



## Article

# Comparative Genomic Analyses of *Flavobacterium psychrophilum* Isolates Reveals New Putative Genetic Determinants of Virulence Traits

Daniel Castillo <sup>1,2</sup> , Valentina L. Donati <sup>3</sup>, Jóhanna Jørgensen <sup>1</sup>, Krister Sundell <sup>4</sup>, Inger Dalsgaard <sup>3</sup>, Lone Madsen <sup>3</sup> , Tom Wiklund <sup>4</sup> and Mathias Middelboe <sup>1,\*</sup>

<sup>1</sup> Marine Biological Section, Department of Biology, University of Copenhagen, 3000 Helsingør, Denmark; daniel.castillo@bio.ku.dk (D.C.); johannajorgensen88@gmail.com (J.J.)

<sup>2</sup> Instituto de Investigación Interdisciplinaria en Ciencias Biomédicas SEK (I3CBSEK), Universidad SEK, Santiago 7520317, Chile

<sup>3</sup> Unit for Fish and Shellfish Diseases, National Institute of Aquatic Resources, Technical University of Denmark, 2800 Kongens Lyngby, Denmark; valdo@aqu.dtu.dk (V.L.D.); inda@aqu.dtu.dk (I.D.); loma@aqu.dtu.dk (L.M.)

<sup>4</sup> Laboratory of Aquatic Pathobiology, Environmental and Marine Biology, Åbo Akademi University, 20500 Turku, Finland; Krister.Sundell@abo.fi (K.S.); tom.wiklund@abo.fi (T.W.)

\* Correspondence: mmiddelboe@bio.ku.dk; Tel.: +45-3532-1991

**Abstract:** The fish pathogen *Flavobacterium psychrophilum* is currently one of the main pathogenic bacteria hampering the productivity of salmonid farming worldwide. Although putative virulence determinants have been identified, the genetic basis for variation in virulence of *F. psychrophilum* is not fully understood. In this study, we analyzed whole-genome sequences of a collection of 25 *F. psychrophilum* isolates from Baltic Sea countries and compared genomic information with a previous determination of their virulence in juvenile rainbow trout. The results revealed a conserved population of *F. psychrophilum* that were consistently present across the Baltic Sea countries, with no clear association between genomic repertoire, phylogenomic, or gene distribution and virulence traits. However, analysis of the entire genome of four *F. psychrophilum* isolates by hybrid assembly provided an unprecedented resolution for discriminating even highly related isolates. The results showed that isolates with different virulence phenotypes harbored genetic variances on a number of consecutive leucine-rich repeat (LRR) proteins, repetitive motifs in gliding motility-associated protein, and the insertion of transposable elements into intergenic and genic regions. Thus, these findings provide novel insights into the genetic variation of these elements and their putative role in the modulation of *F. psychrophilum* virulence.

**Keywords:** *Flavobacterium*; pathogenicity; pan-genome; virulence; evolution; pathogen; genomics; freshwater; T9SS



**Citation:** Castillo, D.; Donati, V.L.; Jørgensen, J.; Sundell, K.; Dalsgaard, I.; Madsen, L.; Wiklund, T.; Middelboe, M. Comparative Genomic Analyses of *Flavobacterium psychrophilum* Isolates Reveals New Putative Genetic Determinants of Virulence Traits. *Microorganisms* **2021**, *9*, 1658. <https://doi.org/10.3390/microorganisms9081658>

Academic Editor: Bart C. Weimer

Received: 2 June 2021

Accepted: 29 July 2021

Published: 3 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Flavobacterium psychrophilum* is a yellow-pigmented, Gram-negative fish pathogenic bacterium with a global distribution in freshwater aquaculture, and causing “bacterial cold water disease” (BCWD) and “rainbow trout fry syndrome” (RTFS) [1]. The disease results in high rates of fry mortality, increased tendency for other infections, and high costs of treatment with antibiotics; causing important economic losses for salmonid aquaculture worldwide [2]. A diverse array of visible phenotypic alterations are attributable to *F. psychrophilum* infection, including necrotic lesions, partial dark skin coloration, exophthalmia, anemia, ascites, and vertebral deformities of the fish [1,2]. These signs have been related to the presence of extracellular proteases [3], adhesion or biofilm formation [4], hemolysis [5], and secreted systems [6,7]. Historically, the initial isolation of *F. psychrophilum* was reported in the USA in the 1940s [1]; since then, *F. psychrophilum* isolates have been obtained from all salmonid producing countries, including European nations [8–10].

Previous studies on *F. psychrophilum* genomics have mainly focused on describing the genomic properties involved in genetic diversification and virulence. First, whole genomic analyses displayed a well conserved core genome with high similarities at nucleotide level in isolates obtained across large geographical scales, suggesting that specific genomic differences in *F. psychrophilum* isolates are mainly driven by gain and loss of mobile genetic elements (MGE) [8,11]. Second, pathogenic-related genes associated with functions such as proteases, adhesion, transport, and motility were found in a collection of isolates from Denmark, Chile, and the USA [11]. Third, genome analyses suggested that *F. psychrophilum* uses the widespread type IX secretion system (T9SS) to secrete many surface adhesins, soluble or cell-associated peptidases, nucleases, and other hydrolytic enzymes [6,7,12,13]. Finally, the genomes of 41 *F. psychrophilum* isolates displayed the presence of small plasmids and genomic islands encoding DNA replication, restriction-modification systems, phage-related elements, and transcriptional regulator genes [8]. Together, these findings suggested that the genomic contents could provide valuable insights into the mechanisms of pathogenicity in *F. psychrophilum*. However, despite the global picture of genomic diversity obtained for this pathogen (>160 genomes available in public databases in 2021), there are still gaps in our understanding of the links between specific genotype and virulence characteristics of *F. psychrophilum*, which are required to unravel the molecular evolution of this pathogenic bacterium. Thus, combining comparative genomics with virulence assays could provide an excellent approach to obtain a more detailed characterization of the gene repertoire or genetic regions that undergo variation (e.g., clusters of tandem repeats genes and amino acid sequences) in relation to changes in virulence properties. Linking potential genomic and phenotypic patterns could contribute to disease prevention approaches, including alternative methods such as phage therapy [14–17].

Studies on genetic variants of fish pathogenic bacteria have demonstrated that the evolution of virulence traits includes several factors such as acquisition of genetic elements by horizontal gene transfer [18], genetic microvariations [19], and genome reduction and gene loss [20]. In this study, we compared the genomic sequences of twenty-five *F. psychrophilum* isolates collected in rainbow trout farms around the Baltic Sea (Denmark, Finland, Sweden, Germany, Russia, and Poland), previously characterized phenotypically by Sundell et al., 2019 [4]. Thus, we explored the relationship between genomic diversity and virulence traits among the *F. psychrophilum* isolates. Together, these approaches provided insights into the local genetic and pathogenic evolution of the fish pathogen *F. psychrophilum* in the Baltic Sea rainbow trout production.

## 2. Materials and Methods

### 2.1. Strain Isolation, Medium Composition, and Growth Conditions

This study used twenty-five *F. psychrophilum* isolates from rainbow trout fish farms in different geographic localities in Denmark, Finland, Sweden, Germany, Poland, and Russia, covering a spatial scale of >2000 km and a temporal scale of >22 years (Table S1) [4]. For comparative purposes, the type strain NCIMB 1947<sup>T</sup> was also included in this study [21]. The isolates were stored at  $-80\text{ }^{\circ}\text{C}$  in TYES broth (tryptone 0.4%, yeast extract 0.04%,  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  0.05%, and  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  0.05%) with 30% glycerol [22]. For culturing the *F. psychrophilum* isolates, cells from the  $-80\text{ }^{\circ}\text{C}$  stocks were inoculated in 50 mL TYES broth and incubated at  $15\text{ }^{\circ}\text{C}$  with agitation (200 rpm) for 48–72 h [11].

### 2.2. DNA Extraction

DNA from *F. psychrophilum* isolates were extracted from cells harvested by centrifugation ( $5000 \times g$ , 15 min,  $4\text{ }^{\circ}\text{C}$ ) using a Wizard<sup>®</sup> Genomic DNA Purification Kit (Promega; Catalogue number A1120). The bacterial DNA of the four isolates selected from the Pacific Biosciences (PacBio) sequencing platform was extracted from 48 h-old bacterial broth cultures using a QIAamp DNA Mini Kit (Qiagen; Catalogue number 51306). The amount of genomic DNA was measured using a Quant-iT<sup>™</sup> PicoGreen<sup>®</sup> quantification kit (Invitrogen, Waltham, MA, USA).

### 2.3. Genome Sequencing, Assembly, and Annotation

The genomic DNA sequences of twenty-five *F. psychrophilum* isolates were obtained using the Illumina HiSeq platform at the FIMM Technology Centre (Finland) and BGI (China) with pair-end read sizes of 100 bp. Library construction, sequencing, and data pipelining were performed in accordance with the manufacturer's protocols. The Illumina data were assembled into contiguous sequences using Geneious software version 10 [23], then, short and low coverage contigs were filtered out. The remaining contigs were aligned using the previously sequenced *F. psychrophilum* strain JIP02/86 as a reference genome (GenBank accession number: AM398681; September 2019) [3]. In addition, plasmid sequences were identified (3.3 kb and 2.1 kb respectively; Table S1) that did not align with the reference genome. In addition, *F. psychrophilum* isolates FPS-R7, FPS-S6, 950106-1/1, and 160401-1/5N were selected for sequencing using the Pacific Biosciences (PacBio) sequencing platform (BGI, China). Library construction, sequencing, and data pipelining were performed in accordance with the manufacturer's protocols. A hybrid assembly was made using Illumina (100 bp) and PacBio read data (average 20 kb) by the Flye assembler program [24]. Circularized assemblies were further polished with the BUSCO [25] and CheckM [26] packages to correct possible single-base and indel errors. To trace the presence of any plasmid, the filtered reads were mapped using SOAP to the bacterial plasmid database [27]. Annotation of the genomic sequences was done using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) [28].

### 2.4. Predictions of Genomic Islands, Virulence-Related Factors and Prophages

We used islandviewer v.4 [29] and MAUVE [30] to predict the putative genomic islands (GIs) (>8 kb, >8 ORFs, associated with integrases or transposases). A virulence database was constructed for *F. psychrophilum*, containing all the putative virulence-related factors recognized previously [3,8,11] and new virulence genes identified by searching against MvirDB (E-value  $\geq 10^{-5}$ ; identity  $\geq 35\%$ ; coverage  $\geq 75\%$ ) [31], virulencefinder 1.2 [32], and RAST [33]. Prophage-related sequences were identified by running bacterial genomes in PHASTER [34]. Putative subcellular localization of ORFs was performed using server tools. Prediction of the localization of bacterial proteins was achieved using PSort V3.0b.75 [35]. Checking of trans-membrane helices (TMH) was performed with TMHMM V2.0c.76 [36]. Predictions of signal peptides were obtained using SignalIP V3.0.77 [37]. Detection and alignment of repeats in protein sequences was accomplished using motif scan [38].

### 2.5. Pan Genome Analysis

In order to predict the genomic diversity in *F. psychrophilum*, the bioinformatics program EDGAR [39] was used to predict the pan genome of all twenty-five *F. psychrophilum* isolates and calculate the pan-genome (total gene repertoire), dispensable genome (genes found in two or more genomes but not in all the sequences), accessory genome (specific genes, only found in one genome), and core genome (common genes, mutually conserved). Pan-genome development was calculated by iterative pairwise comparison of *F. psychrophilum* genomic sequences. Using the metacontig function of EDGAR, we also defined custom groups of *F. psychrophilum* genomes for which the core genome or the pan genome were stored as virtual contigs [11].

### 2.6. Phylogenetic Analysis

To determine the phylogenetic relationship among *F. psychrophilum* isolates based on genomic data, we selected a set of orthologous genes shared by all twenty-five isolates (1866 genes present in a single copy, paralogs not included) using OrthoMCL with an e-value cut off  $10^{-8}$  [40]. The set of 1866 single core and virulent-related genes were first aligned at amino acid level using Clustal W version 2.0 [41]. The alignment of all amino acid sequences from orthologous genes was concatenated using FASconCAT [42]. A gene tree was constructed using PhyML [43].

## 2.7. Accession Numbers

Accession numbers for the twenty-five *F. psychrophilum* isolates and their respective plasmid sequences are listed in Table S1.

## 3. Results

### 3.1. Virulence Properties of *F. psychrophilum* Isolates

We obtained twenty-five *F. psychrophilum* isolates from infected rainbow trout [4]. The *F. psychrophilum* isolates were obtained from spleens and kidneys (Table S1). The type strain NCIMB 1947<sup>T</sup> was also included in this study as a reference strain. In vitro proteolytic activities (collagenase, elastinase, caseinase) and gliding motility were previously measured for all the isolates [4]. Moreover, the virulence phenotype by median lethal dose (LD<sub>50</sub>) of this entire *F. psychrophilum* strain collection had been previously reported [4], and represents a more direct infection process driven by *F. psychrophilum*. The challenge trials (intramuscular injection) with juvenile rainbow trout (mean weight 5 g) divided the isolates into four groups referred to as high (7 isolates; LD<sub>50</sub> < 10<sup>5</sup>), moderate (13 isolates; LD<sub>50</sub> = 10<sup>5</sup>–10<sup>6</sup>), weak (5 isolates; LD<sub>50</sub> > 10<sup>6</sup>), and non-virulent (1 isolate; no mortality observed) (Table S1) [4].

### 3.2. Genomic Characteristics of *F. psychrophilum* Isolates

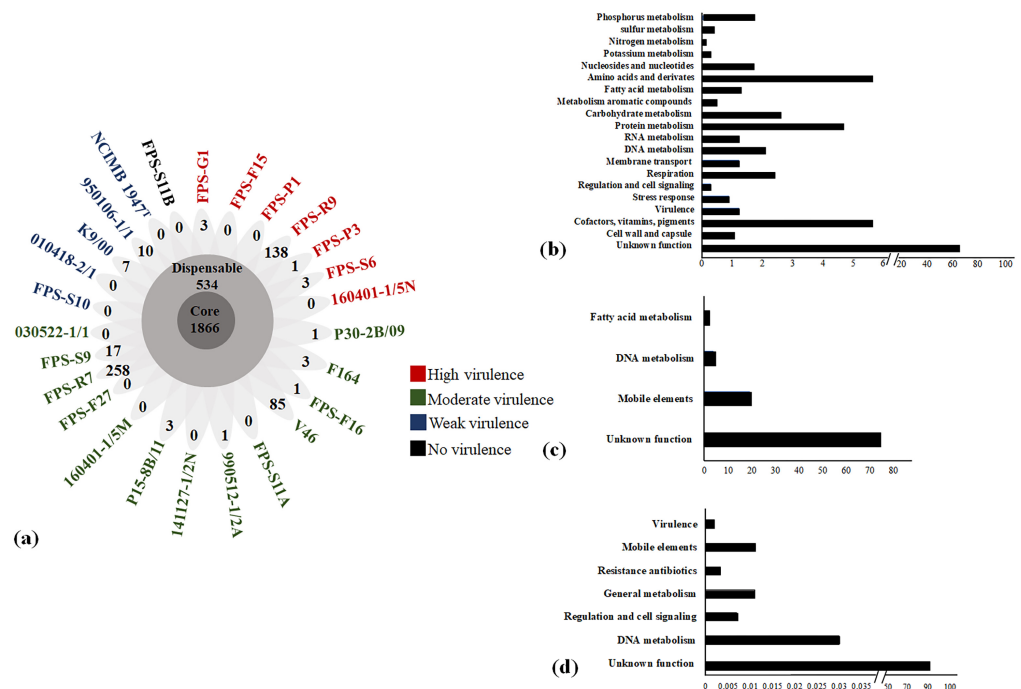
The genomic annotations obtained from the twenty-five *F. psychrophilum* isolates were analyzed together with the genomic annotation of the type isolate NCIMB 1947<sup>T</sup> [17]. The isolates varied in genomic size from 2.71 to 3.20 Mb, with a GC content from 32.3% to 32.6% (Table 1). Analysis of annotated sequences revealed relatively similar coding sequences (CDS) among all the isolates, ranging from 2258 to 2806 (Table 1). One plasmid of either 3.3 or 2.1 kb was present in 15 out of 26 *F. psychrophilum* isolates (Table 1). In contrast to a previous genomic analysis of the type strain NCIMB 1947<sup>T</sup> [8], we did not identify plasmids in this strain in this study, suggesting that the plasmids may have been unstable and lost from the cell.

**Table 1.** Genomic features of *F. psychrophilum* isolates used in this study.

Isolate	Origin	Isolation Year	Genome Size (Mb)	Genes	CDS	GC%	tRNA	Plasmid (kb)
FPS-G1	Germany	2017	2.86	2494	2423	32.4	49	3.3
FPS-F15	Finland	2017	2.86	2496	2425	32.5	49	2.1
FPS-P1	Poland	2016	2.86	2385	2344	32.5	41	No
FPS-R9	Russia	2017	2.86	2387	2347	32.6	34	No
FPS-P3	Poland	2017	2.86	2398	2358	32.4	34	No
FPS-S6	Sweden	2017	2.86	2528	2457	32.5	49	3.3
160401-1/5N	Denmark	2016	2.82	2527	2456	32.5	49	3.3
P30-2B/09	Finland	2009	2.86	2542	2471	32.5	49	3.3
F164	Sweden	1996	2.86	2489	2418	32.3	49	No
FPS-F16	Finland	2017	2.86	2496	2425	32.5	49	No
V46	Finland	2005	2.84	2294	2258	32.4	30	No
FPS-S11A	Sweden	2017	2.86	2484	2413	32.6	49	3.3
990512-1/2A	Denmark	1999	2.86	2486	2415	32.5	49	3.3
141127-1/2N	Denmark	2014	2.85	2488	2417	32.6	49	3.3
P15-8B/11	Finland	2011	2.85	2489	2418	32.5	49	2.1
160401-1/5M	Denmark	2016	2.85	2484	2413	32.5	49	3.3
FPS-F27	Finland	2017	2.86	2490	2419	32.4	49	2.1
FPS-R7	Russia	2017	3.20	2878	2806	32.6	49	No
FPS-S9	Sweden	2017	2.86	2519	2448	32.5	49	No
030522-1/1	Denmark	2003	2.86	2534	2463	32.5	49	3.3
FPS-S10	Sweden	2017	2.86	2495	2425	32.5	49	No
010418-2/1	Denmark	2001	2.86	2546	2475	32.6	49	3.3
K9/00	Finland	2000	2.86	2496	2425	32.5	49	No
950106-1/1	Denmark	1995	2.84	2507	2436	32.6	49	3.3
NCIMB 1947 <sup>T</sup>	USA	Unknown	2.71	2397	2305	32.6	49	No
FPS-S11B	Sweden	2017	2.86	2486	2415	32.6	49	3.3

### 3.3. Relation of *F. psychrophilum* Pan Genome and Virulence Traits

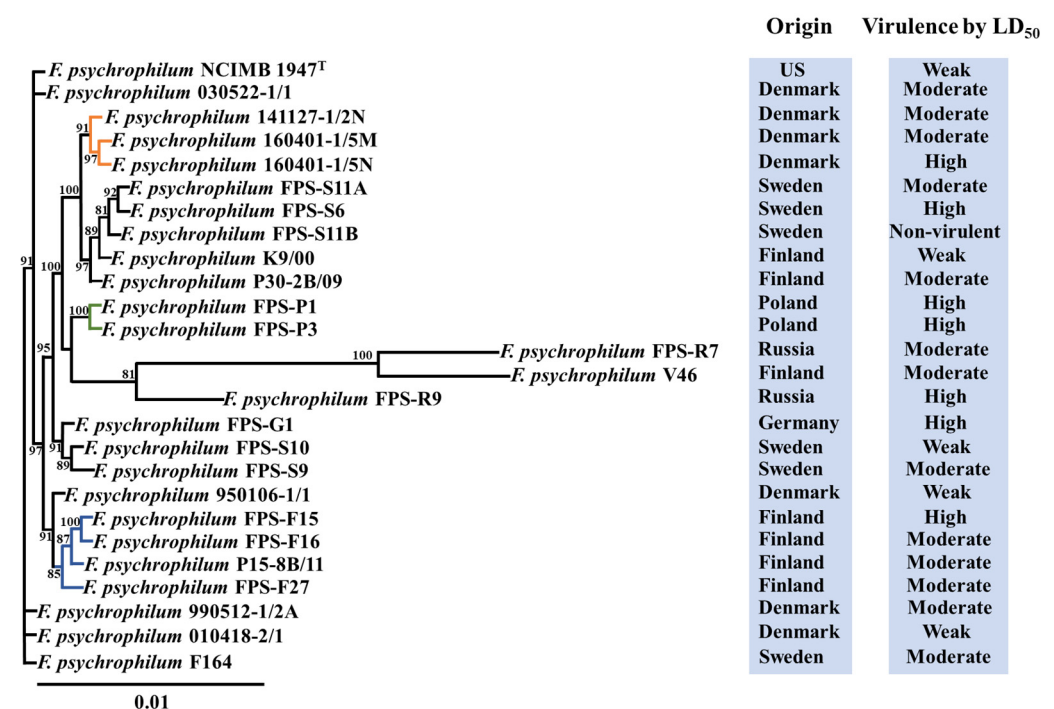
In order to examine the gene repertoire of the twenty-six *F. psychrophilum* isolates (including the type strain NCIMB 1947<sup>T</sup>), a pan genomic analysis was made using the EDGAR platform (Figure 1; Figure S1). The pan genome increased slightly with each addition of a new genome and had at least 4550 ORFs (Figure S1). In contrast, the core genome decreased with the addition of each new genome, reaching 1866 ORFs (Figure S1; Figure 1a). These core-related genes were allocated to putative functional categories using the Clusters of Orthologous Groups of Proteins (COG) database (Figure 1b). The results showed that approximately 65% of ORFs were assigned as hypothetical proteins, 5.6% to cofactor and amino acid derivatives, 4.7% to protein metabolism, and 1.2% to putative virulence-related proteins (Figure 1b). Moreover, the dispensable genome (shell genes present in two or more isolates) reached 534 ORFs and was split between 20% mobile elements (plasmids and prophages) and 13% metabolic-related proteins (Figure 1c). Finally, the remaining 529 ORFs were defined as the *F. psychrophilum* accessory genome (Figure 1a). The number of nonduplicated unique genes in each *F. psychrophilum* isolate varied from 0 to 258 ORFs. The *F. psychrophilum* isolates FPS-R7, FPS-R9, and V46 had the largest numbers of accessory genes, being 258, 138, and 85, respectively (Figure 1a). Functional annotation showed that 91% of ORFs were classified as hypothetical proteins (Figure 1c).



**Figure 1.** Pan-genome of *F. psychrophilum* isolates. (a) The flower plot represents the number of shared (core), dispensable, and accessory genes based on cluster orthologs for each DNA sequence. Petals display numbers of strain-specific genes (accessory genome) found in each genome of *F. psychrophilum*. The array of colors indicates the virulence category as found in the juvenile rainbow trout model [4]. (b) COG subcategories of predicted genes within the core genomes of *F. psychrophilum* isolates. Each category is graphed as a percentage of the total number of genes in the core pool of genes. (c) COG subcategories of predicted genes within the dispensable genome of *F. psychrophilum* isolates. Each category is graphed as a percentage of the total number of genes in the dispensable pool of genes. (d) COG subcategories of predicted genes within the accessory genome of *F. psychrophilum* isolates. Each category is graphed as a percentage of the total number of genes in the accessory pool of genes.

Following the analysis of the *F. psychrophilum* pan genome, we examined possible correlations between the virulence traits and genomic repertoire. First, we inferred the phylogeny of our *F. psychrophilum* strain collection by comparing 1866 core ORFs for each genomic sequence (Figure 2). The phylogenetic tree revealed distinct clustering of the

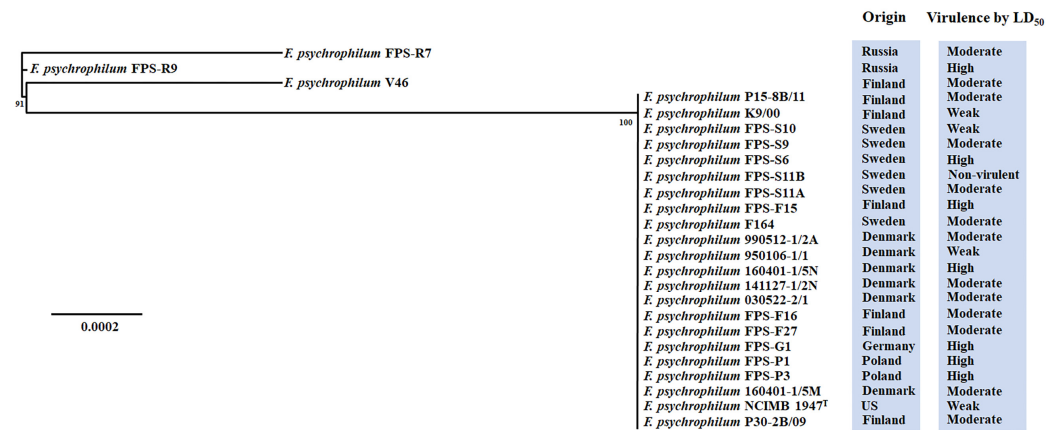
isolates, without clear correlations with the virulence properties or country of isolation (Figure 2). For example, *F. psychrophilum* isolates FPS-G1, FPS-S10, and FPS-S9 clustered together, but originated from two different countries and showed high, moderate, and low virulence, respectively (Figure 2; Table S1). However, some isolates clustered according to the geographical origin of isolation such as: Danish isolates 141127-1/2N, 160401-1/5N, and 160401-1/5M; Polish isolates FPS-P1 and FPS-P3; and Finnish isolates FPS-F15, FPS-F16, FPS-F27, and P15-8B/11 (Figure 2; highlighted colors). In addition, we investigated the dynamic nature of virulence-related gene repertoires across the *F. psychrophilum* isolates. One hundred and nine virulence-related genes had been previously identified in *F. psychrophilum* isolates (Table S2). More than 99% of these virulence-related genes were present in all isolates (Table S2). Genes involved in gliding motility, T9SS, metalloproteases, stress response, adhesion, and virulence were found in all isolates (Table S2). For example, we found a broad distribution of genes whose virulence had previously been confirmed in in vivo experiments: *exbD2* (id: E5164\_11170) of a TonB system [44], collagenase (id: E5164\_07005) [45], *fpgA* type-2 glycosyltransferase (id: E5164\_06305) [46], and the thiol oxidoreductase-like *tlpB* (id: E5164\_02660) [47]. Interestingly, the Swedish isolate FPS-S11B revealed a single mutation in the gliding motility gene *gldB* (id: H0I52\_08290), causing a premature stop codon in the amino acid sequence, which was previously linked to a complete loss of virulence [12].



**Figure 2.** Core genome phylogeny of *F. psychrophilum* isolates. The maximum likelihood tree was obtained from a concatenated amino acid sequence alignment of the orthologous core genes (1866 ORFs) for the twenty-five *F. psychrophilum* isolates and the type strain NCIMB 1947<sup>T</sup>. The numbers above the branches indicate the bootstrap value. Bootstrap values of <80% were removed from the tree. The horizontal bar at the base of the figure represents 0.01 substitution per amino acid site. The virulence properties of the isolates and geographical places of origin were added to facilitate comparison.

On the other hand, only the high-virulent isolates FPS-P1, FPS-R9, and FPS-P3, and the moderate-virulent isolate FPS-R7 had a multicopper oxidase (id: QRE04907.1). Moreover, the moderate-virulent isolates V46 and FPS-R7 harbored a multi antimicrobial extrusion pump (id: QRE04889.1) (Table S2). To reveal the evolution of the virulence, we inferred the genetic diversity of 109 representative virulence-related factors shared by the *F. psychrophilum* isolates (Figure 3). Only the high-virulent isolate FPS-R9 and moderate-virulent isolates V46 and FPS-R7 tended to cluster as mono-phylogenetic groups (Figure 3). The

remaining isolates, which presented from weak- to high-virulent phenotypes clustered together in the same phylogenetic group (Figure 3). These findings suggest an unclear association between virulence traits and well-shared virulence-related gene content in *F. psychrophilum*.



**Figure 3.** Phylogenetic tree of 109 concatenated virulence-related factors shared by all *F. psychrophilum* isolates. The phylogenetic tree was constructed based on the maximum likelihood algorithm, using a concatenated alignment of 109 amino acid sequences inferred from putative virulence-related factors identified in the pan genome analysis (Table S2). The numbers above the branches indicate the bootstrap value. The horizontal bar at the base of the figure represents 0.0002 substitutions per amino acid site. The virulence ranking and origin of isolation for each isolate was added to facilitate comparison.

Similarly, the dispensable genome did not indicate a coherent correlation with virulence traits. For example, two plasmids (3.3 and 2.1 kb) encoding a toxin-antitoxin system were found in both virulent (e.g., FPS-G1, FPS-F15) and weak or non-virulent isolates (950106-1/1 and FPS-S11B, respectively) (Table 1; Figure S2). Similarly, a 6H-like prophage element was found in seven out of 26 isolates (Table S1) [48]. On other hand, only high-virulent isolates FPS-G1 and FPS-P1 had a genomic island associated with transposase and tRNA-Asn (Figure S2).

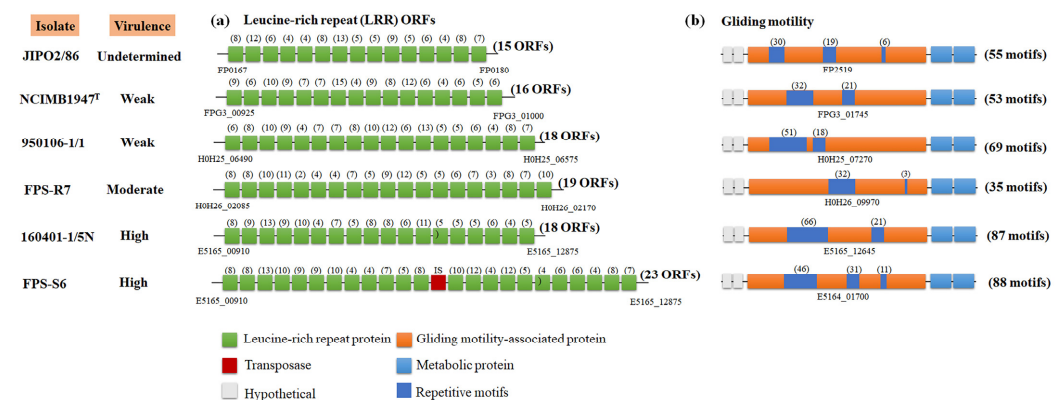
Finally, there was no clear link between virulence traits and the number of accessory genes. Instead, the virulence tended to be distributed among the different *F. psychrophilum* clusters (Figure 1a). For example, isolates FPS-G1 and FPS-F15 showed the highest virulence in the juvenile rainbow trout infection model and contained only three and zero accessory genes, respectively (Figure 1a). Contrarily, *F. psychrophilum* isolate FPS-R7, ranked as one of the lowest moderate-virulent isolates, had 258 accessory genes (Figure 1a). Interestingly, this specific isolate had ten genomic islands, being the GI-10 with a length of around 320 kb and encoding multicopper oxidase (id: H0H26\_04775), cysteine protease (id: H0H26\_04400),  $\beta$ -lactamase (id: H0H26\_04585), multidrug efflux pumps (id: H0H26\_04095- H0H26\_04110), and multiple hypothetical proteins (Table S3).

### 3.4. Precision Long-Read Sequencing for *F. psychrophilum*

Long-read sequencing approaches can produce completely closed genomes, which allow the assembly of complex genomic areas (e.g., transposon elements and highly repetitive regions) [49] and provide an opportunity to identify new genetic elements involved in virulence mechanisms, besides the gene loss function from single mutations. By combining shot-read Illumina and long-read Pac Bio technologies (hybrid assembly), we fully sequenced four *F. psychrophilum* isolates, which displayed different virulence phenotypes in the juvenile rainbow trout infection model. The selected *F. psychrophilum* isolates were the high-virulent isolates FPS-6 and 160401-1/5N, the moderate-virulent isolate FPS-R7 and the weak-virulent isolate 950106-1/1 (Table S1). The non-virulent isolate FPS-S11B was not included in the strains selected for Pac Bio sequencing as the loss of virulence in

this strain had previously been identified to be caused by a single mutation in the gene *gldB* [12].

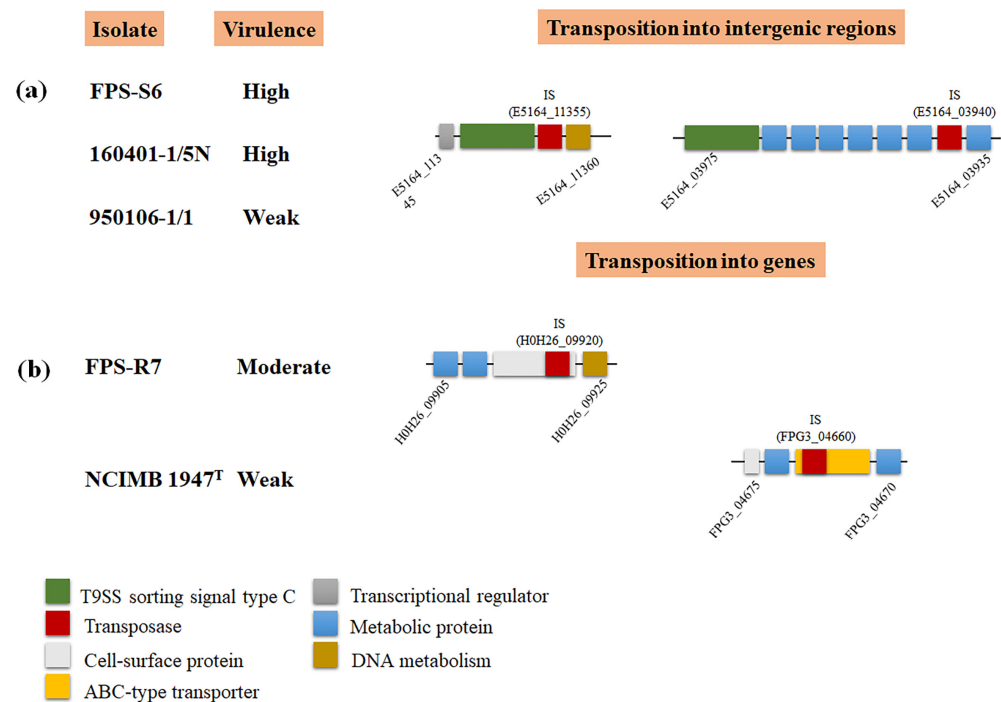
For Illumina data, we mapped the reads to the published reference genome *F. psychrophilum* JIP02/86 (Figure S3; see materials and methods). For all the isolates, we identified two genetic zones with very high coverage (from 1500× to 2200×) caused by repetitive DNA sequences, which had around 10× more coverage than the entire genome (Figure S3). The first zone was in a genetic region of 21 kb, which had leucine-rich repeat (LRR) proteins (Figure S3A), and the second region encoded a putative gliding motility-associated protein (Figure S3B). Based on the hybrid analysis, the *F. psychrophilum* isolates FPS-R7, FPS-S6, 950106-1/1, and 160401-1/5N showed a different number of LRRs in comparison to the reference strain JIP02/86 (Figure 4). For example, the high virulent isolates FPS-S6 and 160401-1/5N harbored 23 and 19 LRR proteins, respectively, and these isolates tended to have a higher number of LRRs than the moderate- and weak-virulent isolates FPS-R7 and 950106-1/1, which had 19 and 18 LRR ORFs, respectively, and the weak-virulent strain NCIMB 1947<sup>T</sup>, which had only 16 LRR ORFs (Figure 4a). Moreover, the putative gliding motility protein varied in repetitive motifs among the *F. psychrophilum* isolates (Figure 4b; Table S4). The high-virulent isolates FPS-S6 and 160401-1/5N had higher numbers of repetitive units, with a total of 87 and 88, respectively, whereas the moderate isolate FPS-R7 and the weak-virulent isolate 950106-1/1 had 35 and 69 units, respectively. The weak-virulent strain NCIMB 1947<sup>T</sup> had 55 motifs (Figure 4b).



**Figure 4.** Schematic distribution of repetitive genetic regions in *F. psychrophilum* isolates. (a) Localization of leucine-rich repeats identified by hybrid assembly of Illumina and PacBio sequencing technologies. (b) Representative amino acid repeat units in gliding motility-associated protein. Genes are represented as squares. Loci tags, isolates, and virulence phenotype are shown in the figure to facilitate comparison. The number of repetitive motifs is presented in parenthesis and their amino acid sequences are described in the Table S4.

The impact of mobile insertion elements (IS) on virulence via gene interruption has been described for *F. psychrophilum*. In the strain JIP02/86, the transposition of the IS element IS256 within the collagenase gene resulted in a pseudogene formation, probably inactivating the enzymatic activity [3,50]. Beyond this, by full genome comparison, we identified 62 transposases across the five *F. psychrophilum* genomic sequences (Table S5). Eighty percent of the transpositions of these elements were in intergenic regions. For example, the high-virulent isolates FPS-S6 and 160401-1/5N and the weak-virulent strain 950106-1/1 had an element transposition between metabolic (id: E5164\_11350) and T9SS sorting signal type C domain-containing (id: E5164\_11355) proteins (Figure 5a). Interestingly, the moderate-virulent strain FPS-R7 and the weak-virulent strain NCIMB 1947<sup>T</sup> were the only isolates with a transposition within ORFs that encoded cell-surface (id: H0H26\_09915) and antibiotic ABC transporter (id: FPG3\_04665) proteins, respectively (Figure 5b; Table S5).





**Figure 5.** Selected examples of transposon insertion in *F. psychrophilum* isolates. (a) Example of IS transposition into intergenic regions. (b) Example of IS in encoding genetic regions. Genes are represented as squares; insertions of transposases into genes are represented by red squares. Gene functional categories are shown by colored boxes. Isolates, loci tags, and virulence phenotype were added in the figure to facilitate comparison.

#### 4. Discussion

Bacterial pathogens exhibit significant variation in their genomic content of virulence-related factors [51]. This reflects the richness of strategies evolved by pathogens to infect host organisms [52]. In this study, we found that the nature and distribution of gene repertoires of *F. psychrophilum* isolates did not clearly reflect their virulence properties (Figure 1). The results presented here revealed a low overall genetic diversity within *F. psychrophilum* recovered from Baltic Sea countries (Figures 2 and 3). However, the complete genome sequences of four selected *F. psychrophilum* isolates, 950106-1/1, FPS-R7, FPS-S6, and 160401-1/5N, obtained using Illumina and PacBio hybrid assembly, demonstrated specific dissimilarities on leucine-rich repeats (LRR) ORFs, repetitive motifs in gliding motility-associated protein, and IS transposition in these isolates (Figures 4 and 5). These results suggest a link between these genetic variations and the modulation of virulence phenotypes, suggesting that these genetic regions may play an important role in the evolution and virulence of *F. psychrophilum*.

Pan-genome analyses have been an effective approach to understand pathogenic bacteria, allowing the association of genotype–phenotype profiles in specific pathogenic groups of bacteria [53,54]. Previous analysis of *F. psychrophilum* isolates from Baltic Sea countries [11] suggested that the core/pan-genome ratio (80% total gene content) was very close to those obtained from other geographical areas [8]. However, when combining genomic with virulence traits using LD<sub>50</sub> in this study, there was no correlation between the presence of accessory genes and the virulence level of the isolates (Figure 1; Table S1). Similarly, the core genome composition and the virulence-related gene distribution phylogeny were not associated with virulence properties (Figures 2 and 3). These results are opposite to the virulence patterns observed in *Vibrio anguillarum* [55], a pathogenic bacterium that resides in marine water and that can cause vibriosis in many fish and shellfish species [56]. The pan genome analysis of *V. anguillarum* showed a clear relationship between gene content and virulence, where the most pathogenic isolates possessed a unique

accessory genome, attributed to pathogenic genomic islands, toxin-carrying prophages, and virulence-related factors [55]. Although the results obtained in this study showed that in *F. psychrophilum* the accessory gene content was not linked with virulence traits by LD<sub>50</sub> (Figure 1), interesting genes were found in a genomic island of 320 kb in the moderately virulent *F. psychrophilum* isolate FPS-R7. This genomic island encoded multicopper oxidase, cysteine protease,  $\beta$ -lactamase, and multiple multidrug efflux pumps (Table S3). These genes have previously been connected to microbial pathogenesis, antibiotic resistance, and stress management [57–60]. However, their direct contribution in the mechanisms of pathogenicity of *F. psychrophilum* was not determined in this study; undoubtedly, these findings open new perspectives for the study of gene function in this bacterium.

The large-scale distribution and genetic homogeneity of core virulence-related factors (Table S2; Figure 3) supported the previous speculation that *F. psychrophilum* isolates all have a similar mode of pathogenicity, based on adhesion, colonization, and destruction of fish tissues [11]. An earlier report showed that all the isolates had in vitro proteolytic activities (e.g., gelatinase, caseinase, elastinase) and gliding motility except the *F. psychrophilum* isolate FPS-S11B [4], while another study established that this specific isolate had a premature stop codon in the gliding motility protein GldB, corroborating that gliding motility genes are linked to virulence-related properties in *F. psychrophilum* [12,13]. However, besides this finding, virulence-related gene profiling was not strictly linked to the virulence phenotype by LD<sub>50</sub> presented by the isolates (Table S1; Figure 3). Therefore, we propose that other genetic factors could be responsible for the virulence traits in *F. psychrophilum*.

Transposase elements are among the simplest mobile genetic elements and widespread in bacteria [61]. However, it is now clear that they play an important role as bacterial mutagenic agents, enabling the host to adapt to environmental challenges [62], colonize new niches [63], and modulate virulence [64]. In this study, we found that the *F. psychrophilum* isolates 950106-1/1, FPS-R7, FPS-S6, and 160401-1/5N harbored transposase elements inserted into genic and intergenic regions (Figures 4 and 5). The direct impact of these dynamic elements on the virulence of *F. psychrophilum* was not investigated in this study; however, it has been described that IS transposition into genic sequences can modulate biofilm formation and the production of extracellular polymeric substances in the pathogenic bacteria *Staphylococcus aureus* [65] and *Enterococcus faecalis* [66]. Moreover, we also found IS transposition into non-coding regions (Figure 5), a previously described mechanism for altering the expression of adjacent genes [61]. Most commonly, the effect of IS transposition is gene repression [67–69]; however, cases have been described illustrating gene activation [70]. Thus, we hypothesize that transposases may promote genetic and phenotypic variability in *F. psychrophilum*.

Repetitive motifs are known to occur in a wide variety of proteins present in bacteria [71]. These motifs represent an alternative module for the assembly of various multiprotein complexes, and thus, repetitive-containing proteins often participate in a wide range of functional roles, including virulence [19]. First, we found tandem leucine-rich repeat (LRR) proteins, which tended to increase in number according to the virulence phenotype of *F. psychrophilum* isolates (Figure 4a). These LRR proteins have been shown to be part of the microbial virulence-related factors involved in the interaction with host cells [72] and the invasion of mammalian host cells [73,74]. For example, the interactions of microbe–host and immune response in the human pathogenic bacteria *Streptococcus pyogenes* [75], *E. faecalis* [76], and *Listeria monocytogenes* [77] are mediated by cell surface proteins with conserved LRR motifs. Thus, we hypothesize that these LRR cell-surface proteins could be involved in the pathogenic mechanisms of *F. psychrophilum*. Second, we found a higher number of repetitive units in a gliding motility-associated protein in the high-virulent isolates 160401-1/5N and FPS-S6 (Figure 4b; Table S4). These gliding motility-related proteins have been described as facilitators of the biological function of the type IX secretion system (T9SS), which in pathogenic *Flavobacterium* species has been verified as a major virulence determinant, playing a role in motility and translocation of cell surface adhesins, peptidases, and other enzymes and virulence-related factors [7,13,78,79]. Several

studies have shown that disruption or mutations of the gliding motility-related genes resulted in defects in motility [80], extracellular enzymatic activity [6], and translocation of gliding motility proteins to the cell surface [81]; thus providing a link between the virulence and motility apparatus in pathogenic members of the *Bacteroidetes* phylum. Therefore, our finding allows us to speculate that repetitive motifs can alter both the structure and function of virulence-related proteins in *F. psychrophilum*. However, there is not a direct coupling between the number of LRR and the virulence phenotypes, and more research is required to define the evolutionary trends of the repeat motif features and their potential virulence functions in this fish pathogenic species.

## 5. Conclusions

Although diverse putative virulence-related genes have been identified in genomic sequences of *F. psychrophilum*, the role played by other genetic factors in the development of BCWS and RTFS diseases is still poorly understood [3,11]. Previous work has indicated that gliding motility and proteolytic activity are required for pathogenicity in *F. psychrophilum* [4]. Although, gene repertoire and comparative genomic analyses did not reveal a clear relationship between genotype and virulence traits in the current study (Figure 1), we found that variability in the number of consecutive leucine-rich repeat (LRR) proteins, repetitive motif dynamics, and IS transposition could be key factors in understanding the evolution and virulence of *F. psychrophilum* (Figures 4 and 5). This new information will form the foundation of future investigations into the pathogenicity mechanisms of *F. psychrophilum* and stimulate various experimental studies, including genetic manipulation by specific gene knock-out [82] or transposition [83], to fully understand the factors governing the virulence in this freshwater pathogen.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9081658/s1>, Figure S1: *F. psychrophilum* pan, core, and accessory genome evolution according to the number of sequenced genomes, Figure S2: Schematic representation of dispensable genome in *F. psychrophilum* isolates, Figure S3: Graphic representation of mapped reads in *F. psychrophilum* isolate JIPO82/6 as a reference genome, Table S1: Characteristics of *F. psychrophilum* isolates used in this study, Table S2: Distribution of putative virulence-related factors in *F. psychrophilum* isolates, Table S3: Distribution of genomic islands in *F. psychrophilum* isolates, Table S4: Amino acid sequences of repetitive motifs in *F. psychrophilum* isolates, Table S5: Distribution of transposase genes in *F. psychrophilum* isolates.

**Author Contributions:** Conceived the idea, D.C. and M.M.; performed bioinformatics and computational analysis, D.C. and V.L.D.; analyzed the data, D.C., V.L.D., J.J., K.S., L.M., I.D., T.W. and M.M. wrote the manuscript, D.C. and M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** BONUS FLAVOPHAGE project supported by BONUS (Art 185), funded jointly by the EU and Innovation Fund Denmark, the Academy of Finland, and the National Centre for Research and Development (NCBiR), Poland. This study was also supported by the Danish Council for Independent Research (Project # DFF-7014-00080).

**Data Availability Statement:** The genomic sequences of all sequenced strains have been deposited in the NCBI database under the accession numbers listed in Table S1.

**Acknowledgments:** We would like to thank Eva Jansson (Sweden), Bettina Schletz and Stephanie Bornstein (Germany), Agnieszka Pękala-Safińska (Poland), and Vladimir Voronin (Russia) for providing their isolates, and the fish farmers for allowing us to sample for bacterial isolation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Nematollahi, A.; Decostere, A.; Pasmans, F.; Haesebrouck, F. *Flavobacterium psychrophilum* infections in salmonid fish. *J. Fish Dis.* **2003**, *26*, 563–574. [[CrossRef](#)] [[PubMed](#)]
2. Barnes, M.E.; Brown, M.L. A review of *Flavobacterium psychrophilum* biology, clinical signs, and bacterial cold water disease prevention and treatment. *Open Fish Sci. J.* **2011**, *4*, 40–48. [[CrossRef](#)]

3. Duchaud, E.; Boussaha, M.; Loux, V.; Bernardet, J.F.; Michel, C.; Kerouault, B.; Mondot, S.; Nicolas, P.; Bossy, R.; Caron, C.; et al. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*. *Nat. Biotechnol.* **2007**, *25*, 763–769. [[CrossRef](#)] [[PubMed](#)]
4. Sundell, K.; Landor, L.; Nicolas, P.; Jørgensen, J.; Castillo, D.; Middelboe, M.; Dalsgaard, I.; Donati, V.L.; Madsen, L.; Wiklund, T. Phenotypic and genetic predictors of pathogenicity and virulence in *Flavobacterium psychrophilum*. *Front. Microbiol.* **2019**, *10*, 1711. [[CrossRef](#)]
5. Högfors-Rönholm, E.; Wiklund, T. Hemolytic activity in *Flavobacterium psychrophilum* is a contact-dependent, two-step mechanism and differently expressed in smooth and rough phenotypes. *Microb. Pathog.* **2010**, *49*, 369–375. [[CrossRef](#)]
6. Castillo, D.; Christiansen, R.H.; Dalsgaard, I.; Madsen, L.; Middelboe, M. Bacteriophage resistance mechanisms in the fish pathogen *Flavobacterium psychrophilum*: Linking genomic mutations to changes in bacterial virulence factors. *Appl. Environ. Microbiol.* **2015**, *81*, 1157–1167. [[CrossRef](#)]
7. Barbier, P.; Rochat, T.; Mohammed, H.H.; Wiens, G.D.; Bernardet, J.F.; Halpern, D.; Duchaud, E.; McBride, M.J. The type IX secretion system is required for virulence of the fish pathogen *Flavobacterium psychrophilum*. *Appl. Environ. Microbiol.* **2020**, *86*, e00799–e00820. [[CrossRef](#)]
8. Duchaud, E.; Rochat, T.; Habib, C.; Barbier, P.; Loux, V.; Guérin, C.; Dalsgaard, I.; Madsen, L.; Nilsen, H.; Sundell, K.; et al. Genomic diversity and evolution of the fish pathogen *Flavobacterium psychrophilum*. *Front. Microbiol.* **2018**, *9*, 138. [[CrossRef](#)]
9. Nilsen, H.; Sundell, K.; Duchaud, E.; Nicolas, P.; Dalsgaard, I.; Madsen, L.; Aspán, A.; Jansson, E.; Colquhoun, D.J.; Wiklund, T. Multilocus sequence typing identifies epidemic clones of *Flavobacterium psychrophilum* in Nordic countries. *Appl. Environ. Microbiol.* **2014**, *80*, 2728–2736. [[CrossRef](#)]
10. Castillo, D.; Christiansen, R.H.; Espejo, R.; Middelboe, M. Diversity and geographical distribution of *Flavobacterium psychrophilum* isolates and their phages: Patterns of susceptibility to phage infection and phage host range. *Microb. Ecol.* **2014**, *67*, 748–757. [[CrossRef](#)]
11. Castillo, D.; Christiansen, R.H.; Dalsgaard, I.; Madsen, L.; Espejo, R.; Middelboe, M. Comparative genome analysis provides insights into the pathogenicity of *Flavobacterium psychrophilum*. *PLoS ONE* **2016**, *11*, e0152515. [[CrossRef](#)]
12. Castillo, D.; Jørgensen, J.; Sundell, K.; Madsen, L.; Dalsgaard, I.; Wiklund, T.; Middelboe, M. Genome-informed approach to identify genetic determinants of *Flavobacterium psychrophilum* phage susceptibility. *Environ. Microbiol.* **2021**, in press.
13. Pérez-Pascual, D.; Rochat, T.; Kerouault, B.; Gómez, E.; Neulat-Ripoll, F.; Henry, C.; Quillet, E.; Guijarro, J.A.; Bernardet, J.F.; Duchaud, E. More than gliding: Involvement of GldD and GldG in the virulence of *Flavobacterium psychrophilum*. *Front. Microbiol.* **2017**, *8*, 2168. [[CrossRef](#)]
14. Stenholm, A.R.; Dalsgaard, I.; Middelboe, M. Isolation and characterization of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*. *Appl. Environ. Microbiol.* **2008**, *74*, 4070–4078. [[CrossRef](#)]
15. Castillo, D.; Higuera, G.; Villa, M.; Middelboe, M.; Dalsgaard, I.; Madsen, L.; Espejo, R.T. Diversity of *Flavobacterium psychrophilum* and the potential use of its phages for protection against bacterial cold water disease in salmonids. *J. Fish Dis.* **2012**, *35*, 193–201. [[CrossRef](#)]
16. Donati, V.L.; Dalsgaard, I.; Sundell, K.; Castillo, D.; Er-Rafik, M.; Clark, J.; Wiklund, T.; Middelboe, M.; Madsen, L. Phage-mediated control of *Flavobacterium psychrophilum* in aquaculture: In vivo experiments to compare delivery methods. *Front. Microbiol.* **2021**, *12*, 628309. [[CrossRef](#)] [[PubMed](#)]
17. Castillo, D.; Middelboe, M. Genomic diversity of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*. *FEMS Microbiol. Lett.* **2016**, *363*, fnw272. [[CrossRef](#)]
18. Gyles, C.; Boerlin, P. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Vet. Pathol.* **2014**, *51*, 328–340. [[CrossRef](#)] [[PubMed](#)]
19. Cervený, L.; Strásková, A.; Danková, V.; Hartlova, A.; Cecková, M.; Staud, F.; Stulík, J. Tetratricopeptide repeat motifs in the world of bacterial pathogens: Role in virulence mechanisms. *Infect. Immun.* **2013**, *81*, 629–635. [[CrossRef](#)]
20. Bliven, K.A.; Maurelli, A.T. Antivirulence genes: Insights into pathogen evolution through gene loss. *Infect. Immun.* **2012**, *80*, 4061–4070. [[CrossRef](#)] [[PubMed](#)]
21. Wu, A.K.; Kropinski, A.M.; Lumsden, J.S.; Dixon, B.; MacInnes, J.I. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum* ATCC 49418(T). *Stand. Genomic Sci.* **2015**, *10*, 3. [[CrossRef](#)]
22. Holt, R.A.; Rohovec, J.S.; Fryer, J.L. Bacterial cold-water disease. In *Bacterial Disease of Fish*, 1st ed.; Inglis, V., Roberts, R.J., Bromage, N.R., Eds.; Blackwell Scientific Publications: Oxford, UK, 1993; pp. 3–22.
23. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)]
24. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [[CrossRef](#)] [[PubMed](#)]
25. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. In *Gene Prediction. Methods in Molecular Biology*, 1st ed.; Kollmar, M., Ed.; Springer Protocols: New York, NY, USA, 2019; Volume 1962, pp. 227–245.
26. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [[CrossRef](#)] [[PubMed](#)]
27. Ruiqi, L.; Yingrui, L.; Karsten, K.; Jun, W. SOAP: Short oligonucleotide alignment program. *Bioinformatics* **2008**, *24*, 713–714.

28. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [[CrossRef](#)] [[PubMed](#)]
29. Bertelli, C.; Laird, M.R.; Williams, K.P.; Simon Fraser University Research Computing Group; Lau, B.Y.; Hoad, G.; Winsor, G.L.; Brinkman, F.S.L. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **2017**, *45*, W30–W35. [[CrossRef](#)]
30. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [[CrossRef](#)]
31. Zhou, C.E.; Smith, J.; Lam, M.; Zemla, A.; Dyer, M.D.; Slezak, T. MvirDB a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res.* **2007**, *35*, D391–D394. [[CrossRef](#)] [[PubMed](#)]
32. Kleinheinz, K.A.; Joensen, K.G.; Larsen, M.V. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage* **2014**, *4*, e27943. [[CrossRef](#)]
33. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genom.* **2008**, *9*, 75. [[CrossRef](#)]
34. Arndt, D.; Marcu, A.; Liang, Y.; Wishart, D.S. Phast, phaster and phastest: Tools for finding prophage in bacterial genomes. *Brief. Bioinform.* **2019**, *20*, 1560–1567. [[CrossRef](#)]
35. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **2010**, *26*, 1608–1615. [[CrossRef](#)]
36. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)]
37. Emanuelsson, O.; Brunak, S.; von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953–971. [[CrossRef](#)] [[PubMed](#)]
38. Pagni, M.; Ioannidis, V.; Cerutti, L.; Zahn-Zabal, M.; Jongeneel, C.V.; Hau, J.; Martin, O.; Kuznetsov, D.; Falquet, L. MyHits: Improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.* **2007**, *35*, W433–W437. [[CrossRef](#)] [[PubMed](#)]
39. Blom, J.; Kreis, J.; Spänig, S.; Juhre, T.; Bertelli, C.; Ernst, C.; Goesmann, A. EDGAR 2.0: An enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* **2016**, *44*, W22–W28. [[CrossRef](#)] [[PubMed](#)]
40. Chen, F.; Mackey, A.J.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* **2006**, *34*, D363–D368. [[CrossRef](#)]
41. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
42. Kück, P.; Meusemann, K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenetics Evol.* **2010**, *56*, 1115–1118. [[CrossRef](#)] [[PubMed](#)]
43. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)] [[PubMed](#)]
44. Álvarez, B.; Álvarez, J.; Menéndez, A.; Guijarro, J.A. A mutant in one of two *exbD* loci of a TonB system in *Flavobacterium psychrophilum* shows attenuated virulence and confers protection against cold water disease. *Microbiology* **2008**, *154*, 1144–1151. [[CrossRef](#)]
45. Nakayama, H.; Tanaka, K.; Teramura, N.; Hattori, S. Expression of collagenase in *Flavobacterium psychrophilum* isolated from cold-water disease-affected ayu (*Plecoglossus altivelis*). *Biosci. Biotechnol. Biochem.* **2016**, *80*, 135–144. [[CrossRef](#)]
46. Pérez-Pascual, D.; Gómez, E.; Guijarro, J.A. Lack of a type-2 glycosyltransferase in the fish pathogen *Flavobacterium psychrophilum* determines pleiotropic changes and loss of virulence. *Vet. Res.* **2015**, *46*, 1–9. [[CrossRef](#)]
47. Alvarez, B.; Secades, P.; Prieto, M.; McBride, M.J.; Guijarro, J.A. A mutation in *Flavobacterium psychrophilum* *tlpB* inhibits gliding motility and induces biofilm formation. *Appl. Environ. Microbiol.* **2006**, *72*, 4044–4053. [[CrossRef](#)]
48. Castillo, D.; Espejo, R.; Middelboe, M. Genomic structure of bacteriophage 6H and its distribution as prophage in *Flavobacterium psychrophilum* strains. *FEMS Microbiol. Lett.* **2014**, *351*, 51–58. [[CrossRef](#)]
49. Loman, N.J.; Quick, J.; Simpson, J.T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **2015**, *12*, 733–735. [[CrossRef](#)] [[PubMed](#)]
50. Sundell, K.; Wiklund, T. Characteristics of epidemic and sporadic *Flavobacterium psychrophilum* sequence types. *Aquaculture* **2015**, *441*, 51–56. [[CrossRef](#)]
51. Leitão, J.H. Microbial Virulence Factors. *Int. J. Mol. Sci.* **2020**, *21*, 5320. [[CrossRef](#)]
52. Diard, M.; Hardt, W.D. Evolution of bacterial virulence. *FEMS Microbiol. Rev.* **2017**, *41*, 679–697. [[CrossRef](#)] [[PubMed](#)]
53. Rouli, L.; Merhej, V.; Fournier, P.E.; Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **2015**, *7*, 72–85. [[CrossRef](#)]
54. Kim, Y.; Gu, C.; Kim, H.U.; Lee, S.Y. Current status of pan-genome analysis for pathogenic bacteria. *Curr. Opin. Biotechnol.* **2020**, *63*, 54–62. [[CrossRef](#)]
55. Castillo, D.; Alvise, P.D.; Xu, R.; Zhang, F.; Middelboe, M.; Gram, L. Comparative genome analyses of *Vibrio anguillarum* strains reveal a link with pathogenicity traits. *Msystems* **2017**, *28*, e00001-17. [[CrossRef](#)] [[PubMed](#)]

56. Frans, I.; Michiels, C.W.; Bossier, P.; Willems, K.A.; Lievens, B.; Rediers, H. *Vibrio anguillarum* as a fish pathogen: Virulence factors, diagnosis and prevention. *J. Fish Dis.* **2011**, *34*, 643–661. [[CrossRef](#)] [[PubMed](#)]
57. Kaur, K.; Sharma, A.; Capalash, N.; Sharma, P. Multicopper oxidases: Biocatalysts in microbial pathogenesis and stress management. *Microbiol. Res.* **2019**, *222*, 1–13. [[CrossRef](#)]
58. Stockbauer, K.E.; Magoun, L.; Liu, M.; Burns, E.H., Jr.; Gubba, S.; Renish, S.; Pan, X.; Bodary, S.C.; Baker, E.; Coburn, J.; et al. A natural variant of the cysteine protease virulence factor of group A *Streptococcus* with an arginine-glycine-aspartic acid (RGD) motif preferentially binds human integrins  $\alpha$ 5 $\beta$ 1 and  $\alpha$ IIb $\beta$ 3. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 242–247. [[CrossRef](#)]
59. Gharrah, M.M.; Mostafa El-Mahdy, A.; Barwa, R.F. Association between virulence factors and extended spectrum beta-lactamase producing *Klebsiella pneumoniae* compared to nonproducing isolates. *Interdiscip. Perspect. Infect. Dis.* **2017**, *2017*, 7279830. [[CrossRef](#)]
60. Alcalde-Rico, M.; Hernando-Amado, S.; Blanco, P.; Martínez, J.L. Multidrug efflux pumps at the crossroad between antibiotic resistance and bacterial virulence. *Front. Microbiol.* **2016**, *7*, 1483. [[CrossRef](#)]
61. Vandecraen, J.; Chandler, M.; Aertsen, A.; Van Houdt, R. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* **2017**, *43*, 709–730. [[CrossRef](#)] [[PubMed](#)]
62. Rojo-Bezares, B.; Estepa, V.; Cebollada, R.; de Toro, M.; Somalo, S.; Seral, C.; Castillo, F.J.; Torres, C.; Saenz, Y. Carbapenem resistant *Pseudomonas aeruginosa* strains from a Spanish hospital: Characterization of metallo-beta-lactamases, porin OprD and integrons. *Int. J. Med. Microbiol.* **2014**, *304*, 405–414. [[CrossRef](#)]
63. Benson, M.A.; Ohneck, E.A.; Ryan, C.; Alonzo III, F.; Smith, H.; Narechania, A.; Kolokotronis, S.O.; Satola, S.W.; Uhlemann, A.C.; Sebra, R.; et al. Evolution of hypervirulence by a MRSA clone through acquisition of a transposable element. *Mol. Microbiol.* **2014**, *93*, 664–681. [[CrossRef](#)]
64. Christie-Oleza, J.A.; Nogales, B.; Martin-Cardona, C.; Lanfranconi, M.P.; Alberti, S.; Lalucat, J.; Bosch, R. ISPst9, an ISL3-like insertion sequence from *Pseudomonas stutzeri* AN10 involved in catabolic gene inactivation. *Int. Microbiol.* **2008**, *11*, 101–110.
65. Arciola, C.R.; Campoccia, D.; Ravaioli, S.; Montanaro, L. Polysaccharide intercellular adhesin in biofilm: Structural and regulatory aspects. *Front. Cell. Infect. Microbiol.* **2015**, *5*, 7. [[CrossRef](#)]
66. Perez, M.; Calles-Enriquez, M.; del Rio, B.; Ladero, V.; Martin, M.C.; Fernandez, M.; Alvarez, M.A. IS256 abolishes gelatinase activity and biofilm formation in a mutant of the nosocomial pathogen *Enterococcus faecalis* V583. *Can. J. Microbiol.* **2015**, *61*, 517–519. [[CrossRef](#)]
67. Graves, J.L., Jr.; Tajkarimi, M.; Cunningham, Q.; Campbell, A.; Nonga, H.; Harrison, S.H.; Barrick, J.E. Rapid evolution of silver nanoparticle resistance in *Escherichia coli*. *Front. Genet.* **2015**, *6*, 42. [[CrossRef](#)]
68. Sun, X.; Dennis, J.J. A novel insertion sequence derepresses efflux pump expression and preadapts *Pseudomonas putida* S12 for extreme solvent stress. *J. Bacteriol.* **2009**, *191*, 6773–6777. [[CrossRef](#)]
69. Cannatelli, A.; Giani, T.; D’Andrea, M.M.; Di Pilato, V.; Arena, F.; Conte, V.; Tryfinopoulou, K.; Vatopoulos, A.; Rossolini, G.M.; Group CS. MgrB inactivation is a common mechanism of colistin resistance in KPC-producing *Klebsiella pneumoniae* of clinical origin. *Antimicrob. Agents Chemother* **2014**, *58*, 5696–5703. [[CrossRef](#)]
70. Boutoille, D.; Corvec, S.; Caroff, N.; Giraudeau, C.; Espaze, E.; Caillon, J.; Plésiat, P.; Reynaud, A. Detection of an IS21 insertion sequence in the *mexR* gene of *Pseudomonas aeruginosa* increasing  $\beta$ -lactam resistance. *FEMS Microbiol. Lett.* **2004**, *230*, 143–146. [[CrossRef](#)]
71. Kajava, A.V. Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* **2012**, *179*, 279–288. [[CrossRef](#)]
72. Kedzierski, Ł.; Montgomery, J.; Curtis, J.; Handman, E. Leucine-rich repeats in host-pathogen interactions. *Arch. Immunol. Exp.* **2004**, *52*, 104–112.
73. Parida, S.K.; Domann, E.; Rohde, M.; Muller, S.; Darji, A.; Hain, T.; Wehland, J.; Chakraborty, T. Internalin B is essential for adhesion and mediates the invasion of *Listeria monocytogenes* into human endothelial cells. *Mol. Microbiol.* **1998**, *28*, 81–93. [[CrossRef](#)]
74. Haraga, A.; Miller, S.I. A *Salmonella enterica* serovar typhimurium translocated leucine-rich repeat effector protein inhibits NF- $\kappa$ B-dependent gene expression. *Infect. Immun.* **2003**, *71*, 4052–4058. [[CrossRef](#)]
75. Loimaranta, V.; Hytönen, J.; Pulliainen, A.T.; Sharma, A.; Tenovuo, J.; Strömberg, N.; Finne, J. Leucine-rich repeats of bacterial surface proteins serve as common pattern recognition motifs of human scavenger receptor gp340. *J. Biol. Chem.* **2009**, *284*, 18614–18623. [[CrossRef](#)]
76. Brinster, S.; Posteraro, B.; Bierne, H.; Alberti, A.; Makhzami, S.; Sanguinetti, M.; Serror, P. Enterococcal leucine-rich repeat-containing protein involved in virulence and host inflammatory response. *Infect. Immun.* **2007**, *75*, 4463–4471. [[CrossRef](#)]
77. Marino, M.; Braun, L.; Cossart, P.; Ghosh, P. Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. *Mol. Cell* **1999**, *4*, 1063–1072. [[CrossRef](#)]
78. McBride, M.J.; Zhu, Y. Gliding motility and Por secretion system genes are widespread among members of the phylum Bacteroidetes. *J. Bacteriol.* **2013**, *195*, 270–278. [[CrossRef](#)]
79. Lasica, A.M.; Ksiazek, M.; Madej, M.; Potempa, J. The type IX secretion system (T9SS): Highlights and recent insights into its structure and function. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 215. [[CrossRef](#)]
80. Rhodes, R.G.; Nelson, S.S.; Pochiraju, S.; McBride, M.J. *Flavobacterium johnsoniae* sprB is part of an operon spanning the additional gliding motility genes *sprC*, *sprD*, and *sprF*. *J. Bacteriol.* **2011**, *193*, 599–610. [[CrossRef](#)]

81. Sato, K.; Naito, M.; Yukitake, H.; Hirakawa, H.; Shoji, M.; McBride, M.J.; Rhodes, R.G.; Nakayama, K. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 276–281. [[CrossRef](#)]
82. Liu, D.; Zhang, T.; Wang, Y.; Muhammad, M.; Xue, W.; Ju, J.; Zhao, B. Knockout of alanine racemase gene attenuates the pathogenicity of *Aeromonas hydrophila*. *BMC Microbiol.* **2019**, *19*, 72. [[CrossRef](#)]
83. Alvarez, B.; Secades, P.; McBride, M.J.; Guijarro, J.A. Development of genetic techniques for the psychrotrophic fish pathogen *Flavobacterium psychrophilum*. *Appl. Environ. Microbiol.* **2004**, *70*, 581–587. [[CrossRef](#)] [[PubMed](#)]