



OPEN

Proximity ligation strategy for the genomic reconstruction of microbial communities associated with the ectoparasite *Caligus rogercresseyi*

Diego Valenzuela-Miranda^{1✉}, Ana Teresa Gonçalves^{1,2}, Valentina Valenzuela-Muñoz¹, Gustavo Nuñez-Acuña¹, Ivan Liachko³, Bradley Nelson³ & Cristian Gallardo-Escarate¹

The sea louse *Caligus rogercresseyi* has become one of the main constraints for the sustainable development of salmon aquaculture in Chile. Although this parasite's negative impacts are well recognized by the industry, some novel potential threats remain unnoticed. The recent sequencing of the *C. rogercresseyi* genome revealed a large bacterial community associated with the sea louse, however, it is unknown if these microorganisms should become a new focus of sanitary concern. Herein, chromosome proximity ligation (Hi-C) coupled with long-read sequencing were used for the genomic reconstruction of the *C. rogercresseyi* microbiota. Through deconvolution analysis, we were able to assemble and characterize 413 bacterial genome clusters, including six bacterial genomes with more than 80% of completeness. The most represented bacterial genome belonged to the fish pathogen *Tenacibaculum ovolyticum* (97.87% completeness), followed by *Dokdonia* sp. (96.71% completeness). This completeness allowed identifying 21 virulence factors (VF) within the *T. ovolyticum* genome and four antibiotic resistance genes (ARG). Notably, genomic pathway reconstruction analysis suggests putative metabolic complementation mechanisms between *C. rogercresseyi* and its associated microbiota. Taken together, our data highlight the relevance of Hi-C techniques to discover pathogenic bacteria, VF, and ARGs and also suggest novel host-microbiota mutualism in sea lice biology.

The sea louse, *Caligus rogercresseyi*, is a marine ectoparasite that has become one of the main constraints for developing sustainable aquaculture in Chile¹. This copepod is the etiological agent of caligidosis, a prevalent and severe disease that has been established as one of the top priorities for Chilean salmon farming². Lice infestation produces severe deleterious effects on the host, including a reduction in growth performance, a reduced feed-conversion efficiency, severe chronic stress, skin wounds, and loss of scales that increase the susceptibility of the infested salmonids to secondary infections^{3,4}. Currently, sea lice control strategies are mainly based on pharmacological treatments, increasing the average production costs by \$1.4 US/kg⁵. Besides this well-known negative impact, novel evidence suggests the presence of a large and diverse microbiota associated with *C. rogercresseyi*⁶. However, the biological role of these microorganisms or their potential importance on salmon aquaculture remains largely elusive.

Over the past years, the idea that microbes play pivotal roles in the development of almost all living organisms has gained consensus⁷. Microbes can drive physiological and evolutionary changes due to their interaction with their host⁸. Particularly in parasites, the microbiota can actively participate during the infective process by producing toxins, redirecting the host immune system, or increasing the parasite's virulence^{7,9,10}. Herein, the parasite's microbiota has been proposed as a target to develop alternative therapeutic strategies to cope with infestations^{11,12}.

¹Interdisciplinary Center for Aquaculture Research (INCAR), University of Concepción, P. O. Box 160-C, Concepción, Chile. ²GreenCoLab-Associação Oceano Verde, University of Algarve, Campus de Gambelas, Faro, Portugal. ³Phase Genomics, Inc., Seattle, USA. ✉email: divalenzuela@udec.cl

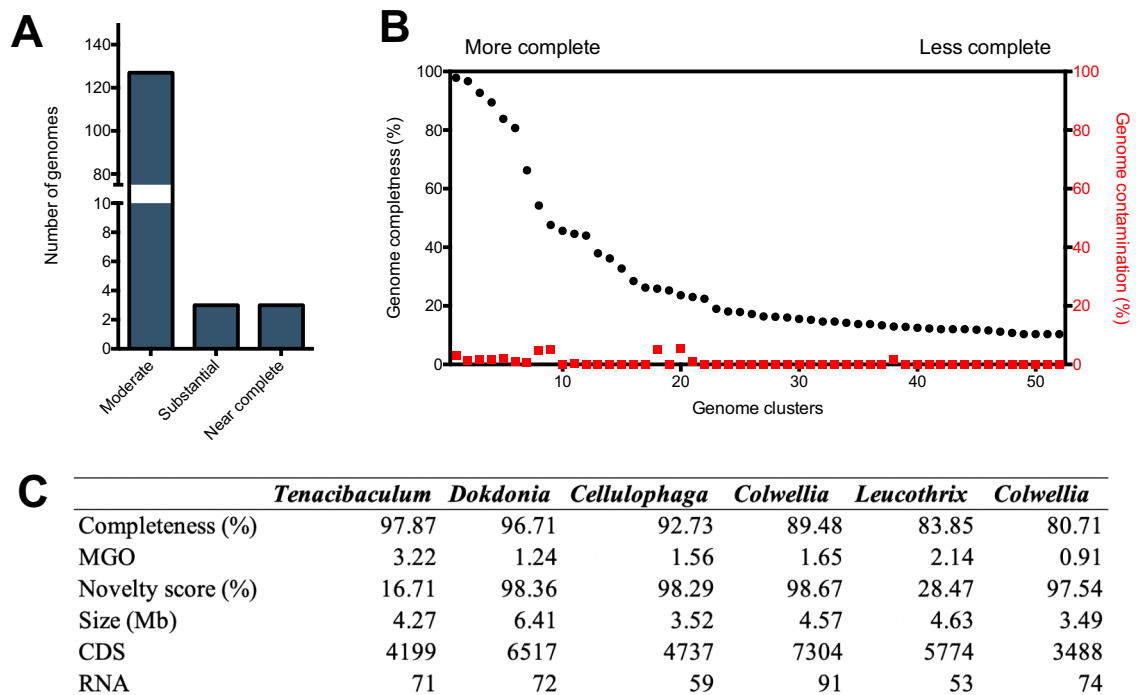


Figure 1. Summary of the genomic reconstruction of bacterial clusters associated with *Caligus rogercresseyi*. (A) Genome cluster classification according with their completeness in near complete (>95%), substantial (>80%), and moderate (<80%). (B) Genome completeness and contamination among the 50 most represented bacterial clusters. (C) Main genomic features of the top 5 represented genomes clusters in sea lice microbiota.

Vector-borne diseases are another major issue regarding parasite-associated microbes in arthropods, responsible for transmitting virus, bacteria, and protozoa among vertebrate hosts¹³. Copepods are widely colonized by bacteria^{14,15}, among which pathogenic species are usually identified^{16,17}. Known fish pathogens have been found in other caligid parasitic copepods, such as *Vibrio* sp. in *Caligus lalandei*¹⁸. Likewise, it has been suggested that the sea louse *Lepeophtheirus salmonis* could act as a vector of several diseases, including infectious hematopoietic necrosis virus (IHNV), the infectious salmon anemia virus (ISAV), and furunculosis^{19–21}. Our research group previously identified bacterial pathogens present in *C. rogercresseyi* associated microbiota, including *Vibrio*, *Tenacibaculum*, and *Aeromonas* genus⁶. However, these species' genomic features are poorly understood, hindering the biological role of the sea lice microbiota and its potential impact on salmon aquaculture.

High-throughput chromosomal conformation capture (Hi-C) coupled with long-read sequencing has offered the possibility to deeply understand microbiomes by reconstructing individual bacterial genomes from complex microbe samples^{22–24}. However, this approach has been mainly applied in the study of mammalian gut microbiota and it is less frequently used in marine environments. This work aimed to reconstruct the genetic background of sea lice microbiota to evaluate possible new threats for salmon aquaculture and to reveal novel insights about the biological roles of these bacteria in *C. rogercresseyi* biology. Through Hi-C and bioinformatic deconvolution, we were able to reconstruct different bacterial genome clusters, allowing the characterization of complete genomes of pathogenic salmon bacteria and identifying virulence factors (VF), and antibiotic resistance genes (ARG) present in sea lice microbiota. Our data also reveal that some bacteria associated with *C. rogercresseyi* might be fulfilling key metabolic roles, suggesting novel host-microbiota mutualism in sea lice biology.

Results

Organism-associated microbiota can fulfill key roles in the development of almost all living organisms. In parasites, this role can comprise both biological and functional roles during pathogenesis. Hi-C sequencing and a bioinformatic deconvolution analysis were used to gain insights about sea lice microbiota's relevance in aquaculture. Through this analysis, we were able to assemble and characterize 413 bacterial genome clusters. These clusters were then analyzed for completeness and genetic contamination, based on the marker gene overrepresentation (MGO). Based on their completeness, the genomes were classified as near-complete (100% to 95%), substantial complete (95–85%), and moderate complete (<85%). The Hi-C sequencing and deconvolution analysis revealed that two genome clusters accounted for over 95% completeness, two others were substantial complete with around 83 and 89% completeness, and with the more significant number of clusters classified as moderate complete (Fig. 1A). Regarding gene overrepresentation, the analysis evidenced a low degree of genetic cross-contamination among bacterial clusters, with no values above 4% (Fig. 1B). Genomic clusters with the most completeness comprised five different genera, including *Tenacibaculum*, *Dokdonia*, *Cellulophaga*, *Colwellia*, and *Leucothrix* (Fig. 1C). The genome cluster belonging to the known fish pathogen *Tenacibaculum* reached over 97% completeness with an estimated genome size of 4.27 Mb. Interestingly, this approach allowed the

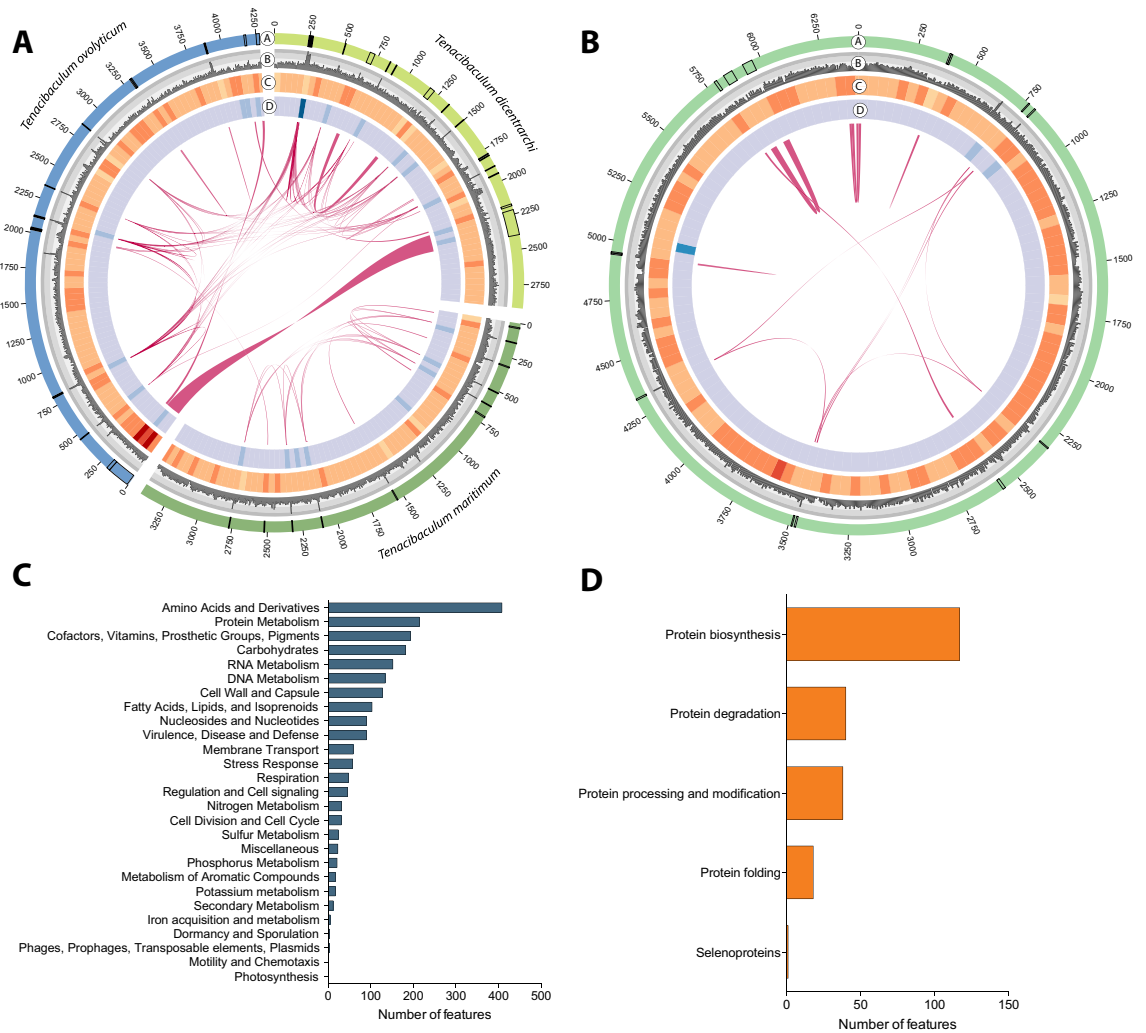


Figure 2. General overview of *Tenacibaculum* and *Dokdonia* sp. genomes. (A) Circos⁶⁵ plot representing different *Tenacibaculum* genomes including the genome size (a), GC content (b), CDS density (c), ncRNA density (d) and syntenic blocks (ribbons). (B) Genome characterization of *Dokdonia* sp. cluster. Circos⁶⁵ plot representing *Dokdonia* sp. genome including the genome size (a), GC content (b), CDS density (c), ncRNA density (d) and syntenic blocks (ribbons). (C) Functional annotation of the CDS characterize in *Dokdonia* sp. genome provided by RAST annotation. (D) Functional annotation of the CDS within amino acids metabolism found in *Dokdonia* sp. genome.

genome reconstruction of a relatively large bacterial genome (>6.41 Mb) with estimated completeness of over 96%. Compared with the RefSeq database, some of the genome clusters accounted for a novelty score above 95% (*Dokdonia*, *Cellulophaga*, and *Colwellia*), thus evidencing that this approach also allowed the reconstruction of previously unknown bacterial genomes.

We further explored the *Tenacibaculum* cluster, as it is a genus known to comprise different pathogenic species that are relevant for Chilean aquaculture²⁵. Based on all the 16S genes fully annotated in the genome, we were able to identify this particular cluster as *T. ovolyticum* (Fig. S1). The circos plot, including different features such as GC content, coding, and non-coding gene density, was constructed to gain a general idea about *T. ovolyticum* genome structure (Fig. 2A). Comparative genome analyses among *T. dicentrarchi*, *T. maritimum*, and *T. ovolyticum* showed that different syntenic blocks between *T. dicentrarchi* and *T. maritimum* are shared with *T. ovolyticum*. Notably, more syntenic conservation was observed between *T. ovolyticum* and *T. dicentrarchi* (Fig. 2A). Given the degree of completeness, we further explored the presence of virulence factors (VFs) in the *T. ovolyticum* genome. Using the Virulence Factor Database (VFDB) as a reference, we were able to identify 21 genes present in the *T. ovolyticum* genome, including genes belonging to the Dot/Icm type IVB secretion system, 60 K heat shock proteins (HtpB), and catalases (*katA*) genes among others (Table 1). Moreover, using the MEGARes database, the results evidenced antibiotic resistance genes (ARGs) within all the bacterial genome clusters. Through this analysis, it was possible to identify 4 ARGs, including the *Tetracycline efflux Na⁺/H⁺ transporter tet(35)*, *Plasmid-mediated quinolone resistance protein (QnrS2)*, *Integron chloramphenicol acetyltransferase (catB9)*, and *Oxa beta-lactamase (Oxa-209)* genes (Table 1).

Gene	Description	E-value	Closely related specie
VF			
basC	Acinetobactin biosynthesis protein	3.53E-100	<i>Acinetobacter baumannii</i>
basG	Acinetobactin biosynthesis protein	1.39E-109	<i>Acinetobacter baumannii</i>
bpIC	Lipopolysaccharide biosynthesis protein	1.49E-107	<i>Bordetella pertussis</i>
bpIC	Lipopolysaccharide biosynthesis protein	2.85E-101	<i>Bordetella pertussis</i>
bpID	UDP-N-acetylglucosamine 2-epimerase	6.00E-124	<i>Bordetella pertussis</i>
bpID	UDP-N-acetylglucosamine 2-epimerase	9.67E-114	<i>Bordetella pertussis</i>
bpIE	Probable glycosyl transferase	2.23E-121	<i>Bordetella pertussis</i>
cap8D	Capsular polysaccharide synthesis enzyme	3.77E-114	<i>Staphylococcus aureus</i> subsp. <i>aureus</i>
CBU_0270	Dot/Icm type IVB secretion system	4.45E-169	<i>Coxiella burnetii</i>
clpC	Endopeptidase Clp ATP-binding chain C	0	<i>Listeria monocytogenes</i>
clpC	Endopeptidase Clp ATP-binding chain C	0	<i>Listeria monocytogenes</i>
clpC	Endopeptidase Clp ATP-binding chain C	1.25E-112	<i>Listeria monocytogenes</i>
galE	UDP-glucose 4-epimerase	5.83E-103	<i>Haemophilus influenzae</i>
htpB	Hsp60, 60 K heat shock protein HtpB	0	<i>Legionella pneumophila</i> subsp. <i>pneumophila</i>
htpB	Hsp60, 60 K heat shock protein HtpB	0	<i>Legionella pneumophila</i> subsp. <i>pneumophila</i>
katA	Catalase	0	<i>Legionella pneumophila</i> subsp. <i>pneumophila</i>
katA	Catalase	0	<i>Legionella pneumophila</i> subsp. <i>pneumophila</i>
pilR	Two-component response regulator	6.78E-104	<i>Pseudomonas aeruginosa</i>
relA	Probable GTP pyrophosphokinase	1.03E-136	<i>Mycobacterium tuberculosis</i>
rffG	dTDP-glucose 46-dehydratase	4.95E-113	<i>Haemophilus influenzae</i>
tviB	Vi polysaccharide biosynthesis protein	1.88E-160	<i>Salmonella enterica</i> subsp. <i>enterica</i>
ARG			
tet(35)	Tetracycline efflux Na ⁺ /H ⁺ transporter	0	<i>Vibrio harveyi</i>
QnrS2	Plasmid-mediated quinolone resistance protein	3.13E-151	<i>Salmonella enterica</i>
catB9	Integron chloramphenicol acetyltransferase	2.03E-77	<i>Vibrio cholerae</i>
OXA-209	Oxa beta lactamase	2.52E-71	<i>Riemerella anatipestifer</i>

Table 1. Genes coding for virulence factors (VF) and antibiotic resistance genes identified in the genomes associated with *C. rogercresseyi*.

The ubiquitous sets of genes determine genome completeness, and a single copy within a phylogenetic lineage could be used as a proxy for bacterial abundance in metagenomic studies. As one of the largest and more complete genomes represented in sea lice microbiota, we explored the *Dokdonia* sp. genome, representing their main features through a circos plot (Fig. 2B). The genome size of *Dokdonia* sp. was estimated in 6.41 Mb, accounting for 6517 CDS and 72 RNAs. The functional annotation evidence a large number of genes associated with different metabolism features, including protein, amino acids, and derivatives metabolism, followed by cofactors, vitamins, pigments, and others (Fig. 2C). Most of the genes were annotated in protein biosynthesis, degradation, processing, and modification (Fig. 2D). Given the large number of protein metabolism genes found in the *Dokdonia* sp. genome, and the recent availability of the sea lice genome (ASM1338718v1)²⁶, we further investigate possible metabolic complementation between sea lice and its associated microbiota. To do this, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used as a reference to annotate the genes in different metabolic pathways, including lipid, carbohydrate, and amino acid metabolism (Fig. 3A). The results evidenced that the more significant number of metabolic genes annotated in the lice microbiota but not present in the sea louse genome belongs to the amino acid metabolism, followed by carbohydrate and lipid metabolism (Fig. 3B). Regarding specific pathways mostly encoded by parasite microbiota, the fatty acid biosynthesis, butanoate metabolism, phenylalanine, tyrosine, and tryptophan biosynthesis were identified for lipid, carbohydrate, and amino acid metabolism, respectively (Fig. 3B). The complete list of all the metabolic KEGG pathways analyzed can be found in the supplementary material 1. As an example, we analyzed the branched chain amino acid (BCA) metabolism (Fig. 4). The results evidenced that the genome of *C. rogercresseyi* contains almost all the genes coding for this pathway. However, key biosynthetic genes, including *Idh*, *iLve*, *HMGCL*, and *OXCT* were absent or not identified. On the other hand, we were able to find analog genes to *Idh*, *iLve*, *HMGCL*, and *OXCT* in the metagenome. Although it is not possible to confirm metabolic complementation between sea lice and its microbiota, these results could suggest novel insights about possible roles of sea lice-associated microbiota in parasite biology.

Discussion

Marine organisms constantly interact with microbial communities present in the environment²⁷. While some of these bacteria might fulfill beneficial roles in one species, they can also become harmful in other ones. Therefore, this study aimed to explore possible roles of sea lice-associated microbiota in *C. rogercresseyi* biology and its potential risk to salmon aquaculture. Previous efforts have been made to understand the microbial communities associated with the northern hemisphere salmon louse (*Lepeophtheirus salmonis*)^{28,29}. However, these approaches

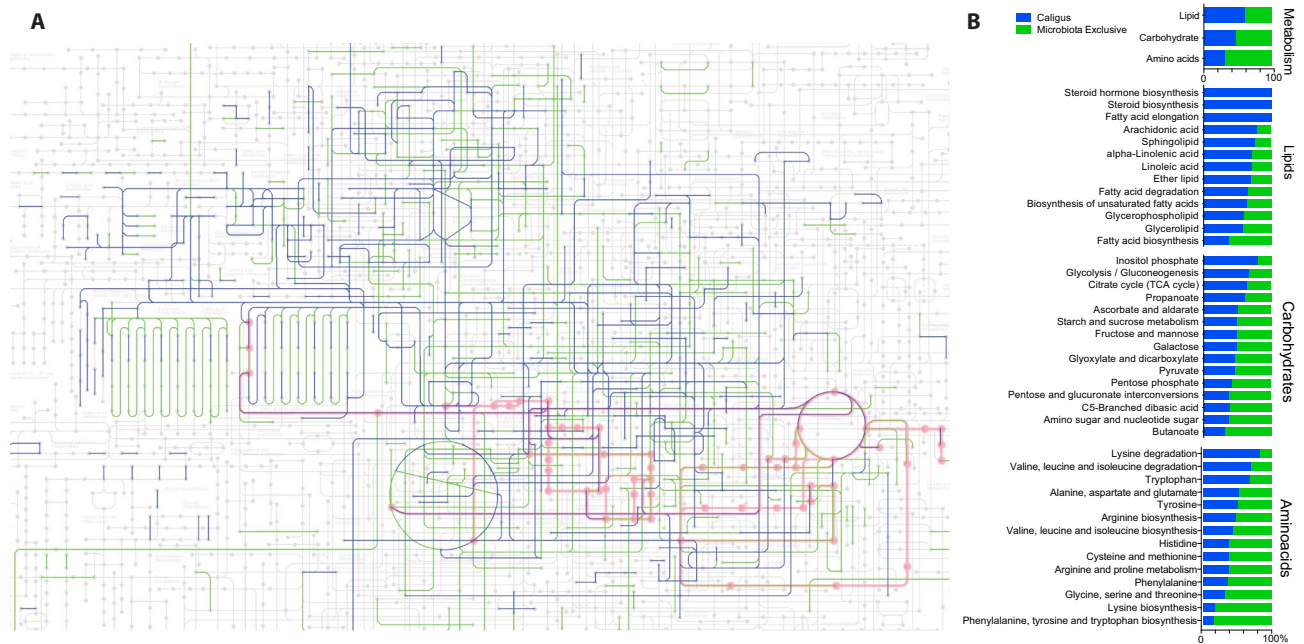


Figure 3. (A) Graphical representation of the metabolic pathways reconstructed for *Caligus rogercresseyi* (blue) and its associated microbiome (green). The amino acid metabolism is highlighted in red. (B) Percentage of genes annotated for the different metabolic pathways encoded in the sea lice genome (blue) and that were exclusively encoded in the microbiome (green). All pathways were constructed using KEGG pathways as reference⁶⁶.

were technically limited based on the need to isolate and culture each bacteria species that want to be identified or the lack of taxonomical resolution at the species or even family level.

Long nanopore reads and computational tools have recently allowed the full-length 16s rRNA sequencing of the unraveled sea lice-associated microbiota, evidencing the geographical variations and the core microbiota at species-specific levels⁶. Here, chromosome proximity ligation (Hi-C) coupled with long-read sequencing is a valuable tool to reconstruct individual bacterial genomes, uncovering novel insights associated with the sea lice microbiome. Thus, our results evidenced the assemble of almost complete genomes from a complex sample with low degrees of cross-contamination. The complete genomes belonged to species within the *Tenacibaculum* and *Dokdonia* genus, previously described as abundant components of salmon and sea lice microbiota^{6,28,29}. As expected, our results evidence that each bacterial genome cluster's completeness degree might be directly related to the abundance of each bacteria within the sample. Although the estimation of each bacteria's abundance was not a goal in this research, the inclusion of a known number of a particular bacterium within the sample can be employed to estimate bacterial abundance using Hi-C data²⁴.

It is known that ectoparasites can become vectors of different pathogens^{18,30}. Previously, it has been suggested that the northern salmon lice *L. salmonis* might fulfill roles as a vector for the transmission of the infectious hematopoietic necrosis virus (IHNV) and the *Aeromonas salmonicida* subsp. *Salmonicida*^{31,32}. However, the role of *C. rogercresseyi* as a vector for fish pathogens is less clear. Given the genome completeness degree, our results confirm previous research evidencing *Tenacibaculum* genus as an abundant component of *C. rogercresseyi* microbiota⁶. Over the last years, this genus has gained major importance as an emergent pathogen causing different outbreaks in salmon aquaculture with presence in Chile, Norway, and Australia^{25,33–35}. Although no evidence has been found about sea lice's role in the transmission of the salmon pathogen *Piscirickettsia salmonis*³⁶, its potential as a transmission vector for *Tenacibaculum* species needs to be further investigated.

Beyond identifying bacterial pathogens, our approach allowed the characterization of 4911 coding genes in the *T. ovolyticum* genome. This information can become key for the application of reverse vaccinology, which has been used to cope with vector-borne parasite diseases³⁷. The use of genomic information allows the rapid discovery of potential antigens, being a promising methodology for developing new fish vaccines³⁸. Virulence factors were also identified in the *T. ovolyticum* genome, including the Dot/Icm type IVB secretion system, 60 K heat shock proteins (HtpB), and catalases (katA) genes, among others, which has been suggested as key factors during the infection of different fish pathogens^{39–42}. Thus, this genomic information availability can also be used to infer pathogenic mechanisms employed by potential fish pathogens present in *C. rogercresseyi* microbiota.

Antibiotic resistance is another major issue both for human health and aquaculture. Therefore, using the ARG database, we identified four different ARGs present in the genome clusters of sea lice microbiota, including *tet(35)*, *QnrS2*, *catB9*, and *OXA-209*. It has been shown that *tet(35)* gene encodes for an efflux pump that confers resistance against tetracyclines⁴³, *QnrS2* is a plasmid-mediated protein that confers resistance against quinolones⁴⁴, *catB9* is a chromosome-encoded chloramphenicol resistance protein⁴⁵ while *OXA-209* gene encodes a beta-lactamase resistance protein⁴⁶. Tetracyclines, quinolones, chloramphenicol, and beta-lactamases antibiotics have been commonly used in aquaculture^{45,47,48}. Particularly in Chile, it was estimated that between 2011 and

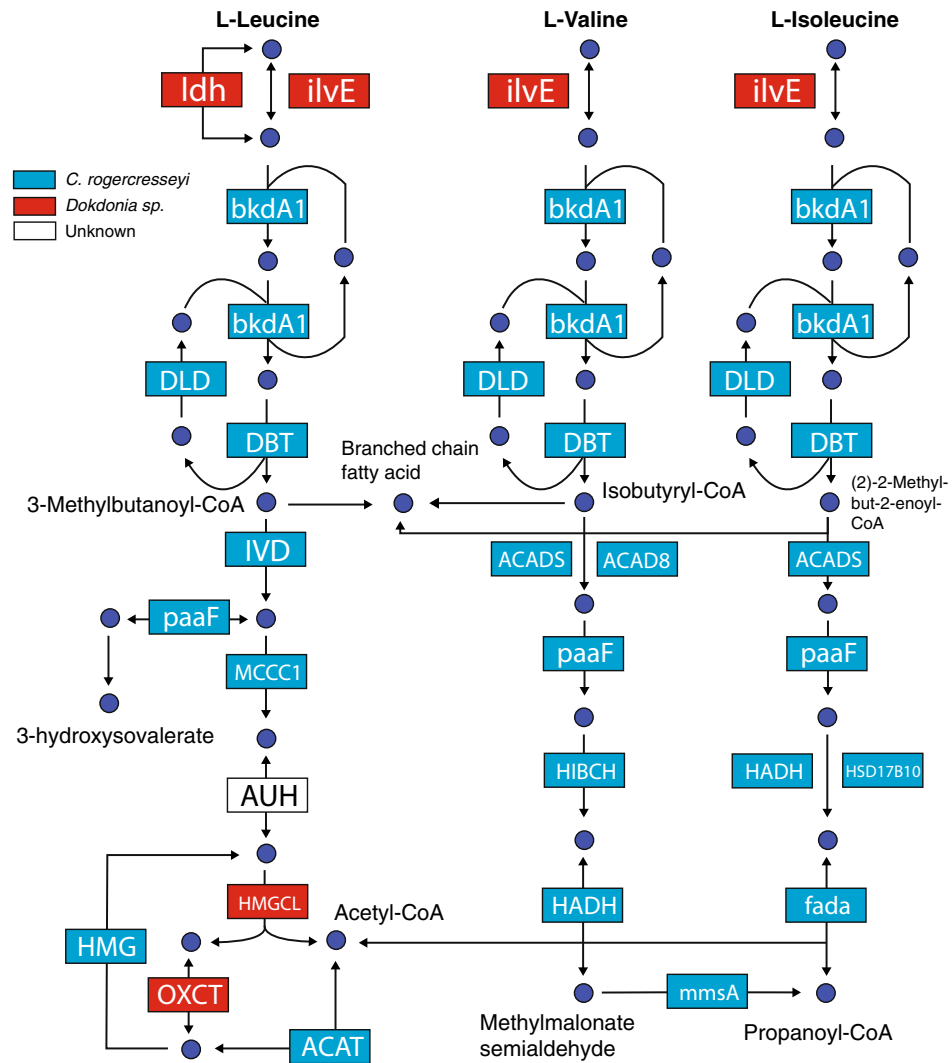


Figure 4. Metabolic reconstruction of the leucine, valine and isoleucine metabolism. Each box represents a gene coding for a protein in the pathway, where the blue boxes represent genes found in *C. rogercresseyi* genome and the red boxes genes found just in *Dokdonia* sp. Genome.

2015, the antibiotics used to produce 1 ton of salmon were 1,500 times higher than in Norway⁴⁹. Although this number has been decreased over the last years, florfenicol and oxytetracycline are still the most used antibiotics in Chilean salmon aquaculture⁵⁰. Here, the presence of *tet(35)* on sea lice-associated microbiota should raise a major concern for the industry since it has been shown that this resistance determinant found from aquaculture sources can be transferred between bacterial species⁵¹. Thus, there is a potential risk that these genes could be transferred between bacteria species composing the sea lice microbiota, including some fish pathogens that can later infect salmon and reduce the efficacy of the currently used antibiotic treatments.

It is known that arthropods-associated microbiota can play pivotal roles in their host development^{52,53}. Here, metabolic interdependence is a widespread phenomenon between bacteria and arthropods that have been previously reported (reviewed by Zientz et al.⁵³). These interactions have been identified for different metabolic pathways, including glycolysis, gluconeogenesis, phospholipids, nucleotide, and sulfur metabolism⁵³. Our results evidenced that the genome of *C. rogercresseyi* lacks different metabolic-related genes, which are putatively encoded in the microbiome. Amino acid metabolism was the most extensive set of microbiota-exclusive genes, with phenylalanine, tyrosine, tryptophan, and lysine biosynthesis the most supported pathways. The absence of these genes in the *C. rogercresseyi* genome suggests that sea lice cannot biosynthesize these amino acids; therefore, they should be obtained through feeding. However, another possibility is that *C. rogercresseyi* might complement the metabolic machinery present in its core associated microbiota to obtain these amino acids. This phenomenon is well-known in aphid parasites. Previous reports have suggested that the bacteria *Buchnera aphidicola* provides essential amino acids to its insect host⁵⁴. Likewise, through RNA-seq and pathway analysis, researchers have found that aphid host can upregulate genes in order to fill the gaps of *Buchnera*'s essential amino acid pathways⁵⁵. On the other hand, the whole-genome sequencing of the symbiont bacteria *Sulcia muelleri* on the insect *Homalodisca coagulata* revealed that the bacteria produced most of the essential amino acids

needed by its host⁵⁶. Moreover, it has been shown that the deprivation of some bacteria in arthropods through antibiotics negatively impacts the host's development and survival⁵⁷. While the nutritional requirements of the sea lice are barely known, our approach suggest the existence of essential amino acids in sea lice, that in turn, are encoded in the microbiome.

Through full 16S nanopore sequencing we have recently identify a core microbiome from sea lice collected from different geographical areas in Chile⁶. Among the identified species, it was suggested that *Dokdonia* sp. could fulfill roles during the infective process of sea lice because of the secretory capacity of bioactive compounds from this bacteria⁶. Notably, one of the most represented genomes in the sea lice microbiome was *Dokdonia* sp. Here, the pathway analysis evidenced that the genome of *Dokdonia* sp. encodes for genes that can fill the gaps for the biosynthesis and degradation of valine, leucine, and isoleucine. Although further studies are needed to confirm the metabolic interdependency between the sea lice and its associated microbiota, a strategy targeting some bacteria like *Dokdonia* sp. might provide novel therapeutic tools to mitigate the parasite's impacts in salmon aquaculture.

Methods

Sample collection and sequencing. Adult specimens of *C. rogercesseyi* were collected in spring from a commercial salmon farm located in Los Lagos, Chile. The site selection was based on our previous established criteria⁶, prioritizing cages with more than 8 months in seawater and centers with less or no pharmacological (both delousing and antibiotics) usage. Collected lice were treated with different antibiotics to minimize environmental bacteria's presence as previously standardized⁶. Thus, adult sea lice were treated with a solution containing 20 mg/ml ampicillin (Sigma-Aldrich, USA), 20 mg/ml Kanamycin (US biological, USA), 1× Penicillin-Streptomycin (GIBCO, USA), and 100 µg/ml Primocin (Invivogen, USA). After 72 h, the specimens were transferred to tubes containing molecular grade EtOH 100% and stored at -80 °C for further processing. Genomic DNA was isolated from a pool of 10 individuals (5 females and 5 males) using the Qiagen DNA purification kit (QIAGEN, Germantown, MD, USA) according to manufacturer's instructions. DNA quality and integrity were assessed through a 1% agarose gel and NanoDrop 1000 Spectrophotometer. Samples with no smear and absorbance ratios above 1.8 were used for further processing. Shotgun metagenomic libraries were prepared using the TruSeq® DNA PCR-Free Library Prep kit (Illumina, San Diego, CA, USA), while Hi-C sequencing libraries were prepared using the Phase Genomics' Animal Hi-C kit (Phase Genomics, Seattle, WA, USA) following the manufacturer's instructions. Both short-read sequencing libraries and Hi-C libraries were sequenced on a HiSeq 4000 (Illumina, San Diego, CA, USA).

Sequencing data analysis. Sequencing analysis and processing were performed as previously described²⁴. Briefly, sequencing adapters from the shotgun metagenomic sequencing data were removed using BBDuk (BBTools developed by the Joint Genome Institute). Sequence reads were mapped against the last version of the sea lice genome⁵⁸ using the BWA-MEM alignment tool. Unmapped reads were de novo assembled using Megahit⁵⁹ and assessed through MetaQuast⁶⁰ considering default parameters. The Hi-C reads were mapped against the de novo assembly through the Burrows-Wheeler alignment tool BWA-MEM⁶¹. The deconvolution of contigs was performed as previously described²⁴. Contigs with less than two restriction sites and shorter than 1000 bp in sizer were removed, whereas the remaining contigs were grouped into genome clusters considering the Hi-C contact data using a proprietary Markov Chain Monte Carlo algorithm. All assembled cluster were deposited on figshare⁶².

Genome clusters characterization and annotation. The first quality control of the generated genome clusters was assessed using CheckM⁶³, which provides a robust estimation of genome completeness and contamination by accounting for ubiquitous genes single-copy within a phylogenetic lineage. Thus, genome clusters with marker gene overrepresentation (MGO) over five were discarded from subsequent analysis. To estimate genome novelty, each metagenome-assembled genomes (MAGs) were compared with the RefSeq genome database using Mash⁶⁴. Thus, a genome was considered as "novel" when its >70% complete, <10% marker gene overrepresentation and >90% novelty score. On the other hand, a "known" genome is defined when >70% completeness, <10% marker gene overrepresentation and <90% novelty score. Genome annotation was performed using the Rapid Annotations using Subsystems Technology (RAST) servers considering default parameters. For each genome cluster, the coding sequences (CDS) were annotated, and BLAST was used for the annotation of virulence factors (VF) and antibiotic resistance genes (ARG) considering as reference the virulence factor database (VFDB) and the MEGARes database respectively. Individual genome cluster features such as GC content, CDS, and RNA density was calculated by a 10,000 bp frame using R. Synteny blocks between closely related species were detected using the Synteny Block ExpLoration tool (Siberia) (<http://bioinf.spbau.ru/sibelia>) and plotted with Circos software version 0.69-6 (<http://circos.ca/>)⁶⁵.

Metabolic pathway analysis. The CDS obtained from the different bacterial clusters were compared with the genes annotated in the last version of the *C. rogercesseyi* genome⁵⁸ to predict possible biological roles of sea lice microbiota. All the CDS were translated to proteins using the CLC genomic workbench (V10, Qiagen). The predicted proteins for both *C. rogercesseyi* and the bacterial clusters were annotated with KEGG Automatic Annotation Server (KAAS)⁶⁶ through a bi-directional best hit (BBH) method to assign orthologs and selecting GHOST X search engine. The resulting pathways were then compared to find metabolic complementarity between the sea lice and the genome clusters composing in sea lice microbiota.

Conclusion

Proximity ligation techniques for the reconstruction of microbial communities in aquaculture provide valuable information for understanding the biological role and potential risk of the sea lice microbiota. Herein, we provide a novel approach to reveal genome features of some bacteria associated with the sea lice, allowing the identification of pathogenic bacteria, virulence factors, antibiotic resistance genes, and the reconstruction of bacterial metabolic pathways associated with *C. rogercresseyi*. We believe that our findings should raise awareness about sea lice microbiota as a pivotal component in salmon aquaculture, providing novel perspectives to develop novel control strategies.

Received: 20 July 2021; Accepted: 16 December 2021

Published online: 17 January 2022

References

- Quinones, R. A., Fuentes, M., Montes, R. M., Soto, D. & Leon-Munoz, J. Environmental issues in Chilean salmon farming: A review. *Rev. Aquac.* **11**, 375–402. <https://doi.org/10.1111/raq.12337> (2019).
- Gallardo-Escárate, C. *et al.* The race between host and sea lice in the Chilean salmon farming: A genomic approach. *Rev. Aquac.* <https://doi.org/10.1111/raq.12334> (2019).
- Gonzalez, L. & Carvajal, J. Life cycle of *Caligus rogercresseyi*, (Copepoda: Caligidae) parasite of Chilean reared salmonids. *Aquaculture* **220**, 101–117. [https://doi.org/10.1016/S0044-8486\(02\)00512-4](https://doi.org/10.1016/S0044-8486(02)00512-4) (2003).
- Johnson, S. C., Treasurer, J. W., Bravo, S., Nagasawa, K. & Kabata, Z. A review of the impact of parasitic copepods on marine aquaculture. *Zool. Stud.* **43**, 229–243 (2004).
- Dresdner, J. *et al.* Impact of *Caligus* treatments on unit costs of heterogeneous salmon farms in Chile. *Aquacult. Econ. Manag.* **23**, 1–27. <https://doi.org/10.1080/13657305.2018.1449271> (2019).
- Goncalves, A. T. *et al.* Nanopore sequencing of microbial communities reveals the potential role of sea lice as a reservoir for fish pathogens. *Sci. Rep.* **10**, 1–12. <https://doi.org/10.1038/s41598-020-59747-0> (2020).
- Hahn, M. A. & Dheilly, N. M. Experimental models to study the role of microbes in host–parasite interactions. *Front. Microbiol.* **7**, 1300. <https://doi.org/10.3389/fmicb.2016.01300> (2016).
- Rosenberg, E. & Zilber-Rosenberg, I. Symbiosis and development: The hologenome concept. *Birth Defects Res. Part C Embryo Today Rev.* **93**, 56–66. <https://doi.org/10.1002/bdrc.20196> (2011).
- Blaxter, M. & Koutsovoulos, G. The evolution of parasitism in Nematoda. *Parasitology* **25**, 1–14 (2014).
- Chaston, J. M. *et al.* The entomopathogenic bacterial endosymbionts *Xenorhabdus* and *Photorhabdus*: Convergent lifestyles from divergent genomes. *PLoS ONE* **6**, e27909. <https://doi.org/10.1371/journal.pone.0027909> (2011).
- Hoerauf, A. *et al.* Tetracycline therapy targets intracellular bacteria in the filarial nematode *Litomosoides sigmodontis* and results in filarial infertility. *J. Clin. Investig.* **103**, 11–18. <https://doi.org/10.1172/JCI4768> (1999).
- Fichorova, R. N. *et al.* Endobiont viruses sensed by the human host—beyond conventional antiparasitic therapy. *PLoS ONE* **7**, e48418. <https://doi.org/10.1371/journal.pone.0048418> (2012).
- Kalluri, S., Gilruth, P., Rogers, D. & Szczur, M. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: A review. *PLoS Pathog.* **3**, 1361–1371. <https://doi.org/10.1371/journal.ppat.0030116> (2007).
- Hansen, B. & Bech, G. Bacteria associated with a marine planktonic copepod in culture. 1. Bacterial genera in seawater, body surface, intestines and fecal pellets and succession during fecal pellet degradation. *J. Plankton Res.* **18**, 257–273. <https://doi.org/10.1093/plankt/18.2.257> (1996).
- Dumontet, S. *et al.* Ecological relationship between *Aeromonas* and *Vibrio* spp. and planktonic copepods in the coastal marine environment in southern Italy. *Comp. Immunol. Microb.* **19**, 245–254. [https://doi.org/10.1016/0147-9571\(96\)00012-4](https://doi.org/10.1016/0147-9571(96)00012-4) (1996).
- Vezzulli, L., Pruzzo, C., Huq, A. & Colwell, R. R. Environmental reservoirs of *Vibrio cholerae* and their role in cholera. *Environ. Microbiol. Rep.* **2**, 27–33. <https://doi.org/10.1111/j.1758-2229.2009.00128.x> (2010).
- Heidelberg, J. F., Heidelberg, K. B. & Colwell, R. R. Bacteria of the gamma-subclass *Proteobacteria* associated with zooplankton in Chesapeake Bay. *Appl. Environ. Microbiol.* **68**, 5498–5507. <https://doi.org/10.1128/Aem.68.11.5498-5507.2002> (2002).
- Sepulveda, F. A., Torres, J. F., Infante, C. D. & Gonzalez, M. T. Potential role of ectoparasites (*Zeuxapta seriola* and *Caligus lalandei*) in the transmission of pathogenic bacteria in yellowtail kingfish *Seriola lalandi*, inferred from cultivable microbiota and molecular analyses. *J. Fish Dis.* **40**, 979–985. <https://doi.org/10.1111/jfd.12582> (2017).
- Jakob, E., Barker, D. E. & Garver, K. A. Vector potential of the salmon louse *Lepeophtheirus salmonis* in the transmission of infectious haematopoietic necrosis virus (IHNV). *Dis. Aquat. Org.* **97**, 155–165. <https://doi.org/10.3354/dao02414> (2011).
- Nese, L. & Enger, O. Isolation of *Aeromonas salmonicida* from salmon lice *Lepeophtheirus salmonis* and marine plankton. *Dis. Aquat. Org.* **16**, 79–81. <https://doi.org/10.3354/dao016079> (1993).
- Nylund, A., Wallace, C. J., Hovland, T., Boxshall, G. A. & A Defaye, D. The possible role of *Lepeophtheirus salmonis* (Krøyer) in the transmission of infectious salmon anaemia. In *Conference Proceedings* (1993).
- Yaffe, E. & Relman, D. A. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat. Microbiol.* **5**, 343. <https://doi.org/10.1038/s41564-019-0625-0> (2020).
- DeMaere, M. Z. & Darling, A. E. bin3C: Exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes. *Genome Biol.* **20**, 1–16. <https://doi.org/10.1186/s13059-019-1643-1> (2019).
- Stalder, T., Press, M. O., Sullivan, S., Liachko, I. & Top, E. M. Linking the resistome and plasmidome to the microbiome. *ISME J.* **13**, 2437–2446. <https://doi.org/10.1038/s41396-019-0446-4> (2019).
- Avendano-Herrera, R., Collarte, C., Saldarriaga-Cordoba, M. & Irgang, R. New salmonid hosts for *Tenacibaculum* species: Expansion of tenacibaculosis in Chilean aquaculture. *J. Fish Dis.* **43**, 1077–1085. <https://doi.org/10.1111/jfd.13213> (2020).
- Gallardo-Escárate, C. *et al.* Chromosome-scale genome assembly of the sea louse *Caligus rogercresseyi*. *Figshare* <https://doi.org/10.6084/m9.figshare.11780658.v1> (2020).
- Wahl, M., Goeckle, F., Labes, A., Dobretsov, S. & Weinberger, F. The second skin: Ecological role of epibiotic biofilms on marine organisms. *Front. Microbiol.* **3**, 292. <https://doi.org/10.3389/fmicb.2012.00292> (2012).
- Barker, D. E., Braden, L. M., Coombs, M. P. & Boyce, B. Preliminary studies on the isolation of bacteria from sea lice, *Lepeophtheirus salmonis*, infecting farmed salmon in British Columbia, Canada. *Parasitol. Res.* **105**, 1173–1177. <https://doi.org/10.1007/s00436-009-1523-9> (2009).
- Dalvin, S. *et al.* Microbial communities associated with the parasitic copepod *Lepeophtheirus salmonis*. *Mar. Genom.* **49**, 100688. <https://doi.org/10.1016/j.margen.2019.05.003> (2020).
- Overstreet, R. M., Jovonovich, J. & Ma, H. W. Parasitic crustaceans as vectors of viruses, with an emphasis on three penaeid viruses. *Integr. Comp. Biol.* **49**, 127–141. <https://doi.org/10.1093/icb/icp033> (2009).

31. Sepúlveda, F. A., Torres, J. F., Infante, C. D. & González, M. T. Potential role of ectoparasites (*Zeuxapta seriolae* and *Caligus lalandei*) in the transmission of pathogenic bacteria in yellowtail kingfish *Seriola lalandi*, inferred from cultivable microbiota and molecular analyses. *J. Fish Dis.* **40**, 979–985. <https://doi.org/10.1111/jfd.12582> (2017).
32. Novak, C. W., Lewis, D. L., Collicutt, B., Verkaik, K. & Barker, D. E. Investigations on the role of the salmon louse, *Lepeophtheirus salmonis* (Caligidae), as a vector in the transmission of *Aeromonas salmonicida* subsp. *salmonicida*. *J. Fish Dis.* **39**, 1165–1178. <https://doi.org/10.1111/jfd.12449> (2016).
33. Avendano-Herrera, R. *et al.* Isolation, characterization and virulence potential of *Tenacibaculum dicentrarchi* in salmonid cultures in Chile. *Transbound. Emerg. Dis.* **63**, 121–126. <https://doi.org/10.1111/tbed.12464> (2016).
34. Klakegg, O., Abayneh, T., Fauske, A. K., Fulberth, M. & Sorum, H. An outbreak of acute disease and mortality in Atlantic salmon (*Salmo salar*) post-smolts in Norway caused by *Tenacibaculum dicentrarchi*. *J. Fish Dis.* **42**, 789–807. <https://doi.org/10.1111/jfd.12982> (2019).
35. Wilson, T. K., Douglas, M. & Dunn, V. First identification in Tasmania of fish pathogens *Tenacibaculum dicentrarchi* and *T. soleae* and multiplex PCR for these organisms and *T. maritimum*. *Dis. Aquat. Organ.* **136**, 219–226. <https://doi.org/10.3354/dao03407> (2019).
36. Labra, A., Bravo, S. & Marshall, S. H. Defining the role of *Caligus rogercresseyi* in transmission and spreading of *Piscirickettsia salmonis*. *Aquaculture* **528**, 735489. <https://doi.org/10.1016/j.aquaculture.2020.735489> (2020).
37. Lew-Tabor, A. E. & Valle, M. R. A review of reverse vaccinology approaches for the development of vaccines against ticks and tick borne diseases (vol 7, pg 573, 2016). *Ticks Tick-Borne Dis.* **7**, 1236–1237. <https://doi.org/10.1016/j.ttbdis.2016.07.008> (2016).
38. Ma, J., Bruce, T. J., Jones, E. M. & Cain, K. D. A review of fish vaccine development strategies: Conventional methods and modern biotechnological approaches. *Microorganisms* **7**, 569. <https://doi.org/10.3390/microorganisms7110569> (2019).
39. Gomez, F. A. *et al.* 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. <https://doi.org/10.1371/journal.pone.0054934> (2013).
40. Valenzuela-Miranda, D., Valentina, V.-M., Gustavo, N.-A. & Cristian, G.-E. Long-term serial culture of *Piscirickettsia salmonis* leads to a genomic and transcriptomic reorganization affecting bacterial in vitro virulence. *Figshare*. <https://doi.org/10.6084/m9.figshare.12047238.v1> (2020).
41. Yu, Y. X. *et al.* Complete genome sequence of *Photobacterium damsela* subsp. *damsela* strain SSPD1601 isolated from deep-sea cage-cultured *Sebastes schlegelii* with septic skin ulcer. *Int. J. Genomics* <https://doi.org/10.1155/2019/4242653> (2019).
42. Perez-Pascual, D. *et al.* The complete genome sequence of the fish pathogen *Tenacibaculum maritimum* provides insights into virulence mechanisms. *Front. Microbiol.* **8**, 1542. <https://doi.org/10.3389/fmicb.2017.01542> (2017).
43. Teo, J. W. P., Tan, T. M. C. & Poh, C. L. Genetic determinants of tetracycline resistance in *Vibrio harveyi*. *Antimicrob. Agents Chem.* **46**, 1038–1045 (2002).
44. Gay, K. *et al.* Plasmid-mediated quinolone resistance in non-Typhi serotypes of *Salmonella enterica*. *Clin. Infect. Dis.* **43**, 297–304 (2006).
45. Santos, L. & Ramos, F. Antimicrobial resistance in aquaculture: Current knowledge and alternatives to tackle the problem. *Int. J. Antimicrob. Agents* **52**, 135–143 (2018).
46. Chen, Y. P., Lee, S. H., Chou, C. H. & Tsai, H. J. Detection of florfenicol resistance genes in *Riemerella anatipestifer* isolated from ducks and geese. *Vet. Microbiol.* **154**, 325–331. <https://doi.org/10.1016/j.vetmic.2011.07.012> (2012).
47. Romero-Soto, I. C. *et al.* Degradation of chloramphenicol in synthetic and aquaculture wastewater using electrooxidation. *J. Environ. Qual.* **47**, 805–811. <https://doi.org/10.2134/jeq2017.12.0475> (2018).
48. Chen, B. *et al.* Complex pollution of antibiotic resistance genes due to beta-lactam and aminoglycoside use in aquaculture farming. *Water Res.* **134**, 200–208. <https://doi.org/10.1016/j.watres.2018.02.003> (2018).
49. Miranda, C. D., Godoy, F. A. & Lee, M. R. Current status of the use of antibiotics and the antimicrobial resistance in the Chilean salmon farms. *Front. Microbiol.* **9**, 1284. <https://doi.org/10.3389/fmicb.2018.01284> (2018).
50. Sernapesca. Informe Sobre Uso de Antimicrobianos en la Salmonicultura Nacional. Valparaíso. <http://www.sernapesca.cl> (2020).
51. Akinbowale, O. L., Peng, H. & Barton, M. D. Diversity of tetracycline resistance genes in bacteria from aquaculture sources in Australia. *J. Appl. Microbiol.* **103**, 2016–2025. <https://doi.org/10.1111/j.1365-2672.2007.03445.x> (2007).
52. Esposti, M. D. & Romero, E. M. The functional microbiome of arthropods. *PLoS ONE* **12**, e0176573. <https://doi.org/10.1371/journal.pone.0176573> (2017).
53. Zientz, E., Dandekar, T. & Gross, R. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* **68**, 745–770. <https://doi.org/10.1128/MMBR.68.4.745-770.2004> (2004).
54. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86. <https://doi.org/10.1038/35024074> (2000).
55. Hansen, A. K. & Moran, N. A. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. USA* **108**, 2849–2854. <https://doi.org/10.1073/pnas.1013465108> (2011).
56. Wu, D. *et al.* Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**, 1079–1092. <https://doi.org/10.1371/journal.pbio.0040188> (2006).
57. Wilkinson, T. L. The elimination of intracellular microorganisms from insects: An analysis of antibiotic-treatment in the pea aphid (*Acyrtosiphon pisum*). *Compar. Biochem. Physiol. Mol. Integr. Physiol.* **119**, 871–881. [https://doi.org/10.1016/S1095-6433\(98\)00013-0](https://doi.org/10.1016/S1095-6433(98)00013-0) (1998).
58. Gallardo-Escárate, C. *et al.* Chromosome-scale genome assembly of the sea louse *Caligus rogercresseyi* by SMRT sequencing and Hi-C analysis. *Sci. Data* **8**, 1–12. <https://doi.org/10.1038/s41597-021-00842-w> (2021).
59. Li, D. H., Liu, C. M., Luo, R. B., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033> (2015).
60. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090. <https://doi.org/10.1093/bioinformatics/btv697> (2016).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> (2009).
62. Valenzuela Miranda, D.G.-E.C., Valenzuela-Muñoz, V., Núñez-Acuña, G. & Goncalves, A. T. Proximity ligation strategy for the genomic reconstruction of microbial communities associated with the ectoparasite *Caligus rogercresseyi*. *Figshare* <https://doi.org/10.6084/m9.figshare.16980616.v1> (2021).
63. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055. <https://doi.org/10.1101/gr.186072.114> (2015).
64. Ondov, B. D. *et al.* Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 1–14. <https://doi.org/10.1186/s13059-016-0997-x> (2016).
65. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645. <https://doi.org/10.1101/gr.092759.109> (2009).
66. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185. <https://doi.org/10.1093/nar/gkm321> (2007).

Acknowledgements

This work was financed by the FONDECYT FONDAP (#15110027), FONDECYT POSTDOC (#3200600), and FONDECYT (#1210852).

Author contributions

The study was conceived by D.V., V.V., G.N., A.T.G. and C.G. V.V. prepared the samples, while sequencing and deconvolution was done by I.L. and B.N. D.V., A.T.G. and C.G. analyzed the data and D.V. drafted the manuscript with the input and approval of all the authors.

Competing interests

IL and BN are employees and shareholders of Phase Genomics, Inc, a company commercializing Hi-C technology. DV, ATG, VV, GN and CG declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-04485-0>.

Correspondence and requests for materials should be addressed to D.V.-M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2022