



# Whole-genome comparative analysis of the pathogen *Piscirickettsia salmonis*



Cristian Bravo, Victor Martinez\*

FAVET-INBIOGEN, Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Avda. Santa Rosa, 11735, Santiago, Chile

## ARTICLE INFO

### Article history:

Received 17 May 2016

Received in revised form 7 October 2016

Accepted 9 October 2016

### Keywords:

*Piscirickettsia salmonis*

SRS

Whole genome sequencing

Comparative genomics

Invasion

Intracellular survival

## ABSTRACT

The intracellular pathogen *Piscirickettsia salmonis* is the etiological agent of piscirickettsiosis, the most important bacterial disease that affects the Chilean salmon industry. Despite its importance, little is known regarding the biology of the pathogen. In this study, recently published sequencing data was used in order to characterize the genome of *P. salmonis*, defining groups of genes associated with bacterial processes such as, invasion and intracellular survival. Moreover, one Chilean *P. salmonis* isolate, which is known to be virulent at *in vitro* and *in vivo* assays, was sequenced, assembled, annotated and functionally characterized. Whole-genome comparisons between public *P. salmonis* isolates confirmed the existence of two different genogroups associated with the LF-89 and EM-90 strains, and the bacterial pan and core genome were defined. Additionally, differences were observed at the genomic level between the *P. salmonis* reference strain and a Norwegian isolate, which is known to produce milder piscirickettsiosis outbreaks. Finally, candidate genes for invasion and intracellular survival were chosen from phylogenetically related bacteria, and annotated in *P. salmonis* using comparative genomics. These results showed the presence of several genes that might be related to bacterial pathogenesis, for example those of the type III, IV and VI secretion systems, in which some amino acidic differences within both genogroups and the Norwegian isolate were established. Altogether, these results will be relevant for understanding the host-pathogen interaction and further studies, aimed at generating new disease control strategies, should be devised using this information.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

*Piscirickettsia salmonis*, the etiological agent of piscirickettsiosis, is a Gram-negative, non-motile, aerobic, intracellular facultative bacterium. This pathogen belongs to the *Piscirickettsiaceae* family of the  $\gamma$ -proteobacteria class (Fryer and Hedrick, 2003; Fryer et al., 1992; Rozas and Enríquez, 2014), and is characterized for producing an aggressive systemic infection which involves several organs and tissues in salmonids, gaining for this reason a great commercial and animal health importance (Mauel and Miller, 2002). Although the general knowledge about the bacterium has increased, and that at least 4 complete genome sequences have been published at the NCBI database, the host-pathogen interaction mechanisms remains unclear. Some studies have recently

identified virulence related genes using comparative genomics. For example the *clpB* (a protease induced by heat stress) and *bipA* (a protein that interrupts signaling pathways) bacterial genes, have increased their expression under the infection of SHK-1 cell line, probably as a defense mechanism to inhospitable intracellular environment (Isla et al., 2014). It has been reported that *P. salmonis* encodes functional genes of a Dot/Icm-like Type IV secretion system with a potential role in the intracellular survival of bacterium (Gómez et al., 2013), and the existence of genes related to the quinolone resistance, particularly associated to a mutation on the *gyrA* gene (Henríquez et al., 2015).

*P. salmonis* is an intracellular bacterial pathogen and as such, it must invade the host cells to unleash its pathogenic potential. Thus, groups of genes probably shared by *P. salmonis* strains, might be associated with the invasion and survival inside the host cells. A recent work has demonstrated that host-clathrin pathway is required for internalization of *P. salmonis*, nevertheless, the bacterial mechanisms and specific genes involved in this process remains unknown (Ramírez et al., 2015).

The invasion genes from phylogenetically related bacteria, such as *iglC* and *pdpA* from *Francisella* (Law et al., 2014), *enhC*, *lpnE*, *rtxA*,

\* Corresponding author at: Victor Martinez, FAVET-INBIOGEN, Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Avda. Santa Rosa, 11735, Santiago, Chile.

E-mail addresses: [cl.bravo@veterinaria.uchile.cl](mailto:cl.bravo@veterinaria.uchile.cl) (C. Bravo), [vmartine@uchile.cl](mailto:vmartine@uchile.cl) (V. Martinez).

*lvhB2* and *htpB* of *Legionella pneumophila* (Newton et al., 2010) and CBU\_1260 (*OmpA*) of *Coxiella* (Martinez et al., 2014) together with genes of the type IV secretion system of *Legionella* and *Coxiella* (Segal et al., 2005) and the type VI secretion system of *Francisella* (Bröms et al., 2010), might be excellent candidate genes for a comparative genomics approach in *P. salmonis*.

In the present study a virulent isolate of *P. salmonis* was sequenced to establish a whole-genome comparative analysis using complete and draft genomic available data from Chilean *P. salmonis* isolates and raw sequencing data from a Norwegian isolate. The pan and core genome of the bacteria were predicted and genes that might be related to the invasion and intracellular survival process were discovered using comparative genomics.

## 2. Materials and methods

### 2.1. Bacterial strain

The *P. salmonis* isolate sequenced in this study (FIUCHILE-89L01) was obtained in Puerto Montt, Chile, from an outbreak of a commercial Atlantic salmon producing company, following standard procedures (Fryer et al., 1990). The bacteria were recovered and spread as *in vitro* culturing on Sf21 cell line (Birkbeck et al., 2004) and has been later cultured routinely in FN2 broth at 18 °C and 125 rpm of shaking as a cell free culture. Moreover, this isolate was confirmed as *P. salmonis* before its sequencing by an specific hydrolysis qPCR assay, as previously described (Corbeil et al., 2005; Dettleff et al., 2015).

### 2.2. Shotgun genome sequencing, assembly and annotation

Genome sequencing was carried out in an Illumina MiSeq sequencer at the FAVET-INBIOGEN laboratory, Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile. The bacterial genomic DNA (gDNA) was extracted with the NucleoSpin® Plant II kit (Macherey-Nagel) in accordance with the manufacturer's instructions and later quantified with Qubit (Life Technologies). Then, paired-end libraries were prepared with the Nextera XT Sample Prep kit (Illumina Inc.). Briefly, gDNA was tagged and tagged in a first step, and then subjected to limited-cycle PCR to add indexes and sequencing adapters. PCR gDNA library was purified using AMPure XP beads (Beckman Coulter) to exclude very short fragments. Quality was assessed using the Fragment Analyzer (Advanced Analytical) and quantified by Qubit. Libraries were sequenced 2 × 250 paired end, and initial basecalls were converted to Fastq files on the same platform.

Raw reads were trimmed using the software BBDMap version 31.40 (<http://sourceforge.net/projects/bbmap/>), and assembled using a hybrid strategy with the softwares CLC Genomics Workbench v7.5.5 (CLC bio), Spades v3.5.0 (Nurk et al., 2013) and A5-miseq v20141120 (Coil et al., 2015). Resulting contigs from each assembly were integrated with the CISA tool (Lin and Liao, 2013) to improve the sequencing quality and contiguity. Genome assembly data were deposited at NCBI genome database, and annotation was carried out using the NCBI Prokaryotic Genome Annotation Pipeline. To complement this annotation, the gene prediction from this pipeline was used to perform a subsystem annotation at the RAST Server platform (Aziz et al., 2008).

### 2.3. Phylogenetic analysis

A whole-genome alignment of 12 genomes (Table 1) of *P. salmonis* from NCBI database (draft, complete and raw reads from a Norwegian isolate) was carried out using the Gegenees tool (Ågren et al., 2012). This tool utilizes a fragmented alignment algorithm to calculate average similarity among the compared genomes using BLASTn. The average similarity was calculated with a BLAST score threshold of 40%, generating a heat plot matrix that was further used to deduce phylogenetic relationships between genomes. Subsequently, the SplitsTree software (Huson, 1998) was used for building an un-rooted phylogenetic tree employing the Neighbor-Joining algorithm. The consensus genome sequence of the Norwegian isolate was obtained after mapping raw reads (GenBank accession nos. **SRX147955** and **SRX147954**) to the *P. salmonis* reference genome (GenBank accession no. **NZ\_CP011849.2**) using the CLC Genomics Workbench v7.5.5 software with a minimum coverage of 10×.

### 2.4. Structural genomic variation in a norwegian *P. salmonis* isolate

The raw reads from a Norwegian *P. salmonis* isolate (GenBank accession nos. **SRX147955** and **SRX147954**), were mapped to the *P. salmonis* reference genome (GenBank accession no. **NZ\_CP011849.2**) and analyzed with the CLC Genomics Workbench v7.5.5 software for SNPs discovery considering the haploid genome. This information was included in a genomic context with the BRIG software (Alikhan et al., 2011).

### 2.5. Functional annotation, calculation of pan and core genome

Predicted coding sequences (CDS) were used to establish functional annotation based on the eggNOG database (Huerta-Cepas et al., 2016) by recovering the cluster of orthologous groups

**Table 1**  
*P. salmonis* analyzed genomes. Sequenced genomes statistics, April 2016.

Strain	Host	GenBank accession	Size (Mb)	GC%	Scaffolds	Genes	Proteins	Release	Level
ATCC VR-1361	<i>Oncorhynchus kisutch</i>	NZ_CP011849.1	3.45006	39.658	4	3455	3151	7/9/15	Complete Genome
PM32597B1	<i>Oncorhynchus kisutch</i>	NZ_CP012508.1	3.50968	39.6192	5	3519	3242	8/31/15	Complete Genome
PM15972A1	<i>Salmo salar</i>	NZ_CP012413.1	3.24567	39.7308	5	3306	3122	8/21/15	Complete Genome
PSCGR01	<i>Oncorhynchus mykiss</i>	NZ_CP013944.1	3.48974	39.6191	5	3521	3211	1/14/16	Complete Genome
CGR02	<i>Salmo salar</i>	NZ_CP013975.1	3.42354	39.6368	4	3435	3146	1/25/16	Complete Genome
LF-89 <sup>b</sup>	<i>Oncorhynchus kisutch</i>	NZ_ASSK02000001	2.80548	39.6	355	2842	2551	5/24/13	Contig
AUSTRAL-005	<i>Salmo salar</i>	NZ_AZYQ02000222	3.5296	39.8	227	3817	3251	2/19/14	Contig
EM-90	<i>Salmo salar</i>	NZ_JRHP01000039	3.03544	40	534	3101	2781	9/26/14	Contig
T-GIM	<i>Oncorhynchus mykiss</i>	NZ_JRHD01000180	2.93655	39.5	342	2906	2730	10/6/14	Contig
FIUCHILE-89L01 <sup>a</sup>	<i>Salmo salar</i>	NZ_LELB01000232	3.09115	39.8	285	3142	2779	6/19/15	Scaffold
LF-89 <sup>c</sup>	Not available	NZ_AMFF02000021	3.02912	40.1	284	2780	2598	9/17/12	Scaffold
FSC773	<i>Salmo salar</i>	SRX147954, SRX147955	-	-	-	-	-	9/22/12	Raw reads

All sequenced genomes available at NCBI database corresponded to Chilean isolates.

<sup>a</sup> This study.

<sup>b</sup> GenBank accession no NZ\_ASSK02000001, University of Texas.

<sup>c</sup> GenBank accession no NZ\_AMFF02000021, Fraunhofer Chile.

of proteins (COGs) using the tool hmmscan of the HMMER algorithm (Eddy, 2011) with an E-value  $\leq 10^{-6}$  for annotation. Moreover, CDS were used to retrieve gene ontology annotation (GO) with the software Blast2GO (Conesa and Götzt, 2008) by seeking out in a local protein non-redundant (nr) database and assigned annotations with an E-value  $\leq 10^{-6}$  as threshold. The obtained GO annotation were plotted with the WEGO tool (Ye et al., 2006).

The CDS from the genomes of *P. salmonis* analyzed in this study were used to predict the bacterial pan and core genome identifying orthologous groups by applying a Markov cluster (MCL) algorithm included in the OrthoMCL v2.0.9 program (Fischer et al., 2011). The parameters to assign orthologous groups were an identity of 80% and E-value cut-off of  $10^{-6}$ . Clusters that contained orthologous groups in all the strains constituted the core genome while those that did not have the corresponding orthologous group in any of the other genomes, were considered as strain specific. Finally, functional categories were assigned to gene clusters by searching against the COG annotation previously generated.

### 2.6. *P. salmonis* pathogenesis related genes selection

The CDS from phylogenetic related bacterial species *L. pneumophila* subsp. *pneumophila* str. Philadelphia-1, *F. tularensis* subsp. *tularensis* str. SCHU S4, and *C. burnetii* str. RSA 493 (RefSeq accession nos. NC\_002942.5, NC\_006570.2, NC\_002971.3 respectively), were recovered in order to generate a list of functional genes related to cell invasion and intracellular survival processes (Table S1). Subsequently, CDS from all *P. salmonis* genomes were used to create a local database and perform a BLASTp alignment using each gene from related species as query sequence (using an E-value of  $10^{-6}$  threshold of selection). Recovered genes from *P. salmonis* genomes were aligned and compared at the amino acid level with the software CLC Genomics Workbench v7.5.5.

## 3. Results

### 3.1. Genome assembly, annotation and phylogenetic analysis

After sequencing, a total of 13,828,866 reads were obtained, which were trimmed by quality obtaining a final set of 12,457,544 reads (90.08% of total sequenced reads) used in the assembly step. The genome assembly produced after the integration with CISA was represented in 285 scaffolds with a genome size of 3.09 Mb and a GC content of 39.8%. This genome was annotated in 3147 genes, 2779 CDS, 14 rRNA, and 57 tRNA, and its sequence was deposited on GenBank with the accession no. NZ\_LELB01000232. Detailed statistics together with a comprehensive assessment of the data available for the remaining genomes used in this study are presented in Table 1.

The comparative analysis using whole-genome information placed the available Chilean genomes in two main groups; the first one included by six genomes (including the isolate sequenced in this study) which grouped together with the LF-89 reference strain. The other group included the EM-90 strain and the remaining five genomes (Figs. 1 and 2). Two draft genomes described as LF-89 strain (GenBank accession nos. NZ\_ASSK02000001 and NZ\_AMFF02000021) on the same database, grouped as EM-90. On the other hand, the Norwegian isolate FSC773 was predicted in a separate branch of the phylogenetic tree.

The search for genomic variations in the Norwegian *P. salmonis* isolate revealed the presence of 64,598 mutations (Table S2), of which 88.65% corresponded to SNPs and of these, the 26.55% were nonsynonymous mutations. Some nonsynonymous SNPs were within the coding region of genes related to invasion and

<i>P. salmonis</i> strain	1	2	3	4	5	6	7	8	9	10	11	12
1: ATCC_VR-1361	100	100	100	100	99	99	90	88	87	88	87	88
2: PM32597B1	100	100	100	100	99	99	91	88	87	88	87	88
3: CGR02	100	100	100	100	99	99	90	88	87	88	87	88
4: PSCGR01	100	100	100	100	99	99	90	88	87	88	87	88
5: FIUCHILE-89L01	99	99	99	99	100	99	90	87	87	87	87	87
6: AUSTRAL-005	99	99	99	99	98	100	90	87	87	87	86	87
7: FSC773	91	91	90	90	91	90	100	87	87	87	87	87
8: PM15972A1	87	87	87	87	88	87	87	100	98	98	98	99
9: LF-89 <sup>1</sup>	87	87	87	87	87	87	87	99	100	99	99	99
10: LF-89 <sup>2</sup>	88	88	88	87	88	87	87	99	99	100	99	99
11: T-GIM	87	87	87	87	87	87	86	98	98	98	100	99
12: EM-90	87	87	87	87	87	87	86	98	98	98	98	100

LF-89<sup>1</sup>: GenBank accession no. NZ\_ASSK02000001, University of Texas.  
LF-89<sup>2</sup>: GenBank accession no. NZ\_AMFF02000021, Fraunhofer Chile.

**Fig. 1.** Phylogenetic analysis of *P. salmonis* genomes. The heat plot shows average similarity values among the analyzed genomes.

intracellular survival (see below). Nonsynonymous SNPs were contextualized at genomic level in Fig. 3.

### 3.2. Functional annotation, calculation of pan and core genome

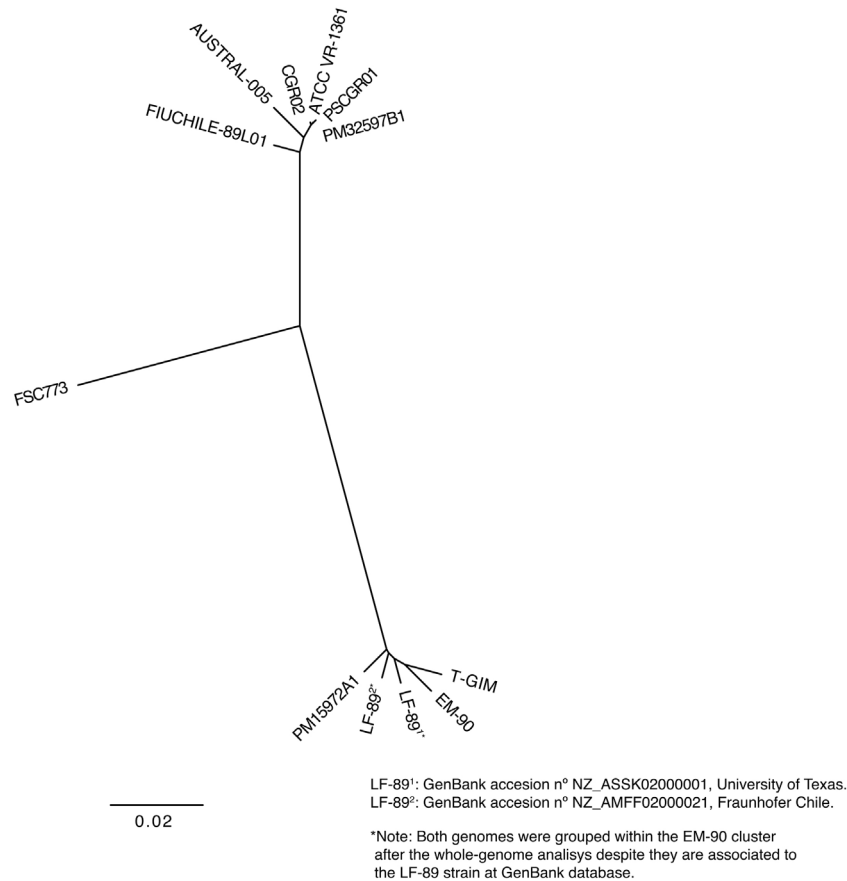
After the hmmscan search, the successful COG assignment was accomplished for 2737 of the 2779 CDS of our isolate, corresponding to a 98.5% of assignment. Thus, this genome was defined in twenty COG categories (Fig. 4), standing out the presence of flagella genes annotated within the category of cell motility ( $n=68$ ) and several genes related to secretion systems annotated within the intracellular trafficking category ( $n=58$ ). A high representation of unknown function cluster ( $n=906$ ) was found, but this cluster incorporated important annotations for example the *icmB*, *icmL*, *icmT*, *icmV*, *icmW*, *dotC* and *dotD* genes of the T4SS; and the *impB*, *impC* and *impG* genes of the T6SS (Table S3, Fig. 4). The complete functional annotation including COGs, UniProtKB, RAST and GO can be revised in Table S3. The GO annotation was plotted in Fig. S1.

For further characterization of the *P. salmonis* genome, the bacterial pan and core genome were predicted. Thus, the pan genome of *P. salmonis* was composed of at least 29,862 genes (including all analyzed genomes) grouped in 3298 homologous clusters, while 1683 genes composed the core genome. After comparing strain-specific genes, the AUSTRAL-005 strain had the highest amount of these ( $n=278$ ), while the isolate sequenced in this study had 61 specific genes (Fig. 5).

### 3.3. Invasion and intracellular survival related genes

In *P. salmonis* the invasion and intracellular survival genes are poorly characterized, hence, a list of candidate genes for each process was selected to search within *P. salmonis* using comparative genomics. As a result, three genes of the invasion list were recovered from the *P. salmonis* genome, including the gene *htpB*, which had a significant identity percentage (73.76%) with the *L. pneumophila* gene. The *htpB* gene (*groEL* in *P. salmonis*) was found in all genomes reaching a hit length of 526 of the 550 aa of the *L. pneumophila* gene. Moreover, amino acid comparison evidenced differences between the previously mentioned genogroups (Table S4A).

After searching in intracellular survival list, a total of forty-three genes were obtained from the *P. salmonis* genome. Out of these, only three genes (*icmF*, *icmH* and *pyrB*) were not present in one genome (Austral-005). This is most likely due to misassembly or uneven coverage in this genome, rather than an equivocal



**Fig. 2.** Whole-genome phylogenetic analysis of *P. salmonis* genomes. Whole-genome comparison-based phylogenetic tree of analyzed *P. salmonis* genomes. The tree was constructed with the neighbor joining algorithm.

representation of these genes. Again, the comparison of these genes evidenced differences at the amino acid level, associated to the *P. salmonis* genogroups previously mentioned (Table S4B).

#### 4. Discussion

Despite the recent publication of several complete genomes of *P. salmonis* (including the reference strain ATCC VR-1361) at NCBI database, no studies have yet used whole genome information more extensively in order to understand the molecular composition of *P. salmonis*. Hence, a comparative genomics approach was used in this study, in order to understand the global gene composition of the *P. salmonis* genome. We were mostly interested in pathogenesis related genes in two specific phases, invasion and intracellular survival. These issues were not previously described in any detail.

As a first step, a *P. salmonis* isolate was sequenced, focusing on the assembly step to improve the genome integration and the assembly quality. Thus, the bioinformatic tool CISA (Lin and Liao, 2013) was selected because it allowed us to integrate contigs generated by different assemblers. Considering this strategy, similar statistic values, that those found for draft genomes at NCBI database, were obtained with our isolate (Table S1).

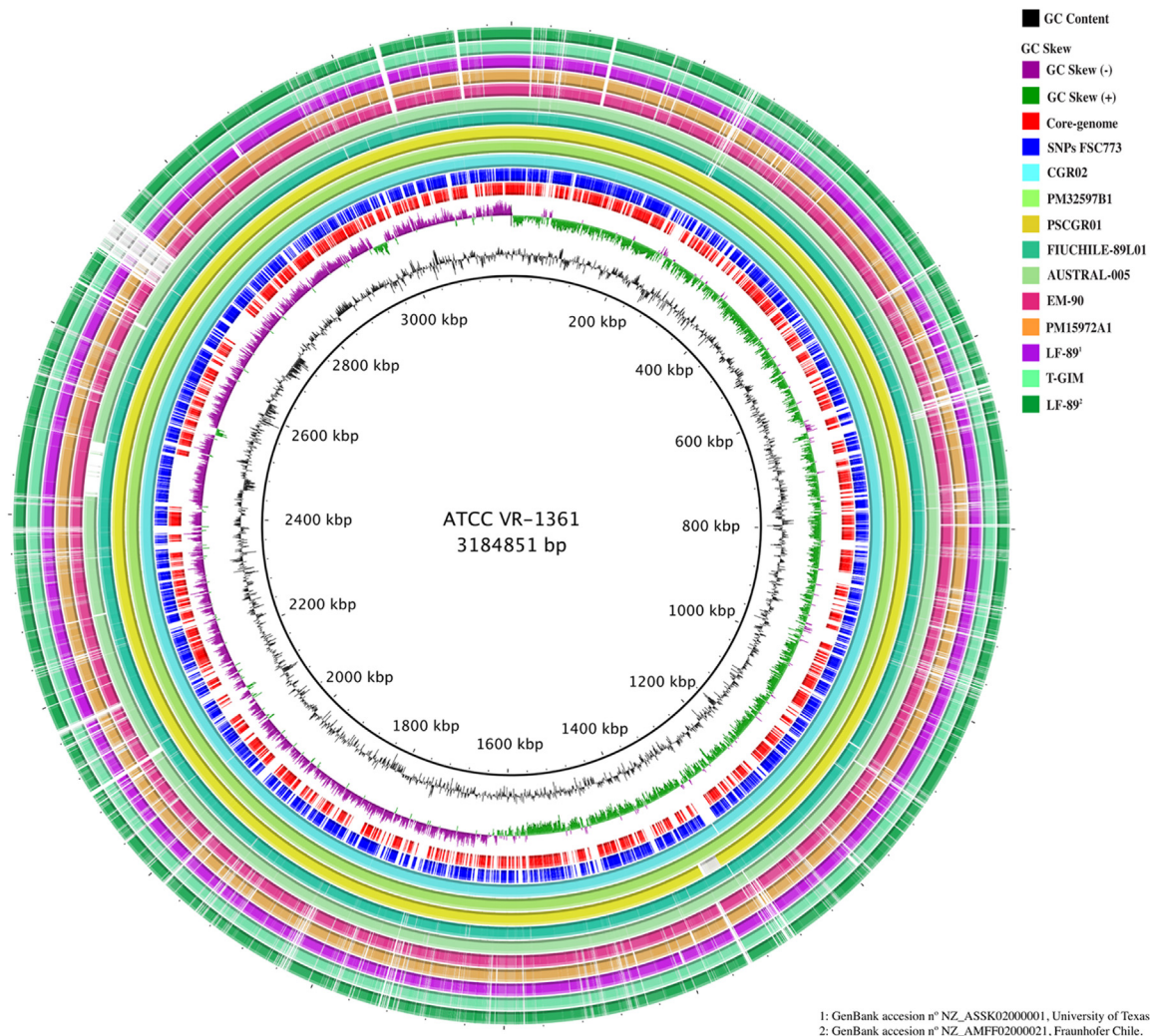
The recent information published at NCBI database allows the characterization of *P. salmonis* as a bacterium with one circular chromosome and three or four circular plasmids. When comparing whole genomes, the existence of two main genogroups corresponding to the LF-89 and the EM-90 strains were confirmed in the Chilean isolates. This is consistent with previous findings, but using conserved regions (Bohle et al., 2014; Mauel et al., 1999;

Otterlei et al., 2016). Additionally, the consensus genome generated from a Norwegian isolate was placed as an outgroup. Structural differences of this isolate were mostly observed as nonsynonymous SNPs within coding regions of pathogenic genes. This genomic composition was probably related to the different geographic origin, but more importantly on differences in pathogenicity of this isolate. This had been observed previously in (House et al., 1999) (See below).

It has been suggested that the EM-90 strain is species specific (Bohle et al., 2014), but in this analysis we found that the same strain was isolated from both *O. kisutch* and *O. mykiss* (GenBank accession nos **NZ\_ASSK02000001** and **NZ\_JRHD01000180**). This strongly suggests that this strain is ubiquitous. Furthermore, it has to be mentioned that two draft genomes belonging to the LF-89 strain (GenBank accession nos **NZ\_ASSK02000001** and **NZ\_AMFF02000021**) were placed within the EM-90 group in the whole-genome comparison.

In *P. salmonis*, differences in virulence between isolates obtained from Chile (LF-89), Canada (ATL-4-91) and Norway (NOR-92) has been previously described, being the NOR-92 isolate the least virulent (House et al., 1999). However, a possible explanation for these results is not clear. In this study, sequencing data from a Norwegian *P. salmonis* isolate was analyzed identifying 15,000 nonsynonymous SNPs mutations, which might be associated with differences in pathogenicity, as observed between Chilean and Norwegian outbreaks. In bacterial pathogens, a variety of SNPs have been discovered, belonging to the category of pathogenicity-enhancing mutations, in regulatory or structural genes, that provide a selective advantage during the course of a single infection, epidemic spread or long-term evolution of





**Fig. 3.** A circular representation of *P. salmonis* whole-genome comparison. Draft and complete genomes were aligned against the ATCC VR-1361 reference genome. Each genome was represented by a ring and the core genome was represented in red color. Blue ring represented the nonsynonymous SNPs of the FSC773 Norwegian isolate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

virulence (Sokurenko et al., 1999). Furthermore, the SNPs described in this study were found within coding regions of the invasion related genes *htpB*, *enhC* and *CBU\_1260*; and the intracellular survival related genes *carA*, *dotB*, *dotC*, *dotD*, *icmB*, *icmE*, *icmF*, *icmJ*, *icmK*, *icmL*, *icmO*, *icmP*, *icmV*, *icmW*, *lpg\_1905*, *mip*, *pilD*, *pyrB*, *rep* and *rpoS*. Further analyses are required in order to understand more deeply these striking results, which may explain differences in virulence between isolates.

As part of the genomic characterization, the pan and core genome of *P. salmonis* were defined as a way to cluster existing genes in all analyzed genomes. Incorporating the CDS available for analyzed genomes, 29,862 and 1683 genes were detected within the pan and core genome respectively. The core genome of *P. salmonis* had a highly heterogeneous composition standing out the cell motility category, in which 52 genes were annotated, corresponding 29 to the flagellar system. This is a remarkable fact, since *P. salmonis* has been described as a non-motile bacteria (Fryer et al., 1992, 1990; Rozas and Enríquez, 2014), although some studies have previously described the presence of flagellar genes (Machuca and Martinez, 2016; submitted), which have been used as immunogenic factors in recombinant proteins against the bacteria (Wilhelm et al., 2006). On the other hand, the presence of these genes could be related to its participation in a T3SS, possibly due the evolutionary relationship between this secretion system

and the flagella (Blocker et al., 2003; Notti et al., 2015). Other findings suggest the presence of a type IV pilus (maybe related to bacterial adhesion and invasion), and several genes associated to both, a T4SS and a T6SS (Bravo and Martinez, 2016; submitted). These genes explain intracellular survival in *L. pneumophila*, *C. burnetti* and *F. tularensis* (Bröms et al., 2010; Segal et al., 2005).

The annotation of the *P. salmonis* genome revealed 3 important genes related to invasion. The most important example is the *groEL* gene, which had the least differences at the amino acidic level between strains. Chaperonin 60 (Cpn60 or GroEL) is one of the most conserved and essential proteins in living organisms, because along with chaperonin 10 (Cpn10 or GroES) takes part in a protein-folding machine required for the appropriate folding of many proteins and for recycling of others misfolded ones. They are also known as heat-shock proteins (HSPs), since heat stress leads to a strong induction of their expression (Kupper et al., 2014). Bacterial GroEL proteins have shown to initiate apoptotic signaling (Equils et al., 2006; Jha et al., 2011), and it has been demonstrated that *Aggregatibacter actinomycetemcomitans* GroEL protein induces apoptosis, preferentially in primary salmon CD8+ T cells (Kant et al., 2015). In *L. pneumophila* the HtpB protein (GroEL) found on the bacterium surface (as confirmed by its susceptibility to trypsin and neutralization by antibodies) turned out to be an invasion factor, mediating the internalization by HeLa cells (Garduno et al., 1998).

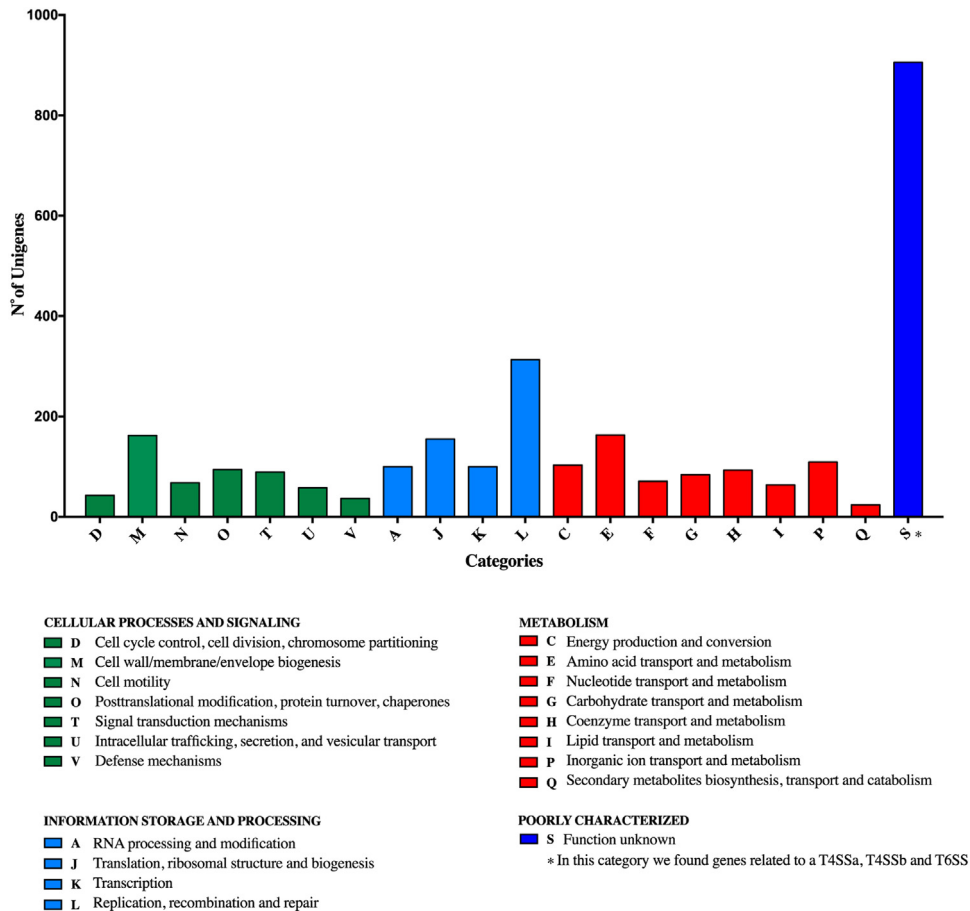


Fig. 4. *P. salmonis* functional annotation. COG distribution of the *P. salmonis* strain FIUCHILE-89L01 genome annotation.

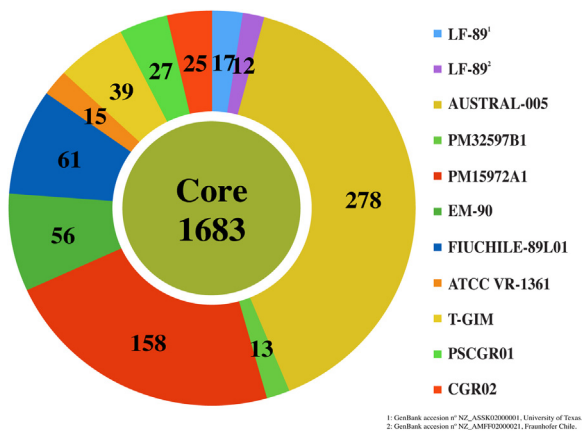


Fig. 5. Representation of *P. salmonis* core genome and specific strain genes. Information was generated with the OrthoMCL software.

Moreover, HtpB could not complement a temperature sensitive GroEL defect in *E. coli* strain CG218, which indicates that both proteins are not functionally equivalent at this level. We found these genes within the core genome of *P. salmonis*.

The gene *CBU\_1260 (ompA)* of *C. burnetii*, which was described as the first invasins of this bacterium, was found within the *P. salmonis* genome. Mutations in *ompA* decreased *Coxiella* internalization and replication within host cells and internalization was efficiently inhibited by pre-treating host cells with purified OmpA or by incubating *Coxiella* with a specific anti-OmpA antibody prior

to host cell infection (Martinez et al., 2014). In the *P. salmonis* genome, this gene was annotated as an outer membrane insertion signal protein and was included in the core genome of the bacteria. Furthermore, in all invasion related genes, differences at amino acidic level between the two *P. salmonis* genogroups were found.

Regarding intracellular survival genes, 19 of the *dot/icm* system were predicted, which 3 of them had been previously described (*dotA*, *dotB* and *icmK*) (Gómez et al., 2013). These genes might be part of the intracellular genomic repertoire required by the bacteria to survive in the host. Genes that were not previously annotated or mentioned were also found in this study. For example, the *iglA* gene and the regulator operon *mglA* which were fundamental for intracellular replication of *F. tularensis* (Sjöstedt, 2006); the *lpg1905* gene, a functional NTPDase, whose inactivation results in the defective replication of *L. pneumophila* in amoebae, epithelial cells, and macrophages (Sansom et al., 2007); and the *mip* gene (macrophage infectivity potentiator), a membrane-associated protein that it is mainly found on the bacterial surface, and that it is necessary for efficient replication within host cells. Among these genes, *mglA*, *lpg\_1905* and *mip* were found in the *P. salmonis* core genome.

Differences at the amino acidic level in intracellular survival genes were also found between the *P. salmonis* strains. A group of genes (*dotB*, *dotC*, *dotD*, *icmJ*, *icmT*, *icmO*, *icmG*, *icmE*) had significant amino acid differences within the LF-89 genogroup in one particular genome (PSCGR01), which was isolated from the kidney of a rainbow trout (*Oncorhynchus mykiss*) specimen; a particularly susceptible fish. These genes could explain in part the differential response of different salmonid species to *P. salmonis*, but further studies are required to confirm these findings.

## 5. Conclusions

Using the whole-genome available data of *P. salmonis* we confirmed the presence of several single nucleotide mutations within coding genes related to pathogenic processes such as invasion and intracellular survival. This study gives more insights at the whole genome level on the two main genogroups of *P. salmonis*, associated to the LF-89 and EM-90 strains. The functional annotation of the sequenced field isolate of this study, has shown the presence of several COG categories, standing out the cell motility and intracellular trafficking, in which annotations related to flagella and secretion systems, could be associated to early host-pathogen interaction mechanisms. A comparative genomics approach was used to recover invasion and intracellular survival related genes from the *P. salmonis* genome, and indeed, the presence of some of these genes in the core genome of the bacteria was clear, suggesting its importance in all analyzed genomes, at the chromosomal level rather than on plasmids. Aminoacidic changes in particular pathogenic genes of the bacteria, might explain in part, distinct disease presentations within the Chilean genogroups and the Norwegian isolate analyzed in this study.

These results may be used in further studies, aimed to improve understanding of the disease and thus, encouraging its control.

## Conflict of interest statement

The authors declare that they have no competing interests.

## Acknowledgements

This work was supported by the Chilean Government through CONICYT (FONDECYT 1120608 and FONDEF IDEA 14110090) to V. Martinez. Cristian Bravo acknowledges the support of scholarship for National PhD students from CONICYT. Authors acknowledges to Vanessa Garrido for contributing in the English review in this document.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.vetmic.2016.10.015>.

## References

- Ågren, J., Sundström, A., Häfström, T., Segerman, B., 2012. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* 7, e39107. doi: <http://dx.doi.org/10.1371/journal.pone.0039107.s018>.
- Alikhan, N.-F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12, 402. doi: <http://dx.doi.org/10.1186/1471-2164-12-402>.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi: <http://dx.doi.org/10.1186/1471-2164-9-75>.
- Birkbeck, T.H., Griffen, A.A., Reid, H.I., Laidler, L.A., Wadsworth, S., 2004. Growth of *Piscirickettsia salmonis* to high titers in insect tissue culture cells. *Infect. Immun.* 72, 3693–3694. doi: <http://dx.doi.org/10.1128/IAI.72.6.3693-3694.2004>.
- Blocker, A., Komoriya, K., Aizawa, S.-I., 2003. Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc. Natl. Acad. Sci. USA* 100, 3027–3030. doi: <http://dx.doi.org/10.1073/pnas.053535100>.
- Bohle, H., Henriquez, P., Grothusen, H., Navas, E., Sandoval, A., Bustamante, F., Bustos, P., Mancilla, M., 2014. Comparative genome analysis of two isolates of the fish pathogen *Piscirickettsia salmonis* from different hosts reveals major differences in virulence-associated secretion systems. *Genome Announc.* 2, e01219-14. doi: <http://dx.doi.org/10.1128/genomeA.01219-14>.
- Bröms, J.E., Sjöstedt, A., Lavander, M., 2010. The role of the Francisella Tularensis pathogenicity island in type VI secretion intracellular survival and modulation of host cell signaling. *Front. Microbio.* 1, 136. doi: <http://dx.doi.org/10.3389/fmicb.2010.00136>.
- Coil, D., Jospin, G., Darling, A.E., 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31, 587–589. doi: <http://dx.doi.org/10.1093/bioinformatics/btu661>.
- Conesa, A., Götz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genom.* 2008, 1–12. doi: <http://dx.doi.org/10.1093/nar/gkl197>.
- Corbeil, S., Hyatt, A.D., Crane, M.S.J., 2005. Characterisation of an emerging rickettsia-like organism in Tasmanian farmed Atlantic salmon *Salmo salar*. *Dis. Aquat. Org.* 64, 37–44. doi: <http://dx.doi.org/10.3354/dao064037>.
- Dettleff, P., Bravo, C., Patel, A., Martinez, V., 2015. Patterns of *Piscirickettsia salmonis* load in susceptible and resistant families of *Salmo salar*. *Fish Shellfish Immunol.* 45, 67–71. doi: <http://dx.doi.org/10.1016/j.fsi.2015.03.039>.
- Eddy, S.R., 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* doi: <http://dx.doi.org/10.1371/journal.pcbi.1002195.g006>.
- Equils, O., Lu, D., Gatter, M., Witkin, S.S., Bertolotto, C., Arditi, M., McGregor, J.A., Simmons, C.F., Hobel, C.J., 2006. *Chlamydia* heat shock protein 60 induces trophoblast apoptosis through TLR4. *J. Immunol.* 177, 1257–1263. doi: <http://dx.doi.org/10.4049/jimmunol.177.2.1257>.
- Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., Stoeckert, C.J., 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr. Protoc. Bioinform.* 1–19. doi: <http://dx.doi.org/10.1002/0471250953.bi0612s35> (Chapter 6, Unit 6.12).
- Fryer, J.L., Hedrick, R.P., 2003. *Piscirickettsia salmonis*: a gram-negative intracellular bacterial pathogen of fish. *J. Fish Dis.* 26, 251–262.
- Fryer, J.L., Lannan, C.N., Garcés, L.H., Larenas, J.J., Smith, P.A., 1990. Isolation of a Rickettsiales-like organism from diseased coho salmon (*Oncorhynchus kisutch*) in Chile. *Fish Pathol.* 25, 107–114.
- Fryer, J.L., Lannan, C.N., Giovannoni, S.J., Wood, N.D., 1992. *Piscirickettsia salmonis* gen. nov., sp. nov., the causative agent of an epizootic disease in salmonid fishes. *Int. J. Syst. Bacteriol.* 42, 120–126. doi: <http://dx.doi.org/10.1099/00207713-42-1-120>.
- Gómez, F.A., Tobar, J.A., Henríquez, V., Sola, M., Altamirano, C., Marshall, S.H., 2013. Evidence of the presence of a functional Dot/Icm type IV-B secretion system in the fish bacterial pathogen *Piscirickettsia salmonis*. *PLoS One* 8, e54934. doi: <http://dx.doi.org/10.1371/journal.pone.0054934.s004>.
- Garduno, R.A., Garduno, E., Hoffman, P.S., 1998. Surface-associated hsp60 chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. *Infect. Immun.* 66, 4602–4610.
- Henriquez, P., Bohle, H., Bustamante, F., Bustos, P., Mancilla, M., 2015. Polymorphism in *gyrA* is associated to quinolones resistance in Chilean *Piscirickettsia salmonis* field isolates. *J. Fish Dis.* 38, 415–418. doi: <http://dx.doi.org/10.1111/jfd.12255>.
- House, M.L., Bartholomew, J.L., Winton, J.R., Fryer, J.L., 1999. Relative virulence of three isolates of *Piscirickettsia salmonis* for coho salmon *Oncorhynchus kisutch*. *Dis. Aquat. Org.* 35, 107–113. doi: <http://dx.doi.org/10.3354/dao035107>.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., Bork, C., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: <http://dx.doi.org/10.1093/nar/gkv1248>.
- Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14, 68–73.
- Isla, A., Haussmann, D., Vera, T., Kausel, G., Figueroa, J., 2014. Identification of the *clpB* and *bipA* genes and an evaluation of their expression as related to intracellular survival for the bacterial pathogen *Piscirickettsia salmonis*. *Vet. Microbiol.* 1–5. doi: <http://dx.doi.org/10.1016/j.vetmic.2014.08.014>.
- Jha, H.C., Srivastava, P., Vardhan, H., Singh, L.C., Bhengraj, A.R., Prasad, J., Mittal, A., 2011. *Chlamydia pneumoniae* heat shock protein 60 is associated with apoptotic signaling pathway in human atherosclerotic plaques of coronary artery disease patients. *J. Cardiol.* 58, 216–225. doi: <http://dx.doi.org/10.1016/j.jcc.2011.07.010>.
- Kant, M., Akgül, B., Nalbant, A., 2015. Endogenous heat shock protein GroEL of *A. actinomycetemcomitans* preferentially targets primary human CD8+ T cells. *Turk. J. Biol.* 39, 104–110. doi: <http://dx.doi.org/10.3906/biy-1404-64>.
- Kupper, M., Gupta, S.K., Feldhaar, H., Gross, R., 2014. Versatile roles of the chaperonin GroEL in microorganism-insect interactions. *FEMS Microbiol. Lett.* 353, 1–10. doi: <http://dx.doi.org/10.1111/1574-6968.12390>.
- Law, H.T., Sriram, A., Fevang, C., Nix, E.B., Nano, F.E., Guttman, J.A., 2014. IgIC and PdpA are important for promoting *Francisella* invasion and intracellular growth in epithelial cells. *PLoS One* 9, e104881. doi: <http://dx.doi.org/10.1371/journal.pone.0104881.s002>.
- Lin, S.-H., Liao, Y.-C., 2013. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8, e60843. doi: <http://dx.doi.org/10.1371/journal.pone.0060843.t003>.
- Martinez, E., Cantet, F., Fava, L., Norville, I., Bonazzi, M., 2014. Identification of OmpA, a *Coxiella burnetii* protein involved in host cell invasion, by multi-phenotypic high-content screening. *PLoS Pathog.* 10, e1004013. doi: <http://dx.doi.org/10.1371/journal.ppat.1004013.s010>.
- Mauel, M.J., Miller, D.L., 2002. *Piscirickettsiosis* and *piscirickettsiosis*-like infections in fish: a review. *Vet. Microbiol.* 87, 279–289.
- Mauel, M.J., Giovannoni, S.J., Fryer, J.L., 1999. Phylogenetic analysis of *Piscirickettsia salmonis* by 16S, internal transcribed spacer (ITS) and 23S ribosomal DNA sequencing. *Dis. Aquat. Org.* 35, 115–123. doi: <http://dx.doi.org/10.3354/dao035115>.



- Newton, H.J., Ang, D.K.Y., van Driel, I.R., Hartland, E.L., 2010. Molecular pathogenesis of infections caused by *Legionella pneumophila*. Clin. Microbiol. Rev. 23, 274–298. doi:http://dx.doi.org/10.1128/CMR.00052-09.
- Notti, R.Q., Bhattacharya, S., Lilic, M., Stebbins, C.E., 2015. A common assembly module in injectisome and flagellar type III secretion sorting platforms. Nat. Commun. 6, 7125. doi:http://dx.doi.org/10.1038/ncomms8125.
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S.R., Woyke, T., Mclean, J.S., Lasken, R., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comput. Biol. 20, 714–737. doi:http://dx.doi.org/10.1089/cmb.2013.0084.
- Otterlei, A., Brevik, Ø.J., Jensen, D., Duesund, H., Sommerset, I., Frost, P., Mendoza, J., McKenzie, P., Nylund, A., Apablaza, P., 2016. Phenotypic and genetic characterization of *Piscirickettsia salmonis* from Chilean and Canadian salmonids. BMC Vet. Res. 12, 1–11. doi:http://dx.doi.org/10.1186/s12917-016-0681-0.
- Ramírez, R., Gómez, F.A., Marshall, S.H., 2015. The infection process of *Piscirickettsia salmonis* in fish macrophages is dependent upon interaction with host-cell clathrin and actin. FEMS Microbiol. Lett. 362, 1–8. doi:http://dx.doi.org/10.1093/femsle/fnu012.
- Rozas, M., Enríquez, R., 2014. Piscirickettsiosis and *Piscirickettsia salmonis* in fish: a review. J. Fish Dis. 37, 163–188. doi:http://dx.doi.org/10.1111/jfd.12211.
- Sansom, F.M., Newton, H.J., Crikis, S., Cianciotto, N.P., Cowan, P.J., d'Apice, A.J.F., Hartland, E.L., 2007. A bacterial ecto-triphosphate diphosphohydrolase similar to human CD39 is essential for intracellular multiplication of *Legionella pneumophila*. Cell. Microbiol. 9, 1922–1935. doi:http://dx.doi.org/10.1111/j.1462-5822.2007.00924.x.
- Segal, G., Feldman, M., Zusman, T., 2005. The Icm/Dot type-IV secretion systems of *Legionella pneumophila* and *Coxiella burnetii*. FEMS Microbiol. Rev. 29, 65–81. doi:http://dx.doi.org/10.1016/j.femsre.2004.07.001.
- Sjöstedt, A., 2006. Intracellular survival mechanisms of *Francisella tularensis*, a stealth pathogen. Microbes Infect. 8, 561–567. doi:http://dx.doi.org/10.1016/j.micinf.2005.08.001.
- Sokurenko, E.V., Hasty, D.L., Dykhuizen, D.E., 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends Microbiol. 7, 191–195.
- Wilhelm, V., Miquel, A., Burzio, L.O., Roseblatt, M., Engel, E., Valenzuela, S., Parada, G., Valenzuela, P.D.T., 2006. A vaccine against the salmonid pathogen *Piscirickettsia salmonis* based on recombinant proteins. Vaccine 24, 5083–5091. doi:http://dx.doi.org/10.1016/j.vaccine.2006.03.027.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34, W293–W297. doi:http://dx.doi.org/10.1093/nar/gkl031.