilum* sample with
590 the second-lowest inoculum. These samples failed to amplify at the 1:10 dilution,
591 possibly due to PCR inhibition. For these two cases, a 1:100 dilution was used for
592 quantification, as amplification clearly occurred in these samples. Three NTCs were
593 assayed, and two had detectable Cq values of 26.15 and 30.35. Based on the standard
594 curve, these Cq values correspond to less than 10^2 bacterial cells in each reaction. This
595 low-level detection is likely due to minor contamination and is not high enough to cause
596 any noticeable change in calculated copies in the actual samples.
597
598 The total bacterial cell count was calculated by dividing the 16S rRNA gene count by the
599 estimated median 16S rRNA gene copies per cell in the environment, which was
600 informed by order-level taxonomic profiling using Illumina MiSeq 16S rRNA gene
601 sequencing. The calculation is summarized as follows: the Flavobacteriales read count
602 was multiplied by the V4 copy number of inoculated *Flavobacterium* species plus the
603 Betaproteobacteriales read count multiplied by the median gene copy number (GCN) for
604 this bacterial order in the ribosomal RNA operon copy number database (rrnDB)
605 (50)The resulting value was then divided by the sum of the Betaproteobacteriales and
606 Flavobacteriales read counts. This number varied from sample to sample depending on

607 the proportion of Flavobacteriales to Betaproteobacteriales (the two dominant bacterial
608 orders in these samples). Higher inoculum samples had a higher estimated average
609 16S rRNA GCN due to *F. psychrophilum* and *F. columnare*, both with higher 16S rRNA
610 GCN than the determined median for the Betaproteobacteriales, making up a larger
611 proportion of the microbial milieu.
612
613 *Controls*
614 Samples that were not inoculated with the target strains were used as additional
615 negative controls, resulting in 7 negative controls for each species. For *F. columnare*,
616 zero reads were detected in the negative control samples. For *F. psychrophilum*, low
617 levels were detected in 1 of 7 negative control samples. Additionally, for MiSeq
618 sequencing, negative and mock community controls were assayed to account for
619 potential contamination and PCR bias issues. Negative controls included PCR and
620 extraction negative controls, neither of which presented a significant number of reads,
621 indicating that contamination was not a major concern at these steps. The mock
622 bacterial community control yielded the expected community composition as specified
623 by Zymo Research.
624
625 *Data availability*: Raw read data are available in the NCBI SRA database under project
626 ID PRJNA732893.
627

628 *Code availability:* Source code, including QIIME 2 processing commands and R
629 commands used, is available at https://github.com/joerggraflab/16S-Flavobacterium-
630 Detection.
631
632 *Acknowledgments:* We thank the UConn Microbial Analysis, Resources, and Services
633 (MARS) facility for performing the sequencing work. We thank Ahmad Hassan and
634 Mariya Riat for helping with the processing of the lake water filtration samples. We
635 thank Dr. Thomas Loch and Dr. Benjamin LaFrentz for providing strains used during the
636 ddPCR assay validation. We thank Dr. Jeremiah Marden for providing feedback and
637 editing the manuscript. This work was funded by the United States Department of
638 Agriculture (USDA) under grant number 8082-32000-006-00-D.
639
640
641
642
643
644
645

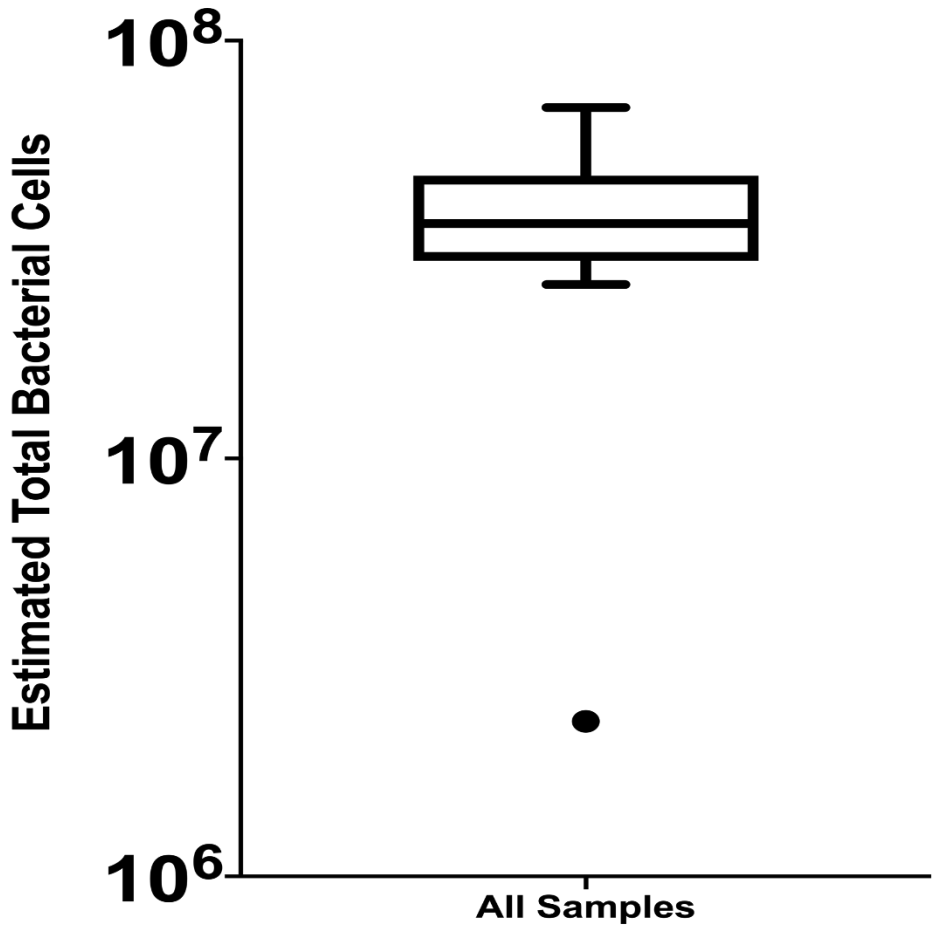


646

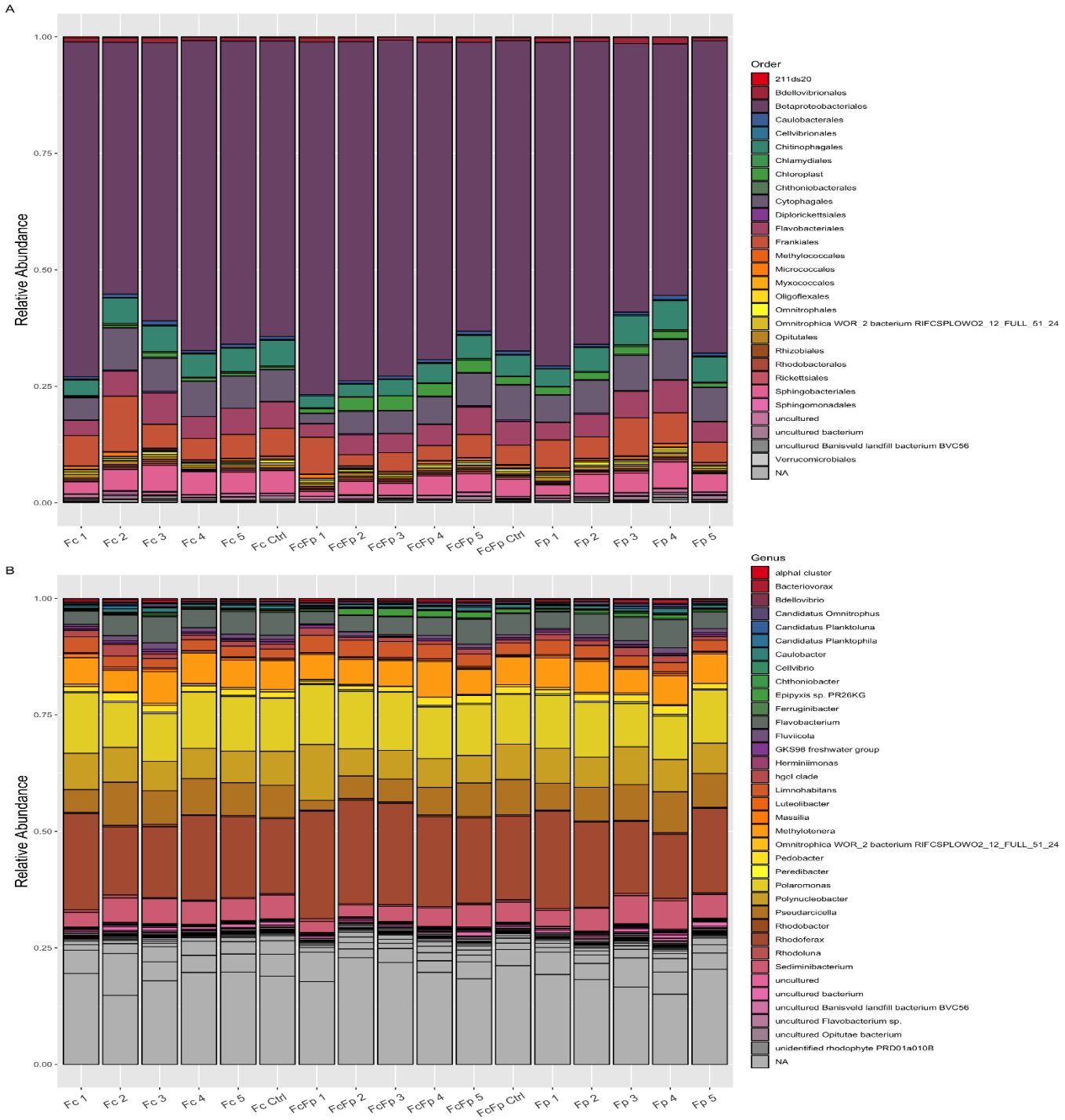
647 **Supplementary Figure 1: Standard Curve for**

648 **Eubacterial 16S rRNA Gene Quantification.** Circles

649 represent standards and x's represent samples.



Supplementary Figure 2: Boxplot of Estimated Total Bacterial Cells as Determined by qPCR. All samples are presented in this boxplot for all dilution series, including negative samples. One outlier is noted that was the negative sample for the dilution series with *F. columnare* alone. The boxplot is presented in Tukey style.



657

658 **Supplementary Figure 3: Order and Genus Level Taxonomic Barplots of Water**

659 **Samples with Spiked Flavobacteria Removed.** (A) Order level taxonomic breakdown

660 for each sample excluding reads from *F. columnare* and *F. psychrophilum*. (B) Genus

level taxonomic breakdown for each sample excluding reads from *F. columnare* and *F. psychrophilum*. Labels on the x-axis correspond to the dilution series. Barplots are presented as relative abundances. The *Flavobacterium* inoculum level decreases from left to right with each series. Fc – *F. columnare* only, FcFp – Co-inoculation, Fp – *F. psychrophilum* only, NA – No assignment could be made at this taxonomic level.

Supplementary Table 1: Calculated *F. columnare* and *F. psychrophilum* Cell Counts for QIIME 2 and ddPCR.

	Qiime2 Calculated Fc Cells	ddPCR Calculated Fc Cells[a]	Qiime2 Calculated Fp Cells	ddPCR Calculated Fp Cells[a]
Fc Ctrl	0	1	0	1067
Fc 5	0	1122	0	0
Fc 4	12514	20022	0	0
Fc 3	137322	97622	0	2267
Fc 2	2238792	1460622	0	0
Fc 1	17000414	30499622	3747	8667
FcFp Ctrl	0	0	0	3267
FcFp 5	0	2622	3833	2667
FcFp 4	22552	23822	14408	15967
FcFp 3	124104	180622	102359	143667
FcFp 2	1675120	1825622	1067381	1276667
FcFp 1	8784734	27799622	2608850	13637667
Fp 5	0	422	6975	4067
Fp 4	0	0	22616	23067
Fp 3	0	0	136545	135667
Fp 2	0	0	2458902	2204667
Fp 1	0	442	14397209	26887667

a. Cell counts shown for ddPCR are baseline adjusted

Supplementary Table 2: Primer and Probe Sequences Used for ddPCR and qPCR.

	Type	Application	Target	Sequence (5'-3')	Reference
AtpA_779F	Forward Primer	ddPCR	*Flavobacterium*	CCA CCA GGA CGT GAR GC	this paper
AtpA_978R	Reverse Primer	ddPCR	*Flavobacterium*	GCA GAA ACG TCA CCC GC	this paper
AtpA_Fc91P	Probe	ddPCR	*F. columnare*	[HEX] AAG CCG TTA ATG ACC CTC CCC C [BHQ1]	this paper
AtpA_Fp61P	Probe	ddPCR	*F. psychrophilum*	[FAM] AAT GAA CCA CCA CCT TTT ACG A [BHQ1]	this paper
27F	Forward Primer	qPCR	Eubacteria	AGA GTT TGA TCM TGG CTC AG	Goffredo et al., 2016
519R	Reverse Primer	qPCR	Eubacteria	GWA TTA CCG CGG CKG CTG	Goffredo et al., 2016

1. Stentiford GD, Sritunyalucksana K, Flegel TW, Williams BAP, Withyachumnarnkul B, Itsathitphaisarn O, Bass D. 2017. New Paradigms to Help Solve the Global Aquaculture Disease Crisis. PLOS Pathogens 13:e1006160.
2. Pridgeon JW, Klesius PH. 2012. Major bacterial diseases in aquaculture and their vaccine development. CAB Reviews https://doi.org/10.1079/PAVSNNR20127048.
3. Cai W, De La Fuente L, Arias CR. 2013. Biofilm formation by the fish pathogen flavobacterium columnare: Development and parameters affecting surface attachment. Applied and Environmental Microbiology 79:5633–5642.
4. Ríos-Castillo AG, Thompson KD, Adams A, Marín de Mateo M, Rodríguez-Jerez JJ. 2018. Biofilm formation of *Flavobacterium psychrophilum* on various substrates. Aquaculture Research 49:3830–3837.
5. Tekedar HC, Karsi A, Reddy JS, Nho SW, Kalindamar S, Lawrence ML. 2017. Comparative genomics and transcriptional analysis of Flavobacterium columnare strain ATCC 49512. Frontiers in Microbiology 8.
6. Cai W, De La Fuente L, Arias CR. 2013. Biofilm formation by the fish pathogen Flavobacterium columnare: development and parameters affecting surface attachment. Applied and environmental microbiology 79:5633–42.
7. NASS. 2019. Trout Production.
8. Sundell K, Wiklund T. 2011. Effect of biofilm formation on antimicrobial tolerance of Flavobacterium psychrophilum. Journal of Fish Diseases 34:373–383.
9. LaFrentz BR, García JC, Waldbieser GC, Evenhuis JP, Loch TP, Liles MR, Wong FS, Chang SF. 2018. Identification of Four Distinct Phylogenetic Groups in Flavobacterium columnare With Fish Host Associations. Frontiers in Microbiology 9.
10. Hoare R, Ngo TPH, Bartie KL, Adams A. 2017. Efficacy of a polyvalent immersion vaccine against Flavobacterium psychrophilum and evaluation of immune response to vaccination in rainbow trout fry (Onchorynchus mykiss L.). Veterinary Research 48.
11. Gómez E, Méndez J, Cascales D, Guijarro JA. 2014. Flavobacterium psychrophilum vaccine development: A difficult task. Microbial Biotechnology 7:414–423.
12. Loch TP, Faisal M. 2015. Emerging flavobacterial infections in fish: A review. Journal of Advanced Research. Elsevier.
13. Austin B, Austin DA. 2007. Isolation/Detection, p. 151–183. *In* Bacterial Fish Pathogens. Springer Netherlands.
14. Crump EM, Perry MB, Clouthier SC, Kay WW. 2001. Antigenic characterization of the fish pathogen Flavobacterium psychrophilum. Applied and Environmental Microbiology 67:750–759.
15. Schrader C, Schielke A, Ellerbroek L, Johne R. 2012. PCR inhibitors - occurrence, properties and removal. Journal of Applied Microbiology 113:1014–1026.
16. Wiklund T, Madsen L, Bruun MS, Dalsgaard I. 2000. Detection of Flavobacterium psychrophilum from fish tissue and water samples by PCR amplification. Journal of Applied Microbiology 88:299–307.
17. Bas S, Kramer M, Stopar D. 2017. Biofilm Surface Density Determines Biocide Effectiveness. Frontiers in Microbiology 8.

18. Rantsiou K, Kathariou S, Winkler A, Skandamis P, Saint-Cyr MJ, Rouzeau-Szynalski K, Amézquita A. 2018. Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment. International Journal of Food Microbiology 287:3–9.
19. Hoffmann C, Stockhausen M, Merkel K, Calvignac-Spencer S, Leendertz FH. 2016. Assessing the feasibility of fly based surveillance of wildlife infectious diseases. Scientific Reports 6.
20. Nkili-Meyong AA, Bigarré L, Labouba I, Vallaeys T, Avarre JC, Berthet N. 2017. Contribution of Next-Generation Sequencing to Aquatic and Fish Virology. Intervirology. S. Karger AG.
21. Kim Y, Aw TG, Teal TK, Rose JB. 2015. Metagenomic Investigation of Viral Communities in Ballast Water. Environmental Science and Technology 49:8396–8407.
22. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. 2014. Analysis, Optimization and Verification of Illumina-Generated 16S rRNA Gene Amplicon Surveys. PLoS ONE 9:e94249.
23. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang K Bin, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik A V., Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 1.
24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581–583.
25. Miotke L, Lau BT, Rumma RT, Ji HP. 2014. High sensitivity detection and quantitation of DNA copy number and single nucleotide variants with single color droplet digital PCR. Analytical Chemistry 86:2618–2624.
26. Tang H, Cai Q, Li H, Hu & P. 2016. Bioscience, Biotechnology, and Biochemistry Comparison of droplet digital PCR to real-time PCR for quantification of hepatitis B virus DNA. Bioscience, Biotechnology, and Biochemistry https://doi.org/10.1080/09168451.2016.1196576.

27. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M. 2013. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nature Methods 10:1003–1005.
28. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. 2015. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Research https://doi.org/10.1093/nar/gku1201.
29. Yang B, Wang Y, Qian PY. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics 17:135.
30. Brandt J, Albertsen M. 2018. Investigation of detection limits and the influence of DNA extraction and primer choice on the observed microbial communities in drinking water samples using 16S rRNA gene amplicon sequencing. Frontiers in Microbiology 9.
31. Garcia C, Pozet F, Michel C. 2000. Standardization of experimental infection with Flavobacterium psychrophilum, the agent of rainbow trout Oncorhynchus mykiss fry syndrome. Diseases of Aquatic Organisms 42:191–197.
32. Kinnula H, Mappes J, Valkonen JK, Sundberg LR, Yu XJ. 2015. The influence of infective dose on the virulence of a generalist pathogen in Rainbow Trout (Oncorhynchus mykiss) and Zebra Fish (Danio rerio). PLoS ONE 10.
33. Auer L, Mariadassou M, O'Donohue M, Klopp C, Hernandez-Raquet G. 2017. Analysis of large 16S rRNA Illumina data sets: Impact of singleton read filtering on microbial community description. Molecular Ecology Resources 17:e122–e132.
34. Gu W, Miller S, Chiu CY. 2019. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annual Review of Pathology: Mechanisms of Disease 14:319–338.
35. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G. 2017. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Archives of Pathology and Laboratory Medicine 141:776–786.
36. Pereira RPA, Peplies J, Brettar I, Höfle MG. 2017. Development of a genus-specific next generation sequencing approach for sensitive and quantitative determination of the Legionella microbiome in freshwater systems. BMC Microbiology 17:79.
37. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. 2018. Evaluating the Information Content of Shallow Shotgun Metagenomics. mSystems 3.
38. Bernardet J-F, Nakagawa Y. 2006. An Introduction to the Family Flavobacteriaceae, p. 455–480. *In* The Prokaryotes. Springer New York.
39. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Research 42.
40. Price MN, Dehal PS, Arkin AP. 2009. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution 26:1641–1650.
41. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5:e9490.

42. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research https://doi.org/10.1093/nar/gkab301.
43. Anacker RL, Ordal EJ. 1959. STUDIES ON THE MYXOBACTERIUM CHONDROCOCCUS COLUMNARIS. Journal of Bacteriology 78.
44. LaFrentz BR, Klesius PH. 2009. Development of a culture independent method to characterize the chemotactic response of Flavobacterium columnare to fish mucus. Journal of Microbiological Methods 77:37–40.
45. Benjamino J, Beka L, Graf J. 2018. Microbiome Analyses for Toxicological Studies. Current Protocols in Toxicology 77:e53.
46. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:1–17.
47. Goffredo M, Mass K, Parks EJ, Wagner DA, McClure EA, Graf J, Savoye M, Pierpont B, Cline G, Santoro N. 2016. Role of gut microbiota and short chain fatty acids in modulating energy harvest and fat partitioning in youth. Journal of Clinical Endocrinology and Metabolism 101:4367–4376.
48. TURNER S, PRYER KM, MIAO VPW, PALMER JD. 1999. Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis. The Journal of Eukaryotic Microbiology 46:327–338.
49. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. Journal of Bacteriology 173:697–703.
50. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. 2015. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Research 43:D593–D598.