

RESEARCH ARTICLE

# Comparative Genome Analysis Provides Insights into the Pathogenicity of *Flavobacterium psychrophilum*

Daniel Castillo<sup>1</sup>, Rói Hammershaimb Christiansen<sup>1,2</sup>, Inger Dalsgaard<sup>2</sup>, Lone Madsen<sup>2</sup>, Romilio Espejo<sup>3</sup>, Mathias Middelboe<sup>1\*</sup>

**1** Marine Biological Section, University of Copenhagen, Helsingør, Denmark, **2** National Veterinary Institute, Technical University of Denmark, Frederiksberg, Denmark, **3** Centro Nacional de Genómica y Bioinformática and Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Macul, Santiago, Chile

\* [mmiddelboe@bio.ku.dk](mailto:mmiddelboe@bio.ku.dk)



OPEN ACCESS

**Citation:** Castillo D, Christiansen RH, Dalsgaard I, Madsen L, Espejo R, Middelboe M (2016) Comparative Genome Analysis Provides Insights into the Pathogenicity of *Flavobacterium psychrophilum*. PLoS ONE 11(4): e0152515. doi:10.1371/journal.pone.0152515

**Editor:** Yung-Fu Chang, Cornell University, UNITED STATES

**Received:** February 2, 2016

**Accepted:** March 15, 2016

**Published:** April 12, 2016

**Copyright:** © 2016 Castillo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Genomes are publicly available and accession numbers are given in Supporting Information.

**Funding:** This work was supported by The Danish Directorate for Food, Fisheries and Agri Business, grant # 3414-09-02613; The Danish Strategic Research Council (ProAqua, grant # 09-072829); and EU-IRSES (AQUAPHAGE, grant #69175). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

*Flavobacterium psychrophilum* is a fish pathogen in salmonid aquaculture worldwide that causes cold water disease (CWD) and rainbow trout fry syndrome (RTFS). Comparative genome analyses of 11 *F. psychrophilum* isolates representing temporally and geographically distant populations were used to describe the *F. psychrophilum* pan-genome and to examine virulence factors, prophages, CRISPR arrays, and genomic islands present in the genomes. Analysis of the genomic DNA sequences were complemented with selected phenotypic characteristics of the strains. The pan genome analysis showed that *F. psychrophilum* could hold at least 3373 genes, while the core genome contained 1743 genes. On average, 67 new genes were detected for every new genome added to the analysis, indicating that *F. psychrophilum* possesses an open pan genome. The putative virulence factors were equally distributed among isolates, independent of geographic location, year of isolation and source of isolates. Only one prophage-related sequence was found which corresponded to the previously described prophage 6H, and appeared in 5 out of 11 isolates. CRISPR array analysis revealed two different loci with dissimilar spacer content, which only matched one sequence in the database, the temperate bacteriophage 6H. Genomic Islands (GIs) were identified in *F. psychrophilum* isolates 950106-1/1 and CSF 259–93, associated with toxins and antibiotic resistance. Finally, phenotypic characterization revealed a high degree of similarity among the strains with respect to biofilm formation and secretion of extracellular enzymes. Global scale dispersion of virulence factors in the genomes and the abilities for biofilm formation, hemolytic activity and secretion of extracellular enzymes among the strains suggested that *F. psychrophilum* isolates have a similar mode of action on adhesion, colonization and destruction of fish tissues across large spatial and temporal scales of occurrence. Overall, the genomic characterization and phenotypic properties may provide new insights to the mechanisms of pathogenicity in *F. psychrophilum*.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

*Flavobacterium psychrophilum* is a yellow-pigmented, Gram-negative fish pathogen with a global distribution, causing "cold water disease" (CWD) and "rainbow trout fry syndrome" (RTFS) in freshwater aquaculture [1]. The disease results in high rates of juvenile mortality, increased predisposition to other infections and high costs of treatment with antibiotics causing significant economic losses for salmonid aquaculture worldwide [2]. Historically, the first isolation of *F. psychrophilum* was described in USA in 1960 [1] and isolation of *F. psychrophilum* as a disease-causing agent in fish farms has been reported since 1980's in several European countries [3–5], Canada [6] and in Chile and Japan [7, 8].

Disease outbreaks can result in necrotic lesions, partial dark skin colorizing, exophthalmia, anemia, ascites, and vertebral deformities of the fish [1]. Although, pathogenesis still needs to be elucidated, proteases [9, 10], adhesins [11], secretion systems [12] and biofilm formation [13] have been proposed to play a role in pathogenicity. Treatment with antibiotics (oxolinic acid, sulfadiazine/trimethoprim, florfenicol) is still required to decrease economic losses [14], and a vaccine is currently at an early stage of development [15]. More recently, the application of specific *F. psychrophilum* phages to reduce the pathogen population has been suggested as an alternative method for treatment of RFTS and CWD [16, 17].

The major typing methods used to distinguish between specific *F. psychrophilum* strains include ribotyping, plasmid profiling and serotyping [18], multilocus sequence typing MLST [19, 20], direct genome restriction enzyme analysis (DGREA) [21] and pulsed field gel electrophoresis (PFGE) [22]. Use of these typing methods revealed a relatively low genetic diversity of *F. psychrophilum* on local and global scales, possibly due to efficient dispersal of the pathogen along with the massive import of fish eggs across the fish-producer countries [21, 23, 24]. From a genomics perspective, the first complete genome sequence of the virulent *F. psychrophilum* strain JIPO2/86 (ATCC 49511) revealed a 2,861,988-bp circular chromosome with 2432 predicted protein-coding genes. Among these predicted proteins, stress response mediators, gliding motility proteins, adhesins and putative secreted proteases are probably involved in the pathogenesis of the bacterium [12]. In addition, one Danish *F. psychrophilum* isolate (strain 950106-1/1) was sequenced recently, displaying the presence of novel protein secretion system, the Por secretion system (PorSS) and the secretion of extracellular enzymes at *in vitro* conditions, which was suggested to participate in adhesion, colonization and destruction of the fish tissues [13]. However, despite the economic and immediate importance of this bacterium, only a few pathogenic isolates have been sequenced and reported in the literature [12, 13, 25].

Comparative genomic analyses lend insight into structural features such as variations related to genomic rearrangements, changes in the gene repertoire, identification of horizontal gene transfer elements and prophage-related sequences, and hence expose particularities on the evolution in this species [26]. These analyses have defined a conserved "core" genome shared among nearly all members of the species interspersed with "accessory" genomic elements that are present in some but absent in other strains [27]. The genome sequencing of strains belonging to the same species offers the possibility of defining their pan-genome, which comprises the core-genome and the accessory genome compartment. This strain-specific accessory genome may also be involved in critical activities of pathogenicity, drug resistance, and stress responses. While these factors may increase the adaptability of pathogens to the particular niches they inhabit, they are not imperative to the survival of the organism. Moreover, some of these genes can be acquired by horizontal gene transfer and have also been shown to be over represented in genomic islands [28]. Such pan-genome analyses have previously proved useful in identification of virulence factors in *Escherichia coli* [29] and *Streptococcus agalactiae* [30]. Thus in the present work, we have sequenced, annotated, and compared the

genomic DNA sequences of six *F. psychrophilum* isolates from Chile and compared those with genomes of five sequenced *F. psychrophilum* isolates from USA, Denmark and France, allowing a comparison of isolates covering a large time scale of isolation (>65 years) and a variety of sources and geographic locations (spatial scale of isolation >12000 km). We focused specifically on characterizing the distribution of potential virulence factors, prophage content and the pan-genome in *F. psychrophilum*, in order to get insights into the mechanisms of pathogenicity and to establish differences in gene content that could contribute to physiological variability among the isolates.

## Materials and Methods

### Strain selection, medium composition and growth conditions

This study used 11 *F. psychrophilum* strains isolated from trout aquaculture in different geographic localities in Denmark, France, Chile and USA, covering a spatial scale of >12000 km and temporal scale of >60 years (S1 Table). *F. psychrophilum* strains MH1, PG2, VQ50, 3, 4 and 5 were isolated from Chile ([17]; this study). The strains 3, 4 and 5 were isolated from a private fish farm in Puerto Montt, Chile, with permission of the owner of the farm. Strain 950106-1/1 was isolated in Denmark [16], strains CSF259-93, FPG101 and FPG3 (ATCC 49418) were isolated in North America [6, 30] and strain JIP02/86 was isolated in France [31] (S1 Table). The isolates were stored at -80°C in broth TYES-B (tryptone 0.4%, yeast extract 0.04%, CaCl<sub>2</sub> 0.05% and MgSO<sub>4</sub> 0.05%) with 15% glycerol. For culturing of the *F. psychrophilum* strains, cells were inoculated in TYES-B medium and incubated at 15°C with agitation for 48–72 hours [16].

*F. psychrophilum* isolates MH1, PG2, VQ50, 3, 4 and 5 were selected for whole genome sequencing and compared with the previously sequenced strains 950106-1/1, JIP02/86, CSF 259–93, FPG3 and FPG101. Strains MH1, PG2, VQ50, 3, 4, 5, CSF 259–93, FPG3 and FPG101 were selected for phenotypic analysis (see below).

### DNA extraction

Bacterial DNA from Chilean *F. psychrophilum* isolates MH1, PG2, VQ50, 3, 4 and 5 were extracted from pelleted and purified isolates using the QIAamp DNA mini Kit (QIAGEN) according to manufacturer's protocol. The amount of genomic DNA was measured using a Nanodrop2000 UV-Vis Spectrophotometer (Thermo scientific).

### Sequencing, assembly and annotation

The genomic DNA sequences of the Chilean *F. psychrophilum* strains MH1, PG2, VQ50, 3, 4 and 5 were obtained using Illumina HiSeq platform (BGI, China) with pair-end read sizes of 100 bp. Library construction, sequencing, and data pipelining were performed in accordance with manufacturer's protocols. The Illumina data were assembled into contiguous sequences using Geneious software version 7.1.4, then short and low-coverage contigs were filtered out. For strains MH1, VQ50, PG2, 3, and 5 the remaining contigs were aligned using the previously sequenced *F. psychrophilum* strain JIPO2/86 (ATCC 49511) as reference genome (GenBank accession number: AM398681; August 2013). The DNA sequences were assembled into one scaffold with an average coverage >76x for each isolate. Gaps of unknown length (100 N's) were added adjacent to the last base, indicating a not-connection between this base and the first base. In addition, for strains MH1 and VQ50, two contigs were identified (3211 bp and 2567 bp respectively; S2 Table), which did not align with the reference genome. These contigs were localized at the end of the DNA sequences using 100 N's to represent gaps of unknown length, according to the gapped format for genome submissions specified in NCBI database.

Thus, these *F. psychrophilum* genomes were assigned as draft genomic DNA sequences (incomplete chromosome) (S2 Table). For strain 4, the number of ambiguous bases and gaps in the genomic DNA sequence was relatively high, and the 23 contigs identified in this strain were therefore submitted to the NCBI database as a Whole Genome Shotgun (WGS) project (WGS). Finally, annotation of the contigs was achieved by the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) [32].

## Predictions of pathogenic and resistance islands, virulence factors and prophages

We used PAI finder [33] and PIPS [34] to predict the putative genomic islands (GIs) and antimicrobial resistance islands (REIs) in *F. psychrophilum* isolates. In addition, a genome comparison carried out by MAUVE [35] was used to identify GIs which were not recognized by PAI finder or PIPS. The virulence database MvirDB [36] was used to predict putative virulence factors. All predicted genes of the 11 *F. psychrophilum* isolates were searched against the MvirDB by blastp with loose criteria (E-value  $\geq 1e-5$ ; identity  $\geq 35\%$ ; coverage  $\geq 80\%$ ). Also, virulencefinder 1.2 [37] was used to screen putative virulence factors using selected database from *Escherichia coli*, *Enterococcus* and *Streptococcus aureus*. Prophage-related sequences were identified by running bacterial genomes in phage\_finder v2.1 [38] and PHAST [39].

## CRISPR array detection

The putative CRISPR loci for 11 isolates were identified with CRISPRfinder [40]. Spacer sequences were aligned to the whole genome sequenced *F. psychrophilum* bacteriophages FpV4 (90 kb), FpV9 (48 kb), FpV21 (90 kb) and 6H [41] using ClustalW algorithm in Geneious 7.1.4 program [42].

## Pan genome analysis

In order to predict the possible genomic changes in *F. psychrophilum*, the bioinformatics program EDGAR [43] was used to predict pan genome of all 11 *F. psychrophilum* isolates and calculate the pan-genome (gene repertoire), accessory genome (specific genes, only found in one genome) and core genome (common genes, mutually conserved). Pan-genome development of was calculated by iterative pairwise comparison of a set of genomes. Using the metacontig function of EDGAR, we also defined custom groups of *F. psychrophilum* genomes for which the core genome or the pan genome have been stored as virtual contigs [43].

## 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 11 isolates (1426 genes present in a single copy, paralogs not included) and *F. columnare* ATCC 49512 (outgroup to root the tree) using OrthoMCL with an e-value cut off  $10^{-10}$  [44]. The set of 1426 single core genes were first aligned at amino acid level using Clustal W version 2.0 [45], then back translated to DNA sequences using PAL2NAL [46]. The alignment of all orthologous genes was concatenated using FASconCAT [47]. Gene tree was constructed using PhyML [48].

## Biofilm formation

Biofilm formation was quantified using the standard assay with crystal violet staining of biofilm and subsequent measurement of the optical density at 595 nm as previously described for *F. psychrophilum* [12]. The ten isolates were grown in half strength TYES-B to mid-exponential

phase ( $10^8$  cells/ml). *F. psychrophilum* phage resistant isolate V1-20 was used as a negative control [13]. The cultures were diluted 1/100 in TYES-B and then 1 ml of each dilution was inoculated in quadruplicate into polystyrene tubes (Becton Dickson, Falcon), which were incubated statically at 15°C for 5 days. Following this incubation, the supernatants were discarded, and the tubes were washed seven times with 1 ml of sterile distilled water. Then 1 ml of a 1% (wt/vol) crystal violet solution (Sigma-Aldrich) was added to each polystyrene tube containing the cells. After 45 min, the crystal violet solution was removed, the wells were washed seven times with 1 ml of sterile distilled water to remove the unbound dye, and then 1 ml of 96% (vol/vol) ethanol was added to release attached cells from the tube surface. Biofilm formation was then quantified by measuring the optical density at 595 nm. Experiments were repeated in four independent assays.

### Motility

The ability of the ten isolates to move by gliding were examined by inspection of motility on glass microslides [13], by using the hanging drop technique [49], as well as by using agar plates as surfaces for the gliding motility [50]. Briefly, all the isolates were grown in TYES-B for 48–72 h at 15°C. Glass capillary microslides (Camlab) were introduced in the bacterial cultures and motility was observed using phase contrast microscopy (Olympus BX61). In an additional approach, one drop of liquid culture of each strain was deposited onto a cover slip, which was turned upside down and placed on tiny stands on a glass slide. Bacterial motility was observed through the cover slip. Finally, aliquots of 5  $\mu$ L bacterial culture were spread on plates with 2x, 1x, 0.5x, and 0.1x and 0.01x diluted TYES-B agar (1.1%). After 72 h incubation, the colony diameters were measured as an estimate of dispersal rate by gliding motility. Experiments were repeated in four independent assays.

### Secretion of extracellular proteins

The activities of the extracellular enzymes proteolysins, and gelatinase were measured in each strain as previously described for *F. psychrophilum*, including the phage resistant isolate V1-20 [13]. 50  $\mu$ l aliquots of bacteria-free supernatant (0.45- $\mu$ m-filtration) from three replicates of a 4–5 day liquid culture of each isolate with cell densities were added to holes punched into agar plates. For total proteolytic activity, cells were added to agar plates containing 2% skim milk and plates were incubated at 15°C overnight. Gelatinase activity was determined using 2% gelatine plates and incubated at 4°C. Experiments were repeated in four independent assays.

### Hemolytic activity

Bacterial hemolytic activity was assessed using the microplate hemolysis assay described by Högfors-Rönholm and Wiklund [51]. Blood was collected in an equal volume of Alsever's solution (Sigma Aldrich) by caudal venipuncture of rainbow trout (about 800 g). Erythrocytes were centrifuged (1000 x g, 5 min, 4°C), washed three times with phosphate buffered saline (PBS, pH 7.2). For the assay, washed and packed erythrocytes were suspended to 5% (v/v) in PBS. An equal amount (30  $\mu$ l) of erythrocyte and bacterial suspensions (including the phage resistant strain V1-20 as a negative control [13]) were mixed in triplicates into a U-well microtiter plate (Greiner bio-one) and incubated for 24 h at 10°C and 400 rpm rotation. Following incubation, 150  $\mu$ l 0.5% NaCl was added to the wells and the plate was centrifuged (1000 x g, 5 min, 4°C). The supernatants were transferred to a F-well microtiter plate (Greiner Bio-one) and the absorbance was measured at 540 nm (A). A negative control (background,  $A^{\text{background}}$ ) with only 0.5% NaCl and erythrocytes and a positive control (total hemolysis,  $A^{100\%}$ ) with distilled water and erythrocytes were included in triplicates on each plate. The hemolytic activity

was calculated according to: Hemolytic activity =  $(A - A^{\text{background}}) / (A^{100\%} - A^{\text{background}})$ . Experiments were repeated in four independent assays.

### Statistical analysis

A Student's t-test (GraphPad Prism 4 software) was used to analyze the statistical significance of the observed differences in activities among the strains. P-values of <0.01 were defined as statistically significant.

### Accession numbers

Accession numbers for *F. psychrophilum* strains are listed in the [S1 Table](#). Bacteriophages 6H, FpV4, Fpv9 and Fpv21 have been assigned GenBank accession numbers: KC959568, KT876724, KT876725 and KT876726 respectively.

## Results

### General features and architecture of the *F. psychrophilum* genomes

The DNA sequences were obtained for *F. psychrophilum* isolates MH1, PG2 and VQ50, which are responsible for salmon and rainbow trout infections in Chile ([S1 Table](#)) [17]. The *F. psychrophilum* strains 3, 4 and 5 were isolated from water samples collected from fish farms in Chile. These six DNA sequences were analyzed together with the genomes of isolates 950106-1/1, JIP02/86, CSF 259–93, FPG3 and FPG101 already available in the databases ([S1 Table](#)). The 11 *F. psychrophilum* isolates varied in size from 2.71 and 2.86 (Mb) with a GC content of 32.4–32.7%. One plasmid was present in the isolate JIP02/86 ([Table 1](#)). Analysis of annotated contigs revealed a relatively similar ORF number among all the isolates (2406–2569 ORFs).

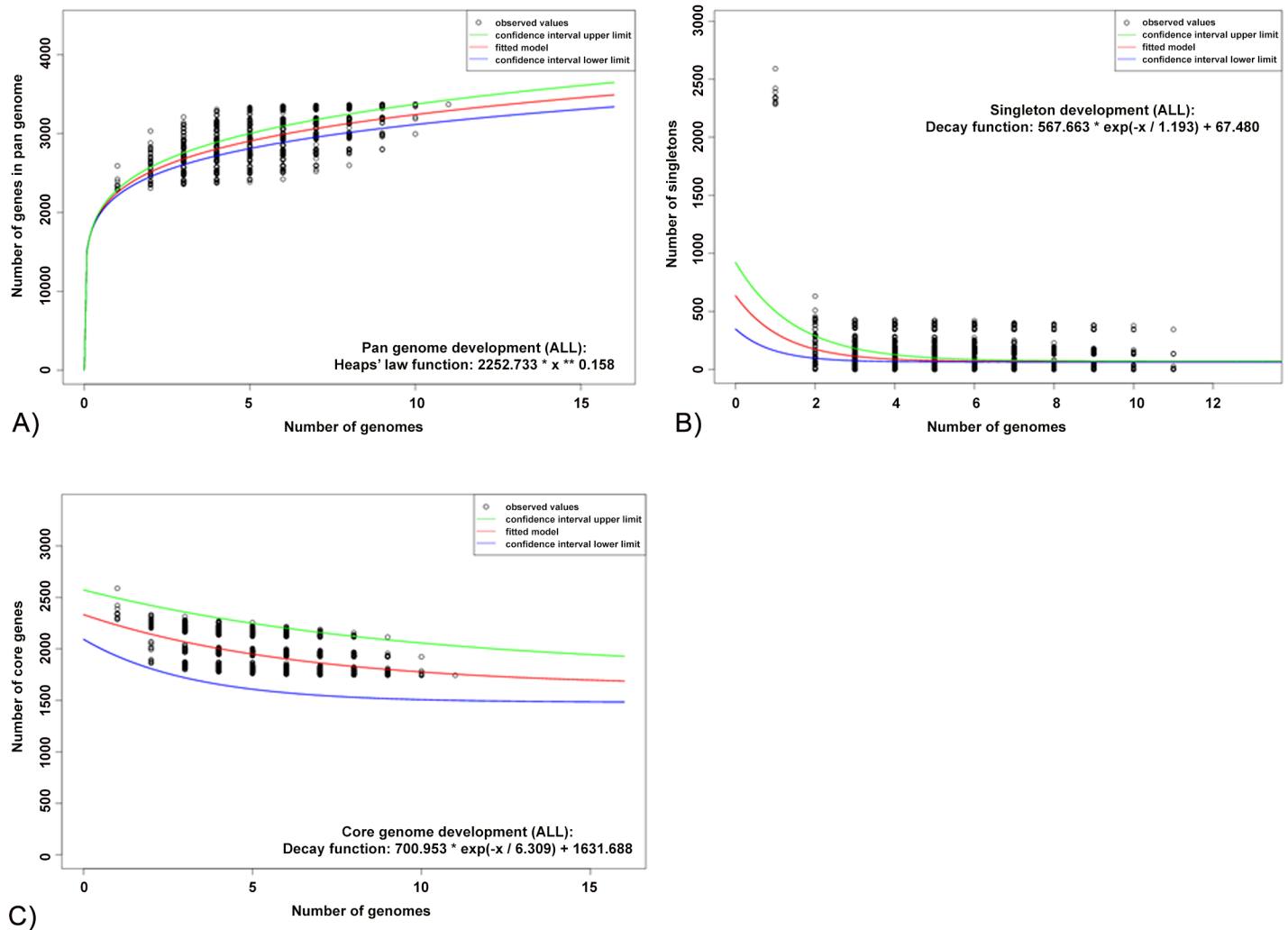
### Pan-genome analysis

Examination of the *F. psychrophilum* pan-genome indicated that the gene repertoire increased with sequential addition of each new genome, and continued to increase for all additions ([Fig 1A](#)). The number of new genes supplied by a novel genome was 201 for the second genome added and 76 for the last, and the model predicted that the gene repertoire of *F. psychrophilum* could hold at least 3373 genes. In order to determine if the *F. psychrophilum* pan-genome was open, the accessory genes were calculated from incorporation of a new genome sequence every time. We therefore applied the exponential decay model to identify unique genes using the

**Table 1. Summary of genome sequence projects of *F. psychrophilum* isolates.**

Isolate	Geographical Origin	Size (Mb)	Genes	CDS	Pseudogenes	%GC	rRNA operons	tRNA	Plasmid	Reference
FPG3 (ATCC 49418)	USA	2.71	2406	2310	28	32.7	6	49	None	[52]
4	Chile	2.71	2494	2344	49	32.4	6	36	None	This study
CSF 259–93	USA	2.90	2569	2467	34	32.5	6	49	None	[25]
FPG101	Canada	2.83	2516	2424	24	32.6	6	49	None	[6]
3	Chile	2.80	2482	2390	24	32.6	6	49	None	This study
5	Chile	2.84	2539	2447	24	32.5	6	49	None	This study
PG2	Chile	2.85	2541	2451	22	32.5	6	49	None	[17]
950106-1/1	Denmark	2.74	2459	2396	21	32.4	1	49	None	[13]
JIP02/86 (ATCC 49511)	France	2.86	2556	2446	43	32.5	6	49	pCP1	[12]
MH1	Chile	2.84	2541	2447	26	32.6	6	49	None	[17]
VQ50	Chile	2.80	2483	2390	25	32.5	6	49	None	[17]

doi:10.1371/journal.pone.0152515.t001



**Fig 1. *F. psychrophilum* pan, core and accessory genome evolution according to the number of sequenced genomes.** A) Total number of genes (pan-genome) for a given number of genomes sequentially added. The pan genome analysis is shown for increasing values for every *F. psychrophilum* genome sequenced based on a power law fit model. B) Number of shared genes (core genome) as a function of the number of genomes sequentially added. The exponential decay model based on the median value for the conserved core genes displays that the core genome had a minimum of 1743 genes in *F. psychrophilum*. C) Number of unique genes (accessory genome) for a given number of genomes sequentially added. Decreasing number of accessory genes per genome was observed with increased of genomes examined. The graphic shows the exponential decay model based on the median value for the accessory genes when increasing numbers of genes were analyzed. An average of 67 new genes could be detected in the *F. psychrophilum* pan genome for each new genome added. For all the plots the upper (green) and lower (blue) lines indicate the first (25th percentile of the data) and third (75th percentile) quartiles respectively. The central (red) line refers the sample median (50th percentile) of random input order of the *F. psychrophilum* genomes.

doi:10.1371/journal.pone.0152515.g001

median value. The model estimated that  $67 \pm 3$  new genes could be revealed for every new *F. psychrophilum* sequence added to the analysis (Fig 1B). This relatively high rate of increase in the pan-genome size with addition of new genomes suggested that *F. psychrophilum* possesses an open pan-genome.

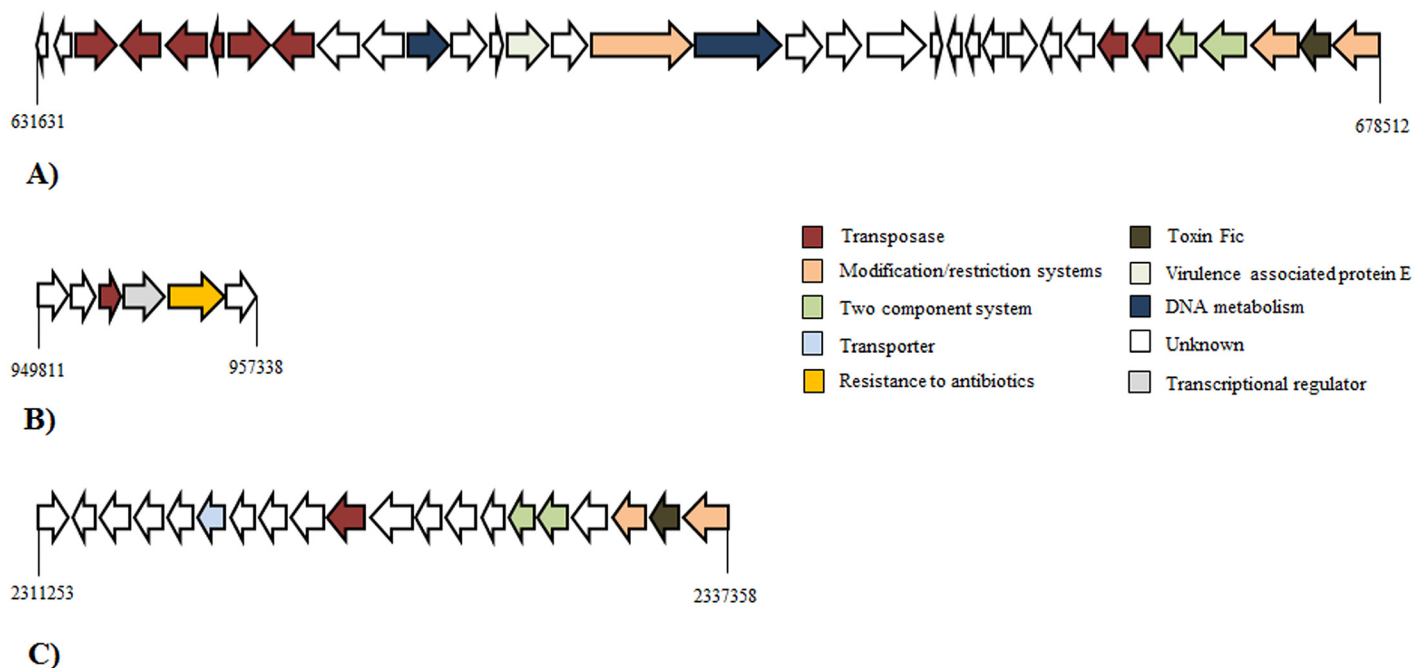
In contrast to the pan-genome, an examination of the *F. psychrophilum* core genome showed that the number of shared genes decreased with the addition of each new genome, as expected (Fig 1C). The *F. psychrophilum* average gene content is  $2375 \pm 54$  genes and the core genome was estimated to contain  $1743 \pm 51$  genes, corresponding to 73% of the genome which would remain relatively constant even if additional isolates genomes were included.

### Virulence factors

For examination of the possible molecular basis for virulence phenotypes, we compared the *F. psychrophilum* genomes with respect to virulence gene content to provide additional insights into the biology and evolution of this fish pathogen. Using databases from MvirDB and virulencefinder did not detect any virulence factors. Previous *F. psychrophilum* genome studies have, however, detected putative gene candidates participating in virulence, such as genes encoding proteases and adhesion [12], transport [52], motility and a specific secretion system [53]. These 44 putative virulence factors among 11 *F. psychrophilum* genomes showed a remarkably similar distribution in the strains independent of geographic locality, year of isolation or source of isolates, except for a putative hemolysin D transporter (*hlyD*), which was found only in the isolates CSF 259–93, FPG101 and 4 (S3 Table).

### Strain-specific genomic islands

Genomic islands (GI) are defined as cluster genes in prokaryotic genomes of probable horizontal origin and commonly encoding mobility related genes, and genes involved in virulence and drug resistance [54]. We examined the genomic distribution of GI in the 11 *F. psychrophilum* isolates in order to find unique acquired regions associated with virulence genes. Only a subset of strains (950106-1/1, CSF 259–93) displayed GIs detected by bioinformatics tools (Fig 2; S4 and S5 Tables). Functions assigned to GI genes ranged from transposases, modification and restriction systems, resistance to antibiotics, virulence factors, toxins, DNA metabolism and unknown functions. Unfortunately, parts of these genes with unknown functions have been poorly characterized, and their biological significance was difficult to infer based on available annotations. The largest GI found was present in the isolate CSF 259–93 (46.8 kb), and contained genes encoding toxin Fic (WP\_034100128.1), virulence associated protein E (VapE),



**Fig 2. Schematic representation of Genomic islands (GI) present in the *F. psychrophilum* isolates.** A) GI found in *F. psychrophilum* CSF 259–93 associated with toxin Fic and virulence factor E family. B) GI found in *F. psychrophilum* CSF 259–93 associated with tetracycline resistance. C) GI found in *F. psychrophilum* 950106-1/1 associated with toxin Fic. The colors were assigned according to the possible role of each ORF.

doi:10.1371/journal.pone.0152515.g002



type II restriction endonucleases and one two- component system associated with multiple transposases (Fig 2A). Moreover, another GI in the same isolate was associated with an antibiotic resistance (Fig 2B). Finally, the Danish isolate 950106-1/1 contained a GI of 28.5 kb, which encoded a toxin Fic (Fig 2C).

Other putative GIs (arbitrarily >3 ORFs; 388 ORFs total) were detected as non-aligned regions using Mauve alignment. Of these genes, 269 belonged to *F. psychrophilum* isolate 4 (69%). For example, a unique region of 9.4 kb with genes encoding transport, multi-drug resistance and heat shock proteins was found in this isolate. Part of these proteins showed amino acid similarity to homologous genes of *Flavobacterium sp*, *Chryseobacterium sp*, *Riemerella colombina* and other bacteria belonging to the phylum *Bacterioidetes* (data not shown). The predicted functions assigned each GIs are showed in the supplementary information (S5 Table).

## Prophages

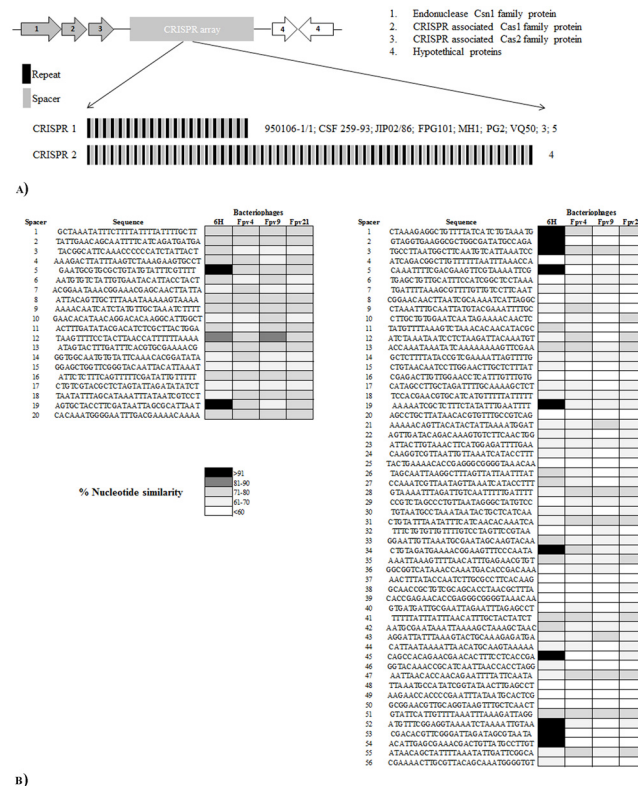
We examined phage-like elements in the genomes in order to determine the prophage carriage in *F. psychrophilum*. Except from the presence of the previously characterized 45-kb prophage 6H [41], which was found in the *F. psychrophilum* isolates 950106-1/1, JIP02/86, MH1, PG2 and 5, the data did not reveal phage-related regions in genomes (data not shown).

## Identification of CRISPR loci in *F. psychrophilum* isolates

In order to characterize the CRISPR/Cas system within *F. psychrophilum* isolates, we examined the 11 DNA sequences for CRISPR arrays. The results showed an identical CRISPR array linked to three putative cas proteins, encoded by genes- *csn1*, *cas1* and *cas2*, that exhibited 20 different spacers of 29–30 bp and 21 direct repeats of 46 bp for *F. psychrophilum* isolates 950106-1/1, CSF 259–93, JIP02/86, FPG101, MH1, PG2, VQ50, 3 and 5 (nominated CRISPR1) (Fig 3A). *F. psychrophilum* strain 4, on the other hand, showed a CRISPR array associated to cas proteins with 57 direct repeats and 56 spacers (nominated CRISPR 2; Contig 4) (Fig 3A). No CRISPR arrays were found in strain FPG3. We examined the nucleotide similarity of *F. psychrophilum* spacers from all CRISPRs with sequences of 4 known *F. psychrophilum* bacteriophages (6H, FpV4, FpV9 and FpV21) and other sequences of extrachromosomal origin in the NCBI database. The results for CRISPR 1 revealed that 19 out 21 spacers did not significantly match (<80% similarity) with genomic sequences from *F. psychrophilum* bacteriophages Fpv4, Fpv9, Fpv21 or with sequences in the NCBI database. The two exceptions were the spacers 5 and 19, which showed 100% and 97% similarity, respectively with sequences of the temperate bacteriophage 6H [41] (Fig 3B). In the same way, spacers from CRISPR 2 showed no significant match with genomes from bacteriophages FpV4-9, -21, or any sequence from database, but 10 spacers displayed 100% similarity with sequences from bacteriophage 6H (Fig 3B). The ORFs targeting the bacteriophage 6H are shown in the additional information (S6 Table).

## Phylogeny

We determined the overall phylogenetic relationship from a concatenated alignment of 1426 single-copy orthologs shared by all 11 *F. psychrophilum* isolates, rooted by *F. columnare* ATCC49512 (outgroup) (Fig 4). Three different clusters can be visualized: Cluster 1 including the American isolate FPG3. Cluster 2 was formed by the Chilean isolate 4. Finally, cluster 3 which contained the North American isolates CSF 259–93 and FPG101, French isolate JIP02/86, Danish isolate 950106-1/1 and the Chilean isolates MH1, VQ50, PG2, 5 and 3, which showed a temporal scale of isolation >28 years (Fig 4). The phylogenetic tree indicated that 5 out of the 6 Chilean *F. psychrophilum* isolates (MH1, VQ50, PG2, 4 and 3) had a common



**Fig 3. *F. psychrophilum* isolates CRISPRs overview.** A) Graphic representation of cas genes, spacer and repeats of CRISPR1 and CRISPR2 in the *F. psychrophilum* isolates. B) Nucleotide comparison of spacers from CRISPR1 (left) and CRISPR2 (right) against genome sequences from bacteriophages 6H, Fpv4, Fpv9 and Fpv21.

doi:10.1371/journal.pone.0152515.g003

ancestor. The American and Chilean *F. psychrophilum* isolates FPG3 and 4, respectively, displayed the most distant lineages in our phylogenetic tree with 98% support (Fig 4).

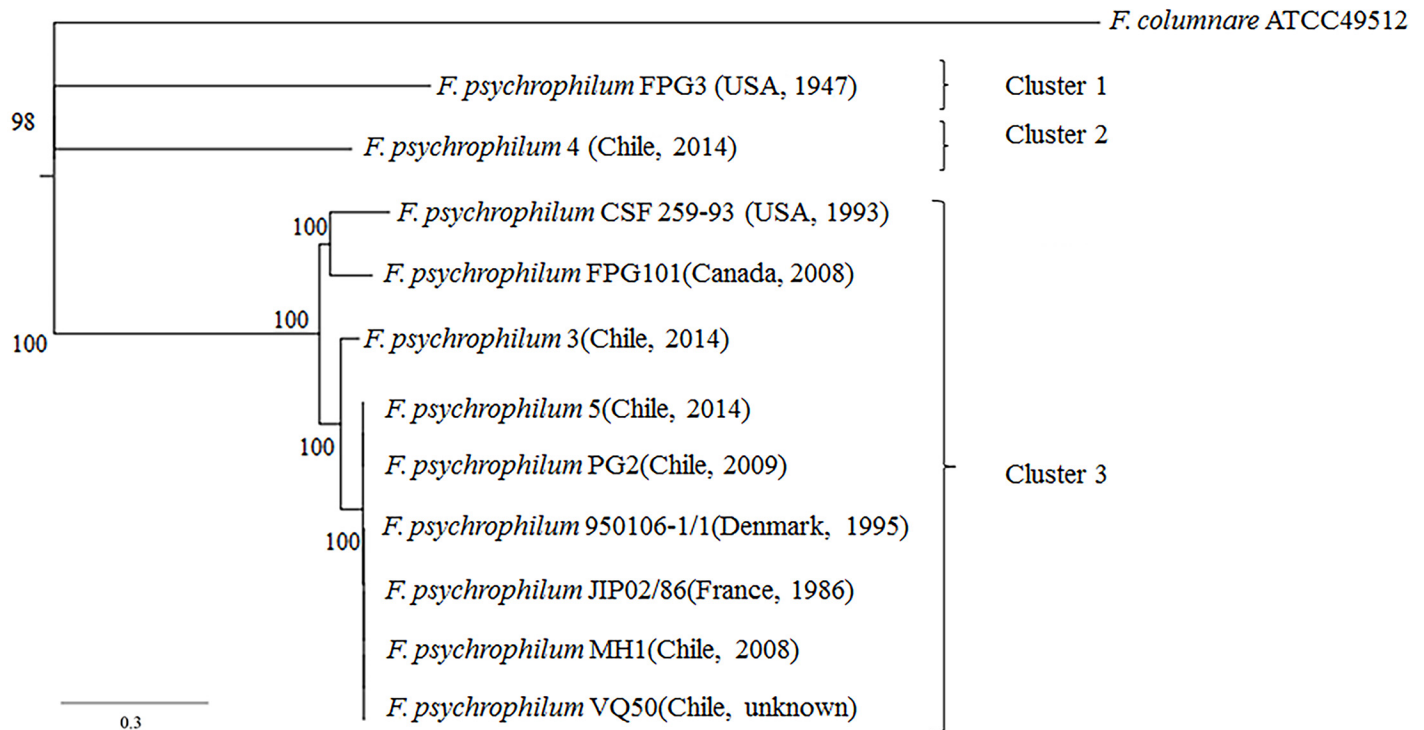
### Phenotypic properties

Ten *F. psychrophilum* isolates were phenotypically analyzed for gliding motility, biofilm formation and secretion of extracellular enzymes (S1 Fig). None of the strains showed motile behavior in any of the conditions tested (data not shown), and no significant differences ( $p > 0.01$ ) in biofilm formation, hemolytic activity, gelatinase activity and total protease activity was observed among the strains (S1 Fig).

### Discussion

#### The *F. psychrophilum* pan-genome

Regression analysis suggested that the *F. psychrophilum* pan-genome can be categorized as an open pan-genome because the gene repertoire could hold at least 3373 genes (Fig 1A) according to the model, and as many new genes ( $67 \pm 3$ ) could be identified for every new genome added to the analysis (Fig 1C). In contrast to that, the size of the core genome decreased with the addition of each new sequenced isolate, representing on average 73% of the pan-genome (Fig 1B). Thus, these results demonstrated that the 11 *F. psychrophilum* isolates only represented a subset of the genetic diversity in this species and suggested that the *F. psychrophilum*



**Fig 4. Phylogenetic tree inferred from concatenated genes.** Maximum likelihood tree obtained from a concatenated nucleotide sequence alignment of the orthologous core genes for the 11 *F. psychrophilum* isolates and *F. columnare* ATCC49512 (outgroup). The horizontal bar at the base of the figure represents 0.3 substitutions per nucleotide site. The percentages of genes that support the branches of the tree are indicated. Geographic location and isolation year of *F. psychrophilum* isolates were added to facilitate comparison.

doi:10.1371/journal.pone.0152515.g004

pan-genome is still evolving by gene acquisition and diversification. Similar results have been described for *Erwinia amylovora* [55], *Streptococcus agalactiae* [30] and *Propionibacterium acnes* [56], whereas analysis of *Bacillus anthracis* suggested a closed pan-genome in which the diversity was covered with only four isolates [57].

The indication of an open pan-genome in *F. psychrophilum* is somewhat surprising given the relatively restricted ecological niche as a fish pathogen. However, little is known about the ecology of *F. psychrophilum* and our study support previous studies demonstrating a high recombination rate among *F. psychrophilum*, likely mediated by mobile elements [21, 23, 58], which can contribute to accessory gene repertory in *F. psychrophilum*.

### Virulence factors and prophage identification in *F. psychrophilum*

Virulence factors of pathogenic bacteria play an important role in colonizing various niches through infection of their host and adaptations to new environmental conditions [59]. All *F. psychrophilum* isolates showed similar content and distribution of adhesion and metalloproteases proteins, which could play important roles in attachment and destruction of host tissues respectively [9, 60]. Moreover, the gliding motility and Por secretion system (PorSS) proteins were detected in all the isolates (S3 Table). These proteins are widespread among members of the phylum *Bacteroidetes* [61] and have been associated with the transport and translocation of virulence factors in the pathogens *F. johnsoniae* and *Porphyromonas gingivalis* [53]. Besides, genome analysis focusing on resistance mechanisms in phage-resistant *F. psychrophilum* clones displayed a link between mutations in the proteins of PorSS system and decreased of virulence properties at *in vitro*, suggesting an important role in virulence for this bacterium [13].

The presence of prophage 6H in 5 of the genomes confirmed a recent study which showed a wide spread and global distribution of this prophage in the genomes of *F. psychrophilum* isolates [41]. Prophage 6H has been associated with putative virulence factors; however, it is not known to what extent prophage genes are participating in bacterial infection. Interestingly, no other prophage-like sequences were detected in *F. psychrophilum* genomes (data not shown). Similar findings have been described for *F. branchophilum*, *F. columnare* and *F. johnsoniae*, where genome sequences are lack of prophage-like elements [62]. Moreover, a large number of lytic phages (>20) have been isolated against *F. psychrophilum* [16], and the absence of these phages as prophages in the sequenced strains suggested that their integration through a lysogenic life cycle is rare. On the other hand, the rapid loss of prophage 6H after exposure to phages in *F. psychrophilum* strain 950106-1/1 [21] also indicates a dynamic and temporary occurrence of prophages due to the integration/excision to and from *F. psychrophilum* genomes. However, the dynamics of prophages is not resolved in this study.

### Genomic islands and strain-specific genes in *F. psychrophilum*

A total of 3 specific GIs were detected in the 11 *F. psychrophilum* isolates using PAIfinder and PIPS tools (Fig 2; S4 Table). *F. psychrophilum* isolates CSF 259–93 and 950106-1/1 showed GIs of 46.8 kb and 25.8 kb respectively, associated to transposases, modification-restriction systems, virulence associated protein E (VapE) and toxin Fic (Fig 2A and 2C). Fic proteins have emerged as a new class effector that interferes with host cell signaling pathways, affecting indirectly cytoskeletal dynamics [63]. Although a similar gene has been identified previously in *F. psychrophilum* strains, they were not associated to mobile elements in these strains [6, 12]. Virulence associated protein E (VapE) was originally identified and associated with virulence in *Dichelobacter nodosus* [64] and *Staphylococcus aureus* [65], but the mechanism by which this protein affects virulence has not been determined. Finally, isolate CSF 259–93 also presented a small GI of 7.5 kb associated to tetracycline resistance gene (Fig 2B), confirming previous observations of resistance to this specific antibiotic [66]. In the current analysis, we defined arbitrarily GIs as genomic regions containing three or more ORFs, which were specific for a subset of the *F. psychrophilum* isolates. A total of 388 strain-specific ORFs were detected among the isolates 950106-1/1; CSF 259–93, FPG3 and 4 (S5 Table). Although genes associated with transposases, DNA metabolism, transport and hydrolases were found among these, hypothetical genes or unknown genes represented the vast majority of these unique regions, and the similarity of these new genes between bacteria belonging to the phylum *Bacteroidetes* suggests a possible acquisition though horizontal transfer.

### CRISPR identification in *F. psychrophilum*

Two different CRISPR systems were found among *F. psychrophilum* isolates used in this study (Fig 3). Previous studies have determined the presence of CRISPR 1 in the *F. psychrophilum* isolates 950106-1/1 and JIP02/86 [12], but the current work demonstrated the presence of this specific CRISPR system also in the isolates CSF 259–93, FPG101, MH1, PG2 VQ50, 3 and 5 (Fig 3A). Interestingly, *F. psychrophilum* isolate 4 contained a different CRISPR array (CRISPR 2) with 56 spacers and 57 direct repeats (Fig 3A). The nucleotide similarity of spacers against a bacteriophage collection showed that both CRISPR systems contained spacers which matched with bacteriophage 6H, but with variable numbers of 6H specific spacers (Fig 4B; S6 Table). These findings support the hypothesis that bacteriophage 6H belongs to an abundant group of cosmopolitan temperate phages which have lysogenized a large fraction of the global *F. psychrophilum* community [41]. The stability of CRISPR1 across the large temporal and geographical distances between the isolation of the CRISPR1-containing strains suggested that this

system is not an active and dynamic phage defense mechanism in *F. psychrophilum*. Alternative roles for CRISPR systems have been described in pathogenic bacteria [67]. For example, when *P. aeruginosa* is lysogenized by a specific bacteriophage, the CRISPR/Cas system interacts with a particular gene in the chromosomally integrated prophage to inhibit the creation of biofilms [68]. We have previously suggested that prophage 6H could play a role in the virulence properties of *F. psychrophilum* [41], but we do not know whether CRISPR in *F. psychrophilum* contributes to the regulation of prophage 6H genes.

## Phylogenetic analyses

The phylogenetic analyses based on a concatenated alignment of 1426 single-copy orthologs shared for all the *F. psychrophilum* isolates revealed close relationships among the diverse isolates of *F. psychrophilum*, across large temporal and geographic scales of isolation (Fig 4). For example, Chilean *F. psychrophilum* isolates PG2, MH1, VQ50, 3 and 5 showed less diversity and grouped together with Danish, French and North American isolates 950106-1/1, JIP02/86, CSF 259-93 and FPG101 respectively (cluster 3) (Fig 4). In the same way, Chilean isolate 4 and American isolate FPG3 clustered in a distant lineages (cluster 1 and 2). These observations are consistent with studies using multi locus sequence typing (MLST), where the Chilean *F. psychrophilum* isolates were closely related to the genotypes most prevalent in European and North American fish farms [24], probably due to massive trade of fish eggs across the fish-producer countries [21, 23, 24]. Further, the high similarities in orthologous genes among strains isolated in France, Denmark and Chile over a 28 year period (Cluster 3, Fig 4) suggest that these core genes are highly conserved and that the strain specific genomic differences in *F. psychrophilum* are mainly due to gain and loss of mobile genetic elements. More extensive genomic samplings are required to allow a more detailed phylogenetic analysis among *F. psychrophilum*, and to establish the relationships of the core genome across the genus level.

## Phenotypic traits in *F. psychrophilum*

Phenotypic characterization showed similar properties of 10 *F. psychrophilum* isolates with respect to biofilm formation, hemolytic activity and secretion of extracellular enzymes (S1 Fig). These results are in agreement with our observations of a homogenous distribution of adhesion and metalloproteases proteins in all *F. psychrophilum* isolates (S3 Table). Moreover, components of Por secretion system (PorSS) were found in all the isolates (S3 Table), suggesting that *F. psychrophilum* is proficient in protein secretion through the PorSS, as is proposed for *F. branchiophilum* [62].

Despite the identification of genes involved in motility (*gld*) (S3 Table) no motility was observed in any of the isolates in the conditions tested (data not shown). However, we cannot rule out that *F. psychrophilum* may be motile, but experimental conditions used here failed to simulate natural conditions where gliding motility proteins could be expressed.

## Conclusion and future directions

We report here the DNA sequences of six Chilean *F. psychrophilum* isolates (MH1, PG2, VQ50, 3, 4 and 5) and compared them with 5 previous genome sequences from North America, Denmark and France. Our genome analysis clearly showed that Pan-genome is open in *F. psychrophilum* (Fig 1) and many more sequences are required to cover the diversity of this bacterial species. The identification and characterization of core and accessory genes is essential for understanding the basic metabolism in *F. psychrophilum*, for providing a better insight on virulence properties, and choice of antibiotic treatment, and for further drug and vaccine development, as have been proposed for universal group B of *Streptococcus* [69]. Future studies

about the estimation of the Pan genome in the genus *Flavobacterium* could provide insight into species differentiation and enhance our understanding of evolution within the *F. psychrophilum* cluster (Fig 4).

The distribution of putative indicators of virulence, previously identified in the *F. psychrophilum* isolates JIP02/86 and FPG3, was similar in all the geographic distant isolates (S3 Table). This large scale dispersion of virulence factors may explain the similar ability to form biofilm and to secrete extracellular enzymes among the strains (S1 Fig). Consequently, we can speculate that *F. psychrophilum* isolates have a similar mode of action on adhesion, colonization and destruction of fish tissues across large spatial and temporal scales of occurrence. However, the presence of an extra Fic toxin in genomic islands identified in the strains CSF 259–93 and 950106-1/1 (Fig 2) could indicate that those isolates differed on some steps of infection, probably causing a major effect of destabilization of the cytoskeleton on host cells. In the same way, several strain-specific genetic elements were also found within the genomes of isolates FPG3 and 4 and could be participating in the infection of the host (S5 Table).

Altogether, our results allowed us to gain a better understanding of *F. psychrophilum* pathogenicity and diversity. Defining the pan genome, distribution of virulence factors, mobile elements and phylogenetic relationship and how this genetic information contributes to the phenotype in *F. psychrophilum*, may help developing diagnostic tools and strategies for the control of this pathogen.

## Supporting Information

**S1 Fig. Phenotypic characterization of *F. psychrophilum* isolates.** A) Biofilm formation. B) Hemolytic activity. C) Gelatinase activity on gelatin plates. D) Total protease activity on skim milk plates.

(DOCX)

**S1 Table. Characteristics of *F. psychrophilum* isolates analyzed in this study.**

(DOCX)

**S2 Table. Genome characteristics of Chilean *F. psychrophilum* isolates.**

(DOCX)

**S3 Table. Distribution of putative virulence factors across *F. psychrophilum* isolates.**

(DOCX)

**S4 Table. Unique genome regions identified in *F. psychrophilum* isolates using PAI finder.**

(DOCX)

**S5 Table. Unique genome regions identified in *F. psychrophilum* isolates using MAUVE program.**

(DOCX)

**S6 Table. Target genes on bacteriophage 6H by CRISPR spacers.**

(DOCX)

## Acknowledgments

This work was supported by a grant from The Danish Directorate for Food, Fisheries and Agri Business, The Danish Strategic Research Council (ProAqua, project # 09–072829) and by the EU-IRSES-funded project AQUAPHAGE.

## Author Contributions

Conceived and designed the experiments: DC MM RHC ID LM RE. Performed the experiments: DC RHC MM. Analyzed the data: DC MM. Contributed reagents/materials/analysis tools: MM. Wrote the paper: DC MM RHC ID LM RE.

## References

1. Nematollahi A, Decostere A, Pasmans F, Haesebrouck F. *Flavobacterium psychrophilum* infections in salmonid fish. *J Fish Dis*. 2003; 26:563–574. PMID: [14653314](#)
2. Lorenzen E, Dalsgaard I, From J, Hansen EM, Herrlyck V, Korsholm H, et al. Preliminary investigations of fry mortality syndrome in rainbow trout. *Bull Eur Assoc Fish Pathol*. 1991; 11:77–79.
3. Bruno DW. *Cytophaga psychrophila* (*Flexibacter psychrophilus*) (Borg), histopathology associated with mortalities among farmed rainbow trout *Oncorhynchus mykiss* (Walbaum) in the UK. *Bull Eur Assoc Fish Pathol*. 1992; 12: 215–216.
4. Wiklund T, Kaas K, Lönnström L, Dalsgaard I. Isolation of *Cytophaga psychrophila* (*Flexibacter psychrophilus*) from wild and farmed rainbow trout (*Oncorhynchus mykiss*) in Finland. *Bull. Eur. Assoc. Fish Pathol*. 1994; 14: 44–46.
5. Toranzo AE, Barja JL. Fry mortality syndrome (FMS) in Spain: isolation of the causative bacterium *Flexibacter psychrophilus*. *Bull. Eur. Assoc. Fish Pathol*. 1993; 13: 30–32.
6. Hesami S, Metcalf DS, Lumsden JS, Macinnes JI. Identification of cold-temperature-regulated genes in *Flavobacterium psychrophilum*. *Appl Environ Microbiol*. 2011; 77:1593–600. doi: [10.1128/AEM.01717-10](#) PMID: [21216906](#)
7. Bustos PA, Calbuyahue J, Montana J, Opazo B, Entrala P, Solervicens R. First isolation of *Flexibacter psychrophilus* as causative agent of rainbow trout fry syndrome (RTFS), producing rainbow trout mortality in Chile. *Bull. Eur. Assoc. Fish Pathol*. 1995; 15: 162–164.
8. Wakabayashi H, Horiuchi M, Bunya T, Hoshiai G. Outbreaks of cold-water disease in coho salmon in Japan. *Fish Pathol*. 1991; 26: 211–212.
9. Bertolini JM, Wakabayashi H, Watral VG, Whipple MJ, Rohovec JS. Electrophoretic detection of proteases from selected strains of *Flexibacter psychrophilus* and assessment of their variability. *J Aquat Anim Health*. 1994; 6: 224–233.
10. Ostland VE, Byrne PJ, Hoover G, Ferguson HW. Necrotic myositis of rainbow trout, *Oncorhynchus mykiss* (Walbaum): proteolytic characteristics of a crude extracellular preparation from *Flavobacterium psychrophilum*. *J Fish Dis*. 2000; 23: 329–336.
11. Kondo M, Kawai K, Kurohara K, Oshima S. Adherence of *Flavobacterium psychrophilum* on the body surface of the ayu *Plecoglossus altivelis*. *Microbes Infect*. 2002; 4: 279–283. PMID: [11909737](#)
12. Duchaud E, Boussaha M, Loux V, Bernardet JF, Michel C, Kerouault B, et al. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*. *Nat Biotechnol*. 2007; 25:763–769. PMID: [17592475](#)
13. Castillo D, Christiansen RH, 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; 81:1157–67.
14. Bruun MS, Schmidt AS, Madsen L, Dalsgaard I. Antimicrobial resistance patterns in Danish isolates of *Flavobacterium psychrophilum*. *Aquaculture*. 2000; 187:201–212.
15. Fredriksen BN, Olsen RH, Furevik A, Souhoka RA, Gauthier D, Brudeseth B. Efficacy of a divalent and a multivalent water-in-oil formulated vaccine against a highly virulent strain of *Flavobacterium psychrophilum* after intramuscular challenge of rainbow trout (*Oncorhynchus mykiss*). *Vaccine*. 2013; 31:1994–1998. doi: [10.1016/j.vaccine.2013.01.016](#) PMID: [23352330](#)
16. Stenholm AR, Dalsgaard I, Middelboe M. Isolation and characterization of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*. *Appl Environ Microbiol*. 2008; 74:4070–8. doi: [10.1128/AEM.00428-08](#) PMID: [18469131](#)
17. Castillo D, Higuera G, Villa M, Middelboe M, Dalsgaard I, Madsen L, et al. 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. doi: [10.1111/j.1365-2761.2011.01336.x](#) PMID: [22324343](#)
18. Madsen L, Dalsgaard I. Comparative studies of Danish *Flavobacterium psychrophilum* isolates: ribotypes, plasmid profiles, serotypes and virulence. *J Fish Dis*. 2000; 23:211–218.
19. Fujiwara-Nagata E, Chantry-Darmon C, Bernardet JF, Eguchi M, Duchaud E, Nicolas P. Population structure of the fish pathogen *Flavobacterium psychrophilum* at whole-country and model river levels in Japan. *Vet Res*. 2013; 44:34. doi: [10.1186/1297-9716-44-34](#) PMID: [23682575](#)

20. Nilsen H, Sundell K, Duchaud E, Nicolas P, Dalsgaard I, et al. Multilocus sequence typing identifies epidemic clones of *Flavobacterium psychrophilum* in Nordic countries. *Appl Environ Microbiol.* 2014; 80:2728–2736. doi: [10.1128/AEM.04233-13](https://doi.org/10.1128/AEM.04233-13) PMID: [24561585](https://pubmed.ncbi.nlm.nih.gov/24561585/)
21. Castillo D, Christiansen RH, 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. doi: [10.1007/s00248-014-0375-8](https://doi.org/10.1007/s00248-014-0375-8) PMID: [24557506](https://pubmed.ncbi.nlm.nih.gov/24557506/)
22. Arai H, Morita Y, Izumi S, Katagiri T, Kimura H. Molecular typing by pulsed-field gel electrophoresis of *Flavobacterium psychrophilum* isolates derived from Japanese fish. *J Fish Dis.* 2007; 30:345–355. PMID: [17498178](https://pubmed.ncbi.nlm.nih.gov/17498178/)
23. Nicolas P, Mondot S, Achaz G, Bouchenot C, Bernardet JF, Duchaud E. Population structure of the fish-pathogenic bacterium *Flavobacterium psychrophilum*. *Appl Environ. Microbiol.* 2008; 74:3702–3709. doi: [10.1128/AEM.00244-08](https://doi.org/10.1128/AEM.00244-08) PMID: [18424537](https://pubmed.ncbi.nlm.nih.gov/18424537/)
24. Avendaño-Herrera R, Houel A, Irgang R, Bernardet JF, Godoy M, Nicolas P, et al Introduction, expansion and coexistence of epidemic *Flavobacterium psychrophilum* lineages in Chilean fish farms. *Vet Microbiol.* 2014; 170:298–306. doi: [10.1016/j.vetmic.2014.02.009](https://doi.org/10.1016/j.vetmic.2014.02.009) PMID: [24636160](https://pubmed.ncbi.nlm.nih.gov/24636160/)
25. Wiens GD, LaPatra SE, Welch TJ, Rexroad C 3rd, Call DR, Cain KD, et al. Complete Genome sequence of *Flavobacterium psychrophilum* strain CSF259-93, used to select rainbow trout for increased genetic resistance against bacterial cold water disease. *Genome Announc.* 2014; 2 pii: e00889–14. doi: [10.1128/genomeA.00889-14](https://doi.org/10.1128/genomeA.00889-14) PMID: [25237017](https://pubmed.ncbi.nlm.nih.gov/25237017/)
26. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics.* 2015; 15:141–161. doi: [10.1007/s10142-015-0433-4](https://doi.org/10.1007/s10142-015-0433-4) PMID: [25722247](https://pubmed.ncbi.nlm.nih.gov/25722247/)
27. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol.* 2008; 11:472–477. PMID: [19086349](https://pubmed.ncbi.nlm.nih.gov/19086349/)
28. Syvanen M. Evolutionary implications of horizontal gene transfer. *Annu Rev Genet.* 2012; 46:341–58. doi: [10.1146/annurev-genet-110711-155529](https://doi.org/10.1146/annurev-genet-110711-155529) PMID: [22934638](https://pubmed.ncbi.nlm.nih.gov/22934638/)
29. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol.* 2008; 190:6881–6893. doi: [10.1128/JB.00619-08](https://doi.org/10.1128/JB.00619-08) PMID: [18676672](https://pubmed.ncbi.nlm.nih.gov/18676672/)
30. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A.* 2005; 102:13950–13955. PMID: [16172379](https://pubmed.ncbi.nlm.nih.gov/16172379/)
31. Bernardet JF, Kerouault B. Phenotypic and genomic studies of "*Cytophaga psychrophila*" isolated from diseased rainbow trout (*Oncorhynchus mykiss*) in France. *Appl Environ Microbiol.* 1989; 55:1796–1800. PMID: [2764577](https://pubmed.ncbi.nlm.nih.gov/2764577/)
32. Tatusova T, DiCuccio M, Badretdin A, Chetvermin V, Ciufu S, Li W. Prokaryotic genome annotation pipeline. In: National Center for Biotechnology Information (NCBI). 2013. <http://www.ncbi.nlm.nih.gov/books/NBK174280/>.
33. Yoon SH, Park YK, Kim JF. PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Res.* 2015; 43: D624–30. doi: [10.1093/nar/gku985](https://doi.org/10.1093/nar/gku985) PMID: [25336619](https://pubmed.ncbi.nlm.nih.gov/25336619/)
34. Soares SC, Abreu VA, Ramos RT, Cerdeira L, Silva A, Baumbach J, et al. PIPS: pathogenicity island prediction software. *PLoS One.* 2012; 7:e30848. doi: [10.1371/journal.pone.0030848](https://doi.org/10.1371/journal.pone.0030848) PMID: [22355329](https://pubmed.ncbi.nlm.nih.gov/22355329/)
35. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004; 14:1394–403. PMID: [15231754](https://pubmed.ncbi.nlm.nih.gov/15231754/)
36. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res.* 2007; 35:D391–4. PMID: [17090593](https://pubmed.ncbi.nlm.nih.gov/17090593/)
37. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol.* 2014; 52:1501–10. doi: [10.1128/JCM.03617-13](https://doi.org/10.1128/JCM.03617-13) PMID: [24574290](https://pubmed.ncbi.nlm.nih.gov/24574290/)
38. Fouts DE. Phage\_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res.* 2006; 34:5839–5851. PMID: [17062630](https://pubmed.ncbi.nlm.nih.gov/17062630/)
39. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PFAST: a fast phage search tool. *Nucleic Acids Res.* 2011; 39(Web Server issue):W347–52. doi: [10.1093/nar/gkr485](https://doi.org/10.1093/nar/gkr485) PMID: [21672955](https://pubmed.ncbi.nlm.nih.gov/21672955/)
40. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 2007; 35:W52–7. PMID: [17537822](https://pubmed.ncbi.nlm.nih.gov/17537822/)
41. 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. doi: [10.1111/1574-6968.12342](https://doi.org/10.1111/1574-6968.12342) PMID: [24267868](https://pubmed.ncbi.nlm.nih.gov/24267868/)



42. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, et al. 2010. In: Geneious v7.1.4. 2010. <http://www.geneious.com>.
43. Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter FJ, Zakrzewski M, et al. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics*. 2009; 10:154. doi: [10.1186/1471-2105-10-154](https://doi.org/10.1186/1471-2105-10-154) PMID: [19457249](https://pubmed.ncbi.nlm.nih.gov/19457249/)
44. Chen F, Mackey AJ, Stoekert CJ Jr, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research*. 2006; 34(Database issue):D363–8. PMID: [16381887](https://pubmed.ncbi.nlm.nih.gov/16381887/)
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23:2947–2948. PMID: [17846036](https://pubmed.ncbi.nlm.nih.gov/17846036/)
46. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006; 34(Web Server issue):W609–12. PMID: [16845082](https://pubmed.ncbi.nlm.nih.gov/16845082/)
47. Kück P, Meusemann K. FASconCAT: Convenient handling of data matrices. *Mol Phylogenet Evol*. 2010; 56:1115–8. doi: [10.1016/j.ympev.2010.04.024](https://doi.org/10.1016/j.ympev.2010.04.024) PMID: [20416383](https://pubmed.ncbi.nlm.nih.gov/20416383/)
48. Guindon S, Dufayard JF, 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; 59:307–21. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010) PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)
49. Bernardet JF, Nakagawa Y, Holmes B, Subcommittee on the taxonomy of *Flavobacterium* and *Cytophaga*-like bacteria of the International Committee on Systematics of Prokaryotes. Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int J Syst Evol Microbiol*. 2002; 52:1049–1070. PMID: [12054224](https://pubmed.ncbi.nlm.nih.gov/12054224/)
50. Laanto E, Bamford JK, Laakso J, Sundberg LR. Phage-driven loss of virulence in a fish pathogenic bacterium. *PLoS One*. 2012; 7:e53157. doi: [10.1371/journal.pone.0053157](https://doi.org/10.1371/journal.pone.0053157) PMID: [23308090](https://pubmed.ncbi.nlm.nih.gov/23308090/)
51. 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. doi: [10.1016/j.micpath.2010.08.002](https://doi.org/10.1016/j.micpath.2010.08.002) PMID: [20817089](https://pubmed.ncbi.nlm.nih.gov/20817089/)
52. Wu AK, Kropinski AM, Lumsden JS, Dixon B, MacInnes JI. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum* ATCC 49418(T.). *Stand Genomic Sci*. 2015; 10: 3. doi: [10.1186/1944-3277-10-3](https://doi.org/10.1186/1944-3277-10-3) PMID: [25685258](https://pubmed.ncbi.nlm.nih.gov/25685258/)
53. Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, et al. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci USA*. 2010; 107:276–281. doi: [10.1073/pnas.0912010107](https://doi.org/10.1073/pnas.0912010107) PMID: [19966289](https://pubmed.ncbi.nlm.nih.gov/19966289/)
54. Whittle G, Shoemaker NB, Salyers AA. The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell Mol Life Sci*. 2002; 59:2044–2054. PMID: [12568330](https://pubmed.ncbi.nlm.nih.gov/12568330/)
55. Mann RA, Smits TH, Bühlmann A, Blom J, Goesmann A, et al. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. *PLoS One*. 2013; 8:e55644. doi: [10.1371/journal.pone.0055644](https://doi.org/10.1371/journal.pone.0055644) PMID: [23409014](https://pubmed.ncbi.nlm.nih.gov/23409014/)
56. Tomida S, Nguyen L, Chiu BH, Liu J, Sodergren E, Weinstock GM, Li H. Pan-genome and comparative genome analyses of *Propionibacterium acnes* reveal its genomic diversity in the healthy and diseased human skin microbiome. *MBio*. 2008; 4:e00003–13.
57. Keim P, Gruendike JM, Klevytska AM, Schupp JM, Challacombe J, Okinaka R. The genome and variation of *Bacillus anthracis*. *Mol Aspects Med*. 2009; 30:397–405. doi: [10.1016/j.mam.2009.08.005](https://doi.org/10.1016/j.mam.2009.08.005) PMID: [19729033](https://pubmed.ncbi.nlm.nih.gov/19729033/)
58. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J*. 2009; 3:199–208. doi: [10.1038/ismej.2008.93](https://doi.org/10.1038/ismej.2008.93) PMID: [18830278](https://pubmed.ncbi.nlm.nih.gov/18830278/)
59. Wu HJ, Wang AH, Jennings MP. Discovery of virulence factors of pathogenic bacteria. *Curr Opin Chem Biol*. 2008; 12:93–101. doi: [10.1016/j.cbpa.2008.01.023](https://doi.org/10.1016/j.cbpa.2008.01.023) PMID: [18284925](https://pubmed.ncbi.nlm.nih.gov/18284925/)
60. Ikegami A, Honma K, Sharma A, Kuramitsu HK. Multiple functions of the leucine-rich repeat protein LrrA of *Treponema denticola*. *Infect. Immun*. 2004; 72:4619–4627. PMID: [15271922](https://pubmed.ncbi.nlm.nih.gov/15271922/)
61. McBride MJ, Zhu Y. Gliding motility and Por secretion system genes are widespread among members of the phylum *bacteroidetes*. *J Bacteriol*. 2013; 195:270–8. doi: [10.1128/JB.01962-12](https://doi.org/10.1128/JB.01962-12) PMID: [23123910](https://pubmed.ncbi.nlm.nih.gov/23123910/)
62. Touchon M, Barbier P, Bernardet JF, Loux V, Vacherie B, Barbe V, et al. Complete genome sequence of the fish pathogen *Flavobacterium branchiophilum*. *Appl Environ Microbiol*. 2011; 77:7656–62. doi: [10.1128/AEM.05625-11](https://doi.org/10.1128/AEM.05625-11) PMID: [21926215](https://pubmed.ncbi.nlm.nih.gov/21926215/)
63. Engel P, Goepfert A, Stanger FV, Harms A, Schmidt A, Schirmer T, et al. Adenylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature*. 2012; 482:107–10. doi: [10.1038/nature10729](https://doi.org/10.1038/nature10729) PMID: [22266942](https://pubmed.ncbi.nlm.nih.gov/22266942/)

64. Bloomfield GA, Whittle G, McDonagh MB, Katz ME, Cheetham BF. Analysis of sequences flanking the vap regions of *Dichelobacter nodosus*: evidence for multiple integration events, a killer system, and a new genetic element. *Microbiology*. 1997; 143:553–562. PMID: [9043132](#)
65. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol Microbiol*. 1998; 9:527–543.
66. Soule M, LaFrentz S, Cain K, LaPatra S, Call DR. Polymorphisms in 16S rRNA genes of *Flavobacterium psychrophilum* correlate with elastin hydrolysis and tetracycline resistance. *Dis Aquat Organ*. 2005; 65:209–216. PMID: [16119889](#)
67. Sampson TR, Weiss DS. CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front Cell Infect Microbiol*. 2014; 4:37. doi: [10.3389/fcimb.2014.00037](#) PMID: [24772391](#)
68. Zegans ME, Wagner JC, Cady KC, Murphy DM, Hammond JH, O'Toole GA. Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J Bacteriol*. 2009; 191:210–209. doi: [10.1128/JB.00797-08](#) PMID: [18952788](#)
69. Maione D, Margarit I, Rinaudo CD, Massignani V, Mora M, Scarselli M, et al. Identification of a universal Group B *Streptococcus* vaccine by multiple genome screen. *Science*. 2005; 309:148–50. PMID: [15994562](#)